Fall 2014

Identification of population average treatment effects using nonlinear instrumental variables estimators : another cautionary note

Cole Garrett Chapman
University of Iowa

IDENTIFICATION OF POPULATION AVERAGE TREATMENT EFFECTS USING NONLINEAR INSTRUMENTAL VARIABLES ESTIMATORS: ANOTHER CAUTIONARY NOTE

by

Cole Garrett Chapman

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Pharmacy
in the Graduate College of
The University of Iowa

December 2014

Thesis Supervisor: Professor John M. Brooks

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Cole Garrett Chapman

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Pharmacy at the December 2014 graduation.

Thesis Committee: ______________________________
John M. Brooks, Thesis Supervisor

____
Padmaja Ayyagari

____
Elizabeth A. Chrischilles

____
Linnea A. Polgreen

____
Mary C. Schroeder

For Anna, Meghan, and my mother.
You may not enjoy reading this, but you've certainly helped me to complete it.

We thrive not when we’ve done it all, but when we still have more to do... Completion is a goal but we hope it is never the end.

Sarah Lewis
Embrace the Win (TED Talk)

ACKNOWLEDGMENTS

This dissertation would not have been possible without the unwavering guidance and patience of countless individuals who have, in one way or another, helped me through the process.

Chief among these individuals is my advisor, Dr. John M. Brooks, whose unfailing energy and willingness to discuss issues relevant to my research were, and continue to be of immeasurable importance.

Drs. Mary C. Schroeder, Padmaja Ayyagari, Elizabeth A. Chrischilles, and Linnea A. Polgreen who generously donated their time and expertise as members of my dissertation committee.

Miss Lois Baker, without whom I may never have completed the administrative processes necessary for graduation.

The faculty and staff of the University of Iowa College of Pharmacy for all of their help and guidance throughout my graduate education.

And finally, my mother, for every reason.

ABSTRACT

Nonlinear two-stage residual inclusion (2SRI) estimators have become increasingly favored over traditional linear two-stage least squares (2SLS) methods for instrumental variables analysis of empirical models with inherently nonlinear dependent variables. Rising adoption of nonlinear 2SRI is largely attributable to simulation evidence showing that nonlinear 2SRI generates consistent estimates of population average treatment effects in nonlinear models, while 2SLS and nonlinear 2SPS do not. However, while it is believed that consistency of 2SRI for population average treatment effects is a general result, current evidence is limited to simulations performed under unique and restrictive settings with regards to treatment effect heterogeneity and conditions underlying treatment choices. This research contributes by describing existing simulation evidence and investigating the ability to generate absolute estimates of population average treatment effects (ATE) and local average treatment effects (LATE) using common IV estimators using Monte Carlo simulation methods across 10 alternative scenarios of treatment effect heterogeneity and sorting-on-the-gain. Additionally, estimates for the effect of ACE/ARBs on 1-year survival for Medicare beneficiaries with acute myocardial infarction are generated and compared across alternative linear and nonlinear IV estimators. Simulation results show that, while 2SLS generates unbiased and consistent estimates of LATE across all scenarios, nonlinear 2SRI generates unbiased estimates of ATE only under very restrictive settings. If marginal patients are unique in terms of treatment effectiveness, then nonlinear 2SRI cannot be expected to generate unbiased or consistent estimates of ATE unless all factors related to treatment effect heterogeneity are fully measured.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This research is a response to growing adoption of sophisticated nonlinear estimation strategies for analysis in models with binary and other inherently nonlinear dependent variables, often under beliefs that these methods are generally superior to more traditional linear models. While these methods may offer advantages under certain settings, validity and interpretability of absolute treatment effect estimates generated using these models is often complicated by added strong assumptions of the nonlinear model. Unfortunately, these assumptions and the implications of violating them appear to remain unclear for many health services researchers. This research is intended as a clear and thorough explanation of assumptions underlying identification with popular linear and nonlinear instrumental variables methods. This research is not intended as a condemnation of sophisticated nonlinear regression methods, but rather as a suggestion that there is presently no "silver bullet" to generate absolute estimates of population average treatment effects. The interpretation of absolute treatment effect estimates generated by instrumental variables estimators is a function of the characteristics of treatment effect heterogeneity in the relevant population, which of the factors related to treatment choice and treatment effect heterogeneity are measured by the researchers, and the instruments specified in the model. This is true regardless of the estimator used for analysis, linear or nonlinear. I also seek to draw attention to the interesting possibilities for inferences from LATE estimates that are often taken for granted. Population averages are often thought of as the ultimate goal, but I am skeptical that this should always be the case if treatment effects are heterogeneous. Under settings of heterogeneous treatment effects where treatment decisions are related to this heterogeneity, population averages offer little insight for potential benefits from increasing or decreasing treatment rates. LATE estimates, on the other hand, may often be considered as an average effect across individuals whose treatment decisions may be most likely affected by policy changes.

CHAPTER 1

INTRODUCTION AND BACKGROUND

Estimation of policy-relevant treatment effects using observational data is a core component of health outcomes and comparative effectiveness research. Large observational data sets have not only become increasingly accessible and user-friendly but may be, in many contexts, the only socially or ethically acceptable treatment variation available. Unfortunately, the non-experimental nature of these data often complicates its use in research. While simple, direct risk-adjustment methods may be sufficient for estimating the effect of a treatment on outcomes when treatment assignment can be assumed to be essentially random—such as in randomized controlled trials (RCTs)—researchers must often consider the likelihood that choice of treatment was based upon unobserved factors associated with outcomes in the data. This complication—commonly referred to as unmeasured confounding, residual confounding, or endogeneity—is a frequent threat to validity in health outcomes research using observational data where factors such as underlying patient health and clinical complexity are not only exceptionally difficult to define and quantify, but unlikely to be measurable. Fortunately, when theory suggests sources of endogeneity between treatment choice and outcomes may exist, instrumental variables methods can offer a potential pathway to causality. However, as more complex instrumental variables methods become readily and easily usable through advanced statistical packages, researchers must be made aware of their limitations.

Instrumental variables methods make it possible to circumvent endogeneity in treatment assignment by taking advantage of exogenous variation in observed factors that are directly related to treatment choice but thought to be unrelated to outcomes or other unmeasured variables related to outcomes. These factors are commonly referred to as instrumental variables (IVs).[1-3] Under these conditions, the treatment variation determined by the IVs can be regarded as similar to randomization.[4,5] In fact, random treatment assignment in RCTs can be considered an instrumental variable that affects treatment but is unrelated to outcomes or other unmeasured factors affecting outcomes. However, unlike RCTs which randomize treatment choice for every patient, instruments often only affect treatment choice for a subpopulation of individuals—a subset commonly referred to as the "marginal patients" or "compliers". Treatment effect estimates generated using only the variation in treatment choice from the instruments are only directly interpretable as the average treatment effect for this subset of marginal patients.[1-4,6] This average treatment effect parameter is referred to as the local average treatment effect (LATE).[2,7]

While the potential for policy-relevant inferences from LATE may at first seem limited, LATE estimates hold an important place in comparative effectiveness and outcomes research. LATE estimates generated from observational data often represent the average effect of treatment for patients with a high degree of uncertainty in treatment choice, perhaps the patients most likely to have treatment decisions influenced by policy changes.[2,4,8-11] Alternatively, researchers may seek estimates of other averages such as the population average treatment effect (ATE)—an estimate of the average treatment effect across all patients with a given condition. But if treatment effects vary or are heterogeneous across a population then estimates of ATE may arguably be less informative then estimates of LATE. For example, consider that only very rarely will a policy change affect the treatment choices for all patients. It is often more likely that policy changes will affect treatment decisions for only those patients for whom the benefits of treatment are less clear. In these cases, if the research objective is to inform policy-makers and treatment decision makers about the potential implications of expanding treatment rates, LATE may be more informative than ATE.

The two-stage predictor substitution (2SPS) method is perhaps the most common approach for IV-based regression models estimating LATE. In the first stage of 2SPS, treatment choice is predicted by one or more IVs and other measured variables associated with treatment choice. The predicted treatment value generated from the first-stage regression is then substituted for observed treatment in the second-stage outcome model and health outcomes are regressed on predicted treatment and other measured factors associated with outcomes. The most prevalent form of 2SPS is two-stage least squares (2SLS), in which both first- and second-stage regressions are linear models estimated using ordinary least squares (OLS). In 2SLS, the coefficient estimate on predicted treatment in the second-stage model is the LATE estimate.[1,2] The statistical properties of LATE estimates generated using 2SLS methods are based on the central limit theorem. With large datasets, distributional assumptions for the model error terms are not required for testing model parameters.[8,12,13] Consistency and interpretability of results generated by nonlinear models, on the other hand, are conditional on distributional assumptions.[12,14]

While LATE is generally considered to be interpretable only for the population of marginal patients, other researchers have argued that, under certain circumstances related to treatment effect heterogeneity and treatment choice, LATE estimates can be used to estimate alternative average treatment effect parameters, such as the ATE. Generalization of LATE to ATE requires that either (1) treatment effects are homogeneous across patients in the population, or (2) if treatment effects are heterogeneous across patients then treatment decision makers are not using information on patients' idiosyncratic gains from treatment when making treatment decisions.[15-17]

However, if the research goal is to estimate an ATE, recent research by Terza and colleagues has suggested that IV-based nonlinear two-stage residual inclusion (2SRI) estimators can be used as an alternative to 2SLS or nonlinear 2SPS methods for generating consistent ATE estimates in models with binary or otherwise limited dependent variables, without requiring the assumptions previously deemed necessary to extend LATE to ATE.[18,19] This research using simulation modeling showed that nonlinear 2SRI methods generate consistent estimates of ATE in models with *inherently* *nonlinear dependent variables* while both linear and nonlinear 2SPS estimators yield generally inconsistent estimates of ATE in these same models.[18-20] For the remainder of this research paper, the term *inherently nonlinear dependent variable* will be used to refer to dependent variables which have a structure (e.g., non-negative, binary, skewed) that is suggestive of a nonlinear relationship between the dependent and independent variables. The current literature does not give a more precise definition of the term "inherently nonlinear" dependent variables.

Based upon simulation results showing that nonlinear 2SRI is less biased than 2SLS or 2SPS for estimating ATE, Terza et al. (2008) suggest that researchers should broadly adopt the nonlinear 2SRI method—in place of 2SLS or nonlinear 2SPS methods—for analysis of empirical models with inherently nonlinear dependent variables and problems of endogeneity.[19] This suggestion, made broadly regardless of project research goals, has taken hold. Terza et al.'s 2008 paper[19] has been cited in excess of 230 times, often as substantiation for use of IV-based nonlinear 2SRI methods in place of linear 2SLS or other IV-based methods.[5,21-41] However, the simulation studies that serve as the basis of this recommendation did not discuss the LATE parameter or compare estimates generated by 2SRI, 2SLS, or 2SPS methods to true values of LATE.[1-3,7] This point is commonly overlooked by empirical researchers citing Terza et al. who imply that 2SLS is generally inconsistent in models with inherently nonlinear dependent variables regardless of the treatment effect parameter of interest.[21,23,28,30-32,34,39-42] If it is theorized that policy-changes may affect treatment choices for only a subset of patients for whom the benefits of treatment, relative to costs, are uncertain, then an estimate of LATE, not ATE, may be the parameter of interest. Estimates of LATE represent the average treatment effect for patients whose treatment choices were affected by the instrument; in certain scenarios it may be argued that these are the patients characterized by uncertainty in treatment choice and that estimates of LATE are valuable for informing policy-makers on the potential impacts of slightly increasing or decreasing treatment rates in a population.[7,11,16,43,44]

The 2SRI method is an extension of an endogeneity test proposed by Hausman (1978) and represents a special case of control function (CF) methods.[45-47] The first-stage regression of 2SRI is identical to that of the 2SPS method. The residual term from the first-stage regression is then estimated and included as a covariate, along with observed treatment, in the second-stage outcome regression. Applications of the 2SRI method to models with inherently nonlinear dependent variables are most often estimated using a nonlinear regression method, such as logit or probit in the case of binary treatment or outcome. For models in which both the first- and second-stage equations are estimated using linear models, the 2SPS (i.e., 2SLS) and 2SRI methods will yield identical estimates.[19,45] It is curious, then, why non-linearity in estimation, in and of itself, enables researchers to suspend the need to justify assumptions that link LATE estimates to ATE. Given the quick adoption of this method by empirical researchers, it is vital to assess the implied assumptions underlying the simulation approaches used to show the positive properties of 2SRI, and assess whether these results are robust to model specification.

When looking deeper into the simulation approach of Terza and colleagues that provide the basis for asserting the superiority of nonlinear 2SRI over other estimators, it becomes apparent that their result reflects only one unique theoretical scenario related to treatment effect heterogeneity and treatment choice. In their scenario (1) the absolute effect of treatment varies or is heterogeneous across patients as a nonlinear function of **all** other factors that affect outcome directly; (2) the only unmeasured variation in treatment choice is the unmeasured confounder itself; (3) the marginal population is approximately uniformly distributed across the distribution of treatment effectiveness; and (4) absolute treatment effect differences across patients do not consistently influence individual treatment choices such that patients with higher benefit from treatment are more likely treated, all else equal. This narrow theoretical scenario clearly is not reflective of the many and varied scenarios that exist in clinical practice. As such, significant knowledge gaps exist with respect to the identification and consistency properties of estimates generated by nonlinear IV estimators when (1) absolute treatment effects differ across patients and this influences treatment choices directly such that those patients with greater benefit from treatment are more likely treated; (2) factors exist that affect treatment effectiveness but are not related to outcomes independent of treatment (e.g., genetics); or (3) factors exist that affect outcomes directly but do not affect treatment effectiveness (e.g., socio-economic status). If, for a given empirical problem, it is theorized that any of these settings may apply, the advantages or disadvantages of nonlinear 2SRI estimators relative to alternative estimators is unknown.

Consider the consequences of a case in which treatment decision makers *do* recognize patient characteristics related to treatment effectiveness and *are* using this information about individual patient idiosyncratic gains from treatment to make treatment decisions—a phenomenon termed "passive personalization" or *essential heterogeneity*.[17,48] Suggestions that physicians do use information about patients idiosyncratic gains from treatment when making treatment decisions have been given in several recent papers by Joshua Angrist, Anirban Basu, and others.[6,7,9,15-17,48-50] The decision to be treated with mastectomy versus breast-conserving radiation therapy for women with breast cancer is an example of this, as Basu and Heckman (2007) reported strong evidence of self-selection into treatment based on heterogeneous treatment effectiveness across women with breast cancer.[17] While 2SLS methods still yield consistent estimates of LATE under these settings, the LATE estimate will only be "locally interpretable" for the subpopulation of marginal individuals whose treatment choices were affected by the instruments and who are unique in terms of treatment effectiveness. This limitation of 2SLS is well established.[1,2,7,17] The limitations for identification and interpretability of ATE or LATE estimates generated by nonlinear 2SRI methods, on the other hand, have not been examined explicitly under settings of essential heterogeneity. It could be expected that nonlinear 2SRI methods will not produce parameter estimates that are interpretable across the entire population and therefore not produce valid estimates of the ATE under settings of essential heterogeneity because treated and untreated subpopulations differ on unmeasured characteristics related to treatment effect heterogeneity. Moreover, even if parameter estimates are unbiased for the subpopulation of marginal individuals, absolute LATE estimates may not be estimable because the subpopulation of marginal individuals cannot be identified directly from the data. To my knowledge, no existing methodological research has examined the ability of nonlinear 2SRI methods to generate unbiased LATE estimates from observational data. The distributional assumptions underpinning nonlinear 2SRI are not necessary for consistent estimation of LATE using linear 2SLS.[12,13]

This research is a first step towards characterizing the settings under which common IV estimators identify alternative average treatment effects. I focus particularly on the ability to generate unbiased estimates of LATE and ATE using 2SLS, nonlinear 2SPS, and nonlinear 2SRI methods in a model with binary outcome and single binary treatment. A key goal of this research is to find settings in which alternative common IV methods may have comparative advantages or disadvantages. Identification of the ATE and LATE will be assessed across scenarios varying by whether: (1) factors exist that affect the effectiveness of treatment but not outcomes directly (e.g., genetics); (2) factors exist that effect outcomes directly but not the effectiveness of treatment (e.g., socio-economic status); and (3) heterogeneity in treatment effectiveness is essential and patients with greater benefit from treatment are more likely to be treated. In each theoretical scenario, true and estimated values of the ATE and LATE will be generated using Monte Carlo simulation methods—a common strategy in research evaluating properties of IV estimators.[7,18-20] Finally, nonlinear 2SRI, nonlinear 2SPS, and linear 2SLS estimators will be applied to a real world problem in estimating the effects of angiotensin converting-enzyme (ACE) inhibitors and angiotensin receptor blockers

(ARBs) on 1-year survival among Medicare beneficiaries with new acute myocardial infarction (AMI). It has been shown that the effectiveness of ACE/ARBs is heterogeneous across AMI patients by presence of certain comorbid conditions and clinical characteristics (e.g., diabetes, left ventricular ejection fraction) and that there may be some degree of selection into treatment with ACE/ARBs based on these expected gains.[51-58] Using Medicare administrative claims data, estimates for the effect of ACE/ARB use after AMI on 1-year survival will be generated using nonlinear 2SRI, nonlinear 2SPS, and linear 2SLS estimators. These estimates will be compared and discussed in context of the simulation results for alternative possible theoretical scenarios that may apply to this clinical problem. This is, to my knowledge, the first methodological research attempting to characterize the robustness and interpretability of nonlinear 2SRI estimators across various scenarios of treatment effect heterogeneity and choice.

Specific Aims

1. Characterize the ability to generate consistent estimates of population average treatment effects (ATE) and local average treatment effects (LATE) using popular linear and nonlinear instrumental variables methods across alternative scenarios of treatment effect heterogeneity and choice.
    a. Investigate ability to generate unbiased estimates of ATE and LATE using nonlinear two-stage residual inclusion (2SRI), nonlinear two-stage predictor substitution (2SPS), and linear two-stage least squares (2SLS) estimators.
    b. Examine these estimators across alternative scenarios of treatment effect heterogeneity and choice.
        i. Heterogeneity in treatment effectiveness is essential or non-essential.
        ii. Factors exist that affect treatment effectiveness but have no effect on outcome independent of treatment.

iii. Factors exist that affect outcome but do not affect treatment effect heterogeneity.

iv. Factors exist that affect both treatment effectiveness and outcome directly.

2. Compare treatment effect estimates generated by alternative instrumental variables methods in an empirical example with observational data and discuss the potential inferences that could be made from these estimates in light of a theoretical model characterizing the settings of treatment effect heterogeneity and treatment choice in the clinical scenario.

   a. Investigate the effects of angiotensin converting-enzyme inhibitors (ACE) and angiotensin receptor blockers (ARB) on 1-year survival amongst Medicare beneficiaries with acute myocardial infarction.

      i. Effectiveness of ACE/ARBs has been shown to vary by characteristics of patients such as presence of diabetes and left ventricular ejection fraction. Evidence suggests diabetes and left ventricular ejection fraction are related to both treatment choice and outcomes directly.

      ii. Survival has been shown to be affected by individual's socioeconomic status (e.g., income and education) but there is no evidence to suggest that these factors may affect the effectiveness of ACE/ARBs on survival.

   b. Generate estimates of ATE and LATE using nonlinear two-stage residual inclusion and nonlinear two-stage predictor substitution. Compare with estimates of LATE generated by linear two-stage least squares.

   c. Discuss interpretability of estimates given theoretical settings, in the context of simulation results.

CHAPTER 2

BACKGROUND AND REVIEW OF THE LITERATURE

Treatment Effect Heterogeneity

As originally discussed by Brooks and Fang (2009), if the objective of a researcher is to make causal inferences about the effect of a treatment (T) on outcome (Y) using observational data, then the researcher must first make assumptions regarding characteristics of treatment effect heterogeneity and circumstances underlying treatment choice.[7,16] Three types of factors must be considered: (1) X_1 factors that affect treatment effectiveness but have no direct effect on outcomes independent of treatment, (2) X_2 factors that affect both treatment effectiveness and outcomes directly, and (3) X_3 factors that affect outcomes directly but do not have any effect on the effectiveness of treatment. Using this notation, and letting $P(Y_i) \equiv P(Y_i = 1)$ be the underlying probability that individual i is "cured" (indicated by observed dichotomous outcome Y_i), a general outcome model is:

$$P(Y_i) = g(T(X_{1i}, X_{2i}), X_{2i}, X_{3i}) \qquad \{E1\}$$

For a given empirical problem, the researcher may theorize that any number of X_1, X_2, and/or X_3 factors exist. If only X_3 factors exist then the treatment effect is homogeneous, or constant, across individuals. This scenario can be illustrated as a simple linear model:

$$P(Y_i) = \beta_0 + \beta_1 T_i + \beta_3 X_{3i}. \qquad \{E2\}$$

β_0 is the baseline probability of positive outcome, Y. β_1 is the true treatment effect and β_3 is the effect of a unit change of X_3 on $P(Y)$. Alternatively, if either X_1 or X_2 factors exist, then treatment effects are heterogeneous across individuals. If only X_1 factors exist then treatment effects are heterogeneous, but no factors related to heterogeneity have a direct effect on outcome, independent of treatment. This scenario can be illustrated linearly as

$$P(Y_i) = \beta_0 + (\beta_{10} + \beta_{11}X_{1i})T_i. \qquad \{E3\}$$

The effect of treatment on outcome for individual i in {E3} is $(\beta_{10} + \beta_{11}X_{1i})$. The treatment effect for any individual is determined by a constant component (β_{10}) and a heterogeneous component $(\beta_{11}X_{1i})$ that is subject to individual's X_1 characteristics. If X_2 factors exist, then treatment effects are heterogeneous and X_2 factors affecting heterogeneity are also related to outcomes directly. This scenario can be illustrated as

$$P(Y_i) = \beta_0 + (\beta_{10} + \beta_{12}X_{2i})T_i + \beta_2X_{2i}. \qquad \{E4\}$$

Once again, the absolute effect of treatment on outcome $(\beta_{10} + \beta_{12}X_{2i})$ is determined by both a fixed and heterogeneous component. The scenario depicted by {E4} is distinct from that of {E3} because X_{2i} has a direct effect on outcome (through β_2), independent of treatment. Combining {E2}-{E4}, a general linear outcome model including X_1, X_2, and X_3 factors is

$$P(Y_i) = \beta_0 + (\beta_{10} + \beta_{11}X_{1i} + \beta_{12}X_{2i})T_i + \beta_2X_{2i} + \beta_3X_{3i}. \qquad \{E5\}$$

An alternative to this linear modeling approach ({E5}) is a nonlinear discrete outcome model. In this model, the observed binary outcome is the result of an index function on a continuous latent (i.e., unobserved by the researcher) variable. The general form of the nonlinear latent index model can be illustrated as

$$Y_i^* = \beta_T T_i + \beta_2X_{2i} + \varepsilon_i, \qquad \{E6\}$$

$$Y_i = \begin{cases} 1 \ if \ \ (Y_i^* > 0) \\ 0 \ if \ \ (Y_i^* \leq 0) \end{cases}. \qquad \{E7\}$$

Y^* is a continuous latent outcome variable, β_T is the effect of treatment (T) on Y^*, β_2 is the effect of X_2 on Y^*, Y is the observed dichotomous outcome, and ε is a random disturbance term drawn from a specified distribution (e.g., $\varepsilon \sim N(0, 1)$ is the probit model). Unlike linear models that benefit from straightforward interpretation of absolute treatment effects, β_T in {E6} cannot be interpreted as the true absolute effect of treatment on probability of outcome. β_T is a relative effect, the absolute effect of treatment must be estimated by a nonlinear function of all other factors affecting Y^* (i.e., X_2 factors).[13] In this model, the true absolute effect of treatment (TE_i) is forced to vary across individuals by their "baseline risk" determined by the covariates and the random disturbance term (ε) which defines the nonlinear function. Assuming that the error term (ε) is drawn from a standard normal distribution, {E6}-{E7} is the probit model and the absolute effect of treatment for individual i is

$$TE_i = \Phi(\beta_T + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i}), \qquad \{E8\}$$

where Φ represents the standard normal distribution function.[18,59] Because {E8} is nonlinear, TE_i varies across individuals by their X_{2i} characteristics. X_{2i} also affects outcome, independent of treatment choice. The absolute effect of an incremental increase of X_{2i} on outcome can be calculated as

$$X_{2_{Effect_i}} = \Phi\big(\beta_T T_i + \beta_2 (X_{2i} + 1)\big) - \Phi(\beta_T T_i + \beta_2 X_{2i}). \qquad \{E9\}$$

Nonlinear latent variable models of the form illustrated by {E6}-{E7} are commonly used to generate data for simulations that demonstrate consistency and identification properties of nonlinear instrumental variables methods.[18,20] However, common nonlinear models of the form of {E6}-{E7} require acceptance of several restrictive assumptions. The first of these assumptions is that this model forces the absolute effect of treatment to depend on all covariates affecting outcome. As illustrated by {E8}, all factors affecting outcome directly in this model are assumed to be "X_2 factors" (as defined above for {E4}) in that they also affect treatment effectiveness. Additional assumptions imposed by this model will be discussed in the next section.

It is possible to expand upon this simple nonlinear model to allow for X_1 and/or X_3 factors to exist or for heterogeneity to be essential. A nonlinear model with X_1 factors can be modeled by specifying X_1 as an interaction term, such that:

$$Y_i^* = \ (\beta_T X_{1i}) T_i + \beta_2 X_{2i} + \varepsilon_i. \qquad \{E10\}$$

β_T is the constant effect of treatment on Y^*. X_1 affects the effectiveness of treatment on outcome but not outcome directly, independent of treatment. The absolute effect of treatment for individual i is

$$TE_i = \Phi(\beta_T X_{1i} + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i}).$$

Unlike X_2 factors, which influence outcome (through β_2) regardless of whether $T = 1$ or $T = 0, X_1$ factors are interacted with T and therefore have no influence on outcome when $T = 0$. Assuming X_1 is a continuous random variable, the absolute effect of an incremental increase in X_{1i} on outcome can be illustrated by

$$X_1\_Effect_i = \Phi(\beta_T(X_{1i} + 1)T_i + \beta_2 X_{2i}) - \Phi(\beta_T(X_{1i})T_i + \beta_2 X_{2i}). \qquad \{E11\}$$

All variables are defined as in {E6}. If X_{1i} is constant across observations then {E10} is analogous to {E6}.

A nonlinear model with X_3 factors—factors affecting outcomes but unrelated to treatment effectiveness—is less straightforward to model because X_3 must be linearly additive to the probability of outcome determined by the nonlinear function generating Y^* in order to be unrelated to treatment effectiveness. A nonlinear model with X_3 factors is therefore a combination of both the linear and nonlinear modeling approaches; this model is similar to the quasi-linear utility model discussed by Varian.[60] The following model illustrates this approach:

$$P(Y_i) = f(T_i, X_{2i}; \beta) + \beta_3 X_{3i}. \qquad \{E12\}$$

The probability of outcome $P(Y_i)$ in {E12} includes both a nonlinear component ($f(T_i, X_{2i}; \beta)$) and linear component ($\beta_3 X_{3i}$). The nonlinear component is analogous to {E6}. The function $f(\blacksquare)$ represents the nonlinear functional form; in the case of a binary outcome with normally distributed error term, $f(\blacksquare)$ will be the standard normal distribution function and $0 < f(T_i, X_{2i}; \beta) < 1$. The effect of treatment on outcome in this model is determined solely by the nonlinear component. The linear component of {E12} includes only the X_3 factors that affect probability of outcome through β_3.

Because X_{3i} is not within the nonlinear function and is not interacted with treatment, it is not related to the effectiveness of treatment on outcome.

Treatment Choice and Conditions Necessary for Sorting-on-the-Gain

Beyond considerations for the types of factors in the outcome model, it must be acknowledged that observed treatment exposures in observational data are actually the result of treatment choices made by providers and patients.[7,16,17,48] These choices reflect a cost-benefit decision made by providers and patients. Costs could include monetary costs, travel costs, and expected side effects that may vary across patients. Benefits may include the value that the patient assigns to being cured and the expected effect of the treatment on outcome, which may be heterogeneous across patients. Therefore, in addition to considering what X_1, X_2, and X_3 factors exist in the outcome model, researchers must consider the treatment choice model and the conditions underlying treatment choice. If it is theorized that treatment decision makers may be "sorting-on-the-gain" with respect to individual's idiosyncratic gains from treatment, then the researcher must be careful in interpreting for whom estimated treatment effects may apply.[17,61]

Others have described sorting-on-the-gain (or what is known as essential heterogeneity[17]) as being present when selection into treatment is based on individual patient idiosyncratic gains from treatment.[17] Essential heterogeneity may be especially common in the analysis of treatment decisions since the choice of treatment is likely to be guided by expectations of individual patient benefits and risks associated with alternative treatment options.[17] When characteristics related to essential heterogeneity are unmeasured in empirical models, individuals with the same observed characteristics may be treated or untreated based upon unobserved factors contributing to different expected gains from treatment.

For the purposes of this research, we define essential heterogeneity more specifically as being present when treatment effectiveness and treatment value are directly related and strictly positively correlated. Under this definition of essential heterogeneity, patients with greater expected benefit from treatment (through higher treatment effectiveness) are more likely treated, all else equal, than those with lesser expected benefit. More formally, letting T_i^* be the value associated with treatment for individual i and TE_i be the effect of treatment on being "cured", essential heterogeneity is present when $\partial T^*/\partial TE > 0$. Following the simple linear models introduced above, essential heterogeneity can be modeled by including the factors that modify treatment effectiveness (i.e., X_1 and/or X_2 factors) in the first-stage treatment choice model. For example, a scenario of essential heterogeneity is represented by:

$$T_i^* = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + u_i \qquad \{E13\}$$

$$T_i = \begin{cases} 1 \ if \ \ (T_i^* > 0) \\ 0 \ if \ \ (T_i^* \leq 0)' \end{cases} \qquad \{E14\}$$

$$P(Y_i) = \beta_0 + (\beta_{10} + \beta_{11} X_{1i} + \beta_{12} X_{2i}) T_i + \beta_2 X_{2i} + \beta_3 X_{3i}. \qquad \{E15\}$$

T_i^* is a latent variable indicating an individual's "value" associated with treatment and T_i is the observed discrete treatment choice of individual i. This interpretation of the latent variable T_i^* in {E13} as an individual's "value" associated with treatment suggests that individuals with positive value associated with treatment are treated and those with zero or negative value are not treated. X_{1i}, X_{2i}, and X_{3i} in {E13} and {E15} are defined as in {E5}. α_j is the coefficient relating X_{ji} to treatment value (T_i^*). X_{4i} are measured factors that affect treatment value, and therefore treatment choice, but have no effect on outcomes (i.e., they are not in {E15})—these factors are candidates for instrumental variables. The effect of treatment on outcome for individual i is $TE_i = (\beta_{10} + \beta_{11} X_{1i} + \beta_{12} X_{2i})$. u_i is a random disturbance term composed of all unmeasured variation in treatment value. Sorting-on-the-gain, as defined for the purposes of this research, is present when corresponding coefficients α_1 and β_{11}, and α_2 and β_{12}, are of the same sign (i.e., either both positive or both negative). If, on the other hand, α_1 and β_{11} are of opposite signs then treatment decision makers would be preferentially treating patients with lower benefit from treatment—those with less to gain from treatment would be more likely treated, all else equal. While past definitions of essential heterogeneity have not explicitly excluded this possibility, this is counterintuitive to ideas of rationality and utility maximization. Why would individuals choose treatment if not for the associated positive effect on a valued outcome?

If treatment effects are heterogeneous but decision makers are *not* sorting-on-the-gain then heterogeneity is non-essential.[17] Under non-essential heterogeneity X_1 and/or X_2 factors are theorized to exist but are either not observed by treatment decision makers prior to treatment choice or decision makers ignore the treatment effect information from these variables.[17] In a simple linear model, non-essential heterogeneity can be modeled by omitting X_1 and/or X_2 factors from the first-stage treatment choice model when these factors affect treatment effectiveness in the outcome model. For example, assuming {E15} is the outcome model, non-essential heterogeneity can be modeled as

$$T_i^* = X_{3i} + X_{4i} + u_i \qquad \{E16\}$$

$$T_i = \begin{cases} 1 \ if \ \ (T_i^* > 0) \\ 0 \ if \ \ (T_i^* \leq 0) \end{cases}. \qquad \{E17\}$$

In this scenario, X_1 and X_2 do not affect treatment value or choice and therefore the distribution of X_1 and X_2 factors can be assumed to be evenly distributed across treated and untreated individuals. As stated in the introduction, when heterogeneity is non-essential—as in {E16}—estimates of LATE can be generalized to estimate ATE.[17,61] This point will be explained in greater detail in the following section.

Understanding the ideas of essential and non-essential heterogeneity now allows us to discuss a second key assumption imposed by the traditional nonlinear model. Assuming the standard nonlinear model {E6}-{E7} is the outcome model, the treatment choice model in the common nonlinear modeling approach can be illustrated by

$$T_i^* = \alpha_2 X_{2i} + \alpha_4 X_{4i} + u_i, \qquad \{E18\}$$

$$T_i = \begin{cases} 1 \ \ if \ \ (T_i^* > 0) \\ 0 \ \ if \ \ (T_i^* \leq 0) \end{cases}. \qquad \{E19\}$$

X_{4i} are factors related to treatment value (T_i^*) but unrelated to outcome in {E7}—these are candidates for instrumental variables. u_i is a random disturbance term. While X_{2i} factors in this common nonlinear model are related to both treatment effectiveness ({E8}) and treatment choice ({E18}), this scenario may not be consistent with our definition of essential heterogeneity. Treatment effectiveness in this common nonlinear model affects treatment choice in a nonlinear manner and, depending upon the parameter values (i.e, β_2 in {E6}), may be negatively associated with treatment value and probability of being treated. This does not make intuitive sense—why would patients with greater benefit from treatment be *less* likely to be treated, all else equal? This nonlinear model ({E6}, {E18}) does not fit the current description of non-essential heterogeneity, either. Scenarios of *non-essential heterogeneity* have been described as circumstances where treatment effectiveness is either unobserved or ignored by decision makers, such that average treatment effects for treated, untreated, or other subpopulations can be expected to be equal.[17] This will not be the case for this nonlinear model where treatment effect heterogeneity is forced to vary across patient subpopulations by a nonlinear function of factors in the outcome equation. Therefore, while distinct factors (X_{2i}) driving heterogeneity in treatment effectiveness enter into the treatment choice equation, treatment effect heterogeneity cannot necessarily be characterized as either essential or non-essential in this non-linear specification.

Characterizing Essentiality in Terza Simulation Model

This point can be illustrated with the nonlinear binary treatment and binary outcome model used by Terza et al. (2007) to demonstrate the superiority of nonlinear IV methods, relative to 2SLS, for estimating ATE in inherently nonlinear models.[18] Treatment choice was modeled as

$$T_i^* = \alpha_{21}X_{21i} + \alpha_{22}X_{22i} + \alpha_{41}X_{41i} + \alpha_{42}X_{42i}, \qquad \{E20\}$$

$$T_i = \begin{cases} 1 \; if \;\; (T_i^* > 0) \\ 0 \; if \;\; (T_i^* \leq 0) \end{cases},$$

and outcome was modeled as

$$Y_i^* = \beta_T T_i + \beta_{21}X_{21i} + \beta_{22}X_{22i} + \varepsilon_i, \qquad \{E21\}$$

$$Y_i = \begin{cases} 1 \; if \;\; (Y_i^* > 0) \\ 0 \; if \;\; (Y_i^* \leq 0) \end{cases}.$$

Variable definitions for this example are detailed in Table 1. The value of treatment for individual i is T_i^* from {E20} and the absolute treatment effect for individual i is

$$TE_i = \Phi(\beta_T + \beta_{21}X_{21i} + \beta_{22}X_{22i}) - \Phi(\beta_{21}X_{21i} + \beta_{22}X_{22i}).$$

If heterogeneity in this simple nonlinear model example is essential then this should be evident by (1) a clear positive correlation between TE and T^*, and (2) true treatment effectiveness being greater for treated individuals than untreated individuals. Figure 1 shows a quadratic prediction fit for the relationship between TE and T^* based on 200,000 simulated observations using the Terza treatment choice and outcome model—{E20} and {E21}, respectively. Figure 2 shows a scatter plot of this same data. From this simple illustration, it is clear that treatment effectiveness across patients is not directly and positively related to the value of treatment, as would be expected under scenarios of essential heterogeneity.

Table 1: Variable Definitions for Terza et al. (2007) Binary Treatment and Outcome Model[18]

Covariate	Definition
X_{21}	Uniform(-1.73, 1.73)
X_{22}	Normal(0, 1)
X_{41}	Uniform(-2.45, 2.45)
X_{42}	Uniform(-2.45, 2.45)
ε	Normal(0, 1)
α_{21}	0.1
α_{22}	1
α_{41}	-0.25
α_{42}	0.2
β_T	2.5
β_{21}	1
β_{22}	1
TE_i	$\Phi(\beta_T + \beta_{21}X_{21i} + \beta_{22}X_{22i}) - \Phi(\beta_{21}X_{21i} + \beta_{22}X_{22i})$
Φ	Standard Normal Distribution Function

Figure 1: Quadratic Fit of Treatment Value and Treatment Effectiveness (N = 200,000)

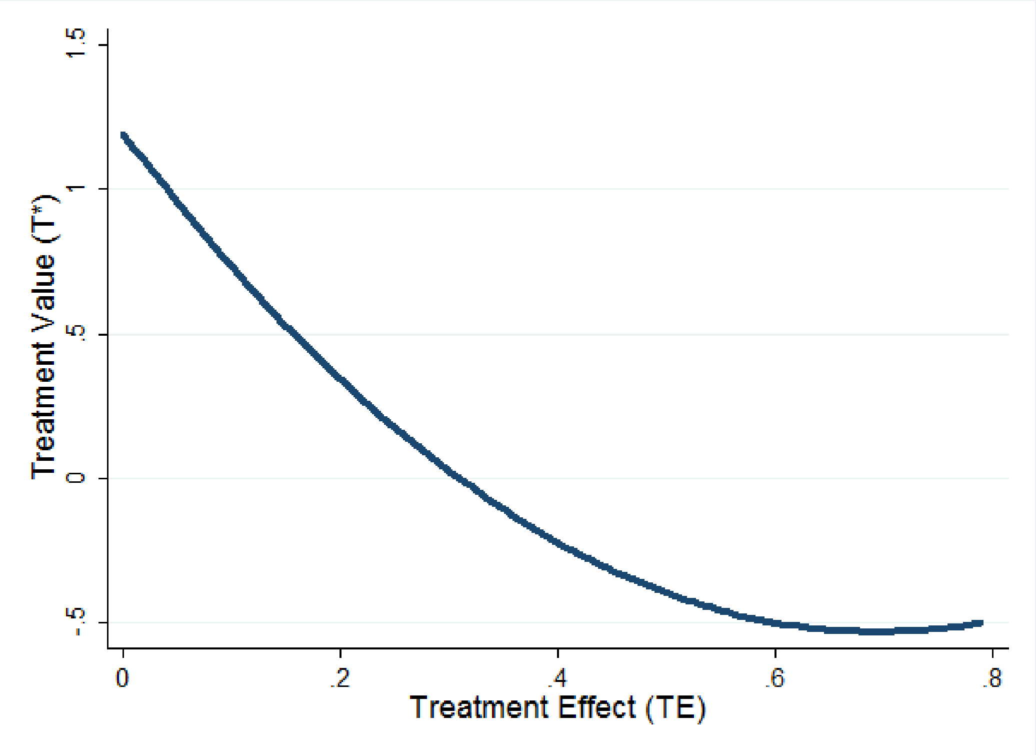
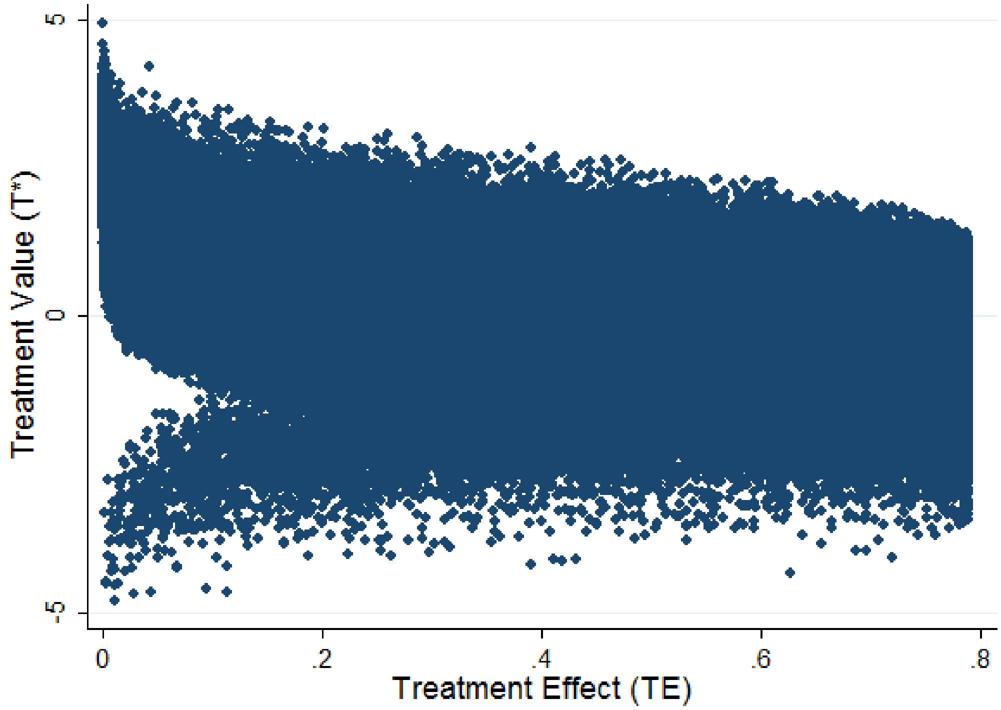


Figure 2: Scatter-Plot of Treatment Value and Treatment Effectiveness (N = 200,000)



While this does not fit the definition for essential heterogeneity, treatment effect heterogeneity in this example cannot be characterized as non-essential either because there is some apparent sorting. Heterogeneity in this standard nonlinear scenario may perhaps be described as "quasi-essential"—treatment effectiveness is related to treatment value, but they are not directly or strictly positively correlated. Patients with higher benefit from treatment do not have higher probability of treatment, as would be expected. Somewhat illogically, treatment decision makers in this simulated example are sorting patients such that the patients more likely to be treated (greater T_i^* value) are the patients with lower treatment effectiveness. Table 2 shows the summary statistics for treatment effectiveness, treatment value, probability of treatment, and probability of positive outcome across deciles of treatment effectiveness. From this table, it is clear that the average value of treatment and probability of being treated decreases as treatment effectiveness increases. Patients in the highest decile of treatment effectiveness also have the lowest mean probability of positive outcome without treatment, $P(Y_i = 1|T_i = 0)$.

Table 2: Summary Statistics by Deciles of Treatment Effectiveness

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	Average Treatment Value (T*)	% Treated	% Y = 1	% Marginal*
1	20,000	.012	1.41	98	99.6	4.5
2	20,000	.067	.67	82	97	18
3	20,000	.16	.32	66	90	22
4	20,000	.27	.070	55	81	20
5	20,000	.39	-.13	47	71	19
6	20,000	.51	-.29	40	59	18
7	20,000	.61	-.42	33	48	17
8	20,000	.70	-.50	29	39	18
9	20,000	.75	-.56	25	33	20
10	20,000	.78	-.58	28	30	22
Total	200,000	.42	-.0015	50	65	18

** Marginal defined as -.25 < T* < .25.*

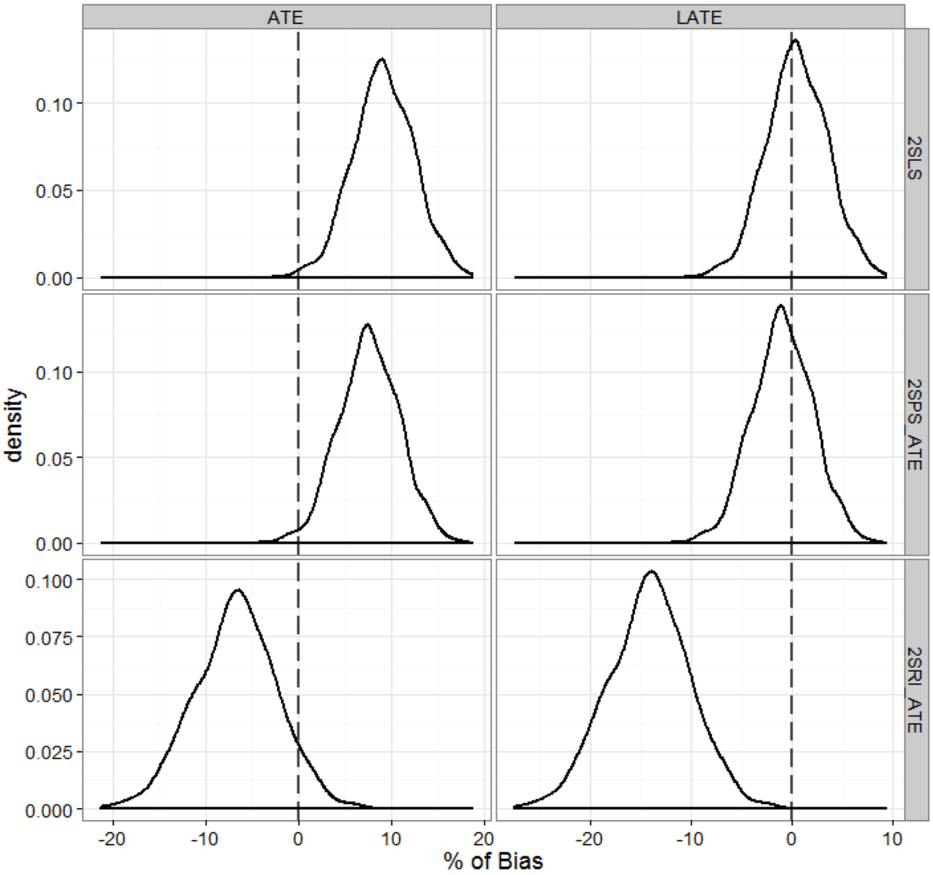
Table 2 also shows the proportion of marginal patients across deciles of treatment effectiveness. Terza et al. do not discuss or provide a definition for marginal patients. For the purposes of this example, marginal patients are defined as individuals satisfying $-.25 < T_i^* < .25$; approximately 18% of the population meets these criteria. These patients are those for whom a small positive or negative change in the value of instruments may change treatment choice. As shown in Table 2, the marginal patients are fairly evenly distributed across the distribution of treatment effectiveness—with the exception of the lowest quintile which has a relatively low proportion of marginal patients. This could be a considered as representing a rather unique clinical scenario. Under essential heterogeneity, it may be more likely that treatment choices are fairly certain for those with very high or low treatment effectiveness and marginal patients are unique in terms of treatment effectiveness.

Results from simulations examining the bias of 2SLS and nonlinear 2SRI using this model are given in Table 3. Results in Table 3 reflect the percentage difference between mean estimates and mean true values across 1,000 simulations of 20,000 observations per simulation. The estimates from 2SRI were less biased for ATE than estimates from 2SLS or 2SPS—a result consistent with the findings of Terza and colleagues—though no method produces completely unbiased estimates of ATE here. Comparing estimates to true LATE, however, shows that 2SLS is consistent and drastically less biased than 2SRI for estimating LATE. 2SPS is also substantially less biased for LATE than 2SRI, and is just slightly more biased than 2SLS. Claims that 2SLS is generally inconsistent in models with inherently nonlinear dependent variables are misleading; 2SLS may not provide consistent estimates of the ATE, but is unbiased and consistent for LATE with minimal assumptions. Density plots showing the bias of estimates for alternative estimators are provided in Figure 3: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 3: Average % Bias of Estimate, by Estimator (Terza et al. model)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	9.11	0.51
2SRI	-6.79	-14.14
2SPS	7.58	-0.90

Figure 3: Density Plot of % Bias of Estimates, by Estimator, for Terza Scenario



Identification in Alternative Models of Treatment Effect Heterogeneity and Choice

Direct risk adjustment estimators (e.g., linear regression by OLS) generate estimates of the average treatment effect on the treated (ATT).[12,16,59] Two-stage moment based estimators using instrumental variables (e.g., 2SLS) generate estimates of local average treatment effects (LATE)—an estimate of the average treatment effect across the subpopulation of "compliers", or "marginal patients" whose treatment choice was determined by the instrumental variables used.[2] Alternative average treatment effect parameters, such as the average treatment effect across everyone (ATE), the average treatment effect across untreated patients (ATU), or the average treatment effect for distinct patients or patient subgroups cannot be identified without further assumptions.[1,2,8,15,16] The extent to which alternative average treatment effect parameters can be identified depends on the following considerations: (1) which factors in X_1, X_2, and X_3 exist; (2) whether heterogeneity in treatment effectiveness, if it exists, is essential (i.e., whether X_1 or X_2 factors exist and affect treatment choice); (3) which factors in X_1, X_2, and X_3 factors are measured and included in the empirical model of the researcher; and (4) whether specified instruments affect treatment choices across all distinct patient subpopulations.[15] The evidence provided by Terza et al. (2007, 2008), however, shows a specific scenario where nonlinear 2SRI methods generate consistent estimates of ATE in models with unmeasured confounders and treatment effect heterogeneity.[18,19] Terza et al. (2002) suggests that estimation of average treatment effects for specific patient subgroups is possible using nonlinear 2SRI, without needing to accept any of these aforementioned assumptions.[62] The remainder of this section discusses identification of treatment effect parameters across alternative possible scenarios of the circumstances underlying treatment choice, treatment effect heterogeneity, and outcomes.

The simplest scenario is that of homogeneous treatment effects. Under this scenario, there are X_3 factors but no X_1 or X_2 factors. This scenario can be illustrated using the following linear model:

$$T_i^* = \alpha_0 + \alpha_3 X_{3i} + \alpha_4 X_{4i} \qquad \{E22\}$$

$$P(Y)_i = \beta_0 + \beta_1 T_i + \beta_3 X_{3i}. \qquad \{E23\}$$

Consistent with previous examples, T_i^* represents the latent treatment value and T_i is the observed treatment choice for individual i. If all X_3 factors are measured and included in the empirical model then direct regression methods will identify the ATT. Because treatment effects are homogenous in this scenario, estimates of ATT can be generalized to alternative treatment effect parameters such as LATE, ATE, ATT, and ATU. If any X_3 factors are unmeasured or not included in the empirical model then simple direct regression estimation of β_1 will be biased for the ATT. In this case, two-stage moment based estimators, such as 2SLS, with a valid instrumental variable (X_4) can be used to generate consistent estimates of the LATE. Because the treatment effect is homogeneous across observations, LATE=ATT=ATE=ATU.[16,17]

Alternatively, a scenario with X_1 factors—factors related to the effect of treatment on outcome but not outcome directly—can be illustrated by

$$T_i^* = \alpha_0 + \alpha_1 X_{1i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} \qquad \{E24\}$$
$$P(Y)_i = \beta_0 + (\beta_{10} + \beta_{11} X_{1i}) T_i + \beta_3 X_{3i}. \qquad \{E25\}$$

Assuming α_1 and β_{11} have the same sign, this is a scenario of essential heterogeneity—X_1 factors affect both treatment effect heterogeneity and treatment choice and $\partial T^*/\partial TE > 0$. If α_1 and β_{11} have opposite signs then patients with higher effectiveness will be less likely to be treated—this scenario is not logical and is excluded from our definition of essential heterogeneity. Provided that all X_{3i} factors in {E25} are measured and included in the empirical model, direct regression methods will identify the ATT.[12,16,59,63] Similarly, two-stage moment-based estimators with valid instruments (X_{4i}) can be used to identify the LATE. Heterogeneity is essential in this scenario; decision makers are sorting on the gain with respect to X_{1i} such that individuals with higher values of X_{1i} will have higher treatment effects and higher probability of treatment. As such, treatment effect estimates will be sensitive to the distribution of X_1 characteristics in the sample and estimates of ATT or LATE cannot generally be used to make inferences about ATE or ATU as X_1 will be distributed differently across these groups.[15,16] If X_1 is measured, it may be possible to make inferences about treatment effect heterogeneity by either stratifying the sample by X_1 (if X_1 is a factor variable) to make inferences about treatment effects within X_1-defined groups or modeling an interaction term between treatment and X_1 in the empirical model to generate an estimate of β_{11}. As purported by Angrist and demonstrated in an example by Brooks, instrumental variables estimators can only possibly identify alternative average treatment effects beyond LATE under settings of essential heterogeneity if (1) the instruments affect treatment choice across all patient subpopulations and (2) all factors driving essential heterogeneity are measured by the researcher.[15,16,49] This limitation has not been discussed, however, in the context of estimating average treatment effect parameters using the nonlinear 2SRI estimator. Terza et al. (2008) suggest that nonlinear 2SRI methods generate generally consistent estimates of ATE in models with inherently nonlinear dependent variables, but 2SRI has not been evaluated in models with X_1 factors or essential heterogeneity.[19]

If, on the other hand, we suppose that X_1 factors do not affect treatment choice, then heterogeneity is *non-essential*. Consider a scenario in which the outcome model is {E25}, but X_{1i} is either unobserved or ignored by treatment decision makers, such that:

$$T_i^* = \alpha_0 + \alpha_3 X_{3i} + \alpha_4 X_{4i}. \qquad \{E26\}$$

This may represent a scenario where, for example, technology or current medical knowledge has not allowed for physicians to be aware of individual's distinct X_{1i} characteristics. While physicians may know that treatment effects vary across individuals, they are not able to predict individual's idiosyncratic gains when making treatment decisions and therefore cannot "sort" patients based on these expected gains. Under non-essential heterogeneity, simple direct regression methods will identify the ATT and two-stage moment based estimators using instrumental variables will identify the LATE. These estimates of ATT and LATE are not subject to the distribution of X_1 in the sample and LATE=ATT=ATE=ATU, even though treatment effects do vary or are heterogeneous across individuals in truth.[17]

Another scenario can be illustrated if we suppose that there are no X_1 factors present but both X_2 factors and X_3 factors exist, such that

$$T_i^* = \alpha_0 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} \qquad \{E27\}$$
$$P(Y)_i = \beta_0 + (\beta_{10} + \beta_{12} X_{2i}) T_i + \beta_2 X_{2i} + \beta_3 X_{3i}. \qquad \{E28\}$$

Assuming α_2 and β_{12} have the same sign, heterogeneity in this example is essential because X_{2i} affects both treatment effect heterogeneity and treatment choice directly such that $\partial T^*/\partial TE > 0.$ X_{2i} also affects outcomes directly and is therefore a confounder. Assuming X_{3i} and X_{4i} are measured, while X_{2i} is observed by the patient/provider making treatment decisions but unmeasured by the researcher, direct regression methods will generate estimates of the ATT that are biased and two-stage moment based IV methods will generate consistent estimates of the LATE. Because heterogeneity is essential and decision-makers are sorting-on-the-gain with respect to individual's idiosyncratic gains (X_{2i} factors) the estimates of LATE will reflect the distribution of X_2 in the group of patients whose treatment choice was sensitive to the instruments and LATE≠ATT≠ATE.[15,16] While Terza et al. (2008) suggests that nonlinear 2SRI methods may generate consistent estimates of ATE in nonlinear models, this has yet to be shown in scenarios of essential heterogeneity.[19] As was demonstrated in the previous section, the common nonlinear models in which 2SRI has been examined do not have the essential heterogeneity property. If X_2 is measured, it may be possible to make inferences about treatment effect heterogeneity by either stratifying the sample by X_2 (if X_2 is a factor variable) to make inferences about treatment effects within X_2-defined groups or by modeling an interaction term between treatment and X_2 in the empirical model to generate an estimate of β_{12}.

If we instead suppose that X_2 factors exist but do not affect treatment choice, then heterogeneity is *non-essential*. Under non-essential heterogeneity the treatment choice model becomes analogous to {E26}. Unmeasured X_2 factors are no longer confounders and estimates from simple direct regression methods and 2SLS methods will be consistent for the ATT and LATE, respectively. Additionally, because there is no sorting-on-the-gain, the distribution of X_2 factors can be assumed to be consistent across treated and untreated populations and LATE=ATT=ATE=ATU.[17]

Similar concepts can be demonstrated using a nonlinear modeling approach. Using the nonlinear binary outcome model, outcome (Y_i) is the observed binary result of an index function on an underlying latent variable (Y_i^*). A simple illustration of this model is given by:

$$\begin{aligned} T_i^* &= \alpha_2 X_{2i} + \alpha_4 X_{4i} + u_i \\ T_i &= 1(T_i^* > 0) \qquad \{E29\} \end{aligned}$$

$$\begin{aligned} Y_i^* &= \beta_1 T_i + \beta_2 X_{2i} + \varepsilon_i \\ Y_i &= 1(Y_i^* > 0). \qquad \{E30\} \end{aligned}$$

In {E29} and {E30}, observed binary treatment status T_i and outcome Y_i are generated by index functions on the underlying latent variables T_i^* and Y_i^*, respectively. u_i and ε_i are random disturbance terms drawn from some theorized distribution, such as the standard normal in the case of the probit model. Unlike previous scenarios, β_1 in {E30} cannot be interpreted as the absolute effect of treatment on outcome (e.g., change in the probability of a cure). The absolute effect of treatment on outcome for the model illustrated by {E30} is a nonlinear function varying with X_{2i}. Given individual i's X_{2i}, the absolute effect of treatment on outcome is the difference between the probability that $Y_i^* > 0$ when $T_i = 1$ and $T_i = 0$, which is a calculated by a function related to the model disturbance term (ε_i). Using an example of the probit model—where $\varepsilon \sim N(0,1)$—the absolute treatment effect at the individual level is

$$TE_i = \Phi(\beta_1 + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i}). \qquad \{E31\}$$

$\Phi(\cdot)$ denotes the standard normal distribution function. In this example, the ATE is the mean of {E31} across the population of observations, ATT is the mean of {E31} across the treated subpopulation, and LATE is the mean of {E31} over the subset of marginal individuals whose treatment choice was influenced by the instruments (X_{4i}).

As discussed previously, even though X_{2i} affects both TE_i and T_i^*, the scenario illustrated by {E29}-{E30} cannot be immediately described as either essential, or non-essential. For the purposes of this research, essential heterogeneity is defined as being present when treatment effectiveness has a direct effect on treatment choice, such that TE_i and T_i^* are positively correlated over their ranges. While there may be some relationship between TE_i and T_i^* in the nonlinear model {E29}-{E30}, such that ATU≠ATT≠ATE, this relationship will be nonlinear and treatment effectiveness does not have a direct or necessarily positive effect on treatment value and therefore the probability of treatment.

If it is theorized that either X_1 or X_3 factors exist, or that decision makers are sorting-on-the-gain such that those with greater expected benefit from treatment are more likely treated, then this common nonlinear model may be inappropriate. Despite these limitations and the potential threats they pose for accuracy and interpretability of treatment effect estimates, this restrictive scenario ({E29}-{E30}) is, to my knowledge, the only scenario under which properties of nonlinear 2SRI models have been demonstrated in simulations with binary dependent variables.[18-20,46,64]

Building on the common nonlinear binary treatment-outcome model illustrated by {E29}-{E30}, X_1 factors that directly modify treatment effectiveness but are not related to outcomes, independent of treatment, can be modeled as an interaction term on treatment choice in the outcome model. Assuming that heterogeneity in non-essential, this model can be illustrated as

$$T_i^* = \alpha_4 X_{4i} + u_i$$

$$T_i = 1(T_i^* > 0) \qquad \{E32\}$$

$$Y_i^* = \beta_1 X_{1i} T_i + \varepsilon_i$$

$$Y_i = 1(Y_i^* > 0). \qquad \{E33\}$$

X_{1i} in {E33} affects the effect of treatment on outcome, but is unrelated to outcome independent of treatment—X_{1i} does not affect outcome when $T_i = 0$. All other variables are defined as in {E29}-{E30}. Heterogeneity in this model is non-essential because X_{1i} does not affect T_i^* and is therefore unrelated to probability of treatment. Similar to the linear model with X_1 factors, risk adjustment estimators will identify an estimate of the ATT and two-stage moment based estimators will identify LATE. Because decision makers are not sorting on the gain, ATE=ATT=ATU=LATE. Alternatively, it may be that X_{1i} affects treatment choice, such that the treatment choice model becomes

$$T_i^* = \alpha_1 X_{1i} + \alpha_4 X_{4i} + u_i. \qquad \{E34\}$$

In this scenario, X_{1i} has a direct effect on treatment value and, as an interaction term on β_1 in {E33}, has a direct effect on treatment effectiveness. If α_1 and β_1 (in {E33}) have the same sign, then treatment effectiveness and treatment value will be positively correlated and this scenario is one of essential heterogeneity. Two-stage moment based IV estimators such as 2SLS will identify LATE but, because the distribution of X_1 factors will differ across treated and untreated subpopulations, it may not be possible to use these LATE estimates to estimate alternative average treatment effect parameters such as the ATE, ATT, or ATU. Our definition of essential heterogeneity does not allow for α_1 and $\beta_1$to be of opposite sign.

Two-Stage Instrumental Variables Methods

The simple theoretical dummy endogenous variable model illustrated below will be the basis for demonstration and comparison of the 2SLS and 2SRI instrumental variables estimators. The model in this example is analogous to the homogeneous treatment effect scenario discussed in the previous section ({E22}-{E23}). Notation used below follows from the previous sections. For person i, let Y_i be the observed continuous health outcome, T_i be observed binary treatment status that is determined by the unobserved latent variable T_i^*. X_{3i} is an unobserved confounding factor, X_{4i} is a set of one or more instrumental variables, and X_{5i} is a set of one or more observed factors affecting outcomes directly but unrelated to treatment choice. This simple binary model can be written as:

$$T_i^* = \alpha_0 + \alpha_3 X_{3i} + \alpha_4 X_{4i} + u_i \qquad \{E35\}$$
$$Y_i = \beta_0 + \beta_1 T_i + \beta_3 X_{3i} + \beta_5 X_{5i} + \varepsilon_i, \qquad \{E36\}$$

where

$$T_i = 1(\alpha_0 + \alpha_3 X_{3i} + \alpha_4 X_{4i} + u_i > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}. \qquad \{E37\}$$

β_1 represents the homogeneous causal effect of treatment (T_i) on outcome (Y_i). The index function {E37} generating T_i as a function of T_i^* follows from the notion that treatment choice is determined by a comparison of the expected benefit from treatment to the expected costs incurred from treatment. If expected benefit exceeds expected costs then net expected utility from treatment is positive (i.e., $T_i^* > 0$) and the individual chooses treatment. Only the result of this cost-benefit decision (T_i) is observed by the researcher, T_i^* is unobserved. This simple model is a common econometric approach for representing discrete choice.[2,16,18,20]

Endogeneity of T_i in {E36} is caused by the one or more unmeasured X_{3i} factors that are directly related to both T_i^* and Y_i, such that there is correlation between the disturbance terms u_i and ε_i (i.e., $E[u_i \varepsilon_i] \neq 0$). This will result in biased estimates of β_1

from simple direct regression methods. However, given minimal assumptions, two-stage moment based estimators can be used to generate consistent estimates of causal local average treatment effects (LATE).[2] The first of these assumptions is that there must exist one or more observable factors (X_4) that are independent of disturbance terms—u and ε—in each equation, such that:

$$E[X_4 u] = 0, \quad E[X_4 \varepsilon] = 0. \qquad \{A1\}$$

In other words, the researcher must have measured X_4 factors available that are conditionally independent of outcomes or other unmeasured variables related to outcomes. If {A1} holds, then any effect of X_4 factors on Y must be through the effect of X_4 on T. A second key assumption is that the covariance between T and X_4 is non-zero, or

$$cov(T, X_4) \neq 0. \qquad \{A2\}$$

Assumption {A2} mandates that α_4 in {E35} be significantly different from zero—that the specified instruments have a significant, non-arbitrary effect on treatment choice. Taken together, {A1} and {A2} suggest that the variation in the observed endogenous treatment (T) determined by X_4 factors is significantly different from zero and exogenous with respect to outcomes.

An additional assumption is Stable Unit Treatment Value Assignment (SUTVA).[2] SUTVA implies that potential outcomes for any individual are unrelated to the treatment status of any other individual. Violation of SUTVA occurs, for example, when there are spillover effects of treatment. A scenario in which SUTVA may be violated is in estimating effects of vaccination: As individuals receive vaccination the probability of being exposed to the disease that the vaccination was intended to prevent may decrease for all vaccinated and unvaccinated individuals through effects of herd immunity. This represents a change in the potential outcomes for these individuals and violation of SUTVA.

The final assumption to be noted is the monotonicity assumption. Monotonicity requires that there is no individual who does the opposite of the assignment implied by the instrument.[2] In other words, the monotonicity assumption asserts that the direction of the effect of X_4 on T^*, whether positive or negative, is consistent across individuals. Angrist et al. (1996) define four "types" of observations: always-takers, never-takers, compliers, and defiers.[2] For the case in which X_4 is binary, let $T_i(X_4) = T$ represent treatment choice (T) for individual i with $X_{4i} = X_4$. For example, $T_i(1) = 1$ indicates that individual i received treatment ($T_i = 1$) for $X_{4i} = 1$. With this notation, always-takers are those who will always choose treatment regardless of the value of the instrument, such that $T_i(1) = T_i(0) = 1$. Never-takers are those who will never choose treatment, regardless of the value of the instrument, such that $T_i(1) = T_i(0) = 0$. Compliers are those whose treatment choice is sensitive to the instruments, such that either $T_i(1) > T_i(0)$ or $T_i(1) < T_i(0)$. The existence of defiers—individuals whose treatment choice goes against the direction suggested by the instrument—is precluded by the monotonicity assumption, which implies that either $T_i(1) \geq T_i(0)$ or $T_i(1) \leq T_i(0)$ for all $i = 1, \dots, N$.[2]

Interpreting the Instrumental Variables Estimand

The potential outcomes framework formalized by Rubin (1974, 1978, 2000) and further discussed in the context of IV models by Imbens et al. (1994) and Angrist et al. (1996) suggests that all individuals have distinct potential outcomes under all counterfactual treatment states, despite their ultimately being observed in only one of these states.[1,2,65-67] For example, in the binary treatment-control model ({E32}-{E33}), each treated individual has an observable outcome in the treated state and an unobservable counterfactual outcome in the control state. Similarly, each untreated individual has an observed outcome in the untreated state and an unobserved counterfactual outcome in the treated state.

This framework suggests that a true causal treatment effect exists for each individual unit of observation that is equal to the difference in potential outcomes between the treated and untreated states. To illustrate this point more clearly, let Y_i^1 be the potential outcome for individual i in the treated state ($T_i = 1$) and Y_i^0 be the potential outcome for individual i in the untreated state ($T_i = 0$). The causal effect of treatment for individual i is

$$\beta_i = Y_i^1 - Y_i^0. \qquad \{E38\}$$

However, because it is generally the case that only a single outcome state can be observed for any single individual, this individual treatment effect cannot be directly observed or calculated. Instead, researchers have focused on the estimation of various average causal treatment effects such as the average treatment effect across the entire population (ATE), the average treatment effect on the treated (ATT), the average treatment effect on the untreated (ATU), and the Local Average Treatment Effect (LATE).

Two-stage moment-based estimators—such as 2SLS—generate estimates of LATE.[2] Estimates of LATE represent the average effect of treatment for individuals whose treatment choices were sensitive to the instrumental variable. Using these methods, the causal effect of treatment for person i is estimated as the causal effect of changes in the instrument X_{4i} on outcome Y_i, or

$$(Y_i^1 - Y_i^0) \cdot \big(T_i(X_{4i} = 1) - T_i(X_{4i} = 0)\big). \qquad \{E39\}$$

Y_i^1 and Y_i^0 represent the outcomes for individual i when the treatment (T_i) equals 1 and 0, respectively. $T_i(X_{4i} = 1)$ and $T_i(X_{4i} = 0)$ represent the treatment decision for individual i when the instrument (X_4) equals 1 and 0, respectively. The estimate represented by {E39} is nonzero only when treatment choice is influenced by the instrument. The IV estimand is therefore the weighted sum of average causal effects for the subpopulation of individuals whose treatment choice was sensitive to the instrument—that is, the individuals for whom $T_i(X_{4i} = 1) \neq T_i(X_{4i} = 0)$. If treatment choice is not changed as a result of a change in the instrument then $\big(T_i(X_{4i} = 1) - T_i(X_{4i} = 0)\big) = 0$ and {E39} will equal zero. Since the monotonicity assumption rules out the existence of defiers, the IV estimand is the average causal effect of treatment for the compliers. The average causal effect across the compliers is the Local Average Treatment Effect (LATE). Assuming a binary IV, the LATE parameter is

$$\beta_{LATE} = E\big[\big(Y_i(1) - Y_i(0)\big)|T_i(1) - T_i(0) = 1\big]. \qquad \{E40\}$$

As discussed in the previous section, β_{LATE} cannot be interpreted as the average treatment effect for the entire population (ATE) or the average treatment effect amongst the treated population (ATT) without further strong assumptions about characteristics of treatment effect heterogeneity and treatment choice. Moreover, since the specific observations belonging to the group of compliers cannot be identified directly from observational data, inferences may not be made for any specific subpopulation defined in terms of the covariates.[2]

Two-Stage Least Squares (2SLS)

Following the example drawn in {E35} and {E36}, and assuming that X_3 factors are unmeasured by the researcher, the 2SLS approach can be operationalized by replacing the endogenous regressor (T_i) by its linear projection

$$L(T|X_4) \equiv X_4{}'E^{-1}(X_4X_4{}')E(X_4T), \qquad \{E41\}$$

where X_4 is a vector of valid instrumental variables. This is the "first stage" of the two-stage least squares method. In the second stage, $L(T_i|Z_i, X_i)$ replaces observed T_i in the outcome equation, such that

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 T_i + \varepsilon_i \\ = {} & \beta_0 + \beta_1\big((T_i - L(T_i|X_{4i})) + L(T_i|X_{4i})\big) + \varepsilon_i \qquad \{E42\} \\ = {} & \beta_0 + \beta_1 L(T_i|X_{4i}) + \big\{\varepsilon_i + \beta_1\big(T_i - L(T_i|X_{4i})\big)\big\}. \qquad \{E43\}^{39} \end{aligned}$$

Because $\beta_1\big((T_i - L(T_i|X_{4i})) + L(T_i|X_{4i})\big)$ is additive in the linear model ({E42}), the variation in T_i that is not predicted by the linear projection ({E41}) can be relegated directly to the error term ($\{\varepsilon_i + \beta_1(T_i - L(T_i|X_{4i}))\}$), which is assumed to be uncorrelated with X_{4i} or $L(T_i|X_{4i})$ through the exclusion restriction ({A1}). If the assumptions for a valid instrument—{A1} and {A2}—hold, then the least squares estimator of Y_i on $L(T_i|X_{4i})$ is consistent for mean β_1 across the compliers (i.e., β_{LATE}).[2,47]

More simply put, the first stage of the 2SLS procedure generates predicted values of the endogenous treatment variable ($L(T_i|X_{4i})$) by regressing observed treatment (T_i) on one or more IVs (X_{4i}) and other variables related to outcomes or treatment choice using the ordinary least squares (OLS) method. $L(T_i|X_{4i})$ replaces observed treatment in the second stage linear regression on outcome and the parameter estimate $\hat{\beta}_1 = \hat{\beta}_{LATE}$ from {E40}.

Two-stage least squares is a special case of the more general class of two-stage predictor substitution estimators (2SPS). The 2SPS method is carried out identically to the 2SLS method described above, with the exception that first- and second-stage regressions may be estimated by a nonlinear regression method (e.g., quantile regression, logistic, probit, negative binomial, etc). The first-stage regression of 2SPS methods is often estimated using linear OLS because, with a linear first-stage model, a correctly specified second-stage nonlinear model will be consistent even if the relationship between treatment and independent variables is not linear in truth.[12,19] Consistency of the second-stage when first-stage models are estimated using nonlinear regressions, on the other hand, depends on correctly specifying the nonlinear first stage.[12,19] The main disadvantage of nonlinear 2SPS is that the distributional and functional form assumptions imposed by a second-stage nonlinear estimator are very difficult to justify.[47] 2SPS has been suggested as a generally inconsistent estimator of the ATE when estimated using nonlinear regression methods.[19,68]

Two-Stage Residual Inclusion (2SRI)

The models and notation below follow from {E35}-{E36}, with the assumption that confounding factors (X_3) are unmeasured by the researcher. The first stage of the 2SRI procedure is carried out as in the first-stage of 2SPS methods: endogenous treatment (T_i) is regressed on one or more instrumental variables (X_{4i}). The residual term from this first stage equation is then calculated ($\hat{u}_i$) and included as a covariate in the second stage regression model with the observed endogenous treatment variable (T_i). For example,

$$T_i = M(\alpha_0 + \alpha_4 X_4) + u_i \qquad \{\text{E44}\}$$
$$Y_i = M(\beta_0 + \beta_1 T_i + \beta_u \widehat{u_\iota}) + \varepsilon_i. \qquad \{E45\}$$

$M(\cdot)$ is a functional form chosen by the researcher (identity, probit, logit, negative binomial, etc…). When $M(\cdot)$ is the identity function (i.e., the model is estimated by linear OLS), the coefficient estimate for β_1 can be interpreted as the absolute LATE estimate. For models where $M(\cdot)$ is a nonlinear function, estimates of β_1 are relative effects and absolute effect estimates must be calculated as average marginal effects of treatment (i.e., average derivatives).

Hausman first suggested the 2SRI approach using linear models as a method to test for endogeneity where significance of the residual term in the outcome model represents a test for the presence of unmeasured confounding.[45] The 2SRI method is a special case of the control function (CF) method. Control function methods assume that some function of the first-stage residuals can be used as an appropriate control for unmeasured confounders.[19,46,64] In the case of 2SRI, the "control function" is assumed to be the raw residual term from the first-stage regression (u_i in {E44}). In other words, the 2SRI method assumes that the residual term from the first-stage regression provides a valid estimate of all variation in unmeasured confounders, such that this can be used as a covariate in the second-stage outcome model to control for variation in the unmeasured confounders and produce unbiased parameter estimates that can then be used to calculate unbiased estimates of ATE. The 2SRI method has been suggested as a generally consistent nonlinear method for estimating the ATE in models with *inherently nonlinear dependent variables*.[19] As such, 2SRI is most commonly applied in empirical settings suggestive of nonlinear relationships between treatment and outcomes. In a completely linear specification of 2SRI (i.e., both first and second stage regression equations are linear, $M(\cdot)$ is the identity function), the 2SRI and 2SLS estimators are analogous and produce identical absolute effect estimates of the LATE.[19,45]

The main and minimal assumption of the 2SRI model is that the conditional mean of the outcome (Y_i) is of the form

$$E[Y_i|T_i, X_i] = M(\beta_1 T_i + \beta X_i). \qquad \{E46\}$$

$M(\cdot)$ is a known nonlinear function, X_i is a vector containing all observed and unobserved confounding factors, and β is a vector of parameters relating X_i to Y_i.[19] The 2SRI method works by assuming that the regression parameters in the first stage treatment choice model (i.e., α_0, α_4) can be consistently estimated such that the regression error term ($\hat{u}_i$) is an estimate of the unmeasured confounders (X_{3i}). Including $\hat{u}_i$ in the second stage outcome regression as a consistent estimate of the unmeasured X_{3i} is assumed to control for the unmeasured confounder. Nonlinear 2SRI estimators require additional distributional and functional form assumptions not imposed by linear IV estimators such as 2SLS, but may provide efficiency gains and potentially grant statistical significance where less efficient, but more robust, linear methods may not.[12,47,64]

Based on theoretical reasoning and simulation evidence, Lee (2012) recommends that both two-stage substitution methods (e.g., 2SLS) and nonlinear control function (e.g., 2SRI) methods should be applied when using two-stage IV methods.[47] If estimates from these two methods differ, Lee recommends adjusting the functional form of the control function method until estimates are similar to those obtained by the more robust 2SLS method. Estimates from the nonlinear 2SRI method could then be used for inference, as the standard errors will be smaller.[47] This logic follows closely from a statement by Joshua Angrist; “when simple and sophisticated estimation strategies do differ, I invariably prefer the simple!”[12] However, these ideas suggest that nonlinear 2SRI and 2SLS are estimating the same average treatment effect concepts—which may not necessarily be the case if nonlinear 2SRI generates estimates of ATE in models where ATE and LATE are not equal.

Methodological Research on Nonlinear 2SRI Estimators

A linear application of the 2SRI method was first suggested by Hausman (1978) as a test for endogeneity in instrumental variables applications.[45] Hausman notes that linear applications of the 2SRI method are analogous to 2SLS and shows that the significance of the residual in the second stage of 2SRI can be used as a large sample test for the presence of unmeasured confounding in the relationship between treatment and outcome.[45] Examples for the use of the 2SRI approach as a test for presence of endogeneity between treatment and outcome remain common in the literature.[24,26,33,69-76] Over the past 30 years, several extensions of this IV-based residual inclusion method have been discussed for specific nonlinear models and applications by Newey (1987), Blundell and Smith (1989), and Wooldridge (2002).[59,77,78] However, Terza et al. (2008) were the first to suggest 2SRI as a general nonlinear two-stage IV method for producing consistent estimates of ATE in models with inherently nonlinear dependent variables.[19]

Terza et al. (2007) brought attention to the nonlinear 2SRI method when they issued a cautionary note that linear 2SLS methods may produce inconsistent estimates of the ATE in models with nonlinearity between dependent and independent variables—that is, models with *inherently nonlinear dependent variables*.[18] The authors recommend that nonlinear IV methods, such as 2SRI, should be used in place of 2SLS in these models. The authors illustrated this point using simulated data in two common nonlinear regression contexts: the first a non-negative dependent variable model and the second a binary response model.

Outcomes for simulations in the non-negative dependent variable model were generated using a variant of the inverse Box-Cox model. Using this simulated data, the authors generated average marginal effect estimates using both the traditional linear 2SLS method and a nonlinear 2SRI method. The 2SRI model is based on the same inverse Box-Cox model that was used to generate the data. The true average marginal effect (AME) of the continuous endogenous regressor on the non-negative outcome was approximated and used as the benchmark for evaluation of estimator performance. This true AME—an average effect across the whole population—is compared to the coefficient estimate generated from 2SLS models and to an estimate of the AME derived from 2SRI estimates. Simulation results showed that estimates from 2SLS were significantly more biased than 2SRI estimates—49% versus 7% biased, respectively, at sample size of 1,000. Moreover, the bias of 2SRI estimates, but not 2SLS estimates, is attenuated to near zero as sample size increases.

For simulations in the binary outcome context, the authors generated data based on a bivariate probit (BVP) sampling design with a binary endogenous regressor. For this simulation, estimates from traditional 2SLS were compared with estimates from a BVP estimation method. The target parameter for estimation is the ATE. An estimate of true ATE generated using the true parameter values is used as a benchmark for comparison to 2SLS estimates and an estimate of the ATE calculated using the parameters estimates generated by the BVP model. Results showed that BVP is less biased then 2SLS for ATE—15% versus 16% at sample size of 1,000—and that bias of BVP estimates, but not 2SLS estimates, decreases as sample size increases.

While these results are intriguing, they are not particularly surprising. These simulations represent absolutely ideal settings with regards to the performance and use of the nonlinear 2SRI and BVP methods. For both simulations, the functional forms applied to the nonlinear IV estimators match exactly those of the data generation processes. Moreover, the methods for producing the absolute treatment effect estimates from the nonlinear IV estimators are exactly analogous to those producing the true parameter estimates used for comparison. The estimates generated using 2SLS, on the other hand, are not necessarily comparable to the population average treatment effects used as benchmarks in this study. 2SLS yields estimates of the LATE, not the ATE or other population average treatment effects and LATE estimates cannot be generalized to estimate population average treatment effect concepts without further assumptions[1,4,6,11,15,17,43,49]—assumptions not discussed by Terza et al. in their study. The apparent bias of 2SLS estimates in this study, relative to 2SRI estimates, may be observed because true LATE does not equal true ATE in this scenario; Terza et al. do not discuss estimates of the LATE in these scenarios.

In a related simulation study of the nonlinear 2SRI method, Terza et al. (2008) compared bias of ATE estimates generated by nonlinear 2SPS and 2SRI methods in two nonlinear models.[19] The results of these simulations show that 2SRI produces consistent ATE estimates but 2SPS is inconsistent for ATE across these models. Based on these results, Terza et al. propose the 2SRI method as a generally consistent IV-based estimator of ATE for models with inherently nonlinear dependent variables. The first, of two, simulations is a duration model with multinomial endogenous treatments. The authors specify the outcome as belonging to a Weibull distribution and estimate treatment effects using identically specified 2SPS and 2SRI methods. Results show that estimates generated by the 2SPS method are significantly more biased than those from 2SRI. Moreover, the bias of 2SRI estimates converges quickly to 0 as sample size increases while bias in 2SPS estimates does not. The second simulation models an ordered logit outcome with count-valued endogenous treatment. Once again, average absolute effect estimates generated by 2SRI are significantly less biased than those generated by 2SPS.

The simulations from Terza et al. (2008) clearly demonstrate the inconsistency of nonlinear 2SPS methods—and the comparative consistency of nonlinear 2SRI—for estimating the ATE in nonlinear models.[19] Once again, however, the authors do not report estimates of the LATE or discuss the relative ability of 2SPS and 2SRI estimators to yield consistent estimates of LATE in these models. Moreover, no simulation models consider the possibility of X_1 or X_3 factors, or represent scenarios of essential heterogeneity as defined in this research paper. The properties of 2SRI and 2SPS methods remain unknown in alternative nonlinear models where factors affect treatment effect heterogeneity but not outcomes directly, factors affect outcomes directly but not the effectiveness of treatment, or individuals idiosyncratic gains from treatment affect treatment choices directly in a manner consistent with essential heterogeneity. Under these alternative scenarios, the parameter estimates generated by nonlinear 2SRI methods may be only locally interpretable and use of these estimates to calculate average derivative effects may be inappropriate.

Methodological Research on Robustness of Nonlinear 2SRI Methods

Based on the simulation evidence detailed above, many have been quick to adopt nonlinear 2SRI methods for models with inherently nonlinear dependent variables. However, some researchers have expressed concern regarding several assumptions inherent to the 2SRI method. Recognizing that the 2SRI method may be, in some instances, a misapplication of the control function approach—which suggests that *some* function of the residuals, not necessarily the raw residual, itself, may be an appropriate control function—Garrido et al. (2012) compared estimates from linear 2SLS and nonlinear 2SRI estimators using various forms of the residual for the 2SRI model.[46] This research was performed using an empirical example for the problem of estimating the effect of palliative care consultations on costs in the Veterans Health Administration.

The authors attempt to generate and compare estimates of LATE and ATT across 2SRI models with alternatively specified control functions of the residuals to estimates of LATE and ATT generated by traditional 2SLS methods. Alternative control functions included response residuals, Pearson residuals, Anscombe residuals, and deviance residuals. The authors also examined using higher-degree polynomials of these residuals as control function.

Estimates of LATE and ATT from traditional 2SLS methods are reported by Garrido et al. as the parameter estimate generated by the 2SLS model. Though theory about the characteristics of treatment effect heterogeneity and choice in their data is not discussed, the authors assumption that ATT is identified using 2SLS implies that the authors must believe treatment effects are either homogenous across individuals or that heterogeneity is non-essential. Absolute effect estimates of ATT and LATE from nonlinear 2SRI models were calculated as the average marginal effect of treatment (binary) across treated individuals and individuals defined by the authors as being member to the subpopulation of "compliers", respectively. The authors define compliers as individuals observed to have either received a consultation when one was ordered or not receive a consultation when one was not ordered. Note that this methodology assumes no always-takers or never-takers exist. If this assumption is violated then estimates of LATE will include treatment effects from other subpopulations and may be inaccurate.

Garrido et al. showed that estimates from the various alternative nonlinear 2SRI models were only modestly sensitive to the type of control function specified.[46] However, ATT and LATE estimates generated by 2SLS models were found to be drastically different from estimates generated by nonlinear 2SRI models. Estimates from 2SLS were generally at least an order of magnitude greater in value than 2SRI estimates. Based on these findings, Garrido et al. recommended nonlinear 2SRI over 2SLS for IV-based estimation in models with skewed dependent variables. However, the methods used leave some issues to be considered.

While the substantial difference across estimates generated by 2SLS and 2SRI methods may be due to bias of 2SLS estimates, estimates from linear and nonlinear models are often similar in magnitude and it is possible that these observed differences are at least partially attributable to bias of the 2SRI estimates, which depend on considerably heavier assumptions. Both first- and second-stage regressions of the 2SRI approach were estimated by nonlinear least squares methods. If parametric assumptions are not strictly held, particularly for the first-stage treatment choice model, then average derivative estimates may not be accurate. The strategy of estimating average marginal treatment effects across specific subgroups of observations leans heavily on strong assumptions that parameter estimates generated using only the variation in treatment choice related to variation in the instrument can be used to obtain absolute treatment effect estimates across all patient subpopulations. In other words, assuming that the true treatment effect differs across all patient subpopulations defined by the covariates, this method assumes that heterogeneity across these groups follows strictly the parametric assumptions imposed on the model. If parametric assumptions are not held, then relative effect estimates may be only locally interpretable for the subpopulation of marginal patients and average derivative estimates may be inaccurate for true absolute effects. To my knowledge, no existing methodological research has examined the ability of nonlinear 2SRI methods to generate unbiased absolute LATE estimates from observational data. Without knowledge of the true treatment effect parameters (true ATT and LATE), it is not possible to conclude whether estimates from the nonlinear 2SRI models, which rely on much stronger assumptions than the linear 2SLS model, are in fact closer to the truth.

Also noting the need for research on the comparative operating characteristics of popular IV-based estimators, O'Malley et al. (2013) compared estimates from traditional 2SLS and nonlinear 2SRI methods using both an empirical application and simulations.[64]

The empirical work is done in the context of evaluating whether new antipsychotic drugs (a binary variable) may be associated with lower long-term mental health related costs (a non-negative, skewed dependent variable) for individuals with schizophrenia in Florida's Medicaid program. The authors' a priori theory is that newer drugs, though more expensive, may lower costs in the long run, and that observed correlations suggesting newer antipsychotics are associated with increasing costs may be attributable to confounding by unobserved factors related to mental health comorbidities, illness severity, and treatment preferences.

Estimation was carried out using OLS, 2SLS, nonlinear 2SPS, nonlinear 2SRI, and likelihood based methods. In all models, log-transformed costs were used as the dependent variable. Nonlinear 2SPS and 2SRI models were operationalized as follows: the first stage regression modeling treatment choice was estimated using a probit model, and the second-stage outcome model was estimated by OLS. Treatment effect estimates generated by 2SLS suggested a negative, though insignificant, effect of new antipsychotics on cost; a relationship consistent with the a priori theory of the authors. Nonlinear 2SPS and 2SRI approaches, on the other hand, each generated positive estimates for the effect of treatment on costs—though the effect was not statistically significant.

To shed light on which of these estimates should be preferred, O'Malley et al. simulated data based on the Florida Medicaid data to examine the operating characteristics of the various estimators when assumptions underlying the methods are violated.[64] The simulated outcome is modeled using a linear process while the binary endogenous treatment variable is the result of a discrete choice process based upon an underlying continuous latent variable. This scenario represents one of theorized homogeneous treatment effects (similar to {E22}-{E23}).

Simulation settings were altered to test the robustness of methods to violation of the exclusion restrictions (i.e., instrument exogeneity assumptions {A1}) and to increasing degrees of correlation between the residual terms of first- and second-stage models. Sensitivity to distributional assumptions was explored by drawing error terms alternatively from normal, t, and gamma distributions. Results showed that 2SLS performed well across all scenarios except when the exclusion restrictions were violated, which resulted in 2SLS estimates that were more biased than OLS estimates. When the distribution of the underlying error term was symmetric, 2SRI models generated more precise (i.e., lower root mean-squared error) estimates than 2SPS or 2SLS. When the distribution of the underlying error term was not symmetric, 2SLS was less biased and more precise than 2SRI.

Based upon these findings, the authors suggest that 2SLS is appropriate except in circumstances where exclusion restrictions are violated—although 2SRI and 2SPS also performed poorly in these scenarios. 2SLS was found to be more robust than 2SRI or 2SPS if there is evidence that the outcome residual term has a skewed distribution. On the other hand, 2SRI may be best when evidence supporting validity of the IV is strong but analysts are working with a small sample size and the study may be insufficiently powered under 2SLS. In general, this study emphasizes the robustness of 2SLS when there is a valid IV available and the ability of nonlinear IV estimators, particularly 2SRI estimators, to increase efficiency under ideal settings. However, these simulations are all performed under a scenario of homogeneous treatment effects, O'Malley et al. is do not consider alternative scenarios of treatment effect heterogeneity or sorting-on-the-gain, where results may differ.

In a thorough discussion of various IV-based estimators useful for limited dependent variables models, Lee (2012) formally discusses the consistency properties, method, and assumptions of alternative IV-based estimators.[47] Lee compares predictor substitution and control function estimators in simulations of both binary and continuous limited dependent variable settings. Across simulations, the control function approach has significantly smaller standard errors than 2SPS models. Lee explores the repercussions of applying a miss-specified functional form for the control function estimator and finds that misspecification causes substantial bias. In concluding remarks, Lee notes that, while control function approaches can be useful for reducing the error term variance—even in the presence of misspecification—it may be more biased than simpler substitution-based estimators when functional form assumptions are violated. In practice, Lee suggests the following approach: First, use both substitution and control function estimators; if estimates from the two methods differ, alter the functional form of the control function until estimates from the two models are comparable, at which point inferences can be made using the more efficient control function estimates. Comparison of linear 2SLS and nonlinear control function estimates is limited, however, if researchers are interested in absolute treatment effects. Unless it is possible to identify the subpopulation of marginal individuals in the data, it may not be possible to generate estimates of LATE from nonlinear 2SRI models. Assuming that nonlinear 2SRI models generate unbiased parameter estimates in the nonlinear outcome model, from which estimates of the ATE can be generated, 2SLS generates estimates of LATE that may not equal the ATE without acceptance of restrictive assumptions about the nature of treatment effect heterogeneity and the circumstances underlying treatment choice.[4,7,11,16,17,43,49]

Surprisingly, no studies of the 2SRI method examine the potential to identify alternative absolute average treatment effects under settings of essential heterogeneity where treatment choice is directly affected by treatment effect heterogeneity. Stuart, Doshi, and Terza (2009) state that, if the parametric assumptions underlying the nonlinear 2SRI model reflect the truth, then all sources of potential heterogeneity in the marginal treatment effects will be appropriately controlled for by the nonlinear function so that estimated average marginal treatment effect estimates are consistent, but that if any assumptions are violated then results may only be “locally” interpretable—that is, interpretable only for the marginal population.[14] In other words, if heterogeneity in treatment effectiveness is solely a product of a known nonlinear function of all covariates in the outcome model, and individuals idiosyncratic gain from treatment is not directly and positively related to treatment choice then absolute treatment effect estimates calculated from 2SRI models are consistent. If these conditions do not hold, however, then parameter estimates generated by nonlinear 2SRI methods may only be accurate for the marginal population. Moreover, if it is the case that parameter estimates are only accurate for the marginal population, then it may not be possible to calculate absolute treatment effect estimates—as an average derivative of the nonlinear function—unless the marginal individuals can be identified in the data. To my knowledge, the ability to generate absolute LATE estimates using 2SRI has not been shown. Stuart, Doshi, and Terza (2009) also note that the local interpretation of estimates generated by nonlinear average derivative estimators, such as nonlinear 2SRI, has not been studied and is not well understood.[14] In general, the limitations of nonlinear 2SRI methods are not well understood and rarely acknowledged by empirical researchers using observational data.

Empirical Work Employing Nonlinear 2SRI with Observational Data

Popularity of the 2SRI method as a nonlinear estimator for IV applications in health services research has grown considerably since 2008. Since Terza and colleagues (2008) published results of simulations showing positive properties of nonlinear 2SRI methods in Health Economics in 2008, their study has been cited over 230 times.[19] Over 60 of these studies were studies in health care and health services research that implement the nonlinear 2SRI methods in place of linear 2SLS or other IV methods.[5,21-41] Many of these studies cite Terza and colleagues work as showing that 2SLS is inconsistent in models with inherently nonlinear dependent variables to support their only using 2SRI in estimation.[21,23,28,30-32,34,39-42] This is surprising for two reasons, (1) Terza et al. (2008) never actually discuss the linear 2SLS estimator in this paper; and (2) Terza and colleagues focus entirely on the ability to generate accurate estimates of population average treatment effects—they do not discuss the LATE parameter, for which it is well established 2SLS yields consistent estimates.[19]

Despite assertions that 2SRI methods should be used to gain efficiency when estimates are comparable to those of more robust linear 2SLS methods, only a small proportion of studies implementing nonlinear 2SRI estimators actually perform this robustness check. It seems that many researchers believe the 2SRI method to be generally superior to 2SLS, despite the very limited evidence. Lack of methodological research examining the properties of 2SRI and making the potential limitations of the method clear, paired with growing adoption of the 2SRI method in empirical research has led to increased pressure placed on researchers to adopt nonlinear 2SRI methods. Because the added assumptions of the 2SRI method and the differences between 2SRI and 2SLS in terms of what average treatment effect concept is being estimated have not been made clear, the stronger assumptions required with nonlinear 2SRI and the potential consequences of those assumptions being violated are not well understood or appreciated. Of 51 published health-related studies that use 2SRI as their main method for estimation, only 6 report results from a comparable 2SLS model.[35-37,39,42,62] Across these 6 studies, 4 studies report using 2SLS as a robustness check and prefer 2SRI for efficiency gains.[35-37,42] None of these 6 studies, however, discuss the idea that nonlinear 2SRI and 2SLS methods may actually estimate different treatment effect concepts. Bonsang (2009), for example, reported statistically significant 2SLS and 2SRI estimates that were in opposite directions of effect and proceeded to disregard the 2SLS estimate as inconsistent, stating that the standard 2SLS approach is inconsistent when applied to nonlinear models.[39] However, at least these 6 studies report 2SLS estimates so that readers may draw their own inferences. For the 45 reviewed papers using nonlinear 2SRI that do not report estimates from 2SLS for comparison, readers cannot even be aware that there may be inconsistencies for inference across methods. The belief that nonlinear 2SRI is generally superior to 2SLS is exampled also by Gibson et al. (2010), whom state that their addition to the literature is to use a consistent method (i.e., nonlinear 2SRI) to evaluate a model that has in the past only been estimated by linear 2SLS, which they suggest is inconsistent.[34]

An additional consideration that is often overlooked in studies implementing nonlinear 2SRI methods is the local interpretation of estimates generated by IV regressions (i.e., LATE). Perhaps because Terza et al. discuss the 2SRI estimator only in the context of identifying population average treatment effects,[18,19] many empirical papers seem to take for granted that parameter estimates generated by these models are based on the marginal population whose treatment choice was sensitive to the instruments; unless these relative effects are truly constant across the population then 2SRI estimates may not be used to calculate the ATE. For example, Hadley and Reschovsky (2012) find that treatment effect estimates depend upon the instruments used in their nonlinear 2SRI model and attribute this to things such as over-identification or imprecision, but fail to consider that if parameter estimates reflective of the marginal patient are not representative of the entire population then estimates are only locally interpretable.[30] Under this scenario, average derivative effects calculated across the entire population may be inaccurate and alternative instruments may identify treatment effects for alternative "local" populations.[30] This is not a unique occurrence; of the 51 reviewed papers that use nonlinear 2SRI as their primary estimator, only 5 studies discuss LATE and how treatment effect estimates from their IV regressions may only be only locally interpretable—4 of these 5 studies also implement a 2SLS approach.[5,14,79-81]

CHAPTER 3
METHODOLOGY

Simulation

The 2SRI model suggested by Terza and colleagues (2007, 2008) is discussed in settings in which treatment effectiveness varies in a continuous manner across patients as a nonlinear function of all factors affecting outcomes directly, but treatment effectiveness does not directly affect treatment choice for decision-makers in a manner consistent with essential heterogeneity, as defined in Section 2.2.[18,19] Under these settings, Terza and colleagues show that nonlinear IV methods such as 2SRI are consistent for population average treatment effects (ATE). However, this scenario is just one of several possible scenarios that could be theorized in the real world. The properties and interpretability of estimates generated by nonlinear 2SRI methods in alternative settings of treatment effect heterogeneity and choice remain untested.

The simulation approach used to assess the settings under which 2SLS, nonlinear 2SPS, and nonlinear 2SRI methods identify the ATE and LATE is adapted from approaches used by Terza et al. (2007) and Brooks & Fang (2009).[16,18] The outcome (Y_i) is a binary variable representing whether an individual is observed to have been cured from a given condition. Patients can be cured (i.e., $Y_i = 1$) without treatment (T_i), but treatment can increase the probability of cure. The magnitude of the effect of treatment on outcome is termed *treatment effectiveness*. Treatment effectiveness is heterogeneous across simulated patients based on factors (X_{1i} and/or X_{2i}) that may be either observed or unobserved by the patient/provider dyad when making treatment decisions. If individual patient treatment effectiveness affects treatment choice for the patient/provider dyad directly, such that those patients with greater benefit from treatment have higher value associated with treatment and are more likely treated, all else equal, then heterogeneity is *essential*. If treatment effectiveness is unrelated to treatment choice then heterogeneity is *non-essential*. The equations determining outcome and treatment choice vary across simulated scenarios.

Estimation of ATE and LATE using alternative linear and nonlinear IV estimators will be considered across 10 unique scenarios. These scenarios will cover alternative theoretical circumstances underlying heterogeneity in treatment effectiveness and treatment choice. The first two scenarios (Scenario 1a and Scenario 1b) follow directly from the nonlinear binary model used by Terza et al. and discussed in previous sections, differing only through minor modifications to parameter values. The goal of these slight changes is to test the sensitivity of results to assumptions that (1) treatment effectiveness is negatively associated with treatment value, and (2) marginal patients exist in a fairly consistent proportion across the distribution of treatment effectiveness. For both Scenarios 1a and 1b, all factors that affect outcome are also related to the effectiveness of treatment on outcome through a nonlinear function but heterogeneity is not essential—it is quasi-essential. Unlike the model of Terza et al., however, parameters for the model of Scenario 1a will be chosen such that treatment effectiveness is positively correlated with treatment value—though there remains no direct relationship between treatment effectiveness and treatment choice. Consistent with the simulation modeled by Terza et al., Scenario 1a can be characterized as having a fairly consistent proportion of marginal individuals across the distribution of treatment effectiveness. However, this notion of having marginal individuals throughout the distribution of treatment effectiveness may be unique. It may be more likely in certain clinical settings that marginal individuals exist primarily in the "middle" of the treatment effectiveness distribution—while those with greater and lesser effectiveness may be always-treated or never-treated, respectively. Scenario 1b will explore the sensitivity of results to these conditions by choosing parameter values such that the proportion of marginal patients is not consistent across the distribution of treatment effectiveness.

The remaining 8 scenarios introduce concepts not considered by existing methodological work examining nonlinear 2SRI, nonlinear 2SPS, and 2SLS methods. Scenario 2 departs from Scenarios 1a and 1b by modeling heterogeneity as non-essential—this is done by assuming factors related to treatment effect heterogeneity are unrelated to treatment choice. Scenario 3 builds on Scenarios 1a, 1b, and 2 by assuming that treatment effectiveness has a direct positive effect on treatment choice, thereby modeling heterogeneity as essential. Scenarios 4, 5, and 6 will depart from Scenarios 1a, 1b, 2, and 3 by modeling factors that are related to the effectiveness of treatment, but unrelated to outcome independent of treatment (i.e., X_1 factors). Heterogeneity in Scenario 4 will be non-essential, while heterogeneity in Scenarios 5 and 6 will be quasi-essential and essential, respectively, in order to examine the properties of IV estimators under alternative settings related to sorting-on-the-gain in models with X_1 factors. Scenarios 7, 8, and 9 will be unique from previous scenarios by modeling factors that are related to outcomes, but unrelated to the effectiveness of treatment on outcomes (i.e., X_3 factors). To examine the properties of IV estimators under alternative settings related to treatment effect heterogeneity and how it is related to treatment choices in models with X_3 factors, heterogeneity will be non-essential in Scenario 7, quasi-essential in Scenario 8, and essential in Scenario 9. Table 4 provides a description of each scenario. Details of the modeling and estimation strategies for each scenario follow.

Table 4: Characteristics of Simulation Scenarios

	Types of Factors in Model	Characteristics of Heterogeneity and Sorting-on-the-Gain
Scenario 1	**X_2 factors exist**: *Related to outcome and effectiveness of treatment on outcome.* No X_1 or X_3 Factors.	**Quasi-Essential**: *Treatment effectiveness not consistently positively correlated with probability of treatment.*
Scenario 1b		**Quasi-Essential:** *Proportion of marginal individuals not consistent across distribution of treatment effectiveness.*
Scenario 2		**Non-Essential**: *Treatment effectiveness unrelated to probability of treatment.*
Scenario 3		**Essential**: *Treatment effectiveness directly and positively related to probability of treatment.*
Scenario 4	**X_1 factors exist**: *Related to effectiveness of treatment on outcome, not related to outcome independent of treatment.* No X_2 or X_3 factors.	**Non-Essential**: *Treatment effectiveness unrelated to probability of treatment.*
Scenario 5		**Quasi-Essential**: *Treatment effectiveness indirectly related to treatment choice through nonlinear function of factors affecting outcome and treatment choice.*
Scenario 6		**Essential**: *Treatment effectiveness directly and positively related to probability of treatment.*

Table 4 (continued)

Scenario 7	**X_3 factors exist**: *Related to outcome, not related to effectiveness of treatment on outcome.* No X_1 or X_2 factors.	**Non-Essential**: *Treatment effectiveness unrelated to probability of treatment.*
Scenario 8		**Quasi-Essential**: *Treatment effectiveness indirectly related to treatment choice through nonlinear function of factors affecting outcome and treatment choice.*
Scenario 9		**Essential**: *Treatment effectiveness directly and positively related to probability of treatment.*

Scenario 1a: Nonlinear, Quasi-Essential, X_2 Factors Only

Scenario 1a will be comparable to the binary outcome model used by Terza et al. (2007) and used as an example in a previous section.[18] Observed cure (Y_i) will be generated from an index function on an underlying continuous latent variable (Y_i^*). The model generating outcome will be

$$Y_i^* = \beta_1 T_i + \beta_{21} X_{21i} + \beta_{22} X_{22i} + \varepsilon_i \qquad \{E47\}$$

$$Y_i = 1(Y_i^* > 0) \equiv \begin{cases} 1 & if \;\; Y_i^* > 0 \\ 0 & if \;\; Y_i^* \leq 0 \end{cases}. \qquad \{E48\}$$

T_i is the observed binary treatment status of individual i. β_1, β_{21}, and β_{22} are parameters relating T_i, X_{21i}, and X_{22i} to latent outcome Y_i^*, respectively. ε_i is the value of a random disturbance term drawn from a standard normal distribution (i.e., $\varepsilon \sim N(0,1)$). In this scenario, we assume that X_{22i} is an unobserved confounding characteristic.

Because observed Y_i is generated from an index function on underlying Y_i^*, the absolute effect of treatment on outcome is heterogeneous across observations as a nonlinear function of all factors affecting outcome (i.e., X_{21i}, X_{22i}). Since the error term in this model is drawn from the standard normal distribution, the effect of treatment on outcome for individual i is

$$TE_i = \Phi(\beta_1 + \beta_{21}X_{21i} + \beta_{22}X_{22i}) - \Phi(\beta_{21}X_{21i} + \beta_{22}X_{22i}). \qquad \{E49\}$$

Φ denotes the standard normal distribution function. The value of {E49} varies with individuals X_{21i} and X_{22i} values; in other words, absolute treatment effects are heterogeneous with these factors.

The treatment choice model used by Terza et al. (2007) in simulations has observed treatment status (T_i) generated as the result of a cost-benefit decision by patients and providers where individuals are treated (T_i = 1) if the value of treatment, net costs, is positive.[18] The underlying treatment value is represented by T_i^*. The treatment choice model will be

$$T_i^* = \alpha_{21}X_{21i} + \alpha_{22}X_{22i} + \alpha_{41}X_{41i} + \alpha_{42}X_{42i} \qquad \{E50\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}. \qquad \{E51\}$$

X_{21i} and X_{22i} are defined as in the outcome equation {E47}. X_{41i} and X_{42i} are factors related to treatment choice but unrelated to outcomes. These factors are candidates for instrumental variables. $\alpha_{21}, \alpha_{22}, \alpha_{41}$, and α_{42} are parameters relating $X_{21i}, X_{22i}, X_{41i}$, and X_{42i} to treatment value T_i^*, respectively. Heterogeneity in Scenario 1a does not fit the description of either essential heterogeneity or non-essential heterogeneity. The effectiveness of treatment ({E49}) does not affect treatment choice directly, but treatment effectiveness does vary across patients in a nonlinear manner related to treatment value such that ATE, ATT, and ATU may not be equal. However, depending upon the coefficient values, the patients with higher treatment benefit may be less likely to be treated. This is the case for the scenario modeled by Terza et al.

(2007).[18] Parameter values and details regarding the distribution of variables are available in Table 5.

In each simulation, the absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{21i}, X_{41i}$, and $X_{42i}$are measured and X_{22i} is unmeasured. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated based on equation {E49}. The true Local Average Treatment Effect (LATE)—the average treatment effects for the marginal population—will be calculated as the average value of {E49} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as those observations satisfying

$$Marginal_i = 1 \quad iff \quad -0.25 < T_i^* < 0.25. \qquad \{E52\}$$

The definition for marginal is based upon the notion that an incremental change in the values of instrumental variables may change treatment choices for marginal patients. As such, the absolute value of bounds defining marginal patients are based upon the absolute values of the true parameter values on the instrumental variables in the treatment choice model.

Scenario 1b: Nonlinear, Quasi-Essential, X_2 Factors Only, Proportion Marginal not Consistent Across TE

Scenario 1b will depart from Scenario 1a by parameterizing the model such that individuals with very high or very low treatment effectiveness are less likely to be marginal, while those with treatment effectiveness near the mean are more likely to be marginal. The treatment choice and outcome models in Scenario 1b are exactly analogous to the treatment choice and outcome models of Scenario 1a ({E50}-{E51}; {E47}-{E48}), and treatment effectiveness is estimated by {E49}. The only manner in which Scenario 1a and 1b differ is the parameter values chosen. Details for parameter values and the distribution of variables are available in Table 5. Like Scenario 1a, treatment effectiveness is "quasi-essential"; while treatment effectiveness is correlated with treatment value, there is no direct relationship and the correlation may not be strictly positive such that $\partial T^*/\partial TE > 0$.

In each simulation, the effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{21i}, X_{41i}$, and $X_{42i}$are measured and X_{22i} is unmeasured. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E55}. The true Local Average Treatment Effect (LATE) will be calculated as the average value of {E55} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0—defined consistently with {E52}.

Scenario 2: Nonlinear, Non-Essential, X_2 Factors Only

Scenario 2 is distinct from Scenarios 1a and 1b by modeling a scenario of non-essential heterogeneity. Under non-essential heterogeneity, treatment decision makers are not using any information related to treatment effectiveness when making treatment decisions. Treatment effectiveness has absolutely no influence on treatment choice. There is no sorting of patients based upon heterogeneity and therefore ATE will be expected to be equal to LATE. The outcome model in Scenario 2 is exactly analogous to the outcome model of Scenario 1a ({E47}-{E48}), and treatment effectiveness is estimated by {E49}. However, X_{21i} and X_{22i}—both factors related to treatment effectiveness—are not present in the treatment choice model for Scenario 2. As such, X_{21i} and X_{22i} are not confounders. Details for parameter values and the distribution of variables in Scenario 2 are available in Table 5.

In each simulation, the effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{21i}, X_{41i}$, and $X_{42i}$are measured and X_{22i} is unmeasured. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E55}. The true Local Average Treatment

Effect (LATE) will be calculated as the average value of {E55} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0—defined consistently with {E52}.

Scenario 3: Nonlinear, Essential, X$_2$ Factors Only

Scenario 3 will depart from Scenarios 1a, 1b, and 2 by modeling heterogeneity as essential. In Scenario 3 patient/provider dyads are using information about individual's idiosyncratic gains from treatment—TE_i from {E49}—when making treatment decisions, such that patients who stand benefit more from treatment are more likely treated. The outcome model in Scenario 3 is exactly analogous to the outcome model of Scenario 1a ({E47}-{E48}), and treatment effectiveness is estimated by {E49}. However, in Scenario 3, individuals own idiosyncratic gain from treatment (TE_i) enters directly into the treatment choice model such that $\partial T^*/\partial TE > 0$ (i.e., $\alpha_{TE} > 0$) and heterogeneity is essential. The treatment choice model therefore becomes

$$T_i^* = \alpha_0 + \alpha_{TE} TE_i + \alpha_{41} X_{41i} + \alpha_{42} X_{42i} + u_i \qquad \{E53\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}. \qquad \{E54\}$$

Where

$$TE_i = \Phi(\beta_1 + \beta_{21} X_{21i} + \beta_{22} X_{22i}) - \Phi(\beta_{21} X_{21i} + \beta_{22} X_{22i}). \qquad \{E55\}$$

T_i, X_{41i}, and X_{42i} are defined as in {E41} of Scenario 1a. u_i is a random disturbance term, drawn from a standard normal ($N(0,1)$) distribution. α_0 is a constant. TE_i is a function of X_{21i} and X_{22i} ({E55}). Details for parameter values and the distribution of variables are available in Table 5.

In each simulation, the effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{21i}, X_{41i}$, and $X_{42i}$are measured and X_{22i} is unmeasured. TE_i will not be included in empirical models for estimation of treatment effects. X_{22i} does not directly affect T_i^* in {E53}, but affects T_i^* through TE_i and is therefore an unmeasured confounder. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E55}. The true Local Average Treatment Effect (LATE) will be calculated as the average value of {E55} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0—defined consistently with {E52}.

Table 5: Parameter and Variable Definitions (Scenarios 1a, 1b, 2, and 3)

	Scenario 1a	Scenario 1b	Scenario 2	Scenario 3
X_{21}	Uniform(-1.5, 1.5)			
X_{22}	Normal(0, 1)			
X_{41}	Uniform(-2, 2)			
X_{42}	Uniform(-2, 2)			
ε	Normal(0, 1)			
u	--	--	Normal(0, 1)	Normal(0, 1)
α_0	--	--	-0.25	-0.25
α_{21}	1			
α_{22}	1			
α_{41}	-0.25			
α_{42}	0.25			
α_{TE}	--	--	--	3
β_1	1.5	2.5	1.5	1.5
β_{21}	0.75	1	0.75	0.75
β_{22}	-0.75	1	-0.75	-0.75
TE_i	$\Phi(\beta_1 + \beta_{21}X_{21i} + \beta_{22}X_{22i}) - \Phi(\beta_{21}X_{21i} + \beta_{22}X_{22i})$			
Φ	Standard Normal Distribution Function			
True ATE	$\frac{1}{N}\sum_{i=1}^{N} TE_i$			

Scenario 4: Nonlinear, Non-Essential, X_1 and X_2 Factors

Scenario 4 will be distinct from Scenarios 1-3 by assuming that X_1 factors—factors that affect treatment effectiveness but are unrelated to outcomes, independent of treatment—exist. Like Scenario 2, heterogeneity in Scenario 4 is non-essential. Heterogeneity in treatment effectiveness in Scenario 4 has absolutely no influence on treatment choice and there is no sorting of patients based upon heterogeneity; ATE will be expected to be equal to LATE in Scenario 4. Building upon the outcome model of Scenarios 1-3, X_{1i} will be modeled as an interaction term modifying the effect of treatment (T_i) on outcome. X_{1i} does not affect outcomes independent of treatment. The outcome model for Scenario 4 will be

$$Y_i^* = X_{1i}\beta_1 T_i + \beta_2 X_{2i} + \varepsilon_i \qquad \{E56\}$$

$$Y_i = 1(Y_i^* > 0) \equiv \begin{cases} 1 & if \;\; Y_i^* > 0 \\ 0 & if \;\; Y_i^* \leq 0 \end{cases}. \qquad \{E57\}$$

The absolute effect of treatment on outcome is heterogeneous across observations as a function of X_{1i} and X_{2i}. Since the error term in this model is drawn from the standard normal distribution, the effect of treatment on outcome for individual i is

$$TE_i = \Phi(X_{1i}\beta_1 + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i}). \qquad \{E58\}$$

As in previous scenarios, observed treatment status (T_i) is the result of a cost-benefit decision by patients and providers where individuals are treated (T_i = 1) if the value of treatment (T_i^*) is positive. The treatment choice model will be

$$T_i^* = \alpha_{41} X_{41i} + \alpha_{42} X_{42i} + u_i \qquad \{E59\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}. \qquad \{E60\}$$

u_i in {E59} is a random disturbance term . Heterogeneity in Scenario 4 is non-essential—the only variation in treatment effectiveness comes from variation in X_{1i}, which does not affect treatment choice. There are no unmeasured confounders in

Scenario 4. Details for parameter values and variable definitions/distributions are given in Table 6.

For Scenario 4, the absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{2i}, X_{41i}$, and $X_{42i}$are measured and X_{1i} is unmeasured. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated by {E58}. The true Local Average Treatment Effect (LATE)—the average treatment effects for the marginal population—will be calculated as the average value of {E58} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as those observations satisfying

$$Marginal_i = 1 \quad iff \quad -0.25 < T_i^* < 0.25. \qquad \{E61\}$$

The definition for marginal is based upon the notion that an incremental change in the values of instrumental variables may change treatment choices for marginal patients. As such, the absolute value of bounds defining marginal patients are based upon the absolute values of the true parameter values on the instrumental variables in the treatment choice model.

Scenario 5: Nonlinear, Quasi-Essential, X_1 and X_2 Factors

Similar to Scenario 4, Scenario 5 will model X_1 factors. However, similar to the "Terza model" and Scenarios 1a and 1b, Scenario 5 will model quasi-essential heterogeneity of treatment effectiveness by including a factor in the treatment choice model that is nonlinearly related to treatment effect heterogeneity. Under quasi-essential heterogeneity treatment effectiveness is only indirectly related to treatment choice and physicians may not be sorting-on-the-gain, this is distinct from essential heterogeneity where treatment effectiveness is directly positively related to treatment value, such that $\partial T^*/\partial TE > 0$. The outcome model in Scenario 5 will be exactly analogous to {E56}-

{E57}, and treatment effectiveness will be estimated by {E58}. However, the treatment choice model in Scenario 5 will be

$$T_i^* = \alpha_0 + \alpha_2 X_{2i} + \alpha_{41} X_{41i} + \alpha_{42} X_{42i} + u_i \qquad \{E62\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \ \ T_i^* > 0 \\ 0 & if \ \ T_i^* \leq 0 \end{cases}.$$

T_i, X_{41i}, and X_{42i} are defined consistently with previous scenarios. X_{2i} is a factor directly related to treatment choice and outcome, and is also nonlinearly related to treatment effectiveness through {E58}. X_{2i} is driving the quasi-essential heterogeneity. α_0 is an intercept term specified with the goal of making true ATE distinct from true LATE. Details for parameter values and variable definitions/distributions are given in Table 6.

The absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{2i}, X_{41i}$, and $X_{42i}$are measured and X_{1i} is unmeasured. Because X_{1i} is not related to outcome directly—independent of treatment choice—it is not an unmeasured confounder. There are no unmeasured confounders in Scenario 5. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E58}. The true Local Average Treatment Effect (LATE) will be calculated as the average value of {E58} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as illustrated by {E61}.

Scenario 6: Nonlinear, Essential, X_1 and X_2 Factors

Similar to Scenarios 4 and 5, Scenario 6 will model X_1 factors. However, Scenario 6 will be distinct from Scenarios 4-5 by including TE_i—the treatment effect for individual i, which is a function of X_{1i} and X_{2i}—in the treatment choice model such that heterogeneity is essential (i.e., $\partial T^*/\partial TE > 0$). This represents decision-makers sorting patients into treatment based upon their expected gains from treatment—those with higher expected gains are more likely to be treated, all else equal. The outcome model in

Scenario 6 will be exactly analogous to {E56}-{E57}, and treatment effectiveness will be estimated by {E58}. The treatment choice model in Scenario 6 will be

$$T_i^* = \alpha_{TE}TE_i + \alpha_{41}X_{41i} + \alpha_{42}X_{42i} + u_i \qquad \{E63\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}.$$

T_i, X_{41i}, and X_{42i} are defined consistently with previous scenarios. α_{TE} and β_1 are defined to be positive. If α_{TE} and β_1 were of opposite sign then this scenario would not fit our definition of essential heterogeneity. Further details for parameter values and variable definitions/distributions are given in Table 6.

The absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{2i}, X_{41i}$, and $X_{42i}$are measured and X_{1i} is unmeasured. TE_i will not be measured or specified in any empirical model. Because X_{1i} is not related to outcome directly—independent of treatment choice—it is not an unmeasured confounder. There are no unmeasured confounders in Scenario 6. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E58}. The true Local Average Treatment Effect (LATE) will be calculated as the average value of {E58} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as illustrated by {E61}.

Table 6: Parameter and Variable Definitions (Scenarios 4, 5, and 6)

	Scenario 4	Scenario 5	Scenario 6
X_1	Uniform(-1, 3)		
X_2	Uniform(-2, 2)		
X_{41}	Uniform(-2, 2)		
X_{42}	Uniform(-2, 2)		
u	Normal(0, 1)		
ε	Normal(0,1)		
α_0	--	0.5	--
α_2	--	1	--
α_{TE}	--	--	2
α_{41}	-0.25		
α_{42}	0.25		
β_1	1.5		
β_2	-0.75		
TE_i	$\Phi(X_{1i}\beta_1 + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i})$		
Φ	Standard Normal Distribution Function		
True ATE	$\frac{1}{N}\sum_{i=1}^{N} TE_i$		

Scenario 7: Quasi-Linear, Non-Essential, X_2 and X_3 Factors Only

Scenario 7 will be distinct from all previous Scenarios by assuming that X_3 factors—factors that affect treatment choice and outcome directly, but are unrelated to the effect of treatment on outcomes—exist. Similar to Scenarios 2 and 4, heterogeneity in Scenario 7 will be non-essential. For Scenario 7, heterogeneity in treatment effectiveness will have absolutely no influence on treatment choice and ATE will be expected to be equal to LATE. The outcome model for Scenario 7 will be quasi-linear, outcomes are modeled using both a linear and nonlinear component. The base probability of outcome will first be estimated as the result of inputting the value of latent outcome (Y_i^*)—generated from a model analogous to {E38}—into the standard normal cumulative density function (Φ). This returns a probability, $P(Y_i|X_{2i})$. X_{3i} will then be linearly added to $P(Y_i|X_{2i})$ to produce a final probability of outcome, $P(Y_i = 1|X_{2i}, X_{3i})$. Observed dichotomous outcome Y_i will then be generated from probability $P(Y_i = 1|X_{2i}, X_{3i})$ using the rbinomial process in STATA. The outcome model for Scenario 7 will be

$$Y_i^* = \beta_1 T_i + \beta_2 X_{2i} + \varepsilon_i \qquad \{E64\}$$

$$P(Y_i = 1|X_{2i}) = \Phi(\beta_1 T_i + \beta_2 X_{2i}) \qquad \{E65\}$$

$$P(Y_i = 1|X_{2i}, X_{3i}) = P(Y_i = 1|X_{2i}) + \beta_3 X_{3i} \qquad \{E66\}$$

$$P_{adj}(Y_i = 1|X_{2i}, X_{3i}) = \frac{P(Y_i = 1|X_{2i}, X_{3i}) + \left|\min\left(P(Y_i = 1|X_{2i}, X_{3i})\right)\right|}{\left|\min\left(P(Y_i = 1|X_{2i}, X_{3i})\right)\right| + \max\left(P(Y_i = 1|X_{2i}, X_{3i})\right)}. \qquad \{E67\}$$

Φ denotes the standard normal distribution function. $P(Y_i = 1|X_{2i}, X_{3i})$ is a probability and must therefore be bound between 0 and 1. However, this may not be the case for {E66}. Therefore $P_{adj}(Y_i = 1|X_{2i}, X_{3i})$ will be calculated as the standardized/adjusted value, bound between 0 and 1. Definitions for model parameters and distributions from which variable values will be drawn are available in Table 7.

Because X_3 factors affect outcome linearly, outside of the nonlinear function ({E64}-{E65}), the absolute effect of treatment on outcome is not affected by X_3. The absolute effect of treatment on outcome is heterogeneous with X_2 and is calculated as

$$TE_i = \Phi(\beta_1 + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i}). \qquad \{E68\}$$

This true treatment effect must be adjusted to account for the standardization process ({E67}) performed on probability of outcome $P_{adj}(Y_i = 1|X_{2i}, X_{3i})$. The adjusted treatment effect is calculated as

$$TE\_adj_i = \frac{TE_i}{\left|\min\left(P(Y_i = 1|X_{2i}, X_{3i})\right)\right| + \max\left(P(Y_i = 1|X_{2i}, X_{3i})\right)}. \qquad \{E69\}$$

As in previous scenarios, observed treatment status (T_i) is generated as the result of a cost-benefit decision by patients and providers where individuals are treated (T_i = 1) if the value of treatment, net costs, is positive. The underlying treatment value is represented by T_i^*. The treatment choice model will be

$$T_i^* = \alpha_0 + \alpha_3 X_{3i} + \alpha_{41} X_{41i} + \alpha_{42} X_{42i} + u_i \qquad \{E70\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \quad T_i^* > 0 \\ 0 & if \quad T_i^* \leq 0 \end{cases}. \qquad \{E71\}$$

u_i is a random disturbance term containing all unmeasured variation in treatment choice. Because treatment effect heterogeneity does not affect treatment choice, heterogeneity in Scenario 7 is non-essential.

The absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{2i}, X_{41i}$, and $X_{42i}$are measured and X_{3i} is unmeasured. X_{3i} is an unmeasured confounder. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E69}. The true Local Average Treatment Effect (LATE)—the average treatment effects for the marginal population—will be calculated as the average value of {E69} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as those observations satisfying

$$Marginal_i = 1 \quad iff \quad -0.25 < T_i^* < 0.25. \qquad \{E72\}$$

The definition for marginal is based upon the notion that an incremental change in the values of instrumental variables may change treatment choices for marginal patients. As such, the absolute value of bounds defining marginal patients are based upon the absolute values of the true parameter values on the instrumental variables in the treatment choice model.

Scenario 8: Quasi-Linear, Quasi-Essential, X_2 and X_3 Factors Only

Similar to Scenarios 7, Scenario 8 models X_3 factors. However, similar to Scenario 5, Scenario 8 will model quasi-essential heterogeneity of treatment effectiveness by including a factor in the treatment choice model that is nonlinearly related to treatment effect heterogeneity. As described previously, under quasi-essential heterogeneity treatment effectiveness is indirectly related to treatment choice and physicians may not be sorting-on-the-gain. The outcome model in Scenario 8 will be exactly analogous to {E64}-{E67}, and treatment effectiveness will be estimated and adjusted by {E68} and {E69}, respectively. However, the treatment choice model in Scenario 8 will be

$$T_i^* = \alpha_0 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_{41} X_{41i} + \alpha_{42} X_{42i} + \alpha_{TE} TE_i + u_i \qquad \{E73\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}.$$

$T_i, X_{2i}, X_{3i}, X_{41i}, X_{42i}$, and u_i will be defined consistently with Scenario 7. X_{2i} is a factor directly related to treatment choice and outcome, and is also nonlinearly related to treatment effectiveness through {E68}. X_{2i} is driving the quasi-essential heterogeneity. α_0 is a constant intercept term. Definitions for model parameters and distributions from which variable values will be drawn are available in Table 7.

The absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{2i}, X_{41i}$, and $X_{42i}$are measured and X_{3i} is unmeasured. X_{3i} is an unmeasured confounder. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E69}. The true Local Average Treatment Effect (LATE) will be calculated as the average value of {E69} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as illustrated by {E72}.

Scenario 9: Quasi-Linear, Essential, X_2 and X_3 Factors Only

Similar to Scenarios 7-8, Scenario 9 will model X_3 factors. However, Scenario 9 will be distinct from Scenarios 7 and 8 by including a measure of the effectiveness of treatment (TE_i) directly in the treatment choice model such that heterogeneity is essential (i.e., $\partial T^*/\partial TE > 0$). This represents decision-makers sorting patients into treatment based upon their expected gains from treatment—those with higher expected gains are more likely to be treated, all else equal. The outcome model in Scenario 9 will be exactly analogous to {E64}-{E67}, and treatment effectiveness is estimated by {E69}. In Scenario 9, individuals own idiosyncratic gain from treatment calculated (TE_i from {E68}) will have a direct positive effect on T_i^* (i.e., $\alpha_{TE} > 0$). The treatment choice model in Scenario 9 will be

$$T_i^* = \alpha_3 X_{3i} + \alpha_{41} X_{41i} + \alpha_{42} X_{42i} + \alpha_{TE} TE_i + u_i \qquad \{E74\}$$

$$T_i = 1(T_i^* > 0) \equiv \begin{cases} 1 & if \;\; T_i^* > 0 \\ 0 & if \;\; T_i^* \leq 0 \end{cases}.$$

Where

$$TE_i = \Phi(\beta_1 + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i}). \qquad \{E75\}$$

Note that, because the adjusted value for treatment effectiveness cannot be calculated until treatment choice is determined, the unadjusted value for treatment effectiveness (TE_i) affects treatment choice. $T_i, X_{2i}, X_{3i}, X_{41i}, X_{42i}$, and u_i will be defined consistently with Scenario 7. Definitions for model parameters and distributions from which variable values will be drawn are available in Table 7.

The absolute effect of treatment on outcome will be estimated using empirical models where $Y_i, T_i, X_{2i}, X_{41i}$, and $X_{42i}$are measured and X_{3i} is unmeasured. X_{3i} is an unmeasured confounder. True values of the average absolute effect of treatment on outcome across the population (ATE) will be calculated using equation {E69}. The true Local Average Treatment Effect (LATE) will be calculated as the average value of {69} across the subpopulation of marginal observations whose treatment value (T_i^*) is proximal to 0. The marginal population will be defined as illustrated by {E72}.

Table 7: Parameter and Variable Definitions (Scenarios 7, 8, and 9)

	Scenario 7	Scenario 8	Scenario 9
X_2	Uniform(-1.5, 1.5)		
X_3	Uniform(-1, 1)		
X_{41}	Uniform(-2, 2)		
X_{42}	Uniform(-2, 2)		
u	Normal(0, 1)		
ε	Normal(0, 1)		
α_0	0.5	0.5	--
α_2	--	1	--
α_3	1		
α_{41}	-0.25		
α_{42}	0.25		
α_{TE}	--	--	1
β_1	1.5		
β_2	0.75		
β_3	-0.2		
TE_i	$\Phi(\beta_1 + \beta_2 X_{2i}) - \Phi(\beta_2 X_{2i})$		
TE_adj_i	$\frac{TE_i}{\lvert \min(P(Y_i = 1 \mid X_{2i}, X_{3i})) \rvert + \max(P(Y_i = 1 \mid X_{2i}, X_{3i}))}$		
Φ	Standard Normal Distribution Function		
True ATE	$\frac{1}{N}\sum_{i=1}^{N} TE\_adj_i$		

Estimation and Simulation Approach

For each scenario, models will be estimated by (1) nonlinear 2SRI, (2) nonlinear 2SPS, and (3) linear 2SLS methods. The first-stage of both the nonlinear 2SRI and nonlinear 2SPS methods will be estimated using linear OLS regressions.[12,19] With a first-stage OLS model, estimates generated by a correctly specified second-stage outcome models will be consistent, regardless of whether the first-stage is linear in truth.[12] The nonlinear second-stage model of the 2SRI and 2SPS methods will be estimated using a probit regression model—the appropriate nonlinear method for models of this form with error term (ε) drawn from a $N(0,1)$ distribution.[18,20,59] X_{41i} and X_{42i} will be specified as instrumental variables in each method.

For all scenarios, final true and estimated values of ATE and LATE will be generated using Monte Carlo simulations. Each simulation will be completed with 1,000 iterations of 20,000 observations per iteration. The absolute treatment effect estimates generated by 2SRI, 2SPS, and 2SLS will be averaged over these 1,000 simulations and compared with mean values of true ATE and LATE. Bias of estimates will be calculated as the absolute percentage difference between the mean values of the parameter estimate from each regression method and the mean value of the true parameter value. STATA 12.1 will be used for simulating all data and for all analyses.[82]

Empirical Application

The objective of this empirical example will be to discuss what inferences can be made from estimates generated by alternative linear and nonlinear instrumental variable (IV) methods when using observational data. IV methods will be used to generate estimates of the effects of renin-angiotensin system antagonists including angiotensin converting enzyme (ACE) and angiotensin receptor blockers (ARBs) on 1-year survival among Medicare beneficiaries with acute myocardial infarction (AMI). Use of ACE/ARBs for patients with AMI is an interesting clinical setting for discussion of the interpretability of estimates generated by alternative IV estimators because the benefits of these treatments are known to be heterogeneous across AMI patients, with patients at greater risk of future cardiovascular events receiving greatest benefit.[83,84] Moreover, many of the factors related to future risk of cardiovascular events, propensity to be treated using ACE/ARBs, and heterogeneity in the effects of ACE/ARBs on outcomes are likely to be unmeasured in observational claims data. Observed geographic variation in rates of ACE/ARB treatment suggests that prescribing of ACE/ARBs may be sensitive to the beliefs, preferences, or practice styles of physicians across areas.[85-87] This geographic variation forms the foundation for our instrumental variables analysis. A theoretical model for the factors likely to be affecting decision to use ACE/ARBs, the effectiveness of ACE/ARBs, and survival for patients with AMI is described below. The validity and interpretability of estimates generated by 2SLS, nonlinear 2SRI, and nonlinear 2SPS estimators will be discussed in light of this theoretical model and the results of simulations outlined in the previous section.

Theoretical Model

It has been shown that the effectiveness of ACE/ARBs is heterogeneous across AMI patients by underlying severity of patients AMI (e.g., left ventricular ejection fraction) and presence of certain comorbid conditions (e.g., diabetes). Moreover, research strongly suggests that providers are knowledgeable as to patients expected idiosyncratic gains from ACE/ARB treatment and that they may use this knowledge to accomplish some degree of selection into treatment based upon patients expected gains.[51-58] In other words, it is likely that sorting-on-the-gain (i.e., passive personalization, essential heterogeneity) is present in this clinical scenario such that those with greater expected benefit from ACE/ARBs are more likely to be treated, all else equal. These heterogeneous treatment benefits are highlight by Gustafsson et al. who, using data on patients with AMI from the Trandolapril Cardiac Evaluation (TRACE) study, found that diabetic patients treated using the ACE-inhibitor trandolapril had significantly greater reduction in risk of all-cause mortality than non-diabetic patients with the same treatment.[55] Niskanen et al. noted heterogeneity in the effectiveness of ACE-inhibitors even within a cohort of diabetic patients, where patients with impaired metabolic control were observed to benefit most from ACE-inhibitor based treatment.[88] The existence of heterogeneity in ACE/ARB benefits and associated selection into treatment is evidenced strongly by the 2005 clinician guidelines published by the American College of Cardiology/American Heart Association (ACC/AHA) Task Force on Practice Guidelines.[89] Guidelines suggest, for example, that ACE/ARBs may be of greater benefit to patients with diabetes, patients at risk of renal failure, patients at higher risk of future heart failure or cardiovascular death, and patients with lower left ventricular ejection fraction.[89] Guidelines recommend that ACE/ARBs be used for these patients, excepting cases of patients who are intolerant to therapies or who have conditions that may complicate their use—such as hyperkalemia or hypotension.[89] Following the notation used throughout previous sections, factors such as diabetes, left ventricular ejection fraction, and AMI severity may be considered examples of X_2 factors—they are related both to effectiveness of treatment and outcomes directly. While diabetes is likely measurable in observational data—diabetes will be a covariate in the empirical model for this research—left ventricular ejection fraction and other factors related to patients underlying health or AMI severity are not measurable using Medicare administrative data. Therefore, left ventricular ejection fraction and patients underlying health and AMI severity may be considered unmeasured confounders (unmeasured X_2 variables) in this empirical model.

It can also be argued that factors exist that are related to both patient outcomes and treatment choice directly, but not related to the effectiveness of treatment on outcome (i.e., X_3 factors). For example, socio-economic status may affect treatment choice for individuals as those with lower income are less able to afford treatment or less likely to choose treatment because the opportunity cost of treatment is greater than for those with higher income.[90] Income can also be expected to affect outcomes directly; low income has been shown to be associated with lower health, perhaps due to a less healthful diet and lower utilization of health care services.[91,92] However, there is no evidence to suggest that income may be associated with the effectiveness of ACE/ARBs for patients with AMI. Individual patient income is not observed in Medicare administrative claims and is unlikely to be available in many observational data sets. Income may therefore be considered an unmeasured X_3 variable in the empirical model.

The theoretical model for ACE/ARB use and survival across patients with AMI suggests that the effectiveness of ACE/ARBs for survival after AMI is heterogeneous with measured diabetes (measured X_2) and unmeasured left ventricular ejection fraction and AMI severity (unmeasured X_2s). It is also theorized that providers are likely aware of whether patients have these conditions when making treatment decisions and use information about patients expected gains from treatment when making treatment decisions such that those with greater expected benefit are more likely treated, all else equal. As such, this scenario is likely one of essential heterogeneity. Patient socioeconomic characteristics (e.g., income) are theorized to affect both choice of treatment and outcomes for patients, but are not expected to be related to the effectiveness of ACE/ARB treatment on survival—these factors are X_3's. This theoretical model most resembles simulation Scenario 9 (X_2 and X_3, essential heterogeneity).

There is substantial observed geographic variation in ACE/ARB use amongst AMI patients in Medicare,[87] suggesting that use of ACE/ARBs may be sensitive to the beliefs and preferences of treatment decision makers. This variation in provider's beliefs and preferences underlies the instrumental variables specified in this research. Variation in providers beliefs and preferences regarding ACE/ARB use that ultimately affects their proclivity to prescribe ACE/ARBs for patients with AMI are examples of X_4 factors—

factors related to treatment choice but not related to outcomes, independent of treatment choice. These factors are candidate instrumental variables. If heterogeneity is essential and providers are sorting patients with greatest expected benefits into treatment, then estimates of LATE generated by two-stage moment based estimators are expected to reflect the effectiveness of treatment for patients whose use of ACE/ARBs is unclear and determined by the proclivity of physicians in their local area—these patients are likely those in the middle of the distribution of effectiveness.[6,7,16,43,49] Consistency of estimates generated by IV methods leans on assumptions that variation in ACE/ARB practice styles across local areas are unrelated to unmeasured factors associated with patient's cardiovascular health or outcomes. Examples of factors that could threaten the validity of IV estimates include characteristics of the health care delivery system, social or environmental characteristics, area health behaviors, pollution, or regional disease prevention efforts. For example, if geographic variation in ACE/ARB use rates reflect the quality of physicians across these areas, then these key assumptions necessary to validate the IV approach may be violated.

Data and Analytical Sample

Data will include all Medicare claims files, beneficiary enrollment information, and Medicare Part D prescription drug event files for Medicare beneficiaries hospitalized with AMI in 2008, and with no previous AMI in the 365-days preceding the incident AMI. The Chronic Care Warehouse definition of AMI as an inpatient stay with an International Classification of Diseases, 9th Revision (ICD-9) code 410.xx (excluding 410.x2) in the first or second diagnosis position will be used to identify AMI events.[93] The period of hospitalization for AMI will be defined to begin on the date of initial hospitalization for AMI and include all institutional stays (acute, long-term care hospital, inpatient stay, or nursing facilities) with overlapping admission and discharge dates. The analytical sample will be limited to individuals with continuous Medicare part A and B fee-for-service enrollment for at least 1-year prior to the AMI hospitalization and at least 1-year after discharge from the hospital for AMI (or until death). The analytical sample will be further restricted to patients discharged alive with continuous Medicare Part D enrollment for at least 6-months prior to AMI-hospitalization and at least 1-year after discharge from their stay for AMI, or until death. To assure that all Part D events are observed during the 30-day post-index treatment exposure period, individuals who used any hospice or skilled nursing facility care, were readmitted to a hospital for inpatient care, or died within 30-days after discharge for AMI will be excluded. Finally, because driving times are used in defining both control variables and IVs, individuals with invalid or missing zip codes or with valid zip codes outside of the continental United States will be excluded.

Measures

The primary dependent variable in this analysis will be 1-year survival, defined using a binary variable equaling 1 if the patient survived 365-days after AMI discharge and 0 otherwise. The independent variable of interest will be whether individuals used an ACE/ARB for treatment after AMI. ACE/ARB use after AMI will be defined using a binary variable set to 1 if part D event files indicate that the patient filled a prescription for an ACE/ARB within 30-days after discharge for AMI, 0 otherwise. Covariates will include comorbid diagnoses and procedures observed in the 365-days prior to the index AMI event, medications used during the 180-days prior to AMI, characteristics of individuals AMI, comorbid diagnoses and procedures observed during the stay for AMI, other relevant medications filled after discharge (i.e., statins, beta-blockers), individual demographic characteristics (i.e., age, race), part D premium levels and phase at diagnosis, whether individuals were dual-eligible for Medicaid in the AMI month, and patient zip code socio-economic characteristics (i.e., per capita income, poverty rate, education level, English speaking percentage, racial mix). Full information on control variable to be included in the analytical models is available in Table A-1 of the Appendix.

Instrumental variables will be Area Treatment Rates (ATRs) of ACE/ARBs in the local area around patient's zip code. These measures represent the proclivity of physicians to use ACE/ARBs. The Driving Area for Clinical Care (DACC) method will be used to define the local areas around Medicare AMI patient's residence zip codes and create ATRs.[94,95] This method creates local areas around each zip code by consecutively adding the next nearest zip code—distance measured using driving times between zip code centroids—until a defined threshold number of patients within the local area is reached. Using this method, the geographic size of local areas reflect differences in health care access across urban and rural areas, as well as the notion that patients in rural areas are expected to travel greater distances for health care.[94] Following the findings of previous work exploring the validity of ATRs as instruments in this clinical setting, instrumental variables will be defined in this study using a threshold of 150 patients to create local areas.[87] Area treatment ratios (ATRs) will be calculated for each zip code-defined local area as the ratio of the number of patients observed to have received an ACE/ARB after AMI over the sum of the estimated probabilities of these patients receiving an ACE/ARB after AMI. Predicted ACE/ARB probabilities will be generated using a multivariate logistic model of ACE/ARB receipt across all patients in the analytical sample. Because predicted probabilities for ACE/ARB receipt will be generated using the full analytical sample, these probabilities reflect a mix of the preferences and styles of all physicians in the sample, while the number observed to have received an ACE/ARB in a local area will be sensitive to the physician practice styles in that area alone. The sensitivity of results to alternative thresholds will be tested using thresholds of 100 and 200 patients.

Analytical Approach

Modeling approaches will include the linear 2SLS, nonlinear 2SPS, and nonlinear 2SRI instrumental variable estimators. Linear 2SLS yields consistent estimates of LATE that are robust to underlying distributional assumptions regarding the model error terms.[8,12] In the first stage of 2SLS, a linear probability model (LPM) of ACE/ARB choice (binary) will be estimated for all patients in the analytical sample. Independent variables will include all covariates and instrumental variables described above and detailed in Table A-1 of the Appendix. The instrument will be specified using 4 binary variables indicating quintiles of individuals ATR values, across all individuals. In the second stage of 2SLS, an LPM model predicting 1-year survival (binary) as a function of the predicted value of ACE/ARB from the first stage treatment choice model and all covariates from the first stage model (excluding instrumental variables) will be estimated. Using the 2SLS approach, the parameter estimate on the second stage predicted ACE/ARB choice probability is the LATE estimate. This estimate of LATE is not thought to be generalizable to the ATE because it is theorized that treatment effect heterogeneity exists and is essential—that is, treatment decision makers are using information about heterogeneity in treatment effectiveness that exists with patient characteristics including diabetes, left ventricular ejection fraction, and other factors related to patients AMI severity.

The first stage of the nonlinear 2SPS model will be estimated using a linear probability model. As long as the nonlinear second stage outcome model is specified correctly, estimating the first stage treatment choice model as an LPM is robust to underlying error distribution assumptions.[12,19] The predicted ACE/ARB probability then replaces observed ACE/ARB choice in the second stage outcome model of 2SPS, which will be estimated using a probit specification. Similarly, the first stage of the nonlinear 2SRI approach will be estimated using a LPM. The residual from this first stage model is then included as a covariate in the second stage outcome model of the 2SRI approach, along with observed ACE/ARB treatment status and all other covariates, excepting instruments, which will be estimated using a probit model. The parameter estimates for effect of ACE/ARB on 1-year survival from probit regressions of 2SRI and 2SPS models are not absolute treatment effect estimates. Absolute effect estimates from these nonlinear models will be estimated as the average absolute marginal effect of treatment across individuals in the sample. The average absolute treatment effect (ATE) for nonlinear models will be estimated as

$$\frac{1}{N}\sum_{i=1}^{N}\Phi\left(\hat{\beta}_1+\hat{\beta}_2X_{2i}+\cdots+\hat{\beta}_kX_{ki}\right)-\Phi\left(\hat{\beta}_2X_{2i}+\cdots+\hat{\beta}_kX_{ki}\right).$$

Φ represents the standard normal distribution function, which follows from the probit specification. $\hat{\beta}_1$ is the parameter estimate on ACE/ARB in each nonlinear model. $\hat{\beta}_2,\ldots,\hat{\beta}_k$ are estimated parameters on covariates $X_{2i},\ldots,X_{ki}$ for individual i. To test the sensitivity of results to alternative nonlinear specifications, nonlinear 2SRI and 2SPS models will also be estimated using a first-stage probit and second-stage probit specification. If parametric assumptions do not hold in any of these nonlinear models, then parameter estimates may be either inconsistent or only "locally interpretable" for the marginal population—in which case ATE estimates may be inaccurate.

Estimates of LATE from nonlinear 2SPS and 2SRI models will be estimated using an approach that attempts to identify the approximate subgroup of marginal patients or "compliers". Marginal patients are theoretically those with the greatest degree of uncertainty in treatment choice. Therefore, marginal patients will be defined as those patients whose predicted probability of receiving ACE/ARB treatment—the predicted value of the dependent variable from ACE/ARB treatment choice models—is around 0.5. These are the patients whose measured characteristics suggest treatment choice is most uncertain. Marginal patients in this study will be defined as those patients whose predicted probability of ACE/ARB treatment is between 0.45 and 0.55. The sensitivity of results will be tested by defining marginal patients with a larger and smaller range around 0.5. As a further robustness check, an additional estimate of LATE will be calculated as the absolute effect of treatment at the mean of all covariates. This alternative estimate of LATE is illustrated by

$$\Phi\left(\hat{\beta}_1 + \hat{\beta}_2\overline{X}_2 + \cdots + \hat{\beta}_k\overline{X}_k\right) - \Phi\left(\hat{\beta}_2\overline{X}_2 + \cdots + \hat{\beta}_k\overline{X}_k\right).$$

Where Φ represents the standard normal distribution function; $\hat{\beta}_1$ is the parameter estimate on ACE/ARB; and $\hat{\beta}_2, \ldots, \hat{\beta}_k$ are estimated parameters on mean values of covariates $\overline{X}_2, \ldots, \overline{X}_k$ across all individuals in the sample.

Because this example will be using observational claims data, it will not be possible to compare estimated ATE and LATE values to true values of ATE or LATE. Therefore, the potential interpretability of estimates generated by all models will be discussed in the context of the theoretical model developed above—where heterogeneity is theorized to be essential.

CHAPTER 4

RESULTS

Simulation Results

Results for each of the 10 alternative simulation scenarios follow. For each scenario, the primary statistic of interest is the extent to which estimates generated by 2SLS, 2SPS, and 2SRI estimators were biased for estimates of true ATE and true LATE. Five regression estimates are generated for each scenario: (1) the parameter estimate generated by linear 2SLS; (2) the average absolute effect of treatment across all observations using parameter estimates generated by nonlinear 2SPS (2SPS_ATE); (3) the average absolute effect of treatment across only marginal observations using parameter estimates generated by nonlinear 2SPS (2SPS_LATE); (4) the average absolute effect of treatment across all observations using parameter estimates generated by nonlinear 2SRI (2SRI_ATE); and (5) the average absolute effect of treatment across only marginal observations using parameter estimates generated by nonlinear 2SRI (2SRI_LATE). Each of these estimates is compared to an estimate of the true absolute effect of treatment across all observations (ATE), and an estimate of the true absolute effect of treatment across the marginal observations (LATE). The percent to which each estimate is biased is calculated as the difference between the mean values of the estimate and the truth, divided by the truth.

Scenario 1a

Scenario 1a is similar to the binary outcome model used by Terza et al. (2007) and examined as an example in an earlier section.[18] However, parameters values for Scenario 1a were chosen such that treatment effectiveness is not negatively correlated with treatment value, though there remains no direct relationship between treatment effectiveness and treatment choice—treatment effectiveness is “quasi-essential”. As with Terza’s original model, Scenario 1a is characterized by having an approximately uniform proportion of marginal individuals across the distribution of treatment effectiveness. Summary statistics for percentage of marginal patients, mean treatment value (T*), and other factors across deciles of treatment effectiveness are provided in Table 8. These statistics are based on a sample of 200,000 observations, generated using the models relevant to Scenario 1a described in detail in the methods section.

Table 8: Summary Statistics by Treatment Effectiveness Decile (Scenario 1a)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal*	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	.05	20	-.43	38	96
2	20,000	.14	14	-.0034	56	90
3	20,000	.22	13	.11	58	82
4	20,000	.3	13	.13	57	75
5	20,000	.37	14	.1	54	68
6	20,000	.42	13	.069	51	61
7	20,000	.47	13	.048	49	57
8	20,000	.51	13	.015	47	52
9	20,000	.53	12	-.017	45	48
10	20,000	.54	13	-.025	45	47
Total	200,000	.36	14	.00044	50	67

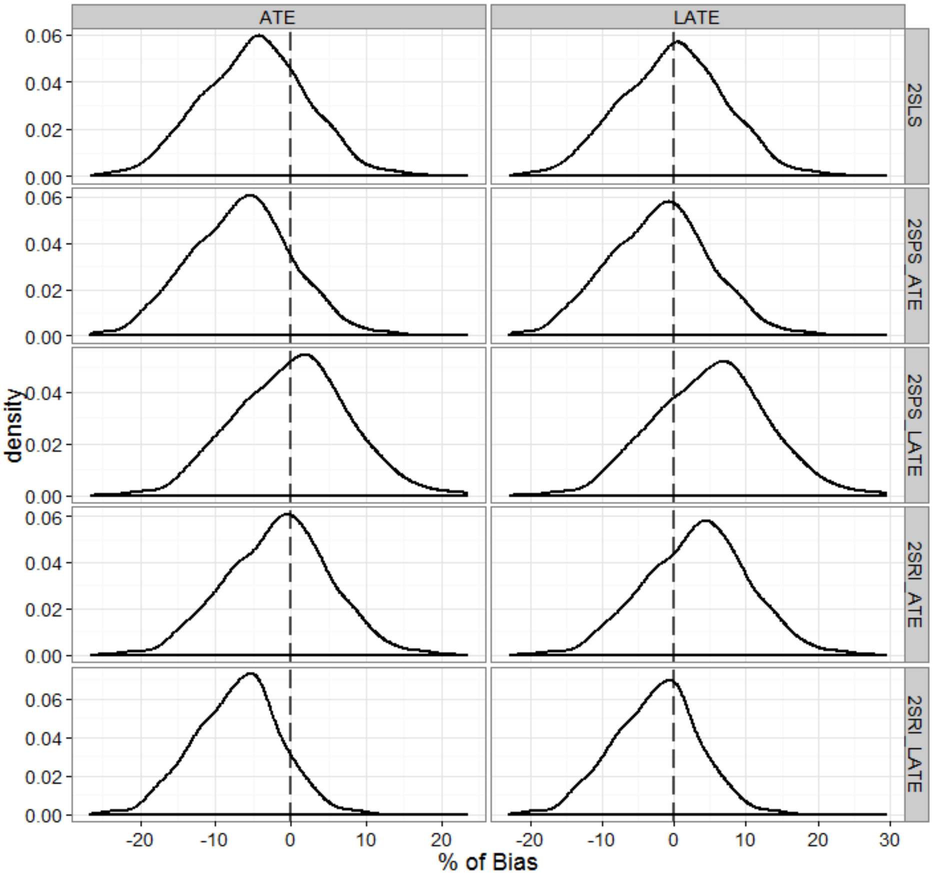
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 1a are provided in Table 9. True ATE was estimated to be 0.36. True LATE was estimated to be 0.34. Results show that nonlinear 2SRI generated less biased estimates of ATE than nonlinear 2SPS or 2SLS. This results is consistent with the suggestions of Terza, Basu, and Rathouz (2008) and Terza, Bradford, and Dismuke (2007), as well as the simulation results from Terza, Bradford, and Dismuke (2007).[18,19] While Terza et al. (2007) state that 2SLS is inconsistent in nonlinear models,[18] 2SLS generated consistent estimates of LATE that are less biased than nonlinear 2SPS or nonlinear 2SRI—even while it is assumed that the population of marginal individuals is identifiable in the data. Note that, while the LATE estimate generated by nonlinear 2SPS is the least biased estimate of ATE, this is an artifact of chance and should not be interpreted as a general result. Density plots showing the bias of estimates for the alternative estimators in Scenario 1a are given in Figure 4: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 9: Average % Bias of Estimate, by Estimator (Scenario 1a)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	-4.79	-0.12
2SPS_ATE	-6.62	-2.04
2SPS_LATE	0.17	5.08
2SRI_ATE	-1.69	3.13
2SRI_LATE	-7.12	-2.56

Figure 4: Density Plot of % Bias of Estimates, by Estimator (Scenario 1a). X_2 Factors, Unmeasured X_2, Quasi-Essential Heterogeneity, Marginal population uniformly distributed across TE.



Scenario 1b

Scenario 1b follows from Scenario 1a and the binary outcome model used by Terza et al. (2007).[18] However, Scenario 1b departs from these models by assuming that marginal individuals are those near the mean of treatment effectiveness, while those with higher and lower treatment effectiveness are less likely to be marginal. This is distinct from Scenario 1a and the Terza model, which assume that marginal individuals are fairly uniformly distributed across the distribution of treatment effectiveness. Like Scenario 1a, treatment effectiveness in Scenario 1b is “quasi-essential”. The only differences between Scenarios 1a and 1b are in the coefficient values of the models. Summary statistics for percentage of marginal patients, mean treatment value (T*), and other factors across deciles of treatment effectiveness are provided in Table 10. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 1b described in detail in the methods section.

Table 10: Summary Statistics by Treatment Effectiveness Decile (Scenario 1b)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	.017	0	2.3	100	100
2	20,000	.081	.025	1.3	99	99
3	20,000	.18	5.2	.78	96	96
4	20,000	.29	20	.31	84	90
5	20,000	.41	33	-.1	62	77
6	20,000	.52	35	-.46	37	59
7	20,000	.62	25	-.76	17	39
8	20,000	.7	14	-.99	5.5	23
9	20,000	.76	4.8	-1.2	.76	15
10	20,000	.78	.57	-1.2	0	11
Total	200,000	.43	14	.00044	50	61

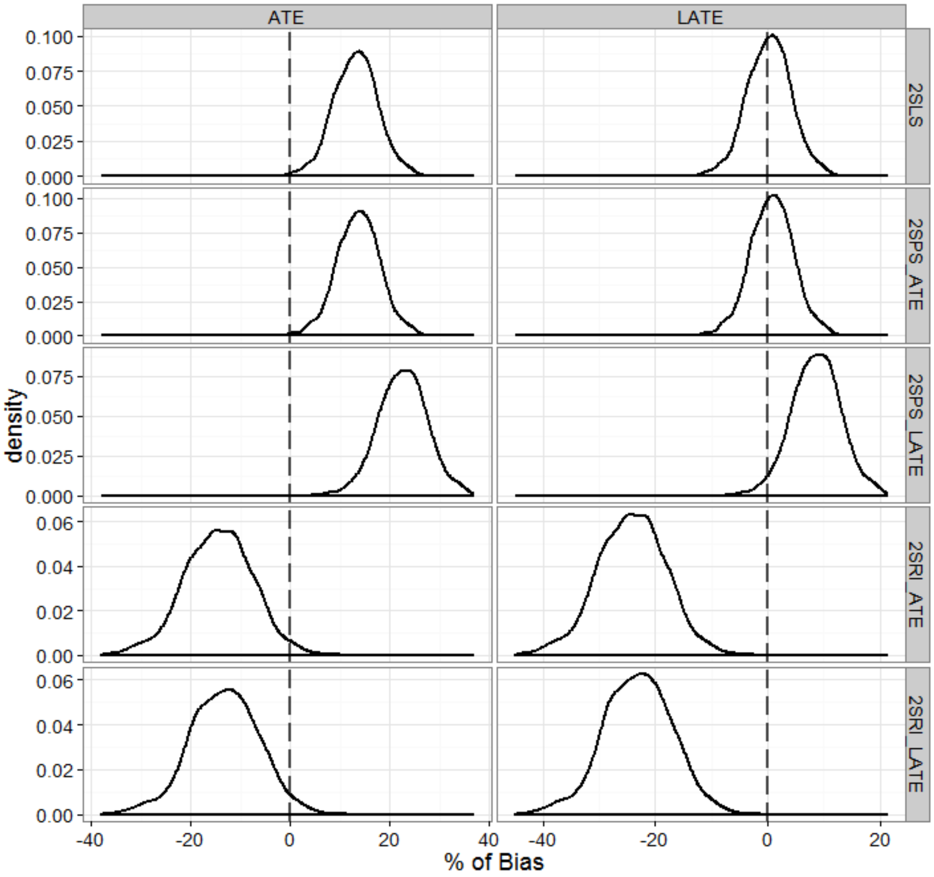
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 1b are provided in Table 11. True ATE was estimated to be 0.43. True LATE was estimated to be 0.49. For Scenario 1b, no estimator generates an unbiased or consistent estimate of the ATE. IV methods use only the variation in treatment choice that is driven by the instruments—variation in treatment choice for the marginal individuals. Because the marginal population is not distributed evenly throughout the distribution of treatment effectiveness, IV methods do not appear able to yield consistent estimates of the ATE. All estimators are approximately 14% biased for the true ATE, excepting the 2SPS_LATE estimate, which is more biased (22.6%). This result is not consistent with the suggestions of Terza, Basu, and Rathouz (2008) that nonlinear 2SRI methods can yield unbiased and consistent estimates of the ATE across models with inherently nonlinear dependent variables. Moreover, only two estimators generate unbiased and consistent estimates of the LATE; 2SLS was 0.29% biased for LATE and 2SPS_ATE was 0.77% biased for LATE. Both nonlinear 2SRI estimates had absolute bias of approximately 24% for LATE. For Scenario 1b, LATE is the only identifiable average absolute treatment effect parameter and 2SLS produced the least biased estimate. Density plots showing the bias of estimates for the alternative estimators in Scenario 1b are given in Figure 5: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 11: Average % Bias of Estimate, by Estimator (Scenario 1b)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	13.18	0.29
2SPS_ATE	13.72	0.77
2SPS_LATE	22.64	8.68
2SRI_ATE	-14.46	-24.19
2SRI_LATE	-13.22	-23.10

Figure 5: Density Plot of % Bias of Estimates, by Estimator (Scenario 1b). X_2 Factors, Unmeasured X_2, Quasi-Essential Heterogeneity, Marginal population not uniformly distributed across TE.



Scenario 2

The outcome model for Scenario 2 is identical to that of Scenarios 1a and 1b. However, treatment effect heterogeneity is non-essential in Scenario 2—there is no correlation between treatment effectiveness and treatment value (i.e., treatment effectiveness is unrelated to probability of treatment). Non-essential heterogeneity was accomplished by omitting X_2 factors from the treatment choice model. This is evident in the summary statistics for percentage of marginal patients, mean treatment value (T*), and other factors across deciles of treatment effectiveness, provided in Table 12. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 2 described in the methods section.

Table 12: Summary Statistics by Treatment Effectiveness Decile (Scenario 2)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	0.05	18	-0.23	41	95
2	20,000	0.14	18	-0.26	41	87
3	20,000	0.22	18	-0.24	41	79
4	20,000	0.30	18	-0.24	41	71
5	20,000	0.36	18	-0.24	42	64
6	20,000	0.42	17	-0.25	41	57
7	20,000	0.47	18	-0.25	41	52
8	20,000	0.51	18	-0.26	40	49
9	20,000	0.53	18	-0.26	41	46
10	20,000	0.54	18	-0.25	41	45
Total	200,000	0.35	18	-0.25	41	65

* Marginal defined as -.25 < T* < .25.

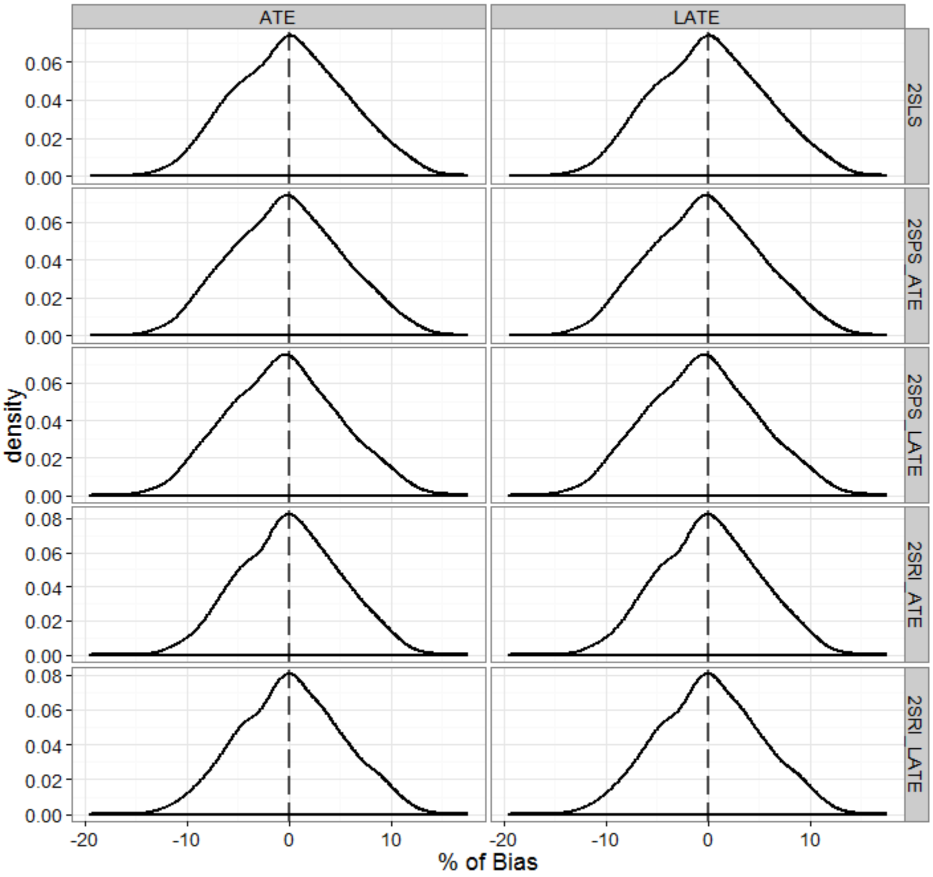
Simulation results for Scenario 2 are provided in Table 13. True ATE and true LATE were both estimated to be 0.36. Because treatment effect heterogeneity is non-essential in Scenario 2, true LATE is approximately equal to true ATE. All estimators generated unbiased and consistent estimates of ATE and LATE. While there is not meaningful difference in the percent bias of estimates across estimators, 2SRI_LATE estimates were least biased for LATE, on average. This greater precision of the probit 2SRI model is likely explained by their being no unmeasured confounders and probit being the correct functional form for the outcome model, which is probit in truth.

Density plots showing the bias of estimates across the alternative estimators in Scenario 2 are given in Figure 6: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 13: Average % Bias of Estimate, by Estimator (Scenario 2)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	0.24	0.21
2SPS_ATE	-0.11	-0.14
2SPS_LATE	-0.49	-0.52
2SRI_ATE	0.16	0.13
2SRI_LATE	0.18	0.15

Figure 6: Density Plot of % Bias of Estimates, by Estimator (Scenario 2). X_2 Factors, Unmeasured X_2, Non-Essential Heterogeneity.



Scenario 3

Scenario 3 introduces the idea of essential heterogeneity: treatment effectiveness has a direct and positive effect on treatment choice, such that $\partial T^{*}/\partial TE > 0$. This is distinct from quasi-essential heterogeneity demonstrated by Scenarios 1a and 1b, which assume treatment effectiveness is only indirectly and nonlinearly associated with treatment choice. Essential heterogeneity in Scenario 3 is evidenced by the summary statistics of factors, including treatment value (T*), across deciles of treatment effectiveness provided in Table 14. It is clear from this table that treatment value is increasing with treatment effectiveness. Moreover, treatment effectiveness is a factor in the treatment choice model that affects treatment value directly and positively. Statistics presented in Table 14 are based on a sample of 200,000 observations, generated following the models relevant to Scenario 3 described in detail in the methods section.

Table 14: Summary Statistics by Treatment Effectiveness Decile (Scenario 3)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	.05	18	-.11	46	96
2	20,000	.14	18	.16	56	89
3	20,000	.22	17	.41	65	84
4	20,000	.3	15	.64	72	79
5	20,000	.37	13	.85	78	77
6	20,000	.42	12	1	83	75
7	20,000	.47	10	1.2	86	74
8	20,000	.51	9.6	1.3	88	73
9	20,000	.53	8.5	1.3	89	72
10	20,000	.54	8.3	1.4	90	72
Total	200,000	.36	13	.81	75	79

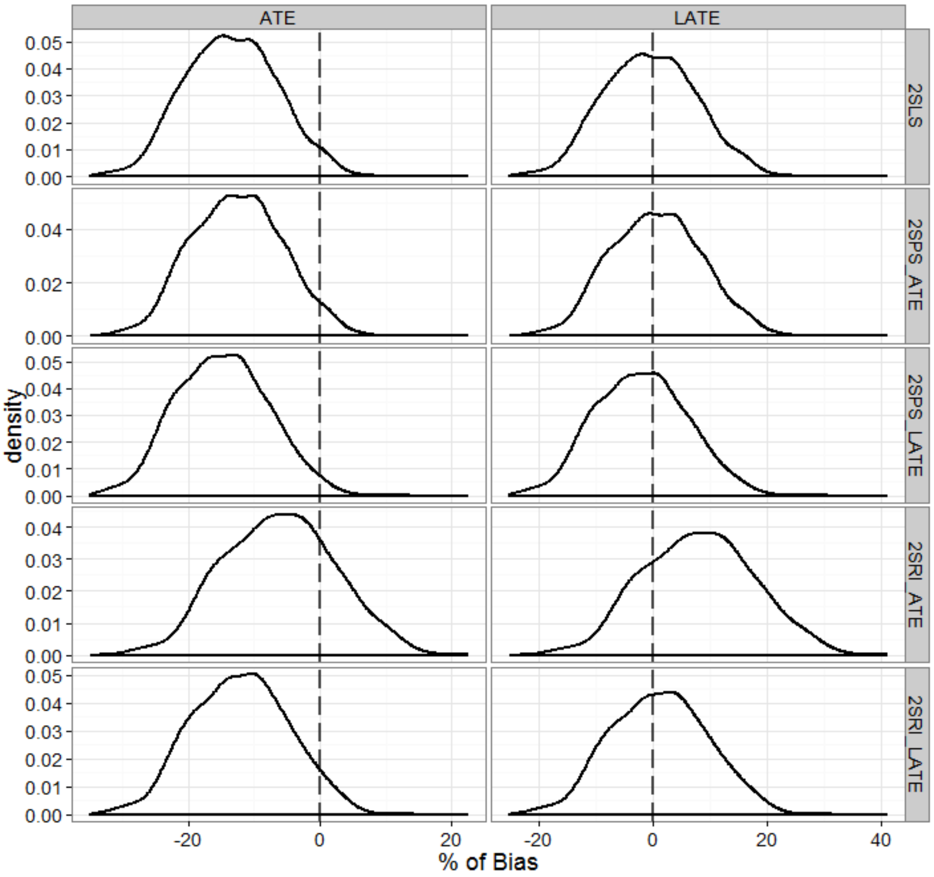
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 3 are provided in Table 15. True ATE was estimated to be 0.36. True LATE was estimated to be 0.31. In Scenario 3 treatment decision makers are "sorting-on-the-gain" and only variation in treatment choice for the marginal individuals is used in estimating treatment effects. Because the marginal population is unique in terms of treatment effectiveness and does not accurately reflect the distribution of treatment effectiveness in the entire population, IV methods cannot produce accurate estimates of the ATE. Table 15 shows that no estimator generates an unbiased or consistent estimate of the ATE. 2SRI_ATE generates the least biased estimate of ATE (-6.2%) but was the most biased estimator of LATE. With the exception of 2SRI_ATE, bias in LATE was relatively small across estimators. LATE estimates generated by 2SLS were less biased than estimates generated by any nonlinear method. Density plots showing the bias of estimates for the alternative estimators in Scenario 3 are given in Figure 7: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 15: Average % Bias of Estimate, by Estimator (Scenario 3)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	-13.48	-0.48
2SPS_ATE	-12.55	0.58
2SPS_LATE	-14.82	-2.02
2SRI_ATE	-6.18	7.91
2SRI_LATE	-12.10	1.10

Figure 7: Density Plot of % Bias of Estimates, by Estimator (Scenario 3). X_2 Factors, Unmeasured X_2, Essential Heterogeneity.



Scenario 4

Scenario 4 introduces the idea of X_1 factors—factors that affect the effectiveness of treatment but not outcomes directly. This is distinct from Scenarios 1-3, which assume all factors are X_2 factors that affect both treatment effectiveness and outcomes. Treatment effect heterogeneity is non-essential in Scenario 4—there is no correlation between treatment effectiveness and treatment value (i.e., treatment effectiveness is unrelated to probability of treatment). This is evident in the summary statistics for percentage of marginal patients, mean treatment value (T*), and other factors across deciles of treatment effectiveness provided in Table 16. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 4 described in detail in the methods section.

Table 16: Summary Statistics by Treatment Effectiveness Decile (Scenario 4)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	-0.35	18	-0.01	50	46
2	20,000	-0.13	18	0.01	51	34
3	20,000	0.00	18	-0.01	50	50
4	20,000	0.09	19	0.00	50	84
5	20,000	0.16	18	0.00	50	79
6	20,000	0.25	18	0.00	50	74
7	20,000	0.36	18	0.00	50	70
8	20,000	0.50	18	0.00	50	66
9	20,000	0.66	18	-0.01	50	61
10	20,000	0.83	18	0.01	50	56
Total	200,000	0.24	18	0.00	50	62

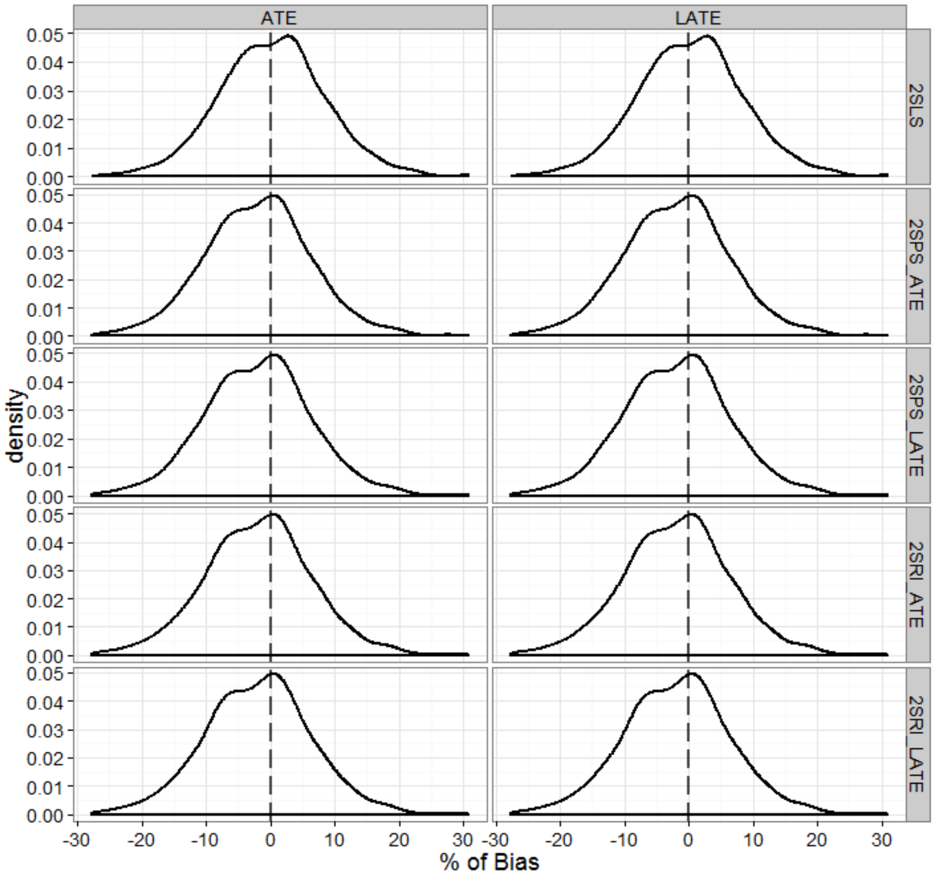
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 4 are provided in Table 17. True ATE and true LATE were estimated to be 0.236. Because treatment effect heterogeneity is non-essential in Scenario 4, true LATE is equal to true ATE. All estimators generated unbiased and consistent estimates of ATE and LATE. However, while differences in bias are small, 2SLS generates the least biased estimates for both ATE and LATE. Density plots showing the bias of estimates for the alternative estimators in Scenario 4 are given in Figure 8: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 17: Average % Bias of Estimate, by Estimator (Scenario 4)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	0.28	0.30
2SPS_ATE	-1.97	-1.95
2SPS_LATE	-1.93	-1.91
2SRI_ATE	-1.93	-1.91
2SRI_LATE	-1.93	-1.91

Figure 8: Density Plot of % Bias of Estimates, by Estimator (Scenario 4). X_1 and X_2 Factors, Unmeasured X_1, Non-Essential Heterogeneity.



Scenario 5

Scenario 5 builds upon Scenario 4 by assuming that X_1 and X_2 factors exist but treatment effect heterogeneity is quasi-essential, as in Scenarios 1a and 1b. Treatment effectiveness is indirectly related to treatment choice through X_2 factors that are directly related to treatment choice and related to treatment effectiveness nonlinearly. Table 18 shows summary statistics for mean treatment value (T*), percent marginal, and other factors across deciles of treatment effectiveness. This table illustrates the nonlinear relationship between treatment effectiveness and treatment value (T*)—which is related to probability of treatment. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 5 described in detail in the methods section.

Table 18: Summary Statistics by Treatment Effectiveness Decile (Scenario 5)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	-0.35	15	-0.05	49	47
2	20,000	-0.13	9	0.88	71	32
3	20,000	0.00	9	0.45	58	51
4	20,000	0.09	10	-0.74	26	81
5	20,000	0.16	13	-0.30	38	77
6	20,000	0.25	15	0.09	50	75
7	20,000	0.36	16	0.49	63	75
8	20,000	0.50	13	0.88	76	78
9	20,000	0.66	10	1.32	87	85
10	20,000	0.83	4	1.94	96	94
Total	200,000	0.24	11	0.50	62	69

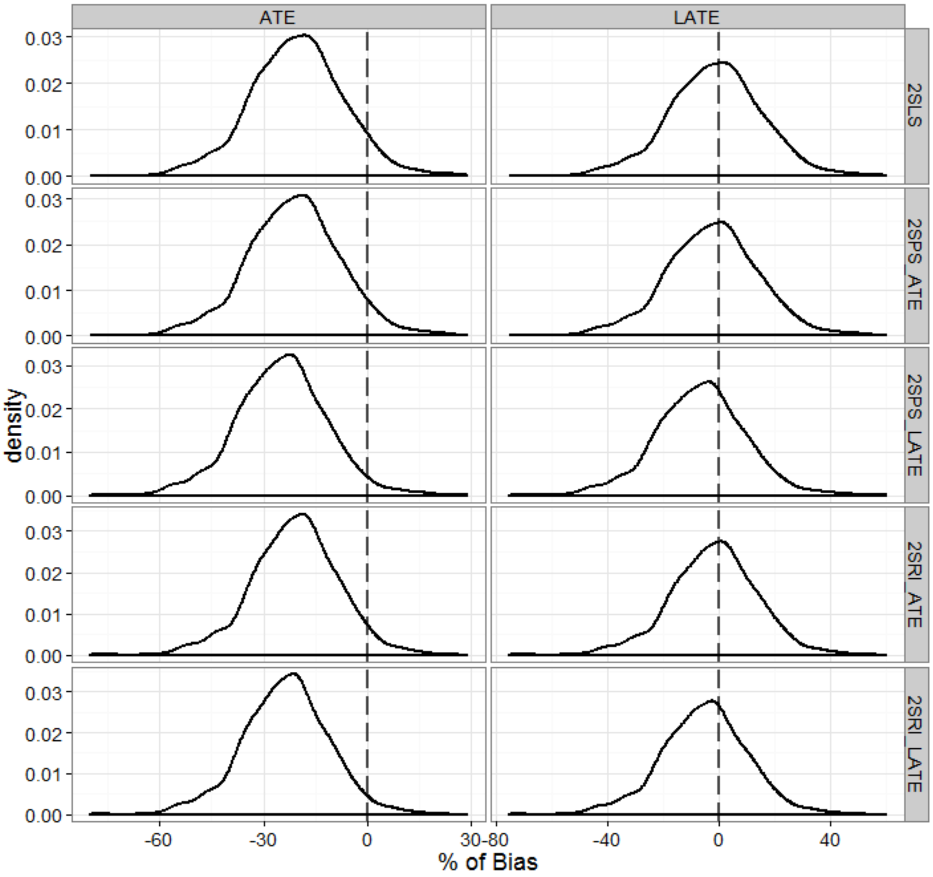
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 5 are provided in Table 19. True ATE was estimated to be 0.24. True LATE was estimated to be 0.19. In Scenario 5, no estimator generates an unbiased or consistent estimate of the ATE. Though heterogeneity is not strictly essential in this scenario, there exists some degree of sorting into treatment choice—albeit nonlinearly with effectiveness—and marginal individuals are not uniformly distributed throughout the distribution of treatment effectiveness. Because IV methods use only variation in treatment choice for the marginal individuals in estimating treatment effects, and the population of marginal individuals is not reflective of the population as a whole, IV methods do not produce accurate estimates of the ATE. 2SLS produced the least biased estimate of LATE (1.08% bias), though 2SPS_ATE and 2SRI_ATE were also relatively unbiased (2.18% biased and 1.73% biased, respectively). Density plots showing the bias of estimates for the alternative estimators in Scenario 5 are given in Figure 9: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 19: Average % Bias of Estimate, by Estimator (Scenario 5)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	-20.31	-1.08
2SPS_ATE	-21.20	-2.18
2SPS_LATE	-24.71	-6.53
2SRI_ATE	-20.84	-1.73
2SRI_LATE	-23.55	-5.10

Figure 9: Density Plot of % Bias of Estimates, by Estimator (Scenario 5). X_1 and X_2 Factors, Unmeasured X_1, Quasi-Essential Heterogeneity.



Scenario 6

As in Scenarios 4 and 5, Scenario 6 assumes that X_1 and X_2 factors exist. However, treatment effect heterogeneity in Scenario 6 is essential—treatment effectiveness has a direct positive effect on treatment choice, such that $\partial T^*/\partial TE > 0$. In other words, treatment decision makers are sorting-on-the-gain and treated individuals are unique, relative to untreated individuals, in terms of the distribution of treatment effectiveness. This is evident in the summary statistics for mean treatment value (T*), and other factors across deciles of treatment effectiveness given in Table 20. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 6 described in detail in the methods section.

Table 20: Summary Statistics by Treatment Effectiveness Decile (Scenario 6)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	-0.35	15	-0.69	27	54
2	20,000	-0.13	18	-0.25	41	35
3	20,000	0.00	18	0.00	50	51
4	20,000	0.09	18	0.18	57	84
5	20,000	0.16	17	0.33	62	81
6	20,000	0.25	16	0.49	67	78
7	20,000	0.36	15	0.71	74	78
8	20,000	0.50	12	0.99	82	81
9	20,000	0.66	9	1.31	89	86
10	20,000	0.83	6	1.66	94	92
Total	200,000	0.24	14	0.47	64	72

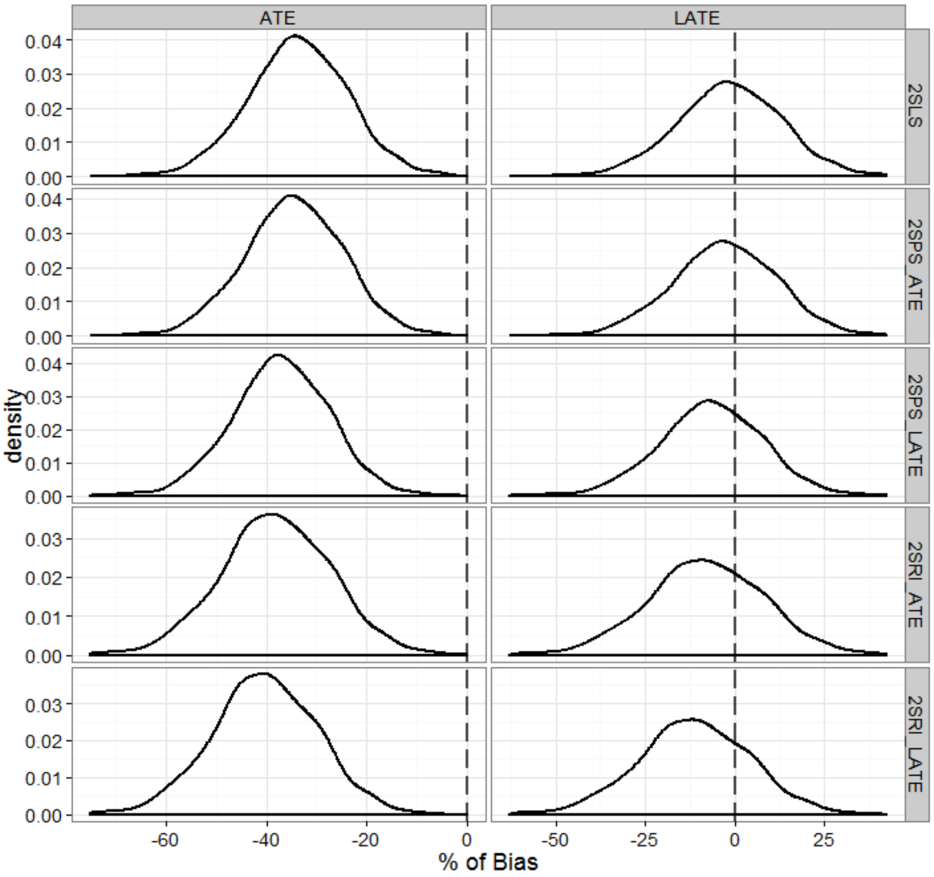
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 6 are provided in Table 21. True ATE was estimated to be 0.236. True LATE was estimated to be 0.159. For Scenario 6, no estimator generates an unbiased or consistent estimate of the ATE. 2SLS estimates were least biased, at 34% absolute bias. Like Scenario 3, treatment decision makers are "sorting-on-the-gain" in Scenario 6. Only variation in treatment choice for the marginal individuals is used in estimating treatment effects and, because the marginal population is unique in terms of treatment effectiveness, IV methods do not produce accurate estimates of the ATE. Moreover, while 2SLS and 2SPS_ATE generated relatively unbiased estimates of LATE, the percentage of bias of LATE estimates generated by nonlinear 2SRI methods was not insubstantial. Density plots showing the bias of estimates for the alternative estimators in Scenario 6 are given in Figure 10: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 21: Average % Bias of Estimate, by Estimator (Scenario 6)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	-33.72	-1.50
2SPS_ATE	-34.61	-2.83
2SPS_LATE	-37.19	-6.66
2SRI_ATE	-38.19	-8.15
2SRI_LATE	-40.38	-11.41

Figure 10: Density Plot of % Bias of Estimates, by Estimator (Scenario 6). X_1 and X_2 Factors, Unmeasured X_1, Essential Heterogeneity.



Scenario 7

Scenario 7 introduces the idea of X_3 factors—factors that affect outcomes directly but are unrelated to the effectiveness of treatment on outcome. This is distinct from Scenarios 1-6, where no factor exists that affects outcome but is unrelated to the effectiveness of treatment on outcome. Treatment effect heterogeneity is non-essential in Scenario 7—there is no correlation between treatment effectiveness and treatment value (i.e., treatment effectiveness is unrelated to probability of treatment). This is evident in the summary statistics for percentage of marginal patients, mean treatment value (T*), and other factors across deciles of treatment effectiveness provided in Table 22. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 7 described in detail in the methods section.

Table 22: Summary Statistics by Treatment Effectiveness Decile (Scenario 7)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	0.12	15	0.50	66	80
2	20,000	0.16	15	0.50	65	78
3	20,000	0.21	15	0.50	65	76
4	20,000	0.27	15	0.51	65	72
5	20,000	0.32	14	0.52	66	69
6	20,000	0.37	15	0.49	65	65
7	20,000	0.40	15	0.50	66	58
8	20,000	0.42	15	0.49	66	51
9	20,000	0.43	15	0.49	65	52
10	20,000	0.43	15	0.48	65	51
Total	200,000	0.31	15	0.50	66	65

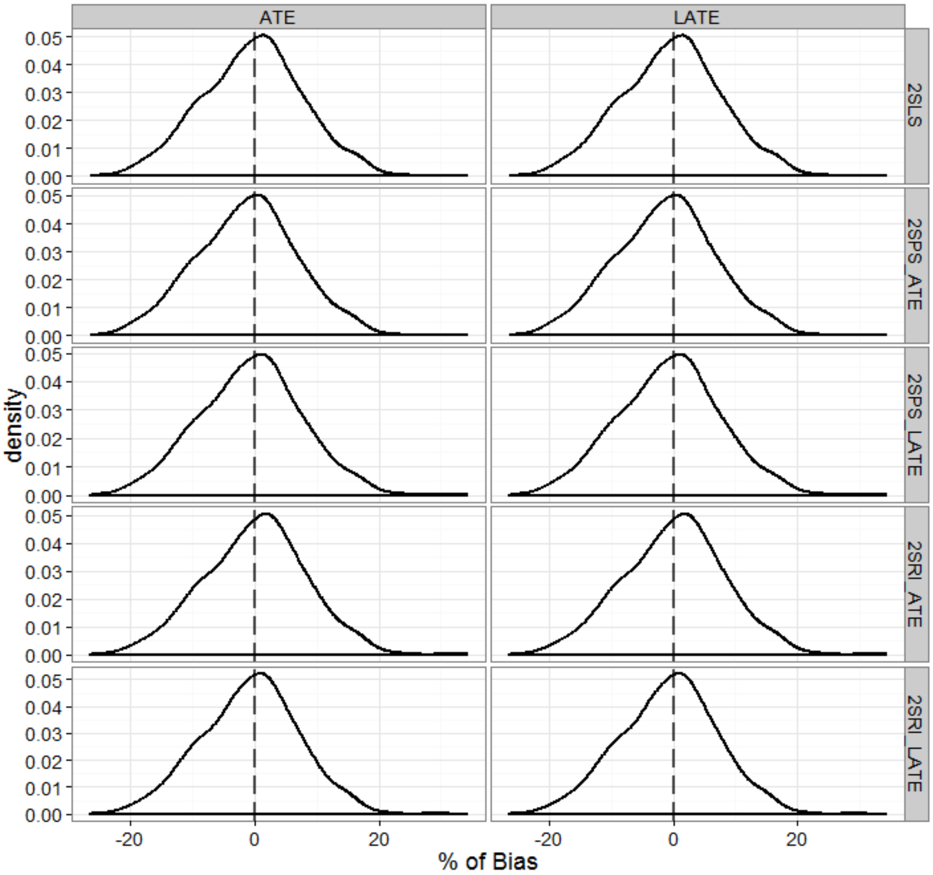
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 7 are provided in Table 23. True ATE and true LATE were both estimated to be 0.315. Because treatment effect heterogeneity is non-essential in Scenario 7, true LATE is equal to true ATE. All estimators generated unbiased and consistent estimates of ATE and LATE. The estimates from nonlinear 2SRI_ATE model are least biased. This greater precision of the probit 2SRI model may be explained by probit being the correct functional form for the outcome model, which is probit in truth. Density plots showing the bias of estimates for the alternative estimators in Scenario 7 are given in Figure 11: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 23: Average % Bias of Estimate, by Estimator (Scenario 7)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	-0.32	-0.31
2SPS_ATE	-1.26	-1.24
2SPS_LATE	-0.70	-0.68
2SRI_ATE	0.06	0.08
2SRI_LATE	-0.78	-0.77

Figure 11: Density Plot of % Bias of Estimates, by Estimator (Scenario 7). X_2 and X_3 Factors, Unmeasured X_3, Non-Essential Heterogeneity.



Scenario 8

Scenario 8 builds upon Scenario 7 by assuming that X_2 and X_3 factors exist but treatment effect heterogeneity is quasi-essential. Treatment effectiveness is indirectly related to treatment choice through X_2 factors that affect treatment choice and are related to treatment effectiveness nonlinearly. Table 24 shows summary statistics for mean treatment value (T*), percent marginal, and other factors across deciles of treatment effectiveness. This table illustrates the nonlinear relationship between treatment effectiveness and treatment value (T*)—which represents probability of treatment. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 8 described in detail in the methods section.

Table 24: Summary Statistics by Treatment Effectiveness Decile (Scenario 8)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	0.12	5	1.85	93	83
2	20,000	0.16	8	1.54	89	82
3	20,000	0.21	10	1.24	84	80
4	20,000	0.27	12	0.96	78	76
5	20,000	0.32	14	0.65	70	70
6	20,000	0.37	16	0.34	61	64
7	20,000	0.40	15	-0.08	48	51
8	20,000	0.42	15	-0.48	35	39
9	20,000	0.43	14	-0.50	34	38
10	20,000	0.43	15	-0.50	34	39
Total	200,000	0.31	12	0.50	63	62

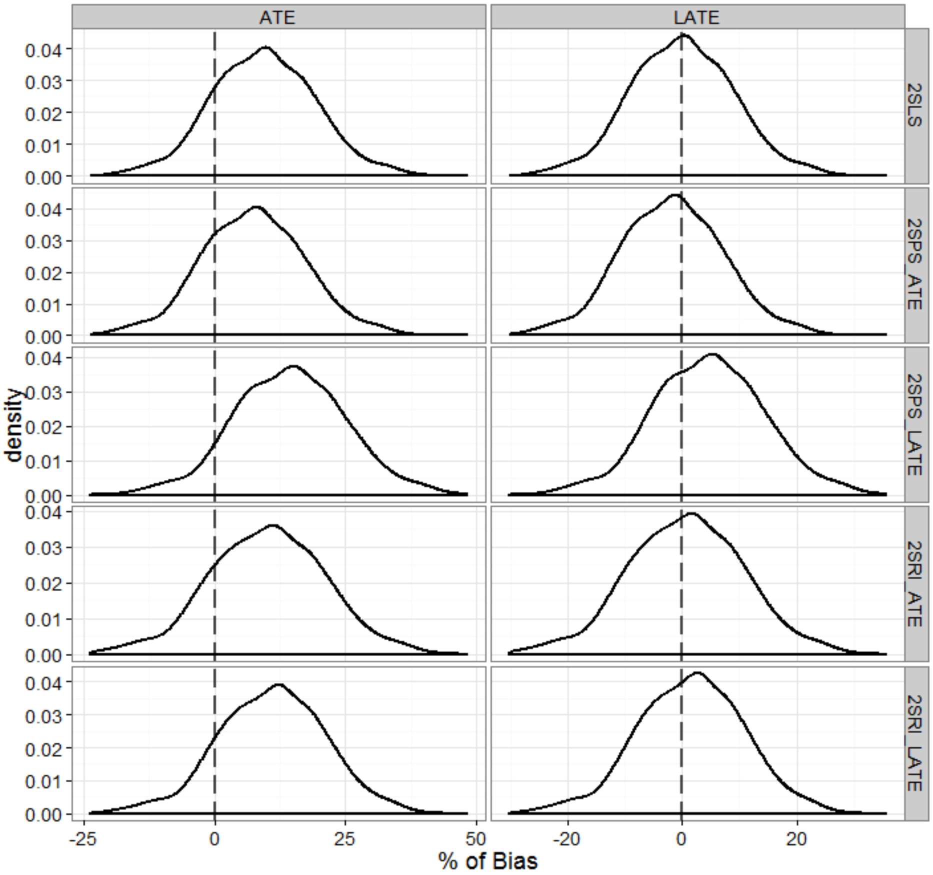
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 8 are provided in Table 25. True ATE was estimated to be 0.31. True LATE was estimated to be 0.34. In Scenario 8, no estimator generates an unbiased or consistent estimate of the ATE. This is, once again, because there is quasi-essential heterogeneity and the marginal patients are unique in terms of treatment effectiveness. 2SLS produced the least biased estimate of LATE (0.00% bias). Density plots showing the bias of estimates for the alternative estimators in Scenario 8 are shown in Figure 12: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 25: Average % Bias of Estimate, by Estimator (Scenario 8)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	9.36	0.00
2SPS_ATE	7.43	-1.77
2SPS_LATE	14.41	4.62
2SRI_ATE	10.07	0.65
2SRI_LATE	10.95	1.46

Figure 12: Density Plot of % Bias of Estimates, by Estimator (Scenario 8). X_2 and X_3 Factors, Unmeasured X_3, Quasi-Essential Heterogeneity.



Scenario 9

As in Scenarios 7 and 8, Scenario 9 assumes that X_2 and X_3 factors exist. However, treatment effect heterogeneity in Scenario 9 is essential—treatment effectiveness has a direct and positive effect on treatment choice, such that $\partial T^*/\partial TE > 0$. In other words, treatment decision makers are sorting-on-the-gain and treated individuals are unique, relative to untreated individuals, in terms of the distribution of treatment effectiveness. This is evident in the summary statistics for mean treatment value (T*), and other factors across deciles of treatment effectiveness given in Table 26. These statistics are based on a sample of 200,000 observations, generated following the models relevant to Scenario 9 described in detail in the methods section.

Table 26: Summary Statistics by Treatment Effectiveness Decile (Scenario 9)

Treatment Effectiveness Decile (1 = Low, 10 = High)	N	Average Treatment Effect (TE)	Unadjusted TE	% Marginal	Average Treatment Value (T*)	% Treated	% Y = 1 (Cured)
1	20,000	0.12	0.15	15	0.45	64	79
2	20,000	0.16	0.20	14	0.59	68	79
3	20,000	0.21	0.27	13	0.80	74	78
4	20,000	0.26	0.33	11	1.02	79	77
5	20,000	0.32	0.40	10	1.20	84	75
6	20,000	0.37	0.46	9	1.37	87	72
7	20,000	0.40	0.51	8	1.52	90	68
8	20,000	0.42	0.53	7	1.58	90	62
9	20,000	0.43	0.54	7	1.59	90	63
10	20,000	0.43	0.55	7	1.63	91	62
Total	200,000	0.31	0.39	10	1.17	82	71

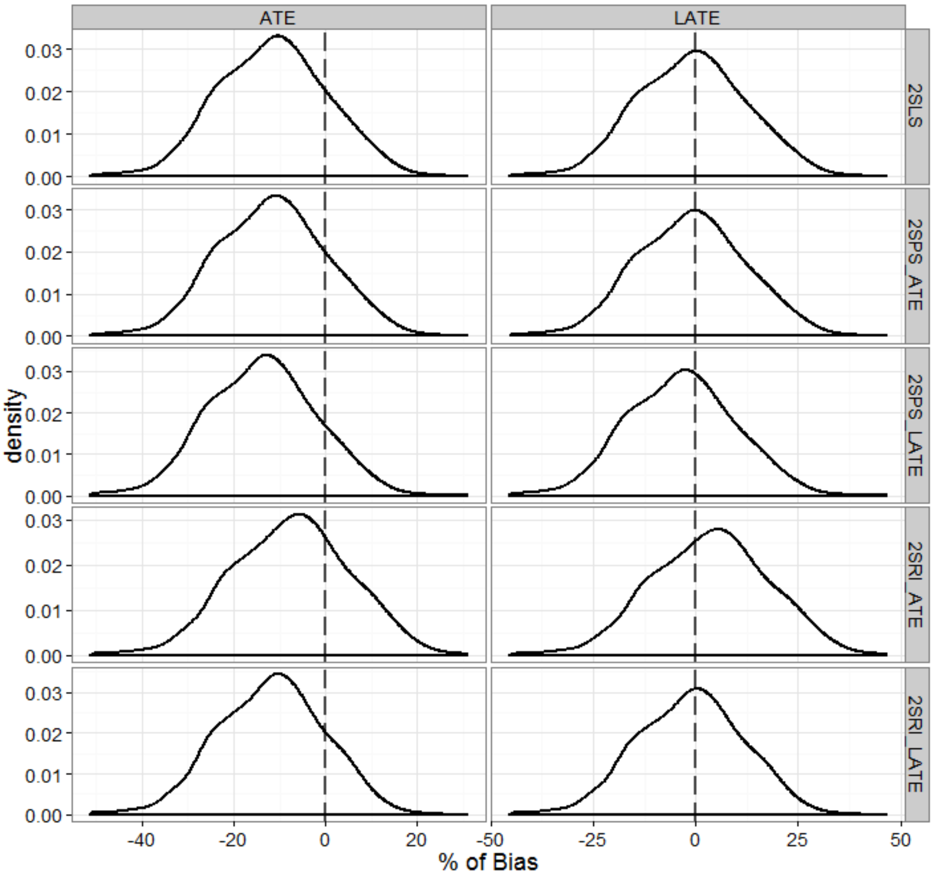
* Marginal defined as -.25 < T* < .25.

Simulation results for Scenario 9 are provided in Table 27. True ATE was estimated to be 0.32. True LATE was estimated to be 0.29. For Scenario 9, no estimator generates an unbiased or consistent estimate of the ATE. 2SRI_ATE is least biased, but absolute bias of these estimates is not negligible at 7.4%. Like Scenarios 3 and 6, treatment decision makers are “sorting-on-the-gain” in Scenario 9. Because only the variation in treatment choice for the marginal individuals is used in estimating treatment effects, and because the marginal population is unique in terms of treatment effectiveness under essential heterogeneity, IV methods cannot produce accurate estimates of the ATE. 2SLS generates the least biased estimate of LATE (-0.88% biased), followed closely by 2SPS_ATE (-1.1% biased) and 2SRI_LATE (-1.4% biased). Density plots showing the bias of estimates for the alternative estimators in Scenario 9 are given in Figure 13: The left column shows bias of each estimator (indicated by label on the left) relative to true ATE, and the right column shows bias of each estimator relative to true LATE.

Table 27: Average % Bias of Estimate, by Estimator (Scenario 9)

Estimator	% Bias for ATE	% Bias for LATE
2SLS	-11.37	-0.88
2SPS_ATE	-11.55	-1.08
2SPS_LATE	-13.73	-3.52
2SRI_ATE	-7.44	3.52
2SRI_LATE	-11.82	-1.39

Figure 13: Density Plot of % Bias of Estimates, by Estimator (Scenario 9). X_2 and X_3 Factors, Unmeasured X_3, Essential Heterogeneity.



Summary of Simulation Results

Taken together, simulation results support ideas that IV methods only identify treatment effects for the marginal individuals and cannot be used to make more general inferences without additional assumptions about the nature of treatment effect heterogeneity and treatment choice. Table 28 summarizes results across all simulations. From Table 28, it is clear that no IV method can be generally expected to generate consistent estimates of the ATE across all possible scenarios.

Table 28: Summary of Simulation Results – All Scenarios

	Non-Essential		Quasi-Essential ("Terza"; Balanced)		Quasi-Essential		Essential	
	% Bias for ATE	% Bias for LATE	% Bias for ATE	% Bias for LATE	% Bias for ATE	% Bias for LATE	% Bias for ATE	% Bias for LATE
X_2 only								
2SLS	0.24	0.21	-4.79	-0.12	13.18	0.29	-13.48	-0.48
2SPS_ATE	-0.11	-0.14	-6.62	-2.04	13.72	0.77	-12.55	0.58
2SPS_LATE	-0.49	-0.52	*0.17	5.08	22.64	8.68	-14.82	-2.02
2SRI_ATE	0.16	0.13	-1.69	3.13	-14.46	-24.19	-6.18	7.91
2SRI_LATE	0.18	0.15	-7.12	-2.56	-13.22	-23.10	-12.10	1.10
X_1 and X_2								
2SLS	0.28	0.30			-20.31	-1.08	-33.72	-1.50
2SPS_ATE	-1.97	-1.95			-21.20	-2.18	-34.61	-2.83
2SPS_LATE	-1.93	-1.91			-24.71	-6.53	-37.19	-6.66
2SRI_ATE	-1.93	-1.91			-20.84	-1.73	-38.19	-8.15
2SRI_LATE	-1.93	-1.91			-23.55	-5.10	-40.38	-11.41
X_2 and X_3								
2SLS	-0.32	-0.31			9.36	0.00	-11.37	-0.88
2SPS_ATE	-1.26	-1.24			7.43	-1.77	-11.55	-1.08
2SPS_LATE	-0.70	-0.68			14.41	4.62	-13.73	-3.52
2SRI_ATE	0.06	0.08			10.07	0.65	-7.44	3.52
2SRI_LATE	-0.78	-0.77			10.95	1.46	-11.82	-1.39

*This result should not be interpreted as suggesting 2SPS_LATE generates unbiased estimates of ATE in this scenario. Rather this result was due to "chance": the biased of 2SPS_LATE estimates coincidentally equal the true ATE in this model specification, alternative model specifications run in sensitivity analyses demonstrate bias of 2SPS_LATE for true ATE.

Under non-essential heterogeneity, all methods generate unbiased estimates of LATE. Because ATE = LATE under non-essential heterogeneity, all IV methods also generate consistent ATE estimates. Conversely, when treatment effects are heterogeneous across patients and treatment decision-makers are sorting-on-the-gain such that those with higher effectiveness are more likely to be treated—that is, heterogeneity is essential—IV methods do not yield estimates of the ATE. Under scenarios of essential heterogeneity: no method generates consistent ATE estimates, 2SLS generates consistent estimates of LATE that are less biased than any nonlinear method examined, and nonlinear 2SRI may generate inconsistent estimates of the LATE. These results are similar in scenarios of quasi-essential heterogeneity except under particular unique settings where no X_1 or X_3 factors exist and marginal patients are uniformly distributed across the distribution of treatment effectiveness.

Similar to the findings of Terza et al., our simulations suggest that nonlinear 2SRI generates consistent estimates of the ATE under settings of quasi-essential heterogeneity when marginal patients are uniformly distributed across the distribution of treatment effectiveness and only X_2 factors exist. This is the “Quasi-Essential Balanced” scenario shown in Table 28 (i.e., Scenario 1a). When marginal individuals are well representative of the population as a whole—in terms of treatment effectiveness—average derivative estimates generated by 2SRI appear to be unbiased for the ATE. However, even under these settings it is not necessarily true that unbiased absolute average treatment effects for subpopulations can be estimated using nonlinear 2SRI—as Terza et al. (2002) have suggested.[62] Even when all marginal individuals are assumed to be directly identifiable in the data, such that an average marginal effect could be calculated across the subpopulation, nonlinear 2SRI generates biased and inconsistent estimates of LATE in our simulation. Note that, while simulation results for Scenario 1a (i.e., the “Terza”, Balanced Quasi-Essential scenario in Table 28) appear to suggest that 2SPS_LATE generates the least biased estimate of ATE, this is misleading. This result is an artifact of coincidence related to the specific model parameters specified in this scenario. Sensitivity analyses using alternative parameter values generated 2SPS_LATE estimates that are biased for ATE.

Though it may be taken for granted, a significant finding from these simulations is that linear 2SLS generates consistent estimates of LATE across all scenarios. This is not a surprising finding—it is well established that 2SLS generates consistent estimates of LATE with minimal assumptions—but it is a point that often goes under-appreciated. LATE estimates seem to be often viewed as a "second-best" option—researchers report LATE estimates with the mentality of "there is confounding, so we will take what we can get". This is unfortunate. For clinical scenarios where treatment effects are heterogeneous and treatment decision makers are sorting patients based upon patients expected idiosyncratic gains, LATE may be the treatment effect parameter of greatest value to clinicians and policy-makers. If treatment decision makers are already able to clearly identify those patients who stand to benefit most and least from treatment, and use this information when making treatment decisions, then the uncertainty lies in the patients for whom this cost-benefit calculation is least clear. These "marginal patients" may often be the patients for whom research efforts may provide the most benefit. Researchers can be confident that 2SLS generates unbiased estimates of LATE, regardless of the circumstances underlying treatment effect heterogeneity and treatment choice. This is not necessarily the case when using nonlinear 2SRI or nonlinear 2SPS. Interestingly, however, results from these simulations suggest that estimation of LATE using nonlinear 2SPS is not encumbered by the need to identify the subpopulation of marginal individuals. Nonlinear 2SPS estimates are unbiased for LATE when average derivate estimates are calculated across the entire population (i.e., 2SPS_ATE is unbiased for LATE), but biased for LATE when average derivative is taken over the marginal subpopulation directly (i.e., 2SPS_LATE is biased for LATE).

Model Performance Under Full Information

Though IV methods are traditionally used as a tool to estimate treatment effects when unmeasured confounding is present, we examine the ability of IV estimators to generate unbiased and consistent estimates of ATE and LATE when all relevant variables are measured to see how these estimators perform under “ideal” settings. With all variables measured, nonlinear IV methods—particularly 2SRI—may correctly model treatment effect heterogeneity such that average derivative estimates are unbiased for true ATE and LATE. It is expected that 2SLS will still generate consistent estimates of LATE, but that these estimates will not be generalizable to estimate ATE without additional assumptions. To examine the ability of alternative estimators to generate consistent estimates of ATE and LATE under settings of full information, the 10 scenarios above were run in models with all variables measured and included as covariates in regressions. Results of these simulations are summarized in Table 29.

Table 29: Summary of Full Information Simulation Results – All Scenarios

	Non-Essential		Quasi-Essential (“Terza”; Balanced)		Quasi-Essential		Essential	
	% Bias for ATE	% Bias for LATE	% Bias for ATE	% Bias for LATE	% Bias for ATE	% Bias for LATE	% Bias for ATE	% Bias for LATE
X_2 only								
2SLS	0.17	0.14	-4.79	-0.12	13.19	0.30	-13.57	-0.58
2SPS_ATE	-0.81	-0.83	-4.76	-0.08	9.52	-2.94	-10.80	2.60
2SPS_LATE	-1.18	-1.20	-4.61	0.07	114.81	90.36	-19.66	-7.59
2SRI_ATE	0.07	0.04	0.06	4.97	0.09	-11.31	0.11	15.15
2SRI_LATE	0.10	0.07	-4.61	0.07	12.84	-0.01	-12.95	0.12
X_1 and X_2								
2SLS	0.25	0.27			-20.37	-1.15	-33.74	-1.53
2SPS_ATE	0.94	0.96			-8.29	13.85	-20.99	17.42
2SPS_LATE	1.00	1.01			-11.82	9.47	-21.02	17.37
2SRI_ATE	1.41	1.43			-11.22	10.21	-24.77	11.79
2SRI_LATE	1.42	1.43			-13.85	6.95	-27.27	8.08
X_2 and X_3								
2SLS	-0.37	-0.36			9.29	-0.06	-11.38	-0.89
2SPS_ATE	-1.08	-1.06			7.88	-1.35	-10.24	0.38
2SPS_LATE	-1.43	-1.41			12.71	3.07	-16.20	-6.29
2SRI_ATE	0.93	0.95			10.10	0.68	-5.80	5.34
2SRI_LATE	-0.29	-0.27			9.47	0.10	-11.40	-0.91

It appears clear from the results of Table 29 that relative estimator performance for estimation of ATE and LATE is not sensitive to whether there are unmeasured variables or confounding. The exception, however, is with nonlinear 2SRI in Scenarios 1-3. Under settings of full information and only X_2 factors (i.e., Scenarios 1-3), nonlinear

2SRI_ATE appears to generate consistent estimates of ATE and 2SRI_LATE appears to be unbiased for LATE, even in scenarios with essential or quasi-essential heterogeneity. This is likely because treatment effect heterogeneity is properly modeled by nonlinear 2SRI when all factors are measured and included as covariates. Nonlinear 2SRI works under the premise that the residual from the first-stage treatment choice model can serve as a valid approximation of the unmeasured confounder. However, when all variables are measured in the treatment choice and outcome model, the residual term is purely noise and has no impact on the model (this is similar to why nonlinear 2SRI performed well in non-essential scenarios for simulations with unmeasured variables discussed in the previous section). Under these settings, the nonlinear 2SRI model becomes comparable to a single stage probit model with full information and average derivative estimates will be consistent. However, this does not hold true when either X_1 or X_3 factors exist because including the nonlinear 2SRI model assumes that these are X_2 factors and models treatment effect heterogeneity incorrectly.

Unlike the nonlinear 2SRI model, nonlinear 2SPS does not identify ATE and LATE in Scenarios 1-3 with full information. While 2SRI carries forward information from the first-stage treatment choice model in the form of a residual term—which can be correctly given an estimated coefficient of zero in the regression model on outcome—2SPS carries forward the information by replacing observed treatment with predicted treatment and therefore forces the information from the first-stage treatment choice model to persist in the outcome model, affecting absolute treatment effect estimates. Under scenarios of full information where only X_2 factors exist, the arguments made by Terza et al. (2008) regarding superiority of 2SRI to 2SPS hold true.[19] As in the limited information scenarios, 2SPS_ATE estimates are unbiased for LATE. For 2SLS and 2SPS, the ability to generate consistent estimates of ATE and LATE is determined primarily by the circumstances regarding treatment effect heterogeneity and how it is related to treatment choices. 2SLS generates consistent estimates of LATE across all

Scenarios, but these estimates can only be generalized to ATE under non-essential heterogeneity or when ATE is equal to LATE by coincidence.

Empirical Result: Effect of ACE/ARBs on Survival

The purpose of this empirical example is to demonstrate the differences in estimates generated by linear 2SLS, nonlinear 2SPS, and nonlinear 2SRI estimators in actual practice. The focus for interpretation of results is on comparisons of ATE and LATE estimates generated by alternative estimators. After all inclusion criteria are applied, the analytical sample includes 63,685 fee-for-service Medicare beneficiaries with a new AMI in 2008.

The instrumental variable used in this analysis is ACE/ARB area treatment ratios (ATRs), a zip code level variable representing the ratio of predicted to actual patients treated using ACE/ARBs in the geographic region around beneficiaries' Medicare-associated zip code. The instrument is specified in the first-stage treatment choice model using binary variables indicating the quintile of the ATR to which a beneficiary's zip code is a member. First-stage instrument exclusion tests (Chow Test) were used to test the assumption that instruments are significantly related to treatment choice, independent of all other measured covariates.[96] The F-Statistic for the Chow test is 54.53, well above the commonly accepted threshold of weak instrument tests of $F > 10$. The instrument exclusion restriction—the assumption that the instruments are unrelated to the outcome or other unmeasured variables related to the outcome, independent of treatment—is not directly testable. However, the exclusion restriction is supported by Table 30, which shows no clear trend in the means of measured covariates related to patient health and outcomes across quintiles of ACE/ARB ATRs.

Table 30: Means of select covariates across quintiles of ACE/ARB Area Treatment Ratios

	Quintiles of Area Treatment Ratio (ATR), 150-person Circles					
	1	2	3	4	5	Total
N	12,738	12,739	12,738	12,734	12,736	63,685
Treatment and Outcome						
Survival Post 365	0.84	0.85	0.85	0.85	0.84	0.85
ACE/ARB	0.44	0.47	0.49	0.51	0.54	0.49
Comorbidities in 365-days prior to AMI						
Charlson pre365	1.8	1.9	1.8	1.8	1.9	1.8
Stroke pre365	0.054	0.052	0.054	0.053	0.057	0.054
Diabetes pre365	0.37	0.37	0.37	0.37	0.38	0.37
Heartfailure pre365	0.3	0.3	0.3	0.3	0.31	0.3
CABG pre365	0.025	0.027	0.025	0.025	0.025	0.025
Stent pre365	0.026	0.029	0.029	0.029	0.029	0.028
Characteristics of AMI						
AWAMI index	0.06	0.062	0.066	0.068	0.06	0.063
NSTEMI index	0.76	0.75	0.75	0.74	0.76	0.75
Comorbidities during stay for AMI						
Charlson index	3.4	3.4	3.3	3.3	3.4	3.4
Stroke index	0.029	0.025	0.028	0.026	0.027	0.027
Diabetes index	0.37	0.37	0.37	0.37	0.37	0.37
Heartfailure index	0.45	0.46	0.45	0.45	0.46	0.45
CABG index	0.39	0.4	0.4	0.4	0.4	0.4
Stent index	0.31	0.32	0.32	0.32	0.32	0.32

Table 30 (continued)

Demographics and Area Socio-Economic Factors						
Female	0.58	0.57	0.57	0.57	0.57	0.57
White	0.86	0.84	0.83	0.82	0.82	0.83
Black	0.074	0.07	0.069	0.086	0.094	0.078
Other race	0.068	0.092	0.1	0.096	0.089	0.09
Age 66-70	0.2	0.2	0.21	0.21	0.2	0.2
Age 71-75	0.19	0.2	0.2	0.2	0.2	0.2
Age 76-80	0.21	0.2	0.2	0.21	0.2	0.2
Age 81-85	0.2	0.19	0.19	0.18	0.18	0.19
Age 85+	0.21	0.21	0.2	0.21	0.21	0.21
% Low Income Area	0.49	0.48	0.51	0.51	0.49	0.5
High No HS Education Area	0.5	0.48	0.5	0.5	0.49	0.49
High % Poverty Area	0.47	0.49	0.5	0.49	0.51	0.49

Table 31 provides estimates for the effect of ACE/ARB use after acute myocardial infarction generated by alternative IV estimators. While no model generated an estimate for the effect of ACE/ARBs on 1-year survival that is statistically significant, principle interest for this example lies in comparisons of absolute effect estimates across and within estimators.

Table 31: Comparing Estimators for Regression Estimates of ACE/ARB on Survival (150-person circle ATR Quintiles for Instruments)

	2SLS	2SPS (Linear-Probit)	2SRI (Linear-Probit)	2SPS (Probit-Probit)	2SRI (Probit-Probit)
ATE		0.023	0.024	0.027	-0.024
LATE[b]	0.017	0.023	0.024	0.027	-0.024
LATE (at means)	0.017	0.022	0.023	0.026	-0.023
N	63,685				
Under ID	217.14				
(P-Value)	*(p<0.0001)*				
Chow Test[a]	54.53				
(P-Value)	*(p<0.0001)*				

[a] F-Test of excluded instrumental variables (ATR quintiles) in first-stage treatment choice equation

[b] Patients marginal if predicted probability of treatment ($P(T)$) satisfied $0.45 < P(T) < 0.55$

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

The parameter estimate generated by 2SLS was 0.017. It is well established that two-stage least squares generates consistent estimates of LATE with minimal assumptions—this idea was supported by the simulation results in the previous section. As such, the 2SLS estimate is considered the “gold standard” for comparisons of LATE in this example. LATE estimates generated by nonlinear 2SPS (0.023) and nonlinear 2SRI (0.024) models are not meaningfully different from the LATE estimate generated by 2SLS when estimated using a linear first-stage model and probit outcome model. Similarly, the LATE estimate generated by nonlinear 2SPS with a probit specification for both the first- and second-stage models is comparable to the 2SLS estimate of LATE.

However, estimates generated by 2SRI methods with a nonlinear (probit) first-stage model are considerably different from estimates yielded by all other models—the probit-probit 2SRI estimate is, in fact, similar in magnitude but opposite in direction of effect. This finding may indicate that the additional assumptions imposed by the nonlinear first-stage model results in inconsistent treatment effect estimates. Estimates of LATE generated by 2SPS and 2SRI models do not differ by the method used to identify the marginal individuals in the data. Treatment effect estimates from models using alternative methods for defining the marginal population are provided in appendix Tables A-2 and A-3.

The theoretical model discussed in relation to this empirical problem was most similar to simulation Scenario 9—essential heterogeneity with both X_2 and X_3 factors present. Simulation results for Scenario 9 suggest that 2SLS, 2SPS_ATE, 2SPS_LATE, 2SRI_ATE, and 2SRI_LATE are all relatively unbiased for the true LATE. 2SLS was least biased for LATE in the simulation. These trends from simulation Scenario 9 fit well to the findings of the empirical results. All estimates—excepting those generated by probit-probit 2SRI—are similar in magnitude, though nonlinear estimates differed slightly from the linear 2SLS estimate. However, no estimator generated consistent estimates of the ATE in simulation Scenario 9—or any simulation scenario of essential heterogeneity. As such, it is unlikely that any estimate reported in Table 31 is reflective of the true ATE, unless we alter our theory regarding the characteristics of treatment effect heterogeneity and treatment choice or accept additional strong assumptions. For example, estimates may be interpreted as ATE if it is assumed that treatment effect heterogeneity is actually non-essential or that the true ATE and true LATE in this population are equal by coincidence.

Estimates of LATE reflect the benefits of ACE/ARBs for patients whose treatment decisions were influenced by their geography. Theory suggests that the benefits of ACE/ARBs are heterogeneous and that treatment decision-makers are "sorting-on-the-gain" with respect to this heterogeneity, such that heterogeneity is "essential"—those with higher effectiveness are more likely to be treated, all else equal. It would be expected that the true average treatment effect for the treated population would be greater than LATE, while true the average treatment effect for the untreated population would be less than LATE. In this empirical example, absolute treatment effect estimates of ATE generated by nonlinear 2SRI and 2SPS were nearly equal to estimates of LATE for each model and specification. If it were assumed that nonlinear 2SRI and nonlinear 2SPS methods yield unbiased estimates of both the ATE and LATE, then the approximately equal estimates of ATE and LATE generated by these nonlinear estimates would suggest that true ATE and true LATE are approximately equal. However, while this is indeed possible, simulation results discussed in the previous section suggest that neither nonlinear 2SRI, nor nonlinear 2SPS, can identify the ATE in scenarios of essential heterogeneity. Moreover, estimates of ATE generated by nonlinear estimators were found to be less biased for true LATE then true ATE in simulation scenarios modeling essential heterogeneity. With these considerations, it is likely that both ATE and LATE estimates for the effects of ACE/ARBs on survival reflect only the treatment benefits for the marginal population and cannot be interpreted as reflecting average treatment effects across the entire population.

These inferences are robust to alternative specifications of the instrumental variables (i.e., larger or smaller geographic regions used to define ATRs), though magnitude of effect estimates is somewhat variable across these alternative model specifications. Tables of results for models using ATRs created from 100-person and 200-person circles are provided in the appendix (Appendix Tables A-2 and A-3, respectively).

CHAPTER 5

DISCUSSION AND CONCLUDING REMARKS

Simulations performed in this research support ideas that IV methods only identify treatment effects for the marginal individuals and cannot be used to make more general inferences without additional assumptions about the nature of treatment effect heterogeneity and treatment choice. For example, if researchers assume that heterogeneity in treatment effectiveness can be characterized as non-essential, then estimates from 2SLS, 2SPS, and 2SRI can be used to make inferences about population average treatment effects.[17,48] Or, if researchers can assume that all factors in the relationships being modeled are X_2 factors and are measured and specified in empirical models, then correctly specified nonlinear 2SRI (or a naïve nonlinear regression) may yield estimates of ATE even under settings of essential heterogeneity. Unfortunately, these may not be reasonable assumptions for many clinical scenarios in health care. Rather, if treatment effects are heterogeneous across patients then it may be likely that physicians are knowledgeable about variation in expected benefits of treatment across distinct patients. This knowledge may stem from their own expertise and training or through experience gained through accumulated clinical practice—a phenomena termed passive personalization.[48] If it is the case that treatment decision-makers are sorting patients into treatment based upon their expected idiosyncratic gains, then IV methods cannot be expected to generally yield estimates interpretable as population average treatment effects. However, across all potential clinical scenarios, 2SLS can be depended upon to generate consistent estimates of the LATE given only minimal assumptions regarding validity of the specified instrumental variables—fundamental assumptions to all IV methods, linear or nonlinear.[2,12] Assertions that nonlinear 2SRI should be adopted because 2SLS results in inconsistent estimates do not hold true in cases where LATE is the parameter of interest.

When marginal individuals are well representative of the population as a whole average derivative estimates generated by correctly specified nonlinear 2SRI models may be unbiased for the ATE. It is not necessarily true, however, that unbiased average treatment effects for subpopulations can be estimated using even correctly specified nonlinear 2SRI—as has been suggested to be a benefit of nonlinear 2SRI, relative to 2SLS, in the past.[62] Terza (2002) reports heterogeneous effects of alcohol abuse on employment, asserting that this is possible because the nonlinear 2SRI method accounts for heterogeneity on both the observables and unobservables.[62] However, the results of our simulations suggest that these conclusions may only be valid if it can be assumed that marginal individuals are uniformly distributed across the distribution of treatment effectiveness—assumptions not discussed by Terza. Moreover, even if the above assumption is satisfied, the conditions necessary to consistently estimate absolute effects for subgroups using nonlinear 2SRI have not been clearly defined. Simulation results from Scenarios 1b, 5, and 6 show that absolute LATE estimates generated by 2SRI, calculated as an average derivative over the known marginal population, were biased for LATE. It may not be possible to consistently estimate subgroup absolute effects when there is heterogeneity on the unobservables. Nonlinear models force treatment effects to be heterogeneous across all patient subpopulation, as defined by the covariates affecting outcome directly. This is an artifact of the nonlinear model, but is an untestable assumption that may not reflect the truth. As such, inferences from nonlinear models for treatment effect differences across populations may be misleading. For example, while nonlinear 2SRI or 2SPS models would suggest that treatment effectiveness varies across X_3 groups in simulation Scenarios 7-9, the X_3 characteristics have no relationship with treatment effectiveness in truth and such inferences would be erroneous.

This research serves as support for the importance of, and need for well-developed theoretical models in empirical work. These theoretical models underlie the validity of inferences made by empirical researchers when discussing and interpreting estimates generated using observational data. A principle motivation for this research has been an apparent lax in interpretation of empirical results, particularly when instrumental variables methods are used. The job of the researcher is not to simply provide estimates, but to responsibly interpret those estimates and provide discussion regarding to whom those estimates may apply, given theory about treatment effect heterogeneity and circumstances underlying treatment choices. While it is likely that some degree of essential heterogeneity exists in a plurality of clinical settings, only 5 of 51 reviewed papers that use nonlinear 2SRI as their primary estimator include discussion stating that treatment effect estimates from their IV regressions may only be only locally interpretable. Many empirical works simply provide estimates and proceed to interpret them as a general effect with no discussion given to whether treatment effects may be heterogeneous or to whom these estimates may apply. Researchers need to be clear about interpretation of estimates, especially from IV estimators. Proper discussion of the theorized treatment choice model and characteristics of treatment effect heterogeneity is an important first step. IV models do not simply remove bias, as a surgeon would remove a tumor. IV methods estimate treatment effects for a specific subpopulation that may be very unique in terms of treatment effects and not reflective of the larger population. While this has been repeatedly stated and stressed by researchers in the IV methods literature, it is often left unsaid in empirical work. This may be a particularly prevalent issue in empirical work employing nonlinear 2SRI methods because the methodological work commonly cited as evidence that nonlinear 2SRI should be used in place of 2SLS includes no discussion of LATE or local interpretability of IV estimates.[18,19]

Perspectives of LATE as an interesting parameter in and of itself, and not a treatment effect parameter that researchers must settle for when there is unmeasured confounding, is something I feel must be stressed. If treatment effects are truly heterogeneous across patients then population average treatment effects may often be far less meaningful or clinically relevant than LATE. Who really is the "average" patient? Ideas of sorting-on-the-gain or "passive personalization" represents the artistry of medicine. The richness and clinical importance of information observed and analyzed by physicians when making treatment decisions, as compared to the very limited information available in observational data, cannot be overemphasized. With this data, treatment decision makers may often be able to use their clinical expertise to identify those patients who stand to benefit most and least from treatment and ultimately would be expected to use this information when making treatment decisions. In such cases, the uncertainty lies in the patients for whom this cost-benefit calculation is least clear. For these patients, naturally random factors unrelated to treatment effectiveness or outcomes may affect treatment decisions. These are the marginal patients whose treatment choices may be influenced by policies seeking to influence treatment rates. These marginal patients may be the patients for whom research efforts may provide the most benefits, and for whom LATE estimates may most accurately represent.

There are several limitations that must be discussed regarding this research. The simulation scenarios examined are not exhaustive of all possible clinical settings. This research is a small step that I hope will further both the understanding and discussion of the ability of IV estimators to produce estimates of ATE and LATE across alternative possible clinical scenarios. Previous research has suggested that nonlinear 2SRI can be generally used to obtain consistent estimates of population average treatment effects in models with inherently nonlinear dependent variables. However, past research did not examine the properties of estimators across alternative scenarios of treatment effect heterogeneity and sorting-on-the-gain. As a result, the interpretability of estimates across many relevant clinical settings was not known and swift adoption of nonlinear 2SRI methods may have resulted in potentially misleading inferences. The simulations conducted in this research aid by clearly describing alternative possible scenarios and demonstrating the limitations of nonlinear 2SRI and potential benefits of linear 2SLS methods across these scenarios.

This research examines only the ability of nonlinear IV methods to generate unbiased estimates of absolute treatment effects. Many clinicians believe that absolute treatment effects are the most valuable for patients and physicians when making treatment decisions.[97] However, the properties of relative effect estimates generated by nonlinear IV models, which may be the concepts of interest for some empirical research applications, are not examined or discussed. Furthermore, results reported for the magnitude of bias of absolute treatment effect estimates are sensitive to parameter values and model specification. Alternative model specifications—including changes to parameter values in data generation models—may result in different results for magnitude of bias.

The empirical example performed presents special limitations because it is not possible to know with certainty the true ATE or true LATE values from observational data. Previous research, as well as the simulation results of this research, suggests that 2SLS generates consistent estimates of LATE with minimal assumptions that are fundamental to all IV estimators. As such, we use the 2SLS estimate as a gold standard in the empirical scenario. However, if assumptions of essential heterogeneity in ACE/ARB use for AMI patients are correct, then no method generates an estimate of the ATE and there is no reliable gold standard for comparison. Additionally, it is not possible to directly identify the marginal individuals whose treatment choices were influenced by the instruments in the observational data. While average derivative estimates of LATE did not differ in any meaningful way across alternative methods used to approximate the subpopulation of marginal individuals, the methods used here are untested and no known research has demonstrated a valid method for identifying the marginal population from observational data.

Though not the immediate focus of this research, it is important to restate the added strong assumptions and potential dangers of using a nonlinear first-stage treatment choice model when employing nonlinear IV methods. Consistency of second-stage treatment effect estimates is robust to a linear first-stage treatment choice model, even if the relationships modeled in the first-stage are not linear in truth.[12] This robustness property does not hold for models with nonlinear first-stage regressions.[12] Given the results of the empirical example for ACE/ARB use on survival, absolute treatment effect estimates generated using nonlinear 2SRI appear to be especially sensitive to nonlinear misspecification of the first-stage. Estimates generated using a probit-probit 2SRI model were opposite in direction of effect, compared to all other estimators. Estimates generated using probit-probit 2SPS models were not significantly different from 2SLS or linear-probit model estimates. However, despite these potential risks, researchers employing 2SRI methods for models with "inherently nonlinear dependent variables" commonly opt for a nonlinear regression in the first stage treatment choice model.[5,21,23,25,26,33,34,62,80] If estimates generated from models with nonlinear first-stage equations differ significantly from estimates generated by linear models, it may be likely that the nonlinear models are generating inconsistent estimates.[12,47] Incomplete understanding of this could easily lead too improper inferences. For example, estimating the effect on informal care receipt on formal care use, Bonsang et al. (2009) report estimates from 2SLS suggesting a statistically significant positive effects while nonlinear 2SRI—with a nonlinear first stage model—estimates suggested statistically significant negative effects.[39] Bonsang and colleagues disregard the 2SLS estimate as inconsistent based upon the assertions of Terza et al. (2008)—stating that "the standard approach of instrumental variables estimation used in linear models provides inconsistent estimates when applied to nonlinear models".[39] However, this is not completely accurate. As shown in our simulations, 2SLS yields consistent estimates of LATE. Nonlinear 2SRI estimates, on the other hand, lean heavily on assumptions that are not discussed by

Bonsang and colleagues. Given these added strong assumptions, it may be likely that the significant negative effect estimated by nonlinear 2SRI models is inconsistent and should not be used to inform policy-makers.

It is my hope that this research highlights the necessity for researchers to consider concepts of treatment effect heterogeneity and "sorting-on-the-gain" when interpreting regression estimates. IV methods yield estimates of LATE that are interpretable for only a specific subset of the population. Generalization of these estimates requires that the researcher justifies additional strong assumptions about the nature of treatment effect heterogeneity and circumstances underlying treatment choices. These assumptions are integral to making responsible inferences from treatment effect estimates using observational data and should be stated and defended through theory laid out by empirical researchers up front. If these assumptions cannot be reasonably accepted, then researchers must consider that LATE is itself may be an interesting parameter and, while ATE may often be more "palatable" because of its nature as a straightforward population average, population average treatment effects may often be less informative or useful than LATE for clinicians and policy-makers.

The pursuit of innovative statistical methodologies that enable researchers to produce consistent estimates of alternative population and patient-centered treatment effects using observational data is a laudable goal and certainly something to continue striving for. The availability and richness of observational data will only increase as we technology advances and we move forward in the age of electronic medicine, surveillance, and record keeping. However, as implementation of advanced regression methods becomes increasingly easy through convenient statistical packages, researchers must be careful to avoid using these methods simply for the sake of using them. Without understanding and careful thought of the added strong assumptions and potential limitations on inferences that can be made from treatment effect estimates generated by these estimators, it can be far too easy to make potentially inappropriate inferences that may ultimately misguide patients, clinicians, or policy-makers.

APPENDIX

COVARIATE DEFINITIONS AND ADDITIONAL RESULTS

Table A-1: Coding scheme for study covariates

Variable	Definition/Source	Source of Data	Values
Patient demographics			
Age	Categories will be: 66-70, 71-75, 76-80, 81-85, Over 85	Beneficiary Summary Files (2008)	1 if in age category, 0 otherwise
Gender	Categories will be M, F	Beneficiary Summary Files (2008)	1 if male, 0 female
Race	Categories will be Unknown, White, Black, Other, Asian, Hispanic, Native American	Beneficiary Summary Files (2008)	1 if in race category, 0 otherwise
Metro/Non-Metro Geographic Location	Will use Rural-Urban Continuum Codes (1-9) developed by the USDA Economic Research Service (ERS): summarized to metro (1-3) and non-metro (4-9) areas.	Beneficiary Summary Files (2008)	1 if in location category, 0 otherwise
Average Life Expectancy	Average life expectancy in quartiles by zip code[98]	U.S. Census, NCHS	1 if in each category, 0 otherwise
Dual Eligible Status	Dual eligibility status at the time of index AMI and recent change in dual eligibility status	Beneficiary Summary Files (2008)	1 if dual eligible in AMI month, 0 otherwise; 1 if dual eligibility status changed within last 12 months

Table A-1 (continued)

Low Income Subsidy	Whether individuals received a low income subsidy available under Medicare part D prescription drug program.	Beneficiary Summary Files (2008)	1 if individuals received a low-income subside; 0 otherwise
Area Characteristics	Median percent of: residents living in poverty, non-English speakers, immigrant residents, low income residents, residents who did not complete high school by zip code	2000 U.S. Census	Beneficiaries in above-median zip codes received a "1" for that variable, and "0" otherwise.
Baseline medical history/comorbidities			
Time Period	Two sets of variables will measure: (1) up to 365 days pre admission and (2) during admission		
AMI (Acute Myocardial Infarction)	At least 1 inpatient claim with ICD-9 codes 410.01, 410.11, 410.21, 410.31, 410.41, 410.51, 410.61, 410.71, 410.81, 410.91 (ONLY first or second ICD-9 on the claim) [99]	Part A (Inpatient)	1 if code occurs, 0 otherwise
Stroke	At least 1 inpatient claim or 2 HOP or Carrier claims with ICD-9 codes 430, 431, 434.00, 434.01, 434.10, 434.11, 434.90, 434.91, 435.0, 435.1, 435.3, 435.8, 435.9, 436, 997.02 (any ICD-9 on the claim); If any of the qualifying claims have: 800 <= ICD-9 Code <= 804.9, 850 <= ICD-9 Code <= 854.1 in any ICD-9 position or ICD-9 V57xx as the principal ICD-9 code, then EXCLUDE. [99] Exclude 435.X per Mini-Sentinel (2011) [100]	Part A (Inpatient) Part B (outpatient, Carrier)	1 if code occurs, 0 otherwise
TIA (Transient Ischemic Attack)	ICD-9 codes 435.x in hospitalization or emergency encounter data [100]	Part A/B	1 if code occurs, 0 otherwise

Table A-1 (continued)

Heart Failure	ICD-9 398.91, 402.01, 402.11, 402.91, 404.01, 404.11, 404.91, 404.03, 404.13, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9 (Any ICD-9 on the claim); At least 1 inpatient, HOP or Carrier claim with ICD-9 codes [99]	Part A (Inpatient) Part B (outpatient, Carrier)	1 if code occurs, 0 otherwise
Atrial Fibrillation	ICD-9 427.31 (ONLY first or second ICD-9 on the claim); At least 1 inpatient claim or 2 HOP or Carrier claims with ICD-9 code[99]	Part A (Inpatient) Part B (outpatient, Carrier)	1 if code occurs, 0 otherwise
CKD (Chronic Kidney Disease)	ICD-9 016.00, 016.01, 016.02, 016.03, 016.04, 016.05, 016.06, 095.4, 189.0, 189.9, 223.0, 236.91, 249.40, 249.41, 250.40, 250.41, 250.42, 250.43, 271.4, 274.10, 283.11, 403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 440.1, 442.1, 572.4, 580.0, 580.4, 580.81, 580.89, 580.9, 581.0, 581.1, 581.2, 581.3, 581.81, 581.89, 581.9, 582.0, 582.1, 582.2, 582.4, 582.81, 582.89, 582.9, 583.0, 583.1, 583.2, 583.4, 583.6, 583.7, 583.81, 583.89, 583.9, 584.5, 584.6, 584.7, 584.8, 584.9, 585, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9, 586, 587, 588.0, 588.1, 588.81, 588.89, 588.9, 591, 753.12, 753.13, 753.14, 753.15, 753.16, 753.17, 753.19, 753.20, 753.21, 753.22, 753.23, 753.29, 794.4 (any ICD-9 on the claim); At least 1 inpatient, SNF or HHA claim or 2 HOP or Carrier claims with ICD-9 codes[99]	Part A (Inpatient, SNF, HHA) Part B (outpatient, Carrier)	1 if code occurs, 0 otherwise
Diabetes	ICD-9 249.00, 249.01, 249.10, 249.11, 249.20, 249.21, 249.30, 249.31, 249.40, 249.41, 249.50, 249.51, 249.60, 249.61, 249.70, 249.71, 249.80, 249.81, 249.90, 249.91, 250.00, 250.01, 250.02, 250.03, 250.10, 250.11, 250.12, 250.13, 250.20, 250.21, 250.22, 250.23, 250.30, 250.31, 250.32, 250.33, 250.40, 250.41, 250.42, 250.43, 250.50, 250.51, 250.52, 250.53, 250.60, 250.61, 250.62, 250.63, 250.70, 250.71, 250.72, 250.73, 250.80, 250.81, 250.82, 250.83, 250.90, 250.91, 250.92, 250.93, 357.2, 362.01, 362.02, 366.41 (any ICD-9 on the claim); At least 1 inpatient, SNF or HHA claim or 2 HOP or Carrier claims with ICD-9 code [99] AND add a dummy for insulin use from pharmacy claims	Part A (Inpatient, SNF, HHA) Part B (outpatient, Carrier) Part D	1 if code occurs, 0 otherwise

Table A-1 (continued)

Hypertension (complicated)	ICD-9 402.10, 402.90, 404.10, 404.90, 405.1, 405.9[101]	Part A/B	1 if code occurs, 0 otherwise
Hypertension (uncomplicated)	ICD-9 401.1, 401.9 [101]	Part A/B	1 if code occurs, 0 otherwise
Hyperlipidemia	ICD-9 272.xx OR a pharmacy claim for a lipid lowering medication[102]	Part A/B/D	1 if code occurs, 0 otherwise
COPD	ICD-9 491.0, 491.1, 491.20, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 494.0, 494.1, 496 (any ICD-9 on the claim)[99]	Part A/B	1 if code occurs, 0 otherwise
Cancer (general)	ICD-9 140.x-172.x, 174.x-195.8, 200.x-208.x[103]	Part A/B	1 if code occurs, 0 otherwise
Metastatic Cancer	ICD-9 196.x-199.x[101]	Part A/B	1 if code occurs, 0 otherwise
Asthma	ICD-9 493.0, 493.1, 493.9[104]	Part A/B	1 if code occurs, 0 otherwise

Table A-1 (continued)

Non-AMI Ischemic Heart Disease (AMI codes removed)	ICD-9 410.00, 410.02, 410.10, 410.12, 410.20, 410.22, 410.30, 410.32, 410.40, 410.42, 410.50, 410.52, 410.60, 410.62, 410.70, 410.72, 410.80, 410.82, 410.90, 410.92, 411.0, 411.1, 411.81, 411.89, 412, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.10, 414.11, 414.12, 414.19, 414.2, 414.3, 414.8, 414.9 PROC 00.66, 36.01, 36.02, 36.03, 36.04, 36.05, 36.06, 36.07, 36.09, 36.10, 36.11, 36.12, 36.13, 36.14, 36.15, 36.16, 36.17, 36.19, 36.2, 36.31, 36.32 HCPCS 33510, 33511, 33512, 33513, 33514, 33516, 33517, 33518, 33519, 33521, 33522, 33523, 33533, 33534, 33535, 33536, 33542, 33545, 33548, 92975, 92977, 92980, 92982, 92995, 33140, 33141 (any ICD-9, PROC or HCPCS on the claim); At least 1 inpatient, SNF, HHA, HOP or Carrier claim with ICD-9, Procedure or HCPC codes[99]	Part A (Inpatient, SNF, HHA) Part B (outpatient, Carrier)	1 if code occurs, 0 otherwise
Unstable Angina	ICD-9 411.xx	Part A/B	1 if code occurs, 0 otherwise
Angioedema	ICD-9 995.1x	Part A/B	1 if code occurs, 0 otherwise
Depression	311.xx, 298.0x, 300.4x, 309.1x; 296.20-296.26; 296.30-296.36; 296.50-296.56; 296.60-296.66; 296.89 Exclude if BETOS code not in D1A, D1B, D1C, D1D, D1E, D1F, D1G, O1A	Part A/B	1 if code occurs, 0 otherwise
CABG	ICD-9 Procedure codes 36.0-36.39 or CPT 33510-33536[105,106]	Part A/B	1 if code occurs, 0 otherwise
Stent	Stent alone CPT codes 92980-92981[106] ICD-9-CM procedure code 36.06,drug-eluting stent (ICD-9-CM procedure code 36.07), or both[107]	Part A/B	1 if code occurs, 0 otherwise

Table A-1 (continued)

Pacemaker Implantation	ICD-9 procedure codes 37.80-37.89[108]	Part A/B	1 if code occurs, 0 otherwise
PTCA	Angioplasty only: ICD-9 procedure codes 36.01, 36.02, 36.05, 36.06[105]	Part A/B	1 if code occurs, 0 otherwise
Charlson Comorbidity Index	Comorbidity Score from 0 to 19 possible points[103]	Part A/B	0-19 points possible
Serious Myopathy	A primary or secondary discharge code for myoglobinuria((ICD-9-CM 791.3), OR a primary code for "other disorders of muscle," (ICD-9-CM 728.89, 729.1, 359.4,359.8, 359.9, 710.4, 728.9, 729.8X, E942.2) OR a secondary code for any of the above codes ”accompanied by a claim for a CK test within 7 days of hospitalization or a discharge code for acute renal failure (CPT codes 82550, 82552, 82554, 80012, 80016, 80018, or 80019).[109]	Part A/B	1 if code occurs, 0 otherwise
Non-serious Myopathy	All codes above, but no claim type or position restrictions, and no CK test required.	Part A/B	1 if code occurs, 0 otherwise
Bradycardia	ICD-9 code 427.8[110]	Part A/B	1 if code occurs, 0 otherwise
Heart Block	ICD-9 code 426.x[110]	Part A/B	1 if code occurs, 0 otherwise
Hyperkalemia	ICD-9 code 276.7[111]	Part A/B	1 if code occurs, 0 otherwise

Table A-1 (continued)

Hepatic Events	Acute/sub-acute necrosis of liver (570.xx) or hepatitis(573.3x) or other disorders of liver (573.8x, 573.9x)[112]		
Cardiac Arrest	ICD-9 427.5[113]	Part A/B	1 if code occurs, 0 otherwise
Ventricular Arrhythmia	ICD-9 427.1x, 427.4x , 427.41, 427.42[114]	Part A/B	1 if code occurs, 0 otherwise
Other Arrhythmia	ICD-9 427.xx, 798.xx	Part A/B	1 if code occurs, 0 otherwise
Cardiogenic Shock	ICD-9 785.51[113]	Part A/B	1 if code occurs, 0 otherwise
Hypotension	ICD-9 458.x[110] or ICD-9 458.0[115] and ICD-9 458.9, 785.5x, and 998.0[116]	Part A/B	1 if code occurs, 0 otherwise
Renal Failure	acute renal failure/acute tubular necrosis (ICD-9 584.xx) or acute glomerulonephritis (ICD-9 580.xx)[112,117]	Part A/B	1 if code occurs, 0 otherwise
Baseline and Post-Discharge medications			
Time Period	180 days prior to index admission date AND 30 days post discharge for ACE/ARBs and beta blockers only		
Drug Classes	All drug classes identified by linking Part D claims to Multum Lexicon Plus dataset (Copyright 2012 Lexi-Comp, Inc. and/or Cenner Multum, Inc)		

Table A-1 (continued)

Nitrates		Part D	1 if prescription filled during time period, 0 otherwise
Clopidogrel		Part D	
ACE Inhibitors		Part D	
ARBs		Part D	
Beta Blockers		Part D	
Lipid-lowing Agents		Part D	
Calcium-channel Blockers		Part D	
Low molecular weight Heparin		Part D	
Warfarin		Part D	
Diuretics (loop diuretics, thiazide		Part D	

Table A-1 (continued)

Other antihypertensive		Part D	
Fenofibrates and other lipid-lowering agents		Part D	
Insulin		Part D	
Other anti-diabetics	Alpha-glucosidase, amylin analogues, biguanides, dipeptidylpeptidase 4 inhibitors, glucagon-like peptide 1 receptor agonists, meglitinides, sulfonylureas, thiazolidinediones, others (epalrestat, exenatide, glybuzole)	Part D	
Diagnosis on admission			
Anterior wall AMI	ICD-9 410.0 410.1[118]	Part A/B	1 if code occurs, 0 otherwise

Table A-1 (continued)

Subendocardial Infarction (NSTEMI)	ICD-9 410.7x[118]	Part A/B	1 if code occurs, 0 otherwise
Other AMI locations	Other AMI locations ICD-9 410.5-410.6, 410.8-410.9[118]	Part A/B	1 if code occurs, 0 otherwise
Complications during admission			
Sepsis	ICD-9 995.91[100]	Part A/B	1 if code occurs, 0 otherwise
Pneumonia	Secondary ICD-9 481 to 483[119]	Part A/B	1 if code occurs, 0 otherwise
Procedures during hospitalization			
Stress Test	ICD-9 89.4x, CPT 93015 or CPT codes 93015-93018, 93350, 78460-78465, 78472-78483, 78494, 78496, 78491-78492 (includes nuclear imaging)[106]	Part A/B	1 if code occurs, 0 otherwise
Cardiac Catheterization	CPT codes 93508, 93510-93529, 93539-93540, 93543, 93545-93552[106]	Part A/B	1 if code occurs, 0 otherwise
Echocardiography	On inpatient claim, ICD-9 procedure code 88.72; on outpatient claim, CPT 93307, 93320, 93325, 93308[120]	Part A (Inpatient) Part B (outpatient)	1 if code occurs, 0 otherwise

Table A-1 (continued)

Insurance variables			
Benefit Phase	Will indicate whether patient was in deductible, "donut hole" or catastrophic phase at index admission	Part D Event Data	1 if code occurs, 0 otherwise
Plan Premium Quartile	Plan premium rates by beneficiary separated into quartiles	Pharmacy characteristics file	1 if in each category, 0 otherwise
Cumulative Beneficiary Responsibility Amount	Cumulative beneficiary responsibility at index admission amount by beneficiary separated into quartiles	PDE	1 if in each category, 0 otherwise
Cumulative Total Cost	Cumulative total cost at admission by beneficiary separated into quartiles	PDE	1 if in each category, 0 otherwise
Utilization variables			
Days in ICU	Number of days in ICU as measured by occurrence of revenue center code 0201	Part A	Number
Days in CCU	Number of days in CCU as measured by occurrence of revenue center code 0210	Part A	Number
Days in IMC	Number of days in IMC as measured by occurrence of revenue center code 0206	Part A	Number
Other Acute Days	Number of other acute days as measured by remainder of days in acute LOS	Part A	Number

Table A-1 (continued)

Other non-acute Institutional Days	Number of other institutional days as measured by remainder of all other days in LOS	Part A	Number
ER Use	Occurrence of revenue center code 0450	Part A	1 if used, 0 otherwise
Transferred to another facility	Occurrence of multiple short term or CAH hospitals in overall stay	Part A	1 if transferred, 0 otherwise

Table A-2: Comparing Estimators for Regression Estimates of ACE/ARB on Survival (100-person circle ATR Quintiles for Instruments)

	2SLS	2SPS (Linear-Probit)	2SRI (Linear-Probit)	2SPS (Probit-Probit)	2SRI (Probit-Probit)
ATE		-0.002	-0.001	0.009	-0.043
LATE (0.05 band)	0.000	-0.002	-0.001	0.009	-0.042
LATE (at means)	0.000	-0.001	-0.0009	0.009	-0.041
LATE (0.02 band)	0.000	-0.002	-0.001	0.009	-0.043
LATE (0.1 band)	0.000	-0.001	-0.0009	0.009	-0.041
N	63,685				
Under ID	261.80				
(P-Value)	*(p<0.0001)*				
Chow Test[a]	65.85				
(P-Value)	*(p<0.0001)*				

[a] F-Test of excluded instrumental variables (ATR quintiles) in first-stage treatment choice equation

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

Table A-3: Comparing Estimators for Regression Estimates of ACE/ARB on Survival (200-person circle ATR Quintiles for Instruments)

	2SLS	2SPS (Linear-Probit)	2SRI (Linear-Probit)	2SPS (Probit-Probit)	2SRI (Probit-Probit)
ATE		0.010	0.011	0.022	-0.041
LATE (0.05 band)	-0.002	0.010	0.011	0.022	-0.040
LATE (at means)	-0.002	0.010	0.010	0.021	-0.039
LATE (0.02 band)	-0.002	0.011	0.011	0.023	-0.041
LATE (0.1 band)	-0.002	0.010	0.010	0.021	-0.039
N	63,685				
Under ID	162.85				
(P-Value)	*(p<0.0001)*				
Chow Test[a]	40.83				
(P-Value)	*(p<0.0001)*				

[a] F-Test of excluded instrumental variables (ATR quintiles) in first-stage treatment choice equation

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

REFERENCES

1. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica*. 1994;62(2):467-475.

2. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*. 1996;91(434):444-455.

3. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*. 1995;90(430):431-442.

4. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? *JAMA: the journal of the American Medical Association*. 1994;272(11):859-866.

5. Hadley J, Yabroff KR, Barrett MJ, Penson DF, Saigal CS, Potosky AL. Comparative effectiveness of prostate cancer treatments: Evaluating statistical adjustments for confounding in observational data. *J Natl Cancer Inst*. 2010;102(23):1780-1793.

6. Harris KM, Remler DK. Who is the marginal patient? understanding instrumental variables estimates of treatment effects. *Health Serv Res*. 1998;33(5 Pt 1):1337.

7. Brooks JM, Chrischilles EA. Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods. *Med Care*. 2007;45(10):S123-S130.

8. Angrist JD, Pischke J. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press; 2008.

9. Angrist JD. Treatment effect heterogeneity in theory and practice*. *The Economic Journal*. 2004;114(494):C52-C83.

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

11. Brooks JM, McClellan M, Wong HS. The marginal benefits of invasive treatments for acute myocardial infarction: Does insurance coverage matter? *Inquiry: a journal of medical care organization, provision and financing*. 2000;37(1):75.

12. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business & Economic Statistics*. 2001;19(1).

13. Greene WH. *Econometric analysis*. 5th ed. Upper Saddle River, New Jersey, 07458: Pearson Education, Inc.; 2003.

14. Stuart BC, Doshi JA, Terza JV. Assessing the impact of drug use on hospital costs. *Health Serv Res*. 2009;44(1):128-144.

15. Angrist JD, Fernandez-Val I. Extrapolate-ing: External validity and overidentification in the late framework. *NBER Working Paper No. w16566*. December 2010.

16. Brooks JM, Fang G. Interpreting treatment-effect estimates with heterogeneity and choice: Simulation model results. *Clin Ther*. 2009;31(4):902-919.

17. 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;16(11):1133-1157.

18. Terza JV, Bradford WD, Dismuke CE. The use of linear instrumental variables methods in health services research and health economics: A cautionary note. *Health Serv Res*. 2007;43(3):1102-1120.

19. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *J Health Econ*. 2008;27(3):531-543.

20. Bhattacharya J, Goldman D, McCaffrey D. Estimating probit models with self-selected treatments. *Stat Med*. 2006;25(3):389-413.

21. Trogdon JG, Nurmagambetov TA, Thompson HF. The economic implications of influenza vaccination for adults with asthma. *Am J Prev Med*. 2010;39(5):403-410.

22. Shih YT, Tai-Seale M. Physicians' perception of demand-induced supply in the information age: A latent class model analysis. *Health Econ*. 2012;21(3):252-269.

23. Self S, Grabowski R. Female autonomy and health care in developing countries. *Review of Development Economics*. 2012;16(1):185-198.

24. Schreyoegg J, Stargardt T. The trade-off between costs and outcomes: The case of acute myocardial infarction. *Health Serv Res*. 2010;45(6):1585-1601.

25. Prada SI, Salkever D, MacKenzie EJ. Level-I trauma center effects on return-to-work outcomes. *Eval Rev*. 2012;36(2):133-164.

26. McGeary KA, French MT. Illicit drug use and emergency room utilization. *Health Serv Res*. 2000;35(1):153-169.

27. Lindrooth RC, Weisbrod BA. Do religious nonprofit and for-profit organizations respond differently to financial incentives? the hospice industry. *J Health Econ*. 2007;26(2):342-357.

28. Li Y, Cai X, Mukamel DB, Cram P. Impact of length of stay after coronary bypass surgery on short-term readmission rate an instrumental variable analysis. *Med Care*. 2013;51(1):45-51.

29. Li Y, Jensen GA. Effects of drinking on hospital stays and emergency room visits among older adults. *J Aging Health*. 2012;24(1):67-91.

30. Hadley J, Reschovsky JD. Medicare spending, mortality rates, and quality of care. *International Journal of Health Care Finance & Economics*. 2012;12(1):87-105.

31. Grabowski DC, Feng Z, Hirth R, Rahman M, Mor V. Effect of nursing home ownership on the quality of post-acute care: An instrumental variables approach. *J Health Econ*. 2013;32(1):12-21.

32. Gore JL, Litwin MS, Lai J, et al. Use of radical cystectomy for patients with invasive bladder cancer. *J Natl Cancer Inst*. 2010;102(11):802-811.

33. Gibson TB, Mark TL, Axelsen K, Baser O, Rublee DA, McGuigan KA. Impact of statin copayments on adherence and medical care utilization and expenditures. *Am J Manag Care*. 2006;12:SP11-SP19.

34. Gibson TB, Song X, Alemayehu B, et al. Cost sharing, adherence, and health outcomes in patients with diabetes. *Am J Manag Care*. 2010;16(8):589-600.

35. Fang H, Miller NH, Rizzo J, Zeckhauser R. Demanding customers: Consumerist patients and quality of care. *B E Journal of Economic Analysis & Policy*. 2011;11(1):59.

36. Fang H, Rizzo JA. Information-oriented patients and physician career satisfaction: Is there a link? *Health Economics Policy and Law*. 2011;6(3):295-311.

37. Fang H, Rizzo JA. Has the influence of managed care waned? evidence from the market for physician services. *International Journal of Health Care Finance & Economics*. 2010;10(1):85-103.

38. Dranove D, Ramanarayanan S, Sfekas A. Does the market punish aggressive experts? evidence from cesarean sections. *B E Journal of Economic Analysis & Policy*. 2011;11(2):6.

39. Bonsang E. Does informal care from children to their elderly parents substitute for formal care in europe? *J Health Econ*. 2009;28(1):143-154.

40. Black L, Spetz J, Harrington C. Nurses who do not nurse: Factors that predict non-nursing work in the US registered nursing labor market. *Nursing Economics*. 2010;28(4):245-254.

41. Baughman RA, Smith KE. Labor mobility of the direct care workforce: Implications for the provision of long-term care. *Health Econ*. 2012;21(12):1402-1415.

42. Fang H, Ali MM, Rizzo JA. Does smoking affect body weight and obesity in china? *Economics & Human Biology*. 2009;7(3):334-350.

43. Angrist J, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *The Journal of Economic Perspectives*. 2001;15(4):69-85.

44. 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;175(1):60-65.

45. Hausman JA. Specification tests in econometrics. *Econometrica*. 1978;46(6):1251-1271.

46. Garrido MM, Deb P, Burgess JF,Jr., Penrod JD. Choosing models for health care cost analyses: Issues of nonlinearity and endogeneity. *Health Serv Res*. 2012;47(6):2377-2397.

47. Lee M. Semiparametric estimators for limited dependent variable (LDV) models with endogenous regressors. *Econometric Reviews*. 2012;31(2):171-214.

48. Basu A, Jena AB, Goldman DP, Philipson TJ, Dubois R. Heterogeneity in action: The role of passive personalization in comparative effectiveness research. *Health Econ*. 2013.

49. Brooks JM. Improving characterization of study populations: The identification problem. In: *Developing a protocol for observational comparative effectiveness research: A user's guide*. Publication No. 12(13)-EHC099 ed. Rockville, MD: AHRQ; 2013:161.

50. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? instrumental variables evidence for stage II patients from iowa. *Health Serv Res*. 2003;38(6p1):1385-1402.

51. Pfeffer MA, Braunwald E, Moyé LA, et al. Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction: Results of the survival and ventricular enlargement trial. *N Engl J Med*. 1992;327(10):669-677.

52. Antman EM, Hand M, Armstrong PW, et al. 2007 focused update of the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction. *J Am Coll Cardiol*. 2008;51(2):210-247.

53. Dagenais GR, Pogue J, Fox K, Simoons ML, Yusuf S. Angiotensin-converting-enzyme inhibitors in stable vascular disease without left ventricular systolic dysfunction or heart failure: A combined analysis of three trials. *The Lancet*. 2006;368(9535):581-588.

54. Pahor M, Psaty BM, Alderman MH, Applegate WB, Williamson JD, Furberg CD. Therapeutic benefits of ACE inhibitors and other antihypertensive drugs in patients with type 2 diabetes. *Diabetes Care*. 2000;23(7):888-892.

55. Gustafsson I, Torp-Pedersen C, Køber L, Gustafsson F, Hildebrandt P. Effect of the angiotensin-converting enzyme inhibitor trandolapril on mortality and morbidity in diabetic patients with left ventricular dysfunction after acute myocardial infarction. *J Am Coll Cardiol*. 1999;34(1):83-89.

56. Borghi C, Bacchelli S, Esposti DD, Ambrosioni E, SMILE Study. Effects of the early ACE inhibition in diabetic nonthrombolyzed patients with anterior acute myocardial infarction. *Diabetes Care*. 2003;26(6):1862-1868.

57. Grundy SM, Howard B, Smith S,Jr, Eckel R, Redberg R, Bonow RO. Prevention conference VI: Diabetes and cardiovascular disease: Executive summary: Conference proceeding for healthcare professionals from a special writing group of the american heart association. *Circulation*. 2002;105(18):2231-2239.

58. Mukamal KJ, Nesto RW, Cohen MC, et al. Impact of diabetes on long-term survival after acute myocardial infarction: Comparability of risk with prior myocardial infarction. *Diabetes Care*. 2001;24(8):1422-1427.

59. Wooldridge JM. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, Massachusetts: The MIT press; 2002.

60. Varian HR, Repcheck J. *Intermediate microeconomics: A modern approach*. Vol. 8 ed. New York, NY: WW Norton & Company; 2010.

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

62. Terza JV. Alcohol abuse and employment: A second look. *J Appl Econometrics*. 2002;17(4):393-404.

63. Heckman JJ, Robb R. Alternative methods for evaluating the impact of interventions: An overview. *J Econ*. 1985;30(1):239-267.

64. O'Malley AJ, Frank RG, Normand S-T. Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia. *Stat Med*. 2011;30(16):1971-1988.

65. Rubin DB. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*. 1978:34-58.

66. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annu Rev Public Health*. 2000;21(1):121-145.

67. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688.

68. Lee LF. Identification and estimation in binary choice models with limited (censored) dependent variables. *Econometrica*. 1979;47(4):977-996.

69. Sloan FA, Feinglos MN, Grossman DS. Receipt of care and reduction of lower extremity amputations in a nationally representative sample of U.S. elderly. *Health Serv Res*. 2010;45(6):1740-1762.

70. Shin J, Moon S. Do HMO plans reduce health care expenditure in the private sector? *Econ Inq*. 2007;45(1):82-99.

71. Shih YT, Elting LS, Halpern MT. Factors associated with immunotherapy use among newly diagnosed cancer patients. *Med Care*. 2009;47(9):948-958.

72. Pizer SD, Gardner JA. Is fragmented financing bad for your health? *Inquiry-the Journal of Health Care Organization Provision and Financing*. 2011;48(2):109-122.

73. Norton EC, Lindrooth RC, Ennett ST. Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. *Health Econ*. 1998;7(5):439-453.

74. Marcen M, Molina JA. Informal caring-time and caregiver satisfaction. *European Journal of Health Economics*. 2012;13(6):683-705.

75. Lu C, Chin B, Lewandowski JL, et al. Towards universal health coverage: An evaluation of rwanda mutuelles in its first eight years. *Plos One*. 2012;7(6):e39282.

76. Furukawa MF. Electronic medical records and the efficiency of hospital emergency departments. *Medical Care Research and Review*. 2011;68(1):75-95.

77. Newey WK. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *J Econ*. 1987;36(3):231-250.

78. Blundell RW, Smith RJ. Estimation in a class of simultaneous equation limited dependent variable models. *Rev Econ Stud*. 1989;56(1):37-57.

79. Tan H, Norton EC, Ye Z, Hafez KS, Gore JL, Miller DC. Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. *Jama-Journal of the American Medical Association*. 2012;307(15):1629-1635.

80. Kaestner R, Khan N. Medicare part D and its effect on the use of prescription drugs and use of other health care services of the elderly. *Journal of Policy Analysis and Management*. 2012;31(2):253-279.

81. Zimmer DM. Health insurance and health care demand among the self-employed. *J Labor Res*. 2010;31(1):1-19.

82. StataCorp LP. Stata statistical software: Release 12. College Station, TX. 2011.

83. Reeder GS. Angiotensin converting enzyme inhibitors and receptor blockers in acute myocardial infarction: Clinical trials. *UpToDate*. 2014.

84. Reeder GS. Angiotensin converting enzyme inhibitors and receptor blockers in acute myocardial infarction: Recommendations for use. *UpToDate*. 2014.

85. Setoguchi S, Glynn RJ, Avorn J, Mittleman MA, Levin R, Winkelmayer WC. Improvements in long-term mortality after myocardial infarction and increased use of cardiovascular drugs after DischargeA 10-year trend analysis. *J Am Coll Cardiol*. 2008;51(13):1247-1254.

86. Jencks SF, Huff ED, Cuerdon T. Change in the quality of care delivered to medicare beneficiaries, 1998-1999 to 2000-2001. *JAMA*. 2003;289(3):305-312.

87. Brooks JM, Tang Y, Chapman CG, Cook EA, Chrischilles EA. What is the effect of area size when using local area practice style as an instrument? *J Clin Epidemiol*. 2013;66(8):S69-S83.

88. Niskanen L, Hedner T, Hansson L, Lanke J, Niklason A, CAPPP Study Group. Reduced cardiovascular morbidity and mortality in hypertensive diabetic patients on first-line therapy with an ACE inhibitor compared with a diuretic/beta-blocker-based treatment regimen: A subanalysis of the captopril prevention project. *Diabetes Care*. 2001;24(12):2091-2096.

89. Hunt SA, Abraham WT, Chin MH, et al. ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the Adult—Summary article A report of the american college. *Circulation*. 2005;112(12):1825-1852.

90. Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. *Circulation*. 2008;117(8):1028-1036.

91. Bernheim SM, Spertus JA, Reid KJ, et al. Socioeconomic disparities in outcomes after acute myocardial infarction. *Am Heart J*. 2007;153(2):313-319.

92. Shen JJ, Wan TT, Perlin JB. An exploration of the complex relationship of socioecologic factors in the treatment and outcomes of acute myocardial infarction in disadvantaged populations. *Health Serv Res*. 2001;36(4):711-732.

93. Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of medicare claims-based diagnosis of acute myocardial infarction: Estimating positive predictive value on the basis of review of hospital records. *Am Heart J*. 2004;148(1):99-104.

94. 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;48(8):710-717.

95. Fang G, Brooks JM, Chrischilles EA. Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication. *Am J Epidemiol*. 2012;175(11):1142-1151.

96. Chow GC. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*. 1960;28(3):591-605.

97. Boyd C, McNabney M, Brandt N, et al. Guiding principles for the care of older adults with multimorbidity: An approach for clinicians: American geriatrics society expert panel on the care of older adults with multimorbidity. *J Am Geriatr Soc*. 2012;60(10).

98. Murray CJ, Kulkarni SC, Michaud C, et al. Eight americas: Investigating mortality disparities across races, counties, and race-counties in the united states. *PLoS Medicine*. 2006;3(9):e260.

99. Chronic Condition Warehouse. 27 chronic condition algorithms. www.ccwdata.org. Updated 2011.

100. Various. Mini-sentinel systematic evaluation of health outcome of interest definitions for studies using administrative data. . 2011.

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

102. Lund BC, Perry PJ, Brooks JM, Arndt S. Clozapine use in patients with schizophrenia and the risk of diabetes, hyperlipidemia, and hypertension: A claims-based approach. *Arch Gen Psychiatry*. 2001;58(12):1172-1176.

103. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. *J Clin Epidemiol*. 2000;53(12):1258-1267.

104. Blais L, Lemière C, Menzies D, Berbiche D. Validity of asthma diagnoses recorded in the medical services database of quebec. *Pharmacoepidemiol Drug Saf*. 2006;15(4):245-252.

105. Glance LG, Dick AW, Osler TM, Mukamel DB. Accuracy of hospital report cards based on administrative data. *Health Serv Res*. 2006;41(4p1):1413-1437.

106. 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;113(3):374-379.

107. 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;299(24):2868-2876.

108. Brown DW, Croft JB, Giles WH, Anda RF, Mensah GA. Epidemiology of pacemaker procedures among medicare enrollees in 1990, 1995, and 2000. *Am J Cardiol*. 2005;95(3):409-411.

109. Andrade SE, Graham DJ, Staffa JA, et al. Health plan administrative databases can efficiently identify serious myopathy and rhabdomyolysis. *J Clin Epidemiol*. 2005;58(2):171-174.

110. Rochon PA, Anderson GM, Tu JV, et al. Use of beta-blocker therapy in older patients after acute myocardial infarction in ontario. *CMAJ*. 1999;161(11):1403-1408.

111. Juurlink DN, Mamdani MM, Lee DS, et al. Rates of hyperkalemia after publication of the randomized aldactone evaluation study. *N Engl J Med*. 2004;351(6):543-551.

112. Cziraky MJ, Willey VJ, McKenney JM, et al. Statin safety: An assessment using an administrative claims database. *Am J Cardiol*. 2006;97(8):S61-S68.

113. Aujesky D, Obrosky DS, Stone RA, et al. A prediction rule to identify low-risk patients with pulmonary embolism. *Arch Intern Med*. 2006;166(2):169-175.

114. Liperoti R, Gambassi G, Lapane KL, et al. Conventional and atypical antipsychotics and the risk of hospitalization for ventricular arrhythmias or cardiac arrest. *Arch Intern Med*. 2005;165(6):696-701.

115. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol*. 2004;57(2):131-141.

116. Romano PS, Schembri ME, Rainwater JA. Can administrative data be used to ascertain clinically significant postoperative complications? *Am J Med Qual*. 2002;17(4):145-154.

117. Griffin MR, Yared A, Ray WA. Nonsteroidal antiinflammatory drugs and acute renal failure in elderly persons. *Am J Epidemiol*. 2000;151(5):488-496.

118. Fang J, Mensah GA, Alderman MH, Croft JB. Trends in acute myocardial infarction complicated by cardiogenic shock, 1979-2003, united states. *Am Heart J*. 2006;152(6):1035-1041.

119. Rello J, Ollendorf DA, Oster G, et al. Epidemiology and outcomes of ventilator-associated pneumonia in a large US database. *CHEST Journal*. 2002;122(6):2115-2121.

120. Okrah K, Vaughan-Sarrazin M, Cram P. Trends in echocardiography utilization in the veterans administration healthcare system. *Am Heart J*. 2010;159(3):477-483.