Supplement 1. Improving Characterization of Study Populations: The Identification Problem

John M. Brooks, Ph.D.
University of Iowa College of Pharmacy, Iowa City, IA

Abstract

The identification process is an a priori assessment of the treatment effect estimates that can be produced by a given research design, and of the assumptions required for these estimates to yield accurate assessments of a given CER objective. This supplement describes the factors that a researcher should consider when proposing a research design to address (or "identify") a given CER research objective. Investigators should assess the characteristics of the patient sample relative to the study objective, identify the subset(s) of patients whose treatment variation is exploited by the research design, and identify the assumptions that are required to ensure that (1) the research design produces unbiased treatment effect estimates for the patient subsets, and that (2) the treatment effect estimates produced provide a valid assessment of the study objective. In short, investigators must ensure that the effect estimates produced by a given research design and analysis answer the research question of interest and are interpreted appropriately. This supplement concludes with a checklist of guidance and key considerations for identifying research objectives for observational CER protocols.

Introduction

Comparative effectiveness research (CER) is defined by the Federal Coordinating Council for Comparative Effectiveness Research as "the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in 'real world' settings." As such, in its most basic sense, CER requires treatment variation across patients in the real world in order to estimate the comparative effects of alternative treatments. The *identification* process is an a priori assessment of the treatment effect estimates that can be produced by a given research design, and of the assumptions required for these estimates to yield accurate assessments of a given CER objective. Identification has been a key component in econometrics since being introduced by Koopmans in 1949,[1] and a formal definition can be found in the textbook by Cameron and Trivedi.[2] Economist Charles Manski states that "studies of identification seek to characterize the conclusions that could be drawn if one could use a sampling process to obtain an unlimited number of observations."[3] Or, as described by Peter Kennedy, "identification is knowing that something is what you say it is."[4]

CER researchers should provide a thorough discussion of the circumstances in which treatment variation isolated within their research designs is sufficient to make inferences relative to their specific CER objective. Part of this discussion will necessarily deal with sample size issues and statistical inference for the parameters estimated. However, at a more basic level, researchers should describe circumstances under which the parameters estimated can actually *identify* their CER research objective. The next section provides background on the importance of identification in CER relative to various possible CER research objectives and introduces the issues that a researcher should consider when assessing whether a proposed research design identifies a given CER research objective. The background section is followed by sections that focus on each issue.

Background

In the traditional CER model in which investigators compare the effectiveness of a treatment (T) versus an alternative (A) for a set of *clinically similar* patients in the real world, specific CER objectives can include assessments of any of the following:

1. The effect of removing access to T (currently used universally) and switching all patients to A.
2. The effect of T relative to A for those patients that used T; for example, T is currently used by a subset of patients and a policy is considered to remove patients' access to T.
3. The effect of T relative to A for those patients that used A; for example, T is currently used by a subset of patients and a policy is considered to switch all users of A to T.
4. The effect of a change in the T utilization rate (which thereby changes the A rate); for example, T is currently used by a subset of patients and the effects of a general change in T utilization rates are considered.
5. The effect of a change in the T utilization rate (which thereby changes the A rate) that results from a given behavioral or policy intervention; for example, T is currently used by a subset of patients and the effects of a T rate change resulting from a copayment change are considered.
6. The effect of any of the above for specific subpopulations of the set of clinically similar patients; for example, T is currently used by a subset of patients over age 75, and the effects of a T utilization rate change that could result from a copayment change for these patients are considered.

Objective 1 involves finding the average treatment effect estimate across the entire population of clinically similar patients. For example, T could be a treatment used currently by all patients and a more expensive alternative has become available. A CER objective could be to evaluate a policy to switch all patients from T to the new alternative.

Objective 2 requires finding the average effect of T relative to A for the subset of patients who were treated with T. For example, if T is currently used by a subset of patients, a CER objective could be to evaluate a policy to remove patient access to T, which only will affect the subset of patients using T.

Alternatively, objective 3 requires finding the average effect of T relative to A for the subset of patients who were treated with A. For example, if T is not used currently by a subset of patients, a CER objective could be to estimate a policy of expanding T usage to all patients.

Objective 4 relates to evaluating the effects of treatment rate changes. Often the relevant question for policymakers is not whether a treatment should be used at all, but whether a treatment is over- or underused in practice. Many years ago, John Wennberg correctly posed objective 4 with the question "Which rate is right?"[5] For example, if 80 percent of patients use a beta blocker after acute myocardial infarction, a CER objective may be to assess the effect of increasing the beta blocker treatment rate to 85 percent. Objective 4 is equivalent to objectives 2 and 3 if the specified T rate change means moving from the existing T utilization to either zero or 100 percent, respectively. Note that objective 4 is defined purposely *without* describing how the T treatment rates would be changed and can perhaps be best conceptualized as the effect of rate changes over time as a new treatment diffuses across a clinically similar population. The patient subset within a clinically similar population that only receives T when it is fully diffused may differ from the patient subset that is apt to receive T when it is newly introduced.

In contrast, objective 5 is defined with respect to the patient subset whose choice of T relative to A can be modified with a specific behavioral or policy intervention. At a specific T utilization rate, the patients defined in objective 5 can be thought of as a subset of the patients defined in objective 4, except that distinct patients may be affected by distinct interventions.[6] For example, an information-based intervention may affect a different patient subset from an intervention focused on increasing access to treatment or an

intervention to change copayment rates. Objective 6 applies to any of the first four objectives with respect to defined subsets of the original clinically similar group (e.g., males vs. females, young vs. old, insured vs. uninsured).

The importance of identification with respect to these various CER objectives is highlighted when one reviews a seminal instrumental variable (IV) study in health care.[7] In an examination of the mortality risk associated with more intensive treatment for acute myocardial infarction (AMI) in the elderly, McClellan and colleagues focused on the ability of IV estimators to reduce confounding bias in observational health care studies. While their study produced IV estimates that suggested that surgical interventions for AMI did *not* lessen patient mortality risk, the authors provided the qualification that their IV estimates should be used as evidence of mortality changes only if population surgery rates were modified (objective 4).[7] Their estimates did not provide evidence of the average benefit of surgery for those that received surgery (objective 2), the average benefit of surgery over all AMI patients (objective 1), or the average benefit of surgery for all those patients not receiving surgery (objective 3). Without a discussion of the patient subset whose surgery effects were *identified* by these IV estimates, their results could have misled decisionmakers. Other authors who have compared treatment effect estimates across estimators using observational data have demonstrated comparisons that lack context without a discussion of the treatment effect concepts identified by each estimator.[8-10]

The concept of identification is closely akin to the ideas of external validity or applicability, in that it asks researchers to address the question "*For whom* can the treatment effect estimates be generalized?"[3, 11-13] However, the classic discussions of these concepts mainly focus on the extent to which estimates from randomized studies can be appropriately applied to patients dissimilar to study populations.[11-13] Alternatively, assessment of real-world treatment effectiveness in CER will often rely on treatment variation generated by the real-world treatment choices found in observational databases. Identification takes a broader view and relates to the extent of inferences that can be made using estimates from various estimators in the context of real-world treatment decisionmaking.

To make a case that a research design has the ability to identify a parameter sufficient to assess a specific CER objective, researchers should describe: (1) the characteristics of the patient sample used in the research relative to the objective; (2) the subset of patients within the sample whose treatment variation was exploited by the research design; (3) the assumptions required to ensure that the research design produces unbiased average treatment effect estimates for this patient subset; and (4) the assumptions required so that the treatment effect estimates produced will provide a valid assessment of the researcher's CER objective. Each of these issues is discussed further in separate sections below.

To support the reader, Table S1.1 provides a summary of key concepts and acronyms used throughout the sections below.

Table S1.1. Definitions of key concepts relevant to the identification process

Concept	Definition
Identification process	An a priori assessment of the treatment effect estimates that can be produced by a given research design. This process involves understanding the assumptions required for estimates to yield accurate assessments of the research question of interest.
On the "support"	A research objective is described as being on the "support" of a research database if the patient population of interest is included in the database.
Instrumental Variable (IV)	A variable that strongly predicts exposure but is neither directly nor indirectly related to the outcome. Instrumental variable analyses estimate local average treatment effects (LATE).
Risk Adjustment (RA) methods	Methods such as regression-based methods and propensity score–based approaches that produce estimates interpreted as the average treatment effect for the treated (ATT).
Estimator	A rule for calculating a statistic that estimates a population parameter of interest.
Average Treatment Effect across all patients (ATE)	An estimate of the average treatment effect for all patients within a study population.
Average Treatment effect in Treated patients (ATT)	An estimate of the treatment effectiveness for the distinct subset of patients in a study population who were exposed to the treatment under study. Risk adjustment (RA) methods produce these estimates.
Average Treatment effect in Untreated patients (ATU)	An estimate of the treatment effectiveness for the distinct subset of patients in a study population who were not exposed to the treatment under study.
Local Average Treatment Effect (LATE)	An estimate of the average treatment effect for those patients within a study population whose treatment choices were affected by a set of instrumental variables.
Local Average Treatment Effect for patients whose treatment choices were affected by a Policy change (LATE-P)	An estimate of the average treatment effect for those patients within a study population whose treatment choices were affected by a specific policy change.

Properties of the Study Population

At the very foundation of identification, the CER objectives that can be identified using a given research design will be limited by the characteristics of the patients whose data are available for the research. If a CER objective is defined for a patient population with specific characteristics, the objective is described as being on the "support" of the research database if these patients are included in the research database.[3] For example, a research database containing only those patients with fee-for-service insurance limits the ability of researchers to identify average treatment effects for the entire population, patients without insurance, or patients in managed care programs. Likewise, randomized trial designs have limited ability to identify average treatment effects for those patients not studied (i.e., patients not meeting trial inclusion criteria or refusing to participate). If data are retrospectively collected, changes in treatments over time may limit the ability to identify the effectiveness of current treatment choices. This issue is especially relevant when assessing the effectiveness of treatments

whose benefits take many years to observe. For example, 10 years of followup may be necessary to demonstrate survival differences between surgery and radiation treatments for early-stage prostate cancer. However, at the end of the study it may be unclear as to whether the study identified the comparative effectiveness of *current* surgical and radiation technologies.

In the study by McClellan and colleagues cited above, the authors estimated the effectiveness of surgical treatments for AMI for fee-for-service Medicare beneficiaries. It is unclear whether this study identified the effectiveness of surgery for younger AMI patients or those with insurance coverage distinct from Medicare. In a followup IV study using data for younger AMI patients from Washington State, Brooks et al. showed that surgery effectiveness estimates from AMI patients with more generous insurance coverage would understate the effectiveness of surgery for AMI patients with less generous coverage.[14]

Relationship of Estimation Methods to Patient Subsets

Once a research database is specified and the study population is defined by inclusion criteria, the researcher must then make the case that the parameter estimates produced the estimators chosen are sufficient to identify the CER objective. It has been shown that the estimators available to estimate treatment effectiveness produce average estimates for distinct subsets of patients in the study population. Risk adjustment (RA) methods, including regression-based methods and propensity score-based approaches,[15-17] produce estimates that are properly interpreted as the average treatment effect for the treated patients in a population (ATT).[18-22] In contrast, IV estimators yield estimates of an average treatment effect for those patients whose treatment choices were affected by a set of instrumental variables or "instruments."[7, 14, 23-25] Because of this limitation, IV estimates are described as estimates of local average treatment effects (LATE).[25]

If the CER objective is to assess treatment effectiveness for the subset of patients who were treated (objective 2), a risk-adjusted estimate of ATT may be suitable to address this objective. As will be discussed further below, identification would also require the researcher to justify the RA estimator assumptions that are necessary to avoid bias. If the CER objective is to assess average treatment effectiveness for the subset of patients whose treatment choices were modifiable by an instrument (the LATE for that instrument), an IV estimator is appropriate. A LATE estimate is potentially suitable for evaluating objective 5 if the instrument chosen is related to the specified behavioral or policy intervention being evaluated. For example, suppose a CER objective is to estimate the outcome change that will result from a policy of subsidizing treatment T relative to treatment A. An instrument is a measured factor related to treatment choice, but assumed not to have a direct relationship with outcomes or other unmeasured factors related to outcomes. A researcher could use observed variation in relative copayment rates for T and A for patients across distinct insurance plans as the basis for an instrument. The IV estimates produced using this instrument would be the average treatment effects for the subset of patients whose treatment choices are mutable with respect to financial incentives and may be suitable to identify the policy objective. In addition, as with RA estimators, identification with IV estimators requires the researcher to justify the IV assumption set that the instrument does not have a direct relationship with outcomes or other unmeasured factors related to outcomes.

The McClellan AMI study produced both RA estimates using analysis of variance (ANOVA) estimators, and IV estimates using measures of patient geographic access to key providers as instruments. McClellan's RA estimates of ATT showed statistically significant reductions in mortality associated with surgery for Medicare beneficiaries with AMI, whereas their IV LATE estimates showed no mortality reduction from surgery. Conditional on the validity of assumptions underlying each estimator, the RA estimates directly identified a parameter suitable to assess the effects of surgery for those that had surgery (objective 2), whereas the IV LATE estimate identified a parameter potentially suitable to assess objective 5 for a policy related to modification of provider access (e.g., providing greater geographic

dispersion of catheterization labs). This estimate combination suggests that, for the most part, the surgery rates for AMI Medicare patients in the late 1980s through the early 1990s reflected proper sorting of surgery across patients—that the patients who received treatment benefited, but that expanding treatment rates would yield little additional benefit. These estimates *do not* directly identify any other CER objectives without further assumptions.

Assumptions Required To Yield Unbiased Estimates

For RA estimators to yield unbiased estimates of ATT, it must be assumed that unmeasured factors affecting treatment choice are unrelated to outcomes (or are "ignorable") after conditioning on measured factors.[16, 26] Similarly, for IV estimators to yield a consistent estimate of LATE, an instrument must not be directly or indirectly related to outcomes. In the McClellan study, unbiased estimates of ATT from their ANOVA RA estimator rested on the assumption that all unmeasured factors affecting surgery choice had no direct or indirect relationship with mortality. Likewise, for the McClellan study to have produced consistent estimates of LATE, it must be assumed that the instruments used in the study had no direct relationship with mortality and were also unrelated to any remaining unmeasured factors related to both surgery choice and mortality.

Identification of Research Objectives Other Than ATT or LATE

If the CER objective requires estimation of a treatment effect for a patient population not represented in the research database, or if it requires a parameter distinct from ATT or LATE, identification requires the researcher to assess the validity of extrapolating estimates to their CER objective. Extrapolation will require assumptions that should be directly stated and thoughtfully defended based on clinical plausibility and treatment choice theory.

However, if the CER objective is to estimate a treatment effect parameter distinct from ATT or LATE, identification requires that the researcher explicitly provide the assumptions that are necessary for estimates of ATT or LATE to be validly applied to the set of patients described by the research objective. Examples of other treatment effect parameters that may be needed across CER objectives include the average treatment effect on the untreated (ATU) for objective 3, the average treatment effect across all patients in the population (ATE) for objective 1, or the average treatment effect for the subset of patients whose treatment choices would be affected by a specific policy change (LATE-P) for objective 5. Key assumptions to stipulate before extrapolating ATT or LATE estimates to other CER objectives are related to:

- the heterogeneity or homogeneity of treatment effects across patients; and
- the reasons why treated and untreated patients were observed to make different treatment choices.

For example, to assume that an unbiased estimate of ATT is a valid estimate of ATU, a researcher would need to provide a compelling theory as to why the untreated patients did not chose a given treatment for reasons other than expected differences in treatment effectiveness. An unbiased estimate of ATT would provide sufficient information to identify ATU if the researcher can make the case that either: (1) treatment effects are homogeneous across patients and factors unrelated to treatment effectiveness are the cause of disparate treatment choices in the population or (2) treatment effects are heterogeneous across a population but that providers do not react to the treatment effect heterogeneity when making treatment choices. Condition 2 is the notion of "nonessential heterogeneity."[20, 27] Under either of these conditions, it could also be argued that an estimate of ATT identifies the ATE in a population and the average treatment effects that would result from a policy intervention affecting treatment choice (LATE-P). In contrast, if treatment choice was based on expected treatment effectiveness and the patients who were expected to gain most from treatment received treatment (essential

heterogeneity),[27-28] ATT estimates would likely overstate and not identify the true ATE, ATU, and LATE-P in a population. Similar assumptions are required for LATE estimates from a given instrument set to be used to identify ATT, ATU, ATE, and LATE-P. To make the case for the validity of these assumptions, researchers have to provide a theory to suggest why the patients whose treatment choices varied with the value of their instrument are indistinct from the set of patients underlying these parameters.

In the study by McClellan and colleagues, the authors implied that providers considered the effectiveness of surgery for each AMI patient when making surgery recommendations and that the AMI patients most likely to benefit from surgery were those that received surgery. As such, the authors cautioned against assuming their IV estimates could be used to identify ATE, ATU, or ATT. However, the authors suggested that their IV estimates using instruments based on provider access provide more suitable answers to address the question of whether surgery rates from AMI patients should increase (objective 4) in comparison to existing randomized controlled trial (RCT) evidence. Essentially, the authors argued that their IV estimates identified the treatment effect parameter required to assess objective 4.

The Appendix to this supplement contains a general model of treatment choice and outcomes that can be used to clarify the model assumptions required to identify CER objectives using estimates of ATT from RA estimators or estimates of LATE from IV estimators. The general model contains a series of factors related to treatment effectiveness, treatment choice, and outcomes directly. Twelve hypothetical empirical scenarios are derived by assumptions that relate to the existence of these factors. Scenarios differ by whether treatment effects are assumed to be homogeneous or heterogeneous, whether treatment decisions are based on treatment effect heterogeneity, and which of the model factors are measured.

Checklist: Guidance and key considerations for identifying a research objective in an observational CER protocol

Guidance	Key Considerations	Check
Describe the characteristics of the patient sample used in the research relative to the objective.	Is extrapolation required, and what assumptions are needed to support this?	☐
Describe how the estimates from the proposed estimation methods (i.e., RA or IV methods) address the CER objective.	Does the researcher acknowledge to whom the estimates for the method directly apply?	☐
Describe the assumptions required to ensure that the research design produces unbiased average treatment effect estimates for this patient sample.	Does the researcher acknowledge the assumptions required from each estimator to yield unbiased or consistent estimates?	☐
Describe the assumptions required so that the treatment effect estimates produced will provide a valid assessment of the researcher's CER objective.	Does the researcher state whether the clinical and behavioral assumptions necessary for their estimates identify their CER objective?	☐

Appendix: Treatment Choice/ Outcome Model Specifications, Estimators, and Identification

If a researcher is to make inferences on the effects of treatment (T) on outcome (Y) using observational data:

$$E(Y|T+\Delta t) - E(Y|T),$$

a researcher must make assumptions based on the data-generating process for both treatment choice and outcomes, relative to the factors that affect either treatment choice or outcomes. The section below contains a general model that is used to describe the alterative scenarios of CER objective identification. Figure S.1.1 depicts this model. The general model is defined in terms of factors (Xs) with differential relationships between treatment choice (T) and outcome (Y):

1. $Y = g(T(X_1,X_2), X_2, X_3, X_5)$
2. $T = f(X_1,X_2,X_3,X_4)$ where:

X_1 = factors related to treatment effectiveness, have no direct effects on outcome, and may affect treatment choice (perhaps through their effects on treatment effectiveness);

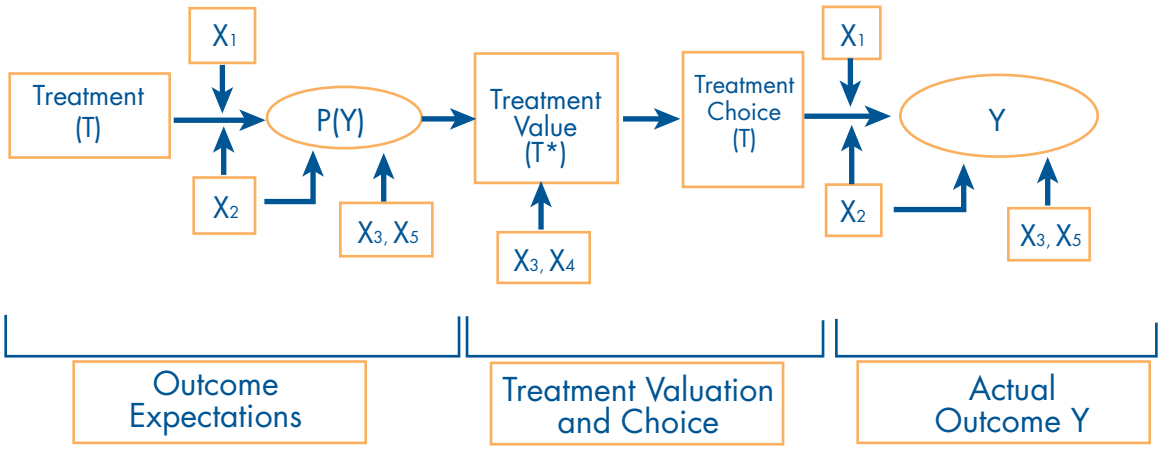
X_2 = factors related to treatment effectiveness, have direct effects on outcome, and may affect treatment choice (perhaps through their effects on treatment effectiveness);

X_3 = factors unrelated to treatment effectiveness, but have direct effects on outcome, and direct effects on treatment valuation;

X_4 = factors having no direct effects on outcome, but have direct effects on treatment valuation; and

X_5 = factors having direct effects on outcome, but do not affect treatment valuation.

Figure S1.1. Model of Treatment Choice and Outcome (adapted from Brooks and Fang[28])*



*This figure is copyrighted by Elsevier Inc. and reprinted with permission.

In a given empirical scenario, the ability to identify and estimate various possible average effects of T on Y (average treatment effect [ATE]; average treatment effect on the treated [ATT]; average treatment effect on the untreated [ATU]; local average treatment effect for a specific instrument [LATE]) is a function of: (1) the assumed relationships between treatment choice and outcomes; (2) which of the factors are measured and unmeasured; and (3) the extent of variation in observed factors. The discussion below details the characteristics required for identification of CER concepts using risk adjustment (RA) and instrumental variable (IV) estimators across variants of this general model. For each factor "X_i", distinctions are made for measured (X_{iM}) and unmeasured (X_{iU}) factors.

Model Scenarios

1. Treatment effect is homogeneous (no X_1 and X_2 factors exist), and no factors affecting treatment choice (T) have a direct effect on outcome (Y).

$Y = g(T,X_{5M},X_{5U})$

$T = f(X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M}:

   i. Sufficient variation in X_4 so that different treatment choices are observed in the data.

   ii. An assumption of no correlation between X_4 and factors in X_{5U} will yield an unbiased estimate of ATT. ATE and ATU are "identified" by this ATT estimate through the assumed homogeneity of treatment effect.

b. IV estimation statistically controlling for X_{5M} and using X_{4M} as an instrument:

   i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data for patients stratified by X_{4M}.

   ii. An assumption of no correlation between X_{4M} and factors in X_{5U} will yield a consistent estimate of LATE specific to the patients whose treatment choices were affected by the factors within X_{4M}. ATE and ATU are "identified" by this LATE estimate through the assumed homogeneity of treatment effect.

2. Treatment effect is homogeneous (no X_1 and X_2 factors exist). Certain factors affecting treatment choice have direct effects on outcome (X_3).

$Y = g(T,X_{3M},X_{3U},X_{5M},X_{5U})$

$T = f(X_{3M},X_{3U},X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{3M} and X_{5M}:

   i. Sufficient variation in X_4 so that different treatment choices are observed in the data after controlling for X_{3M} when estimating the outcome equation.

   ii. Assumptions that no X_{3U} variables exist and that there are no correlations between X_4 and factors in X_{5U} will yield an unbiased estimate of ATT. ATE and ATU are "identified" by the ATT estimate through assumed homogeneity of treatment effect.

b. IV estimation statistically controlling for X_{3M} and X_{5M} and using measured X_{4M} as an instrument:

   i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data across X_{4M} strata after controlling for X_{3M}.

   ii. Assumptions of no correlation between X_{4M} and factors in X_{3U} and X_{5U} will yield a consistent estimate of LATE specific to the set of patients whose T choices were affected by X_{4M} after controlling for X_{3M} and X_{5M}. ATE and ATU are "identified" by this LATE estimate through the assumed homogeneity of treatment effect.

3. Treatment effect is heterogeneous, and the factors affecting treatment effectiveness have no direct effect on Y (X_1 exists; no X_2 factors exist). Moreover, heterogeneity is nonessential: Decisionmakers do not have sufficient knowledge of the X_1 factors affecting heterogeneity to affect treatment choice, and X_1 factors are unmeasured by the researcher.

$Y = g(T(X_{1U}),X_{5M},X_{5U})$

$T = f(X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M}:
   i. Sufficient variation in X_4 so that different treatment choices are observed in the data.
   ii. Assumption of no correlations between X_4 and X_{5U} will yield an unbiased estimate of ATT. ATE and ATU are "identified" by estimating ATT through the assumption that providers do not have knowledge of how X_{1U} relates to treatment effectiveness. However, if X_4 was somehow correlated with X_{1U}, average X_{1U} would vary between treated and untreated patients and the RA estimate of ATT would not be biased; however, it would not be possible to identify either ATE or ATU.

b. IV estimation statistically controlling for X_{5M} and using X_{4M} as an instrument:
   i. Sufficient variation in X_{4M} so different treatment choices are observed in the data across X_{4M} strata.
   ii. Assumption of no correlation between X_{4M} and factors in X_{5U} yields a consistent estimate of LATE specific to the set of patients whose T choices were affected by factors within X_{4M}. ATE and ATU are "identified" by this LATE estimate through the assumption that providers do not have knowledge of how X_{1U} relates to treatment effectiveness. However, if X_{4M} factors are somehow correlated with X_{1U}, then the patients whose treatment choices vary with X_{4M} will differ from the rest of the patient population with regard to X_{1U}. In this case, the IV LATE estimate would not identify either ATE or ATU.

4. Treatment effect is heterogeneous, and factors affecting treatment effectiveness have no direct effect on Y (X_1 exists; no X_2 factors exist). Moreover, heterogeneity is nonessential: Decisionmakers do not have sufficient knowledge of the X_1 factors affecting heterogeneity to affect treatment choice. However, certain suspected X_{1M} factors are measured by the researcher.

   $Y = g(T(X_{1M},X_{1U}),X_{5M},X_{5U})$

   $T = f(X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M} for patient groups stratified by X_{1M}:
   i. Sufficient variation in X_4 exists so that different treatment choices are observed in the data within each stratum of X_{1M}.
   ii. Assumption of no correlation between X_4 and X_{5U} in each X_{1M} stratum will yield an unbiased estimate of ATT within each X_{1M} stratum. ATE and ATU are "identified" by estimating ATT through the assumption that providers do not have knowledge of how X_{1U} relates to treatment effectiveness within each X_{1M} stratum.

b. IV estimation for patient groups stratified by X_{1M} and statistically controlling for X_{5M} and using X_{4M} as an instrument:
   i. Sufficient variation in X_{4M} exists so that different treatment choices are observed in the data across X_{4M} strata within each X_{1M} stratum.
   ii. Assumption of no correlation between X_{4M} and X_{5U} in each X_{1M} stratum will yield a consistent estimate of LATE specific to the set of patient whose T choices were affected by measured factors within X_{4M}. ATE and ATU are "identified" by this LATE estimate through the assumption that providers do not have knowledge of how X_{1U} relates to treatment effectiveness within each X_{1M} stratum.

5. Treatment effect is heterogeneous, and all factors affecting treatment effectiveness have no direct effect on Y (X_1 exists; no X_2 factors exist). Moreover, heterogeneity is essential: Decisionmakers have knowledge of certain X_1 factors affecting treatment effectiveness that is sufficient to affect treatment choice, but these factors are unmeasured by the researcher.

   $Y = g(T(X_{1U}),X_{5M},X_{5U})$

   $T = f(X_{1U},X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M}:
   i. Sufficient variation in X_4 so that different treatment choices are observed in the data.

ii. Assumption of no correlation between X_4 and X_{5U} yields an unbiased estimate of ATT. Because X_{1U} is used in treatment choice, the distribution of X_{1U} will differ between the treated patients and untreated patients. Therefore, the ATE and ATU are unidentified by the ATT estimate.

b. IV estimation statistically controlling for X_{5M} and using X_{4M} as an instrument:

i. Sufficient variation in X_{4M} so different treatment choices are observed in the data across X_{4M} strata.

ii. Assumption of no correlation between X_{4M} and X_{5M} yields a consistent estimate of LATE specific to the set of patient whose T choices were affected by X_{4M}. Because the value of X_{1U} will define the subset of patients for whom X_{4M} factors affect their treatment choices (e.g., X_{4M} will less likely affect the treatment choices for patients with extreme X_{1U} values), the distributions of X_{1U} will differ among treated, untreated, and those patient used to estimate LATE. Therefore, the LATE estimate would not identify ATT, ATU, or ATE.

6. Treatment effect is heterogeneous, and factors affecting treatment effectiveness have no direct effect on Y(X_1 exists; no X_2 factors exist). Moreover, heterogeneity is essential: Decisionmakers have knowledge of the X_1 factors affecting heterogeneity sufficient to affect treatment choice, and all X_1 factors are measured by the researcher.

$Y = g(T(X_{1M}), X_{5M}, X_{5U})$

$T = f(X_{1M}, X_{4M}, X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M} within each X_{1M} stratum:

i. Sufficient variation in X_4 exists within each X_{1M} stratum so that different treatment choices are observed within each X_{1M} stratum.

ii. Assumed no correlation between X_4 and X_{5U} in each X_{1M} stratum yields an unbiased estimate of ATT within each X_{1M} stratum. Within each X_{1M} stratum, the ATE and ATU are "identified" by estimating ATT through the assumed homogeneity of treatment effect within the X_{1M} stratum.

b. Estimation for patient groups stratified by X_{1M} and statistically controlling for X_{5M} and using X_{4M} as an instrument:

i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data across X_{4M} strata within each X_{1M} stratum.

ii. Assumed no correlation between X_{4M} and X_{5U} in each X_{1M} stratum yields a consistent estimate of LATE specific to the set of patient whose T choices were affected by X_{4M}. ATE and ATU are "identified" within each X_{1M} stratum by estimating this LATE through the assumed homogeneity of treatment effect within each X_{1M} stratum. Moreover, with X_{1M} measured it would be possible to identify population-level values of ATT, ATE, and ATU, using LATE estimates based on X_{4M}.[27, 29, 30]

7. Treatment effect is heterogeneous, and factors affecting treatment effectiveness have no direct effect on Y (X_1 exists; no X_2 factors exist). Moreover, heterogeneity is essential: Decisionmakers have knowledge of the X_1 factors affecting heterogeneity sufficient to affect treatment choice. Only certain X_1 factors are measured by the researcher.

$Y = g(T(X_{1M}, X_{1U}), X_{5M}, X_{5U})$

$T = f(X_{1M}, X_{1U}, X_{4M}, X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M} within each X_{1M} stratum:

i. Sufficient variation in X_4 or X_{1U} exists within each X_{1M} stratum so that different treatment choices are observed within each X_{1M} stratum.

ii. Assumed no correlation between X_4 andX_{1U} and X_{5U} in each X_{1M} stratum yields unbiased estimates of ATT in each X_{1M} stratum. However, within each X_{1M} stratum, ATE and ATU that are not identified as X_{1U} will be distributed differently for treated and untreated patients within each X_{1M} stratum.

b. IV estimation for patient groups stratified by X_{1M} and statistically controlling for X_{5M} and using X_{4M} as an instrument:

i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data across X_{4M} strata within each X_{1M} stratum.

ii. Assumed no correlation between X_{4M} and X_{5U} in each X_{1M} stratum yields consistent estimates of LATE in each X_{1M} stratum that are specific to the set of patient whose T choices were affected by X_{4M}. ATE and ATU are not "identified" within each X_{1M} stratum, as the distribution of X_{1U} will vary between treated and untreated patients within each X_{1M} stratum.

8. Treatment effect is heterogeneous, and the factors affecting treatment effectiveness have direct effects on Y (no X_1 factors exist; X_2 factors exist). Moreover, heterogeneity is nonessential: Decisionmakers do not have sufficient knowledge of the X_2 factors affecting heterogeneity to affect treatment choice and X_2 factors are unmeasured by the researcher.

$Y = g(T(X_{2U}), X_{2U},X_{5M},X_{5U})$

$T = f(X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M}:

i. Sufficient variation in X_4 so that different treatment choices are observed in the data.

ii. Assumed no correlations between X_4 and X_{2U} and X_{5M} yields an unbiased estimate of ATT. ATE and ATU are "identified" by estimating ATT through the assumption that X_4 and X_{2U} are uncorrelated. If X_4 was correlated with X_{2U}, average X_{2U} would vary between treated and untreated patients, and the RA estimate of ATT would be biased (which is distinct from scenario 3).

b. IV estimation statistically controlling for X_{5M} and using X_{4M} as an instrument:

i. Sufficient variation in X_{4M} so different treatment choices are observed in the data across X_{4M} strata.

ii. Assumed no correlation between X_{4M} and X_{2U} and X_{5U} yields a consistent estimate of LATE specific to the set of patient whose T choices were affected by factors within X_{4M}. ATE and ATU are "identified" by estimating this LATE through the assumption that X_{4M} and X_{2U} factors are uncorrelated. If X_{4M} factors are correlated with X_{2U}, then the IV LATE estimate would be inconsistent.

9. Treatment effect is heterogeneous, and factors affecting treatment effectiveness have direct effect on Y (no X_1 factors exists; X_2 factors exist). Moreover, heterogeneity is nonessential: Decisionmakers do not have sufficient knowledge of the X_2 factors affecting heterogeneity to affect treatment choice. However, certain suspected X_{2M} factors are measured by the researcher.

$Y = g(T(X_{2M},X_{2U}), X_{2M},X_{2U},X_{5M},X_{5U})$

$T = f(X_{4M},X_{4U})$

a. Direct RA estimation of Y equation statistically controlling for X_{5M} for patient groups stratified by X_{2M}:

i. Sufficient variation in X_4 exists so that different treatment choices are observed in the data within each stratum of X_{2M}.

ii. Assumed no correlation between X_4 and X_{2U} and X_{5U} in each X_{2M} stratum yields unbiased estimates of ATT within each X_{2M} stratum. Within each X_{2M} stratum, the ATE and ATU are "identified" by the ATT estimate through the assumed lack of provider knowledge of treatment effect heterogeneity related to X_{2U} when making treatment choices within each X_{2M} stratum.

b. IV estimation for patient groups stratified by X_{2M} and statistically controlling for X_{5M} and using X_{4M} as an instrument:

i. Sufficient variation in X_{4M} exists so that different treatment choices are observed in the data across X_{4M} strata within each X_{2M} stratum.

ii. Assumed no correlation between X_{4M} and X_{2U} and X_{5U} in each X_{2M} stratum yields consistent estimates of LATE, specific in each X_{2M} stratum for the set of patient whose treatment choices were affected by factors within X_{4M}. ATE and ATU are "identified" by LATE within each X_{2M} stratum through the assumed lack of provider knowledge of treatment effect

heterogeneity related to X_{2U} when making treatment choices within each X_{2M} stratum.

10. Treatment effect is heterogeneous, and all factors affecting treatment effectiveness have no direct effect on Y (no X_1 factors exists; X_2 factors exist). Moreover, heterogeneity is essential: Decisionmakers have knowledge of certain X_2 factors affecting treatment effectiveness that is sufficient to affect treatment choice, but these factors are unmeasured by the researcher.

    $Y = g(T(X_{2U}),X_{2U},X_{5M},X_{5U})$

    $T = f(X_{2U},X_{4M},X_{4U})$

    a. Direct RA estimation of Y equation statistically controlling for X_{5M}:

        i. Sufficient variation in X_4 so that different treatment choices are observed in the data.

        ii. Because X_{2U} is unmeasured and is related to both Y and T, the RA estimator will be a biased estimate of ATT. Accordingly, ATE and ATU will be unidentified by the biased ATT estimate.

    b. IV estimation statistically controlling for X_{5M} and using X_{4M} as an instrument:

        i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data across X_{4M} strata.

        ii. Assumed no correlation between X_{4M} and X_{2U} and X_{5U} yields consistent estimates of LATE specific to the patients whose treatment choices were affected by X_{4M}. Because the value of X_{2U} will define the subset of patients for whom X_{4M} factors affect their treatment choices (e.g., X_{4M} will less likely affect the treatment choices for patients with extreme X_{2U} values), the distributions of X_{2U} will differ among treated, untreated, and those patients used to estimate LATE. Therefore, LATE, while consistent, would not identify ATT, ATU, or ATE.

11. Treatment effect is heterogeneous, and factors affecting treatment effectiveness have no direct effect on Y (no X_1 factors exists; X_2 factors exist). Moreover, heterogeneity is essential: Decisionmakers have knowledge of the X_2 factors affecting heterogeneity sufficient to affect treatment choice, and all X_2 factors are measured by the researcher.

    $Y = g(T(X_{2M}) ,X_{2M},X_{5M},X_{5U})$

    $T = f(X_{2M},X_{4M},X_{4U})$

    a. Direct RA estimation of Y equation statistically controlling for X_{5M} within each X_{2M} stratum:

        i. Sufficient variation in X_4 exists within each X_{2M} stratum so that different treatment choices are observed within each X_{2M} stratum.

        ii. Assumed no correlation between X_4 and X_{5U} in each X_{2M} stratum yields unbiased estimate of ATT within each X_{2M} stratum. Within each X_{2M} stratum, the ATE and ATU are "identified" by estimating ATT through assumed homogeneity of treatment effect within each X_{2M} stratum.

    b. IV estimation for patient groups stratified by X_{2M} and statistically controlling for X_{5M} and using X_{4M} as an instrument:

        i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data across X_{4M} strata within each X_{2M} stratum.

        ii. Assumed no correlation between X_{4M} and X_{5U} in each X_{2M} stratum yields consistent estimates of LATE in each X_{2M} stratum specific to the patients whose treatment choices were affected by X_{4M}. ATE and ATU are "identified" within each X_{2M} stratum by this LATE estimate through assumed homogeneity of treatment effect within each X_{2M} stratum.

12. Treatment effect is heterogeneous, and factors affecting treatment effectiveness have no direct effect on Y (no X_1 factors exists; X_2 factors exist). Moreover, heterogeneity is essential: Decisionmakers have knowledge of the X_2 factors affecting heterogeneity sufficient to affect treatment choice. Only certain X_2 factors are measured by the researcher.

    $Y = g(T(X_{2M},X_{2U}),X_{2M},X_{2U},X_{5M},X_{5U})$

    $T = f(X_{2M},X_{2U},X_{4M},X_{4U})$

    a. Direct RA estimation of Y equation statistically controlling for X_{5M} within each X_{2M} stratum:

        i. Sufficient variation in X_4 or X_{2U} exists within each X_{2M} stratum so that different

treatment choices are observed within each X_{1M} stratum.

ii. Because X_{2U} is related to both Y and T and is unmeasured, the RA estimator yields a biased estimate of ATT within each X_{2M} stratum. Accordingly, ATE and ATU will be unidentified by the biased ATT estimate within each X_{2M} stratum.

b. IV estimation for patient groups stratified by X_{2M} and statistically controlling for X_{5M} and using X_{4M} as an instrument:

i. Sufficient variation in X_{4M} so that different treatment choices are observed in the data across X_{4M} strata within each X_{2M} stratum.

ii. Assumed no correlation between X_{4M} and X_{2U} and X_{5U} in each X_{2M} stratum yields consistent estimates of LATE in each X_{2M} stratum specific to the patients whose treatment choices were affected by X_{4M}. ATE and ATU are not "identified" within each X_{2M} stratum as the distribution of X_{2U} will vary between treated and untreated patients within each X_{2M} stratum.

References

1. Koopmans TC. Identification problems in economic model construction. Econometrica. 1949;17:125-44.
2. Cameron AC, Trivedi PK. Microeconometrics: Methods and Applications. New York: Cambridge University Press; 2005.
3. Manski CF. Identification for Prediction and Decision. Cambridge, MA: Harvard University Press; 2007.
4. Kennedy P. A Guide to Econometrics. 4th Edition ed. Cambridge, MA: The MIT Press; 1998.
5. Wennberg JE. Which rate is right? N Engl J Med. 1986;315(13):810-5.
6. Heckman JJ. Building bridges between structural and program approaches to evaluating policy. J Econ Lit. June 2010;68(2):356-98.
7. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial-infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA. 1994 Sep 21;272(11):859-66.
8. Stukel TA, Fisher ES, Wennberg DE, et al. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007 Jan 17;297(3):278-85.
9. Zeliadt SB, Potosky AL, Penson DF, et al. Survival benefit associated with adjuvant androgen deprivation therapy combined with radiotherapy for high and low-risk patients with nonmetastatic prostate cancer. Int J Radiat Oncol Biol Phys. 2006;66(2):395-402.
10. Lu-Yao GL, Albertson PC, Moore DF, et al. Survival following primary androgen deprivation therapy among men with localized prostate cancer. JAMA. 2008 Jul 9;300(2):173-81.
11. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-experimental Designs for Generalized Causal Inferences. Boston: Houghton Mifflin Company; 2002.
12. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov;64(11):1198-207.
13. Guyatt GH, Sinclair J, Cook DJ, et al. Users' guides to the medical literature: XVI. How to use a treatment recommendation. JAMA. 1999 May 19;281(19):1836-43.
14. Brooks JM, McClellan M, Wong HS. The marginal benefits of invasive treatments for acute myocardial infarction: does insurance coverage matter? Inquiry. 2000 Spring;37(1):75-90.
15. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983 Apr 1;70(1):41-55.
16. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med. 1997 Oct 15;127(8 Pt 2):757-63.
17. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health. 2000;21:121-45.
18. Angrist JD. Treatment effect heterogeneity in theory and practice. Economic Journal. 2004;114(494):C52-C83.

19. Brooks JM, Chrischilles EA. Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods. Med Care. 2007 Oct;45(10 Supl 2):S123-30.

20. Heckman JJ, Urzua S, Vytlacil E. Understanding instrumental variables in models with essential heterogeneity. Review of Economics and Statistics. 2006;88(3):389-432.

21. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22:189-212.

22. Heckman JJ, Robb R. Alternative Methods for Evaluating the Impact of Interventions. In: Heckman JJ, Singer B, eds. Longitudinal Analysis of Labor Market Data. New York: Cambridge University Press; 1985:156-245.

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

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

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

26. Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41-55.

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

28. Brooks JM, Fang G. Interpreting treatment-effect estimates with heterogeneity and choice: simulation model results. Clin Ther. 2009 Apr;31(4):902-19.

29. Angrist JD, Fernandez-Val I. Extrapolate-ing: External Validity and Overidentification in the LATE Framework. National Bureau of Economic Research Working Paper. Cambridge, MA; 2010.

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