NBER WORKING PAPER SERIES

ESTIMATING PERSON-CENTERED TREATMENT (PET) EFFECTS USING INSTRUMENTAL VARIABLES

Anirban Basu

Working Paper 18056
http://www.nber.org/papers/w18056

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2012

The author acknowledges support from the National Institute of Health Research Grants, RC4CA155809 and R01CA155329. The author acknowledges helpful comments from seminar participants at the Department of Economics at the University of Washington. The author thanks John Gore and Edward Vytlacil for their comments on an earlier version of the paper and takes responsibility on all errors in the paper. Excellent programming assistance from William Kreuter is acknowledged. This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the Applied Research Program, NCI; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database The collection of the California cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract N01-PC-35136 awarded to the Northern California Cancer Center, contract N01-PC-35139 awarded to the University of Southern California, and contract N02-PC-15105 awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #U55/CCR921930-02 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the author(s) and endorsement by the State of California, Department of Public Health the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred. The authors acknowledge the efforts of the Applied Research Program, NCI; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

Estimating Person-Centered Treatment (PeT) Effects Using Instrumental Variables
Anirban Basu
NBER Working Paper No. 18056
May 2012, Revised July 2012
JEL No. C21,C26,D04,I12

ABSTRACT

This paper builds on the methods of local instrumental variables developed by Heckman and Vytlacil (1999, 2001, 2005) to estimate person-centered treatment (PeT) effects that are conditioned on the person's observed characteristics and averaged over the potential conditional distribution of unobserved characteristics that lead them to their observed treatment choices. PeT effects are more individualized than conditional treatment effects from a randomized setting with the same observed characteristics. PeT effects can be easily aggregated to construct any of the mean treatment effect parameters and, more importantly, are well-suited to comprehend individual-level treatment effect heterogeneity. The paper presents the theory behind PeT effects, studies their finite-sample properties using simulations and presents a novel analysis of treatment evaluation in health care.

Anirban Basu
Department of Health Services
School of Public Health
University of Washington
1959 NE Pacific St
Box - 357660
Seattle WA 98195
and NBER
basua@uw.edu

INTRODUCTION

Much of the literature on treatments effects has focused on estimating effect parameters that inform population level or policy-level decisions. Even when distributional impacts of treatments and policies are studied, the impacts are viewed as informing a social decision maker to help choose across alternative options (Heckman 2001). However, in the presence of heterogeneous treatment effects, it is natural to expect that individual choices of treatments may vary from the socially optimal treatment that is identified based on some average social welfare criterion. More importantly, treatment effect information that can help change future individual-level behavior on treatment choices would automatically influence social choice of treatments through positive self-selection. Hence, estimating treatment effects that can inform individual-level decision making can be of great social value.

This conundrum manifests in its most acute form in the health care setting. In traditional clinical outcomes research, the focus has always been on finding average effects either through large clinical trials or observational datasets. Estimating treatment effect heterogeneity has mostly been relegated to post-hoc analysis, rather than becoming the central goal of the analysis.[1] Yet, the clinical setting is an obvious place where individual-level decision making is most relevant as a physician-patient dyad tries to decide on the best line of treatment for that patient. There is a growing recognition based on fundamental theoretical principles that more nuanced and possibly individualized estimates of treatment effects between alternative medical interventions can lead to increased welfare through more efficient use of medical technologies (Basu 2009, 2011). In contrast, failing to generate such individualized estimates and also producing results on population average effects without recognizing the underlying heterogeneity could lead to welfare losses including faster growth in health care expenditures (Basu et al., 2011; Basu 2011).[2]

---

[1] In randomized settings, heterogeneity analyses are often accomplished using post-hoc subgroup analyses (ref provenge). In some of our recent work, we have shown that such approaches are likely to be futile since these subgroups are often defined based on broad characteristics (e,g, gender) that only explains a very small fraction of the individual-level variance in treatment effects (Basu et al, 2012).

[2] In fact such insights and assertions line up well with the political economy of outcomes research funding in the United States, which witnessed the creation of the Patient-Centered

In the evaluation literature, such nuanced treatment effects are most popularly characterized by conditional average treatment effects (CATE) where an average treatment effect is estimated conditional on certain values of observed covariates over which treatment effects vary. For example, if age is the only observed risk factor, one can establish a conditional effect of surgery versus active surveillance on mortality for patients of age 60 years diagnosed with clinically localized prostate cancer. This is an average effect for all 60 year olds in this condition. However, does this estimate apply to all men with clinically localized prostate cancer at age 60 years? Certainly not, as there may be many other factors that determine heterogeneity in treatment effects in this population. For example, clinical stage and grade of cancer not only determines overall survival but may also determine differential effects from alternative treatments. To the extent that all potential *moderators* of treatments effects are observed to the analyst of the data, a nuanced CATE can be established conditioning on values of all of these factors.

In most applied work however, not all moderators of treatment effects are observed. One reason is that many of these moderators are yet to be discovered and hence remain unknown to scientific knowledge. They are typically represented by the pure stochastic error term in statistical analysis of data. However, there are some moderators that fall within the purview of scientific knowledge but remain unmeasured in the data at hand. This is usually the case for most randomized studies that rely on randomization to equate the distribution of all these factors across the randomization arms and forgo measurement of several factors in the interest of time and expenses.

In observational studies, these unmeasured moderators of treatment effects play a vital role in generating essential heterogeneity as often they are observed by individuals and acted upon by some while making treatment selection (Heckman 1997; Heckman and Vytlacil, 1999).[3] An entire genre of methods, including methods based on local instrumental variable (LIV) approaches, have been developed to

Outcomes Research Institute (PCORI) through the 2010 Patient Protection and Affordable Care Act.

[3] In fact Basu (2011) made the argument that the traditional "selection on gains" rationale used in the education and labor literature is not the only mechanism to assert essential heterogeneity. Even if gains are unpredictable and selection is based on baseline factors, as long as those factors are not completely independent of the gains, essential heterogeneity is induced.

estimate policy-relevant and structurally stable mean treatment effect parameters in the presence of essential heterogeneity (Heckman and Vytlacil 1999, 2001, 2005). Basu et al. (2007, 2011) introduced these methods to the health economics literature where essential heterogeneity is widespread and instrumental variable methods are gaining meteoric popularity. Carneiro and Lee (2009) extended the LIV methods to estimate the marginal distributions of expected potential outcomes that are geared towards studying distributional impacts of population level policies.

LIV methods can seamlessly explore treatment effect heterogeneity across both observable characteristics and unobserved confounders and also be used to establish CATE based on observed factors. In this paper, we develop and present a new individualized treatment effect concept called *Person-Centered* Treatment (PeT) effects, which can also be estimated using LIV methods. This new treatment effect concept is more personalized than CATE as it takes into account individual treatment choices and the circumstances under which people are making those choices in an observational data setting in order to predict their individualized treatment effects. In our prostate cancer example suppose that we not only have data on age of the prostate cancer patients but also the treatment they choose and the distances of their residence from the hospitals that offer surgical procedures. Assume that these distances impart a cost for accessing surgery and therefore influence treatment selection but do not affect the potential outcomes for these patients under either treatment, i.e. they are instrumental variables. Under such circumstance, 60-year old patients, who live far from hospital and still choose surgery is likely to have a different distribution of unobserved confounders than 60-year old patients who live close to the hospital and choose surgery. Therefore, by taking into account treatment choices and the observed circumstances under which those choices were made, we can enrich CATE to form a Person-centered Treatment (PeT) effect that provides a conditional treatment effect that is averaged over a personalized conditional distribution of unobserved confounders and not their marginal distribution as in CATE.

There are several intuitive aspects about the PeT effects:

1. They help to comprehend individual-level treatment effect heterogeneity better than CATEs.
2. They are better indicators for the degree of self-selection than CATE. Specifically, they are better predictors of true treatment effects at the individual

level both in terms of the positive predictive value and the negative predictive value.

3. They can explain a larger fraction of the individual-level variability in treatment effects than the CATEs. That the marginal distribution of PeT effects is a better proxy for the true marginal distribution of individual effects that that of CATEs.
4. All mean treatment effect parameters can be easily computed from PeT effects without any further weighting. So, they also form integral components for population–level decision making.

All of these features of PeT effects will be studied here. We begin in Section 2 with the definition, identification and estimation of PeT effects. Section 3 presents a simulation study showing the how PeT effects inform individual-level and the mean treatment effect parameters across a variety of outcomes and sample sizes. In Section 4, we illustrate the use of estimated PeT effects of surgery versus active surveillance on 7-year survival and costs among patients diagnosed with clinical localized prostate cancer. Discussions follows in Section 5.

2. PERSON-CENTERED TREATMENT (PeT) EFFECTS

Structural Models for Outcomes and Choices

We start by formally developing structural models of outcomes and treatment choice following Heckman and Vytlacil (1999, 2001, 2004). For the sake of simplicity we will restrict our discussion to two treatment states – the *treated* state denoted by $j = 1$ and the *untreated* state denoted by $j = 0$, The corresponding potential individual outcomes in these two states are denoted by Y_1 and Y_0. We assume,

$$Y_1 = \mu_1(X_O, X_U, \vartheta) \qquad \text{and} \qquad Y_0 = \mu_0(X_O, X_U, \vartheta) \tag{1}$$

Where X_O is a vector of observed random variables, X_U is a vector of unobserved random variables which are also believed to influence treatment selection (they are the unobserved confounders) and ϑ is an unobserved random variable that capture all remaining unobserved random variables.

Assumption 1. $(X_O, X_U) \coprod \vartheta$ and $X_O \coprod X_U$ where $\coprod$ denotes statistical independence.

We assume individual choose to be in state 1 or 0 (prior to the realization of the outcome of interest) according to the following equation:

$$D = 1 \quad \text{if} \quad \mu_D(X_0, Z) - U_D > 0 \qquad (2)$$

Where Z is a (non-degenerate) vector of observed random variables (instruments) influencing the decision equation but not the potential outcome equations, μ_D is an unknown functions of X_0 and Z, and U_D is a random variable that captures X_U and all remaining unobserved random variables influencing choice. By definition, $U_D \coprod \vartheta$, which also defines the distinction between X_U and ϑ in (1). Equation (1) and (2) represent the nonparametric models that conform to the Imbens and Angrist's (1994) independence and monotonicity assumptions needed to interpret instrumental variable estimates in a model of heterogeneous returns (Vytlacil, 2002). As in Heckman and Vytlacil (1999, 2001, 2005), we can rewrite (2) as

$$D = 1 \quad \text{if} \quad P(x_0, Z) > V \qquad (3)$$

Where $V = F_{U_D|X_O,Z}\left[U_D \mid X_O, Z\right]$, $P(x_0,Z) = F_{U_D|X_O,Z}\left[\mu_D(x_O, Z) \mid X_O, Z\right]$. Therefore, for any arbitrary distribution of U_D conditional on X_O and Z, by definition, $V \sim$ Unif[0, 1] conditional on X_O and Z.

Assumption 2. Assume that a) $\mu_D(x_0, Z)$ is a nondegenrate random variable conditional on $X_O = x_O$; b) (X_U, ϑ, U_D) are independent of Z conditional on $X_O = x_O$; c) The distribution of U_D conditional on (X_O, Z) and that of $\mu_D(x_0, Z)$ conditional on $X_O = x_O$ are absolutely continuous with respect to Lebesgue measure; d) Y_1 and Y_0 have finite moments and e) $\Pr(D = 1) > 0$.

An individual–level treatment effect is given as

$$\text{TE} = (Y_1 - Y_0) \qquad (4)$$

Obviously, we never observe both the potential outcomes for each individual. Our observed outcome Y is given as

$$Y = Y_1 \quad \text{if } D = 1 \text{ and } Y = Y_0 \quad \text{if } D = 0 \tag{5}$$

Therefore, the goal of the analysis is to obtain estimates of Y_1 for subjects with $D = 0$ and of Y_0 for subjects with $D = 1$. These outcomes are known as counterfactual outcomes as they represent the potential outcomes had the subjects chose a different treatment than they have in practice. Differences in the counterfactual outcomes across individual subjects will depend of X_O, X_U, and ϑ. Several individual level treatment effect parameters can be defined that reflect these variations.

Treatment Effect Definitions

Individualized Expected Treatment Effect (IETE): Since ϑ is typically not only unmeasured but also unknown (as otherwise would have been used for treatment selection), the most precise individualized expected treatment effect (IETE) that one can hope for in terms of predictions is given by:

$$\text{IETE} = \xi(x_O, x_U) = E_{\vartheta | X_O, X_U}(Y_1 - Y_0 \mid x_0, x_U) = E_{\vartheta}(Y_1 - Y_0 \mid x_0, x_U) \tag{6}$$

Throughout this paper, we will denote IETE as $\xi(x_O, x_U)$ and it will serve as a reference to which our proposed individual treatment effect parameter and other parameters will be compared. The typical population-level mean treatment effect parameters, the Average Treatment Effect (ATE), the Effect on the Treated (TT) and the Effect on the UnTreated (TUT), can be derived by appropriate aggregation of $\xi(x_O, x_U)$ over the relevant subgroups.

$$\text{ATE} = E_{X_O}\left\{E_{X_U|X_O}\left\{\xi(x_O, x_U) \mid x_O, x_U\right\} \mid x_O\right\} = E_{X_O}\left\{E_{X_U}\left\{\xi(x_O, x_U) \mid x_O, x_U\right\} \mid x_O\right\}$$

$$\text{TT} = E_{X_O|D=1}\left\{E_{X_U|X_O,D=1}\left\{\xi(x_O, x_U) \mid x_O, x_U, D = 1\right\} \mid x_O, D = 1\right\}$$

$$\text{TUT} = E_{X_O|D=0}\left\{E_{X_U|X_O,D=1}\left\{\xi(x_O, x_U) \mid x_O, x_U, D = 0\right\} \mid x_O, D = 0\right\} \tag{7}$$

Note that the second equality for ATE follows from Assumption 1.

Conditional Average Treatment Effect (CATE): Since X_O are the only observed variables from the outcomes equation, a conditional average treatment effect (CATE)

(Heckman 1997) can be formed which is the average treatment effect conditioned on levels of X_O only.

$$\text{CATE} = E_{X_U|X_O}\{\xi(x_O, x_U) \mid x_O, x_U\} = E_{X_U}\{\xi(x_O, x_U) \mid x_O, x_U\} \qquad (8)$$

where the second equality follows from Assumption 1. We will denote CATE as $\xi(x_O)$. This is the treatment effect parameter that an ideal experiment can give where only X_O are observed. Note that the outer expectation in CATE averages over the marginal distribution of X_U. Although the ATE can be obtained by trivial aggregation of CATEs over all individuals (as in (7)), aggregation of CATE over the treated or the untreated individuals do not produce the TT or the TUT parameters respectively.

Marginal Treatment Effect (MTE): The marginal treatment effect is perhaps the most nuanced estimable effect (Heckman 1997; Vytlacil 1999, 2001). It identifies an effect for an individual who is at the margin of choice such that one's levels of X_O and Z are just balanced by one's level of V (which includes X_U), i.e. $P(x_0, Z) = V$. MTE can be expressed as

$$\text{MTE}(x_O, z) = E_{X_U|X_O, P(Z)=V}\{\xi(x_O, x_U) \mid x_O, p(z)\} \qquad (9)$$

Note that, unlike CATE, the expectation in MTE averages over the conditional distribution of X_U conditioned on meeting the definition for marginal patients. Heckman and Vytlacil (1999, 2001) have provided the weights needed to aggregate MTEs to form the mean treatment effect parameters. These weights need to be calculated from the data at hand.

Person-centered Treatment (PeT) Effect: Despite the granularity of MTEs, it may be hard to use MTEs directly as representation of individual treatment effects as they themselves lack individual identity. This is because it is hard (if not impossible) to pinpoint an individual to whom an MTE estimate can be applied to. Instead another treatment effect, which we call the Person-centered Treatment (PeT) effect (denoted as Δ), can be written as:

$$\Delta = E_{X_U|X_O, P(Z), D}\{\xi(x_O, x_U) \mid x_O, p(z), D = d\} \qquad (10)$$

where the expectation of unobserved confounders is made conditional on person-specific estimates of X_O, $P(Z)$ and D. Naturally, PeT effects are more nuanced than CATEs. Note that this parameter was originally defined by Heckman and Vytlacil (1999). However, they use this parameter as a stepping stone for defining structurally stable mean effects on treated parameter whose definition do not depend on data (Y, X, Z). The PeT effect in (10) would take on different values corresponding to two values of $Z =(z, z')$, $z \neq z'$, with (Y, X,D) being constant. However, this is exactly the variation we are after when we are envisioning PeT effects. The fact that two otherwise observably similar persons choose the same treatment under two values of Z informs us that their personalized treatment effects may be different.

Conceptually, a PeT effect is also a weighted version of MTEs. For any given individual, the PeT effects identifies the specific margins where that individual may belong given its individual values of X_O, $P(Z)$ and D. It then averages the MTEs over those margins, but not all as in ATE. As we prove below, a PeT effect is basically the X-Z-conditional Effect on the Treated (x-z-CTT) for persons undergoing treatment and is the X-Z-conditional Effect on the Untreated (x-z-CTUT) for persons not undergoing treatment. Because conditioning is done based on identifiable individual-level characteristics, a PeT effect can be identified for each individual in the data.

Uses of PeT effects

All mean treatment effect parameters can be easily computed from the PeT effects without any further weighting. For example:

$$\begin{aligned}
&\text{ATE} = E_{X_O,P(Z),D}(\Delta)\\
&\text{TT} = E_{X_O,P(Z)|D=1}(\Delta)\\
&\text{TUT} = E_{X_O,P(Z)|D=0}(\Delta)
\end{aligned} \qquad (11)$$

In fact, any policy parameter that shifts a certain subgroup of individuals, characterized by shifting the distribution of X_O, to take up or give up treatment can be predicted. Therefore, these patient-centered treatment effects can form integral components for population–level decision making.

More importantly, distributions of treatment effects are useful for policy makers who care about distributional effects of policies (Heckman and Robb, 1985). For individual decision maker, such distributional effects are of central importance. Although difficult to establish, the most useful metrics to study distributional impacts of policies and treatments are the full marginal and joint distributions of potential outcomes. Previous work by Imbens and Rubin (1997) and Abadie (2002, 2003) have developed estimators for the marginal distributions of potential outcomes under the local average treatment effect (LATE) framework, where the instrument corresponds to the specific policy question that is being studied. Carniero and Lee (2009) extends the LIV framework of Heckman ad Vytlacil (1999, 2001, 2005) to identify distributions of potential outcomes and to develop a semiparametric estimator for the entire marginal distribution of potential outcomes. However, when it comes to understanding individualized decision making, estimating the marginal distribution of potential outcomes is not enough. They carry no information to help identify the quantile of the marginal distribution of counterfactual outcomes where an individual may lie had he taken an alternative treatment (Carneiro et al 2001). One must have knowledge about the full joint distribution of potential outcomes, which can only be established under much more stronger assumptions (Heckman and Honoré 1990, Heckman and Smith 1993, Heckman et al. 1997).[4]

In the absence of identification of the joint distribution of potential outcomes, however, the marginal distribution of the PeT effect can be crucial for understanding individual level decision making. The PeT effects can be used to more accurately comprehend individual-level treatment effect heterogeneity that CATEs fail to convey. First, they may be better predictors of true treatment effects at the individual level both in terms of the positive predictive value (Pr($\xi(x_O, x_U) \geq 0 \mid \Delta \geq 0$)) and the negative predictive value (Pr($\xi(x_O, x_U) < 0 \mid \Delta < 0$)) than the CATEs (we will study this using simulations). Second, PeT effects are more likely to explain a larger fraction of the individual-level variability in treatment effects than the CATEs. Both play a big role in not only identifying person characteristics to guide treatment allocations but also in guiding future research to focus on collection of relevant measures of X_O and X_U.

[4] Heckman and Honoré (1990) uses parametric assumptions. Heckman and Smith (1993) and Heckman et al. (1997) assume that the persons at the *qth* percentile in the density of $Y0$ are at the *qth* percentile of $Y1$. More recently, using additional measurements in micro data, factor structure models have been used to establish the joint distribution of potential outcomes (Aakvik et al. 1999, Carniero et al. 2003).

Naturally, in the absence of essential heterogeneity, the PeT effects converge to CATEs.

Identification of PeT effects

<u>Theorem 1</u>. *Consider the nonparametric selection and outcome models in (1) and (2). Under Assumption 1 and 2,*

$$E_{X_U|X_O,P(Z),D} E_\vartheta(Y_1 - Y_0 \mid x_O, P(z), D = 1) = P(z)^{-1} \int_0^{P(z)} \left(\left. \frac{E_\vartheta(Y \mid x_O, P = p)}{\partial p} \right|_{p=\upsilon} \right) d\upsilon$$

$$E_{X_U|X_O,P(Z),D} E_\vartheta(Y_1 - Y_0 \mid x_O, P(z), D = 0) = (1 - P(z))^{-1} \int_{P(z)}^{1} \left(\left. \frac{E_\vartheta(Y \mid x_O, P = p)}{\partial p} \right|_{p=\upsilon} \right) d\upsilon$$

provided that $E_\vartheta(Y \mid X_O, P = p)$ *is continuously differentiable with respect to p for almost every* x_O.

<u>Proof</u>. The identification for PeT effects follows identification of marginal treatment effects (MTEs). Assumption 1(a) and (c) ensure that P is nondegenrate, continuously distributed random variable conditional on X_O. Assumption 2(d) is needed to ensure that the expectations considered are finite. First, following Heckman and Vytlacil (1999, 2001, 2005), the marginal treatment effect is identified as

$$\begin{aligned}
E_\vartheta(Y \mid X_0, Z) &= E(DY_1 + (1-D)Y_0 \mid X_O, Z) \\
&= E_\vartheta(Y_0 \mid X_O) + E_\vartheta(D(Y_1 - Y_0) \mid X_O, Z) \\
&= E_\vartheta(Y_0 \mid X_O) + \Pr(D = 1 \mid X_O, Z) \cdot E_\vartheta((Y_1 - Y_0) \mid X_O, V < P) \\
&= E_\vartheta(Y_0 \mid X_O) + \int_o^p E_\vartheta((Y_1 - Y_0) \mid X_O, V = \upsilon) d\upsilon
\end{aligned}$$

Where the second and third equalities follow from Assumptions 1(b) and the fourth equality comes from the fact that V is uniformly distributed on [0,1] conditional on X_O and Z. Therefore, differentiating both sides with respect to p, we have

$$\frac{\partial E_{\vartheta}(Y \mid x_0, Z)}{\partial p} = E_{\vartheta}((Y_1 - Y_0) \mid X_O, V = \upsilon) = MTE(x_O, \upsilon) \qquad (12)$$

It then follows,

$$E_{X_u|X_O,P(Z),D} E_{\vartheta}(Y_1 - Y_0 \mid x_0, P(z), D = 1)$$

$$= E(Y_1 - Y_0 \mid x_0, V < P(z)) = P(z)^{-1} \int_0^{P(z)} MTE(x_O, \upsilon) d\upsilon \qquad (13)$$

Similarly, conditional effect on the untreated (CTUT) is obtained by integrating *MTEs* over values of *V* that are greater than *p*. ■

The identification of PeT effects comes out directly from the identification results of Heckman and Vytlacil (1999, 2001, 2005). However, while Heckman and Vytlacil (1999, 2001, 2005) are mainly concerned with average treatment effects in the population, we use their results to identify individualized expected treatment effects and their marginal distribution in the population.

The PeT effects can be trivially aggregated over observed distribution of (X_O, $P(Z)$, D) in order to estimate mean treatment effect parameters such as the Effect on the Treated (TT), Effect on the Untreated (TUT) and the Average Treatment Effect (ATE). These derivations are provided in Heckman and Vytlacil (1999).

Semi-parametric estimation

In order to avoid certain disadvantages of full nonparametric estimation of the models in (1) and (2), we propose a partially separable outcomes model as follows:

$$Y = \mu(X_O, X_U, D; \beta) + \vartheta \qquad (14)$$

where $\mu(X_O, X_U, D)$ is an unknown non-linear function of observable (X_O) and unobservable (X_U) characteristics and treatment indicator (D); ϑ are purely random error term. Conditional on specific levels of X_O and X_U, idiosyncratic expected gains (or losses) from treatment over control is given by $\mu(x_O, x_U, D = 1; \beta)$ - $\mu(x_O, x_U, D = 0; \beta)$. These idiosyncratic gains or losses may vary either over observed characteristics X_O or over unobserved characteristics X_U or both, giving rise to treatment effect heterogeneity. The terms, *observable* and *unobservable*, pertain to the analyst's perspective and these covariates enter the structural model symmetrically in determining potential outcomes (Mullahy 1997). We will refer to this formulation of the symmetric structural non-linear model as the *pure* non-linear model. It encompasses the broad categories of all parametric and semiparametric generalized linear models (McCullagh and Nelder, 1989) that include models for limited dependent variables.

In addition to the assumption of $X_O, X_U \amalg \vartheta$, $X_O \amalg X_U$ and that of Assumption 1, we make the following additional assumptions:

Assumption 3: $E(\mu(X_O, X_U, D; \beta) \mid P = p, Z) = \varpi(X_O, K(P); \alpha)$, is continuously differentiable with respect to p, where $K(P)$ a non-linear kernel for P.

Estimation of PeT effects proceeds in four steps:

1. An estimate P is constructed using a semiparametric regression of D on X_O and Z (Das et al, 2003).
2. α is estimated using local polynomial approximation of $\varpi(X_O, K(P); \alpha)$ over P (Robinson, 1988; Fan and Gijbels, 1996). Here, $K(P)$ is represented by the polynomial approximation. Such approximation can be estimated using GMM estimators using the well-known quasi score equations (Wedderburn, 1974). For N individuals,

$$G_\alpha = \sum_{i=1}^{N} G_\alpha^i = \sum_{i=1}^{N} (Y_i - \mu_i) V_i^{-1} (\partial \mu_i / \partial \alpha) = 0, \qquad (15)$$

where i denotes individuals. α is estimated by solving $G_\alpha = 0$, yielding estimator $\hat{\alpha}_N$. Under mild regularity conditions, $\hat{\alpha}_N \xrightarrow{p} \alpha$ as $N \to \infty$ and $(\hat{\alpha}_N - \alpha)$ is asymptotically normal with mean 0 and covariance matrix $\boldsymbol{A_N}$ given by:

$$\boldsymbol{A_N} = \left[E(-\partial G_\alpha / \partial \alpha)\right]^{-1} \frac{N}{(N-1)} \left(\sum_{i=1}^{N} E(G_\alpha^i G_\alpha^{iT}) \right) \left[E(-\partial G_\alpha / \partial \alpha)\right]^{-T}. \qquad (16)$$

Replacing α by $\hat{\alpha}_N$ and $E(G_\alpha^i G_\alpha^{iT})$ with $G_\alpha^i G_\alpha^{iT}$ in (9) yields a sandwich estimator of the variance-covariance of $\hat{\alpha}_N$ (Huber, 1972; Liang and Zeger, 1986).

3. Obtain estimates for MTE following Assumption (2):
$$M\hat{T}E(X_0, V = v) = \partial \varpi(X_O, K(P); \hat{\alpha}) / \partial p\big|_{p=v} \qquad (17)$$
4. Construct PeT effects for each individual as: $\hat{\Delta}(x_O, p, D) = I(D=1) \cdot E_{V|D^*=1}\left(M\hat{T}E(x_0, V)\right) + I(D=0) \cdot E_{V|D^*=0}\left(M\hat{T}E(x_0, V)\right)$, where
$$D^* = D^*(p, v) = \left[\Phi(p) + \Phi(1-v)\right] > 0 \qquad (18)$$ [5]

The proof follows directly realizing that V~Unif[0,1].

Variance estimates for PeT effects at the individual level can be readily obtained by bootstrap, which is in line with obtaining variance estimates of CATEs. For each replicate of the bootstrap with-replacement sample, the average effect for each person is saved. In any given replicate, only those persons who are sampled would have an estimate. However, multiple bootstrap replicates should be able to cover all individuals. The total number of bootstrap replicates needed can be monitored by monitoring the minimum number of times each individual is sampled across replicate datasets.

Simulations

Set up

We study the effects of a binary treatment variable on three different types of outcomes. First is a typical normally distributed outcome. Second is a binary outcome and the third is a count data outcome. For each outcome, we specify a data generating process that incorporates essential heterogeneity.[6] We then estimate the PeT effects across individuals using LIV approaches assuming that we observe (Y, D, X_O, Z) only. We compare these PeT effects to the true values of CATEs. Also we compare the PeT-based estimates of mean treatment effect parameters to their true values. We also compute the traditional IV effects for comparison.

[5] We thank James Heckman and Philipp Eisenhauer for suggesting this approach to numerical computation.
[6] Note that essential heterogeneity is not being generated by direct selection of gains but rather through factor X_U that is shared between treatment choice and potential outcomes models.

Treatment choice model:

$D = I(\Lambda > 0)$ $\quad \Lambda = 1 + 1.0{*}X_O - 1.0{*}X_U + 1.0{*}Z + \varepsilon_\Lambda$, where ε_Λ ~Normal (0, 1)

Potential Outcomes Data Generating Mechanism (DGPs):

Normal Outcome:

$Y_1 = \mu_1 + \varepsilon_{Y1}$, $\mu_1 = -0.5 + 0.5{*}X_O - 0.5{*}X_U$

$Y_0 = \mu_0 + \varepsilon_{Y0}$, $\mu_0 = -1.0 - 0.5{*}X_O + 0.5{*}X_U$

where $\varepsilon_{\Gamma 1}$, $\varepsilon_{\Gamma 0}$ ~Normal (0,1) and $\varepsilon_{\Gamma 1} \amalg \varepsilon_{\Gamma 0}$, $\amalg$ denoting statistical independence. Here $\mu_1(x_O) = 0.5 + 1.5{*}X_O$ and $\mu_0(x_O) = 1.0 + 1.0{*}X_O$.

Binary Outcome:

$Y_1 = (\Gamma_1 > 0)$, $\quad \Gamma_1 = -0.5 + 0.5{*}X_O - 0.5{*}X_U + \varepsilon_{\Gamma 1}$,

$Y_0 = (\Gamma_0 > 0)$, $\quad \Gamma_0 = -1.0 - 0.5{*}X_O + 0.5{*}X_U + \varepsilon_{\Gamma 0}$,

where $\varepsilon_{\Gamma 1}$, $\varepsilon_{\Gamma 0}$ ~Normal (0, 1) and $\varepsilon_{\Gamma 1} \amalg \varepsilon_{\Gamma 0}$. Here $E(Y_1|X_O, X_U) = \mu_1 = \Phi(-0.5 + 0.5{*}X_O - 0.5{*}X_U)$ and $E(Y_0|X_O, X_U) = \mu_0 = \Phi(-1.0 - 0.5{*}X_O + 0.5{*}X_U)$. Also, $\mu_1(x_O) = \Phi((-0.5 + 0.5{*}X_O)/\sqrt{1.25})$ and $\mu_0(x_O) = \Phi((-1.0 - 0.5{*}X_O)/\sqrt{1.25})$.

Skewed Non-negative Outcomes:

Y_1 ~Gamma (a_1, b_1), $\quad b_1 = \exp(-\ln(a_1) - 0.5 + 0.5{*}X_O - 0.5{*}X_U)$

Y_0 ~Gamma(a_0, b_0), $\quad b_0 = \exp(-\ln(a_0) - 1.0 - 0.5{*}X_O + 0.5{*}X_U)$,

where $E(Y_j| X_0, X_U) = \mu_j = a_j{*}b_j$. a_1, a_0 are the inverse-dispersion parameters such that $Var(Y_j | X_0, X_u) = a_j^{-1} \cdot \mu_j^2$, $j = 0,1$. We assume $a_j = 2$, j=0,1. $Y_1 \amalg Y_0 | X_0, X_u$.

Also, $\mu_1(x_O) = a_1^2$*exp(0.5 + 1.5*X_O)*exp(0.125) and $\mu_0(x_O) = a_0^2$*exp(1.0 + 1.0*X_O)*exp(0.125)

X_O, X_U and Z are generated as independent Normal (0,1) variates. 1000 replicate samples each of sample sizes 5000, and 20000 were drawn from each DGP. [7] Results, as described below were averaged over replicates. Furthermore, for each replicate of data, 500 bootstrapped samples were drawn to assess the empirical variance of the estimators we study.

<u>Treatment Effects</u>

For each DGP, we estimate PeT effects as $\hat{\Delta} = \hat{\mu}_1 - \hat{\mu}_0$ using the LIV approach. We will also construct true values for CATE, where conditioning is done on X_O: $\xi(x_O) = \mu_{1i}(x_O) - \mu_{0i}(x_O)$.[8] We compute ATE, TT and TUT using estimated Δ and compare them to their respective true values. We also estimate the IV effect using traditional IV methods and compare them to the true values of the mean treatment effect parameters. Additionally, we report the Monte Carlo standard deviations for each parameter estimate and their 95% coverage probabilities.

In order to show that CATE(X_O) are consistently recoverable from the PeT effects, we estimate CATE(X_O) by averaging $\hat{\Delta}_i$ over deciles of X_O and comparing then to true values.

Finally, in order to evaluate the accuracy in predicting individual level effects, we will compare the distribution of Δ_i and $\xi(x_O)$ to the true distribution of expected individual-level treatment effects $\xi(x_O, x_U) = \mu_1 - \mu_0$ using the following metrics:

1. Corr($\hat{\Delta}$, $\xi(x_O, x_U)$) versus Corr($\xi(x_O)$, $\xi(x_O, x_U)$),

2. R^2 of $\xi(x_O, x_U)$ on $\hat{\Delta}$ versus R^2 of $\xi(x_O, x_U)$ on $\xi(x_O)$,

---

[7] These sample sizes are most typical of health economic analyses. For example, a Medline search of all IV applications in the past three years revealed that 70% of then had sample size greater than 5000. In fact, in field of CER, with the emergence of more integrated data, these sample sizes are only likely to increase.

[8] Note that we do not estimate CATEs but rather construct them based on true values in order to make the comparisons with estimated PeT effects conservative.

3. Positive Predictive Values (PPV): $\Pr(\xi(x_O, x_U) \geq 0 \mid \hat{\Delta} \geq 0)$ vs $\Pr(\xi(x_O, x_U) \geq 0 \mid \xi(x_O) \geq 0)$

4. Negative Predictive Values (NPV): $\Pr(\xi(x_O, x_U) < 0 \mid \hat{\Delta} < 0)$ vs $\Pr(\xi(x_O, x_U) < 0 \mid \xi(x_O) < 0)$

Results

Table 1 compares the performance of traditional IV estimates and constructed CATE(x_O) values to PeT estimates of mean treatment effect parameters. As expected, in the presence of essential heterogeneity, the traditional IV estimates do not correspond to any of the mean treatment effect parameters. In case of binary and non-negative outcomes, the IV estimates have signs opposite to the true value of Average Treatment Effect (ATT) and Effect on the Treated (TT). Increase in sample size has no influence on these results.

CATE(x_O) values when aggregated over all individuals provides an unbiased and consistent estimators for the ATE. But when aggregated over subjects choosing or not choosing treatment, it provides a biased estimator for TT or TUT respectively. Although this is expected since TT and TUT are influenced by levels of X_U that are not accounted for in CATE(x_O) values, this limitation of CATE(x_O) has strong implications for heterogeneity estimated from randomized studies. If these studies fail to measure certain factors, which would then by acted on by subjects in the population, then heterogeneous treatment effects estimated from randomized setting cannot be used to forecast population impacts of access to these treatments.

Finally, the PeT effects were found to be consistent estimators of all of the mean treatment effect parameters and maintained appropriate coverage probabilities at 95% even at larger sample sizes.

Table 2 presents the performance of CATE(x_O) values versus PeT effects as compared to true individualized effects $\xi(x_O, x_U)$. Across the board, we find that the PeT effects are better correlated with $\xi(x_O, x_U)$, explain a large amount of variance of $\xi(x_O, x_U)$ and have higher PPV and NPV compared to CATE(x_O) values. All differences were statistically significant even in datasets with sample size of 5000.

Figure 1 illustrates how PeT effects can be trivially aggregated across deciles of a covariate to form CATEs for that decile. The figures show that, for all types of outcomes, the PeT based aggregation provides unbiased estimates for the true CATEs, maintaining appropriate coverage probabilities. Results arising out of sample size of 20,000 are not shown as they convey the same message.

The simulations presents strong evidence that the PeT estimates can provide consistent and nuanced individual-level treatment effects in observational data.

3. DISTRIBUTIONAL IMPACTS OF PROSTATE CANCER TREATMENTS ON 7 YEAR COSTS AND SURVIVAL.[9]

Background

We study the distributional effects of alternative treatment modalities on health and economic outcomes in prostate cancer patients using PeT Effects. Note that although this empirical example is set to look at an evaluation in health care, the methods employed have broad applicability to a wide variety of evaluations across many different fields.

Prostate Cancer Treatment Evaluations.

Prostate cancer (PCa) is the most commonly detected non-cutaneous malignancy among American men (Landis et al. 1999) with more than 186,000 cases diagnosed in 2008 and more than 28,000 men dying from the disease (Jemal et al. 2008). As the cohort of "baby boomers" age, the incidence and prevalence of PCa will likely continue to increase as long as contemporary screening patterns continue. Here we compare two treatment strategies: Surgery versus active surveillance (AS), in terms of 7-year costs and survival for elderly men diagnosed with early-stage (clinically-localized) prostate cancer. The broad rationale for looking at these patients and these treatment modalities can be found elsewhere (Hadley et al. 2010). Most importantly, many have

[9] Previous results presented were found to be driven by a programming error. Results are corrected and updated as of June 15, 2012.

argued that the rapid growth in costs of prostate cancer treatments does not fit in line with the clinical benefits that the sole randomized study in this area have shown (Holmberg et al, 2002). Result of that randomized study showed that among elderly patients, surgery and AS produces 8-year survival probabilities of 77.4% and 78.6%, respectively (p= 0.78) (Bill-Axelon et al. 2008).[10]

However, many factors render this RCT evidence to be obsolete. Besides the fact that the above RCT was not powered to look at differences among the elderly group of patients, life expectancies for elderly individuals have dramatically improved over the last two decades. Between 1975 and 2005, 15-year survival probabilities for 65-year old men have increased by 17 percentage points in the US (Muenning and Glied 2010). This indicates that the survival gains from eliminating cancer are likely to be more than those twenty-years ago, even when the underlying disease progression from diagnosis had remained the same. Moreover, with more aggressive screening regimen implemented during the late 1980s and early 1990's, and especially with the advent of prostate-specific antigen (PSA) screening, distribution of PCa diagnosed among elderly men in the late 1990s were less advanced than those diagnosed during the pre-PSA era. Last, but not the least, the quality of surgery has arisen over the past two decades as evident from the declining morbidity from such procedures. Therefore, exploring the casual effect of surgery versus AS among elderly patients with PCa using recent data becomes important.

Data

Our data comes from the 1995 – 2009 SEER-Medicare linked dataset. SEER is an epidemiologic surveillance system consisting of population-based tumor registries designed to track cancer incidence and survival in the United States. The SEER-Medicare data links claims for health services collected by Medicare for its beneficiaries to the SEER registry (Cooper et al. 2002; Viring et al. 2002). We extracted data for patients of age 66 years or older and who were diagnosed with prostate cancer between 1995 and 2002. The data contains zip codes for patient residences which were used to link to Hospital Referral Regions (HRR) identifiers and HRR- year-specific

[10] Certainly, benefits in other dimensions, such as quality of life, are not captured in these studies and also in our analyses. We delegate this to future work.

characteristics based on the Dartmouth Atlas Data[11]. We used the linked claims data from these patients for up to December 2009 or their death if that happen before December 2009. We have 7 years follow-up data for everyone in our sample. The key variables in our sample are categorized as (a) Outcomes Variables (Y); Treatment (D); Independent Risk Factors (X_O); Instrumental Variable (Z). These categories are common to any type of evaluation analysis.

(a) Outcomes Variables: We look at two outcomes. On the benefits side we use a binary indicator for 7-year overall survival. On the costs side, we use the total undiscounted 7-year expenditures on health care expressed in 2009 dollars. Expenditures accumulate over all types of medical costs reimbursed by Medicare or a third party payer and patients' out-of-pocket costs.

(b) Treatment (D): Comparison is made between the use of surgery (without any form of radiation of hormone therapy) in the first six months of diagnosis versus active surveillance that is defined as no use of surgery, hormone therapy or radiation in the first six months of diagnosis along with at least two PSA tests within first year of diagnosis. Treatment indicator takes a value of one for surgery.

An indicator of surgery is likely to be endogenous for three reasons: True severity of cancer is unobserved as we only have data on the cross-sectional characteristics of the tumor at diagnosis, but not how the tumor is growing or prostate-specific antigen (PSA) levels (used to detect prostate cancer) is rising. Higher severity may be positively correlated with surgery receipt and also negatively correlated with survival, but positively correlated with costs.[12] These correlations render the naïve effects on surgery to be biased downward and that on costs to be biased upward. Second, general frailties of the patients are unobserved, which again would follow the same correlations as tumor severity.[13] Third, psychological anxiety of being diagnosed with cancer would be positively correlated with both surgery receipt and costs and utilizations. Its correlation with survival remains to be ambiguous.

---

[11] http://www.dartmouthatlas.org/
[12] Decreased survival within a fixed window of time is usually associated with higher costs due to expenditure spikes at the end of life (Brown, 2002).
[13] Although one may expect that higher frailty would be negatively correlated with surgery, our first stage regression shows that patients with more number of hospitalization and more comorbidities are more likely to get surgery.

(c) Independent Risk Factors (X_O): These include clinical stage and grade of cancer for patients at diagnosis using standard definitions (Meltzer et al. 2001), demographics, indicator for metropolitan area, Elixhauser comorbidity indices based on hospitalization in year preceding diagnosis, year and state fixed effects, zip-code level area characteristics on racial makeup, density and education levels. We also adjust for HRR-level characteristics using logged versions of population size, and per 100,000 patients' supply of hospital beds, physician, specialists and urologists.

(d) Instrumental Variable (Z): We use HRR-specific rates of active surveillance in prostate cancer patients in the year prior to the diagnosis of a patient. Such an instrument has been used in the past in the context of prostate cancer (Hadley et al. 2010); however, concerns exist about the contamination in area-level variations that would violate the exclusion restriction for two reasons: first, such variations may be correlated with variations in case-mix of patients; second, contamination may exist due to productivity spillovers that make areas with more efficient deliveries of treatments correlated with higher rates of treatment (Chandra and Staiger, 2007) . We try to address both of these concerns and mitigate the effect of such contaminations on the IV. In order to address the first concern, we control for many concurrent area-level fixed effects and variations as mentioned above. Contamination due to productivity spillovers (Chandra and Staiger, 2007) are directly controlled by adjusting for the number of urologist per capita as the urologist are the main specialists delivering surgery for prostate cancer patients. We study the properties of our IV after controlling for these factors and believe that it meets the requirements for a valid and strong instrumental variable.

Methods

We study the strength of the IV in a logistic model for surgery along with all other independent risk factors. To explore plausible contamination in the IV due patient level characteristics, we run a separate logistic model for treatment with only the IV as a regressor. We then compare the imbalance in the patient-level independent risk factors across treatment categories with the imbalance in the same across the median of the IV-only predicted propensity to choose surgery. A valid IV would necessarily appear to reduce such imbalances. We explore these comparisons mainly for individual level demographic and illness severity factors after converting them to their respective z-scores.

Next, MTE's and PeT effects are estimated using standard LIV methods described in our estimation and simulation sections. For the binary survival outcome we use a logistic regression. For the expenditure outcome, we use a semiparametric generalized linear model with log link and Gamma variance. Various goodness-of-fit tests were employed to ensure good model fit to these data. We study both the mean treatment effect parameters and also the joint distribution of PeT effects across survival and costs and the implications of such distributions for treatment choices.

Results & Discussions

Our final analytic sample consists of 13,495 patients, of whom 9,913 (73.5%) received surgery. As evident from the first-stage regression results in Table 1, likelihood of receiving surgery increases with ages younger and older than 74 years, T1 stage, advancing grade, and increased number of hospitalization in previous year. The instrumental variable was found to be strongly predictive of surgery receipt conditional on other factors (F-stat: 10.9, $p<0.0001$).

Figure 2 illustrates that the IV may be particularly suitable in reducing residual confounding in this application since it is able to reduce imbalance in observed factors considerably.[14] The identified support of the IV-based predicted propensity score (PS) ranges from 0.07 to 0.995.

Polynomials of propensity scores were not found to be significant in either of the LIV models. The final models for either outcomes contained covariates, interaction of covariates with PS and PS.[15] This indicates that essential heterogeneity is small for these outcomes.[16] This is presumably because we capture a very rich array of

[14] An LPM version of the IV model rejects under-identification of the IV ($p<0.0001$) and passes the weak identification test based on its F-stat.
[15] The models passed all goodness-of-fit tests. No systematic biases were detected from residual analyses.
[16] Note that since we use non-linear models, absence of polynomial of PS does not mean absence of essential heterogeneity in the additive scale, which is our scale of interest.

observed factors and estimate significant treatment effect heterogeneity across those factors. In essence, in this application PeT effects become similar to CATEs, where conditioning is done on the entire vector of observed factors. The mean treatment effect estimates are given in Table 4. The average treatment effect was estimated to be -$30,056 and 7.4%pts for costs and survival respectively, which were not significant. The average survival effects, although not significant, indicates the potential for substantial benefits of surgery over AS in this population, which are in stark contrast to the results from the largest and only randomized trial comparing these two treatments that was conducted on patients diagnosed with PCa about a quarter century ago (Holmberg et al 2002).

Figure 3 illustrates the joint distribution of PeT effects for 7-year survival and costs in an incremental cost-effectiveness plane where the X-axis represents PeT effects on survival and the Y-axis the PeT effects on costs. Each dot on the plane represents a patient. The size of the treatment effect marker for each patient is driven by the z-score of their respective treatment effect. Patients with more significant effects have larger markers. The correlation between estimated PeT effects on costs and survival was small: 0.03 (95% CI: -0.20, 0.25). Only 21% of patients were found to have negative incremental survival from surgery. Surgery was found to be a *dominant* treatment in 61% of patients (South-East quadrant of graph) as it incurs lower costs and increased survival.

There is little evidence of positive self-selection in practice. This is not surprising given that a majority of patients appear to benefit from surgery. Surgery rates were 74% among patients for whom surgery produces negative effects versus 73% among those who would benefit from surgery. This is reflected in the estimates for the effect on the treated (TT) and the untreated (TUT) (Table 4). Both TT and TUT are identical to ATE and neither reach statistical significant for costs or survival. The heterogeneity of treatment effects illustrated in Figure 3, however, indicates that there may be much room for improvement. In a hypothetical world of perfect selection (Meltzer et al. 2003), where patients who would get hurt by surgery are removed from being eligible for comparing these two modalities of treatment, the ATE and TT of surgery would climb to 12.6%pt (95% CI: 2.5, 37.0) and 12.7%pt (95% CI: 1.5, 40.5) respectively for 7-year survival (Table 4). These estimates can also be used to establish the value of more targeted approach to treatment allocation. However, compared to the ATE and TT estimates without selection, the ATE and TT estimates with perfect selection indicate only modest cost savings and better survival (Table 4), which do not reach statistical significance.

Finally, PeT effects can be used to explore the dimensions (factors) along which treatment selections are efficient (i.e. they conform to gains) and where they are inefficient (i.e. they conform to losses). Compared to patients for whom survival effects are significantly positive (at 10% level), patients with significant negative survival effects with surgery had significantly higher rates of well grade cancer, higher number of pre-period hospitalization and higher rates of every comorbidities listed in Table 3 except for peripheral vascular disease. Future work can use the estimated PeT effects and its uncertainties to develop prediction algorithms for treatment effects.

4. CONCLUSIONS

This paper interprets a treatment effect parameter, originally defined by Heckman and Vytlacil (1999), to represent Person-centered Treatment (PeT) effects. Heckman and Vytlacil (1999) use this parameter to establish relationship between the mean treatment effect parameters such as LATE, ATE, TT and TUT with the Marginal Treatment Effect (MTE) parameter but do not use it further. A PeT effect is derived as an alternate weighting of MTEs and is shown to represent individualized treatment effects that not only conditions on the individual's observed characteristics but also averages over a conditional distribution of unobserved characteristics (in contrast to their marginal distributions as in CATEs) that conditions on treatment choice made by an individual and the circumstances under which that choice were made. The paper presents the theory behind PeT and proposes semiparametric estimators to estimate PeT effects using instrumental variables.

Finite sample simulations show that, in the present of essential heterogeneity (Heckman 1997), the PeT effects may explain a significantly larger fraction of individual-level treatment effect heterogeneity compared to CATEs. Therefore, in the absence of data that can help identify the full joint distribution of potential outcomes, PeT effects can serve as a valuable addition to the evaluation literature looking at the distributional impacts of treatment access and policies. Moreover, they truly mimic individual level treatment effects as they can be trivially aggregated across all patients, or across patients who did or did not choose treatment in order to construct estimates of ATE, TT or TUT respectively.

The introduction of PeT effects and its role in identifying treatment effect heterogeneity lines up well with the political economy of health care evaluations. Despite the age-old practice of evaluating health care technologies using randomized trial and more recently with observational data that were used to estimate average treatment effects (and often local average effects), the Affordable Care Act of 2010 specifically ask for producing estimates at a more nuanced and individualized level. It created a Patient Centered Outcomes Research Institute (PCORI) as an independent, non-profit research organization to conduct research to provide information about the best available evidence to help patients and their health care providers make more informed decisions. Its mission is to help people make informed health care decisions – and improves health care delivery and outcomes – by producing and promoting high integrity, evidence-based information – that comes from research guided by patients, caregivers and the broader health care community (PCORI Mission Statement, 2011). PCORI is positioned to be one of the largest funders of outcomes research in the United States in the coming years and has so far asserted that one of the primary focus in patient-centered outcomes research (PCOR) should be answering the question for patients: "Given my personal characteristics, conditions and preferences, what should I expect will happen to me?".

While CATEs can provide answers to these questions, estimating CATEs directly based on multiple observed covariates can be tricky. In contrast, PeT effects can serve as outcomes that can be used to develop predictive algorithms for CATEs based on combinations of patient and other observed characteristics in the data. Such an approach would be most valuable for allocating Category II and III treatments, as defined by Chandra and Skinner (2011), since uncertainties in their comparative effectiveness either precludes them from access in some settings or facilitates rapid adoption that leads to welfare loss. Furthermore, since PeT effects allow for estimating more nuanced individual treatment effects, understanding the difference in variance between PeT effects and CATEs can help establish the value of future research that can identify factors relevant for treatment effect heterogeneity that are not collected in the current databases (Basu and Meltzer 2007).

In summary, PeT effects can serve as a useful treatment concept for a variety of evaluations both at the policy and at the individual level.

REFERENCES

Aakvik, A, Heckman JJ, Vytlacil E. 1999. Semiparametric program evaluation: lesson from an evaluation of a Norwegian training program. University of Chicago, unpublished manuscript.

Basu A, Heckman J, Navarro-Lozano S, Urzua S. 2007. Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients. *Health Economics* **16(11)**: 1133 -1157.

Basu A. 2009. Individualization at the heart of comparative effectiveness research: The time for i-CER has come. *Medical Decision Making* **29(6)**: N9-N11.

Basu A. 2011a. Economics of individualization in comparative effectiveness research and a basis for a patient-centered healthcare. 2011 *NBER Working Paper* No. w16900. *Journal of Health Economics* **30(3)**; 549-559.

Basu A. 2011b. Estimating decision-relevant comparative effects using instrumental variables. *Statistics in Biosciences* **3(1)**: 6-27.

Basu A, Jena A. Philipson T. 2011. Impact of comparative effectiveness research on health and healthcare spending. 2010 NBER Working Paper No. w15633. *Journal of Health Economics* **30(4)**: 695-76.

Bill-Axelon A, Holmberg L, Filen F, et al. 2009. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandanavian prostate cancer group 4 randomized trial. *Journal of the National Cancer Institute* **100**:1144-1154.

Brown M, GF Riley, N Schussler, Etzioni R. 2002. Estimating health care costs related to cancer treatment from SEER-Medicare data. *Medical Care* **40(8 Suppl.)**:IV-104-117.

Carniero P, Lee S. 2009. Estimating distribution of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics* **149**: 191-208.

Carniero P, Hansen KT, Heckman JJ. 2001. Removing the veil of ignorance in assessing distributional impacts of social policies. *Swedish Economic Policy Review* **8**: 273-301.

Carniero P, Hansen KT, Heckman JJ. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 2003; **44(2):** 361-422.

Chandra A, Staiger D. Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy* 2007; 115(11):103-140.

Chandra A, Skinner JS. Technology growth and expenditure growth in health care. *NBER Working Paper* 16953; 2011.

Claxton K. The Irrelevance of Inference: A Decision-Making Approach to the Stochastic Evaluation of Health Care Technologies. *Journal of Health Economics* 1999; 18: 341-364.

Cooper GS, Viring B, Klabunde CN, Schussler N, Freeman J, Warren JL. Use of SEER-Medicare data for measuring cancer surgery. *Medical Care* 2002; 40[Suppl]: IV-43 – IV-48.

Garber, Alan M. and Charles E. Phelps. "Economic Foundations of Cost-Effectiveness Analysis," *Journal of Health Economics* vol. 16, no. 1, February 1997, pp. 1-32.

Holmberg L, Bill-Axelson A, Helgesen F et al. A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. *New England Journal of Medicine* 2002; 347(11): 781-9.

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

Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. 2008. Cancer Statistics, 2008. *CA: A Cancer Journal for Clinicians* **58**: 71-96.

Landis SH, Murray T, Bolden S, Wingo PA. 1999. Cancer Statistics, 1999. *CA: A Cancer Journal for Clinicians* **49(1)**: 8-31.

Meltzer, DO, B Egleston, I Abdalla. 2001. Patterns of prostate cancer treatment by clinical stage and age in the United States. *American Journal of Public Health* **91(1)**: 126-128.

Meltzer D, Huang E, Jin L, Shook M, Chin M. 2003. Major bias in cost-effectiveness analysis due to failure to account for self-selection: impact in intensive therapy for type 2 diabetes among the elderly. *Medical Decision Making* (abstract) **23(6)**: 576.

Muening PA, Glied SA. 2010. What changes in survival rates tell us about US health care. *Health Affairs* **29(11)**: 1-9.

Heckman JJ. 1996. Comments on Angrist, Imbens, and Rubin: Identification of Causal Effects Using Instrumental Variables. *Journal of American Statistical Association* **91**: 434.

Heckman JJ. Honoré B. 1990. The empirical content of the Roy Model. *Econometrica* **58**: 1121-1149.

Heckman JJ, Smith J. 1993. Assessing the case for randomized evaluations of social programs, in K. Jensen and P. Madsen (eds.), Measuring Labor Market Outcomes, Ministry of Labor, Copenhagen.

Heckman JJ, Vytlacil EJ. 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* **96(8)**: 4730-34.

Heckman JJ. 2001. Accounting for heterogeneity, diversity and general equilibrium in evaluating social programmes. The Economic Journal 111: F654-F699.

Heckman JJ, Vytlacil E. 2001. *Local instrumental variables*. In C. Hsiao, K. Morimue, and J.L. Powell (Eds.) Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in the Honor of Takeshi Amemiya, Cambridge University Press: New York, 1-46.

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

Heckman JJ, Clements N, Smith J. 1997. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in program impacts, *Review of Economic Studies* **64**: 487-535.

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

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

Vanness DJ, Mullahy J. 2006. Perspectives on Mean-based Evaluation of Health Care. In Jones A. (Eds) The Elgar Companion to Health Economics, Edward Elgar Publishing: Cheltenham.

Viring BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. 2002. Studying radiation therapy using SEER-Medicare linked data. *Medical Care* **40[Suppl]**: IV-49 – IV-54.

Table 1: Simulation results on average effects.

OUTCOMES	True Values	N=5,000 Mean (sd*) [avg. se**] {Coverage Pr***}			N=20,000 Mean (sd*) [avg. se**] {Coverage Pr***}		
NORMAL		IV	CATE(x_O)-Based	PET	IV	CATE(x_O)-Based	PET
ATE	.50	.06 (.10) [.10] {.01}	.52 (.02) [.03] {.91}	.47 (.12) [.13] {.96}	.03 (.05) [.05] {0}	.51 (.01) [.01] {.87}	.48 (.06) [.06] {.94}
TT	1.01	.06 (.10) [.10] {0}	.75 (.03) [.03] {0}	.91 (.19) [.21] {.95}	.03 (.05) [.05] {0}	.76 (.01) [.01] {0}	.95 (.10) [.10] {.90}
TUT	-.64	.06 (.10) [.10] {0}	-.01 (.03) [.03] {0}	-.53 (.14) [.15] {.90}	.03 (.05) [.05] {0}	-.05 (.01) [.01] {0}	-.60 (.07) [.07] {.93}
BINARY		IV	CATE(x_O)-Based	PET	IV	CATE(x_O)-Based	PET
ATE	0.13	-.05 (.04) [.04] {.01}	.14 (.01) [.01] {.73}	.13 (.04) [.04] {.97}	-.06 (.02) [.02] {0}	.14 (.002) [.002] {.17}	.13 (.02) [.02] {.97}
TT	0.27	-.05 (.04) [.04] {0}	.2 (.01) [.01] {0}	.26 (.05) [.05] {.93}	-.06 (.02) [.02] {0}	.2 (.002) [.002] {0}	.26 (.02) [.03] {.94}
TUT	-0.18	-.05 (.04) [.04] {.09}	0 (.01) [.01] {0}	-.15 (.05) [.05] {.94}	-.06 (.02) [.02] {0}	-.01 (.003) [.004] {0}	-.16 (.02) [.02] {.92}
NON-NEG		IV	CATE(x_O)-Based	PET	IV	CATE(x_O)-Based	PET
ATE	0.31	-.25 (.09) [.09] {0}	.32 (.02) [.02] {.92}	.31 (.05) [.12] {.98}	-.29 (.05) [.05] {0}	.31 (0) [.01] {.89}	.31 (.02) [.05] {.99}
TT	0.61	-.25 (.09) [.09] {0}	.46 (.02) [.02] {0}	.58 (.08) [.10] {.97}	-.29 (.05) [.05] {0}	.46 (.01) [.01] {0}	.59 (.04) [.04] {.97}
TUT	-0.36	-.25 (.09) [.09] {.74}	-.01 (.02) [.01] {0}	-.31 (.08) [.18] {.98}	-.29 (.05) [.05] {.60}	-.03 (.01) [.01] {0}	-.33 (.04) [.08] {.99}

CATE(x_O) are constructed based on true values.
* Standard deviation across 1000 Monte Carlo replicates.
** Based on variance estimate from 500 bootstrap samples for each Monte Carlo replicate data and averaged over all Monte Carlo replicates.
*** Coverage indicator of true values based on 95% CI estimate from 500 bootstrap samples for each Monte Carlo replicate data and the indicator averaged over all Monte Carlo replicates.

Table 2: Simulation results on distributional effects.

OUTCOMES	N=5,000 Mean (sd*)				N=20,000 Mean (sd*)			
NORMAL	Corr. with §(x_{Oi}, x_{Ui})	R^2 on §(x_{Oi}, x_{Ui})	PPV	NPV	Corr. with §(x_{Oi}, x_{Ui})	R^2 on §(x_{Oi}, x_{Ui})	PPV	NPV
CATE(x_O)-Based	.70 (.01)	.49 (.01)	.79 (.01)	.70 (.01)	.7 (.003)	.5 (.005)	.79 (.004)	.71 (.005)
PET estimates ($\hat{\Delta}_i$)	.77 (.01)	.59 (.02)	.83 (.01)	.76 (.03)	.77 (.004)	.6 (.005)	.83 (.007)	.77 (.014)
Difference [p-val]	.07 (.01) [<.001]	.1 (.015) [<.001]	.04 (.015) [.009]	.06 (.031) [.071]	.07 (.003) [<.001]	.1 (.005) [<.001]	.04 (.007) [<.001]	.06 (.015) [<.001]
BINARY	Corr. with §(x_{Oi}, x_{Ui})	R^2 on §(x_{Oi}, x_{Ui})	PPV	NPV	Corr. with §(x_{Oi}, x_{Ui})	R^2 on §(x_{Oi}, x_{Ui})	PPV	NPV
CATE(x_O)-Based	.7 (.008)	.48 (.011)	.79 (.008)	.7 (.012)	.7 (.003)	.49 (.005)	.79 (.004)	.71 (.005)
PET estimates ($\hat{\Delta}_i$)	.76 (.016)	.58 (.024)	.83 (.018)	.76 (.037)	.78 (.004)	.6 (.006)	.83 (.009)	.77 (.017)
Difference [p-val]	.07 (.016) [<.001]	.10 (.024) [<.001]	.04 (.018) [.035]	.05 (.038) [.156]	.07 (.004) [<.001]	.11 (.006) [<.001]	.04 (.009) [<.001]	.06 (.017) [<.001]
NON-NEGATIVE	Corr. with §(x_{Oi}, x_{Ui})	R^2 on §(x_{Oi}, x_{Ui})	PPV	NPV	Corr. with §(x_{Oi}, x_{Ui})	R^2 on §(x_{Oi}, x_{Ui})	PPV	NPV
CATE(x_O)-Based	.69 (.009)	.47 (.012)	.79 (.008)	.70 (.012)	.69 (.004)	.48 (.006)	.79 (.004)	.71 (.005)
PET estimates ($\hat{\Delta}_i$)	.72 (.012)	.52 (.017)	.83 (.014)	.76 (.025)	.73 (.005)	.53 (.008)	.83 (.007)	.77 (.012)
Difference [p-val]	.04 (.011) [<0.001]	.05 (.015) [<0.001]	.04 (.013) [.002]	.06 (.025) [.027]	.04 (.004) [<.001]	.05 (.006) [<.001]	.04 (.006) [<.001]	.06 (.013) [<.001]

CATE(x_O) are constructed based on true values. PPV: Positive Predictive Value; NPV: Negative Predictive Value.
* Standard deviation across 1000 Monte Carlo replicates.

Table 3: First-stage results from logistic regression on Surgery indicator.

Covariates	Logit coefficients (std. err.) [z-stat]
IV	
ivrate_activesurv	-1.496 (0.5) [-3.02]++
DEMOGRAPHICS	
Age (centered at 74)	-0.176 (0.01) [-30.95]++
Age^2	0.0124 (0) [17.89]++
T1-stage (Ref: T2)	1.05 (0.05) [22.11]++
Grade – Well (Ref: Undetermined)	1.402 (0.14) [9.67]++
Grade – Moderate	1.424 (0.13) [11.04]++
Grade – Poor	2.261 (0.14) [16.14]++
White (Ref: Other)	-0.425 (0.15) [-2.76]++
Black	-0.347 (0.19) [-1.82]+
Hispanic	-0.089 (0.23) [-0.39]
Metropolitan area of residence	-0.052 (0.09) [-0.58]
ILLNESS SEVERITY	
1 hospitalization last year (Ref: No hosp)	0.283 (0.09) [3.02]++
2 hospitalizations last year	0.288 (0.15) [1.87]+
>2 hospitalizations last year	0.545 (0.21) [2.6]++
Congestive heart failure	0.338 (0.21) [1.59]
Valvular disease	-0.113 (0.23) [-0.48]
Peripheral vascular disease	0.02 (0.21) [0.1]
Paralysis	0.638 (0.34) [1.89]+
Other neurological disorders	-0.22 (0.23) [-0.97]
Chronic Lung Disease	0.13 (0.14) [0.9]
Diabetes	0.05 (0.16) [0.32]
Diabetes with chronic complications	0.226 (0.36) [0.63]
Hypothyroidism	0.232 (0.26) [0.88]
Obesity	-0.03 (0.36) [-0.08]
Fluid and electrolyte disorders	0.136 (0.15) [0.88]
Deficiency Amemias	0.258 (0.2) [1.28]
Alcohol abuse	0.116 (0.35) [0.34]
Depression	0.167 (0.31) [0.54]
Hypertension with complications	-0.053 (0.11) [-0.48]
ZIPCODE-LEVEL 2000 CENSUS XTICS	YES
YEAR FIXED EFFECTS	YES
STATE FIXED EFFECTS	YES
HRR-SPECIFIC XTICS	YES

+ p-val< 0.10; ++ p-val < 0.05.

Table 4: Mean treatment effects based on estimated PeT effects (Surgery versus Active surveillance)

Effects	7 –year Costs, 2009 $ Mean (95% CI*)	7-year Surv. Pr., %Pt Mean (95% CI*)
Average Treatment Effect (ATE)	-30,056	7.4
	(-115,807, 19,355)	(-17.7, 40.2)
Effect on the Treated (TT)	-28,191	7.4
	(-115,877, 20,451)	(-17.1, 43.2)
Effect on the UnTreated (TUT)	-35,255	7.4
	(-109,419, 16,543)	(-19.3, 30.8)
TT - TUT	7064	0.01
	(-8,969, 15,340)	(-5.8, 13.3)
With Perfect Selection on Survival PeTs		
Average Treatment Effect	-30,641	**12.6**
	(-119,640, 22,116)	**(2.5, 37.0)**
Effect on the Treated	-28,332	**12.7**
	(-119,927, 22,859)	**(1.5, 40.5)**
Effect on the UnTreated	-28,332	**12.4**
	(-117,645, 17,882)	**(5.1, 26.7)**
Gains with Perfect Selection		
ATE(Sel) – ATE	-585	5.2
	(-14,394, 4,885)	(-3.4, 23.3)
TT(Sel) - TT	-141	5.3
	(-14,035, 5,475)	(-3.0, 21.9)

* 95% CI based on bias-corrected estimates from 1000 bootstrap replicate.
Bold face indicates exclusion of zero from 95% CI.

Figure 1: PeT-based conditional treatment effects for N=5000.

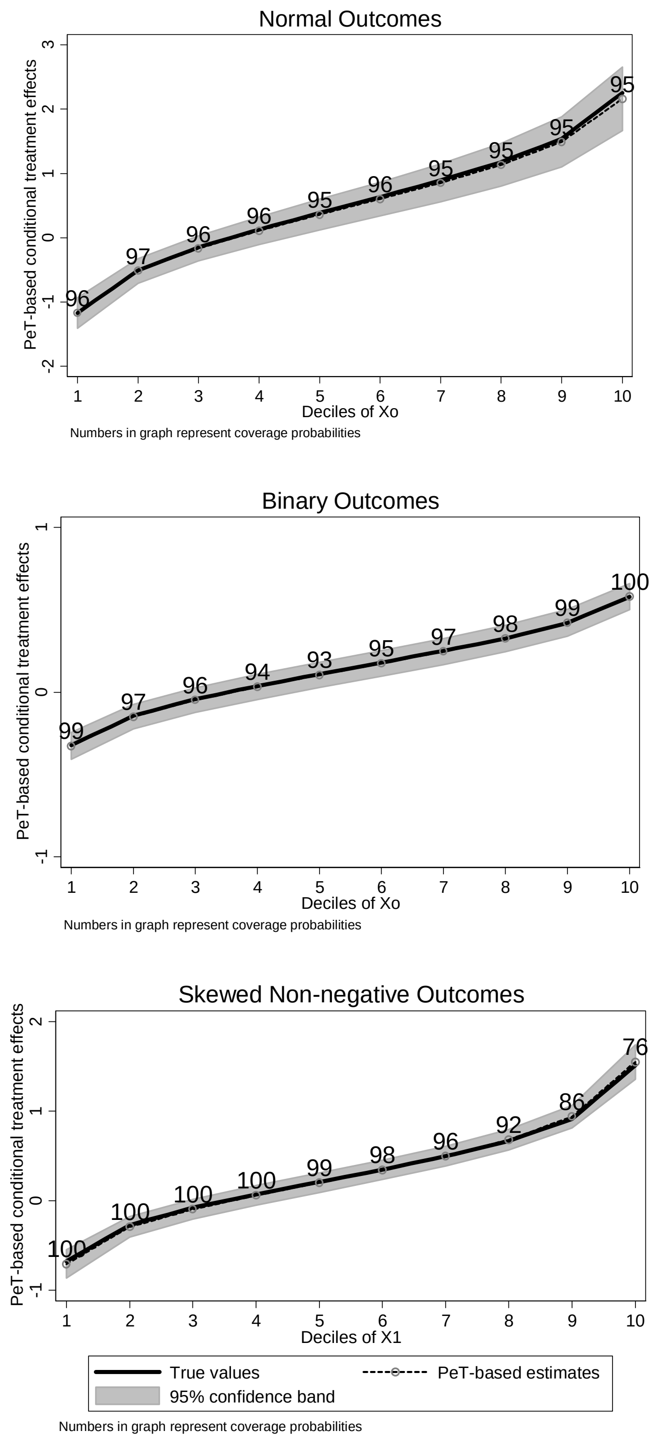

Figure 2: Covariate imbalance across treatments versus across instrumental variable.

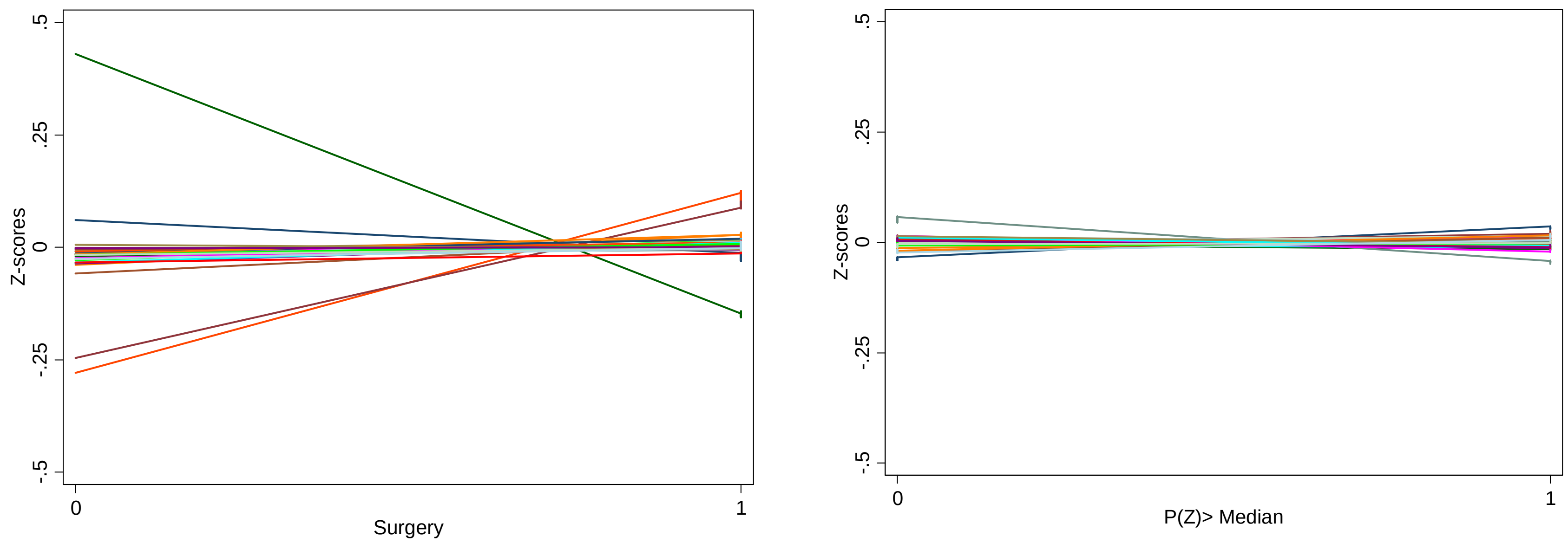

Figure 3: Distribution of PeT effects on survival and costs, differentially illustrated by significance.