

NBER WORKING PAPER SERIES

ECONOMICS OF INDIVIDUALIZATION IN COMPARATIVE EFFECTIVENESS
RESEARCH AND A BASIS FOR A PATIENT-CENTERED HEALTH CARE

Anirban Basu

Working Paper 16900
<http://www.nber.org/papers/w16900>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2011

This work was supported by the Alan Williams Health Economics Fellowship from the University of York and a National Institute of Mental Health Research grant 1R01MH083706. The author is grateful to Karl Claxton, James Heckman, Thomas McGuire, David Meltzer, Harold Pollack, Rena Conti, Mark Sculpher, Sergio Urzua and two anonymous reviewers for their constructive comments on a previous draft or presentations of this work. The opinions expressed here are those of the author only, and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by Anirban Basu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Economics of Individualization in Comparative Effectiveness Research and a Basis for a Patient-Centered Health Care

Anirban Basu

NBER Working Paper No. 16900

March 2011

JEL No. C11,D61,I18

ABSTRACT

The United States aspires to use information from comparative effectiveness research (CER) to reduce waste and contain costs without instituting a formal rationing mechanism or compromising patient or physician autonomy with regard to treatment choices. With such ambitious goals, traditional combinations of research designs and analytical methods used in CER may lead to disappointing results. In this paper, I study how alternate regimes of comparative effectiveness information help shape the marginal benefits (demand) curve in the population and how such perceived demand curves impact decision-making at the individual patient level and welfare at the societal level. I highlight the need to individualize comparative effectiveness research in order to generate the true (normative) demand curve for treatments. I discuss methodological principles that guide research designs for such studies. Using an example of the comparative effect of substance abuse treatments on crime, I use novel econometric methods to salvage individualized information from an existing dataset.

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

1. Introduction

Decisions in health care policy and clinical medicine most often involve evaluation of alternatives: Which treatment will produce the optimal clinical outcome in a given patient? What type of insurance coverage will maximize social welfare? The fundamental problem of ex-ante evaluation is to predict outcomes under alternative policies or interventions. A large literature, dating from the early 1920s, focuses on developing methods to address such problems of evaluation (Neyman, 1923; Fisher, 1935). In recent years, health care-related methods and applications that belong to this genre of evaluation have been broadly grouped under the umbrella term “comparative effectiveness research” (CER). CER conducts head-to-head comparisons to determine which drugs, devices, and procedures are most effective and carry the lowest risk. The 2009 American Recovery and Reinvestment Act dedicates \$1.1 billion to this end.

The Congressional Budget Office, the Medicare Payment Advisory Commission (MedPAC), and other independent stakeholders define comparative effectiveness research as “a rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients” (MedPAC Report, 2008). Notably, “the set of patients” is usually defined by broad and easily observable criteria. In a recent report submitted to the President and Congress, the Federal Coordinating Council on Comparative Effectiveness Research states: “Clinicians and patients need to know not only that a treatment works on average but also which interventions work best for specific types of patients (e.g. the elderly, racial and ethnic minorities)” (FCC Report, June 30, 2009). Addressing heterogeneity in outcomes and especially in the comparative effects of alternative treatments is becoming increasingly popular in such evaluations (Sculpher, 2008).

Recent legislation,¹ however, calls for pursuing even finer levels of heterogeneity, including “genetic and molecular sub-types or quality of life preferences”, through CER. Such an approach may be the key to the success of translating comparative effectiveness information into practice. The popular mantra of “what works” associated with CER may not take us far enough unless we try to discover “what works for whom” in an individualized manner. In this paper, I

¹ Patient Protection and Affordable Care Act of 2009, H.R. 3590, 111th Congress §6301 (2010).

highlight the economics of individualization in CER and discuss methodological principles behind individualization.

In our analysis, the economics of CER information is driven by the physician-patient dyad. This dyad is assumed to work together and choose among alternative treatments based on its perceived comparative information on benefits of treatments (throughout the paper, I regard benefits to represent benefits net of harms). Comparative effectiveness and other evaluative research reveal information on the benefits of treatments, which helps the physician-patient dyads form perceptions of comparative effects and guides their treatment choice behavior. At the population level, aggregate demand is affected. In line with the goals of comparative effectiveness research, the question we will ask is how the alternative treatments in question will be utilized and how optimal co-insurance rates would vary from a social insurer's perspective under alternative informational regimes of CER.

The benefits anticipated by the physician-patient dyad before taking a treatment may not always coincide with the true benefits experienced by a patient. Anticipated benefits rely on CER results, which reveal incomplete information, such as average benefits. In contrast to anticipated benefits, a normative ("true") marginal benefits curve can be defined by the actual realization of the benefits of treatments across individuals. Therefore, unless CER generates information to precisely predict these ex-post outcomes for each patient, the true marginal benefit schedule will differ from market demand.

In Section 2, I highlight the link between CER information and decision-making and discuss the potential deadweight losses induced by different levels of information arising out of CER.

In Section 3, I discuss methodological approaches to CER that can generate more nuanced² comparative information than what is typically produced by this line of research. Specifically, I argue and provide rationale for an explicit and intensive program within CER that is devoted towards discovering effects at the individual level.

To highlight the potential of novel econometric approaches to identify individual heterogeneity in treatment effects, I draw from the economics literature

² That is, richer information on heterogeneity rather than higher precision of the average effects.

on program evaluation (Heckman and Robb, 1985; Heckman 1990, 1992; Manski and Garfinkel, 1992; Heckman and Smith, 1998; Dehejia, 2005) and present an illustrative example of the comparative effectiveness of substance abuse treatments. I discuss how CER programs can build on such analyses by collecting additional data that can help target specific treatments to patients. A broader discussion on how individualized comparative effects can be generated follows in Section 4.

2. Comparative effectiveness information and decision making

Let us begin with a problem of evaluating the comparative effectiveness of a new treatment compared to a control/standard treatment for a population of N patients indexed by i . Standard treatment may also include the do-nothing option. Let the individual-level true treatment effect represent the benefits (net of harms) of the new treatment over the control and is denoted by b_i .³ Let p denote the price of the new treatment which is also the marginal cost for manufacturing the new treatment.⁴ (We will later relax the assumption of a constant marginal cost curve and allow the price to be endogenous and vary with aggregate demand.)

Patients are members of risk classes, Ω_k , $k \leq N$, which determine heterogeneity in treatment effects across individuals through a production function $b()$, i.e. $b_i = b(\Omega_k)$. Risk classes may be defined by the combination of various risk factors such as severity of illness, patient demographics, their genetic makeup, preferences and many others factors. In order to predict an individual-level treatment effect, one requires knowledge about both the patient's risk class Ω_k and the production function $b()$. CER is responsible for estimating the production function.

There are two types of decision makers, 1) the patient-physician dyad, which we will refer to as the individual decision maker, is assumed to always have knowledge about their risk class; and 2) an insurer or social planner who decides the coinsurance rate for providing health insurance coverage for the new treatment.

³ b_i is composed of two parts, one that represents the potential health benefits of treatment and the other that represents the potential harms of treatment including adverse events and the costs of accessing the treatment (e.g. time costs etc). This is also in line with the Net Health Benefit's approach in cost-effectiveness analysis (Stinnett and Mullahy, 1998).

⁴ Here p can be seen as a normalized price, where the price of the standard/control treatment is assumed to be zero.

2.1. The First best scenario

Under complete information, both the insurer and the individuals are aware of the risk classes and the production function and are able to perfectly predict b_i . Here, individuals will choose treatment only if $b_i - p \geq 0$, when they are exposed to the full price of treatments. This is efficient. If individuals had full insurance they would choose treatment if $b_i \geq 0$. Since the insurer can fully anticipate this individual behavior, she can provide full coverage for treatment only for those individuals who would experience benefits greater than cost and not provide coverage for the rest. Thus, there is no efficiency loss due to moral hazard.

Difficulties in achieving the first best scenario are mostly driven by three conditions or their combinations. These conditions are the asymmetry of information between decision makers, imperfect information to the decision makers and incomplete information.

2.2. Scenario with asymmetry of information

Asymmetry of information arises when individuals are assumed to be aware of Ω_k and $b()$ and to be able to combine them to predict b_i perfectly, but the insurer cannot. Pauly and Blavin (2008) discuss the case of insurer having no information on Ω_k . In our terms, the insurer cannot identify individuals who belong to specific risk classes even when they are aware of the production function $b()$. Consequently, the insurer cannot exclude patients from coverage who would get treatment benefits lower than the cost of treatment (i.e. $b_i - p < 0$). This leads to moral hazard (Pauly, 2008) and the insurer may offer coverage with a fractional coinsurance rate (c), which is the fraction of price a patient must pay in order to receive treatment. Figure 1 illustrates this situation. The coinsurance rate is designed to deter individuals (i.e. for whom $b_i - c \cdot p < 0$) from taking treatments so that the resulting moral hazard is traded off with the benefits of risk protection that insurance coverage will provide (Manning and Marquis, 1996).

One important aspect, which is typically kept implicit in most analyses under asymmetry of information, is that patients who perceive $b_i < 0$ (i.e. the patients at the extensive margin), would automatically decline treatment even when such treatment is offered for free.

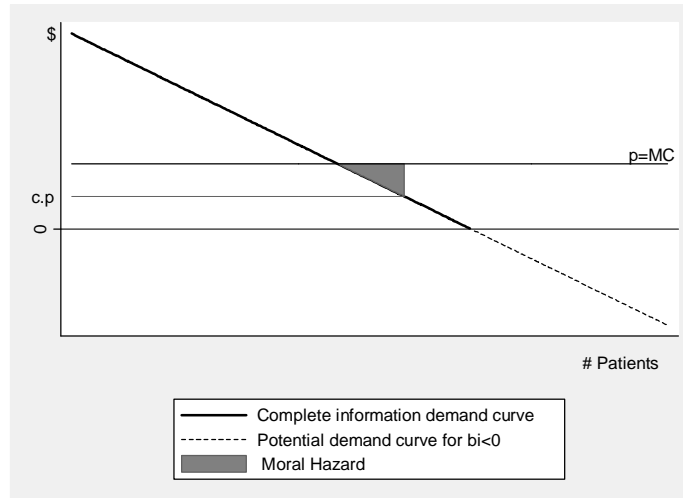


Figure 1: Moral hazard with asymmetric information.

In Figure 1, the potential demand curve for patients with $b_i < 0$ is shown with dotted lines. Theoretically, it corresponds to having to pay patients money to accept treatments that produce negative benefits. Such behavior is documented in the literature on compensating wage-differentials where consumers accept higher wages to accept riskier jobs (Thaler and Rosen 1975). However, in health care, such behavior is non-existent and therefore when individuals have complete information about b_i , there is no demand for treatment among those with $b_i < 0$. That is patients will be demand-responsive to treatment (Chakley and Khalil, 2005). The implications of such self-selection will be apparent below when these patients fail to make such selections based on available information.⁵

2.2. Scenario with imperfect information

Pauly and Blavin (2008) extends their analyses to study the scenario where patients do "not have demand curves that map the marginal benefit curves generated by clinical evidence". That is, even though knowledge on $b()$ is available, patients have imperfect information and are unable to precisely predict b_i . Individuals end up either under or overestimating the benefits of treatment

⁵ Such self selection behavior is studied extensively in the literature on program evaluation and structural modeling (Heckman and Robb 1985; Heckman 1992; Manski and Garfinkel 1992; Heckman and Smith 1998). In the context of health care, Meltzer et al. (2003) discuss how such self-selection behavior based on patient preference heterogeneity may affect cost-effectiveness analyses.

conditional on their risk class. Pauly and Blavin show how optimal coinsurance rates varies under these conditions and discuss potential ways to address the inefficiency, which includes value-based insurance designs that can lower copayments when patients underestimate benefits (Pauly and Held 1990) and vice versa.

2.3. Scenarios with incomplete information

Building on Pauly and Blavin's analyses, we study the case when clinical evidence reveals incomplete information about the true (normative) marginal benefit curve. We assume that individuals possess perfect information on clinical evidence, but this evidence may be incomplete.⁶ We argue that moral hazard due to insurance coverage is aggravated when clinical evidence only produces information on average benefits.

CER estimating average effects: Let individuals and the insurer rely on a CER study that reveals information about the average incremental benefits of the new treatment over control: $B = E_i\{b(\Omega_k)\}$. This expectation is over all patients including those who would be harmed by treatment ($b_i < 0$). Individuals and the insurer do not have any additional information about the variability of comparative effectiveness in the population. Consequently, individuals have a demand curve that is simply an indicator, $I(B > c \cdot p)$, and that is same for all individuals. If new treatment is cost-effective on average, (i.e. $B > p$), then for any co-pay $c \leq 1$, all patients will choose to receive the new treatment producing a large aggregate demand for the treatment and an associated large amount of moral hazard as shown in Figure 2.

Since average comparative effectiveness information leads to higher demand responsiveness to price and lower demand responsiveness to treatments, outcomes-based payments may not be worthwhile (Chakley and Khalil, 2005). Consequently, value-based insurance designs, which promote the use of treatments that are cost-effective on average (Fendrick et al., 2001), will also promote this inefficiency due to moral hazard.

⁶ This assumption can be relaxed and results from Blavin and Pauly (2008) and Pauly and Held (1990) can be applied over and above the results we obtain here.

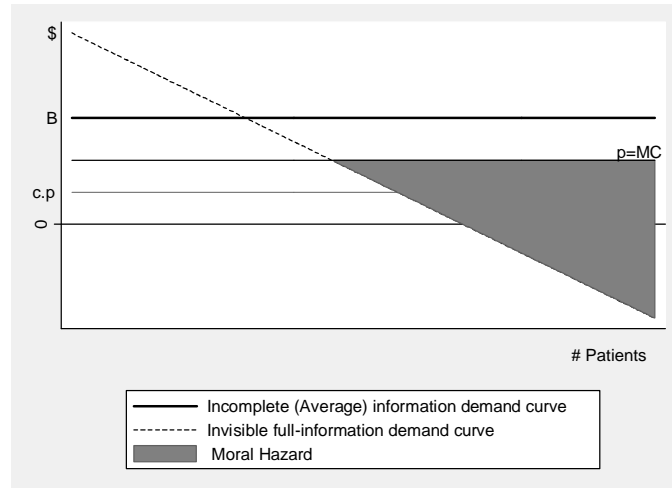


Figure 2: Moral hazard with incomplete (average) CER information.

Given various constraints on time and production and restricted market entry for medical treatments, the marginal cost curve may be increasing over quantity in many realistic situations. In such a scenario, the moral hazard loss is worsened by the increase in equilibrium prices that accompanies the increase in aggregate demand (Jena and Philipson, 2009; Basu and Philipson, 2010) (Figure 3). In fact, the equilibrium price may end up being higher than p and even B . Consequently, besides a larger moral hazard, a portion of the consumer surplus and wealth will be transferred to the producers. Combined together, they would incur a large strain on the demand of health insurance due to the high premiums that the insurer needs to charge in order to offset such inefficiencies.

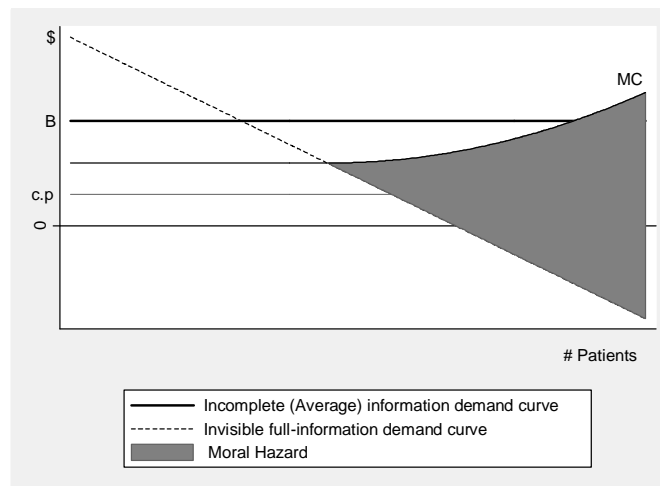


Figure 3: Moral hazard with incomplete (average) CER information and upward sloping marginal cost curve.

Individualization using nuanced subgroup-specific CER: If CER can evolve to reveal more information on risk classes and their associated treatment effects, individual demand curves will begin to map to the true normative marginal benefit curves. At the limit, when sufficient heterogeneity in comparative effects is revealed it is possible to approach the first best solution described above.

Thus with individualized information, expected demand curves begin to be less price responsive and more responsive to treatments. This enhances the role of VBID both as an outcome-based payment mechanism (Chakley and Khalil, 2005) and also in scenarios when clinical evidence map the true marginal benefits curve but the patients have imperfect information and their perception of benefits fall short of it (Pauly and Blavin, 2008).

2.3. Scenarios with combinations of incomplete, imperfect and asymmetry of information

Consider the situation when the insurer relies on a CER that reveals incomplete information about the average benefits of alternative treatments, B . The insurer also believes that individuals do not possess any private information that can enable them to predict individualized treatment effects. However, individuals may have additional information on $b()$ so that that can anticipate b_i based on their own risk class. This creates asymmetry of information between insurer and individuals. Two specific situations where such anticipation is plausible are: 1) where medical knowledge and experience enable physicians to help patients anticipate b_i , rather imperfectly, based on levels of their risk classes, even though no formal CER estimates on such heterogeneity exist. 2) Individuals have an opportunity to reveal b_i perfectly, using trial and error methods on both treatments within a short period of time, without any noticeable effects on their long-term benefits (i.e. costless trial and error method) and then self-select into treatments that maximize their individual benefit. Such situations are quite common in the use of pharmaceutical treatments for chronic diseases, in which clinicians help patients navigate through a series of choices to select a drug that can provide the maximum benefit to that individual.

Under either of these situations, the expected demand curve perceived by the insurer is generally flatter (typical of that in Figure 2) than the empirical

expected demand curve in the population set by individual -level behavior. Since the insurer then anticipates higher levels of moral hazard than would be true with complete information, she is more likely to charge a higher co-insurance rate and/or demand a higher premium in this population. At the limit, an insurer may be reluctant to provide any coverage in some of these situations even though providing it would lead to Pareto improvement.

A corollary for this result is that when the insurer is faced with the decision of whether to curb coverage on certain products, it will often be useful to learn about how individuals are making treatment choices in the presence of coverage and the benefits they are realizing on choosing or not choosing a treatment. This will enable the insurer to align her perceived expected demand curve with the information set (which may include imperfect information to the individuals) on which individual treatment choices are based. Not only the insurer would benefit from knowing the complete marginal schedule of benefits but also from learning about whether individuals have imperfect anticipations of these benefits (Dehejia, 2005). Our empirical example on the comparative effects of substance abuse treatments on crime highlights the role of econometric methods in enabling such alignment. It can also help generate and disseminate more nuanced comparative effectiveness information that could help in individualization.

3. Econometrics for individualized comparative effectiveness research (iCER)

Substance abuse treatments (SATs) can reduce crime, improve health, increase employment, and reduce spread of infectious diseases such as HIV (IOM 2000). There is strong evidence that SATs are effective in lowering abuse of various substances over relatively long periods of time (McClellan et al., 1996; Simpson et al., 2002). Crime reductions, however, account for much of the economic benefits of SAT (French et al. 2002; McCollister and French 2003; Dismuke et al. 2004; Sindelar et al. 2004). Types of SAT vary.

The goal of our empirical analysis is to generate information about the benefits (reduction in the number of armed robberies) of one treatment over another at the individual level. This is in contrast to the goals of a traditional comparative effectiveness study that aims to ascertain which treatment would produce the largest reduction in crime on average. The distribution of the estimated individualized incremental benefits between two treatments can be used to create the marginal benefits schedule across all

patients. We can also identify patients on this schedule who choose to receive specific treatments and therefore form an alternative schedule of marginal benefits only among patients who chose a particular treatment. Comparing the schedule of marginal benefits across all patients versus that among patients who choose a specific treatment can help an insurer learn about the level of imperfect information that individuals possess and therefore help her to make a more efficient decision on coverage.

3a Data

The individual-level data comes from the 1992-97 National Treatment Improvement Evaluation Study (NTIES), a large sample size of voluntary (i.e. those who are not mandated to receive treatments by the judicial system) SAT clients across three principal modalities. NTIES has a higher follow-up response rate (82%) than any comparable client-level follow-up treatment survey (Gerstein and Johnson 2000; Flynn et al. 2001; Gerstein and Johnson 2001). The sample is drawn from units supported by the Center for Substance Abuse Treatment. Compared with nationally representative client surveys, NTIES included a higher percentage of criminal justice clients. It is therefore well suited to analyses of a criminally active client population (Zarkin et al. 2002).

NTIES is especially useful because its pre-SAT and post-SAT measures record whether respondents have committed armed robbery and other offenses, and the number of any such offenses based on a self-report, in ranges (0, 1, 2-5, 6-20, 21-100, 100+). We convert this into a count variable using the mid-points of the intervals (0=0; 1=1; 2-5 = 3.5; 6-20=13.5; 21-100 = 60; 100+ = 100).

In line with our previous analyses with this data (Basu et al., 2008), we restricted our analysis to self-volunteering clients. Other clients include referrals from judicial systems and it is not clear which criteria were used to make SAT selections. We dropped observations from one long-term treatment modality, which included only 8 respondents. We compare three treatment modalities: residential short-term (N=675), residential long-term (N=854), and ambulatory outpatient (N=1,580) treatment.

3b Potential outcomes and treatment effects between two treatments

We built on the potential outcomes literature in economics and statistics (Roy 1951; Rubin 1974, 1978; Holland, 1986) and write a structural model for pre and post outcomes. Let $Y_{t=0}$ = pre-period outcomes; $Y_{t=1}^{k=0,1}$ = post-period potential outcomes with treatment k , and the production functions for these outcomes be:⁷

$$\begin{aligned}
 Y_0 &= X\alpha_0 + \theta\lambda_0 + \varepsilon_0 & Y_1^0 &= X\alpha_1^0 + \theta\lambda_1^0 + \varepsilon_1^0 \\
 & & Y_1^1 &= X\alpha_1^1 + \theta\lambda_1^1 + \varepsilon_1^1
 \end{aligned}
 \tag{ 1 }$$

X represents a vector of observed characteristics (factors) and θ represents a vector of unobserved characteristics (factors). Treatment selections occur between the pre and post periods. Both X and θ affect treatment selection and therefore are represent observed and unobserved confounders respectively . Combinations of levels of X and θ represents the risk classes denoted by Ω_k in Section 2.

The parameter vectors α and λ represent the coefficients (factor loadings) on the observed and unobserved factors respectively. By construction, $(\lambda_1^0 - \lambda_0)$ & $(\lambda_1^1 - \lambda_0)$ represent the within-treatment (over time) heterogeneity in outcomes while $(\lambda_1^1 - \lambda_1^0)$ represents the between-treatment heterogeneity effects in the post period (or how the over-time effects on outcome vary across treatments).

The main parameter of interest to the traditional CER literature has been the Average Treatment Effect (ATE) or the Conditional Average Treatment Effect (CATE) that is conditional on the levels (x) of observed confounders X (Vanness and Mullahy 2006) .

$$\begin{aligned}
 \text{CATE}(x) &= E(Y_1^1 - Y_1^0 \mid X = x) = x(\alpha_1^1 - \alpha_1^0) \\
 \text{ATE} &= \sum_x \text{CATE}(x) \cdot f(x) = \sum_x x(\alpha_1^1 - \alpha_1^0) \cdot f(x)
 \end{aligned}
 \tag{ 2 }$$

Other relevant mean treatment effect parameter from a policy perspective involves the Conditional Effect on the Treated (CTT) and the Conditional Effect on the Untreated (CTUT).

$$\begin{aligned}
 \text{CTT}(x) &= E(Y_1^1 - Y_1^0 \mid D = 1, X = x) = x(\alpha_1^1 - \alpha_1^0) + E(\theta(\lambda_1^1 - \lambda_1^0) \mid D = 1) \\
 \text{CTUT}(x) &= E(Y_1^1 - Y_1^0 \mid D = 0, X = x) = x(\alpha_1^1 - \alpha_1^0) + E(\theta(\lambda_1^1 - \lambda_1^0) \mid D = 0),
 \end{aligned}
 \tag{ 3 }$$

⁷ Here, linearity is assumed for simplicity, without losing the general point of the discussion.

where $D = I(\text{treatment choice})$. Population average or unconditional versions of the Effect on the Treated (TT) and the Untreated (TUT) can be obtained by integrating over the distribution of observed characteristics (Heckman and Robb 1985).

Finally, individualized mean effect parameter is given by the marginal treatment effect parameter, which conditions on levels, x and θ , of both observed and unobserved confounders (Heckman and Vytlacil 1999):

$$\text{MTE}(X = x, \theta = \theta) = E(Y_1^1 - Y_1^0) = x(\alpha_1^1 - \alpha_1^0) + \theta(\lambda_1^1 - \lambda_1^0) \quad (4)$$

MTEs represent average effects for thinner margins in the population defined by levels of both X and θ . They represent b_i , the building blocks for the normative marginal benefits curve, described in Section 2.

It is clear from the above expressions that in the absence of between-treatment heterogeneity, i.e., $\lambda_1^1 = \lambda_1^0$, all of the $\text{TT}(x)$, $\text{TUT}(x)$ and the $\text{MTE}(x, \theta)$ parameters converge to the $\text{CATE}(x)$ parameter. This is an important condition as it simplifies the evaluation problem considerably. Meeting this condition implies that treatment effects are not heterogeneous over unobserved confounders and treatment-effect heterogeneity can be studied entirely based on observed confounders.

3c Traditional econometric approach

The most popular method to analyze such pre-post data is to use fixed effects modeling, which works under the assumption that durable and unobserved client characteristics that may be correlated with treatment selection are eliminated through a difference-in-difference (DID) approach (Bertrand et al . 2004). However, it is not clear, what parameter a DID estimator ends up estimating.

More generally, observed outcome in the post-period is given by $Y_1 = DY_1^1 + (1 - D)Y_1^0$. This representation is Quandt's switching regression framework (Quandt 1958, 1972). Therefore, a difference-in-difference estimator, which estimates the expected difference between pre and post outcomes for each treatment will produce:

$$\begin{aligned}
& E(Y_1 - Y_0 | X = x, D = 1) - E(Y_1 - Y_0 | X = x, D = 0) \\
& = E(Y_1^1 - Y_0 | X = x, D = 1) - E(Y_1^0 - Y_0 | X = x, D = 0) \\
& = (x\alpha_1^1 - x\alpha_0) + E(\theta(\lambda_1^1 - \lambda_0) | D = 1) - (x\alpha_1^0 - x\alpha_0) - E(\theta(\lambda_1^0 - \lambda_0) | D = 0) \\
& = (x\alpha_1^1 - x\alpha_1^0) + E(\theta(\lambda_1^1 - \lambda_0) | D = 1) - E(\theta(\lambda_1^0 - \lambda_0) | D = 0) \\
& = (x\alpha_1^1 - x\alpha_1^0) + E(\theta(\lambda_1^1 - \lambda_1^0) | D = 1) + (\lambda_1^0 - \lambda_0) \cdot \{E(\theta | D = 1) - E(\theta | D = 0)\}
\end{aligned}$$

≠CTT(x) or CTUT(x) or CATE(x) or MTE(x,θ) (5)

Thus, in the presence of both within- and between-treatment effect heterogeneity, the DID estimator produces results with no clear interpretations (LaLonde, 1986; Heckman, 1978; 1990). In the absence of within-treatment heterogeneity ($\lambda_1^0 = \lambda_0$), CTT(x) is identified. While in the absence of both within- & between-treatment heterogeneity ($\lambda_1^1 = \lambda_1^0 = \lambda_0$), CATE(x) is identified.

3d A structural model based approach to analyzing comparative effectiveness of SATs

To address the limitations of the traditional econometric approach, we directly estimate parameters from a structural model for potential outcomes that allows for both within- and between-treatment heterogeneity in outcomes. When self-selection is present, a patient's choice of treatment is based on his anticipation of benefits conditional on the idiosyncratic levels of his observed and unobserved characteristics. This anticipation may be imperfect and not always coincide with the realized (potential) benefits. On average, however, such anticipation would lead more patients to choose the better treatment for themselves. We use a latent factor to represent the unobserved durable characteristics of the patients, based on which patients select treatments. We specify a model for treatment choice that is based on patient's anticipated benefits and models for potential outcomes that are based on the patient's realized benefits.

Treatment choice model:

Individuals (i) maximize a latent random utility in order to choose treatment: $D_i = \text{Argmax}_k \{L_{ik}\}$; $k = 0$ (Outpatient), 1 (Short-term), 2 (Long-term).

Following the traditional random utility model in economics (MacFadden, 1974, Manski, 1975), the treatment specific random utility is given as:

$$L_{ik} = \beta_k^1 + \beta_k^2 \cdot X_i + \beta_k^3 \cdot \theta_{1i} + v_{ik}, \quad (6)$$

where each coefficient in $(\beta_2^1, \beta_2^2, \beta_2^3)$ is normalized to zero. X represent the vector of observed characteristics. We have added the '~' to differentiate the notation of a vector for a scalar quantity. For example, the unobserved characteristics (to the analyst) are scaled to be represented by one latent factor θ_1 , and v_k denotes the random error.⁸

Pre-SAT outcomes model:

$$Y_{i0} = \exp(\alpha_{10} + \alpha_{20} \cdot X_i + \alpha_{30} \cdot \theta_{1i}) + \varepsilon_{i0}, \quad (7)$$

where we use a log-linear model for the non-negative outcomes.⁹

Post-SAT outcomes model:

$$Y_{i1}^k = \exp(\alpha_{11}^k + \alpha_{21}^k \cdot X_i + \alpha_{31}^k \cdot \theta_{1i}) + \varepsilon_{i1}^k, \quad k = \text{Treatments}; 0, 1, 2; \quad (8)$$

The goal of our analysis is to recover the joint distribution of pre-post changes in crime under alternative treatments: $F(\Delta_i^0, \Delta_i^1, \Delta_i^2)$, where $\Delta_i^k = Y_{i1}^k - Y_{i0}$, which would require us to identify $F(Y_{i0}, Y_{i1}^k)$, $k=1, 2, 3$. Note, however, that we do not directly observe Y_{i1}^k , but only $Y_{i1} = I(D_i = k) \cdot Y_{i1}^k$. Therefore, we only observe the conditional joint distributions $F(Y_{i0}, Y_{i1}^k | D_i = k)$, which are different than the unconditional distributions $F(Y_{i0}, Y_{i1}^k)$ due to confounders that affect these outcomes and also determine selection into treatment. In our above formulations, we represent a reduced form of these confounders using the latent factor θ_{1i} . Once we identify θ_{1i} and account for its effect on outcomes, we can not only solve the selection bias (i.e. make $v^k \perp \varepsilon_0 \perp \varepsilon_1^k | \theta_1, X$, \perp denoting statistical independence) but can also recover the unconditional distribution $F(Y_{i0}, Y_{i1}^k)$ or $F(\mu_{i0}, \mu_{i1}^k)$, where $\mu_{i0} = E(Y_{i0} | X_i, \theta_i, \alpha_0)$ and $\mu_{i1}^k = E(Y_{i1}^k | X_i, \theta_i, \alpha_1^k)$. However, in order to identify θ_{1i} independent from its effects (α_{30} and α_{31}^k) on outcomes, we must rely on the observed conditional joint distribution of $F(Y_{i0}, L_{ik} | D_i = k)$ and $F(Y_{i1}^k, L_{ik} | D_i = k)$, for $k = 0, 1$. These added conditional covariances help us identify the following vector of parameters: $\{\beta_3^1, \beta_3^2, \alpha_{30}, \alpha_{31}^0, \alpha_{31}^1, \alpha_{31}^2, \text{Var}(\theta_1)\}$. These parameters, along with the regression coefficients on observed covariates, which are identified using traditional

⁸ The dependence of choices on pre-SAT outcomes is accounted for by the structural model of pre-SAT outcomes in (20) that is also dependent of X and θ_1 .

⁹ This is in line with standard statistical models used for count data. In fact, the specification in a structural model for potential outcomes is usually not driven by theory, as is the case for any standard outcomes regression model. Therefore, goodness of fit tests are conducted by testing the adequacy of fit.

regression approaches, can be used to construct the unconditional distribution $F(\mu_{i_0}, \mu_{i_1}^k)$. The semi-parametric identification strategy is detailed in Appendix A.

For estimation, we let $Y_{i_0} \sim \text{Poisson}(\mu_{i_0})$, $Y_{i_1} \sim \text{Poisson}(I(D_i = k) \cdot \mu_{i_1}^k)$, and gave non-informative priors on all parameters: $\theta_{i_1} \sim \text{Normal}(0, \sigma_{\theta_1}^2)$, $\sigma_{\theta_1} \sim \text{Uniform}(0, 10)$ and all regression coefficients $\sim \text{Normal}(0, 10E-6)$. Treatment choice model was estimated using: $D_i \sim \text{Multinomial}(\pi_0, \pi_1, \pi_2)$ with logit link, $\beta_3^2 = 1$. All other logit coefficients $\sim \text{Normal}(0, 10E-6)$. For sensitivity analysis, we also analyzed with alternative prior specifications for $\theta_{i_1} \sim \text{Uniform}(-10, 10)$. Posterior distribution of parameters was obtained via MCMC (Gibbs-Metropolis-Hastings) using WinBUGs software. Three independent chains were generated using an MCMC sampling technique. Each of these chains was started with random starting values dispersed across the support of the parameter space. The first 10,000 iterations were discarded (burn-ins).

3f Results

Table 1 presents the descriptive statistics for individuals who were enrolled in different SATs. Patients who chose outpatient treatments were slightly older than the other two modalities. Long-stay patients were more likely to be females, while short-stay patients were less likely to be blacks. Various other differences in observed characteristics exist between the modalities. The baseline number of robberies committed was much higher for patients choosing long-stay treatment than for either outpatient or short-stay treatments.

The MCMC results show reasonably quick convergence (assessed via the Gelman-Rubin statistic (Gelman and Rubin, 1992) and good mixing of the independent chains. Moderate autocorrelation was found between subsequent draws for each chain, and consequently thinning of 20 was applied to obtain our posterior samples. The posterior means for the latent factor were -0.0001 for the overall data and 0.0005, 0.005 and -0.008 for subjects selecting outpatient, long-term and short-term treatments respectively. The posterior standard deviation was 0.08 and by design, the same in all groups. Correlation between the latent factor and the observed covariates ranged from -0.05 (primary addiction to alcohol) to 0.06 (primary addiction to cocaine). Results were found to be robust to alternative specification of priors for the latent factor.

Table 2 presents the estimated coefficients from the treatment choice model. Among other factors predicting choice, we find that patients with higher levels of the latent factor are significantly more likely to choose short-stay treatment over outpatient treatment (given that higher levels of the latent factor are artificially made to associate with a greater likelihood of choosing long-stay treatment over outpatient treatment). Table 2 also presents coefficient estimates from the outcomes models. We find that the latent factor significantly affects outcomes in both the pre and the post period and under all treatment modalities. We find significant within-treatment (over time) heterogeneity in effects under outpatient treatment. We also find significant between-treatment heterogeneity in effects.

Table 3 shows a cross-table for the predicted number of robberies. The rows signify the observed treatment groups while the columns signify the potential treatment groups. That is, outcomes in cell (A, B) represent what would have happened if the group of individuals who had originally received treatment A had received treatment B instead. First we find that the model predictions tally well with the observed quantities both in the subgroups and also overall. The pre-period goodness of fit is especially a testament for the validity of our model, as the predictions on armed robbery were able to discriminate between the three treatment groups without a treatment-specific indicator or treatment-specific coefficients in the pre-period model. That is, this discriminatory ability is entirely generated by the latent factor.

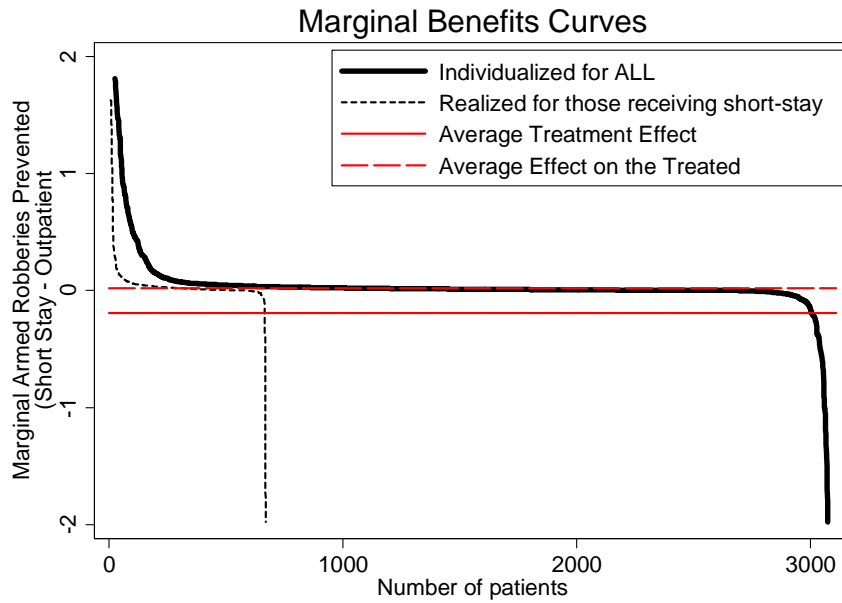
Interesting and policy-relevant comparative effectiveness results (TT and ATE) come out of this analysis that would have been masked using the traditional econometric approaches (Table 3). Among patients choosing outpatient treatment, outpatient treatment produces the largest reduction in armed robberies per patient (-0.57 (0.02), $p < 0.0001$).¹⁰ Compared to the second best treatment in this group, the short-stay treatment, these reductions were bigger by -0.10 (se = 0.19, $p = 0.60$) robberies per patient.

Among patients choosing short-stay residential treatment, the short-stay treatment produces the largest reduction in robberies per patient (-0.53 (0.03) $p < 0.0001$). However, it is estimated that outpatient treatment could have also produced a similar magnitude of reductions, the difference between the two modalities being not significant. Figure 4a illustrates the different marginal benefits

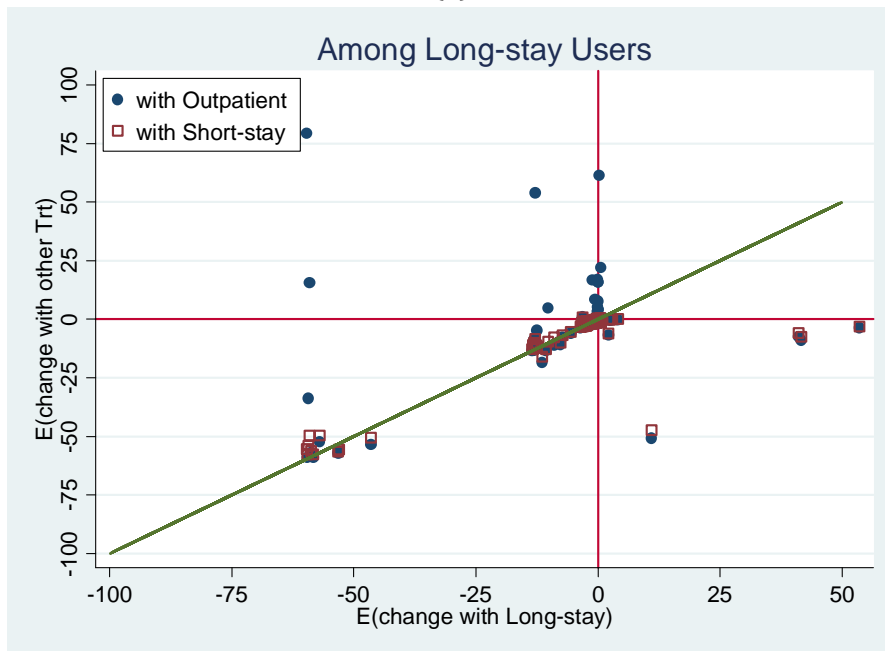
¹⁰ Note that the p-value reported is the Bayesian analog of the frequentist p-value.

curves for short-stay versus outpatient under alternative information regimes. Based on average benefit (incomplete) information, the curve perceived by the insurer would be flat and indicate that on average short-stay treatment would prevent fewer robberies than outpatient treatment. However, current analyses may help the insurer realize that the normative marginal benefits curve (formed using estimated MTEs) is downward sloping. Moreover, the individualized marginal benefits curve realized amongst those who select short-stay treatment (presumably based on imperfect information on MTEs) indicates that a majority of subjects who would have obtained small benefits or negative benefits from short-stay treatment over outpatient do not choose short-stay. Nevertheless, there remain opportunities to improve.

Finally, among patients choosing the long-stay residential treatment, outpatient treatment (-1.06 (0.04) $p < 0.0001$) and not the long-stay treatment (-0.87 (0.04), $p < 0.0001$) is expected to produce the biggest reduction in robberies per patient on average, the difference being statistically significant (-0.20 (0.03), $p < 0.0001$). The average treatment effect was highest for the outpatient treatment, signifying that if one treatment were to replace all other treatments in this population, it must be the outpatient modality. In fact, based on the current choices that these patients are making, the average “as-treated” effect, which produces a reduction of -0.64 robberies per patient, is less (by 0.05 robberies per patient, (se= 0.01, $p < 0.0001$)) than the reductions that could be achieved by having outpatient treatment as the only treatment available. From the SP's point of view, if heterogeneity is not revealed, the average effects argue for a restriction on coverage on short-term and long-term residential treatment. However, looking at the treatment effects on treated for individual treatments, it is apparent that the inefficiency in “as-treated” results compared to the average treatment effect is generated solely by patients choosing long-term residential treatment. Interestingly, however, not all patients choosing the long-term treatment are making suboptimal choices.



(a)



(b)

Figure 4: (a) Marginal benefits curve for short-stay versus outpatient under alternative information regimes. (b) Distribution of individual level treatment effects among patients who chose to receive long-stay residential treatment.

Figure 4b shows the joint distribution of predicted individual-level effects (MTEs) under the three treatment modalities for patients who in practice chose the long-term treatment. Each dot represents a patient and dots below the 45-degree line represent a patient who made suboptimal choices by selecting long-term treatment. As evident from the figure, there is considerable heterogeneity in effects. In fact, we estimate that 52% of patients receiving long-term treatment make suboptimal choices and that the value of achieving individualization that could overcome these inefficient choices is about 0.36 robberies per person. Corresponding numbers for patients choosing outpatient or short-term treatment are 98% and 0.08 or 59% and 0.08 respectively. In these later modalities, even though patients appear to make suboptimal choices, the net gain achieved if individualization could be attained is relatively low. Overall, the value of individualized care over all treatment modalities could achieve a further reduction in 0.16 robberies per patient. Compare that to a policy that only relies on the average treatment effect and offers outpatient treatment to everyone, and achieves only an additional reduction of 0.05 robberies per patient compared to the average as-treated effects.

4. Discussions

The \$2.5 trillion in health care costs in United States constituted approximately 18 percent of GDP in 2009, and at current rates of growth, health care will exceed one-fourth of GDP by 2025 (Sisko et al., 2009; CBO, 2008). As part of a portfolio of innovative public policies that are aimed to address this challenge, public investments in comparative effectiveness research (CER) are viewed as worthwhile endeavors by payers, providers and patients (ACES Testimony, 2009). Identifying options that work on average creates winners and losers by influencing physician decision making or through coverage decisions that favor the winners, and may substantially induce demand for the winners (Meltzer et al. 2007). This can, in turn, lead to increases in prices for those options, and therefore cannot guarantee decreases in overall expenditures (Basu and Philipson, 2010). More importantly, patient welfare may increase on average but is not likely to achieve the full potential that can be tapped through individualization (Basu & Meltzer, 2007; Basu, 2009).

In this paper, it is argued that divulging individual-level heterogeneity in treatment effects is essential to realize the full potential of CER. The core of this argument lies in understanding how CER information influences the behavior of individual patients and their providers, which in turn produces a market demand curve, as well as in helping a population-level policymaker anticipate a market demand curve for a treatment. CER focused on producing average effects results in an empirical market demand curve and also a perceived demand curve that is flatter or more elastic than it truly should be. Consequently, the decision making that follows may lead to perverse results from a welfare perspective. Individualization, on the other hand, reveals the true (normative) downward-sloping demand curve and enables decision makers to make improved choices and policies, which can assure that health is neither compromised nor distributed inequitably.

The potential impact on patients' well-being due to individualized CER (*i*-CER) can very well be orders of magnitude larger than is possible with the current state of CER (Basu 2009). For example, individualization based on preference heterogeneity in prostate cancer patients was found to produce benefits that are worth 100 times more than identifying better treatment based on average preferences (Basu and Meltzer 2007). Similar empirical evidence on the evaluation of SATs are presented in this paper.

Ideas of individualization and its appropriate uses can have strong implications for how comparative effectiveness research is practiced in general. Despite the growing consensus that individualization should be a central part of CER (Garber and Tunis, 2009), there is little discussion about how such individualized effects can be estimated. Addressing treatment effect heterogeneity using "hypothesis testing"-driven approaches by subgroups, which has become the traditional approach to comparative effectiveness research, can be an extremely inefficient process of information generation. This is because, more often than not, treatment effects vary over subgroups that are defined by the complex interactions of many patient characteristics. Thus the number of sub-groups over which treatment effect heterogeneity should potentially be studied are numerous and cannot possibly be tested (even after adjustments of p-values) using *a priori* specification of subgroups.

A key approach to such a problem is to be able to develop prediction algorithms for individual-level treatment effect heterogeneity. Such algorithms can be constructed without identifying low-dimensional individualized characteristics

such as genomic information, but rather by collapsing multi (high)-dimensional outcomes and behavior into individual-level latent characteristics, which can be used to establish individualized treatment effects. These prediction algorithms can be viewed as a hypothesis generation exercise at the individual level. However, these algorithms have two extremely useful implications for comparative effectiveness research. First, any attempt to individualize care based on prediction algorithms must begin with a hypothesis generation exercise and therefore these results can provide valuable resources to clinicians and policymakers, who in their absence much rely on traditional comparative effectiveness results on averages. The necessity of an algorithmic approach lies in the feasibility of translating enormous amounts of information to the bedside, without overwhelming physicians. Second, these results will provide key input to any confirmatory randomized trial evaluating and improving such prediction algorithms.

For example, although our empirical analysis reveal considerable treatment effect heterogeneity across different SAT modalities, the cost of implementing an individualized schedule of treatment assignment may be costly. An algorithmic approach to CER can help reduce the cost of implementation in this regard and can therefore produce value. Identifying the patient characteristics that lead to suboptimal choices can help to reallocate by allowing more careful consideration of treatment choices. Table 4 identifies the factors responsible for suboptimal treatment choices among patients who chose the long-term modality. Note that this information is only obtainable because we have identified individual-level treatment effects and therefore can identify suboptimal treatment choices ex-post. Table 4 suggests that secondary addiction to alcohol or cocaine was the only factor that predicted long-term residential treatment to be a suboptimal choice. The latent factor was not significantly associated with suboptimal choices. However, future research, which collects a wider range of patient-level characteristics, can identify components of this latent factor and can further enhance the individualization of comparative effectiveness in this field. Future research should also study the simultaneous affect of these treatments on a multi-dimensional vector of outcomes that are relevant for these patients (Sindelar et al., 2004).

Current data from many observational studies and some clinical trials (such as the NIMH sponsored CATIE trial, Stroup et al., 2003) contain information to identify the joint distribution of potential outcomes but only when paired with novel methods such as those that use (latent) factor models (Munkin and Trivedi, 2003;

Heckman and Navarro, 2003) and control-functions (Florens et al., 2008). The bases for these approaches lie in using patients' clinical history and behavior as outcomes of underlying (latent) processes and characteristics rather than as determinants of the comparative effects and thereby identify these latent factors and use them to establish individualized comparative effectiveness. Investing in pragmatic trials (Tunis et al., 2003) by focusing more on adaptive trials (Cheng and Berry, 2007), adaptive assignments within trials (Murphy, 2005), and cross-over designs (Senn, 2002) could help generate data that potentially have the identifying information for individual effects—but such an approach still needs to be paired with the right analytical tools to salvage this information. Observational studies can play a crucial role in this process as they carry rich information about choices. Investing in longitudinal databases, measurements of cognitive and non-cognitive capabilities of patients that are predictive of behavior, and outcomes and development of novel methods in this field will be the key to achieving this ambitious paradigm of CER. Therefore, a commitment to further develop and refine these methodologies seems to be of significant scope within the CER agenda.

A crucial challenge lies in translating the enormous volume and rich evidence on individualized care to the bedside. Leveraging health information technology to enable seamless translation will also be crucial for this purpose. The current administration's investments in both comparative effectiveness and health information technology are a major step toward achieving a more efficient health care system. It remains to be seen whether researchers, providers and payers can take this opportunity to deliver on the expectations.

Finally, individualized CER can guide research prioritization in the future. It can identify misalignments between individual and social decision making, in terms of anticipation of potential outcomes and behavior at both levels, and direct future research to eliminate such misalignments and enhance social welfare. Even when decision making at the two levels becomes aligned, it is important to understand the dependence in potential outcomes that is driven by unobserved or non-forecastable heterogeneity at the individual level. Future research that transforms a portion of today's uncertainty about comparative effects (deemed to be first-order) into identifiable/observable heterogeneity (e.g. genetic and epigenetic factors) in comparative effects can improve welfare through better individual and social decision making in the future.

Identification

Identification is required for the following set of parameters:

$\{\beta_3^1, \beta_3^2, \alpha_{30}, \alpha_{31}^0, \alpha_{31}^1, \alpha_{31}^2, \text{Var}(\theta_1)\}$. The rest of the regression coefficients are

identified based on the variances of Y , X and their covariances using traditional regression approaches. What follows is conditional on estimating these other regression coefficients. The following details the sequence of identification of all the remaining parameters in the models:

Identification Strategy	Normalization	Identify
$\text{Cov}(L^2, Y_0) \propto E(\beta_3^2 \cdot \theta_1 \cdot \exp(\alpha_{30} \cdot \theta_1))$ $\approx \beta_3^2 \cdot \text{Var}(\theta_1) \cdot \alpha_{30}$ (By 2 nd order Taylor series approximation)	1) $\beta_3^2 = 1$	$\text{Var}(\theta_1) \cdot \alpha_{30}$
Similarly, $\text{Cov}(L^1, Y_0) \approx \beta_3^1 \cdot \text{Var}(\theta_1) \cdot \alpha_{30}$	-	β_3^1
$\text{Cov}(Y_1^0, Y_0 D=0) \approx \alpha_{31}^0 \cdot \text{Var}(\theta_1 D=0) \cdot \alpha_{30}$	2) $\text{Var}(\theta_1 D=0) = \text{Var}(\theta_1)$	α_{31}^0
$\text{Cov}(L^1, Y_1^0) \approx \beta_3^1 \cdot \text{Var}(\theta_1 D=0) \cdot \alpha_{31}^0$	-	$\text{Var}(\theta_1), \alpha_{30}$
$\text{Cov}(L^1, Y_1^1) \approx \beta_3^1 \cdot \text{Var}(\theta_1 D=1) \cdot \alpha_{31}^1$	3) $\text{Var}(\theta_1 D=1) = \text{Var}(\theta_1)$	α_{31}^1
$\text{Cov}(L^1, Y_1^2) \approx \beta_3^1 \cdot \text{Var}(\theta_1) \cdot \alpha_{31}^2 +$	-	α_{31}^2

+ $\text{Var}(\theta_1 | D=2) = \text{Var}(\theta_1)$ because of normalizations 2) and 3).

Table 1: Descriptives of patients under each SAT

Descriptives	Outpatient N=1580	Short -stay N=675	Long-stay N=854
Age (years), mean (sd)	32.41 (9.15)	30.97 (7.45)	30.27 (8.04)
Female	27.5%	29.3%	52.2%
Non-Hispanic Black	62.2%	44.6%	62.8%
Hispanics	14.6%	16.6%	10.0%
Currently married	20.1%	25.8%	15.5%
Ever homeless in last 12 months	16.8%	19.9%	26.0%
Released from jail/prison in last 12 months	38.1%	44.4%	34.2%
Child in Household	24.6%	18.7%	24.8%
No income in last 12 months	10.5%	10.4%	10.9%
No wage income in last 12 months	49.9%	34.1%	55.6%
Wage income < \$3000 in last 12 months	19.2%	17.2%	19.3%
Medicaid	29.2%	8.9%	26.0%
Primary addiction to cocaine	38.9%	32.6%	60.5%
Primary addiction to alcohol	38.1%	19.9%	15.3%
Secondary addiction to cocaine/alcohol	39.3%	39.6%	53.2%
Pre-period armed robberies, mean (sd)	0.66 (6.05)	0.63 (5.34)	1.25 (7.03)
Post-period armed robberies, mean (sd)	0.09 (0.89)	0.1 (1.06)	0.38 (4.17)

Table 2: Posterior mean and standard deviation estimates for regression coefficients.

	Treatment Choice Model		Outcomes Models			
	Short-stay vs Outpatient	Long-stay vs Outpatient	Pre-period	Post-period Outpatient	Post-period Short-stay	Post-period Long-stay
Age (years)	-0.03 (0.01)+	-0.003 (0.01)	-0.28 (0.04)+	-0.08 (0.02)+	-0.11 (0.06)	-0.42 (0.05)+
Female	1.16 (0.11)+	0.62 (0.12)+	-2.59 (0.56)+	-1.71 (0.43)+	2.41 (1.04)+	-0.16 (0.57)
Non-Hispanic Black	-0.72 (0.13)+	-0.7 (0.13)+	0.32 (0.51)	-0.31 (0.35)	-1.2 (0.82)	-1.41 (0.54)+
Hispanics	-0.72 (0.17)+	-0.24 (0.16)	-0.32 (0.61)	0.39 (0.39)	0.12 (1.22)	0.01 (0.7)
Currently married	-0.13 (0.13)	0.41 (0.13)+	-0.33 (0.63)	-1.85 (0.65)+	-4.93 (1.61)+	2.74 (0.79)+
Ever homeless	0.45 (0.12)+	0.41 (0.13)+	2.62 (0.48)+	1.76 (0.31)+	1.2 (1.01)	4.05 (0.55)+
Released from jail/prison	-0.03 (0.11)	0.05 (0.11)	1.87 (0.44)+	0.82 (0.3)+	2.46 (1.07)+	2.72 (0.49)+
Child in household	-0.22 (0.12)	-0.33 (0.13)+	-0.69 (0.62)	0.2 (0.38)	-2.44 (1.78)	0.1 (0.69)
No income	0.04 (0.12)	-0.86 (0.13)+	0.63 (0.52)	0.15 (0.37)	2.75 (1.23)+	3.69 (0.66)+
No wage income	0.02 (0.14)	-0.69 (0.15)+	0.11 (0.55)	0.94 (0.38)+	4.23 (1.19)+	2.25 (0.69)+
Wage income < \$3000	0.06 (0.17)	0.11 (0.19)	0.81 (0.61)	-0.16 (0.48)	-2.57 (1.35)	-3.95 (0.72)+
Medicaid	-0.48 (0.12)+	-1.32 (0.16)+	0.95 (0.52)	0.27 (0.33)	-24.75 (14.08)	-1.96 (0.52)+
Pr. addiction to cocaine	0.65 (0.13)+	-0.67 (0.13)+	0.22 (0.51)	-0.27 (0.35)	-0.87 (0.74)	6.25 (0.75)+
Pr. addiction to alcohol	-0.73 (0.14)+	-1.49 (0.14)+	-1.29 (0.59)+	-0.91 (0.37)+	-26.64 (13.71)	4.66 (0.73)+
Sec. addiction to cocaine/alcohol	0.44 (0.1)+	-0.06 (0.1)	1.81 (0.43)+	0.61 (0.29)+	-0.44 (1.04)	0.04 (0.47)
Theta	3.08 (1.06)+	1	37.09 (11.39)+	19.27 (6.03)+	33.05 (9.9)+	30.72 (9.88)+
_cons	-0.8 (0.17)+	0.68 (0.15)+	-11.98 (1.14)+	-6.09 (0.7)+	-13.36 (2.99)+	-19.61 (1.94)+
+: Bayes p-val <0.05			diff with pre in Theta coeff	-17.8 (5.9)+	-4.0 (7.4)	-6.37 (3.55)

Table 3: Observed and predicted armed robberies (posterior mean and standard deviations)

Observed Treatment Received ↓	Estimates	Pre-period	Post-period			Overall
			Outpatient	Short-stay Residential	Long-stay Residential	
Outpatient Drug-free	Observed	0.66	0.09	-	-	
	Predicted	0.66 (0.02)+	0.09 (0.007)+	0.19 (0.19)	0.49 (0.12)+	
	E(change)	-	-0.57 (0.02)+	-0.47 (0.19)+	-0.17 (0.12)	
Short-stay Residential	Observed	0.63	-	0.10	-	
	Predicted	0.63 (0.03)+	0.12 (0.02)+	0.10 (0.01)+	0.31 (0.24)	
	E(change)	-	-0.51 (0.03)+	-0.53 (0.03)+	-0.32 (0.24)	
Long-stay Residential	Observed	1.24	-	-	0.38	
	Predicted	1.25 (0.04)+	0.19 (0.03)+	0.71 (0.58)	0.38 (0.02)+	
	E(change)	-	-1.06 (0.04)+	-0.54 (0.58)	-0.87 (0.04)+	
OVERALL	Observed	0.81	-	-	-	0.17
	Predicted	.81 (.02)+	0.12 (0.01)+	0.31 (0.23)	0.42 (.09)+	0.17 (0.01)+
	E(change)		-0.69 (0.02)+	-0.51 (0.23)+	-0.39 (0.09)+	-0.64 (0.02)+

+ Bayes p-value < 0.05; E(change) = Post - Pre

Table 4: Factor responsible for suboptimal choices among patients who chose long-term residential treatment.

	logit(Suboptimal Choices)		
	OR	(95% CI)	[p-value]
Age (years)	0.99	(0.98, 1.01)	[0.461]
Female	1.02	(0.74, 1.41)	[0.891]
Non-Hispanic Black	0.81	(0.55, 1.2)	[0.297]
Hispanics	0.86	(0.48, 1.55)	[0.625]
Currently married	1.23	(0.86, 1.75)	[0.262]
Ever homeless	1.07	(0.77, 1.5)	[0.677]
Released from jail/prison	0.98	(0.7, 1.36)	[0.897]
Child in household	1.20	(0.83, 1.74)	[0.331]
No income	1.23	(0.75, 2.01)	[0.412]
No wage income	0.90	(0.64, 1.26)	[0.529]
Wage income < \$3000	0.85	(0.55, 1.33)	[0.485]
Medicaid	1.06	(0.79, 1.43)	[0.693]
Pr. addiction to cocaine	0.89	(0.59, 1.34)	[0.563]
Pr. addiction to alcohol	0.84	(0.54, 1.29)	[0.415]
Sec. addiction to cocaine/alcohol	1.35	(1.02, 1.78)	[0.037]
Theta	1.01	(0.21, 4.83)	[0.987]

References:

- ACES Testimony. Testimony on comparative effectiveness research by the Alliance of Comparative Effectiveness Stakeholders (ACES) to the Dr. Carolyn M. Clancy, Director Agency for Healthcare Research and Quality (AHRQ), 2009 April 3. <http://www.ahrq.gov/about/nac/aces.htm> (Accessed Jan 31, 2011)
- Arrow KJ, Debreu G. Existence of an equilibrium for a competitive economy. *Econometrica* 1954; **22**:265-290.
- Arrow KJ. Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review* 1963; **53(5)**: 941-973.
- Basu A. Individualization at the heart of comparative effectiveness research: The time for *i*-CER has come. *Medical Decision Making* 2009; **29(6)**: N9-N11.
- Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. *Med Decis Making*. 2007; **27(2)**:112-127.
- Basu A, Paltiel DA, Pollack HA. Social costs of robbery and the cost-effectiveness of substance abuse treatment. *Health Economics* 2008; **17(8)**: 926-948.
- Basu A, Philipson T. Impact of comparative effectiveness research on health and healthcare spending. *NBER Working Paper* No. w15633, 2010.
- Bertrand M, Duflo E, Mullainathan S. How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics* 2004; **119(1)**: 249-275.
- Chakley M, Khalil F. Third party purchasing of health services: Patient choice and agency. *Journal of Health Economics* 2005; **24**: 1132 - 1153.
- Cheng Y, Berry DA. Optimal adaptive randomized designs for clinical trials. *Biometrika*. 2007; **94(3)**: 673-689.
- Chernew, ME., Rosen, A., Fendrick, AM. Value-based insurance design. *Health Affairs* 2007; **26(2)**: w195-w203.
- Congressional Budget Office. January, 2008. Technological change and the growth of health care spending. Available at: <http://www.cbo.gov/ftpdocs/89xx/doc8947/01-31-TechHealth.pdf>. Accessed June 30, 2009.
- Dehejia RH. Program evaluation as a decision problem. *Journal of Econometrics* 2005; **125**: 141-173.
- Dismuke, C., M. French, et al. (2004). "Out of touch or on the money: Do the clinical objectives of addiction treatment coincide with economic evaluation results?" *Journal of Substance Abuse Treatment* **27**: 253-263.
- Federal Coordinating Council for Comparative Effectiveness Research, Report to the President and the Congress, June 30, 2009. <http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf> (accessed June 30, 2009)

- Fendrick AM, Smith DG, Chernew ME, Shah SN. A benefit-based copay for prescription drugs: patient contribution based on total benefits, not drug acquisition cost. *Am J Manag Care*. 2001; **7**:861-867
- French, M., J. Salomé, et al. (2002). "Benefit-Cost Analysis of Addiction Treatment: Methodological Guidelines and Empirical Application Using the DATCAP and ASI." *Health Services Research* **37(2)**: 433-455.
- Flynn P, Simpson D, Anglin M, et al. (2001). Comment on nonresponse and selection bias in treatment follow-up studies. *Substance Use and Misuse* **36(12)**:1753-1757.
- Fisher RA. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935.
- Florens JP, Heckman JJ, Meghir C, Vytlacil E, Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects. *Econometrica*. 2008; **76(5)**: 1191-1206.
- Garber AM, Tunis S. Does comparative effectiveness research threaten personalized medicine. *The New England Journal of Medicine* 2009; **360(19)**: 1925-1927.
- Gelman A, Rubin DB. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Sciences*. **7**: 457-511.
- Gerstein D, Johnson R. (2000). Nonresponse and selection bias in treatment follow-up studies. *Substance Use and Misuse*. **35(6-8)**:971-1014
- Gerstein D, Johnson R. (2001). In reply to Patrick M. Flynn and colleagues. *Substance Use and Misuse*. **36(12)**:1753-1757
- Heckman JJ. Dummy endogenous variables in a simultaneous equations system. *Econometrica* 1978; **46(4)**: 931-959.
- Heckman JJ. Varieties of selection bias. *American Economic Review* 1990; **80**: 313-318.
- Heckman JJ. Evaluating welfare and training programs. In. Manski and Garfinkel.1992.
- Heckman JJ, Robb R. Alternative methods for evaluating the impact of interventions. In. Heckman J, Singer B (eds), *Longitudinal Analysis of Labor Market Data*. Econometric Society Monograph No. 10. Cambridge University Press, Cambridge, 1985.
- Heckman JJ, Smith J. Evaluating the welfare state. In storm S. (ed.), *Econometrics in the 20th Century: The Ragnar Frisch Centenary*. Econometric Society Monograph Series. Cambridge University Press, Cambridge, 1998.
- Heckman JJ, Navarro S. Dynamic Discrete Choice and Dynamic Treatment Effects. *Journal of Econometrics*. 2007; **136(2)**: 341-396.
- Heckman JJ, Vytlacil EJ. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 1999; **96(8)**: 4730-34.

- Holland P. Statistics and causal inference. *Journal of the American Statistical Association* 1986; **81**: 945-970.
- Jena A, Philipson TJ. Endogenous Cost-Effectiveness Analysis in Health Care Technology Adoption. *NBER Working Papers* 15032, 2009.
- LaLonde R. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 1986; **76(4)**: 604-620.
- Manning, W.G., Marquis, M.S. Health insurance: the trade-off between risk pooling and moral hazard. *Journal of Health Economics* 1996; **15 (5)**, 609–639.
- Manski, C. Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics* 1975; **3**: 205-228.
- Manski C, Garfinkel I. Evaluating welfare and training programs. Harvard University Press, Cambridge. 1992.
- McLellan AT, Woody GE, Metzger D, et al. Evaluating the effectiveness of addiction treatments: reasonable expectations, appropriate comparisons. *Milbank Quarterly* 1996;74(1):51-85.
- McCollister, K., French M. The relative contribution of outcome domains in the total economic benefit of addiction interventions: a review of first findings. *Addiction* 2003; **98**:1647-1659.
- McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers of Econometrics*. P. Zarembka, ed., New York: Academic Press, 1974.
- MedPAC Report to Congress: Reforming the Delivery System. June 2008. Chapter 5: “Barack Obama and Joe Biden’s Plan To Lower Health Care Costs and Ensure Affordable, Accessible Health Coverage For All,” accessed December 5, 2008.
- Meltzer D, Huang E, Jin L, Shook M, Chin M. 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) 2003; **23(6)**: 576.
- Meltzer D, Basu A, Conti R. The Economics of comparative effectiveness studies: Societal and private perspectives and their implications for prioritizing public investments in comparative effectiveness research. *Pharmacoeconomics* 2010; **28(10)**:843-853.
- Munkin MK, Trivedi PK. Bayesian analysis of self-selection model with multiple outcomes using simulation-based estimation: an application to the demand for healthcare. *Journal of Econometrics* 2003; **114**: 197–220.
- Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine* 2005; **24**:1455–1481.

- Neyman J. Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principe" 1923. French. [English translation of excerpts by Dabrowska D, Speed T. *Statistical Sciences* 1990; **5**: 463-472]
- Pauly, MV. The economics of moral hazard: Comment. *The American Economic Review* 1968; **58(3)**: 531-537.
- Pauly MV, Held PJ. Benign moral hazard and cost-effectiveness analysis of insurance coverage. *Journal of Health Economics* 1990; **9(4)**: 447-461.
- Pauly MV. Adverse selection and moral hazard: Implications for health insurance markets. In Sloan F, Kasper H (eds.), *Incentives and Choice in Health and Health Care*, MIT Press, Cambridge MA, 2008.
- Pauly MV, Blavin FE. Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignoring. *Journal of Health Economics* 2008; **27**: 1407-1417.
- Quandt RE. The estimation of parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 1958; **53(284)**: 873-80.
- Quandt RE. A new approach to estimating switching regression. *Journal of the American Statistical Association* 1972; **67(338)**:306-10.
- Roy AD, Some thoughts on the distribution of earnings. *Oxford Economic Papers* 1951; 3: 135-146.
- Rubin D. Estimating causal effects of treatment in randomized and non-randomized studies *J. Educ. Psych* 1974; **66**: 688-701.
- Rubin D. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 1978; **6**: 34-58.
- Sculpher MJ. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics* 2008; **26**:799-806.
- Senn S, Hoboken, NJ, Cross-Over Trials in Clinical Research. (2nd ed.) Wiley, 2002.
- Simpson DD, Joe GW, Broome KM. A national 5-year follow-up of treatment outcomes for cocaine dependence. *Archives of General Psychiatry*, 2002; **59**, 538-544.
- Sindelar, J., M. Jofre-Bonet, et al. Cost-effectiveness analysis of addiction treatment: paradoxes of multiple outcomes. *Drug and Alcohol Dependence* 2004; **73**: 41-50.
- Sisko A, Truffer C, Smith S, Keehan S, Cylus J, Poisal JA, Clemens MK, Lizonitz JHealth Spending Projections Through 2018: Recession Effects Add Uncertainty to the Outlook. *Health Affairs*, Web Exclusive 2009; w346-w357
- Stinnett AA, Paltiel DA. Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics* 1996; **15(5)**: 641-653.

- Stroup TS, McEvoy JP, Swartz MS, Byerly MJ, Glick ID, Canive JM, McGee MF, Simpson GM, Stevens MC, Lieberman JA. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophrenia Bulletin* 2003; **29(1)**:15-31.
- Thaler R, Rosen S. The value of saving a life: evidence from the labor market. In Nestor E. Terleckyj, ed., *Household production and consumption*. National Bureau of Economic Research, New York, 1975.
- Tunis SR, Stryer DB, Clancy CM. Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy. *Journal of the American Medical Association* 2003; **290**: 1624-1632
- 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.
- Zarkin G, Dunlap L, Bray J, et al. The effect of treatment completion and length of stay on employment and crime in outpatient drug-free treatment. *Journal of Substance Abuse Treatment* 2002; **23(4)**:261-267