



Published in final edited form as:

Health Econ. 2014 March ; 23(3): 359–373. doi:10.1002/hec.2996.

HETEROGENEITY IN ACTION: THE ROLE OF *PASSIVE* PERSONALIZATION IN COMPARATIVE EFFECTIVENESS RESEARCH

ANIRBAN BASU^{a,*}, ANUPAM B. JENA^b, DANA P. GOLDMAN^c, TOMAS J. PHILIPSON^d, and ROBERT DUBOIS^e

^aDepartment of Health Services, University of Washington, Seattle and the National Bureau of Economic Research, Cambridge MA, Seattle, WA, USA

^bDepartment of Health Care Policy, Harvard Medical School, Boston, MA, USA

^cLeonard D. Schaeffer Center for Health Policy and Economics, University of Southern California, California, USA

^dIrving B. Harris School of Public Policy Studies, University of Chicago, Chicago, USA

^eNational Pharmaceutical Council, Washington D.C., USA

Abstract

Despite the goal of comparative effectiveness research (CER) to inform patient-centered care, most studies fail to account for the patient-centeredness of care that already exist in practice, which we denote as passive personalization (PP). Because CER studies describe the average effectiveness of treatments rather than heterogeneity in how individual patients respond to therapies, clinical or coverage policies that respond to CER results may undermine PP in clinical practice and generate worse outcomes. We study this phenomenon empirically in the context of use of antipsychotic drugs in Medicaid patients with schizophrenia using novel instrumental variable methods. We find strong support for PP in clinical practice and demonstrate that the average effects from a CER study cannot be replicated in practice because of the presence of PP. In contrast, providing physicians with evidence to further personalize treatment can produce significant benefits.

Keywords

comparative effectiveness; heterogeneity; passive personalization; schizophrenia; antipsychotic drugs

1. INTRODUCTION

Growing public funding and interest in comparative effectiveness research (CER) have fueled discussions of how patient-centered health care can be best achieved through CER (Garber and Tunis 2009; Clancy and Collins 2010; Conway and Clancy 2010; Wu *et al.*, 2010). Despite the broader CER priority placed on improving individualized/personalized patient care (Institute of Medicine, 2009; Wu *et al.*, 2010), most randomized controlled trials (RCTs) and observational studies describe the average effectiveness of treatments rather than how individual patients respond to therapies. Identifying therapies which have heterogeneous treatment effects and predicting those effects for individual patients is important to meeting the goals of CER (Trusheim *et al.*, 2007; Gabler *et al.*, 2009; Garber and Tunis 2009; Basu, 2011a, 2011b; Basu *et al.*, 2011).

The focus of CER on the average effectiveness of treatments has important implications for the impact of CER on patient health. Clinical guidelines and insurer coverage policies based on average treatment effectiveness may lead providers to choose treatments that are less effective for an individual patient, though more effective for the population on average (Kravitz *et al.*, 2004; Trusheim *et al.*, 2007; Basu, 2011a, 2011b). For instance, new CER may lead physicians and patients to switch treatments despite already being on a potentially maximal regimen. Similarly, although CER assists clinicians in choosing the most effective initial treatment for a disease, idiosyncratic patient information not accounted for in RCTs—for example, additional comorbidities, sociodemographic factors, and health-related behaviors—may favor use of a treatment that is less effective on average but more effective for a particular patient.

Although traditional CER studies attempt to capture treatment heterogeneity by studying the responses of patients in different subgroups, most subgroups are not narrowly enough defined to reflect differences in how individual patients respond to therapies. This is best illustrated by individualized therapies for cancer which work primarily in patients with specific genotypes, regardless of whether they fall into commonly defined subgroups based on age, race, and gender (Gautschi *et al.*, 2008). However, to what extent combinations of patient characteristics, such as their demographics, comorbidities, preferences, genetics, and even environmental contexts within which care is delivered, can help predict such individualized care, even in the absence of genetic information, has been rarely studied (Kaplan *et al.*, 2010).

Clinical practice, however, often does attempt to individualize care even in the absence of direct evidence guiding how such variation in individual choices may be best implemented. Such personalization may be denoted as *passive personalization* (PP) where, in the absence of explicit and active research to discover identifiers, patients and physicians ‘learn by doing’ mostly because of the repeated use of similar products on similar patients. In other words, clinicians engage in across-patient and within-patient learning and are able to match individual patients with treatment more closely than what the evidence would suggest for an average patient. These decisions certainly rely on a host of characteristics that are directly observed by the physicians, but not all of them are observed by us, the analyst of the data. For example, even in the presence of a handful of over-the-counter headache medications,

many people may have a very good sense of which medications work best for them in order to control headache. This is because most people have had the opportunity to engage in some form of trial and error to arrive at their preferred medication.

Despite the recognition that the goal of CER is to inform patient-centered care (Garber and Tunis 2009; Institute of Medicine, 2009; Wu *et al.*, 2010; Basu, 2011a, 2011b) and that substantial variation in treatment effectiveness exists across patients (Kaplan *et al.*, 2010), data are lacking on the quantitative importance of this variation, the extent of PP, and the implication for patients of basing treatment choices on the average effectiveness of treatments. There is large literature on research methods in economics that directly allows answering these questions (Heckman and Vytlačil 1999, 2001 2005; Heckman 2001). Recently, these methods were applied in the context of health (Basu *et al.*, 2007) and were extended to estimate person-centered treatment (PeT) effects (Basu 2013), which can be used to study the extent of PP.

In this paper, we applied these methods to a large database of Medicaid patients with schizophrenia, initiating treatment with an atypical antipsychotic drug (AAD). We examined the importance of treatment heterogeneity in CER by retrospectively analyzing how the effectiveness of AADs (measured in hospitalizations avoided) varied across this population. AADs represent the primary treatment for patients with schizophrenia and are among the largest drug classes in Medicaid (Bruen and Ghosh, 2004). Although heterogeneity in response to AADs has been recognized since their introduction (Meltzer 1986; Insel, 2011), major RCTs of AADs have lacked information on this heterogeneity (Lieberman *et al.*, 2005; Jones *et al.*, 2006). For example, the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) suggested that risperidone was associated with significantly lower time-to-discontinuation and quality of life (Lieberman *et al.*, 2005; Rosenheck and Lieberman, 2007) but did not provide information on the individual distribution of treatment effects across therapies. CATIE was also not powered to look at utilization measures such as hospitalization, which represent a composite dimension of clinical outcomes in patients with schizophrenia (Kreamer *et al.*, 2009; Meltzer *et al.*, 2009).

In this work, by using novel instrumental variable (IV) analyses, we predicted the effectiveness of a group of AADs that are currently generic (risperidone and olanzapine) compared with a group of AADs that are currently branded (ziprasidone, quetiapine, and aripiprazole) for each individual patient, allowing us to estimate for that patient whether the chosen therapy was most effective for them. During the time span of our data, 2002–2005, all the AADs were branded. We examined how often physicians chose the most effective among the two groups of therapy for a patient and estimated the impact on patient health of initiating all patients on the most effective group for the population on average. We also explored the role of subgroup analyses in explaining individual level variation in treatment effects.

2. METHODS

2.1. Conceptual framework

Figure 1a displays the hypothetical distribution in effectiveness of two treatments for schizophrenia (generic versus branded atypical antipsychotics) in preventing hospitalization. Both groups are more effective than placebo but vary in their effectiveness for a given patient. For patients below the 45-degree line, generics are associated with more hospitalizations compared with branded AADs and vice versa. In an RCT, the generics would be relatively less effective on average in preventing hospitalization (represented by the diamond).

If physicians respond to results from an RCT of these treatments by starting all new patients on branded AADs rather than generics, overall hospitalizations may rise or fall depending on how accurately, before the RCT results came out, physicians selected treatments that were optimal for each patient. For example, if physicians were unable to anticipate a patient's treatment response, choosing generic AADs for all patients would minimize hospitalizations because branded AADs are less effective on average in our stylized example. However, if physicians correctly matched patients to optimal treatments on the basis of patient characteristics, RCT evidence could increase overall hospitalizations if, following those results, all patients were treated with branded AADs rather than some by generics. The extent to which physicians can correctly individualize treatments for patients determines whether hospitalizations will be greater or less than would occur if all patients were treated with the most effective drug on average.

Applying this framework to Medicaid patients with schizophrenia, we estimated the individual patient-level distribution of effects on hospitalizations when starting patients on alternative groups of AADs. We estimated how frequently physicians chose the treatment predicted by our model to be most effective and modeled the impact on hospitalizations of initiating AAD therapy with the most effective drug for the population on average.

3. ECONOMETRIC METHODS

3.1. Study sample

We assembled pharmacy and medical claims of continuously enrolled Medicaid patients with schizophrenia who initiated an AAD during 2003 or 2004 in 24 states. Patients were identified by the presence of at least one ICD-9 diagnosis code for schizophrenia between 2002 and 2005. The first pharmacy claim date for an AAD (risperidone, olanzapine, quetiapine, ziprasidone, or aripiprazole) during 2003 or 2004 was defined as their index start date. Only 'clean starters' were included by excluding those with a pharmacy claim for an AAD in the 6 months preceding the index start date. The sample was restricted to patients who: (i) were continuously enrolled in Medicaid for 12 months before and after the index start date; (ii) were not eligible for Medicare; and (iii) were alive for the entire 2-year period in the sample. The data were de-identified and exempted from review by the institutional review board at University of Southern California.

3.2. Study variables

We estimated the effect of AAD-group choice on: (i) all-cause and (ii) schizophrenia-related hospitalizations, accounting for detailed patient comorbidities and demographic characteristics. Hospitalizations were measured in the 12 months following the index date. These hospitalization outcomes are excellent metrics to study the extent of personalization because, in this population, hospitalizations are the main drivers of expenditures. Moreover, the primary reason for hospitalizations is exacerbations of ‘positive’ symptoms in schizophrenia, such as hallucinations and delusions, which are directly controlled by these second generation drugs. Personalization occurs in terms of being able to control these positive symptoms, and the results manifest themselves in terms of averted hospitalizations. In addition, various side effects of treatment also lead to hospitalizations. Personalization also occurs in reducing side effects.

Atypical antipsychotic drugs were categorized dichotomously into generics (risperidone and olanzapine) versus branded (ziprasidone, quetiapine, and aripiprazole) groups on the basis of the statuses of these AADs in 2012. Additional independent variables included patient age, gender, race, Elixhauser comorbidity indicators (Table I) on the basis of medical claims in the year prior to initiating an AAD, indicators for hospitalizations in the prior year (schizophrenia-related, any psychiatric, or all-cause hospitalizations), and indicators for prior use of an AAD more than 6 months prior. Prescribing patterns of a patient’s physician were reflected by indicators for whether the physician prescribed AADs 1, 2, 3, 4, 5, 6 or more than six times in the prior 6 months.¹ Clinical measures of psychiatric symptom severity were unavailable and are most likely correlated with both hospitalizations and AAD choices.

To address unobserved confounding, we identified two IVs: (i) the frequency with which a patient’s physician prescribed (currently) generic AADs during the 6 months prior to the patient’s index start date and (ii) the average rate of (currently) generic AAD use in a patient’s zip code in the year prior to the index date. Both variables were expected to be associated with the use of AADs in the generic group but not otherwise associated with a patient’s risk of hospitalization (Brookhart and Schneeweiss 2007). This is because the population served by US Medicaid consists of low-income patients, and their incomes are less likely to be correlated to the physician propensity to prescribe a branded AAD. Moreover, the physicians prescribing AADs are not responsible for determining hospital admissions. Note that if PP was perfect, then physician preferences would not be a good instrument, but we would be able to detect such a situation easily as that would also mean that physician preferences are not a strong predictor of future use of a particular brand of drug. To control for residual area-level confounders, we controlled for state-level fixed effects. The IVs were validated by examining whether observed patient characteristics and the distribution of specific non-risperidone AAD choice varied above and below the median predicted probability of risperidone use.

¹All patients in our analytic sample saw a physician who wrote at least one prescription for AAD in the last 6 months. Over 87% of our sample patients saw physicians who wrote more than six prescriptions for AADs in the last 6 months.

4. ECONOMETRIC METHODS

4.1. Introduction on PeT effects

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 neglected. For example, in randomized settings, heterogeneity analyses are often accomplished using post hoc subgroup analyses. In the evaluation literature, such nuanced treatment effects are most popularly characterized by conditional average treatment effects (CATE), where an average treatment effect (ATE) 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 each of these factors. In practice, however, this is seldom done. Rather, CATEs are established over univariate risk factors one at a time. We will study the utility of such an approach in our example.

Importantly, in most applied work, 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 they are often observed by individuals and acted upon by some while making treatment selection (Heckman 1997; Heckman and Vytlačil, 1999). An entire genre of methods, including methods based on local IV (LIV) approaches, have been developed to estimate policy-relevant and structurally stable mean treatment effect parameters in the presence of essential heterogeneity (Heckman and Vytlačil 1999, 2001, 2005). The LIV methods identify marginal treatment effects (MTEs), which are the building blocks for all mean treatment effect parameters. Basu *et al.* (2007, 2011) introduced these methods to the health economics literature where essential heterogeneity is widespread and IV methods are gaining meteoric popularity. Carniero 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.

Local instrumental variable methods can seamlessly explore treatment effect heterogeneity across both observable characteristics and unobserved confounders and also be used to establish CATE on the basis of observed factors. In a recent paper, Basu (2013) developed a new individualized treatment effect concept called 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 make these choices in an observational data setting to predict their individualized treatment effects. In our schizophrenia example, suppose that we not only have data on age of the patients but also the treatment they choose and their physician's preferences for prescribing certain treatments. Assume that these physician preferences impart a barrier to freely choose treatments and therefore influence treatment selection but do not affect the potential outcomes for these patients under either treatment, that is, they are IVs. Under such circumstance, 60-year old patients, who go to doctors that are more inclined to prescribe generics and still get a prescription for a branded drug are more likely to have a different distribution of unobserved confounders than 60-year-old patients who go to branded-drug-prescribing doctors and receive a branded drug. Thus, by taking into account treatment choices and the observed circumstances under which those choices were made, one can enrich CATE to form a 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 of 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 than 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.

Technical details, included derivations, identification, and validations using simulation are published elsewhere (Basu 2013). In the succeeding text, we provide a summary of the estimation of the PeT effects for our data at hand.

4.2. Formal models for person-centered treatment effects

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 *generic receipt (treated)* state denoted by $j = 1$ and the *branded receipt (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,

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

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 captures all remaining unobserved random variables. $(X_O, X_U) \perp\!\!\!\perp \vartheta$ and $X_O \perp\!\!\!\perp X_U$ where $\perp\!\!\!\perp$ denotes statistical independence.

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

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

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_O and Z , and U_D is a random variable that captures X_U and all remaining unobserved random variables influencing choice. By definition, $U_D \perp\!\!\!\perp \vartheta$, which also defines the distinction between X_U and ϑ in 1. Equation 1 and 2 represent the nonparametric models that conform to Imben's and Angrist's (1994) independence and monotonicity assumptions needed to interpret IV estimates in a model of heterogeneous returns (Vytlačil, 2002). As in Heckman and Vytlačil (1999, 2001, 2005), we can rewrite 2 as

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

where $V = F_{U_D|X_O, Z}[U_D|X_O, Z]$, $P(x_O, Z) = F_{U_D|X_O, Z}[\mu_D(X_O, Z)]$, and F represents a cumulative distribution function. Therefore, for any arbitrary distribution of U_D conditional on X_O and Z , by definition, $V \sim \text{unif}[0, 1]$ conditional on X_O and Z .

Under regular IV assumptions, Heckman and Vytlačil (1999) show that MTEs can be identified by

$$\frac{\partial E_{\vartheta}(Y|x_O, Z)}{\partial p} = E_{\vartheta}((Y_1 - Y_0)|X_O, V=v) = MTE(x_O, v), \quad (4)$$

where $Y = D*Y_1 + (1 - D)*Y_0$ is the observed outcomes. An MTE is perhaps the most nuanced estimable effect (Heckman 1997; Heckman and Vytlačil, 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), that is, $P(x_O, Z) = V$.

Basu (2013) extends the LIV methods to identify PeT effects, which, for persons who choose treatment, follow

$$\begin{aligned} E_{X_U|X_O, P(Z), D} E_{\vartheta}(Y_1 - Y_0|x_O, P(z), D=1) &= E(Y_1 - Y_0|x_O, V < P(z)) \\ &= P(z)^{-1} \int_0^{P(z)} MTE(x_O, v) dv \end{aligned} \quad (5)$$

Similarly, the conditional effect for a person who did not choose treatment is obtained by integrating *MTEs* over values of V that are greater than p .

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_0 , $P(Z)$, and D . It then averages the *MTEs* over those margins but not all as in CATE. Therefore, a PeT effect is basically the X - Z -conditional effect on the treated (TT) for persons undergoing treatment and is the X - Z -conditional effect on the untreated (TUT) for persons not undergoing treatment. Further details can be found in Basu (2013).

4.3. Estimation of person-centered treatment effects

Estimation of PeT effects follows an LIV approach prescribed by Heckman and Vytlačil (1999). In the face of treatment effect heterogeneity, the advantages of LIV approach over traditional IV methods such as 2SLS and 2SRI are widely documented. LIV is also a two-stage approach, where the first stage comprises running a propensity score model for receiving the generic group of AADs where IVs were included as covariates along with other observed confounders. In the second stage, the predicted probability of starting on generic group, estimated from the first stage, was used in conjunction with other observed confounders to form a control function that approximates both the observed and the unobserved part of the outcomes equation. Specifically, for hospitalization outcomes, this control function was represented by semi-parametric generalized linear models with flexible links and variance functions (Extended Estimating Equation, Basu and Rathouz 2005). Various goodness-of-fit tests were used to ascertain the goodness of fit for the models to the data.

The first derivative of this control function with respect to the estimated propensity score estimated at any particular value of propensity score (p) produces an estimate of *MTE* ($MTE(x_0, v)|_{v=p}$) conditional on the levels of observed factors and unobserved confounders denoted by x_0 and v respectively. For each person, a PeT effect is then calculated by averaging *MTEs* over values of v that would suffice 3, given that person's values for $P(x_0, z)$ and D .

The PeT effects can be trivially aggregated over the observed distribution of (X_0 , $P(Z)$, and D) to estimate mean treatment effect parameters such as the TT, TUT and the ATE. These derivations are provided in Heckman and Vytlačil (1999).

A 1000-replicate bootstrapped sample was used to derive 95% CI for these effects.

4.4. Variation in patient-level treatment effectiveness across generic versus branded groups of atypical antipsychotic therapies

Person-centered treatment effects are illustrated to envision the spread of individual level heterogeneity. To study the utility of subgroup analysis, where typically the subgroups are defined on the basis of the levels of a single factor, the percentage of total variance in estimated PeT effects that are explained by variance across subgroups of observed factors (e.g., gender and age) are calculated.²

4.5. Modeling the impact of clinical decisions based on comparative effectiveness research

Estimated PeT effects represent the difference in the predicted number of hospitalizations for each patient under two scenarios, initiating therapy with generics versus branded AADs. How often physicians selected treatments that were predicted by our model to lead to fewer hospitalizations for a given patient are examined. The average treatment effects of generics for those patients who received generics (TT), and separately for those who chose branded AADs (TUT) are calculated. If physicians accurately chose patients who might benefit from generics over other branded, the average increase in hospitalizations with generic use among those initiated on generics would be less than those initiated on branded AADs. This can be construed as significant evidence for PP.

4.6. comparative effectiveness research-influenced clinical decision/policy effects

The average number of hospitalizations if all patients in our sample were initiated on the most effective AAD group on average are calculated and compared with the average number of hospitalizations under: (i) the status quo (in which patients and physicians selected a particular AAD on their own) and (ii) a scenario in which patients were matched to the AAD group predicted by our model to be most effective for them.

Let $E(Y)$ represent the observed mean number of hospitalizations. These estimates are calculated as follows:

$$E(Y_1) = Pr(\text{Generic Receipt}) * E(Y|\text{Generic Receipt}) \\ + Pr(\text{Branded Receipt}) * \{E(Y|\text{Branded Receipt}) + E(\text{PeT}|\text{Branded Receipt})\}$$

$$E(Y_0) = Pr(\text{Generic Receipt}) * \{E(Y|\text{Generic Receipt}) - E(\text{PeT}|\text{Generic Receipt})\} \\ + Pr(\text{Branded Receipt}) * E(Y|\text{Branded Receipt})$$

Finally, the optimal average number of hospitalizations in which patients were matched to the AAD group predicted by our model to be most effective for them is calculated as

$$E(Y_{OPT}) = E\{I(\text{Generic Receipt} \ \& \ \text{PeT} > 0) * (Y - \text{PeT}) \\ + I(\text{Generic Receipt} \ \& \ \text{PeT} \leq 0) * Y \\ + I(\text{Branded Receipt} \ \& \ \text{PeT} > 0) * Y \\ + I(\text{Branded Receipt} \ \& \ \text{PeT} \leq 0) * (Y + \text{PeT})\}$$

where $I()$ is an indicator function.

5. RESULTS

Between 2003 and 2004, 31,079 patients with schizophrenia initiated therapy with (currently) generic AADs (i.e., risperidone or olanzapine) and 47,452 with (currently) branded AADs (ziprasidone, quetiapine, and aripipazole). There were several significant differences in observed demographics, prior health care utilization, and Elixhauser

²The R-squares on regression of the estimated PeT effects on categorized levels of a single factor provide these metrics.

comorbidities (Table I), which include average proportion of females (47.5% years for generic versus 55.4% for branded AADs, $p < 0.0001$), and the probability of a schizophrenia-related hospitalization in the year prior to the index start date (28.9% vs. 26.9%, $p < 0.001$), prior use of AAD monotherapy (13.2% vs. 11.7%, $p < 0.001$), average numbers of Elixhauser's comorbidities (1.7 vs. 1.86, $p < 0.001$). Significant differences in rates of specific comorbidities were also found (Table I). Overall, patients receiving branded AADs were likely to have more comorbidities than those receiving generic AADs, although rates of substance abuse were higher in the later group.

Patients receiving generic AADs experienced significantly higher numbers of overall and schizophrenia-related hospitalizations within 1 year.

Both IVs were significantly associated with initial use of generic group of AADs ($p < 0.001$ for each individual IV and jointly). Observed patient characteristics and even the distribution of specific AAD choice among patients receiving branded group of AADs balanced well across patients above and below the median of the IV-based predicted probability of generic group use (e.g., $p = 0.20$ for chi-squared test of distribution of AAD choice).

5.1. Impact of initial atypical antipsychotic drug group on subsequent hospitalization

5.1.1. Average treatment effect—Compared with branded AADs, starting all patients with schizophrenia on generics was predicted to significantly increase the average number of overall hospitalizations by 0.35 (95% CI 0.02, 0.67) and nonsignificantly reduce schizophrenia-related hospitalizations by 0.07 (95% CI -0.28, 0.10) (Table II).

5.1.2. Effect on the treated—Among those patients initiated on the generic group, the average annual number of overall hospitalizations was predicted to nonsignificantly increase by 0.17 (95% CI -0.17, 0.44), and the schizophrenia-related hospitalizations was predicted to significantly reduce by 0.15 (95% CI -0.38, -0.03) with the generic group versus the branded group. In fact, compared with the ATEs, the TTs indicate that the selective use of generic AADs results in significant reduction in both types of hospitalizations compared with a nonselective or random use of these drugs. This provides strong evidence of PP in clinical practice.

5.1.3. Effect on the untreated—In comparison, among patients initiated on branded AADs, the average annual number of overall hospitalizations was predicted to significantly increase by 0.61 (95% CI 0.29, 1.05), whereas schizophrenia-related hospitalizations were not predicted to change with generic group versus the branded group of AADs.

5.2. Distribution of person-centered treatment effects on hospitalizations

Besides the average effects documenting evidence on PP, a large variation on PeT effects was observed on both overall and schizophrenia-related hospitalization. In Figure 2a, the PeT effects on overall hospitalization (y -axis) are plotted against the same on schizophrenia-related hospitalization (x -axis). The correlation between the PeT effects across these two outcomes was significantly positive ($\rho = 0.45$, 95% CI: 0.23, 0.67), indicating that patients who would benefit from the generic group of AADs on schizophrenia-related hospitalization

are also more likely to benefit on overall hospitalizations. Indeed, as Figure 2a illustrates, 69% of patients fall in the north-east and the south-west quadrants. The PeT effects also show that about 59% (95% CI: 49%, 69%) and 68% (95%CI: 52%, 84%) of the patient population would benefit from starting on generic group of AADs in terms of reduced overall and schizophrenia-related hospitalizations, respectively. However, the remaining patient population is expected to benefit more from starting on branded groups of AADs.

Interestingly, on the basis of these estimated PeT effects, physicians were NOT more likely than average to initiate the treatment that was predicted by our model to be most effective. For example, among patients initiated on either generic or branded drugs, 59% (95% CI 39.1–62.3%) were predicted to incur more overall hospitalizations if instead started on the branded group (identical to the 59% in the overall population). In this scenario, PP in the use of AADs is therefore driven by the magnitude of effects at the patient level rather than the frequency of positive or negative effects. This is illustrated in Figure 2b, where the same plot from Figure 2a is presented categorized by actual treatment choices. Such evidence are in line with the fact that physicians may be able to discern a better signal-to-noise ratio and therefore more effectively personalize treatments for patients who would experience higher magnitude effects than those where effects are small.

5.3. Ability of subgroups to explain variation in treatment effectiveness

Subgroups explained little of the variation in the estimated PeT effects (Table III). For example, the percent of overall variance in the treatment effect of risperidone explained by gender, race, or age was less than 4%. Allowing for all possible combinations of the Elixhauser disease indicators (creating 5905 subgroups) explained at best 88% of the variance in individual treatment effects on overall hospitalization but much lower for schizophrenia-related hospitalization.

5.4. Impact of clinical/policy decisions based on comparative effectiveness research

A typical Medicaid-representative CER study, which randomly assigns patients to either the generic or the branded group would have documented an average comparative effect of reduction of 0.35 hospitalization by branded AADs over generic AADs (Table II). However, the average annual number of overall hospitalizations when all patients were initiated on branded AADs was 1.73 (95% CI 1.59–1.87), similar to (and not significantly different from) the observed number of annual hospitalizations under the status quo that captures the PP by physicians (1.83, 95% CI 1.81 –1.85) (Table IV). That is, in practice, following the results of a CER study in this case would have resulted in only one-third of the benefit that the CER study suggests. This would happen because many patients would be harmed by following the results of the CER study, especially those who benefit for the generic AADs and were receiving them before the CER study.

Nevertheless, patient-centeredness in CER can be of tremendous value. The average annual number of hospitalizations could be reduced to 1.32 (95% CI 1.26–1.40), a 28% decrease from status quo, if patients were matched to the AAD predicted to be most effective for them (significant difference with status quo, $p < 0.001$). This highlights the potential value

of further personalization of care purely on the basis of factors based on which physicians are already making treatment selections.

6. DISCUSSION

Using a large database of Medicaid patients with schizophrenia initiating treatment with atypical antipsychotic therapy, we estimated significant patient-level heterogeneity in rates of hospitalization associated with use of currently generic versus branded group of AADs. On average, initiation on branded AADs was associated with fewer hospitalizations compared with generics. Compared with average hospitalization rates observed in practice, clinical guidelines or insurer coverage policies requiring patients with schizophrenia to initiate therapy with branded AADs would have resulted in minimal benefit, but substantial increase in costs due to the price of branded drugs. Conversely, if coverage policies required initial use of generics (as they are cheaper), the number of hospitalizations would likely increase, more than what the average effect would suggest. The deviation from average effects in either case is driven by the extent of PP in practice that is unaccounted for in typical comparative studies. In our study, physicians were more likely to prescribe an AAD for a given patient that was predicted to produce a larger magnitude of benefit than the average. This indicates that the treatment choice of physicians was not random.

Our analysis highlights that when patients vary in their response to different therapies, individualized decision-making may improve or worsen outcomes compared with when all patients are treated with the most effective drug on average. If physicians can correctly identify which patients should receive which treatments—as they may be able to do in some diseases and for some patients but not others—overall outcomes may be unchanged or even improved if self-selection of treatments is allowed to occur. As in other studies (Kantoff *et al.*, 2010), subgroups themselves explained little of the heterogeneity in individual patient responses. For example, we predicted that a clinical trial with 5905 subgroups (based on all possible combinations of Elixhauser comorbidities) would only explain 88% of the variation in treatment effects. Any practical attempt to predict individual responses to therapy must therefore explore and develop predictive algorithms on the basis of the combinations of clinical and sociodemographic factors.

Our analysis has several limitations. Although physicians in our study appeared to be somewhat successful in matching patients with schizophrenia to the most effective therapy, our analysis did not explore which characteristics of patients could be used to predict individual responses to therapy. However, the estimated PeT effects can be readily used as outcomes to develop prediction algorithms that would help clinicians with treatment selection based on observed characteristics. It can also inform whether other important unobserved characteristic needs to be measured in order to refine these algorithms.

The estimated heterogeneity in treatment effectiveness for patients with schizophrenia may also not generalize to other chronic diseases, and the predicted accuracy of physician treatment choices in schizophrenia may also not be typical. A more robust understanding of how much treatment decisions based on average effectiveness improve patient health beyond the individualized treatment choices physicians already make is important to

assessing the value of CER. Despite IV validation tests, our statistical approach to predicting patient-level heterogeneity in response to various antipsychotics is ultimately not the same as directly observing heterogeneous treatment effects within an individual patient with the help of observed biomarkers. Although cross-over RCTs allow for the demonstration of patient heterogeneity by observing the same patient in two different treatment scenarios (Kravitz *et al.*, 2004), they are rarely used for this purpose. Identifying which treatments are likely to exhibit heterogeneous treatment effects is important since the relative benefit of choosing a treatment based on average effectiveness is greater only when significant heterogeneity does not exist or one treatment is superior to another for nearly all patients.

This rationale is also compelling for cost-effectiveness analysis. We do not explore the effect of PP on costs, but such work is important. Understanding the effect of PP on both costs and effectiveness could shed new light to the assessment of existing technologies by incorporating quantitative estimate of how these technologies are being used in practice.

The American Recovery and Reinvestment Act of 2009 and the Patient Protection and Affordable Care Act made an unprecedented commitment to CER. Despite the recognition that the goal of CER is to ultimately improve the quality of care through evidence-based personalized medicine, most existing CER studies describe the average effectiveness of treatments across broad groups of patients. The value of CER lies in allowing physicians and patients to incorporate data from the experiences of others into personalized decision-making. Implementing therapies that are best on average for all patients may lead to suboptimal outcomes. Recognizing the importance of treatment heterogeneity and understanding the impact of treatment standardization is a necessary step forward to harnessing the value of CER.

ACKNOWLEDGEMENTS

This research was funded by the National Pharmaceutical Council with additional support from the National Institute on Aging through its support of the RAND Roybal Center for Health Policy Simulation (7P30AG024968). Dr. Basu acknowledges financial support from the National Institutes of Health grants R01MH083706, RC4CA155809 and R01CA155329.

REFERENCES

- Basu A. Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *Journal of Health Economics*. 2011a; 30(3):549–559. [PubMed: 21601299]
- Basu A. Estimating decision-relevant comparative effects using instrumental variables. *Statistics in Biosciences*. 2011b; 3(1):6–27. [PubMed: 22010051]
- Basu A. Person-centered treatment (PeT) effects using instrumental variables. National bureau of economic research working paper No w18056. *Journal of Applied Econometrics*. 2013 (In press).
- Basu A, Rathouz P. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*. 2005; 6(1):93–109. [PubMed: 15618530]
- Basu A, Heckman J, Navarro-Lozano S, Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics*. 2007; 16(11):1133–1157. [PubMed: 17910109]
- Basu A, Jena AB, Philipson TJ. The impact of comparative effectiveness research on health and health care spending. *Journal of Health Economics*. 2011; 30(4):695–706. [PubMed: 21696840]

- Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *International Journal of Biostatistics*. 2007; 3(1):14.
- Bruen, B.; Ghosh, A. [accessed December 21, 2009] Medicaid prescription drug spending and use. Kaiser Commission on Medicaid and the Uninsured Issue Paper. 2004. <http://www.kff.org/medicaid/upload/Medicaid-Prescription-Drug-Spending-and-Use.pdf>
- Carniero P, Lee S. Estimating distribution of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics*. 2009; 149:191–208.
- Clancy C, Collins FS. Patient-centered outcomes research institute: the intersection of science and health care. *Science Translational Medicine*. 2010; 2(37):37cm18.
- Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. *Journal of the American Medical Association*. 2010; 303(10):985–986. [PubMed: 20215615]
- Gabler NB, Duan N, Liao D, et al. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials*. 2009; 10:43. [PubMed: 19545379]
- Garber AM, Tunis SR. Does comparative-effectiveness research threaten personalized medicine? *New England Journal of Medicine*. 2009; 360(19):1925–1927. [PubMed: 19420360]
- Gautschi O, Mack PC, Davies AM, Jablons DM, Rosell R, Gandara DR. Pharmacogenomic approaches to individualizing chemotherapy for non-small-cell lung cancer: current status and new directions. *Clinical Lung Cancer*. 2008; 9(3):S129–S138. [PubMed: 19419927]
- Heckman JJ. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources*. 1997; 32(3):441–462.
- Heckman JJ. Accounting for heterogeneity, diversity and general equilibrium in evaluating social programmes. *The Economic Journal*. 2001; 111:F654–F699.
- Heckman JJ, Vytlacil E. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*. 1999; 96(8):4730–4734.
- Heckman, JJ.; Vytlacil, E. Local instrumental variables. In: Hsiao, C.; Morimue, K.; Powell, JL., editors. *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in the Honor of Takeshi Amemiya*; New York: Cambridge University Press; 2001. p. 1-46.
- Heckman JJ, Vytlacil E. Structural equations, treatment effects and econometric policy evaluation. *Econometrica*. 2005; 73(3):669–738.
- Imbens G, Angrist J. Identification and estimation of local average treatment effects. *Econometrica*. 1994; 62(2):467–475.
- Insel, TR. Rethinking schizophrenia. Director’s Blog. National Institute of Mental Health; 2011. <http://www.nimh.nih.gov/about/director/publications/rethinking-schizophrenia.shtml> [Accessed Sep 14, 2011]
- Institute of Medicine. [Accessed May 21, 2009] Initial national priorities for comparative effectiveness research: committee on comparative effectiveness research prioritization board on health care services. 2009. <http://www.iom.edu/CMS/3809/63608.aspx>
- Jones PB, Barnes TR, Davies L, et al. Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: cost utility of the latest antipsychotic drugs in schizophrenia study (cutlass 1). *Archives of General Psychiatry*. 2006; 63:1079–1087. [PubMed: 17015810]
- Kantoff PW, Higano CS, Shore ND, et al. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *New England Journal of Medicine*. 2010; 363:411–422. [PubMed: 20818862]
- Kaplan SH, Billimek J, Sorkin D, Ngo-Metzger Q, Greenfield S. Who can respond to treatment?: Identifying patient characteristics related to heterogeneity of treatment effects. *Medical Care*. 2010; 48(6):S9–S16. [PubMed: 20473205]
- Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Quarterly*. 2004; 82(4):661–687. [PubMed: 15595946]

- Kreamer HC, Glick I, Klein DF. Clinical trial design lessons from the CATIE study. *American Journal of Psychiatry*. 2009; 166:1222–1228. [PubMed: 19797435]
- Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*. 2005; 353:1209–1223. [PubMed: 16172203]
- Meltzer HY. Novel approaches to the pharmacotherapy of schizophrenia. *Drug Development Research*. 1986; 9(1):23–40.
- Meltzer D, Basu A, Meltzer HY. Comparative effectiveness research for antipsychotic medications: how much research is enough? *Health Affairs*. 2009; 28(5):w794–w808. [PubMed: 19622539]
- Rosenheck RA, Lieberman JA. Cost-effectiveness measures, methods, and policy implications from the clinical antipsychotic trials of intervention effectiveness (CATIE) for schizophrenia. *Journal of Clinical Psychiatry*. 2007; 68:e05. [PubMed: 17335313]
- Trusheim MR, Brendt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nature Reviews Drug Discovery*. 2007; 6:287–293.
- Vytlačil E. Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*. 2002; 70(1):331–341.
- Wu AW, Snyder C, Clancy CM, et al. Adding the patient perspective to comparative effectiveness research. *Health Affairs*. 2010; 29(10):1863–1871. [PubMed: 20921487]

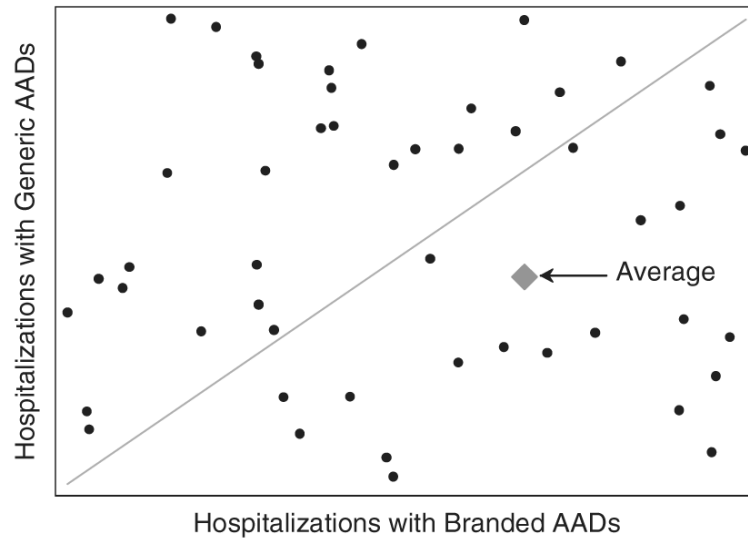


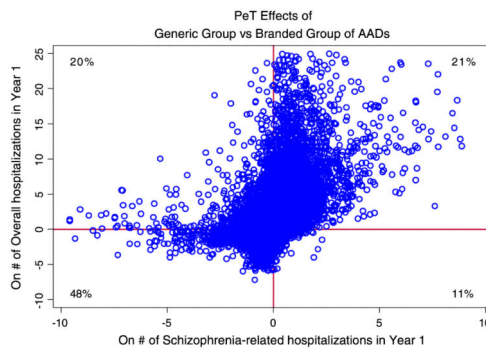
Figure 1. Hypothetical individual-level joint distribution of hospitalizations across generic versus branded atypical antipsychotic drug (AADs)

Author Manuscript

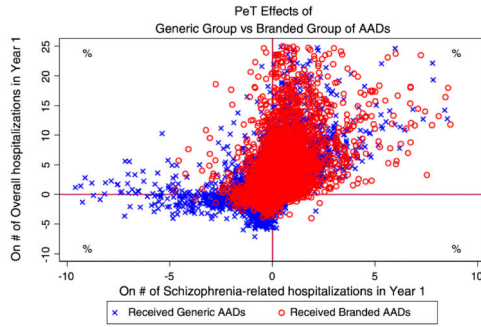
Author Manuscript

Author Manuscript

Author Manuscript



(a) Overall distributions



(b) Distributions categorized by actual treatment choices

Figure 2. Distribution of person-centered treatment (PeT) effects on overall and schizophrenia-related hospitalizations across atypical antipsychotic therapies

Table I

Characteristics of patients, according to initial atypical antipsychotic therapy

Variable	Generic AADs (N = 31,079)	Branded AADs (N = 47,452)	p-value
	Mean	Mean	
Average age, year (SD)	40.8 (11.9)	40.4 (11.5)	0.93
Female, %	47.5	55.4	<0.001
Race, %			<0.001
White	18.0	23.4	
Black	11.3	10.1	
Other	2.6	2.5	
Unknown	68.1	64.0	
Health care utilization prior to index start date of AAD, %			
Any schizophrenia hospitalization	28.9	26.9	<0.001
Any psychiatric hospitalization	56.8	56.3	0.14
Any hospitalization	59.7	59.2	0.12
Any prior AAD mono-therapy	13.2	11.7	<0.001
Any prior AAD poly-therapy	1.1	1.4	0.001
Comorbidities, %			
Average no. Elixhauser indicators (SD)	1.71 (1.9)	1.86 (1.9)	<0.001
Congestive heart failure	2.3	2.4	0.25
Pulmonary circulation disorder	1.1	1.2	0.70
Hypertension, uncomplicated	18.1	20.8	<0.001
Hypertension, complicated	1.6	1.6	0.80
Paralysis	1.5	1.3	0.11
Other neurological disorders	8.8	9.6	<0.001
Chronic pulmonary disease	14.6	16.6	<0.001
Diabetes without complications	8.6	11.6	<0.001
Diabetes with complications	1.2	1.5	0.001
Hypothyroidism	3.1	4.4	<0.001
Liver disease	3.5	3.6	0.81
Chronic peptic ulcer	1.5	1.3	0.01
Obesity	5.5	8.4	<0.001
Weight loss	1.8	1.4	<0.001
Fluid and electrolyte disorders	7.6	8.2	0.006
Blood loss anemia	6.4	6.5	0.81
Alcohol abuse	19.9	18.6	<0.001
Drug abuse	26.2	25.1	<0.001
Psychoses	26.8	28.4	<0.001
Depression	11.4	13.9	<0.001
No. of hospitalizations	1.85 (3.30)	1.80 (3.30)	0.06
No. of schizophrenia hospitalizations	0.77 (1.63)	0.70 (1.59)	<0.001

AAD, atypical antipsychotic drug.

Table II

Predicted impact of generic group atypical antipsychotic drugs (AADs) compared with branded group AADs on average number of hospitalizations in 12 months following initiation of therapy

Group	All hospitalizations	Schizophrenia-related hospitalizations
	Mean (95% CI)	Mean (95% CI)
All patients (ATE)	0.35 (0.02, 0.67)	-0.07 (-0.28, 0.10)
Patients initiating therapy with generic group (TT)	0.17 (-0.17, 0.44)	-0.15 (-0.38, -0.03)
Patients initiating therapy with branded group (TUT)	0.61 (0.29, 1.05)	0.002 (-0.13, 0.22)
TT—ATE	-0.18 (-0.13, -0.28)	-0.08 (-0.04, -0.12)

ATE, average treatment effect; TT, effect on the treated; TUT, effect on the untreated.

Table III

Percent variation in person-centered treatment effects explained by subgroups

Subgroup	#of subgroups	Variance in treatment effects explained by subgroups (%)	Variance in treatment effects explained by subgroups (%)
Gender	2	1.2	1.0
Race	4	4.5	1.0
Age	47	4.2	2.1
Total number of comorbidities	15	46.7	18.4
Unique combinations of comorbidities	5,905	87.8	67.1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table IV

Predicted hospitalizations in 12 months following initiation of atypical antipsychotic therapy under various therapeutic scenarios

Scenario	Average annual number of hospitalizations (95% CI)	% change from Status quo	<i>p</i> -value
Status quo	1.83 (1.81–1.85)	–	–
All patients started on branded group of AADs	1.73 (1.59–1.87)	–5.5	0.15
All patients started on generic group of AADs	2.07 (1.91–2.23)	13.1	0.001
All patients started on optimal predicted therapy	1.32 (1.26–1.40)	–27.9	<0.001

AAD, atypical antipsychotic drug.

Notes: *p*-values reflect comparisons of average annual number of hospitalizations under various scenarios to status quo.

Branded group of AADs include ziprasidone, quetiapine, and aripipazole. Generic groups of AADs include risperidone and olanzapine.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript