

Cost-Saving Effects of Olanzapine as Long-Term Treatment for Bipolar Disorder

Yuting Zhang

Ph.D., Department of Health Policy and Management, University of Pittsburgh, Pittsburgh, PA, USA.

Abstract

Background: Promoters of new medications often argue that using newer drug can reduce use of non-drug medical services and therefore reduce total healthcare spending. This cost-offset argument is plausible both in theory and in practice, but rigorous research on specific drugs or drug categories is needed to make targeted and efficient policy and management decisions.

Aims of the Study: I examined the drug-offset hypothesis for bipolar disorder, an important yet under-studied clinical condition where effective medication treatments can service as substitutes for non-drug medical treatments. I compared two first line long-term treatments, a new atypical antipsychotic medication, olanzapine, and a traditional mood stabilizer, lithium.

Methods: I used private sector insurance claims data collected from a nationally representative sample of U.S. health plans between January 1998 and December 2001. I first selected a cohort of patients with bipolar disorder who were continuously enrolled for at least two years. I then used a propensity-score method to match individuals taking each drug on observed variables that are known to affect medication choices. The central challenge for estimation is that drug treatments are not randomly assigned among patients with bipolar disorder. To identify a causal link between choice of drugs and non-drug medical spending, I employed three different advanced econometrics techniques to assess the robustness of findings; namely interrupted time series, differencing strategies, and an instrumental variables approach.

Results: I found that compared to *similar* lithium users, olanzapine users spent approximately \$330 more on monthly average non-drug medical services during the first year after initiation of drug treatment. The higher spending for olanzapine users was accounted for by both higher rates of re-hospitalization and more outpatient visits. In addition, olanzapine cost \$153 per month while lithium cost \$16 per month. Including the direct cost of the drugs, compared to similar patients taking lithium, patients with bipolar disorder taking olanzapine spent \$5,600 more annually on health care services.

Discussion: These findings do not support the hypothesis that new drugs “pay for themselves” by reducing the need for other health care services in the case of olanzapine for bipolar disorder. This does not mean that the new drug is not “cost-effective” because increased “benefits” associated with the drug in terms of the improved quality of life may be worth the increased costs. However the findings do indicate that “cost-offsets” must be measured and not taken for granted. Incorporating such drug-offset evidence into policy and business decisions can facilitate appropriate clinical practices and improve efficiency of resource allocation. The methods used in this study to test for cost-offsets can be applied to other clinical areas and drug classes.

Received 10 October 2007; accepted 15 May 2008

Introduction

Between 1990 and 2006, prescription drug spending in the United States increased from \$50 billion to \$219 billion, and currently accounts for 13% of total healthcare expenditures.* Use of newer drugs accounts for one third of this expenditure growth.† If newer medications are highly effective, they may reduce total healthcare costs by improving control of chronic illnesses and reducing use of services such as hospital and emergency department care. If these savings in other services can offset, or more than offset, the incremental increase in prescription drug costs, net cost savings to society could result. These potential savings are termed “drug-offset” effects.

The drug-offset effect has important policy and business implications. Policy makers have used evidence on drug-offset in different stages of a new drug’s life cycle, such as approval (e.g., in Australia and the Canadian province of Ontario), pricing (e.g., in Australia and Japan), and drug benefit design (in most industrialized countries for both private and public health plans). Promoters of new drugs often claim that new drugs can save money despite their high

* Correspondence to: Dr. Yuting Zhang, 130 DeSoto Street, A664 Crabtree Hall, Dept. of Health Policy and Management, University of Pittsburgh, Pittsburgh, PA 15261, USA.
Tel.: +1-412-383 5340
Fax: +1-412-624 3146
E-mail: ytzhang@pitt.edu

Source of Funding: The National Institute of Mental Health, the Alfred P. Sloan foundation, and the Pharmaceutical Policy Research Fellowship provided jointly by Harvard Medical School and Harvard Pilgrim Health Care.

* Predicted by the Centers for Medicare and Medicaid Services, “National Health Expenditure Projections 2006-2016,” <http://www.cms.hhs.gov/NationalHealthExpendData/downloads/proj2006.pdf> (accessed 14 March 2007).

† Henry J. Kaiser Family Foundation, “Prescription Drug Trends Fact Sheet-June 2006 Update,” <http://www.kff.org/rxdrugs/3057.cfm> (accessed 14 March 2007).

costs because they lead to a reduction on overall health care costs.* However, decision makers currently have little evidence regarding drug-offset effects. Furthermore, the evidence on cost-offsets for one drug used to treat one medical condition cannot be generalized to other drugs and conditions, because of wide variations in costs, effectiveness and appropriate use of drugs. Drug-offset effects should be examined for a specific drug or drug class used to treat a condition, rather than at an aggregated level across medical conditions.

The drug-offset effect is more likely to occur in a medical condition with effective medication treatments and associated with a higher rate of hospitalization. This paper examines the drug-offset hypothesis for bipolar disorder, an important yet under-studied clinical condition where drug treatments would be effective substitutes for non-drug medical services. The paper also tests several sophisticated statistical methods to investigate drug-offset effects. These methods can be generalized to examine cost-saving effects for other new drugs or drug classes.

Drug-offset Hypothesis for Bipolar Disorder and Drug-offset Evidence

Drug-Offset Hypothesis for Bipolar Disorder

Bipolar disorder affects one to two percent of the U.S. general population¹ and costs \$45 billion annually of which 40% are direct medical costs and 60% indirect economic costs.² Patients with bipolar disorder shift moods between severe depression with increased risk of suicide and the intense highs of mania that lead to functional impairment and anxiety, and therefore the illness is associated with high rates of hospitalization. The rate of hospitalization is 20%, higher than the average rate at 10% for other medical conditions.† The primary treatment for bipolar disorder is medication, including atypical antipsychotic agents, anticonvulsant agents and lithium. These medications are effective in controlling both manic and depressive episodes and preventing relapses to hospitalization.^{3,4} Most relapses occur within the first six months after an acute episode and thus treatment guidelines recommend that patients stay on drugs for a few months. There are only three first-line choices for long-term treatment of bipolar disorder: lithium, olanzapine* or valproate according to the American Psychiatric Association guidelines.⁵ Lithium was approved by the Food and Drug Administration (FDA) in 1970; olanzapine was approved by the FDA for acute care of bipolar disorder in March 2001 and for long-term treatment in 2004. While valproate, a drug initially approved in 1978 to treat seizure disorders, has never been approved to treat bipolar disorder, it has been used to treat the mania and rapid-cycling form of bipolar disorder. Valproate is not the focus of this study because it is not completely comparable with the other two due to its lack of approval and lesser effectiveness in treating depressant phase of bipolar disorder. Thus, this study examines the drug-offset effect of olanzapine (new) relative to lithium (old) as long-term treatments for bipolar disorder.

Lithium, a traditional mood stabilizer, is inconvenient to take and associated with severe side effects. Patients who take

lithium are required to visit clinics regularly for blood tests used to adjust dosage, which often changes with diet, other medications, or even the seasons. Rates of premature discontinuation of lithium are high⁶ because of its side effects such as neurotoxicity (including paraesthesia, ataxia, tremor and cognitive impairment), acne, diarrhea, dysfunction of kidney and drug interaction with Angiotensin-Converting Enzyme (ACE) inhibitors, Non-steroidal anti-inflammatory drugs (NSAIDs) or diuretics. Costs of the illness skyrocket among patients who terminate treatment due to increased rates of rehospitalization, suicide, functional impairment and full-episode recurrences following discontinuation of medication treatment.⁷ Therefore both efficacy and *tolerability* of medication regimens are important in reducing total costs of treatment. Olanzapine, a new second-generation antipsychotic medicine, is generally regarded as better tolerated and more convenient to take than lithium. Olanzapine was first approved in 1996 for schizophrenia, but it had been prescribed off-label to treat bipolar disorder before the FDA's approval for bipolar indication in March 2000. If the prophylactic use of olanzapine leads to a higher level of adherence to treatment regimens, and subsequent effectiveness in preventing relapses, the total cost of treatment might fall. Thus, I hypothesized patients who used olanzapine as first-line drug prevention therapy should have lower expenditures on non-drug medical services, compared with *similar* patients who used lithium.

Current Drug-offset Evidence

Literature on the drug-offset effect is sparse. At the aggregate level, Lichtenberg^{8,9} used the 1996 Medical Expenditure Panel Survey (MEPS) data and noted that overall new drugs more than pay for themselves in the form of reduced non-drug medical spending. Some recent studies have questioned the robustness of Lichtenberg's original studies. Zhang and Soumerai used the 1996-1998 US MEPS data to replicate the original results.¹⁰ Through step-by-step modifications to the model and assumptions, they found that the original results were not sustained. Moreover, grouping all drugs together as the Lichtenberg study does may result in unreliable estimates of drug-offset effects due to the wide variation in efficacy, safety, cost, and appropriateness of use of *both* old and new drugs.

An alternative approach is to examine the ability of new drugs to reduce non-drug medical utilization for a specific disease. One clinical area where drug-offset effects are plausible is atypical antipsychotic agents (AAs), compared with traditional antipsychotic agents, due to the new drug class's better side-effects profile. Using Medicaid recipients from the state of California for the 1993-2001 period, Duggan¹¹ evaluated the link between the introduction of a new drug class and reduced costs in other medical services for patients with schizophrenia. He used three different identification strategies to demonstrate a consistent result: Medicaid spending on AAs increased 610% during the study period, but the shift to new drug treatments did not reduce spending on other medical services. Frank, McGuire and Normand used geographic variations in the timing of approval for different AAs to identify the causal link between the use of new drug and non-drug spending.¹² They did not find a significant cost-saving effect of AAs used to

treat schizophrenia.

AAs are commonly prescribed to treat bipolar disorder, but there is little literature to examine the drug-offset hypothesis of AAs for this subgroup of patients, especially using real-world utilization data. I examined the drug-offset hypothesis of olanzapine, compared with lithium, as maintenance treatment for bipolar I disorder.

The current study contributes to the drug-offset effect literature in three ways. First, while using real-world observational data from administrative claims, I implemented a careful study design and “pre-processing” techniques to create a study sample that mimics a randomized controlled trial. I also used strict clinical inclusion and exclusion criteria to eliminate complications due to switching, premature discontinuation and interactions of drugs. Second, I compared various identification methodologies to obtain consistent results. These methods can be generalized to investigate cost-saving effects for other drugs and conditions. Third, I provide economic evidence on the comparative effectiveness of olanzapine and lithium as a long-term treatment for bipolar disorder.

Method

Estimation Strategies to Identify the Drug-offset Effect

Using private insurance claims and enrollment data, I constructed a rolling cohort of patients with bipolar disorder who were *newly treated* with either olanzapine or lithium for at least three months and *continuously enrolled* for at least one year before and after initiation of drug treatment. Outcomes (non-drug medical expenditures) are repeatedly measured for the period twelve months before and after treatment initiation.

It is difficult to investigate the effect of new prescription drugs on other types of health care expenditures given that drug treatment is not randomly assigned. Individuals who take a certain drug are likely to differ from their counterparts who do not, in both observed and unobserved characteristics. Failing to control for these characteristics will result in biased estimates (selection problems). I first used a propensity-score method to match patients taking each drug based on observed characteristics known to affect medication choices. Compared with traditional regression controlling for factors expected to influence drug choice, the propensity-score method can force a direct test that the distribution of characteristics in the treated and untreated groups overlaps. Lack of overlap yields imprecise estimates.^{13,14} The propensity score method can also guard against model misspecifications in response models,¹⁵ if the matching achieves the same (multivariate) distribution of all risk factors in the treated and control groups. I used one-to-one “optimal matching,” which provides optimal balance of dispersion and uniformity compared with alternative methods such as “greedy matching” and exact matching.^{16,17}

Propensity-score matching cannot however guard against estimation biases caused by unobserved characteristics that affect both medication choice and non-drug spending. To eliminate these potential biases, I used three identification strategies: interrupted time series, differencing strategies, and an instrumental variables approach. I started with an interrupted time series to examine whether non-drug expenditures after the initiation of drug treatment differ between olanzapine users and lithium users. Such an approach controls the previous trends in non-drug spending and evaluates longitudinal effects of the drug treatment on non-drug spending. As a population-level estimation strategy, however, interrupted time series cannot estimate individual-level characteristics that might affect the outcome variables. I addressed this problem through two individual-level estimate approaches, as discussed below.

The first individual-level estimation approach is a differencing strategy comparing the change in non-drug spending for olanzapine and lithium users after and before drug treatment. Differencing takes advantages of time repeated measures within individuals to eliminate the endogeneity problem caused by individual-specific and time-invariant factors, especially unobserved variables that are not controlled by propensity score matching.* However, differencing strategies cannot mitigate the unobserved, patient-unrelated heterogeneity, nor can they address heterogeneity caused by time-variant characteristics. For example, physician practice patterns simultaneously affect patients’ choice of drugs and other types of health care services. If an older physician is more likely to prescribe older drugs and more likely to refer patients to inpatient care services than a younger physician, I will observe a correlation between the use of an old drug and higher spending in inpatient care services. If patients change their providers over time, then the endogeneity problem due to physician practice patterns cannot be mitigated by the differencing strategy. This problem motivates an instrumental variables approach.

The instrumental variables approach can solve the endogeneity problem due to omitted variables, measurement errors or reverse causality. A good instrumental variable should be highly correlated with the choice of drug therapy but not correlated with unobserved variables that might affect the use of non-drug medical services. I used the timing of the U.S. Food and Drug Administration’s approval of olanzapine for bipolar indication as an instrument to identify the drug-offset effect.

Timing of the FDA approval of olanzapine for bipolar indication can serve as a good instrumental variable because it is highly correlated with choice of drugs and not correlated with unobserved factors that might affect non-drug spending. First, the timing of approval affects medication choice. Before olanzapine was approved in March 2000 to treat acute bipolar disorder, it was prescribed off-label. After the FDA’s approval, the number of new users of olanzapine increased, both for acute and maintenance treatment for bipolar disorder. While the number of patients newly diagnosed with bipolar disorder remained constant in the study period from 1998-2001, the percentage of new long-term (at least three

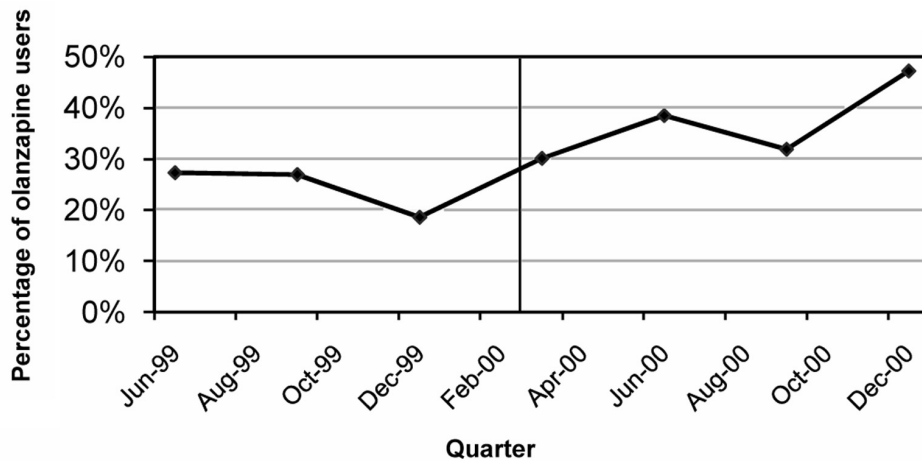


Figure 1. Percentage of New Continuous Users of Olanzapine among Patients Newly Diagnosed with Bipolar Disorder

Notes: The vertical line is March 2000, when olanzapine was approved by the FDA for bipolar indication. More patients with bipolar disorder initiated olanzapine rather than lithium after the FDA approval.

months) users of olanzapine increased substantively after the FDA's approval (see **Figure 1**). Second, the timing of approval is not associated with physician behavior patterns or other unobserved variables that might affect patients' non-drug medical expenditures. Off-label use of drugs in mental health area is very common and not necessarily associated with severity of illness.¹⁸ Instead, choice of drugs used to treat bipolar disorder is largely affected by patients' comorbidity, tolerance of side effects, concerns of convenience and stigma associated with drugs. For example, a patient with a family history of diabetes might be willing to take lithium to avoid weight gain and metabolic effects associated with olanzapine use, whereas a patient with kidney dysfunction might take olanzapine to avoid side effects from lithium. Therefore timing of the FDA approval is a good instrument that satisfies the two main conditions. I specify these three empirical strategies – propensity score matching, differencing strategy and instrumental variables approach in Data Analytic Procedures section.

MarketScan Database, Study Sample and Pre-processing Data

MarketScan Data 1998-2001

I analyzed the MarketScan data for the 1998-2001 period. The annual MarketScan database collects private-sector health data from approximately 200 nationally-representative payers in the U.S. The MarketScan data consists of three databases including a benefit plan design database, a commercial claims and encounter database, and a Medicare supplemental database. This study focuses on private non-elderly enrollees solely covered in the first two databases. The benefit plan design database includes benefit plan types and service coverage information. The commercial claims and encounter database includes inpatient and outpatient claims, pharmacy claims, detailed patient information, and enrollment tables. The inpatient admissions claims data includes information on principal diagnoses and procedures,

service date, and payments to facilities and physicians for over-night services provided in a hospital. The outpatient claims data includes encounters and claims provided in a doctor's office, a hospital outpatient facility, an emergency room or other outpatient facility. The pharmacy claims data contains information on national drug codes (NDC, a coding system to identify prescription drugs), dosage forms, days of supply, prescription fill dates and financial types such as copayment, deductible and total gross payment.

Sample Selection

In selecting the study sample, I included individuals who: (i) had a maintenance drug treatment, defined as continuous use of either olanzapine or lithium for at least three months within the 1998-2001 period (13,764 patients); (ii) were continuously enrolled in the health plan network one year prior to and one year following the initiation of each drug treatment (4,677 patients); (iii) were diagnosed with bipolar disorder* but not schizophrenia,† which is defined as having at least one inpatient or two outpatient diagnoses during the two-year continuous enrollment period (1,372 patients); and (iv) had not used any atypical antipsychotics or any other mood stabilizers within six months prior to the initiation of either olanzapine or lithium (933 patients, with 221 olanzapine users and 711 lithium users).

“Pre-Processing” Data Using Propensity Score Matching

If patients who took olanzapine and patients who took lithium were identical except for the drug treatment, I could simply compare the non-drug spending for the two groups. **Table 1** shows that olanzapine users and lithium users were different with respect to some observed demographic characteristics. Specifically, Health Maintenance Organization (HMO) enrollees were more likely to use lithium than olanzapine, compared with Preferred Provider Organization (PPO) enrollees. Patients from the South were more likely to take lithium than patients from other regions. Furthermore, the frequency of being hospitalized for various

Table 1. Descriptive Statistical Summary of the Unmatched Sample by User Group (219 Olanzapine users and 689 Lithium users).

Characteristics	Olanzapine	Lithium	P-Value
Female	57%	57%	0.796
North East	28%	24%	0.117
North Central	24%	23%	0.500
West	2%	3%	0.089
South	40%	47%	0.015*
Ffs	36%	39%	0.257
Hmo	4%	8%	0.026*
POS NO CAP	11%	7%	0.017*
POS CAP	32%	34%	0.401
PPO	17%	12%	0.019*
Age	43(15)	40(15)	0.0001*
Major Diagnosis #	6(3)	5(3)	< 0.0001*
Substance Abuse	0(2)	0(2)	0.810
Epilepsy	0(1)	0(0)	0.246
Heart Disease	1(3)	0(4)	0.586
Diabetes	1(5)	1(4)	0.146
Psychotherapy	6(10)	4(9)	0.0004*
Bipolar Visits	11(15)	10(12)	0.259
Mental Visits	19(20)	15(16)	0.0050*
Bipolar IP Visits	0(1)	0(1)	0.027*
Bipolar OP Visits	11(15)	10(12)	0.294
Mental IP Visits	0(1)	0(1)	0.0008*
Mental OP Visits	18(20)	15(16)	0.0064*
Bipolar Spending\$	2,541(5,843)	1,960(4,702)	0.084
Bipolar IP \$	1,511(5,062)	1,158(3,939)	0.220
Bipolar OP \$	1,029(2,277)	802(1,688)	0.076
Mental Spending\$	3,835(6,971)	2,616(5,422)	0.0022*
Mental IP \$	2,046(5,662)	1,410(4,358)	0.048*
Mental OP \$	1,789(3,029)	1,206(2,008)	0.0005*
Total Medial \$	9,521(15,194)	5,385(8,869)	< 0.0001*

* indicates characteristics are significantly different between olanzapine and lithium users at p-value of 0.05.

Notes:

- (i) All variables are measures in the one-year baseline period before drug initiation. "IP" is inpatient care and "OP" is outpatient care.
- (ii) The first 11 rows in the table are demographic variables, sex, enrollment region, health insurance type and age. Health insurance plans include fee-for-service (FFS), HMO, point-of-service (POS with or without capitation), and preferred provider organization (PPO).
- (iii) High s.d. demonstrates high volatility in mental health spending. s.d. measures the variability of data around the mean, which is square root of sample variance, and s.e. measures how well this sample mean estimates population mean.

reasons within a year before drug treatment was significantly higher for olanzapine users than lithium users.

To ensure that two groups (primary olanzapine vs. lithium users) are comparable at baseline, I used propensity score matching, conducted in two stages: first, I calculated the estimated probability (i.e., propensity score) of being an olanzapine user based on some observed variables known to affect medication choice; second, I used the propensity score as a benchmark to match the closest individual within the lithium group to a member of the olanzapine group. The variables used in calculating propensity score in the first stage include demographic variables (age, sex and region), health insurance type (including HMO, PPO with capitation or without capitation, FFS, and POS) and patients' health status (measured by the number of health service visits and indicators of major conditions in the year before drug treatment).

I used two R libraries, MATCHIT and OptMatch, to perform the matching process to obtain an optimal matched sub-sample that balances uniformity and dispersion (i.e., minimum weighted distance measured by propensity score). MATCHIT uses semi-parametric and non-parametric matching methods to preprocess data and it reduces model dependence on functional forms and improves parametric statistical models later used in the analyses. Statistical models after appropriately preprocessing with MATCHIT, without other modification, can produce inferences that are more robust and less sensitive to modeling assumptions.¹⁹ The MATCHIT results in 212 lithium users matched with 212 olanzapine users.

The comparison of patients' characteristics for the matched sample (see **Table 2**) shows that two groups are well balanced with respect to all the observed covariates in the year before initiation of drug treatment. This sample of 212

Table 2. Descriptive Statistical Summary by User Group in the Baseline (Matched sample: 212 Olanzapine users and 212 Lithium users).

Characteristics	Olanzapine	Lithium	P-Value
Female	54%	58%	0.408
North East	29%	32%	0.418
North Central	28%	25%	0.422
West	2%	3%	0.527
South	32%	35%	0.492
FFS	33%	33%	0.891
HMO	5%	6%	0.839
POS NO CAP	14%	12%	0.548
POS CAP	32%	28%	0.378
PPO	16%	22%	0.113
AGE	39(17)	40(15)	0.364
Major Diagnosis#	6(3)	6(3)	0.430
Substance Abuse	1(6)	0(3)	0.505
Epilepsy	0(1)	0(0)	0.133
Heart Disease	0(1)	0(3)	0.616
Diabetes	1(4)	1(4)	0.481
Psychotherapy	7(12)	7(11)	0.962
Bipolar Visits	12(16)	13(14)	0.647
Mental Visits	20(19)	20(18)	0.862
Bipolar IP Visits	0(1)	0(1)	0.756
Bipolar OP Visits	12(16)	13(14)	0.658
Mental IP Visits	0(1)	0(1)	0.603
Mental OP Visits	19(19)	19(18)	0.882
Bipolar Spending\$	2,448(4,228)	2,721(5,284)	0.558
Bipolar IP \$	1,343(3,383)	1,713(4,880)	0.365
Bipolar OP \$	1,105(2,134)	1,008(1,549)	0.594
Mental Spending\$	3,922(5,606)	3,430(5,836)	0.377
Mental IP \$	2,127(4,413)	1,919(5,050)	0.653
Mental OP \$	1,795(2,799)	1,511(1,878)	0.220
Total Medical \$	7,410(8,959)	7,216(9,178)	0.826
Propensity Score	0.272(0.112)	0.270(0.110)	0.843

* indicates characteristics are significantly different between olanzapine and lithium users at p-value of 0.05.

Notes:

- (i) All variables are measures in the one-year baseline period before drug initiation. "IP" is inpatient care and "OP" is outpatient care.
- (ii) The first 11 rows in the table are demographic variables, sex, enrollment region, health insurance type and age. Health insurance plans include fee-for-service (FFS), HMO, point-of-service (POS with or without capitation), and preferred provider organization (PPO).
- (iii) High s.d. demonstrates high volatility in mental health spending. s.d. measures the variability of data around the mean, which is square root of sample variance, and s.e. measures how well this sample mean estimates population mean.

lithium users and 212 olanzapine users will be used for most estimation analyses except for instrumental variable approach.

Outcome and Covariates Measures

The outcome variables are non-drug spending in three levels of related services: total medical services, mental health services, and bipolar disorder related services. The key explanatory variable is a dummy indicator that equals one for olanzapine users and zero for lithium users.

The covariates include patients' demographic variables, health status indicators, and medication treatment patterns. Patients' demographic variables are sex, age categories (less than 17, 18-34, 35-44, 45-54, 55-64), insurance plan types, and enrollment region (North-Central, Northeast, West, and South). Patients' previous health status is measured by the

number of inpatient care visits and outpatient care visits, the number of psychotherapies, the number of major diagnostic categories, and the number of incidents for the following conditions: heart disease, epilepsy (ICD-9 codes 345.0-345.9), substance abuse (ICD-9 codes 304 or 305), alcoholic psychoses (ICD-9 codes 291.0-291.9) and diabetes (ICD-9 code 250). Clinical evidence of olanzapine's effectiveness is well established for medium and high dosages (5mg, 10mg, and above) but not for a low dosage of 2.5mg. In order to control for the different effects due to different dosage levels, I created an indicator to identify patients who only took the low dosage form of olanzapine and control for it in the regressions.

Data Analytic Procedures

Interrupted Time Series Analysis

YUTING ZHANG

I used an interrupted time series model to estimate the time trend and level change in medical spending, excluding an interrupted duration.²⁰ I computed average health care expenditures for patients taking olanzapine or lithium for 24 time points, with 12 points before and 12 points after drug initiation. The model specification is presented in equation (1).

$$Y_t = \beta_0 + \beta_1 treatment_t + \beta_2 month_t + \beta_3 post_t + \beta_4 postmonth_t + \beta_5 treatment_t * month_t + \beta_6 treatment_t * post_t + \beta_7 treatment_t * postmonth_t + error_t \quad (1)$$

where, treatment is a dummy variable that equals 1 for olanzapine users and 0 for lithium users.

Month measures time in months from the start of the study period; it ranges from 1 to 24.

Post is an indicator for time t occurring after intervention.

Postmonth measures the numbers of months after the intervention. It is coded 0 for all months before intervention, and equals to 1 to 12 for months after intervention.

The remaining variables are interaction terms between treatment and time series variables.

Based on the definitions above, β_0 estimates the average monthly expenditures of lithium users for the baseline point (i.e., twelve months before they initiated lithium). β_1 estimates the difference in average monthly expenditures between olanzapine users and lithium users at the baseline point. β_2 estimates the incremental change by month in the average expenditure of lithium users before intervention, i.e., the baseline trend. β_3 estimates the change in the monthly-average expenditures for lithium users immediately after the intervention. β_4 estimates the incremental change in the trend in the monthly-average expenditures after intervention. The interaction terms estimate the incremental difference between olanzapine users and lithium users in each time series variable. For example, β_6 estimates the difference between olanzapine users and lithium users in the change in non-drug spending after and before drug treatment. This is essentially a difference-in-difference estimate.

Autocorrelation is detected through a plot of raw residuals against time. In the model, I adjusted for serial autocorrelation, and the Durbin-Watson statistic for the final regression model is close to 2, indicating the model is adjusted for autocorrelation. Throughout the analysis, I used maximum likelihood methods to estimate the effects, starting with a full model that includes all interaction terms as shown in equation (1). I then used a stepwise method to estimate reduced models with the exclusion of insignificant interaction terms, eventually reaching the most parsimonious model which only includes significant terms besides the main treatment.

Differencing Strategies – Individual-Level Estimates

One challenge with individual-level estimates comes from some unique characteristics associated with health care expenditure data: (a) a nontrivial fraction of zero outcomes in the sample, and (b) a positively skewed empirical distribution of the nonzero realizations.

The literature has suggested three classes of models to address unique features of health care expenditure data: least-

squares (LS) estimators for the $\ln(y)$ with retransformation, the generalized linear models (GLM) with log link function and gamma distribution, and generalized Gamma model.²¹⁻²⁵ The LS-based methods can be biased in the face of heteroscedasticity if not appropriately retransformed.²¹⁻²⁴ Heteroscedasticity is detected across all time trend related variables in my sample, which makes an appropriate retransformation challenging. The GLM method can yield consistent yet possibly imprecise estimates if the log-scale error is heavy-tailed. Following the algorithm suggested by Manning and Mullahy,^{22,23} and after considering the tradeoff between bias and precision, I chose to use GLM with log link function and gamma distribution. A Box-Cox test detects that the logarithm transformation fits the data well. A modified Park test²⁶ determines the mean-variance relationship and leads to the conclusion that a Gamma distribution assumption fits the data appropriately.*

I conducted four sensitivity analyses for variance-covariance structures within individuals: (i) Huber-White robust variance with a correction for clustering adjusted by patient clusters, (ii) a patient fixed-effect generalized linear model, (iii) a generalized estimation equation to fit population-averaged panel-data model, and (iv) a generalized linear mixed model to allow random correlation effects across time within individuals. All four models assume a log link function for a power transformation of the outcome variable and a Gamma distribution for the mean-variance relationship.

An Instrumental Variables Approach

I used the timing of the FDA approval of olanzapine as treatment for bipolar disorder as an instrument. I created a dummy variable (IV) that equals one if the patient initiated the drug (either olanzapine or lithium) after March 2000 (“post” users) and zero otherwise (“pre” users). The most recent study showed that two-stage residual inclusion (2SRI) estimation results in consistent results in non-linear models.²⁷ I used the total sample before propensity score matching to avoid losing precision due to the combination of propensity score matching and instrumental variable approach. The 2SRI instrumental variables approach is conducted in two stages. In first stage, I used Logistic model to estimate the effect of timing of approval on the probability of being an olanzapine user as opposed to being a lithium user, controlling for other covariates discussed in the sub-section of Outcome and Covariates Measures. The first-stage F-test was 27.63 ($P < 0.05$), indicating a strong instrument.²⁸ I calculated residuals from the first-stage logistic model. In second stage, I included the 1st-stage residual, original endogenous variable (being an olanzapine user or a lithium user), and other covariates in the regression model. To adjust for skew and kurtotic features of the health expenditure data, I used GLM model with Gamma distribution and log link function in second stage.

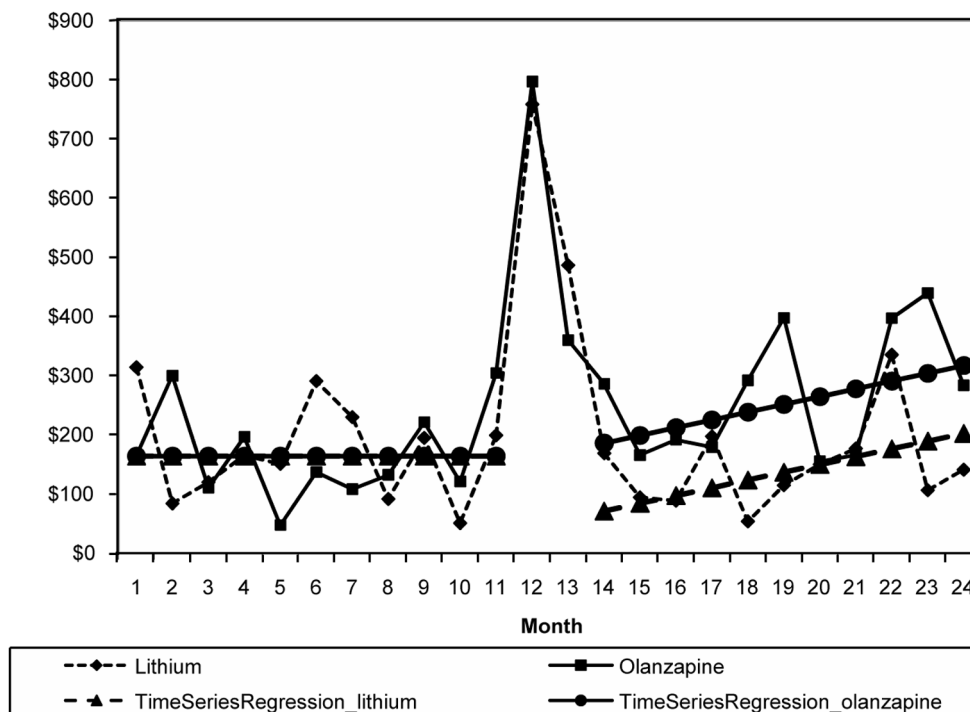


Figure 2. Monthly Non-drug Spending on Bipolar Disorder Related Services.

Notes: Month 13 is the initiation of drug treatment. Each point on the jagged is the monthly average nondrug spending among matched patients. Straight lines are predicted regression lines estimated by interrupted time series models. Time-series regression lines for lithium and olanzapine are overlapped in the month 1 through 11. Some patients were receiving weekly psychotherapy before they started medication treatment, which partially explains the non-drug spending before the drug initiation. Onetime doctor's visit cost a patient from \$75 to \$150 in the period from 1998 to 2001.

Results

Interrupted Time series

Figure 2 demonstrates the time trend of monthly-average non-drug medical spending for bipolar disorder related services. Two points are immediately observable: (i) average non-drug spending was the highest in the month immediately before the drug initiation; (ii) average non-drug spending of olanzapine users was higher than that of lithium users after patients started the medication treatment. Similar patterns were observed for mental health related services and total medical services.

The first observation implies that there is a significant increase in the use of inpatient care in the days leading up to an individual's first prescription for the drugs of interest. For the year after drug initiation, patients' service use was stabilized. This stabilization results in reduced non-drug expenditures compared to the period immediately before drug initiation. The second observation implies that olanzapine users did not spend less on non-drug services than lithium users. This difference was not caused by different severity level of bipolar disorder, as shown in **Table 2** that olanzapine users and lithium users are very comparable in previous spending on bipolar disorder related services.

The interrupted time series results are reported in **Table 3**. Three columns show the results in three levels of outcome

expenditures: non-drug total medical expenditures, non-drug expenditures in mental-health related services and non-drug expenditures in bipolar disorder related services. Lithium users and olanzapine users were comparable in all three levels of previous non-drug expenditures. After initiation of drug therapy, non-drug expenditures on total medical services among lithium users decreased by \$415, while non-drug total expenditures of olanzapine users increased by \$156. The difference between the two drug groups in change in non-drug spending after and before drug treatment was therefore \$571. But the difference-in-difference gap shrank by \$41 each month throughout the study period.* At the end of the first year after drug therapy, the difference reduced to \$125. When spread out over the year, the average monthly difference-in-difference estimate was \$328, which implies that olanzapine patients spent \$328 more in monthly average non-drug medical services during the first year after drug therapy than lithium patients did. Similar results were observed for non-drug expenditures on bipolar disorder related services and mental health related services. After initiation of drug therapy, olanzapine users spent \$180 more on non-drug mental health related services and \$108 more on bipolar disorder related services.

Results from Differencing Strategies

The results from the generalized linear model with a Huber-White robust variance are reported in Column 1 of **Table 4**.

YUTING ZHANG

Table 3. Interrupted Time Series Regression Model to Estimate the Impact of the Drug Treatment on Monthly Non-drug Spending.

	Total Medical Services	Mental Health Services	Bipolar Disorder Services
Intercept	360.116** (5.783)	167.668** (4.603)	173.848** (9.649)
Treatment	-29.773 (-0.533)	16.564 (0.454)	-31.074 (-1.223)
Month	27.981** (3.516)	8.719** (2.039)	
Post	-414.876** (-3.576)	-132.785** (-2.008)	-115.715** (-2.902)
Post Monthly Change			13.943** (3.159)
Treatment* Post	600.896** (4.450)	163.686** (3.190)	139.005** (3.928)
Treatment* Post Monthly Change	-40.526** (-2.557)		

* indicates significance at the 0.1 level, ** indicates significance at the 0.05 level.
Notes: Robust t-statistics are in parenthesis.

The coefficient of the interaction between the treatment and post variables is significant, implying that the change in non-drug expenditures after and before drug therapy differs between olanzapine users and lithium users. Olanzapine users spent more in non-drug expenditures after drug therapy, and the change in spending for olanzapine users was 1.64 times larger than that of lithium users. For an average lithium user, non-drug expenditures in the month immediately after drug therapy were \$414, versus \$740 for an average olanzapine user in that month. This \$326 difference in non-drug expenditures remained the same throughout the first year after drug therapy.

The results are consistent based on sensitivity analysis for variance-covariance structures, reported in columns 2 through 4 in **Table 4**. Specifically, a generalized linear model with patient fixed effects estimated that olanzapine users spent \$356 per month more than lithium users throughout the first year after drug therapy.

Results from Instrumental Variables Estimation

The results from the instrumental variables estimation are reported in **Table 5**. Olanzapine users spent 1.87 times as high as lithium users did in yearly non-drug expenditures after initiation of medication treatment. This result is bigger than the effects from GLM, but not statistically significant. Instrumental variables approach only estimates a Local Average Treatment Effect (LATE) for the subgroup of patients whose choice of drugs is affected by the instrument, timing of approval. These “marginal patients” would use lithium if they initiated drug treatment before March 2000 (timing of approval) but would choose olanzapine if they initiated treatment after the approval. These patients do not include those who used olanzapine as off-label use.

Summary of Findings

Compared to lithium users, olanzapine users spent approximately \$330 more in monthly average non-drug medical services during the first year after initiation of drug treatment. Both higher rates of rehospitalization and more outpatient visits explained the higher spending for olanzapine users. In addition, olanzapine cost \$153 per month* and lithium cost \$16 per month. Factoring in the direct cost of the drugs, patients with bipolar disorder taking olanzapine spent \$470 more in total health care spending than similar patients taking lithium, or approximately \$5,600 more per year.

Conclusions and Implications

Recently, pharmaceutical industry representatives have cited cost offsets of new medications as a justification for their higher prices.* The concept is plausible both in theory and in practice, but rigorous research on specific drugs or drug categories is needed to make targeted and efficient policy and management decisions.

I conducted a careful study design and employed various advanced empirical methods to test the drug-offset hypothesis for a disease and a new medication where cost-offset effects are plausible. Using three empirical strategies, I consistently found that patients with bipolar disorder did not spend less on non-drug services after initiating olanzapine relative to lithium. Patients taking olanzapine spent approximately \$5,600 more per year in total healthcare expenses compared to *similar* patients taking lithium. Thus, in the case of olanzapine use for bipolar disorder, these findings do not support the hypothesis that new drugs “pay for themselves” by reducing the need for other health care services. These findings suggest several possibilities: (i) the superior *pecuniary* effectiveness of olanzapine relative to

Table 4. Generalized Linear Models to Estimate the Impact of Drug Treatments on Total Non-drug Spending.

	(1)	(2)	(3)	(4)
	GLM Gamma	Fixed Effect	GEE Panel	Generalized Linear MIX
Treatment	1.060 (0.39)	1.117 (0.15)	1.120 (0.78)	1.006 (0.05)
Base Monthly Trend	1.012 (0.61)	1.006 (0.68)	1.013 (0.62)	1.043** (3.67)
Post	0.904 (0.46)	0.908 (1.07)	1.027 (0.11)	0.822 (1.42)
Post Monthly Change	0.987 (0.49)	0.982 (1.61)	0.974 (0.85)	0.938** (3.41)
Post Treatment Level Effect	1.642* (2.41)	1.536** (5.47)	1.559* (2.17)	1.747** (3.14)
Age 18-34	0.944 (0.27)		0.748 (1.32)	0.786 (1.40)
Age 35-44	0.699* (2.09)		0.625** (2.58)	0.717* (2.26)
Age 45-54	0.683* (2.14)		0.637* (2.44)	0.663** (2.75)
Age 55-64	1.136 (0.56)		0.882 (0.55)	0.820 (1.24)
Age >=65	1.476 (0.76)		1.412 (0.59)	1.426 (0.91)
Female	1.077 (0.62)		1.193 (1.44)	1.136 (1.30)
South	0.794 (1.88)		0.787 (1.94)	0.780* (2.53)
West	1.402 (0.70)		1.433 (0.76)	0.862 (0.57)
FFS	1.257 (1.84)		1.206 (1.60)	1.434** (3.40)
HMO	1.055 (0.24)		0.939 (0.31)	1.152 (0.77)
Pre # Major Diagnosis	1.136** (6.00)		1.143** (5.93)	1.219** (11.35)
Pre Heart Disease	1.065 (1.02)		1.054 (0.96)	1.055 (1.21)
Pre Diabetes	1.016 (1.38)		1.023 (1.66)	1.013 (1.21)
Pre # Psychotherapy	1.016** (4.23)		1.017** (4.12)	1.022** (5.36)
Pre Epilepsy	1.097 (0.66)		1.083 (0.90)	1.169 (1.79)
Months On Drug	0.999 (0.16)		1.005 (0.92)	1.002 (0.66)
Post Monthly Treatment				0.972 (1.24)
Observations	9366	9366	9366	9366
Number of patients	424	424	424	424

* significant at 5%; ** significant at 1%.

Notes: Robust t statistics in parentheses.

Table 5. Instrument Variable Regressions on the Impact of Drug Treatments on Total Non-drug Medical Spending.

	1st Stage Logit	2nd Stage GLM
Instrument-Post FDA	2.449** (1.754-3.420)	
Treatment		1.897 (0.907-3.967)
1 st Stage Pearson Residual		0.917 (0.680-1.236)
Female	0.835 (0.599-1.165)	0.896 (0.704-1.139)
Age 18-34	0.704 (0.387-1.280)	0.901 (0.583-1.392)
Age 35-44	0.565* (0.333-0.956)	0.702 (0.477-1.035)
Age 45-54	0.684 (0.413-1.132)	0.614* (0.417-0.905)
Age 55-64	0.763 (0.422-1.379)	1.296 (0.841-1.998)
Age >=65	1.808 (0.331-9.891)	1.175 (0.286-4.833)
FFS	0.711 (0.476-1.061)	1.302 (0.978-1.731)
HMO	0.385** (0.192-0.772)	1.008 (0.653-1.557)
Pre Heart Disease	1.017 (0.950-1.089)	0.990 (0.924-1.061)
Pre Diabetes	0.974 (0.917-1.034)	1.016 (0.985-1.049)
Pre # Psychological Therapy	1.027** (1.011-1.043)	1.020** (1.007-1.034)
Pre Substance Abuse	1.011 (0.973-1.049)	1.002 (0.965-1.041)
Pre # Major Diagnosis Codes	1.061 (0.999-1.126)	1.064** (1.017-1.113)
Northeast	0.797 (0.503-1.264)	1.027 (0.735-1.434)
West	0.558 (0.191-1.630)	0.707 (0.341-1.468)
South	0.687 (0.461-1.022)	0.838 (0.634-1.107)
Observations	911	909

* significant at 5%; ** significant at 1%

Notes: Robust 95% confidence intervals in parentheses.

lithium does not fully justify its higher price; (ii) the improved tolerability of olanzapine promotes adherence to other medical care (such as regular physician visits, one component shown in the study that drove up non-drug spending after initiation of drug treatment); and/or (iii) the higher price of olanzapine might be justified by important *non-pecuniary* benefits, such as higher quality of life due to better tolerance or reduced sick days, which are not measured in this study. If the value of such unmeasured non-pecuniary

benefits is above \$5,600 per year for some patients, using olanzapine relative to lithium improves consumer welfare.

There are some limitations of this study. The claims data do not have information on rebates from pharmaceutical manufactures to insurances, so I could not observe the complete transaction prices of drugs. Furthermore, this study only evaluates a partial effect of using new drugs, cost-saving effects rather than entire cost-effectiveness effects. Also the study sample was selected from private sector insurance enrollees. They are either employed in a large corporation or the dependents of the employees. For employees, their condition of bipolar disorder might be less severe compared to patients with bipolar disorder who are covered in disability insurance or elders under Medicare. Thus the results cannot be generalized to an unemployed and elderly population.

Nevertheless this study demonstrates that econometric analysis of observational data can provide useful alternatives to clinical trial data. Unlike clinical trial data, insurance claims data reflect individuals' revealed choices in the real world so this analysis provides information on comparative effectiveness that is more generalizable for policy and management decisions than trial data. The robustness test of various methods used in the study can be applied to study a broad array of cost saving hypotheses, for example, the cost-offset effects of other drug classes, the cost-saving effect of a diabetes management program, or a consumer-directed marketing program.

Evaluating the incremental cost-effectiveness of new versus old drugs can help us understand the clinical and economic justifications for novel therapies. The incremental cost-effectiveness analyses include both the examination of the difference in the total healthcare costs associated with the new drug as well as the change in the quality of life (or other clinical outcomes) – and thus the “cost-offset analyses” is integral to a cost-effectiveness analyses. If innovative medications can prevent hospitalizations and other costly services, payers should encourage their use, in spite of higher initial costs, to minimize overall health system expenditures. However, until a study is actually done, it is not possible to determine whether a cost-offset occurs. With the continued rapid growth in prescription drug spending and pressure for cost control from public and private payers, we face a pressing need for the use of sophisticated drug-offset analyses and a better understanding of drug-offset effects.

Acknowledgement

I am grateful for many valuable comments from Richard Frank, Haiden Huskamp, Tom McGuire, Alyce Adams, Dennis Ross-Degnan, Judy Lave, Julie Donohue, Vardges Levonyan, Rachel Garfield, and Stephen Soumerai. Thanks also go to seminar participants for their helpful comments at the Harvard Health Policy Research Seminar, the Public Finance Seminar at Harvard Economic Department, and the Pharmaceutical Policy Research Seminar at Harvard Medical School.

References

1. Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK. The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry* 1993; **50**(2): 85-94.
2. Sajatovic M. Bipolar disorder: disease burden. *Am J Manag Care* 2005; **11**(3 Suppl): S80-84.
3. Bowden CL. Atypical antipsychotic augmentation of mood stabilizer therapy in bipolar disorder. *J Clin Psychiatry* 2005; **66** (Suppl 3): 12-19.
4. Bauer M. Review: lithium reduces relapse rates in people with bipolar disorder. *Evid Based Ment Health* 2004; **7**(3): 72.
5. APA. *Practice Guideline for the treatment of patients with bipolar disorder* (Second Edition). American Psychiatric Association guideline, http://www.psych.org/psych_pract/treatg/pg/bipolar_revisebook_indexcfm, (accessed September 28, 2006) 2002.
6. Bowden CL, Lecrubier Y, Bauer M, et al. Maintenance therapies for classic and other forms of bipolar disorder. *J Affect Disord* 2000; **59** (Suppl 1): S57-S67.
7. Bowden CL. *New Developments in the Treatment of Bipolar Disorder*. American Psychiatric Association 155th Annual Meeting, <http://www.medscape.com/viewarticle/436383> (Accessed on 21 April 2006) 2006.
8. Lichtenberg FR. Are the benefits of newer drugs worth their cost? Evidence from the 1996 MEPS. *Health Aff* 2001; **20**(5): 241-251.
9. Lichtenberg FR. Do (more and better) drugs keep people out of hospitals? *Am Econ Rev* 1996; **86**(2): 384-388.
10. Zhang YSS. Do Newer Prescription Drugs Pay For Themselves? A Reassessment of the Evidence. *Health Aff* 2007; **26**(3): 7.
11. Duggan M. Do new prescription drugs pay for themselves? The case of second-generation antipsychotics. *J Health Econ* 2005; **24**(1): 1-31.
12. Frank RG, McGuire TG, Normand SL. *Cost Offsets of New Medications for Treatment of Schizophrenia*. Unpublished paper 2006.
13. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983; **70**(1): 41-55.
14. Joffe MM, Rosenbaum PR. Invited Commentary: Propensity Scores. *Am J Epidemiol* 1999; **150**(4): 327-333.
15. Drake C. Effects of Misspecification of the Propensity Score on Estimators of Treatment Effect. *Biometrics* 1993; **49**(4): 1231-1236.
16. Gu x, Rosenbaum PR. Comparison of multivariate matching methods: structures, distances, and algorithms. *Journal of Computational and Graphical Statistics* 1993; **2**: 405-420.
17. Hansen B, Klopfer S. Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, <http://www.stat.berkeley.edu/~bbh/hansenKlopfer2005.pdf>, (accessed September 1, 2006) 2006; 15.
18. Radley DC, Finkelstein SN, Stafford RS. Off-label Prescribing Among Office-based Physicians. *Arch Int Med* 2006; **166**(4): 5.
19. Ho D, Imai K, King G, Stuart E. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 2007; **15**(199-236): <http://gking.harvard.edu/files/abs/matchp-abs.shtml>, accessed on January 21, 2008.
20. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002; **27**(4): 299-309.
21. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ* 2005; **24**(3): 465-488.
22. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *J Health Econ* 1998; **17**(3): 283-295.
23. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ* 2001; **20**(4): 461-494.
24. Duan N. Stochastic microenvironment models for air pollution exposure. *J Expo Anal Environ Epidemiol* 1991; **1**(2): 235-257.
25. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures. *J Health Econ* 2004; **23**(3): 525-542.
26. Park R. Estimation with heteroscedastic error terms. *Econometrica* 1966; **34**: 888.
27. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *J Health Econ*; In Press.
28. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; **63**(3): 557-586.