HC
7/28/06
10:05 AM

Newer is not always better:
Re-evaluating the benefits of newer pharmaceuticals*

Michael Law, M.Sc.† and Karen Grépin, S.M.‡

WORKING PAPER:
This draft paper is intended for review and comments only.
It is not intended for citation, quotation, or other use in any form.

July 27, 2006

*The authors acknowledge and appreciate the very helpful advice of (in alphabetical order): Ernst Berndt, Jacqueline Chattopadhyay, Ben Hansen, Dan Hopkins, Gary King, Elena Llaudet, Frank Lichtenberg, Don Rubin, Ian Yohai and Alan Zaslavsky

†Candidate, Ph.D. Program in Health Policy, Harvard University, mlaw@fas.harvard.edu
‡Candidate, Ph.D. Program in Health Policy, Harvard University, grepin@fas.harvard.edu

Abstract

Whether newer pharmaceuticals justify their higher cost by saving other types of health spending (such as hospitalization and physician care), so-called "offset" effects, is an important health policy question. We aimed to replicate the analysis of Lichtenberg (2001), which suggests the savings from newer drugs substantially outweigh their additional cost. We find our replicated results are highly dependent on the model and the dataset used: substituting either a model less sensitive to expenditure outliers or a newer data release results in the effect disappearing; substituting both causes it to reverse in direction. Further, we propose an alternative method to estimate offset effects in the same dataset using propensity score matching and a two-part expenditure model, which we estimate for essential hypertension. With this model, using a newer drug is associated with $179 higher yearly drug costs, but the change in non-drug spending is indistinguishable from zero, providing no evidence for the presence of offsets for newer hypertension drugs.

Introduction

For many years, prescription drug expenditure in the United States has grown at a rate higher than that of overall health care spending.[1] This spending growth has been partly driven by an increase in the average price paid per prescription, which is partly the result of new drug approvals, which tend to have higher initial prices relative to older drugs for similar conditions.[2] An important question from a health policy perspective is whether or not these newer drugs are worth their higher costs. There are several ways in which newer drugs might justify their higher cost. For example, newer drug purchases could reduce non-drug health expenditures such as hospitalizations, improve a patient's quality of life, or produce benefits outside the health sector (e.g. increase the productivity of workers). In this paper we investigate the first of these possibilities, the so-called "offset" of non-drug health care expenditures through the use of newer medicines.

The studies examining this question have not produced a conclusive answer as to whether such offsets from newer drugs exist. Two general research approaches have been used in past work. One approach has investigated the savings from newer drugs used to treat a single condition relative to another older treatment option in a randomized clinical trial setting.[3] While there are many advantages of this approach, its main limitation is that it usually only allows the comparison of two or a few treatment alternatives, which may not reflect the range of clinical options available in practice. Moreover, many clinical trials are conducted against a placebo, which gives no indication about the specific offset of newer medicines versus older medicines.

A second approach has investigated whether offsets from newer medicines exist using aggregate level observational data. However, the most notable papers using this approach have reached dramatically different conclusions. First, Lichtenberg (2001), examined the effect of using newer drugs on non-drug health care expenditures using the 1996 Medical

Expenditure Panel Survey (MEPS) and found evidence of a significant offset (herein L01).[4] In an updated working paper, Lichtenberg also tested the same basic model using additional years of the MEPS data.[5] However, Zhang and Soumerai (forthcoming), however, suggest these results are not robust to alternative model specifications.[6] Also using the MEPS data, but over a longer time period, Miller et al. (2006) investigated the presence of offsets from newer cardiovascular drugs.[7] After including a number of controls for potentially confounding variables, they find no evidence of an offset effect. Finally, Duggan (2005) examined newer antipsychotic medicines in Medicaid patients and found no evidence of drug offsets and mixed evidence of any improvement in health for these patients using three different estimation strategies.[2]

In addition to the two general approaches mentioned above, a related literature has also investigated the changes in cost of care for a number of medical conditions through constructing price indices. These take into consideration changes in clinical practice and quality, as well as changes in prices of inputs, the case-mix of patients, and the quantity of services delivered. Using data on the treatment of heart attacks from 1983 to 1994, Cutler et al. (1998) concluded that the real cost of treating a heart attack has declined over time.[8] Similarly, using administrative claims data from a large retrospective database, Berndt. et al (2002) found that the real cost of treating major depression has also declined during the early 1990s.[9] Both of these disease conditions had seen the development of a number of new pharmaceutical treatments which may have contributed to these decreases in the cost of treatment per case, but this evidence is only suggestive of drug offsets from newer drugs.

The major methodological difficulty in trying to disentangle the effects of newer drugs using observational data is that drug selection for a given patient is not random. There are likely systematic biases in how people are assigned to treatment: for example, sicker patients receiving more powerful medicines, or older painkillers being given to patients when other treatment options have been exhausted. *A priori*, it is not clear what direction the biases

will go, but the non-random selection of drug treatment means that any estimate of the effects are likely to be unreliable without proper consideration of these selection effects.

We further investigated the existence of drug offsets using observational data. We used as a starting point for our analysis the data and methods employed in L01. After replicating these results, we found that using either a different model less sensitive to outliers or the newest data release causes the offset effect to disappear. We then develop an alternative model to account for the non-random selection effects in the same data sets using a propensity score method and were unable to find evidence of an offset for essential hypertension, the most prescribed-for condition in the dataset. Our findings, therefore, fail to support the existence of offsets for newer drugs for hypertension.

Data Sources

Our analysis uses the same dataset as L01, the 1996 MEPS.[10] The 1996 household component of the survey collected comprehensive data on the personal characteristics and medical care utilization from 22,601 respondents. We obtained copies of the initial release of the 1996 data, the same used in L01, as well as a subsequently released version updated to fix discrepancies in the original release. Unfortunately, we were unable to obtain a copy of the original drug age table used in L01. To substitute, we used data from the Food and Drug Administration's (FDA) online Drugs@FDA database and Mosby's DrugConsult 2006 to estimate drug age based on the original FDA approval date for the active ingredient, the same approach used in L01.[11,12] Combination therapies were coded using the average date of approval for each active ingredient. Matching drugs to dates in the databases was performed using both National Drug Codes and drug names, with discrepancies between the two methods manually checked. Records with missing drug names or National Drug Codes were handled with listwise deletion, which to the best of our knowledge was the method employed in L01.

For our analysis of expenditures related to essential hypertension, we extracted all the records for individuals who reported hypertension in the year, defined by the 3-digit ICD-9 code 401. Total yearly drug expenditures for each respondent were calculated by summing the total expenditures on all drugs received for hypertension during the entire sample period.* Similarly, total non-drug health care expenditures were calculated using all utilization records except for prescription drugs linked to hypertension. This includes records from six databases covering inpatient procedures, office-based visits, outpatient procedures, dental visits, emergency room visits and other medical procedures.

Observations with missing data for any of our model variables were eliminated using listwise deletion. This approach resulted in a very minimal data loss; of the 1917 condition records for hypertension only 2.2% had missing data. We also removed the records of 15 individuals who had multiple condition records for hypertension (n = 31 records) to avoid double-counting events, resulting in a final dataset of 1844 individuals. We defined taking a 'newer' drug as receipt of any drug for hypertension that was still likely under patent in 1996, using "12 years since the date of approval" of the drug as a cut-off for a drug being under patent, since the literature suggests that 12 years is approximately the average effective patent length for a drug approved during the 1980s.[13] By this method, 919 of the individuals in our dataset were 'treated' with newer therapies, whereas 925 were 'controls', as they were using only older medications.

Replication Results

We first attempted to replicate the results of L01 using the same data and model specification. The base model in L01 uses each prescription as the unit of analysis and includes 737 other variables covering a range of demographic and clinical characteristics along with the

*This includes both patient cost-sharing and insurance payments. The data excludes over-the-counter medicines purchased without a prescription.

logarithm of the drug's age as the variable of interest. The following model was used:

$$E(y|\lambda, X) = \gamma ln(\lambda) + \beta X$$

Where y represents the outcome of interest, including total non-drug expenditure for the condition, γ is the coefficient of interest, λ represents the drug age in years and X represents the matrix of control covariates. The original results suggest that each log year of drug age was associated with $18 in additional prescription expenditure, but this was significantly offset by an annual savings of approximately $71 in non-drug related medical expenditures (see Tables 1 and 2). Using this data, we were able to very closely replicate these results with estimates of $17.92 and $67.81, respectively. We expect the remaining differences are likely due to differences in our drug age dataset or our handling of missing data.

Replication of L01 led us to identify what we consider are 3 major limitations with this approach. First, interpreting the original analysis, L01 concludes that the "reduction of $71.09 in non-drug spending is much greater than the $18 increase in prescription cost". However, the unit of analysis in L01 was a single prescribed drug event. The estimate for prescription expenditure used only the expenditure for each individual prescription, but the total non-drug expenditures were calculated for the associated condition over the entire year. Thus, the comparisons would be equivalent only if a person had a single prescription each year. Our analysis of the data indicates there were 4.79 prescriptions per condition in the original analytic dataset. Therefore, even if the original data and specification were less problematic, it was incorrect to conclude that offsets were present (4.79 scripts $\times$ $18/script = $86.22 additional total drug cost > $71.09 additional total other health care spending).

Second, L01 employed Ordinary Least Squares (OLS), which assumes that conditional on covariates the residual components of health expenditures are normally distributed. However, there is a large literature in health economics discussing the distinctly non-normal nature

of most health expenditure data.[14,15] Evidence from our replication of the original model suggests the normality assumption is not being met. The left side of Figure 1 shows a QQ-plot of the residuals of the MEPS dataset, which should be a straight line along the diagonal if the normality assumption is met. The plot shows dramatic deviations from normality resulting from the significant outliers present in the dataset and that the straightforward application of OLS is problematic.

A simple modification that helps account for outliers as well as zero expenditure observations is to transform the model by taking the log of expenditures and adding one dollar to all observations. Thus, the new model estimated is:

$$E(ln(y+1)|\lambda, X) = \gamma ln(\lambda) + \beta X$$

where the coefficients remain the same as in the base model above. The results from this change, shown in Table 2, suggest the offset effect is indistinguishable from zero when modeled on a log scale. Moreover, the right side of Figure 1 shows the residuals from this model, which conform much more closely to a normal distribution, indicating this model better fits the normality assumption. A further concern with the original model is the assumption that the data points are independent from one another. However, this is clearly not the case with the original dataset, which contains multiple observations for not only the same individual, but the same drug and same condition. While this concern is addressed somewhat in a later update to the paper, the model used in this subsequent paper still employs an untransformed dependent variable, making it potentially subject to the specification problems discussed above.[5]

A third limitation is that L01 employs an early release of the 1996 MEPS data, which has since been corrected. We repeated the analysis using the updated dataset and found that the original model for prescription expenditures is approximately the same, $18.93, as shown in

Table 1. However, Table 2 shows the estimate log drug age on total non-drug expenditure is reduced to $0.21 and is statistically indistinguishable from zero. Finally, we ran the modified log model discussed above on the updated dataset. The results in the final line of Table 2 show the estimated coefficient for log drug age reverses in direction, suggesting newer drugs are associated with higher non-drug health care costs-the reverse of the offset hypothesis. Thus, using the newer dataset essentially eliminates the earlier finding, even with no other modifications to the model.

In summary, we reanalyzed the results of L01 and found the results to be highly dependent on the model, specification and dataset used. Our modifications to the approach of L01 suggest there is not conclusive evidence for offsets from newer drugs in the 1996 MEPS data. This result is consistent with the findings of two other independent analysis approaches.[6,7]

An Alternative Approach: Analysis of Hypertension

Building on the analysis above, we suggest an alternative method for examining whether offset effects exist in this dataset. First, we propose that analysis of offsets should follow the example of Duggan (2005) and work within a single disease condition.[2] The L01 approach pooled data across all medical diagnoses. We believe this is problematic because it assumes an extra unit of logged drug age should have the same effect on expenditures across all conditions, but it is not obvious why this would be the case. It seems more reasonable that we should expect very different effects for different conditions, as the drugs available for different conditions are vastly different in age and effectiveness, particularly if a given condition has seen the approval of a revolutionary new drug. If the effect does differ, the estimate for offsets will depend critically on the mix of conditions within the particular sample being analyzed, which is difficult to both replicate and generalize.

We propose to investigate the effect of treatment for a single condition, essential hyper-

tension (ICD-9 number 401). We selected hypertension for three primary reasons. First, to determine the effect of taking a newer therapy, there needs to be variation in prescribing which is unrelated to the health care costs for the patient in question. This is likely to be the case for hypertension therapy, which has a variety of different drug class options and considerable debate regarding their relative effectiveness.[16] As a result, we expect there will be substantial variation between physicians in their propensity to prescribe different medications, be it new or old, for patients with the same disease severity.

Second, there needs to be variation in the age of the therapies given, which is the case with hypertension in the MEPS data. A profile of the top 30 drugs prescribed to MEPS respondents for hypertension is presented in Table 3. It shows substantial variation in the age of the drugs being used with reasonable frequency and demonstrates the diversity in prescribing, with no drug representing more than 7% of total hypertension prescriptions. Third, hypertension is a very prevalent condition associated with high total health care expenditures. Moreover, it is the most prescribed-for condition in the 1996 MEPS. Therefore, if an offset effect resulted from the use of newer antihypertensives, it would be substantively important for policy reasons.

Producing a Propensity Score Matched Dataset

We propose analyzing the data using a propensity score matching procedure to address non-random selection to treatment.[17] This technique matches observations on the estimated probability that an individual receives a 'new' treatment as a product of observable characteristics, such as age, sex and co-morbidities. Because we conceptualize 'treatment' as seeing a physician who prescribes newer medicines on average, we match on a range of variables which might potentially influence prescribing decisions or the propensity to see a given physician. After estimating this probability of receiving a newer therapy using logistic regression,

we match observations that are 'treated' and 'untreated' based on this score and produce a dataset composed of comparable observations which one could observe in a randomized experiment. Following matching, or 'pre-processing', to make the results more robust, we fit a parametric model to estimate the effect of taking a newer medication on total drug costs for the year and total non-drug expenditures to examine the presence or absence of any offset.[18]

We conceptualized 'treatment' to be receipt of any hypertension drug that had been approved by the FDA after 1984, since the literature suggests that 12 years is approximately the average effective patent length for a drug approved during the 1980s.[13] We believe this definition of older vs. newer medicines is a more meaningful definition than the L01 specification as it eliminates the problem of having to assume that a log-year of additional drug age has the same impact throughout the entire lifespan of a drug. Moreover, it is more likely correlated with higher drug prices. It was also the most feasible definition we could implement given the difficulty in retrospectively determining whether the many hundreds of drugs prescribed for hypertension were still on or off market exclusivity when prescribed.[†]

To facilitate matching we employed optimal matching, which minimizes the overall distance between treated and control observations in the matched dataset.[19] We used two R libraries, OptMatch and MatchIt to perform the matching operation.[20–22] The propensity score was estimated using numerous models until good balance on all variables and interactions used to estimate propensity to receive treatment was achieved. Testing many specifications for this model is legitimate, given that the goal is to achieve the best balance possible and no outcome variables are examined until the final dataset is determined.[23] We determined that optimal matching within a caliper of 0.2 standard deviations provided the best balance on observable characteristics. The variables used in our estimation of the

[†]In addition to the 12-year specification, we conducted a sensitivity analysis with both 10 and 14 years as the cutoff for 'newer' drugs. The results are very similar and lead to the same substantive conclusions regarding offset effects.

propensity score are presented in Table 4.

The results of our matching produced a dataset of 815 observations in each of the 'treated' and 'control' groups, for a final sample size of 1630 observations. Matches were not found for 104 'treated' and 110 'controls', so they are excluded from the matched dataset. Table 5 shows the balance across the two groups in observable characteristics. Following matching the distribution of these variables should be similar, which is clearly shown by the very small differences between the groups. As shown in Figure 2, there was large spread in both groups in the propensity for treatment. It should be noted that all subsequent results apply only to the subset of matched observations.

Analysis Using a Two-Part Expenditure Model

The second part of our alternative approach method involves fitting a two-part model to model total non-drug expenditures in the matched dataset. This type of model helps account for the unique distributional features of aggregate health expenditures. Health expenditures for a population typically will have a very large number of observations with zero expenditure, and a large positive skew for the few individuals with very high expenditures. Our two-part model first predicts the probability of any health expenditure using a logistic regression, then fits a Generalized Linear Model (GLM) with a log-link to the observations with a positive expenditure, which has been shown to perform well for expenditure data in several previous econometric studies.[24,25] The GLM model with a log-link uses a quasi-likelihood approach to directly model the following expectation:

$$E(y|X) = exp(X\beta)$$

where y is as above and X is a matrix of covariates, including the 'treatment' variable. We determined the variance structure for the GLM using the method in Bunting and Zaslavsky.[25] This method examines the model predictions versus observed values for deciles of expenditure level. From this, we determined the power function with variance proportional to the mean best fit of the MEPS data for both drug and non-drug expenditures. After fitting both parts of the model, predictions for individual i can be obtained using a simple rule of probability:

$$E(y_i|X_i) = Pr(y_i > 0|X_i)E(y_i|X_i, y_i > 0)$$

where the first term is the estimated probability from the logistic regression and the second is the estimated expenditure from the GLM, both conditional on the matrix of covariates X. For the expected yearly drug cost estimates, only the second part of the model was used because only 13 individuals had no drug expenditure.[‡]

We employed a range of demographic and clinical characteristics to model both parts. The variables and their specification are outlined in Table 6. Previous research on hypertension expenditure has suggested that only using medical events coded as hypertension can underestimate expenditure.[26] Thus, we examined the costs specifically linked to hypertension as our first outcome. Our second outcome adds the cost of acute myocardial infarction (ICD-9 410), a major complication of hypertension. Finally, we model the total expenditure associated with the more comprehensive definition of hypertension-related complications created by Hodgson et al.[26§]

Our primary quantity of interest was the Average Treatment Effect (ATE) of taking a 'new' therapy on both total yearly drug costs and total yearly non-drug spending related to

[‡]The GLM can handle zero values, but the two-part model may perform better when there is a large number of zeros, as there is for total non-drug expenditure. We speculate that the zero value observations may have resulted from free physician samples.

[§]This list includes ischemic heart disease, other forms of heart disease and heart failure, stroke, atherosclerosis, and certain vascular disease and includes ICD-9 codes 410-414, 424-438, 440-441.

hypertension. To calculate an average treatment effect, we simulated 25,000 draws of the model parameters from the multivariate normal approximation to the likelihood function and determined expectations for each individual given their covariates under their 'treated' and 'untreated' counterfactuals. The mean difference of these values estimates the expected effect of taking a newer drug on both expenditure outcomes. We also estimated the Average Treatment effects for the Treated (ATT) using the same method for only the treated observations.[18] Confidence bounds for both estimates were obtained by calculating the 2.5 and 97.5th quantiles of these draws.

The simulation results for the effect of receiving a new drug on both total drug costs and total non-drug expenditures are shown in table 7. The results on lines 1 and 2 show that receiving one or more 'newer' drugs during the year is associated with an increased yearly drug expenditure of $179, with a 95% confidence range of $47-$355. The result for the treated observations is very similar. The estimated offset effect of receiving a new drug on total yearly non-drug expenditures is estimated to be a $8 savings, with a 95% confidence interval which ranges from a $92 savings to an extra cost of $44. Our results for the treated observations are again very similar. Similarly inconclusive results are seen for the other expenditure categories, with including myocardial infarction leading to an estimated increase in expenditure and a more comprehensive list leading to a decrease, both statistically indistinguishable from zero. Thus, we find no evidence that taking newer hypertension therapies offsets associated non-drug hypertension expenses.

Discussion

While we believe our alternative approach represents a methodological improvement, our findings should also be interpreted with a few caveats. First, a key limitation of using the 1996 MEPS dataset to investigate potential offsets is the limited time frame over which data

are available for any individual. Hence, our estimate of the offset effect would underestimate the effect if offsets accrued over subsequent time periods after the drug is taken. To identify any real offset effect, one would want to look at this over the entire duration of a disease condition. Longitudinal panel data would be better to investigate this effect. Duggan (2005) uses such a long-term approach in his estimates for antipsychotics.[2] Moreover, the size of the dataset may be a limitation in discerning the impact of relatively rare high-cost events such as heart attacks.

Second, a key assumption of the propensity score method is that after matching, treatment assignment is unconfounded by any other factors. As we only estimate the score from several observable patient characteristics, it is possible that selection on unobservable characteristics is occurring and biasing our results. Ideally, one would want to find a natural experiment or exogenous source of variation in the age of the drug used. The introduction of new drugs provides just such an opportunity to identify such quasi-experiments. However, one drawback to using observational data is the fact that a particularly effective drug will likely capture much of a market in short order, making matching on the propensity score both more difficult and possibly more confounded by unobserved characteristics (such as side-effect reactions). Hence, the method we describe above would only be reliable when there is variation in the drugs that are being prescribed that is not correlated with patient costs.

Third, our analysis considers only the pure offset of prescription expenditure with other non-drug related health costs. We recognize that this is a very narrow view of the potential benefits of a new drug; other non-pecuniary aspects should also be considered in evaluating the true benefit of a newer medicine. If a newer drug improves health outcomes or other economic outcomes when contrasted with older drugs, any additional costs of therapy may well be justified. For example, newer therapies may have a safer side effect profile or come in more convenient dosages. A net cost savings on health care expenditure should not be

the only test by which new therapies are judged, but would add to arguments for drug development.

The implications of our analysis are twofold. First, our re-evaluation of Lichtenberg (2001) suggests its findings are not robust to modifications in the model specification or dataset and thus should be interpreted very cautiously.[4] Secondly, we developed a more robust method for analyzing offset effects using propensity score matching and a two-part expenditure model. As we believe that drug offsets are likely to be condition-specific if they exist, we used this approach for a single condition, essential hypertension. Our findings provide strong evidence that the use of newer drugs for hypertension is associated with higher drug costs. However, we find no evidence that these higher costs were offset by savings in non-drug health expenditures over a one year period. Whether newer drugs for hypertension have a long-term effect on non-drug health care costs remains a topic for future research and debate.

Reference List

[1] Cynthia Smith, Cathy Cowan, Stephen Heffler, and Aaron Catlin. National health spending in 2004: Recent slowdown led by prescription drug spending. *Health Affairs*, 25(1):186–195, 2006.

[2] M. Duggan. Do new prescription drugs pay for themselves? the case of second-generation antipsychotics. *Journal of Health Economics*, 24(1):1–31, 2005.

[3] P. J. Neumann, E. A. Sandberg, C. M. Bell, P. W. Stone, and R. H. Chapman. Are pharmaceuticals cost-effective? a review of the evidence. *Health Affairs*, 19(2):92–109, 2000.

[4] F. R. Lichtenberg. Are the benefits of newer drugs worth their cost? evidence from the 1996 meps. *Health Affairs*, 20(5):241–51, 2001.

[5] F. R. Lichtenberg. Benefits and costs of newer drugs: An update. NBER Working Paper 8996, 2002.

[6] Yuting Zhang and Stephen B. Soumerai. Do newer prescription drugs pay for themselves? a reassessment of the evidence (manuscript under review). 2006.

[7] G. Edward Miller, John F. Moeller, and Randall S. Stafford. New cardiovascular drugs: Patterns of use and association with non-drug health expenditures. *Inquiry*, 42(4): 397412, 2006.

[8] David M. Cutler, Mark McClellan, Joseph P. Newhouse, and Dahlia Remler. Are medical prices declining? evidence from heart attack treatments. *Quarterly Journal of Economics*, 113(4):991–1024, 1998.

[9] Ernst R. Berndt, Anupa Bir, Susan H. Busch, Richard G. Frank, and Sharon-Lise T. Normand. The medical treatment of depression, 1991-1996: productive inefficiency, expected outcome variations, and price indexes. *Journal of Health Economics*, 21(3): 373–396, 2002.

[10] Agency for Healthcare Research and Quality. Medical expenditure panel survey (1996), February 22, 2006 .

[11] The Food and Drug Administration. Drugs@fda database, May 5, 2006 .

[12] Mosby's. Drug consult. 2006.

[13] Henry Grabowski and John Vernon. Effective patent life in pharmaceuticals. *International Journal of Technology Management*, 19(1/2):98–120, 2000.

[14] Naihua Duan, Willard G. Manning, Carl N. Morris, and Joseph P. Newhouse. A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics*, 1(2):115–126, 1983.

[15] A. Basu, W. G. Manning, and J. Mullahy. Comparing alternative models: log vs cox proportional hazard? *Health Economics*, 13(8):749–65, 2004.

[16] J. M. Flack and K. Bledsoe. Combination antihypertensive drug therapy: a therapeutic option long overdue. *American Family Physician*, 61(10):2974, 2977, 2000.

[17] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[18] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. 2005.

[19] X. Gu and P. R. Rosenbaum. Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2 (4):405–420, 1993.

[20] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

[21] Ben B. Hansen and Stephanie Olsen Klopfer. Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, 15, 2006.

[22] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matchit: Nonparametric preprocessing for parametric causal inference, 2006.

[23] Donald B. Rubin. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3-4):169, 2001.

[24] W. G. Manning and J. Mullahy. Estimating log models: to transform or not to transform? *Journal of Health Economics*, 20(4):461–94, 2001.

[25] M. B. Buntin and A. M. Zaslavsky. Too much ado about two-part models and transformation? comparing methods of modeling medicare expenditures. *Journal of Health Economics*, 23(3):525–542, 2004.

[26] T A Hodgson and L Cai. Medical care expenditures for hypertension, its complications, and its comorbidities. *Medical Care*, 39(6):599–615, 2001.

Model	Dataset	Estimate	SE	t	p
Lichtenberg (2001)	Original	-18.00	0.184	97.74	<.001
Replication	Original	-17.92	0.186	-96.28	<.001
	Updated	-18.93	0.202	-93.89	<.001

Table 1: Coefficients for ln(Drug Age) using ordinary least squares regression on Prescription Expenditure for 3 models: the Original Paper, Our Replication and Replication using the updated MEPS dataset.

Model	Outcome	Dataset	Estimate	SE	t	p
Lichtenberg (2001)	Total Costs	Original	71.09	15.16	4.69	<.001
Replication	Total Costs	Original	67.81	23.37	2.90	0.004
		Updated	0.21	25.36	0.01	0.993
	ln(totalcost+1)	Original	-0.01	0.009	-0.68	0.496
		Updated	-0.06	0.010	-5.44	<.001

Table 2: Coefficients for ln(Drug Age) in OLS on Total Non-Drug Health Care Expenditure for 5 models: the original paper, replication, replication using the updated MEPS data and using ln(totalcost+1) as the outcome with both the original and updated data

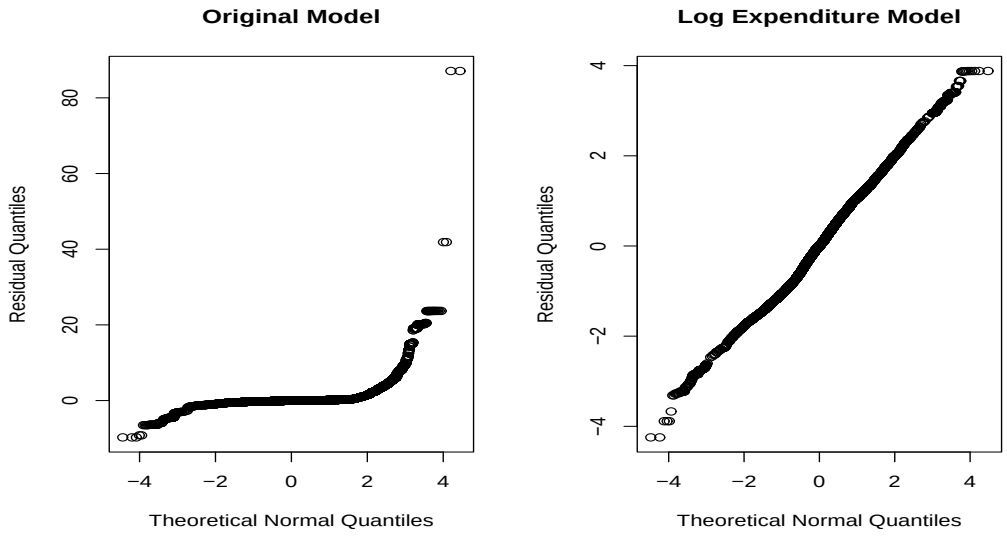

Figure 1: Model residuals for two OLS models using the original dataset-the untransformed outcome variable (left) and log-transformed variable (right)

Drug Name	**Category**	**New?**	**Freq**	**%**
Lisinopril	Ace Inhibitor	Yes	1131	6.59%
Nifedipine	Calcium Channel Blocker	No	1085	6.32
Verapamil Hydrochloride	Calcium Channel Blocker	No	1035	6.03
Atenolol	Beta Blocker	No	894	5.21
Enalapril Maleate	Ace Inhibitor	No	862	5.02
Amlodipine Besylate	Calcium Channel Blocker	Yes	682	3.97
Hydrochlorothiazide (HCTZ)	Diuretic	No	634	3.69
Diltiazem Hydrochloride	Calcium Channel Blocker	No	601	3.50
Benazepril Hydrochloride	Ace Inhibitor	Yes	494	2.88
Metoprolol Fumarate	Beta Blocker	No	413	2.40
Triamterene	Diuretic	No	379	2.21
Potassium Chloride	Supplement	No	367	2.14
Furosemide	Diuretic	No	339	1.97
Captopril	Ace Inhibitor	No	334	1.94
Quinapril Hydrochloride	Ace Inhibitor	Yes	322	1.87
Clonidine	Alpha Blocker	No	300	1.75
Indapamide	Diuretic	No	272	1.58
Propranolol Hydrochloride	Beta Blocker	No	267	1.55
Terazosin Hydrochloride	Alpha Blocker	Yes	259	1.51
Metoprolol Succinate	Beta Blocker	No	202	1.18
Ramipril	Ace Inhibitor	Yes	194	1.13
Methyldopa	Alpha Blocker	No	160	0.93
Lisinopril & HCTZ	Ace Inhibitor & Diuretic	Yes	150	0.87
Fosinopril Sodium	Ace Inhibitor	Yes	144	0.84
Spironolactone	Diuretic	No	140	0.82
Prazosin Hydrochloride	Alpha Blocker	No	128	0.75
Labetalol Hydrochloride	Alpha/Beta Blocker	No	123	0.72
Felodipine	Calcium Channel Blocker	Yes	117	0.68
Losartan Potassium	Angiotensin Receptor Blocker	Yes	112	0.65
Amiloride Hydrochloride	Diuretic	No	94	0.55

Table 3: Top 30 Prescriptions for Hypertension in the MEPS dataset by Name, Category, 'New' Classification, Frequency and Percentage of Scripts. The table demonstrates large variability in the types and ages of hypertension drugs prescribed to MEPS respondents.

Variable	Form	Variable	Form
Age	Quintile	Income	Linear
Live in MSA	Binary	US Census Region	4 Categories
Poverty Category	Linear	Have Usual Care Source	Binary
Health Rating	Linear	Ever Uninsured	Binary
Ever Privately Insured	Binary	Ever Medicaid Eligible	Binary
Ever Medicare Eligible	Binary	Hispanic Ethnicity	Binary
Black Ethnicity	Binary	Male	Binary
Married	Binary	High School Graduate	Binary
College Degree	Binary	Currently Employed	Binary
Number of Conditions	Count	Number of Conditions*Age	Interaction
Age*Health Rating	Interaction	Number of Conditions*Health Rating	Interaction
Age*Income	Interaction	Number of Conditions*Income	Interaction
Age*Male	Interaction	Number of Conditions*Male	Interaction

Table 4: Variables and their functional form used in calculation of the propensity score

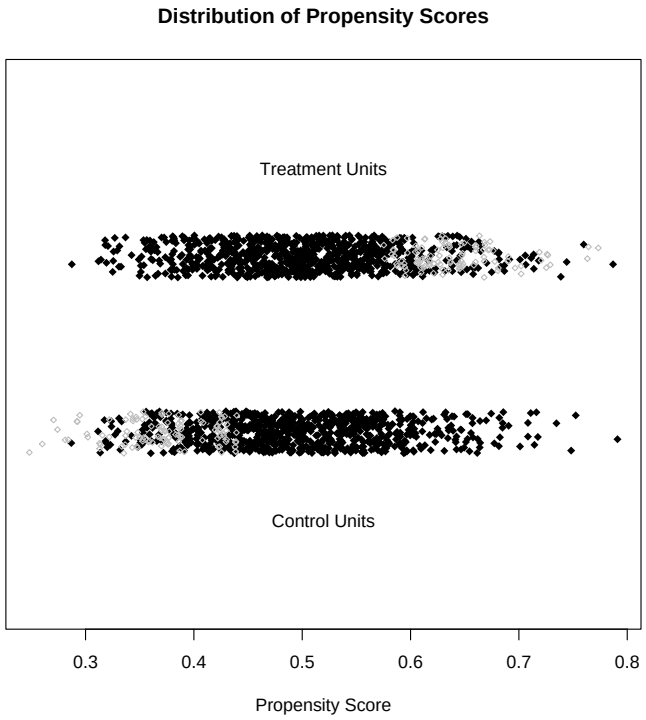

Figure 2: Distribution of the estimated propensity for taking new therapies in both the new and old drug takers. Black dots indicate matched observations, grey are discarded. The plot demonstrates large overlap in the distributions between the two groups.

Variable	**Mean Newer**	**Mean Older**	**Difference**
Propensity Score	0.50	0.50	0.00
Age Quintile	2.97	2.97	0.00
Income Category	1.53	1.54	−0.01
Live in MSA	0.74	0.73	0.01
NorthEast Region	0.20	0.19	0.01
MidWest Region	0.24	0.24	0.00
South Region	0.38	0.38	0.00
West Region	0.18	0.19	−0.01
Poverty Category	3.50	3.51	−0.01
Usual Source of Care	0.96	0.96	0.00
Health Rating	2.95	2.96	−0.01
Ever Uninsured	0.05	0.05	0.00
Ever have Private Insurance	0.68	0.69	−0.01
Ever Medicaid Eligible	0.02	0.02	0.00
Ever Medicare Eligible	0.16	0.15	0.01
Hispanic	0.51	0.51	0.00
Black	0.10	0.09	0.01
Other Race	0.19	0.18	0.01
Male	0.40	0.40	0.00
Married	0.59	0.60	−0.01
High School Graduate	0.49	0.49	0.00
College Degree	0.19	0.19	0.00
Employed	0.40	0.40	0.00
Number of Reported Conditions	6.58	6.58	0.00
Sample Size	815	815	-

Table 5: Matched sample variable means, by receipt of a new drug. The similar variable means above suggest good balance has been achieved in the matched sample for these observable covariates.

Variable	Form	Variable	Form
New Drug User	Indicator	Age	Quintiles
Income	Linear	Live in MSA	Binary
Health Rating	Linear	Ever Uninsured	Binary
Ever Privately Insured	Binary	Ever Medicaid Eligible	Binary
Ever Medicare Eligible	Binary	Gender	Binary
Hispanic Ethnicity	Binary	Black Ethnicity	Binary
High School Graduate	Binary	College Degree	Binary
Number of Conditions	Count	Age*Gender	Interaction

Table 6: Variables and their functional form used in the two-part model to estimate total prescription expenditure and non-drug health care expenditure

Variable	Quantity	Mean	2.5%	97.5%
Hypertension Drug Expenditure	ATE	179.04	46.94	354.88
	ATT	179.24	47.10	354.00
Hypertension Non-Drug Expenditure	ATE	-7.77	-92.44	44.21
	ATT	-7.75	-91.16	43.62
+ Acute Myocardial Infarction	ATE	12.24	-132.98	211.71
	ATT	13.50	-134.50	214.81
+ Hypertension Complications	ATE	-51.24	-527.84	206.67
	ATT	-51.49	-516.78	203.83

Table 7: Average Treatment Effect (ATE) and Average Treatment Effect for the Treated (ATT) for the matched sample of taking a newer medication on total drug cost and total expenditure on other health care for essential hypertension. The results suggest an increase in total yearly drug costs for the recipients of newer drugs for both groups, but do not demonstrate a substantial change in total non-drug expenditure.