Local control for identifying subgroups of interest in observational research: persistence of treatment for major depressive disorder

DOUGLAS E. FARIES,¹ YI CHEN,¹ ILYA LIPKOVICH,¹ ANTHONY ZAGAR,¹ XIANCHEN LIU^{1,2}
& ROBERT L. OBENCHAIN³

¹ Eli Lilly and Company, Indianapolis, IN, USA

² Indiana University Department of Psychiatry, Indianapolis, IN, USA

³ Risk Benefit Statistics, LLC, Carmel, IN, USA

Key words

cluster analysis, patient heterogeneity, classification trees, counterfactual

Correspondence

Douglas E. Faries, Research Fellow, Eli Lilly & Company, Lilly Corporate Center, Indianapolis, IN 46285, USA. Telephone (+1) 317-276-9801
Email: d.faries@lilly.com

Abstract

Caregivers are regularly faced with decisions between competing treatments. Large observational health care databases provide a golden opportunity for research on heterogeneity in patient response to guide caregiver decisions, due to their sample size, diverse populations, and real-world setting. Local control is a promising tool for using observational data to detect patient subgroups with differential response on one treatment relative to another. While standard data mining approaches find subgroups with optimal responses for a particular population, detecting subgroups that reveal treatment differences while also adjusting for confounding in observational data is challenging. Local control utilizes unsupervised clustering to form non-parametric patient-level counterfactual treatment differences and displays them as an observed distribution of effect-size estimates. Classification and regression trees (CART) then find the factors that drive the greatest outcome differentiation between treatments. In this manuscript, we demonstrate the use of this two-step strategy using local control plus CART to identify depression patients most (least) likely to benefit from treatment with duloxetine relative to extended-release venlafaxine. Prior medication costs and age were found to be factors most associated with differential outcome, with prior medication costs remaining as an important factor after sensitivity analyses using a second dataset. *Copyright* © 2013 John Wiley & Sons, Ltd.

Introduction

The use of large medical databases, such as health care claims databases and electronic medical records, has been growing in importance in medical research. Such

observational databases are potentially rich resources for data mining that address a variety of medical questions (Motheral *et al.*, 2003; Crystal *et al.*, 2007). They provide data on large numbers of patients representative of

potentially diverse usual care settings. Due to the large sample sizes and a heterogeneous population of patients, such databases are valuable for studying key patient subgroups of interest (Crystal *et al.*, 2007). However, the ability to draw comparative inferences from observational research is challenged by selection bias and potential for unmeasured confounding. Patients are not randomized to treatments, and thus comparisons between treatment groups are subject to bias due to many patient and physician factors that influence treatment selection in usual care practice. Statistical adjustment for measured confounders is possible, using techniques such as propensity scoring. However, because one cannot adjust for unmeasured confounders, observational research occupies a lower level, relative to randomized clinical trials, in the hierarchy of evidence for causal inference (Benson and Hartz, 2000; Concato *et al.*, 2000; Vandenbroucke, 2008). The challenges of adjusting observational data for bias and confounding have led to a long history of methodological research (Rosenbaum and Rubin, 1983; D'Agostino, 1998; Rosenbaum, 2010) – and recently to guidances on comparative effectiveness research (Des Jarlais *et al.*, 2004; Vandenbroucke *et al.*, 2007; Berger *et al.*, 2009; Cox *et al.*, 2009; Johnson *et al.*, 2009).

While clinical trial research has focused on establishing the efficacy of an intervention in narrowly defined patient populations, recently the importance of patient heterogeneity has been more vigorously investigated (Davidoff, 2009; Ruberg *et al.*, 2010). Average superiority of drug A over drug B does not mean that drug A is superior to drug B for each patient; the analyses example presented later in this text clearly illustrates this. The outcomes for some individual patients may deviate considerably from the estimate for the population mean. Multiple types of factors, ranging from genetic, phenotypic, pharmacokinetic, environmental to social, may influence individual outcomes (Ruberg *et al.*, 2010). Thus, the problem of predicting individual response to multiple potential interventions that could be administered is a complex and challenging task. However, this is a challenge that patients and health care providers are demanding of researchers. Therefore, research investigating patient heterogeneity is of growing importance in the field of medicine.

One class of data mining approaches for subgroup identification is based on classification and regression tree (CART) methodology (Breiman *et al.*, 1984). A tree is formed by recursively splitting patient subgroups (starting with the entire population) by choosing the best cutoff point(s) on a variable selected from a pre-specified list of patient characteristics (X-covariates). The splitting criterion is defined so as to make the resulting child

subgroups as homogeneous as possible with respect to the outcome variable.

In heterogeneity research to identify the best of several potential treatment options for patient, it is desirable to identify subgroups of subjects with substantially disparate treatment effects. That is, not to just identify subgroups that optimize outcome for a single cohort, but subgroups of individuals who are more likely to have better (worse) outcomes on a given intervention relative to competing interventions. Thus, this question requires more than the standard applications of method such as CART, which optimize for a given outcome measure in a given population. That is, they find subgroups with the best response on treatment A, but such individuals may also respond just as well to treatment B. The interest here is identification of subgroups of subjects who would respond better to treatment A relative to how they would respond on treatment B.

One approach for identifying subgroups with desired treatment effects is based upon potential (or counterfactual) outcomes concepts, much like imputation of missing values. A variable is defined as a potential or counterfactual outcome if it describes the outcome a patient would have experienced given a potential exposure that they may not have actually experienced (Hernan, 2004). For instance, we may observe a patient's outcome after initiating treatment A, but their unknown (or missing) outcome had they initiated treatment B would be considered a counterfactual outcome. To appropriately estimate counterfactual outcomes in observational research, one must address confounding bias – that is, patients on treatment A really need to be compared with highly similar patients on treatment B. Matching on estimated propensity scoring is now a widely practiced approach to adjust for measured confounders in observational research (Rosenbaum and Rubin, 1983). However, here we focus on the use of cluster analysis of observed pre-treatment covariates to assure that only the most relevant “local” (within cluster) treatment comparisons are being made.

In the first step, potential outcomes for all available treatments are evaluated for each subject using appropriate data mining tools to form subgroups of highly similar patients (e.g. random forests Foster *et al.*, 2011 or local control Obenchain, 2010). For instance, local control utilizes cluster analysis methodology (unsupervised) to form mutually exclusive and exhaustive patient subgroups in X-space. Cluster membership is then a balancing score (Rosenbaum and Rubin, 1983) that can be observed essentially non-parametrically, avoiding reliance on a parametric model to estimate the true propensity. Treatment comparisons made within clusters of well-matched patients are then conditionally unbiased and thus can

produce appropriate counterfactual outcome estimates even with observational data. As step 2, patient-level estimated (counterfactual) treatment differences can then be “evaluated” using a model (e.g. CART-based) to determine which patient baseline characteristics determine subgroups with the largest (smallest) expected treatment differences.

In this manuscript, we demonstrate use of this new two-step strategy using local control plus CART to study patient heterogeneity in persistence to dual serotonin and norepinephrine reuptake inhibitors (SNRIs) for major depressive disorder (MDD) in primary care. Local control is used in step 1 to compute counterfactual treatment differences for each patient based on the heterogeneity in outcomes observed in many small clusters of well matched patients, while CART is subsequently used to identify fewer and larger subgroups that predict the counterfactual treatment differences observed in step 1. In particular, we follow recently completed research (Liu *et al.*, 2011) on group comparisons between medication persistence on duloxetine (DLX) and venlafaxine XR (VLX) to study factors identifying patients who may benefit most (least) from using one therapy relative to the other. Liu *et al.* (2011) compared the average adherence and persistence of four different treatments (duloxetine, venlafaxine, escitalopram, and generic selective serotonin reuptake inhibitors [SSRIs]) as well as identifying predictors of adherence. However, they did not investigate the important question of identifying factors or subgroups which predicted a treatment difference in adherence. For instance, their research suggested that patients with chronic pain conditions or alcohol and drug dependence had lower medication adherence, but did not investigate whether patients in this (or any) subgroup had consistently lower adherence on all medication groups (and thus the results on the overall means would apply) or whether they would benefit in particular from one treatment over another (the means do not apply to everyone).

Methods

Data source and sample selection

This retrospective analysis of claims data employed medical, pharmacy, and enrollment information from the Thomson Reuters MarketScan® Commercial Claims and Encounters Databases (Ann Arbor, MI, USA) which includes de-identified administrative claims of employees, spouses, and dependents with employer-sponsored commercial insurance. MarketScan is publicly available as a fee-for-service database and has been previously used for a number of retrospective and prospective analysis

projects. See Liu *et al.* (2011) for additional details regarding the database and construction of the analytical file used in this analysis. Patients were included in the study if they had a first prescription fill for DLX or VLX in 2006, had no prescriptions of the same study medication in the three months prior to the index date, nor any prescriptions prior to the three-month window for which the days supply would suggest availability of remaining medications within the three-month window, and had one or more inpatient or outpatient claims associated with diagnoses of MDD (ICD-9-CM: 296.2 and 296.3) either one year prior to or one month after the study medication was initiated. Patients had to be 18 to 64 years of age, be commercially insured, and have continuous enrollment over the prior 12-month and post 12-month periods. Patients who initiated on both DLX and VLX during 2006 at different times were classified in the cohort they initiated first.

Outcome and pre-treatment measures

The outcome measure for this research was medication persistence, defined as the number of days from the index date to the earliest of the three following dates: the ending date of the last prescription, the date of the first gap of more than 15 days between prescriptions, or the end of the 12-month study period (Cantrell *et al.*, 2006; Cramer *et al.*, 2008). Recent work has shown poor medication persistence is associated with increased health care resource utilization for patients with major depression (Cui *et al.*, 2012). Pre-treatment variables included as potential variables in the clustering and classification tree analyses later were demographic characteristics (age, gender, geographic region of residence), health plan type, comorbidities (chronic pain conditions [headache, rheumatoid arthritis, osteoarthritis, lower back pain, fibromyalgia, neuropathic pain], sleep disorders, psychiatric disorders [dysthymia, anxiety, alcohol or drug dependence, schizophrenia, bipolar, organic psychosis], diabetes, hypertensive disease, and the charlson comorbidity index), medication use categories (antidepressants, sedatives, psychotropics, mood stabilizers, psychostimulants, opioids, other pain medications), health care resource utilization (number of hospitalizations, office visits, ER visits, psychiatric services, medication costs, medical services costs, and total costs). Comorbidities were based on the one year period prior to initiation with DLX or VLX. Prior use (one year period) of medications included specified therapeutic classes as well as selected individual drugs known to be used in treating depression, psychosis, bipolar disorders, anxiety, sleep disorders, or chronic

pain or known to be associated with initiation of DLX (Kroenke *et al.*, 2009; Liu *et al.*, 2011; Stafford *et al.*, 2001; Ye *et al.* 2011).

Statistical methods

Patient demographics and baseline patient characteristics for the DLX and VLX cohorts were summarized and between cohort comparisons were made using *t* and chi-square tests. Local control methodology was used to assess patient heterogeneity in regards to treatment differences in medication persistence (Obenchain, 2010, 2011). Unsupervised clustering on all pre-treatment X-variables was utilized to assign each of the 13,673 patients to one of 1500 clusters, so the average cluster size was to be approximately 10. Ward's minimum variance method, which at each step minimizes the within-cluster sum of squares overall all partitions that can be constructed by merging any two clusters, was used to create the clusters via SAS (PROC CLUSTER). "Informative" clusters are defined as clusters containing at least one DLX-treated patient and at least one VLX treated patient. Local treatment differences (LTDs) were computed as the difference (DLX minus VLX) in therapy duration mean values within each informative cluster. Patients in non-informative clusters, defined as clusters containing patients from only one treatment group, were discarded from the analysis, analogous to patients for whom no match could be found in a propensity score matching analysis. The observed LTD within each cluster was then used as the counterfactual treatment difference estimate for every patient within that cluster.

CART methodology was used to construct a regression tree for predicting the patient-level counterfactual treatment difference estimates using all pretreatment X-variables as potential splitting variables using PROC Arbor within SAS Enterprise Miner. A minimum subgroup size of 5% of the overall sample was utilized to produce more meaningful subgroups, guard against overfitting, and reduce the need for pruning. The tree selection process was iterative and focused on selecting a tree that was interpretable from a subject matter point of view. Specifically, while multiple nodes may produce greater differential between cohorts, interpretation of nodes nested within multiple higher level nodes can become complex and the tree was pruned to a maximum of two splits. The relative importance statistic (Breiman *et al.*, 1984) was used to confirm the value of the variables in the final tree and identify important variables that may be masked by another correlated variable and can fail to appear in a final pruned tree. The variable importance statistics for variable

X is computed in PROC ARBOR following the original L. Breiman's method (Breiman *et al.*, 1984) by adding up reductions in impurity across all internal nodes of the tree, weighted by a degree of participation of X in a given split. Note that the reduction in impurity is based on the reduction in the error sum of squares when splitting a node into two child nodes. The weight is set to one if X is the primary splitter in that node, it is equal to the agreement r ($0 \leq r \leq 1$) between the split by the primary splitter and split by X that mimics the primary split if X is one of the surrogate splitters, and weight equals zero otherwise.

Another important issue is evaluating the statistical significance of the CART-identified subgroups compared to the mean level of the outcome. That is, to evaluate whether such or more extreme subgroups could be identified by chance as an artifact of searching and selecting through a large number of potential predictors, even if none of them had any association with the outcome. To address this issue, "null" data were constructed by randomly permuting the outcome and applying the same CART procedure (with the same splitting and pruning parameters) as was applied to the original data. One thousand replications of this process were conducted, and the outcomes in the tree node associated with the largest treatment difference were recorded. The proportion of times a tree fitted to such permuted data would have resulted in a subgroup with the mean (standardized) treatment difference as large as the one observed was computed as the permutation-based *p*-value.

To validate the findings, the key factors identified from the analysis were assessed using a 2007 sample of the same database (MarketScan; patients initiating in 2007 using the same criteria). Specifically, treatment differences based on these factors were assessed using the 2006 tree structure and the new data to see whether findings were directionally similar.

Results

A total of 13,673 patients with continuous enrollment in the health plan for 12 months prior to and post index date were included in the study. Each patient was characterized as initiating the post index study period with either DLX ($N=7567$) or VLX ($N=6106$). Summary statistics for demographic characteristics and other pre-treatment variables of these patients are reported in Table 1. Due to the large sample size and non-randomized nature of the data, many statistical differences were expected and were observed. In general, patients initiating DLX tended to have a more costly treatment history with more comorbidities and medication use. In the 12 months

Table 1. Pretreatment variables of patients initiated on duloxetine and venlafaxine XR

	Duloxetine (<i>n</i> = 7567)	Venlafaxine XR (<i>n</i> = 6106)	<i>p</i> -Value
<i>Demographic characteristics</i>			
Female gender (%)	75.1	71.5	<0.0001
Age group, year (%)			
18–25	4.8	9.1	<0.0001
26–35	9.3	12.5	
36–45	23.7	23.8	
46–55	38.4	33.1	
56–64	23.9	21.4	
Mean, years (SD)	47.2 (10.6)	45.1 (11.9)	<0.0001
<i>Health plan type (%)</i>			
Comprehensive	11.8	10.1	<0.0001
HMO	18.6	21.3	
POS	11.5	12.9	
PPO	56.0	53.3	
Other	2.1	2.4	
<i>Region of residence (%)</i>			
Northeast	8.8	10.1	<0.0001
North central	30.7	30.7	
South	40.8	35.9	
West	19.4	22.8	
Unknown	0.3	0.5	
<i>Comorbid conditions</i>			
Diabetes mellitus (%)			
Yes	11.5	10.0	0.003
Hypertensive diseases (%)			
Yes	25.4	21.5	<0.0001
Sleep disorders			
Yes	12.4	9.6	<0.0001
Number of pain conditions			
Mean (SD)	0.8 (0.92)	0.5 (0.79)	<0.0001
Number of psychiatric disorders			
Mean (SD)	0.5 (0.73)	0.5 (0.73)	0.4225
<i>Prior use of medications</i>			
Number of antidepressants			
Mean (SD)	1.4 (0.96)	1.3 (0.99)	<0.0001
Number of sedatives			
Mean (SD)	1.3 (1.26)	1.0 (1.18)	<0.0001
Number of psychotropics			
Mean (SD)	0.4 (0.77)	0.3 (0.68)	<0.0001
Number of opioids			
Mean (SD)	1.1 (1.20)	0.8 (1.05)	<0.0001
Number of other pain medications			
Mean (SD)	1.3 (1.48)	0.9 (1.26)	<0.0001
<i>Baseline health care resource utilization</i>			
Number of hospitalizations			
Mean (SD)	0.3 (0.94)	0.3 (0.85)	0.0006
Number of ER services			

(Continues)

Table 1. (Continued)

	Duloxetine (<i>n</i> = 7567)	Venlafaxine XR (<i>n</i> = 6106)	<i>p</i> -Value
Mean (SD)	0.9 (2.24)	0.8 (1.95)	<0.0001
Number of office visits			
Mean (SD)	9.5 (7.66)	7.9 (7.08)	<0.0001
Number of psychiatric services			
Mean (SD)	10.7 (14.62)	9.6 (14.17)	<0.0001
Baseline drug cost			
Mean (SD)	4250.0 (5010.86)	2912.2 (4991.06)	<0.0001
Baseline medical cost			
Mean (SD)	12,110.0 (26,686.04)	9824.2 (26,302.31)	<0.0001
Baseline total cost			
Mean (SD)	16,360.0 (28,078.01)	12,736.4 (27,588.66)	<0.0001

Note: ER = emergency room; HMO = health maintenance organization; *n* = number of patients; POS = point of service; PPO = preferred provider organization; SD = standard deviation.

Baseline healthcare resources, prior medications, and comorbid conditions were evaluated over the 12 months prior to initiation.

The *p*-values based on chi-square tests for binary variables and *t*-tests for continuous variables.

after medication initiation, average length of therapy (persistence) was statistically significantly longer with DLX (mean = 158.5 days, standard deviation [SD] = 133.9, median = 95.0) than VLX (mean = 149.6 days, SD = 129.9, median = 90.0).

Using the pre-treatment variables in Table 1, 1500 clusters were formed in X-space (six of the 13,673 patients were excluded from cluster analyses because of missing X-values). Two hundred and eighty clusters containing 673 patients (540 DLX patients and 133 VLX patients) were found to be non-informative and were discarded from the analysis. As a whole, the excluded group tended to have more comorbidities with greater resource use and costs than the population remaining in the analysis. These discarded patients had a mean persistence of 153.6 days (153.1 days for 540 DLX patients and 155.8 days for 133 VLX patients). The distribution of the LTDs from 12,994 patients within the 1220 informative clusters is shown in Figure 1. Each patient within an informative cluster is assigned the LTD value for his/her cluster. It is apparent that there are a substantial number of patients with counterfactual treatment difference estimates favoring each treatment. In fact, roughly 46% of the observed LTDs are negative, favoring VLX, while 53% are positive, favoring DLX.

Figure 2 displays the pruned regression tree for predicting counterfactual treatment differences. The initial node represents the full sample (among informative

Figure 1. Distribution of local treatment differences (days of persistence). The local treatment differences (in days) are the observed mean cohort differences within informative clusters. Each informative cluster (*N* = 1220) produces a single value in the histogram, regardless of the number of patients within the cluster. Positive values indicate longer persistence with duloxetine relative to venlafaxine.

clusters) of 12,994 patients, and the overall mean treatment difference in persistence of +7.65 days (positive values indicate longer duration for DLX). The follow-up nodes indicate that the estimated treatment difference is over two weeks favoring DLX for patients with higher baseline drug costs (indicating greater intensity/duration

Figure 2. Classification and regression tree analysis of treatment differences in days of medication persistence. Pruned CART tree: outcome measure is the counterfactual treatment difference in days of treatment (persistence) for each patient. The “Average” represents the mean of the counterfactual treatment differences for all patients in the node. A positive value indicates longer persistence with duloxetine relative to venlafaxine. BDRUGCOST = Prior (to initiation of duloxetine or venlafaxine) six-month medication-related healthcare costs.

of prior treatment and/or treatment for other comorbidities) or for older aged (>56 years) patients. Conversely, treatment differences were small for younger patients without high prior costs. The permutation test indicated the maximum differences were unlikely under the null scenario ($p < 0.001$). The variable importance statistic indicated that baseline total costs, baseline drug costs, and age were the key splitting variables. Due to the high correlation between overall cost and drug costs variables ($r = +0.66$), only one cost appears in the final model.

The associations between treatment differences and persistence for each of the two factors (pre-treatment drug costs and age) were assessed using the follow-up data as a sensitivity analysis. Using the same inclusion and exclusion criteria, except having a first prescription fill for DLX or VLX in 2007, a sample of 13,655 patients was obtained. Using the same splits as in the regression tree based on the 2006 data, pre-treatment drug costs were found to have a directionally similar result. That is, the mean treatment difference in the more costly patients showed greater differences favoring DLX (mean = 13.6 days), while in the less costly patients the opposite was true (mean = -4.1 days). Baseline costs, both medical and drug, also had among the higher importance statistics. However, no association between age and treatment

difference in persistence was observed within the 2007 data, and lower importance statistics confirmed the lack of impact.

Discussion

This research utilized a health care claims database and new unsupervised clustering methodology to examine patient heterogeneity for persistence among patients with MDD in usual care. Regression tree methodology then suggested that two factors, prior medication costs and age, were the most important factors predicting individual patient treatment differences between DLX and VLX. Total prior year medication cost is likely a marker for the treatment complexity of the patient – indicating treatment for multiple comorbidities and/or longer duration and more intensive MDD treatment. Age was a factor whose potential influence was not replicated when year 2007 data were examined.

Prior research has investigated select factors regarding potential associations with outcomes for patients with MDD treated with DLX (Mallinckrodt *et al.*, 2005; Wohlreich *et al.*, 2007). However, there is little research investigating factors associated with differential effects between treatment strategies. Perahia *et al.* (2008) present results from two head-to-head randomized

clinical trials of DLX and VLX and found no significant treatment differences in the primary benefit risk measure nor standard clinical efficacy measures. A greater percentage of VLX patients completed the 12-week trials compared to DLX patients. Perahia *et al.* (2008) did not report any information on patient heterogeneity in clinical response or duration. Using usual care data, Ye *et al.* (2011) found that patients with more comorbidities and higher prior health care costs were preferentially prescribed DLX relative to VLX. Liu *et al.* (2011) compared the persistence of DLX and VLX in usual care settings but also did not assess factors associated with differential responses. This current work adds to the analysis of Liu *et al.* (2011) by presenting data mining aimed at understanding factors associated with patient differential persistence to two SNRIs as utilized in primary care.

Local control methodology utilizes unsupervised clustering to estimate a counterfactual treatment difference for each patient. Because local treatment differences are computed within relatively homogeneous (in terms of baseline X-characteristics) clusters of patients, the method is applicable to observational research. By estimating and utilizing the counterfactual treatment differences, tools such as CART can then identify heterogeneity of responses between two treatments for an outcome of interest, as opposed to simply heterogeneity of response to a single treatment in a population. Counterfactual treatment differences for each patient can also be estimated using other techniques, such as regression models containing treatment interaction terms or matching patients on propensity score estimates. Several researchers have developed methods that utilize the basic CART algorithm as well as random forests (Breiman, 2001) to identify effects that correspond to treatment by covariate interactions. Computationally intensive approaches, like those of (Negassa *et al.*, 2005) and (Su *et al.*, 2009), partition the entire covariate space into subsets of patients with heterogeneous survival patterns. A similar procedure for direct subgroup identification called Subgroup Identification based on Differential Effect Search (SIDES) has been proposed (Lipkovich *et al.*, 2011). Unlike previously mentioned tree-based approaches, in this procedure the emphasis is not on model building for prediction of treatment effect for any given subject but rather on direct search for subgroups exhibiting treatment effect in the desired direction, while restricting the search space using various penalty parameters and controlling for the overall type I error rate associated with the search using permutation-based multiplicity adjustment. More work is needed to assess the advantages

and disadvantages of this local control approach relative to alternative data mining techniques for this task. Additional research is also needed to assess the tradeoff between using a larger number of smaller clusters to eliminate bias versus fewer clusters to avoid non-informative clusters and encompass a larger patient population. Similarly, while Ward's minimum variance method was used to drive the formation of clusters here, research on the optimal choice of clustering method would be of value.

There are several other key limitations to this work. First, measurement bias is a general limitation to causal inference especially when utilizing retrospective observational data. The data was not collected for research purposes and information is obtained in a usual care and not in a blinded fashion as is typical for clinical trials. Coding biases and lack of validation of diagnoses are known potential issues in claims database research and recent International Society for Pharmacoeconomic and Outcomes Research (ISPOR) good research practices documents discuss these issues in detail (Berger *et al.*, 2009; Cox *et al.*, 2009). The methodology discussed in our work nor the large sample sizes in such claims databases resolves these data limitations. The database also lacks information on clinical severity, genetic information, and socio-economic influences. Such variables may be important influencers of MDD medication persistence but were not considered in this analysis. In fact, results from observational studies are particularly susceptible to bias from unmeasured confounders due to lack of randomization of patients to treatments. Additional replication would be highly desirable. Replication is a hallmark of research designed to avoid spurious findings. Unfortunately, there is no clear gold standard method for replication in retrospective database work. While data from additional years of this database were utilized, replication using the same database may simply repeat the same biases due to unmeasured confounders (Vandenbrouke, 2008). Utilizing different sources of data with different measures and populations would be of great value. Lastly, while we investigated factors prior to initiation of the new medication, we did not attempt to account for any post-initiation influences, such as concomitant medications including other medications for depression. The findings must be interpreted in light of the earlier limitations.

In summary, there is a growing need for analyses providing information to assist medical decision-makers regarding individual treatment choices. The local control methodology illustrated here, based on unsupervised clustering, is one promising approach for such data mining in observational research.

Acknowledgements

The authors thank Ms Teri Tucker of i3Statprobe for her assistance in preparing this manuscript and Dr Chakib Battiou of Eli Lilly and Company for his assistance with analysis programming.

Declaration of interest statement

This study was funded by Eli Lilly and Company. Dr's Faries, Chen, Zagar, Lipovich, and Liu were full-time employees of Eli Lilly when this work was conducted. Dr Obenchain received consulting fees from Eli Lilly and Company.

References

- Benson K., Hartz, A.J. (2000) A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine*, **342**(25), 1878–1886.
- Berger M.L., Mamdani M, Atkins D., Johnson M. L. (2009) Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part I. *Value in Health*, **12**(8), 1044–1052.
- Breiman L. (2001) Random forests. *Machine Learning*, **45**(1), 5–32.
- Breiman L., Friedman J., Stone C.J., Olshen R.A. (1984) *Classification and Regression Trees*, New York, Chapman & Hall/CRC.
- Cantrell C.R., Eaddy M.T., Shah M.B., Regan T.S., Sokol M.C. (2006) Methods for evaluating patient adherence to antidepressant therapy: a real-world comparison of adherence and economic outcomes. *Medical Care*, **44**(4), 300–303.
- Cramer J.A., Roy A., Burrell A., Fairchild C.J., Fuldeore M.J., Ollendorf D.A., Wong P.K. (2008) Medication compliance and persistence: terminology and definitions. *Value in Health*, **11**(1), 44–47.
- Cox E., Martin B.C., Van Staa T., Garbe E., Siebert U., Johnson M.L. (2009) Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value in Health*, **12**(8), 1053–1061.
- Concato J., Shah N., Horwitz R.I. (2000) Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, **342**(25), 1887–1892.
- Crystal S., Akincigil A., Bilder S., Walkup J.T. (2007) Studying prescription drug use and outcomes with medicaid claims data: strengths, limitations, and strategies. *Medical Care*, **45**(10), S58–S65.
- Cui Z., Faries D.E., Gelwicks S., Novick D., Liu X. (2012) Early discontinuation and suboptimal dosing of duloxetine treatment in patients with major depressive disorder: analysis from a US third party payer perspective. *Journal of Medical Economics*, **15**(1), 1–11.
- D'Agostino R.B. Jr. (1998) Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics Medicine*, **17**(19), 2265–2281.
- Davidoff F. (2009) Heterogeneity is not always noise: lessons from improvement. *Journal of the American Medical Association*, **302**(23), 2580–2586.
- Des Jarlais D.C., Lyles C., Crepaz N., the TREND Group. (2004) Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health*, **94**(3), 361–366.
- Foster J.C., Taylor J.M.C., Ruberg S.J. (2011) Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, **30**, 2867–2880.
- Hernan M.A. (2004) A definition of causal effect for epidemiological research. *Journal of Epidemiology & Community Health*, **58**(4), 265–271.
- Johnson M.L., Crown W., Martin B.C., Dormuth C.R., Siebert U. (2009) Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part III. *Value in Health*, **12**(8), 1062–1073.
- Kroenke K., Krebs E.E., Bair M.J. (2009) Pharmacotherapy of chronic pain: a synthesis of recommendations from systematic reviews. *General Hospital Psychiatry*, **31**(3), 206–219.
- Lipkovich I., Dmitrienko A., Denne J., Enas G. (2011) Subgroup Identification based on Differential Effect Search (SIDES) – a recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics Medicine*, **30**(21), 2601–2621.
- Liu X., Chen Y., Faries D. (2011) Adherence and persistence with branded antidepressants and generic SSRIs among managed care patients with major depressive disorder. *ClinicoEconomics*, **3**, 63–72.
- Mallinckrodt C.H., Watkin J.G., Liu C., Wohlreich M.H., Raskin J. (2005) Duloxetine in the treatment of major depressive disorder: a comparison of efficacy in patients with and without melancholic features. *BMC Psychiatry*, **5**, 1.
- Motheral B., Brooks J., Clark M.A., Crown W.H., Davey P., Hutchins D., Martin B.C., Stang P. (2003) A checklist for retrospective database studies – report of the ISPOR Task Force on Retrospective Databases. *Value in Health*, **6**(2), 90–97.
- Negassa A., Ciampi A., Abrahamowicz M., Shapiro S., Boivin J.F. (2005) Tree-structured subgroup analysis for censored survival data: validation of computationally inexpensive model selection criteria. *Statistics and Computing*, **15**(3), 231–239.
- Obenchain R.L. (2010) The local control approach using JMP. In Faries D., Leon A.C., Haro J.M., Obenchain R.L. (eds) *Analysis of Observational Health Care Data Using SAS*, pp. 151–192, Cary, NC: SAS Press.
- Obenchain R.L. (2011) SAS Macros for “Local Control” Phases One and Two. Copyright (c) 2009, Foundation for the National Institutes of Health; Apache 2.0 License. (OMOP Version 1.4). <http://members.iquest.net/~softx> [23 August 2011].
- Perahia D.G., Pritchett Y.L., Kajdasz D.K., Bauer M., Jain R., Russell J.M., Walker D.J., Spencer K.A., Froud D.M., Raskin J., Thase M.E. (2008) A randomized, double-blind comparison of duloxetine and venlafaxine in the treatment of patients with major depressive disorder. *Journal of Psychiatric Research*, **42**, 22–34.
- Rosenbaum P.R. (2010) *Design of Observational Studies*. New York, Springer Science + Business Media, LLC.
- Rosenbaum P.R., Rubin D.B. (1983) The central role of the propensity score in observational

- studies for causal effects. *Biometrika*, **70**(1), 41–55.
- Ruberg S., Chen L., Wang Y. (2010) The mean doesn't mean as much anymore: finding subgroups for tailored therapeutics. *Clinical Trials*, **7**(5), 574–583.
- Stafford R.S., MacDonald E.A., Finkelstein S.N. (2001) National patterns of medication treatment for depression, 1987 to 2001. *Primary Care Companion Journal of Clinical Psychiatry*, **3**(6), 232–235.
- Su X., Tsai C.L., Wang H., Nickerson D.M., Li B. (2009) Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, **10**, 141–158.
- Vandenbroucke J.P. (2008) Observational research, randomised trials, and two views of medical science. *PLoS Medicine*, **5**(3), e67.
- Vandenbroucke J.P., von Elm E., Altman D.G., Gotzsche P.C., Mulrow C.D., Pocock S.J., Poole C., Schlesselman J.J., Egger M, for the STROBE Initiative. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Medicine*, **4**(10), e297.
- Wohlreich M.M., Wiltse C.G., Desai D., Ye W, Robinson R.L., Kroenke K, Kornstein S.G., Griest J.H. (2007) Duloxetine in practice-based clinical settings: assessing effects on the emotional and physical symptoms of depression in an open-label, multicenter study. *Primary Care Companion Journal of Clinical Psychiatry*, **9**(4), 271–279.
- Ye W, Zhao Y, Robinson R.L., Swindle R.W. (2011) treatment patterns associated with duloxetine and venlafaxine use for major depressive disorder *BMC Psychiatry*, **11**, 19.