January 2011

Comparing Antipsychotic Treatments For Schizophrenia: A Health State Approach

Lewei Lin

Comparing Antipsychotic Treatments for Schizophrenia: A Health State Approach

A Thesis Submitted to the

Yale University School of Medicine

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

by

Lewei (Allison) Lin

2011

Abstract

Comparing Antipsychotic Treatments for Schizophrenia: A Health State Approach

Lewei (Allison) Lin, (Sponsored by: Robert Rosenheck), Department of Psychiatry, Yale University School of Medicine, New Haven, CT

HYPOTHESIS/AIMS: Apply health state analysis to a large clinical trial dataset of patients with schizophrenia to provide added insights into clinical and quality of life characteristics. We also evaluated the hypothesis that there is no difference in the distribution of health states of patients across several antipsychotic medications. METHODS: This study was a secondary analysis of data from the CATIE trial, a multi-site double blind clinical trial of antipsychotic treatments for schizophrenia. We applied K-means clustering to the CATIE data, creating discrete clusters with symptom and side effect characteristics that were then validated using a panel of quality of life measures. A comparison was made across medications for differences in cluster distributions at baseline and 6 months. RESULTS: 1049 patients from the CATIE trial dataset were included for initial cluster analysis. By examining cluster profile plots, it was determined that 5 was the optimum number of health states. Using intent to treat, the model was applied to compare 6-month outcomes for patients on perphenazine, olanzapine, risperidone and quetiapine. Chi square tests of independence showed significant difference ($p=0.0090$) in the distribution of patients across health states for the 4 medications at 6 months. Chi squared pairwise comparisons were significant for only perphenazine vs. risperidone ($p = 0.012 < \alpha$ of 0.025 with Hochberg correction) and for olanzapine vs. risperidone ($p= 0.0010 < \alpha$ of 0.05) but not for any other pairs. At baseline, almost 20% of patients were in the worst health state (HS+Dp+Ak), but decreased at 6 months, with the greatest decreases in the pherphenazine (9.2% decrease) and olanzapine (11.1%) groups compared to risperidone (4.7%) and quetiapine (6.7%). There was a large increase in the best health state (LS+LSE) for patients taking perphenazine (15.0%), olanzapine (18.5%) and quetiapine (12.0%) but less for patients taking risperidone (4.5%). CONCLUSION: This study demonstrated health state analysis is a useful tool that provides information on the overall clinical state of patients and can potentially be used to help guide clinicians in treatment decisions.

Acknowledgements

I would like to extend sincere thanks to my advisor, Dr. Robert Rosenheck for his patience and dedication to this project. Thank you for the guidance, the rapid fire e-mails, and the many conversations regarding this paper and the field in general. I greatly appreciated the opportunity to work with you and hope we continue to have opportunities in the future.

I would also like to thank Catherine Sugar, the Director of the Semel Institute Statistics Core and Assistant Professor of Biostatistics at UCLA. She is the creator of the health state analysis technique and performed the cluster analysis for this study. I thank you for your dedication over the entire course of this project and for providing statistical guidance along the way.

I would like to thank my dear friend, Rebecca Lessem, a newly minted economist, who taught me STATA through the course of this project. I cannot imagine anyone else who would have so willingly and tirelessly provided answers to all of my questions. And again, I hope there will be many joint projects in our future.

And finally I would like to thank Joern Putschke, for his patience and willingness to listen and for his desire to help in any way he can.

Table of Contents

Introduction 5
Comparative effectiveness research today 5
A need for CEA studies of antipsychotics 6
How can CEA be more useful to clinical care? 13
Need for new methods – health state analysis as an alternative approach 14
Statement of aims and hypotheses 16
Methods 17
Study design 17
Measures 18
Analyses 19
Results 22
Sample Characteristics 22
Generating clusters 22
Concurrent Validation 26
Discussion 30
Conclusion 39
References 41
Appendix 1 48

Introduction

Comparative effectiveness research today

With the advent of many new treatments for psychiatric illnesses in the last fifty years, there has also been greater uncertainty about the relative effectiveness amongst treatments and the impact on a patient's overall well-being. Comparative effectiveness research is potentially a powerful approach to improve quality of care and outcomes. With the passing of the American Recovery and Reinvestment Act of 2009, policy makers not only acknowledged its importance, but also allocated $1.1 billion to support new research and to create the Federal Coordinating Council for Comparative Effectiveness Research, an organization to oversee its development (1).

In 2007, the Institute of Medicine defined comparative effectiveness as the "comparison of one diagnostic or treatment option to one or more others" and others have defined it as the study of risks and harms associated with alternative options of health care and their impact on patients (2) (1). Comparative effectiveness studies have been a recent focus of research trials because of their potential to improve clinical care. However, the IOM has estimated that only half of the treatments that are considered standard of care have shown to be clinically effective in research trials and that the degree to which one treatment is more effective than another is in most cases unclear due to the lack of evidence (3). The roots of this problem may lie deep in the US healthcare system. For example, the Food and Drug Administration requires evidence on the safety

profile and effectiveness of new treatments as compared to placebo but does not typically require comparisons to be made with previously approved alternatives. There is increasing clinical research comparing treatments but it is still recognized that there is not enough comparative research that can be directly clinically relevant (4).

A need for CEA studies of antipsychotics

Comparative effectiveness research may be particularly important in evaluating treatments for schizophrenia. Since the discovery of chlorpromazine in the late 1950's, there has been an explosion of new antipsychotic medications on the market. In 2007, there were over sixty approved antipsychotics worldwide (5). These medications are usually divided into two categories, first generation antipsychotics, including chlorpromazine, were those discovered prior to clozapine. In the 1980's, reports were published about clozapine indicating unprecedented efficacy in treating refractory schizophrenia (6). All drugs discovered after its release was marketed as a new and distinctly superior category with fewer extrapyramidal symptoms and were labeled atypical or second generation antipsychotics (SGA's).

Initially it was thought that the SGA's would be at least as effective in treating the positive symptoms of schizophrenia (hallucinations, delusional thinking, and thought disorganization), but would result in fewer side effects. In addition, SGA's were thought to be promising for treating negative symptoms (anhedonia, flat affect, paucity of speech)

cognitive impairments and mood symptoms. Based on these hopes, large professional bodies around the world including the American Psychiatric Association, the World Psychiatric Association, the United Kingdom's National Institute for Health and Clinical Excellence, and the Texas Medication Algorithm Project recommended the newer medications in favor of the old (7) (8) (9) (10). Some thought the SGA's would render FGA's obsolete (11).

Subsequently, use of SGA's have soared along with costs. In 1994, the annual expenditures on antipsychotic medications in the U.S. was $1.4 billion and less than 5% of schizophrenic patients received SGA's (12). Ten years later, 90% of patients received SGA's and the annual expenditure totaled over $10 billion, with 70% paid by the U.S government through Medicaid (12). A similar story unfolded globally. Global expenditures on antipsychotic medications increased from $0.5 billion annually in the 1990's to more than $15 billion annually a decade later (13). This increase was driven by use of SGA's which cost 5-30 times more than FGAs. Such costs could be justified if patients experienced significant symptom reduction and overall improvement in quality of life, but there is increasing skepticism about overall benefit. Such uncertainty has led to a growing literature of systematic reviews and large scale clinical trials comparing effectiveness of antipsychotic medications.

One of the first comprehensive systematic reviews was published by Geddes et al. in 2000 (14). That analysis included results of 52 randomized trials published before

Dec. 1998 comparing SGA's (amisulpride, clozapine, olanzapine, quetiapine, risperidone, and sertindole) with FGA's (most commonly haloperidol or chlorpromazine). Outcome measures included symptom measures and side effects. The study concluded that SGA's were equally effective as a group and were no more effective than FGA's when they controlled for higher than recommended doses of FGA's in trials. SGA's still showed a moderate benefit in EPS symptoms, but there were no significant difference in dropout rates.

In contrast, a review published by Davis et al in 2003 concluded that SGAs were not a homogenous group in terms of effectiveness. That analysis included 142 randomized controlled trials (published and unpublished data) up to May 2002 comparing ten SGA's with FGA's in patients with schizophrenia or schizoaffective disorder (15). The main outcome measure was symptom scores. The study concluded that four SGA's (clozapine, amisulpride, risperidone, and olanzapine) were significantly more effective than FGA's. The remaining SGA's (aripiprazole, quetiapine, remoxipride, sertindole, and ziprasidone) were no more effective than FGA's. And unlike the Geddes et al. study, they found no evidence that the dose of FGAs affected the results.

Leucht and colleagues published the most recent review in 2009 comparing nine SGA's with FGA's on a variety of outcome measures, including symptoms, relapse rates, quality of life, EPS, weight gain and sedation (16). Their review included 150 doubled blind studies published before Aug 2005. They concluded that the same 4 SGAs

highlighted in the Davis et al review (amisulpride, olanzapine, clozapine, and risperidone) were significantly more effective in treating both positive and negative symptoms with small to medium effect sizes and that aripiprazole, quetiapine, sertindole, ziprasidone, and zotepine were no more effective than FGA's. In addition, SGA's resulted in fewer EPS symptoms compared to haloperidol even at low doses, but the difference in EPS compared with less potent FGA's was less clear. However, SGA's (except for aripiprazole and ziprasidone) were associated with greater weight gain. They also concluded that few studies included data on quality of life and relapse rates. Leucht and colleagues also compared the results of their review with the others reported here as well as the Cochrane Reviews and concluded that all of the reviews demonstrated that amisulpride, clozapine, olanzapine and risperidone were superior to FGAs but for the other SGAs, there were no significant differences (17) (16).

The need for more objective and definitive evidence comparing effectiveness of different antipsychotics on the market also led to two large scale government sponsored multi-site trials on the long term effectiveness of medications to treat schizophrenia. The Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1) from the United Kingdom was a large (N=227) randomized clinical trial with a study design that mimicked real world prescribing practice. Clinicians were allowed to prescribe any SGA (excluding clozapine in the first arm of the study) and any FGA medication and both patients and clinicians were not blind to treatment assignment. This design resulted in a high 12 month followup rate of 81%, but found SGAs showed no

significant advantage over FGAs after one year on measures of quality of life, symptoms, discontinuation, or side effects.

The Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) for schizophrenia was sponsored by the National Institute of Mental Health and conducted between 2001 and 2004. 1460 patients were followed for up to 18 months at 57 sites in the U.S and clinicians and patients were blinded to treatment assignment (18). CATIE compared the effectiveness of one first generation antipsychotic, perphenazine, with the four second generation antipsychotics available on the market at the time. Primary CATIE results found olanzapine to be associated with significantly longer time to discontinuation compared to quetiapine and risperidone but the difference was not significant in the perphenazine comparison. Time to all cause treatment discontinuation was significantly longer for clozapine in a treatment resisitant arm of the study compared to quetiapine and risperidone but not to olanzapine. And the overall discontinuation rate was high – 74% of patients discontinued their study medication before the end of trial at 18 months. There were no significant advantages of second generation antipsychotics over perphenazine on symptoms, neurologic side effects, quality of life, employment, or neurophsychological functioning (19). Perphenazine was less costly by $300-500 per month compared to all SGA's when accounting for medication costs and costs of health services utilized (20).

These reviews and clinical trials have been some of the largest to date, but they have not been without criticism. More than 80% of the studies included in Leucht et al.'s meta analysis had a study period of 12 weeks or less, even though long term treatment of schizophrenic patients is needed to prevent relapse. Moreover, most of the trials included in the analysis are from early drug studies sponsored by pharmaceutical companies, which can skew the evidence in favor of those drugs over comparisons (21). In addition the choice of comparison drugs in 95 out of 150 of Leucht's analysis was haloperidol, a high potency FGA. Few studies chose medium potency drugs, such as perphenazine, which is thought to be associated with lower rates of EPS (22).

Although it was hoped that results from these two large scale studies would provide conclusive answers, the results actually fueled more debate. On one side, led by the primary investigators of the studies, the results were interpreted to suggest that SGA's confer no additional clinical benefit over FGA's. On the other side, the argument is that the study methodologies were flawed and caution should be used in applying the results. Critics of the CATIE study point to several components. In particular, they question the results indicating no difference in tardive dyskinesia (TD) across the medications. In the CATIE trial, patients with tardive dyskinesia (about 15% of the sample) were excluded from being randomly assigned to perphenazine because it was generally thought in 2000 by an expert panel that patients with TD should not be exposed to any FGA (23). Critics say this exclusion may have biased the samples so that those assigned to perphenazine were not as severely ill as those assigned to other medications. However the study

investigators have pointed out that this issue was already addressed in the study design and analysis. In the analysis, perphenazine patients were only compared to equivalent patients who did not have tardive dyskinesia at baseline.

Although there were some differences between the results of CATIE and CUtLASS and those of the meta analyses, overall, the evidence suggests that FGA's and SGA's are not homogenous groups of medications and that each drug should be judged on its own efficacy and risk of side effects. Furthermore, with the results of such large scale multi-site government sponsored studies, there is also enough evidence to suggest that SGA's as a group are no more effective than FGA's, even though the debate will still likely continue. And as more is learned about the mechanisms of SGA's, it is apparent that the mechanisms of drugs in this category differ significantly from one another. For example, it was thought that the property of blocking serotonin receptors accounts for their improved efficacy, but many of the most potent serotonin blockers are not more efficacious and amisulpride, which is not a serotonin blocker, is more effective than many other SGA's (15).

Yet, as some recent studies have shown, the knowledge gained from these large scale, carefully designed trials is not being disseminated widely to clinicians and is not having widespread impact on prescribing practices. A European survey of psychiatric trainees show that 96% of trainees prefer "atypical" antipsychotics for the acute treatment of schizophrenia and cite efficacy as the primary (76%) factor for their prescription

choices (24). Side effect profile (21%) and cost (3%) played less of a role. Thus, it is still the case that many clinicians reach first for SGA's for schizophrenia under the assumption that the drugs are more effective and have reduced side effects

How can CEA be more useful to clinical care?

One barrier in disseminating information from comparative effectiveness studies may be that the results are not always directly applicable to treatment selection for a specific patient with a complex clinical profile. Clinical trials often incorporate methods such as subgroup analysis to look for effects that may pertain to specific patient populations, but such analysis is often not comprehensive and there are statistical challenges in their interpretation (1). However, even the most well designed clinical trials often fail to provide conclusions that can guide the crucial question of "what is the best treatment for this particular patient?"

Outcomes of trials are usually reported using dimensional scales and comparisons are made using averages along dimensions of health including symptoms, side effects, and overall quality of life. This approach cannot capture important and complex relationships between different dimensions of health (25). And in clinical practice, treaters must routinely assess the relative benefit of symptoms and side effects as they emerge in multidimensional health states. Comparative effectiveness research has been urged to develop methods to provide insight into overall clinical impact of medications because an individual patient can experience both benefit and harm from a given treatment (26).

Need for new methods – health state analysis as an alternative approach

Health state modeling is an analytic method that could potentially improve the practical value of comparative effectiveness research. In this data driven approach, cluster analysis is used to classify patients. Cluster analysis is a way of dividing data into meaningful or useful groups by capturing the natural structure of the data. This technique is widely used in many fields including psychology, biology, statistics, pattern recognition, machine learning and data mining (27).

Measurements at particular points in time can then be grouped into discrete multi-dimensional health states such that patients in a given state are as similar as possible over several dimensions of symptoms and side effects. Clinically relevant change can then be measured as the probability of moving individuals from one health state to another rather than the traditional approach of measuring net increase or decrease over time on multiple preset scales. One can then compare randomized treatments by the differing probabilities of causing desirable transitions. For example, patients in two treatment groups can have similar levels of overall health, but one treatment increase the likelihood that patients have severe symptoms and few side effects while the other group leads to severe side effects but few symptoms. Health state modeling could clearly distinguish between these two types of outcomes.

Health state modeling has the additional benefit of providing a simple way to estimate long run outcomes using data from studies of a finite duration using Markov chain theory. The transition probability from one health state to another can be accumulated over several time periods and across the entire patient sample to determine the distribution of patients in each health state at equilibrium with different treatments (28). These long run health states can be used to estimate health related utility levels or Quality Adjusted Life Years for use in cost-effectiveness or cost benefit analysis (29).

This method has recently been used to compare outcomes of patient populations from different mental health programs in a paper by James et al (30). In this study, they compare the health state modeling approach to a conventional mixed effects regression model in the analysis of longitudinal data of patients treated at a Veterans Affairs (VA) medical center versus patients treated at a Community Mental Health Center (CMHC). Although both methods produced similar conclusions – patients treated at the CMHC were more stable over time compared to VA treated patients, the health state analysis gave additional insight into which subgroups of the VA population were developing more severe symptoms. The analysis showed that patients in the best and worst health states changed little over time, but patients with mild symptoms with hallucinations and patients with severe positive and negative symptoms were the most likely to deteriorate.

In another recent clinical trial, health state analysis was used to compare effectiveness and cost effectiveness of haloperidol and clozapine (25). A secondary

analysis was conducted of a randomized controlled trial comparing haloperidol and clozpaine in a hospitalized VA patient population. The health state model consisted of clusters with different symptom and side effect profiles. The study concluded that clozapine, compared to haloperidol, differentially increased the proportion of patients with mild symptoms and decreased the proportion of patients with severe positive symptoms, but the proportion of patients with negative symptoms stayed relatively constant over time. And projecting long run outcomes and health costs using a Markov model, the study authors conclude that clozapine is more cost effective compared to haloperidol.

In this study, we apply health state modeling as a secondary analysis of data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial. We use the primary data from this large study, which tried to represent a real world cohort of patients through its broad inclusion and exclusion criteria. Although the original analyses have provided insight into treatment effects on specific dimensions of health, the overall impact of medications remains ambiguous. In this study, we re-analyze the data from CATIE using a health state approach.

Statement of aims and hypotheses

Aim 1

Determine if health state analysis can potentially provide a richer picture of the clinical and quality of life characteristics of this patient population.

Hypothesis 1

Health state analysis results in additional insights into the clinical and quality of life characteristics of patients compared to traditional dimensional approaches.

Aim 2

Apply health state analysis to compare the impact of various common antipsychotic medications on the health state outcomes of patients.

Hypothesis 2

There is no difference in the distribution of health states of patients across medications over the treatment time period.

Methods

Study design

CATIE was conducted between January 2001 and December 2004 at 56 U.S. sites. The study was designed to compare effectiveness and cost-effectiveness of one first generation antipsychotic, perphenazine and the second generation antipsychotics olanzapine, quietiapine, risperidone and ziprasidone. Details of the study and exclusion criteria have been reported elsewhere (31) (32). An algorithm assigned patients to a series of treatment phases. In phase 1, patient were assigned to medications under double-blind conditions and were followed up to 18 months or until treatment was discontinued for any reason. Patients with tardive dyskinesia (15% of the sample) were excluded from the randomization that included perphenazine, and these patients with a history of TD

baseline were excluded from our analyses. Patients who discontinued the first medication they were assigned were invited to start another second-generation antipsychotic and open treatment was offered to patients who refused the second random assignment or who failed the second randomization. In the present analysis, we used a subset of data on patients treated with perphenazine, olanzapine, quetiapine and risperidone because ziprasidone was introduced after much of the sample had been recruited.

Measures

The data consisted of symptom measuring instruments including the Positive and Negative Syndrome Scale (33). The PANSS has 3 subsections: positive symptoms such as hallucinations and delusions, negative symptoms such as blunted affect, and general emotional disturbance such as anxiety and depression. The Calgary Depression Symptom Scale for Schizophrenia (34) further assesses specific neurovegatative and subjective aspects of depression including sleep, suicidality, and hopelessness. In this analysis, we used summed scores from the subsections of the PANSS and the total Calgary score. Extrapyramidal side effects were assessed using the Abnormal Involuntary Movement Scale (35), the Barnes Akithesia Scale (36)and the Simpson-Angus Scale (37). The BMI was also included as a measure of metabolic side effects. With the exception of BMI, all of these instruments use Likert scales to measure symptom severity with higher scores indicating more severe symptoms.

A broad selection of quality of life measures were used as validation measures for overall severity and impact of illness associated with each health state. These were

assessed as part of the original CATIE study and in this analysis, also serve as a measure of relative illness severity and psychosocial functioning associated with patients in each cluster. The measures include the Lehman Quality of Life summary item (38), a semi-structured interview that assesses the overall life circumstances of patients with mental illness. The SF12 mental and physical subscales (39) are summary scores from a patient administered health related quality of life survey that has been widely used in many patient populations. The Heinrichs Carpenter Quality of Life Interview (40) was originally developed to assess the schizophrenia deficit syndrome and contains subscales measuring interpersonal relations, occupation, intrapsychic foundations (motivation, anhedonia, emotional interaction) and common objects and activities. Finally, the Visual Analogue Scale is a patient's self reported measure of overall health. Data were collected at baseline, 1, 3, 6, 9, 12, 15, and 18 months.

Analyses

We use K-means clustering on the standardized symptom and side effects measures to obtain discrete health states. K-means is a partitioning algorithm that treats each observation in the dataset as an object with a location in multi-dimensional space defined by the characteristics that are measured (28). The algorithm creates clusters of observations, where each cluster is defined by the patient members that have been assigned to it and by its centroid, or the center point of all the members in that cluster. K-means is an iterative algorithm that minimizes the sum of distances from each member in the cluster to its cluster centroid and over all clusters. The process involves moving

objects between clusters until the sum of distances has reached a global minimum. This results in non-empty non-overlapping clusters where members in each cluster are as close together as possible and where members in different clusters are as far apart as possible.

The most important technical issue to consider in k-means cluster analysis is the selection of k, the number of clusters used. K is a user defined value that can take on any integer value. It is important to have as few clusters as possible for ease in interpreting results and to allow for enough subjects in each cluster for adequate statistical power. However, it is also important to have enough clusters so that patients who appear clinically different are not grouped into the same cluster. In this study, we use cluster profile plots as a visual tool to aid in selection of clusters. For each cluster, we plot the average score for each outcome score across all patients who belong to that cluster. We produce plots for different values of k and compare the patterns of plots to choose the greatest value of k where each cluster plot appears clinically distinct. We increase k as long as addition of clusters results in separation of cluster centers along at least one of the clinical axes. This plot provides a picture of the average patient in each cluster and can also be used to provide a narrative description of these patients for purposes of utility measurement. In addition, we use quality of life measures as a way of validating and differentiating the severity of the health states.

After choosing the number of clusters and removing patients with missing values, we ran the k-means cluster algorithm on the symptom and side effect data to produce the

health state model. We then used the health state model to compare outcomes across medications over time. Specifically, we compared the proportion of patients in each health state at baseline and then at time 6 months across medications. We examined the data cross-sectionally using chi-square tests of independence to determine if there were significant differences of health state patterns across the treatment groups at each time point.

Next, we looked at long run differences in patterns of health states across medications. Before applying Markov chain theory, we tested assumptions of stationarity of the data, to determine whether the transition probabilities from one state to another remain fixed over time. With data that remains stationary, one can calculate a stationary distribution, which is the fraction of patients residing in each health state modeled after many time periods.

In this study, the initial clustering was performed by Dr. Catherine Sugar, a professor of biostatistics at UCLA who helped develop the health state technique. The planning for the cluster analysis and selection of the optimum number of clusters were done in partnership with her and all data interpretation was performed by this paper's author.

Results

Sample Characteristics

Appendix 1 depicts the demographic and clinical characteristics of patients at the time of initial random assignment in this study. The average age is 39.3. 74% of patients are male. 60.1% of patients are White and 60.6% of patients have never been married. The average burden of psychotic symptoms as measured by the PANSS total is 75.5. Patients on average are overweight, bordering on obese with BMI of 29.8. The average monthly cost of all healthcare received is $2,299.

Generating clusters

In order to determine the optimal number of health states, all available observations with complete information on outcomes measures across all time points were used. By examining cross-tabulation tables and cluster profile plots using different number of clusters, it was determined that 5 clusters allowed for the greatest amount of differentiation amongst clinically meaningful clusters with the most parsimonious and interpretable set of clusters. Fig. 1 shows the cross-tabulation of cluster memberships between 4 and 5 cluster models. In the figure, the rows represent the 4 cluster model and columns represent the 5 cluster model. So the value in the first row, second column shows the number of people who were in cluster 1 in the 4 cluster model and cluster 2 in the 5 cluster model. The numbering of clusters is arbitrary in this case – for example, the majority of patients in cluster 2 in the 4 cluster model are in cluster 3 in the 5 cluster

model. But patients in cluster 4 become distributed in clusters 4 and 5 in the 5 cluster model. This indicates that most patients stay assigned to the same cluster, but that most of the patients in cluster 4 become reassigned to form clusters 4 and 5 in the 5 cluster model.

Figure 1. Cross tabulation of 4 cluster and 5 cluster models. Values represent number of patients in each cluster under 4 cluster (rows) and 5 cluster (columns) models.

	1	2	3	4	5	Row Total
1	196	92	0	0	0	288
2	39	35	350	3	0	427
3	3	13	55	172	0	243
4	1	1	2	36	36	76
Column Total	239	141	407	211	36	1034

We examine cluster profile plots in order to determine the clinical differences between these clusters. Figure 2 shows 4 plots corresponding to each cluster in the 4 cluster model. In each plot, the values shown are the re-centered scores calculated by subtracting the global mean from each score. This allows one to compare values across all outcome measures for which the raw scales scores are measured on different scales. For example, cluster 1 is comprised of patients with low symptoms and low neurological side effects but high BMI scores. In the 5 cluster model, shown in Fig. 3, patients in cluster 1 have low symptoms, neurological side effects and low BMI scores, while cluster 2 is a distinctly new cluster with low symptoms and side effects but strikingly high BMI. The BMI in cluster 2 is over 10 points higher than all other clusters. Thus, the 4 cluster

model has too few clusters to capture the clinical diversity in the patient sample. The addition of clusters beyond 5 produced some additional distinctions in specific movement side effects, but the additional clusters come at the cost of interpretability, parsimony, and power for long run Markov analysis.

Figure 2. Cluster plots for 4 cluster model with re-centered scores on Y axis and question number corresponding to specific questions on all 8 symptom and side effects scales on X axis.

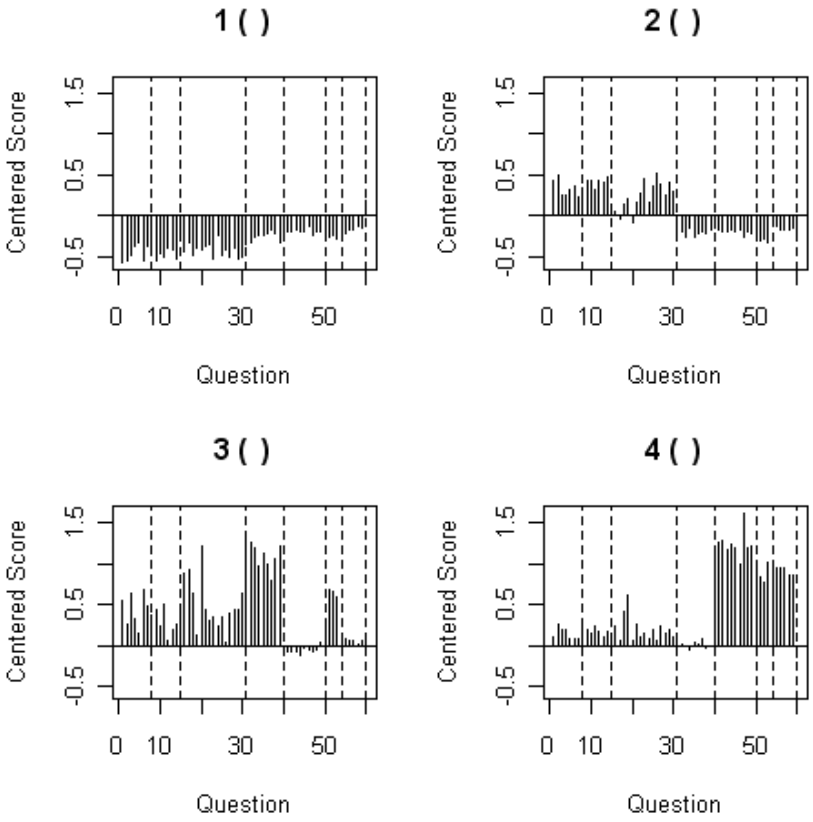
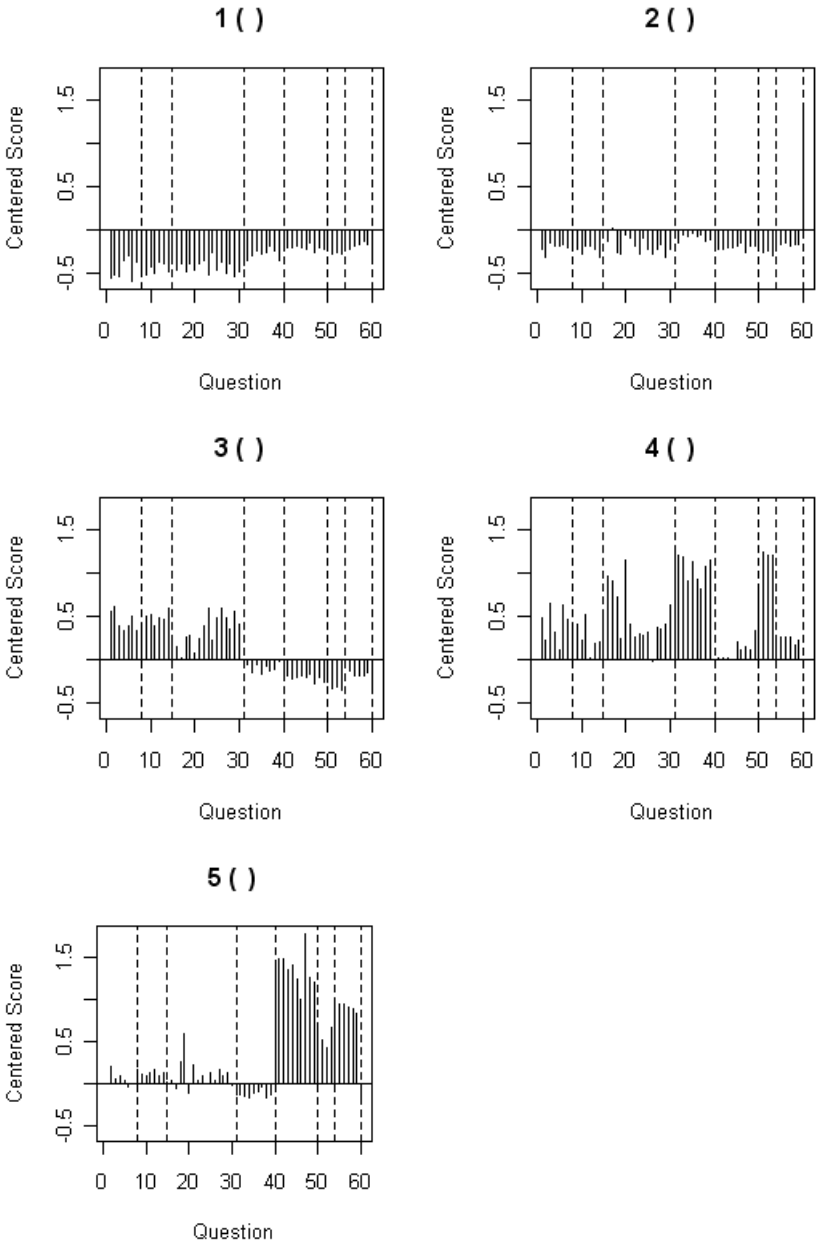


Figure 3. Cluster plots for 5 cluster model with re-centered scores on Y axis and question number corresponding to specific questions on all 8 symptom and side effects scales on X axis.



Objective descriptions of the health states were then created from the mean scores of outcome measures of each health state and are shown in Table 1. For example, patients in cluster 4 in the 5 cluster model had high scores on all sections of the PANSS, highest depression scores, and high akithesia compared to other patients in the sample. In summary, the health states can be characterized as: 1) low symptoms and low side effects (LS+LSE) 2) low symptoms and obesity (LS+Ob) 3) high symptoms and low side effects (HS+LSE) 4) high symptoms with depression and akithesia (HS+Dp+Ak), and 5) moderate symptoms and high side effects (MS+HSE)

Table 1. Mean (std. dev.) of 8 outcome measures across 5 clusters

Cluster	n	PANSS Pos.	PANSS Neg.	PANSS Gen.	Calgary	AIMS	Barnes	Simpson Angus	BMI
1	2674	12.2	14.9	26.0	10.9	1.0	0.6	0.6	27.8
		(3.6)	(4.7)	(5.2)	(2.5)	(2.1)	(1.5)	(1.1)	(4.3)
2	1310	14.6	16.9	30.7	12.2	1.0	0.6	0.7	40.6
		(4.2)	(4.6)	(6.2)	(3.3)	(2.1)	(1.4)	(1.2)	(5.7)
3	2244	20.4	23.3	40.0	12.2	1.0	0.5	0.8	28.0
		(5.0)	(5.4)	(6.9)	(3.1)	(2.2)	(1.1)	(1.2)	(5.2)
4	1095	20.2	21.4	43.3	19.4	2.9	4.2	1.9	30.3
		(5.0)	(5.8)	(8.0)	(4.5)	(3.9)	(2.9)	(2.1)	(7.1)
5	852	16.8	20.0	35.6	12.0	10.5	2.8	3.8	29.2
		(5.2)	(5.8)	(8.0)	(3.1)	(5.8)	(2.9)	(3.1)	(5.8)

Concurrent Validation

In order to assess the differences in quality of life and overall psychosocial functioning, we evaluated scores on 5 summary quality of life or social functioning

measures and report the mean values for each cluster in Table 2. We use these scores as an additional validation step in the choice of clusters. Patients in LS+LSE have highest (least symptomatic) mean scores across all validation measures which is consistent with the cluster characteristics of having the lowest symptoms and side effects. Patients in cluster 4 (HS+Dp+Ak) have the lowest (worst) scores on all validation measures, consistent with the high burden of symptoms and side effects in this group.

Table 2. Mean (std. dev.) of quality of life validation variables across 5 clusters

	Clusters	VAS[1]	QOL31[2]	HCTOT[3]	PCS12[4]	MCS12[5]
1	LS+LSE	73.4	4.8	3.3	50.9	46.3
		(21.2)	(1.2)	(1.1)	(8.5)	(10.2)
2	LS+Ob	68.2	4.6	3.0	46.7	43.5
		(22.3)	(1.3)	(1.0)	(10.1)	(10.6)
3	HS+LSE	64.1	4.3	2.4	49.6	41.4
		(25.4)	(1.3)	(1.0)	(9.3)	(11.1)
4	HS+Dp+Ak	47.1	3.6	2.3	45.0	32.1
		(25.2)	(1.3)	(0.9)	(11.3)	(9.9)
5	MS+HSE	66.3	4.6	2.6	48.2	43.8
		(24.0)	(1.3)	(1.0)	(9.6)	(10.1)

[1]VAS: Visual Analog Scale (1-100)
[2]QOL31: Lehman Quality of Life (1-7)
[3]HCTOT: Heinrichs Carpenter Scale (0-6)
[4]PCS12: SF-12 physical health summary scale (0-100)
[5]MCS12: SF-12 mental health summary scale (0-100)

We used the 5 cluster health state model in a cross sectional analysis to compare 6-month outcomes across the four medications, perphenazine, olanzapine, risperidone and quetiapine. We performed intent to treat analysis, but similar to the original CATIE analysis, we excluded patients with a history of tardive dyskinesia at the baseline

randomization from randomization to all of these drugs because those patients were excluded from the perphenazine arm of the trial. Chi square tests of independence between medication and health state at baseline showed no statistical difference ($p = 0.19$) at baseline across the medications. At 6 months, there was significant difference ($p=0.0090$) in the distribution of patients across health states for the 4 medications. Pairwise comparisons are performed between each pair of medications at 6 months (with 6 pairs in total). Chi squared test was significant only for perphenazine vs. risperidone ($p = 0.012 < \alpha$ of 0.025 with Hochberg false discovery rate correction) and for olanzapine vs. risperidone ($p= 0.0010 < \alpha$ of 0.05) but not for any other pairwise comparisons (41).

To assess specific differences in cluster distribution, we calculated the percent of patients in each health state at baseline and then at 6 months, shown in Table 3. At baseline, 19.7% of patients were in the worst health state, HS+Dp+Ak, and the health state HS+LSE had the largest percentage of patients (over 38.9%). At 6 months, there was an across the board decrease in patients in the worst health state (HS+Dp+Ak), with the greatest decreases in the pherphenazine group (9.2% decrease) and olanzapine (11.1% decrease) groups compared to risperidone (4.7% decrease) and quetiapine (6.7% decrease) as seen in Table 4. There was a large increase in the best health state (LS+LSE) for patients taking perphenazine (15.0%), olanzapine (18.5%) and quetiapine (12.0%) but less for patients taking risperidone (4.5%). Patients in MS+HSE comprised the smallest proportion of patients and this group also had the least evidence of change over the 6 months.

Table 3. Percentage of patients in each cluster at baseline and 6 months

	Clusters	0 mths	6 mths	diff
1	LS+LSE	24%	36%	12%
2	LS+Ob	14%	19%	5%
3	HS+LSE	39%	27%	-12%
4	HS+Dp+Ak	20%	13%	-7%
5	MS+HSE	4%	6%	2%

Table 4. Percentage of patients in each cluster at baseline and 6 months across medications

	perphenazine			olanzapine			risperidone			quetiapine		
	0 mths	6 mths	diff	0 mths	6 mths	diff	0 mths	6 mths	diff	0 mths	6 mths	diff
1	24%	39%	15%	26%	45%	19%	19%	24%	4%	23%	35%	12%
2	15%	19%	4%	10%	14%	4%	15%	19%	4%	14%	23%	9%
3	38%	29%	-9%	37%	23%	-14%	42%	35%	-7%	41%	24%	-17%
4	17%	8%	-9%	24%	13%	-11%	21%	16%	-5%	19%	12%	-7%
5	6%	6%	0%	3%	6%	3%	3%	6%	3%	3%	6%	3%

Increases in the health state LS+Ob were similar across perphenazine (3.92%), olanzapine (3.83%) and risperidone (4.40%) groups but were higher in the quetiapine group (9.19%). Within the LS+Ob group, the average BMI of the olanzapine group increased by 2.23%, while the BMI of all other medication groups decreased. Since olanzapine was associated with greater weight gain in the original analysis (31), we further examined continuous measures of BMI and found the difference in BMI from 6 months to baseline was largest for those patients taking olanzapine (mean 1.46, std dev.

2.64) and smallest for perphenazine (mean -0.077, std. dev. 2.35). Thus changes in obesity were not well captured by our clusters.

In this analysis, 34% of patients were re-randomized into a different arm of the study or dropped out before 6 months. We compared the characteristics patients who stayed on the original medication at 6 months and those who did not. There was no significant difference on any characteristic and a MANOVA performed on the 5 quality of life variables was not significant between the two groups ($p = 0.52$). Further analysis of the phase 1 only sample (excluding observations after discontinuation of the initially assigned drug) showed that Chi square comparisons across the 4 medications and between any medication and perphenazine was not significant ($p = 0.19$ across all 4 medications). Likewise, Chi square comparison at 6 months was not significant across all medications ($p = 0.11$). Comparisons between pairs of medications indicate only olanzapine vs. risperidone was significant ($p = 0.043 < \alpha$ of 0.05 with Hochberg correction). As expected, when examining only the phase 1 sample, there was a greater increase in the best health state (LS+LSE, 15.9%) and greater decrease in the worst health state (HS+Dp+Ak, 9.7%). The change in the LS+Ob cluster remains very similar 5.1% compared to 5.0% in the entire sample.

Discussion

The goal of this study was to determine if the application of health state analysis to data from a large clinical trial could provide additional insights into the overall impact

of a set of widely used pharmacological treatments. To this end, we demonstrated that health state approach can provide a richer picture of the overall clinical status of patients than simple linear measures and can document transitions between states that can be followed over time and compared between treatments. This approach could potentially help researchers better communicate their findings to clinicians. It may also provide a better description to patients of their expected symptom and side effect profiles if they choose to remain adherent to a given treatment. We have also shown how health state analysis can be used to account for overall changes in quality of life.

This is the first comparative effectiveness study, to our knowledge, that applies health state analysis to clinical trial data of several commonly used antipsychotic medications for the treatment of schizophrenia and is one of a small group of studies employing health state analysis in general. Although the results of our study generally confirm the results of original CATIE analyses that employed traditional methods to compare outcomes using linear dimensions of health, this study provides additional support for the view that second generation antipsychotic medications provide no additional benefit over the first generation antipsychotic used in this trial.

An initial aim of this study was to determine if health state analysis could be applied to data from a large clinical trial to identify simple and clinically meaningful representations of a complex patient population with schizophrenia. To this end, we created a model with 5 distinct health states that encompassed data from 4 categories of

symptom measures and 4 side effect measures. Assuming 8 total dimensions of health with a minimum of 3 levels in each dimension, a traditional full factorial design describing the same population would have required at least 24 health states, even though many of those states may not be occupied. Such large number of states would have required a much larger sample of patients in order to obtain meaningful results.

With health state analysis, the selection of clusters is purely data driven by an algorithm that minimizes within group differences and maximizes between group differences. However, one can easily judge if the individual health states are distinctly clinically meaningful. In this analysis, the 5 cluster model created health states with average values on the symptom and side effect scales that could be easily described with distinct clinical characteristics. Although, the description of health states using this method introduces a level of clinically informed judgment, the benefit of having such a parsimonious model allowing further analysis outweighs it. Furthermore, we conducted additional validation analyses using a panel of commonly used quality of life measures. We show that the LS+LSE health state is associated with the highest scores and highest quality of life, which is intuitive clinically. Furthermore, the HS+Dp+Ak health state is associated with the lowest scores on all four validation scales, with lower scores than other health states (MS+HSE) which also have high symptom and side effect burdens. The other three health states have varying scores on the different validation measures depending on the associated attributes. For example, the LS+obesity health state has

relatively low scores on the physical component of the SF12 quality of life instrument, but a comparable score to MS+HSE on the QOL31, an overall measure of quality of life.

The next aim of our study was to determine if health state analysis could potentially provide more information than traditional dimensional comparisons. In studies using these traditional methods, outcomes are usually reported as average changes on an outcome measure along each dimension separately. For example, a medication could result in an average improvement of 10% on symptom measures and also result in 5% increase in side effects. However, it is difficult to determine if the same patients are experiencing both symptom improvement and side effects or if one subpopulation is experiencing the majority of clinical benefit and another population is experiencing predominant increase in side effects. In our analysis, the health states encompass both symptoms and side effects so we can easily assess the overall impact of treatment over time.

In this analysis, we showed that at baseline, the majority of patients were in health states with a high symptom and side effect burden (39% in HS+LSE and 20% in HS +Dp+Ak). However, there was also a significant proportion of patients at baseline in the LS+LSE health state. Over 6 months of treatment, averaging transitions across all study medications, there was a large increase in the LS+ LSE health state and a large decrease in the HS+LSE and HS+Dp+Ak indicating that in general, the same patients who experienced a decrease in symptoms also experienced a decrease in side effects.

However, there was also a sizable increase in the proportion of patients in the low symptoms and high obesity (BMI) health state (5% increase) in the course of 6 months. And at 6 months, over a third of patients still experienced moderate to high levels of symptoms and side effects. Interestingly, the proportion of patients in the MS+HSE state remained relatively low and constant over the 6 months (3.5% increase to 5.8% at 6 months). This health state may have represented a small group of patients who were treatment resistant or were predisposed to experience a high level of side effects.

The final aim of this study was to determine whether there were differences in the distribution of changes in health states across different medications. Despite a group of very large meta analyses indicating that only a few SGA's potentially provide better outcomes and several large multi-site national trials indicating that SGA's do not provide better outcomes over the FGA's used in those trials, there is still considerable controversy about the relative effectiveness of the newer SGA medications. In addition, many clinicians still believe SGA's are more efficacious in general, although those opinions are likely based as much on intensive marketing as on knowledge of outcomes research results (24). Thus, there is a need for additional comparative effectiveness research that could provide evidence to further guide clinicians in their choice of treatments.

In our intent to treat analysis, we showed that the distribution of patients across clusters at 6 months was significantly different across the four medications ($p = 0.009$). After comparing each pair of medications amongst the four, we find that the health state

distributions were significantly different only between perphenazine and risperidone ($p = 0.012 < \alpha$ of 0.025 after Hochberg correction) and between olanzapine and risperidone ($p = 0.001 < \alpha$ of 0.05 after Hochberg correction). Comparing the percentage of patients in the five clusters at 6 months, it was clear that both perphenazine and olanzapine resulted in a larger increase in the LS+LSE (best) health state and a larger decrease in the HS+Dp+Ak (worst) health state in comparison to risperidone. Risperidone was also associated with the smallest decrease in the HS+LSE health. Risperidone resulted in similar changes in the LS+Ob and MS+HSE health states so that overall, it appeared that perphenazine and olanzapine were significantly superior medications to risperidone, at least at the dosages used in this double blind trial. There were no significant differences between any other pair of medications.

Compared to recent large meta analyses, our results align most closely with that of Geddes and colleagues (14) who found that SGA's as a group were no more effective than FGA's. However, in contrast to our analysis, they also found that SGA's show a moderate benefit in EPS symptoms. The reviews by both Davis and colleagues (15)and Leucht and colleagues(16) both concluded that olanzapine and risperidone were significantly more effective than FGA's and were also associated with fewer EPS, although the Leucht review noted that the evidence was predominantly based on comparisons with the high potency drug haloperidol, and were far less clear for comparisons with low or intermediate potency FGA's. The use of perphenazine, an intermediate potency antipsychotic, may account for some of the differences in findings.

In our analysis, perphenazine was associated with a similar decrease in HS+Dp+Ak and MS+HSE, the two health states with high movement side effects, compared with the three SGA's in this study. In addition, as acknowledged by Leucht and colleagues, the effectiveness design of the CATIE and CUtLASS studies was quite different from the efficacy designs used by most of the studies included in the meta analyses, and thus more informative about real-world practice (16). Most previous studies, like many of those used in the meta analyses addressed safety and efficacy whereas the two large scale clinical trials focused more on real world effectiveness by including diverse sites and allowing far broader inclusion and fewer exclusion criteria.

Unsurprisingly, our results are closely aligned with those of the original CATIE analysis. The primary results indicated that olanzapine was associated with a longer time to discontinuation compared to risperidone and quetiapine, but not compared to perphenazine (31). There were no significant differences between any SGA and perphenazine on measures of symptoms, side effects, or quality of life. Although the results indicated similar conclusions overall, there are some differences. For example, our results showed that perphenazine and olanzapine were associated with a significantly different distribution of health states compared to risperidone. Risperidone was associated with the smallest changes in LS+LSE, HS+Ak+Dp, and HS+LSE. These changes suggested that risperidone was associated with the lowest probability of transitioning a patient to an improved health state. This data could potentially provide more concrete guidance for clinicians in their choice of treatment. Although risperidone

may be associated with some improvements as seen on average changes on various outcomes measures, our results suggested that the medication was associated with the lowest probability of significant clinical improvement as defined by the health states in this study.

An additional goal of this study was that health state analysis could be used to obtain long run predictions on the distribution of patients across health states under the different medications. In order to obtain these predictions, we tested the data for the assumptions of Markov chain theory which requires that the data fit both the stationary criteria, where the transition probabilities from one state to another would remain fixed over time, and also the assumption that the transition probabilities could depend on nothing other than the previous health state. However, since our data did not appear to meet these criteria, Markov analysis could not be conducted. Specifically when evaluating data across the entire 18 month study period, it appeared that transition probabilities were dependent on more than just the previous health state. We surmise that this finding is likely to reflect the high dropout rate over the course of 18 months during the CATIE trial. In fact, over 74% of the patients discontinued their initial study medication prior to the end of the trial. In order to test the impact of dropout on the results of our first 6 month analysis, we evaluated the sample for differences in clinical and demographic features between those who stopped their study medication at 6 months and those who stayed and found no significant differences in between the two groups. However, it was still possible that the pattern of patients dropping out changes over the

course of the 18 months and lended significant variability to the pattern of transition probability over time. Thus, in this analysis we focused only on changes in health state over the first 6 months, a clinically meaningful interval, during which drop outs were limited.

The health state approach as applied to the CATIE data has several limitations. First, the generation of clusters was completely data driven and was based on the data for the sample at hand. As a result, the observed clusters may not be generalizable to other clinical samples. We used data from a large clinical trial designed to assess differences in effectiveness of medications under real world conditions. Although the study had broad inclusion and exclusion criteria, the sample may still represent a subset of patients with chronic schizophrenia who may have failed previous treatments. A second limitation was the challenge of applying statistical tests to the cluster distributions we obtained. Although we use chi square test to determine if the distribution of health states is the same for patients receiving different treatments, we must also analyze the percent of patients within the different health states to determine if one pattern of health states is preferable to another. In addition, the health states themselves may not pick up on subtle differences in outcome. For example, our results showed that the changes in proportion of patients in LS+Ob was similar for patients treated with perphenazine, olanzapine and risperidone, but was higher for those in the quetiapine group. However, we also found that the difference in BMI from baseline to 6 months was highest for olanzapine and was significantly higher than perphenazine ($p<0.01$). These results combined with the

original CATIE trial results showing a greater increase in BMI for patients treated with olanzapine, suggest that there can be heterogeneity within a cluster that may mask some significant differences in outcome across treatment groups.

This example also illustrates the challenge of directly comparing health state analysis with outcomes of traditional dimensional methods. Because health state analysis clusters patients along various dimensions of health, it can be challenging to compare the results to traditional methods that compare patients along single dimensions of health. Finally, with health state analysis, there is an element of judgment in determining the appropriate number of health states that balances detail against simplicity. In this analysis, the final selection of number of health states was strongly influenced by need for ease of interpretability and parsimony.

Conclusion

In this study, we show that health state analysis is a useful tool in the comparison of multiple treatments for a clinically complex illness such as schizophrenia. It not only strengthens previous results, but can also provide additional insights into the differences in overall health outcomes of patients. From our analysis, we conclude that SGA's are not a homogenous group in terms of effectiveness. And as a group, they are no more superior to perphenazine, an intermediate potency FGA, which is associated with a large proportion of patients with improved symptom and side effect profiles. These results are

now added to a growing body of work indicating that SGA's as a group are not better than many FGA's on symptoms and side effects taken as a whole. The release of many new SGA's in the last decade with strikingly similar mechanisms, symptom and side effect profiles may make it more challenging for clinicians to decide on the most appropriate treatment for patients. However, with the nation's rapidly rising healthcare costs, it is imperative that clinicians look to well substantiated evidence based treatments.

References

1. Kraemer, H.C., Frank, E., and Kupfer, D.J. 2006. Moderators of treatment outcomes: clinical, research, and policy importance *JAMA* 296:1286-1289.

2. Institute of Medicine. 2007. Learning What Works Best: The Nations Need for Evidence on Comparative Effectiveness in Health Care.

3. Institute of Medicine. June 30, 2009. Initial National Priorities for Comparative Effectiveness Research.:12.

4. Elizabeth Docteur, R.B. Feb 2010. How Will Comparative Effectiveness Research Affect the Quality of Health Care? *Timely Analysis of Immediate Health Policy Issues*:1-15.

5. Tandon, R., Belmaker, R.H., Gattaz, W.F., Lopez-Ibor, J.J.,Jr, Okasha, A., Singh, B., Stein, D.J., Olie, J.P., Fleischhacker, W.W., Moeller, H.J. et al. 2008. World Psychiatric Association Pharmacopsychiatry Section statement on comparative effectiveness of antipsychotics in the treatment of schizophrenia *Schizophr. Res.* **100**:20-38.

6. Kane, J., Honigfeld, G., Singer, J., and Meltzer, H. 1988. Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine *Arch. Gen. Psychiatry* **45**:789-796.

7. Lehman, A.F., Lieberman, J.A., Dixon, L.B., McGlashan, T.H., Miller, A.L., Perkins, D.O., Kreyenbuhl, J., American Psychiatric Association, and Steering Committee on Practice Guidelines. 2004. Practice guideline for the treatment of patients with schizophrenia, second edition *Am. J. Psychiatry* **161**:1-56.

8. General Assembly of the World Psychiatric Association. Consensus Statement on the use and usefulness of second generation antipsychotic medication. **2011**.

9. Anonymous 2002/030 NICE recommends newer antipsychotic drugs as one of the first line options for schizophrenia **2011**.

10. Miller, A.L., Hall, C.S., Buchanan, R.W., Buckley, P.F., Chiles, J.A., Conley, R.R., Crismon, M.L., Ereshefsky, L., Essock, S.M., Finnerty, M. et al. 2004. The Texas Medication Algorithm Project antipsychotic algorithm for schizophrenia: 2003 update *J. Clin. Psychiatry* **65**:500-508.

11. Anonymous 1999. Treatment of schizophrenia 1999. The expert consensus guideline series *J. Clin. Psychiatry* **60 Suppl 11**:3-80.

12. Duggan, M. 2005. Do new prescription drugs pay for themselves? The case of second-generation antipsychotics *J. Health Econ.* **24**:1-31.

13. Hoenberg, K., Goetz, K. 2006. Antipsychotics: Analysis of Disease Markets and Emerging Agents.

14. Geddes, J., Freemantle, N., Harrison, P., and Bebbington, P. 2000. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis *BMJ* **321**:1371-1376.

15. Davis, J.M., Chen, N., and Glick, I.D. 2003. A meta-analysis of the efficacy of second-generation antipsychotics *Arch. Gen. Psychiatry* **60**:553-564.

16. Leucht, S., Corves, C., Arbter, D., Engel, R.R., Li, C., and Davis, J.M. 2009. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis *Lancet* **373**:31-41.

17. Adams, C.E., Coutinho, E.S., Davis, J., Duggan, L., Leucht, S., Li, C., and Tharyan, P. 2008. Cochrane Schizophrenia Group *Schizophr. Bull.* **34**:259-265.

18. Swartz, M.S., Stroup, T.S., McEvoy, J.P., Davis, S.M., Rosenheck, R.A., Keefe, R.S., Hsiao, J.K., and Lieberman, J.A. 2008. What CATIE found: results from the schizophrenia trial *Psychiatr. Serv.* **59**:500-506.

19. Rosenheck, R.A., and Sernyak, M.J. 2009. Developing a policy for second-generation antipsychotic drugs *Health. Aff. (Millwood)* **28**:w782-93.

20. Rosenheck, R.A., Leslie, D.L., Sindelar, J., Miller, E.A., Lin, H., Stroup, T.S., McEvoy, J., Davis, S.M., Keefe, R.S., Swartz, M. et al. 2006. Cost-effectiveness of second-generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia *Am. J. Psychiatry* **163**:2080-2089.

21. Tungaraza, T., and Poole, R. 2007. Influence of drug company authorship and sponsorship on drug trial outcomes *Br. J. Psychiatry* **191**:82-83.

22. Tyrer, P., and Kendall, T. 2009. The spurious advance of antipsychotic drug therapy *Lancet* **373**:4-5.

23. Swartz, M.S., Stroup, T.S., McEvoy, J.P., Davis, S.M., Rosenheck, R.A., Keefe, R.S., Hsiao, J.K., and Lieberman, J.A. 2008. What CATIE found: results from the schizophrenia trial *Psychiatr. Serv.* **59**:500-506.

24. Jauhar, S., and Research Group of the European Federation of Psychiatric Trainees. 2009. Are new drugs for schizophrenia better than old ones? *Lancet* **373**:1249; author reply 1249-50.

25. Sugar, C.A., James, G.M., Lenert, L.A., and Rosenheck, R.A. 2004. Discrete state analysis for interpretation of data from clinical trials *Med. Care* **42**:183-196.

26. Kraemer, H.C., and Frank, E. 2010. Evaluation of comparative treatment trials: assessing clinical benefits and risks for patients, rather than statistical effects on measures *JAMA* **304**:683-684.

27. Tan, Pang-Ning, Steinback, Michael, Kumar, Vipin. 2006. Cluster analysis: Basic concepts and algorithms. In *Introduction to Data Mining.* Addison-Wesley. 488.

28. Howard M. Taylor, Samuel Karlin. 1994. *An Introduction to Stochastic Modelling.* Academic Press. Boston.

29. Martha R. Gold, Joanna E. Siegel, Louise B. Russell, Milton C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine.* Oxford University Press.

30. James, G.M., Sugar, C.A., Desai, R., and Rosenheck, R.A. 2006. A comparison of outcomes among patients with schizophrenia in two mental health systems: a health state approach *Schizophr. Res.* **86**:309-320.

31. Lieberman, J.A., Stroup, T.S., McEvoy, J.P., Swartz, M.S., Rosenheck, R.A., Perkins, D.O., Keefe, R.S., Davis, S.M., Davis, C.E., Lebowitz, B.D. et al. 2005. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia *N. Engl. J. Med.* **353**:1209-1223.

32. Stroup, T.S., McEvoy, J.P., Swartz, M.S., Byerly, M.J., Glick, I.D., Canive, J.M., McGee, M.F., Simpson, G.M., Stevens, M.C., and Lieberman, J.A. 2003. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development *Schizophr. Bull.* **29**:15-31.

33. Kay, S.R., Fiszbein, A., and Opler, L.A. 1987. The positive and negative syndrome scale (PANSS) for schizophrenia *Schizophr. Bull.* **13**:261-276.

34. Addington, D., Addington, J., and Maticka-Tyndale, E. 1994. Specificity of the Calgary Depression Scale for schizophrenics *Schizophr. Res.* **11**:239-244.

35. Guy W. 1976. Abnormal Involuntary Movements. In *Assessment Manual for Psychopharamcology.* DHEW No. ADM 76-338 edition. National Institute of Mental Health. Rockville, MD.

36. Barnes, T.R. 1989. A rating scale for drug-induced akathisia *Br. J. Psychiatry* **154**:672-676.

37. Simpson, G.M., and Angus, J.W. 1970. A rating scale for extrapyramidal side effects *Acta Psychiatr. Scand. Suppl.* **212**:11-19.

38. Lehman A. 1988. A quality of life interview for the chronically mentally ill. *Evaluation and Program Planning*:51-62.

39. Ware, J.,Jr, Kosinski, M., and Keller, S.D. 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity *Med. Care* **34**:220-233.

40. Heinrichs, D.W., Hanlon, T.E., and Carpenter, W.T.,Jr. 1984. The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome *Schizophr. Bull.* **10**:388-398.

41. Benjaminni J. Hochberg T. 1995. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Bulletin*:289-300.

42. Rosenheck, R.A., Leslie, D.L., Sindelar, J., Miller, E.A., Lin, H., Stroup, T.S., McEvoy, J., Davis, S.M., Keefe, R.S., Swartz, M. et al. 2006. Cost-effectiveness of second-generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia *Am. J. Psychiatry* **163**:2080-2089.

Appendix 1

Clinical and demographic characteristics across medications of sample used for initial clustering*.

		Total Sample		Olanzapine		Perphenazine		Quetiapine		Risperidone	
		Mean/N	%/SD	Mean/N	%/SD	Mean/N	%/SD	Mean/N	%/SD	Mean/N	%/SD
		N=1049		N=263		N=256		N=261		N=269	
Age		39.3	10.9	39.36	10.56	39.97	11.06	39.15	10.88	38.78	11.05
Male		777	74.0%	190	72.2%	196	76.6%	192	73.5%	199	74.0%
Race/Ethnicity											
	White	631	60.1%	153	58.2%	151	59.0%	167	64.0%	160	59.5%
	Black	368	35.1%	96	36.5%	90	35.1%	84	32.1%	98	36.4%
	Hispanic	129	12.3%	37	14.1%	24	9.3%	39	14.9%	29	10.8%
Marital Status											
	Married	131	12.5%	30	11.4%	43	16.8%	27	10.3%	31	11.5%
	Separated/ Divorced	219	20.8%	61	23.2%	50	19.4%	55	20.9%	53	19.8%
	Never Married	636	60.6%	159	60.4%	146	57.0%	167	64.0%	164	61.0%
PANSS Total		75.5	17.5	75.7	18.2	74.2	18.0	74.8	17.0	77.2	16.5
	Positive	18.4	5.6	18.4	5.5	17.9	5.9	18.3	5.4	19.0	5.6
	Negative	20.2	6.5	20.3	6.7	20.3	6.3	19.8	6.5	20.4	6.4
	General	36.9	9.3	37.0	9.8	36.0	9.5	36.7	9.2	37.8	8.6
	Depression	1.6	0.6	1.6	0.6	1.6	0.6	1.6	0.6	1.6	0.6
Side Effects											
	Simpson Angus	0.18	0.29	0.16	0.27	0.19	0.32	0.16	0.25	0.20	0.29
	Barnes	0.47	0.84	0.58	0.97	0.43	0.79	0.46	0.76	0.47	0.81
	AIMS	0.12	0.27	0.15	0.31	0.12	0.27	0.11	0.24	0.12	0.23
BMI		29.80	7.09	29.24	6.86	29.63	6.93	30.22	7.05	30.09	7.48
Health costs (previous month)											
	All medication	$422	$325	$419	$344	$420	$314	$418	$331	$433	$313
	Inpatient/ Residential	$1,512	$3,715	$1,828	$3,988	$1,127	$2,530	$1,442	$3,642	$1,636	$4,381
	Outpatient	$365	$935	$379	$864	$392	$1,173	$410	$1,066	$281	$513
	Total	$2,299	$3,831	$2,628	$4,078	$1,940	$2,811	$2,271	$3,813	$2,352	$4,389

* Data courtesy of Supplemental Table 1 of (42)