Discrete state analysis for interpretation of data from clinical trials

Catherine A. Sugar, Ph.D.
Marshall School of Business, University of Southern California

Gareth M. James, Ph.D.
Marshall School of Business, University of Southern California

Leslie A. Lenert, M.D.
Staff Researcher and Physician, Veterans Health Administration San Diego Healthcare System

Robert A. Rosenheck, M.D.
Departments of Psychiatry and Epidemiology and Public Health, Yale Medical School and VA Northeast Program Evaluation Center, West Haven, CT

Correspondence Address: Catherine A. Sugar
University of Southern California
Bridge Hall 400
Los Angeles, California 90089-0809
Tel: (213) 740 7957
Fax: (213) 740 7313
Email: sugar@usc.edu

Acknowledgments: This work was partially funded by the NIMH program in Clinical Antipsychotic Trials of Intervention Effectiveness in Schizophrenia and Alzheimer's Disease (CATIE) (N01-MH9001)(J. Lieberman, PI).

Running Title: Discrete state analysis of clinical data

Word Count: 5180

Complete Author Information

Catherine A. Sugar, Ph.D.

University of Southern California
Bridge Hall 400
Los Angeles, California 90089-0809
Tel: (213) 740 7957
Fax: (213) 740 7313
Email: sugar@usc.edu
Expertise: Cluster analysis, functional data analysis, multivariate statistics, biostatistics

Gareth M. James, Ph.D.

University of Southern California
Bridge Hall 401 P
Los Angeles, California 90089-0809
Tel: (213) 740 9696
Fax: (213) 740 7313
Email: gareth@usc.edu
Expertise: Functional data analysis, classification, cluster analysis

Leslie A. Lenert, M.D., M.S.

HSRD Section, MC 111N1
VA San Diego HCS
3350 La Jolla Village Dr.
San Diego, CA 92161
Tel: (858) 552 4325
Fax: (858) 552 4321
Email: llenert@ucsd.edu
Expertise: Medical informatics, utility theory, decision support, preference measurement

Robert A. Rosenheck, M.D.

Director, Northeast Program Evaluation Center
Professor of Psychiatry and Public Health, Yale Medical School
VA Connecticut Healthcare System
950 Campbell Ave.
West Haven, CT 06516
Tel: (203) 937-3850
Fax: (203) 937-3433
Email: robert.rosenheck@yale.edu
Expertise: Health services research, cost-effectiveness analysis, psychiatry, mental health, service use

Discrete state analysis for interpretation of data from clinical trials

Abstract (250 words)

Objective: To demonstrate a multivariate health state approach to analyzing complex disease data that allows projection of long-term outcomes using clustering, Markov modeling, and preference weights.

Subjects: Patients hospitalized 30-364 days with refractory schizophrenia at 15 Veterans Affairs medical centers.

Study Design: Randomized clinical trial comparing clozapine, an atypical antipsychotic and haloperidol, a conventional antipsychotic.

Methods: Health status instruments measuring disease-related symptoms and drug side-effects were administered in face-to-face interviews at baseline, 6 weeks, and quarterly follow-up intervals for one year. Cost data were derived from Veterans Affairs records, supplemented by interviews. K-means clustering was used to identify a small number of health states for each instrument. Markov modeling was used to estimate long-term outcomes.

Results: Multivariate models with 7 and 6 states, respectively, were required to describe patterns of psychiatric symptoms and side effects (movement disorders). Clozapine increased the proportion of clients in states characterized by mild psychiatric symptoms and decreased the proportion with severe positive symptoms, but showed no long-term benefit for negative symptoms. Clozapine dramatically increased the proportion of patients with no movement side effects and decreased incidences of mild akathesia. Effects on extrapyramidal symptoms and tardive dyskinesia were far less pronounced and slower to develop. Markov modeling confirms the consistency of these findings.

Conclusions: Analyzing complex disease data using multivariate health state models allows a richer understanding of trial effects and projection of long-term outcomes. While clozapine generates substantially fewer side effects than haloperidol, its impact on psychiatric aspects of schizophrenia is less robust and primarily involves positive symptoms.

Keywords: Health state models, cost-benefit analysis, longitudinal studies, cluster analysis, schizophrenia.

1. Introduction

In clinical trials different aspects of physical and psychological health typically are measured using both disease specific and more general health status instruments consisting of dozens of item responses. Multiple items may be summarized by combining them into continuous scales reflecting symptomology, functioning or quality of life. Such composite measures can be analyzed using a variety of standard univariate statistical techniques. When evaluating complex diseases, such methods are frequently applied to several composite scores. However, this approach potentially ignores important interrelationships between different dimensions of health. In this study we develop a multivariate health state modeling approach to the analysis of complex clinical trials which seeks to harness more of the inherent structural richness of such data. The patient population is partitioned into a set of health states via cluster analysis, rather than by the factorial design traditionally used in health index models such as the Health Utilities Index **[1,2]**, the EQ-5D **[3,4]**, and the Quality of Well Being Scale **[5,6]**. It is desirable for patients in the same health state to be as similar as possible or equivalently to have as little variability as possible over the dimensions of health that describe the patient

population. Clustering allows the data to choose the optimal locations of the health states. As a result, the clinical status of a patient population can often be as accurately represented, in terms of within group variability, with many fewer states than a comparable factorial design. In addition because it is data driven clustering is particularly well suited to capturing complex interrelationships. Clinical change is not measured in terms of a simple net increase or decrease in the mean on a preset continuous scale. Instead, the effects of a medication are assessed in terms of its probability of moving individuals from any given health state to another, over time. A treatment's benefit for patients from a given cluster is greater if it has a higher probability of moving them to a superior state. Naturally, the data driven nature of clustering means that one must be careful to check whether the resulting health state models still apply when generalizing them to new populations.

Health state models have several additional advantages. One is that they provide a natural way to estimate the long term effectiveness of treatments using data from clinical trials which are necessarily of finite duration. Results from Markov chain theory allow one to calculate the long run fraction of individuals residing in each health state for each treatment group and thus compare the effectiveness of different medications. As with any approach that involves extrapolation beyond the study period, results are based on the assumption that the treatments and patterns observed during the trial will continue indefinitely. Another advantage of health state models is that they facilitate utility estimation. It is relatively straightforward to generate descriptions of the prototypical patients in each state and survey both the general population and patients to estimate the

corresponding utilities. These preference weights can be used to express the results of a trial in terms of changes in Quality Adjusted Life Years. QALY scores can be combined with financial data and long-run distributions to assess the efficiency of investments in health at a societal level.

In this paper, we use health state modeling to perform a secondary analysis of data from a comprehensive double-blind trial **[7]** conducted at 15 Veterans Affairs (VA) medical centers comparing haloperidol (HALDOL, Ortho-McNeil Pharmaceuticals, Spring House, PA) and clozapine (CLOZARIL, Novartis Pharmaceuticals Corporation, East Hanover, NJ), two medications for treating schizophrenia. Clozapine was the first of a class of new, more effective, medications referred to as "atypical antipsychotics" because of their distinctive lack of movement side effects and has shown special promise in the treatment of patients with refractory schizophrenia **[8]**. The 12 month study in **[7]** provided the first comprehensive assessment of the impact of clozapine on social, vocational and community functioning and societal costs, in addition to measuring traditional clinical factors such as side effects, positive and negative symptoms, and general psychological distress. The initial presentation of results was based on univariate comparisons of means for a handful of scales. While this analysis provided an easily interpretable overall assessment it did not take into account complex interactions among the scales. Further, the lack of discrete health states made it difficult to elicit utilities or assess the long term effects of each medication. In this study we apply a health state model to the same data set to achieve all of these objectives.

2. Methods

2.1 Data

In this paper we extend the analysis of the cohort from **[7]** which consisted of 423 patients treated at 15 veterans health centers around the country. Within each center patients were randomized to receive clozapine or haloperidol. The data consisted mainly of scores on standard health status instruments measuring a broad spectrum of emotional, interpersonal, and physical functioning. Our analysis focuses on 2 areas, mental health and extra-pyramidal medication side effects. For the first we use the Positive and Negative Syndrome Scale (PANSS) **[9]**. This instrument has 3 subsections, **positive symptoms** such as hallucinations, delusions and hostility, **negative symptoms** such as blunted affect, withdrawal, passivity, and difficulty in abstract thinking, and **general emotional disturbances** such as anxiety, depression and guilt. To assess extra-pyramidal side effects, we combined items from 3 commonly used instruments, the Abnormal Involuntary Movement Scale (AIMS) which measures tardive dyskinesia i.e. unconscious movements, **[10]**; the Barnes Akathesia Scale (BAS) which focuses on involuntary restlessness **[11]**; and the Simpson-Angus Scale (SAS) which deals with syndromes of pseudo-parkinsomism, involuntary tremors and stiffness of muscles, and salivation **[12]**. All these instruments use Likert scales to measure severity of symptoms with higher scores indicating more severe impairment.

Data were collected by trained research assistants at 6 time-points (baseline, 6 weeks and 3, 6, 9, and 12 months) and were available for 87% of planned follow up observations.

Because patients tended to lack complete questionnaires rather than answers to single questions, we eliminated from further study any patient-time combination with missing data. During the study some subjects responded poorly to a medication and changed to an alternative treatment. Patients who switched from haloperidol to clozapine (n=49 [22%]) were treated as members of the control group before they changed medications and members of the treatment group afterwards. Crossovers from clozapine to haloperidol, or to another conventional medication, (n=83 [40%]) were handled analogously. Subjects who went off all medications or switched to a third form of treatment (n=157 [37% overall]) were analyzed on an intent to treat basis, meaning that they remained in the group to which they were originally assigned. In addition there was evidence of significant differences in ratings among the 15 study sites. We fit mixed effects models for each question using patient response as the dependent variable and time, treatment and study site as independent variables and subtracted off the estimated site effects. This made the responses comparable across sites. Further details concerning the study population, study and services delivered can be found in **[7]**.

2.2 Identifying dimensions of health

Clustering raw questionnaire data usually produces very unstable health states because of the large number of items. Dimension reduction techniques allow one to capture most of the important information in an instrument, while eliminating much of the variability. Hence, a critical first step in constructing any health state model is to identify a small set of variables or dimensions of health that captures the information necessary to differentiate among members of the population of interest. One standard approach is to

perform univariate analyses based on summary statistics. In our study, this might consist of the total PANSS score and a composite measure of side-effect severity obtained by combining the AIMS, SAS and BAS. Although a total score provides an easily interpretable overview of the data it is not necessarily the only or even the most important characteristic of health captured by a particular questionnaire. Previous studies have shown that the instruments used here measure multiple dimensions of health. Our choice of appropriate composite scores was further complicated by the fact that refractory patients differ substantially from the general population of those with schizophrenia. We used principal components analysis **[13]** to identify a small number of dimensions that capture the important information in the PANSS and side effects scales. We included all components for which the proportion of variance explained was higher than the average variance per dimension.

2.3 Forming the health states

Next we derived a final health state model using the variables resulting from the principal components analysis. The traditional approach has been to construct a factorial design in which each variable or dimension of health is divided into evenly spaced levels forming a grid. The resulting hyper-rectangles correspond to health states and their Euclidean centers represent the prototypical patients in those states. There are several problems with such a design. First, for even a small number of variables it produces a large number of health states. More importantly, there is no *a priori* reason why the natural groupings of patients should follow a symmetric grid. Thus a factorial design results in a large number of health states that are either empty or are poorly centered around their "typical" patient.

As a result, many disease specific symptoms and consequences of treatment may be missed or ignored **[14]**.

Instead, we used k-means cluster analysis **[15]** to construct a parsimonious and data-driven collection of health states. The k-means algorithm works by partitioning the data space into non-empty, non-overlapping regions so that points in the same cluster are close together while those in different clusters are as widely separated as possible. Specifically, for a given number of clusters, k, the algorithm finds the set of centroids that minimizes the *distortion* or sum of squared distances between each observation and its closest cluster center. This approach is generally extremely efficient, requiring many fewer health states to adequately differentiate the members of the patient population. The cluster centroids are much more representative of prototypical patients because they have been defined as the means of the data within their respective states. Finally, because the cluster analysis approach is data driven, the resulting health states can be asymmetrically shaped, capturing important interactions among the characteristics that define the population.

There are a number of technical issues to consider when using cluster analysis to develop a health state model including preprocessing and scaling of the data, and initialization of the clustering algorithm. A more detailed discussion of these points is provided in **[16]**. Since the health status instruments used in this study all had items measured on comparable scales and the observations were spread fairly uniformly in the data space, none of these issues presented a serious problem here. The most important decision was

choosing the number of clusters to fit to the data. Clustering will always partition a data space into mathematically non-overlapping sets. However, it is important that enough clusters are used so that medically distinct patients are not grouped together producing compromise health states. Statistical methods based on distortion can be used to identify the number of groups in a data set **[17]**. However, such techniques must usually be combined with contextual information to ensure that the model is sufficiently parsimonious for practical use in cost-effectiveness analyses while remaining sensitive to important clinical differences between patient groups.

For this data set, statistical techniques did not provide a definitive indication of the number of clusters. Thus we developed several graphical tools which experts can use to choose the medically optimal number of health states. The first approach, involves using cluster mean plots to examine the distribution of centers that are formed for varying values of k. In general it is worth adding additional clusters when doing so identifies a new health state that is clinically distinct in terms of one or more important dimensions of health. Visually this corresponds to separation of the cluster centers along at least one of the principal component axes. Cluster profile plots provide another useful tool by allowing one to visually asses the characteristics of a given health state. For each cluster one plots the average score for each questionnaire item among all patients falling in that health state. It is worth adding additional clusters as long as they have clinically distinct profiles. Note that once the optimal number of clusters has been chosen, cluster profile plots can be used to easily create objective health state descriptions.

After choosing the number of health states and removing patients with missing values we ran the k-means algorithm separately for the PANSS and side-effects data to produce the health state models. The observations for all subjects and time combinations were used in the clustering. This ensured that the full range of states encountered in all phases of treatment was included in the model.

2.4 Longitudinal Analyses

Next we used the cluster generated health state models to analyze differences between patients on haloperidol and clozapine. First we examined the data cross-sectionally checking whether patients had different patterns of health state membership by performing chi-square tests of independence between medication and health state for the PANSS and side effects scales at each of the six time points. Note that the assignments depend on the estimated health state model. Since there is uncertainty and variability in the estimation process this causes a dependence between observations which violates an assumption of the chi-square test. In principle, one could use an appropriate bootstrap resampling technique to more precisely estimate the null distribution of the test statistic and the associated p-values. However, the dependence is low so in this study that this is unlikely to significantly alter the results.

Next we examined long run differences between the medications for the study population. Provided that patients' probabilities of movement from one state to another remain fixed over time one can calculate a stationary distribution which is simply the

fraction of patients that will reside in each health state once a state of equilibrium is reached. We estimated separate stationary distributions for the clozapine and haloperidol groups and checked for long run differences between them by using a permutation test **[18]** (Chapter 15) analogous to a chi-squared test that we developed for this problem. Details of the calculations are provided in Appendix A.

Finally, we compared the long-run costs of treating patients with clozapine versus haloperidol. Yearly total societal costs, including inpatient and outpatient psychiatric, substance abuse and medical-surgical services; study-related medication costs; and non-health costs related to criminal justice involvement, transfer payments, and employment productivity (a negative cost) were available for each patient **[7]**. However, since patients occupied multiple states during the study, direct measurements of health state costs were unavailable. Instead we fit a multiple linear regression using the number of weeks patients spent in each of the PANSS health states as the predictors (without an intercept), and cost as the response. The fitted coefficients provided estimates of the weekly cost of maintaining a patient in a given health state. The approximately 24% of patients with missing observations were removed. Ideally, utilities or QALY weights would be obtained by having subjects rate the health states identified in this study. However, since resources did not allow for this approach we instead mapped our states onto ones obtained using similar methods **[19, 20]** for which utilities had been measured via the standard gamble approach in a general public sample, as recommended by **[21]**. After calculating the cost and QALY for an average patient in each health state we computed long run weighted averages for the two medications based on the previously obtained

stationary distributions. Finally, permutation tests were used to check whether the differences in costs and QALYs were statistically significant.

3 Results

3.1 Dimensions of health

Principal components analysis identified 5 important dimensions of health on the PANSS and 4 on the side effects scale. The PANSS components measure, in order of variability explained, overall mental health, a contrast between negative and positive symptoms, subjective emotional distress, hostility, and thought disturbances. Previous studies using general populations of schizophrenic patients have also found 5 important dimensions in the PANSS. The side effects components represent overall severity, a contrast between akathesia and tardive dyskinesia, extrapyramidal syndromes excluding akathesia, and a contrast between facial and extremity movements. Interestingly, these components break down largely along the 3 questionnaires that make up the side effects scale, indicating that each instrument captures different information. For detailed descriptions of all the components see Appendix B.

3.2 The health state models

Using cluster mean and profile plots led us to select models with 7 multidimensional health states for the PANSS and 6 for the side effects scale. For example, Figure 1 gives the cluster mean plots corresponding to fitting 4 and 6 state models to the side effects data. For ease of comparison, the clusters are labeled according to their score on the first

principal component. Note that in the 4 cluster model the third principal component does not differentiate among any of the groups but in the 6 cluster model it clearly separates out cluster 4. This suggests that 4 clusters is too few to capture variability in the patient population. Further analyses showed no significant additional differentiation in the means beyond 6 clusters. Figure 2 shows the corresponding profile plots for a 6 cluster fit to the side effects data. For each cluster we have plotted the average score for each questionnaire item among all patients falling in that health state. The item scores have been centered by subtracting off their global means. For example, cluster 1 has very low averages over all questions, indicating that patients have relatively few side effects. In contrast patients in cluster 6 have severe disorders. Note that cluster 4, which was well separated in the third principal component on the cluster mean plot, has the worst average scores among all states on the Simpson-Angus scale, which measures bodily rigidity and tremors. The profile plots show clear distinctions among all the clusters which suggests that there are at least 6 medically interpretable states of side effects health for this population. However, no further differentiation was achieved when additional clusters were added, indicating that 6 is the optimal number of side effects states. Similar analyses were performed for the PANSS.

Once the optimal number of clusters has been chosen, summary statistics can be combined with profile plots to easily create objective health state descriptions. The mean principal component scores and standard deviations for patients in each cluster of the final health state models are presented in Table 1. The PANSS states can be characterized as: 1) mild symptoms on the entire PANSS (MS), 2) moderate symptoms across the

PANSS but with high global subjective distress (MS+HGSD), 3) moderate symptoms across the PANSS but with high grandiosity (MS+HG), 4) severe negative symptoms with low subjective distress (SNS+LSD), 5) severe negative symptoms with high subjective distress (SNS+HSD), 6) severe positive symptoms (SPS), and 7) low subjective distress but severe symptoms on all other questions (SS+LSD). The side effects states can be characterized as: 1) no side effects (NSE), 2) mild tardive dyskinesia (MTD), 3) mild akathesia (MA), 4) extra-pyramidal symptoms (EPS), 5) frank tardive dyskinesia (FTD), and 6) severe tardive dyskinesia and akathesia (STD+SA). See Appendix C for detailed health state descriptions

3.3 Health state transitions and long-term outcomes

A primary purpose of producing the health state models was to examine differences between patients on haloperidol and clozapine. We first performed a cross-sectional analysis to check whether patients on the two medications have different distributions across the health states and if so at what times. Tables 2 (PANSS) and 3 (side effects) give these distributions along with p-values for chi-squared tests of independence between medication and health state at baseline and each follow up period. As one would hope, there are no statistically significant differences between the medications at baseline. However, the side effects scale shows differences that are both highly significant and increasing for all follow up periods. For instance, at all time points after baseline there are over 20% more clozapine than haloperidol patients in state 1 (NSE), corresponding to no side effect problems, and at all time points 3 months or more after baseline there are approximately 10% fewer clozapine patients in state 3 (MA). Overall,

effects for the PANSS are more delayed with statistically significant differences occurring only at 6 and 9 months. However, tests only comparing the medications for differences in states 1 and 7, which are respectively the best and worst health states as measured by overall severity of symptoms, show significant differences at the 5% level for all follow up periods. This suggests that clozapine does a better job of moving patients out of the worst and into the best health states but that there is insufficient power to detect this effect when all 7 states are considered simultaneously. A closer examination of Table 2 reveals that clozapine's greatest impact is on the reduction of positive symptoms. From 6 weeks on there are consistently more clozapine than haloperidol patients (up to 7.5% difference) in state 1 (MS). In contrast, there are on average nearly 8% fewer clozapine patients in state 6 (SPS) over the same period, and there is a much smaller but consistent difference for state 7 (SS+LSD). Interestingly and unexpectedly late in the trial as many as 7% more clozapine patients are found in state 4 (SNS+LSD). There is no evidence of consistent differences between groups in the remaining states.

Next we considered transition probabilities which simply give the likelihood of a patient moving from any one health state to another during a 3 month period. Transition probabilities for the PANSS and side effects health states on each medication are given in Tables 4 though 7. They suggest that patients on clozapine are much more likely to stay in the best side effects health states. For example, a clozapine patient in state 1 (NSE) has an 80.2% chance of remaining there compared to only 61.2% for a haloperidol patient.

This finding is corroborated by the stationary distribution shown in Table 8 which estimates that, for the side effects scale, in the long-run 61% of patients on clozapine will reside in state 1 (NSE) compared to only 38% of those on haloperidol. The long run analysis also suggests a notable 6% difference in the prevalence of state 6 (STD+SA). The p-value for a difference in the long run distributions between medications on the side effects scale is less than 0.001, providing extremely strong evidence of the superiority of clozapine.

There is also some evidence that clozapine is more efficacious as measured by the PANSS, although this finding is far less robust. In the long-run analysis, about 4% more clozapine than haloperidol patients end up in state 1 (MS), mild symptoms, and 12% fewer in states, 6 (SPS) and 7 (SS+LSD), severe positive symptoms and severe symptoms with low subjective distress. Surprisingly, 6% more clozapine patients are predicted to be in state 4 (SNS+LSD), severe negative symptoms with low subjective distress. However, these findings could be artificial since the p-value for a difference between medications on the PANSS scale is 12.1%. Finally note, by comparing Tables 2 and 3 with Table 8, that the distributions at the 12 month follow-up are very similar to the stationary distributions, suggesting that the population had almost reached equilibrium by the end of the study.

Overall, there is clear evidence that clozapine has added benefits for otherwise refractory patients. Clozapine is a considerably more expensive drug than haloperidol, but medication costs make up only a fraction of society's financial burden from treating

patients with schizophrenia. If clozapine reduces expenses from other sources, such as hospitalization and lost earning potential, it may be more cost-effective overall. To address this issue we estimated the average total cost of care for patients in each of the 7 PANSS health states. We then computed weighted average costs for the 2 medications for the 6 study periods and in the long-run. The weekly cost for a patient in each PANSS health state is shown in the first row of Table 9. As one might expect, the better states have lower health costs. For example, patients in state 7 (SS+LSD) have yearly expenses of $84,240 compared to only $51,844 for those in state 1 (MS). Table 10 gives annualized costs for patients on each medication calculated both using the stationary distributions and by extrapolating from each of the observed time periods. For both treatment groups costs are declining over time and appear close to equilibrium by the end of the study. However, those for patients on clozapine are consistently slightly lower than for those on haloperidol with a test for a significant long-run difference yielding a p-value of 4%. This suggests that even though clozapine is more expensive up front, in the long run it may actually result in lower overall health costs as well as improved quality of life. These results are consistent with those of previous studies **[7]**. A similar analysis was performed for the side effects health states but it was found that they had no significant correlation with costs.

Finally we consider the long run difference in utility levels or QALYs for patients on the two medications. The second row of Table 9 gives QALY scores for each PANSS health state. For instance, the score for state 7 (SS+LSD) is less than half that of state 1(MS), meaning that a patient would prefer to live half a year with mild symptoms than a full

year with severe symptoms. Using the previously obtained stationary distributions we find an average long run QALY value of 0.733 for patients on clozapine and 0.716 for those on haloperidol. This small but clinically meaningful difference is similar in magnitude to that found by **[22]** using a more ad hoc measure of utility. However, the difference is not statistically significant.

4. Discussion

Health state models have several distinct advantages over traditional univariate approaches to analyzing clinical trial data for complex diseases such as schizophrenia. First, they give a parsimonious multivariate representation of the population. For instance, the model created for this study involves 7 PANSS and 6 movement clusters for a total of 42 discrete health states. Even assuming a minimal 3 dimensions of health for both the PANSS and side effects scale and 3 levels in each dimension, a traditional full factorial design would require 729 health states, although they would not all necessarily be occupied. This parsimony is critical for subsequent phases of analysis since if the number of states is too great even an extremely large trial would provide insufficient data to estimate the quantities of interest. Second, health state models provide a convenient framework for performing longitudinal analyses. One can estimate the long run fraction of people in each health state in addition to the cross-sectional distributions of patients during the study period. By performing such analyses on treatment subgroups it is easy to compare the benefits of different medications in both the short and long term. The partitioning of the population into health states leads to a more richly informative analysis of the efficacy of treatments than a more standard univariate approach. For

example, one can find that a medication works well for the average patient but still leaves a high percentage of people in "undesirable" states. Finally, stationary distributions can be combined with a wide variety of outcome variables, such as costs or QALYs, to calculate the long-run effects of treatments. While care is always required when extrapolating beyond the range of the data, this approach allows one to make objective long term health policy decisions by balancing treatment effectiveness against societal costs on a quantitative basis.

From our study there is clear evidence that clozapine significantly reduces extrapyramidal side effects, particularly akathesia, in both the immediate and long term, compared to haloperidol Differences on the PANSS scale are less dramatic and slower to develop. There is evidence to suggest that, in the long-run, clozapine reduces severe positive symptoms, but it appears to have little differential effect on negative symptoms. We also found that clozapine may produce a small but clinically meaningful long-run improvement in QALYs compared with haloperidol. Finally, despite its initial expense, treatment with clozapine results in lower net costs to society.

In this analysis, we elected to produce separate health states for the PANSS and side effects scales instead of combining them into a single model. If there had been a strong relationship between the PANSS and side effects health state memberships one potentially could have produced a unified model with fewer than the 42 health states that we used. However, since the conditional distributions of PANSS state membership given side effects state were all similar forming a single health state model over both scales

would have involved producing on the order of 40 clusters. The corresponding health state descriptions would not have been easily interpretable, nor was there sufficient data to accurately estimate the associated transition probabilities, stationary distributions and costs. Thus we opted to fit the scales separately which only required 13 distinct clusters. In general, we recommend initially fitting scales separately and examining the conditional distributions of state memberships across one scale given different membership combinations on the others. Substantial differences in these distributions provide evidence of interaction, suggesting that the scales should be clustered jointly. Whether such an approach is worthwhile depends on the availability of sufficient data and whether the interaction effects one would capture are clinically meaningful.

4.1 Limitations

The approach described in this paper has several limitations. First, since clustering is data driven, the resulting models may not generalize easily to other populations. For instance, the health state model derived for refractory VA patients may not apply well to the general population of patients with schizophrenia. A second limitation arises from the difficulty of determining the appropriate number of clusters. This generally requires some degree of clinical judgment, and therefore introduces a subjective component to the analysis. Of course, this is also the case with traditional factorial designs. Thirdly, extrapolation of the trial results using stationary distributions requires assuming that the health care processes operating during the trial will continue indefinitely. Although this seems reasonable, it may not take into account patient mortality or disease progression. In addition, it may be difficult to identify violations of the Markovian assumption using a

relatively small number of time points or subjects. Fourthly, since the utilities used in our outcome analysis were obtained from a different study we cannot be sure they map perfectly to our health states. This could be remedied in a future study by having subjects rate the health states we identified. Finally, the data have several limitations including the fact that a large number of patients switched medications or were missing financial data. Additionally, at the time of the original study **[7],** some of the side effects of clozapine such as weight gain and hypoglycemia were not properly recognized and hence not measured.

4.2 Conclusions

A discrete state, multi-dimensional approach to data analysis has a number of advantages in interpretation of clinical trial data. It allows a richer understanding of treatment effects, and the projection of long-run outcomes. In addition, health state modeling provides a simple framework for elicitations and facilitates the application of cost-effectiveness analyses.

Acknowledgments

This work was partially funded by the NIMH program in Clinical Antipsychotic Trials of Intervention Effectiveness in Schizophrenia and Alzheimer's Disease (CATIE) (N01-MH9001)(J. Lieberman, PI). The original study on which this work was based was VA Cooperative Study in Health Services number 17, "The Clinical and Economic Impact of Clozapine on Refractory Schizophrenia", funded by VA research services.

Appendix

A. Technical Appendix : **Calculating the limiting distribution**

A limiting stationary distribution is guaranteed to exist provided the transitions between health states are Markovian **[23]**, meaning that the probability of moving from one's current state to any other state depends only on the current state and not on what states one has occupied in the past. Markovian data, can be summarized using its transition matrix, P. The (i,j)th entry of a transition matrix is the probability that an individual currently in State i will move to State j in the next time period. Standard Markov theory shows that the stationary distribution is simply the first eigenvector of P^T after normalizing the eigenvector so that its entries sum to 1 **[23]**. To verify that our transition data were consistent with the assumption of a Markovian structure we compared estimated transition matrices over different time periods. For example, one can compute the matrix of transitions from the start to the end of the first quarter and compare this to the transitions during the fourth quarter. If the structure is Markovian these two transition matrices should be equal up to errors in the estimates. To test for differences in the transition matrices we calculated a pooled estimate of the transition probabilities, generated random data for the first quarter and the last quarter according to these probabilities, calculated two new estimated transition matrices from the new data and recorded the sum of squared deviations of each entry in the first matrix from the corresponding entry in the second matrix. This procedure was repeated 200 times and these deviations compared to that from the originally observed data. The fraction of deviations that were larger than that for the original data provided a p-value for this hypothesis test. If the data were not Markovian one would expect to find a large observed

difference in the transition matrices between the first and last time periods and hence a small p-value. However, the p-value was far greater than 10% indicating that there was no evidence that that data was not Markovian. This is an example of a bootstrap resampling technique **[18]**.

Hence, we created estimated transition matrices for each medication by combining all movements from one state to another between 3 months and 6 months, 6 months and 9 months and 9 months and 12 months. Patients that changed drugs during one of these time intervals were not used for that period. From the transition matrices, we were able to estimate final stationary distributions for patients on clozapine and for patients on haloperidol. We then tested for long run differences between treatments by using a permutation test analogous to a chi-squared test. Specifically, we randomly permuted the treatment variable, recalculated the transition matrices and stationary distributions, and computed the sum of squared differences between the probabilities for the two stationary distributions. This procedure was repeated 1000 times to simulate the null distribution corresponding to no difference between medications and empirical critical points were used to determine p-values. An almost identical procedure was used to test for differences between long run costs for treatments except that the stationary distributions were multiplied by the estimated health state costs and summed. The same procedure was used to test for differences in QALYs.

B. Interpretation of principal components

B.1 Side effects

PC 1 : The first component is roughly an average of all the side effects questions with somewhat less emphasis on the Simpson-Angus instrument than the other two scales. It measures overall the degree to which a patient experiences side effects problems. High positive scores mean severe problems. This component explains 31% of the variability in the data.

PC 2 : The second component is a contrast between the akathesia and AIMS scales. It puts positive weights on the akathesia questions and negative weights on the AIMS. A high positive score means severe akathesia problems but low tardive dyskinesia and vice versa. This component explains 13% of the variability in the data.

PC 3 : The third component separates out the Simpson-Angus scale with the exception of the akathesia and salivation questions. High negative scores mean problems with extra-pyramidal syndromes such as gait, rigidity, tremor and salivation. This component explains 8% of the variability in the data.

PC 4 : The final component focuses on the AIMS scale. It seems to be largely a contrast between facial/oral movements (which get negative scores) and the other questions, especially those about the extremities, which get positive scores. The other two scales have little weight. High positive scores mean problems with extremity movements and high negative scores mean problems with facial movements. This component explains 6% of the variability in the data.

B.2 PANSS

PC 1 : The first component is fundamentally an average although lower weights are put on some of the general emotional concerns questions such as depression and anxiety. High positive scores indicate severe problems. This component explains 23% of the variability in the data.

PC 2 : The second component is a contrast between positive and negative symptoms. High positive scores indicate problems with positive but not negative symptoms. High negative scores mean the reverse. This component explains 11% of the variability in the data.

PC 3 : The third component is a mixture of positive and negative weights on several questions. However, the questions about depression, anxiety, guilt and somatic concern are significantly more negative. High negative scores on this component indicate the patient has problems with general negative feelings. This component explains 8% of the variability in the data.

PC 4 : The fourth component measures hostility. Excitement, hostility, tension, un-cooperativeness, and poor impulse control all get higher positive weights, so high positive scores correspond to greater hostility. This component explains 6% of the variability in the data.

PC 5 : The final component corresponds to thought disturbances. High negative weights are put on questions like conceptual disorganization, problems with abstract thinking, lack of judgment and so forth. This component explains 5% of the variability in the data.

C. Interpretation of health states

C.1 Side effects

Health state 1 is the best and state 6 the worst in terms of overall severity of extrapyramidal side effects, although state 5 is also fairly bad. States 3 and 6 correspond to akathesia problems while patients in state 5 have problems with abnormal involuntary movements. Finally, state 4 corresponds to problems on the Simpson-Angus scale.

State 1. **No side effects problems (NSE).** These people are below average on all the side effects questions so they are relatively speaking in good shape. Typical average scores per question are around 0.25 to 0.5.

State 2. **Mild tardive dyskinesia (MTD).** These people have worse scores than average on the AIMS, average scores on the Simpson-Angus, and better than average scores on the akathesia questions. Questions on the AIMS average close to 1.

State 3. **Mild akathesia (MA).** These people are average or slightly better than average on all questions except the akathesia scale where they are markedly worse than average. Typical scores on the akathesia questions range from 1 to 1.5.

State 4. **Extra-pyramidal symptoms (EPS).** These people are right on average in every area except the first eight Simpson-Angus questions on which they are significantly worse than average. The typical scores on the Simpson-Angus questions range from 1.5 to 2.

State 5. **Frank tardive dyskinesia (FTD).** These people are worse than average on most questions but only really strongly so on the AIMS where their average scores range from 1.5 all the way to 3.

State 6. **Servere tardive dyskinesia and severe akathesia (STD+SA).** These people fare poorly across the board on the side effects questions, with particularly severe akathesia problems and moderately severe AIMS, although the AIMS is not as bad as state 5. Typical akathesia scores average around 2.

C.2 PANSS

Health state 1 is the best and state 7 the worst overall as measured by total PANSS score. States 4, 5, and 7 correspond to negative values on the second principal component, meaning more negative symptoms than positive. The other states have the reverse pattern. States 2 and 5 correspond to high negative scores in the third principal component meaning significant problems with depression and other indicators of subjective distress.

State 1. **Mild symptoms (MS).** These people have better than average scores on all PANSS questions. Typical question scores are around 2.

State 2. **Moderate symptoms and high global subjective distress (MS+HGSD).** These people are better than average on most questions except that they have higher than average levels of anxiety, depression and other general emotional disturbances.

State 3. **Moderate symptoms with high grandiosity (MS+HG).** These people are worse than average on positive symptoms and slightly better than average on

other questions. Typical scores on most of the positive symptom questions are from 3 to 4.

State 4. **Severe negative symptoms with low subjective distress (SNS+LSD).** This state is the reverse of state 3. The patients are better than average on positive symptoms and depression related issues and worse than average on negative symptoms and some of the general questions on similar topics.

State 5. **Severe negative symptoms with high subjective distress (SNS+HSD).** These people have about average positive symptoms, and are worse than average on negative symptoms and depression. They are similar to state 4 with the addition of depressive problems.

State 6. **Severe positive symptoms (SPS).** These people have severe positive symptoms, are average on negative symptoms, and have moderately bad problems across the board on the general symptoms.

State 7. **Severe symptoms with low subjective distress (SS+LSD).** These people have severe impairments on all items except those related to the depression and anxiety on which they are roughly average.

References

1. Feeny D, Furlong W, et al. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics. 1995; 7: 490-502.
2. Feeny D, Furlong W, et al. Multiattribute approach to the assessment of health-related quality of life: Health Utilities Index. Med Pediatr Oncol. 1998; Suppl 1: 54-59.

3. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997; 35: 1095-1108.
4. Kind P, Dolan P, et al.Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ. 1998; 316: 736-741.
5. Andresen EM, Rothenberg BM, et al. (1998). Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. Med Care. 1998; 36:1349-1360.
6. Kaplan RM and Anderson JP. A general health policy model: update and applications. Health Serv Res. 1988; 23: 203-235.
7. Rosenheck R, Cramer J, et al. A Comparison of Clozapine and Haloperidol in Hospitalized Patients with Refractory Schizophrenia. N Engl J Med. 1997; 337: 809-815.
8. Kane JM, Honigfeld G, et al. Clozapine for the treatment-resistant schizophrenic: a double-blind comparison with chlorpromazine. Arch Gen Psychiatry. 1988; 45: 789-796.
9. Kay SR, Fiszbein A, et al. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophr Bull. 1987; 13:261-276.
10. Guy W. Abnormal involuntary movements. In: Guy W, ed. ECDEU assessment manual for psychopharmacology. Rockville, Md.: National Institute of Mental Health (DHEW publication no. ADM 76-338.); 1976
11. Barnes TRE A rating scale for drug-induced akathesia. Br J Psychiatry. 1989; 154: 672-676.

12. Simpson GM and Angus JWS. A rating scale for extrapyramidal side effects. Acta Psychiatry Scand. 1970; Supp **212**: 11-19.
13. Seber G. Multivariate Observations. Wiley Series in Probability and Mathematical Statistics. Probability and Mathematical Statistics, New York: Wiley; 1984.
14. Wu AW, Jacobson KL, et al. Validity and responsiveness of the euroqol as a measure of health-related quality of life in people enrolled in an AIDS clinical trial. Qual Life Res. 2002; 11:273-282.
15. Hartigan JA and Wong MA. Algorithm AS 136 : A k-means clustering algorithm. Applied Statistics. 1978; 28: 100-108.
16. Sugar CA, Sturm R, et al. Empirically Defined Health States for Depression from the SF-12. Health Serv Res. 1998; 33: 911-928.
17. Sugar CA and James GM. Finding the number of clusters in a data set : An information theoretic approach. Journal of the American Statistical Association (Conditionally accepted) 2003
18. Efron B and Tibshirani R. An Introduction to the Bootstrap. London : Chapman and Hall; 1993.
19. Lenert LA, Sturley AE. Use of the Internet to study the utility values of the public. Proc AMIA Symp 2002; 440-444.
20. Lenert LA, Sturley AE, Rupnow M. Toward improved methods for measurement of utility: automated repair of errors in elicitations. Med Decis Making 2003; 23(01).

21. Gold M, Patrick D, Torrance G, Fryback D, Hadorn D, Kamlet M, et al. Cost-effectiveness in Health and Medicine. New York: Oxford Press; 1996.
22. Rosenheck R, Cramer J, et al. Multiple Outcome Assessment in a Study of the Cost-Effectiveness of Clozapine in the Treatment of Refractory Schizophrenia. Health Serv Res. 1998; 33: 1237-1261.
23. Taylor HM and Karlin S. An Introduction to Stochastic Modeling. San Diego: Academic Press; 1994.

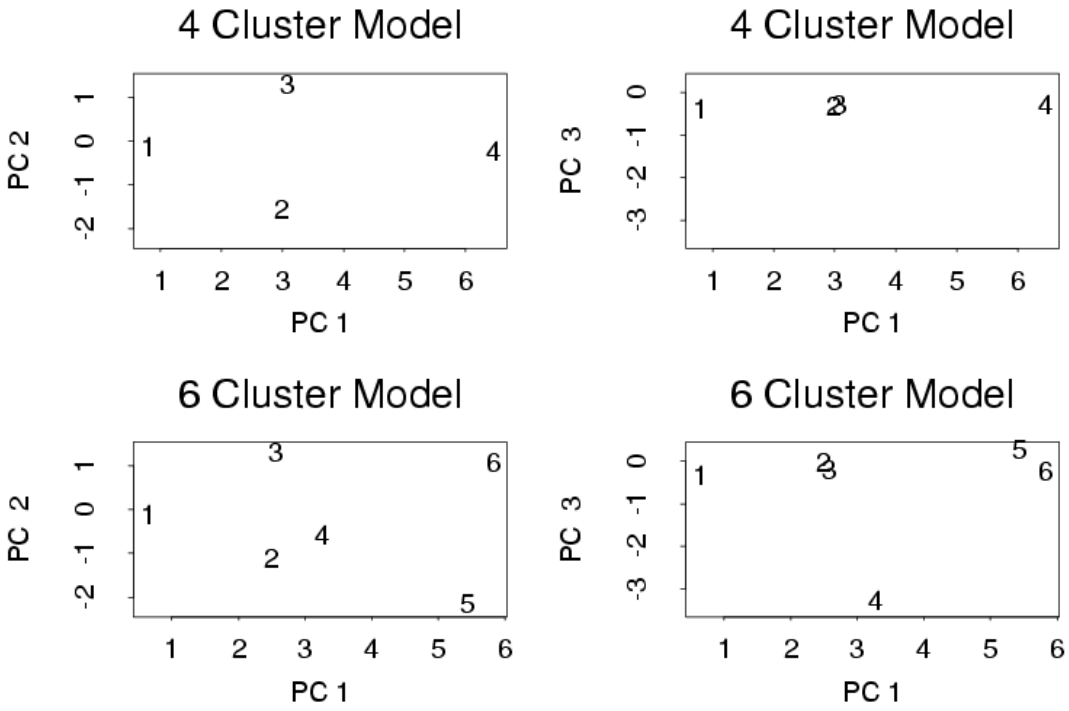

Figure 1: Cluster centers plotted in the first three dimensions for the four and six cluster models for the side effects data. The first principal component is on the x-axis, the second component on the y-axis for the first column and the third component on the y-axis for the second column.

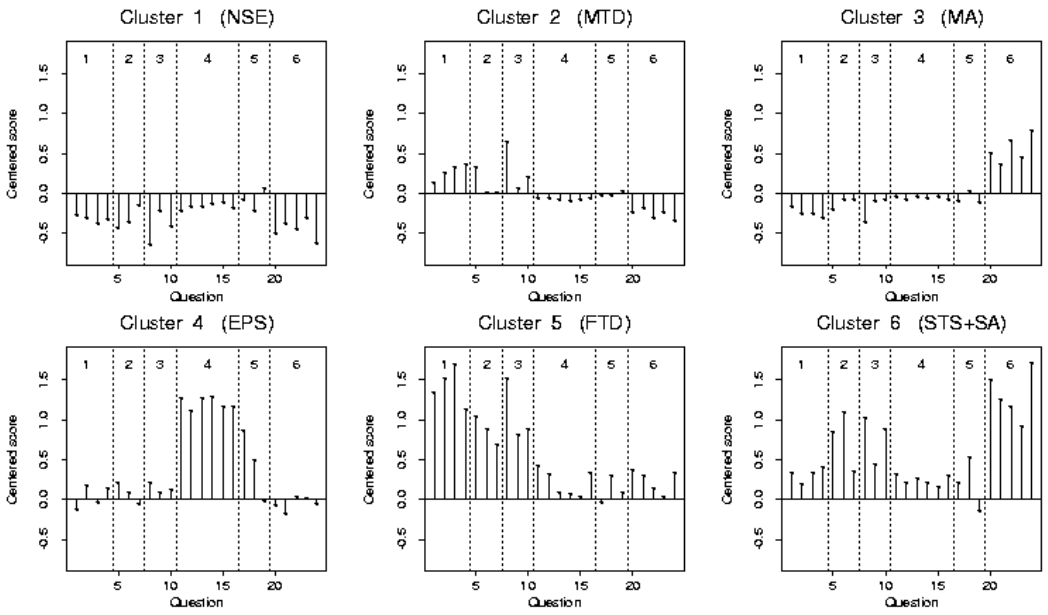

Figure 2: Cluster profile plots for the side effects data. The bands correspond to 1) AIM Q1-4: facial/oral movements, 2) AIM Q5-7: extremity and trunk movements, 3) AIM Q8-10: global severity, 4) SAS Q1-6: rigidity of gait, arms, head, 5) SAS Q7-9: glabellar tap, tremor and salivation and 6) SAS Q10, BAS Q1-4: akathesia. The bars represent average scores for each question with overall population mean subtracted off to show relative severity.

Dimension	Health State						
PANSS	1 Mild symptoms	2 Moderate symptoms and high global subjective distress	3 Moderate symptoms with high grandiosity	4 Severe negative symptoms with low subjective distress	5 Severe negative symptoms with high subjective distress	6 Severe positive symptoms	7 Severe symptoms with low subjective distress
PC 1	9.99	12.93	14.22	14.33	15.70	17.39	19.75
	(1.48)	(1.31)	(1.30)	(1.44)	(1.38)	(1.35)	(1.72)
PC 2	-0.46	0.82	1.67	-2.68	-1.55	2.11	-1.48
	(1.55)	(1.30)	(1.33)	(1.51)	(1.42)	(1.46)	(1.70)
PC 3	-3.44	-6.02	-2.81	-2.88	-6.22	-3.96	-3.47
	(1.40)	(1.23)	(1.22)	(1.28)	(1.35)	(1.54)	(1.49)
Side effects	1 No side effects	2 Mild tardive dyskinesia	3 Mild akathesia	4 Extra-pyramidal symptoms	5 Frank tardive dyskinesia	6 Severe tardive dyskinesia + severe akathesia	
PC 1	0.63	2.49	2.57	3.26	5.44	5.82	
	(0.65)	(0.72)	(0.81)	(1.41)	(1.26)	(1.46)	
PC 2	-0.13	-1.12	1.31	-0.58	-2.16	1.06	
	(0.53)	(0.78)	(0.82)	(1.20)	(1.21)	(1.07)	
PC 3	-0.36	-0.06	-0.21	-3.28	0.25	-0.25	
	(0.61)	(0.67)	(0.67)	(1.15)	(1.08)	(1.09)	

Table 1 : Mean scores on the first 3 principal components for patients in each health state for the PANSS and side effects scales. Standard deviations are given in parentheses. There were no significant differences between the health states on the fourth component of the side effects scale or the fourth and fifth components on the PANSS

PANSS			Health State						
			1 Mild symptoms	2 Moderate symptoms and high global subjective distress	3 Moderate symptoms with high grandiosity	4 Severe negative symptoms with low subjective distress	5 Severe negative symptoms with high subjective distress	6 Severe positive symptoms	7 Severe symptoms with low subjective distress
Time	p-value	Drug							
Baseline	0.267	clozapine	2.0	22.4	14.4	6.0	19.9	19.4	15.9
		haloperidol	0.0	22.7	15.0	5.0	16.4	26.4	14.5
6 Weeks	0.200	clozapine	16.2	16.8	13.9	15.6	16.2	9.8	11.6
		haloperidol	9.2	18.4	15.4	11.4	15.4	15.4	14.9
3 Months	0.10	clozapine	22.0	16.4	15.8	17.5	10.2	9.0	9.0
		haloperidol	16.4	15.5	12.3	15.5	11.8	18.6	10.0
6 Months	<0.001	clozapine	23.2	19.0	21.4	16.7	6.0	7.1	6.5
		haloperidol	20.1	11.4	15.7	14.8	12.2	15.3	10.5
9 Months	0.02	clozapine	26.2	15.7	13.4	16.9	10.5	9.9	7.6
		haloperidol	18.7	13.9	18.7	9.6	11.0	20.1	8.1
12 Months	0.237	clozapine	23.5	18.8	15.3	18.2	9.4	9.4	5.3
		haloperidol	20.3	17.0	20.8	11.8	8.0	14.2	8.0

Table 2 : Cross-sectional distributions. The rows give the percentage of clozapine or haloperidol patients in each PANSS health state at each of the 6 study time points. The p-values correspond to significance tests of differences between treatments at each time point.

Side effects			Health State					
			1 No side effects	2 Mild tardive dyskinesia	3 Mild akathesia	4 Extra-pyramidal symptoms	5 Frank tardive dyskinesia	6 Severe tardive dyskinesia + severe akathesia
Time	p-value	Drug						
Baseline	0.89	clozapine	26.4	15.9	25.9	10.0	3.5	18.4
		haloperidol	24.4	17.5	24.9	9.2	6.0	18.0
6 Weeks	<0.001	clozapine	52.9	18.8	16.5	2.9	5.3	3.5
		haloperidol	32.7	21.2	20.2	6.2	7.2	12.5
3 Months	<0.001	clozapine	55.2	21.8	13.8	1.7	3.4	4.0
		haloperidol	26.7	18.7	28.9	3.7	9.6	12.3
6 Months	<0.001	clozapine	57.7	18.4	14.1	1.2	6.1	2.5
		haloperidol	35.1	19.3	23.4	4.1	5.8	12.3
9 Months	<0.001	clozapine	54.8	28.3	9.6	0.6	3.0	3.6
		haloperidol	34.7	23.1	17.7	5.4	10.2	8.8
12 Months	<0.001	clozapine	60.2	21.7	8.7	0.6	5.6	3.1
		haloperidol	37.6	22.7	19.9	6.4	9.9	3.5

Table 3 : Cross-sectional distributions. The rows give the percentage of clozapine or haloperidol patients in each side effects health state at each of the 6 study time points. The p-values correspond to significance tests of differences between treatments at each time point.

Side Effects: Transition probabilities for clozapine patients		To State					
		1 No side effects	2 Mild tardive dyskinesia	3 Mild akathesia	4 Extra-pyramidal symptoms	5 Frank tardive dyskinesia	6 Severe tardive dyskinesia + severe akathesia
From State	1	80.2	11.3	7.4	0.4	0.8	0.0
	2	33.0	54.0	6.0	1.0	4.0	2.0
	3	48.3	19.0	27.6	0.0	3.4	1.7
	4	0.0	33.3	0.0	66.7	0.0	0.0
	5	5.6	22.2	5.6	0.0	50.0	16.7
	6	6.3	25.0	18.8	0.0	18.8	31.3

Table 4 : Transition probabilities for the side effects health state model for patients on clozapine. Each row gives a patients percentage chances of moving from their current state to one of the 6 possible health states in a 3 month period and sum to 100%.

Side Effects: Transition probabilities for haloperidol patients		To State					
		1 No side effects	2 Mild tardive dyskinesia	3 Mild akathesia	4 Extra-pyramidal symptoms	5 Frank tardive dyskinesia	6 Severe tardive dyskinesia + severe akathesia
From State	1	61.2	17.1	17.8	0.8	0.0	3.1
	2	29.2	38.2	12.4	2.2	9.0	9.0
	3	20.8	12.1	47.3	1.1	1.1	7.7
	4	15.0	10.0	0.0	70.0	5.0	0.0
	5	3.2	16.1	6.5	0.0	58.1	16.1
	6	14.3	26.2	16.7	0.0	14.3	28.6

Table 5 : Transition probabilities for the side effects health state model for patients on haloperidol. Each row gives a patient's percentage chances of moving from their current state to one of the 6 possible health states in a 3 month period and sum to 100%.

PANSS : Transition probabilities for clozapine patients		To State						
		1 Mild symptoms	2 Moderate symptoms and high global subjective distress	3 Moderate symptoms with high grandiosity	4 Severe negative symptoms with low subjective distress	5 Severe negative symptoms with high subjective distress	6 Severe positive symptoms	7 Severe symptoms with low subjective distress
From State	1	56.4	18.8	7.7	12.0	2.6	2.6	0.0
	2	28.8	51.3	3.8	3.8	7.5	3.8	1.3
	3	14.5	8.4	54.2	7.2	1.2	14.5	0.0
	4	17.9	8.3	7.1	53.6	9.5	1.2	2.4
	5	7.5	15.0	5.0	12.5	55.0	2.5	2.5
	6	2.5	7.5	25.0	12.5	2.5	35.0	15.0
	7	3.2	0.0	16.1	16.1	0.0	19.4	45.2

Table 6 : Transition probabilities for the PANSS health state model for patients on clozapine. Each row gives a patient's percentage chances of moving from their current state to one of the 7 possible health states in a 3 month period and sum to 100%.

PANSS : Transition probabilities for haloperidol patients		To State						
		1 Mild symptoms	2 Moderate symptoms and high global subjective distress	3 Moderate symptoms with high grandiosity	4 Severe negative symptoms with low subjective distress	5 Severe negative symptoms with high subjective distress	6 Severe positive symptoms	7 Severe symptoms with low subjective distress
From State	1	62.2	17.9	6.6	4.7	4.7	3.8	0.0
	2	22.8	41.8	13.9	7.6	5.1	6.4	2.5
	3	11.8	11.8	42.4	7.1	0.0	22.4	4.7
	4	13.9	5.6	11.1	43.1	13.9	2.8	9.7
	5	5.6	11.3	11.3	14.1	39.4	9.9	8.5
	6	5.0	11.0	20.0	3.0	8.0	43.0	10.0
	7	5.7	0.0	18.9	13.2	18.9	24.5	35.8

Table 7 : Transition probabilities for the PANSS health state model for patients on haloperidol. Each row gives a patient's percentage chances of moving from their current state to one of the 7 possible health states in a 3 month period and sum to 100%.

	Percentage of treatment group in each state						
PANSS health states	1 Mild symptoms	2 Moderate symptoms and high global subjective distress	3 Moderate symptoms with high grandiosity	4 Severe negative symptoms with low subjective distress	5 Severe negative symptoms with high subjective distress	6 Severe positive symptoms	7 Severe symptoms with low subjective distress
clozapine	27.5	20.1	15.2	16.8	9.3	7.4	3.6
haloperidol	23.9	16.5	18.1	10.9	8.0	15.5	7.1
Side effects health states	1 No side effects	2 Mild tardive dyskinesia	3 Mild akathesia	4 Extra-pyramidal symptoms	5 Frank tardive dyskinesia	6 Severe tardive dyskinesia + severe akathesia	
clozapine	61.1	22.7	8.9	1.4	4.1	1.9	
haloperidol	37.7	20.9	21.3	3.3	8.3	8.4	

Table 8 : Estimated stationary distributions for patients on clozapine or haloperidol.

	Average cost per week and QALYs in each PANSS state						
	1 Mild symptoms	2 Moderate symptoms and high global subjective distress	3 Moderate symptoms with high grandiosity	4 Severe negative symptoms with low subjective distress	5 Severe negative symptoms with high subjective distress	6 Severe positive symptoms	7 Severe symptoms with low subjective distress
Financial cost ($)	997 (118)	947 (127)	944 (131)	1176 (136)	1306 (148)	1516 (158)	1620 (165)
QALY	0.88 (0.006)	0.75 (0.012)	0.75 (0.012)	0.63 (0.018)	0.63 (0.018)	0.63 (0.018)	0.42 (0.012)

Table 9 : Financial costs and QALYs for a typical patient for each of the 7 PANSS health states.

Standard errors are provided in parentheses.

	Baseline	6 Weeks	3 Months	6 Months	9 Months	12 Months	Long run
haloperidol ($)	65,762	63,444	62,711	61,982	61,686	59,633	59,696
clozapine ($)	65,012	61,476	59,614	57,315	59,434	58,398	57,096

Table 10 : Estimated annualized costs for patients on each medication at 6 periods during the 1 year study as well as long run.