Modeling Depression Progression Dynamics from Electronic Health Record

Jiaqi Huang

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2016

Committee:

Shan Liu, Chair

Zelda B. Zabinsky

W. Art Chaovalitwongse

Program Authorized to Offer Degree:
Industrial and Systems Engineering

University of Washington

Abstract

Modeling Depression Progression Dynamics
from Electronic Health Record

Jiaqi Huang

Chair of the Supervisory Committee:
Assistant Professor Shan Liu
Department of Industrial and Systems Engineering

To assess and monitor the progression dynamics of patients' depression severity conditions, Markov models are refined from other disease progression modeling methodologies to identify the characteristics and evolvement of disease severity state transitions among a cohort of patients. Explored in this thesis is an integrated evaluation approach on the Markov models with emphasis on their application of modeling depression progression dynamics. Using the Patient Health Questionnaire (PHQ) - 9 data from electronic health record, b-spline curving and k-means clustering methods convert individuals' irregular PHQ-9 measurements to complete longitudinal depression trajectories and group the entire dataset into custom subgroups. Multi-State Discrete Time Markov model (MSM), Hidden Markov model (HMM), Semi-Markov model (semi-M) and Hidden Semi-Markov model (HSMM) are applied to each of the five subgroups to model the depression progression dynamics and identify the characteristics of depression severity state transitions. Purpose served in this approach, including providing insights on long-term severity progression outcomes, can be demonstrated among measurements of stationary probability, expected first passage time, and the proportion of time in a depression state where appropriate computed from the subgroup-specific transition probability matrices based on four Markov

models. The effectiveness and accuracy of the models are then further validated under bootstrap and cross validation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to gratefully thank my supervisor, Dr. Shan Liu for her guidance, understanding, patience and most importantly, the excellent example she has provided as an enthusiastic young scholar. I appreciate all her scientific advice and knowledge, insightful discussions and suggestions to make my research experience productive and stimulating. The joy and enthusiasm she has for her research was contagious and motivational for me, even during tough times over the past year.

My gratitude also goes out to my reading committee members, Prof. Zelda B. Zabinsky and Prof. W. Art Chaovalitwongse, for their effort of reviewing this thesis and valuable comments. I want to thank Prof. Zelda B. Zabinsky for her outstanding teaching. She introduced me to mathematical optimization logic and her teaching inspired me to work on mathematical modeling in career pursuit. I wish to thank Prof. W. Art Chaovalitwongse for his continued support and insightful discussions of independent study research projects. He was my primary resource for getting my modeling questions answered and was instrumental in helping me crank out those projects.

Additionally, I am very grateful for the friendship of all of the members of this research group, especially Yin Lin, with whom I worked closely and puzzled over many of the problems.

My time at UW was made enjoyable in large part due to the many friends that became a part of my life. I am grateful for time spent with Weiwei Shang, Jingyi Lu,

Yi Zhou, my training partner Xuelu Yang, and all the memorable trips to the seaside as well as into the mountains.

I especially thank my parents for supporting me in all my pursuits. I would not have made it this far without their encouragement, unwavering love and faith in me. I thank my beloved grandmother, who has passed on half a year ago. I missed her so much and wished she could have lived longer for my graduation and my wedding in the future.

DEDICATION

To my loving parents and friends

Chapter 1

INTRODUCTION

1.1 Research Motivation & Background

Major depression is one of the leading contributors to disease burden worldwide [1]. Depression is a common mental disorder that presents with depressed mood, loss of interest or pleasure, decrease energy, feelings of guilt or low self-worth, disturbed sleep or appetite, and poor concentration, which, even worse, often comes with symptoms of anxiety. Depression ranked as "the fourth leading cause of burden among all diseases, accounting for 4.4% of total disability-adjusted life years, and the leading cause of years lived with disability, accounting for 11.9% of disability years" [2] in World Health Organization's report of 2001. According to World Health Organization, depressive disorders were then ranked as the third in 2004 and will move into the first place by 2030.

In United States, according to the National Comorbidity Survey Replication conducted in 2003, 16.2% of adults are affected by major depressive disorder for lifetime, and 6.6% for 12 months, while for respondents with 12-month major depressive disorders, only an estimated 57.3% received some type of treatment in the 12 month prior to interview [3]. Since 1990s, several large programs were launched to promote awareness of depression [4–6]. The growing awareness and recognition of the seriousness of depression led to the development of depression detection [7], treatment [8] and national public health initiatives [9].

Lack of effective strategies for primary prevention, population level initiatives for major depression have typically focused on the provision of treatment, increasing access to treatment and increasing the quality of treatment received. However, prospective and ef-

fective prevention of major depression can reduce episode incidence in those patients at risk of recurrence, the risk of worsening the depression severity, and reduce the duration of depressive episodes. Types of prevention strategy vary in type of target group of intervention: universal (entire population), selective (high-risk group) or indicated (individuals with depressive symptoms not meeting all criteria for a depressive disorder) [10]. People with sub-threshold/minor depression tend to be at an increasing risk of developing major depression in comparison to individuals not meeting the criteria of minor depression [11]. Additionally, minor depression often correlated with impaired functioning, reduced quality of life and work, and additional medical and non-medical costs [12].

Major depression has been shown to be of high relevance to high medical utilization, unusual set of symptoms such as fatigue and pain and medical costs that are around twice as high as controls [13]. High direct costs are occurred not only in mental health visits, but in every perspective of medical treatment including primary care visits, laboratory tests, emergency visits, and inpatient medical days. From an economic aspect, indirect costs of depression due to work absence, low productivity, and negative effects on family roles are twice as high as the direct costs [14, 15]. Moreover, research indicates that nearly 50% of patients with major depression referred out of primary care to mental health systems of care do not follow through with specialty treatment [16, 17]. Possible reasons include lack of insurance coverage and additional time, stigma about treatment and discommodity of visiting a specialty clinic, which requires better disease management programs in the primary care system by implementing effective mental disease management strategies. Consequently, there is a need to understand and monitor the dynamics of depressive symptoms in population level so that to benefit patients and society in burden alleviation, reduce unnecessary medical costs as well as reduce indirect costs.

The most commonly used depression diagnostic and severity measurement tool is the

patient health questionnaire at 9 items (PHQ-9, see Appendix A), which consists of the actual nine criteria on the diagnosis of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [18] depressive disorders [19]. The PHQ-9 is entirely self-administered version of the primary care evaluation of mental disorders (PRIME-MD), which has been well validated and comparable to the clinician-administered PRIME-MD. As a depression severity measure, the PHQ-9 score ranges from 0 to 27. Each of the 9 items can be scored from 0 ("not at all") to 3 ("nearly every day"). PHQ-9 scores of 5, 10, 15, and 20 representing the thresholds for mild, moderate, moderately severe, and severe depression respectively [20]. The suggested treatment actions in response to these various levels of PHQ-9 depression severity are shown in Table 1.1. If five or more of the nine depressive symptom criteria have been present at least "more than half the days "in the past 2 weeks, and one of the symptoms is depressed mood or anhedonia, major depression is diagnosed.

Table 1.1: PHQ-9 Scores and Proposed Treatment Actions

PHQ -9	**Depression Severity**	**Proposed Treatment Actions**
1 to 4	Healthy	None
5 to 9	Mild	Watchful waiting; repeat PHQ-9 at follow-up
10 to 14	Moderate	Treatment plan, considering counseling, follow-up and/or pharmacotherapy
15 to 19	Moderately Severe	Immediate initiation of pharmacotherapy and/or psychotherapy
20 to 27	Severe	Immediate initiation of pharmacotherapy and, if severe impairment or poor response to therapy, expedited referral to a mental health specialist for psychotherapy and/or collaborative management

As aforementioned, relevant depression severity states are categorically defined. A better observation and interpretation of PHQ-9 scores would help guiding decisions for appropriate and early therapy arrangements before increased risk of developing major depression. Decisions involving depressive diagnoses are currently based on clinical judgment and national guidelines for depressive disorder care. This thesis focuses on developing models to

understand the evolvement of depression severity conditions, aiming at providing analyses and methods that can proactively inform clinical studies, policy development and relevant cost-effectiveness analysis. Besides assessing and monitoring the dynamics of major depression, the duration and occurrence possibility of major depressive episodes are other variables of interest. At population level, such findings would provide essential implications for medical benefit plan design, disability plan management, and occupational health professionals' training for policy makers and decision makers.

1.2 Research Purpose and Contributions

The purpose of this thesis is to address the challenge of effectively modeling and analyzing population-level depression severity progression dynamics, and providing insights on the duration and occurrence possibility of major depressive episodes.

At the population level, the rate of disease progression varies between individuals. To our knowledge, the complex dynamics of individual's depression trajectory has already presented challenges for clinicians to interpret and perform diagnosis/prognosis (individual-level challenge), the widely reported heterogeneity of the depression population further exacerbates the problem of setting "one size fits all" [21] screening guidelines (population-level challenge).

How do we accurately estimate patients' rates of depression progression accounting for heterogeneity in patient characteristics, and allowing for flexibility in assigning population-level distributions of individuals at different depression severity stages, so as to project outcomes and guide decisions regarding to population-level depression severity conditions and potential evolvement, has been the main challenge in the modeling process. Cluster the symptoms of depression among a cohort of patients, which would be expected based on their distinctiveness, can be a tangible approach to address this problem. In terms of testing prediction and simulation results hence informs clinical studies, modeling based on sepa-

rate groups serves to clarify clinical data employing categorical diagnoses and intervention strategy.

Although long-term clinical studies are able to provide significant clinical results, such studies are not practical for every possible group with its own characteristics. The increasing cost and time render these clinical trials unrealistic. In such situations, mathematical models may generate possible results in a relatively short amount of time and lower costs. More specifically, Markov models may prove quite useful to identify the characteristics of disease severity state transitions. A major process of modeling progression using the Markov models is to estimate the transition probabilities. Well constructed Markov models along with the correspondence transition probability matrices provide a set of feasible long-term outcomes to justify initiatives and changes to mental disease management strategies, relevant policies and surveillance.

The remainder of this thesis is organized as follows. In Chapter 2, we describe relevant literature, including disease progression dynamics models, and application of Markov models in modeling the disease progression. Chapter 3 analyzes the trajectory patterns of each subject by using B-spline, provides an efficient cluster approach with k-means clustering algorithm and develops probabilistic progression of PHQ-9 counts for patients with ongoing treatment under four Markov models. Transition probability matrices produced from Markov models are then used in Chapter 4, where we calculate the long-term outcome parameters. We explore stationary probabilities, expected first passage time for a specific depression severity state, the proportion of time in a certain depression severity state across all subgroups using transition probability matrices generated in Chapter 3. Then we extend the analysis to a set of possible ranges for these long-term parameters via bootstrap validation and cross validation. Chapter 5 discusses the conclusions, limitations and future extensions of this thesis.

Chapter 2

LITERATURE REVIEW

Depression is a recurrent, potentially chronic, which progresses pretty slowly over a long period of time, and disabling condition. The capability of detecting the development and progression of such disease at an early stage is instrumental to deliver future clinical insights. Though current literature presents few options for modeling the depressive disorders, numerous recent research interested in the development of progression dynamics modeling and Markov modeling applications in medical decision areas. In this chapter, we divide our literature review into disease progression dynamics modeling and Markov models.

2.1 Modeling Disease Progression Dynamics

The modeling of a target disease progression incorporates developing and estimating mathematical functions to describe the quantitative relationships for individuals or population from prior clinical experience, and predict the time course of disease status and drug effects. With characterized disease progression trajectory, disease progression modeling facilitates the disease prognosis improvement [22], drug effects evaluation [23, 24], and clinical trial design [25,26] hence enlighten the initiatives and revisions of disease management policies or primary care strategies. Non-linear mixed-effects models are often employed to provide an estimate of typical disease progression using a population approach with untreated states [27]. Since the publication of Chan and Holford [28], who were the first to introduce natural disease progression and drug action in principal of clinical pharmacology [29] for diseases such as Parkinson's disease, Alzheimer's disease, respiratory disease, osteoporosis and diabetic

nephropathy.

The concept of conducting model-based evaluations over disease progression is not new. In 1981, Holford and Sheiner promotes a "zero progression" model representing disease status that does not change except through therapeutic intervention [30]. A family of models was proposed by Post et al. [31] to describe the disease status and progression of degenerative diseases, such as type 2 diabetes mellitus, and Parkinson's disease, as function of the disease process and treatment effects. In modeling the progression of diabetes mellitus, another paper considered a mechanism based technique by tracking the interaction between several key indicators [26]. Based on literature meta-analysis, a model describing the longitudinal changes of patients with mild to moderate Alzheimer's disease is presented [32]. A fused sparse group lasso approach was then employed in modeling disease progression with known biomarkers [33]. Exarchos et al. focused on modeling the progression of Coronary Atherosclerosis using a dynamic Bayesian network to predict possible manifestations of the disease [34]. Multi-task learning framework was applied to the prediction of disease progression measure by the cognitive scores and selecting biomarkers [35].

However, current literature on modeling of the longitudinal course of depression progression is quite limited. An essential method in this context is the Markov model which will be reviewed in the next section. To our knowledge, an inverse Bateman model and a limited linked cosine model were used to describe short-term time course of disease progression during a clinical trial with the rate of recovery [36]. Shang et al. evaluated inverse Bateman model and the K-PD model as well as a two-transit-compartment model for their ability and flexibility in describing the time course of placebo response [37].

2.2 Markov Models

Given the nature of depressive disorders, short term and intermittent observation of the depression symptoms generally can't harvest an accurate diagnosis. Long-term analysis of

depression symptoms is required to detect the possible severity state transitions. Markov models provide effective analysis to picture the longitudinal course of depression in terms of mutually exclusive depression severity states and the transitions among them. Patten integrated Markov model to decompose prevalence into incidence, recovery and mortality as weekly transition probabilities, and discovered that recovery was high in the initial weeks of the depression episodes, and declined by a fixed proportion with each passing week [38]. Major depressive disorder is modeled and analyzed using the Markov chain model to obtain a mean to find the future emotional state of the depression based on the subject's current mental status [39]. In another attempt, Markov model is employed to estimate future health effects and costs of an intervention scenario and a current practice scenario in the prevention of major depression [10]. A cohort simulation using state-transition Markov models was performed to simulate movement of a hypothetical cohort of workers to capture the health service and employment related costs [40]. Bhattacharya discussed human depression dynamics under the construction of a two-state Markov chain model [41], along with different aspects of treatment of medical depression.

However, does the observed values of the PHQ-9 depression questionnaire draw actual inference of individuals's actual depression severity states? Hidden Markov models provide a tangible solution to this matter. Separate models govern the progression through underlying states and the correspondence observation process of the underlying states respectively. Jackson et al. described the use of Hidden Markov models for the underlying staged function decline of lung transplant recipients in 2002 [42]. In the following year, they presented a general Hidden Markov model for simultaneously estimating the transition rates and probabilities of stage misclassification with fitted covariates [43]. Hidden Markov models have less frequently been applied in disease progression modeling. Several applications including Alzheimer's disease [44], characterizing CD4 cell decline on HIV patients [45], analyzing the

progression of liver cirrhosis to HCC [46].

These Markov models deal with only discrete-time Markov models. Is it a reasonable assumption that individuals only transit under discrete time intervals? A disease process often evolves in continuous-time, and patients are often monitored at irregular and differing intervals. Most applications put their emphasis on a continuous-time Markov model or Hidden continuous-time Markov model, where the time spent in each state takes non-negative values and exponentially distributed. Hendriks et al. studied the progression of HIV infection in a cohort study of homosexual men using continuous-time Markov models based on CD4 counts and anti-CD3 reactivity [47]. Bureau et al. presented the application of continuous-time Markov models on the oral lesion hairy leukoplakia in a cohort of HIV infected men and cervical human papillomavirus (HPV) infection in a cohort of young women [48].

The characteristic of having an exponential distribution for sojourn distribution is still too restrictive in clinical disease progression modeling process. In this paper, we provide modeling and analyses based on semi-Markov process and Hidden semi-Markov process, which can be considered as an extension of ordinary Markov processes and Hidden Markov processes with completely random sojourn distributions. The only example reported on modeling disease progression involving semi-Markov models is proposed by Froucher et al. They applied a semi-Markov model based on generalized Weibull distribution to the evolution of HIV infected patients [49].

Chapter 3
MARKOV MODELS CONSTRUCTION

This chapter discusses methodologies used in this thesis to explore the impact of a depression intervention at a population level. The population-level data processing procedure, individual's disease trajectory fitting method, subgroup clustering method as well as construction processes of four Markov models are presented.

3.1 Data Processing

3.1.1 Data Source

Our data are provided by the Mental Health Research Network (MHRN), a consortium of 13 health system research centers supported by a cooperative agreement from the National Institute of Mental Health [50]. The MHRN data contain electronic health records (EHR) of patients registered for the depression questionnaire survey PHQ-9. We track the progression of patients' PHQ-9 counts as the primary variable to study the influence of medical intervention. Based on Table 1.1, PHQ-9 scores are categorized into five categories: 1 - 4, 5 - 9, 10 - 14, 15 - 19 and 20 - 27. These health states of our interest consist of a state space, denoted as {1, 2, 3, 4, 5}. For the purposes of consistency in the rest of paper, when describing the severity states of each subject, the state of $'1'$ represents 'Healthy (H)', $'2'$ represents 'Mild (Mi)', $'3'$ represents 'Moderate (Mo)', $'4'$ represents 'Moderately Severe (MS)', and $'5'$ represents 'Severe (S)'. The random variable X_t represents the severity state of patients at time t, $\{X_t \in \{1, 2, 3, 4, 5\}, t \geq 0\}$.

Our cohort contains 509,938 patients in total. Instead of considering the entire datasets,

our analyses focus on selected patients with ongoing treatment in the past 180 days starting from baseline to analyze population-level progression patterns. From this set of patients, patients with no more than four-time observations based on the standard time interval of two weeks are excluded. As to this constraint, if there's more than one observation in each time interval, which is two weeks in this paper, we average the PHQ-9 scores of these observations and regard them as one count. By doing so, the 6,067 patients with ongoing treatment we chose help us to better examine the possible effects of medical intervention on depression severity progression patterns. We box-plot the number of observations and the observed bi-week for each patient in 6,067 patients dataset as shown in Figure 3.1. Still, an obvious difference exists regarding the overall frequencies of these two parameters, which is worthy of further investigation.

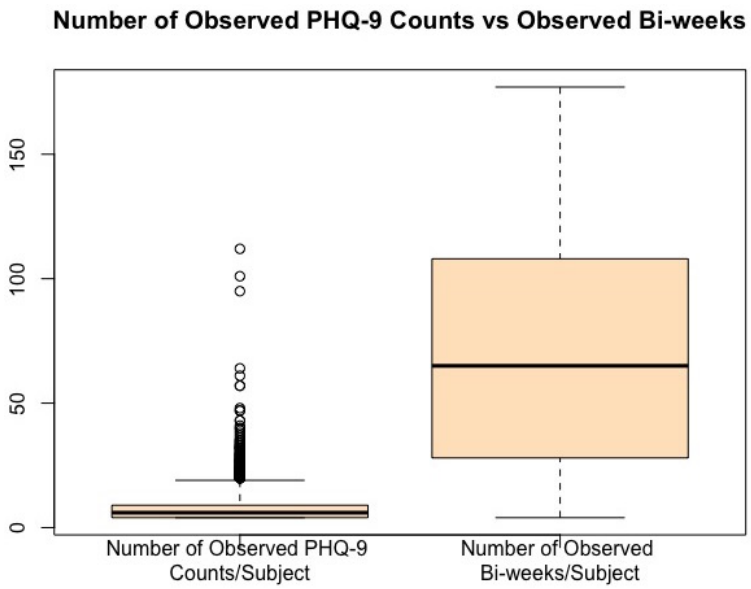

Figure 3.1: Number of Observed PHQ-9 Counts/Subject vs. Number of Observed Bi-Weeks/Subject _ 6067 Dataset

To put together an appropriate dataset of high clinical value, we then select subjects in the original MHRN dataset that have more than six observations within a 20 bi-week time

window as a new dataset with high monitoring frequency. In this case, we find 610 patients in total. As the numbers of observed bi-weeks within this dataset are all 20 bi-weeks, we only box-plot the number of observations for each patient in this dataset as presented in Figure 3.2.

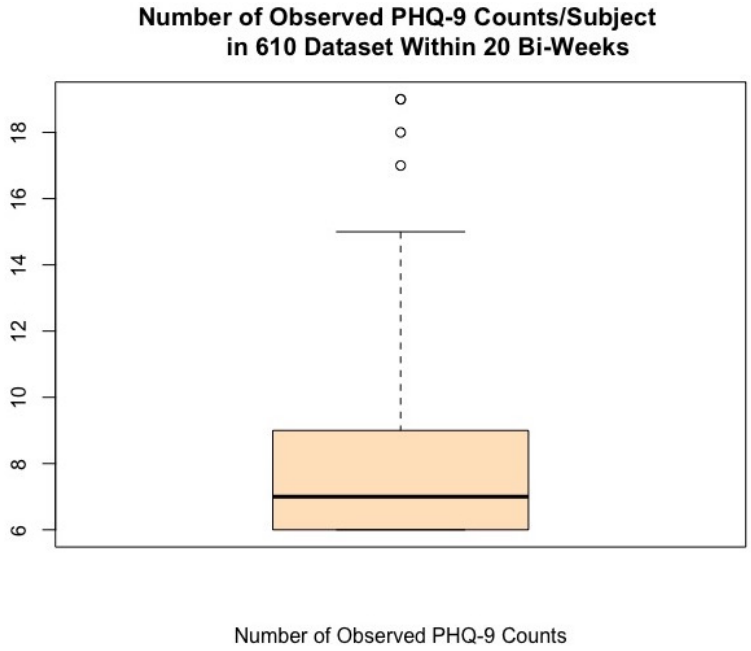

Figure 3.2: Number of Observed PHQ-9 Counts/Subject Within 20 Bi-Weeks _ 610 Dataset

Techniques and methodologies adopted in this thesis are performed on both datasets to investigate patterns for patients with different visiting frequencies. All statistical work in this paper is done in the open-source statistical package R, available at http://www.r-project.org/.

3.1.2 Trajectory Pattern Generation Using Smoothing B - Spline

Multiple challenges exist in constructing transition probability matrices, such as irregular observation times and incomplete data. Both Figure 3.1 and 3.2, in the other aspect, demon-

strate irregularity in the time between consecutive PHQ-9 counts. In fact, many months may elapse between two consecutive PHQ-9 measurements. Hence, we use b-splines to fit a continuous curve to the irregular data, and sample PHQ-9 counts at a bi-week regular time window to obtain a complete longitudinal history of patients and capture the variability of patient's disease trajectory. A similar approach was conducted by Shechter [51] to model the natural history of CD4 progression for HIV care.

B-spline is a curve-fitting method, constructed from polynomial pieces, joined at certain values of x, the knots [52]. Given the knots, it's simple to compute the b-splines recursively at any desired degree of the polynomial. However, the choice of knots is one of the keys to the goodness of fit. Too many knots are expected to result in over fitting of the data. Under-fitting will be produced with too few knots. A smoothing b-spline allows us to make a tradeoff between how smooth the curve is and how close the curve comes to the actual data, which imposes a penalty on the "roughness" of the function to avoid over fitting of the data thus reducing the sum of squared errors or residuals.

Assume vector y contains n values to be smoothed, and t records the times corresponding to each observation, which is often called curvature t. W is a symmetric positive definite weight matrix. We have

$$\text{PENSSE}_\lambda(x \mid y) = [y - x(t)]'W[y - x(t)] + \lambda \text{PEN}(x), \tag{3.1}$$

where $x(t) = \sum_k^K c_k \phi_k(t) = c'\phi(t)$ is the basis function expansion, c is the least-squares estimate of the coefficient vector that we want to partition on, $\phi_k(t_j)$ is the vector of b-spline basis function. For the classic penalty, $\text{PEN}_2(x) = \int [D^2 x(t)]^2 dt = \int [D^2 c'\phi(t)]^2 dt$ measures the roughness of splines that is defined on the derivatives of splines. Smoothing parameter λ controls the amount of roughness. As $\lambda \to 0$, roughness matters less and less, and as

$\lambda \to \infty$, roughness matters more and more. Rather than λ, a parameter called the "degree of freedom" has a direct mathematical relationship with the smoothing parameter λ. The minimal allowed value of *df* is three, and we can increase *df* to the number of data point. By decreasing *df* we can construct a smoother b-spline.

We run cross-validation to find the appropriate value of λ and *df* over Mean Square Error. According to the result, we choose $df = 4, \lambda = 10^3$ for 6067 patients dataset. For 610 patients dataset, the individual trajectory of each patient is fit by non-smoothed b-spline curve. Figure 3.3 and Figure 3.4 present the simulated PHQ-9 trajectories of the same 6 randomly selected patients from 6067 dataset and from 610 dataset separately.

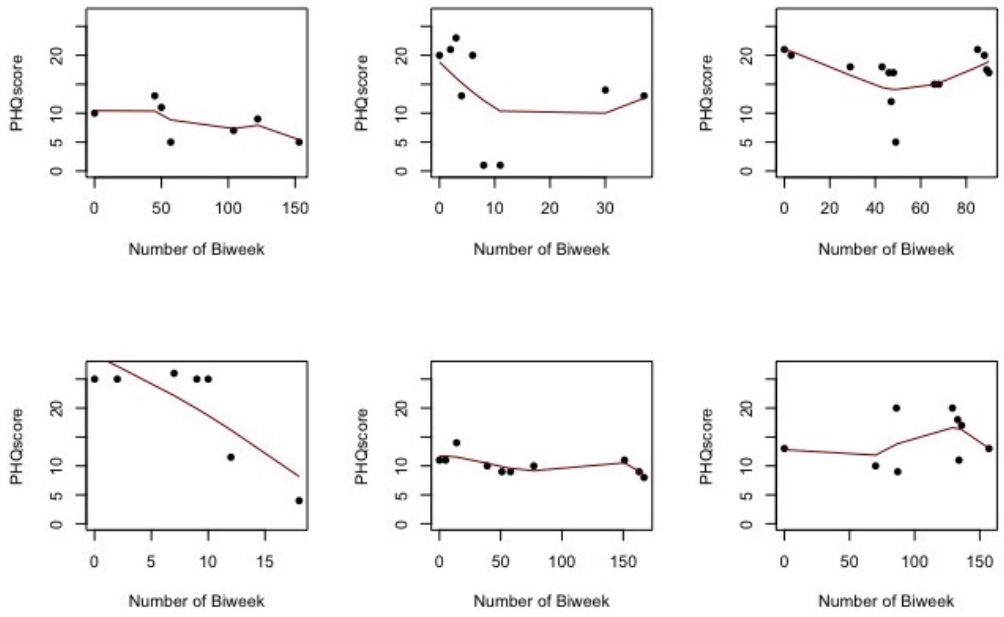

Figure 3.3: Six Random Selected Subjects from 6067 Patients Dataset

3.1.3 Patients Grouping Using K - Means

Previous research on depressive trajectories of depressive symptoms (PHQ-9) found that a cohort of patients from Australian family practice fit to five sub-groups in terms of their trajectory patterns [53]. The trajectories for the five-class model are presented in Figure 3.4.

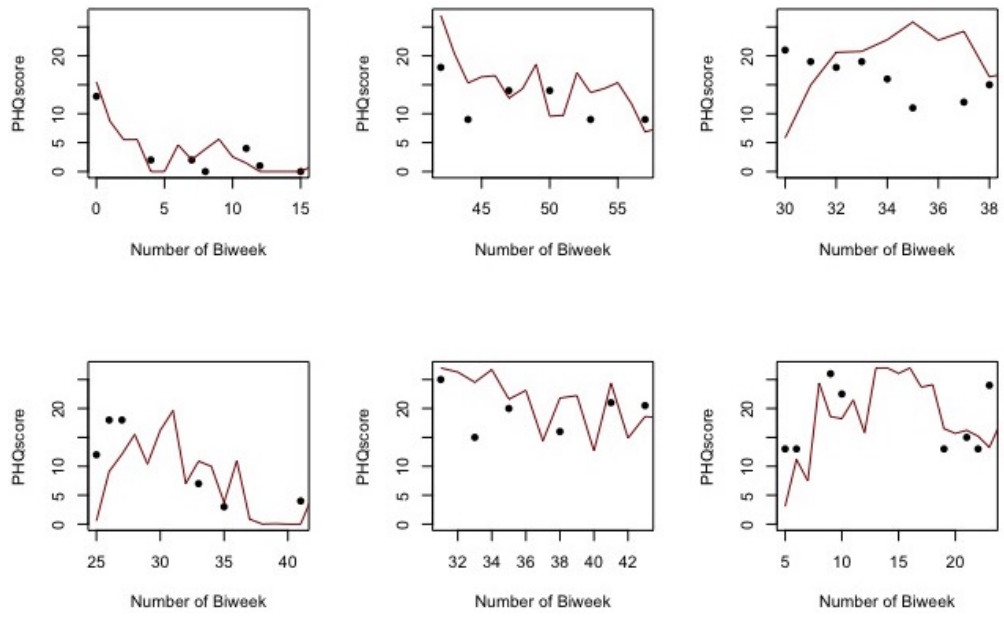

Figure 3.4: Six Random Selected Subjects from 610 Patients Dataset

Patients are sorted into Severe group, Decreasing Severity group, Moderate group, Increasing Severity group and Mild group. We are curious to know whether our cohort can be organized to meaningful structures hence characterize the heterogeneity of the depression population effectively.

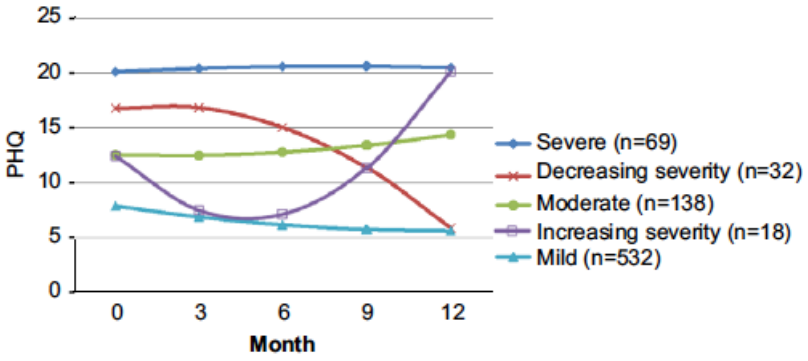

Figure 3.5: Trajectory of Depressive Symptoms (PHQ-9) from Gunn[4]

Cluster analysis is a statistical procedure that can be used to group objects of a similar

kind into respective categories. One of the oldest and most widely used algorithms is the k-means cluster algorithm with a non-iterative procedure, which requires the prior knowledge of the number of clusters. Therefore, prior to apply the k-means cluster analysis, the method of 10-fold cross validation is adopted to obtain an estimate of k. We randomly split the 6067 subjects into 10 folds, and successively choose one fold as testing data. Then apply k-means clustering model on the remaining $10-1$ folds, and validate the model on the testing data by clustering the test subjects. After calculating the average distance and weighted sum distance of these test subjects to their centers, the performance of cluster with different number of clusters are obtained. According to validation results, we start the program with 5 random clusters using 5 coefficient values computed from Section 3.1.1 to minimize variability within clusters and maximize variability between clusters by moving objects in and out of clusters. Clustering results for both 6067 [54] and 610 patient dataset are summarized in Table 3.1 below.

Table 3.1: K-means Clustering Results

No. Subjects	**Mild**	**Increasing**	**Moderate**	**Decreasing**	**Severe**
6067 Dataset	1,783	1,175	539	1,327	1,243
610 Dataset	134	104	130	130	112

As the result of k-means clustering analysis, usually we would estimate the means for each cluster to assess how distinct our k clusters are. We simulate the mean of all coefficients for each cluster to 100 biweeks time window as shown in Figure 3.6 (6067 dataset) and Figure 3.7 (610 dataset). Conjoining the number of patients in each cluster and the simulated mean curve, in 6,067 dataset, nearly two out of three subjects are most likely to belong to the mild group with an essential flat trajectory of what might be considered as mild group. 8.9% and 20.5% of subjects have moderate and severe depressive symptom trajectories. A

further 19.4% have a increasing severity tendency ending at moderately severe depression; conversely, the remaining 21.9% of subjects possess increasing severity tendency ending at mild depression at 100 biweeks. Similarly, we have five subgroups in 610 dataset.

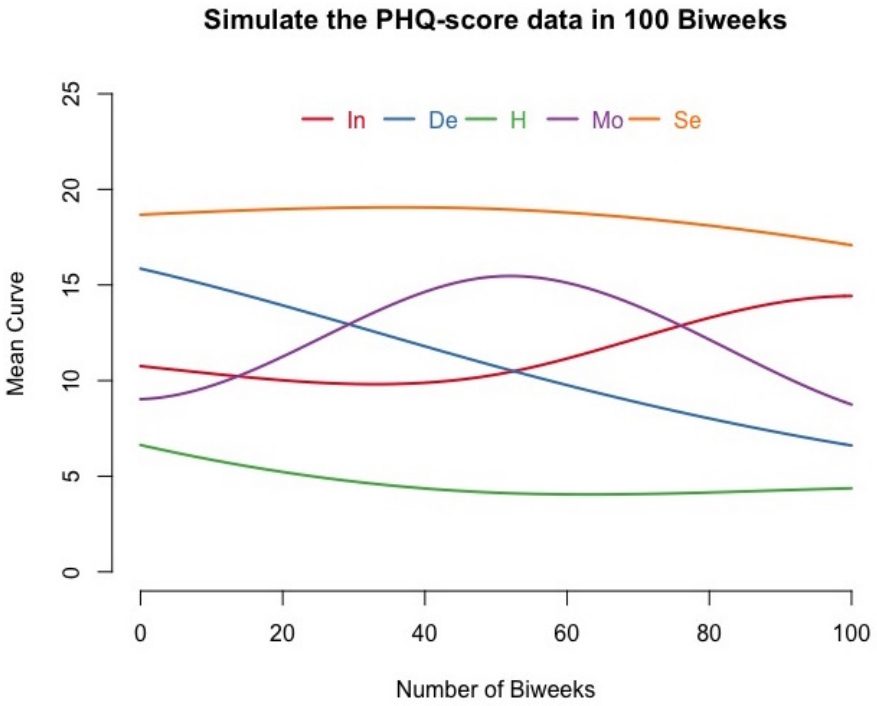

Figure 3.6: Simulated Mean Curve Found by K-means _ 6067 Patients Dataset

3.2 Model Construction Methodology

Random processes that evolve over time in a probabilistic manner are called stochastic processes. As a special kind of stochastic processes, Markov models indicate a sequence of stochastic events where probabilities involving. How the process will evolve in the future is independent of all past states of the process, depending only on the current state. Markov models are widely employed in medical decision making to characterize the possible prognoses experienced by a given cohort of patients.

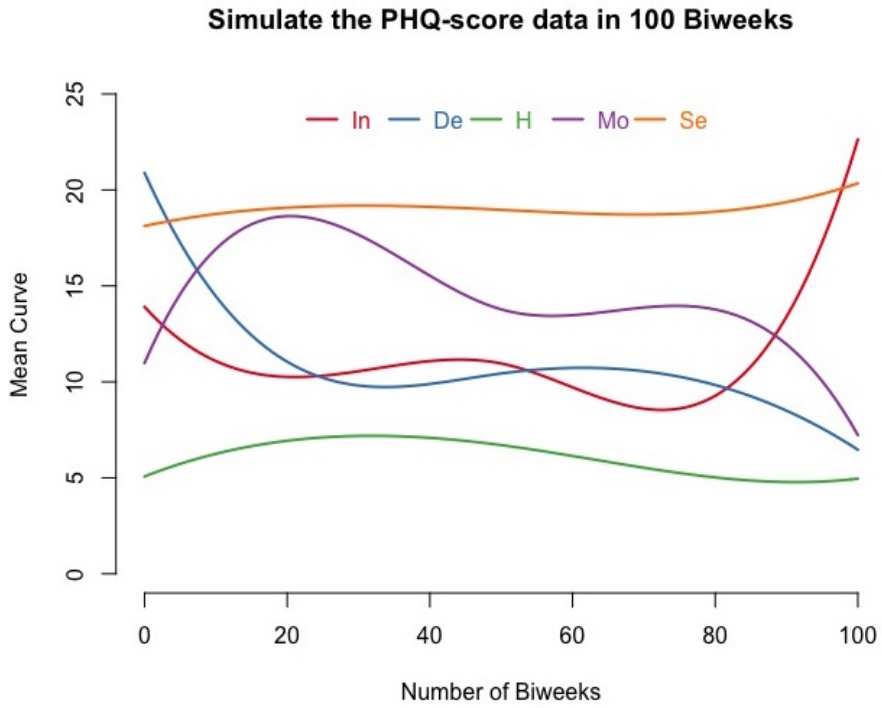

Figure 3.7: Simulated Mean Curve Found by K-means _ 610 Patients Dataset

3.2.1 Markov Property

Prior to building Markov models, one basic assumption is that the stochastic process $\{X_t\}$ is a Markov chain, which holds the Markov Property

$$\begin{aligned} P\{X_t + 1 = j \mid X_0 = k_0, X_1 = k_1, ..., X_t - 1 = k_t - 1, X + t = i\} \\ = P\{X_t + 1 = j \mid X_t = i\}, \end{aligned} \tag{3.2}$$

for $t = 0, 1, ...$ and every sequence $i, j, k_0, k_1, ..., k_t - 1$, indicating that the conditional probability of any future state is independent of the past states and depends only on the current states. To fully specify the Markov model, let us denote $\pi_i = P(X_0 = i)$ indicating the distribution of the initial state. The number of time steps spent in a given state is called sojourn time. The probability of spending u consecutive time steps in state i is denoted by

$d_i(u)$, which is the sojourn density, where

$$\begin{aligned} d_i(u) = P(X_{t+u+1} \neq i, X_{t+u} = i, X_{t+u-1} = i, \ldots, X_{t+2} = i \mid X_{t+1} = i, X_t \neq i) \\ = p_{ii}^{u-1}(1 - p_i) \end{aligned} \tag{3.3}$$

Sojourn time is geometrically distributed for any Markov chain.

From the given set of depression severity sequence, one-step transition probabilities

$$\widehat{p_{ab}} = (\sum_i N(X_i = a, X_{i+1} = b)/(\sum_i N(X_i = a)), i \geq 0, a, b \in \{1, 2, 3, 4, 5\}, \tag{3.4}$$

where N denotes the number of counts for a specific state, is estimated by using the Discrete-Markov model described in Section 3.2.2. We proceed in this fashion to calculate two-step empirical transition probabilities as

$$\widetilde{p_{ac}} = (\sum_i N(X_i = a, X_{i+2} = c)/(\sum_i N(X_i = a)), i \geq 0, a, c \in \{1, 2, 3, 4, 5\}, \tag{3.5}$$

and estimate the two-step transition probabilities as

$$\begin{aligned} \widehat{p_{ac}} = P[X_{i+2} = c \mid X_i = a] = \sum_c P[X_{i+2} = c \mid X_i + 1 = b] \cdot P[X_{i+1} = b \mid X_i = a] \\ = \widehat{p_{ab}} \cdot \widehat{p_{ab}}, i \geq 0, a, b, c \in \{1, 2, 3, 4, 5\}. \end{aligned} \tag{3.6}$$

Perform Pearson χ^2 test between two-step model probabilities $\widehat{p_{ac}}$ and two-step empirical probabilities $\widetilde{p_{ac}}$,

$$T_a = N(X_i = a) \sum_c (\widehat{p_{ac}} - \widetilde{p_{ac}})^2 / \widehat{p_{ac}}, a, c \in \{1, 2, 3, 4, 5\}. \tag{3.7}$$

If the p-value is greater than the given significance level, which is 0.05 in this paper, we cannot

reject the hypothesis that the Markov property holds for the specific transition. Test results are presented in Table 3.2, where all the p-values are greater than 0.05 for 10 subgroups in both datasets. We cannot reject the hypothesis that the Markov property holds for the transitions in our dataset.

Table 3.2: Markov Property Test Results _ P-value of Pearson Chi Square Test

P-Value	**Mild**	**Increasing**	**Moderate**	**Decreasing**	**Severe**
6067 Dataset	0.361	0.299	0.301	0.373	0.375
610 Dataset	0.760	0.923	0.915	0.858	0.886

3.2.2 Discrete-Time Markov Model

After obtaining the B-spline curves in Section 3.1.2, a regular-interval sampled, complete longitudinal history of PHQ-9 measurements can be formed through the interpolation of additional PHQ-9 counts between the observed PHQ-9 values. For some of the splines, curves may slightly exceed PHQ-9 score limits (0-27). When this occurs, we simply capture the estimates at PHQ-9 score limits. Take a simple example: suppose a PHQ-9 count on spline at one bi-week is -0.3, we take this count as within the range of 1-5, which is state $'1$, Healthy$'$. We then build discrete-time Markov model as follows.

The one-step transition probabilities,

$$\widehat{p_{ab}} = (\sum_i N(X_i = a, X_{i+1} = b)/(\sum_i N(X_i = a)), i \geq 0, a, b \in \{1, 2, 3, 4, 5\}, \tag{3.8}$$

are summarized by counting. Within a sub-group we obtained from Section 3.2.3., all transitions are tabulated from the PHQ-9 count at the beginning of one bi-week to the PHQ-9 count at the beginning of the next bi-week that occur as we proceed from the start of one patient's curve to the end of that curve. If the current PHQ-9 measurement is the last record

of a patient, we tabulate the next transition from the beginning of the next patient's curve within the same subgroup. Table 3.3 below presents the aggregate biweekly transitions from baseline until the last recorded PHQ-9 score for increasing group based on 6067 and 610 subjects dataset respectively.

Table 3.3: Aggregated Transition Counts for Increasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	3490	188	0	0	0
Mi	87	25221	715	0	0
Mo	0	364	45862	567	0
MS	0	0	259	12233	125
S	0	0	0	40	1286

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	298	43	0	0	0
Mi	99	640	61	0	0
Mo	1	152	661	50	0
MS	0	2	137	217	14
S	0	0	5	86	134

(b) (610 Subjects)

The categories in the first column represent possible ranges of the patient's PHQ-9 count at the beginning of one bi-week, and the categories in the first row show the same categories at the beginning of the next bi-week. Every time we found a transition from one PHQ-9 count to another, we add a 1 in the table cell accordingly showing a transition between the two categories. This procedure is replicated for each patient within the subgroup. We aggregate the calculations across all patients, and divide each cell count by the row sums to estimate the discrete-time transition probability matrix within this subgroup. The estimated transition probability matrices are presented in Table 3.4.

The simulated Markov traces generated from the transition probability matrices from Discrete-time Markov models indicate a consistent long-term probabilistic transitions tendency with the simulated mean curve obtained from K-means at group level, which is a increasing severity tendency for the selected increasing group. Figure 3.8 and Figure 3.9 are plotted from 6,067 and 610 dataset separately.

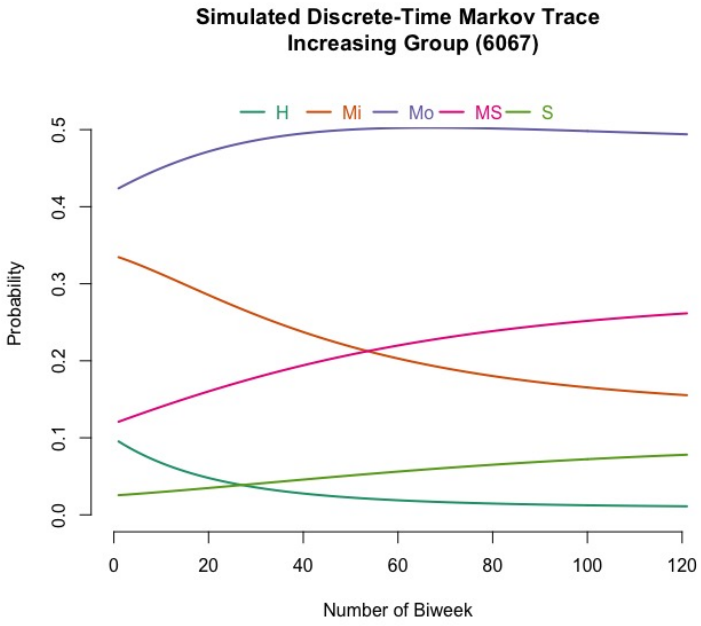

Figure 3.8: Simulated Discrete-Time Markov Trace _ Increasing Group (6067)

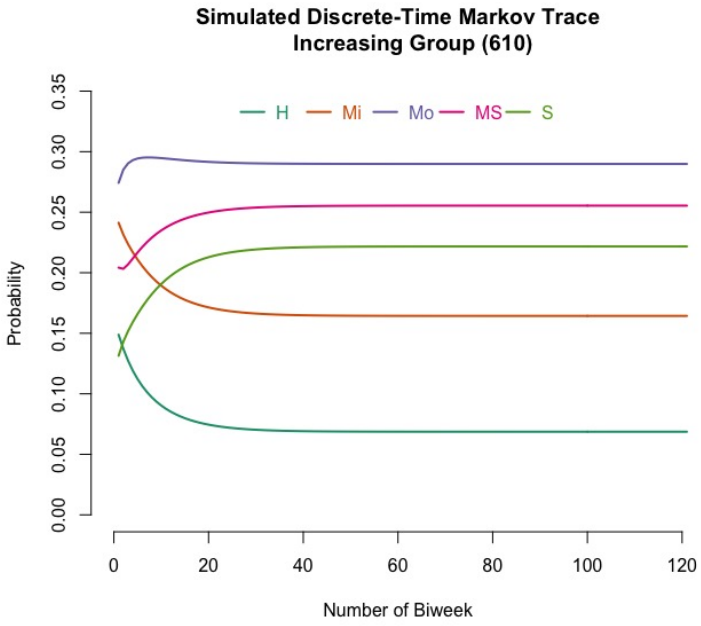

Figure 3.9: Simulated Discrete-Time Markov Trace _ Increasing Group (610)

Table 3.4: Discrete-Time Markov Transition Probability Matrices for Increasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.949	0.051	0	0	0
Mi	0.003	0.969	0.027	0	0
Mo	0	0.008	0.980	0.012	0
MS	0	0	0.021	0.970	0.010
S	0	0	0	0.030	0.970

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.762	0.228	0.010	0	0
Mi	0.097	0.665	0.215	0.023	0
Mo	0.002	0.129	0.691	0.165	0.014
MS	0	0.007	0.201	0.598	0.194
S	0	0	0.011	0.230	0.759

(b) (610 Subjects)

3.2.3 Hidden Markov Model

In discrete-time Markov models, states occurring at each instant in time we employed in modeling are observable. In Hidden Markov Models (HMM), states of a Markov chain are not directly observed. We only observe a sequence of observations at time t denoted as O_t which correlate with the unobserved states S_t but not state itself. The process is visualized in Figure 3.10 [55], where O_{T-1} are observed variables, and X_{T-1} are unobserved hidden states.

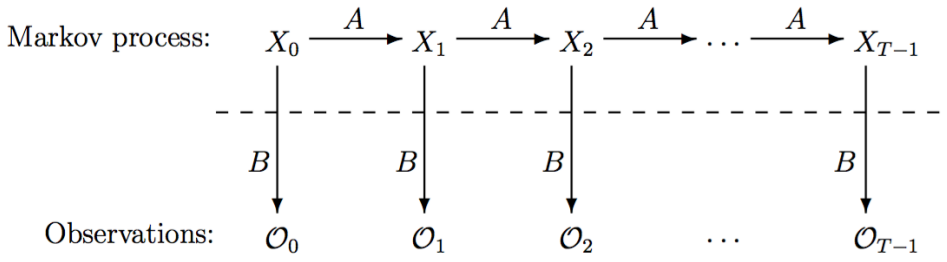

Figure 3.10: Visual Representation of a Hidden Markov Process[2]

We assume our observations are held at discrete, equally-spaced two-week time intervals, which is the same as Section 3.2.2. In addition to the basic assumption that the observation

at time t was produced by the process whose state X_t is hidden from the observer, hidden Markov model assumes that the current state X_t of the hidden process is independent of all the past states prior to $t-1$, which satisfies the Markov property. The joint distribution of this sequence of states and observations can be formulated as below:

$$P(X_{1:\,T}, O_{1:\,T}) = P(X_1)P(O_1 \mid S_1)\prod_{t=2}^{T} P(X_t \mid X_{t-1})P(O_t \mid X_t), \tag{3.9}$$

where $X_{1:\,T}$ denotes $X_1, \ldots, X_T$. Furthermore, hidden Markov model assumes that the state variable X_t of hidden process is discrete.

An HMM is characterized by the number of states in the model, the number of distinct observation symbols per state, and a triple $\theta = (\pi, A, B)$ [56]. $\pi = \{\pi_i\}$, the initial state distribution for each state $i, i \in \{1, 2, 3, 4, 5\}$, is computed from the initial states distributions of each patient within the same sub-group. Denote q_t as the state at time t.

$$\pi_i = P[q_1 = S_i], i \in \{1, 2, 3, 4, 5\}.$$

A, the initial state transition probability matrix, is mostly obtained from literature or calculated based on real data. In our case, we use transition probability matrices obtained from discrete-time Markov model as our initial state transition distributions. B, the initial emission matrix, indicates the observation symbol probability distribution in state j.

$$b_i(x_t) = P(X_t = x_t \mid S_t = i) \tag{3.10}$$

b_i can be obtained from literature or modeled in many different forms, such as a Gaussian, mixture of Gaussians, or a neural network. Our first emission matrix is derived from a distribution of PHQ-9 scores according to depression diagnostic status from Nicholson as

shown in Table 3.5 [20].

Table 3.5: Distribution of PHQ-9 Scores According to Depression Diagnostic Status

PHQ -9, Level of Depression Severity	**0-4**	**5-9**	**10-14**	**15-19**	**20-27**
No Depressive Disorder (%)	73.4	19.6	4.9	1.7	0.4
Other Depressive Disorder (%)	12.3	35.4	26.1	21.5	4.6
Major Depressive Disorder (%)	2.4	9.8	19.5	34.1	34.1

To maintain consistency with states, we add two levels of depression severity to Nicholson's distribution table. Mild depressive disorder distribution is calculated from an average of no depressive disorder distribution and moderate depressive disorder distribution. Moderately Severe depressive disorder distribution is calculated from an average of moderate depressive disorder distribution and major depressive disorder distribution. The emission matrix used in our HMM is presented as Table 3.6 below.

Table 3.6: Emission Matrix for Hidden-Markov models

PHQ -9, Level of Depression Severity	**0-4**	**5-9**	**10-14**	**15-19**	**20-27**
No Depressive Disorder (%)	73.4	19.6	4.9	1.7	0.4
Mild Depressive Disorder (%)	42.9	27.5	15.5	11.6	2.5
Moderate Depressive Disorder (%)	12.3	35.4	26.1	21.5	4.6
Moderately Severe Depressive Disorder (%)	7.4	22.6	22.8	2.8	19.4
Major Depressive Disorder (%)	2.4	9.8	19.5	34.1	34.1

Given the form of HMM aforementioned, one of the three basic problems of interest that need to be solved in real-world applications is how do we adjust the model parameters $\theta = (\pi, A, B)$ to maximize $P(O \mid \theta)$. Given the appropriate values that we have 5 states as $\{1, 2, 3, 4, 5\}$, 5 observation symbols per state as $\{"0-4", "5-9", "10-14", "15-19", "20-27"\}$, initial states distribution obtained from our data, initial states probability transition matrix from discrete-time Markov model as well as emission matrix as described

in Table 3.6, this initial HMM is used as a generator to produce an observation sequence $O = O_1 O_2 \dots O_T$. Now we call Baum-Welch algorithm to recalculate the parameters of the initial HMM model using these estimates. Table 3.7 demonstrated transition probability matrices for increasing group in both 6,067 and 610 dataset. By simply looking at the value of transition probabilities, we found that the state transition probability matrices are quite similar with our initializations, which are the transition probability matrices produced by Discrete-time Markov model.

Table 3.7: Hidden Markov Transition Probability Matrices for Increasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.939	0.061	0	0	0
Mi	0.003	0.971	0.026	0	0
Mo	0	0.008	0.980	0.012	0
MS	0	0	0.021	0.970	0.009
S	0	0	0	0.031	0.969

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.764	0.224	0.012	0	0
Mi	0.109	0.665	0.198	0.028	0
Mo	0.002	0.121	0.717	0.152	0.009
MS	0	0.004	0.195	0.629	0.172
S	0	0	0.002	0.175	0.823

(b) (610 Subjects)

The simulated Markov traces generated from the transition probability matrices from Hidden Markov models indicate a consistent long-term probabilistic transitions tendency with the simulated mean curve obtained from K-means at the group level, which is an increasing severity tendency for the selected increasing group. It seems that patients depression severity progression will reach a stable state after nearly 20 bi-weeks. However, in terms of the value of probability for each state varies across two datasets. It's more likely that patients in 610 dataset with high visiting frequency tend to be of higher sickness as well.

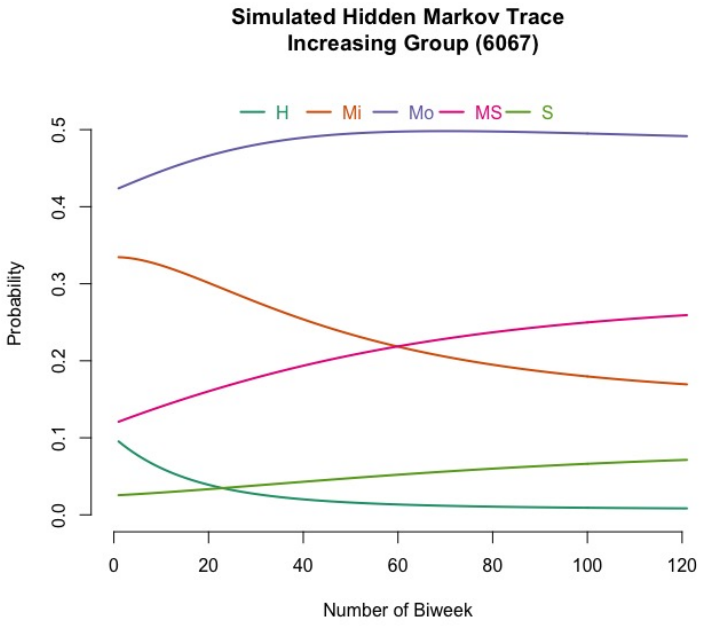

Figure 3.11: Simulated Hidden Markov Trace _ Increasing Group (6067)

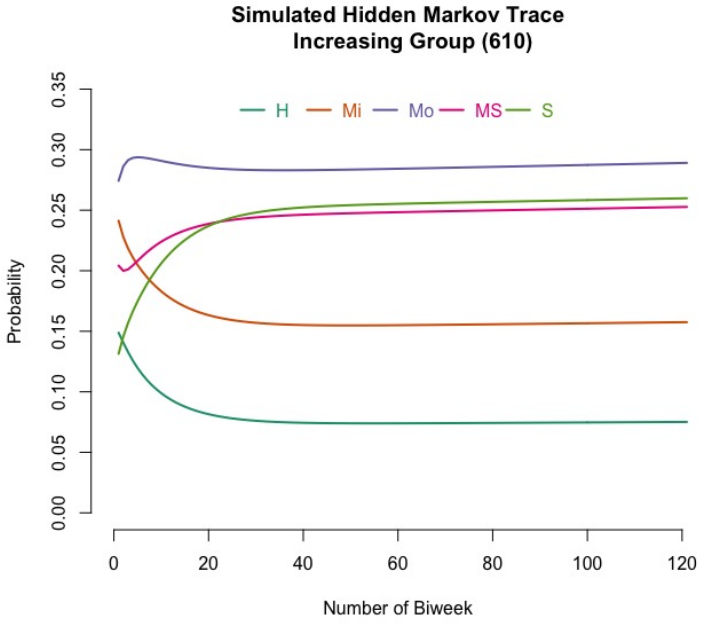

Figure 3.12: Simulated Hidden Markov Trace _ Increasing Group (610)

3.2.4 Semi Markov Model

In many of the depression severity progression analyses, the constraint that the process is only observed at discrete time points is far too restrictive. Patients change from one state to another might take a random amount of time. Given this case, this process can be modeled by the semi-Markov process, which the probability that patient transits from one state to another in accordance with a Markov chain but depends on the length of time that patient has been spent in that state.

Semi-Markov process holds that only the current state is relevant for the transition probabilities, which is still memoryless. However, in addition to the present state, we also want to know how long the patient has spent in that state. For a finite state space, specifically in our case with depression severity states, $I = \{1, 2, 3, 4, 5\}$, let $\{(Y_n, T_n), n \geq 0\}$ denotes a Markov renewal sequence, where T_n denotes the time of entrance to the state $Y_n, n \geq 0, Y_n \in I$. Pyke [57] defined the *kernel* of the semi-Markov process as the matrix $G(x) = [G_{ij}(x)]$, where

$$\begin{aligned} P\{Y_{n+1} = j, T_{n+1} - T_n \leq x \mid Y_n = i, T_n, Y_{n-1}, T_{n-1}, \ldots, Y_0, 0\} = \\ P\{Y_{n+1}, T_{n+1} - T_n \leq x \mid Y_n = i\} \equiv G_{ij}(x) \end{aligned} \tag{3.11}$$

Therefore, the joint probability of observing $Y_{n+1} = j$ in a waiting time of $T_{n+1} - T_n \leq x$, conditioned on the previous history, satisfies the Markov property. Let $X(t) = Y_{N(t)}$, where $N(t)$ be the state with the last completed state spell before t, $N(t) = \sup\{n \geq 0 : S_n \leq t\}$. The stochastic process $\{X(t), t \geq 0\}$ is called as a semi-Markov process. Let F_{ij} denote the probability distribution of sojourn times, which is related to the semi-Markov kernel through

the transition probabilities of the embedded Markov chain,

$$F_{ij} = P(T_{n+1} - T_n \leq x \mid Y_n = i, Y_{n+1} = j) = \frac{G_{ij}(x)}{p_{ij}}. \tag{3.12}$$

Our sojourn time distribution modeling strategies based on two distributions, exponential distribution $\mathcal{E}(\sigma_{ij})$ and Weibull distribution [58] $\mathcal{W}(\sigma_{ij}, \nu_{ij})$. Let $\alpha_{ij}(x)$ denote the hazard rate of semi-Markov process.

Using exponential distributions, the hazard rate is constant over time without memory, which is given by

$$\alpha_{ij}(x) = \frac{1}{\sigma_{ij}}, \forall x \geq 0, \forall \sigma_{ij} > 0. \tag{3.13}$$

The hazard function is defined by two parameters from Weibull distribution,

$$\alpha_{ij}(x) = \nu_{ij} x^{\nu_{ij}-1} (\frac{1}{\sigma_{ij}})^{\nu_{ij}}, \forall x \geq 0, \forall \sigma_{ij} > 0, \forall \nu_{ij} > 0. \tag{3.14}$$

where σ_{ij} is a scale parameter and ν_{ij} is a shape parameter. With $\nu_{ij} = 1$, we obtain the exponential distribution.

Our study is performed based on the $SemiMarkov$ package from Listwon [59]. Instead of using PHQ-9 counts from b-spline as described in Section 3.2.2, in consideration of their sojourn time in state, transitions between states analyzed in this section are based on observed PHQ-9 values. To initialize the semi-Markov process, we formulate our depression data into a table in long format first with the following informations: the ID of each patient, state left by the process, state entered by the process and sojourn time in state left by the process. Each row represents one transition, and there is possibly more than one row per patient. Within each subgroup, the rows are grouped by individuals and ordered chronologically. In a semi-Markov process, we must obtain the length of time that has been spent in that state.

Given this, the transitions between the same states are not allowed. While encountering such transitions, we combined that row with the next transition to obtain a new transition from state i to state j with $i \neq j$.

Next, we prepare a quadratic 5×5 matrices describing the possible transitions and distributions of waiting time based on each subgroup, given that we have 5 states describing the depression severity states. When the transition from i to j is not possible, the element ij of the matrix will be FALSE. According to semi-Markov process, the diagonal of this matrix will all be FALSE. We run a hypothesis test for testing whether the sojourn time distributions associated with other transitions are Weibull distributed or Exponential distributed using Kolmogorov-Smirnov Goodness-of-Fit test [60]. If p-value is larger than 0.05, we will not reject the hypothesis that the sojourn times come from an Exponential distribution or Weibull distribution. The element ij of the matrix will be a character "E" or "W" representing the sojourn time distribution accordingly. Transition probability matrices produced from the method in this subsection for increasing group in both 6,067 and 610 dataset are demonstrated in Table 3.8 below.

Table 3.8: Semi-Markov Transition Probability Matrices for Increasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.279	0.29	0.274	0.157
Mi	0.129	0	0.433	0.292	0.145
Mo	0.092	0.362	0	0.329	0.217
MS	0.11	0.277	0.431	0	0.182
S	0.113	0.26	0.299	0.327	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.457	0.258	0.157	0.127
Mi	0.249	0	0.495	0.172	0.084
Mo	0.123	0.424	0	0.283	0.170
MS	0.098	0.298	0.431	0	0.172
S	0.171	0.233	0.280	0.315	0

(b) (610 Subjects)

3.2.5 Hidden Semi Markov Model

Hidden semi-Markov models (HSMMs) are an extension of hidden Markov models (HMMs). As discussed in Section 3.2.3, the sojourn time of standard HMMs is geometrically distributed. However, see for discussions in Section 3.2.4, in some real-world problems the probability of a state change depends on the time spent in the current state, which makes the HMM has limitations in applications. With more general sojourn time distributions, HSMMs allow the underlying process to be a semi-Markov chain. Each state possesses a variable duration associated with the number of observations produced while in the state. This means that, unlike in HMMs where a state can emit one observation per state, in HSMMs, a state can emit a sequence of observations. The length of the observation sequence per state depends on the duration variable of each state. In our data set, the actual observation cycle for PHQ-9 values is irregular. We may hypothesize that the five levels of depression severity given in Table 3.2 are suitable states for out hidden Markov model, but these states might not have geometrically distributed sojourn time. A HSMM model, may be a suitable model for our data.

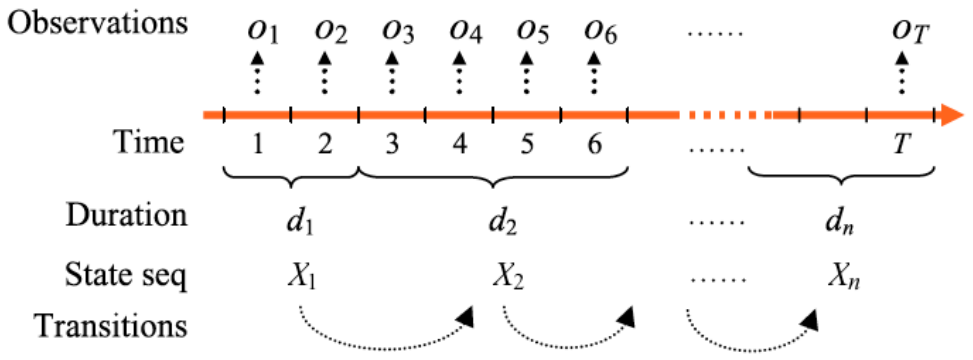

Figure 3.13: Visual Representation of a Hidden semi-Markov Process[10]

The process is visualized in Figure 3.13 [61] as below, where O_T are observed variables,

and X_n are unobserved hidden states. X_1 lasts for $d_1 \geq 2$ time units and produces two observations (O_1, O_2) with the emission probability $b_{X_1,d_1}(O_{1:2})$. Therefore, a HSMMs can be specified by $\theta = (\pi, P, b, d)$. As we only observe the observation sequence without the state sequence, we call EM algorithm to maximize the incomplete data likelihood of HSMM.

Our study is performed based on the *mhsmm* package from O'Connell [62]. We fit a HSMM to the health record data, using the states that are consistent with Table 3.2, {Healthy, Mild, Moderate, Moderately Severe, Major}. We begin by setting reasonable starting values for emission distribution. Based on observed PHQ-9 data, as we are unable to use predefined emission matrix from Table 3.2 within this package for HSMMs. We fit an arbitrary normal emission distribution using maximum likelihood estimation as follows. We assume that our state transitions indicate correlations between the value of PHQ-9 scores and hidden states. Suppose PHQ-9 values $\{"0-4", "5-9", "10-14", "15-19", "20-27"\}$ left by the process represents the corresponding depression severity status {Healthy, Mild, Moderate, Moderately Severe, Major}, and the succeeding PHQ-9 values left by the process are true observations. Select all observations that belong to the same preceding state, and fit the emission distribution for each state separately. We proceed this fashion to all five states.

Furthermore, as we are unsure of the parameters for real states transitions, we assume that the original sojourn times between observed PHQ-9 values are real state duration variables corresponding to five severity states . We have found that Gamma distribution works well for our data using maximum likelihood estimation. Given the other three appropriate parameters that we have initial states distribution obtained from our data, as well as initial states probability transition matrix from discrete-time Markov model, this initial HSMM is used as a generator to give an observation sequence $O = O_1 O_2 \dots O_T$. Now we call EM algorithm to recalculate the parameters of the initial HSMM model using these estimates .

Transition probability matrices generated from the Hidden semi-Markov for increasing

group in both 6,067 and 610 dataset are presented in Table 3.9 below.

Table 3.9: Hidden Semi-Markov Transition Probability Matrices for Increasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.357	0.313	0.243	0.086
Mi	0.135	0	0.469	0.287	0.108
Mo	0.079	0.366	0	0.347	0.209
MS	0.075	0.244	0.450	0	0.231
S	0.055	0.190	0.317	0.438	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.768	0.097	0.063	0.073
Mi	0.267	0	0.536	0.155	0.042
Mo	0.025	0.463	0	0.402	0.109
MS	0.018	0.197	0.486	0	0.299
S	0.092	0.107	0.189	0.611	0

(b) (610 Subjects)

Chapter 4

PERFORMANCE COMPARISON AND VALIDATION

Performance analysis on transition probability matrices obtained from various Markov models are employed to reveal the long-term characteristic on depression progression of the given cohort. Four parameters including steady-state probability, expected first passage time, proportion of time spending in a certain state and spectral norm are computed in Section 4.1. Section 4.2 introduces the Bootstrap validation and cross validation strategy and demonstrates the validation statistics.

4.1 Performance Analysis

4.1.1 Steady-State Probability

While calculating the n-step transition probabilities, if n is large enough, all the rows of the matrix have identical entries. The probability that the system is in each state j no longer depends on the initial state of the system, which means, after a a large number of transitions, this probability is limiting and independent of the initial state. Referring to the transition probability matrices obtained from 3.2.1 and 3.2.2, we estimate the probability of patients being in a certain health state in steady state as follows. For a finite-state, discrete-time Markov chain that is irreducible and aperiodic, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j > 0$ exists and is independent of i, where π_j is defined as steady-state probability [63]. π_j satisfies the following steady-state equations

$$\sum_{j=0}^{M} \pi_j = 1. \tag{4.1}$$

$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}, j = 0, 1, ..., M. \tag{4.2}$$

In matrix form, this equation can be expressed as $\pi = \pi P$, where $\pi = (\pi_0, \pi_1, ..., \pi_M)$. Steady-state probability, also known as stationary probability, defines that the probability of finding the process in a certain state j at time $n = 1, 2, ...n$. After n transitions, this probability tends to the value π_j, which is $P\{X_n = j\} = \pi_j$, and no longer depends on the probability distribution of the initial state. For continuous time transition probability functions, these limiting probabilities also satisfy the equations above. However, stationary probability does not indicate that the process settles down into one state. On the contrary, the process continues transiting from state to state, and at any step n the transition probability from state i to state j is still p_{ij}.

Table 4.1 illustrates the calculation results of stationary probabilities for increasing group in 6,067 and 610 dataset (check Appendix E for full version of computation results). Figure 4.1 and Figure 4.2 vividly present the proportion of the possible long-term states calculated using different Markov models.

Table 4.1: Stationary Probabilities for Increasing Group

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.095	0.334	0.424	0.121	0.026
MSM	0.009	0.136	0.479	0.283	0.093
HMM	0.007	0.15	0.48	0.28	0.084
Semi	0.099	0.234	0.278	0.237	0.153
HSM	0.081	0.226	0.292	0.252	0.15

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.141	0.237	0.310	0.194	0.119
MSM	0.069	0.164	0.29	0.255	0.222
HMM	0.073	0.152	0.28	0.244	0.251
Semi	0.141	0.268	0.284	0.187	0.121
HSMM	0.093	0.264	0.288	0.235	0.12

(b) (610 Subjects)

If taking a closer look at the probabilities, across four models, after many biweeks, the probabilities of finding patients in Moderately Severe and Severe states, are within the range of [0.237, 0.283], [0.084, 1.153] in 6,067 dataset, while [0.187, 0.255], [0.12, 0.251] in 610

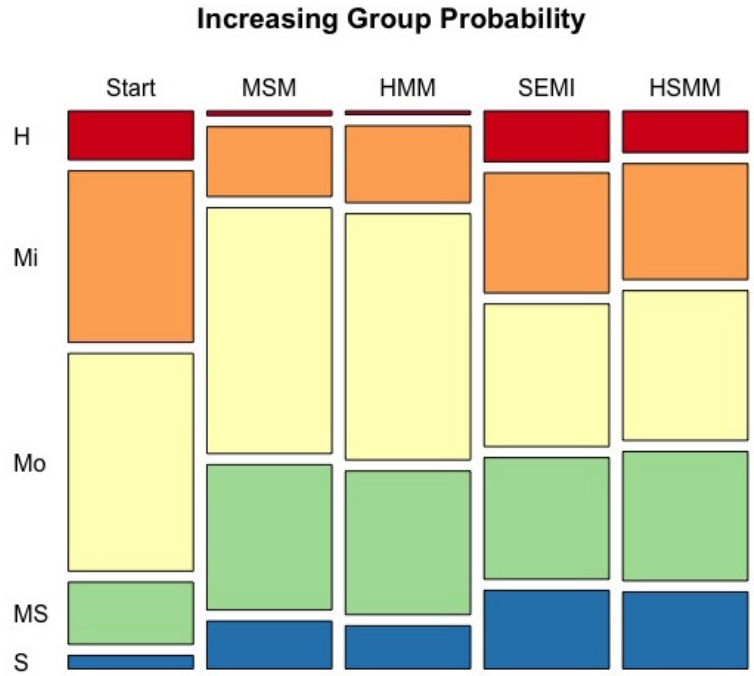

Figure 4.1: Stationary Probability Distribution _ Increasing Group (6067)

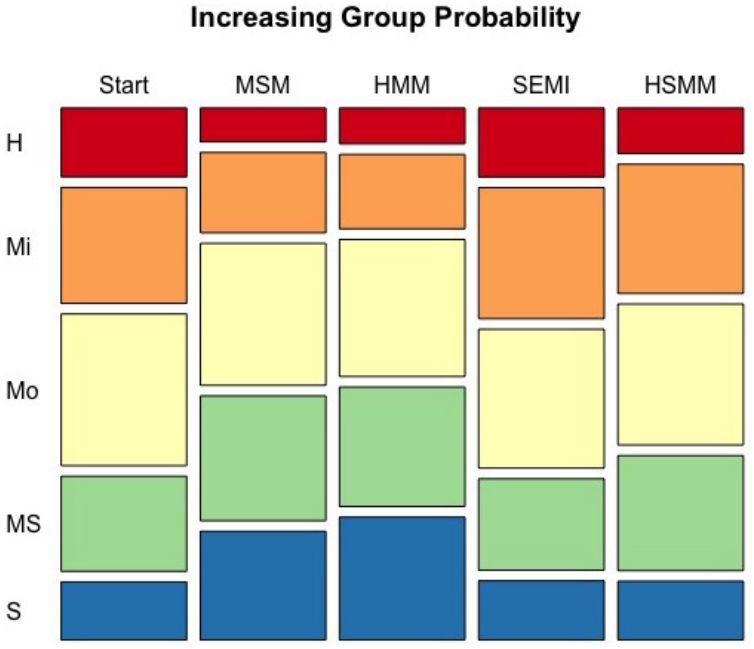

Figure 4.2: Stationary Probability Distribution _ Increasing Group (610)

dataset for increasing group, separately. It is logical to deduce that patients in 610 dataset with higher visiting frequency are more likely to be of higher depression seriousness. To further confirm our implication, we go over all the groups within two datasets, comparing the range of their possibility in each state. The tendency is consistent. However, another possible reason for this lower depression severity in 6,067 dataset might be that individuals' trajectories produced from a smoothed b-spline, where transitions of Discrete-time Markov model and Hidden Markov model tabulated from. It seems to be rather difficult for us to justify the severity of impact between these two causes.

Recall that we have five subgroups of patients as Mild group, Increasing Severity group, Moderate group, Decreasing severity group, and Severe group, who have distinct group-level depression severity progression. From the group-level view point, patients in increasing group and severe group gain no better severity conditions with ongoing treatment. It would be of more beneficial to initiate and develop new treatment guidelines targeting patients in these type of group.

Viewed as a whole, in comparison of the performance of each model across five subgroups, all results from Hidden Markov models are similar to the results from Discrete-time Markov model. Similar patterns are founded on the estimated transition probability matrices as well, indicating that the HMM might unforeseen hidden state transition patterns. Moreover, it seems that two semi-Markov models are more unlikely to detect the heterogeneity across different subgroups in 6,067 dataset, where, in general, the frequency of visiting occurrence is extremely low. In 6,067 dataset, stationary probabilities tend to be evenly distributed, indicating that certain transitions with super long state change sojourn time are probably partially observed. These facts imply that our current assumptions towards Hidden Markov models and semi-Markov models might not be solid based on our dataset.

4.1.2 First Passage Time

In addition to steady-state probabilities, we calculate the number of transitions made by patients in making transitions from depression severity state i to state j for the first time. When $j = i$, the first passage time is the number of transitions until the patients returns to the their initial depression severity state i. In this condition, the first passage time is called recurrence time for state i. In our case, our patients make transitions at standard time interval, which is two weeks. The expected first passage time from state i to state j denoted as μ_{ij} can be used to compute the expected number of bi-weeks until the patients regain their health or become severe. The expectation μ_{ij} can be calculated as

$$\mu_{ij} = 1 + \sum_{k \neq j} p_{ik} u_{kj}. \tag{4.3}$$

When $j = i, u_{ii}$ is the expected number of transitions until patients returns to the initial depression severity state i, which is expected recurrence time for severity state i. We obtain stationary probabilities $(\pi_0, \pi_1, ..., \pi_M)$ from section 3.3.1. These expected recurrence times can be calculated based on steady-state probabilities as

$$\mu_{ii} = \frac{1}{\pi_i}, i = 0, 1...., M. \tag{4.4}$$

Table 4.2 presents the expected first passage time for increasing group in both dataset. Calculated from Discrete-time Markov and Hidden Markov models respectively, we have each cell representing the expected number of steps to transit from all five distinct "From" states based on three "To" states, the Healthy state, Moderately Severe state and Severe state entered by the process.

We can observe that 6,067 dataset produce an unrealistically long time window for most

Table 4.2: First Passage Time for Increasing Group

From	**H**	**Mi**	**Mo**	**MS**	**S**
To H					
MSM	111	2184	2413	2478	2511
HMM	143	2319	2538	2601	2632
To MS					
MSM	166	146	107	4	33
HMM	166	150	109	4	32
To S					
MSM	489	470	431	324	11
HMM	514	497	457	348	12

(a) (6067 Subjects)

From	**H**	**Mi**	**Mo**	**MS**	**S**
To H					
MSM	14	56	74	82	86
HMM	14	53	73	83	88
To MS					
MSM	18	14	10	4	5
HMM	20	16	10	4	6
To S					
MSM	31	27	22	14	5
HMM	35	31	26	17	4

(b) (610 Subjects)

of the transitions. Though tendencies across two dataset are identical, results from 6,067 dataset might be of no clinical value. We run a small test with a less smoothed b-spline within 6.067 patients dataset. The number of steps for each first passage time is significantly decreased. From this aspect, the proposed models are sensitive to the smoothness of b-spline and are more likely to have limitations with less fluctuating trajectories. Due to limited time, in this thesis, we will not place great emphasis on the selection of a sufficient smoothness, which, however, does require future in-depth analysis. Hence in analyzing the result for the first passage time, we will mostly focus on 610 dataset.

On average, patients in Increasing group are expected to reach moderately severe state in less then 20 biweeks, which is less than a year even with ongoing treatment, additionally reach severe state in less then 35 biweeks, which is less then a year and a half. However, the expected time to regain health takes much longer. Taking a look at all tables in Appendix F, records indicate that for 610 dataset, patients in four of the subgroups are expected to reach moderately severe in less than 20 biweeks, where mild group is the only exception. This implies that current visiting frequency and treatment policies and strategies for this

group of patients might not be sufficient. The initiatives of new monitoring and treatment plans are urgent.

4.1.3 Proportion of Time in State i

For semi-Markov process, instead of calculating the first passage time, we calculate the long-run proportion of time in state i, which is the long-term proportion of time that patients stay in a certain state i in our case. Whenever a patient enters state $i, i \in I = \{1, 2, 3, 4, 5\}$: it will enter the next state j with probability $P_{ij}, i, j \in I$; the time until the transition from i to j occurs has distribution F_{ij}. For a semi-Markov process,

$$F_{ij}(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 1. \end{cases} \tag{4.5}$$

For semi-Markov process, denote H_i as the distribution of time it spends in state i before making a transition, and let μ_i denote its mean. By conditioning on the next state, we have

$$H_i(t) = \sum_j P_{ij} F_{ij}(t), \tag{4.6}$$

$$\mu_i = \int_0^\infty x \, \mathrm{d}H_i(x). \tag{4.7}$$

We define T_{ii} as the time between consecutive transitions into state i and let $\mu_{ii} = E[T_{ii}]$. Considering the irreducible semi-Markov process, if T_{ii} has a non-lattice distribution with finite mean, we have P_i exists and independent of the initial state, denoted [63] as long-run proportion of time that the process is in state i.

$$P_i = \lim_{t \to \infty} P\{Z(t) = i \mid Z(0) = j\} = \frac{\mu_i}{\mu_{ii}} \tag{4.8}$$

Furthermore, if $\mu_{ii} < \infty$, then, with probability 1,

$$\frac{\mu_i}{\mu_{ii}} = \lim_{t\to\infty} \frac{\text{amount of time in } i \text{ during } [0,t]}{t} = \frac{\pi_i \mu_i}{\sum_j \pi_j \mu_j}. \tag{4.9}$$

Based on the equations above, the proportion of time in state i, P_i can be calculated from μ_i and π_i. Table 4.3 and Table 4.4 present the proportion of time in state i for increasing group from semi-Markov and Hidden semi-Markov model in both dataset respectively. Each column represents each state and each row represents distinct subgroup. Figure 4.3 and Figure 4.4 illustrate the proportion more vividly.

Table 4.3: Proportion of Time in State i _ Semi-Markov Models

	H	**Mi**	**Mo**	**MS**	**S**
In	0.175	0.306	0.261	0.171	0.086
De	0.149	0.262	0.317	0.195	0.077
H	0.433	0.385	0.128	0.037	0.017
Mo	0.155	0.235	0.282	0.227	0.1
Se	0.03	0.082	0.23	0.365	0.293

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
In	0.129	0.267	0.305	0.178	0.121
De	0.165	0.314	0.31	0.113	0.098
H	0.283	0.35	0.21	0.135	0.021
Mo	0.126	0.242	0.278	0.233	0.121
Se	0.014	0.141	0.27	0.341	0.233

(b) (610 Subjects)

Table 4.4: Proportion of Time in State i _ Hidden Semi-Markov Models

	H	**Mi**	**Mo**	**MS**	**S**
In	0.146	0.302	0.28	0.186	0.086
De	0.167	0.268	0.261	0.195	0.109
H	0.421	0.386	0.135	0.044	0.014
Mo	0.188	0.241	0.237	0.219	0.115
Se	0.06	0.097	0.233	0.348	0.262

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
In	0.085	0.263	0.309	0.223	0.12
De	0.126	0.287	0.297	0.17	0.12
H	0.193	0.306	0.273	0.221	0.006
Mo	0.13	0.253	0.275	0.24	0.102
Se	0.03	0.143	0.288	0.369	0.17

(b) (610 Subjects)

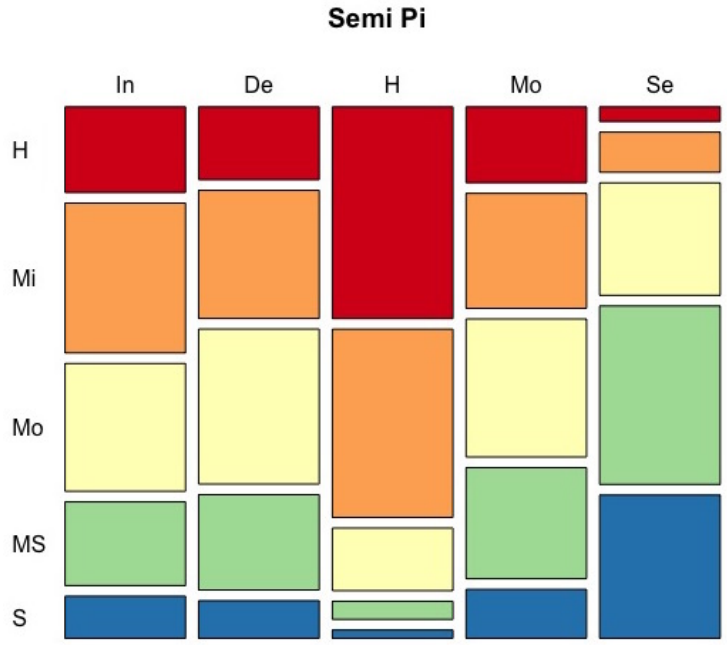

Figure 4.3: Semi-Markov Proportion of Time in State i _ Increasing Group (6067)

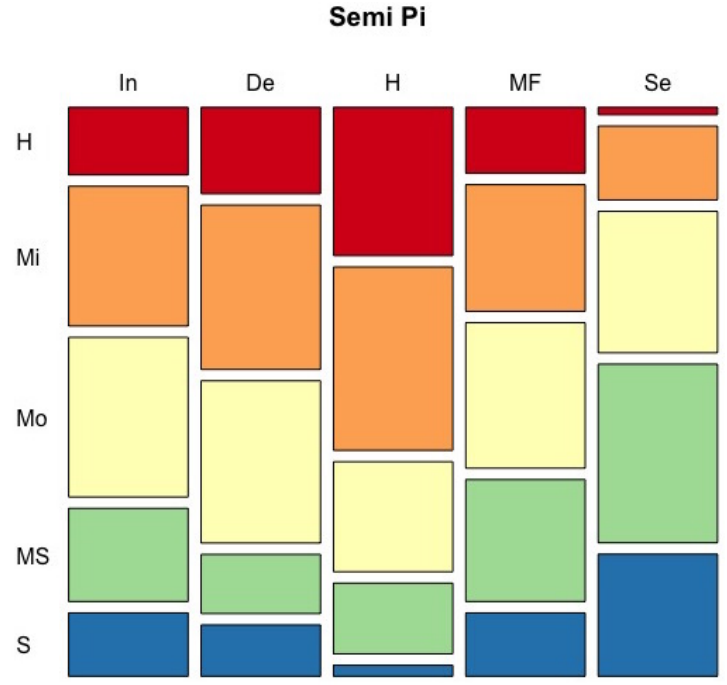

Figure 4.4: Semi-Markov Proportion of Time in State i _ Increasing Group (610)

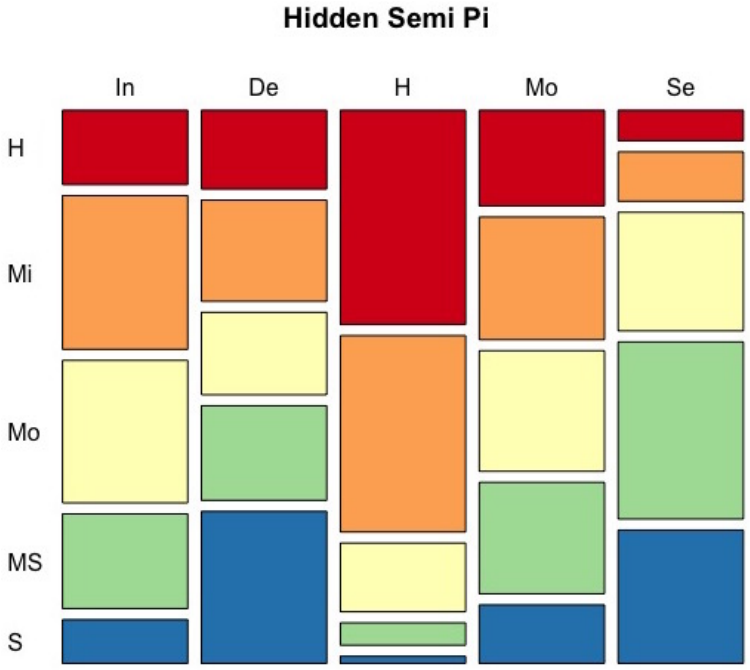

Figure 4.5: Hidden Semi-Markov Proportion of Time in State i _ Increasing Group (6067)

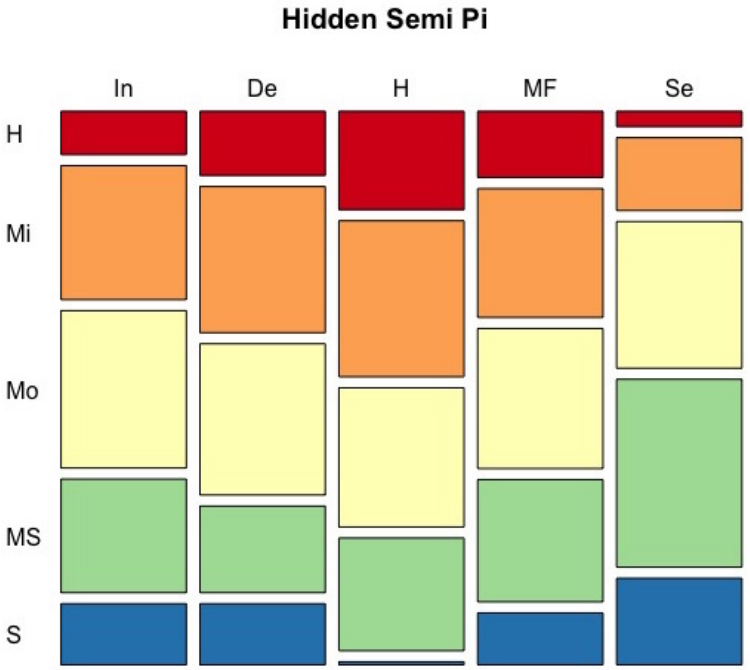

Figure 4.6: Hidden Semi-Markov Proportion of Time in State i _ Increasing Group (610)

Overall, both distributions are consistent with the group level characteristics. Patients in healthy and decreasing severity group spend a larger proportion of time in healthy states, while patients in severe and decreasing severity group spend a larger proportion of time in severe and moderately severe states. Recall that in section 4.1.1, we find that semi-Markov models might not be sufficient to detect heterogeneity across different subgroups according to stationary probabilities. Depending on results in this section, on the contrary, we find that the semi-Markov models reveal the heterogeneity from a sojourn time viewpoint. Patients in severe group spend more than 50% of time with a relatively severe depression symptoms. Correspondence development of health, social and work care guidelines and policies should be initiated targeting this group. Moreover, further investigation can be conducted with a group of patients that has a more realistic sequence of observations.

4.1.4 Spectral Norm

In linear algebra, a norm is a function that assigns a strictly positive length or size to each vector in a vector space. The most popular of all norm is the l_2 norm, also known as a Euclidean norm, which is used as a standard quantity for measuring a vector difference. Matrix norm is a natural extension of the notion of a vector norm to matrices. We say that the matrix norm is subordinate to the vector norm. In special case, the natural norm induced by the L2-norm, the induced matrix norm is the spectral norm, 2-norm. Let A^H be the conjugate transpose of the square matrix A, hence $(a_{ij})^H = (\overline{a}_{ji})$. The spectral norm can be defined as the square root of the maximum eigenvalue of $A^H A$.

$$\| A \|_2 = \sqrt{(\text{maximum eigenvalue of } A^H A)} = \sqrt{\lambda_{max}(A^H A)} = \max_{|x|_2 \neq 0} \frac{| Ax |_2}{| x |_2} \tag{4.10}$$

The spectral norm of a matrix A is the largest singular value of A. In matrix, the distance

between two matrices can be calculated using norm. As we have

$$\| A \|_2 \leq \| A \|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} | a_{ij} |^2} \leq \sqrt{r} \| A \|_2, \tag{4.11}$$

thus we calculate the distance between two matrices using spectral norm.

Table 4.5: Spectral Norm

	In	**De**	**Mi**	**Mo**	**Se**
MSM-HMM	0.01	0.22	0.31	0.01	0.01
SM-HSM	0.17	0.81	0.38	0.7	0.41
MSM-SM	1.38	1.45	1.66	1.39	1.6
HMM-SM	1.38	1.45	1.66	1.39	1.6
MSM-HSM	1.42	1.67	1.67	1.61	1.69
HMM-HSM	1.42	1.65	1.67	1.61	1.69

(a) (6067 Subjects)

	In	**De**	**Mi**	**Mo**	**Se**
MSM-HMM	0.09	0.24	0.51	0.09	0.49
SM-HSM	0.44	0.39	0.78	0.39	0.74
MSM-SM	0.99	1.28	1.42	1.01	1.27
HMM-SM	0.96	1.24	1.4	1.01	1.29
MSM-HSM	1.18	1.46	1.52	1.19	1.41
HMM-HSM	1.13	1.43	1.52	1.18	1.38

(b) (610 Subjects)

Table 4.5 presents the comparison of spectral norm calculation across the distances of two transition probability matrices generated from two different methods. Results in both datasets indicate that distance between discrete-time Markov models with and without hidden states, and distance between semi-Markov models with and without hidden states are closer. Distance between matrices from discrete-time Markov models and semi-Markov models is relatively far, which might imply that the assumption of sojourn distribution generates greater impact towards our results.

4.2 Validation

4.2.1 Bootstrap

The bootstrap is a flexible and powerful statistical technique that can be used to quantify the uncertainty associated with a given estimator or statistical learning method. It can provide

an estimate of the standard error of a coefficient, or a confidence interval for that coefficient. It is not the same as the term "bootstrap" used in computer science meaning to "boot" a computer from a set of core instructions, though the derivation is similar. Bootstrap allow us to mimic the process of obtaining new data sets by repeatedly sampling observations from the original dataset with replacement, rather than from the new population, so that we can estimate the variability of our parameters without generating additional samples. The resampled dataset is the same size as our original dataset, where some observations from original dataset might appear more than once while some not at all.

Bootstrap validation analysis is performed on 610 subjects. The new bootstrap datasets are randomly sampled with replacement within each subgroup the same size as subgroup in original 610 dataset. The PHQ-9 observations for each patient is a time series. We can't simply sample the observations with replacement based on PHQ-9 values. Instead, patients ID within the same subgroup are taken as blocks of consecutive observed PHQ-9 values to sample from with replacement. The we combine all sampled blocks together to obtain a bootstrap dataset. Each bootstrap dataset is used to obtain an estimate of α.

Denoting the first bootstrap dataset for a specific subgroup by Z^{*1}, a new bootstrap estimate for α can be produced using Z^{*1}, which denote as $\hat{\alpha}^{*1}$. The resampling procedure is repeated B times, 10,000 times in our validation process, to generate B bootstrap data sets, $Z^{*1}, Z^{*2}, \ldots, Z^{*B}$ and corresponding $\hat{\alpha}^{*1}, \hat{\alpha}^{*2}, \ldots, \hat{\alpha}^{*B}$. The standard error of these bootstrap estimates is estimated as

$$\mathrm{SE}_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}(\hat{\alpha}^{*r} - \hat{\alpha}^{*})^2}, \tag{4.12}$$

which serves as an estimate of the standard error of $\hat{\alpha}$ estimated from the original dataset. The approximate confidence intervals for our estimated parameter are obtained from the

5% and 95% quantiles of the 1000 values, which represent an approximate 90% confidence interval for the true α. This confidence interval is called a Bootstrap Percentile confidence interval.

4.2.2 Cross Validation

With limited available dataset, researchers want to learn a model from the whole dataset so as to maximally exploit the information behind the data, where leaves no available data to evaluate the accuracy of the model. Cross validation is often used to estimate the performance of a model by resampling and holding out a validation subset from the training observations, and then employing statistical learning method to these held out observations to obtain additional information, like the prediction error, the standard deviation, and the bias of parameters of interest, about the fitted model.

Recall that in Section 3.2.3, the observed PHQ-9 counts serve as both observations and actual state transitions in the construction of HMM. In this section, a 2-fold 5-repeated cross validation is performed on 610 subjects to disjoin the observation transitions and state transitions hence estimate the model performance. The original 610 dataset is randomly split into two equal-sized subsets: a training set k, where k equals to 2 in our case, to initialize the HMM, and a validation or hold-out set, the combined *k-1* parts, containing the remaining observations serving as the observation group to recalculate the parameters of the initial HMM. Then in turn, this process is done for each part $k-1, 2, ...K$. Similar with resampling process of bootstrap, patients ID within the same subgroup are taken as blocks of consecutive observed PHQ-9 values to sample from, instead of simply sampling the PHQ-9 values within each subgroup.

4.2.3 Validation Statistics

Given that 6,067 dataset using smooth *b*-spline might produce an unrealistically flat individual's disease trajectory, we omit this dataset in the validation part.

In bootstrap validation analysis, we repeated the simulation process for 10,000 times over each subgroup, and estimating the stationary probabilities 10,000 times using Discrete-time Markov model. Table 4.6 presents the original stationary probabilities, the mean stationary probability over all 10,000 estimates, the confidence intervals as well as the standard error of the mean of the original stationary probabilities for increasing group from 610 dataset. During the validation process, we calculate the standard deviation as well. With all the standard deviations being less than 0.0001, we will not discuss them in this thesis. Bootstrap validation results for the rest subgroups are attached in Appendix G.

To reduce variability, we repeat the 2-fold cross validation process for 5 times over each subgroup for Hidden-Markov process. Table 4.7 presents the original stationary probabilities, the 5% and 95% quantiles, the 90% confidence intervals as well as the standard error of the mean of the original stationary probabilities and the standard deviations for increasing group from 610 dataset. The rest of the cross validation results are attached in Appendix G.

Table 4.6: Bootstrap Validation Statistics _ Increasing group from 610 Subjects

Increasing	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.069	0.164	0.29	0.255	0.222
Mean	0.069	0.164	0.289	0.255	0.223
CI	(0.043, 0.092)	(0.133, 0.195)	(0.249, 0.332)	(0.216, 0.291)	(0.149, 0.279)
Std.Err	0.012	0.016	0.021	0.019	0.033

From Table 4.6, we can find that the mean stationary probabilities over 10,000 times are almost the same with our original outputs for all five subgroups. Across five subgroups, the standard error are less than 0.04 except for probabilities for healthy state in decreasing

group, which equals to 0.042. The overall results imply that the proposed model might be a tangible approach for Discrete-time Markov model with a less smoothed trajectory. Further validations can be designed to test the sensitivities of our proposed model and the boundaries.

Table 4.7: Cross Validation Statistics _ Increasing group from 610 Subjects

Increasing	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.073	0.152	0.28	0.244	0.251
Quantile	(0.061, 0.086)	(0.123, 0.166)	(0.212, 0.253)	(0.183, 0.26)	(0.3, 0.375)
CI	(0.06, 0.085)	(0.138, 0.181)	(0.307, 0.348)	(0.228, 0.305)	(0.127, 0.202)
Std.Dev	0.008	0.015	0.014	0.024	0.029
Std.Err	0.003	0.005	0.004	0.008	0.009

Taking a look at the cross validation results across five subgroups, we can find that several original stationary probabilities fall within neither 5% and 95% quantiles nor the 90% confidence interval. Across five subgroups, the standard errors are less than 0.04 except for probabilities for healthy state in moderate group, which equals to 0.053. Compared to bootstrap validation results, where all the standard deviation are less than 0.0001, standard deviation results from cross validation process are relatively large, ranging from 0.00025 to 0.012. The overall results imply that the original approach to initiate and estimate our HMM has its limitations in disjoining the observation transitions and hidden states transitions which might unforeseen hidden state transition patterns. Further validations can be designed through a more realistic sequence of observations and solider state transition knowledges.

Chapter 5

CONCLUSIONS

5.1 Summary

Abundant evidence on cost-effectiveness analyses presented that major depression is of high relevance to high medical utilization, direct medical costs and even higher indirect economic costs, let along the suffering patients with pervasive and persistent low mood. However, prospective and effective prescription of major depression can reduce episode incidence in those patients at risk of recurrence, reduce the risk of worsening the depression severity as well as reduce the duration of depressive episodes. This thesis focused on the modeling and analyzing of depression progression dynamics at population level to address insights on the duration and occurrence possibility of major depressive episodes, which help in the understanding of depression progression in the American population with undergoing treatment, and shed some lights on justifying initiatives and changes to mental disease management strategies, relevant medicare policies and population level depression surveillance guidance.

Provided by the MHRN, we track the progression of patients' PHQ-9 counts from EHR data as the primary variable for further analysis. Two cohorts of patients with ongoing treatment were selected, where one dataset includes 6,067 patients with no less than 4 PHQ-9 counts over the entire observation period, plus a dataset that have more than six observations within a 20 bi-week time window. Techniques and progression modeling methodologies are then performed on both dataset in comparing the patterns for patients with different visiting frequencies. The irregular observed PHQ-9 counts are translated to a complete longitudinal history of the patient's disease trajectory by a b-spline curving method. Individuals in 6,067

dataset are fit by smoothed b-spline with a degree of freedom equals to 4 and smoothing parameter equals 1,000, while individuals in 610 dataset are fit by non-smoothed b-spline. Prior to depression progression modeling, we perform a k-means cluster analysis to detect the heterogeneity of the given cohort in terms of their trajectory patterns. Patients are then sorted into five sub-groups in both datasets: Severe, Decreasing Severity, Moderate, Increasing Severity and Mild group, which are consistent with previous cohort clustering research from Gunn et al.

In our proposed four Markov models, transition probability matrices are generated from the modeling of probabilistic progression of patient's severity state transitions. In accordance with the characteristic of different types of Markov model, we make correspondence assumptions. The simulated Markov traces generated from Markov models indicate a consistent long-term probabilistic transitions tendency with the simulated mean curve obtained from K-means at group level. We have to admit that it is relatively hard to address explicit implications directly from transition probability matrices. Therefore, estimations of several featured parameters was performed to analyze the transition probability matrices, which aim to provide insight on the duration and occurrence possibility of major depressive episodes. From this base, our computation of stationary probability, expected first passage time and the proportion of time in a certain state extends the estimation region to any depression severity state with 5 categories.

To demonstrate the calculation accuracy of long-term parameters, bootstrap validation is adopted to mimic the process of obtaining new datasets by repeatedly sampling observations from the original dataset with replacement, rather than generate new population, which lower both the validation cost and time. Parameters like mean, standard deviation and percentile confidence interval are derived from a simulation of 10,000 times bootstrapping.

5.2 Limitations

Several limitations lie within our study. Bounded to the limited data source, the observed PHQ-9 counts serve as both observations and actual state transitions in Hidden Markov model, which might unforeseen state transition patterns since almost all the estimated transition probability matrices are similar to the initializations. It is difficult for us to justify the validity of the model performance from those calculated long-term parameters generated from this stochastic matrix, unless a more realistic sequence of observations are then provided. Though cross-validation analysis is employed to demonstrate the modeling accuracy, still, observations and states are sampled from patients with homogeneity within the same subgroup.

Semi-Markov models take time intervals between observed PHQ-9 counts as real observed state transitions, and Hidden semi-Markov models take those as time intervals between transition of observations. In comparison of stationary probabilities, it seems that semi-Markov models are unlikely to detect the heterogeneity across different subgroups in 6,067 dataset, where, in general, the frequency of visiting occurrence is extremely low. It is logical to deduce that certain transitions with super long state change sojourn time are probably partially observed.

In validation part, in estimating prediction error using bootstrap, we take each bootstrap dataset as our training sample. The original sample is taken as a validation sample. Unlike cross-validation that each of the K validation folds is distinct from the other $K-1$ training folds, each bootstrap sample has significant overlap with the validation sample, which may cause an underestimation of the true prediction error using bootstrap. By only using predictions for those observations that did not occur in the current bootstrap sample can be a tangible approach to partly fix this problem.

5.3 Recommendations for Future Research

Referring to the actual PHQ-9 counts, patients' PHQ-9 measurements vary significantly over time. However, in 6,067 dataset, individuals' trajectories produced from smoothing b-spline, in most case, are almost flat, which are more likely to neglect depression severity fluctuations or major depression episodes. 610 dataset incorporates patients with higher visiting frequency is curved using b-spline without smoothness. We add an additional normal distributed, $N(0, 3)$, random error, and run a small test on these trajectories prior to adopt unsmoothed trajectories. Tests generate relatively more extreme results, which bring up the necessity to conduct future sensitivity analysis and examination towards the selection of appropriate and suitable b-spline parameters.

In addition, numerous evidence [64–67] from previous research indicate that the age and sex difference might affect the rate of progression, which might in relevance with their occupation and social roles. Calculation of the transition intensities can be specified for each covariate of interest hence compare the new intensities with the "averaged" modeling result without covariates, which leads to a feasible estimation on the influences of a certain covariate towards patients' depression progression and the risk of major depression episodes more specifically.

BIBLIOGRAPHY

[1] T. B. Üstün, J. L. Ayuso-Mateos, S. Chatterji, C. Mathers, and C. J. Murray, "Global burden of depressive disorders in the year 2000," *The British Journal of Psychiatry*, vol. 184, no. 5, pp. 386–392, 2004.

[2] W. H. Organization, *The World Health Report 2001: Mental Health: New Understanding, New Hope*. World Health Organization, 2001.

[3] R. C. Kessler, P. Berglund, O. Demler, R. Jin, D. Koretz, K. R. Merikangas, A. J. Rush, E. E. Walters, and P. S. Wang, "The epidemiology of major depressive disorder: results from the national comorbidity survey replication (ncs-r)," *Jama*, vol. 289, no. 23, pp. 3095–3105, 2003.

[4] D. G. Jacobs, "National depression screening day: educating the public, reaching those in need of treatment, and broadening professional understanding," *Harvard Review of Psychiatry*, vol. 3, no. 3, pp. 156–159, 1995.

[5] M. Dwight-Johnson, J. Unutzer, C. Sherbourne, L. Tang, and K. B. Wells, "Can quality improvement programs for depression in primary care address patient preferences for treatment?," *Medical Care*, vol. 39, no. 9, pp. 934–944, 2001.

[6] A. J. Dietrich, T. E. Oxman, J. W. Williams, H. C. Schulberg, M. L. Bruce, P. W. Lee, S. Barry, P. J. Raue, J. J. Lefever, M. Heo, *et al.*, "Re-engineering systems for the treatment of depression in primary care: cluster randomised controlled trial," *Bmj*, vol. 329, no. 7466, p. 602, 2004.

[7] U. P. S. T. Force *et al.*, "Screening for depression: recommendations and rationale.," *Annals of Internal Medicine*, vol. 136, no. 10, p. 760, 2002.

[8] N. C. C. for Mental Health (UK *et al.*, "Depression: the treatment and management of depression in adults (updated edition)," British Psychological Society, 2010.

[9] W. Katon, M. Von Korff, E. Lin, J. Unützer, G. Simon, E. Walker, E. Ludman, and T. Bush, "Population-based care of depression: effective disease management strategies to decrease prevalence," *General Hospital Psychiatry*, vol. 19, no. 3, pp. 169–178, 1997.

[10] M. van den Berg, F. Smit, T. Vos, and P. H. van Baal, "Cost-effectiveness of opportunistic screening and minimal contact psychotherapy to prevent depression in primary care patients," *PLoS One*, vol. 6, no. 8, p. e22884, 2011.

[11] P. Cuijpers and F. Smit, "Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies," *Acta Psychiatrica Scandinavica*, vol. 109, no. 5, pp. 325–331, 2004.

[12] P. Cuijpers and F. Smit, "[subclinical depression: a clinically relevant condition?]," *Tijdschrift Voor Psychiatrie*, vol. 50, no. 8, pp. 519–528, 2007.

[13] G. Simon, J. Ormel, M. VonKorff, and W. Barlow, "Health care costs associated with depressive and anxiety disorders in primary care," *American Journal of Psychiatry*, vol. 152, no. 3, pp. 352–357, 1995.

[14] A. Stoudemire, R. Frank, N. Hedemark, M. Kamlet, and D. Blazer, "The economic burden of depression," *General Hospital Psychiatry*, vol. 8, no. 6, pp. 387–394, 1986.

[15] P. E. Greenberg, R. C. Kessler, H. G. Birnbaum, S. A. Leong, S. W. Lowe, P. A. Berglund, and P. K. Corey-Lisle, "The economic burden of depression in the united states: how did it change between 1990 and 2000?," *Journal of Clinical Psychiatry*, vol. 64, no. 12, pp. 1465–1475, 2003.

[16] H. C. Schulberg, J. L. Coulehan, M. R. Block, J. Lave, E. Rodriguez, C. P. Scott, M. J. Madonia, S. Imber, and J. Perel, "Clinical trials of primary care treatments for major depression: issues in design, recruitment and treatment," *The International Journal of Psychiatry in Medicine*, vol. 23, no. 1, pp. 29–42, 1993.

[17] C. M. Callahan, H. C. Hendrie, R. S. Dittus, D. C. Brater, S. L. Hui, and W. M. Tierney, "Improving treatment of late life depression in primary care: a randomized clinical trial," *Journal of the American Geriatrics Society*, vol. 42, no. 8, pp. 839–846, 1994.

[18] D.-. A. P. Association *et al.*, "Diagnostic and statistical manual of mental disorders," *Arlington: American Psychiatric Publishing*, 2013.

[19] K. Kroenke and R. L. Spitzer, "The phq-9: a new depression diagnostic and severity measure," *Psychiatric Annals*, vol. 32, no. 9, pp. 509–515, 2002.

[20] K. Kroencke, R. Spitzer, and J. Williams, "The phq-9: Validity of a brief depression severity measure [electronic version]," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–13, 2001.

[21] J. M. Pyne, K. M. Rost, F. Farahati, S. P. Tripathi, J. Smith, D. K. Williams, J. Fortney, and J. C. Coyne, "One size fits some: the impact of patient treatment attitudes on the cost-effectiveness of a depression primary-care intervention," *Psychological Medicine*, vol. 35, no. 06, pp. 839–854, 2005.

[22] L. B. Sheiner and D. B. Rubin, "Intention-to-treat analysis and the goals of clinical trials," *Clinical Pharmacology & Therapeutics*, vol. 57, no. 1, pp. 6–15, 1995.

[23] D. Mould, N. Denman, and S. Duffull, "Using disease progression models as a tool to detect drug effect," *Clinical Pharmacology and Therapeutics*, vol. 82, no. 1, pp. 81–86, 2007.

[24] B. A. Ploeger and N. H. Holford, "Washout and delayed start designs for identifying disease modifying effects in slowly progressive diseases using disease progression analysis," *Pharmaceutical Statistics*, vol. 8, no. 3, pp. 225–238, 2009.

[25] H. C. Kimko, S. S. Reele, N. H. Holford, and C. C. Peck, "Prediction of the outcome of a phase 3 clinical trial of an antischizophrenic agent (quetiapine fumarate) by simulation with a population pharmacokinetic and pharmacodynamic model," *Clinical Pharmacology & Therapeutics*, vol. 68, no. 5, pp. 568–577, 2000.

[26] W. De Winter, J. DeJongh, T. Post, B. Ploeger, R. Urquhart, I. Moules, D. Eckland, and M. Danhof, "A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 33, no. 3, pp. 313–343, 2006.

[27] S. Beal, "The nonmem system," *The American Statistician*, vol. 34, pp. 118–119, 1980.

[28] P. Chan and N. Holford, "Drug treatment effects on disease progression," *Annual review of Pharmacology and Toxicology*, vol. 41, no. 1, pp. 625–659, 2001.

[29] D. Mould, "Models for disease progression: new approaches and uses," *Clinical Pharmacology & Therapeutics*, vol. 92, no. 1, pp. 125–131, 2012.

[30] N. H. Holford and L. B. Sheiner, "Understanding the dose-effect relationship," *Clinical Pharmacokinetics*, vol. 6, no. 6, pp. 429–453, 1981.

[31] T. M. Post, J. I. Freijer, J. DeJongh, and M. Danhof, "Disease system analysis: basic disease progression models in degenerative disease," *Pharmaceutical Research*, vol. 22, no. 7, pp. 1038–1049, 2005.

[32] K. Ito, S. Ahadieh, B. Corrigan, J. French, T. Fullerton, T. Tensfeldt, A. D. W. Group, *et al.*, "Disease progression meta-analysis model in alzheimer's disease," *Alzheimer's & Dementia*, vol. 6, no. 1, pp. 39–53, 2010.

[33] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via fused sparse group lasso," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 1095–1103, ACM, 2012.

[34] K. P. Exarchos, T. P. Exarchos, C. V. Bourantas, M. I. Papafaklis, K. K. Naka, L. K. Michalis, O. Parodi, and D. I. Fotiadis, "Prediction of coronary atherosclerosis progression using dynamic bayesian networks," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 3889–3892, IEEE, 2013.

[35] J. Zhou, J. Liu, V. A. Narayan, J. Ye, A. D. N. Initiative, *et al.*, "Modeling disease progression via multi-task learning," *NeuroImage*, vol. 78, pp. 233–248, 2013.

[36] N. Holford, J. Li, L. Benincosa, and M. Birath, "Population disease progress models for the time course of hamd score in depressed patients receiving placebo in anti-depressant clinical trials," in *XI Annual Meeting of the Population Approach Group in Europe. Abstr*, vol. 311, 2002.

[37] E. Y. Shang, M. A. Gibbs, J. Landen, N. G. Denman, M. Krams, T. Russell, and D. R. Mould, "Model based analysis of placebo response in major depression.," in *Journal of Clinical Pharmacology*, vol. 46, pp. 1073–1073, Sage Publications Inc 245 Teller Rd, Thousand Oaks, CA 91320 USA, 2006.

[38] S. B. Patten and R. C. Lee, "Describing the longitudinal course of major depression using markov models: Data integration across three national surveys," *Population Health Metrics*, vol. 3, no. 1, p. 11, 2005.

[39] K. S. Oskooyee, A. M. Rahmani, and M. M. R. Kashani, "Predicting the severity of major depression disorder with the markov chain model," in *Proceedings of the International Conference on Bioscience, Biochemistery and Bioinformatics*, vol. 5, 2011.

[40] F. Cocker, J. M. Nicholson, N. Graves, B. Oldenburg, A. J. Palmer, A. Martin, J. Scott, A. Venn, and K. Sanderson, "Depression in working adults: comparing the costs and health outcomes of working when ill," *PloS one*, vol. 9, no. 9, p. e105430, 2014.

[41] S. Bhattacharya, "Markov chain model to explain the dynamics of human depression," *Journal of Nonlinear Dynamics*, vol. 2014, 2014.

[42] C. H. Jackson and L. D. Sharples, "Hidden markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients," *Statistics in medicine*, vol. 21, no. 1, pp. 113–128, 2002.

[43] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto, "Multistate markov models for disease progression with classification error," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, no. 2, pp. 193–209, 2003.

[44] R. Sukkar, E. Katz, Y. Zhang, D. Raunig, and B. T. Wyman, "Disease progression modeling using hidden markov models," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 2845–2848, IEEE, 2012.

[45] C. Guihenneuc-Jouyaux, S. Richardson, and I. M. Longini, "Modeling markers of disease progression by a hidden markov process: application to characterizing cd4 cell decline," *Biometrics*, vol. 56, no. 3, pp. 733–741, 2000.

[46] N. Bartolomeo, P. Trerotoli, and G. Serio, "Progression of liver cirrhosis to hcc: an application of hidden markov model," *BMC Medical Research Methodology*, vol. 11, no. 1, p. 1, 2011.

[47] J. C. Hendriks, G. A. Satten, I. M. Longini, H. A. van Druten, P. T. A. Schellekens, R. A. Coutinho, and G. J. van Griensven, "Use of immunological markers and continuous-time markov models to estimate progression of hiv infection in homosexual men.," *AIDS*, vol. 10, no. 6, pp. 649–656, 1996.

[48] A. Bureau, S. Shiboski, and J. P. Hughes, "Applications of continuous time hidden markov models to the study of misclassified disease outcomes," *Statistics in Medicine*, vol. 22, no. 3, pp. 441–462, 2003.

[49] Y. Foucher, E. Mathieu, P. Saint-Pierre, J. Durand, and J. Daures, "A semi-markov model based on generalized weibull distribution with an illustration for hiv disease," *Biometrical Journal*, vol. 47, no. 6, p. 825, 2005.

[50] K. P. D. of Research (Northern California), "Mental health research network. (n.d.)." Available at `http://hcsrn.org/mhrn/en/` accessed 11 May 2016.

[51] S. M. Shechter, *When to initiate, when to switch, and how to sequence HIV therapies: A Markov decision process approach*. PhD thesis, University of Pittsburgh, 2006.

[52] P. H. Eilers and B. D. Marx, "Flexible smoothing with b-splines and penalties," *Statistical Science*, pp. 89–102, 1996.

[53] J. Gunn, P. Elliott, K. Densley, A. Middleton, G. Ambresin, C. Dowrick, H. Herrman, K. Hegarty, G. Gilchrist, and F. Griffiths, "A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care," *Journal of Affective Disorders*, vol. 148, no. 2, pp. 338–346, 2013.

[54] S. Liu, "Analysis of depression trajectory patterns using collaborative learning," in *37th Annual Meeting of the Society for Medical Decision Making*, Smdm, 2015.

[55] M. Stamp, "A revealing introduction to hidden markov models," *Department of Computer Science San Jose State University*, 2004.

[56] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[57] R. Pyke, "Markov renewal processes: definitions and preliminary properties," *The Annals of Mathematical Statistics*, pp. 1231–1242, 1961.

[58] W. Weibull, "Wide applicability," *Journal of Applied Mechanics*, vol. 103, p. 33, 1951.

[59] A. Listwon and P. Saint-Pierre, "Semimarkov: an r package for parametric estimation in multi-state semi-markov models," 2013.

[60] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[61] S.-Z. Yu, "Hidden semi-markov models," *Artificial Intelligence*, vol. 174, no. 2, pp. 215–243, 2010.

[62] J. O'Connell, S. Højsgaard, *et al.*, "Hidden semi markov models for multiple observation sequences: The mhsmm package for r," *Journal of Statistical Software*, vol. 39, no. 4, pp. 1–22, 2011.

[63] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*. Boston: McGraw-Hill, 2010.

[64] A. Jorm, "Sex and age differences in depression: a quantitative synthesis of published research," *Australian and New Zealand Journal of Psychiatry*, vol. 21, no. 1, pp. 46–53, 1987.

[65] M. M. Weissman, R. Bland, P. R. Joyce, S. Newman, J. E. Wells, and H.-U. Wittchen, "Sex differences in rates of depression: cross-national perspectives," *Journal of Affective Disorders*, vol. 29, no. 2, pp. 77–84, 1993.

[66] J. M. Twenge and S. Nolen-Hoeksema, "Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: A meta-analysis.," *Journal of Abnormal Psychology*, vol. 111, no. 4, p. 578, 2002.

[67] T. Covassin, R. J. Elbin III, E. Larson, and A. P. Kontos, "Sex and age differences in depression and baseline sport-related concussion neurocognitive performance and symptoms," *Clinical Journal of Sport Medicine*, vol. 22, no. 2, pp. 98–104, 2012.

Appendix A

PHQ-9 SAMPLE QUESTIONAIRE

Nine Symptom Checklist

Over the last 2 weeks, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things.....	0	1	2	3
2. Feeling down, depressed, or hopeless.....	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much.....	0	1	2	3
4. Feeling tired or having little energy.....	0	1	2	3
5. Poor appetite or overeating.....	0	1	2	3
6. Feeling bad about yourself - or that you are a failure or have let yourself or your family down.....	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television.....	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual.....	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way.....	0	1	2	3

(For office coding: Total Score ____ = ___ + ___ + ___)

If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult
☐	☐	☐	☐

From the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues. For research information, contact Dr. Spitzer at rls8@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc. Copyright© 1999 Pfizer Inc. All rights reserved. Reproduced with permission.

Figure A.1: PHQ-9

Appendix B

ADDITIONAL AGGREGATED TRANSITION COUNTS

Table B.1: Aggregated Transition Counts for Decreasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	4178	42	0	0	0
Mi	488	23654	106	0	0
Mo	1	1147	29571	52	0
MS	0	1	764	8818	10
S	0	0	0	199	1062

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	298	43	0	0	0
Mi	99	640	61	0	0
Mo	1	152	661	50	0
MS	0	2	137	217	14
S	0	0	5	86	134

(b) (610 Subjects)

Table B.2: Aggregated Transition Counts for Mild Group

	H	**Mi**	**Mo**	**MS**	**S**
H	65572	547	0	0	0
Mi	954	53976	154	0	0
Mo	0	387	4856	11	0
MS	0	0	58	428	3
S	0	0	0	3	51

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	1049	94	0	0	0
Mi	103	863	68	0	0
Mo	0	69	290	29	0
MS	0	0	24	86	2
S	0	0	0	1	2

(b) (610 Subjects)

Table B.3: Aggregated Transition Counts for Moderate Group

	H	**Mi**	**Mo**	**MS**	**S**
H	3100	128	0	0	0
Mi	132	16951	362	0	0
Mo	0	382	26793	254	0
MS	0	0	250	9906	33
S	0	0	0	33	674

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	182	53	10	0	0
Mi	83	324	95	5	0
Mo	3	136	480	115	1
MS	0	6	132	456	76
S	0	0	1	84	358

(b) (610 Subjects)

Table B.4: Aggregated Transition Counts for Severe Group

	H	**Mi**	**Mo**	**MS**	**S**
H	20	4	0	0	0
Mi	5	541	40	0	0
Mo	0	76	12403	333	0
MS	0	0	512	42119	334
S	0	0	0	495	22228

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	12	22	2	0	0
Mi	2	86	52	0	0
Mo	0	26	476	136	2
MS	0	0	120	1366	202
S	0	0	0	190	1786

(b) (610 Subjects)

Appendix C

ADDITIONAL TRANSITION PROBABILITY MATRICES

Table C.1: Discrete-Time Markov Transition Probability Matrices for Decreasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.990	0.010	0	0	0
Mi	0.020	0.976	0.004	0	0
Mo	0	0.037	0.961	0.002	0
MS	0	0	0.080	0.919	0.001
S	0	0	0	0.158	0.842

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.874	0.126	0	0	0
Mi	0.124	0.800	0.076	0	0
Mo	0.001	0.176	0.765	0.058	0
MS	0	0.005	0.370	0.586	0.038
S	0	0	0.022	0.382	0.596

(b) (610 Subjects)

Table C.2: Discrete-Time Markov Transition Probability Matrices for Mild Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.992	0.008	0	0	0
Mi	0.017	0.980	0.003	0	0
Mo	0	0.074	0.924	0.002	0
MS	0	0	0.119	0.875	0.006
S	0	0	0	0.056	0.944

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.918	0.082	0	0	0
Mi	0.100	0.835	0.066	0	0
Mo	0	0.178	0.747	0.075	0
MS	0	0	0.214	0.768	0.018
S	0	0	0	0.333	0.667

(b) (610 Subjects)

Table C.3: Discrete-Time Markov Transition Probability Matrices for Moderate Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.960	0.040	0	0	0
Mi	0.008	0.972	0.021	0	0
Mo	0	0.014	0.977	0.009	0
MS	0	0	0.025	0.972	0.003
S	0	0	0	0.047	0.953

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.743	0.216	0.041	0	0
Mi	0.164	0.639	0.187	0.010	0
Mo	0.004	0.185	0.653	0.156	0.001
MS	0	0.009	0.197	0.681	0.113
S	0	0	0.002	0.190	0.808

(b) (610 Subjects)

Table C.4: Discrete-Time Markov Transition Probability Matrices for Severe Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.833	0.167	0	0	0
Mi	0.009	0.923	0.068	0	0
Mo	0	0.006	0.968	0.026	0
MS	0	0	0.012	0.980	0.008
S	0	0	0	0.022	0.978

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.333	0.611	0.056	0	0
Mi	0.014	0.614	0.371	0	0
Mo	0	0.041	0.744	0.213	0.003
MS	0	0	0.071	0.809	0.120
S	0	0	0	0.096	0.904

(b) (610 Subjects)

Table C.5: Hidden Markov Transition Probability Matrices for Decreasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.991	0.009	0	0	0
Mi	0.020	0.976	0.004	0	0
Mo	0	0.036	0.962	0.002	0
MS	0	0	0.110	0.889	0.001
S	0	0	0	0.313	0.687

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.882	0.118 0	0 0		
Mi	0.122	0.795	0.083	0	0
Mo	0.002	0.298	0.484	0.216	0
MS	0	0.011	0.165	0.749	0.074
S	0	0	0.022	0.244	0.734

(b) (610 Subjects)

Table C.6: Hidden Markov Transition Probability Matrices for Mild Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.992	0.008	0	0	0
Mi	0.017	0.981	0.003	0	0
Mo	0	0.069	0.928	0.003	0
MS	0	0	0.139	0.859	0.002
S	0	0	0	0.272	0.728

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.921	0.079	0	0	0
Mi	0.103	0.81	0.087	0	0
Mo	0	0.27	0.546	0.185	0
MS	0	0	0.162	0.773	0.064
S	0	0	0	0.141	0.859

(b) (610 Subjects)

Table C.7: Hidden Markov Transition Probability Matrices for Moderate Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.958	0.042	0	0	0
Mi	0.008	0.973	0.020	0	0
Mo	0	0.014	0.977	0.009	0
MS	0	0	0.026	0.970	0.003
S	0	0	0	0.043	0.957

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.195	0.033	0	0	
Mi	0.156	0.651	0.188	0.005	0
Mo	0.003	0.163	0.682	0.151	0
MS	0	0.006	0.17	0.724	0.1
S	0	0	0	0.163	0.837

(b) (610 Subjects)

Table C.8: Hidden Markov Transition Probability Matrices for Severe Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0.840	0.160	0.000	0.000	0
Mi	0.008	0.925	0.067	0	0
Mo	0.000	0.006	0.968	0.026	0.000
MS	0	0.000	0.012	0.981	0.008
S	0	0.000	0.000	0.023	0.977

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0.567	0.424	0.009	0	0
Mi	0.036	0.705	0.259	0	0
Mo	0	0.035	0.772	0.193	0
MS	0	0	0.059	0.836	0.105
S	0	0	0	0.079	0.921

(b) (610 Subjects)

Table C.9: Semi Markov Transition Probability Matrices for Decreasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.424	0.280	0.175	0.121
Mi	0.336	0	0.405	0.171	0.088
Mo	0.183	0.520	0	0.216	0.080
MS	0.117	0.358	0.411	0	0.114
S	0.120	0.229	0.328	0.323	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.634	0.264	0.054	0.048
Mi	0.381	0	0.488	0.074	0.057
Mo	0.135	0.526	0 0.182	0.158	
MS	0.110	0.294	0.483	0	0.113
S	0.095	0.133	0.350	0.422	0

(b) (610 Subjects)

Table C.10: Semi Markov Transition Probability Matrices for Mild Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.674	0.209	0.077	0.040
Mi	0.694	0 0.208	0.062	0.037	
Mo	0.406	0.507	0	0.058	0.029
MS	0.355	0.401	0.203	0	0.041
S	0.330	0.365	0.176	0.130	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.668	0.120	0.179	0.032
Mi	0.550	0 0.378	0.064	0.008	
Mo	0.279	0.519	0	0.174	0.027
MS	0.245	0.284	0.407	0	0.064
S	0.302	0.400	0.198	0.101	0

(b) (610 Subjects)

Table C.11: Semi Markov Transition Probability Matrices for Moderate Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.302	0.242	0.286	0.171
Mi	0.197	0	0.402	0.271	0.131
Mo	0.140	0.369	0	0.352	0.140
MS	0.122	0.261	0.446	0	0.172
S	0.119	0.245	0.303	0.333	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.478	0.266	0.180	0.075
Mi	0.293	0	0.380	0.203	0.124
Mo	0.118	0.392	0	0.344	0.146
MS	0.114	0.256	0.451	0	0.180
S	0.045	0.185	0.283	0.487	0

(b) (610 Subjects)

Table C.12: Semi Markov Transition Probability Matrices for Severe Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.028	0.804	0.069	0.100
Mi	0.058	0	0.238	0.375	0.329
Mo	0.038	0.128	0	0.458	0.376
MS	0.028	0.085	0.332	0	0.555
S	0.036	0.097	0.221	0.645	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.110	0.554	0.223	0.113
Mi	0.035	0	0.329	0.312	0.324
Mo	0.017	0.093	0	0.655	0.235
MS	0.005	0.125	0.497	0	0.373
S	0.028	0.417	0.196	0.360	0

(b) (610 Subjects)

Table C.13: Hidden Semi-Markov Transition Probability Matrices for Decreasing Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.713	0.250	0.030	0.007
Mi	0.508	0	0.420	0.058	0.014
Mo	0.190	0.624	0	0.148	0.038
MS	0.024	0.112	0.205	0	0.659
S	0.009	0.025	0.068	0.898	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.879	0.071	0.016	0.033
Mi	0.423	0	0.482	0.065	0.029
Mo	0.029	0.48	0	0.341	0.151
MS	0.017	0.142	0.567	0	0.274
S	0.051	0.065	0.244	0.64	0

(b) (610 Subjects)

Table C.14: Hidden Semi-Markov Transition Probability Matrices for Mild Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.692	0.219	0.072	0.017
Mi	0.687	0	0.224	0.067	0.021
Mo	0.385	0.511	0	0.077	0.027
MS	0.273	0.351	0.228	0	0.149
S	0.170	0.227	0.164	0.440	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.818	0.058	0.122	0.002
Mi	0.522	0	0.415	0.062	0
Mo	0.083	0.419	0	0.493	0.005
MS	0.076	0.126	0.768	0	0.029
S	0.061	0.069	0.294	0.576	0

(b) (610 Subjects)

Table C.15: Hidden Semi-Markov Transition Probability Matrices for Moderate Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.750	0.175	0.062	0.013
Mi	0.493	0	0.357	0.119	0.031
Mo	0.128	0.377	0	0.355	0.140
MS	0.035	0.120	0.370	0	0.475
S	0.012	0.053	0.198	0.737	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.643	0.2	0.113	0.043
Mi	0.368	0	0.443	0.154	0.035
Mo	0.096	0.474	0	0.368	0.063
MS	0.066	0.175	0.456	0	0.302
S	0.026	0.052	0.134	0.788	0

(b) (610 Subjects)

Table C.16: Hidden Semi-Markov Transition Probability Matrices for Severe Group

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.067	0.895	0.021	0.017
Mi	0.258	0	0.339	0.250	0.153
Mo	0.141	0.270	0	0.363	0.225
MS	0.020	0.068	0.276	0	0.636
S	0.016	0.059	0.161	0.764	0

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
H	0	0.7	0.277	0.02	0.002
Mi	0.176	0	0.518	0.2	0.105
Mo	0.024	0.213	0	0.683	0.08
MS	0.001	0.116	0.501	0	0.381
S	0.003	0.235	0.103	0.66	0

(b) (610 Subjects)

Appendix D

ADDITIONAL MARKOV TRACES

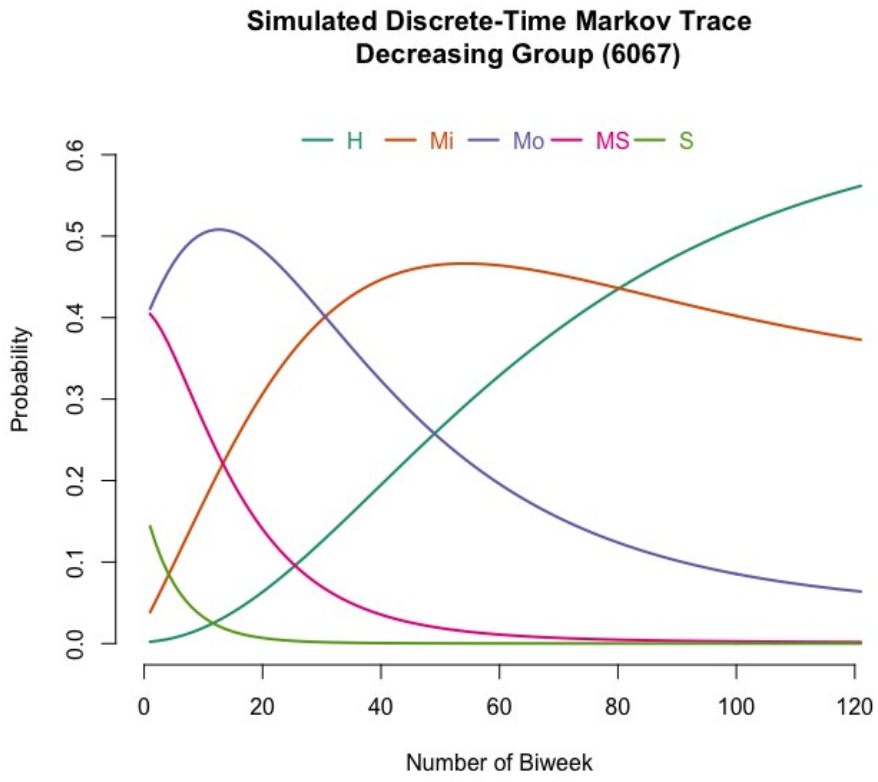

Figure D.1: Simulated Discrete-Time Markov Trace _ Decreasing Group (6067)

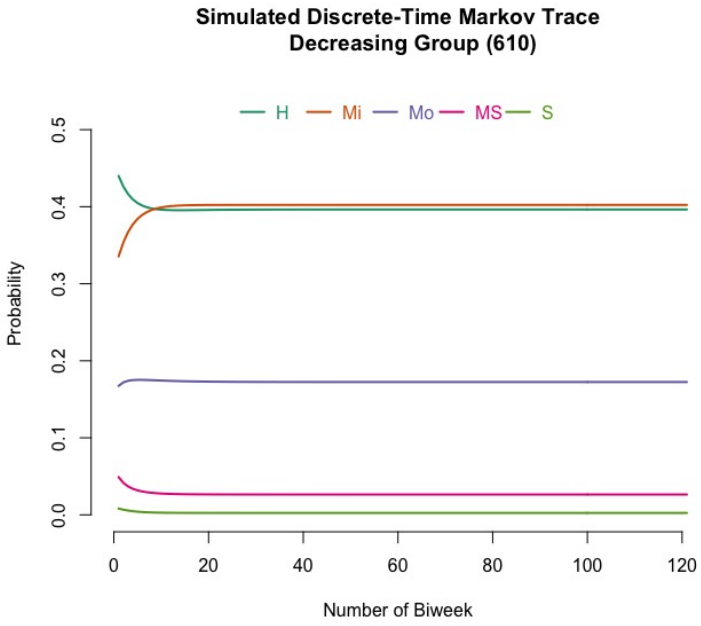

Figure D.2: Simulated Discrete-Time Markov Trace _ Decreasing Group (610)

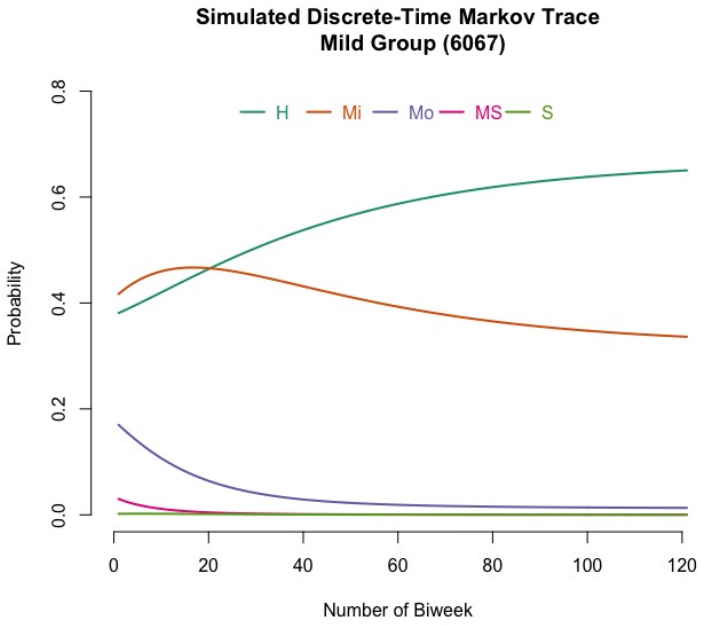

Figure D.3: Simulated Discrete-Time Markov Trace _ Mild Group (6067)

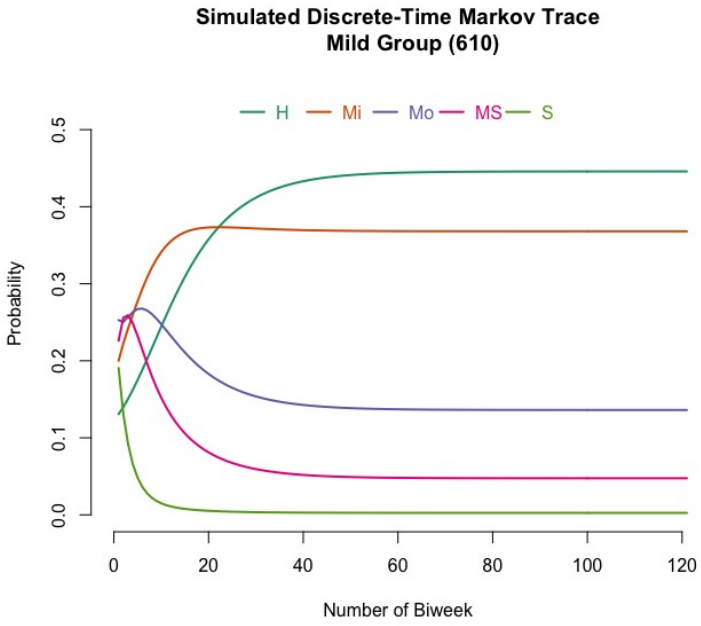

Figure D.4: Simulated Discrete-Time Markov Trace _ Mild Group (610)

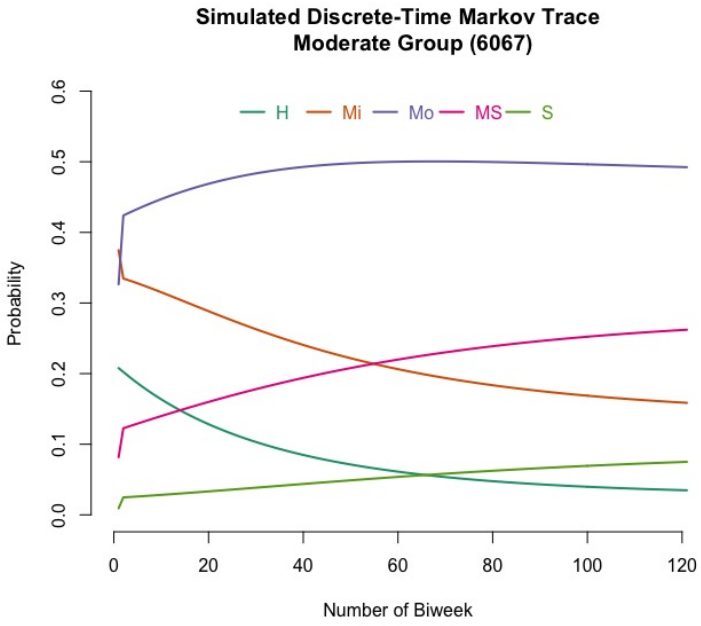

Figure D.5: Simulated Discrete-Time Markov Trace _ Moderate Group (6067)

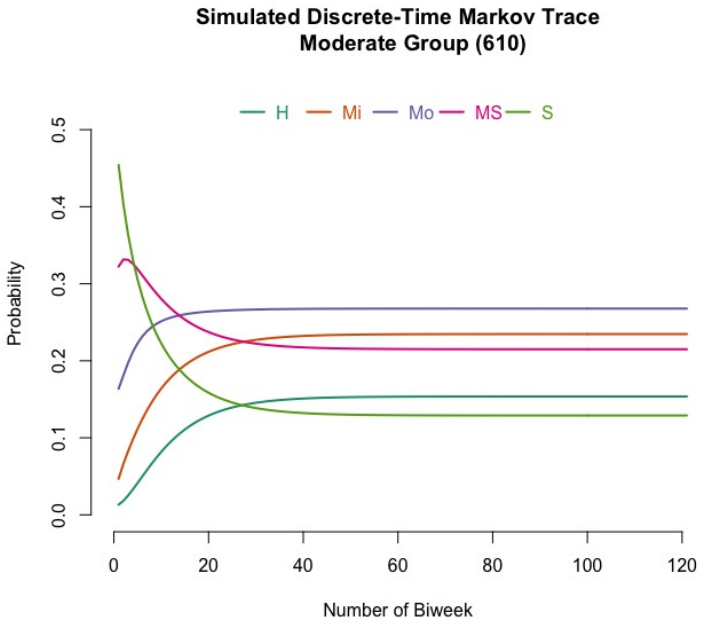

Figure D.6: Simulated Discrete-Time Markov Trace _ Moderate Group (610)

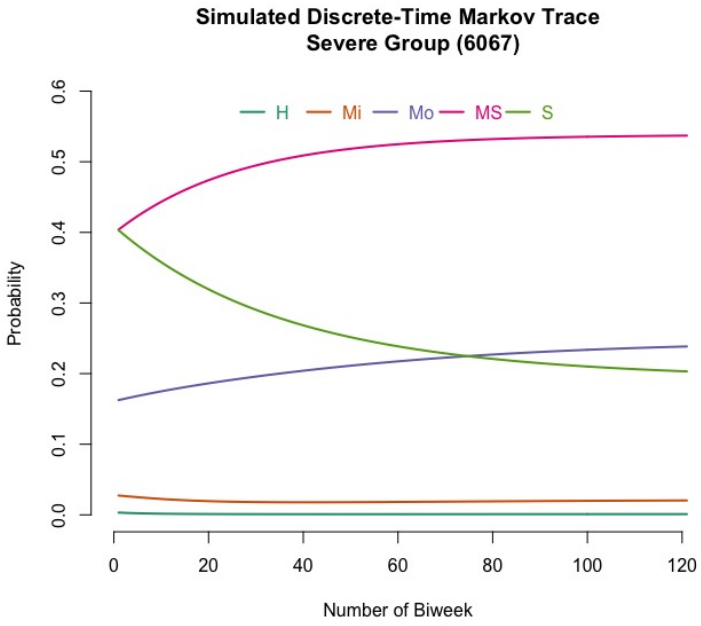

Figure D.7: Simulated Discrete-Time Markov Trace _ Severe Group (6067)

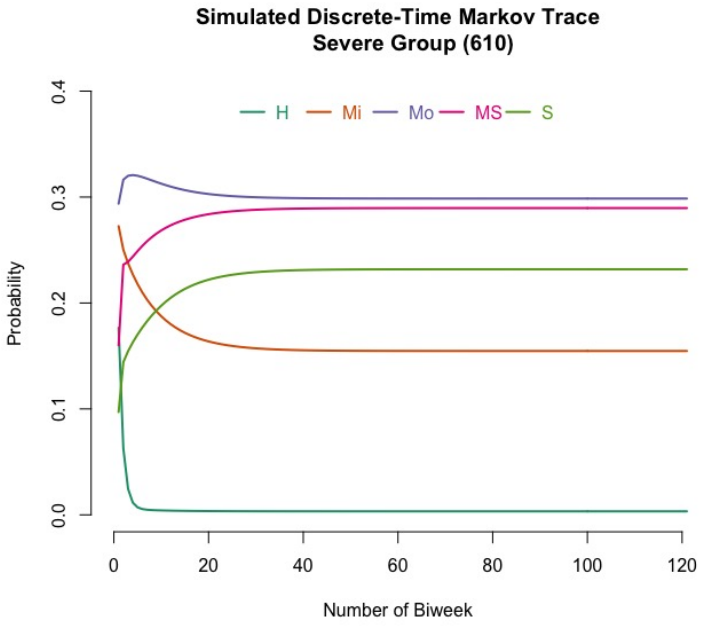

Figure D.8: Simulated Discrete-Time Markov Trace _ Severe Group (610)

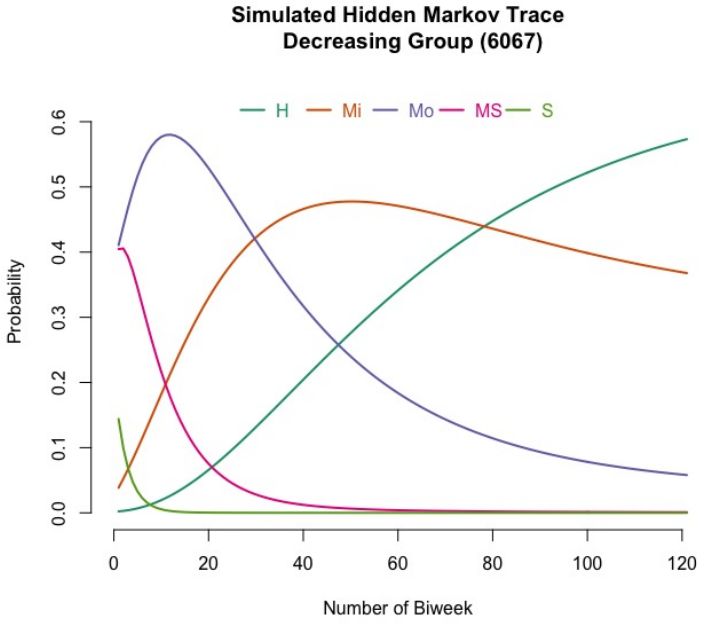

Figure D.9: Simulated Hidden Markov Trace _ Decreasing Group (6067)

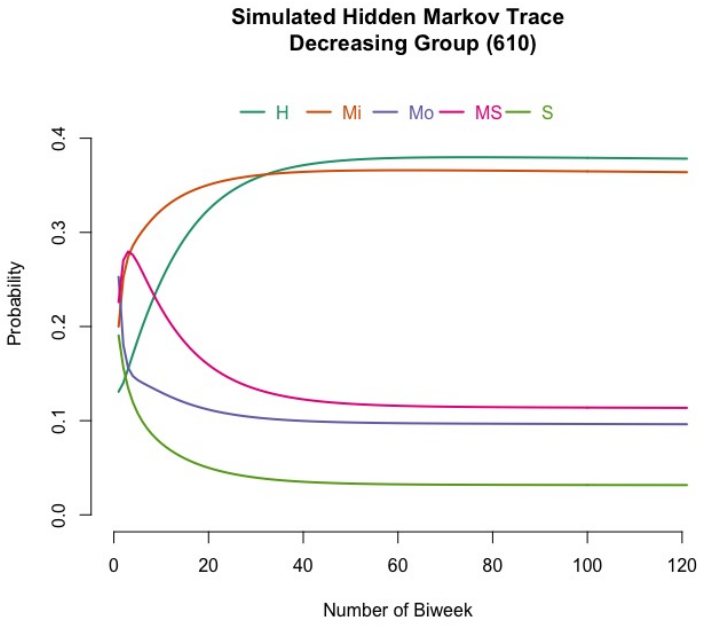

Figure D.10: Simulated Hidden Markov Trace _ Decreasing Group (610)

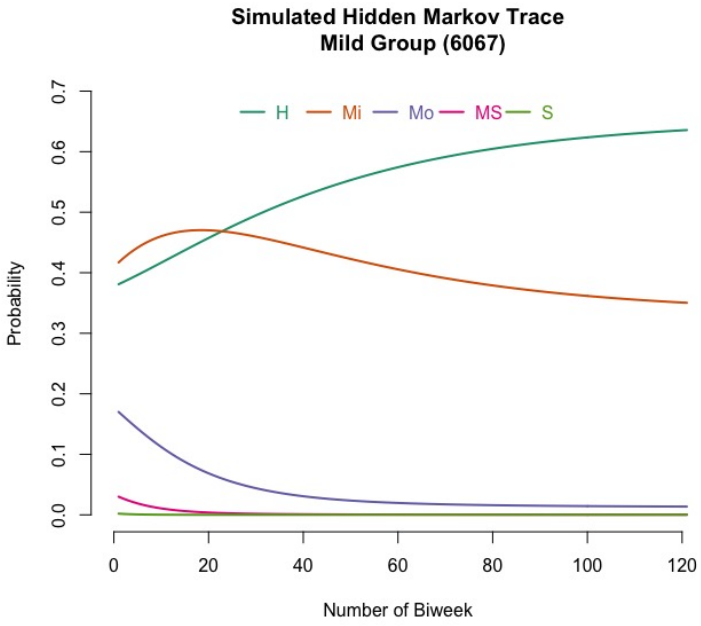

Figure D.11: Simulated Hidden Markov Trace _ Mild Group (6067)

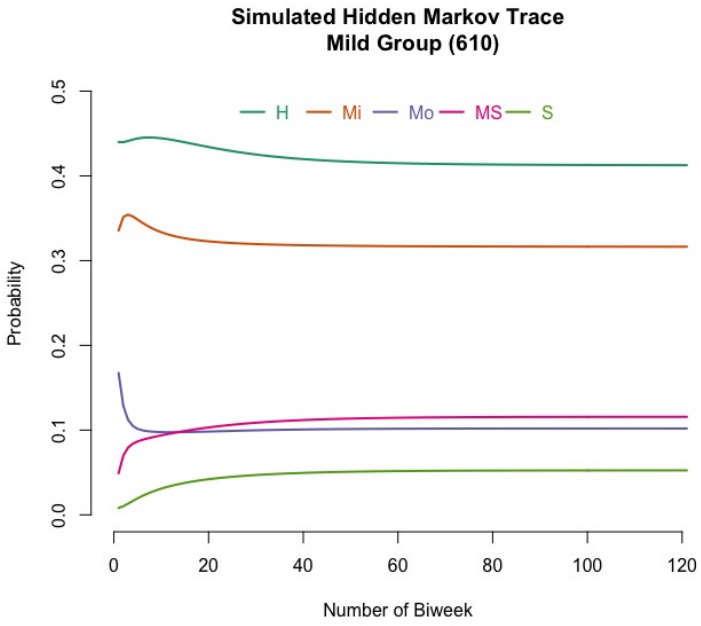

Figure D.12: Simulated Hidden Markov Trace _ Mild Group (610)

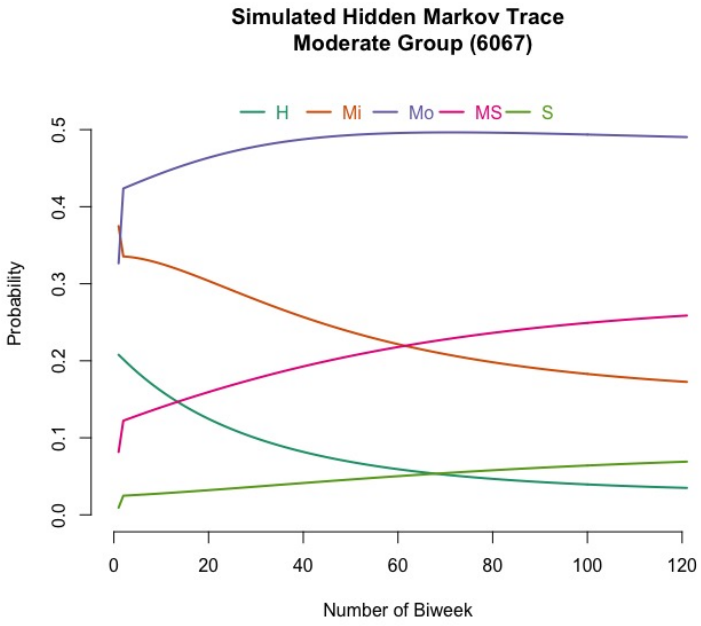

Figure D.13: Simulated Hidden Markov Trace _ Moderate Group (6067)

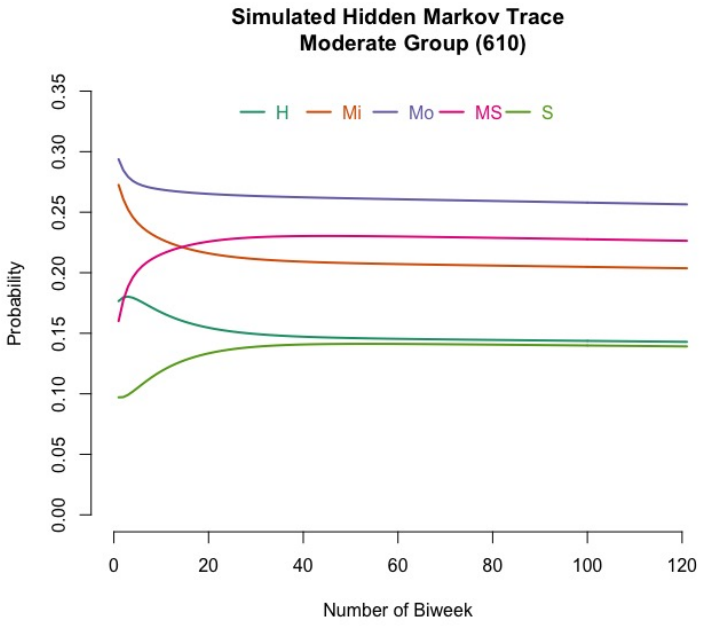

Figure D.14: Simulated Hidden Markov Trace _ Moderate Group (610)

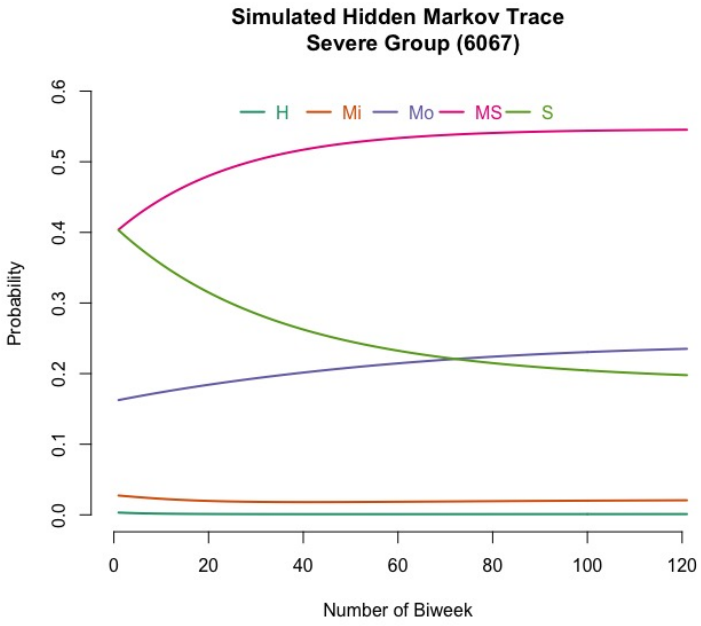

Figure D.15: Simulated Hidden Markov Trace _ Severe Group (6067)

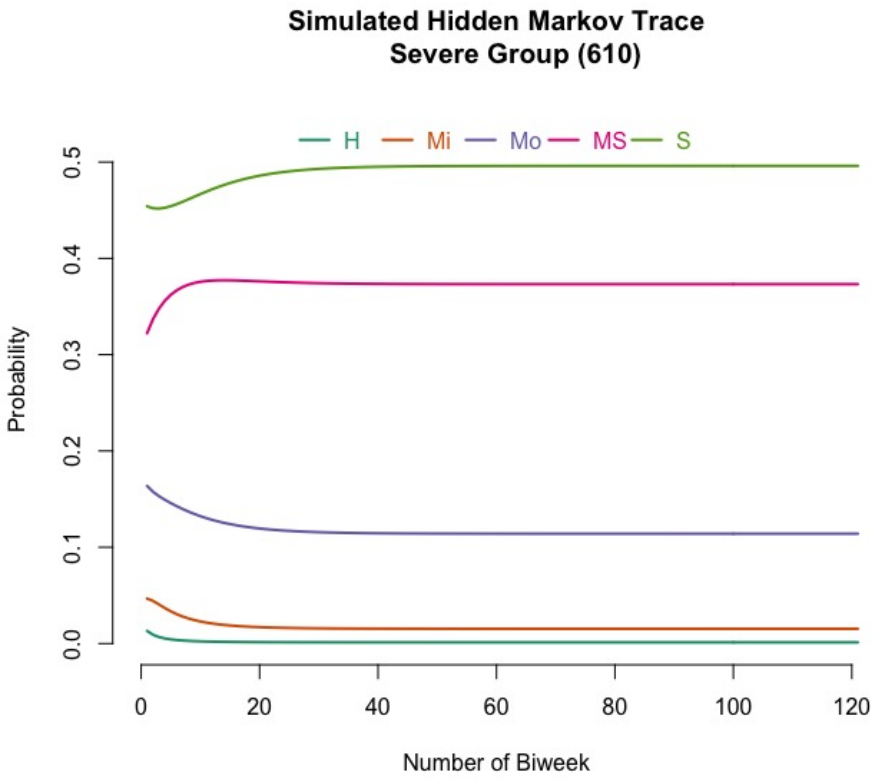

Figure D.16: Simulated Hidden Markov Trace _ Severe Group (610)

Appendix E

ADDITIONAL STATIONARY PROBABILITY RESULTS

Table E.1: Stationary Probabilities for Decreasing Group

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.002	0.038	0.411	0.405	0.144
MSM	0.644	0.318	0.037	0.001	0
HMM	0.656	0.311	0.033	0	0
Semi	0.179	0.296	0.268	0.169	0.089
HSM	0.197	0.297	0.217	0.166	0.123

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.146	0.308	0.324	0.14	0.083
MSM	0.396	0.402	0.172	0.026	0.002
HMM	0.382	0.369	0.099	0.117	0.033
Semi	0.182	0.317	0.292	0.123	0.087
HSM	0.139	0.29	0.28	0.185	0.106

(b) (610 Subjects)

Table E.2: Stationary Probabilities for Mild Group

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.381	0.417	0.17	0.03	0.002
MSM	0.668	0.319	0.012	0	0
HMM	0.654	0.333	0.013	0	0
Semi	0.36	0.368	0.171	0.065	0.036
HSM	0.347	0.366	0.18	0.077	0.03

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.434	0.379	0.144	0.042	0.001
MSM	0.446	0.368	0.136	0.047	0.003
HMM	0.412	0.317	0.102	0.116	0.053
Semi	0.29	0.35	0.219	0.115	0.026
HSM	0.201	0.311	0.29	0.191	0.008

(b) (610 Subjects)

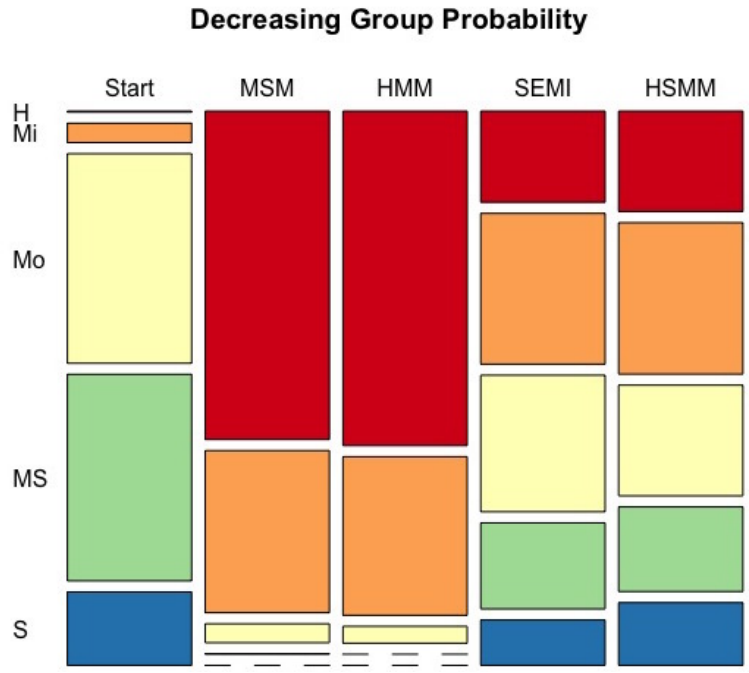

Figure E.1: Stationary Probability Distribution _ Decreasing Group (6067)

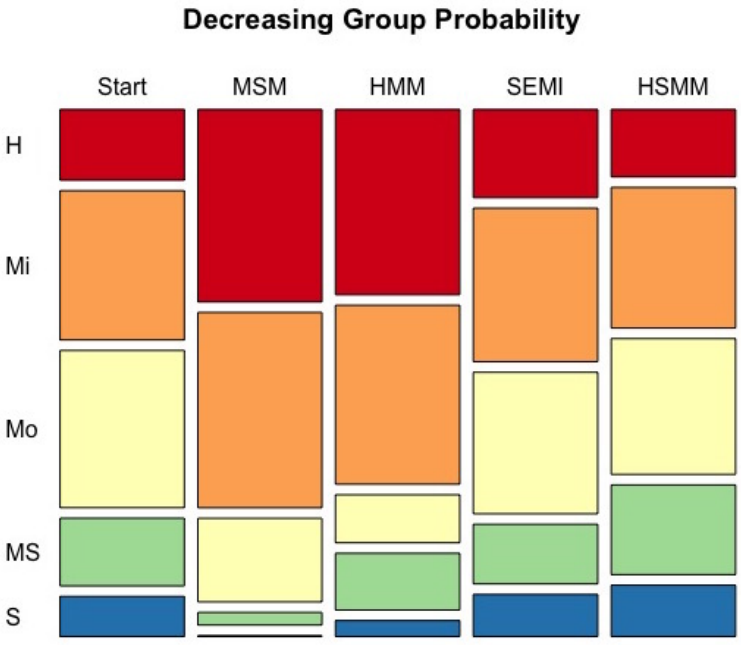

Figure E.2: Stationary Probability Distribution _ Decreasing Group (610)

Table E.3: Stationary Probabilities for Moderate Group

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.208	0.375	0.327	0.082	0.009
MSM	0.058	0.305	0.454	0.171	0.012
HMM	0.059	0.327	0.447	0.155	0.013
Semi	0.128	0.232	0.27	0.238	0.131
HSM	0.156	0.238	0.227	0.229	0.15

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.111	0.197	0.277	0.249	0.166
MSM	0.154	0.235	0.268	0.215	0.129
HMM	0.148	0.21	0.265	0.233	0.144
Semi	0.136	0.25	0.267	0.226	0.121
HSM	0.14	0.261	0.264	0.233	0.102

(b) (610 Subjects)

Table E.4: Stationary Probabilities for Severe Group

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.003	0.027	0.163	0.404	0.403
MSM	0.001	0.021	0.247	0.538	0.192
HMM	0.001	0.022	0.243	0.546	0.188
Semi	0.035	0.089	0.23	0.339	0.307
HSM	0.07	0.105	0.231	0.32	0.273

(a) (6067 Subjects)

	H	**Mi**	**Mo**	**MS**	**S**
Start	0.008	0.032	0.144	0.369	0.448
MSM	0	0.014	0.126	0.381	0.479
HMM	0.001	0.015	0.113	0.372	0.498
Semi	0.018	0.164	0.267	0.315	0.235
HSM	0.037	0.166	0.285	0.341	0.171

(b) (610 Subjects)

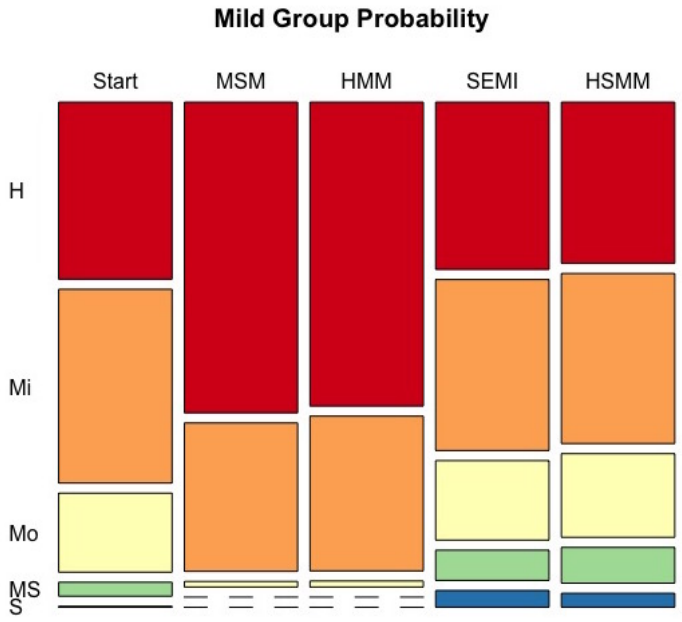

Figure E.3: Stationary Probability Distribution _ Mild Group (6067)

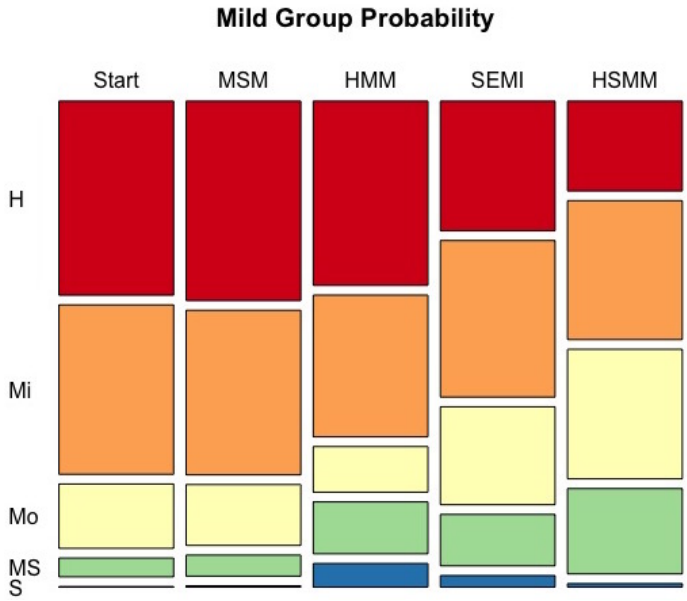

Figure E.4: Stationary Probability Distribution _ Mild Group (610)

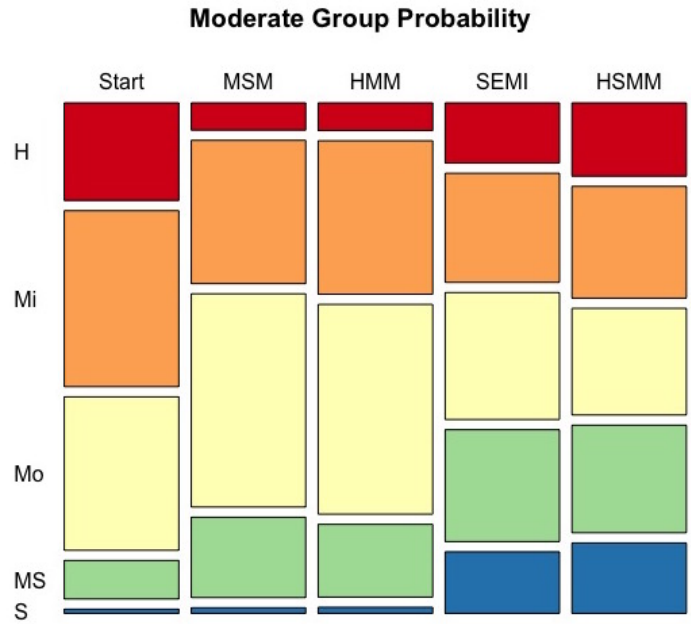

Figure E.5: Stationary Probability Distribution _ Moderate Group (6067)

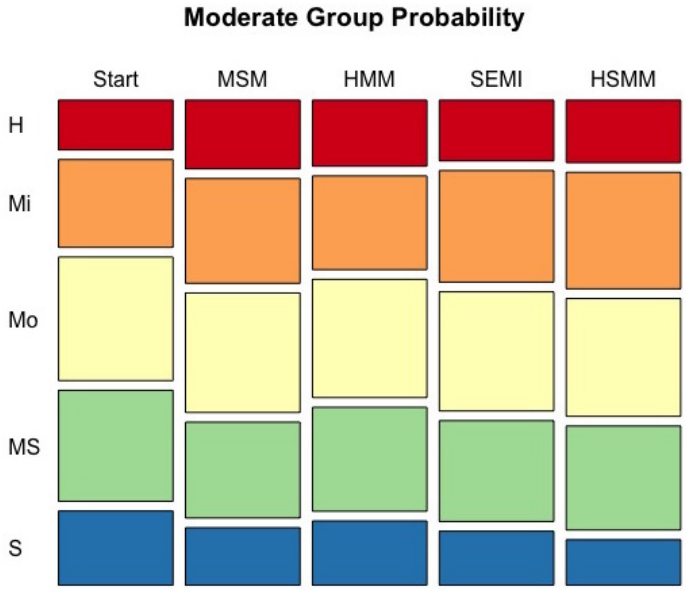

Figure E.6: Stationary Probability Distribution _ Moderate Group (610)

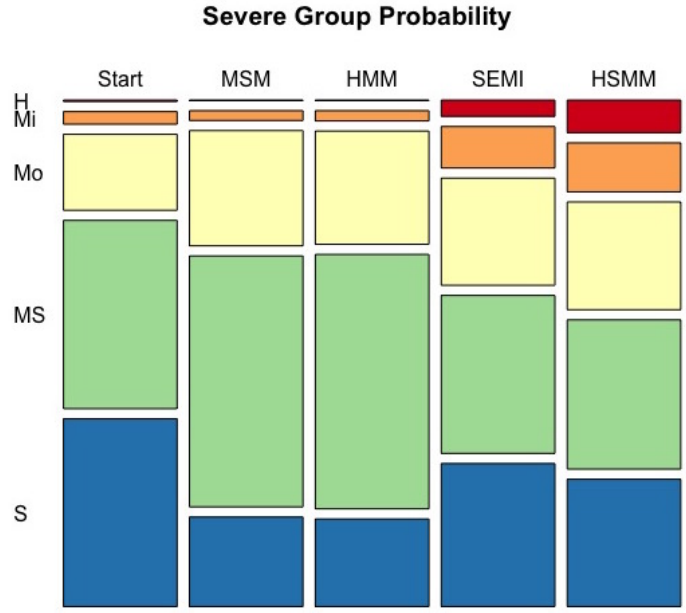

Figure E.7: Stationary Probability Distribution _ Severe Group (6067)

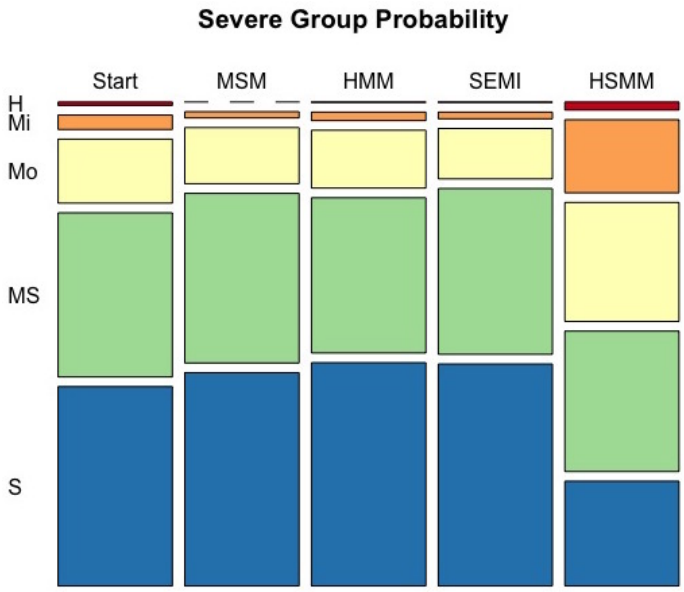

Figure E.8: Stationary Probability Distribution _ Severe Group (610)

Appendix F

ADDITIONAL FIRST PASSAGE TIME RESULTS

Table F.1: First Passage Time for Decreasing Group

From	**H**	**Mi**	**Mo**	**MS**	**S**
To H					
MSM	2	56	83	96	102
HMM	2	56	84	93	96
To MS					
MSM	16648	16547	15856	1000	6
HMM	20741	20635	19811	Inf	3

(a) (6067 Subjects)

From	**H**	**Mi**	**Mo**	**MS**	**S**
To H					
MSM	3	12	19	21	24
HMM	3	14	22	28	31
To MS					
MSM	3	12	19	21	24
HMM	3	11	18	21	22

(b) (610 Subjects)

Table F.2: First Passage Time for Mild Group

From	**H**	**Mi**	**Mo**	**MS**	**S**
To H					
MSM	1	60	74	83	101
HMM	2	62	77	84	88

(a) (6067 Subjects)

From	**H**	**Mi**	**Mo**	**MS**	**S**
To H					
MSM	2	15	23	28	31
HMM	2	18	28	37	44

(b) (610 Subjects)

Table F.3: First Passage Time for Moderate Group

From	H	Mi	Mo	MS	S
To H					
MSM	17	409	509	553	574
HMM	17	382	478	519	542
To MS					
MSM	277	252	194	6	21
HMM	288	264	204	6	24
To S					
MSM	2058	2032	1975	1781	83
HMM	2139	2115	2056	1851	77

(a) (6067 Subjects)

From	H	Mi	Mo	MS	S
To H					
MSM	6	20	31	38	43
HMM	7	23	37	46	52
To MS					
MSM	24	21	14	5	5
HMM	23	20	15	5	5
To S					
MSM	59	56	49	35	8
HMM	62	59	52	37	7

(b) (610 Subjects)

Table F.4: First Passage Time for Severe Group

From	H	Mi	Mo	MS	S
To MS					
MSM	63	57	42	2	46
HMM	64	58	42	2	44
To S					
MSM	256	250	235	193	5
HMM	255	249	233	191	5

(a) (6067 Subjects)

From	H	Mi	Mo	MS	S
To MS					
MSM	9	8	5	3	10
HMM	12	10	6	3	13
To S					
MSM	20	19	16	11	2
HMM	25	23	19	13	2

(b) (610 Subjects)

Appendix G

ADDITIONAL VALIDATION RESULTS

Table G.1: Bootstrap Validation Statistics _ Decreasing Group from 610 Subjects

Decreasing	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.396	0.402	0.172	0.026	0.002
Mean	0.396	0.402	0.173	0.027	0.003
CI	(0.311, 0.472)	(0.346, 0.462)	(0.126, 0.213)	(0.013, 0.035)	(0, 0.003)
Std.Err	0.042	0.03	0.022	0.006	0.001

Table G.2: Bootstrap Validation Statistics _ Mild Group from 610 Subjects

Mild	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.446	0.368	0.136	0.047	0.003
Mean	0.44	0.367	0.139	0.049	0.004
CI	(0.371, 0.529)	(0.319, 0.416)	(0.09, 0.173)	(0.018, 0.067)	(0, 0.005)
Std.Err	0.04	0.025	0.021	0.012	0.002

Table G.3: Bootstrap Validation Statistics _ Moderate Group from 610 Subjects

Moderate	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.154	0.235	0.268	0.215	0.129
Mean	0.155	0.234	0.267	0.215	0.129
CI	(0.103, 0.198)	(0.2, 0.27)	(0.235, 0.301)	(0.178, 0.252)	(0.093, 0.162)
Std.Err	0.025	0.018	0.017	0.019	0.018

Table G.4: Bootstrap Validation Statistics _ Severe Group from 610 Subjects

Severe	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0	0.014	0.126	0.381	0.479
Mean	0	0.014	0.126	0.381	0.478
CI	(0 , 0)	(0.001, 0.023)	(0.086, 0.162)	(0.334, 0.428)	(0.412, 0.546)
Std.Err	0	0.006	0.019	0.024	0.035

Table G.5: Cross Validation Statistics _ Decreasing group from 610 Subjects

Decreasing	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.382	0.369	0.099	0.117	0.033
Quantile	(0.314, 0.458)	(0.314, 0.387)	(0.094, 0.13)	(0.083, 0.147)	(0.018, 0.036)
CI	(0.306, 0.45)	(0.351, 0.424)	(0.068, 0.104)	(0.087, 0.151)	(0.03, 0.048)
Std.Dev	0.053	0.023	0.015	0.022	0.006
Std.Err	0.017	0.007	0.005	0.007	0.002

Table G.6: Cross Validation Statistics _ Mild group from 610 Subjects

Mild	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.412	0.317	0.102	0.116	0.053
Quantile	(0.385, 0.49)	(0.285, 0.333)	(0.092, 0.123)	(0.093, 0.122)	(0.031, 0.061)
CI	(0.334, 0.439)	(0.301, 0.349)	(0.081, 0.112)	(0.11, 0.139)	(0.045, 0.075)
Std.Dev	0.035	0.016	0.011	0.01	0.011
Std.Err	0.011	0.005	0.003	0.003	0.004

Table G.7: Cross Validation Statistics _ Moderate group from 610 Subjects

Moderate	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.148	0.21	0.265	0.233	0.144
Quantile	(0.112, 0.194)	(0.169, 0.222)	(0.199, 0.27)	(0.202, 0.246)	(0.155, 0.258)
CI	(0.102, 0.184)	(0.198, 0.251)	(0.26, 0.331)	(0.22, 0.264)	(0.03, 0.133)
Std.Dev	0.025	0.018	0.026	0.016	0.039
Std.Err	0.008	0.006	0.008	0.005	0.012

Table G.8: Cross Validation Statistics _ Severe group from 610 Subjects

Severe	**Healthy**	**Mild**	**Moderate**	**MS**	**Severe**
Original	0.001	0.015	0.113	0.372	0.498
Quantile	(0, 0.003)	(0.007, 0.023)	(0.099, 0.15)	(0.345, 0.43)	(0.425, 0.543)
CI	(0, 0.002)	(0.007, 0.023)	(0.076, 0.127)	(0.314, 0.399)	(0.453, 0.571)
Std.Dev	0.001	0.006	0.018	0.028	0.037
Std.Err	0	0.002	0.006	0.009	0.012