Multivariate Longitudinal Data Analysis with Mixed Effects Hidden Markov Models

by

Jesse D. Raffa

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics - Biostatistics

Waterloo, Ontario, Canada, 2012

© Jesse D. Raffa 2012

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Jesse D. Raffa

Abstract

Longitudinal studies, where data on study subjects are collected over time, is increasingly involving multivariate longitudinal responses. Frequently, the heterogeneity observed in a multivariate longitudinal response can be attributed to underlying unobserved disease states in addition to any between-subject differences. We propose modeling such disease states using a hidden Markov model (HMM) approach and expand upon previous work, which incorporated random effects into HMMs for the analysis of univariate longitudinal data, to the setting of a multivariate longitudinal response. Multivariate longitudinal data are modeled jointly using separate but correlated random effects between longitudinal responses of mixed data types in addition to a shared underlying hidden process. We use a computationally efficient Bayesian approach via Markov chain Monte Carlo (MCMC) to fit such models. We apply this methodology to bivariate longitudinal response data from a smoking cessation clinical trial. Under these models, we examine how to incorporate a treatment effect on the disease states, as well as develop methods to classify observations by disease state and to attempt to understand patient dropout. Simulation studies were performed to evaluate the properties of such models and their applications under a variety of realistic situations.

Acknowledgements

I would like to acknowledge everyone who made my program of study in Waterloo possible. I would first like to thank my thesis advisor Joel Dubin for his patience and assistance both intellectual and financial over my tenure at the University of Waterloo. I also received financial assistance from several sources during my PhD program, including from the Canadian Institutes of Health Research (CIHR), the Ontario Graduate Scholarship (OGS) Program, the Statistical Society of Canada (SSC), and other programs at the University of Waterloo, among others. I would also like to thank Dr. Stephanie O'Malley (Yale University) for providing the dataset which is the primary motivation for the analyses presented in my thesis, in addition to my thesis committee members: Dr. Shoja'eddin Chenouri (University of Waterloo), Dr. Richard Cook (University of Waterloo), Dr. Michael Escobar (University of Toronto), and Dr. Steve Manske (University of Waterloo). I have made a number of good friends while at Waterloo, and would like to thank in particular my officemates Longyang Wu and Catherine Donnelly for being able to bounce crazy ideas off them. I am also grateful to the friends I have met outside of school, many of whom have helped me balance my life in a way that I didn't realize was possible until recently.

To my parents, Pat and Barbara Raffa.

Contents

List of Figures **ix**

List of Tables **xiii**

List of Abbreviations **xiv**

1 Analysis of Longitudinal Data **1**

1.1 Longitudinal Data in Clinical Studies 1

1.2 Analysis Approaches 1

1.2.1 Marginal models 2

1.2.2 Mixed effects models 3

1.2.3 Transition models 4

1.3 Joint or Multivariate Responses in Longitudinal Data Analysis 6

1.4 Extensions to Mixed Effects and Transition Models 10

2 Smoking Cessation Clinical Trials and the Naltrexone Augmentation Study **11**

2.1 Smoking Cessation Studies 11

2.1.1 Study outcomes 12

2.1.2 Measurements of smoking status 13

2.1.3 Assuming dropouts as treatment failure 15

2.2 The Naltrexone Augmentation of Nicotine Patch Study 15
2.3 Statistical Methods of Smoking Cessation Data 21
2.4 Overview and Motivation 22

3 Multivariate Longitudinal Data Analysis with Mixed Effects Hidden Markov Models 24

3.1 Introduction 24
3.2 Mixed effects hidden Markov models for multivariate longitudinal responses 26
3.2.1 Model Framework 27
3.2.2 Motivating Example 30
3.3 Bayesian Inference 31
3.3.1 Prior Specification 31
3.3.2 Updating the Hidden States 32
3.3.3 Sampling from the Posterior 33
3.3.4 Introducing a Treatment Effect for the Hidden Process 36
3.3.5 Label Switching 38
3.4 Simulation Study 39
3.5 Results from the smoking cessation motivating example 46
3.6 Discussion 51
3.7 Appendix: Markov Chain Monte Carlo Convergence Diagnostic Plots . . . 53
3.7.1 Trace and Density Estimation Plots 53
3.7.2 Auto-correlation function plots 57

4 Using Multivariate MHMMs for Classification of the Disease States 64

4.1 Background 64
4.2 Motivation 65
4.3 Model Considerations 66
4.3.1 Methodology 66

4.3.2 Results 70
4.4 Classifying the Hidden States 76
4.4.1 Methodology 77
4.4.2 Results 78
4.5 Assessing the Conditional Independence Assumption 90
4.6 Simulation Study 91
4.7 Discussion 100

5 Using MHMMs for Prediction and other Missing Data Issues 104
5.1 Background 104
5.2 Prediction of Hidden States 105
5.2.1 Methods 105
5.2.2 Results 106
5.3 Missing Data Issues 110
5.3.1 Methods 112
5.3.2 Results 117
5.4 Discussion 117

6 Mixed Effects Hidden Markov Models for Multivariate Longitudinal Data: Summary and Future Work 122
6.1 Summary 122
6.1.1 Modelling Framework 122
6.1.2 Classification and Prediction of Disease States in Smoking Cessation 125
6.1.3 Missing Data Issues 126
6.2 Future Work 126
6.2.1 Model Extensions 127
6.2.2 Model Selection 128
6.2.3 Missing Data Approaches 128

6.3 Conclusions 128

Bibliography **130**

List of Figures

1.1 Simple first-order Markov chain transition model for subjects i. Y_{it} represents the longitudinal response for subject i at time t 5

2.1 Agreement between self-reported (SR) smoking count (any during a given week), and having a positive CO test for that week for several CO thresholds 16

2.2 Agreement between SR smoking count (more than 5 total cigarettes reported in the week), and having a positive CO test for that week for several CO thresholds (in ppm) 18

2.3 The effect of last day of recorded self-reported smoking by study week on end of week CO reading (in ppm). 19

2.4 ROC plots for Prediction of each of the measure of assessing smoking status. 20

3.1 Mixed Effects Hidden Markov Model structure for subject i, with hidden process $\boldsymbol{Z}_i$, observed process $\boldsymbol{Y}_i$, and random effects, $\boldsymbol{b}_i$ 27

3.2 Trace and density estimate plots for parameters (from top to bottom): π_1 (var1), π_2 (var2), p^0_{12} (var3) and p^0_{13} (var4). 54

3.3 Trace and density estimate plots for parameters (from top to bottom): p^0_{21} (var1), p^0_{23} (var2), p^0_{31} (var3) and p^0_{32} (var4). 55

3.4 Trace and density estimate plots for parameters (from top to bottom): τ_{11} (var1), τ_{12} (var2), τ_{13} (var3) and τ_{21} (var4). 56

3.5 Trace and density estimate plots for parameters (from top to bottom): τ_{22} (var1), τ_{23} (var2), Σ_{11} (var3) and Σ_{21} (var4). 57
3.6 Trace and density estimate plots for parameters (from top to bottom): Σ_{22} (var1), $\frac{1}{\sigma_\epsilon^2}$ (var2), p_{12}^1 (var3) and p_{13}^1 (var4) 58
3.7 Trace and density estimate plots for parameters (from top to bottom): p_{21}^1 (var1), p_{23}^1 (var2), p_{31}^1 (var3) and p_{32}^1 (var4). 59
3.8 Auto-correlation function plots for the MCMC sampling approach for parameters: π_1 (1), π_2 (2), p_{12}^0 (3), p_{13}^0 (4), p_{21}^0 (5) and p_{23}^0 (6). 60
3.9 Auto-correlation function plots for the MCMC sampling approach for parameters: p_{31}^0 (7), p_{32}^0 (8), τ_{11} (9), τ_{12} (10), τ_{13} (11) and τ_{21} (12). 61
3.10 Auto-correlation function plots for the MCMC sampling approach for parameters: τ_{22} (13), τ_{23} (14), Σ_{11} (15), Σ_{21} (16), Σ_{22} (18) and $\frac{1}{\sigma_\epsilon^2}$ (19). . . . 62
3.11 Auto-correlation function plots for the MCMC sampling approach for parameters: p_{12}^1 (20), p_{13}^1 (21), p_{21}^1 (22), p_{23}^1 (23), p_{31}^1 (24) and p_{32}^1 (25). . . . 63

4.1 QQ-normal plot and histogram for the posterior mean of $\boldsymbol{b}_i$ for the self reported cigarette count-only univariate response MHMM model 73
4.2 QQ-normal plot and histogram for the posterior mean of $\boldsymbol{b}_i$ for the CO-only univariate response MHMM model 74
4.3 QQ-normal plot and histogram for the posterior mean of $\boldsymbol{b}_i$ for the bivariate response MHMM model 75
4.4 ROC curve for $\hat{\phi}_{it1} = P(Z_{it} = 1|y, ...)$ for each study visit (t) for the self reported cigarette count-only univariate response MHMM model. 81
4.5 ROC curve for $\hat{\phi}_{it1} = P(Z_{it} = 1|y, ...)$ for each study visit (t) for the CO-only univariate response MHMM model. 82
4.6 ROC curve for $\hat{\phi}_{it1} = P(Z_{it} = 1|y, ...)$ for each study visit (t) for the bivariate response MHMM model. 83

4.7 ROC curve for $\hat{\phi}_{it3} = P(Z_{it} = 3|y, ...)$ for each study visit (t) for the self reported cigarette count-only univariate response MHMM model. 84
4.8 ROC curve for $\hat{\phi}_{it3} = P(Z_{it} = 3|y, ...)$ for each study visit (t) for the CO-only univariate response MHMM model. 85
4.9 ROC curve for $\hat{\phi}_{it3} = P(Z_{it} = 3|y, ...)$ for each study visit (t) for the bivariate response MHMM model. 86
4.10 Example longitudinal plot for subjects with one or more visits such that the most probable hidden state according to the MHMM bivariate longitudinal model was state three, despite it being a non-smoking visit according to the study definition. 87
4.11 Example longitudinal plot for subjects with one or more visits such that the most probable hidden state according to the MHMM bivariate longitudinal model was state one, despite it being a smoking visit according to the study definition. 88
4.12 Box plot for the Poisson SR count response $(r = 1)$ random effect at each study visit stratified by the most probable hidden state, $\hat{\delta}_{it}$. 92
4.13 Box plot for the normal CO level response $(r = 2)$ random effect at each study visit stratified by the most probable hidden state, $\hat{\delta}_{it}$. 93

5.1 ROC curves for $\hat{\omega}_{i61}$ and $\hat{\omega}_{i63}$ in the complete data analysis 108
5.2 ROC curves for $\hat{\omega}_{i61}$ and $\hat{\omega}_{i63}$ in the five observation time subset analysis . 109
5.3 Patient disposition for the smoking cessation study 111
5.4 Histogram of the distribution of $\hat{\phi}_{it1} = P(Z_{it} = 1|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ overall, at $t = 6$ and at dropout. 113
5.5 Histogram of the distribution of $\hat{\phi}_{it2} = P(Z_{it} = 2|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ overall, at $t = 6$ and at dropout. 114
5.6 Histogram of the distribution of $\hat{\phi}_{it3} = P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ overall, at $t = 6$ and at dropout. 115

List of Tables

3.1 Simulation results for 200 runs of Situation (1) 41
3.2 Simulation results for 200 runs for each of Situations (2) and (4) 43
3.3 Simulation results for 200 runs of Situation (3) 44
3.4 Simulation results for 200 runs for each of Situation (5) and (6) 47
3.5 Simulation results for 200 runs of Situation (7) 48
3.6 Results for Models I-IV from the Motivating Smoking Cessation Example 50

4.1 Comparison of the bivariate and the two univariate models 72
4.2 Breakdown of most probable disease states by different study-defined disease states under separate univariate and bivariate models 79
4.3 Comparison of the bivariate and the two univariate models under simulation through 100 simulation runs for Situation (1) 95
4.4 Comparison of the bivariate and the two univariate models under simulation through 100 simulation runs for Situation (2) 97
4.5 Hidden state classification results for the simulations 99

5.1 Comparison of full observation model and the five week model 107
5.2 Bivariate MHMM results for three different missing data scenarios under the binomial count SR model 118
5.3 Bivariate MHMM results for three different missing data Scenarios under the Poisson count SR model 119

List of Abbreviations

AUC - Area Under the Curve

CO - Carbon Monoxide

GEE - generalized estimating equation

iid - independent and identically distributed

HMM - Hidden Markov Model

MAR - Missing at Random

MNAR - Missing Not at Random

MCMC - Markov Chain Monte Carlo

MHMM - Mixed Effects Hidden Markov Model

MVN - Multivariate Normal

NRT - Nicotine Replacement Therapy

ppm - parts per million

ROC - Receiver Operator Characteristic

SR - Self-reported

Chapter 1

Analysis of Longitudinal Data

1.1 Longitudinal Data in Clinical Studies

In the past 50 years, there has been an explosion in the number of clinical studies evaluating the efficacy or effectiveness of medical interventions, or examining the natural history of a multitude of diseases. In some studies, the intervention or duration of illness may be relatively short, requiring very few observations on each subject, recorded cross-sectionally over a short time period. However, in most studies, the course of treatment or duration of disease are long enough to measure some aspects of the disease, or response to treatment over a period of time on each subject. When such repeated measurements are collected over a period of time, the data are said to be longitudinal.

1.2 Analysis Approaches

The benefits of using longitudinal over cross-sectional data, including the ability to measure change over time, are numerous and well-documented (Diggle et al., 2002). However, the correlation between observations of the same subject often complicates the analysis

of longitudinal data. Independence of observations is a ubiquitous assumption used in the majority of statistical analysis, and while usually suitable in cross-sectional analyses, the repeated nature of the observations within a subject in longitudinal studies makes it quite likely that observations within each subject are correlated. How to account for this within-subject correlation has resulted in several different approaches to the analysis of longitudinal data. Each differs slightly in the underlying assumptions, and resulting interpretation, but are usually aimed at: describing the change of the response variable over time, estimating the effect of any explanatory variables, and often characterizing the variability of the response attributable to a subject and/or the nature of the correlation within a subject between the repeated observations.

With these aims in mind, let Y_{it} be the response variable for subject $i = 1, 2, \ldots, N$ at time $t = 1, 2, \ldots, n_i$, where each subject i has n_i repeated measurements. Let $\boldsymbol{x}_{it}$ be p dimensional vector of explanatory variables (possibly time-varying) for subject i at time t. Three general classes of analysis approaches have been described in the literature (Diggle et al., 2002), which address the aims of longitudinal data analysis in different ways. A brief overview of the approaches follows.

1.2.1 Marginal models

Marginal models are one such approach to the analysis of correlated data in general, and have been used extensively in longitudinal data analysis. This marginal approach focuses primarily on modelling the marginal expectation of the response, $E(Y_{it})$, as a function of the explanatory variables. This marginal expectation of Y_{it} is modelled to depend on the explanatory variables through a known link function, such that if $E(Y_{it}) = \mu_{it}$, then $g(\mu_{it}) = \boldsymbol{x}_{it}^{'}\boldsymbol{\beta}$, where g is known as a link function.

Similarly, the marginal variance of Y_{it}, $Var(Y_{it}) = v(\mu_{it})\phi$, is assumed to depend only on a function of the marginal mean of Y_{it} and a scale parameter, ϕ. Lastly, the correlation between Y_{it} and $Y_{it'}$ is assumed to be a function of the marginal means, and perhaps some

additional parameters, $\boldsymbol{\alpha}$, such that $Corr(Y_{it}, Y_{it'}) = \rho(\mu_{it}, \mu_{it'}, \boldsymbol{\alpha})$.

Under this approach which specifies the marginal mean and variance of Y_{it} in addition to its correlation structure of the response variable within each subject, a generalized estimating equation (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986) for β is given by:

$$S_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right)' \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0 \tag{1.1}$$

where $\boldsymbol{V}_i$ is often referred to as the working covariance matrix of Y_i, which depends on both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The solution of 1.1 is solved iteratively. Given initial estimates for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ϕ, updates of $\boldsymbol{\beta}$ can be solved by 1.1, which can then be used to update $\boldsymbol{\alpha}$ and ϕ, and repeated until convergence. Several methods have been discussed in the literature on how to do the latter update of $\boldsymbol{\alpha}$, but this iterative process will yield estimates of $\boldsymbol{\beta}$ which are as efficient (asymptotically) as estimates when $\boldsymbol{\alpha}$ is known. Further work has yielded methods where estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is done through separate estimating equations (Prentice, 1988; Zhao and Prentice, 1990).

1.2.2 Mixed effects models

Mixed effects models introduce random effects which are unobservable variables which capture some aspects of the heterogeneity observed among subjects (Laird and Ware (1982)). The central assumption of mixed effects models is that conditional on these random effects, $(\boldsymbol{b}_i)$, the longitudinal responses, Y_{it}, are independent, though this assumption can be relaxed to include situations where $Y_{it}|\boldsymbol{b}_i$ and $Y_{it'}|\boldsymbol{b}_i$, $t \neq t'$ are correlated. Furthermore, we assume that these random effects are mutually independent, and arise from a common (possibly multivariate) distribution, and are independent of any explanatory variables.

We take a similar initial approach to that of marginal models, where if Y_{it} is from an exponential family, and given a known link function, g, we introduce random effects, $\boldsymbol{b}_i$

such that:

$$g\big[E(Y_{it}|\boldsymbol{b}_i)\big] = \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \boldsymbol{w}_{it}^{'}\boldsymbol{b}_i$$

where $\boldsymbol{x}_{it}$ and $\boldsymbol{w}_{it}$ are p and q dimensional vectors of explanatory variables.

In specific circumstances, the random effects can be viewed as nuisance parameters, and conditional likelihood methods are available, but are not appropriate when subject-specific effects are of interest. In contrast, the marginal distribution of the response variable can be obtained by integrating out the random effects of the joint distribution of $\boldsymbol{Y}_i$ and $\boldsymbol{b}_i$, such that the likelihood function can be represented as:

$$L(\boldsymbol{\Theta}; \boldsymbol{y}) = \prod_{i=1}^{N} \int_{\boldsymbol{b}_i} \prod_{t=1}^{n_i} f(y_{it}|\boldsymbol{b}_i; \boldsymbol{\beta}) f(\boldsymbol{b}_i; \boldsymbol{\Sigma}) d\boldsymbol{b}_i \tag{1.2}$$

where $\boldsymbol{\Theta}$ is the combined parameters of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ where the latter is the covariance matrix of the random effects. If y_{it} are Gaussian, and the random effects are assumed to arise from a (possibly multivariate) Gaussian distribution (Laird and Ware, 1982), then 1.2 has a closed-form solution, and estimation of $\boldsymbol{\Theta}$ is relatively straightforward. On the other hand, in most other circumstances, 1.2 has no closed-form solution, and the integration must be evaluated numerically.

1.2.3 Transition models

Transition models are another method used to model longitudinal data. Transition models focus on modeling the response, y_{it}, conditional on the subjects history, denoted $\mathcal{H}_{it}$, where for subject i $\mathcal{H}_{it} = \{y_{ik}, k = 1, 2, \ldots, t-1\}$. Such models are often modeled as a q-order Markov chain, where the conditional distribution of y_{it} depends only on the most previous q responses for subject i, $\{y_{it-1}, y_{it-2}, \ldots, y_{it-q}\}$. An example of a transition model for a first-order Markov chain is presented in Figure 1.1, where the current observation Y_{it} is only dependent on the most previous observation.

Even under the Markov assumption, except in special cases, it is usually not possible to specify the full likelihood directly, and instead, the conditional likelihood is usually used:

$$\boxed{Y_{i1}} \longrightarrow \boxed{Y_{i2}} \longrightarrow \boxed{Y_{i3}} \longrightarrow \boxed{Y_{i4}} \longrightarrow \cdots$$

Figure 1.1: Simple first-order Markov chain transition model for subjects i. Y_{it} represents the longitudinal response for subject i at time t

$$L^C(\boldsymbol{\Theta}; \boldsymbol{y}) = \prod_{i=1}^{N} \prod_{t=q+l}^{n_i} f(y_{it}|\mathcal{H}_{it}).$$

Focus of such models has often been on binary Markov chains, where, for example, a first-order Markov chain for a binary response can be modeled as:

$$logit(Pr(Y_{it} = 1|Y_{it-1} = y_{it-1}, \boldsymbol{x}_{it})) = \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + y_{it-1}\boldsymbol{x}_{it}^{'}\boldsymbol{\alpha}, \tag{1.3}$$

where here $\boldsymbol{\beta}$ represents a vector of fixed effects, and $\boldsymbol{\alpha}$ represents any interaction between the the previous response, y_{it-1}, and one or more of the covariates. Using 1.3 we can fully specify a transition matrix for a binary Markov chain, such that for any subject with covariate values $\boldsymbol{x}_{it}$, we have a transition matrix of the form:

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix},$$

where each component can be determined using $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, such that:

$$p_{c1} = \frac{\exp(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + c\boldsymbol{x}_{it}^{'}\boldsymbol{\alpha})}{1 + \exp(\boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + c\boldsymbol{x}_{it}^{'}\boldsymbol{\alpha})}. \tag{1.4}$$

Even when fitting higher order Markov models, fitting such models can be done using ordinary logistic regression, provided that the transition events are uncorrelated. If there is heterogeneity between individuals in terms of the transition probabilities, then an appropriate model may be for example:

$$logit(Pr(Y_{it} = 1|Y_{it-1} = y_{it-1}, \boldsymbol{x}_{it}, b_i)) = b_i + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + y_{it-1}\boldsymbol{x}_{it}^{'}\boldsymbol{\alpha}, \tag{1.5}$$

where b_i is a subject-specific random effect as presented in section 1.2.2, where $b_i \sim N(0, \sigma^2_{b_i})$. In this case the heterogeneity between subjects is modeled through the introduction of random effects, and subsequently makes the transition within a person correlated. Marginal methods have also been used in this setting to account for the within-subject correlation (e.g., Azzalini (1994)).

1.3 Joint or Multivariate Responses in Longitudinal Data Analysis

Thus far, we have only examined the longitudinal models for one response variable. In many circumstances, more than one response variable is followed longitudinally, and analysing both jointly may be beneficial. Until recently, methods for multiple longitudinal outcomes have largely been based on simple approaches where each outcome is analyzed separately, or by reducing the dimension of the multiple outcomes through a factor analysis or principal components type of approach. The former approach is reasonably easy to implement, with the approaches already discussed, but ignores both the correlation between longitudinal outcomes and/or other features such as measurement error likely to exist in one or more of the outcomes. Reducing the dimension of the multiple outcomes is also easy to implement, and can quite often capture much of the correlation between outcomes. In such models, inference is focused on the principal components, and not the original outcome variables, often affecting interpretation of the model results and can also be difficult to implement in situations where the longitudinal data are unbalanced, or when the outcomes are measured at different time points.

Modelling multivariate responses can be difficult. If the joint distribution of the two or more responses is known, then inference is relatively straightforward. Such a circumstance would allow for inference to be conducted marginally, as well as jointly. In general, the joint distribution of two or more longitudinal responses is rarely known, and can be

difficult to obtain for data arising from different data types (e.g. continuous and discrete). Furthermore, inference is complicated in situations when outcomes are measured at different time points and/or are unbalanced.

Several modeling approaches have been developed which jointly model the longitudinal responses by specifying in a number of different ways the joint distribution of the longitudinal responses (e.g., Chapter 13-16, Fitzmaurice et al. (2006)). One possibility is the use of a conditional model, where the joint likelihood of the two or more responses is factorized. For two longitudinal responses, the factorization would typically look like:

$$f(y_1, y_2) = f(y_1|y_2)f(y_2) = f(y_2|y_1)f(y_1)$$

where y_1 and y_2 are vectors of two different longitudinal responses. Marginal inference is often difficult, as the parametric form may not be the same for the marginal and conditional models. Additionally, specification of the factorization can be very complicated in the setting of more than two responses.

Another frequently used method of modelling multivariate longitudinal responses is the introduction of latent variables which often simplify the joint distribution by assuming conditional independence. In a simple scenario, two responses, Y_1 and Y_2 are said to be conditionally independent, when given some latent random variable (or random effect) U, such that:

$$f(y_1, y_2) = \int_b f(y_1, y_2|b)f(u)db = \int_b f(y_1|b)f(y_2|b)f(b)db \tag{1.6}$$

The conditional independence assumption, $f(y_1, y_2|b) = f(y_1|b)f(y_2|b)$ is central to many different multivariate longitudinal models.

These shared random effects can be introduced into both models for Y_1 and Y_2, for example that for *each* subject:

$$\begin{aligned} Y_{1t} &= \beta_1 + b + \beta_2 t + \epsilon_{1t} \\ Y_{2t} &= \beta_3 + \gamma b + \beta_4 t + \epsilon_{2t} \end{aligned} \tag{1.7}$$

where in this example, $\epsilon_{1t} \overset{\text{iid}}{\sim} N(0, \sigma_1^2)$, $\epsilon_{2t} \overset{\text{iid}}{\sim} N(0, \sigma_2^2)$ and $b \sim N(0, \sigma_b^2)$. The parameter γ is used to scale the random effects and define the dependence between responses. In this example, the correlation within each response, and between each response are:

$$\begin{aligned}
Corr(Y_{1t}, Y_{1s}) &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_1^2} \\
Corr(Y_{2t}, Y_{2s}) &= \frac{\gamma^2\sigma_b^2}{\gamma^2\sigma_b^2 + \sigma_2^2} \\
Corr(Y_{1t}, Y_{2s}) &= \frac{\gamma\sigma_b^2}{\sqrt{\sigma_b^2 + \sigma_1^2}\sqrt{\gamma^2\sigma_b^2 + \sigma_2^2}} \\
&= \sqrt{Corr(Y_{1t}, Y_{1s})}\sqrt{Corr(Y_{2t}, Y_{2s})}
\end{aligned} \tag{1.8}$$

Another frequently used method is to again introduce random effects, but instead of sharing the random effect across the longitudinal responses, use separate, but correlated random effects in the longitudinal responses (e.g., Gueorguieva and Agresti (2001)). The conditional independence assumption noted in 1.6 still applies, but 1.7 becomes:

$$\begin{aligned}
Y_{1t} &= \beta_1 + b_1 + \beta_2 t + \epsilon_{1t} \\
Y_{2t} &= \beta_3 + b_2 + \beta_4 t + \epsilon_{2t}
\end{aligned} \tag{1.9}$$

where in this example, $\epsilon_{1t} \overset{\text{iid}}{\sim} N(0, \sigma_1^2)$, $\epsilon_{2t} \overset{\text{iid}}{\sim} N(0, \sigma_2^2)$ and $(b_1, b_2)' \sim MVN(\mathbf{0}, \mathbf{\Sigma})$. Here, (b_1, b_2) is multivariate normal, with a covariance matrix, $\mathbf{\Sigma}$, which allows for the random effects, b_1 and b_2 to be correlated.

The correlation within and between responses is similar to 1.8 but, the correlation between responses is now:

$$\begin{aligned}
Corr(Y_{1t}, Y_{2s}) &= Corr(b_1, b_2)\sqrt{Corr(Y_{1t}, Y_{1s})}\sqrt{Corr(Y_{2t}, Y_{2s})} \\
&\leq \sqrt{Corr(Y_{1t}, Y_{1s})}\sqrt{Corr(Y_{2t}, Y_{2s})}
\end{aligned}$$

In these latter two settings, examples of which are described in 1.7 and 1.9, extensions

to responses of different types (e.g. both continuous and discrete responses), and extensions to more than two response variables are relatively straightforward. Furthermore, the conditional independence assumption under most circumstances makes the interpretations of model parameters the same marginally as they would be jointly. In the shared random effect case, the correlation between responses is only dependent on the correlation within responses, whereas in the separate random effects case, the correlation structure between responses is attenuated by the use of separate random effects, and is therefore not dependent solely on the within-response correlation of each response.

In both the shared and separate random effects cases, the conditional independence assumption stated in 1.6 requires integration to compute the full joint distribution of the response variables. As was the case in the simple, univariate response setting discussed in section 1.2.2, there are settings when integrating out the random effects results in a closed-form solution, but it is often the case that numerical integration is required. When the number of longitudinal response variables is relatively small, the numerical integration can often be accomplished without too much trouble, but as the number of responses increases, the dimension of the integral increases, often making computation difficult. To get around this computational difficulty, some have advocated modelling all pairs of response variables, and subsequently using a pseudo-likelihood approach (Fieuws and Verbeke, 2006).

We have presented several analysis approaches which allow for modeling of multivariate longitudinal data by introduction of latent variables, which can model both between- and within- variable correlation. However, using a latent variable approach is not the only way to model such data, and others (e.g., Gray and Brookmeyer (1998, 2000)) have used marginal approaches to model more than one longitudinal response.

1.4 Extensions to Mixed Effects and Transition Models

As can be seen from the discussion above, there exists many different approaches to the modeling of longitudinal data. In particular, the features of mixed effects often allow for investigation of subject-specific longitudinal responses and has also immediate extensions to multivariate longitudinal data, while transition models are often useful in situations where values for previous response(s) predict a future response. As can be seen in 1.5, random effects can be easily introduced into transition models for binary responses, allowing for subject-specific transition probabilities. We will be using this idea to incorporate a latent or hidden transition model which governs two or more observed longitudinal responses which are also modeled using separate, but correlated random effects as discussed in 1.9. These types of models will be used to model data arising from a longitudinal smoking cessation study described in Chapter 2, where heterogeneity can be attributed to between-subject differences, but also dynamic changes in disease state.

Chapter 2

Smoking Cessation Clinical Trials and the Naltrexone Augmentation Study

2.1 Smoking Cessation Studies

The inhalation of tobacco smoke poses many public health concerns, including the development of cancer and cardiovascular disease. In the United States, tobacco smoking (from herein referred to as simply smoking) is the leading preventable cause of death in adults. Though the medical community has recommended for some time that chronic smokers quit smoking tobacco, the prevalence of smoking on adults has exceeded 20% in most developed countries. This is primarily due to the addictive nature of one of tobacco's principle ingredients, nicotine (Benowitz, 1996).

Since becoming a major public health concern in the latter half of the 20th-century, clinical trials examining the efficacy and/or effectiveness of both behavioral and pharmaceutical interventions for smoking cessation have been conducted. For pharmaceutical

interventions, there have been many randomized clinical trials where one or more pharmaceutical agents are administered and compared amongst themselves and/or a control group.

One class of these agents is known as nicotine replacement therapies (NRT), which reduces nicotine withdrawal by replacing the nicotine consumed through smoking most often transdermally by patch or orally via a gum. Several investigations have attempted to augment NRT by co-administration with other agents, and have reported cessation rates higher than patch alone (Stead et al., 2008). Despite the development of such interventions, many individuals fail to quit smoking over the smoking cessation study period. Smoking within the first few weeks has been shown to be a strong predictor of failure to quit smoking during the remainder of the study (Kenford et al., 1994).

2.1.1 Study outcomes

In most such trials, the protocols are structured for a relatively short period of frequent (e.g., once per week) follow-up for several weeks, sometimes followed by period of less frequent (once every several months) longer term follow-up, the latter often done after treatment has been terminated as per the study design. Evaluations of treatment efficacy are often focused on the short period of frequent follow-up, and most studies investigating the efficacy of pharmaceuticals evaluate smoking abstinence as their primary study outcome.

The definition of smoking abstinence is often complicated, and can vary between studies. Smoking abstinence is usually monitored longitudinally over the study period through one or more measures of smoking status. These measures may be biochemical or self-reported.

These longitudinal measurements along with the duration and/or timing of smoking abstinence are often reduced to a single binary study outcome of smoking abstinence. Often these definitions are very strict, requiring abstinence throughout the entire study

period on all outcomes, or complete abstinence after some predetermined time (often one to two weeks post baseline). Additionally, missing data are often handled in a strict manner, where patients who dropout of the study for any reason are assumed to have smoked.

2.1.2 Measurements of smoking status

While typically reduced into a single binary study outcome, smoking status is often monitored longitudinally by one or more types of methods. Such methods can be classified as patient reported or biochemical.

Patient reported: self-reported smoking counts

Patient reported smoking counts are commonly collected in many settings, to describe smoking cessation, as well as the effect of smoking exposure on health outcomes. Patients most often record a total number of cigarettes smoked over a well-defined time period (a day or week). These are often recorded in a diary, and are often subject to well-described epidemiological biases, and have been shown to generally underreport the prevalence and amount smoked (Patrick et al., 1994). Self-reported counts are usually whole numbers, and are often relatively consistent day-to-day within a patient. These counts are often near notable fractions or numbers of packs (e.g. 1 pack is 20 cigarettes, it is common to see common fractions of 20 cigarettes (10 - half a pack, 30 = 1.5 packs, etc)) (Wang and Heitjan, 2008). Since, as previously mentioned, self-reported smoking counts are generally believed to represent underreporting of the number of cigarettes smoked, smoking status is usually defined by any smoking reported during some defined time period.

Biochemical: carbon monoxide

Biochemical monitoring is often done through detection of carbon monoxide through oral respiration, where the subject exhales into an instrument which estimates the concentration of carbon monoxide (CO) in parts per million (ppm). Smoking status is usually determined

by a pre-defined threshold, where exceeding this threshold is used to provide evidence that the participant smoked. Commonly used thresholds are 8 or 10 ppm. While CO is often found in small quantities in humans due to other exposures to CO, smokers have significantly higher levels of CO during periods when they have smoked. A limitation of using CO in smoking cessation studies is due to its relatively short half-life (5-6 hours), meaning that a CO assessment on a Tuesday would likely fail to detect the presence of a smoking event even a few days prior, on Friday or Saturday for instance. Further, any non-smoking related environmental exposure to CO in close time proximity to the study visit may be detected and falsely interpreted as evidence of smoking.

Biochemical: serum cotinine

Serum cotinine is another biochemical marker of smoking, which detects a metabolite of nicotine within the blood. It is considered a sensitive and specific method to detect smoking status, as nicotine exposure is not common in the environment in the absence of tobacco products. Serum cotinine is often considered the gold standard for measuring smoking status, since it has a longer half life in plasma than is typical for CO, and has generally been found to have higher sensitivity and specificity to detect smoking when compare CO (Jarvis et al., 1987). With that said, confirming smoking status is a difficult task, and would require 24 hour video surveillance, or something similar to achieve 100% certainty of an individual's status at a given time. Further, in smoking cessation studies, it is common to use NRT via gum or patch for at least a portion of the study, and unfortunately, serum cotinine has limited utility as a determinant of smoking status during any period which permits NRT. In addition, obtaining serum cotinine requires blood work, which may be considered burdensome to many potential study participants.

2.1.3 Assuming dropouts as treatment failure

As is the case in many clinical studies, subject dropout is common in smoking cessation studies. It is a very frequent assumption to consider these individuals as treatment failures (i.e., smokers) when determining smoking status, regardless of the reason for dropout. This is usually considered to be a conservative assumption, which we do not believe in some situations is justified, and will be discussed in more detail in section 5.3 in this thesis.

2.2 The Naltrexone Augmentation of Nicotine Patch Study

O'Malley et al. (2007) conducted a study similar to those described in section 2.1, where naltrexone augmentation of NRT was evaluated for smoking cessation. At baseline all subjects enrolled in the study received NRT, and were randomized in a double-blind fashion to receive placebo, or one of three levels of naltrexone (25, 50 and 100 mg) in a roughly 1:1:1:1 allocation. Subjects were followed six weeks with weekly study visits, where CO was measured and self-reported smoking count journals were collected. At the conclusion of the six weeks, the study drugs were discontinued, and follow-up visits were scheduled at 3, 6 and 12 months. O'Malley et al. (2007) used a composite definition of smoking abstinence, which used both self-reported and CO to define smoking status, and examined smoking abstinence over several time periods. *Continuous abstinence* was defined as no self-reported smoking from baseline to week 6, and all weekly assessments of CO below the study's 10 ppm CO threshold, and *prolonged abstinence* was defined similarly, but allowed a grace period from baseline to the second study visit, where smoking information recorded in that period was ignored.

In total, 673 subjects underwent screening, with 400 proceeding to randomization. Of these 400, 385 received therapy, with 295 completing the six weeks of therapy. It is

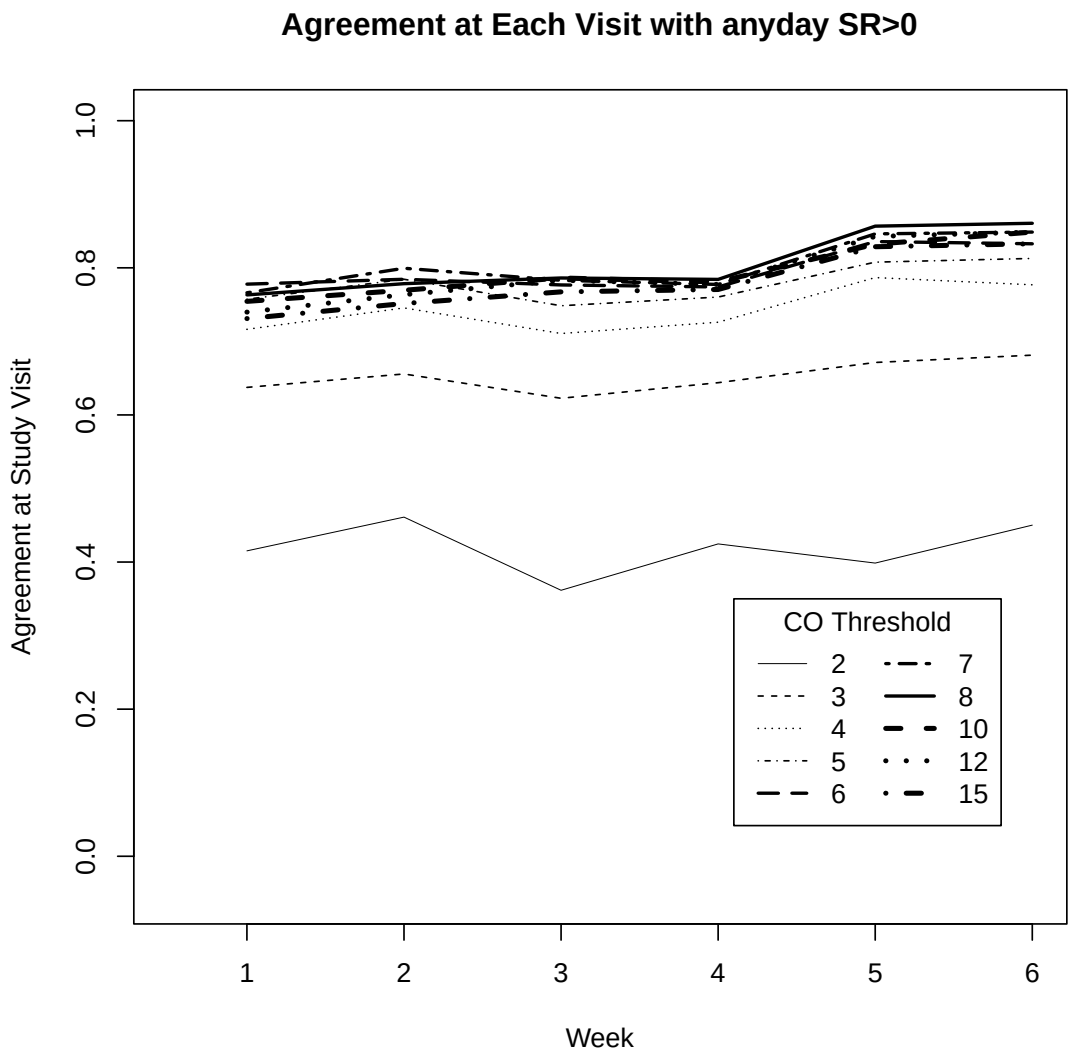

Figure 2.1: Agreement between self-reported (SR) smoking count (any during a given week), and having a positive CO test for that week for several CO thresholds

noteworthy that 354 of the 385 receiving therapy had some sort of evaluable ($\geq$ 1 week) follow-up as it pertains to both CO and self-reported smoking counts. The analysis that follows primarily focuses on these 354 subjects. Abstinence rates varied depending on the definition and duration of abstinence, as well the treatment subgroup, with 41-52% achieving prolonged abstinence and 39-52% maintaining continuous abstinence. Although, the authors note a possible treatment effect in the group with the highest dose of naltrexone when compared to placebo in the treatment completers, naltrexone was not observed to have any statistically significant treatment effect at any other dose or definition of abstinence. A possible beneficial effect on weight was also noticed for some subjects taking naltrexone.

Using the data from this study (O'Malley et al., 2007), we examined the agreement between the self-reported and CO determinants of smoking status across several different thresholds of CO. In Figure 2.1, smoking status as determined by self-report (SR) is defined as self-reported smoking of any cigarettes during the week, and the agreement with the CO-determined smoking status is investigated across several CO thresholds. As can be seen in this figure, the agreement between the two measurements is highest for the higher thresholds of CO, as would be expected, however, the maximum level of agreement is approximately 80%, and does not appear to depend on the visit week. Similarly in Figure 2.2, a higher level of SR smoking status ($>$ 5 cigarettes on any day of the week specified) results in a generally higher level of agreement between the SR definition and the different CO thresholds, with the maximum agreement approximately 90-95% for the higher CO thresholds.

In Figure 2.3, we can see the effect of the short half-life for CO. Each set of boxplots represents one study week, with the right most boxes in each plot representing the days closet to the CO measurement (i.e. day 7 is the day of the CO measurement, and day 6 would be the day before, etc). Day 0 represents individuals who did not report smoking during the week. A stark increase in both the variability and median is noted on the

most right hand side of these plots, illustrating the limited period of sensitive detection of smoking through CO.

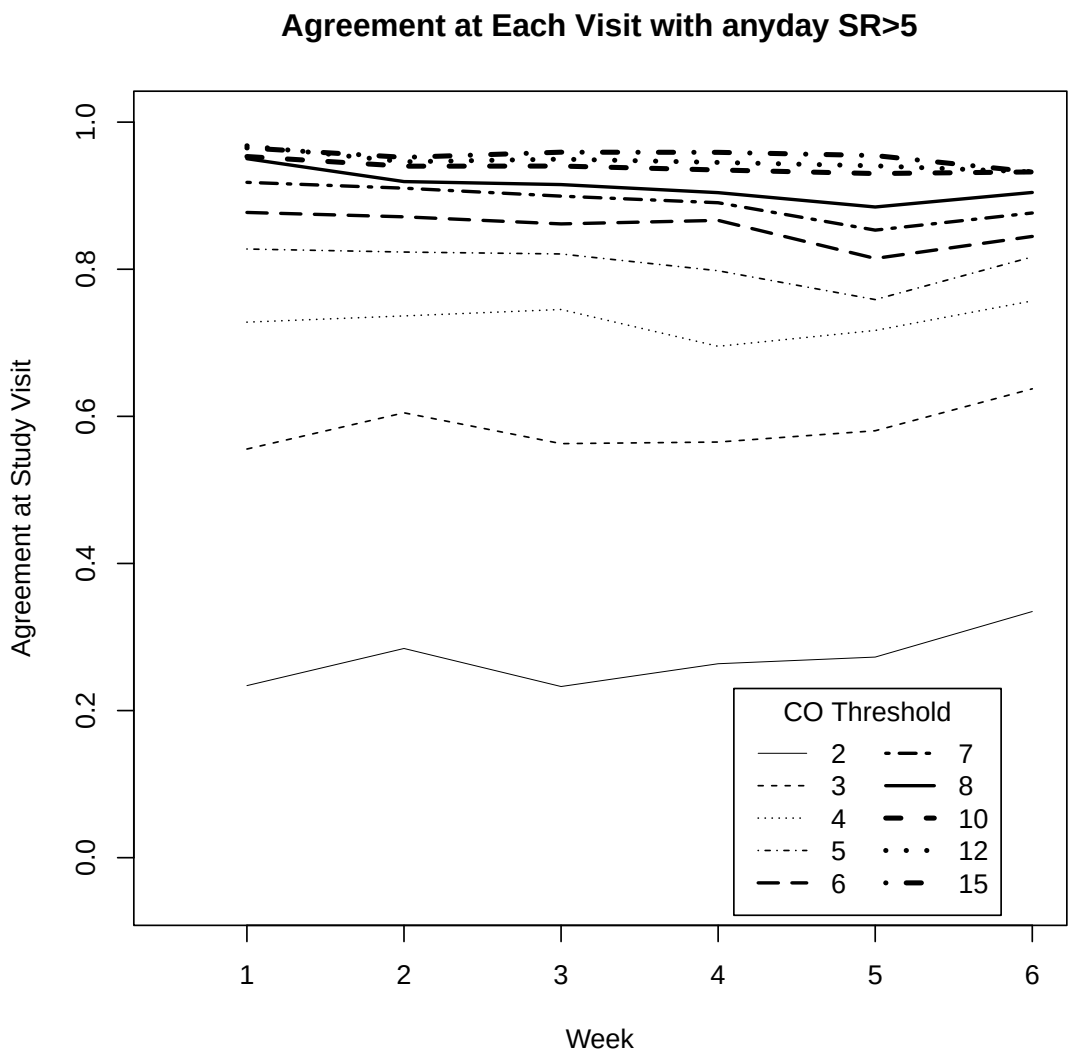

Figure 2.2: Agreement between SR smoking count (more than 5 total cigarettes reported in the week), and having a positive CO test for that week for several CO thresholds (in ppm)

We can also assess the agreement between measures, by treating one as a gold standard,

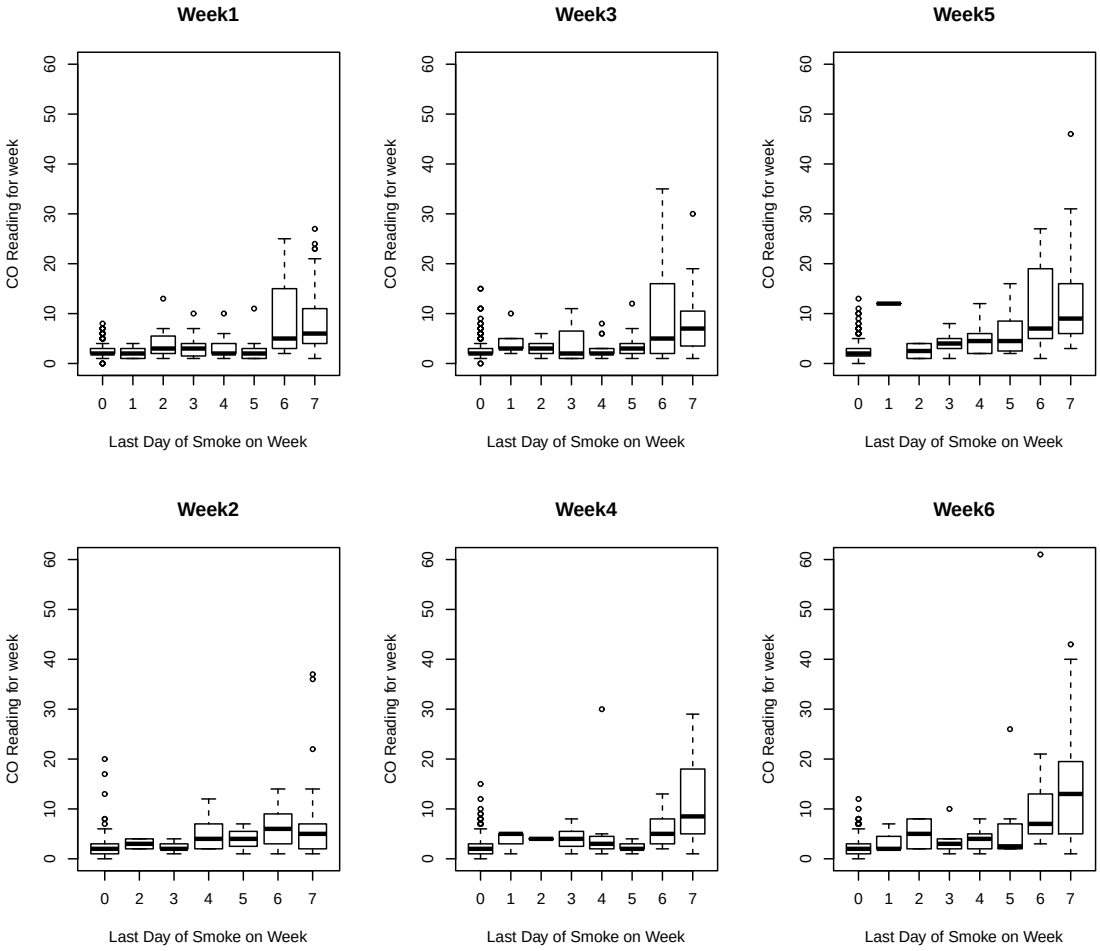

Figure 2.3: The effect of last day of recorded self-reported smoking by study week on end of week CO reading (in ppm).

Note: Day 0 represents no self-reported smoking during the specified study week, and day 7 would be the day of the CO measurement.

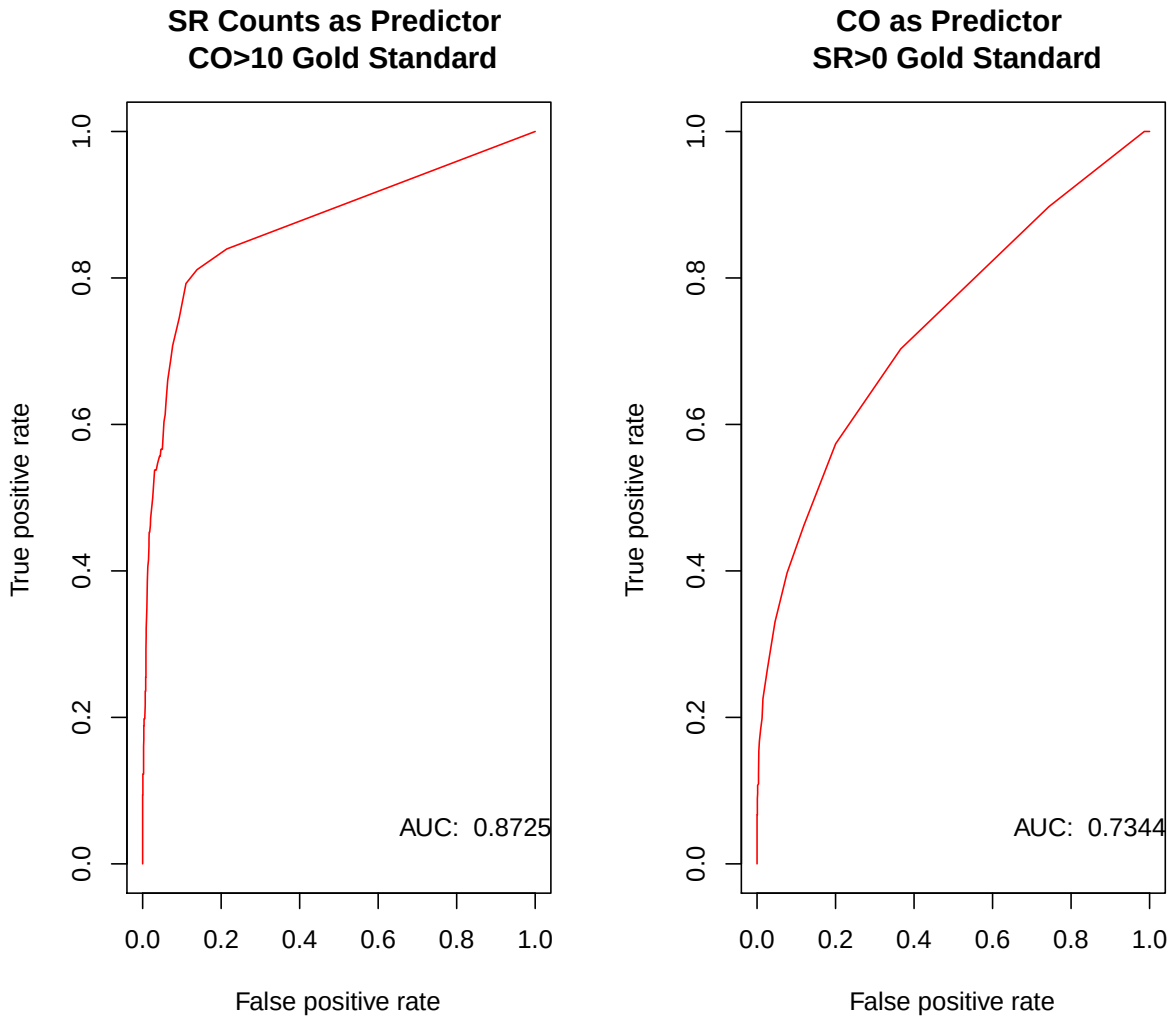

Figure 2.4: ROC plots for Prediction of each of the measure of assessing smoking status.

and assessing the prediction of the other measure by receiver operator characteristic (ROC) curves. These curves for each of the SR and CO measures are presented in Figure 2.4, where we see that the predictive power for one is relatively strong, but not perfect.

2.3 Statistical Methods of Smoking Cessation Data

In most efficacy studies of smoking cessation the primary analyses focus on comparing the proportion of those subjects who achieved and/or maintained abstinence over some pre-determined time period among the treatment groups. These analyses can be typically accomplished through relatively simple methods, but given the complex nature of the longitudinal data, some methodological literature has been dedicated to smoking cessation studies. Some of the most relevant literature is described below.

In Borrelli et al. (2002), the issue of study participant dropout is examined in smoking cessation studies. As discussed briefly in section 2.1.3, the definition for abstinence often assumes that those individuals who dropout of the study prior to the completion of the treatment period are deemed non-abstainers (i.e., they have smoked). Borrelli fits a multinomial logistic regression model examining the characteristics of three groups: the quitters (abstainers), the smokers, and the dropouts. They reported that dropouts differ from smokers in several ways, but most notably, that individuals who believe the therapy works (self-efficacy) were more likely to dropout when compared to smokers, while there was no difference in terms of self-efficacy of smokers and quitters. They also report that the rate of dropout may differ among treatments, and treating dropouts as smokers may miss important features that treating them separately would capture.

Qin et al. (2009) reports on modelling two binary smoking observations (i.e., two observation time points where smoking was observed or not) by using two latent class models. The joint model of both observation time points conditional on latent class membership was assumed to be conditionally independent of the dropout of the second observation.

This model is then compared to other methods (shared parameter model and weighted GEE) for handling the missing data due to dropout, and found to perform generally better.

Liu, Daniels and Marcus (2009) reports of jointly modeling a binary longitudinal response and a continuous response variable in the context of a smoking cessation trial. The focus of their analysis is on a binary longitudinal response (smoking status) and a continuous longitudinal response (body weight). They use a probit form of the binary process which involves an underlying latent variable which simplifies modeling the two responses as multivariate normal data. Particular attention is paid to the association between smoking status and body weight, in which case it is noted that the treatment weakens the association between smoking and weight gain.

2.4 Overview and Motivation

As we have described, smoking cessation studies generate longitudinal data in which data are collected in order to define smoking status. This is a very blunt way of summarizing what can be problematic measures of smoking status. Despite this, the dichotomizing of individuals into groups of smoking status, and often ignoring that individuals may be transitioning between smoking statuses is a common way to treat these data, both in the medical and statistical literature. Furthermore, two separate groups may not fully describe the nature of the treatment effect in such studies. This concern is described in more detail in section 3.3.4.

There is a need to describe the longitudinal data from smoking cessation studies in a much more efficient manner which allows for investigating change over time and in general, more effective use of the longitudinal data generated from such studies. With this in mind we will explore how to use the measurements used to define smoking status as a multivariate longitudinal response, that is governed by a hidden process which could closely resemble

smoking status. In Chapter 3, we will describe a modeling framework to accomplish this, where the heterogeneity in a multivariate longitudinal response attributable to between-subject differences in addition to heterogeneity arising from changes in disease state can be effectively modeled. In Chapter 4, we will discuss the benefits of utilizing multivariate approaches for these models, when compared to the univariate alternatives. We will also attempt to assess the utility and meaning of the hidden states to determine if they resemble some meaningful state that a patient may encounter in a smoking cessation study. In Chapter 5, we will describe methodologies for predicting hidden states, and we will also examine how this hidden process approach may be used to handle missing data and dropout which can be problematic in studies such as smoking cessation. In the last chapter, we will summarize the findings in the previous chapters and discuss areas of potential future work.

Chapter 3

Multivariate Longitudinal Data Analysis with Mixed Effects Hidden Markov Models

3.1 Introduction

Longitudinal studies, where data on study subjects are collected over time, is increasingly involving multivariate longitudinal data where multiple responses are collected repeatedly over the study period. Approaches to the analysis of multivariate longitudinal data have included marginal (e.g., Xu and MacKenzie (2012)) and conditional models (e.g., Gueorguieva and Agresti (2001)). Other procedures generally aim to simplify the dependence between responses and therefore often reduce the computational burden involved in conducting inference in such models (e.g., Fieuws and Verbeke (2006)) or reducing the dimension of the longitudinal response (e.g., Scott, James and Sugar (2005)). The advantages of using multivariate longitudinal data are well documented and include a more complete use of the data collected and the ability to estimate and utilize the dependence between longitudinal responses, among others.

While these collected multivariate responses may generally correlate with some meaningful clinical endpoint, they may be inadequate in capturing the relevant endpoint and the study intervention's effect on this endpoint. For example, in studies of addiction, the relevant clinical endpoint is often the relapse or abstinence rates, which are conceptually understood, but are not effectively measurable. Estimation of these rates is usually done by attempting to determine the underlying disease states at each study visit via thresholds of often several measures of abstinence through biochemical or self-reported data. Both classes of measures used to estimate the disease states are limited by their sensitivity and specificity due to the potential for misclassification and measurement error.

This type of scenario is quite common in clinical trials, where an underlying or hidden disease state is modelled via one or more surrogate endpoints or biomarkers, most often by establishing thresholds of one or more of these responses to define the hidden disease states. For example, as discussed in Chapter 2, O'Malley et al. (2007) conducted a clinical trial evaluating a drug's efficacy for smoking cessation and collected two surrogate endpoints which had thresholds to define a disease state and the true outcome of interest, smoking abstinence. This approach often ignores: any unexplained heterogeneity in the responses by, for example, measurement error or between-subject heterogeneity, the dependence structure of the responses, and any information pertaining to the subject's response or disease state history.

Instead of defining thresholds or cutoffs to define disease states, it may be more efficient to model the data as generated from these hidden or latent disease states. Latent class models have successfully been used for multivariate longitudinal data (e.g., Proust-Lima et al. (2009)), but are generally limited by their static latent states. Recently, hidden Markov models (HMMs) have been used in the context of longitudinal studies (Altman, 2007; DeSantis and Bandyopadhyay, 2011; Maruotti and Ryden, 2009; Scott, James and Sugar, 2005; Shirley et al., 2010). Scott, James and Sugar (2005) developed a model for multivariate continuous longitudinal data by reducing the dimension of the response, and

modelling the reduced response as a HMM. Dimensional reduction can be challenging in a general setting, where responses may be of different data types (e.g., binary and continuous). Alternatively, Altman (2007) develops a general framework for using HMMs to model univariate longitudinal data by introducing random effects into both the observed response and hidden state components of the HMM. This work has subsequently been used by others (DeSantis and Bandyopadhyay, 2011; Maruotti, 2011; Shirley et al., 2010).

Using the general framework described by Altman (2007), we extend her work to the setting of multivariate longitudinal data. This is done by modelling the observed longitudinal responses jointly through separate, but correlated, random effects, and a common hidden Markov process. We apply this type of analysis to data generated from a smoking cessation clinical trial examining the efficacy of a pharmacological intervention.

In this paper we briefly review previous work done with mixed effects hidden Markov models (MHMMs) and extend this to the setting of multivariate longitudinal data in section 3.2. In section 3.3 we justify and outline a Bayesian procedure for computing estimates of such models using Markov chain Monte Carlo (MCMC). In sections 3.4 and 3.5, respectively, we perform a simulation study to understand the properties of the estimates generated from such models in a variety of situations, as well as perform an analysis from a motivating smoking cessation clinical trial example. In section 3.6 we will discuss the results of our models, simulations and possible limitations and extensions.

3.2 Mixed effects hidden Markov models for multivariate longitudinal responses

In this section we will describe a general framework for using mixed effects hidden Markov models (MHMMs) to describe longitudinal data. HMMs are characterized by two dependent processes. The first being a hidden process, which is usually assumed to adhere to the Markov assumption. The second, the observed process, which, when conditioned on

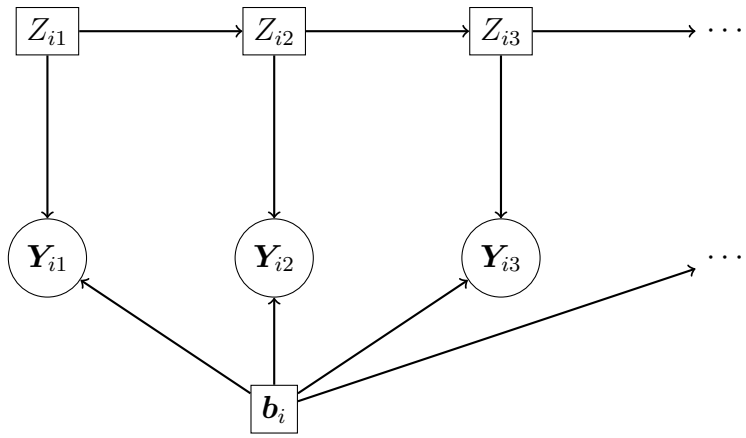

Figure 3.1: Mixed Effects Hidden Markov Model structure for subject i, with hidden process $\boldsymbol{Z}_i$, observed process $\boldsymbol{Y}_i$, and random effects, $\boldsymbol{b}_i$

the current value of the hidden process, is independent of any future or past value of the hidden or observed processes (see Figure 1). HMMs have a great number of applications (Zucchini and MacDonald, 2009), and as mentioned in the introduction, have recently been applied to the analysis of longitudinal data (Altman, 2007; DeSantis and Bandyopadhyay, 2011; Maruotti and Ryden, 2009; Scott, James and Sugar, 2005; Shirley et al., 2010).

3.2.1 Model Framework

Let Z_{it} denote the hidden process for subject $i, (i = 1, 2, \ldots, N)$ at some common time points $t, (t = 1, 2, \ldots, n_i)$. Further, denote $\boldsymbol{Z_i}$ to comprise all values of Z_{it} for subject i, and $\boldsymbol{Z}$ to be all values of $\boldsymbol{Z_i}$ for all i. Assume Z_{it} arises from an $M-$state Markov chain with transition matrix $\boldsymbol{P}$ and initial probability vector, $\boldsymbol{\pi}$. This paper will primarily focus on HMMs arising from first-order time homogeneous Markov chain hidden processes, where $P(Z_{it} = k|Z_{i,t-1} = j) = p_{jk}$ for all $t > 1$, and $P(Z_{i1} = k) = \pi_k$, where $j, k = 1, \ldots, M$.

Let $\boldsymbol{y}_{it}$ denote a $R-$dimensional multivariate longitudinal response for subject i at time t, such that $\boldsymbol{y}_{it} = [y_{i1t}, y_{i2t}, \ldots, y_{iRt}]'$. Similar to above, let $\boldsymbol{y_i}$ denote all the observations

for subject i across all longitudinal responses and time points, and $\boldsymbol{y}$ be the collection of all $\boldsymbol{y_i}$ for $i = 1, 2, \ldots, N$.

Suppose that the joint longitudinal responses arise from an HMM such that $\boldsymbol{Y}_{it}$ and $\boldsymbol{Y}_{it'}$ were independent given the HMM's current state, Z_{it} and $Z_{it'}$, respectively, for $t \neq t'$. Further assume that there exists an additional set of $i.i.d$ subject-specific random effects, $\boldsymbol{b}_i$, where conditional on the current hidden states, Z_{it}, and random effects, $\boldsymbol{b}_i$, each y_{irt} is independent (see Figure 1) with a distribution, $f_{rk}(y_{irt}|z_{it} = k)$, in the exponential family with canonical parameter, θ_{irtk}, and link function $h_r(.)$, where:

$$h_r(\theta_{irtk}) = \eta_{irt} = \boldsymbol{H}'_{irtk}\boldsymbol{\tau}_r + \boldsymbol{x}'_{irt}\boldsymbol{\beta}_{rk} + \boldsymbol{w}'_{irtk}\boldsymbol{b}_i \ . \tag{3.1}$$

Here $\boldsymbol{\tau}_r$ represents the fixed effect intercept term for each hidden state, as defined through an M-dimensional hidden state vector, $\boldsymbol{H}_{irtk}$, which defines some contrast dependent on Z_{it}. Typically, we will define $\boldsymbol{H}_{irtk}$ such that for hidden state k, $\boldsymbol{H}'_{irtk}\boldsymbol{\tau}_r$ is the cumulative sum of the first k elements of $\boldsymbol{\tau}_r$ (see page 36 for more details). Further, $\boldsymbol{x_{irt}}$ is a vector of covariates at time t for subject i when $Z_{it} = k$ for the rth longitudinal response, and the associated fixed effect coefficient parameters for these covariates are represented by $\boldsymbol{\beta}_{rk}$, which also could vary over different values of the hidden process. Finally, $\boldsymbol{w_{irtk}}$ is the row vector of the model matrix for the random effect for the rth longitudinal response for subject i at time t while in hidden state k.

When $R = 1$, i.e., the longitudinal response is univariate, this is one of the models proposed by Altman (2007) where she outlines a procedure to do maximum likelihood estimation based on an approximation to the marginal likelihood:

$$L(\boldsymbol{\Theta}; \boldsymbol{y}) = \prod_{i=1}^{N} \int_{\boldsymbol{b}_i} \boldsymbol{\alpha}_{in_i} \boldsymbol{1}' f_{\boldsymbol{b}_i}(\boldsymbol{b_i}|\boldsymbol{\Theta}) d\boldsymbol{b_i}, \tag{3.2}$$

by Gauss-Hermite quadrature, where $\boldsymbol{\Theta}$ comprises of all parameters of the MHMM, including parameters associated with the hidden process, $\boldsymbol{P}$ and $\boldsymbol{\pi}$, parameters associated with the canonical link of the observed processes in (3.1), $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, parameters associated with

the random effects, $\boldsymbol{\Sigma}$, and additional parameters associated with the observed process, unrelated to (3.1), $\boldsymbol{\xi}$. Here $\boldsymbol{\alpha}_{in_i}$ is a quantity often referred to as the *forward probabilities* (Baum et al., 1970) for subject i, and defined for $t = 1$ as $\boldsymbol{\alpha}_{i1} = \boldsymbol{\pi}\boldsymbol{G}(y_{i11})$, where,

$$\boldsymbol{G}(y_{i1t}) = \boldsymbol{diag}[f_{11}(y_{i1t}|z_{it} = 1, \boldsymbol{b}_i, \boldsymbol{\Theta}), f_{12}(y_{i1t}|z_{it} = 2, \boldsymbol{b}_i, \boldsymbol{\Theta}), \ldots, f_{1M}(y_{i1t}|z_{it} = M, \boldsymbol{b}_i, \boldsymbol{\Theta})],$$

for any $t > 0$ and recursively for $t = 2, 3, \ldots, n_i$, as $\boldsymbol{\alpha_{it}} = \boldsymbol{\alpha_{i,t-1}}\boldsymbol{P}\boldsymbol{G}(y_{i1t})$.

Extending this structure for $R > 1$ involves only modifying $\boldsymbol{G}(y_{i1t})$, and using the conditional independence assumption, such that

$$\boldsymbol{G}(\boldsymbol{y}_{it}) = \boldsymbol{diag}[\prod_{r=1}^{R} f_{r1}(y_{irt}|z_{it} = 1, \boldsymbol{b}_i, \boldsymbol{\Theta}), \prod_{r=1}^{R} f_{r2}(y_{irt}|z_{it} = 2, \boldsymbol{b}_i, \boldsymbol{\Theta}), \ldots, \prod_{r=1}^{R} f_{rM}(y_{irt}|z_{it} = M, \boldsymbol{b}_i, \boldsymbol{\Theta})],$$

for $t > 0$.

A common scenario involving $R > 1$ would be to have separate but correlated random effects for each distinct longitudinal response. Under the scenario of a single random intercept in each response, $\boldsymbol{w_{irtk}}$ would simply be a vector of all zeroes, except the rth entry which would be one. Alternatively, a shared random effect across responses could also be used, and is accommodated by this parameterization, but the focus of this paper will be on the case of separate, but correlated random effects. To evaluate $L(\boldsymbol{\Theta}; \boldsymbol{y})$ under these situations becomes increasingly difficult as R increases, as each evaluation often involves the computation of N $R-$ (or more) dimensional integrals, most often with no closed-form solutions. Although better numerical approximations to the integral may be possible, the integrand is not a relatively simple function of $\boldsymbol{b}_i$ due to the inclusion of the hidden process. Nonetheless, we have been able to fit such models when $R = 2$ by a Gauss-Hermite quadrature approximation, but we have found in general as R increases there is a need for computationally sophisticated methods to fit these models. We will therefore focus the attention of this paper to other methods for computing such models.

3.2.2 Motivating Example

A motivating example arises from a pharmacotherapy smoking cessation study, where smokers were randomized to one of three active treatment groups or a placebo, with all participants receiving nicotine replacement therapy (O'Malley et al., 2007). After randomization, subjects were followed for six weeks, with longitudinal follow-up occurring weekly. At each study visit, patients self-reported daily cigarette use over the previous week, and a carbon monoxide (CO) test was administered as physiological monitoring of cigarette intake.

In this example, we will be modelling a bivariate longitudinal response consisting of a count, Y_{i1t}, of the number self-reported days out of n_{it} in t^{th} week where subject i reported to smoke at least one cigarette, and Y_{i2t} which consists of the transformed ($\log[Y+1]$) CO level at the tth week's study visit. Under the framework outlined in section 2, we model these data as follows:

$$
\begin{aligned}
Y_{i1t}|Z_{it} = k, \boldsymbol{b}_i &\sim Binomial(n_{it}, \theta_{i1tk}) \text{ where,} \\
logit(\theta_{i1tk}) &= \boldsymbol{H}'_{i1tk}\boldsymbol{\tau}_1 + \boldsymbol{x}'_{i1t}\boldsymbol{\beta}_{1k} + \boldsymbol{w}'_{i1tk}\boldsymbol{b}_i \\
Y_{i2t}|Z_{it} = k, \boldsymbol{b}_i &\sim N(\theta_{i2tk}, \sigma^2_\epsilon) \text{ where,} \\
\theta_{i2tk} &= \boldsymbol{H}'_{i2tk}\boldsymbol{\tau}_2 + \boldsymbol{x}'_{i2t}\boldsymbol{\beta}_{2k} + \boldsymbol{w}'_{i2tk}\boldsymbol{b}_i \ .
\end{aligned}
$$

This is only one form of how this complex multivariate longitudinal data could be described. It has been hypothesized (Hughes et al., 2003; Killeen, 2011) that smoking cessation interventions are described by a three state Markov model with a state space of: regular use, withdrawal and long term abstinence. While these states are generally understood, determining the state of a given subject at a given time remains difficult, as self-reported smoking (Y_{i1t}) is known to generally underreport both the prevalence and intensity of smoking (Patrick et al., 1994), and the higher CO levels (Y_{i2t}) observed when smoking have a limited half-life where they are detectable (perhaps as little as 4-6 hours). Further, in most analyses of these types of data, thresholds of one or more of these longitudinal re-

sponses define a binary status (smoker/non-smoker) ignoring the intermediate withdrawal state. Under the models described in this section, we will examine the properties of both the hidden and observed processes.

3.3 Bayesian Inference

Although it is possible to use maximum likelihood estimation by approximating (3.2) by numerical integration, it, in general, remains difficult to apply such methods to the framework discussed in section 3.2.1 when the dimension of the longitudinal response exceeds two, or in many other common scenarios. Indeed, when examining more complicated random effect structures for a single dimensional longitudinal response under a MHMM, Altman (2007) relies primarily on stochastic integration to fit such models. While it may be possible to continue with this approach as R or the number of random effects increase, we have generally found the Bayesian approach to be more efficient from a computational perspective, as well as relatively easy to adapt to many other more complicated situations and models.

3.3.1 Prior Specification

Specifying a prior for $\boldsymbol{\Theta}$ is largely dependent on the specific circumstances for the target of inference. It is often convenient for sampling to choose independent conjugate priors for each component of $\boldsymbol{\Theta}$. In other circumstances the choice of prior may often depend on the distributional assumptions of each dimension of the longitudinal response, the chosen link function, (3.1), or other requirements such as subject-expert knowledge, if available. In general, it is often convenient to factor $[\boldsymbol{\Theta}]$ as:

$$[\boldsymbol{\Theta}, \boldsymbol{b}] = [\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{P}, \boldsymbol{\pi}, \boldsymbol{b}, \boldsymbol{\Sigma}, \boldsymbol{\xi}] = [\boldsymbol{\beta}, \boldsymbol{\tau}] \times [\boldsymbol{P}] \times [\boldsymbol{\pi}] \times [\boldsymbol{b}|\boldsymbol{\Sigma}] \times [\boldsymbol{\Sigma}] \times [\boldsymbol{\xi}] \ . \tag{3.3}$$

While there may be certain circumstances where it is merited to use the joint prior of $\boldsymbol{\xi}$ (such as when two or more responses are normally distributed conditional on their random effects and hidden states), it is often useful to factorize each component of $[\boldsymbol{\xi}]$ as $\prod_{r=1}^{R}[\boldsymbol{\xi_r}]$. Once the priors have been specified, sampling from the posterior distribution, $[\boldsymbol{\Theta}|\boldsymbol{y}]$, can be done through Gibbs sampling (Geman and Geman, 1984) or similar approaches. If one could sample the hidden states, $\boldsymbol{z}$, at each iteration of the Gibbs sampler, one could then sample each component of $\boldsymbol{\Theta}$, conditional on the current simulated value of the other parameters, $\boldsymbol{y}$, and the hidden states, $\boldsymbol{z}$. After this is completed, the hidden states, $\boldsymbol{z}$, must be updated, which will be discussed just below in section 3.3.2. This procedure is then repeated until stationarity is reached, effectively sampling from the posterior, $[\boldsymbol{\Theta}|\boldsymbol{y}]$.

3.3.2 Updating the Hidden States

Taking a Gibbs sampling approach for sampling from the posterior involves alternating between sampling from $[\boldsymbol{Z}|\boldsymbol{y}, \boldsymbol{\Theta}, \boldsymbol{b}]$, and $[\boldsymbol{\Theta}, \boldsymbol{b}|\boldsymbol{y}, \boldsymbol{z}]$. Sampling from the latter is relatively straightforward, and there are many possible approaches to this which are discussed in section 3.3.3. On the other hand, updating the hidden states has been well studied in the Bayesian context using MCMC methods, and Scott (2002) examines two different updating schemes in the setting of HMMs. The preferred option involves sampling from $[\boldsymbol{Z}_i|\boldsymbol{y}, \boldsymbol{\Theta}]$ through a stochastic recursion (Scott, 2002) outlined below, instead of sampling $[Z_{it}|\boldsymbol{y}, \boldsymbol{\Theta}, \boldsymbol{b}_i, \ldots, Z_{i,t-2}, Z_{i,t-1}, Z_{i,t+1}, Z_{i,t+2}, \ldots]$ separately. This method, while more computationally intensive, has been found to mix better in some situations (Cappe, Moulines and Ryden, 2005; Scott, 2002), and considering the addition of another hidden component to our models, mainly $\boldsymbol{b}_i$, improving mixing even at the expense of reasonably efficient computation is a desired attribute.

This update procedure involves computation of *forward probabilities* discussed in section 2.1. Each sequence of hidden states $Z_{i1}, Z_{i2}, \ldots, Z_{in_i}$ can be done separately for each i. In general, $\alpha_{it}(j) = P(Z_{it} = j, \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}|\boldsymbol{b}_i, \boldsymbol{\Theta})$, where $\alpha_{it}(j)$ is the jth element of $\boldsymbol{\alpha_{it}}$.

Calculating $\boldsymbol{\alpha_{it}}$ is done recursively, and is a function of only $\boldsymbol{\Theta}$, $\boldsymbol{b_i}$ and $\boldsymbol{y}$. For $t = n_i$, we note that $\alpha_{in_i}(j) = P(Z_{in_i} = j, \boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\Theta}) \propto P(Z_{in_i} = j|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$. Therefore, sampling Z_{in_i} can be done by calculating $\alpha_{in_i}(j)$ for each hidden state j, and then normalizing by dividing by their sum over j, $\sum_{j=1}^{M} \alpha_{in_i}(j)$. Sampling the remaining $n_i - 1$ hidden states for subject i can be reduced to sampling from $[Z_{it}|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta}, Z_{i,t+1}, \ldots, Z_{in_i}]$ for $t = n_i - 1, n_i - 2, \ldots, 1$. This is done by using an argument originating from Chib (1996), and described in details by Zucchini and MacDonald (2009) and adapted to the situation we encounter such that:

$$
\begin{aligned}
&P(Z_{it}|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta}, Z_{i,t+1}, \ldots, Z_{in_i}) \propto \\
&P(Z_{it}|\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}) P(\boldsymbol{y}_{i,t+1}, \ldots, \boldsymbol{y}_{in_i}, Z_{i,n_i+1}, \ldots, Z_{in_i}|\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}, Z_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}) \propto \\
&P(Z_{it}|\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}) P(Z_{i,t+1}|Z_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}) \times \\
&P(\boldsymbol{y}_{i,t+1}, \ldots, \boldsymbol{y}_{in_i}, Z_{i,n_i+2}, \ldots, Z_{in_i}|\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}, Z_{it}, Z_{t+1}, \boldsymbol{b}_i, \boldsymbol{\Theta}) \propto \\
&P(Z_{it}|\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}) P(\boldsymbol{y}_{i,t+1}, \ldots, \boldsymbol{y}_{in_i}, Z_{i,n_i+2}, \ldots, Z_{in_i}|Z_{t+1}, \boldsymbol{b}_i, \boldsymbol{\Theta}) P(Z_{i,t+1}|Z_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}) \propto \\
&P(Z_{it}, \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}|\boldsymbol{b}_i, \boldsymbol{\Theta}) P(Z_{i,t+1}|Z_{it}, \boldsymbol{b}_i, \boldsymbol{\Theta}).
\end{aligned}
$$

The last line is familiar, as both of the quantities in the product have been described previously, where $P(Z_{it} = j, \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{it}|\boldsymbol{b}_i, \boldsymbol{\Theta}) = \alpha_{it}(j)$, and $P(Z_{i,t+1} = k|Z_{it} = j, \boldsymbol{b}_i, \boldsymbol{\Theta}) = p_{jk}$ from the transition probability matrix, $\boldsymbol{P}$. Hence, after Z_{in_i} has been sampled, the remainder of the hidden states can be sampled in reverse order ($t = n_i - 1, n_i - 2, \ldots, 1$) by computing $P(Z_{it} = j|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta}, Z_{i,t+1} = k) \propto [\alpha_{it}(j)]p_{jk}$ for $j = 1, \ldots, M$, and normalizing each by their sum over all j. This process can be repeated for each i separately, and once updated, sampling from the other components of the posterior can be conducted.

3.3.3 Sampling from the Posterior

Under the prior structure discussed in section 3.3.1, we can take a Gibbs sampling approach, where groups of parameters with priors in conjugate families are sampled from directly. With all other variables, alternative methods can be used such as adaptive rejection sampling (Gilks and Wild, 1992), or Metropolis within Gibbs sampling (Smith and

Roberts, 1993). Given the current values of $\boldsymbol{\Theta}$ and $\boldsymbol{b}$, section 3.3.2 discusses how to sample from $[\boldsymbol{Z}|\boldsymbol{y}, \boldsymbol{\Theta}, \boldsymbol{b}]$. This is generally one of the most computationally intensive steps.

Under the MHMM assumptions, and applying the form of (3.3), we can get relatively simple forms of the posterior distributions: $[\boldsymbol{P}|\boldsymbol{\Theta}_{-\boldsymbol{P}}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}] = [\boldsymbol{P}|\boldsymbol{z}]$, $[\boldsymbol{\pi}|\boldsymbol{\Theta}_{-\boldsymbol{\pi}}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}] = [\boldsymbol{\pi}|\boldsymbol{z}]$, and $[\boldsymbol{\Sigma}|\boldsymbol{\Theta}_{-\boldsymbol{\Sigma}}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}] = [\boldsymbol{\Sigma}|\boldsymbol{b}]$, where the notation $[A|B_{-A}]$ is used to represent the conditional distribution of A given all elements in B *except* A. If conjugate priors are chosen for $\boldsymbol{P}$, $\boldsymbol{\pi}$, and $\boldsymbol{\Sigma}$ as we have done below, this often involves directly sampling from Dirichlet and inverse Wishart distributions (discussed further below).

The full conditional distributions of the other sets of parameters, $\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{b}, \boldsymbol{\xi}$ will generally depend on some or all of the other parameters, in addition to the observed data, $\boldsymbol{y}$ and current simulated hidden states, $\boldsymbol{z}$. We have found it convenient to sample from the $\boldsymbol{\beta}_r$ and $\boldsymbol{\tau}_r$ separately for each r. Under this approach, $[\boldsymbol{\beta}_r, \boldsymbol{\tau}_r|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}, \boldsymbol{\xi}_r] \propto [\boldsymbol{\beta}_r, \boldsymbol{\tau}_r] \times \prod_{i=1}^{N} \prod_{t=1}^{n_i} f_{rk}(y_{irt}|\boldsymbol{b}_i, \boldsymbol{\beta}_r, \boldsymbol{\tau}_r, \boldsymbol{\xi}_r, z_{it} = k)$, and $[\boldsymbol{b}_i|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\Sigma}] \propto$
$[\boldsymbol{b}_i|\boldsymbol{\Sigma}] \times \prod_{t=1}^{n_i} \prod_{r=1}^{R} f_{rk}(y_{irt}|\boldsymbol{b}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\xi}_r, z_{it} = k)$. Lastly, sampling from $[\boldsymbol{\xi}|.]$ will largely be application specific, but will not generally depend on $\boldsymbol{P}$ or $\boldsymbol{\pi}$.

Sampling from the Posterior for the Motivating Example

Sampling from the posterior for the example in section 3.2.2 is relatively straightforward. As above, we will initially choose a relatively uninformative Dirichlet conjugate prior for all treatment groups or strata, s, with prior hyperparameters $\boldsymbol{\zeta}_{\pi}^{(s)}$ and $\boldsymbol{\zeta}_{u}^{(s)}$ set to $[1, 1, 1]$ for $\boldsymbol{\pi}^{(s)}$ and the uth row of $\boldsymbol{P}$, respectively, and in this setting we will assume that $\boldsymbol{b}_i \overset{\text{iid}}{\sim} MVN(\boldsymbol{0}, \boldsymbol{\Sigma})$. We also chose a conjugate inverse $Wishart(\nu_0, \boldsymbol{\Psi}_0)$ prior for $\boldsymbol{\Sigma}$, with $\nu_0 = 1$ and let $\boldsymbol{\Psi}_0$ be the 2×2 identity matrix. Lastly, we chose a flat prior for $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, such that $[\boldsymbol{\beta}, \boldsymbol{\tau}] \propto 1$. In this example, there is also one extra parameter for $r = 2$, where $\xi_{21} = \frac{1}{\sigma_\epsilon^2}$, which will also come from a conjugate family, $Gamma(a_0, b_0)$, with $a_0 = 0.001, b_0 = 0.0002$.

Once $\boldsymbol{Z}_i$ is updated for each i as outlined in section 3.3.2, the following can be sampled

from the full conditional distributions:

$$[\boldsymbol{\Sigma}|\boldsymbol{b}] \sim InverseWishart(N+\nu_0, \boldsymbol{\Psi}_0 + \sum_{i=1}^{N} \boldsymbol{b}_i\boldsymbol{b}_i^{'}),$$

$$[\frac{1}{\sigma_\epsilon^2}|\boldsymbol{y},\boldsymbol{\beta}_2, \boldsymbol{\tau}_2, \boldsymbol{b}, \boldsymbol{z}] \sim$$

$$Gamma(\frac{(\sum_{i=1}^{N} n_i + a_0)}{2}, \quad (b_0 + \frac{(\sum_{i=1}^{N}\sum_{t=1}^{n_i}(y_{i2t} - \boldsymbol{H}_{i2tz_{it}}^{'}\boldsymbol{\tau}_2 - \boldsymbol{x}_{i2t}^{'}\boldsymbol{\beta}_{2z_{it}} - \boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i)^2)}{2}),$$

$$[\boldsymbol{\pi}^{(s)}|\boldsymbol{z}] \sim Dirichlet(\boldsymbol{\zeta}_\pi^{(s)} + \boldsymbol{q}^{(s)}),$$

where $\boldsymbol{q}^{(s)}$ is a M-dimensional vector with m^{th} element $\sum_{i=1}^{N}\sum_{t=1}^{n_i} \mathbb{1}(z_{i1} = m, v_i = s)$,

$[\boldsymbol{P}_u^{(s)}|\boldsymbol{z}] \sim Dirichlet(\boldsymbol{\zeta}_u^{(s)} + \boldsymbol{q}_u^{(s)})$, where $\boldsymbol{P}_u^{(s)}$ is the uth row of $\boldsymbol{P}^{(s)}$,

where $\boldsymbol{q}_u^{(s)}$ is a M-dimensional vector with m^{th} element $\sum_{i=1}^{N}\sum_{t=2}^{n_i} \mathbb{1}(z_{it} = m, z_{i,t-1} = u, v_i = s)$,

$$[\boldsymbol{\tau}_1, \boldsymbol{\beta}_1|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}] \propto \prod_{i=1}^{N}\prod_{t=1}^{n_i}\left(\frac{\exp(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i}{1+\exp(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i)}\right)^{y_{i1t}} \times$$
$$\left(1 - \frac{\exp(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i}{1+\exp(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i)}\right)^{n_{it}-y_{i1t}},$$

$$[\boldsymbol{\tau}_2, \boldsymbol{\beta}_2|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}, \sigma_\epsilon^2] \propto \prod_{i=1}^{N}\prod_{t=1}^{n_i}\exp\left(-\frac{(y_{i2t} - \boldsymbol{H}_{i2tz_{it}}^{'}\boldsymbol{\tau}_2 - \boldsymbol{x}_{i2t}^{'}\boldsymbol{\beta}_{2z_{it}} - \boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i)^2}{2\sigma_\epsilon^2}\right)$$
$$\sim MVN((\tilde{\boldsymbol{B}}^{'}\tilde{\boldsymbol{B}})^{-1}\tilde{\boldsymbol{B}}^{'}\tilde{\boldsymbol{y}_2}, \sigma_\epsilon^2(\tilde{\boldsymbol{B}}^{'}\tilde{\boldsymbol{B}})^{-1}), \text{ where}$$
$$\tilde{\boldsymbol{B}} = \Big[\boldsymbol{H}|\boldsymbol{X}\Big]$$

and is a $\left(\sum_{i=1}^{N} n_i\right) \times (P+c)$ matrix(c being the length of $\boldsymbol{x_{i2t}}$)

with rows comprised of, $[\boldsymbol{H}_{i2tz_{it}}^{'}, \boldsymbol{x}_{i2t}^{'}]$,

and $\tilde{\boldsymbol{y}_2}$ is a $\left(\sum_{i=1}^{N} n_i\right)$ length vector with elements $y_{i2t} - \boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i$.

$$\text{Also, } [\boldsymbol{b}_i|\boldsymbol{y_i}, \boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\epsilon^2] \propto \prod_{t=1}^{n_i} \left(\frac{\exp\left(\boldsymbol{H}'_{i1tz_{it}}\boldsymbol{\tau}_1 + \boldsymbol{x}'_{i1t}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}'_{i1t}\boldsymbol{b}_i\right)}{1 + \exp\left(\boldsymbol{H}'_{i1tz_{it}}\boldsymbol{\tau}_1 + \boldsymbol{x}'_{i1t}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}'_{i1t}\boldsymbol{b}_i\right)} \right)^{y_{i1t}} \times$$

$$\left(1 - \frac{\exp\left(\boldsymbol{H}'_{i1tz_{it}}\boldsymbol{\tau}_1 + \boldsymbol{x}'_{i1t}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}'_{i1t}\boldsymbol{b}_i\right)}{1 + \exp\left(\boldsymbol{H}'_{i1tz_{it}}\boldsymbol{\tau}_1 + \boldsymbol{x}'_{i1t}\boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}'_{i1t}\boldsymbol{b}_i\right)} \right)^{n_{it}-y_{i1t}} \times$$

$$\exp\left(-\frac{(y_{i2t} - \boldsymbol{H}'_{i2tz_{it}}\boldsymbol{\tau}_2 - \boldsymbol{x}'_{i2t}\boldsymbol{\beta}_{2z_{it}} - \boldsymbol{w}'_{i2t}\boldsymbol{b}_i)^2}{2\sigma_\epsilon^2} \right) \times \left[\exp\left(-\frac{1}{2}\boldsymbol{b}'_i\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i \right) \right],$$

We use a random walk Metropolis within Gibbs steps to sample from the full conditionals of $[\boldsymbol{b}_i|.]$ and $[\boldsymbol{\beta}_r, \boldsymbol{\tau}_r|.]$ for $r = 1$ separately, where $r = 2$ has a multivariate normal full conditional distribution which can be sampled from directly. The interpretation of $\boldsymbol{\tau}_r$ depends on the defined model matrix, $\boldsymbol{H}_{irt}$ common to all i and t, but not necessarily r (although we will not examine this here). For instance, the contrasts suggested by Scott (2002) are similar to the following: $\boldsymbol{H}'_{irt1} = [1, 0, 0]$, $\boldsymbol{H}'_{irt2} = [1, 1, 0]$, and $\boldsymbol{H}'_{irt3} = [1, 1, 1]$ for $Z_{it} = 1, 2$ and 3, respectively, and we will use these throughout the remainder of the paper. To perform sampling for the two steps which require Metropolis within Gibbs sampling, a multivariate normal proposal distribution is used, and updated during a burn-in period. Approximately 1.5 million iterations are used, and thinned to 50,000 samples. Trace plots, density estimates, and auto-correlation plots as an assessment of stationarity and mixing are presented in section 3.7.

3.3.4 Introducing a Treatment Effect for the Hidden Process

Thus far we have only discussed the situation where covariates are introduced to the parameters associated with the observed component of the MHMM. While this may be useful in many situations, often the desired treatment contrast may manifest itself through some function of the transition probabilities. Indeed, in situations where $\boldsymbol{y}$ are surrogate endpoints or biomarkers, defining a treatment effect (via $\boldsymbol{\beta}$ in (3.1)) in terms of observed longitudinal responses may be difficult depending on the treatment's effect on: the underlying hidden process, the observed longitudinal responses, or the relationship between these observed and hidden responses. In situations where the hidden states resemble the

underlying *true* disease outcomes, as they often can, it may be more advantageous to think about the treatment's effect on the hidden states.

With the above said, we can define simple treatment effects for a finite number of categorical time-invariant covariates by defining a subject-specific transition matrix, $\boldsymbol{P}_i = \sum_{s=1}^{S} \mathbb{1}(v_i = s)\boldsymbol{P}^{(s)}$, where v_i is a variable indicating treatment group membership, and $\boldsymbol{P}^{(s)}$ is the transition matrix for the sth treatment group. Similarly, a treatment effect can be defined in terms of the initial probabilities, where $\boldsymbol{\pi}_i = \sum_{s=1}^{S} \mathbb{1}(v_i = s)\boldsymbol{\pi}^{(s)}$, where $\boldsymbol{\pi}^{(s)}$ are the initial probabilities for treatment group s. Often, the HMM is assumed to be stationary, and/or the initial probability vector is considered a vector of nuisance parameters, but there may be circumstances where this quantity has a particular meaning. This scenario is probably most interesting in the setting of a randomized clinical trial where the hidden states of the study subjects may be considered very homogeneous at baseline. For instance, at enrollment in the motivating example, all subjects were required to be current smokers, so in principle the initial probability vector has a very specific interpretation – being that it is the first week's transition probability from hidden state three (the highest smoking state), or something conceptually similar.

Given a large enough sample size in each stratum with a reasonably high frequency of transitions between states within each strata, it is possible to add as many treatment groups as needed. Other methods would be required to sample from the posterior in cases where the desired covariate of interest was non-categorical.

In our motivating example from section 3.2.2, we will examine four different scenarios involving different ways to look at the treatment effect within the hidden process. Scenarios include:

I) No treatment effect, where $\boldsymbol{P}_i = \boldsymbol{P}$ and $\boldsymbol{\pi}_i = \boldsymbol{\pi}$ (common) for all i,

II) Treatment effect in only the initial state probabilities ($\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$), with common transition matrix, $\boldsymbol{P}^{(0)} = \boldsymbol{P}^{(1)} = \boldsymbol{P}$,

III) Treatment effect in only the transition probabilities ($\boldsymbol{P}^{(0)}$ and $\boldsymbol{P}^{(1)}$), with common initial state probabilities, $\boldsymbol{\pi}^{(0)} = \boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}$, and

IV) Treatment effect in both the initial state probabilities ($\boldsymbol{\pi}^{(0)}$ and $\boldsymbol{\pi}^{(1)}$), and the transition probabilities ($\boldsymbol{P}^{(0)}$ and $\boldsymbol{P}^{(1)}$).

3.3.5 Label Switching

A well known phenomenon in using MCMC approaches to mixture models including HMMs is that of label switching. This phenomenon generally occurs because of the invariance of the likelihood under any relabeling of the mixture components. Several approaches have been proposed to alleviate this problem, including putting identifiability constraints on the parameter space, and decision theoretic approaches (Stephens, 2000). Under a MHMM, we have found label switching to occur only under certain situations. As we shall see, in our motivating example the modes of most components of the observed multivariate longitudinal responses are well separated by the hidden states, and we do not generally observe label switching. This is consistent with an observation noted by Scott, James and Sugar (2005), who also used Bayesian methodology to examine a HMM with a longitudinal response, but without between-subject heterogeneity in the observed response; they did not observe label switching in their study either. Scott, James and Sugar (2005) attributes this mainly to the increased dimensionality of the observed longitudinal response.

In other settings, we do find label switching to occur, most often when the modes are less well-separated, and the between-subject heterogeneity is high. When this occurs, after the sampling is completed, and as is frequently done, we relabel the states by ordering one of the observed response population means. In the examples examined in this paper we have a bivariate observed longitudinal response, and generally we have not found that choosing one response's parameters over the other to significantly effect the results. Generally, the ordering of the response's parameters disagreed at a relatively low frequency (<1%),

and we picked the continuous response ($r = 2$) on which to reorder the labels. In other situations, such as when the dimension of the longitudinal responses increase, this may no longer be the case, and other strategies may be required.

3.4 Simulation Study

A simulation study was conducted to assess the properties of the estimators derived from the modeling framework and MCMC sampling methods described above and to assess robustness in a few situations of possible concern. Data were simulated for a model similar to those described in the example discussed in section 3.2.2, where the hidden process, $\boldsymbol{z}$, and the random effects, $\boldsymbol{b}$, are first sampled independently (unless stated otherwise), and given these data, the observed multivariate longitudinal response, $\boldsymbol{y}$ is simulated. Patients are randomly allocated in a 3:1 (265:89) manner to treatment and placebo, which have identical population level models with the exception of their respective transition probability matrices (scenario III in section 3.3.4). As is the case in the motivating example, the observed process is sampled at six time points ($n_i = 6$ for all i). For the purposes of this simulation study, data are assumed to not be missing, and 200 simulations from each situation will be fit using the methodology described in section 3.3.3. The following situations were examined:

1. The data are generated from a MHMM with well-separated modes ($\boldsymbol{\tau}_1 = [-5, 3, 3]$, $\boldsymbol{\tau}_2 = [1.2, 0.2, 1.1]$) with a large between-subject heterogeneity in the binomial response via independent multivariate normal random effects. The random effects are $i.i.d$ and have a covariance matrix $\boldsymbol{\Sigma} = \left[\begin{smallmatrix} 8.5 & 0.5 \\ 0.5 & 0.1 \end{smallmatrix}\right]$ with the inverse of the residual error, $\frac{1}{\sigma_\epsilon^2}$, in the normally distributed response set at 10.0. The hidden process is generated from a common initial probability vector, $\boldsymbol{\pi} = [0.6, 0.25, 0.15]$, and placebo transition matrix $\boldsymbol{P}^{(0)} = \left[\begin{smallmatrix} 0.85 & 0.10 & 0.05 \\ 0.60 & 0.20 & 0.20 \\ 0.10 & \mathbf{0.10} & \mathbf{0.80} \end{smallmatrix}\right]$ with treatment transition matrix, $\boldsymbol{P}^{(1)} = \left[\begin{smallmatrix} 0.85 & 0.10 & 0.05 \\ 0.60 & 0.20 & 0.20 \\ 0.10 & \mathbf{0.30} & \mathbf{0.60} \end{smallmatrix}\right]$.

2. The data are generated from a MHMM with less well-separated modes ($\boldsymbol{\tau}_1 = [-3, 1.5, 1.5]$, $\boldsymbol{\tau}_2 = [1.2, 0.5, 0.5]$) with a smaller between-subject heterogeneity in the binomial response via independent multivariate normal random effects. The random effects are again $i.i.d$ multivariate normal and have a covariance matrix $\boldsymbol{\Sigma} = \left[\begin{smallmatrix} 5.0 & 0.5 \\ 0.5 & 0.1 \end{smallmatrix}\right]$. The other sets of parameters, $\boldsymbol{P}^{(0)}$, $\boldsymbol{P}^{(1)}$, $\boldsymbol{\pi}$, and $\frac{1}{\sigma_\epsilon^2}$ remain the same as situation (1).

3. The data are generated similar to situation (2), but the distribution of the random effects, $\boldsymbol{b}_i$, are dependent on the initial hidden response, Z_{i1}. For the subjects in hidden state 1 at time $t = 1$, the random effects are generated from a mean zero multivariate normal distribution with covariance matrix, $\boldsymbol{\Sigma}_1 = \left[\begin{smallmatrix} 4.5 & 0.5 \\ 0.5 & 0.1 \end{smallmatrix}\right]$. Conversely, in those subjects in hidden states 2 or 3 at $t = 1$, the random effects are generated from a mean zero multivariate T distribution with 8 degrees of freedom, and covariance matrix, $\boldsymbol{\Sigma}_2 = \left[\begin{smallmatrix} 5.5 & 0.5 \\ 0.5 & 0.1 \end{smallmatrix}\right]$. As in situation (2), with the exception of $\boldsymbol{\tau}$ the rest of the parameters remain set at the same values as is situation (1).

Situation (1) closely describes the motivating example, where the modes of the hidden states are well-separated, and there is a great deal of heterogeneity in the binomial response. As can be seen in Table 3.1, the estimates generated from sampling from the posterior, $\pi(\boldsymbol{\Theta}|\boldsymbol{y})$, are generally well-behaved, and the posterior means are generally close to the true parameter values, while the posterior standard deviations are generally close to the standard deviation of the posterior means over the simulations. With the exception of Σ_{22}, the 0.025 and 0.975 posterior quantiles cover the true parameter value very close to 95% of the time, with overall mean coverage (all parameters) of 94.7%. In the exception (Σ_{22}), the reduced coverage may be due in part to the bias introduced by our prior.

Situation (2) generally resembles situation (1), but the population modes are less well-separated. As would be expected, for the relevant parameters ($\boldsymbol{\tau}, \boldsymbol{P}, \boldsymbol{\pi}$) the average posterior standard deviations have generally increased with the increased level of uncertainty pertaining to the hidden states. Additionally, some significant bias can be observed, in particular for p_{21}^0 and p_{32}^0 the bias is quite large. In both of these parameters (and most

Table 3.1: Simulation results for 200 runs of Situation (1)

Parameter	True Value	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage
π_1	0.60	0.5961	0.0437	0.0436	0.960
π_2	0.25	0.2555	0.0423	0.0440	0.955
p^0_{12}	0.10	0.1075	0.0391	0.0385	0.955
p^0_{13}	0.05	0.0534	0.0175	0.0178	0.950
p^0_{21}	0.60	0.5601	0.1115	0.0993	0.955
p^0_{23}	0.20	0.2033	0.0735	0.0681	0.960
p^0_{31}	0.10	0.1134	0.0445	0.0440	0.935
p^0_{32}	0.10	0.1135	0.0466	0.0452	0.960
p^1_{12}	0.10	0.1042	0.0237	0.0245	0.945
p^1_{13}	0.05	0.0510	0.0101	0.0093	0.975
p^1_{21}	0.60	0.5935	0.0631	0.0580	0.960
p^1_{23}	0.20	0.1976	0.0387	0.0383	0.945
p^1_{31}	0.10	0.1043	0.0366	0.0356	0.960
p^1_{32}	0.30	0.3070	0.0455	0.0454	0.955
τ_{11}	-5.00	-4.9905	0.2292	0.2374	0.945
τ_{12}	3.00	2.9859	0.1755	0.1772	0.925
τ_{13}	3.00	3.0311	0.2162	0.2262	0.940
τ_{21}	1.20	1.2028	0.0217	0.0236	0.940
τ_{22}	0.20	0.1983	0.0384	0.0424	0.940
τ_{23}	1.10	1.0987	0.0362	0.0381	0.930
Σ_{11}	8.50	8.5833	0.9981	1.0238	0.955
Σ_{12}	0.50	0.5100	0.0760	0.0648	0.985
Σ_{22}	0.10	0.1092	0.0103	0.0108	0.855
$\frac{1}{\sigma^2_\epsilon}$	10.00	10.0868	0.3916	0.3939	0.945

where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

others), the bias is in the direction towards the prior mean, which is 1/3 for p_{21}^0 and p_{32}^0. Indeed, the posterior for these parameters in many, but not all, of these simulated instances is very flat, and much of the information appears to come from the prior. It should be noted however, that both p_{21}^0 and p_{32}^0 are parameters associated with the placebo group, which has a reduced sample size when compared to the treatment group. In the corresponding treatment group parameters, p_{21}^1 and p_{32}^1, while we still observe some bias, it is considerably less than that observed in the smaller placebo group. This is probably not too surprising, given that based on a sample size of 89 and six time points, the true expected frequency of some of these hidden state transitions are as few as 9. Some caution is therefore merited when using these models when the frequency of such transitions may be low, but can be effectively monitored from the Z_{it} samples collected during the MCMC algorithm. In cases where the row totals of the transition frequencies are low, or have high variance, this may be a situation where significant bias may occur. In such cases, it may be necessary to reduce the hidden state space, so that the transition frequencies are high enough, or alternatively incorporate more information into the prior. With this in mind, we repeated this simulation with a more informative prior, such that the hyperparameters, $\boldsymbol{\zeta}_u^{(s)}$, for the transitions probabilities, are [8,1,1], [6,2,2] and [1,1,8] for the first, second and third rows of $\boldsymbol{P^{(s)}}$, respectively, which we use across both sets of parameters associated with the placebo and treatment groups (i.e., for both $s = 0$ and 1). This analysis is also presented in Table 2 and is denoted "situation (4)", and generally reduces any bias observed in the problematic parameters. In spite of the noted bias in many of the transition probabilities noted in Table 2 for situation (2), it is noteworthy that the coverage probabilities of the 95% quantile range are generally well behaved, albeit many of these intervals are very wide.

As can be seen in Table 3, in situation (3) we encounter many of the same problems noted in situation (2). Generally, the parameter estimates are not any more biased than those in situation (2). This suggests that incorporating additional prior information may

Table 3.2: Simulation results for 200 runs for each of Situations (2) and (4)

		Situation (2) - Relatively Uninformative Prior on $\boldsymbol{P}$				Situation (4) - Relatively Informative Prior on $\boldsymbol{P}$			
Parameter	True Value	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage
π_1	0.60	0.5574	0.0812	0.0704	0.950	0.5867	0.0602	0.0539	0.975
π_2	0.25	0.2814	0.0748	0.0645	0.960	0.2544	0.0586	0.0535	0.975
p^0_{12}	0.10	0.1228	0.0810	0.0632	0.980	0.0971	0.0498	0.0397	0.955
p^0_{13}	0.05	0.0614	0.0275	0.0222	0.960	0.0594	0.0230	0.0204	0.960
p^0_{21}	0.60	0.4881	0.1594	0.1338	0.945	0.5988	0.1145	0.0648	0.995
p^0_{23}	0.20	0.2140	0.1153	0.0874	0.990	0.2004	0.0876	0.0585	0.990
p^0_{31}	0.10	0.1249	0.0636	0.0472	0.995	0.1057	0.0507	0.0394	0.985
p^0_{32}	0.10	0.1572	0.0955	0.0757	0.975	0.1015	0.0613	0.0387	0.995
p^1_{12}	0.10	0.1282	0.0643	0.0519	0.960	0.0993	0.0393	0.0322	0.940
p^1_{13}	0.05	0.0597	0.0196	0.0178	0.945	0.0570	0.0163	0.0149	0.945
p^1_{21}	0.60	0.5682	0.1105	0.0857	0.975	0.6189	0.0889	0.0662	0.990
p^1_{23}	0.20	0.1828	0.0708	0.0606	0.965	0.1776	0.0612	0.0472	0.975
p^1_{31}	0.10	0.1144	0.0543	0.0449	0.975	0.1118	0.0482	0.0433	0.965
p^1_{32}	0.30	0.3193	0.0883	0.0754	0.980	0.2629	0.0691	0.0628	0.930
τ_{11}	-3.00	-3.0301	0.1653	0.1594	0.940	-3.0008	0.1520	0.1560	0.940
τ_{12}	1.50	1.4131	0.2801	0.2519	0.955	1.4398	0.2523	0.2250	0.970
τ_{13}	1.50	1.5476	0.2864	0.2461	0.980	1.5169	0.2564	0.2280	0.965
τ_{21}	1.20	1.2002	0.0302	0.0254	0.960	1.2023	0.0269	0.0250	0.970
τ_{22}	0.50	0.4524	0.0963	0.0877	0.930	0.4789	0.0831	0.0708	0.940
τ_{23}	0.50	0.5289	0.0927	0.0818	0.925	0.4996	0.0823	0.0699	0.970
Σ_{11}	5.00	5.0733	0.5312	0.5238	0.960	5.0411	0.5287	0.4962	0.955
Σ_{12}	0.50	0.5067	0.0650	0.0645	0.945	0.5050	0.0644	0.0611	0.965
Σ_{22}	0.10	0.1096	0.0123	0.0124	0.905	0.1090	0.0120	0.0109	0.890
$\frac{1}{\sigma^2_\epsilon}$	10.00	9.9991	0.5526	0.5331	0.930	9.9520	0.5335	0.5364	0.935

where π_k is the initial probability of hidden state k, p^s_{jk} is the j,k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\boldsymbol{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2. S.D. - standard deviation

Table 3.3: Simulation results for 200 runs of Situation (3)

Parameter	True Value	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage
π_1	0.60	0.5683	0.0802	0.0761	0.950
π_2	0.25	0.2716	0.0751	0.0588	0.985
p^0_{12}	0.10	0.1403	0.0879	0.0737	0.960
p^0_{13}	0.05	0.0619	0.0296	0.0247	0.980
p^0_{21}	0.60	0.4926	0.1586	0.1413	0.945
p^0_{23}	0.20	0.1970	0.1138	0.0847	0.985
p^0_{31}	0.10	0.1258	0.0671	0.0613	0.980
p^0_{32}	0.10	0.1609	0.0995	0.0721	0.985
p^1_{12}	0.10	0.1296	0.0643	0.0545	0.965
p^1_{13}	0.05	0.0558	0.0191	0.0173	0.955
p^1_{21}	0.60	0.5652	0.1119	0.0884	0.975
p^1_{23}	0.20	0.1810	0.0713	0.0565	0.970
p^1_{31}	0.10	0.1089	0.0538	0.0410	0.990
p^1_{32}	0.30	0.3163	0.0884	0.0728	0.985
τ_{11}	-3.00	-3.0422	0.1740	0.1758	0.930
τ_{12}	1.50	1.4147	0.2792	0.2725	0.940
τ_{13}	1.50	1.5726	0.2875	0.2587	0.965
τ_{21}	1.20	1.1990	0.0323	0.0300	0.970
τ_{22}	0.50	0.4575	0.0956	0.0970	0.920
τ_{23}	0.50	0.5482	0.0941	0.1644	0.920
Σ_{11}	-	5.8255	0.6166	0.6550	-
Σ_{12}	-	0.6190	0.0772	0.0992	-
Σ_{22}	-	0.1370	0.0147	0.0209	-
$\frac{1}{\sigma^2_\epsilon}$	10.00	10.0380	0.5644	0.5286	0.970

where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2. S.D. - standard deviation

also resolve the biases observed in this situation. Since this situation is a departure from a modeling assumption (the independence of $\boldsymbol{b}_i$ and $\boldsymbol{Z}_i$), it is particularly reassuring to note that the parameters associated with the observed process have very small bias, and the 95% quantile coverage probabilities are generally well-behaved. Further, we may expect this departure to also bias the estimates of π_k, but this is not any more severe than what was observed in situation (2). We would expect a similar reduction of the bias in this situation, as was observed when a more informative prior is used on the transition probability matrix, $\boldsymbol{P}$ as was seen when contrasting situations (2) and (4) in Table 2.

We performed an additional three simulation situations to better understand the bias observed in situations (2) and (3). The following situations were examined:

(5) The data are generated from a MHMM with well-separated modes ($\boldsymbol{\tau}_1 = [-3, 1.5, 1.5]$, $\boldsymbol{\tau}_2 = [1.2, 0.5, 0.5]$) with a large between-subject heterogeneity in the binomial response via independent multivariate normal random effects. The random effects are $i.i.d$ and have a covariance matrix $\boldsymbol{\Sigma} = \left[\begin{smallmatrix} 2.0 & 0.35 \\ 0.35 & 0.1 \end{smallmatrix}\right]$ with the inverse of the residual error, $\frac{1}{\sigma_\epsilon^2}$, in the normally distributed response set at 10.0. The hidden process is generated from a common initial probability vector, $\boldsymbol{\pi} = [0.6, 0.25, 0.15]$, and placebo transition matrix $\boldsymbol{P}^{(0)} = \left[\begin{smallmatrix} 0.85 & 0.10 & 0.05 \\ 0.60 & 0.20 & 0.20 \\ 0.10 & \mathbf{0.10} & \mathbf{0.80} \end{smallmatrix}\right]$ with treatment transition matrix, $\boldsymbol{P}^{(1)} = \left[\begin{smallmatrix} 0.85 & 0.10 & 0.05 \\ 0.60 & 0.20 & 0.20 \\ 0.10 & \mathbf{0.30} & \mathbf{0.60} \end{smallmatrix}\right]$.

(6) Identical to (5) in all respects except the number of observation times is doubled for each patient, such that $n_i = 12$ for all i.

(7) Identical to (5) except, $\boldsymbol{\tau}_1 = [-5, 3, 3]$ and $\boldsymbol{\tau}_2 = [1.2, 0.1, 1.10]$.

The results from these three additional simulations are presented in Tables 3.4 and 3.5. In each case, the default Dirichlet distribution prior with hyperparameters $[1, 1, 1]$ is used for $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and the rows of each of $\boldsymbol{P}^{(0)}$ and $\boldsymbol{P}^{(1)}$, which was a case we encountered some significant bias in our estimates in situations (2) and (3). Situation (5) resembles situation (2) most closely, but has reduced between-subject heterogeneity ($\boldsymbol{\Sigma_{11}}$) on the binomial

response. While the bias does appear to slightly decrease for both p_{21}^0 and p_{32}^0, the major difference is noted when comparing the results to situation (6), where the number of observation times have been doubled, resulting in a quite dramatic decrease in bias. This is reassuring that given a large enough sample size, the bias is virtually non-existent where in other similar situations, it had been significant.

Lastly, situation (7) preserves the random effects covariance structure and hidden state transition probability matrices while increasing the separation of the observed modes in each of the responses by the hidden states, such that we can compare it directly to situation (5). We note significantly less bias in this situation when compared to situation (5) for parameters like p_{21}^0 and p_{32}^0, as well as situation (1), which has the same population modes, but larger between-subject heterogeneity in the observed response.

Throughout situations (1) to (7) we note consistent bias in the parameter $\mathbf{\Sigma_{22}}$, often with lower than expected coverage obtained by the 95% posterior distribution quantiles. This is likely due to the prior choice for $\mathbf{\Sigma_{22}}$ which has a mean which is approximately ten times higher than its true value. We anticipate with a better choice in prior, that this bias and low coverage would be reduced, if not eliminated, but this illustrates a possible sensitivity to prior specification, and some care should be taken when choosing a prior distribution for this and other parameters, especially when the relevant sample size is relatively small for the specific parameter (e.g., p_{21}^0).

3.5 Results from the smoking cessation motivating example

In section 3.2.2 we described a motivating example from a smoking cessation clinical trial (O'Malley et al., 2007) examining a pharmacological intervention's effect on the rates of maintained smoking abstinence over six weeks. When introducing methods to model the treatment effects (section 3.3.4), we enumerated four different ways to incorporate

Table 3.4: Simulation results for 200 runs for each of Situation (5) and (6)

		Situation (5) $n_i = 6$				Situation (6) $n_i = 12$			
Parameter	True Value	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage
π_1	0.60	0.5711	0.0717	0.0780	0.950	0.5975	0.0438	0.0440	0.950
π_2	0.25	0.2692	0.0668	0.0679	0.940	0.2505	0.0439	0.0448	0.950
p^0_{12}	0.10	0.1222	0.0738	0.0718	0.990	0.1048	0.0320	0.0278	0.980
p^0_{13}	0.05	0.0625	0.0264	0.0263	0.970	0.0542	0.0140	0.0142	0.945
p^0_{21}	0.60	0.5105	0.1549	0.1252	0.960	0.5828	0.1003	0.1034	0.945
p^0_{23}	0.20	0.2060	0.1120	0.0870	0.980	0.1977	0.0748	0.0747	0.940
p^0_{31}	0.10	0.1168	0.0590	0.0464	0.990	0.1089	0.0330	0.0323	0.950
p^0_{32}	0.10	0.1554	0.0903	0.0763	0.975	0.1055	0.0413	0.0408	0.950
p^1_{12}	0.10	0.1275	0.0574	0.0660	0.955	0.1059	0.0217	0.0204	0.965
p^1_{13}	0.05	0.0554	0.0175	0.0146	0.975	0.0514	0.0086	0.0083	0.955
p^1_{21}	0.60	0.5735	0.1039	0.0950	0.970	0.5955	0.0571	0.0570	0.955
p^1_{23}	0.20	0.1932	0.0699	0.0582	0.965	0.2017	0.0409	0.0399	0.940
p^1_{31}	0.10	0.1058	0.0493	0.0377	0.975	0.1016	0.0280	0.0251	0.980
p^1_{32}	0.30	0.3253	0.0841	0.0700	0.975	0.3046	0.0436	0.0395	0.965
τ_{11}	-3.00	-3.0137	0.1283	0.1173	0.980	-3.0095	0.0937	0.0963	0.950
τ_{12}	1.50	1.4418	0.2585	0.2332	0.950	1.4799	0.1324	0.1310	0.965
τ_{13}	1.50	1.5315	0.2575	0.2330	0.950	1.5131	0.1329	0.1291	0.940
τ_{21}	1.20	1.1963	0.0302	0.0266	0.970	1.1971	0.0212	0.0205	0.945
τ_{22}	0.50	0.4759	0.0875	0.0826	0.960	0.4913	0.0446	0.0467	0.915
τ_{23}	0.50	0.5146	0.0872	0.0793	0.940	0.5052	0.0455	0.0455	0.930
Σ_{11}	2.00	2.0128	0.2349	0.2300	0.940	2.0278	0.1921	0.1767	0.970
Σ_{12}	0.35	0.3598	0.0461	0.0434	0.955	0.3572	0.0368	0.0346	0.970
Σ_{22}	0.10	0.1103	0.0123	0.0113	0.890	0.1064	0.0095	0.0094	0.900
$\frac{1}{\sigma^2_\epsilon}$	10.00	10.0694	0.5480	0.5867	0.940	10.0003	0.3349	0.3418	0.970

where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2. S.D. - standard deviation

Table 3.5: Simulation results for 200 runs of Situation (7)

Parameter	True Value	Average of Posterior Means	Average Posterior S.D.	S.D. of Posterior Means	95% Posterior Quantile Coverage
π_1	0.60	0.5894	0.0472	0.0473	0.955
π_2	0.25	0.2600	0.0457	0.0461	0.950
p^0_{12}	0.10	0.1161	0.0423	0.0423	0.935
p^0_{13}	0.05	0.0521	0.0170	0.0175	0.940
p^0_{21}	0.60	0.5670	0.1073	0.0979	0.955
p^0_{23}	0.20	0.2095	0.0712	0.0692	0.955
p^0_{31}	0.10	0.1081	0.0433	0.0399	0.970
p^0_{32}	0.10	0.1192	0.0457	0.0424	0.945
p^1_{12}	0.10	0.1090	0.0273	0.0289	0.945
p^1_{13}	0.05	0.0517	0.0100	0.0104	0.940
p^1_{21}	0.60	0.5956	0.0609	0.0594	0.965
p^1_{23}	0.20	0.1960	0.0376	0.0365	0.960
p^1_{31}	0.10	0.1007	0.0375	0.0341	0.955
p^1_{32}	0.30	0.3051	0.0453	0.0393	0.970
τ_{11}	-5.00	-5.0249	0.2266	0.2293	0.970
τ_{12}	3.00	2.9754	0.1938	0.1845	0.965
τ_{13}	3.00	3.0420	0.1798	0.1682	0.965
τ_{21}	1.20	1.2014	0.0213	0.0209	0.965
τ_{22}	0.10	0.1926	0.0374	0.0365	0.960
τ_{23}	1.10	1.1014	0.0353	0.0380	0.940
Σ_{11}	2.00	2.0139	0.2700	0.2692	0.930
Σ_{12}	0.35	0.3436	0.0422	0.0424	0.960
Σ_{22}	0.10	0.1069	0.0098	0.0091	0.915
$\frac{1}{\sigma^2_\epsilon}$	10.00	10.0168	0.3755	0.3771	0.950

where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2. S.D. - standard deviation

treatment effects for the hidden states. We fit each of these models as outlined in section 3.3 with a varying number of parameters (18-26 parameters), and present the results in Table 3.6.

We generally note very little effect of treatment with the exception of perhaps a difference between p_{32}^0 and p_{32}^1. This quantity represents the transition probability of individuals transition from the smoking state to the intermediate withdrawal state. Ideally, this would be high, suggesting the readiness of individuals to stop smoking, even if they had restarted during the study period. Surprisingly however, the rate is higher in the placebo group, suggesting that those on treatment actually do worse than those on placebo. From our MCMC samples, we can estimate quantities like, $P(p_{32}^1 > p_{32}^0)$, which is estimated as 0.072 in model III and 0.067 in model IV. This provides some evidence that the treatment may inhibit abstinence in those subjects who remained or returned to smoking after the first week of the study period. Without formally performing model selection or hypothesis testing on the two different treatment effects, this single effect remains somewhat unconvincing.

Using model I, we can estimate the transformed means of the CO levels for each of the hidden states by using $\boldsymbol{\tau_2}$ and the chosen structure of $\boldsymbol{H}_{irtk}$. In our case, $\boldsymbol{H}_{irtk}$ corresponds to the cumulative sum of the elements of $\boldsymbol{\tau_2}$, resulting in estimates of the means of 1.18, 1.53 and 2.63 for the first, second and third hidden states, respectively. By inverting the $\log(Y + 1)$ transformation, the estimates on the untransformed scale are 2.26, 3.63 and 12.95, for hidden states one, two and three, respectively. In most studies a CO threshold of 8 or sometimes 10 ppm is used as evidence that a subject is smoking, suggesting our results provide some face validity. Similarly, for the daily self-reported binomial response, an individual with this response's corresponding random effect equal to zero would have an estimated probability of smoking on a single day of 0.005, 0.22 and 0.66 for the first, second and third hidden states, respectively.

Table 3.6: Results for Models I-IV from the Motivating Smoking Cessation Example

Parameter	Model I	Model II	Model III	Model IV
π_1^0	0.76 [0.76, 0.039, (0.68,0.83)]	0.76 [0.77, 0.066, (0.62,0.88)]	0.75 [0.75, 0.042, (0.66,0.82)]	0.75 [0.76, 0.073, (0.59,0.88)]
π_2^0	0.18 [0.17, 0.037, (0.11,0.26)]	0.18 [0.17, 0.063, (0.07,0.32)]	0.19 [0.18, 0.040, (0.12,0.27)]	0.19 [0.19, 0.071, (0.07,0.35)]
π_1^1	-	0.75 [0.75, 0.043, (0.66,0.82)]	-	0.74 [0.75, 0.045, (0.65,0.82)]
π_2^1	-	0.18 [0.18, 0.041, (0.11,0.27)]	-	0.19 [0.18, 0.044, (0.11,0.28)]
p_{12}^0	0.035 [0.035, 0.0091, (0.02,0.06)]	0.035 [0.035, 0.009, (0.02,0.056)]	0.024 [0.021, 0.014, (0.004,0.06)]	0.024 [0.022, 0.015, (0.004,0.06)]
p_{13}^0	0.006 [0.006, 0.0031, (0.001,0.013)]	0.006 [0.0051, 0.003, (0.001,0.013)]	0.013 [0.012, 0.008, (0.003,0.03)]	0.013 [0.011, 0.008, (0.002,0.03)
p_{21}^0	0.26 [0.26, 0.056, (0.16,0.38)]	0.26 [0.26, 0.055, (0.16,0.38)]	0.24 [0.23, 0.091, (0.09,0.44)]	0.27 [0.26, 0.096, (0.10, 0.48)]
p_{23}^0	0.20 [0.20, 0.048, (0.12,0.31)]	0.20 [0.20, 0.049, (0.12,0.31)]	0.18 [0.17, 0.075, (0.06,0.35)]	0.18 [0.17, 0.078, (0.06,0.36)]
p_{31}^0	0.055 [0.051, 0.029, (0.011,0.12)]	0.054 [0.050, 0.028, (0.01,0.12)]	0.063 [0.047, 0.059, (0.002,0.22)]	0.066 [0.048, 0.060, (0.002,0.23)]
p_{32}^0	0.18 [0.17, 0.056, (0.08,0.30)]	0.17 [0.17, 0.056, (0.08,0.29)]	**0.30** [0.29, 0.12, (0.10,0.56)]	**0.30** [0.29, 0.12, (0.10, 0.55)]
p_{12}^1	-	-	0.043 [0.041, 0.011, (0.02, 0.07)]	0.043 [0.042, 0.011, (0.02, 0.07)]
p_{13}^1	-	-	0.005 [0.004, 0.003, (0.0003,0.01)]	0.004 [0.004, 0.003, (0.0002,0.01)]
p_{21}^1	-	-	0.27 [0.26, 0.065, (0.15,0.40)]	0.27 [0.27, 0.065, (0.16, 0.41)]
p_{23}^1	-	-	0.21 [0.20, 0.057, (0.11,0.33)]	0.20 [0.20, 0.058, (0.11,0.33)]
p_{31}^1	-	-	0.076 [0.070, 0.037, (0.02,0.16)]	0.075 [0.070, 0.037, (0.02, 0.16)]
p_{32}^1	-	-	**0.12** [0.12, 0.060, (0.02,0.25)]	**0.12** [0.11, 0.060, (0.02, 0.25)]
τ_{11}	-5.30 [-5.29, 0.27, (-5.87,-4.80)]	-5.31 [-5.30, 0.26, (-5.85,-4.83)]	-5.35 [-5.34, 0.28, (-5.93, -4.83)]	-5.35 [-5.34, 0.27, (-5.92,-4.85)]
τ_{12}	4.05 [4.04, 0.29, (3.50,4.64)]	4.03 [4.03, 0.29, (3.46,4.59)]	4.00 [4.00, 0.29, (3.45, 4.57)]	3.95 [3.95, 0.29, (3.39, 4.53)]
τ_{13}	1.92 [1.93, 0.40, (1.11,2.68)]	1.95 [1.96, 0.39, (1.16,2.70)]	2.01 [2.01, 0.40, (1.23,2.80)]	2.11 [2.12, 0.39, (1.33,2.87)]
τ_{21}	1.18 [1.18, 0.023, (1.14,1.23)]	1.18 [1.18, 0.023, (1.14,1.23)]	1.19 [1.19, 0.023, (1.14,1.23)]	1.19 [1.19, 0.024, (1.14, 1.24)]
τ_{22}	0.35 [0.35, 0.081, (0.20,0.52)]	0.33 [0.33, 0.079, (0.18,0.49)]	0.31 [0.31, 0.079, (0.16,0.47)]	0.30 [0.30, 0.078, (0.15,0.45)]
τ_{23}	1.10 [1.10, 0.061, (0.99,1.23)]	1.10 [1.10, 0.059, (0.99,1.22)]	1.11 [1.11, 0.062, (0.99,1.23)]	1.12 [1.12, 0.063, (1.00,1.25)]
Σ_{11}	6.91 [6.81, 1.14, (4.95,9.38)]	6.83 [6.73, 1.15, (4.93,9.40)]	6.78 [6.69, 1.14, (4.82, 9.30)]	6.94 [6.84, 1.15, (4.95, 9.46)]
Σ_{21}	0.32 [0.31, 0.077, (0.17,0.48)]	0.32 [0.31, 0.078, (0.17,0.48)]	0.32 [0.32, 0.080, (0.17,0.49)]	0.33 [0.33, 0.080, (0.19,0.50)]
Σ_{22}	0.099 [0.098, 0.011, (0.08,0.12)]	0.099 [0.099, 0.011, (0.08,0.12)]	0.10 [0.10, 0.012, (0.08,0.12)]	0.10 [0.10, 0.012, (0.08,0.13)]
$\frac{1}{\sigma_\epsilon^2}$	10.47 [10.45, 0.52, (9.50,11.54)]	10.44 [10.42, 0.52, (9.48,11.53)]	9.99 [9.07, 0.49, (9.00,10.99)]	9.91 [9.91, 0.49, (9.00,10.92)]

The displayed table are Posterior Mean [(Posterior Median, Posterior standard deviation, (0.025 quantile, 0.975 quantile)], where π_k^s is the initial probability of hidden state k for the sth treatment group, p_{jk}^s is the j,k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ_ϵ^2 is the residual error associated with response 2.

3.6 Discussion

Overall, we have outlined a general framework for modeling multivariate longitudinal data in the setting where heterogeneity in the longitudinal responses is explained by random subject-specific differences or heterogeneity introduced from a given subject transitioning through multiple disease states. We believe this could have a great number of applications in the modeling of longitudinal data arising from the study of addiction, where the underlying disease states are conceptually understood to contribute significantly to the underlying heterogeneity of measured responses, but ultimately are difficult to measure directly. Modeling a multivariate longitudinal response through a MHMM has a degree of flexibility in accounting for the correlation and heterogeneity between- and within-responses by the incorporation of the hidden process along with the random effects, that is not present by using only one of these components in isolation. Others have applied HMMs to the setting of addiction (DeSantis and Bandyopadhyay, 2011; Shirley et al., 2010), as well as other diseases where a similar argument can be made pertaining to the underlying disease states, such as multiple sclerosis (Altman, 2007) and schizophrenia (Scott, James and Sugar, 2005). While the development of these methods are in their infancy, we believe they can play an important role in understanding how a treatment works on true clinical outcomes through understanding how a treatment may affect the modeled hidden states. Further, these types of models may help to better plan future clinical studies, by, for example, providing better estimates of sample size requirements (Altman, 2012).

The methods as presented in this paper have several limitations which can be addressed in future work. First, in univariate responses, Altman (2007) allows for between-subject heterogeneity in the hidden state process in addition to the observed process. We have not considered this case, mainly due to the limited number of observation times (six) in the motivating smoking cessation study. Scott, James and Sugar (2005) take a Bayesian approach allowing for subject-specific transition matrices under a hierarchal model, and we believe a similar approach could be adapted to models presented in this paper, if

the dataset permitted it. In other situations, such as when there is a large number of observation times, a small hidden state space, little between-subject heterogeneity or well-separated modes in the observed process, or specific types of hidden process transition probability matrices, it may be possible to consider the inclusion of random effects in the hidden process. As we saw during the simulation study in situation (2), one must be aware of the limitations of such models, when so many forms of heterogeneity are being accounted for simultaneously. If a homogeneous common transition matrix cannot be applied to a specific dataset, it may be more worthwhile looking at somewhat simpler alternatives like increasing the order of the hidden process, which may be more feasible than using subject-specific transition matrices in certain situations.

Secondly, we were limited in our ability to perform formal hypothesis testing, or other model selection techniques. Bayes factors and their analogues are usually the basis of Bayesian hypothesis testing, but are generally reliant upon estimating the integrated likelihood of the models in question. The generally poor convergence of estimators such as the harmonic mean commonly used to estimate the integrated likelihood, along with the complex nature of our likelihood, make hypothesis testing reliant on the integrated likelihood difficult. Others have proposed information criteria, such as the deviance information criteria (DIC), which have proven difficult to use in models using missing data approaches, such as mixture and random effects models (Celeux et al., 2006). In our case, not only do we utilize random effects, our model is a special case of mixture model (a hidden Markov model). Hypothesis testing and model selection are crucial to wide use of these models, and is an area of future work.

In our motivating example using a smoking cessation study, we note a possible difference between the hidden state 3-2 transition probabilities in the placebo and treatment groups. This was a somewhat surprising result, as we would normally expect the treatment group to have a beneficial effect, when here it appears it may not. Given the relatively good performance of our method in the simulation study in this paper, which are fairly

representative of the data generated from the motivating study, we are confident this noted difference is not due to a computational issue with our methodology. It is quite possible that the difference estimated in models III and IV for p_{32}^1 and p_{32}^0 is due to some unmodeled missing data mechanism, and is an area we are currently investigating. Also, we note that in O'Malley et al. (2007) a treatment effect on smoking cessation for the entire study sample was not found.

Lastly, although maximum likelihood methods could be used for fitting models such as the ones presented in this paper, we have found, in general, that a Bayesian approach using MCMC is vastly more computationally efficient. We were able to use such methods, where, for example, a two hidden state ($M = 2$) model was fit by approximating the integration in equation (1) by Gauss-Hermite quadrature, and maximizing the approximated likelihood. This required at least 140 quadrature points, and took several days (>3) using a state of the art computer. Alternatively, the models presented in Table 4, some of which have over twice the number of parameters as the model fit by maximum likelihood, took approximately 9 hours to compute on the same computer.

This improvement in computational efficiency, along with the flexibility and general applicability of such models, provides a great opportunity for future work and applications to other disease areas and problems.

3.7 Appendix: Markov Chain Monte Carlo Convergence Diagnostic Plots

3.7.1 Trace and Density Estimation Plots

In order to assess stationarity and mixing of the MCMC sampler, we generated trace and density estimate plots for each of the 24 main parameters ($\mathbf{\Theta}$) from a randomly selected run of our simulation runs in section 3.4 from situation (2). These plots were generated

using the *coda* package in R and are presented in Figures 3.2-3.7.

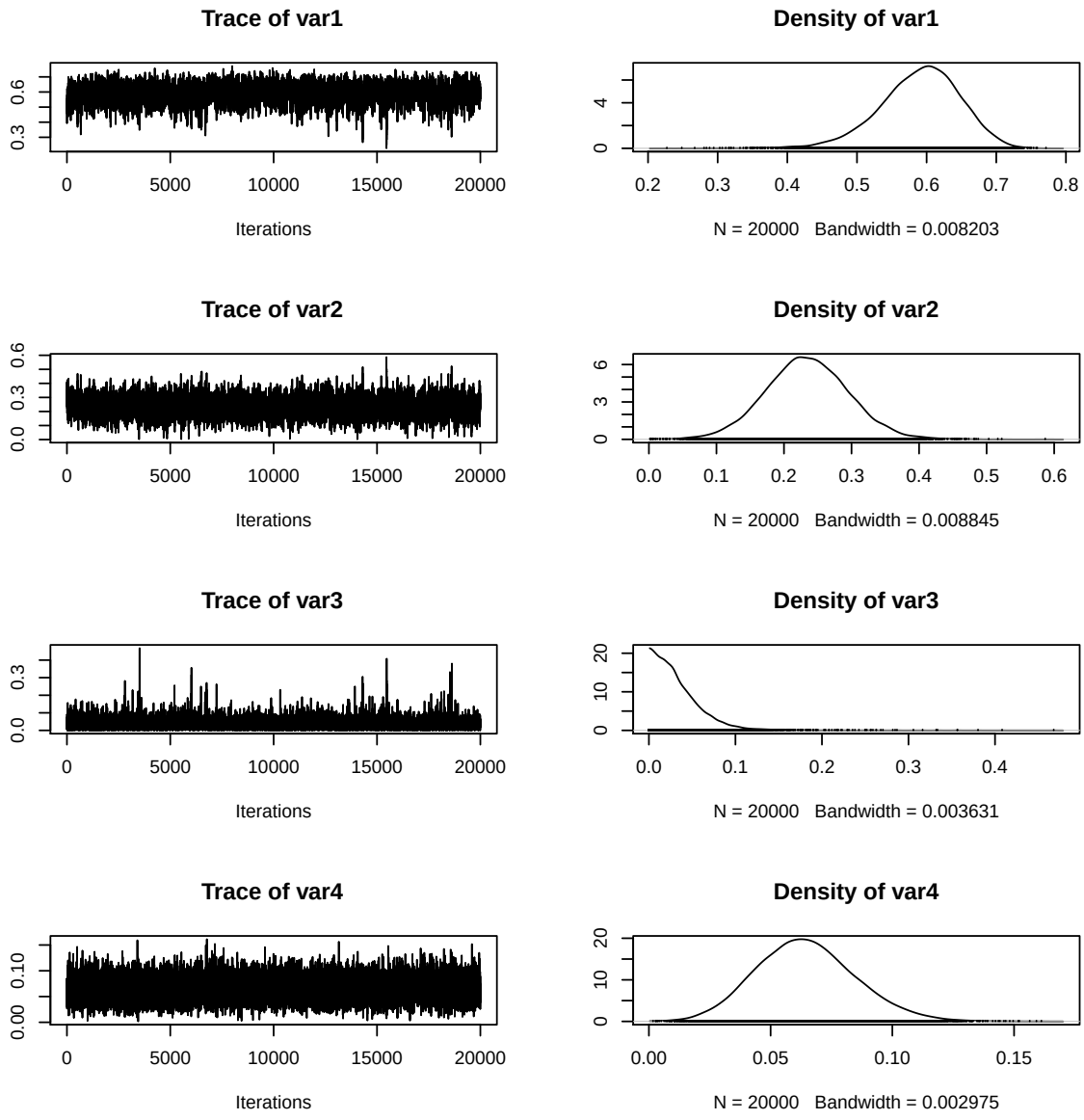

Figure 3.2: Trace and density estimate plots for parameters (from top to bottom): π_1 (var1), π_2 (var2), p_{12}^0 (var3) and p_{13}^0 (var4).

Overall, there does not appear to be any strong indication that the Markov Chain has

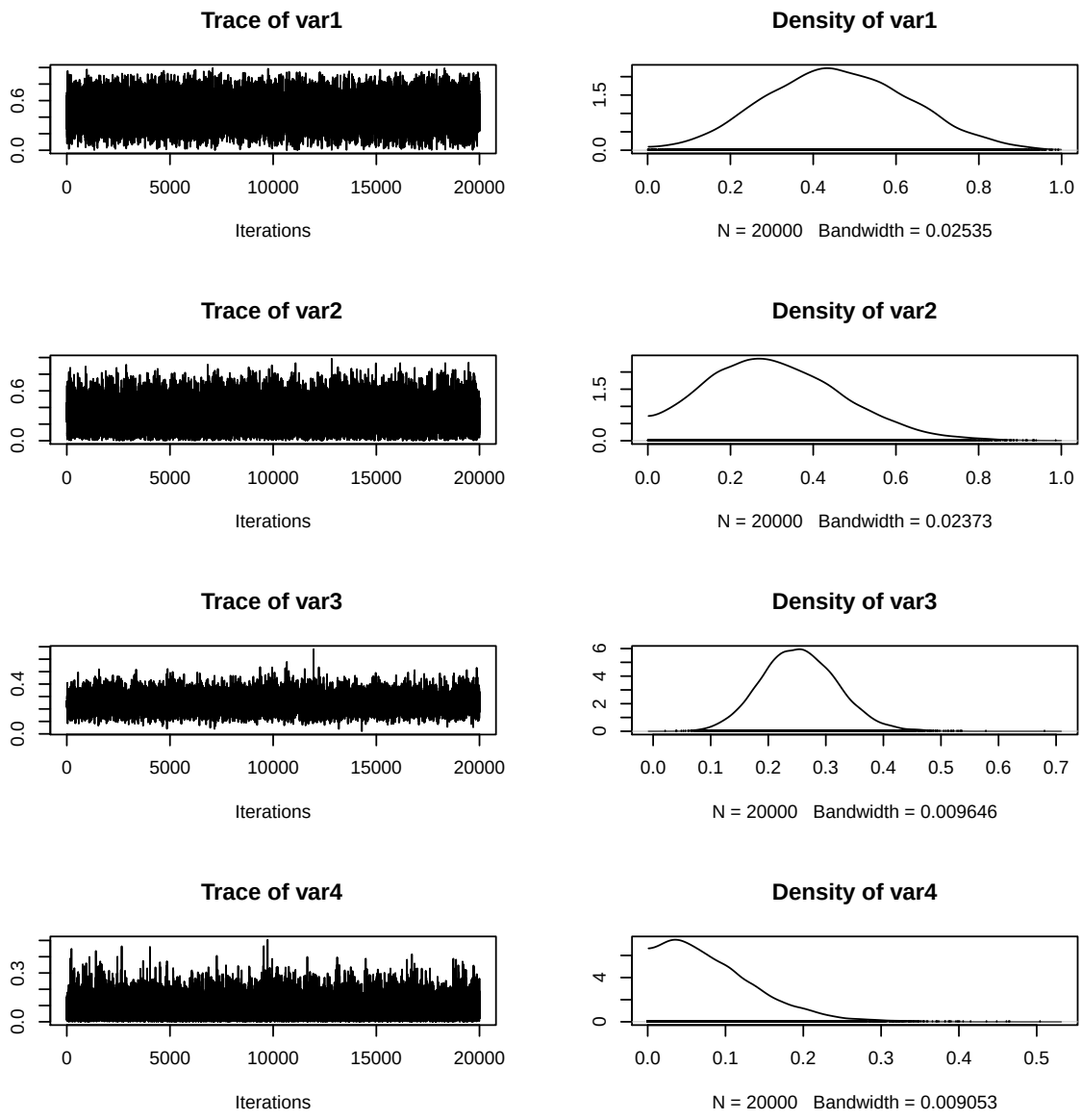

Figure 3.3: Trace and density estimate plots for parameters (from top to bottom): p^0_{21} (var1), p^0_{23} (var2), p^0_{31} (var3) and p^0_{32} (var4).

not reach stationarity. The sampling procedure seems to mix reasonably well. Further, we do not note any multi-modality or phenomena consistent with label switching, although

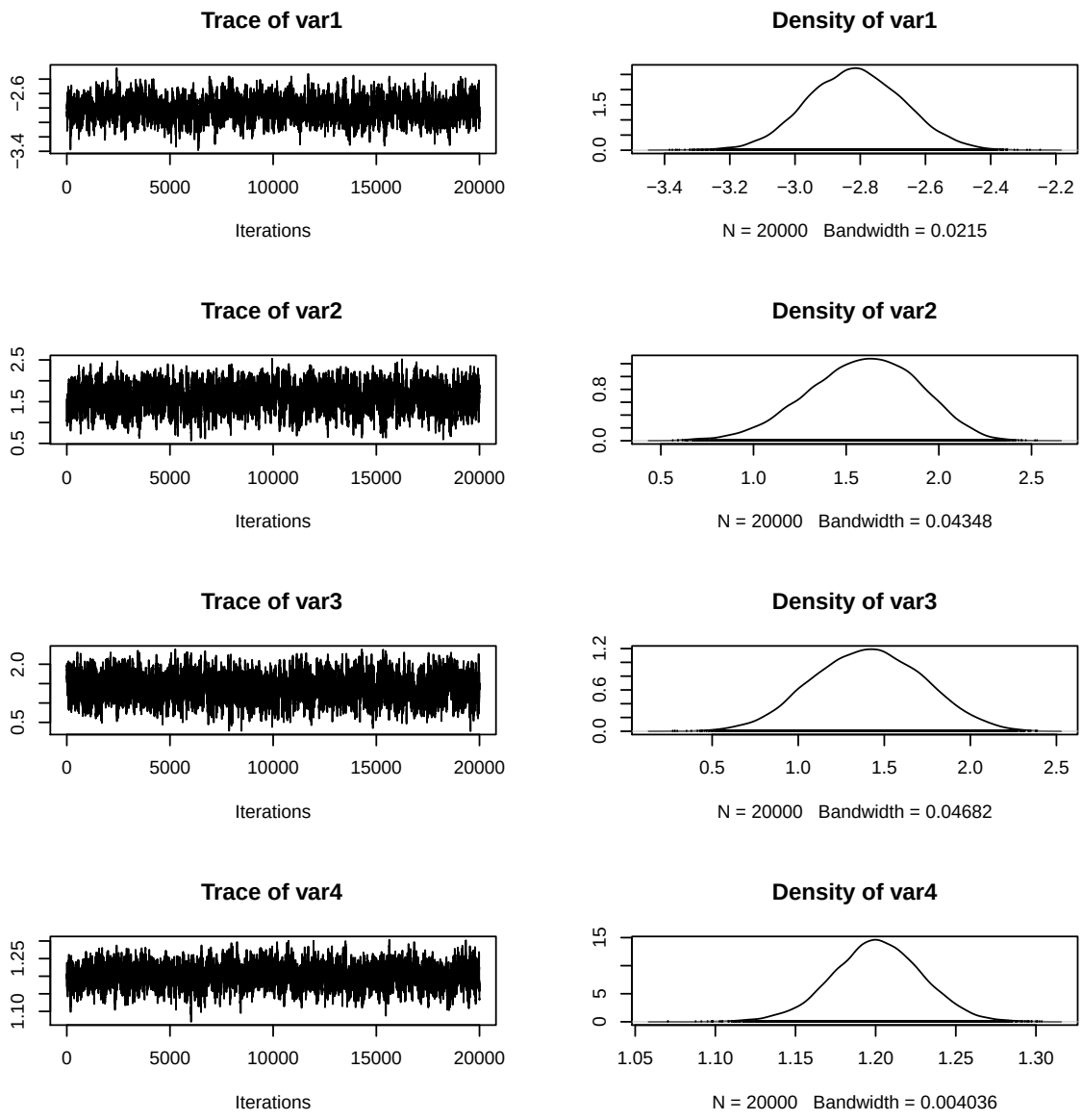

Figure 3.4: Trace and density estimate plots for parameters (from top to bottom): τ_{11} (var1), τ_{12} (var2), τ_{13} (var3) and τ_{21} (var4).

in this same situation in other simulated runs (from different simulated datasets), label switching did occur.

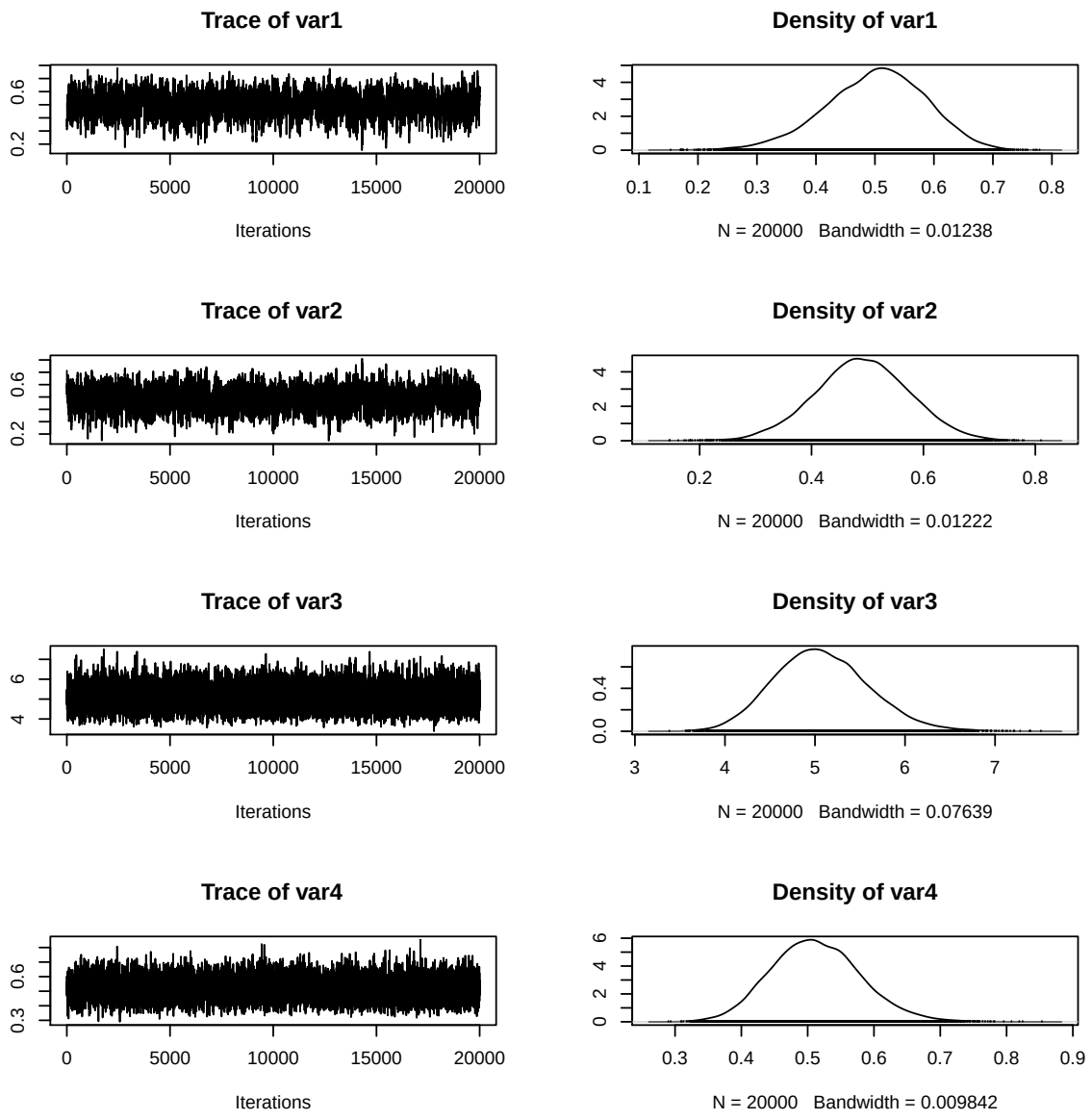

Figure 3.5: Trace and density estimate plots for parameters (from top to bottom): τ_{22} (var1), τ_{23} (var2), Σ_{11} (var3) and Σ_{21} (var4).

3.7.2 Auto-correlation function plots

Using the same MCMC sample as done in section 3.7.1, we also generated estimates for the auto-correlation function within each posterior sample to assess mixing.

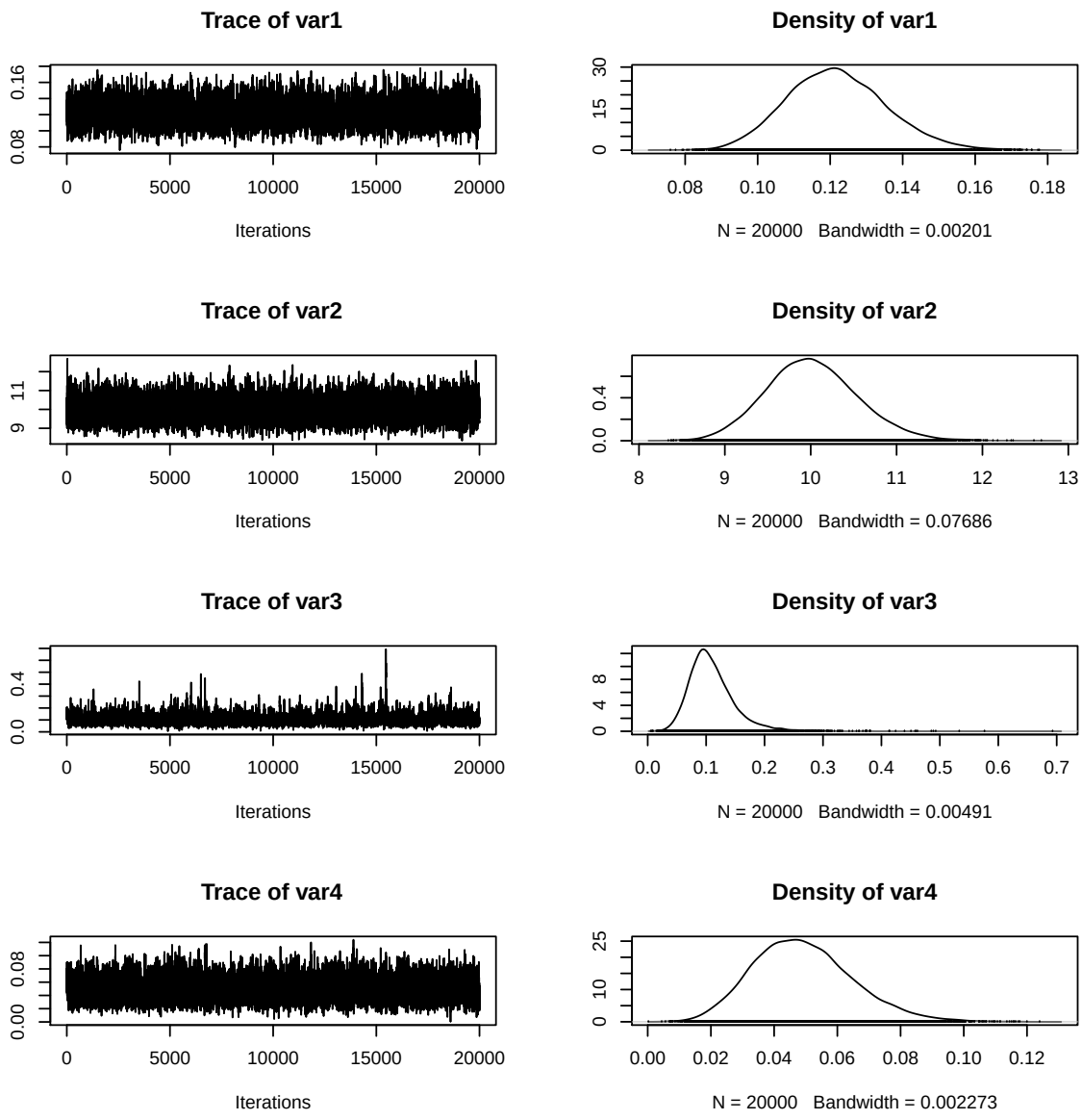

Figure 3.6: Trace and density estimate plots for parameters (from top to bottom): Σ_{22} (var1), $\frac{1}{\sigma_\epsilon^2}$ (var2), p_{12}^1 (var3) and p_{13}^1 (var4) .

Overall, the estimated auto-correlation function decreases with lag significantly, and the auto-correlation within a variable is quite negligible by lag 5 in many such parameters.

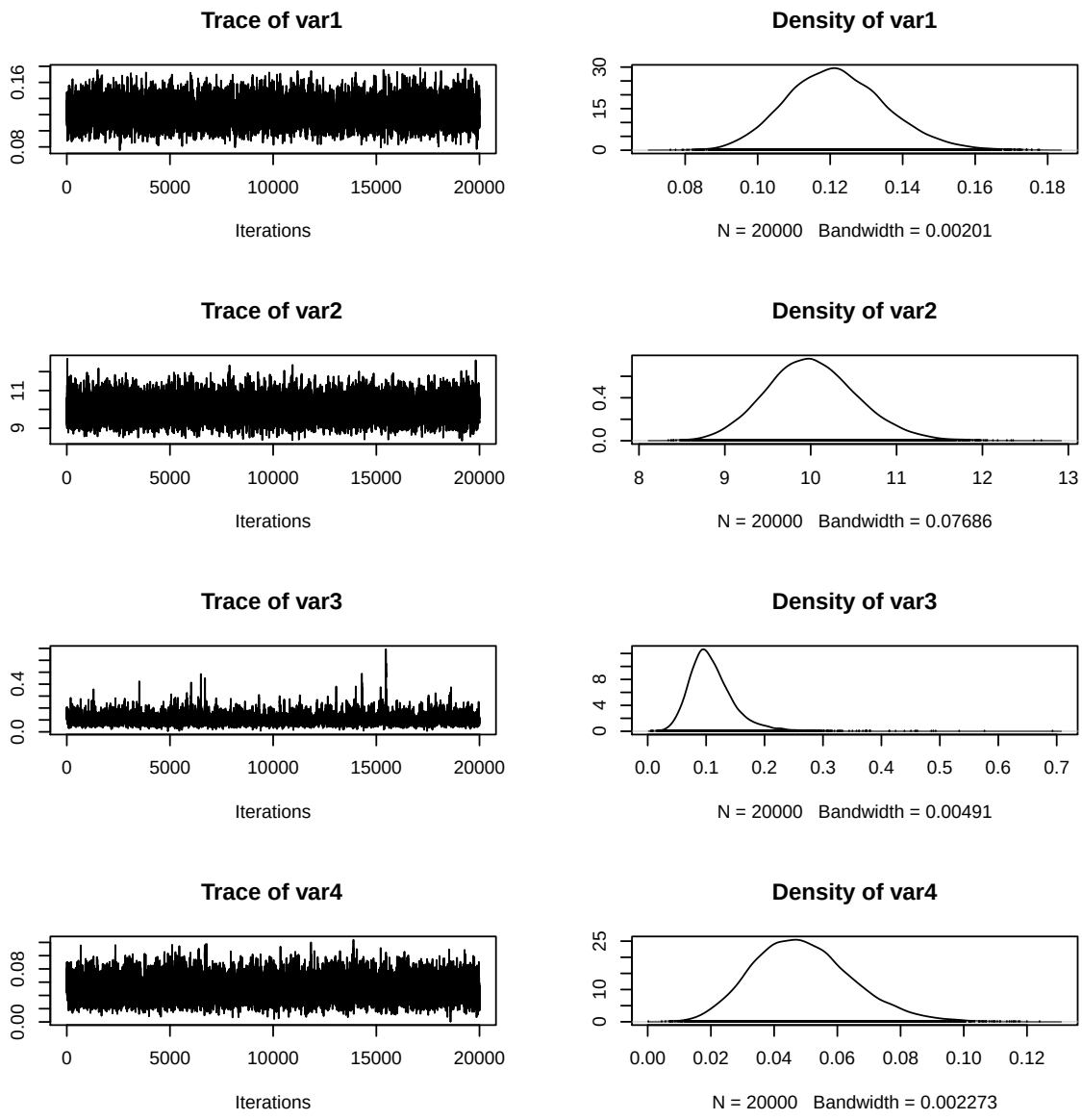

Figure 3.7: Trace and density estimate plots for parameters (from top to bottom): p^1_{21} (var1), p^1_{23} (var2), p^1_{31} (var3) and p^1_{32} (var4).

The most persistent auto-correlations are generally observed in components of $\boldsymbol{\tau}$, which is not surprising, as the full conditional distributions of these parameters are dependent on

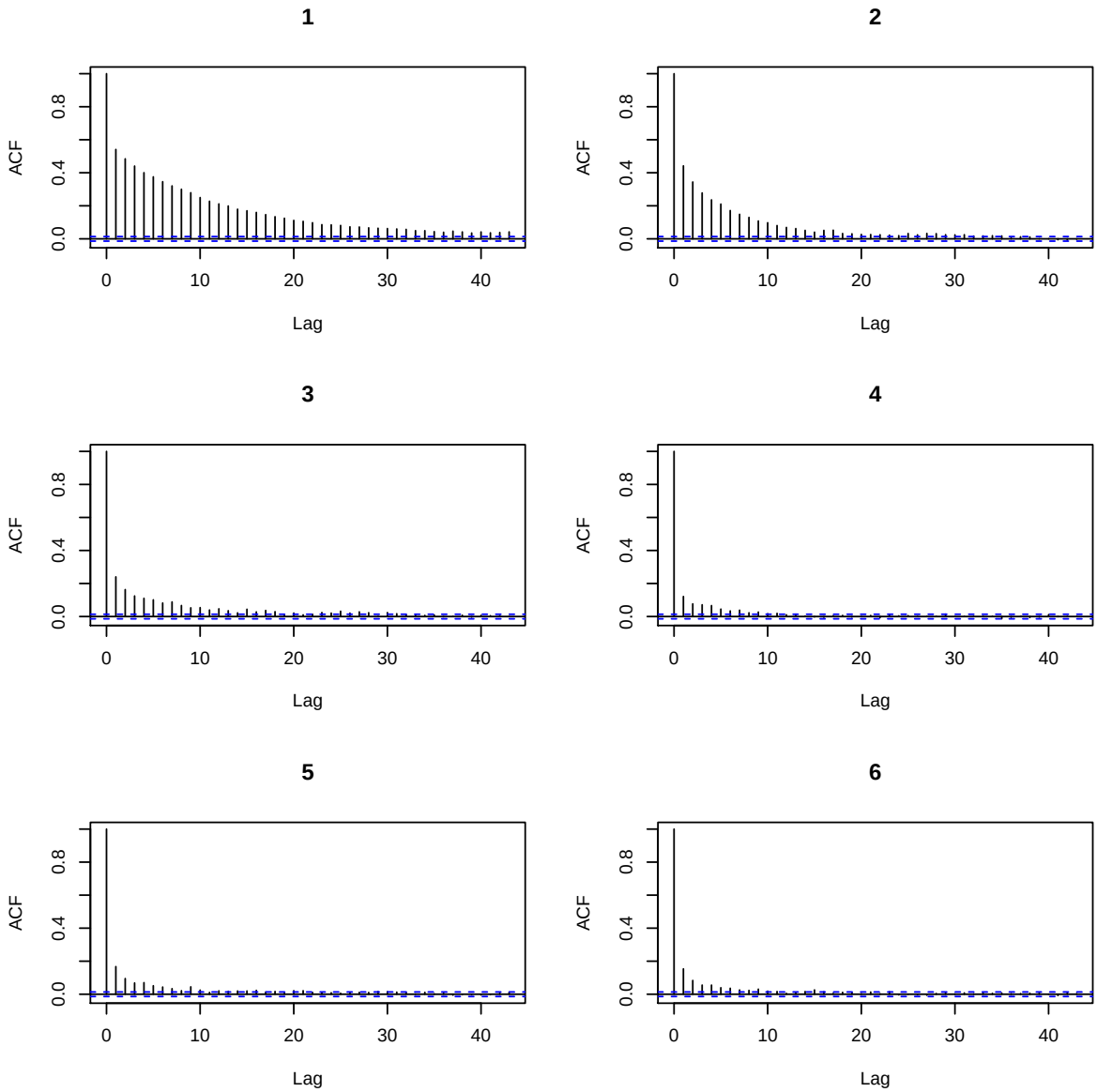

Figure 3.8: Auto-correlation function plots for the MCMC sampling approach for parameters: π_1 (1), π_2 (2), p^0_{12} (3), p^0_{13} (4), p^0_{21} (5) and p^0_{23} (6).

both the random effects $\boldsymbol{b}_i$, and the hidden states, $\boldsymbol{Z_i}$.

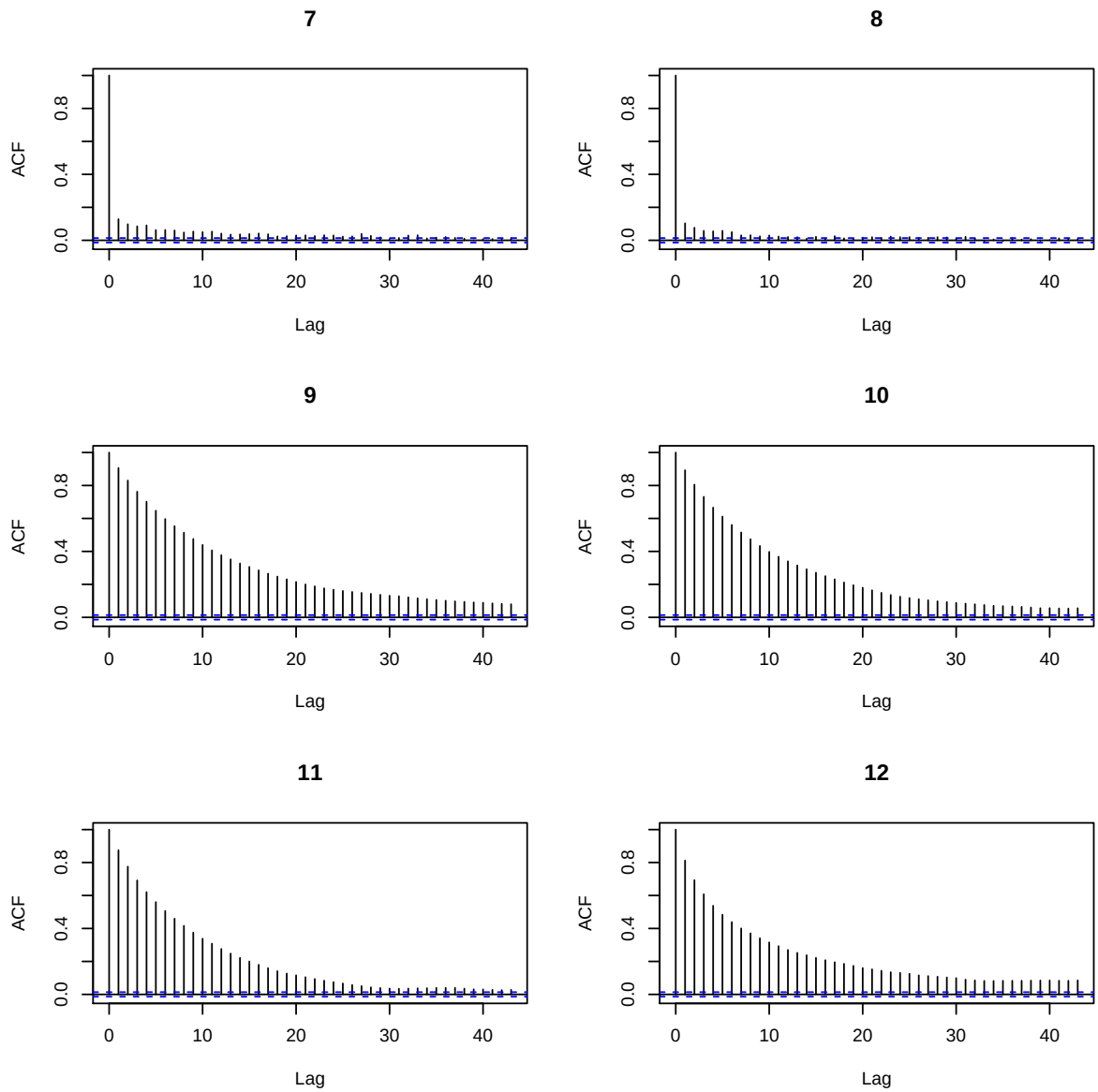

Figure 3.9: Auto-correlation function plots for the MCMC sampling approach for parameters: p^0_{31} (7), p^0_{32} (8), τ_{11} (9), τ_{12} (10), τ_{13} (11) and τ_{21} (12).

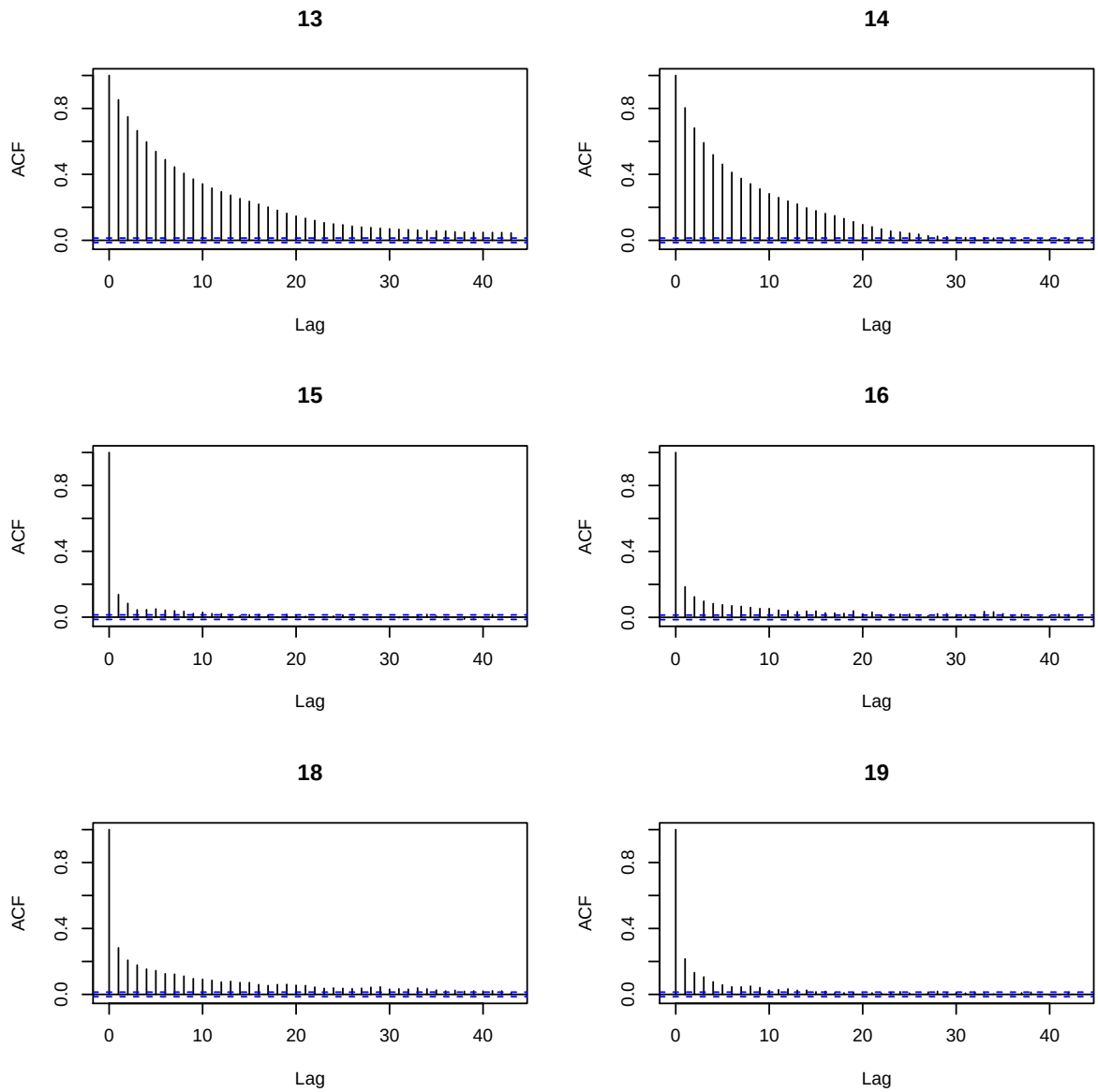

Figure 3.10: Auto-correlation function plots for the MCMC sampling approach for parameters: τ_{22} (13), τ_{23} (14), Σ_{11} (15), Σ_{21} (16), Σ_{22} (18) and $\frac{1}{\sigma_\epsilon^2}$ (19).

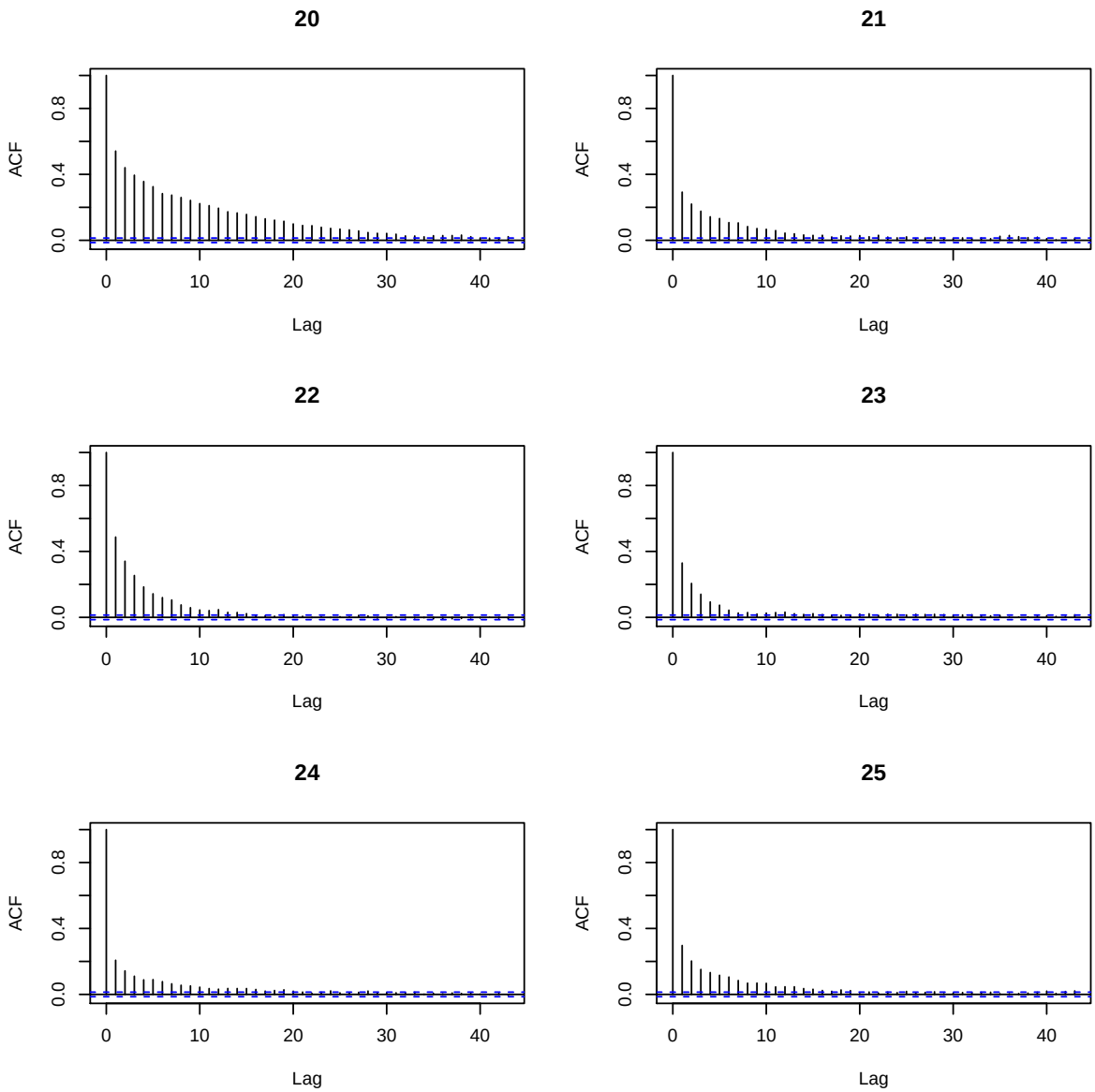

Figure 3.11: Auto-correlation function plots for the MCMC sampling approach for parameters: p^1_{12} (20), p^1_{13} (21), p^1_{21} (22), p^1_{23} (23), p^1_{31} (24) and p^1_{32} (25).

Chapter 4

Using Multivariate MHMMs for Classification of the Disease States

4.1 Background

In Chapter 3, we outlined a modeling framework that extended the work of others (Altman, 2007) from modeling univariate longitudinal data as a hidden Markov model (HMM) with between-subject heterogeneity in the observed process to the setting of multivariate longitudinal data. Thus far, we have focused on drawing inference from the population-level parameters governing these models. Under a MHMM, these parameters can be much more informative than those derived from models which ignore either the dynamic hidden states, or the between-subject heterogeneity. Indeed, much of the computational complexity of these models is generated due to the observation- or subject-specific differences generated from the hidden Markov process, or the random effects. In this chapter we will examine if the hidden states can be used effectively to draw inference to meaningful disease states. Further we will examine the utility of a multivariate model over the possible univariate alternatives.

4.2 Motivation

As is the case in Chapter 3, we will return to our motivating example of a smoking cessation clinical trial, which evaluated a pharmacological intervention to assist people in abstaining from smoking. The results in that chapter generally corresponded with the clinical information used to design such clinical trials. In particular, we found that the population modes of the observed longitudinal response corresponding to their respective underlying (but still at that point hypothetical) hidden disease states were quite well separated. Reassuringly, under model (I) in Table 3.6 in that chapter, for an individual with the relevant random effect set at zero, the three modeled hidden disease states correspond with probabilities of smoking on a given day of 0.005, 0.22 and 0.66, respectively. These roughly correspond with barely smoking at all, smoking one or two days a week, and smoking about five days a week, respectively. Similarly, the carbon monoxide (CO) population means when untransformed by inverting the ($\log[Y+1]$) transformation, correspond with CO levels of 2.26, 3.63 and 12.95 ppm for hidden states, one, two and three, respectively.

It is important to contrast these estimates with thresholds used in the actual study to define the study's two disease states – abstinence and smoking. While the study time period where these definitions are applicable can change, in general, *any* self-reported smoking or a CO observation exceeding ten ppm (or, in some studies, that threshold is set lower, at eight ppm) defines smoking, regardless of any other possible mitigating factors. Under our model, the population modes of the lowest smoking (hidden state 1) and the highest smoking (hidden state 3) states have a close correspondence with the study definitions. Ignoring between-subject heterogeneity of these observed responses, the rates of misclassification should be minimal between these hidden and study disease states. On the other hand, the intermediate hidden state (state 2) has population-level means corresponding with smoking in the self-reported count, and abstinence in the CO response. Some further justification is required before further interpreting inference drawn from the hidden states to those of the diseases states typically found in smoking cessation studies.

This would include how well each hidden state corresponds with some known evaluable definition of disease state. Further, discrepancies may suggest other adjustments to the modeling approach or potential changes to disease state definitions.

Additionally, while the benefits of modeling multivariate longitudinal data are well described, it is unclear under the modeling framework we described in Chapter 3 if improvements are seen by including additional responses. We will begin with this element, and examine if a multivariate model following a MHMM with separate but correlated random effects for each response can be justified over the univariate response alternatives.

4.3 Model Considerations

4.3.1 Methodology

In this chapter we will continue to work with one of the examples from the motivating example in Chapter 3. In Chapter 3, we initially fit using maximum likelihood estimation (via a Gauss-Hermite approximation to the marginal likelihood) a model which considered a bivariate longitudinal response involving cigarette smoking counts and a continuous transformed ($\log[Y + 1]$) longitudinal CO levels. Under the Gauss-Hermite approximation, we were able to fit a two hidden-state Markov model, with separate but correlated zero-mean multivariate normal random effects in the observed longitudinal process by extending the work of others in the univariate settings (Altman, 2007). Increasing the hidden state space would have also significantly increased the computing time, as not only would the computation of the forward probabilities be more computationally complex, approximating the gradient or Hessian (when required, for instance, in computing approximate standard errors for the estimates) would require significantly more evaluations of the likelihood.

We will return to using the cigarette count response in this chapter as further illustration on how the Bayesian methodology in Chapter 3 can be expanded to handle several different types of responses. Further, we believe that in many situations the cigarette counts

will contain more information than using the number of days in the previous week a person smoked, the latter approach used in Chapter 3, where focus was more on demonstrating the methodology as compared to the strategy in this chapter of a more comprehensive consideration of modeling the smoking cessation data.

As was the case in Chapter 3, let Z_{it} denote the hidden process for subject $i, (i = 1, 2, \ldots, N)$ at some discrete time points $t, (t = 1, 2, \ldots, n_i)$. Further, denote $\boldsymbol{Z_i}$ to comprise all values of Z_{it} for subject i, and $\boldsymbol{Z}$ to be all values of $\boldsymbol{Z_i}$ across all i. Assume Z_{it} arises from an $M-$state Markov chain with transition matrix $\boldsymbol{P}$ and initial probability vector, $\boldsymbol{\pi}$. Again, we will primarily focus on HMMs arising from first-order time homogeneous Markov chain hidden processes, where $P(Z_{it} = k|Z_{i,t-1} = j) = p_{jk}$ for all $t > 1$, and $P(Z_{i1} = k) = \pi_k$, where $j, k = 1, \ldots, M$. As was the case in Chapter 3, we will consider models with a hidden state space of three ($M = 3$).

Let $\boldsymbol{y}_{it}$ denote a bivariate longitudinal response for subject i at time t, such that $\boldsymbol{y}_{it} = [y_{i1t}, y_{i2t}]'$, where the first response is a count of cigarettes in week t over a a_{it} day period, and as was the case in Chapter 3, the second response is a CO level reading. Similar to above, let $\boldsymbol{y_i}$ denote all the observations for subject i across all longitudinal responses and time points, and $\boldsymbol{y}$ be the collection of all $\boldsymbol{y_i}$ for $i = 1, 2, \ldots, N$.

We will follow the same framework as Chapter 3, but apply the new weekly cigarette count response (Y_{i1t}) as a Poisson count with logarithmic link function in place of the number of days smoked response such that:

$$Y_{i1t}|Z_{it}, \boldsymbol{b}_i \sim Poisson(\theta_{i1t}) \text{ where,} \tag{4.1}$$

$$\log(\theta_{i1t}) = \boldsymbol{H}'_{i1tk}\boldsymbol{\tau}_1 + \boldsymbol{x}'_{i1t}\boldsymbol{\beta}_{1k} + \boldsymbol{w}'_{i1tk}\boldsymbol{b}_i + \log(a_{it}) \text{ and,}$$

$$Y_{i2t}|Z_{it}, \boldsymbol{b}_i \sim N(\theta_{i2t}, \sigma^2_\epsilon) \text{ where,} \tag{4.2}$$

$$\theta_{i2t} = \boldsymbol{H}'_{i2tk}\boldsymbol{\tau}_2 + \boldsymbol{x}'_{i2t}\boldsymbol{\beta}_{2k} + \boldsymbol{w}'_{i2tk}\boldsymbol{b}_i \ .$$

As was the case in Chapter 3, Y_{i2t} is the transformed ($\log[Y+1]$) CO-levels for subject i and time t, and $\boldsymbol{\tau}_r$ represents the fixed effect intercept term for each hidden state, as defined through an M-dimensional hidden state vector, $\boldsymbol{H}_{irtk}$ (which is dependent on Z_{it}), $\boldsymbol{x_{irt}}$ is a vector of covariates at time t for subject i when $Z_{it} = k$ for the rth longitudinal response. The associated fixed effect coefficient parameters for the covariates are represented by $\boldsymbol{\beta}_{rk}$, which also could vary over different values of the hidden process. We will again repeat a similar set of assumptions, where $\boldsymbol{b}_i \overset{\text{iid}}{\sim} MVN(0, \boldsymbol{\Sigma})$, and will use separate, but correlated random effects, where $\boldsymbol{w}_{i1tk} = [1, 0]$ and $\boldsymbol{w}_{i2tk} = [0, 1]$ for all i, k and t. Here, an offset term of $\log(a_{it})$ is used to account for the slightly different number of days between study visits for each individual (usually six or seven).

As we demonstrated in Chapter 3, such a model can be fit by Gibbs sampling. The procedure which updates the hidden states outlined in section 3.3.2 can be applied after updating the forward probabilities to use a Poisson response for $r = 1$ in place of the binomial response used in Chapter 3.

Using Gibb sampling, we sample from the posterior, $[\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{P}, \boldsymbol{\pi}, \boldsymbol{\Sigma}, \sigma^2_\epsilon, \boldsymbol{b}, \boldsymbol{z}|\boldsymbol{y}]$, by sampling from the following conditionals: $[\boldsymbol{\tau}_1, \boldsymbol{\beta}_1|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{z}]$, $[\boldsymbol{\tau}_2, \boldsymbol{\beta}_2|\boldsymbol{y}, \sigma^2_\epsilon, \boldsymbol{b}, \boldsymbol{z}]$, $[\boldsymbol{P}|\boldsymbol{z}]$, $[\boldsymbol{\pi}|\boldsymbol{z}]$, $[\boldsymbol{\Sigma}|\boldsymbol{b}]$, $[\frac{1}{\sigma^2_\epsilon}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{b}, \boldsymbol{z}]$, $[\boldsymbol{b}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{b}, \boldsymbol{z}, \boldsymbol{\sigma^2_\epsilon}]$, and $[\boldsymbol{Z}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{P}, \boldsymbol{\pi}, \sigma^2_\epsilon, \boldsymbol{b}]$.

Once $\boldsymbol{Z}_i$ is updated for each i as outlined in section 3.3.2, the following can be sampled as:

$$[\boldsymbol{\Sigma}|\boldsymbol{b}] \sim InverseWishart(N + \nu_0, \boldsymbol{\Psi}_0 + \sum_{i=1}^{N} \boldsymbol{b}_i \boldsymbol{b}_i^{'}),$$

$$[\frac{1}{\sigma^2_\epsilon}|\boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\tau}_2, \boldsymbol{b}, \boldsymbol{z}] \sim$$
$$Gamma(\frac{(\sum_{i=1}^{N} n_i + a_0)}{2}, \quad (b_0 + \frac{(\sum_{i=1}^{N} \sum_{t=1}^{n_i} (y_{i2t} - \boldsymbol{H}^{'}_{i2tz_{it}} \boldsymbol{\tau}_2 - \boldsymbol{x}^{'}_{i2t} \boldsymbol{\beta}_{2z_{it}} - \boldsymbol{w}^{'}_{i2t} \boldsymbol{b}_i)^2)}{2}),$$

$$[\boldsymbol{\pi}^{(s)}|\boldsymbol{z}] \sim Dirichlet(\boldsymbol{\zeta}^{(s)}_\pi + \boldsymbol{q}^{(s)}),$$

where $\boldsymbol{q}^{(s)}$ is a M-dimensional vector with m^{th} element $\sum_{i=1}^{N}\sum_{t=1}^{n_i}\mathbb{1}(z_{i1}=m, v_i=s)$,

$$[\boldsymbol{P}_u^{(s)}|\boldsymbol{z}] \sim Dirichlet(\boldsymbol{\zeta}_u^{(s)}+\boldsymbol{q}_u^{(s)}), \text{ where } \boldsymbol{P}_u^{(s)} \text{ is the } u^{th} \text{ row of } \boldsymbol{P}^{(s)},$$

where $\boldsymbol{q}_u^{(s)}$ is a M-dimensional vector with m^{th} element $\displaystyle\sum_{i=1}^{N}\sum_{t=2}^{n_i}\mathbb{1}(z_{it}=m, z_{i,t-1}=u, v_i=s)$,

$$[\boldsymbol{\tau}_1, \boldsymbol{\beta}_1|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}] \propto \prod_{i=1}^{N}\prod_{t=1}^{n_i}\left\{\exp\left(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1+\boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}}+\boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i+\log(a_{it})\right)\right\}^{y_{i1t}} \times$$
$$e^{-\exp\left(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1+\boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}}+\boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i+\log(a_{it})\right)},$$

$$[\boldsymbol{\tau}_2, \boldsymbol{\beta}_2|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}, \sigma_\epsilon^2] \propto \prod_{i=1}^{N}\prod_{t=1}^{n_i}\exp\left(-\frac{(y_{i2t}-\boldsymbol{H}_{i2tz_{it}}^{'}\boldsymbol{\tau}_2-\boldsymbol{x}_{i2t}^{'}\boldsymbol{\beta}_{2z_{it}}-\boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i)^2}{2\sigma_\epsilon^2}\right)$$
$$\sim MVN((\tilde{\boldsymbol{B}}^{'}\tilde{\boldsymbol{B}})^{-1}\tilde{\boldsymbol{B}}^{'}\tilde{\boldsymbol{y}}_{\boldsymbol{2}}, \sigma_\epsilon^2(\tilde{\boldsymbol{B}}^{'}\tilde{\boldsymbol{B}})^{-1}), \text{ where}$$
$$\tilde{\boldsymbol{B}} = \left[\boldsymbol{H}|\boldsymbol{X}\right]$$

and is a $\left(\sum_{i=1}^{N} n_i\right) \times (P+c)$ matrix(c being the length of $\boldsymbol{x_{i2t}}$)

with rows comprised of, $[\boldsymbol{H}_{i2tz_{it}}^{'}, \boldsymbol{x}_{i2t}^{'}]$,

and $\tilde{\boldsymbol{y}}_{\boldsymbol{2}}$ is a $\left(\sum_{i=1}^{N} n_i\right)$ length vector with elements $y_{i2t}-\boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i$.

$$\text{Also, } [\boldsymbol{b}_i|\boldsymbol{y_i}, \boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\epsilon^2] \propto \prod_{t=1}^{n_i}\left\{\exp\left(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1+\boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}}+\boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i+\log(a_{it})\right)\right\}^{y_{i1t}} \times$$
$$e^{-\exp\left(\boldsymbol{H}_{i1tz_{it}}^{'}\boldsymbol{\tau}_1+\boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1z_{it}}+\boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i+\log(a_{it})\right)} \times \exp\left(-\frac{(y_{i2t}-\boldsymbol{H}_{i2tz_{it}}^{'}\boldsymbol{\tau}_2-\boldsymbol{x}_{i2t}^{'}\boldsymbol{\beta}_{2z_{it}}-\boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i)^2}{2\sigma_\epsilon^2}\right)\times$$
$$\left[\exp\left(-\frac{1}{2}\boldsymbol{b}_i^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i\right)\right],$$

with $[\boldsymbol{\tau}_1, \boldsymbol{\beta}_1|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}]$ and $[\boldsymbol{b}_i|\boldsymbol{y_i}, \boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\epsilon^2]$ sampled using a Metropolis within Gibbs step.

Based on this Gibbs sampling approach, we fit this model, and present the posterior means and standard deviations for each of the parameters in Table 4.1. Further, to compare this bivariate modeling approach to the two component univariate models, we also

fit these univariate models (cigarette counts only and CO levels only) separately. These models can be done quite easily by ignoring the unused response's parameters, and adjust how sampling of the hidden states, Z_{it}, and random effects, $\boldsymbol{b}_i$ are conducted. For example, for the cigarette-only count model, we can eliminate sampling from the $[\boldsymbol{\tau}_2, \boldsymbol{\beta}_2 | \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b}, \sigma_\epsilon^2]$ and $[\frac{1}{\sigma_\epsilon^2} | \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\tau}_2, \boldsymbol{b}, \boldsymbol{z}]$ full conditional distributions, and modify the random effects covariance structure, $\boldsymbol{\Sigma}$ to be one-dimensional with its corresponding prior (can remain inverse-Wishart, or its equivalent Gamma distribution form), with corresponding one-dimensional zero-mean, normal prior. Updating the hidden states involves modifying the forward probability formulas, only by ignoring the normally distributed (CO) response, which can be done quite easily. Lastly, sampling from $[\boldsymbol{b}_i | \boldsymbol{y_i}, \boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\epsilon^2]$ can be done by modifying the bivariate full conditional to be:

$$[\boldsymbol{b}_i | \boldsymbol{y_i}, \boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\epsilon^2] \propto \prod_{t=1}^{n_i} \left\{ \exp\left(\boldsymbol{H}_{i1tz_{it}}^{'} \boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'} \boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}_{i1t}^{'} \boldsymbol{b}_i + \log(a_{it}) \right) \right\}^{y_{i1t}}$$
$$\times e^{-\exp\left(\boldsymbol{H}_{i1tz_{it}}^{'} \boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'} \boldsymbol{\beta}_{1z_{it}} + \boldsymbol{w}_{i1t}^{'} \boldsymbol{b}_i + \log(a_{it}) \right)} \times \left[\exp\left(-\frac{1}{2} \boldsymbol{b}_i^{'} \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_i \right) \right].$$

4.3.2 Results

The results of the two univariate models are presented along side the bivariate models in Table 4.1. Of particular note, are the stark differences between the transition matrices and population intercepts of the univariate self-reported (SR) count-only model, and the joint bivariate model. In the case of the univariate SR count-only model, the population modes $\boldsymbol{\tau}_1$ are quite a bit lower, which for an individuals with a the relevant random effect equal to zero correspond with a smoking rates of 0.002, 0.030 and 0.085 cigarettes per day in the first, second and third hidden states, respectively. This is much lower than one would expect these rates to be, where for instance only 20 (5.6%) study participants reported smoking fewer than ten cigarettes on the day of randomization. In this case, most of the heterogeneity of the observed response is modeled via the random effects, where we observe a large between-subject heterogeneity, with $\Sigma_{11} = 9.6023$. Further, when we predict the

one-dimensional random effects, $\boldsymbol{b}_i$ by their posterior means, and look at the distribution in Figure 4.1, we find one which is heavily skewed to the right, with many subjects predicted as having a random effect of around -2. This group of individuals are subjects with very homogeneous self-reported cigarette count longitudinal trajectories, where they claim they were abstinent throughout the entire study period. On the other hand, the patients who do transition to the non-abstinent states are modelled mainly through the heterogeneity in the random effects. For instance, for the two subjects with random effects predicted to exceed six, both subjects dropped out after the first study visit. In the one week of available data, the first subject was predicted to have daily self-reported smoke rates of 0.841, 15.385 and 44.153 in the first, second and third hidden states, respectively. They reported smoking 310 cigarettes over 7 days, for an average daily total of 44.286 cigarettes/day. The second subject had predicted daily self-reported cigarette rates of 0.708, 12.953 and 37.174, and reported smoking on average 37.333 cigarettes per day. In both cases, the predicted rates in the highest hidden state correspond well with the observed values, but the separation of the population means corresponding with the hidden states are not very qualitatively different. Further, the distribution of the posterior mean of the random effects is heavily skewed, suggesting poor model fit.

We also conducted an analysis using only the univariate CO-only response. The results of this analysis (see Table 4.1) are generally consistent with those presented in Chapter 3, with a couple noted exceptions. First, we note that the second and third hidden state population modes are quite a bit higher than what was reported in Chapter 3. In Chapter 3, we estimated the modes to be 2.26, 3.63 and 12.95 for the first, second and third hidden states, respectively, but here they are estimated to be 2.07, 6.21 and 19.72, respectively, which is much higher in all but the lowest hidden state. The between-subject heterogeneity as modeled through Σ_{22} is about half of that reported in Chapter 3's analyses. In general terms, the posterior means of the random effects are reasonably symmetric about zero, and roughly normal.

Table 4.1: Comparison of the bivariate and the two univariate models

Parameter	SR Count Only	CO Only	Bivariate Model
π_1	0.4353 (0.0763)	0.8084 (0.0293)	0.7130 (0.0297)
π_2	0.2572 (0.0617)	0.1545 (0.0283)	0.2146 (0.0289)
p^0_{12}	0.1444 (0.0634)	0.0342 (0.0150)	0.0442 (0.0174)
p^0_{13}	0.0643 (0.0408)	0.0115 (0.0067)	0.0136 (0.0077)
p^0_{21}	0.1985 (0.0963)	0.1556 (0.0746)	0.2072 (0.0719)
p^0_{23}	0.4081 (0.1190)	0.0834 (0.0481)	0.1745 (0.0598)
p^0_{31}	0.2369 (0.0911)	0.1810 (0.1411)	0.1241 (0.0786)
p^0_{32}	0.1635 (0.1135)	0.3420 (0.1771)	0.0877 (0.0796)
p^1_{12}	0.1144 (0.0394)	0.0424 (0.0096)	0.0349 (0.0084)
p^1_{13}	0.0640 (0.0283)	0.0067 (0.0036)	0.0083 (0.0037)
p^1_{21}	0.1822 (0.0696)	0.0940 (0.0355)	0.2515 (0.0486)
p^1_{23}	0.3746 (0.0822)	0.0700 (0.0268)	0.2271 (0.0437)
p^1_{31}	0.2515 (0.0602)	0.1511 (0.0781)	0.0650 (0.0307)
p^1_{32}	0.2435 (0.0603)	0.1647 (0.0914)	0.2007 (0.0501)
τ_{11}	-6.4294 (0.4740)	-	-5.2899 (0.2120)
τ_{12}	2.9067 (0.1597)	-	4.4679 (0.2035)
τ_{13}	1.0543 (0.0495)	-	1.3101 (0.0434)
τ_{21}	-	1.1207 (0.0210)	1.1851 (0.0221)
τ_{22}	-	0.8546 (0.0446)	0.2959 (0.0474)
τ_{23}	-	1.0557 (0.0752)	0.8782 (0.0569)
Σ_{11}	9.6023 (1.8137)	-	2.3318 (0.4179)
Σ_{12}	-	-	0.1242 (0.0459)
Σ_{22}	-	0.0526 (0.0083)	0.1048 (0.0111)
$\frac{1}{\sigma^2_\epsilon}$	-	11.4430 (0.5777)	8.6795 (0.3676)

Each value is the posterior mean (posterior standard deviation), where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

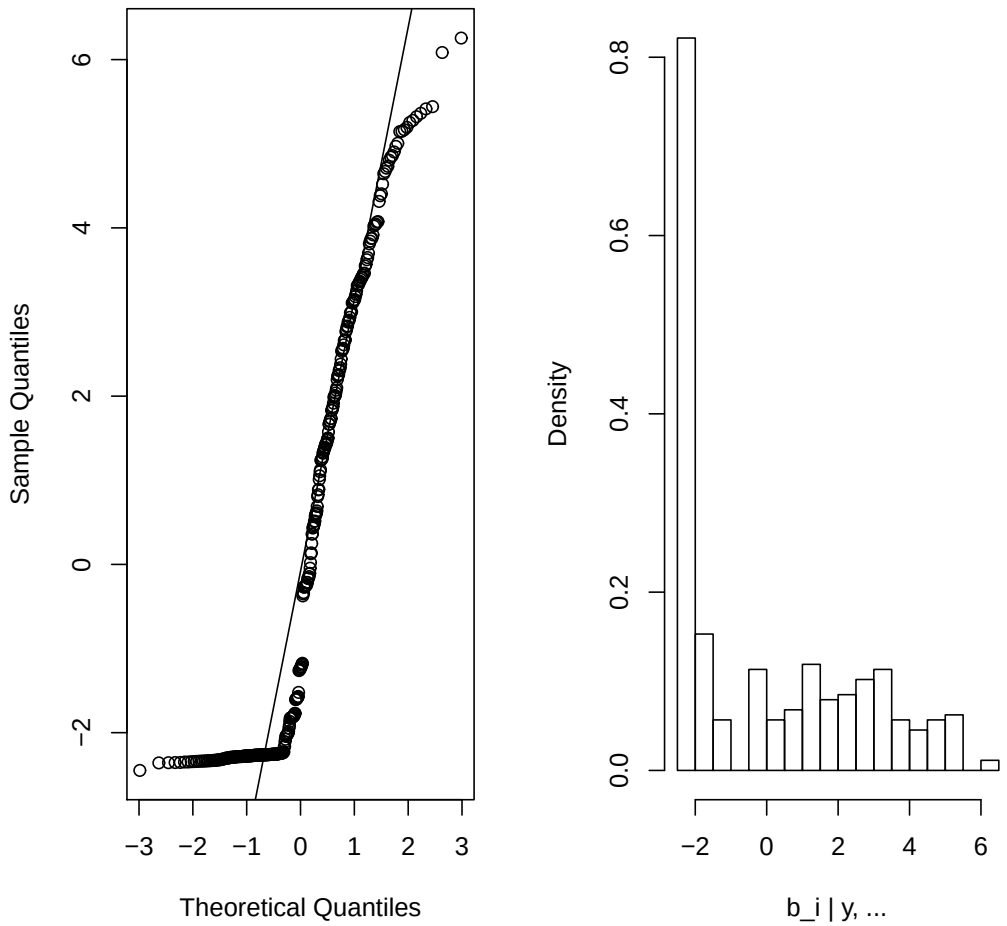

Figure 4.1: QQ-normal plot and histogram for the posterior mean of $\boldsymbol{b}_i$ for the self reported cigarette count-only univariate response MHMM model

An interesting comparison of the sources of heterogeneity in the CO response can be conducted by using some pre-treatment data collected in this study. The study investigators collected up to two screening CO levels, and one baseline measurement, where the subjects were known to be currently smoking. This provides a very homogeneous set of

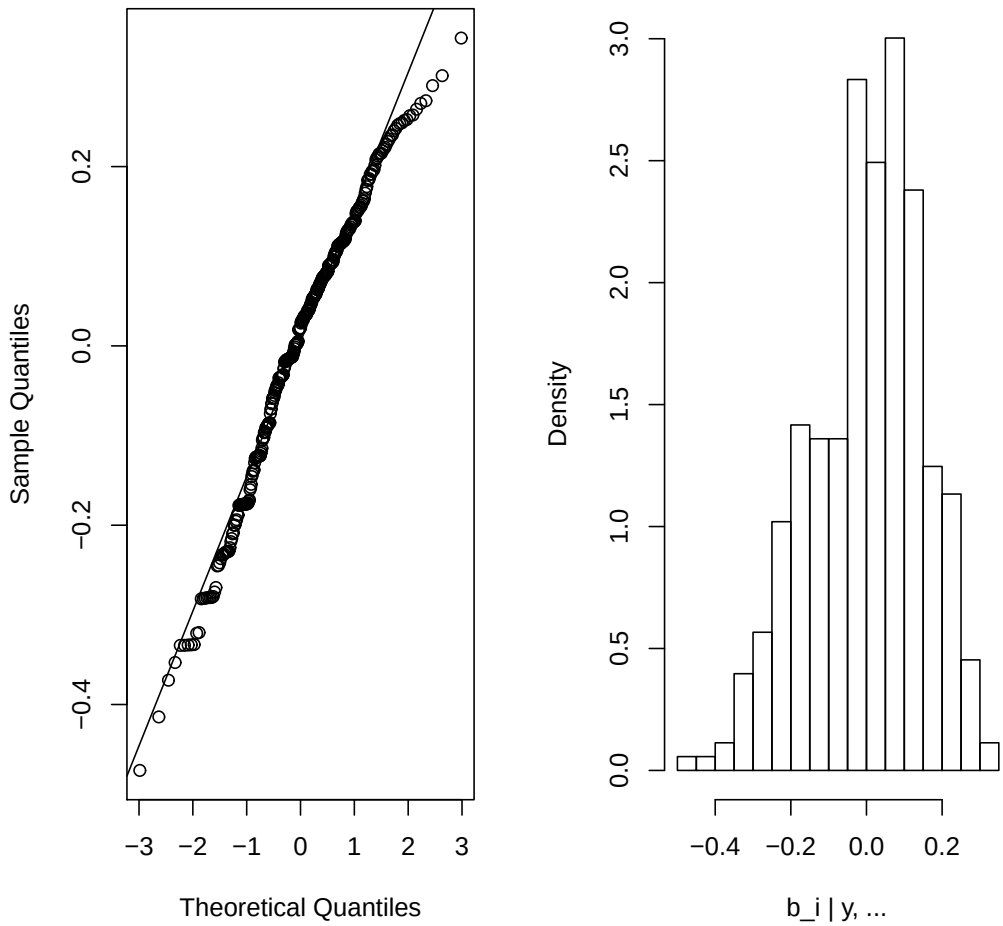

Figure 4.2: QQ-normal plot and histogram for the posterior mean of $\boldsymbol{b}_i$ for the CO-only univariate response MHMM model

observations in terms of disease states which we can compare the variance components estimated under a MHMM to. We fit a simple linear mixed effects model with a random intercept and fixed time effect with the *nlme* package in *R* to these data such that:

$$Y_{i2t} = \gamma_0 + b_i + \gamma_1 t_{it} + \epsilon_{it} \tag{4.3}$$

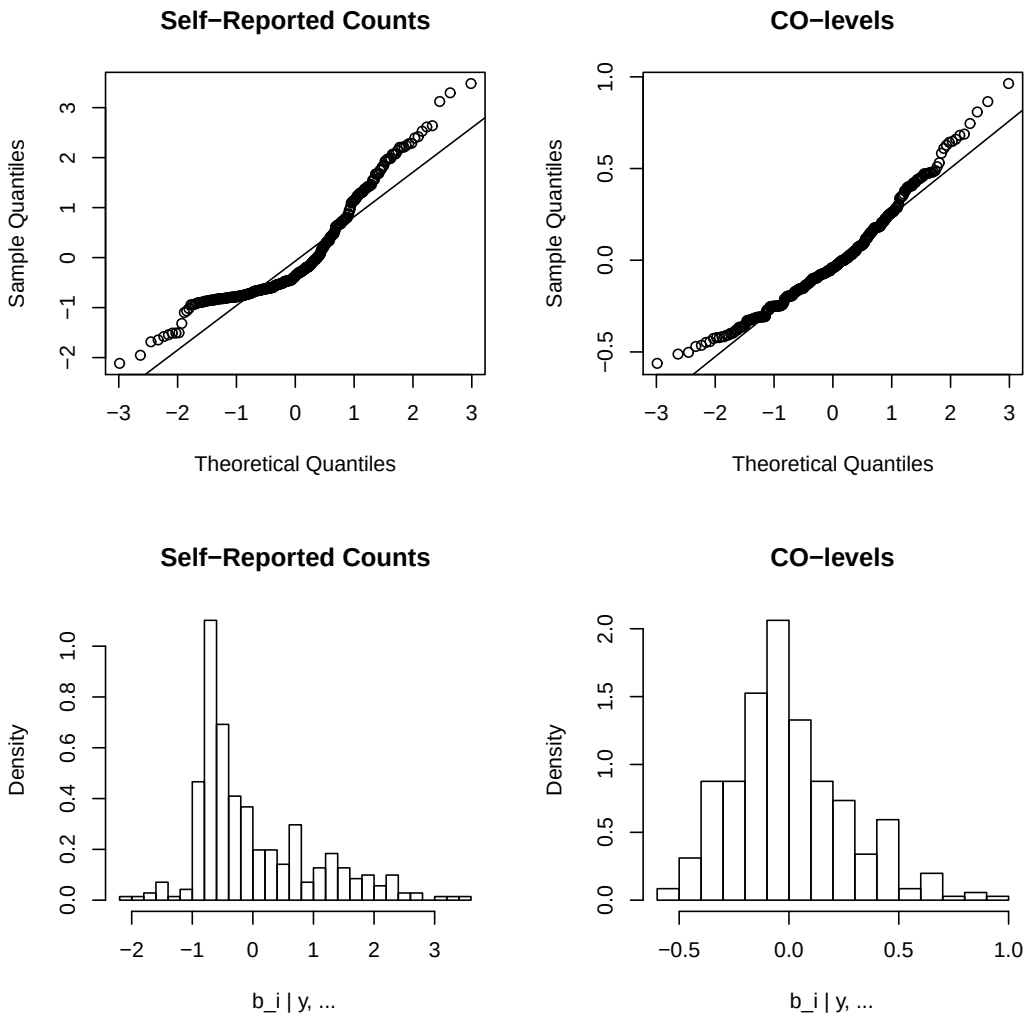

Figure 4.3: QQ-normal plot and histogram for the posterior mean of $\boldsymbol{b}_i$ for the bivariate response MHMM model

for $t \in \{-2, -1, 0\}$, $b_i \overset{\text{iid}}{\sim} N(0, \Sigma_{22})$ and $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2)$. We assessed these estimates to those derived from the CO-only response model in Table 4.1. Under (4.3) the pre-treatment data provides estimates of $\hat{\gamma}_0 = 3.1925$, $\hat{\gamma}_1 = -0.0005$, $\hat{\Sigma}_{22} = 0.0929$, and $\frac{1}{\hat{\sigma}_\epsilon^2} = 16.1296$. The estimate of Σ_{22} corresponds closely with those we have estimated in the MHMM-

based models. On the other hand, $\hat{\gamma}_0 = 3.1925$ corresponds with a untransformed CO level of 23.35 which is almost double that which we have previously reported, suggesting that CO levels in this pre-treatment period may not represent those we observe during the treatment period. In fact, we rarely see CO-levels above 20 during the treatment period, with only 21 longitudinal observations exceeding this threshold. Lastly, under (4.3), $\frac{1}{\hat{\sigma}_\epsilon^2}$ is a little higher than what we estimate elsewhere, suggesting there is smaller residual error during the pre-treatment period, which probably isn't surprising, as we would expect less of this source of heterogeneity under a more constant disease status, where the response may be slightly lagged behind disease status. This analysis was meant to be illustrative only and in principle could include other covariate effects, in addition to a random time effect, but was fit to (4.3) mainly for a simple comparison to the models in Table 4.1.

4.4 Classifying the Hidden States

In general, during a clinical study when disease states are classified by using cutoffs of some surrogate endpoint or biomarker, it may be difficult to determine the true state of the clinical outcome precisely. Likewise, when using hidden state approaches such as the ones outlined through this chapter and Chapter 3, except in cases of simulated data, it will be difficult to evaluate the utility of inference about these hidden states as a classifier for the true underlying disease states. As discussed in section 4.2, these hidden states may in some manner closely resemble the thresholds on the observed responses. With that said, one would expect that inference drawn from hidden state approaches would roughly correspond with the disease states determined by cutoff approaches (i.e., they would agree most of the time). When there lacks agreement between the two approaches, this might illustrate inadequacies of either the modeling approach, the cutoff protocol, or both. With this in mind, in this section, we will discuss a framework for evaluating inferences pertaining to the hidden states, when cutoffs are used to define hidden states.

We will evaluate the agreement of these cutoff-defined disease states, treating the study-defined disease states as a 'gold-standard,' noting that in smoking cessation studies, the only true gold standard would be 24-hour surveillance, which is both impractical and may not generalize to other populations or settings.

4.4.1 Methodology

The inference we can conduct for the hidden states is mainly generated from a function of the state or locally decoded probabilities:

$$\hat{\phi}_{itk} = P(Z_{it} = k|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta}) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}(Z_{it}^{(j)} = k) \ , \tag{4.4}$$

where $Z_{it}^{(j)}$ is the simulated value of the jth iteration of the MCMC algorithm for J total iterations. From each observation time, we get a vector of M probabilities, and will denote this $\boldsymbol{\phi}_{it} = [\phi_{it1}, \phi_{it2}, \ldots, \phi_{itM}]$ with $\sum_{m=1}^{M} \phi_{itm} = 1$ for each i and t. There are other ways to compute these quantities, but this one does not require any additional computation during the MCMC algorithm, and can be quickly and efficiently processed from the MCMC output once the simulation is completed.

We also define the most probable state as a maximum a posteriori estimator for each hidden state as:

$$\hat{\delta}_{it} = \underset{k \in \{1,2,\ldots,M\}}{\operatorname{argmax}} \hat{\phi}_{itk}, \tag{4.5}$$

which can take on any of the discrete values in the hidden state space, such that $\hat{\delta}_{it} \in \{1, 2, \ldots, M\}$, with its value representing the state with highest probability compared to each of the alternatives individually.

4.4.2 Results

Using the study-defined disease as a so-called 'gold standard', we will examine the utility of using $\hat{\boldsymbol{\phi}}_{it}$ and $\hat{\delta}_{it}$ as a framework to do inference about the disease states. First, we look at how $\hat{\delta}_{it}$ corresponds with different possible study definitions of smoking. These results are presented in Table 4.2, where each weekly observation is classified according to several study definitions as being a smoking or a non-smoking observation. These results are cross-tabulated against the constant $\hat{\delta}_{it}$ values for each observation. We note that generally the models based on a univariate self-reported response or the bivariate response model perform the best, with the CO-only model generally performing the poorest across almost all scenarios. We report two different calculations of accuracy, with *agreement* assessed in all observations, equating $\hat{\delta}_{it} = 2$ with smoking observations, and *strict agreement* which assesses only those observations with $\hat{\delta}_{it} = 1$ or 3. The relatively strong performance of the univariate SR only model is somewhat surprising as the results in Table 4.1 were slightly unexpected, and the model fit appeared to be poor (see Figure 4.1). We believe this may be due to the absence of CO data, whereby a model fit using only SR is going to be very sensitive to any non-zero SR observations which defines many of these study definitions. In the bivariate case, many of the misclassified observations appear to be at least partly due to $\hat{\delta}_{it} = 1$ allowing some observations with a small SR count of 1-7 cigarettes in a week, and strict definitions of smoking abstinence would obviously preclude these observations. We will explore these discrepant cases in more detail later.

Next, using $\hat{\phi}_{itk}$ as a classifier we will assess classification performance by receiver operating characteristic (ROC) curves. Figures 4.4 - 4.9 present the ROC curves for each of the three models considered thus far (the two univariate models and one bivariate model) broken down by study visit, with area under the curve (AUC) reported for each. Figures 4.4-4.6 use $\hat{\phi}_{it1}$ as a classifier for the study definition of smoking abstinence used in the motivating smoking cessation study (SR=0 *and* CO$\leq$10). Figures 4.7-4.9 use $\hat{\phi}_{it3}$ as a classifier for the study definition of smoking used in the motivating study (SR>0 *or* CO>10).

Table 4.2: Breakdown of most probable disease states by different study-defined disease states under separate univariate and bivariate models

Definition of Smoking	Univariate SR Only			Univariate CO Only			Bivariate		
		NS	S		NS	S		NS	S
SR>0 or CO>10	$\hat{\delta}_{it}=1$	1291	64	$\hat{\delta}_{it}=1$	1183	240	$\hat{\delta}_{it}=1$	1295	118
	$\hat{\delta}_{it}=2$	16	102	$\hat{\delta}_{it}=2$	127	171	$\hat{\delta}_{it}=2$	17	196
	$\hat{\delta}_{it}=3$	3	299	$\hat{\delta}_{it}=3$	0	54	$\hat{\delta}_{it}=3$	4	151
	Strict Agreement: 96.0% Agreement: 95.3%			Strict Agreement: 83.8% Agreement: 79.3%			Strict Agreement: 92.2% Agreement: 92.5%		
SR>1 or CO>10		NS	S		NS	S		NS	S
	$\hat{\delta}_{it}=1$	1321	34	$\hat{\delta}_{it}=1$	1291	132	$\hat{\delta}_{it}=1$	1388	25
	$\hat{\delta}_{it}=2$	31	102	$\hat{\delta}_{it}=2$	146	152	$\hat{\delta}_{it}=2$	43	170
	$\hat{\delta}_{it}=3$	85	299	$\hat{\delta}_{it}=3$	0	54	$\hat{\delta}_{it}=3$	12	143
	Strict Agreement: 92.8% Agreement: 91.5%			Strict Agreement: 91.1% Agreement: 84.3%			Strict Agreement: 97.6% Agreement: 95.8%		
SR>0 or CO>8		NS	S		NS	S		NS	S
	$\hat{\delta}_{it}=1$	1273	82	$\hat{\delta}_{it}=1$	1183	240	$\hat{\delta}_{it}=1$	1280	133
	$\hat{\delta}_{it}=2$	16	102	$\hat{\delta}_{it}=2$	109	189	$\hat{\delta}_{it}=2$	16	197
	$\hat{\delta}_{it}=3$	3	299	$\hat{\delta}_{it}=3$	0	54	$\hat{\delta}_{it}=3$	2	153
	Strict Agreement: 94.9% Agreement: 94.3%			Strict Agreement: 83.7% Agreement: 80.3%			Strict Agreement: 91.4% Agreement: 91.8%		
SR>7 or CO>10		NS	S		NS	S		NS	S
	$\hat{\delta}_{it}=1$	1334	21	$\hat{\delta}_{it}=1$	1373	50	$\hat{\delta}_{it}=1$	1401	12
	$\hat{\delta}_{it}=2$	65	53	$\hat{\delta}_{it}=2$	196	102	$\hat{\delta}_{it}=2$	145	68
	$\hat{\delta}_{it}=3$	170	132	$\hat{\delta}_{it}=3$	0	54	$\hat{\delta}_{it}=3$	29	126
	Strict Agreement: 88.5% Agreement: 85.6%			Strict Agreement: 96.6% Agreement: 86.1%			Strict Agreement: 97.4% Agreement: 89.9%		
SR>0 or CO>20		NS	S		NS	S		NS	S
	$\hat{\delta}_{it}=1$	1307	48	$\hat{\delta}_{it}=1$	1183	240	$\hat{\delta}_{it}=1$	1307	106
	$\hat{\delta}_{it}=2$	17	101	$\hat{\delta}_{it}=2$	143	155	$\hat{\delta}_{it}=2$	19	194
	$\hat{\delta}_{it}=3$	3	299	$\hat{\delta}_{it}=3$	1	53	$\hat{\delta}_{it}=3$	7	148
	Strict Agreement: 96.9% Agreement: 96.2%			Strict Agreement: 83.7% Agreement: 78.4%			Strict Agreement: 92.8% Agreement: 92.9%		

SR is weekly self-reported smoking; CO is the weekly CO level; NS are non-smoking observations; S are smoking observations; $\hat{\delta}_{it}$ is the most probable hidden state classifier from (4.5).

Generally, the bivariate model has a higher AUC in all but one of the instances, with the univariate model using only the CO levels having the poorest performance. We also note a lack of concavity in Figure 4.7, and believe this to be at least partially attributable to the intermediate hidden state.

In Figure 4.10 we represent all four cases (on three distinct subjects) noted in the SR>0 or CO>10 case for the bivariate model fit in Table 4.2 in longitudinal panels for each patient where the most probable hidden state was state three, but the study definition defined the visit as being non-smoking. In each row we have each of the longitudinal responses and locally decoded probability, $P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ of being in the third hidden state at each study visit. In the first subject, we have a subject where the last two CO measurements are significantly elevated relative to the first three measurements, despite the patient reporting they did not smoke, and CO levels remaining below ten. This would suggest that the hidden states are somewhat sensitive to relative changes in CO levels over the study period, and is again seen in the next two subjects, where the first one is not accompanied by a corresponding SR smoking event, but the last one is. We see a possible strength in this analysis approach, where unexplained significant increases in CO levels relative to other periods are more easily detected when compared to using a hard >10 threshold as was done in the study. Additionally, while only one of these cases was corresponded with SR smoking, the magnitude that $P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ was seen to increase was also the greatest in this instance. These are the only four cases where the model would have predicted smoking by classifying each state on the maximum locally decoded probability.

Alternatively, Figure 4.11 is an illustrative, but representative sample of five subjects which had a maximally locally decoded probability state of hidden state one, while the subject was determined to have smoked according to the study definition. In many instances these discrepancies can be explained by (often single isolated) study weeks where a relatively small number of cigarettes are smoked. Often these subjects report a single

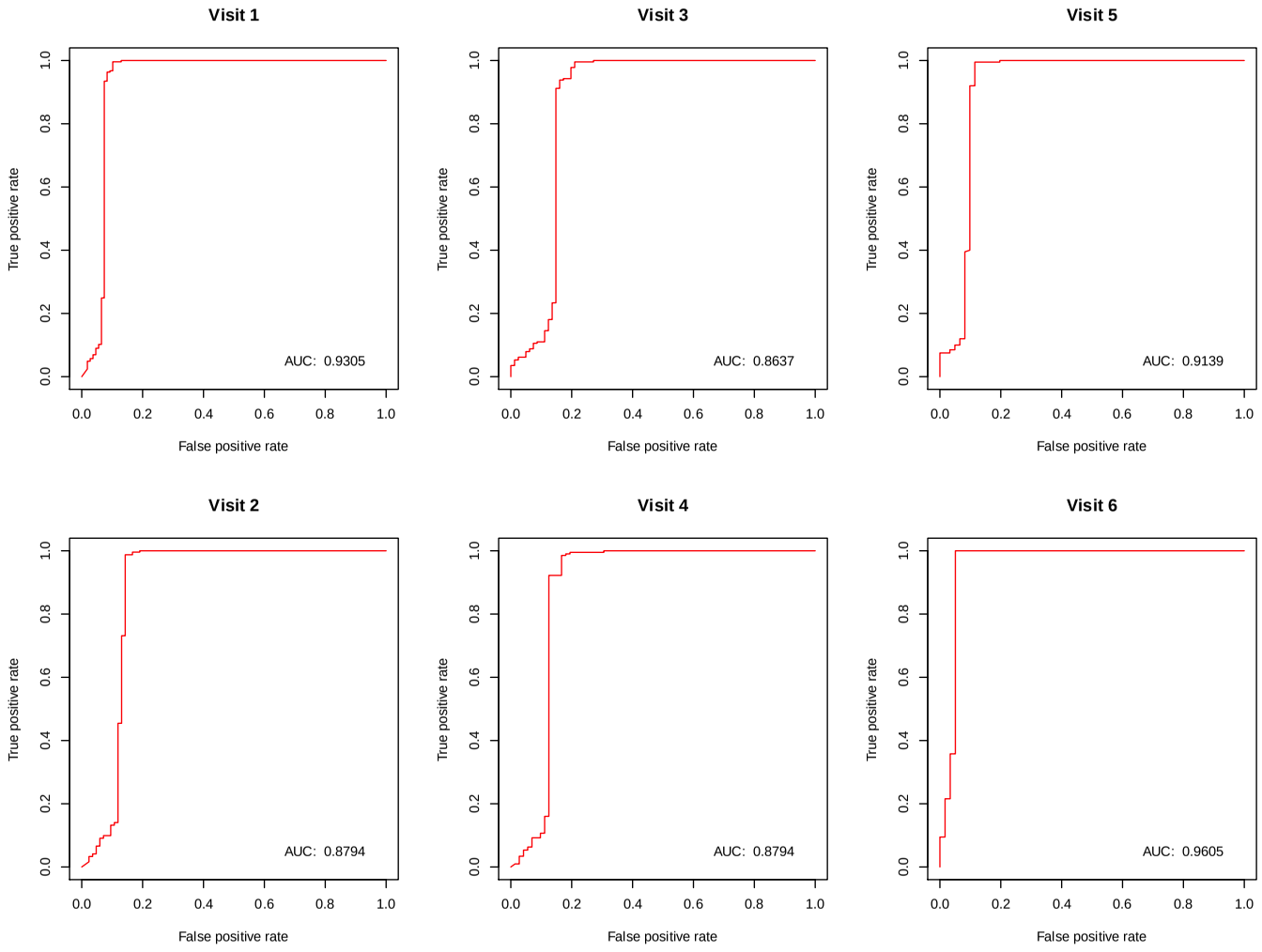

Figure 4.4: ROC curve for $\hat{\phi}_{it1} = P(Z_{it} = 1|y, ...)$ for each study visit (t) for the self reported cigarette count-only univariate response MHMM model.

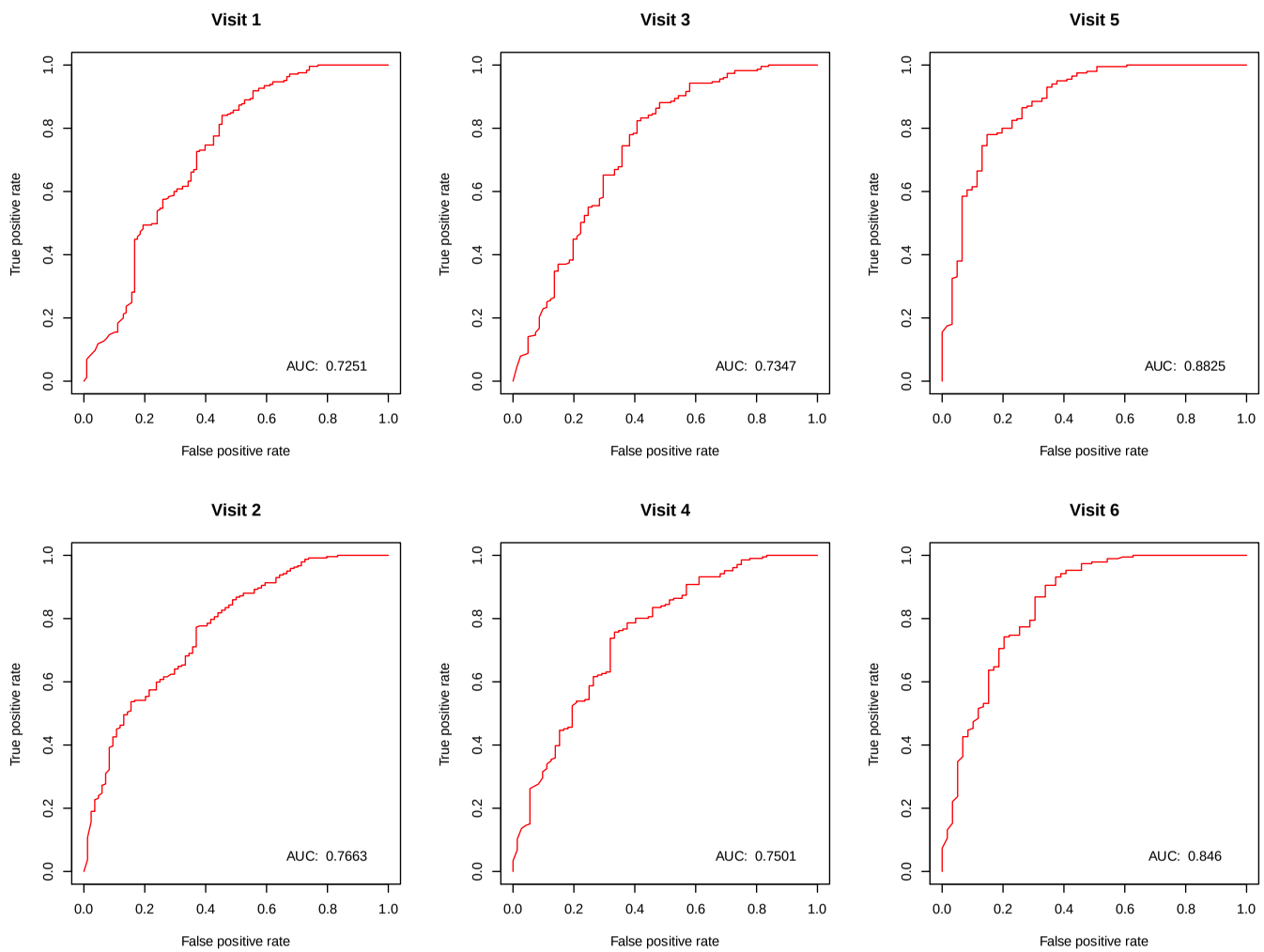

Figure 4.5: ROC curve for $\hat{\phi}_{it1} = P(Z_{it} = 1|y, ...)$ for each study visit (t) for the CO-only univariate response MHMM model.

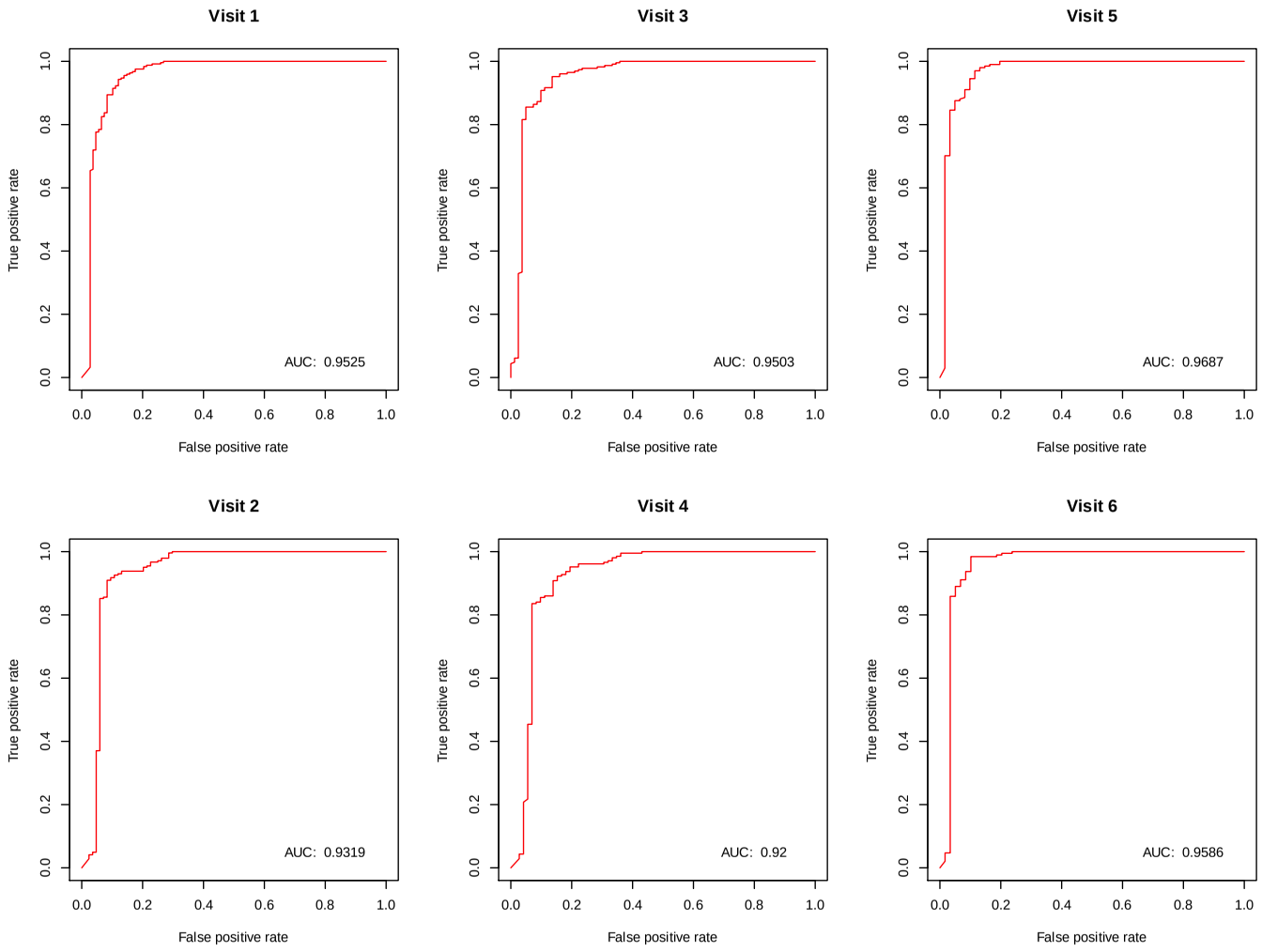

Figure 4.6: ROC curve for $\hat{\phi}_{it1} = P(Z_{it} = 1|y, ...)$ for each study visit (t) for the bivariate response MHMM model.

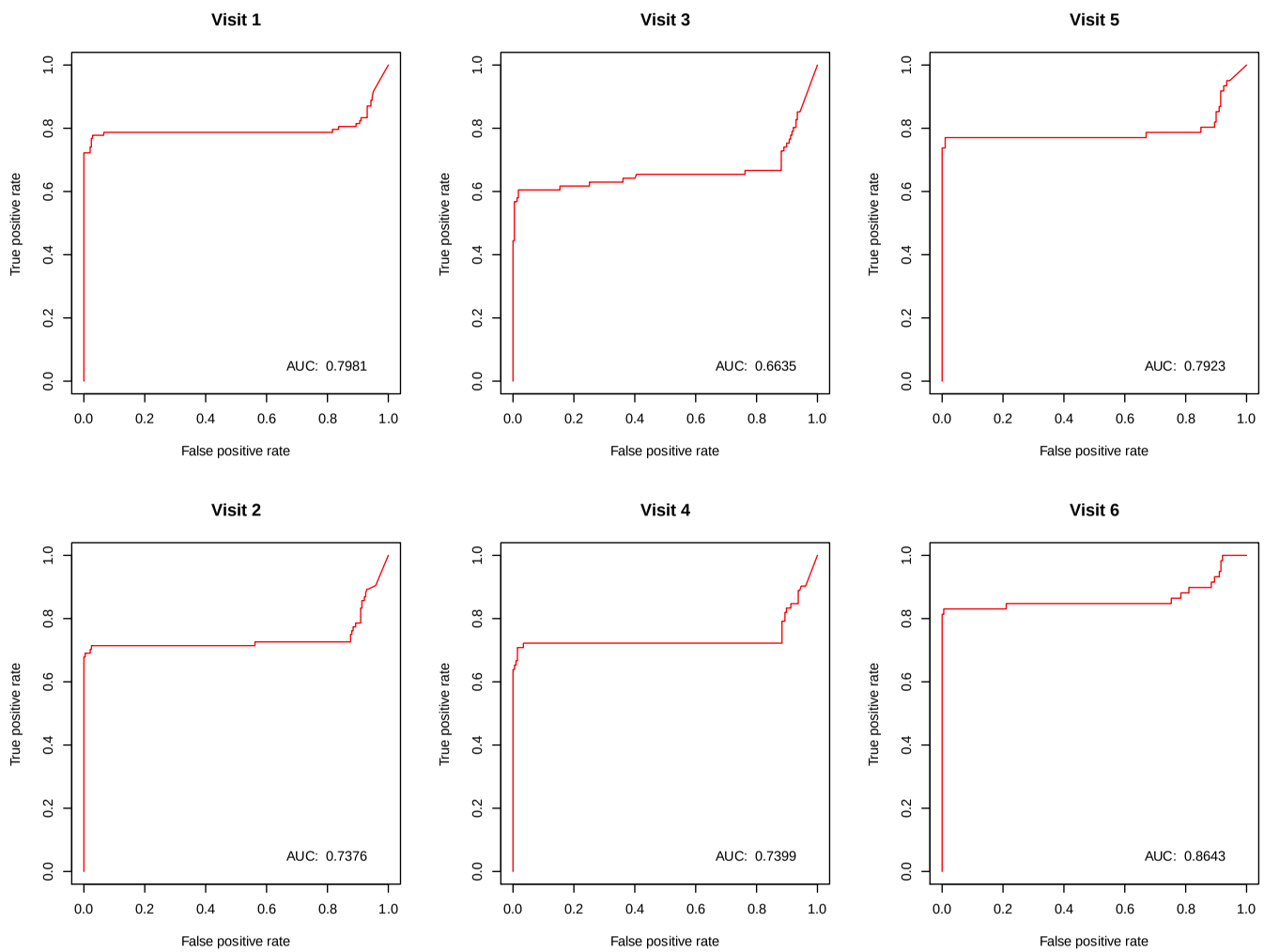

Figure 4.7: ROC curve for $\hat{\phi}_{it3} = P(Z_{it} = 3|y, ...)$ for each study visit (t) for the self reported cigarette count-only univariate response MHMM model.

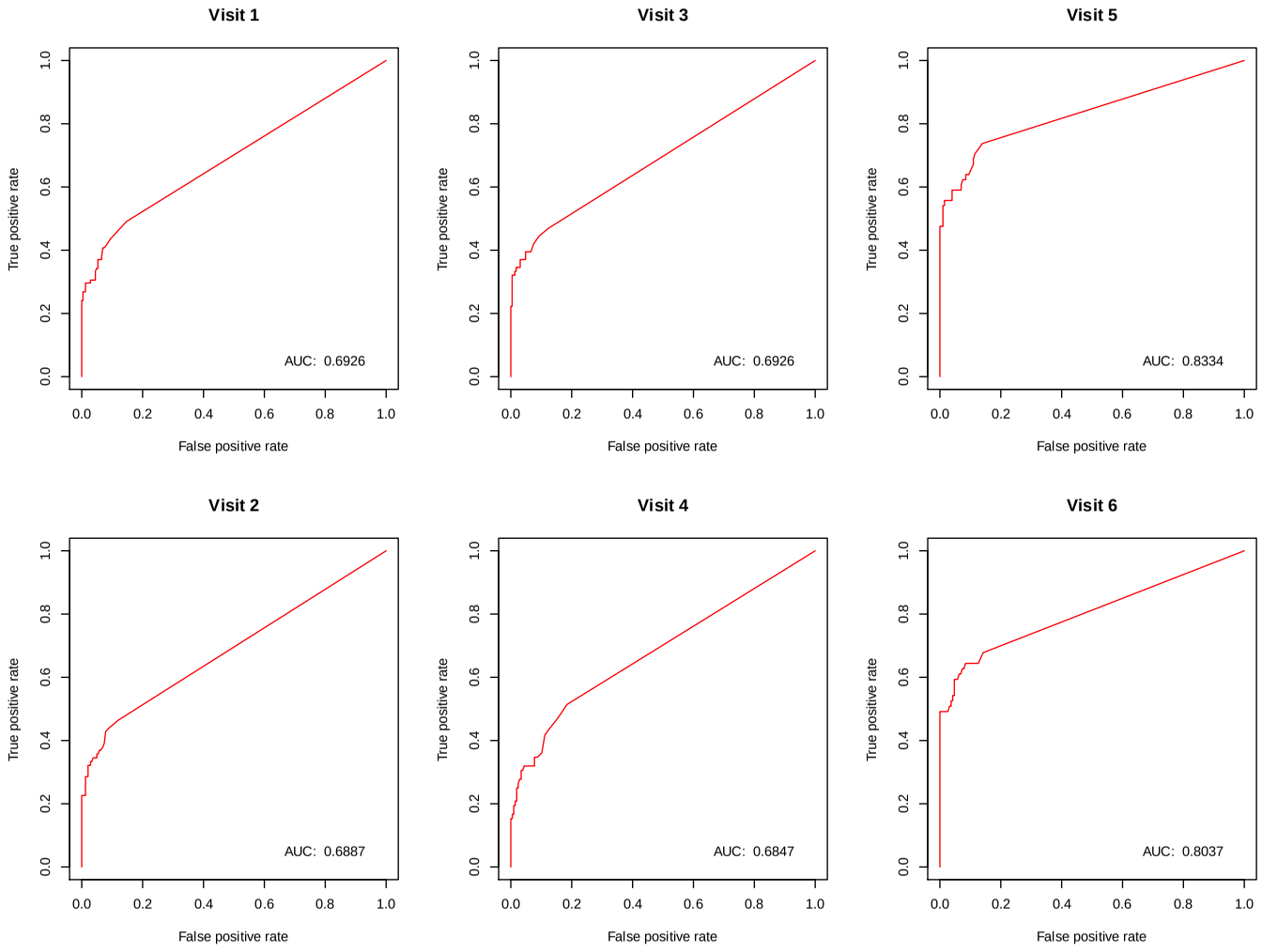

Figure 4.8: ROC curve for $\hat{\phi}_{it3} = P(Z_{it} = 3|y, ...)$ for each study visit (t) for the CO-only univariate response MHMM model.

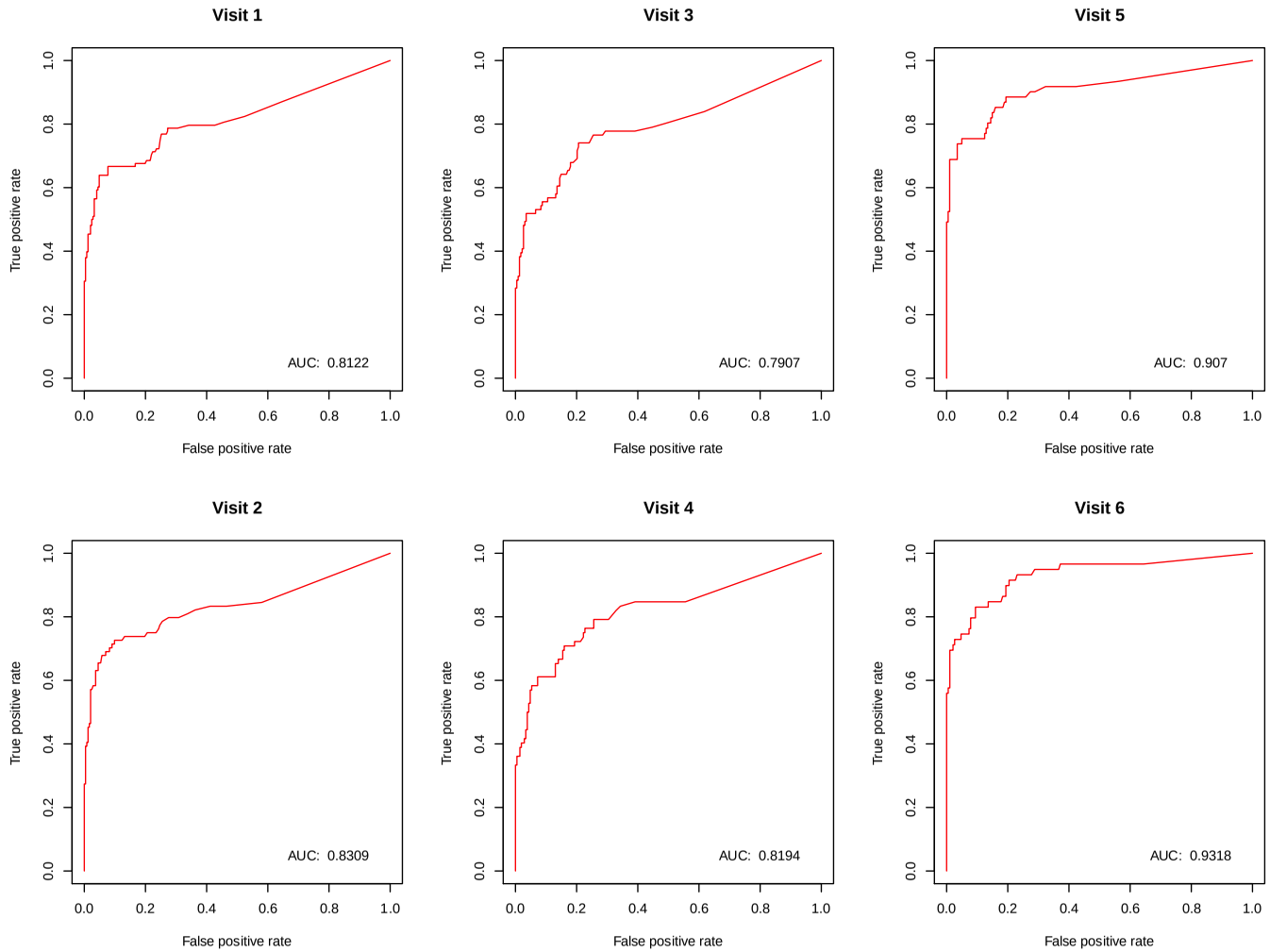

Figure 4.9: ROC curve for $\hat{\phi}_{it3} = P(Z_{it} = 3|y, ...)$ for each study visit (t) for the bivariate response MHMM model.

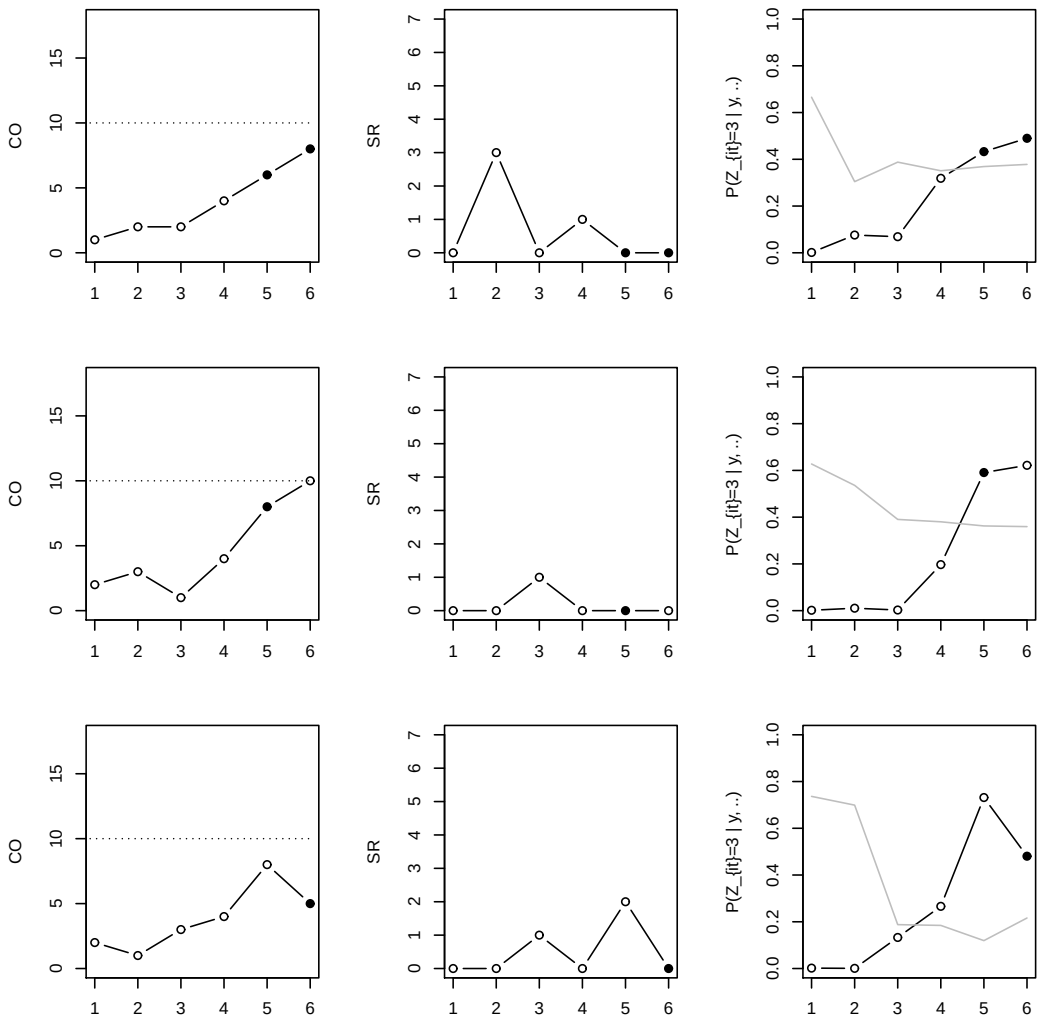

Figure 4.10: Example longitudinal plot for subjects with one or more visits such that the most probable hidden state according to the MHMM bivariate longitudinal model was state three, despite it being a non-smoking visit according to the study definition.
Solid circles represent cases where state three was the most probable state, but the study definition defines this as being a non-smoking visit. The grey line in the rightmost plots represents $P(Z_{it} = 1|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$, whereas the black line represents $P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{\Theta})$. The dotted line in the leftmost plot represents the study-defined CO cutoff of ten.

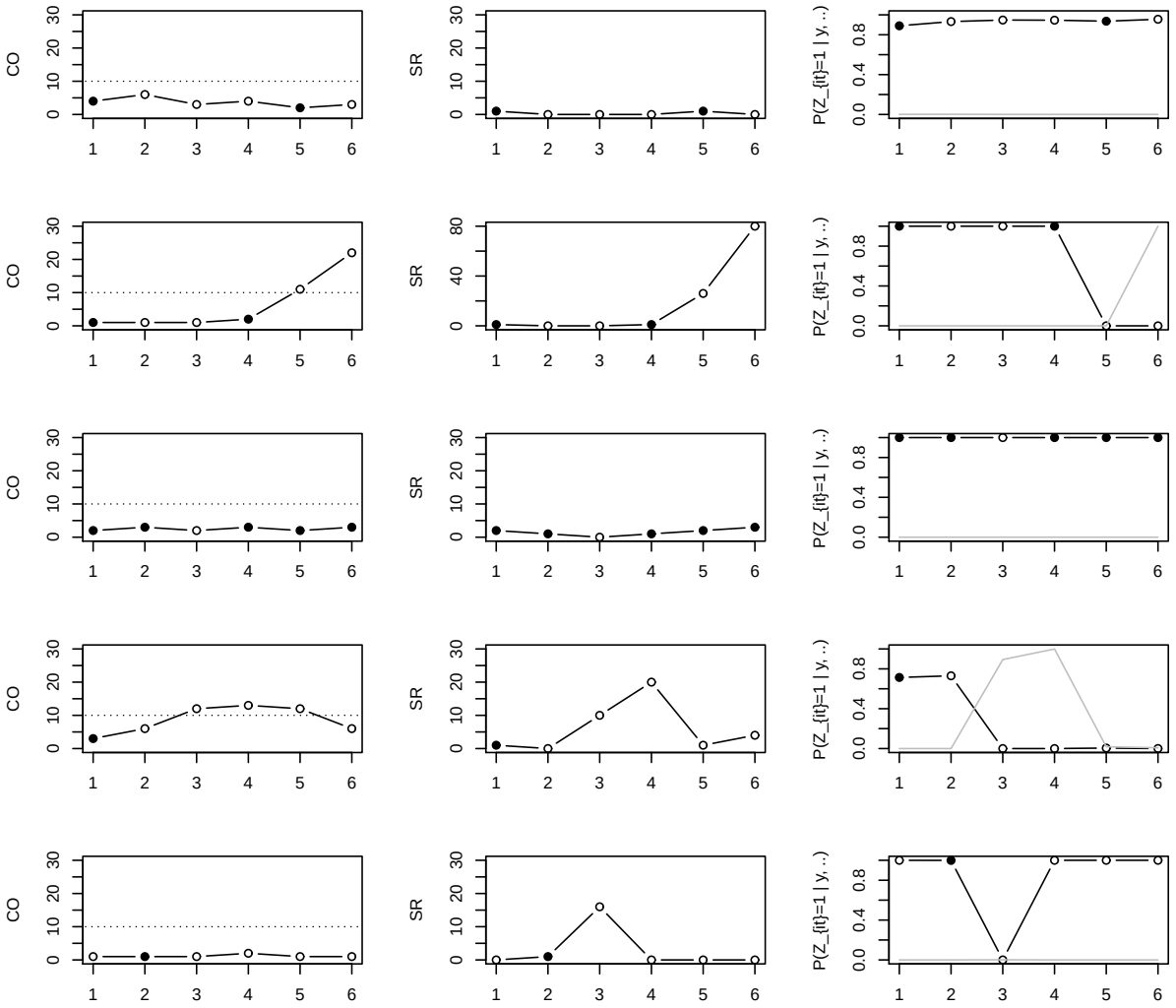

Figure 4.11: Example longitudinal plot for subjects with one or more visits such that the most probable hidden state according to the MHMM bivariate longitudinal model was state one, despite it being a smoking visit according to the study definition.

Solid circles represent cases where state one was the most probable state, but the study definition defines this as being a smoking visit. The grey line in the rightmost plots represents $P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$, whereas the black line represents $P(Z_{it} = 1|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$. The dotted line in the leftmost plot represents the study-defined CO cutoff of ten.

cigarette smoked during the week, have maintained low CO levels during the week, and report smoking abstinence in the following visits. A case of this happens in each of the five examples in Figure 4.11, most often occurring during the first two weeks. Frequently, study definitions of smoking abstinence allow for a grace period during the first couple of study visits, and these would be largely ignored. While often occurring during the first few weeks, these discrepancies between the most probable hidden state and study-defined state do occur in weeks three through six, where, for example, the third row of plots illustrates a subject with all but one week with relatively low self-reported smoking, while maintaining relatively low CO levels. In this sense, the hidden state decoding is somewhat forgiving to low self-reported smoking while low CO levels are maintained, and no continued increases are noted. This is again evident in Table 4.2 where modifying the study definition to include weeks where only one cigarette was smoked increases the agreement and strict agreement from 92.5% and 92.2% to 95.8% and 97.6%, respectively. Of the 118 observations with a discrepancy between the most probable state being state 1, with a study-defined status of smoking, 93 are explained by observations of one cigarette smoked in a week defining smoking status. About half (13 observations) of these remaining 25 discrepant observations, result when a subject smoked seven or fewer cigarettes in a week, and the model classified them in hidden state one. The remaining 12 observations are accounted by slightly elevated CO levels, with six of these remaining discrepancies occurring at a CO level of 11, and all occurring below 17.

An interesting case is that of the last row (subject) in Figure 4.11, where at week 3 the subject reports smoking 16 cigarettes, but maintains a low CO level, and reports no smoking during the rest of the study visits. This one week of smoking is accompanied by a stark decrease to nearly zero for $P(Z_{it} = 1|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$, but relatively no change in $P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$, indicating that this observation would be one which has a most probable hidden state classification of two. This is both reassuring and problematic. It is reassuring, because changes in one of the component smoking measures can be detected

effectively, such that the model would not classify this observation as a hidden state 1 observation. We had already seen a case where relative changes in CO levels were detected as smoking, even when self-reported smoking remained non-existent, and this illustrates the other case, where CO levels remained low, but self-reported smoking was detected as non-abstinent. On the other hand, conventional wisdom would generally conclude that this was a clear case of someone relapsing, but the bivariate MHMM classifies this observation as being most probably in the intermediate state. This is probably due to two separate but important issues. The first is likely that the subject returns to self-reported abstinence in the immediate following weeks. The second is likely that the relapse is not accompanied by a CO level increase, most probably because the smoking occurred out of the window of detection. We rest on the side of reassurance, even if we are not entirely convinced that this observation is emblematic of intermediate smoking state visit, mainly because to classify this observation as a non-smoking state visit would have been a worse classification. Much of this is caused by the lack of a study-defined disease state which would closely correspond with hidden state two, and it is difficult to draw comparisons between this hidden state and findings in the study. As we argue above, the second hidden state resembles the smoking state much more so than the non-smoking state both conceptually, and empirically, that concluding an observation such as this one being in hidden state two is less of an error than concluding that it is in hidden state one, which there is much evidence against.

4.5 Assessing the Conditional Independence Assumption

Throughout this thesis, we have been relying upon several conditional independence assumption to simplify the modeling approach. One assumption we would like to be able to assess is the independence of the random effects from the hidden states. One way of doing this is by plotting the posterior mean for each subject's random effect by the most

probable hidden state ($\hat{\delta}_{it}$) at each study visit. These plots are done in Figures 4.12 for the Poisson SR count response ($r = 1$) and 4.13 for the Normal CO level response ($r = 2$), and are done for the bivariate model examined in sections 4.3.2 and 4.4.2

As can be seen in Figures 4.12 and 4.13, there may be a small dependence of the random effects on the random effect. In general, the predicted random effects for observations with $\hat{\delta}_{it} = 1$ have their median below zero, for both responses. This is relatively small when compared to the overall variance of the random effects distributions, and while the variance of the random effect distribution does seem to fluctuate over the study visits, there is no systematic pattern dependent on the most probable hidden state. While boxplots for $\hat{\delta}_{it} = 1$ do seem to have a larger number of outliers as defined by the boxplots, it also has the largest number of observations, which may explain this in part.

4.6 Simulation Study

In order to assess the classification properties of these methods, we conducted a simulation study. We consider two different situations, with the first situation representing a baseline to compare the latter to. Our first situation is the standard MHMM model where none of the model assumptions are violated. The second situation resembles a scenario when the data are contaminated by some unmodelled external process which is dependent on the hidden process. In this second case, two of the hidden states are contaminated at some rate such that they resemble the third hidden state. This would correspond to an area of concern for our motivating dataset, where some observations may have under-reported SR cigarette counts, or high CO-levels were undetected given its short detection half life.

With this in mind, every run in each of the situations was fit first with the two separate component univariate longitudinal MHMM models, and then with the bivariate model, with 100 runs of each done according to the methodology described in section 4.3.1. The specific details of the two situations are as follows:

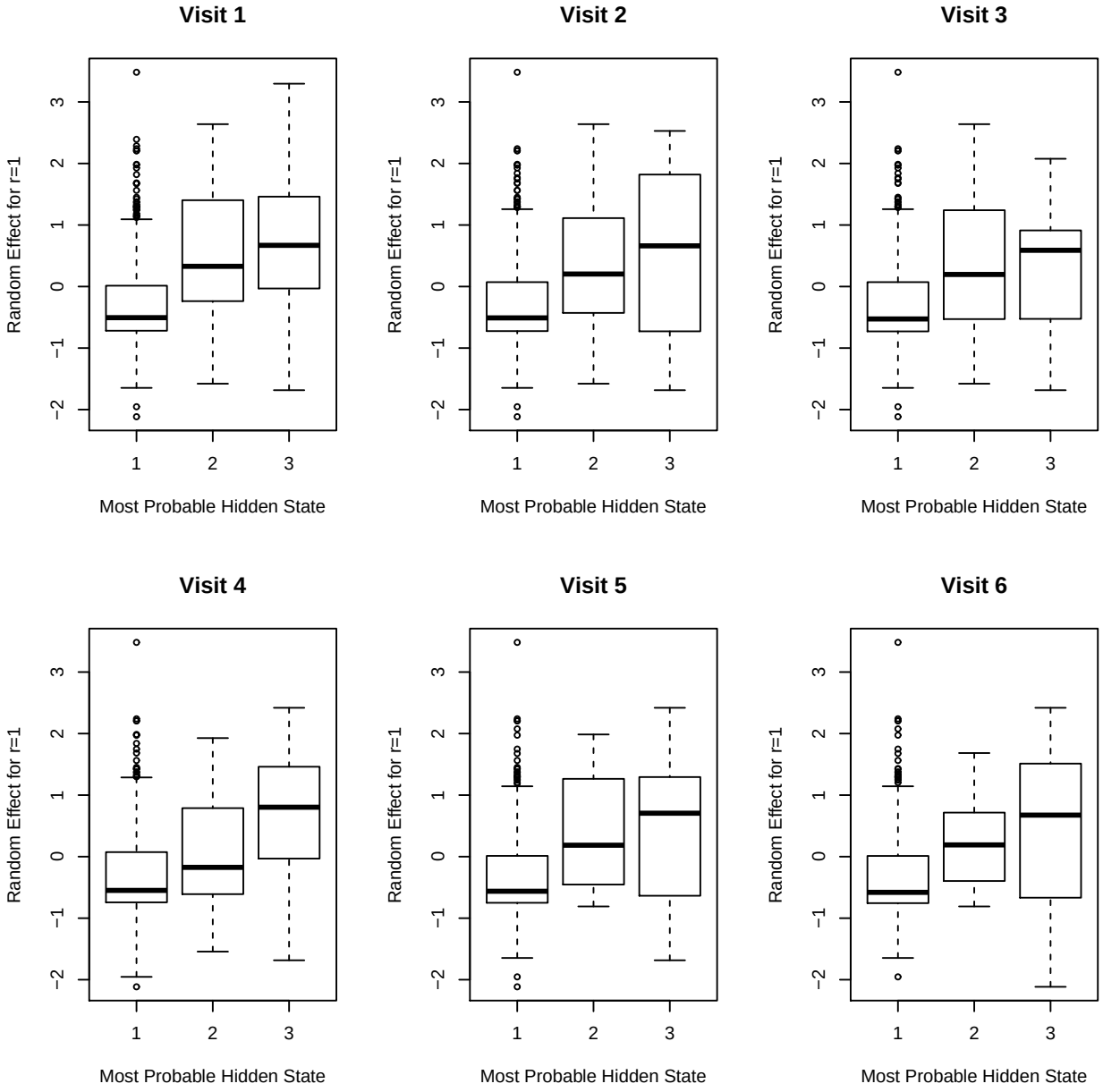

Figure 4.12: Box plot for the Poisson SR count response ($r = 1$) random effect at each study visit stratified by the most probable hidden state, $\hat{\delta}_{it}$.

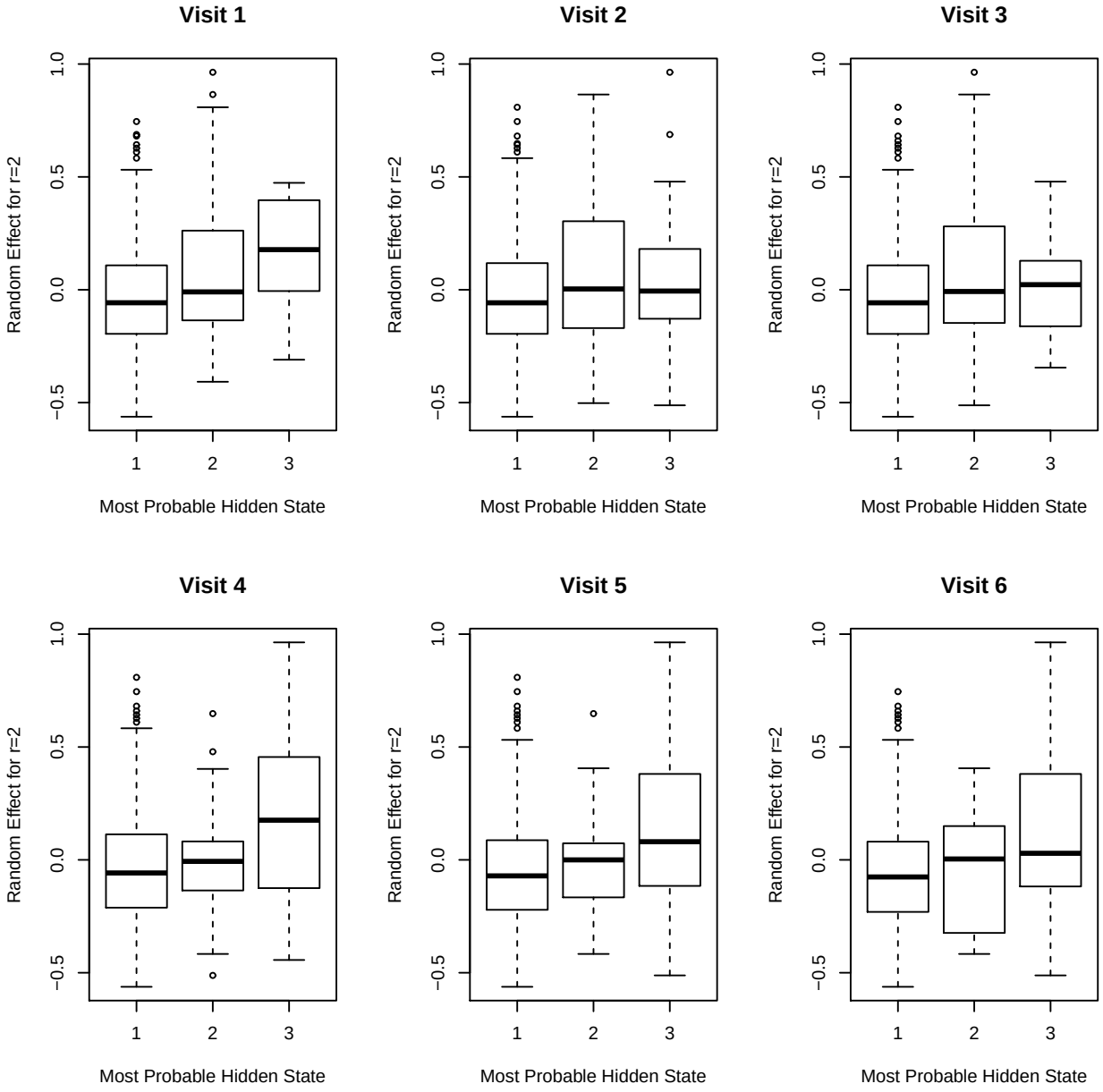

Figure 4.13: Box plot for the normal CO level response ($r = 2$) random effect at each study visit stratified by the most probable hidden state, $\hat{\delta}_{it}$.

1. The data are generated from a MHMM according to (4.1) and (4.2) with $\boldsymbol{\tau}_1 = [-3, 2, 1.5]$ and $\boldsymbol{\tau}_2 = [1.2, 0.5, 0.8]$. As was the case in Chapter 3, we will consider the case where the hidden states are generated as a first-order Markov Chain with a treatment effect in the transition probabilities such that, $\boldsymbol{P}^{(0)} = \begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.20 & 0.60 & 0.20 \\ 0.10 & \mathbf{0.10} & \mathbf{0.80} \end{bmatrix}$. and $\boldsymbol{P}^{(1)} = \begin{bmatrix} 0.85 & 0.10 & 0.05 \\ 0.20 & 0.60 & 0.20 \\ 0.10 & \mathbf{0.30} & \mathbf{0.60} \end{bmatrix}$, with $\boldsymbol{\pi} = [0.60, 0.25, 0.15]$. The random effects are $i.i.d$ zero mean multivariate normal with covariance matrix, $\boldsymbol{\Sigma} = \begin{bmatrix} 2.0 & 0.2 \\ 0.2 & 0.1 \end{bmatrix}$, and the inverse of the variance of the residual error in the second response is $\frac{1}{\sigma_\epsilon^2} = 10.0$

2. The data for the hidden states are generated as in situation (1) with the same transition probability matrices and initial probability vector. Conditional on the hidden states, we generated another latent (but unmodeled) source of variation, which contaminates the model, such that $D_{irt}|Z_{it} = k \overset{\text{iid}}{\sim} Bernoulli(\kappa_{rk})$, with $\kappa_{r1} = 0$, and $\kappa_{rk} = 0.1$ for $k > 1$ and $r = \{1, 2\}$, such that (4.1) and (4.2) is modified to have:

$$
\begin{aligned}
Y_{i1t}|Z_{it}, \boldsymbol{b}_i &\sim Poisson(\theta_{i1t}) \text{ where,} \\
\log(\theta_{i1t}) &= \boldsymbol{H}_{i1tk}^{'}\boldsymbol{\tau}_1 + \boldsymbol{x}_{i1t}^{'}\boldsymbol{\beta}_{1k} + \boldsymbol{w}_{i1t}^{'}\boldsymbol{b}_i + \log(a_{it}) + (-2)D_{i1t} \\
Y_{i2t}|Z_{it}, \boldsymbol{b}_i &\sim N(\theta_{i2t}, \sigma_\epsilon^2) \text{ where,} \\
\theta_{i2t} &= \boldsymbol{H}_{i2tk}^{'}\boldsymbol{\tau}_2 + \boldsymbol{x}_{i2t}^{'}\boldsymbol{\beta}_{2k} + \boldsymbol{w}_{i2t}^{'}\boldsymbol{b}_i + (-0.5)D_{i2t} \ .
\end{aligned}
$$

Note that the D_{irt} are sampled independently for each r conditional on Z_{it}, and the rest of the parameters are set according to the specification in situation (1).

The resulting parameters estimates from each of these situations under univariate and bivariate MHMMs are presented in Tables 4.3 - 4.4.

In Table 4.3 we examine situation (1), and note bias in many parameter estimates under both types of univariate response models, and sometimes this can be quite severe. In all parameters, the bivariate model, which incorporates both of the univariate models' responses exhibits less bias, and in some cases significantly less bias. The one exception

Table 4.3: Comparison of the bivariate and the two univariate models under simulation through 100 simulation runs for Situation (1)

Parameter	True Value	SR Count Only	CO Only	Bivariate Model
π_1	0.60	0.4953 (0.0880)	0.6105 (0.1514)	0.5919 (0.0498)
π_2	0.25	0.2930 (0.0820)	0.2290 (0.1417)	0.2509 (0.0504)
p^0_{12}	0.10	0.2149 (0.1118)	0.1823 (0.1748)	0.1052 (0.0457)
p^0_{13}	0.05	0.1063 (0.0595)	0.0700 (0.0466)	0.0548 (0.0211)
p^0_{21}	0.20	0.2857 (0.1115)	0.3260 (0.2064)	0.2240 (0.0984)
p^0_{23}	0.20	0.2580 (0.0939)	0.2716 (0.1790)	0.2344 (0.0870)
p^0_{31}	0.10	0.1224 (0.0500)	0.1196 (0.0689)	0.1053 (0.0413)
p^0_{32}	0.10	0.1447 (0.0688)	0.1385 (0.1180)	0.1166 (0.0509)
p^1_{12}	0.10	0.1681 (0.0643)	0.1691 (0.1616)	0.1000 (0.0282)
p^1_{13}	0.05	0.0940 (0.0382)	0.0647 (0.0369)	0.0515 (0.0132)
p^1_{21}	0.20	0.2385 (0.0813)	0.3196 (0.1724)	0.2038 (0.0552)
p^1_{23}	0.20	0.2329 (0.0703)	0.2193 (0.1215)	0.2109 (0.0440)
p^1_{31}	0.10	0.1213 (0.0585)	0.1322 (0.0838)	0.1066 (0.0342)
p^1_{32}	0.30	0.2882 (0.0753)	0.2670 (0.1176)	0.2942 (0.0483)
τ_{11}	-3.00	-3.4546 (0.4085)	-	-3.0245 (0.1947)
τ_{12}	2.00	2.0566 (0.2933)	-	2.0220 (0.1830)
τ_{13}	1.50	1.4987 (0.1393)	-	1.5047 (0.0928)
τ_{21}	1.20	-	1.2674 (0.0445)	1.2080 (0.0302)
τ_{22}	0.50	-	0.4125 (0.2364)	0.4795 (0.0572)
τ_{23}	0.80	-	0.7558 (0.2155)	0.7899 (0.0475)
Σ_{11}	2.00	2.4849 (0.3758)	-	1.9692 (0.2659)
Σ_{12}	0.20	-	-	0.2032 (0.0446)
Σ_{22}	0.10	-	0.1220 (0.0153)	0.1043 (0.0129)
$\frac{1}{\sigma^2_\epsilon}$	10.00	-	8.8300 (0.5682)	9.5031 (0.4939)

Each value is the average posterior mean (average posterior standard deviation) of the simulated sample. π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

would be that of the estimate for τ_{13}, where the univariate SR count only model has slightly smaller bias than the bivariate model. In both cases, the bias is very small, and not problematic in either case. These univariate results bear some resemblance to the bias we observed in some of the simulations conducted in Chapter 3 under the bivariate model. Further exploration of this bias, suggested that it was a function of how well-separated the hidden state modes of the observed response were, the amount of between-subject heterogeneity, and the number of observation times per subject. The results here, generally support these conclusions, although it is difficult to quantify the role of the mode separation and between-subject heterogeneity under these different univariate response models. The inflated between-subject heterogeneity as estimated through Σ_{11} and Σ_{22} is also interesting. This would be generally consistent with the results we presented in Table 4.1 where we saw the univariate SR only model describe the heterogeneity through the random effects, and not to nearly the degree expected through the hidden states. With the limited number of observation times per subjects and multiple sources of heterogeneity being modeled, this class of model seems to shrink the components of $\boldsymbol{P}$ toward the prior mean, and compensate by increasing the variance of the random effects. As we noted in Chapter 3, placing an informative prior on the components of $\boldsymbol{P}$ seemed to significantly reduce the amount of bias observed. Lastly, we note that under the bivariate model, the bias is significantly less. In addition to increasing the number of observation times (n_i), and using an informative prior, this study would also suggest that adding an additional response may assist in reducing any bias observed.

In Table 4.4 under situation (2), we note a similar difference between the the bivariate and univariate models. With the exception of p^0_{32} and τ_{12}, the bivariate model has lower bias than each of the univariate models. We generally note that when bias was observed in Table 4.3, we tend to see higher bias in Table 4.4 introduced by the contamination process. With this contamination being hidden state dependent, under the univariate model fits it is perhaps not surprising to see additional bias introduced to some of the the transition

Table 4.4: Comparison of the bivariate and the two univariate models under simulation through 100 simulation runs for Situation (2)

Parameter	True Value	SR Count Only	CO Only	Bivariate Model
π_1	0.60	0.5088 (0.1045)	0.6360 (0.0812)	0.6351 (0.0605)
π_2	0.25	0.2907 (0.0958)	0.2154 (0.0752)	0.2230 (0.0578)
p^0_{12}	0.10	0.2424 (0.1203)	0.2077 (0.0955)	0.1122 (0.0538)
p^0_{13}	0.05	0.1177 (0.0516)	0.0730 (0.0255)	0.0547 (0.0236)
p^0_{21}	0.20	0.3099 (0.1107)	0.3421 (0.0844)	0.2464 (0.0973)
p^0_{23}	0.20	0.2906 (0.1103)	0.2706 (0.0973)	0.2702 (0.1054)
p^0_{31}	0.10	0.1374 (0.0575)	0.1224 (0.0424)	0.1048 (0.0427)
p^0_{32}	0.10	0.1980 (0.0924)	0.1539 (0.0602)	0.1550 (0.0641)
p^1_{12}	0.10	0.2037 (0.0919)	0.1662 (0.0817)	0.0917 (0.0399)
p^1_{13}	0.05	0.1022 (0.0444)	0.0644 (0.0195)	0.0526 (0.0143)
p^1_{21}	0.20	0.2614 (0.0980)	0.3380 (0.0795)	0.2267 (0.0695)
p^1_{23}	0.20	0.2348 (0.0851)	0.2469 (0.0774)	0.2302 (0.0574)
p^1_{31}	0.10	0.1539 (0.0616)	0.1528 (0.0536)	0.1260 (0.0373)
p^1_{32}	0.30	0.3160 (0.0805)	0.2844 (0.0798)	0.3088 (0.0604)
τ_{11}	-3.00	-3.5384 (0.5381)	-	-2.9805 (0.2382)
τ_{12}	2.00	1.9934 (0.4420)	-	1.9324 (0.2245)
τ_{13}	1.50	1.5799 (0.1357)	-	1.5371 (0.0998)
τ_{21}	1.20	-	1.2775 (0.0267)	1.2345 (0.0278)
τ_{22}	0.50	-	0.3812 (0.1350)	0.4578 (0.0927)
τ_{23}	0.80	-	0.7673 (0.1096)	0.7568 (0.0834)
Σ_{11}	2.00	2.5681 (0.4148)	-	2.0325 (0.3065)
Σ_{12}	0.20	-	-	0.2048 (0.0497)
Σ_{22}	0.10	-	0.1177 (0.0128)	0.1034 (0.0116)
$\frac{1}{\sigma^2_\epsilon}$	10.00	-	8.4617 (0.4531)	8.4754 (0.4269)

Each value is the average posterior mean (average posterior standard deviation) of the simulated sample. where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

probabilities, while the other modeled sources of heterogeneity (Σ_{11}, Σ_{22} and σ_ϵ^2) have all increased. With the exception of a few of the placebo transition probabilities (e.g., p_{23}^0 and p_{32}^0), both of which are associated with the contaminated hidden states, and $\frac{1}{\sigma_\epsilon^2}$, the bias in the bivariate MHMM is relatively small, and is reassuring, that even when the data are contaminated by some unmodeled process, the resulting estimates are generally well-behaved.

Lastly, we present some results pertaining to hidden state classification from the two simulation situations for each of the univariate and bivariate models in Table 4.5. The results are generally what one would expect, with the bivariate models generally producing better hidden state classification results than the univariate models. Further, classification was more efficient in the situation without the contaminated data, but still performed remarkably well in spite of this. In general, classification of observations in the intermediate state (hidden state 2) was most difficult, irrespective of model or contamination. We additionally stratified the analysis of situation (2) on whether the observation was contaminated. In the non-contaminated observations, we found that the results were slightly better than the overall AUC assessments. In the contaminated observations, we found that the $\hat{\phi}_{it2}$ is generally a poor predictor for hidden state two as measured by AUC in all analyses. Surprisingly, the bivariate model does slightly worse than the univariate CO model, but this is primarily caused by the fact that the bivariate analysis is 'doubly' contaminated by both responses, but the univariate models are conducted in the subset of observations which is contaminated in either of the responses, including the one it is not considering. When we further restrict the analysis to those observations which are contaminated within their respective responses, and remove from this analysis those that are contaminated in the other response, the mean AUC_2 C reduces to 0.3986 and 0.4626 in the SR and CO univariate models, respectively.

Table 4.5: Hidden state classification results for the simulations

Measure	SR Univariate (1)	SR Univariate (2)	CO Univariate (1)	CO Univariate (2)	Bivariate (1)	Bivariate (2)
AUC_1	0.7811 (0.0203)	0.7523 (0.0246)	0.8885 (0.0184)	0.8743 (0.0166)	0.9205 (0.0104)	0.9021 (0.0130)
AUC_2	0.6317 (0.0288)	0.6072 (0.0343)	0.7190 (0.0610)	0.7022 (0.0483)	0.8305(0.0174)	0.7893 (0.0254)
AUC_3	0.8398 (0.0173)	0.8065 (0.0217)	0.9688 (0.0053)	0.9613 (0.0062)	0.9787 (0.0050)	0.9671 (0.0070)
AUC_1 NC	-	0.7687 (0.0241)	-	0.8866(0.0165)	-	0.9135 (0.0119)
AUC_2 NC	-	0.6144 (0.0397)	-	0.7145 (0.0609)	-	0.8122 (0.0269)
AUC_3 NC	-	0.8396 (0.0209)	-	0.9710 (0.0058)	-	0.9793 (0.0051)
AUC_2 C	-	0.5204 (0.0696)	-	0.5908 (0.1682)	-	0.5381 (0.1135)
AUC_3 C	-	0.6579 (0.0444)	-	0.8897 (0.0267)	-	0.8963 (0.0304)
Agreement	0.6350 (0.0702)	0.6133 (0.0808)	0.7328 (0.0234)	0.7205 (0.0184)	0.7938 (0.0150)	0.7647 (0.0189)

AUC results from using the true hidden simulated states as labels with $\hat{\delta}_{it}$ and $\hat{\phi}_{itk}$ as classifiers

AUC_k area under the ROC curve for true hidden state k and $\hat{\phi}_{itk}$ NC non-contaminated observations; C contaminated observations; Agreement $= \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{t=1}^{n_i} \mathbb{1}(\hat{\delta}_{it} = m, Z_{it} = m)$

4.7 Discussion

Overall, in this chapter we examined the potential utility of using a multivariate approach to modeling data arising from data where heterogeneity in the responses can be attributed to between-subject differences and to transition between disease states. We focused on two areas of concern. The first pertains to the appropriateness of a MHMM multivariate modeling approach to model data from settings such as a smoking cessation clinical trial. The second concern is related to suitability of conducting inference about the hidden states, and whether such inferences are meaningful. Overall, we find that use of a multivariate model has several advantages over the use of separate univariate models under the same MHMM framework. Further, we discover that both the locally decoded hidden state probabilities ($\hat{\phi}_{itk}$), along with most probable hidden state ($\hat{\delta}_{it}$) closely correspond with the study-defined disease state definitions, and illustrate possible deficiencies for these definitions.

Although we described in detail in Chapter 3 a modeling framework for multivariate longitudinal response data under MHMM, it was not clear what benefits such an approach would have over univariate or other modeling approaches. As we saw in section 4.3 there are some clear benefits of this approach to modeling data such as those generated from our motivating smoking cessation example. When comparing the model estimates generated from the univariate SR and bivariate MHMMs, the hidden states in the bivariate case clearly have more meaning, when compared to the univariate SR MHMM, which models most of the heterogeneity in the SR response as being attributable to between-subject differences. This result is generally observed in the simulation results in section 4.6, where we often observe bias in the placebo group's probability transition matrix, $\boldsymbol{P}^{(\mathbf{0})}$, and hidden state dependent intercepts, $\boldsymbol{\tau}$. We think this partly due to the limited number of observation times available for each patient, but note that the bivariate model's estimates are generally significantly less biased than either of the SR or CO univariate response models. Lastly, under a multivariate MHMM there are some significant improvements

of the posterior distribution of the random effects, which are severely skewed in the univariate SR case, but is much less so under a multivariate approach. The asymmetry we observe in the SR response random effect, is primarily caused by a very homogeneous set of observations from patients self-reporting complete abstinence throughout the entire six-week study period. We believe that if the study were continued for a while longer, some of these subjects would relapse, and this would reduce some of the asymmetry noted. Alternatively, other non-Gaussian or mixture model approaches could be used, but also have limitations. Non-parametric approaches have been used to model the random effects in univariate MHMMs (Maruotti and Ryden, 2009), but have been generally limited to univariate random effects. Under our current approach of separate but correlated random effects, this may prove more difficult. The second approach involving mixture models may be more promising, but one would likely want to consider the possibility that the mixtures were dependent on the hidden states we have been modeling. With the limited sample size in this dataset, we believe this would also be difficult, especially considering we believe the random effect distribution asymmetry is caused by primarily a very homogeneous set of observations to begin with. The results in section 4.3 along with the simulation results of section 4.6 demonstrate the advantages of using a multivariate model over the component univariate models.

The second concern pertains to hidden state inference. In general, the modeling framework we have outlined in Chapters 3 and 4 is going to describe the heterogeneity in the response under the model assumptions. The hidden states need not describe any specific disease or other discrete states, but in general, both the plausibility and utility of the modeling approach will be improved greatly if they have some kind of interpretation. Often, as was the case in the motivating smoking cessation clinical trial, study disease states will be defined by these longitudinal responses, instead of being viewed as the disease states generating the longitudinal response, which would ideally be more appropriate. In a case such as this, there is no gold standard as to which to calibrate the hidden states to, as

the corresponding longitudinal response used to define the study's disease states are prone to error. With that said, a general agreement between the hidden states and the study-defined disease states would be ideal. We do see such an agreement, with very high areas under the ROC curve, and a generally high overall agreement, when including or excluding the intermediate second hidden state which is not study-defined. When we look at cases of disagreement, we see situations which may be illustrative to inform further work to improve our model, or provide guidance to conduct future smoking cessation studies. For the most part, the MHMM approach is more sensitive to relative changes in CO levels within a subject, and less sensitive to small isolated cases of SR smoking (most often a single cigarette in a week) when contrasted with the study-defined definitions which are very strict about any SR smoking, and do not consider relative changes in CO levels, with only absolute values exceeding the threshold defining smoking. The differences may illustrate future work that can be done to improve modeling approaches, by modeling directly a stricter definition of SR smoking by explicitly defining hidden state one as having a SR rate of 0. Alternatively, this may also provide some guidance for future studies which may improve study-defined disease outcomes which use relative changes in CO, or less stringent definitions of SR smoking, when isolated or small relapses occur.

With that said, the agreement of the hidden states done by either $\hat{\delta}_{it}$ or $\hat{\phi}_{itk}$ and the study-defined disease states is quite high. This is quite impressive given that this modeling approach was largely an unsupervised task, where the data were used to generate the hidden states, without any explicit labelling of observations. Further, we have been able to model more effectively an intermediate state which has been hypothesized to exist for subjects attempting to cease smoking, and may be important in describing a treatment effect. While other methods have been developed to explain heterogeneity in a longitudinal response, these approaches have generally used approaches which classify entire trajectories on one response or several responses of the same type into static groups (e.g., Lin et al. (2000)). In many settings static groups will be appropriate, but in other settings such as

smoking cessation, it seems unlikely that this approach would work well. Moreover, in the general setting, these approaches have been further limited by their ability to use responses of different data types, such as the discrete and continuous data we have encountered throughout Chapters 3 and 4.

Overall, we believe the multivariate longitudinal data modeling framework developed in Chapter 3 provides a good basis to conduct inference, where the hidden states closely correspond to well understood disease states. Using a multivariate MHMM over the univariate alternatives allows one to conduct inference more efficiently, and generally get more meaningful results.

Chapter 5

Using MHMMs for Prediction and other Missing Data Issues

5.1 Background

In Chapter 3 we described a modeling framework which could describe heterogeneity in multivariate longitudinal response data from between-subject differences in addition to a dynamic hidden process. In Chapter 4, we assessed the utility of inference drawn from the hidden states from these models to describe measurable clinical outcomes, and identified several areas where the models may provide more insightful inference about the true clinical outcome of interest instead of the measurable one. However, thus far we have largely ignored two important issues encountered frequently in longitudinal data analysis. The first issue we will examine will focus on the prediction of hidden states. The second topic is somewhat related, and is related to justification and implementation of a method to account for patient dropout in such studies. We will examine both of the issues in the context of the same motivating clinical trial discussed throughout the previous chapters.

5.2 Prediction of Hidden States

5.2.1 Methods

In Chapter 4, we described one way to conduct inference about the hidden states under a MHMM, by using the locally decoded probabilities,

$$\hat{\phi}_{itk} = P(Z_{it} = k|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta}) = \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}(Z_{it}^{(j)} = k) \ , \tag{5.1}$$

where, Z_{it} is the hidden state for subject $i = 1, \ldots, N$, at time $t = 1, \ldots, n_i$, $\boldsymbol{y}_i$ is the complete set of R longitudinal responses for subject i over all t, $\boldsymbol{b}_i$ is a vector of random effects for subject i, $\boldsymbol{\Theta}$ comprises all population-level parameters of the MHMM, and J are the total number of samples generated from the MCMC algorithm described in Chapters 3 and 4.

Prediction of the hidden states in the future can take several forms, but will generally be some function of the current hidden states (or their decoded probabilities) and the probability transition matrix for the hidden Markov process, $\boldsymbol{P}$. For example, one estimator for $P(Z_{i,t+h} = k|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ based on the MCMC output would be,

$$\hat{\omega}_{itk} = P(Z_{i,t+h} = k|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta}) = \frac{1}{J}\sum_{j=1}^{J} q_{z_{it}k}^{(j)} \ , \tag{5.2}$$

where $q_{lk}^{(j)}$ is the $(l,k)^{th}$ entry of $\boldsymbol{P}^h$ for the jth MCMC sample. If $t + h \leq n_i$, and the data are complete, then one is conditioning on $\boldsymbol{y}_{i,t+h}$, and the situation is not of much direct interest, as inference would be more efficiently conducted through $\hat{\phi}_{itk}$. If $t+h > n_i$, then a more interesting situation would occur and could be utilized when one was trying to predict post-follow-up hidden state distributions, or impute hidden states from data which was missing resulting from study dropout. To illustrate this directly, we will evaluate $\hat{\omega}_{itk}$ as a predictor for the the study-defined smoking definition of study visit $t + 1$.

5.2.2 Results

As mentioned previously, using $\hat{\omega}_{itk}$ to predict hidden states at time points which are conditioned on is not so interesting in most situations. In order to evaluate $\hat{\omega}_{itk}$ for time points in the future, we will fit the bivariate MHMM model discussed in Chapter 4 to a subset of the data considered in those models. In this analysis, we will run the model fitting procedure by MCMC restricting the observation times to only the first five weeks, instead of six. This analysis is presented in Table 5.1 alongside the estimates provided in Table 4.1 from Chapter 4.

In general, the subset analysis involving a maximum of five observation times, has similar estimates to those provided in the previously performed analysis involving all six study visits. In Table 5.1 we obtain what we would expect, that in general the analysis based on only five time points generally has larger posterior standard deviation, representing more uncertainty with less data.

We then use the subset model involving only five observation times, to predict the hidden states in study visit six in the 250 subjects with the available data to perform these analyses. Likewise to illustrate the differences, we also present the complete data (all six observation times) analysis, which we would expect to have slightly higher predictive power, based on the fact that one is conditioning on the observation one is trying to in essence base prediction on.

As seen in Figure 5.1-5.2, the predictive value of $\hat{\omega}_{i6k}$ when conditioned on the longitudinal response at $t = 6$ is higher than that of the subset analysis which only includes the first five observation times. Overall, the ability to do prediction when no information is known about future observations, is still quite good, with both AUCs for $\hat{\omega}_{i61}$ and $\hat{\omega}_{i63}$ exceeding 0.9.

Table 5.1: Comparison of full observation model and the five week model

Parameter	Bivariate Model Full Obs. Time.	Bivariate Model Max(n_i) = 5 Subset
π_1	0.7130 (0.0297)	0.7228 (0.0314)
π_2	0.2146 (0.0289)	0.1932 (0.0304)
p^0_{12}	0.0442 (0.0174)	0.0511 (0.0214)
p^0_{13}	0.0136 (0.0077)	0.0182 (0.0102)
p^0_{21}	0.2072 (0.0719)	0.1888 (0.0783)
p^0_{23}	0.1745 (0.0598)	0.2129 (0.0746)
p^0_{31}	0.1241 (0.0786)	0.2201 (0.1159)
p^0_{32}	0.0877 (0.0796)	0.1973 (0.1356)
p^1_{12}	0.0349 (0.0084)	0.0348 (0.0096)
p^1_{13}	0.0083 (0.0037)	0.0119 (0.0050)
p^1_{21}	0.2515 (0.0486)	0.2688 (0.0550)
p^1_{23}	0.2271 (0.0437)	0.2124 (0.0484)
p^1_{31}	0.0650 (0.0307)	0.0900 (0.0409)
p^1_{32}	0.2007 (0.0501)	0.2429 (0.0640)
τ_{11}	-5.2899 (0.2120)	-5.1700 (0.2257)
τ_{12}	4.4679 (0.2035)	4.4182 (0.2017)
τ_{13}	1.3101 (0.0434)	1.2419 (0.0447)
τ_{21}	1.1851 (0.0221)	1.1895 (0.0227)
τ_{22}	0.2959 (0.0474)	0.2816 (0.0529)
τ_{23}	0.8782 (0.0569)	0.8216 (0.0529)
Σ_{11}	2.3318 (0.4179)	2.2239 (0.4090)
Σ_{12}	0.1242 (0.0459)	0.1258 (0.0463)
Σ_{22}	0.1048 (0.0111)	0.1023 (0.0113)
$\frac{1}{\sigma^2_\epsilon}$	8.6795 (0.3676)	8.7250 (0.4158)

Each value is the posterior mean (posterior standard deviation). where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

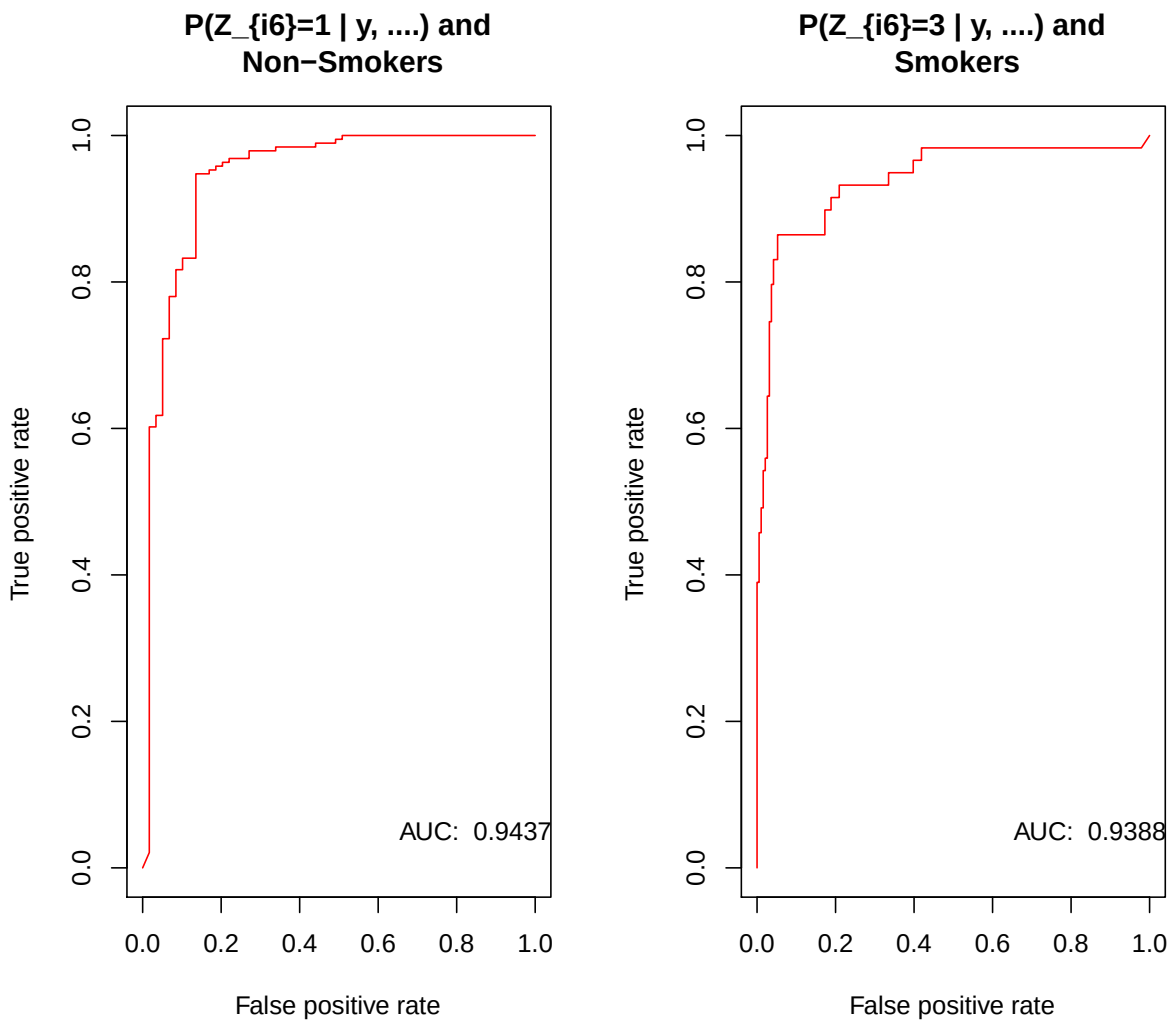

Figure 5.1: ROC curves for $\hat{\omega}_{i61}$ and $\hat{\omega}_{i63}$ in the complete data analysis

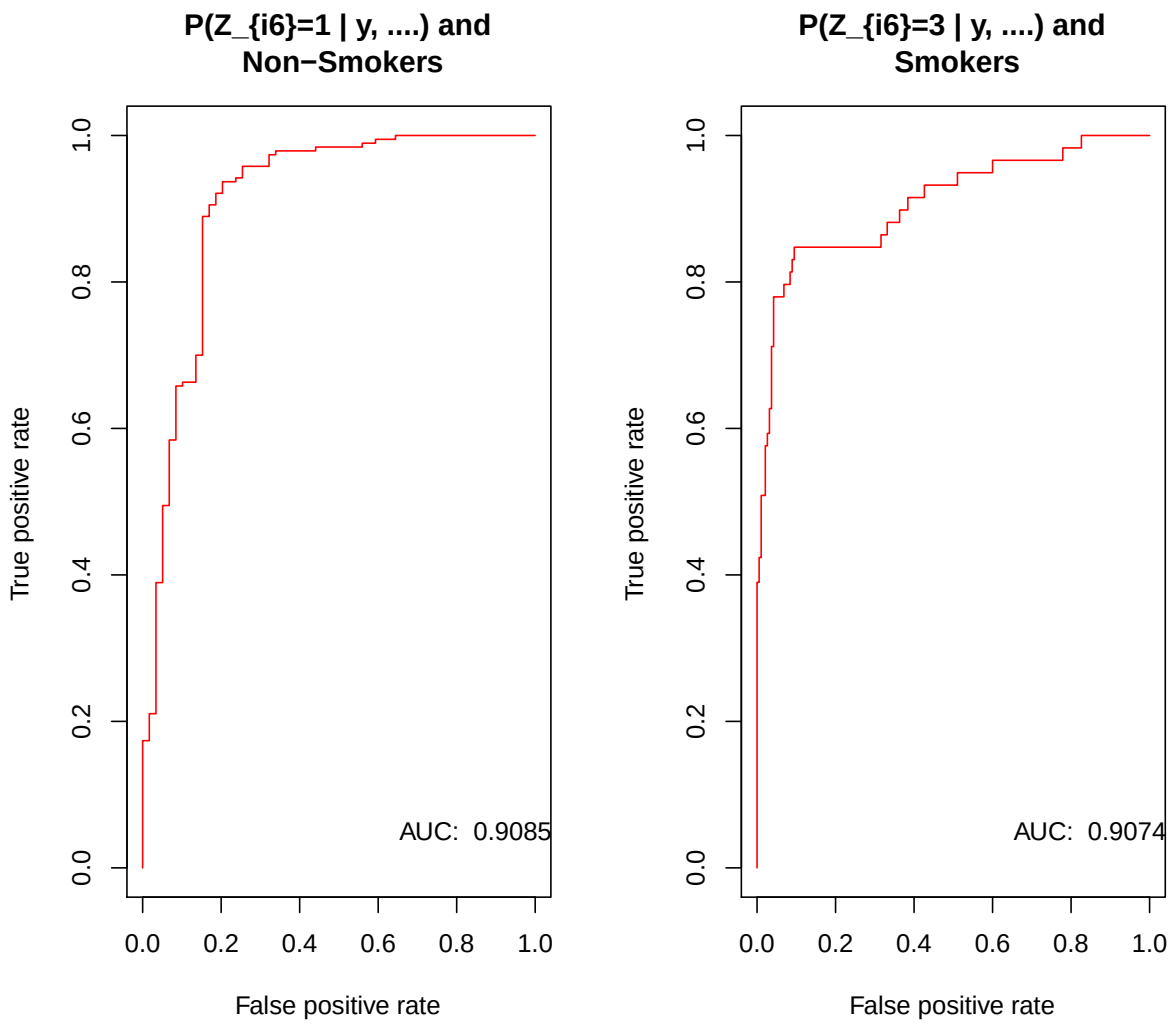

Figure 5.2: ROC curves for $\hat{\omega}_{i61}$ and $\hat{\omega}_{i63}$ in the five observation time subset analysis

5.3 Missing Data Issues

As mentioned in section 5.1, the motivating smoking cessation study had a somewhat significant number of subjects who were enrolled in the study who dropped out before the end of the six week treatment period. In Figure 5.3 we present the patient disposition for the study, where 385 subjects were randomized to receive placebo or one of three active treatments. Of these 385, 31 had no evaluable post-treatment initiation data, or data which contained only one part of the bivariate response we were considering for the MHMM models. This would occur when the patient submitted their first week's SR data (or part thereof), or had their CO measurements taken at the first post-treatment visit, but did not complete the other component. While this information may be useful in some contexts, in both of these cases, we have excluded these patients from analysis in Chapters 3-5, because the data was very limited to draw inference from if it existed at all. The patients may play an important role in the study of smoking cessation treatment, but they generally will not be able to help us conduct inference in the models we are considering.

Of the remaining 354 subjects we have considered in Chapters 3-5, we had 250 subjects with complete data. Of the 104 subjects who did not have complete data, 50 had one or more intermittently missing data. In many instances, these data were recorded as missing due to a technical or other problem, and we therefore believe the mechanism by which they are missing is largely ignorable. Not including the 31 subjects who dropped out near or at week one with no evaluable data, we had an additional 54 subjects who made it through just 37 days or fewer of the 42-day (6-week) treatment period.

While the problems with missing data extend to the intermittently missing data, we will primarily focus on those 54 subjects who dropped out prior to the end of the study. In those subjects, who had intermittent data, as others point out (Altman, 2007), under a missing at random (MAR) assumption, the only modification that needs to be made to the methods described in Chapters 3 and 4 is to how the hidden states ($\boldsymbol{Z}$) are sampled. If observations are missing between times t and $t^{'}$ with $1 \leq t^{'} < t \leq n_i$, then the forward probabilities

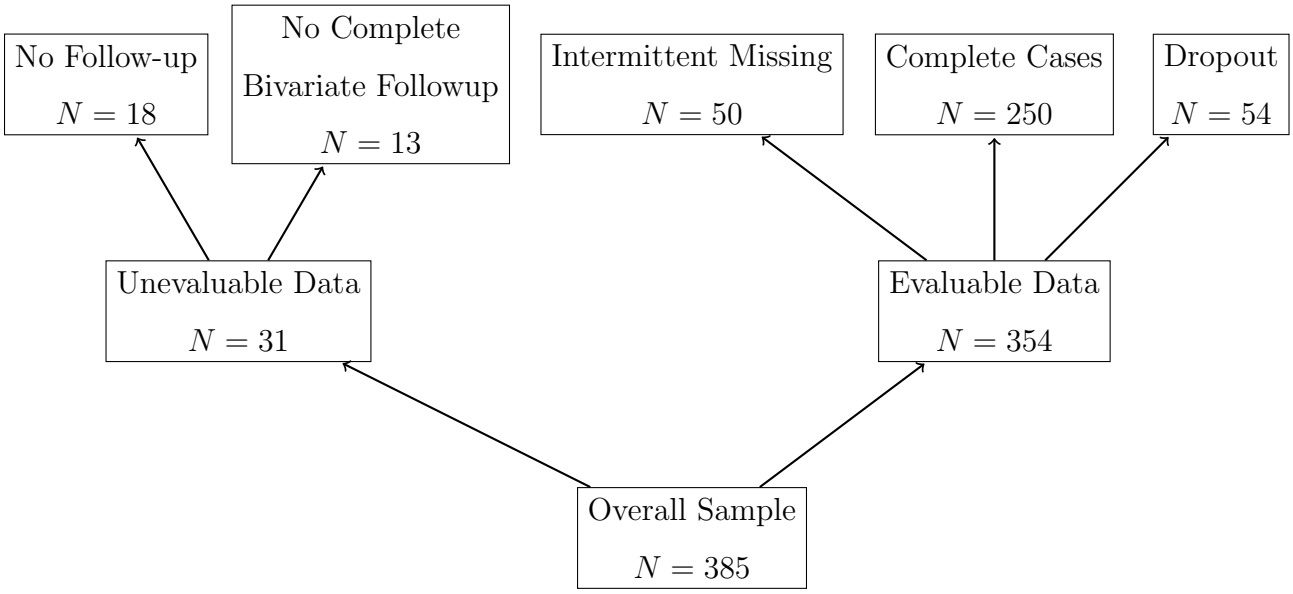

Figure 5.3: Patient disposition for the smoking cessation study

become, $\boldsymbol{\alpha_{it}} = \boldsymbol{\alpha}_{i,t'}\boldsymbol{P}^h\boldsymbol{G}(\boldsymbol{y}_{it})$ where, $h = t - t'$, and $\boldsymbol{P}^h$ is the $h-$step transition matrix, and $\boldsymbol{G}(\boldsymbol{y}_{it})$ is as defined in Chapter 3. Similarly, the stochastic recursion used to update the hidden states can be modified in a similar fashion. There is sufficient reason to believe that the intermittently missing data are missing at random, as very often the missingness was recorded being due to some technical error, or other external issue unrelated to the subject's responses. On the other hand, the missing data mechanism for those subjects which dropped out, is less clear, and will be focus of further analysis.

Often subjects who drop out are assumed to be smokers, and analyses proceed under this assumption, even when for instance, evidence of long term abstinence at 3 or 6 months is obtained. Others (Borrelli et al., 2002) have cited concerns that those individuals who drop out may be different in a number of ways when compared to either smokers or non-smokers. One way of assessing these hypotheses, would be to examine the distributions of $\hat{\phi}_{itk}$ for each k at the time of dropout. If the distributions of $\hat{\phi}_{itk}$ stratified by those who completed the study, and those who dropped-out differed in many respects, then

this may provide some evidence that the subjects who dropped out were a distinguishable group, and the dropout mechanism is not ignorable. If however, the distributions of $\hat{\phi}_{itk}$ were approximately similar to each other, then this would provide some evidence that conditional on Z_{it} at the time of dropout, the dropout mechanism may be ignorable. In Figures 5.4-5.6, we plot histograms of $\hat{\phi}_{itk}$ in all observations (overall), in all observations at $t = 6$, and in those observations at the time of dropout.

In Figures 5.4-5.6 we note a common "U"-shaped curve throughout, with most of the values of $\hat{\phi}_{itk}$ being at $[0, 1]$ interval boundaries. While there are a relatively higher number of observations with $\hat{\phi}_{itk}$ values between 0.1 and 0.9 for $k = 1$ or 2, the sample sizes are relatively small, and it is difficult to attribute these. While the shapes of these distributions are generally the same, this does not mean that the hidden state distributions are identical for each of the groups considered. For instance, for the $t = 6$ observations, the mean values of $\hat{\phi}_{itk}$ are 0.787, 0.0767, and 0.136 for $k = 1$, 2, and 3, respectively. At the dropout time, the mean value of $\hat{\phi}_{itk}$ is 0.574, 0.214 and 0.212 for $k = 1$, 2, and 3, respectively.

In addition to the discussion related to the hidden states at the time of dropout, there may be additional justification for a MAR assumption for the subjects who dropped out. In an unpublished analysis (Dubin, 2012), subjects with low adherence to therapy in the previous week, or with moderate to severe adverse events were found more likely to dropout. While some of the adverse events may be due to the treatment, many are related to withdrawal symptoms or craving, and many of these may be modeled effectively by the hidden states. This along with the analysis of the distributions of $\hat{\phi}_{itk}$ may make a MAR assumption reasonably plausible. Further justification or exploration of the missing data mechanism is probably merited, and should be undertaken in this and other datasets.

5.3.1 Methods

Under an ignorable missing data mechanism, we can modify the methodologies we have described previously in Chapters 3 and 4. Without loss of generality, but primarily be-

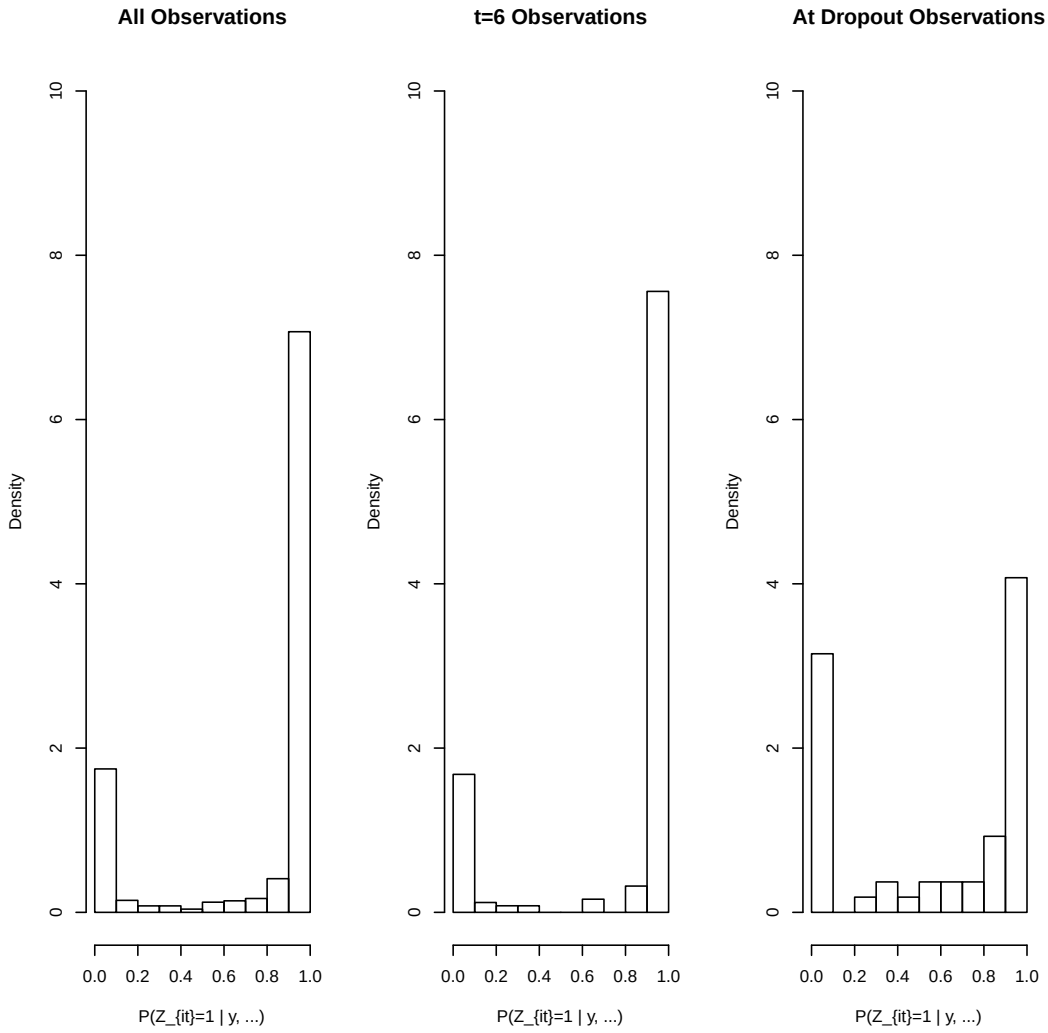

Figure 5.4: Histogram of the distribution of $\hat{\phi}_{it1} = P(Z_{it} = 1|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ overall, at $t = 6$ and at dropout.

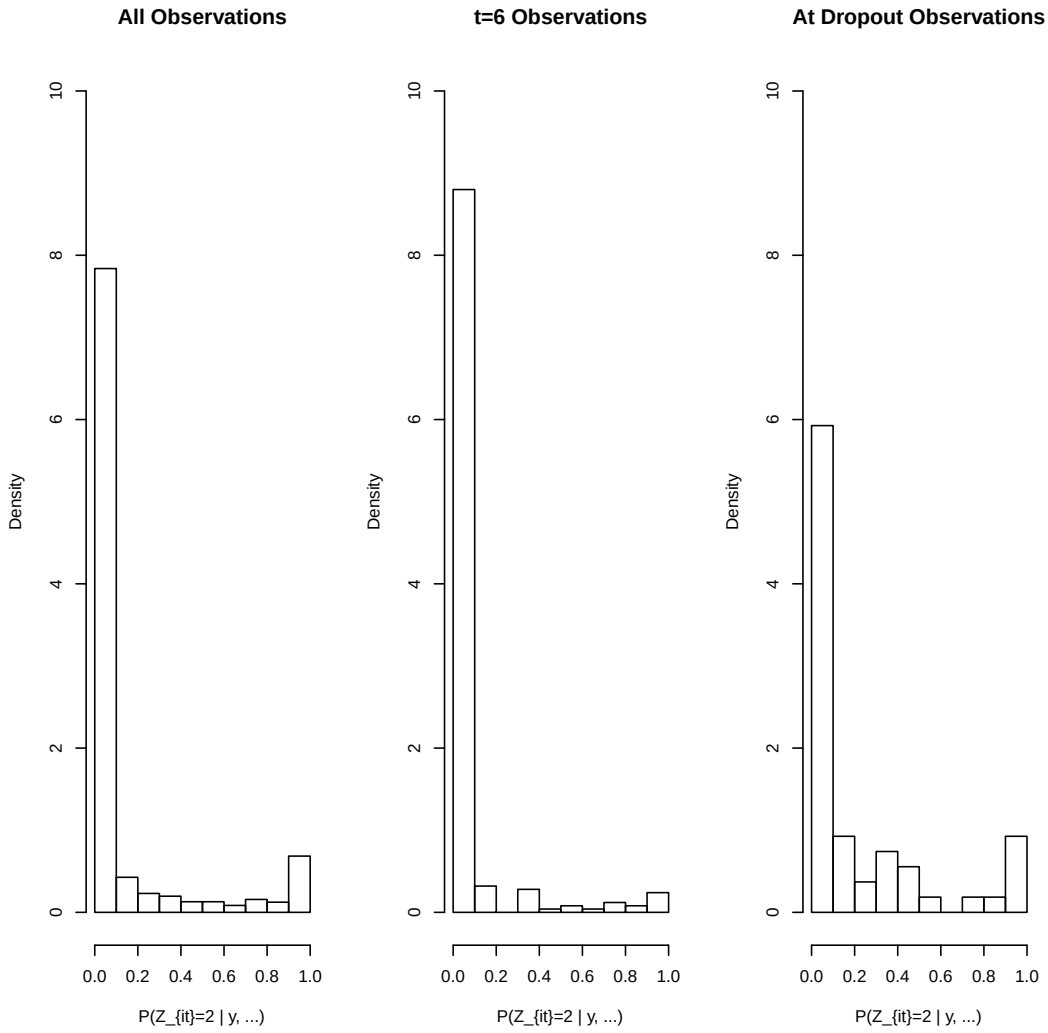

Figure 5.5: Histogram of the distribution of $\hat{\phi}_{it2} = P(Z_{it} = 2|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ overall, at $t = 6$ and at dropout.

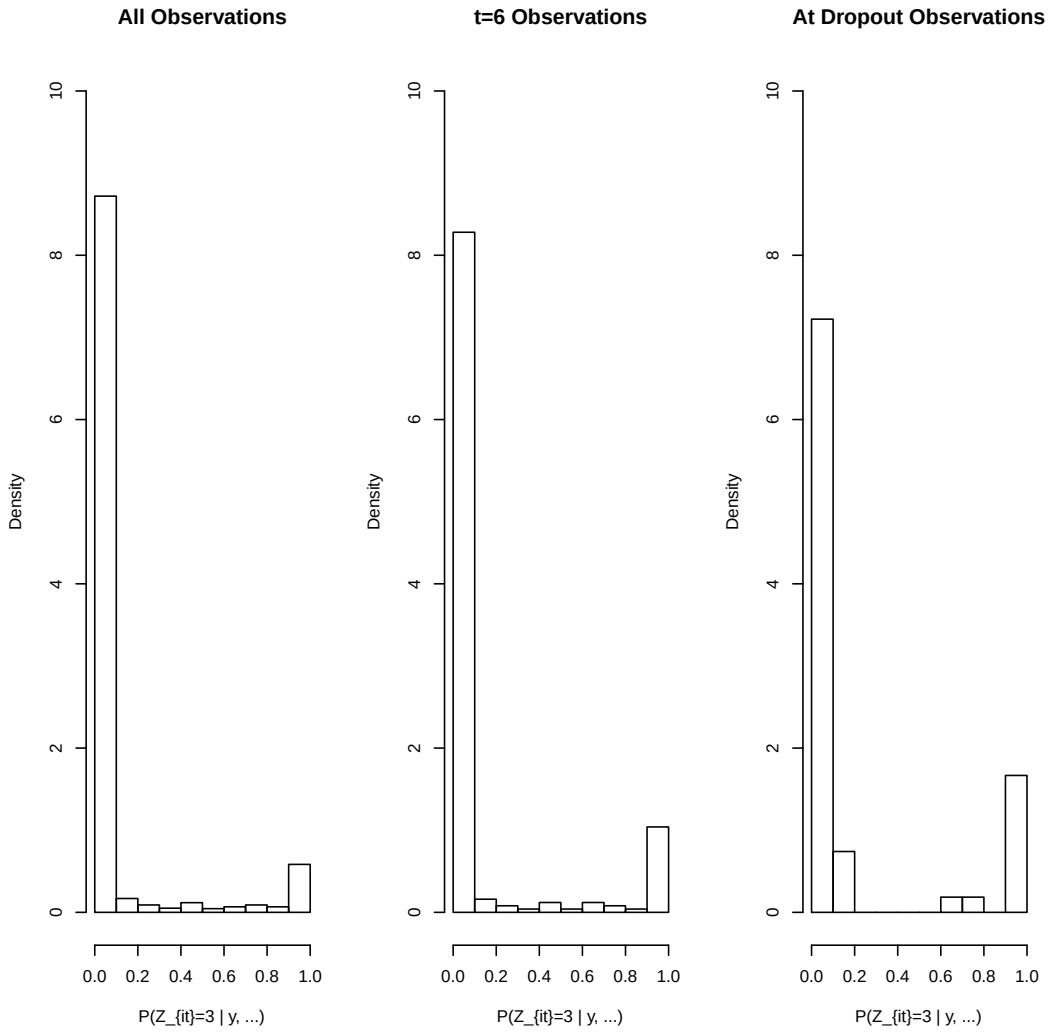

Figure 5.6: Histogram of the distribution of $\hat{\phi}_{it3} = P(Z_{it} = 3|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\Theta})$ overall, at $t = 6$ and at dropout.

cause our focus is on those subjects who drop out of the study, assume that there are no intermittently missing data, and the missing data are only generated from those subjects who drop out of the study. Let $\boldsymbol{y}_i^{OBS}$ be the observed data for the multivariate longitudinal response for subject i observed up to time n_i, and let $\boldsymbol{y}_i^{MISS}$ be the missing data for the multivariate longitudinal response for subject i, such that $\boldsymbol{y}_i^C = [\boldsymbol{y}_i^{OBS}, \boldsymbol{y}_i^{MISS}]$ are both combined, and refer to the complete longitudinal response data. Similarly, for the same common data points as the longitudinal response, let $\boldsymbol{z}_i^{OBS}$ be the hidden states for all observations for $t = 1, \ldots, n_i$, and let $\boldsymbol{z}_i^{MISS}$ be the missing data for the hidden states for subject i, such that $\boldsymbol{z}_i^C = [\boldsymbol{z}_i^{OBS}, \boldsymbol{z}_i^{MISS}]$ are both combined, and refer to the complete hidden state data. Assume, that $V > n_i$ for all i, where V is the maximum follow-up for all subjects.

Ideally, we would like to sample from $[\boldsymbol{\Theta}, \boldsymbol{b_i}|\boldsymbol{y}^C]$, and conduct our inference from the posterior conditional on the complete data. Since we do not have access to the completed data, we can impute the missing data to complete it, and approximate sampling from this posterior. To do this, we add to the approach described in Chapters 3 and 4 to sample from $[\boldsymbol{z}^{MISS}|\boldsymbol{z}^{OBS}, ...]$ and $[\boldsymbol{y}^{MISS}|\boldsymbol{y}^{OBS}, ...]$. This is relatively straightforward if the missing data mechanism is ignorable, and is based on the conditional independence assumptions contained in the modeling framework. At each iteration, $[\boldsymbol{z}_i^{OBS}|\boldsymbol{y}_i^{OBS}, \boldsymbol{\Theta}, \boldsymbol{b}_i]$ is sampled as described in section 3.3.2 using the observed longitudinal response data only. We note that, $[\boldsymbol{z}_i^{MISS}|\boldsymbol{z}_i^{OBS}, \boldsymbol{y}_i^{OBS}, \boldsymbol{\Theta}, \boldsymbol{b_i}] = [\boldsymbol{z}_i^{MISS}|\boldsymbol{z}_i^{OBS}, \boldsymbol{P}]$, and $Z_{it'}$ $(n_i < t' \leq V)$ can be sampled from $P(Z_{it'} = k|Z_{i,t'-1} = j) = p_{jk}^{(s)}$, the $(j,k)^{th}$ entry from the sth treatment group's transition matrix, and then iteratively for times $t'+1, \ldots, V$, conditioning on the previous sampled state. Once $\boldsymbol{z}_i^{MISS}$ has been sampled, we can sample from $[\boldsymbol{y}_i^{MISS}|\boldsymbol{y}_i^{OBS}, \boldsymbol{z}_i^C, \boldsymbol{\Theta}, \boldsymbol{b}_i]$ which can be done for each response independently, the components of $\boldsymbol{y}_i^{MISS}$ can be sampled for time point $t > n_i$ such that $Y_{irt}|Z_{it} = k \sim f_{rk}(y_{irt}|\boldsymbol{\Theta})$, where $f_{rk}(.)$ is the density or mass function for response r in hidden state k. At each iteration of sampling from the posterior by Gibbs sampling, $\boldsymbol{y}^{MISS}$ and $\boldsymbol{z}^{MISS}$ can be generated to complete

their respective datasets after which $\boldsymbol{y}^C$ and $\boldsymbol{z}^C$ can be used to sample from each of the other full conditionals: $[\boldsymbol{\tau}, \boldsymbol{\beta}|\boldsymbol{y}^C, \boldsymbol{b}, \boldsymbol{z}^C]$, $[\boldsymbol{P}|\boldsymbol{z}^C]$, $[\boldsymbol{\pi}|\boldsymbol{z}^C]$, $[\boldsymbol{\Sigma}|\boldsymbol{b}]$, $[\frac{1}{\sigma_\epsilon^2}|\boldsymbol{y}^C, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{b}, \boldsymbol{z}^C]$ and $[\boldsymbol{b}|\boldsymbol{y}^C, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{b}, \boldsymbol{z}^C, \boldsymbol{\sigma_\epsilon^2}]$.

We used this methodology to impute the missing data for each of the binomial and Poisson count models considered in Chapter 3 and 4, respectively. Those results will then be compared to the complete case and the available data analyses.

5.3.2 Results

As one can seen in Tables 5.2 and 5.3, the results from the complete case, available data and imputed missing data do not differ substantially, in either models based on the SR count by a binomial (number of days) model, or a Poisson (cigarette count) model. As would be expected, the complete case analysis generally has the highest posterior standard deviation, given it has the smallest number of observations and subjects used. In general, the imputed missing data analysis has a slightly higher posterior standard deviation in most parameter estimates, representing increased uncertainty due to the missing data.

5.4 Discussion

Overall, we have seen that prediction of the hidden states can be done in a similar fashion to how classification of hidden states was done in Chapter 4. Prediction based on estimators such as $\hat{\omega}_{itk}$ is relatively effective in discriminating between future smoking and non-smoking observations as determined by the study-definition, at least for the subsequent week. Further analysis of properties of $\hat{\omega}_{itk}$ to assess its prediction over longer periods of time, is an area of future investigation. Generalizing these predictions of longer-term abstinence may be possible, but we believe it to be much more difficult, as the transition probability matrix and perhaps even the Markov assumption may not be be sustained after the observation/treatment period of six weeks is completed. To illustrate this, the sta-

Table 5.2: Bivariate MHMM results for three different missing data scenarios under the binomial count SR model

Parameter	Complete Cases $N = 250$	Available Data $N = 354$	Imputed Missing Data $N = 354$
π_1	0.7662 (0.0459)	0.7569 (0.0408)	0.7495 (0.0422)
π_2	0.1862 (0.0439)	0.1787 (0.0386)	0.1829 (0.0405)
p^0_{12}	0.0273 (0.0168)	0.0234 (0.0140)	0.0236 (0.0143)
p^0_{13}	0.0099 (0.0069)	0.0124 (0.0074)	0.0132 (0.0078)
p^0_{21}	0.3734 (0.1191)	0.2799 (0.0978)	0.2506 (0.0917)
p^0_{23}	0.1714 (0.0821)	0.1881 (0.0763)	0.1847 (0.0768)
p^0_{31}	0.0691 (0.0631)	0.0661 (0.0605)	0.0601 (0.0559)
p^0_{32}	0.2983 (0.1175)	0.3076 (0.1180)	0.3016 (0.1167)
p^1_{12}	0.0413 (0.0120)	0.0421 (0.0114)	0.0425 (0.0113)
p^1_{13}	0.0033 (0.0031)	0.0045 (0.0033)	0.0048 (0.0035)
p^1_{21}	0.2617 (0.0650)	0.2669 (0.0640)	0.2698 (0.0641)
p^1_{23}	0.2210 (0.0626)	0.2041 (0.0581)	0.2108 (0.0592)
p^1_{31}	0.0833 (0.0411)	0.0768 (0.0378)	0.0744 (0.0364)
p^1_{32}	0.1180 (0.0630)	0.1158 (0.0606)	0.1235 (0.0598)
τ_{11}	-5.5827 (0.3067)	-5.3024 (0.2700)	-5.3464 (0.2727)
τ_{12}	3.8310 (0.3094)	3.9877 (0.2854)	3.9975 (0.2921)
τ_{13}	2.1909 (0.3809)	2.1059 (0.3937)	1.9559 (0.4172)
τ_{21}	1.1755 (0.0252)	1.1902 (0.0229)	1.1845 (0.0229)
τ_{22}	0.2987 (0.0758)	0.3116 (0.0819)	0.3118 (0.0810)
τ_{23}	1.0737 (0.0621)	1.1195 (0.0614)	1.1037 (0.0614)
Σ_{11}	6.5650 (1.2101)	6.8945 (1.1307)	6.8228 (1.1296)
Σ_{12}	0.3108 (0.0842)	0.3311 (0.0792)	0.3183 (0.0777)
Σ_{22}	0.1059 (0.0128)	0.1011 (0.0112)	0.0986 (0.0110)
$\frac{1}{\sigma^2_\epsilon}$	10.3074 (0.5051)	9.9166 (0.4831)	10.0166 (0.4944)

where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

Table 5.3: Bivariate MHMM results for three different missing data Scenarios under the Poisson count SR model

Parameter	Complete Cases $N = 250$	Available Data $N = 354$	Imputed Missing Data $N = 354$
π_1	0.7475 (0.0372)	0.7130 (0.0297)	0.7258 (0.0298)
π_2	0.1892 (0.0368)	0.2146 (0.0289)	0.1969 (0.0286)
p^0_{12}	0.0409 (0.0197)	0.0442 (0.0174)	0.0414 (0.0161)
p^0_{13}	0.0106 (0.0072)	0.0136 (0.0077)	0.0126 (0.0073)
p^0_{21}	0.2661 (0.0990)	0.2072 (0.0719)	0.2880 (0.0847)
p^0_{23}	0.1564 (0.0689)	0.1745 (0.0598)	0.2488 (0.0747)
p^0_{31}	0.1239 (0.0780)	0.1241 (0.0786)	0.1203 (0.0698)
p^0_{32}	0.0721 (0.0698)	0.0877 (0.0796)	0.0849 (0.0737)
p^1_{12}	0.0374 (0.0100)	0.0349 (0.0084)	0.0359 (0.0088)
p^1_{13}	0.0089 (0.0044)	0.0083 (0.0037)	0.0085 (0.0039)
p^1_{21}	0.2362 (0.0531)	0.2515 (0.0486)	0.2582 (0.0498)
p^1_{23}	0.2361 (0.0479)	0.2271 (0.0437)	0.2348 (0.0445)
p^1_{31}	0.0786 (0.0344)	0.0650 (0.0307)	0.0643 (0.0299)
p^1_{32}	0.1724 (0.0504)	0.2007 (0.0501)	0.1973 (0.0489)
τ_{11}	-5.4652 (0.2731)	-5.2899 (0.2120)	-5.1772 (0.2132)
τ_{12}	4.2017 (0.2915)	4.4679 (0.2035)	4.3274 (0.2214)
τ_{13}	1.3120 (0.0632)	1.3101 (0.0434)	1.3068 (0.0441)
τ_{21}	1.1671 (0.0248)	1.1851 (0.0221)	1.1876 (0.0222)
τ_{22}	0.2708 (0.0526)	0.2959 (0.0474)	0.2801 (0.0469)
τ_{23}	0.8727 (0.0575)	0.8782 (0.0569)	0.8791 (0.0526)
Σ_{11}	2.7983 (0.6228)	2.3318 (0.4179)	2.5771 (0.4644)
Σ_{12}	0.1102 (0.0539)	0.1242 (0.0459)	0.1399 (0.0453)
Σ_{22}	0.1036 (0.0117)	0.1048 (0.0111)	0.1052 (0.0109)
$\frac{1}{\sigma^2_\epsilon}$	9.3642 (0.4317)	8.6795 (0.3676)	8.7176 (0.3707)

where π_k is the initial probability of hidden state k, p^s_{jk} is the j, k transition probability for the sth treatment group for the hidden states, τ_{rk} is the intercept term for response r and hidden state k, $\mathbf{\Sigma}$ is the covariance matrix of the random effects, and σ^2_ϵ is the residual error associated with response 2.

tionary distribution based on estimates of $\boldsymbol{P}$ in either treatment group would be roughly $[0.7, 0.1, 0.2]$, where typically the longterm abstinence rates are much less (about 20%), though, admittedly, this is really comparing apples to oranges, as often in the initial study period the participants are on assigned treatment, whereas in post-treatment follow-up the participants are typically off treatment.. Some caution is therefore warranted before attempting to generalize these results to long-term smoking status.

In section 5.3 we discussed missing data issues, and examined the hypothesis put forth by Borrelli et al. (2002) that subjects who drop out of smoking cessation intervention studies are distinct from individuals who are classified as smokers or non-smokers in such studies. An alternative hypothesis could be that the distinct population Borrelli et al. (2002) observed was actually a mixture of diseases states. We believe we have provided some evidence for this latter alternative hypothesis. In general, we see that subjects, regardless of the time at which we observe them, have generally a high propensity to be in one of the three hidden disease states used in our analyses. Rarely do individuals have a mixed propensity to be classified in multiple hidden disease states. This observation seems to generalize to individuals at the time of dropout, and we therefore believe that those who drop out are actually a mixture of all three hidden states at the time of dropping out. We would estimate the proportions of this mixture to be about 0.574, 0.214 and 0.212, which represents a significantly higher prevalence of hidden state two and three than would be observed in subjects who did not dropout. We believe that this may explain the observation by Borrelli et al. (2002) about subjects who drop out. We also point out that a common practice which assumes that subjects who drop out of smoking cessation studies as smokers is likely not entirely appropriate. In general, we note that those who drop out are more likely to not be in hidden state one which generally identifies with abstainers, but at the time of dropout, we generally observe more abstainers than non-abstainers, which calls into question the legitimacy of this assumption.

Lastly, we examined how to incorporate the additional uncertainty caused by the miss-

ing data. We suggest an imputation approach which assumes ignorability of the missing data mechanism, which we believe to be appropriate, since information about the current disease state seems to be very predictive of future disease states, and this information seems to generalize to both subjects who dropped out and those who did not in a similar fashion. This is only one way of conducting analysis to address missing data issues. Other imputation (e.g., Liu, Taylor and Belin (2000)) and other approaches (e.g., Ibrahim, Chen and Lipsitz (2001)) could also be used under a MHMM to account for missing data, with careful attention paid to both the observed and hidden components of the MHMM framework. Implementation and comparisons of different modelling approaches is an area of future work.

We believe MHMMs provide a unique opportunity to conduct inference as it relates to prediction of hidden states, and also address issues related to missing data. As we have shown, this approach provides a good framework for evaluating issues such as ignorability of the missing data mechanism, and other hypotheses related to subjects who drop out.

Chapter 6

Mixed Effects Hidden Markov Models for Multivariate Longitudinal Data: Summary and Future Work

6.1 Summary

6.1.1 Modelling Framework

In Chapter 3, we described a modeling framework to describe multivariate longitudinal data, where heterogeneity within the response can be described by between-subject differences, in addition to heterogeneity caused by dynamic changes to an underlying hidden disease state. We restricted our consideration of dynamic models for the hidden disease states transitions to M-state first-order discrete time Markov models. We believe that this class of models will describe many situations well, including the study design we considered – a randomized clinical trial for the study of the treatment of addiction.

The modeling framework built upon the work of others (e.g., Altman (2007); Gueorguieva and Agresti (2001)) in both the univariate longitudinal response literature for

hidden Markov models, and the multivariate longitudinal data literature, where utilizing separate, but correlated random effects between responses has been used in several settings. As we have shown, computation can be done by maximizing the marginal likelihood, after integrating over the random effects distribution, but in most situations the computational burden of fitting such models will be extensive. Under this approach computation time significantly increases as a function of overall sample size, hidden state space size, and most importantly the number of responses being considered. In general, it is possible to fit such models by maximum likelihood estimation with a moderately sized sample, but for large hidden state spaces or responses with dimension greater than two, the computation time is significant. Since both of these latter scenarios involving more responses or a larger hidden state space are common issues one will want to examine under this modeling framework, we believe other methodologies will need to be utilized in the general setting.

Other approaches to modeling MHMMs, including semi-parametric approaches (Maruotti and Ryden, 2009) may be useful in univariate responses or for independent random effects distributions, but with these limitations, they will not generally be applicable to most situations encountered modeling a multivariate longitudinal response under a MHMM. With all this in mind, we chose to implement computation through a Bayesian approach using Markov chain Monte Carlo. This approach is relatively straightforward to implement, and can be extended quite easily to other more complex settings.

Under the various simulation studies in Chapter 3, this methodology generally performed well, in terms of producing unbiased estimators for the model parameters, and having approximately frequentist coverage on posterior intervals. In certain situations, bias was noted in some of the parameter estimates, and was generally the result of a flat posterior. This occurred mostly in components of the hidden state probability transition matrix, and we attribute this mainly to the low expected number of transitions observed in smaller strata (i.e., in our case the placebo group which had 1/3 the number of subjects as those on active treatment). The occurrence of a flat prior and subsequent biased

estimation can be caused by a number of different factors examined in Chapter 3, and is something one needs to be aware of when fitting such models. This phenomena did not occur in the full analysis of the smoking cessation dataset, we believe largely due to the well-separated modes of the observed response conditional on the hidden response.

In Chapter 4, under different simulation scenarios, there appears to be a clear benefit of modeling the data as a multivariate response. When compared to the two separate univariate response models, a bivariate model which modeled both responses simultaneously was shown to be superior in many respects, but in particular, the amount of bias observed in many of the parameters was significantly reduced. This would suggest that adding a longitudinal response which has some of the heterogeneity explained by the same underlying disease state to multivariate response MHMMs may improve inference drawn from such a model.

While we have considered one example where a multivariate longitudinal response is described by a MHMM, there may be other settings where this is applicable as well. In general, in the study of substance abuse and addiction, it is common to monitor subjects over time by self-reported use and biochemical confirmation (e.g., urinalysis), and this would be another obvious application of our approach. Psychiatric diseases as well often administer a battery of questionnaires to determine severity and presence of mental illness, and often these diseases are believed to have multiple dynamic disease states. It may also be the case, where only a univariate response is available for a particular disease area, and although a MHMM would apply, no additional responses are available which are known to describe the disease state. It is often the case that in many of these chronic illnesses, the dynamic disease states severely affect quality of life in the individuals. In such a setting, it would seem logical to model the univariate longitudinal response alongside the (possibly multivariate) quality of life response under a MHMM for multivariate response framework we describe in Chapter 3.

6.1.2 Classification and Prediction of Disease States in Smoking Cessation

In Chapter 4, we examined how well the hidden states describe already utilized disease state definitions in a smoking cessation clinical trial. We generally relied on estimates from the locally decoded probabilities of the hidden states. These study-defined disease states are typically characterized by thresholds of one or more longitudinal responses. While far from a 'gold-standard', it would be desirable that the hidden states we model throughout this thesis roughly correspond to these known definitions, and when discrepancies do occur, they illustrate differences between the hidden and study-defined disease states. This is generally the case, where we see a high correspondence between the hidden states and those defined in the study protocol. Discrepancies typically occur when the MHMM model detects relative increases in CO-levels as evidence of non-smoking, whereas the study defined thresholds only consider absolute levels of CO. Similarly, the MHMM hidden states are generally forgiving for isolated self-reported smoking observation (usually fewer than 7 cigarettes in a week), when not accompanied by an increase in CO-levels, or sustained SR counts through subsequent weeks.

We believe these discrepancies describe two areas where one may want to improve the study definitions, or alternatively, enforce stricter adherence to abstinence in the lowest hidden state within the model approach itself. The close agreement between these two estimates of disease state, is reassuring, and generally leads one to believe that a MHMM is useful in the setting of smoking cessation intervention studies. The MHMM approach also has the additional benefit of modeling the intermediate second hidden state, believed to describe withdrawal. This state is much more difficult for study investigators to define and measure, but may play an important role in assisting people in quitting smoking (Killeen, 2011).

In Chapter 5, we showed how prediction of hidden states can be done similarly to the

classification of the hidden states. In general, prediction of future hidden states can be done relatively efficiently, and sampling the hidden states will play an important role in imputation approaches to addressing the issue of missing data.

6.1.3 Missing Data Issues

Missing data as a whole, but in particular study dropout is a problem frequently encountered in longitudinal data analysis, and we encounter it when analysing the dataset from our motivating example. A MHMM approach allows for a unique opportunity to test for non-ignorability of the dropout mechanism, by examining the distribution of the locally decoded probabilities of the hidden states at different time points, including the time leading up to the study dropouts. This along with other information was used as a justification for a missing at random (MAR) assumption, which we used to describe a methodology to impute the missing data (both the observed longitudinal data in addition to the hidden states) during the analysis procedure. While future work may be required to further provide justification for the MAR assumption in our setting, the approach we described could be adapted to whenever such an assumption was appropriate.

6.2 Future Work

We believe that future work falls into three categories. The first obvious area of work would be to apply and extend the modeling framework to other settings and situations. The second topic is related to model selection and hypothesis testing, and is an important topic we have largely ignored. Lastly, we have outlined a very basic procedure for imputing missing data, and we believe future work could involve examining other situations and assumptions.

6.2.1 Model Extensions

We have outlined a framework for modeling multivariate longitudinal data as a MHMM, but have generally limited our analysis to a bivariate response. In practice, as mentioned in section 6.1.1, multivariate longitudinal data can consist of many more responses. While we did observe some definite benefits of analyzing a bivariate response over two univariate responses under a MHMM, it remains unclear if these benefits will extend to further dimensions of responses, or other situations. Increasing the dimension of the longitudinal response will also increase the computational complexity of the MCMC procedure, and it remains to be seen if the current modeling approach is feasible computationally for large response numbers. Improvements to sampling approaches may be necessary to accomplish this, or further modeling assumptions, for instance, pertaining to the variance components of the random effects, may be required.

Altman (2007) has proposed further extensions for univariate longitudinal response MHMMs to allow for between-subject heterogeneity of the hidden response also modeled through random effects. Scott, James and Sugar (2005) proposes a Bayesian hierarchical approach for describing heterogeneity in the hidden process, that could be adapted to the modeling framework we have described. We believe these are important generalizations that allow for increased flexibility in the modeling approach, but were unable to consider it for our application due to the limited observation time ($n_i \leq 6$) available for each subject. Use of such a modeling approach is increasingly difficult as the hidden state space increases, and large observation time would be required, as a model which allows for heterogeneity in all transition probabilities would require an additional $M(M-1)$ random effects in addition to any random effects used in the observed part of the model.

6.2.2 Model Selection

Throughout Chapters 3-5, we generally neglected any direct quantitative comparison of models, largely due to our limited ability to do model selection and hypothesis testing. Others (Celeux et al., 2006) have noted that using information criteria such as the deviance information criteria (DIC) are difficult in models such as random effect or mixture models. In the setting of a MHMM, the models considered are both a random effect and mixture model. In addition to this problem, often such approaches will require extensive computation, and may not be feasible in many settings. Including covariates in these models is of particular interest, as the choice to include certain covariates in the observed or hidden components of the model is not always obvious. For the reasons we have outlined, providing model selection methodology or clarification on how to use already established methods is an area of future research.

6.2.3 Missing Data Approaches

We provided one approach to accommodate missing data via dropout in Chapter 5 based on a missing at random assumption. While we believe this assumption to be appropriate for the dataset we considered, in general a sensitivity analysis should be conducted to consider other assumptions such as a missing not at random (MNAR) dropout. Future work could include the consideration of such alternative modeling assumptions (e.g., Ibrahim, Chen and Lipsitz (2001)), in addition to the sensitivity of the MAR assumption to intermittently missing data which occurred throughout the study.

6.3 Conclusions

MHMMs have been applied to the analysis of univariate longitudinal data, but can be extended to the setting of multivariate longitudinal data analysis, provided that conditional independence assumptions are appropriate. We have taken a Bayesian approach to conduct

inference, and we believe the approach we have outlined can be adapted and extended to many other important situations. Future work will focus on these extensions, in addition to appropriate methods to conduct model selection and addressing subject dropout in such models.

MHMMs are another class of model which can be used to explain heterogeneity within a longitudinal response which would largely be inappropriately accounted for by standard methodologies used for the analysis of longitudinal data. In the setting where data are generated from individuals transitioning through multiple disease states, we believe that MHMMs are an important tool for understanding the natural history of the disease, and identifying the role treatment plays in modifying the disease states.

Bibliography

Alberts P. S., McFarland H. F., Smith M. E. and Frank J. A. (1994). Time series for modelling counts from a relapsing-remitting disease: application to modelling disease activity in multiple sclerosis. *Statistics in Medicine* **13**, 453–466.

Altman R. M. (2007). Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data setting. *Journal of the American Statistical Association* **102,** 201–210.

Altman R. M., Petkau A. J., Vrecko D. and Smith A. (2012). MRI-based clinical trials in relapsingremitting MS: new sample size calculations based on a longitudinal model. *Statistics in Medicine* **31,** 449–469.

Azzalini A. (1994) Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* **81**, 767-775.

Baum L. E., Petrie T., Soules G. and Weiss N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* **41**, 164-171.

Benowitz N. L. (1996). Pharmacology of nicotine: addiction and therapeutics *Annual Review of Pharmacology and Toxicology* **36,** 597–613.

Borrelli B., Hogan J. W., Bock B. Pinto B., Roberts M. and Marcus B. (2002). Predictors

of Quitting and Dropout Among Women in a Clinic-Based Smoking Cessation Program. *Psychology of Addictive Behaviours* **16,** 22–27.

Breslow N. E. and Clayton D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association.* **88**, 9-25.

Cappe O., Moulines E. and Ryden T. (2005). *Inference in Hidden Markov Models.* New York: Springer.

Celeux G., Forbes F., Robert C. P., and Titterington D. M. (2006). *Deviance information criteria for missing data models* **1,** 651–673.

DeSantis S. M. and Bandyopadhyay D. (1996). Calculating posterior distributions and model estimation in Markov mixture models. *Journal of Econometrics* **75,** 79–97.

DeSantis S. M. and Bandyopadhyay D. (2011). Hidden Markov models for zero-inflated Poisson counts with an application to substance use. *Statistics in Medicine* **30,** 1678–1694.

Diggle P. J., Heagerty P., Liang, K. Y. and Zeger S. L. (2002). *Analysis of Longitudinal Data.* Oxford University Press, Oxford, UK.

Dubin J. A. (2012). Personal Communication, November 19, 2012.

Fieuws S. and Verbeke G. (2006). Pairwise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles. *Biometrics* **62,** 424–431.

Fitzmaurice G., Davidian M., Verbeke G., and Molenberghs G. (*ed.*) (2009). *Longitudinal Data Analysis* Chapman & Hall, Boca Raton, FL.

Geman S. and Geman D. (1984). Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6,** 721–741.

Gilks W. R. and Wild P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **41,** 337–348.

Gray S.M. and Brookmeyer R. (1998). Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* **54**, 976–88.

Gray S.M. and Brookmeyer R. (2000). Multidimensional longitudinal data: Estimating a treatment effect from continuous, discrete or time to event response variables. *Journal of the American Statistical Association* **95**, 396–406.

Gueorguieva R. V. and Agresti A. (2001). A Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Responses. *Journal of the American Statistical Association* **96,** 1102–1112.

Hughes J. R., Keely, J. P., Niaura R. S., Ossip-Klein D. J., Richmond R. L. and Swan G. E. (2003) Measures of abstinence in clinical trials: Issues and recommendations. *Nicotine and Tobacco Research* **5,** 13–25.

Ibrahim J. G., Chen M. H. and Lipsitz S. R. (2001) Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88,** 551–564.

Jarvis M. J., Tunstall-Pedoe H., Feyerabend C., Vesey C., and Saloojee Y. (1987) Comparison of tests used to distinguish smokers from nonsmokers. *American Journal of Public Health* **77,** 1435–1438.

Kenford S. L., Fiore M. C., Jorenby D. E., Smith S. S., Wetter D. and Baker T. B. (1994) Predicting Smoking Cessation. *JAMA* **271**, 589–594.

Killeen P. R. (2011). Markov Model for Smoking Cessation. *Proceedings of the National Academy of Science* **108,** 15549–15556.

Laird N. M. and Ware J. H. (1982). Random Effects Models for Longitudinal Data. *Biometrics* **38**, 963–974.

Liang K. Y. and Zeger S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lin H., McCulloch C. E., Turnball B. W., Slate E. H. and Clark L. C. (2000). A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* **19**, 1303–1318.

Liu M., Taylor J. M. G. and Belin T. R. (2000). Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics* **56,** 1157–1163.

Liu X., Daniels M. J. and Marcus B. (2009). Joint Models for the Association of Longitudinal Binary and Continuous Processes with Application to a Smoking Cessation Trial. *Journal of the American Statistical Association* **104** 429–438.

Maruotti A. and Ryden T. (2009). A semiparametric approach to hidden Markov models under longitudinal observations. *Statistics and Computing* **19,** 381–393.

Maruotti A. (2011). Mixed Hidden Markov Models for Longitudinal Data: An Overview. *International Statistical Review* **79,** 427–454.

O'Malley S. S., Cooney J. L., Krishnan-Sarin S., Dubin J. A., McKee S. A., Cooney N. L., Blakeslee A., Meandzija B., Romano-Dahlgard D., Wu R., Makuch R. and Jatlow P. (2006). A Controlled Trial of Naltrexone Augmentation of Nicotine Replacement Therapy for Smoking Cessation. *Archives of Internal Medicine* **166,** 667–674.

Patrick D.L., Cheadle A., Thompson D.C., Diehr P., Koepsell T. and Kinne S. (1994). The validity of self-reported smoking: a review and meta-analysis. *American Journal of Public Health* **84,** 1086–1093.

Prentice R.L. (1988). Correlated Binary Regression with Covariate specific to Each Binary Observation. *Biometrics* **44**, 1033–48.

Proust-Lima C., Joly P., Dartigues J. F. and Jacqmin-Gadda H. (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics & Data Analysis* **53,** 1142–1154.

Qin L., Weissfeld L.A., Shen C., and Levine M.D. (2009). A Two-Latent-Class Model for Smoking Cessation Data with Informative Dropouts. *Communications in Statistics* **38**, 2604–2619

Scott S. L. (2002). Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century. *Journal of the American Statistical Association* **97,** 337–351.

Scott S. L., James G. M. and Sugar C. A. (2005). Hidden Markov Models for Longitudinal Comparisons. *Journal of the American Statistical Association* **100,** 359–369.

Shirley K., Small D., Lynch K. G., Maisto S. A. and Oslin D. W. (2010). Hidden Markov models for alcoholism treatment trial data. *Annals of Applied Statistics* **4,** 366–395.

Smith A. F. M. and Roberts G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo Methods (with discussion). *Journal of the Royal Statistical Society Series B* **55,** 3–24.

Stead L.F., Perera R., Bullen R., Mant D. and Lancaster T. (2008). Nicotine replacement therapy for smoking cessation. *Cochrane Database of Systematic Reviews* 1–123.

Stephens M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society Series B* **62,** 795–809.

Wang H. and Heitjan, D. F. (2008). Modeling heaping in self-reported cigarette counts. *Statistics in Medicine* **27**, 3789–3804.

Xu J. and MacKenzie G. (2012). Modelling covariance structure in bivariate marginal models for longitudinal data *Biometrika* **99,** 649–662.

Zeger S. L. and Liang K. Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics* **42**, 121–130.

Zhao L. P. and Prentice R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77**, 642–8.

Zucchini W. and MacDonald I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R.* Boca Raton, FL: Chapman and Hall/CRC.