INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Accounting for Misclassification in Binary Longitudinal Data

by

Rhonda Jean Rosychuk

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 1999

National Library of Canada

Acquisitions and Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada

Acquisitions et services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-44779-0

Canada

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

This thesis proposes new methodology for alternating binary longitudinal responses collected at discrete time points which may not correctly classify the state of the unobservable true process. The model consists of two distinct parts and enables estimation of the probabilities associated with two types of misclassification when supplementary information is available at each observation time. The misclassification part models misclassification probabilities as logistic functions of misclassification predictors available at each observation time and the true process part is modeled as a continuous-time counter model with time-independent covariates.

Parameter identifiability and estimability issues are investigated when a Type I counter model has exponential state sojourn time distributions. Characteristic functions are used to identify distinct sets of parameter values which yield the same likelihood value for a data set. Estimability issues are discussed when the sampling interval is inadequate or the model is misspecified.

The effect of misclassification on the parameter estimates in the case of constant inter-observation times is next considered. In the absence of covariates and supplementary information, approximated estimators of Type I counter transition probabilities are constructed based on linear functions of known misclassification probabilities. Estimators ignoring misclassification are compared with the approximated and maximum likelihood estimators.

Standard model assessment techniques comparing observed and expected transition counts are applied and may not adequately detect model departures. A simulation study motivates a Type I counter model where one sojourn time has an exponential distribution and the other has a gamma distribution for the data sets considered. Via approximation, the equilibrium probabilities at three consecutive time points are used with misclassification probabilities to calculate expected frequencies. The discrepancy between the expected and observed data at these time points is minimized to obtain parameter estimates and estimates for the true state joint equilibrium probabilities.

Acknowledgements

I wish to thank my supervisor Mary Thompson and co-supervisor Jerry Lawless for sharing their insights, guiding my efforts and providing financial support. I greatly value Mary's dedication to my success, her willingness to mentor, her steadfast support and her sense of humour through many weekly meetings over the years. I also extend sincere thanks to Steve Brown and Richard Cook for providing helpful suggestions which have improved this thesis.

My time at the University of Waterloo has been a positive experience. The faculty, students and staff in the Department of Statistics have provided a supportive and stimulating environment. My life has been enriched by many people over the years and I wish to acknowledge the support of a few special individuals:

> My parents, Gloria and Ronald Rosychuk, and my brother, Jason Rosychuk, who have encouraged and challenged me to be who I am today.
>
> My husband, Jason Stuber, who has believed in me and has been willing discuss many mathematical and computational issues.
>
> My cousins, Lana Millar and Rob Virkutis, for many enjoyable visits and discussions as the Ontario branch of the Melnyk Family.
>
> My friends and colleagues, Karen Kopciuk, Jacinte Jean, Janet Grad, Min Zhan, Matt Schonlau, Andreas Sashegyi, Daniel Fong and Claire Bilodeau, for sharing their experiences and eagerly dispensing advice.

To my parents,

Ronald and Gloria Rosychuk

Contents

List of Tables

List of Figures

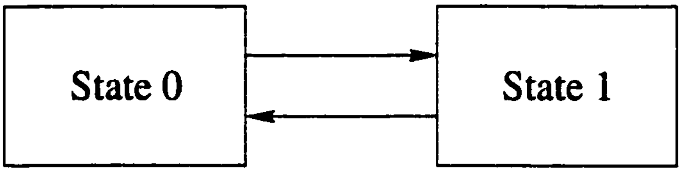
Chapter 1

Introduction

Researchers are often faced with problems in which the data they would like to collect are not directly observable. Instead, data are collected on a process thought to be related to the true unobservable process. In the biostatistics setting, for example, a true disease process may be monitored by an imperfect diagnostic indicator. The observed process may misclassify the true state of the process at a particular time, if the relationship between the observed and the true process is not completely understood. If misclassification is ignored in the modeling, incorrect conclusions may be drawn about the true unobservable process. These incorrect conclusions may influence the treatment a patient receives and may prevent greater understanding of the true disease process.

This thesis considers an underlying stochastic process which has two states and the process alternates between these states at different times. Such a process can be represented by Figure 1.1 and is referred to as the true process. Another response, thought to be related to the true process, is observed and may be used to predict the state of the true process. The observed process might be the outcome of a diagnostic test or some other supplementary information.

Figure 1.1: Schematic diagram of a two-state alternating process.



At a particular time point, the observed state may not agree with the true state. Disagreement between the two processes indicates that the observed state is misclassified. Misclassification may result if the relationship between the observed and true process is not completely deterministic. Misclassification error can happen in two ways: either the true state is 0 and the observed state is 1 or the true state is 1 and the observed state is 0. The probabilities associated with each of these errors, called misclassification probabilities, may also be related to a particular variable referred to as the misclassification predictor. This predictor is also collected at the same times as the observed state.

Biostatistical data provide an example of the types of processes involved. Many chronic diseases follow a relapsing-remitting disease process. A relapsing-remitting disease is one which sufferers experience periods when the disease is worsening (relapsing) and periods when it is improving (remitting). Symptoms are severe during the relapsing state and mild or absent during the remitting state. The true process, in this type of situation, may be the true relapsing or remitting state of the disease process. While the actual disease process is not observable directly, the outcome of a diagnostic test may help determine the true state. The misclassification predictor could indicate whether or not the patient has experienced an exacerbation recently. If an exacerbation has been experienced lately and the outcome of the diagnostic test suggests the disease is worsening, the observed and true state may be more likely to agree than if the diagnostic test suggests an improving period.

The goal is to use the misclassification predictor to clarify the relationship between the true process and the possibly misclassified observed response. Longitudinal data are collected on the observed response, misclassification predictor and covariates to enable estimation of both misclassification probabilities and regression parameters associated with the covariate information.

The model proposed consists of two distinct parts: the misclassification part specifies probabilities related to the misclassification predictor and the true process part is modeled with time-independent covariates. Each of these parts could be modeled in various ways. The approach taken here is to model the misclassification probabilities as logistic functions of the misclassification predictor available at each observation time. A stochastic process approach is taken for the true process. Counter model methodology is considered, particularly appealing because of parsimonious parameterization and its suggestion of a mechanism. Counter models are based on physical machines used to count incoming particles. Different types of events may "trigger" a transition from state 0 to state 1 in the true process. Triggers play the role of the incoming particles and may induce a transition to state 1 if the current state of the process is state 0. The state sojourn time distributions are both considered to be exponential. Chapter 2 provides an overview of the literature and describes the theoretical framework for the model proposed.

Two biostatistical data sets are available to motivate the methodology and illustrate its application. The first set involves a clinical trial of relapsing-remitting Multiple Sclerosis patients. Data include baseline covariates, information from patient brain scans as well as whether a patient has experienced a worsening of symptoms. The second data set, taken from the literature, measures the detection of a parasite in a study of Kenyan children. The data consists of weekly measurements of the parasite's presence or absence in each child. The data sets and preliminary analyses are described in Chapter 3.

Incorporating the possibility of misclassification in a model makes estimation more difficult. Models for misclassification, or the related topic of measurement error, often encounter problems with parameter estimation. For the proposed likelihood-based model, the related issues of identifiability and estimability of model parameters are considered in Chapter 4.

The observed data do not properly describe the true process unless the modeling accounts for the possibility of misclassification. Researchers may incorrectly ignore the possibility of misclassification and draw conclusions based on potentially biased estimates. Chapter 5 examines the effect misclassification has on transition probability estimates and investigates the bias that may result from naively concluding that the observed data is correctly classified.

Model assessment is an important aspect of the model building process. When a transitional modeling approach is taken for binary longitudinal data, the model specification is usually assessed by comparing expected and observed transition counts. Chapter 6 investigates the performance of such assessment tools when applied to a transition model with misclassification. The results of this chapter lead to consideration of a gamma distributed sojourn time for state 1 in Chapter 7. A semi-Markov model is constructed for the true process where state 0 has an exponential sojourn time distribution and state 1 has a gamma distributed sojourn time. Additionally, the methodology described in Chapter 2 is extended to allow for more than two states.

The final chapter summarizes the findings and discusses further research issues.

Chapter 2

Model Description

2.1 Overview

A model with distinct specifications for the misclassification and true process is developed. The probabilities for the two types of misclassification are modeled as logistic functions of misclassification predictors available at each observation time. The true process is modeled as a Type I or Type II counter model with exponential state sojourn times and time-independent covariates. These models are appealing because of parsimonious parameterization and the suggestion of a mechanism. The relationship between the covariates and each state transition rate is assumed to be log-linear. The model yields a recursive likelihood function for parameter estimation and jackknife standard error estimates are considered.

2.2 Introduction

The model proposed for misclassified binary data consists of two distinct parts. One part describes the movement between states of the unobservable true process. The other part relates the observed process, for which data are collected, to the unobservable true process through classification and misclassification probabilities. The unobservable true process is modeled as a counter model, with the open and blocked states of the counter corresponding to the two states of the true process. The probability of misclassification for an observed state is modeled as a logistic function of misclassification predictors.

Background for misclassification models and counter models is given in Section 2.3. Sections 2.4 and 2.5 describe the misclassification and the unobservable true process parts of the model, respectively. The full model with covariates is considered in Section 2.6 and the formula for jackknife standard error estimates is given in Section 2.7.

2.3 Literature Review

2.3.1 Misclassification of the Response

Misclassification in the context of longitudinal data has not been widely addressed in the literature. Often the total number of subjects making a transition from one state to another, called a gross flow, is adjusted for misclassification without consideration of the covariates which may be involved. When the longitudinal data are categorical and progressive, authors including Albert, Hunsberger and

Biro (1997), Espeland, Platt and Gallagher (1989) and Gu (1996) have proposed misclassification models. In many such cases, the misclassification mechanism is of primary interest. In a binary progressive process, subjects may all move from one state to the other and the accurate detection of the change is the focus.

Several authors have considered Hidden Markov Models (HMMs) (see MacDonald and Zucchini (1997), Le, Leroux and Puterman (1992) and Juang and Rabiner (1991), for example). In these models the state sequence generated by a first-order Markov process cannot be observed directly. The observed process is assumed to be stochastic with a set of observation probability measures. The probability of an observed state sequence can be expressed without assuming knowledge of the underlying process from which it was generated. Many of the specific references that follow can be considered special cases of a Hidden Markov Model.

Kasagi et al. (1988) adjust for disease status misclassification under a two-time follow-up study. Two separate diagnostic tests for the presence or absence of disease are considered. Test A has imperfect sensitivity (false negatives) while Test B has imperfect specificity (false positives). All subjects are first tested with Test A. Subjects with a negative result are administered Test B. This testing protocol is conducted at two examination times. Maximum likelihood estimates are calculated to evaluate the incidence and reversion rates of the disease. The modeling does not include any covariates.

More recently, Cook, Ng and Meade (1999) consider a discrete-time HMM with several diagnostic tests applied repeatedly over time. Classification distributions are constructed using log-linear models which are conditional on the true state.

These models allow tests to behave dependently, conditional on the true state. The likelihood constructed is recursive and can provide transition probability and initial state estimates for first-order and second-order Markov models along with classification probability estimates. The methodology can also be adapted to handle the case of missing data. Model assumptions are assessed over time by examining the bootstrap distribution of conditional residuals. The methodology is applied to binary chest radiograph data where the hidden Markov processes investigated are first order, second order and time-nonhomogeneous and different physician diagnoses are considered as individual diagnostic tests.

A logistic model for survey data is examined by Skinner (1992). The focus is on the proportion of subjects of the population who move from one state to another on successive occasions. A logistic model describes the relationship between transition probabilities and covariates for the situation with two states and two occasions. An adjusted estimator of the regression coefficients is calculated based on information about the nature of misclassification from validation studies or under the assumption of unbiased measurement errors. Kuha and Skinner (1997) give a review of the methods for survey data with misclassification errors.

Singh and Rao (1995) also consider survey data which are subject to classification error from one point in time to the next. Methods developed by Poterba and Summers (1986) and Abowd and Zellner (1985) based on reinterview-interview data and the assumption of independent classification errors (ICE) are modified to ensure marginal totals of the adjusted flow table at two subsequent time points agree with published "stocks". Reinterview-interview data are usually divided into a sample where reinterviewers reconcile discrepancies between the original interview and the reinterview with respondents and a sample where no reconciliations are carried out. Reconciled samples, with or without the unreconciled samples, are used to estimate the correct classification and misclassification probabilities that govern an individual's responses at two subsequent time points. The modifications do not require any final marginal adjustment. The robustness of the ICE assumption is examined under different departures from the assumption. The inclusion of covariates is not considered.

Satten and Longini (1996) consider measurement error in CD4 cell counts used as a marker for the progression of the human immunodeficiency virus (HIV). The CD4 counts are discretized into states. The true CD4 count has a corresponding true state and the true state is assumed to follow a continuous-time Markov chain. The logarithm of the observed counts is equal to the logarithm of the true CD4 counts plus a normal error. Likelihood based estimation is carried out using the discretized states. A progressive model is considered as well as a bidirectional model. In the bidirectional model, a subject may either move forward to the next lowest CD4 count state or move backwards to the next highest CD4 count state. The objective of the authors is to make population rather than individual predictions.

Albert, Hunsberger and Biro (1997) examine monotonic ordinal responses which are subject to misclassification. The underlying monotonic response and the diagnostic misclassification process are modeled separately. Diagnostic error is assumed to be independent over visits and the underlying monotonic process is assumed to follow a first-order Markov chain with a proportional odds parameterization. The misclassification mechanism is modeled as a function of the distance between the observed and the true state labels and can either be symmetric around the true state or asymmetric. An EM algorithm is developed to handle maximum likelihood estimation when covariates are introduced in the state transition probabilities or in the misclassification mechanism and when the observed ordinal states are missing at random. The method is used to analyze sexual maturation data.

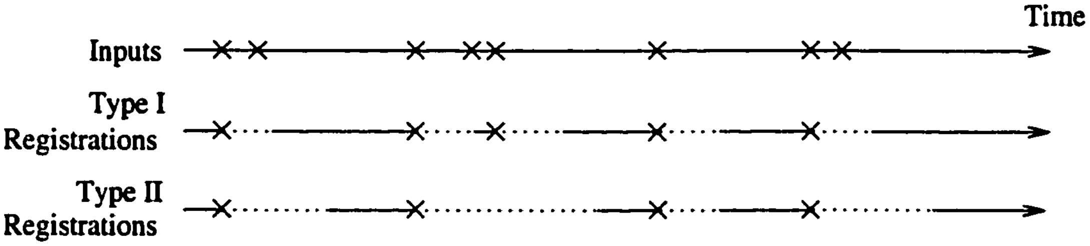
The case of imperfect detectability of parasitic infections is discussed by Nagelkerke, Chunge and Kinoti (1990). False negatives are assumed to occur, whereas false positives do not occur. The unobservable true states are assumed to follow a Markov model with constant infection and cure rates. Parameters are estimated via a maximum likelihood approach. Goodness-of-fit tests are suggested for the model involving the cure rate, infection rate and false negative probability parameters. The methods and results of this paper are further examined in the sections that lie ahead.

2.3.2 Counter Models

Several authors, such as Takács (1956), Smith (1958) and Pyke (1958), have addressed problems involving counter models. Counter models are based on physical devices used to count incoming particles. The Geiger-Müller counter is a classic example. Not all incoming particles are counted (registered). The registering of a particle induces a period of time for which the counter is blocked. During this blocked time or dead time, any incoming particles cannot be registered. Thus, only a subsequence of the incoming particles are registered. The incoming particles are called the input process and the registered particles are referred to as the thinned process.

Several kinds of blocking can be considered. The more common mechanisms are called Type I (or non-paralyzable) and Type II (or paralyzable). Figure 2.1 illustrates the difference between these two counter model types when the dead time is constant. For a Type I counter, only particles arriving during an unblocked time are registered and induce a dead time. During this dead time, any other incoming particles are not registered and do not alter the dead time. If the incoming particles during the dead time extend the dead time as if they had been registered, the model is a Type II counter.

Figure 2.1: Counter models of Type I and II with constant dead time. The incoming particles and dead times are denoted by × and $\cdots$, respectively.



Albert and Nelson (1953) consider a generalization which includes both Type I and II counters as special cases. The generalization is essentially a Type II counter model with the exception that incoming particles during the dead time may prolong the dead time with a certain probability. Other modifications have been considered by authors including Srinivasan and Rajamannar (1970) and Dvurecenskij and Ososkov (1985).

A few papers have been produced with counter model applications. Singh (1964) considers a Type I model for couple fertility. Conception corresponds to the thinned process which is followed by a rest period (dead time) when no other conceptions can be made. A constant probability of conceiving in one time unit during the entire observed time and for all non-sterile, non-abstinent couples is assumed. The rest period is assumed to be constant for all couples and taken to be one year. Since conception data are unavailable, the analysis takes births as the thinned process with the assumption of a one-to-one correspondence between conceptions and births.

Nguyen and Murthy (1984) consider two warranty policies and their associated costs. A warranty where the manufacturer pays for all replacements and repairs during the warranty period relates to a Type I counter with a constant dead time. With a *pro rata* warranty, the customer purchases a replacement at a reduced rate if the item fails during the warranty and the replacement item is covered by an identical warranty. The *pro rata* warranty is a Type II counter with constant dead time.

The probability distribution of traffic counts is addressed by Oliver (1961). A Type I counter is formed with vehicles as the input process and the intervehicle spacings constitute the dead time.

Sequences of bases of DNA are investigated in Biggins and Cannings (1987). DNA sequences can be thought of as long sequences of letters. A particular subsequence of letters is called a word. Individual letters index the input process, the occurrence of a word is the thinned process and the length of the word represents the dead time.

As will be seen later on, it is natural to think of certain disease history models in terms of counter processes. Surprisingly, there appears to be very little explicit use of this idea in the literature, although general Markov and semi-Markov models are often invoked (see Andersen et al. (1993), for example).

2.4 Misclassification Specification

First suppose the two-state process is a homogeneous Markov process and consider the binary responses for one subject. Suppose there are n observations for the subject at times $t_1, t_2, \ldots, t_n$ such that $t_1 < t_2 < \cdots < t_n$. Let O_j take values 0 or 1 depending on the state determined by the observed, possibly misclassified, response. Hereafter, O_j will be referred to as the observed state. The state of the unobservable true process at time t is denoted by $\xi(t)$. At observation time t_j, the state of the unobservable true process is labeled ξ_j.

The true and observed states either agree or disagree. The observed state is misclassified if it differs from the true state. Suppose there is some supplementary information available at each observation time which may help clarify the relationship between the observed and true process. Variables fitting this description are referred to as misclassification predictors. Observations for q misclassification predictors are collected at each observation time, where q is a nonnegative integer. The random variables for the misclassification predictors can be written in a matrix as

$$C = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ C_{11} & C_{12} & \cdots & C_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{q1} & C_{q2} & \cdots & C_{qn} \end{pmatrix}$$

where the rows are the different misclassification predictor variables and the columns are the misclassification predictors collected at each observation time. The first row of ones corresponds to an intercept term for the misclassification probabilities defined below. These misclassification predictors may be continuous or discrete variables.

For simplicity, consider the situation with only one misclassification predictor, C, being collected at observation j, $j = 1, \dots, n$. The labels for the two types of classification and misclassification probabilities appear in Table 2.1. An appropriate functional form for the misclassification probabilities is assumed to be logistic,

Table 2.1: Labels for the misclassification probabilities. The probability of being misclassified in state b at observation j when the true classification is state a is $v_{ab}(C_j) = \Pr(O_j = b \mid \xi_j = a, C_j)$ for $a, b \in \{0, 1\}$.

		Observed State (O_j)	
		0	1
True State	0	$v_{00}(C_j) = 1 - v_{01}(C_j)$	$v_{01}(C_j)$
(ξ_j)	1	$v_{10}(C_j)$	$v_{11}(C_j) = 1 - v_{10}(C_j)$

where

$$v_{01}(C_j) = \frac{e^{\alpha_0+\alpha_1 C_j}}{1+e^{\alpha_0+\alpha_1 C_j}} \quad \text{and} \quad v_{10}(C_j) = \frac{e^{\alpha_0^*+\alpha_1^* C_j}}{1+e^{\alpha_0^*+\alpha_1^* C_j}}. \tag{2.1}$$

Assuming that the probabilities for the two types of misclassification may be non-symmetric, $\alpha_0 \neq \alpha_0^*$ and $\alpha_1 \neq \alpha_1^*$.

Let $O^{(j)}$ denote the history of the observed process up to and including observation j, $O^{(j)} = (O_1, O_2, \ldots, O_j)$, $j = 1, \ldots, n$. The probability for the observed state depends on the history, the misclassification probabilities, and the continuous-time probability transition matrix, $P(t)$, of the true process. The true process is assumed to have the homogeneous Markov property, so that $\Pr(\xi_j \,|\, \xi_{j-1}, \ldots, \xi_1) = \Pr(\xi_j \,|\, \xi_{j-1})$, and the entries of the continuous-time probability transition matrix for observations taken $\Delta t_j = t_j - t_{j-1}$ time units apart are denoted by $P_{ab}(\Delta t_j) = \Pr(\xi_j = b \,|\, \xi_{j-1} = a)$ for $a, b \in \{0, 1\}$, $j = 2, \ldots, n$.

Further, let the history of the misclassification predictor be denoted by $\mathcal{C}^{(j)} = (C_1, C_2, \ldots, C_j)$. If γ_j denotes the probability $\Pr(\xi_j = 1 \,|\, O^{(j)}, \mathcal{C}^{(j)})$, then the probability of observing a 1 conditional on the past observed responses and misclassification predictors is

$$\begin{aligned}\Pr(O_j = 1 \,|\, O^{(j-1)}, \mathcal{C}^{(j)}) = & \, v_{11}(C_j) \Big[(1-\gamma_{j-1}) \, P_{01}(\Delta t_j) + \gamma_{j-1} \, P_{11}(\Delta t_j)\Big] \\ & + v_{01}(C_j) \Big[(1-\gamma_{j-1}) \, P_{00}(\Delta t_j) + \gamma_{j-1} \, P_{10}(\Delta t_j)\Big]\end{aligned} \tag{2.2}$$

and $\Pr(O_j = 0 \,|\, O^{(j-1)}, \mathcal{C}^{(j)}) = 1 - \Pr(O_j = 1 \,|\, O^{(j-1)}, \mathcal{C}^{(j)})$ has a similar form. The probability for each potentially misclassified observed state depends on the probability of transitioning from the last state, which may also be misclassified.

When considering the observed state, the Markov property no longer holds. The probability of an observed state is dependent on the last observed state as well as earlier states through γ_{j-1}. Bayes' rule can be used to calculate γ_j recursively since an explicit expression is difficult to derive. Hence, γ_j is calculated as

$$\begin{aligned}
\gamma_j &= \Pr(\xi_j = 1 \mid O^{(j)}, C^{(j)}) \\
&= \frac{\Pr(O_j \mid \xi_j = 1, O^{(j-1)}, C^{(j)}) \Pr(\xi_j = 1 \mid O^{(j-1)}, C^{(j)})}{\sum_{b=0}^{1} \Pr(O_j \mid \xi_j = b, O^{(j-1)}, C^{(j)}) \Pr(\xi_j = b \mid O^{(j-1)}, C^{(j)})} \\
&= \begin{cases} \dfrac{v_{10}(C_j)\left[(1-\gamma_{j-1}) P_{01}(\Delta t_j) + \gamma_{j-1} P_{11}(\Delta t_j)\right]}{\sum_{b=0}^{1} v_{b0}(C_j)\left[(1-\gamma_{j-1}) P_{0b}(\Delta t_j) + \gamma_{j-1} P_{1b}(\Delta t_j)\right]} & \text{if } O_j = 0 \\ \dfrac{v_{11}(C_j)\left[(1-\gamma_{j-1}) P_{01}(\Delta t_j) + \gamma_{j-1} P_{11}(\Delta t_j)\right]}{\sum_{b=0}^{1} v_{b1}(C_j)\left[(1-\gamma_{j-1}) P_{0b}(\Delta t_j) + \gamma_{j-1} P_{1b}(\Delta t_j)\right]} & \text{if } O_j = 1 \end{cases}
\end{aligned} \tag{2.3}$$

and with $\pi_1 = \Pr(\xi_j = 1)$ as the constant probability for state 1, $j = 1, \ldots, n$, and $\pi_0 = 1 - \pi_1$,

$$\gamma_1 = \begin{cases} \dfrac{v_{10}(C_1)\,\pi_1}{v_{00}(C_1)\,\pi_0 + v_{10}(C_1)\,\pi_1} & \text{if } O_1 = 0 \\ \dfrac{v_{11}(C_1)\,\pi_1}{v_{01}(C_1)\,\pi_0 + v_{11}(C_1)\,\pi_1} & \text{if } O_1 = 1. \end{cases}$$

The contribution to the likelihood for one observed sequence is,

$$\Pr(O_1, O_2, \ldots, O_n) = \Pr(O_1 | C_1) \prod_{j=2}^{n} \Pr(O_j \mid O^{(j-1)}, C^{(j)}) \tag{2.4}$$

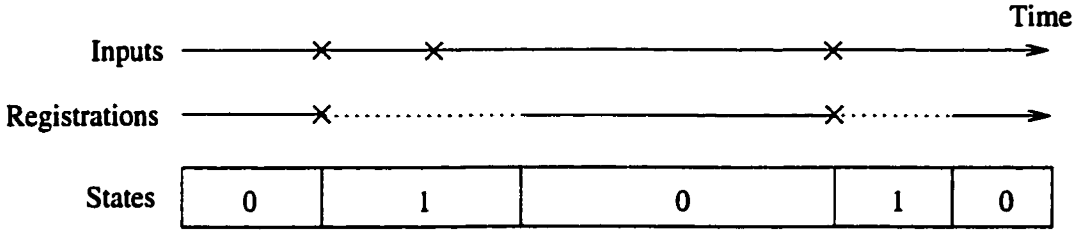
where

$$\Pr(O_1 | C_1) = \begin{cases} v_{00}(C_1)\,\pi_0 + v_{10}(C_1)\,\pi_1 & \text{if } O_1 = 0 \\ v_{01}(C_1)\,\pi_0 + v_{11}(C_1)\,\pi_1 & \text{if } O_1 = 1. \end{cases} \tag{2.5}$$

2.5 Unobservable True Process

The states of the unobservable true process are linked with the open or blocked status of a counter. Figure 2.2 graphically displays the relationship between the two concepts. The time the counter is open corresponds to the time spent in state 0. Similarly, if the counter is blocked, then state 1 is occupied. The registration of an input causes a transition to state 1 and the state is occupied during the blocked time. Once the blocked time is over, the counter becomes open again and a transition is made back to state 0. The input process may also be referred to as triggering events. Not all triggers will cause a transition from state 0 to 1 just as not all inputs are registered. Triggers which do cause a state transition represent the thinned process. Examples of such triggers in the biostatistical setting could be infections, stress or fatigue with states healthy (0) and sick (1).

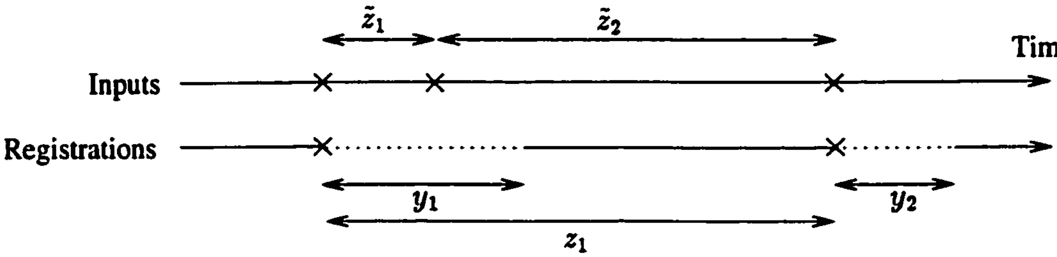
Figure 2.2: The relationship between the alternating process and the counter model. The blocked time is denoted by the dashed line while the × indicate an input.



The distributions for the input process and dead times can be specified to suit the problem at hand. The most tractable situations are to consider Poisson inputs and either constant or exponentially distributed dead times. The case of Poisson inputs and exponentially distributed dead times, often addressed in the literature (see Cox and Isham (1980), for example), is considered here. As in the previous section, only one subject is discussed.

The input counting process $N(t)$, $N(t)=$number of inputs in time interval $[0, t]$, is assumed to be a renewal process with interval sequence $\{\tilde{Z}_k\}$. The Poisson input assumption implies that $\tilde{Z}_k$ has an exponential distribution with intensity (renewal density) $\lambda_{\tilde{Z}}(\tilde{z}) = \rho$, density $f_{\tilde{Z}}(\tilde{z}) = \rho\, e^{-\rho \tilde{z}}$ and survivor function $\mathcal{F}_{\tilde{Z}}(\tilde{z}) = e^{-\rho \tilde{z}}$, $\tilde{z} \geq 0$. The sequence of independent identically distributed blocked times is $\{Y_k\}$ which have density $g(y) = \eta\, e^{-\eta y}$, distribution function $G(y) = 1 - e^{-\eta y}$, and survivor function $\mathcal{G}(y) = e^{-\eta y}$, $y \geq 0$. The state process is labeled $\xi(t)$ and is an alternating renewal process. The thinned process is a renewal process. Let $\{Z_k\}$ refer to the interval sequence of the thinned process, $Z_k \sim f_Z$. Figure 2.3 gives a pictorial representation for the random variables in a simple example.

Figure 2.3: Random variables involved in the counter model, $\tilde{Z}$, Z and Y. Incoming particles and dead times are denoted by × and $\cdots$, respectively.



Two common counter models are considered with Poisson inputs and exponentially distributed blocked times. Section 2.5.1 shows the necessary calculations for the Type I counter model while the Type II model is discussed in Section 2.5.2.

2.5.1 Type I Counter

At the instant the counter becomes unblocked, a transition is made to state 0 and the counter awaits the next trigger. Use of the Poisson assumption for the input process implies that once the counter becomes unblocked, the waiting time for the next trigger is exponentially distributed with rate ρ. Similarly, at the instant the counter becomes blocked, the blocked time has an exponential distribution with rate η. These two rates characterize the conditional intensities for the Type I counter, forming the transitional intensity matrix

$$Q^I = \begin{pmatrix} -\rho & \rho \\ \eta & -\eta \end{pmatrix},$$

where ρ and η are not functionally dependent. The transition intensity matrix along with the probability transition matrix form a system of first order linear differential equations,

$$\frac{\partial}{\partial t} P^I(t) = Q^I P^I(t)$$

with solution $P^I(t) = e^{Q^I t} = \sum_{n=0}^{\infty} (Q^I t)^n / n!$. The matrix Q^I is diagonalized to obtain $Q^I = \mathbf{B}\mathbf{D}\mathbf{B}^{-1}$ where

$$\mathbf{D} = \begin{pmatrix} -(\rho+\eta) & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \rho & 1 \\ -\eta & 1 \end{pmatrix}.$$

The probability transition matrix can then be calculated as

$$\begin{aligned} P^I(t) &= \mathbf{B}\left(\sum_{n=0}^{\infty}\frac{\mathbf{D}^n t^n}{n!}\right)\mathbf{B}^{-1} \\ &= \frac{1}{\rho+\eta}\begin{pmatrix} \rho\, e^{-t(\rho+\eta)}+\eta & \rho\left(1-e^{-t(\rho+\eta)}\right) \\ \eta\left(1-e^{-t(\rho+\eta)}\right) & \rho+\eta\, e^{-t(\rho+\eta)} \end{pmatrix} \end{aligned} \tag{2.6}$$

and the steady state probability for state 1 is $\pi_1 = \rho/(\rho+\eta)$.

The first entry of the probability transition matrix is the probability of being in state 0 at time t given that the process was in state 0 at time $t = 0$. This probability transition matrix is identical to the two-state homogeneous Markov process. Note that such a specification implies constraints on the sums of the diagonals and off-diagonals. Data consistent with this model must have $P_{01}(t) + P_{10}(t) < 1$ and $P_{00}(t) + P_{11}(t) > 1$ for any t. The constraint trace$(P^I(t)) > 1$ is referred to as the embeddability criterion discussed by several authors; for example Kingman (1962), Singer and Spilerman (1976), Singer and Cohen (1980) and Kalbfleisch and Lawless (1985). If the embeddability criterion is not satisfied, a discrete-time approach should be considered. See Section 5.4.1 for a detailed outline of the argument.

2.5.2 Type II Counter

As in the Type I counter model, the waiting time for a trigger when the Type II counter is not blocked is exponentially distributed with rate ρ. However, the conditional intensity from state 1 to state 0 now must depend on the number and timing of triggers that occur during the dead time. In this case, an interval of dead time may be followed by another blocked period before returning to state 0. Calculating the probability transition matrix directly is less complicated than working from the intensity matrix in the case of Poisson inputs.

First consider the probability that the counter is open at time t given that the counter is open at time $t = 0$. Suppose exactly M, $M \in \{0, 1, \dots\}$, inputs occur in $(0, t]$. Let $t_1, t_2, \dots, t_M$ be the times of the M inputs with resulting dead times Y_m, $m = 1, \dots, M$. If for all m, $Y_m < t - t_m$, then the counter will be open at time t. Hence,

$$\begin{aligned} P_{00}(t) &= \sum_{M=0}^{\infty} \int_0^t dt_1 \int_{t_1}^t dt_2 \cdots \int_{t_{M-1}}^t dt_M \, \rho^M e^{-\rho t} G(t - t_1) \cdots G(t - t_M) \qquad (2.7) \\ &= \exp\left\{ -\rho \int_0^t du \, \mathcal{G}(t - u) \right\} \\ &= \exp\left\{ -\frac{\rho}{\eta}(1 - e^{-\eta t}) \right\}. \qquad (2.8) \end{aligned}$$

Note that (2.7) holds for the more general case when the dead time distributions are not exponential.

If the counter starts a dead time of length Y_0 at $t = 0$ and the counter is open at time t, then $Y_0 < t$ and inputs must occur in $(0, t]$ in the same manner as stated above. The resulting transition probability is

$$P_{10}(t) = G(t) P_{00}(t) = (1 - e^{-\eta t}) \exp\left\{ -\frac{\rho}{\eta}(1 - e^{-\eta t}) \right\}. \qquad (2.9)$$

Note that this quantity must be interpreted carefully, since the process is not

Markov, and only the beginning of a dead time is a regeneration point. Using (2.8) and (2.9) gives the probability transition matrix

$$P^{II}(t) = \begin{pmatrix} P_{00}(t) & 1 - P_{00}(t) \\ (1 - e^{-\eta t})\, P_{00}(t) & 1 - (1 - e^{-\eta t})\, P_{00}(t) \end{pmatrix}. \tag{2.10}$$

2.6 Incorporating Covariate Information

Some further notation is required to include responses from each subject at the different follow-up times. Longitudinal data are collected on I subjects. Suppose there are n_i observations at times $t_{i1}, t_{i2}, \ldots, t_{in_i}$ such that $t_{i1} < t_{i2} < \cdots < t_{in_i}$ for subject i, $i = 1, \ldots, I$. The time between consecutive observations for subject i is Δt_{ij}, where $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ for $j = 2, \ldots, n_i$, $i = 1, \ldots, I$. The observed, possibly misclassified, response for subject i at observation j is labeled O_{ij}, with history $O_i^{(j)} = (O_{i1}, \ldots, O_{ij})$, and the true state is ξ_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. For notational simplicity, assume that only one misclassification predictor C is available for subject i at observation j, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. The history of the misclassification predictors for subject i is $C_i^{(j)} = (C_{i1}, C_{i2}, \ldots, C_{ij})$, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. The subject-specific misclassification probabilities in (2.1) depend on C_{ij} and become

$$v_{01}(C_{ij}) = \frac{e^{\alpha_0 + \alpha_1 C_{ij}}}{1 + e^{\alpha_0 + \alpha_1 C_{ij}}} \quad \text{and} \quad v_{10}(C_{ij}) = \frac{e^{\alpha_0^* + \alpha_1^* C_{ij}}}{1 + e^{\alpha_0^* + \alpha_1^* C_{ij}}}.$$

Further, define the misclassification parameter vectors $\boldsymbol{\alpha}_{01} = (\alpha_0, \alpha_1)^T$ and $\boldsymbol{\alpha}_{10} = (\alpha_0^*, \alpha_1^*)^T$. This specification implies that the misclassification probabilities depend on the misclassification predictor in exactly the same way for all subjects at all observation times.

Let $\mathbf{X}_i$ be a random vector of baseline covariates for subject i with observed values, $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})$, for some positive integer p. Assume that the rates for the input process and blocked time for a particular individual are based on these baseline covariates and are denoted by $\rho(\mathbf{x}_i)$ and $\eta(\mathbf{x}_i)$, respectively. The relationship between the covariates and the rates is assumed to be log-linear,

$$\rho(\mathbf{x}_i) = e^{\mathbf{x}_i \boldsymbol{\beta}_\rho} \qquad \text{and} \qquad \eta(\mathbf{x}_i) = e^{\mathbf{x}_i \boldsymbol{\beta}_\eta}$$

with regression parameters $\boldsymbol{\beta}_\rho = (\beta_{\rho 0}, \ldots, \beta_{\rho p})^T$ and $\boldsymbol{\beta}_\eta = (\beta_{\eta 0}, \ldots, \beta_{\eta p})^T$. For each individual, the probability transition matrices $\boldsymbol{P}^I$ and $\boldsymbol{P}^{II}$ now also depend on the covariate vector. The parameters ρ and η are replaced by $\rho(\mathbf{x}_i)$ and $\eta(\mathbf{x}_i)$ for subject i, $i = 1, \ldots, I$. Dropping the superscripts indexing the type of counter model, let $\boldsymbol{P}(t, \mathbf{x}_i)$ denote the probability transition matrix for subject i, $i = 1, \ldots, I$,

$$\boldsymbol{P}(t, \mathbf{x}_i) = \begin{pmatrix} P_{00}(t, \mathbf{x}_i) & P_{01}(t, \mathbf{x}_i) \\ P_{10}(t, \mathbf{x}_i) & P_{11}(t, \mathbf{x}_i) \end{pmatrix}$$

which depends on the parameter vector $\boldsymbol{\Theta} = (\alpha_{01}, \alpha_{10}, \boldsymbol{\beta}_\rho, \boldsymbol{\beta}_\eta)$. Subjects with identical covariates will have identical probability transition matrices resulting in potentially fewer than I distinct matrices.

The steady state probability for subject i is labeled $\pi_1(\mathbf{x}_i) = \Pr(\xi_{ij} = 1, \mathbf{x}_i)$ for state 1,$i = 1, \ldots, I$. In the Type I counter model, this steady state probability is $\pi_1(\mathbf{x}_i) = \rho(\mathbf{x}_i)/(\rho(\mathbf{x}_i) + \eta(\mathbf{x}_i))$. With the definition $\gamma_{ij}(\mathbf{x}_i) = \Pr(\xi_{ij} = 1 \mid O_i^{(j)}, C_i^{(j)}, \mathbf{x}_i)$, (2.2) becomes

$$\begin{aligned}
&\Pr(O_{ij} = 1 \mid O_i^{(j-1)}, C_i^{(j)}, \mathbf{x}_i) \\
&\qquad = v_{11}(C_{ij}) \Big[(1 - \gamma_{i,j-1}(\mathbf{x}_i))\, P_{01}(\Delta t_{ij}, \mathbf{x}_i) + \gamma_{i,j-1}(\mathbf{x}_i)\, P_{11}(\Delta t_{ij}, \mathbf{x}_i)\Big] \\
&\qquad\quad + v_{01}(C_{ij}) \Big[(1 - \gamma_{i,j-1}(\mathbf{x}_i))\, P_{00}(\Delta t_{ij}, \mathbf{x}_i) + \gamma_{i,j-1}(\mathbf{x}_i)\, P_{10}(\Delta t_{ij}, \mathbf{x}_i)\Big].
\end{aligned}$$

The likelihood involving all the observations for each subject is

$$L(\Theta) = \prod_{i=1}^{I} \Pr(O_{i1} \mid C_{i1}, \mathbf{x}_i; \Theta) \prod_{j=2}^{n_i} \Pr(O_{ij} \mid O_i^{(j-1)}, C_i^{(j)}, \mathbf{x}_i; \Theta) \tag{2.11}$$

where

$$\Pr(O_{i1} \mid C_{i1}, \mathbf{x}_i; \Theta) = \begin{cases} v_{00}(C_{i1})\, \pi_0(\mathbf{x}_i) + v_{10}(C_{i1})\, \pi_1(\mathbf{x}_i) & \text{if } O_{i1} = 0 \\ v_{01}(C_{i1})\, \pi_0(\mathbf{x}_i) + v_{11}(C_{i1})\, \pi_1(\mathbf{x}_i) & \text{if } O_{i1} = 1. \end{cases}$$

When the misclassification probabilities are constants such that $v_{01} = 0$ and $v_{10} > 0$, and the unobservable process is modeled as a two-state homogeneous Markov chain, the model here becomes the model proposed by Nagelkerke, Chunge and Kinoti (1990).

Not all possible ranges of misclassification probabilities are permissible in the model. In particular, if $v_{01}(C_{ij}) + v_{10}(C_{ij}) = 1$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$, then it turns out that $v_{00}(C_{ij}) = v_{10}(C_{ij})$ and $v_{11}(C_{ij}) = v_{01}(C_{ij})$. The probability of the observed state is completely independent of the unobservable true state. If

$O_i^{(0)}$ and $C_i^{(0)}$ are defined to be empty sets, then

$$\Pr(O_{ij} \mid O_i^{(j-1)}, C_i^{(j)}, \mathbf{x}_i) = \begin{cases} v_{10}(C_{ij}) & \text{if } O_{ij} = 0 \\ v_{01}(C_{ij}) & \text{if } O_{ij} = 1 \end{cases}$$

for $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. The likelihood will be flat with respect to the regression parameters. Mathematically, a situation with $v_{01}(C_{ij}) + v_{10}(C_{ij}) = 1$ is not informative, although this situation may be possible theoretically. In practice, it is unlikely to have exact equality when multiple misclassification predictors may be available.

Note that the unobservable process modeled as a counter model is completely distinct from any misclassification specification. The counter model approach could be used directly if misclassification was not an issue. The distributions for the state sojourns depend on the baseline covariates for each subject.

The misclassification probabilities at a particular observation time only depend on the misclassification predictors collected at that observation. Hence, the misclassification probabilities only depend on the value of the misclassification predictors and will be the same for any subject at any observation time with the same misclassification predictor values. The misclassification predictors could come from an auxiliary series or could be functions of $O_i^{(j-1)}$. For example, it would be possible to have misclassification probabilities depend on the time since last apparent transition as well as the misclassification predictors. However, such misclassification probabilities should not depend on the true process parameters. When the misclassification predictors come from functions of $O_i^{(j-1)}$, the misclassification could be termed differential with respect to the previous history of the process (see Chen (1989), for example). Conversely, the misclassification is non-differential with respect to the previous history of the process if an auxiliary series is used for the misclassification predictors.

Although the model has several parameters requiring estimation, these parameters are able to quantify the effect each covariate has on each transition rate as well as the effect each misclassification predictor has on each misclassification probability.

2.7 Jackknife Standard Error Estimation

The likelihood given in (2.11) is recursive and has several parameters even when only a few covariates are involved. Calculating the partial derivatives necessary to obtain an information matrix for parameter standard error estimates is prohibitive. A jackknife estimate of parameter standard errors can be calculated instead based on dropping one subject or a group from the estimation at a time. Many authors have considered jackknife estimates for bias and variance including Quenouille (1949), Tukey (1958), Efron (1979) and Wolter (1985).

Let $\widehat{\boldsymbol{\Theta}}$ be the maximum likelihood estimate of $\boldsymbol{\Theta}$ based on the data from all subjects and let $\widehat{\boldsymbol{\Theta}}_{(i)}$ be the maximum likelihood estimate of $\boldsymbol{\Theta}$ based on the data without subject i, $i = 1, \ldots, I$. The drop-1 jackknife estimate of $\boldsymbol{\Theta}$ is written as

$$\widehat{\boldsymbol{\Theta}}_{(\cdot)} = \frac{1}{I}\sum_{i=1}^{I} \widehat{\boldsymbol{\Theta}}_{(i)}.$$

and the jackknife estimated variance-covariance matrix for the parameters is

$$\widehat{V}(\widehat{\Theta}) = \frac{I-1}{I} \sum_{i=1}^{I} \left(\widehat{\Theta}_{(i)} - \widehat{\Theta}_{(\cdot)}\right) \left(\widehat{\Theta}_{(i)} - \widehat{\Theta}_{(\cdot)}\right)^T .$$

More generally, suppose the subjects are partitioned into groups $g_1, \ldots, g_k$ with m subjects each. Let $\widehat{\Theta}_{(g_\ell)}$ be the maximum likelihood estimate of Θ based on the data without the subjects in group g_ℓ , $\ell = 1, \ldots, k$. The drop-m jackknife estimated variance-covariance matrix for the parameters is

$$\widehat{V}_k(\widehat{\Theta}) = \frac{k-1}{k} \sum_{\ell=1}^{k} \left(\widehat{\Theta}_{(g_\ell)} - \widehat{\Theta}_{(\cdot)}\right) \left(\widehat{\Theta}_{(g_\ell)} - \widehat{\Theta}_{(\cdot)}\right)^T$$

where

$$\widehat{\Theta}_{(\cdot)} = \frac{1}{k} \sum_{\ell=1}^{k} \widehat{\Theta}_{(g_\ell)}.$$

Drop-1 jackknife variance-covariance estimates will be calculated when the computational time is feasible. Otherwise, simple random sampling without replacement will determine the groups of size m to be dropped for variance-covariance estimates.

The jackknife estimate of standard errors only requires computation of $\widehat{\Theta}_{(g_\ell)}$ for each group of size m. One alternative method to consider is the bootstrap (Efron (1979)). Bootstrap methods may require 100 or 200 replicates for standard error estimation (Efron and Tibshirani (1993)). These values are substantially larger than the numbers of subjects contained in the data sets described in the next chapter, and thus, the jackknife requires far less computation. In addition to the computational advantage for the data examined hereafter, the jackknife considers data sets which are similar to the original data and may help to identify influential observations. Further investigations could be done to see if bootstrap methods give more stable variance estimates.

Chapter 3

Data Description and Preliminary Analyses

3.1 Overview

A relapsing-remitting Multiple Sclerosis clinical trial and a longitudinal study of childhood parasitic infection are introduced. Type I counter models with misclassification are fit using each data set. The models considered assume known misclassification probabilities or estimate misclassification probability parameters and regression parameters. The Multiple Sclerosis data set yields extreme parameter estimates for some models with misclassification predictors and covariates.

3.2 Introduction

Two biostatistical data sets are available to demonstrate methodology. The first is a substudy of a larger clinical trial for the evaluation of a new treatment in relapsing-remitting Multiple Sclerosis (MS). The second data set is the parasitic infection (PI) data modeled in Nagelkerke, Chunge and Kinoti (1990). The descriptions of these data sets appear in Sections 3.3 and 3.4 along with preliminary analyses based on a Type I counter model for the unobservable process. In these analyses, models are considered which have assumed known misclassification probabilities and models are considered for which the misclassification probabilities are estimated.

3.3 Multiple Sclerosis Clinical Trial

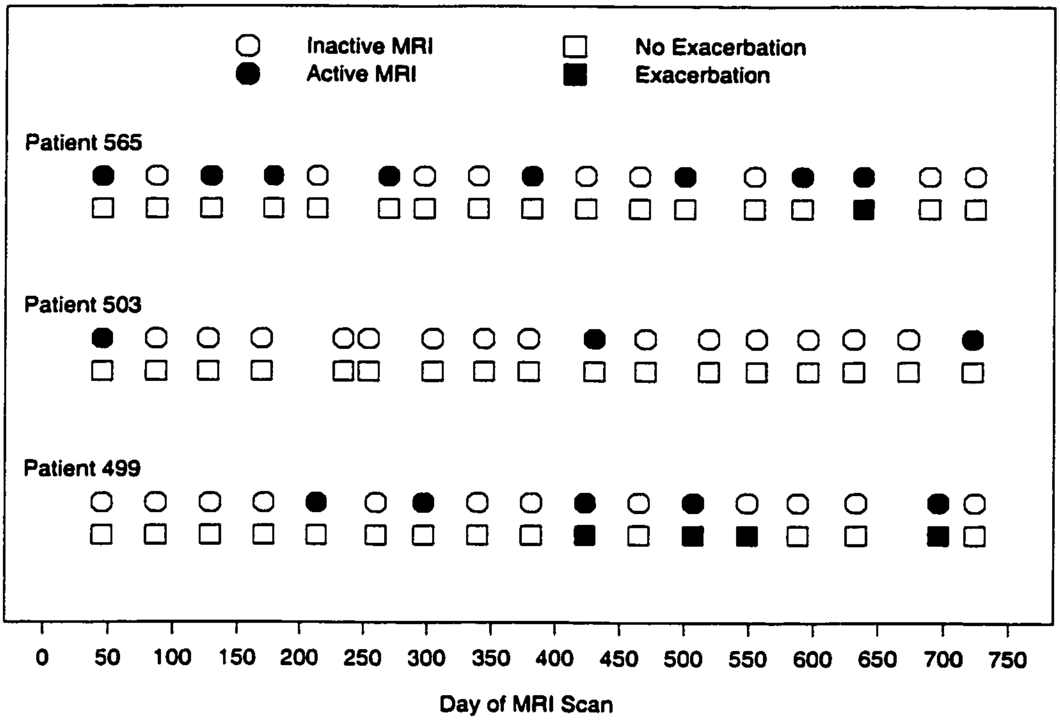
A multicenter, randomized, double-blind, placebo-controlled drug trial was conducted on 372 ambulatory patients with relapsing-remitting MS. Patients experience unforeseeable attacks (exacerbations) which are followed by periods of remission. Interferon beta-lb (IFNB) was the drug treatment under investigation with placebo, 1.6 million international units (MIU) IFNB and 8 MIU IFNB treatment arms. Enrollment started in June 1988 and patients were followed for two years. At the end of two years, patients had the option of continuing treatment for a maximum of 2.5 more years. A substudy of the trial, involving fifty-two subjects, was conducted at the University of British Columbia (UBC). These patients had cranial magnetic resonance imaging (MRI) every six weeks for two years. Data were collected at the UBC MS Clinic as part of the UBC Frequent MRI Substudy of the

Betaseron Clinical Trial sponsored by Berlex Laboratories, Richmond, California and provided by Professor A. John Petkau, Department of Statistics, UBC.

Fifty patients from the UBC substudy constitute the data available. Several variables were measured at the baseline scan (MRI). The patient's age in years, gender, duration of the disease in years, treatment group, origin and burden of disease were recorded as baseline covariates. The Expanded Disability Status Scale (EDSS) for each patient at baseline was also collected. This scale is an ordinal scale, ranging from 0 to 10 with increments of 0.5, and a clinical standard in MS neurological impairment assessment. A scale of 0 corresponds to normal neurological examinations, 5–6 means the patient requires walking assistance of varying degrees and 10 represents death due to MS (Kurtzke (1983)). All patients had EDSS scores of 5.5 or less. Origin indicates if a patient was a British Columbia or Washington State resident. Burden of disease, measured at each scan time, is the total area of MS lesions from all slices of an MRI scan.

At each follow-up scan, three variables were considered response variables: burden of disease, active scans and exacerbations. Each scan is evaluated as active (1) or inactive (0) based on comparison with the immediately preceding scan. A scan was considered active if a new lesion appeared, a lesion recurred at the same position of an earlier lesion that had subsequently disappeared, or if an existing lesion enlarged "significantly", when compared to its immediate predecessor. An exacerbation is defined as the appearance of a new symptom or worsening of an old symptom attributable to MS. Examples of such symptoms include disturbances in visual acuity, bowel and bladder incontinence, and motor disturbances involv-

Figure 3.1: Active scans and exacerbations at MRI scan times for three MS study subjects.



ing walking (Stienman (1996)). At each scan, it is noted if an exacerbation had occurred (1) or not (0) in the period since the last scan. Figure 3.1 displays the activity status and exacerbation status of three study subjects. Paty et al. (1993) and Petkau and White (1995) provide more information about the UBC substudy.

Not all data are complete: some patients missed scans intermittently and 5 patients dropped out after 12 or more visits. Although a subject may have missed one or more of the intermediate scans, information may still be available to determine whether or not an exacerbation occurred, either between the last scan and the missed scan time or the missed scan time and the next scan.

Individual cranial MRIs are not well correlated with standard impairment scales of clinical status (Lohr and Schroeder (1990)). MRIs in this study, along with clinical assessment, are used to monitor the trial. The relationship between the activity status of a scan and the true relapsing or remitting status of the patient is of interest. The definition of a relapse is not precise and involves both concepts of activity and exacerbation. According to McAlpine's definition of relapse reported in Matthews (1991), Kesselring (1997) defines a relapse as

> ... the occurrence of a new symptom, or the recurrence of a symptom previously present. Symptoms should be considered to be an expression of a relapse if they can be explained as due to a new lesion within the CNS [Central Nervous System] or to a reactivation of lesions previously present ...

The beginning of a relapse may be seen in the MRI scan and the clinical symptoms of an exacerbation may follow. The scan activity may misclassify the true relapsing or remitting state if new, recurring or significantly enlarging lesions are not identified by a technician or if new or recurring lesions are not consistent with the appearance or the recurrence of a symptom. Examining how the exacerbation and activity status relate may lead to an etiological understanding of the disease process and enable better evaluation of potential treatments.

The goals are to model the misclassification in the activity status and the true relapsing-remitting status. Studies, such as Sibley, Bamford and Clark (1985) and Panitch et al. (1992), have suggested that intercurrent clinical viral infections trigger exacerbations and may be the most dominant environmental influence in the disease process (Sibley et al. (1991)). These "triggers" motivate the use of counter models. The relapsing and remitting states are denoted by 1 and 0, respectively. **The exacerbation status will only be used in the misclassification part of the model and will be considered the misclassification predictor.** In this case, the likelihood in (2.11) will be a partial likelihood (Cox (1975)) since the misclassification predictor is an internal time-varying covariate, with relationship to the true process being modeled independently of the true process parameters. If an exacerbation has occurred in the time since the last scan, the patient may still be in a relapsing period and an active scan would be consistent with the true status of the patient. The modeling will incorporate the baseline covariates listed above.

3.3.1 Type I Counter Analysis

Of the 50 patients, 38 had complete data for all 17 MRI scans. This reduced data set is referred to as the complete data set. The activity of an MRI scan is considered to be the observed state which is possibly misclassified and the occurrence or not of an exacerbation between scans is the misclassification predictor. For subject i define

$$\begin{aligned}
O_{ij} &= \begin{cases} 1 & \text{if scan } j \text{ is active} \\ 0 & \text{if scan } j \text{ is not active} \end{cases} \\
C_{ij} &= \begin{cases} 1 & \text{if an exacerbation has occurred between scans } j-1 \text{ and } j \\ 0 & \text{if an exacerbation has not occurred between scans } j-1 \text{ and } j \end{cases} \\
\xi_{ij} &= \begin{cases} 1 & \text{if the subject is in a relapsing period at scan } j \\ 0 & \text{if the subject is not in a relapsing period at scan } j \end{cases}
\end{aligned}$$

for $i = 1, \ldots, 38$, $j = 1, \ldots, 17$, and scan 0 refers to the baseline scan.

For these patients, the first scan is active for 7 individuals while it is inactive for 31. The four possible transitions, 0-0, 0-1, 1-0 and 1-1, have counts 400, 78, 75 and 55, respectively. The continuous covariates, age, duration and initial burden (burden of disease at the baseline scan) for the complete data set are standardized to facilitate numerical stability. The means and standard deviations for these covariates appear in Table 3.1. The inter-scan times ranged from 20 days to 70 days with 50% of the times falling between 42 and 44 days.

Table 3.1: Means and standard deviations for continuous covariates in the complete case MS data set.

Covariate	Mean	Std.Deviation
Age	36.079	7.205
Duration	9.158	6.223
Initial Burden	2124.789	2344.583

All computations are provided by the author's C program which incorporates the linear algebra package developed by Stuber (1996). A direction set (Powell's) method provided in Press et al. (1992) is used to maximize the log-likelihood for the models considered in the following sections. This method is preferred when derivatives are not easily calculated and is fairly simple to implement. Given the recursive nature of the observed state dependence on all previous states through γ_{ij}, finding the derivatives would involve complicated expressions. The algorithm is started at several different starting conditions and is stopped when the log-likelihood function fails to increase by more than 10^{-6} on one iteration.

The models considered in the next few sections generally take far less than

1000 function evaluations for each starting point when at most one covariate is included. Of course, the computational burden increases with more covariates. The lack of derivative calculations prevents evaluation of standard errors by use of an information matrix.

Models which do not depend on the covariates or exacerbation status are the simplest cases to consider first. Estimating both misclassification probabilities and both transition rates is problematic in such cases. More than one set of parameter values can give the same probabilities for possible data sets. This situation is referred to as an identifiability problem and issues of identifiability are deferred to Chapter 4. For now, consider models which have either known values for each misclassification probability or only one type of misclassification possible.

Assumed Misclassification Probabilities

Consider three classes of models where the misclassification probabilities are known, small values. Misclassification is assumed to be independent of the exacerbation data. Misclassification probabilities, input rate and blocked period rate are assumed to be the same for all subjects across all scan times. No covariate data are included at this time. The three classes of models are referred to as perfect sensitivity, perfect specificity, and imperfect sensitivity and specificity.

In the perfect sensitivity model, if the true relapsing-remitting status is relapsing, the observed state will be active. The misclassification probabilities are

$$\begin{aligned} v_{01}(C_{ij}) &= \Pr(O_{ij} = 1 \,|\, \xi_{ij} = 0) = \frac{e^{\alpha_0}}{1 + e^{\alpha_0}} = v_{01} \\ v_{10}(C_{ij}) &= \Pr(O_{ij} = 0 \,|\, \xi_{ij} = 1) = 0 \end{aligned}$$

for $i = 1, \ldots, 38$, $j = 1, \ldots, 17$.

If the true relapsing-remitting status is remitting, the observed state will not be active in the perfect specificity model. For this case, the misclassification probabilities are

$$v_{01}(C_{ij}) = 0 \qquad \text{and} \qquad v_{10}(C_{ij}) = \frac{e^{\alpha_0^*}}{1 + e^{\alpha_0^*}} = v_{10}$$

for $i = 1, \ldots, 38$, $j = 1, \ldots, 17$. This model corresponds with the situation assumed by Nagelkerke, Chunge and Kinoti (1990).

The imperfect sensitivity and specificity model allows both misclassification probabilities to be nonzero. Hence,

$$v_{01}(C_{ij}) = \frac{e^{\alpha_0}}{1 + e^{\alpha_0}} = v_{01} \qquad \text{and} \qquad v_{10}(C_{ij}) = \frac{e^{\alpha_0^*}}{1 + e^{\alpha_0^*}} = v_{10}$$

for $i = 1, \ldots, 38$, $j = 1, \ldots, 17$.

Since an underlying counter model for the MS data is assumed, in all three cases with a 28-day time scale, the probability transitions are equal to

$$\begin{aligned} P_{01}(\Delta t_{ij}/28) &= \frac{\rho}{\rho + \eta}\left(1 - e^{-\frac{\Delta t_{ij}}{28}(\rho+\eta)}\right) \quad j \neq 1 \\ P_{10}(\Delta t_{ij}/28) &= \frac{\eta}{\rho + \eta}\left(1 - e^{-\frac{\Delta t_{ij}}{28}(\rho+\eta)}\right) \quad j \neq 1 \end{aligned}$$

where $\rho = \exp(\beta_{\rho 0})$, $\eta = \exp(\beta_{\eta 0})$, and $\pi_1 = \rho/(\rho + \eta)$, for $i = 1, \ldots, 38$ and $j = 1, \ldots, 17$.

The parameter estimates, rates and estimated mean times for the three types of models with assumed misclassification probabilities appear in Table 3.2. Estimated mean times are calculated as the reciprocal of the state transition rates. For the perfect sensitivity and specificity models, as the respective misclassification probability increases, the parameter estimates and rates become smaller and the estimated mean times increase (Satten and Longini (1996)). The estimated mean times are smallest when $v_{01} = v_{10} = 0.01$ and largest when $v_{01} = v_{10} = 0.05$ in the imperfect sensitivity and specificity model. For a specific value of v_{01}, the estimated mean times increase as v_{10} increases. Similarly, the estimated mean times increase as v_{01} increases, for a specific value of v_{10}. Symmetry in the misclassification is not apparent since the estimated mean times are larger when $v_{01} = 0.05$ and $v_{10} = 0.01$ than when $v_{01} = 0.01$ and $v_{10} = 0.05$. In all cases, the estimated mean time in the remitting state is at least three times larger than the estimated mean time in the relapsing state.

The log-likelihood function is plotted for some specified values of the misclassification probabilities in Figure 3.2. As the misclassification probabilities increases, the log-likelihood function becomes flatter and is stretched less in the $\beta_{\eta 0} - \beta_{\rho 0}$ (or η/ρ) direction than in the $\beta_{\eta 0} + \beta_{\rho 0}$ (or $\rho\eta$) direction. This feature indicates that it may be hard to estimate both ρ and η well individually whereas η/ρ may be more easily estimated.

Unknown Misclassification Probabilities

The misclassification probability is quite large in the perfect specificity model when v_{10} is not assumed known. Both state occupancy times are about the same as the total time on study. Clearly, these results indicate that subjects are not making any

Table 3.2: MS estimates under assumed misclassification probabilities for the perfect sensitivity, perfect specificity, and imperfect sensitivity and specificity models. State 0 is remitting, state 1 is relapsing, and EMT is the estimated mean time in days. Jackknife standard errors are given in parentheses.

Perfect Sensitivity

	Assumed Misclassification Probabilities			
Quantity	v_{01}=0.01	v_{01}=0.05	v_{01}=0.10	v_{01}=0.15
$\beta_{\rho 0}$	-1.704(0.249)	-2.111(0.337)	-2.718(0.448)	-3.296(0.529)
ρ	0.182	0.121	0.066	0.037
State 0 EMT	154	231	424	756
$\beta_{\eta 0}$	-0.383(0.244)	-0.562(0.277)	-0.832(0.310)	-1.045(0.339)
η	0.682	0.570	0.435	0.352
State 1 EMT	41	49	64	80
log-likelihood	-321.064	-319.809	-318.950	-320.849

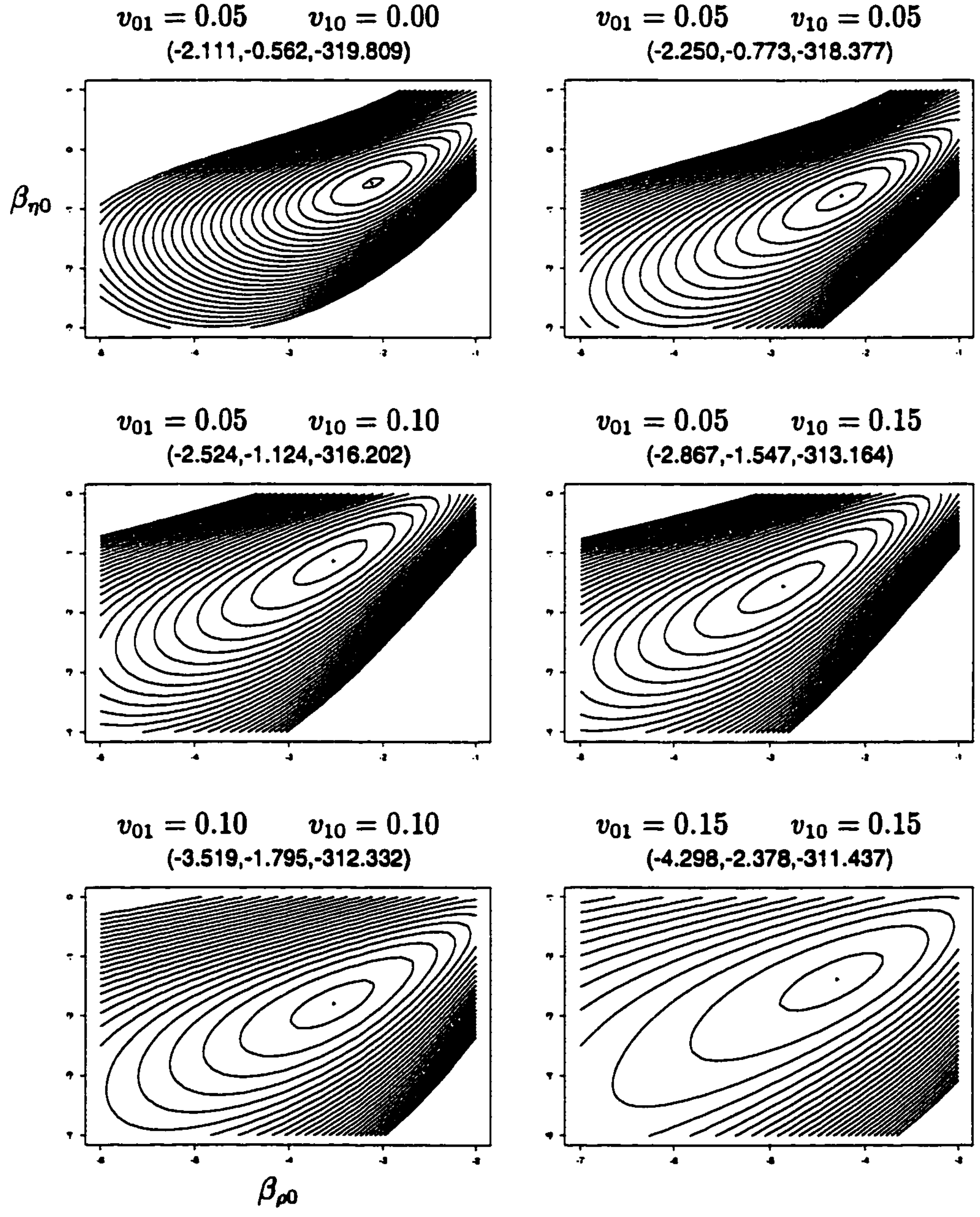
Perfect Specificity

	Assumed Misclassification Probabilities			
Quantity	v_{10}=0.01	v_{10}=0.05	v_{10}=0.10	v_{10}=0.15
$\beta_{\rho 0}$	-1.626(0.237)	-1.659(0.255)	-1.718(0.287)	-1.803(0.330)
ρ	0.197	0.190	0.179	0.165
State 0 EMT	142	147	156	170
$\beta_{\eta 0}$	-0.368(0.244)	-0.460(0.274)	-0.600(0.325)	-0.774(0.390)
η	0.692	0.631	0.549	0.461
State 1 EMT	40	44	51	61
log-likelihood	-321.222	-320.609	-319.658	-318.437

Imperfect Sensitivity and Specificity

	Assumed Misclassification Probabilities			
Quantity	v_{01}=0.01 v_{10}=0.01	v_{01}=0.01 v_{10}=0.05	v_{01}=0.05 v_{10}=0.01	v_{01}=0.05 v_{10}=0.05
$\beta_{\rho 0}$	-1.712(0.253)	-1.754(0.277)	-2.132(0.349)	-2.250(0.437)
ρ	0.181	0.173	0.119	0.105
State 0 EMT	155	162	236	266
$\beta_{\eta 0}$	-0.405(0.251)	-0.506(0.288)	-0.597(0.293)	-0.773(0.392)
η	0.667	0.603	0.550	0.462
State 1 EMT	42	46	51	61
log-likelihood	-230.912	-320.226	-319.562	-318.377

Figure 3.2: Plots of the MS log-likelihood function for specified values of the misclassification probabilities. Each contour after the first represents a drop of 2 in the log-likelihood function. In parentheses, $\hat{\beta}_{\rho 0}$ and $\hat{\beta}_{\eta 0}$ are given along with the maximum value of the log-likelihood function. ($\beta_{\rho 0} = \ln \rho$, $\beta_{\eta 0} = \ln \eta$)



$v_{01} = 0.05 \quad v_{10} = 0.00$ (-2.111,-0.562,-319.809)

$v_{01} = 0.05 \quad v_{10} = 0.05$ (-2.250,-0.773,-318.377)

$v_{01} = 0.05 \quad v_{10} = 0.10$ (-2.524,-1.124,-316.202)

$v_{01} = 0.05 \quad v_{10} = 0.15$ (-2.867,-1.547,-313.164)

$v_{01} = 0.10 \quad v_{10} = 0.10$ (-3.519,-1.795,-312.332)

$v_{01} = 0.15 \quad v_{10} = 0.15$ (-4.298,-2.378,-311.437) state transitions during the study time and suggest that there is a problem with the model. The perfect sensitivity model gives more reasonable estimates, perhaps because the model is close to identifiable. The results for these two models appear in Table 3.3.

Table 3.3: MS estimates under perfect sensitivity and perfect specificity with constant input and blocked rates (ρ and η). Misclassification probabilities are estimated. State 0 is the remitting and state 1 is relapsing. Jackknife standard errors are given in parentheses.

Perfect Sensitivity

Parameter	Estimate (se)	Quantity	Estimate	Estimated Mean Time	
$\beta_{\rho 0}$	-2.667 (0.557)	ρ	0.070	State 0	400 days
$\beta_{\eta 0}$	-0.810 (0.342)	η	0.445	State 1	63 days
α_0	-2.245 (0.410)	v_{01}	0.096		
				log-likelihood:	-318.940

Perfect Specificity

Parameter	Estimate (se)	Quantity	Estimate	Estimated Mean Time	
$\beta_{\rho 0}$	-3.086 (0.437)	ρ	0.046	State 0	609 days
$\beta_{\eta 0}$	-3.233 (0.656)	η	0.039	State 1	718 days
α_0^*	0.382 (0.316)	v_{10}	0.594		
				log-likelihood:	-303.959

Misclassification with a Single Covariate

Now consider the full misclassification model which is dependent on the exacerbation status and for which the transition rates depend on one covariate, x. The model with covariates becomes

$$
\begin{aligned}
v_{01}(C_{ij}) &= \Pr(O_{ij} = 1 \,|\, \xi_{ij} = 0) = \frac{e^{\alpha_0 + \alpha_1 C_{ij}}}{1 + e^{\alpha_0 + \alpha_1 C_{ij}}} \\
v_{10}(C_{ij}) &= \Pr(O_{ij} = 0 \,|\, \xi_{ij} = 1) = \frac{e^{\alpha_0^* + \alpha_1^* C_{ij}}}{1 + e^{\alpha_0^* + \alpha_1^* C_{ij}}} \\
\rho(x_i) = e^{\beta_{\rho 0} + \beta_{\rho 1} x_i} \qquad & \eta(x_i) = e^{\beta_{\eta 0} + \beta_{\eta 1} x_i} \qquad \pi_1(x_i) = \frac{\rho(x_i)}{\rho(x_i) + \eta(x_i)} \\
P_{01}(\Delta t_{ij}/28, x_i) &= \frac{\rho(x_i)}{\rho(x_i) + \eta(x_i)} \left(1 - e^{-\frac{\Delta t_{ij}}{28}(\rho(x_i) + \eta(x_i))}\right) \quad j \neq 1 \\
P_{10}(\Delta t_{ij}/28, x_i) &= \frac{\eta(x_i)}{\rho(x_i) + \eta(x_i)} \left(1 - e^{-\frac{\Delta t_{ij}}{28}(\rho(x_i) + \eta(x_i))}\right) \quad j \neq 1
\end{aligned}
\tag{3.1}
$$

for $i = 1, \ldots, 38$ and $j = 1, \ldots, 17$. The input and blocked rates will be different for different subjects based on individual subject covariates. Only one covariate is considered at a time and the resulting estimates appear in Table 3.4. The continuous covariates are standardized for better numerical stability. The EDSS is an ordinal scale and is not addressed at this time. The estimated mean times are calculated for continuous covariates at median values -0.080, -0.033, and -0.217 for age, initial burden, and duration, respectively.

Several features can be seen in the table. Many of the effect sizes are much more extreme than expected. In proportional hazards regression, effect sizes are rarely beyond -3 or +3. In the model discussed here, the relationship between the covariates and the input process and blocked times has a similar form as in proportional hazards regression. Particularly for the standardized continuous variables, the coefficient values seem too extreme. For example, the estimate of $\beta_{\rho 1}$ in the age model is -6.227. This value implies that a one standardized unit increase in age results in 0.2% of the trigger rate, an increase in the estimated mean time remitting

Table 3.4: MS single covariate model estimates under imperfect sensitivity and specificity. Age, initial burden (IB), and duration are standardized. Estimated mean state occupancy times (EMTs) are given in days for the median values of the continuous covariates and at each level of the categorical covariates. (L=low dose, H=high dose, F=female, B=BC resident)

Parameters	Age	IB	Duration	Sex	Treatment	Origin
$\rho(x)$ $\beta_{\rho 0}$	-10.056	-4.800	-6.212	-39.231	-4.228	-0.482
$\beta_{\rho 1}$	-6.227	0.277	-2.794	35.054F	0.285L -0.973H	-3.625B
State 0 EMT	395,484	3,434	7,615	$>10^{17}$ 1,825F	1,920 1,444L 5,081H	45 1,701B
$\eta(x)$ $\beta_{\eta 0}$	-9.294	-4.287	-5.852	-39.040	-4.830	-1.324
$\beta_{\eta 1}$	-5.514	-0.905	-2.385	35.762F	1.814L 0.975H	-3.522B
State 1 EMT	195,469	1,976	5,806	$>10^{17}$ 743F	3,506 571L 1,322H	105 3,562B
$v_{01}(C_{ij})$ α_0	-2.485	-2.544	-2.797	-2.491	-2.587	0.218
α_1	-1.092	-1.879	-78.152	-1.200	-21.774	-0.147
$v_{01}(0)$	0.077	0.073	0.057	0.076	0.070	0.554
$v_{01}(1)$	0.027	0.012	$<10^{-10}$	0.024	$<10^{-10}$	0.518
$v_{10}(C_{ij})$ α_0^*	-0.016	0.096	0.228	-0.110	0.081	2.559
α_1^*	0.151	0.106	0.076	0.228	0.079	20.374
$v_{10}(0)$	0.496	0.524	0.557	0.473	0.520	0.928
$v_{10}(1)$	0.534	0.550	0.575	0.529	0.540	1.000
log-likelihood	-288.220	-290.782	-291.302	-292.823	-291.125	-290.318

of more than a factor of 500. The situation is more dramatic when sex is considered as the only covariate. The rates for males are essentially zero.

Most of the estimated mean times are very large, suggesting that the estimated process is not transitioning. The misclassification probabilities are mostly small for $v_{01}(\cdot)$ and large for $v_{10}(\cdot)$. Several causes may be contributing to such extreme results. The type of counter model may be incorrect as well as the exponential assumption. The single covariate models may not provide enough information on the process. The link between the covariates and rates may be inappropriate. Issues of model specification are further discussed in Chapter 6.

The misclassification probabilities appear to be different for different values of the exacerbation status in some of the models. In general, if the observed activity status agrees with the exacerbation status then the misclassification probabilities are lower than when these two indicators disagree. The estimates for $v_{10}(C_{ij})$ are very similar for both values of the exacerbation indicator as well as quite large for all of the models fit. These estimates are fairly consistent with the results from the perfect specificity model in Table 3.3. The estimates of $v_{01}(C_{ij})$ are smaller for most of the models than the perfect sensitivity model previously fit.

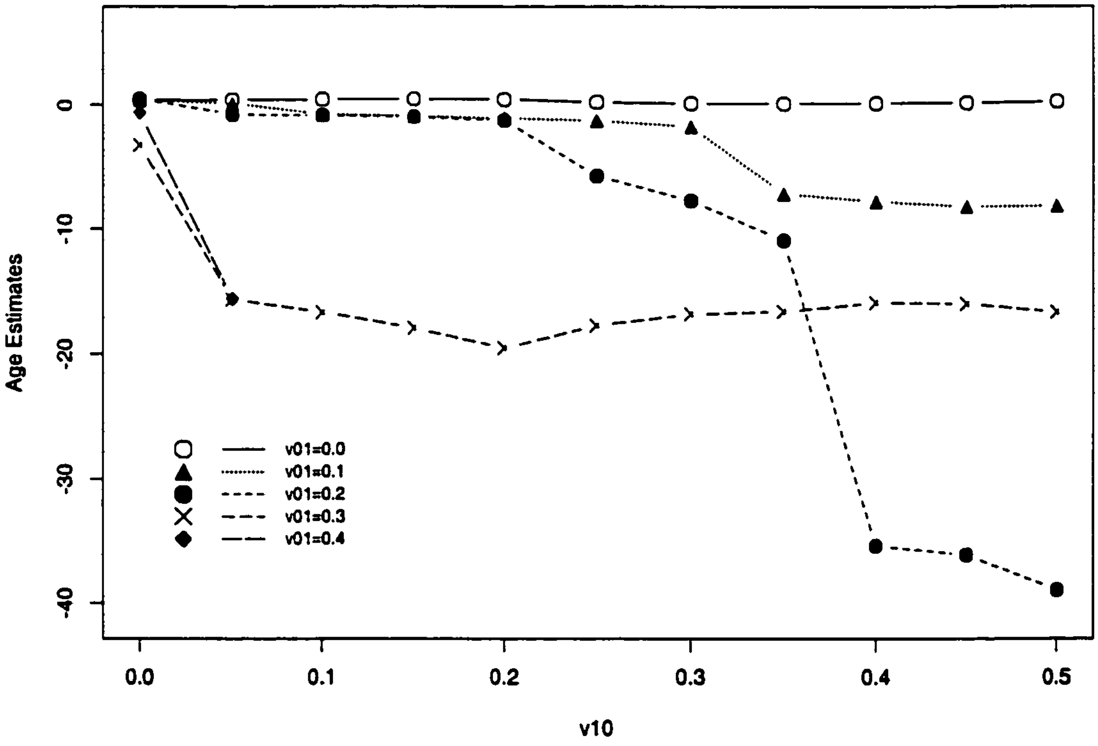
Sensitivity Analysis in Age Model

The results in the previous sections seem to give unreasonable estimates for the mean time spent in a state. An examination of the behaviour of regression parameter estimates for different misclassification probabilities is conducted next. This comparison is carried out with the standardized covariate age model.

Consider the model given in (3.1) depending on the standardized covariate age along with a simplification. Specified misclassification probabilities which are independent of the misclassification predictor are investigated. The model is fit with different assumed values of v_{01} and v_{10}. The misclassification probabilities are allowed to take on values in 0.05 increments ranging from 0 to 0.95. In total, 390 combinations of v_{01} and v_{10} are specified and the regression parameters are estimated. However, not all combinations produce unique estimates. The closer $v_{01} + v_{10}$ is to 1, the less information there is about the other parameters. The likelihood function becomes fairly flat when $v_{01} + v_{10}$ is close to 1 and the maximization algorithm may find more than one maximum. Recall that the maximization algorithm is started at several different starting conditions and is stopped when the log-likelihood function fails to increase by more than 10^{-6} on one iteration.

Plots of regression parameter estimates for varying degrees of misclassification can be examined. The effect of different misclassification probabilities on $\hat{\beta}_{\eta 1}$ for age is displayed in Figure 3.3 as an example. It is clear that for larger values of v_{01} and v_{10} the estimates seem more extreme. For values of v_{01} less than 0.2, the estimates appear to be generally smaller in magnitude for lower values of v_{10}. For larger values, the estimates may be large negative numbers. The value of the estimate is clearly highly dependent on the degree of misclassification present. The intercept estimates, $\hat{\beta}_{\rho 0}$ and $\hat{\beta}_{\eta 0}$, also depend closely on the misclassification probabilities. These estimates become more extreme as $v_{01} + v_{10}$ approaches 1. For values of v_{01} less than 0.2, $\hat{\beta}_{\eta 1}$ seems to decrease for small values of v_{10}. Clearly, these estimates depend on small misclassification probabilities in a more controlled manner than for larger misclassification probabilities. The dependence on small misclassification

Figure 3.3: MS age model estimates of $\hat{\beta}_{\eta 1}$ for age under different specified misclassification probabilities, v_{01} and v_{10}. Only unique estimates are plotted.



probabilities will be further examined in Chapter 5.

3.4 Parasitic Infection Longitudinal Data

A study with young children was conducted in the mid-1980's to examine the incidence and prevalence of a flagellated protozoal parasite which lives in the small intestine. The parasite, called *Giardia lamblia*, is well known to be difficult to detect. Parasitic infection of this type may cause health problems such as diarrhoea and loss of appetite. The parasite is transmitted between subjects when infected individuals excrete viable cysts in feces which are subsequently swallowed by uninfected individuals. The prevalence is relatively high among children. In a village near Nairobi, Kenya, eighty-four children between the ages of 11 and 18 months were studied for 44 consecutive weeks.

Each week, the stools of the children were examined for the presence of the parasite. The data is not complete since not all children were examined each week for various reasons. Nagelkerke, Chunge and Kinoti (1990) selected the longest run of at least 10 uninterrupted weekly examinations for each child in their analysis. This selection resulted in 58 children who have between 10 and 44 observations. The four possible transitions, 0-0, 0-1, 1-0 and 1-1, have counts 378, 119, 112 and 321, respectively. Covariate information and misclassification predictors are not available for this study. Chunge (1989) gives the full details of the study and Appendix A.2 lists the parasitic infection (PI) data for the 58 children.

3.4.1 Assumed Misclassification Probabilities

The same misclassification probability models considered for the MS data are also examined here for the PI data. Since examinations are made weekly, the inter-observation times Δt_{ij} are all equal to 7 days, $j = 2, \ldots, n_i$ $i = 1, \ldots, 58$. For subject i,

$$O_{ij} = \begin{cases} 1 & \text{if a parasite is detected at observation } j \\ 0 & \text{if a parasite is not detected at observation } j \end{cases}$$

$$\xi_{ij} = \begin{cases} 1 & \text{if infected at observation } j \\ 0 & \text{if uninfected at observation } j \end{cases}$$

defined the observed and true states.

The parameter estimates, rates, and estimated mean times for the perfect sensitivity, perfect specificity, and imperfect sensitivity and specificity models with assumed misclassification probabilities appear in Table 3.5. Again, as the misclassification probability increases, the parameter estimates and rates become smaller and the estimated mean times increase. Comparing with the MS results, the size of the misclassification probability does not seem to have as strong an effect on the parameter estimates.

Plots of the log-likelihood function are given in Figure 3.4 for several specified values of the misclassification probabilities. These plots also show the likelihood is stretched more in the $\beta_{\eta 0} + \beta_{\rho 0}$ direction than in the $\beta_{\eta 0} - \beta_{\rho 0}$ direction as was seen in Figure 3.2 for the MS data set. For the PI data set, the log-likelihood function appears to be less flat than in the MS data set.

Table 3.5: PI estimates under assumed misclassification probabilities for the perfect sensitivity, perfect specificity, and imperfect sensitivity and specificity models. State 0 is uninfected, state 1 is infected, and EMT is the estimated mean time in days. Jackknife standard errors are given in parentheses.

Perfect Sensitivity

Quantity	Assumed Misclassification Probabilities			
	v_{01}=0.01	v_{01}=0.05	v_{01}=0.10	v_{01}=0.15
$\beta_{\rho 0}$	-3.128(0.126)	-3.382(0.160)	-3.660(0.187)	-3.901(0.207)
ρ	0.044	0.034	0.026	0.020
State 0 EMT	23	29	39	49
$\beta_{\eta 0}$	-2.994(0.142)	-3.115(0.150)	-3.301(0.150)	-3.425(0.148)
η	0.050	0.044	0.037	0.033
State 1 EMT	20	23	27	31
log-likelihood	-559.847	-555.727	-552.851	-553.043

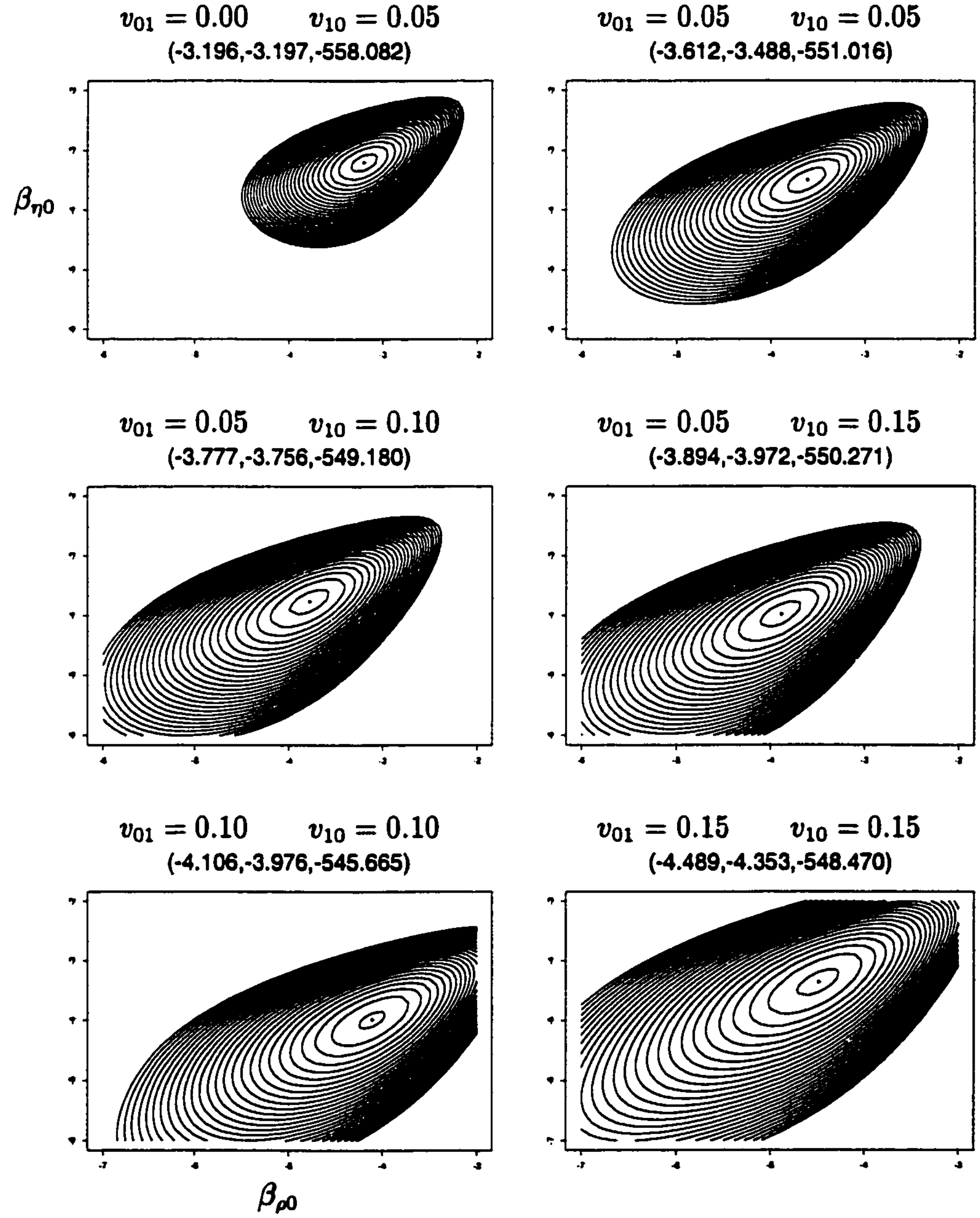
Perfect Specificity

Quantity	Assumed Misclassification Probabilities			
	v_{10}=0.01	v_{10}=0.05	v_{10}=0.10	v_{10}=0.15
$\beta_{\rho 0}$	-3.093(0.120)	-3.196(0.134)	-3.295(0.143)	-3.371(0.145)
ρ	0.045	0.041	0.037	0.034
State 0 EMT	22	24	27	29
$\beta_{\eta 0}$	-3.006(0.148)	-3.197(0.179)	-3.401(0.198)	-3.578(0.206)
η	0.049	0.041	0.033	0.028
State 1 EMT	20	24	30	36
log-likelihood	-560.281	-558.082	-557.293	-558.762

Imperfect Sensitivity and Specificity

Quantity	Assumed Misclassification Probabilities			
	v_{01}=0.01 v_{10}=0.01	v_{01}=0.01 v_{10}=0.05	v_{01}=0.05 v_{10}=0.01	v_{01}=0.05 v_{10}=0.05
$\beta_{\rho 0}$	-3.159(0.130)	-3.274(0.148)	-3.429(0.166)	-3.612(0.206)
ρ	0.042	0.038	0.032	0.027
State 0 EMT	24	26	31	37
$\beta_{\eta 0}$	-3.047(0.150)	-3.246(0.183)	-3.215(0.161)	-3.488(0.219)
η	0.048	0.039	0.040	0.031
State 1 EMT	21	26	25	33
log-likelihood	-559.074	-556.586	-554.642	-551.016

Figure 3.4: Plots of the PI log-likelihood function for specified values of the misclassification probabilities. Each contour after the first represents a drop of 2 in the log-likelihood function. In parentheses, $\hat{\beta}_{\rho 0}$ and $\hat{\beta}_{\eta 0}$ are given along with the maximum value of the log-likelihood function. ($\beta_{\rho 0} = \ln \rho$, $\beta_{\eta 0} = \ln \eta$)



$v_{01} = 0.00 \quad v_{10} = 0.05$
(-3.196,-3.197,-558.082)

$v_{01} = 0.05 \quad v_{10} = 0.05$
(-3.612,-3.488,-551.016)

$v_{01} = 0.05 \quad v_{10} = 0.10$
(-3.777,-3.756,-549.180)

$v_{01} = 0.05 \quad v_{10} = 0.15$
(-3.894,-3.972,-550.271)

$v_{01} = 0.10 \quad v_{10} = 0.10$
(-4.106,-3.976,-545.665)

$v_{01} = 0.15 \quad v_{10} = 0.15$
(-4.489,-4.353,-548.470)

3.4.2 Estimated Misclassification Probabilities

When the nonzero misclassification probability is estimated, the results for the perfect sensitivity and perfect specificity models appear in Table 3.6. The perfect specificity model was considered in Nagelkerke, Chunge and Kinoti (1990) and the estimates are similar. The jackknife standard error estimates are larger than the information based standard errors given in Nagelkerke, Chunge and Kinoti (1990). In the context of the study, a parasite could not be detected in the sample if the child was uninfected. The perfect sensitivity model is given here for illustration only. The fact that the infected estimated mean time is the same in either model is an interesting feature. Attempting to estimate both v_{01} and v_{10} gives many different estimates with the same likelihood function value. Such issues are further discussed

Table 3.6: PI estimates under perfect sensitivity and perfect specificity with constant input and blocked rates. Misclassification probabilities are estimated. State 0 is the uninfected and state 1 is infected. Jackknife standard errors are given in parentheses.

Perfect Sensitivity					
Parameter	Estimate (se)	Quantity	Estimate	Estimated Mean Time	
$\beta_{\rho 0}$	-3.766 (0.242)	ρ	0.023	State 0	43 days
$\beta_{\eta 0}$	-3.357 (0.155)	η	0.035	State 1	29 days
α_0	-1.980 (0.267)	v_{01}	0.121		
			log-likelihood:	-552.570	
Perfect Specificity					
Parameter	Estimate (se)	Quantity	Estimate	Estimated Mean Time	
$\beta_{\rho 0}$	-3.281 (0.157)	ρ	0.038	State 0	27 days
$\beta_{\eta 0}$	-3.369 (0.221)	η	0.034	State 1	29 days
α_0^*	-2.293 (0.357)	v_{10}	0.092		
			log-likelihood:	-557.261	

in Chapter 4.

Chapter 4

Parameter Identifiability and Estimability Issues

4.1 Overview

Parameter identifiability and estimability issues are investigated when the true process is modeled as a Type I counter model. Using characteristic functions, exactly two distinct sets of parameter values for the Type I counter model are shown to yield the same distribution function for the data when the same covariates and misclassification predictors are used to explain each state rate and misclassification probability. Identifiability is shown to be possible in this case if the misclassification probabilities can be restricted to be less than 0.5. Estimability of parameters is discussed when the sampling interval is inadequate or the model is misspecified. Provided that the sampling interval is appropriate and the model is correctly specified, all parameters should be estimable.

4.2 Introduction

Unknown parameters in a model are nonidentifiable if more than one set of parameter values gives the same distribution function for the observation. Conversely, unknown parameters are identifiable if each possible distribution function for the data corresponds to a unique set of parameter values. Identifiability of model parameters is an important consideration for valid estimation and inference. Inferences cannot be complete in principle if two or more explanations of the process are indistinguishable.

Estimability is a concept closely related to identifiability and usually referred to in the linear hypothesis context. For a particular realization of the data, the unknown parameters are termed nonestimable if, for the largest likelihood function values, the likelihood function is flat in some directions of the allowable parameter space. If the observed data do not provide enough information to distinguish parameter values under the model, then the likelihood function will be close to flat as the parameter values vary. Unknown parameters are said to be estimable if in the region of largest likelihood values, the likelihood function is not flat in any direction of the allowable parameter values. Identifiability can be considered as a property of the parameterization while estimability is a property of both the parameterization and a particular data set. Parameter nonestimability and nonidentifiability share the same consequences for inferences.

Problems in measurement error and hidden Markov type models include parameter nonidentifiability. In measurement error regression problems, continuous explanatory variables are measured with error and categorical explanatory variables are misclassified. The models with continuous explanatory variables generally assume that all variables are normally distributed and such an assumption leads to nonidentifiability of the regression parameters (Reiersol (1950), Bekker (1986)). When a dichotomous explanatory variable is subject to misclassification, Bekker, Van Montfort and Mooijaart (1991) show that all parameters are identifiable if a range restriction is made to the probability that the dichotomous variable is misclassified. In practice the problem of nonidentifiability is solved by incorporating supplementary data such as validation studies as well as adding constraints to the parameters (see Carroll and Stefanski (1990), for example).

In hidden Markov models, since only a function of the state in a finite state Markov chain is observed, parameter identifiability is a basic problem. Ito, Amari and Kobayashi (1992) give an algorithm for determining when two seemingly different hidden Markov processes are equivalent. In practice, states may be labeled to correspond with increasing parameter estimates in order to ensure parameter identifiability (see MacDonald and Zucchini (1997), for example). The model proposed by Nagelkerke, Chunge and Kinoti (1990) (a special case of a hidden Markov model) did not allow both types of misclassification and parameter identifiability or estimability were not specifically addressed.

Section 4.3 discusses situations which require restrictions for parameter identifiability. Even under such a restriction, the parameters may not all be estimable when the sampling interval is too long or when the underlying model is misspecified. Section 4.4 addresses these two situations as well as estimability when the sampling interval is adequate and the model is correctly specified.

4.3 Restrictions for Parameter Identifiability

Consider first the Type I counter model with both types of misclassification and exponential sojourn time distributions. When the misclassification probabilities depend on the same set, possibly empty, of misclassification predictors and the transition probabilities depend on the same set of covariates, there are exactly two sets of parameter values which give the same distribution for the observations. Section 4.3.1 gives two sets of parameter values which imply the same distribution for the observed process and Section 4.3.2 proves that these two sets of parameter values uniquely determine the distribution under the assumption that the transition probabilities lie in the interval $(0, 1)$. The features of Type I counter misclassification models which do not require such restrictions for parameter identifiability are discussed in Section 4.3.3.

4.3.1 Two Sets of Parameter Values Implying the Same Distribution

Consider the Type I model where both transition rates depend on the same set of covariates in a log-linear fashion and both misclassification probabilities have a logistic link to the same set of misclassification predictors. Recall from Section 2.6 that $\boldsymbol{\Theta} = (\boldsymbol{\alpha}_{01}, \boldsymbol{\alpha}_{10}, \boldsymbol{\beta}_{\rho}, \boldsymbol{\beta}_{\eta})$ are regression parameters for $v_{01}(\boldsymbol{C})$, $v_{10}(\boldsymbol{C})$, $\rho(\mathbf{x})$ and $\eta(\mathbf{x})$, respectively. The distribution for the observed data, and hence the likelihood function, will be the same for two sets of parameter values $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ if the two parameter vectors are identical, the trivial solution, or if

$$\boldsymbol{\Theta}_1 = (\boldsymbol{\alpha}_{01}, \boldsymbol{\alpha}_{10}, \boldsymbol{\beta}_\rho, \boldsymbol{\beta}_\eta) \qquad \text{and} \qquad \boldsymbol{\Theta}_2 = (-\boldsymbol{\alpha}_{10}, -\boldsymbol{\alpha}_{01}, \boldsymbol{\beta}_\eta, \boldsymbol{\beta}_\rho). \tag{4.1}$$

The non-trivial solution highlights the symmetry in the model. If the labels for the states are interchanged, then the transition probabilities are interchanged and the misclassification probabilities become proper classification probabilities under the new state labels.

Solution (4.1) can be seen by examining features of the relationships between the misclassification predictors and the misclassification probabilities as well as the covariate and transition probability relationships. The logistic link for the misclassification probabilities has the property that

$$v_{01}(\boldsymbol{C}_{ij}; \boldsymbol{\Theta}_1) = \frac{e^{\boldsymbol{C}_{ij}^T \boldsymbol{\alpha}_{01}}}{1 + e^{\boldsymbol{C}_{ij}^T \boldsymbol{\alpha}_{01}}} = \frac{1}{1 + e^{-\boldsymbol{C}_{ij}^T \boldsymbol{\alpha}_{01}}} = v_{11}(\boldsymbol{C}_{ij}; \boldsymbol{\Theta}_2)$$

along with $v_{00}(\boldsymbol{C}_{ij}; \boldsymbol{\Theta}_1) = v_{10}(\boldsymbol{C}_{ij}; \boldsymbol{\Theta}_2)$, where $\boldsymbol{C}_{ij}$ is the vector of misclassification predictor values for subject i collected at observation j, $i = 1, \ldots, I$, $j = 1, \ldots, n_i$. To simplify notation, the dependency on $\boldsymbol{C}_{ij}$ is dropped and relations for the misclassification probabilities under the two parameter vectors are

$$v_{01}(\boldsymbol{\Theta}_1) = 1 - v_{10}(\boldsymbol{\Theta}_2) \qquad \text{and} \qquad v_{10}(\boldsymbol{\Theta}_1) = 1 - v_{01}(\boldsymbol{\Theta}_2). \tag{4.2}$$

The steady state probability for state 1, π_1, under the two parameter vectors can be written as

$$\pi_1(\Theta_1) = \frac{e^{\mathbf{x}^T\beta_\rho}}{e^{\mathbf{x}^T\beta_\rho} + e^{\mathbf{x}^T\beta_\eta}} = 1 - \frac{e^{\mathbf{x}^T\beta_\eta}}{e^{\mathbf{x}^T\beta_\rho} + e^{\mathbf{x}^T\beta_\eta}} = 1 - \pi_1(\Theta_2). \tag{4.3}$$

Similarly, the transition probabilities under the two sets of parameter values become

$$P_{01}(t;\Theta_1) = P_{10}(t;\Theta_2) \quad \text{and} \quad P_{10}(t;\Theta_1) = P_{01}(t;\Theta_2). \tag{4.4}$$

Using (4.2) and (4.4), it is easily seen that $\Pr(O_{i1} = 1;\Theta_1) = \Pr(O_{i1} = 1;\Theta_2)$ for $i = 1,\dots,I$. Under the two sets of parameter values, the recursion relations are related as

$$\gamma_j(\Theta_1) = 1 - \gamma_j(\Theta_2). \tag{4.5}$$

This relationship can be proved by induction. First, for $n = 1$,

$$\begin{aligned}
\gamma_1(\Theta_1) &= \frac{(1-O_1)\,v_{10}(\Theta_1)\,\pi_1(\Theta_1)}{v_{00}(\Theta_1)\,\pi_0(\Theta_1) + v_{10}(\Theta_1)\,\pi_1(\Theta_1)} + \frac{O_1\,v_{11}(\Theta_1)\,\pi_1(\Theta_1)}{v_{01}(\Theta_1)\,\pi_0(\Theta_1) + v_{11}(\Theta_1)\,\pi_1(\Theta_1)} \\
&= \frac{(1-O_1)\,v_{00}(\Theta_2)\,\pi_0(\Theta_2)}{v_{10}(\Theta_2)\,\pi_1(\Theta_2) + v_{00}(\Theta_2)\,\pi_0(\Theta_2)} + \frac{O_1 v_{01}(\Theta_2)\,\pi_0(\Theta_2)}{v_{11}(\Theta_2)\,\pi_1(\Theta_2) + v_{01}(\Theta_2)\,\pi_0(\Theta_2)} \\
&= 1 - \frac{(1-O_1)\,v_{10}(\Theta_2)\pi_1(\Theta_2)}{v_{00}(\Theta_2)\,\pi_0(\Theta_2) + v_{10}(\Theta_2)\,\pi_1(\Theta_2)} - \frac{O_1\,v_{11}(\Theta_2)\,\pi_1(\Theta_2)}{v_{01}(\Theta_2)\,\pi_0(\Theta_2) + v_{11}(\Theta_2)\,\pi_1(\Theta_2)} \\
&= 1 - \gamma_1(\Theta_2).
\end{aligned}$$

Next, assume that the relationship holds for $j = 1,\dots,n-1$ observations. Define functions $\mathcal{A}_{ab}$ and $\mathcal{B}_{ab}$ such that,

$$\begin{aligned}\mathcal{A}_{ab}(\Delta t_j;\Theta_1) =&v_{ab}(\Theta_1)\big[(1-\gamma_{j-1}(\Theta_1))P_{01}(\Delta t_j;\Theta_1)+\gamma_{j-1}(\Theta_1)P_{11}(\Delta t_j;\Theta_1)\big]\\ \mathcal{B}_{ab}(\Delta t_j;\Theta_1) =&v_{ab}(\Theta_1)\big[(1-\gamma_{j-1}(\Theta_1))P_{00}(\Delta t_j;\Theta_1)+\gamma_{j-1}(\Theta_1)P_{10}(\Delta t_j;\Theta_1)\big]\end{aligned}$$

for $a,b \in \{0,1\}$. For $j = 1,\ldots,n-1$ under the two sets of parameter values, $\mathcal{A}_{11}(\Delta t_j;\Theta_1) = \mathcal{B}_{01}(\Delta t_j;\Theta_2)$ and $\mathcal{A}_{10}(\Delta t_j;\Theta_2) = \mathcal{B}_{00}(\Delta t_j;\Theta_1)$ since $\gamma_j(\Theta_1) = 1 - \gamma_j(\Theta_2)$. For $j = n$ observations,

$$\begin{aligned}\gamma_j(\Theta_1) &= \frac{(1-O_j)\,\mathcal{A}_{10}(\Delta t_j;\Theta_1)}{\mathcal{B}_{00}(\Delta t_j;\Theta_1)+\mathcal{A}_{10}(\Delta t_j;\Theta_1)} + \frac{O_j\,\mathcal{A}_{11}(\Delta t_j;\Theta_1)}{\mathcal{B}_{01}(\Delta t_j;\Theta_1)+\mathcal{A}_{11}(\Delta t_j;\Theta_1)}\\ &= \frac{(1-O_j)\,\mathcal{B}_{00}(\Delta t_j;\Theta_2)}{\mathcal{A}_{10}(\Delta t_j;\Theta_2)+\mathcal{B}_{00}(\Delta t_j;\Theta_2)} + \frac{O_j\,\mathcal{B}_{01}(\Delta t_j;\Theta_2)}{\mathcal{A}_{11}(\Delta t_j;\Theta_2)+\mathcal{B}_{01}(\Delta t_j;\Theta_2)}\\ &= 1 - \frac{(1-O_j)\,\mathcal{A}_{10}(\Delta t_j;\Theta_2)}{\mathcal{B}_{00}(\Delta t_j;\Theta_2)+\mathcal{A}_{10}(\Delta t_j;\Theta_2)} + \frac{O_j\,\mathcal{A}_{11}(\Delta t_j;\Theta_2)}{\mathcal{B}_{01}(\Delta t_j;\Theta_2)+\mathcal{A}_{11}(\Delta t_j;\Theta_2)}\\ &= 1-\gamma_j(\Theta_2)\end{aligned}$$

and the relationship holds for all observations $j = 1,\ldots,n$. In fact, this equality is obvious when the interpretation of the two sets of parameter values is considered. The model under Θ_2 is the same as the model under Θ_1 except the state labels are interchanged. Misclassification probabilities under Θ_1 become classification probabilities under Θ_2 and vice versa. For the true process, $\xi_j = 1$ under Θ_1 is the same as $\xi_j = 0$ under Θ_2. Hence, $\gamma_j(\Theta_1) = 1 - \gamma_j(\Theta_2)$ for any j.

As a consequence of (4.2)–(4.5), the conditional probabilities for the observed process are equal under the two sets of parameter values, $\Pr(O_{ij}|O_i^{(j-1)},\Theta_1) = \Pr(O_{ij}|O_i^{(j-1)},\Theta_2)$ for $i = 1,\ldots,I$ and $j = 2,\ldots,n_i$. These results show that the likelihood functions resulting from the two sets of parameter values $\boldsymbol{\Theta}_1 = (\alpha_{01}, \alpha_{10}, \boldsymbol{\beta}_\rho, \boldsymbol{\beta}_\eta)$ and $\boldsymbol{\Theta}_2 = (-\alpha_{10}, -\alpha_{01}, \boldsymbol{\beta}_\eta, \boldsymbol{\beta}_\rho)$ are equal,

$$\begin{aligned} L(\boldsymbol{\Theta}_1) &= \prod_{i=1}^{I} \Pr(O_{i1}; \boldsymbol{\Theta}_1) \prod_{j=2}^{n_i} \Pr(O_{ij} | O_i^{(j-1)}, \boldsymbol{\Theta}_1) \\ &= \prod_{i=1}^{I} \Pr(O_{i1}; \boldsymbol{\Theta}_2) \prod_{j=2}^{n_i} \Pr(O_{ij} | O_i^{(j-1)}, \boldsymbol{\Theta}_2) = L(\boldsymbol{\Theta}_2). \end{aligned}$$

In particular, if $(\widehat{\alpha}_{01}, \widehat{\alpha}_{10}, \widehat{\boldsymbol{\beta}}_\rho, \widehat{\boldsymbol{\beta}}_\eta)$ are maximum likelihood estimates for the parameters, then $(-\widehat{\alpha}_{10}, -\widehat{\alpha}_{01}, \widehat{\boldsymbol{\beta}}_\eta, \widehat{\boldsymbol{\beta}}_\rho)$ are also maximum likelihood estimates.

One of the two sets of parameter values can be eliminated if a simple restriction is added to the model. The logistic form for the misclassification probabilities guarantees that the misclassification probabilities are always between 0 and 1. If the misclassification probabilities are further restricted to the interval $[0, 0.5)$, then not both of the parameter values $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ can have misclassification probabilities within the smaller interval. If $0 \leq v_{01}(\boldsymbol{\Theta}_1) < 0.5$ and $0 \leq v_{10}(\boldsymbol{\Theta}_1) < 0.5$, then $0 \leq v_{11}(\boldsymbol{\Theta}_2) < 0.5$ and $0 \leq v_{00}(\boldsymbol{\Theta}_2) < 0.5$. Since $v_{11}(\boldsymbol{\Theta}_2) = 1 - v_{10}(\boldsymbol{\Theta}_2)$ and $v_{00}(\boldsymbol{\Theta}_2) = 1 - v_{01}(\boldsymbol{\Theta}_2)$, the ranges of $v_{11}(\boldsymbol{\Theta}_2)$ and $v_{00}(\boldsymbol{\Theta}_2)$ imply that $0.5 \leq v_{01}(\boldsymbol{\Theta}_2) \leq 1$ and $0.5 \leq v_{10}(\boldsymbol{\Theta}_2) \leq 1$, which are not admissible misclassification probabilities.

This restriction is common in error-in-variables problems where the imperfect explanatory variable is dichotomous (Bekker, Van Montfort and Mooijaart (1991)). Requiring the misclassification probabilities to lie within the restricted range is a reasonable assumption. If the true misclassification probabilities do not, then either the observed data set is not a good indicator of the true underlying process and should not be used or the labels for the states should be interchanged. In terms of the misclassification predictors and parameters, the range restriction forces $C_j^T \boldsymbol{\alpha}_{01} < 0$ and $C_j^T \boldsymbol{\alpha}_{10} < 0$, for $j = 1, \ldots, n_i$.

Table 4.1 shows the results when the constraint is applied to the PI and MS data sets. Both misclassification probabilities are found to be significantly different than zero for the PI data set. In the case of the MS data set, the estimate for v_{10} is near 0.5 and the estimated mean times are large.

Table 4.1: Estimates under restricted imperfect sensitivity and specificity model with constant input and blocked rates for the PI and MS data sets. Jackknife standard error estimates are given in parentheses.

PI Data Set					
Parameter	Estimate (se)	Quantity	Estimate	Estimated Mean Time	
$\beta_{\rho 0}$	-4.169(0.306)	ρ	0.015	State 0	65 days
$\beta_{\eta 0}$	-4.013(0.300)	η	0.018	State 1	55 days
α_0	-2.071(0.253)	v_{01}	0.112		
α_0^*	-2.202(0.327)	v_{10}	0.100		
			log-likelihood:	-545.547	

MS Data Set					
Parameter	Estimate (se)	Quantity	Estimate	Estimated Mean Time	
$\beta_{\rho 0}$	-4.489(0.852)	ρ	0.011	State 0	2394 days
$\beta_{\eta 0}$	-3.673(0.881)	η	0.025	State 1	1102 days
α_0	-2.542(0.322)	v_{01}	0.073		
α_0^*	-0.155(0.072)	v_{10}	0.461		
			log-likelihood:	-296.419	

4.3.2 Exactly Two Distinct Sets of Parameter Values Imply the Same Distribution

Section 4.3 indicated one way for two distinct sets of parameter values to give rise to the same likelihood function. This section proves that exactly those two sets of parameter values produce the same distribution for the data when the transition probabilities are not equal to 0 or 1 and there are at least three observations. Characteristic functions form the basis of the proof.

Suppose that more than one set of parameter values gives the same distribution for the data. Let two distinct sets of parameter values, $\boldsymbol{\Theta}_1 = (\boldsymbol{\alpha}_{01}, \boldsymbol{\alpha}_{10}, \boldsymbol{\beta}_\rho, \boldsymbol{\beta}_\eta)$ and $\boldsymbol{\Theta}_2 = (\tilde{\boldsymbol{\alpha}}_{01}, \tilde{\boldsymbol{\alpha}}_{10}, \tilde{\boldsymbol{\beta}}_\rho, \tilde{\boldsymbol{\beta}}_\eta)$ generate the same distribution. To avoid the degenerate case mentioned in Section 2.6, suppose $\boldsymbol{\alpha}_{01} \neq -\boldsymbol{\alpha}_{10}$ and $\tilde{\boldsymbol{\alpha}}_{01} \neq \tilde{\boldsymbol{\alpha}}_{10}$. Two sets of parameter values generate the same distribution if their characteristic functions are the same. Dropping the subscript indexing subject, the characteristic function for one subject with n observations is given by

$$\varphi_n(s_1, \ldots, s_n; \boldsymbol{\Theta}_1) = \sum_{\mathcal{K}_n} e^{\iota s_1 O_1 + \iota s_2 O_2 + \cdots + \iota s_n O_n} \Pr(O_1; \boldsymbol{\Theta}_1) \prod_{j=2}^{n} \Pr(O_j | O^{(j-1)}; \boldsymbol{\Theta}_1)$$

where $\mathcal{K}_n$ is the set of all possible binary sequences of length n and $\iota = \sqrt{-1}$. For the equality $\varphi_n(s_1, \ldots, s_n; \boldsymbol{\Theta}_1) = \varphi_n(s_1, \ldots, s_n; \boldsymbol{\Theta}_2)$ to hold,

$$\Pr(O_1 = o_1; \boldsymbol{\Theta}_1) \prod_{j=2}^{n} \Pr(O_j = o_j | O^{(j-1)} = o^{(j-1)}; \boldsymbol{\Theta}_1)$$
$$= \Pr(O_1 = o_1; \boldsymbol{\Theta}_2) \prod_{j=2}^{n} \Pr(O_j = o_j | O^{(j-1)} = o^{(j-1)}; \boldsymbol{\Theta}_2)$$

for each possible observed sequence $o^{(n)} = (o_1, \dots, o_n)$.

Consider the simple case when only the first two observations are examined and the misclassification probabilities only depend on intercept parameters, $v_{01}(\boldsymbol{C}_{ij}) = e^{\alpha_0}/(1+e^{\alpha_0})$ and $v_{10}(\boldsymbol{C}_{ij}) = e^{\alpha_0^*}/(1+e^{\alpha_0^*})$. Define terms h_{0j}, h_{1j}, $p_0(t_1)$ and $p_1(t_1)$ such that $h_{0j} = (1-O_j)\,v_{00} + O_j\,v_{01}$ and $h_{1j} = (1-O_j)\,v_{10} + O_j\,v_{11}$ for $j = 1, 2$, and $p_0(t_1) = \frac{(1-O_1)\,v_{10}\,\pi_1(\mathbf{x})}{v_{00}\,\pi_0(\mathbf{x}) + v_{10}\,\pi_1(\mathbf{x})}$ and $p_1(t_1) = \frac{O_1\,v_{11}\,\pi_1(\mathbf{x})}{v_{01}\,\pi_0(\mathbf{x}) + v_{11}\,\pi_1(\mathbf{x})}$. The characteristic function for two observations is

$$
\begin{aligned}
\varphi_2(s_1, s_2; \boldsymbol{\Theta}_1) = \sum_{\mathcal{K}_2} e^{\iota s_1 O_1 + \iota s_2 O_2} \Big(h_{01}\,\pi_0(\mathbf{x}) + h_{11}\,\pi_1(\mathbf{x})\Big) \\
\times \Bigg\{ \Big[\big(p_0(t_1) + p_1(t_1)\big)\big(1 - P_{01}(\Delta t_2) - P_{10}(\Delta t_2)\big) + P_{01}(\Delta t_2) \Big] \\
\times (2O_2 - 1)\,(1 - v_{01} - v_{10}) + h_{02} \Bigg\}
\end{aligned}
$$

where

$$
\begin{aligned}
P_{01}(t) &= \frac{\rho(\mathbf{x})}{\rho(\mathbf{x}) + \eta(\mathbf{x})} \left(1 - e^{-t(\rho(\mathbf{x}) + \eta(\mathbf{x}))}\right) \\
P_{10}(t) &= \frac{\eta(\mathbf{x})}{\rho(\mathbf{x}) + \eta(\mathbf{x})} \left(1 - e^{-t(\rho(\mathbf{x}) + \eta(\mathbf{x}))}\right)
\end{aligned}
$$

with rates $\rho(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}_\rho)$ and $\eta(\mathbf{x}) = \exp(\mathbf{x}^T \boldsymbol{\beta}_\eta)$ and steady state probability $\pi_1(\mathbf{x}) = \rho(\mathbf{x})/(\rho(\mathbf{x}) + \eta(\mathbf{x}))$. Under parameter values $\boldsymbol{\Theta}_2$, the relevant quantities are denoted by $\widetilde{\rho}(\mathbf{x})$, $\widetilde{\eta}(\mathbf{x})$, $\widetilde{v}_{01}$ and $\widetilde{v}_{10}$.

The conditions under which $\varphi_2(s_1, s_2; \boldsymbol{\Theta}_1) = \varphi_2(s_1, s_2; \boldsymbol{\Theta}_2)$ can be found by solving for $\rho(\mathbf{x})$ and $\eta(\mathbf{x})$ in terms of the unknowns $\widetilde{\rho}(\mathbf{x})$, $\widetilde{\eta}(\mathbf{x})$, v_{01}, $\widetilde{v}_{01}$, v_{10} and $\widetilde{v}_{10}$. The resulting expressions from Maple (Waterloo Maple Inc. (1996)) for $\rho(\mathbf{x})$ and

$\eta(\mathbf{x})$ are,

$$\begin{aligned}\rho(\mathbf{x}) =& \Big\{(v_{01}-\tilde{v}_{01})\,\tilde{\eta}(\mathbf{x}) + (\tilde{v}_{10}+v_{01}-1)\,\tilde{\rho}(\mathbf{x})\Big\} \\ &\times \Big\{ -\ln\tilde{\eta}(\mathbf{x}) - 2\ln(\tilde{v}_{01}+\tilde{v}_{10}-1) - \ln\tilde{\rho}(\mathbf{x}) + (t_2-t_1)\,(\tilde{\rho}(\mathbf{x})+\tilde{\eta}(\mathbf{x})) \\ &\quad + \ln\left[(v_{01}-\tilde{v}_{01})\,\tilde{\eta}(\mathbf{x}) + (\tilde{v}_{10}+v_{01}-1)\,\tilde{\rho}(\mathbf{x})\right] \\ &\quad + \ln\left[(v_{10}-\tilde{v}_{10})\,\tilde{\rho}(\mathbf{x}) + (\tilde{v}_{01}+v_{10}-1)\,\tilde{\eta}(\mathbf{x})\right]\Big\} \\ &\times \Big\{(t_2-t_1)\,(\tilde{\rho}(\mathbf{x})+\tilde{\eta}(\mathbf{x}))\,(v_{10}+v_{01}-1)\Big\}^{-1}\end{aligned} \tag{4.6}$$

and

$$\eta(\mathbf{x}) = \frac{(\tilde{v}_{01}+v_{10}-1)\,\tilde{\eta}(\mathbf{x}) + (v_{10}-\tilde{v}_{10})\,\tilde{\rho}(\mathbf{x})}{(v_{01}-\tilde{v}_{01})\,\tilde{\eta}(\mathbf{x}) + (\tilde{v}_{10}+v_{01}-1)\,\tilde{\rho}(\mathbf{x})} \times \rho(\mathbf{x}). \tag{4.7}$$

Any solutions of $\varphi_n(s_1,\dots,s_n;\Theta_1) = \varphi_n(s_1,\dots,s_n;\Theta_2)$ are also solutions of $\varphi_j(s_1,\dots,s_j;\Theta_1) = \varphi_j(s_1,\dots,s_j;\Theta_2)$ for all $j \in \{1,2,3,\dots,n-1\}$. In particular, the solutions to $\varphi_3(s_1,s_2,s_3;\Theta_1) = \varphi_3(s_1,s_2,s_3;\Theta_2)$ must also solve $\varphi_2(s_1,s_2;\Theta_1) = \varphi_2(s_1,s_2;\Theta_2)$. The solution for the 3 observation case can be found by substituting (4.6) and (4.7) into $\varphi_3(s_1,s_2,s_3;\Theta_1) = \varphi_3(s_1,s_2,s_3;\Theta_2)$ and using Maple to solve for v_{01} and v_{10} in terms of $\tilde{\rho}(\mathbf{x})$, $\tilde{\eta}(\mathbf{x})$, $\tilde{v}_{01}$ and $\tilde{v}_{10}$. The only distinct misclassification probability parameter sets satisfying $\varphi_3(s_1,s_2,s_3;\Theta_1) = \varphi_3(s_1,s_2,s_3;\Theta_2)$ are

$$v_{01} = 1-\tilde{v}_{10} \qquad \text{and} \qquad v_{10} = 1-\tilde{v}_{01}. \tag{4.8}$$

Substituting (4.8) into (4.6) and (4.7) results in $\rho(\mathbf{x}) = \tilde{\eta}(\mathbf{x})$ and $\eta(\mathbf{x}) = \tilde{\rho}(\mathbf{x})$. Therefore, the only distinct parameter sets with the same characteristic function for three observations are

$$\begin{aligned} v_{01} &= 1 - \tilde{v}_{10} & \rho(\mathbf{x}) &= \tilde{\eta}(\mathbf{x}) \\ v_{10} &= 1 - \tilde{v}_{01} & \eta(\mathbf{x}) &= \tilde{\rho}(\mathbf{x}). \end{aligned} \tag{4.9}$$

Since (4.9) must be true for a subject with any value of the covariate vector $\mathbf{x}$, the solution is equivalent to

$$\begin{aligned} \alpha_{01} &= -\tilde{\alpha}_{10} & \boldsymbol{\beta}_\rho &= \tilde{\boldsymbol{\beta}}_\eta \\ \alpha_{10} &= -\tilde{\alpha}_{01} & \boldsymbol{\beta}_\eta &= \tilde{\boldsymbol{\beta}}_\rho \end{aligned} \tag{4.10}$$

implying that $\boldsymbol{\Theta}_2 = (-\alpha_{10}, -\alpha_{01}, \boldsymbol{\beta}_\eta, \boldsymbol{\beta}_\rho)$ is the only distinct solution as given in (4.1).

Now consider the case where both misclassification probabilities are not constants and depend on misclassification predictors. For notational simplicity, assume that only one misclassification predictor is available, and has value C_j at observation time t_j, $j = 1, \ldots, n$. The characteristic function for the first observation is

$$\begin{aligned} \varphi_1(s_1; \boldsymbol{\Theta}_1) = &(1 - v_{01}(C_1))\,\pi_0(\mathbf{x}) + v_{10}(C_1)\,\pi_1(\mathbf{x}) \\ &+ e^{\iota s_1}\left[v_{01}(C_1)\,\pi_0(\mathbf{x}) + (1 - v_{10}(C_1))\,\pi_1(\mathbf{x})\right]. \end{aligned}$$

The characteristic functions under the two sets of parameter values agree for the first observation if and only if

$$(1 - v_{01}(C_1))\,\pi_0(\mathbf{x}) + v_{10}(C_1)\,\pi_1(\mathbf{x}) = (1 - \tilde{v}_{01}(C_1))\,\tilde{\pi}_0(\mathbf{x}) + \tilde{v}_{10}(C_1)\,\tilde{\pi}_1(\mathbf{x})$$

This equation can be rewritten to solve for $\pi_1(\mathbf{x})$,

$$\pi_1(\mathbf{x}) = \frac{\left(1 - v_{01}(C_1) - \tilde{v}_{10}(C_1)\right)\tilde{\pi}_1(\mathbf{x}) + \left(\tilde{v}_{01}(C_1) - v_{01}(C_1)\right)\tilde{\pi}_0(\mathbf{x})}{\left(\tilde{v}_{10}(C_1) - v_{10}(C_1)\right)\tilde{\pi}_1(\mathbf{x}) + \left(1 - v_{10}(C_1) - \tilde{v}_{01}(C_1)\right)\tilde{\pi}_0(\mathbf{x})}\,\pi_0(\mathbf{x}) \tag{4.11}$$

Since the left-hand side of (4.11) does not depend on C_1, the right-hand side must also be independent of C_1. In particular, it must be the case that

$$\frac{\left(1 - v_{01}(C_1) - \tilde{v}_{10}(C_1)\right)\tilde{\pi}_1(\mathbf{x}) + \left(\tilde{v}_{01}(C_1) - v_{01}(C_1)\right)\tilde{\pi}_0(\mathbf{x})}{\left(\tilde{v}_{10}(C_1) - v_{10}(C_1)\right)\tilde{\pi}_1(\mathbf{x}) + \left(1 - v_{10}(C_1) - \tilde{v}_{01}(C_1)\right)\tilde{\pi}_0(\mathbf{x})} = w(\mathbf{x}) \tag{4.12}$$

where $w(\mathbf{x})$ does not depend on C_1. This relationship can be further simplified by rewriting $\tilde{\pi}_0(\mathbf{x})$ in terms of $\tilde{\pi}_1(\mathbf{x})$

$$\frac{\left(1 - \tilde{v}_{01}(C_1) - \tilde{v}_{10}(C_1)\right)\tilde{\pi}_1(\mathbf{x}) + \tilde{v}_{01}(C_1) - v_{01}(C_1)}{-\left(1 - \tilde{v}_{01}(C_1) - \tilde{v}_{10}(C_1)\right)\tilde{\pi}_1(\mathbf{x}) + 1 - v_{10}(C_1) - \tilde{v}_{01}(C_1)} = w(\mathbf{x}). \tag{4.13}$$

Suppose that the misclassification predictor has at least two levels, where c and d are two of these levels. Since (4.13) is independent of the value of the misclassification predictor

$$\begin{aligned}&\frac{\left(1 - \tilde{v}_{01}(c) - \tilde{v}_{10}(c)\right)\tilde{\pi}_1(\mathbf{x}) + \tilde{v}_{01}(c) - v_{01}(c)}{-\left(1 - \tilde{v}_{01}(c) - \tilde{v}_{10}(c)\right)\tilde{\pi}_1(\mathbf{x}) + 1 - v_{10}(c) - \tilde{v}_{01}(c)} \\ &\qquad = \frac{\left(1 - \tilde{v}_{01}(d) - \tilde{v}_{10}(d)\right)\tilde{\pi}_1(\mathbf{x}) + \tilde{v}_{01}(d) - v_{01}(d)}{-\left(1 - \tilde{v}_{01}(d) - \tilde{v}_{10}(d)\right)\tilde{\pi}_1(\mathbf{x}) + 1 - v_{10}(d) - \tilde{v}_{01}(d)}.\end{aligned} \tag{4.14}$$

Provided that $\mathbf{x}$ is not a constant, the terms involving the powers of $\widetilde{\pi}_1(\mathbf{x})$ can be equated to give equalities

$$\begin{aligned}
&\left(1-\widetilde{v}_{01}(c)-\widetilde{v}_{10}(c)\right)\left(1-v_{01}(d)-v_{10}(d)\right) \\
&\qquad\qquad = \left(1-\widetilde{v}_{01}(d)-\widetilde{v}_{10}(d)\right)\left(1-v_{01}(c)-v_{10}(c)\right) \qquad (4.15) \\
&\left(1-v_{10}(d)-\widetilde{v}_{01}(d)\right)\left(\widetilde{v}_{01}(c)-v_{01}(c)\right) \\
&\qquad\qquad = \left(1-v_{10}(c)-\widetilde{v}_{01}(c)\right)\left(\widetilde{v}_{01}(d)-v_{01}(d)\right). \qquad (4.16)
\end{aligned}$$

Certainly, these equations are satisfied with the trivial solution and the distinct solution given in (4.1). Of interest are any other values of the parameters which yield equality. Solving (4.16) for $v_{01}(c)$ gives

$$v_{01}(c) = \frac{\left(1-\widetilde{v}_{01}(c)-v_{10}(c)\right)v_{01}(d)+\left(1-v_{10}(d)\right)\widetilde{v}_{01}(c)-\left(1-v_{10}(c)\right)\widetilde{v}_{01}(d)}{1-v_{10}(d)-\widetilde{v}_{01}(d)} \tag{4.17}$$

in terms of the other unknowns. Substituting (4.17) for $v_{01}(c)$ in (4.15) and solving for $v_{01}(d)$ gives

$$v_{01}(d) = 1 - v_{10}(d)$$

using Maple. This solution implies $v_{01}(C_1) + v_{10}(C_1) = 1$, which is the non-admissible degenerate case. Thus, the only distinct parameter sets are given in (4.1).

Generally speaking, if each misclassification probability depends on a larger set of misclassification predictors, the same methods can be used. The cases under consideration have either constant misclassification probabilities or the misclassification probabilities depending on the same misclassification predictors and the transition rates depending on the same covariates. These cases have shown that the characteristic functions for three or even fewer observations were identical only for the solution (4.1). Thus, the only distinct parameter set which generates the same distribution for the data is given in (4.1).

4.3.3 Situations where Parameter Restrictions are Unnecessary

The results of Section 4.3.2 apply to a specific model where the underlying true process is assumed to follow a Type I counter model, with exponential sojourn times based on the same covariates and misclassification probabilities which have a logistic link to the same misclassification predictors. These specifications resulted in parameter nonidentifiability unless parameter constraints are incorporated. However, parameters are identifiable in some special cases of the aforementioned model, and in other situations for which the relationships identified in (4.2) and (4.4) do not hold.

First consider the Type I counter model with exponential sojourn times depending on the same covariates and constant misclassification probabilities. Suppose that one of the misclassification probabilities is 0, say $v_{01} = 0$. Substituting $v_{01} = 0$ into (4.6) and (4.7), the values $\rho(\mathbf{x})|_{v_{01}=0}$ and $\eta(\mathbf{x})|_{v_{01}=0}$ are obtained. A calculation, similar to that which produced (4.9), shows that characteristic functions for 3 observations agree only when the two sets of parameter values are equal. Hence, in the model with covariates and no misclassification predictors, the model parameters are identifiable when one of the misclassification probabilities is identically zero.

Similarly, when one misclassification probability is zero and the other depends on misclassification predictors, the characteristic functions agree only when the two sets of parameter values are the same. Changing one misclassification probability to zero in (4.12) reveals that only the trivial solution generates the same distribution. All model parameters are identifiable in this case as well.

In the Type I counter model with exponential sojourn times where the same misclassification predictors do not appear in both v_{01} and v_{10} or the transition rates depend on different covariate sets or both, only the trivial solution exists. Let the transition rates depend on two different covariate sets $\mathbf{x}_\rho$ and $\mathbf{x}_\eta$, say. Equation (4.9) implies that $\mathbf{x}_\rho^T \boldsymbol{\beta}_\rho = \mathbf{x}_\eta^T \tilde{\boldsymbol{\beta}}_\eta$ and $\mathbf{x}_\eta^T \boldsymbol{\beta}_\eta = \mathbf{x}_\rho^T \tilde{\boldsymbol{\beta}}_\rho$ resulting in $\mathbf{x}_\eta^T (\tilde{\boldsymbol{\beta}}_\eta - \boldsymbol{\beta}_\eta) = \mathbf{x}_\rho^T (\boldsymbol{\beta}_\rho - \tilde{\boldsymbol{\beta}}_\rho)$. Since this equality must be true for a subject with any values of the distinct covariate vectors $\mathbf{x}_\rho$ and $\mathbf{x}_\eta$, the only solution is $\boldsymbol{\beta}_\rho = \tilde{\boldsymbol{\beta}}_\rho$ and $\boldsymbol{\beta}_\eta = \tilde{\boldsymbol{\beta}}_\eta$. As a result of this conclusion, no other set of parameter values gives the same characteristic function and the model parameters are identifiable. An analogous deduction can be made by considering the misclassification probabilities modeled on two different sets of the misclassification predictors. These statements also apply to the Type II counter model.

If one of the sojourn times for the Type I counter is not exponential, then the transition probabilities will not have the relationship given in (4.4). The relationship shown in (4.4) will not be true if the link between the misclassification probabilities and the misclassification predictors is not logistic. These two features are necessary and sufficient conditions for the parameters to be nonidentifiable. Therefore, in situations where the underlying process follows a Type I counter model with a non-exponential sojourn time, or a non-symmetric link is used for the misclassification probabilities, the parameters are identifiable and no parameter restrictions are necessary.

4.4 Data Considerations for Parameter Estimability

Providing estimates of parameters which explain phenomena is a goal of any model. Now that parameter identifiability is established, possibly under misclassification probability range restrictions, parameter estimability issues are the focus of attention. Sections 4.4.1 and 4.4.2 describe two specific situations where estimability is a problem and Section 4.4.3 discusses estimability in the absence of such situations.

4.4.1 Sampling Interval Considerations

The relationship between the true state sojourn times and the observation times of the possibly misclassified process is an important consideration in parameter estimability. If the sampling interval is too long relative to the sojourn times of the true states, even if the misclassification probabilities are constrained for parameter identifiability, not all transition rate regression parameters and misclassification predictor parameters are estimable.

If the sampling times are far apart with respect to the state transition rates, observations $2, 3, \ldots, n$ are like observations from Bernoulli trials when stationarity is assumed. Consider the case when covariate and misclassification predictor variables are not available to supplement the data. The probability of observing state 1 is the probability the true state was 1 and was correctly classified or the true state was 0 and was misclassified, $\pi_1 v_{11} + \pi_0 v_{01}$. Since the number of subjects and observations are fixed, only one degree of freedom is available. Hence, the new parameter, $\pi = \pi_1 v_{11} + \pi_0 v_{01}$ is estimable but individually π_1, v_{01} and v_{10} are not. In the perfect specificity or perfect sensitivity models, only $\pi_1 v_{11}$ or $\pi_0 v_{00}$ can be estimated, respectively.

With covariates and misclassification predictors available, the situation is similar. Only the complete expression $\pi_1(\mathbf{x}) v_{11}(\mathbf{c}) + \pi_0(\mathbf{x}) v_{01}(\mathbf{c})$ for each observed combination of $\mathbf{x}$ and $\mathbf{c}$ in the data is estimable. Individual regression parameters involved in the transition rates and misclassification probabilities are not estimable. Constraining the range of the misclassification probabilities does not permit estimation of the regression and misclassification parameters. An estimate of $\pi_1(\mathbf{x}) v_{11}(\mathbf{c}) + \pi_0(\mathbf{x}) v_{01}(\mathbf{c})$ for a particular combination of $\mathbf{x}$ and $\mathbf{c}$ can give several admissible values for $\pi_1(\mathbf{x})$, $v_{01}(\mathbf{c})$ and $v_{10}(\mathbf{c})$. For example, suppose $\pi_1(0) v_{11}(0) + \pi_0(0) v_{01}(0) = 0.4$. If $\pi_1 = 0.25$, $v_{11}(0) = 0.85$ and $v_{01}(0) = 0.25$, then these terms combine to give 0.4. The values $\pi_1 = 0.6$, $v_{11}(0) = 0.65$ and $v_{01}(0) = 0.025$ also give the same result. In both cases, the misclassification probabilities are less than 0.5.

A small simulation study was conducted to investigate the behaviour of estimates with different state sojourn times. Data sets were generated to have the same numbers of subjects and visits as the MS data set. As well, the observation times for the simulated data were taken to be the same as the MS data set. The state sojourn times were generated from exponential distributions with means of 28, 120 or 250 days and misclassification probabilities were set as either 0.02 or 0.10. These sojourn times correspond to about 0.67, 3 and 6 times the length of the scheduled 42 days between visits for the MS patients. Covariates and misclassification predictors were not incorporated in the simulations.

The data were simulated in two separate parts. First, exponential sojourn times were generated according to the settings listed above to construct subject histories and the observation times were used to convert these histories to the true states. For each subject at each observation time, a uniform random number u_{ij}^* between 0 and 1 was generated to determine if the true state was misclassified, $i = 1, \ldots, 38$, $j = 1, \ldots, 17$. An observation which was a 0 became misclassified as a 1 if $u_{ij}^* < v_{01}$ and an observation that was a 1 became misclassified as a 0 if $u_{ij}^* < v_{10}$.

Selected results for specific combinations of v_{01} and v_{10} are shown in Table 4.2. These results give the average maximum likelihood estimates for 10 simulated data sets per simulation setting combination. Note that since both misclassification probabilities are nonzero, the constraint described previously in Section 4.3.1 is used here. In parentheses, the standard errors are also given.

The results when one of the sojourn times has a mean of 28 days show how poorly the mean times and misclassification probabilities can be estimated when the sampling interval is larger than the mean of the state sojourn time. These cases show how variable the estimate can be from simulated data set to simulated

Table 4.2: Average estimated mean times (EMTs) and misclassification probability estimates for simulations of size 10 generated with specified mean times (MTs) and misclassification probabilities. Standard errors for the average are in parentheses and the most common inter-observation time is 42 days.

Settings				Estimate (se)			
MT 0	MT 1	v_{01}	v_{10}	EMT 0	EMT 1	$\hat{v}_{01}$	$\hat{v}_{10}$
28	28	0.02	0.02	65.642 (25.466)	148.923 (119.022)	0.104 (0.046)	0.071 (0.050)
28	28	0.02	0.10	1547.798(1441.212)	241.237 (140.350)	0.147 (0.061)	0.196 (0.059)
28	28	0.10	0.10	$> 10^5$ ($> 10^5$)	$> 10^5$ ($> 10^5$)	0.208 (0.071)	0.246 (0.064)
28	120	0.02	0.02	40.535 (7.617)	181.935 (41.688)	0.115 (0.051)	0.043 (0.015)
28	120	0.02	0.10	59.336 (13.486)	462.766 (317.962)	0.240 (0.073)	0.087 (0.029)
28	120	0.10	0.02	45.979 (10.195)	160.107 (53.577)	0.288 (0.066)	0.044 (0.021)
28	250	0.02	0.02	40.477 (5.066)	330.882 (148.766)	0.266 (0.066)	0.020 (0.011)
28	250	0.02	0.10	44.189 (7.413)	277.586 (77.022)	0.101 (0.056)	0.058 (0.019)
28	250	0.10	0.02	45.134 (11.949)	1238.716(1018.072)	0.213 (0.072)	0.029 (0.013)
28	250	0.10	0.10	31.208 (7.063)	123.031 (37.157)	0.245 (0.081)	0.032 (0.017)
120	28	0.10	0.10	302.529 (83.753)	70.405 (11.419)	0.121 (0.023)	0.265 (0.070)
120	120	0.02	0.02	126.032 (5.263)	121.437 (9.667)	0.013 (0.006)	0.030 (0.013)
120	120	0.02	0.10	132.340 (17.507)	119.964 (20.408)	0.033 (0.012)	0.083 (0.019)
120	120	0.10	0.10	152.797 (27.119)	138.714 (17.540)	0.091 (0.032)	0.119 (0.021)
120	250	0.02	0.02	139.050 (5.634)	212.223 (10.055)	0.011 (0.006)	0.020 (0.006)
120	250	0.02	0.10	140.504 (8.644)	184.271 (9.776)	0.037 (0.013)	0.077 (0.008)
120	250	0.10	0.02	140.862 (10.536)	212.773 (13.780)	0.070 (0.017)	0.022 (0.007)
120	250	0.10	0.10	163.757 (12.757)	262.124 (15.772)	0.091 (0.015)	0.118 (0.006)
250	250	0.02	0.02	300.517 (12.354)	207.289 (12.244)	0.020 (0.007)	0.025 (0.008)
250	250	0.02	0.10	299.852 (18.701)	225.924 (17.773)	0.026 (0.007)	0.089 (0.010)
250	250	0.10	0.10	301.553 (23.523)	210.102 (17.826)	0.092 (0.012)	0.105 (0.015)

data set. Not surprisingly, some of these simulated data sets required several maximization attempts to obtain the maximum likelihood estimates presented here. If the estimated mean times are not too far off, the estimates for the misclassification probabilities can be substantially over-estimated. For simulations when the mean times are at least 120 days in each state, both state estimated mean times and the misclassification probabilities are fairly well estimated. Since most of the inter-observation times are around 42 days, the results suggest that collecting data a minimum of three times during the occupancy of each state seems to enable estimation of the state rates as well as estimation of small misclassification probabilities for the simulation settings considered here.

The true state sojourn times are unknown. Investigators plan the sampling protocol, based on previous studies and experience, attempting to make observations sufficiently frequent relative to the perceived mean sojourn times. The data may help determine if the sampling times were not sufficiently frequent. Estimation which produces transition probabilities very close to the steady state probabilities may signal that observations were collected too far apart.

4.4.2 Estimability in the Misspecified Model

Estimating parameters can be problematic when the model for the data is incorrect. Several different parameter sets may generate the same likelihood values. Discussion focuses on the MS data set which provides an example of this type of parameter nonestimability.

Recall from Section 3.3.1, that many of the parameter estimates appeared to be unreasonable for the MS data, implying that the model was incorrect. The estimates for the model based on the covariate sex were particularly extreme and are reproduced in Table 4.3. Table 4.3 also contains estimates for the model subject to the misclassification probability constraints given in Section 4.3.

When the parameters are constrained, even more extreme estimates for the rates are seen. The likelihood for this model has fairly flat regions which result in different claimed maxima for different iterations. Recall from Section 3.3.1 that the maximization algorithm terminates when the log-likelihood function fails to increase by more than 10^{-6} on one iteration. The estimates for α_0^* and α_1^* only differ by the sign, forcing $v_{10}(1) = 0.5$. This estimate lies on the boundary of admissible values for misclassification probabilities. The constrained misclassification probability model has nonestimable parameters for this data.

In both models, the male estimates of $\rho(\mathbf{x})$ and $\eta(\mathbf{x})$ are extreme with each approximately smaller than 10^{-17}. Such extreme values are effectively zero. Many values of $\widehat{\beta}_{\rho 0}$ and $\widehat{\beta}_{\eta 0}$ will also produce rates which are effectively zero. For example, if in either model, 1 is subtracted from $\widehat{\beta}_{\rho 0}$, $\widehat{\beta}_{\rho 1}$ and 1 is added to $\widehat{\beta}_{\eta 0}$ and $\widehat{\beta}_{\eta 1}$, without changing the misclassification parameters, the same likelihood value is obtained. This feature represents an estimability problem when a misspecified model produces extreme parameter estimates. In the region of maximal likelihood function values, the likelihood function is constant in some directions of the parameters.

Such nonestimability suggests that the data set is highly unlikely under the specified model and another model needs to be specified. Model misspecification is an important concern regardless of its effect on the estimability of parameters.

Table 4.3: Imperfect sensitivity and specificity model for MS data based on the covariate sex. Estimated mean state occupancy times (EMTs) are given in days for males and females (F). In the constrained model, $v_{01} < 0.5$ and $v_{10} < 0.5$.

Unconstrained Model					
Transition			Misclassification		
Parameters		Estimates	Parameters		Estimates
$\rho(x)$	$\beta_{\rho 0}$	-39.231	$v_{01}(C_{ij})$	α_0	-2.491
	$\beta_{\rho 1}$	35.054F		α_1	-1.200
State 0 EMT		$>10^{17}$		$v_{01}(0)$	0.076
		1,825F		$v_{01}(1)$	0.024
$\eta(x)$	$\beta_{\eta 0}$	-39.040	$v_{10}(C_{ij})$	α_0^*	-0.110
	$\beta_{\eta 1}$	35.762F		α_1^*	0.228
State 1 EMT		$>10^{17}$		$v_{10}(0)$	0.473
		743F		$v_{10}(1)$	0.529
			log-likelihood:		-292.823

Constrained Model					
Transition			Misclassification		
Parameters		Estimates	Parameters		Estimates
$\rho(x)$	$\beta_{\rho 0}$	-61.055	$v_{01}(C_{ij})$	α_0	-2.483
	$\beta_{\rho 1}$	56.911F		α_1	-1.199
State 0 EMT		$>10^{27}$		$v_{01}(0)$	0.077
		1,764F		$v_{01}(1)$	0.025
$\eta(x)$	$\beta_{\eta 0}$	-60.857	$v_{10}(C_{ij})$	α_0^*	-0.131
	$\beta_{\eta 1}$	57.641F		α_1^*	0.131
State 1 EMT		$>10^{27}$		$v_{10}(0)$	0.467
		698F		$v_{10}(1)$	0.500
			log-likelihood:		-292.886

Issues in model misspecification are further discussed in Chapter 6.

4.4.3 Parameter Estimability in the Properly Sampled and Specified Model

The previous two sections considered situations where estimability is a problem and attention now shifts to situations when estimability is likely. We look for cases where the transition counts and misclassification predictor pair counts contain sufficient information to ensure parameter estimability in the two types of counter models.

When the model is not Markov, transition counts are not sufficient statistics. No data reduction to sufficient statistics seems possible in the two types of misclassification counter models. The entire history of observed process values, misclassification predictors and covariates will classify the subjects into distinct groups. Provided that these groups number at least one more than the number of parameters, parameter estimability should not be a problem. *A fortiori*, if the transition counts together with the misclassification predictor pair counts and covariates themselves provide enough information for parameter estimability, then certainly the entire data set contains enough information to estimate all parameters.

In the absence of misclassification predictors and covariates, transition pair counts under a stationary assumption are not sufficient for estimability of all parameters. In a data structure with N transition pairs, it is easily seen that the expected numbers of transition pairs 0-0, 0-1, 1-0, 1-1 must sum to N, and must be equal for 0-1 and 1-0. Thus there are 2 independent expected transition counts, and 4 parameters v_{01}, v_{10}, ρ and η, in the case where the underlying process is

Markov. Estimation of the parameters must use more than transition pair counts from the data.

However, covariates and/or misclassification predictors in the model mean that transition counts can be subdivided in informative ways. First, consider a Type I counter model with a binary misclassification predictor and for simplicity, one covariate. For each covariate value $x = x_0$, the probabilities associated with the pairwise observations under a stationarity assumption can be listed as in Table 4.4. Note that the argument is the covariate and the time interval is suppressed in the transition probabilities for this section. The relationship $\pi_0(x)\,P_{01}(x) = \pi_1(x)\,P_{10}(x)$ in the Type I counter model is responsible for the equal probabilities seen in the table. Since the probabilities of the observed process pairs are conditional on the value of the misclassification predictor pairs, the rows sum to 1.

The number of parameters which need to be estimated in the model can easily be calculated. Four misclassification parameters are involved: $v_{01}(0)$, $v_{01}(1)$, $v_{10}(0)$ and $v_{10}(1)$. If the covariate x is continuous, 4 parameters specify the intercept and slope terms in the transition rates $\rho(x)$ and $\eta(x)$. If the covariate x is discrete with l levels, the transition rates $\rho(x)$ and $\eta(x)$ involve $2 \times l$ parameters.

Consider now only the information contained in the transition counts, misclassification predictor pair counts and covariates. Let $N(x_0)$ denote the total number of observation pairs with covariate value $x = x_0$. The number of pairs that correspond to misclassification predictors 0-0, 0-1, 1-0 and 1-1 are labeled $N_{00}(x_0)$, $N_{01}(x_0)$, $N_{10}(x_0)$ and $N_{11}(x_0)$, respectively. The expected counts for each combination of misclassification predictor and observation pairs are given in Table 4.5. The number of independent expected counts in the table is 7 and the number of (functionally independent) observed counts will be at least as large if all N_{00}, N_{01}, N_{10} and N_{11} are large enough. Two such tables exist when the covariate has 2 levels, with 14 independent expected counts available to estimate 8 parameters. The parameters should be estimable.

More generally, for a covariate with l levels, there are $7 \times l$ independent expected counts to estimate the $4 + 2 \times l$ parameters. A continuous covariate x would have potentially a large number of such tables and the number of independent expected counts likely would be much larger than the number of parameters. In any case, the information contained in the transition counts with the misclassification predictor pair counts and covariate should be adequate for estimation.

The same statement should hold when more than one covariate enters the model or the misclassification predictor has more than two levels or more than one misclassification predictor is considered. For example, if the misclassification predictor has 3 levels, the number of independent expected counts for each covariate level is 15 and only two more misclassification parameters are required than in the dichotomous case.

Parameter estimability is a property of the parameterization and data collected. If nonestimability of parameters is a problem, several steps can be taken to gain some insight. Different parameterizations can be considered which may be estimable for a particular data set. In the Type I counter model with misclassification, for example, the rate parameterization may be changed to parameters representing the steady state probability for state 1 and the sum of the original rate parameters,

Table 4.4: Probabilities associated with binary misclassification predictor (MP) and observed process pairs under a stationarity assumption for a particular covariate value $x = x_0$.

	Observed Process			
MP	**0-0**	**0-1**	**1-0**	**1-1**
0-0	$\pi_{11}(x_0)$ $= v_{00}^2(0)\pi_0(x_0)P_{00}(x_0)$ $+ v_{10}(0)v_{00}(0)\pi_1(x_0)P_{10}(x_0)$ $+ v_{00}(0)v_{10}(0)\pi_0(x_0)P_{01}(x_0)$ $+ v_{10}^2(0)\pi_1(x_0)P_{11}(x_0)$	$\pi_{12}(x_0)$ $= v_{00}(0)v_{11}(0)\pi_0(x_0)P_{01}(x_0)$ $+ v_{10}(0)v_{11}(0)\pi_1(x_0)P_{11}(x_0)$ $+ v_{00}(0)v_{01}(0)\pi_0(x_0)P_{00}(x_0)$ $+ v_{10}(0)v_{01}(0)\pi_1(x_0)P_{10}(x_0)$	$\pi_{13}(x_0)$ $= \pi_{12}(x_0)$	$\pi_{14}(x_0)$ $= v_{11}^2(0)\pi_1(x_0)P_{11}(x_0)$ $+ v_{01}(0)v_{11}(0)\pi_0(x_0)P_{01}(x_0)$ $+ v_{11}(0)v_{01}(0)\pi_1(x_0)P_{10}(x_0)$ $+ v_{01}^2(0)\pi_0(x_0)P_{00}(x_0)$
0-1	$\pi_{21}(x_0)$ $= v_{00}(0)v_{00}(1)\pi_0(x_0)P_{00}(x_0)$ $+ v_{10}(0)v_{00}(1)\pi_1(x_0)P_{10}(x_0)$ $+ v_{11}(0)v_{01}(1)\pi_1(x_0)P_{10}(x_0)$ $+ v_{10}(0)v_{10}(1)\pi_1(x_0)P_{11}(x_0)$	$\pi_{22}(x_0)$ $= v_{00}(0)v_{11}(1)\pi_0(x_0)P_{01}(x_0)$ $+ v_{10}(0)v_{11}(1)\pi_1(x_0)P_{11}(x_0)$ $+ v_{00}(0)v_{01}(1)\pi_0(x_0)P_{00}(x_0)$ $+ v_{10}(0)v_{01}(1)\pi_1(x_0)P_{10}(x_0)$	$\pi_{23}(x_0)$ $= \pi_{32}(x_0)$	$\pi_{24}(x_0)$ $= v_{11}(0)v_{11}(1)\pi_1(x_0)P_{11}(x_0)$ $+ v_{01}(0)v_{11}(1)\pi_0(x_0)P_{01}(x_0)$ $+ v_{11}(0)v_{01}(1)\pi_1(x_0)P_{10}(x_0)$ $+ v_{01}(0)v_{01}(1)\pi_0(x_0)P_{00}(x_0)$
1-0	$\pi_{31}(x_0)$ $= \pi_{21}(x_0)$	$\pi_{32}(x_0)$ $= v_{11}(0)v_{00}(1)\pi_1(x_0)P_{10}(x_0)$ $+ v_{01}(0)v_{00}(1)\pi_0(x_0)P_{00}(x_0)$ $+ v_{11}(0)v_{10}(1)\pi_1(x_0)P_{11}(x_0)$ $+ v_{01}(0)v_{10}(1)\pi_0(x_0)P_{01}(x_0)$	$\pi_{33}(x_0)$ $= \pi_{22}(x_0)$	$\pi_{34}(x_0)$ $= \pi_{24}(x_0)$
1-1	$\pi_{41}(x_0)$ $= v_{00}^2(1)\pi_0(x_0)P_{00}(x_0)$ $+ v_{10}(1)v_{00}(1)\pi_1(x_0)P_{10}(x_0)$ $+ v_{00}(1)v_{10}(1)\pi_0(x_0)P_{01}(x_0)$ $+ v_{10}^2(1)\pi_1(x_0)P_{11}(x_0)$	$\pi_{42}(x_0)$ $= v_{00}(1)v_{11}(1)\pi_0(x_0)P_{01}(x_0)$ $+ v_{10}(1)v_{11}(1)\pi_1(x_0)P_{11}(x_0)$ $+ v_{00}(1)v_{01}(1)\pi_0(x_0)P_{00}(x_0)$ $+ v_{10}(1)v_{01}(1)\pi_1(x_0)P_{10}(x_0)$	$\pi_{43}(x_0)$ $= \pi_{42}(x_0)$	$\pi_{44}(x_0)$ $= v_{11}^2(1)\pi_1(x_0)P_{11}(x_0)$ $+ v_{01}(1)v_{11}(1)\pi_0(x_0)P_{01}(x_0)$ $+ v_{11}(1)v_{01}(1)\pi_1(x_0)P_{10}(x_0)$ $+ v_{01}^2(1)\pi_0(x_0)P_{00}(x_0)$

Table 4.5: Expected counts for the misclassification predictor (MP) and observed process pairs under a stationarity assumption for $x = x_0$.

	Observed Process				
MP	00	01	10	11	Total
00	$N_{00}(x_0)\,\pi_{11}(x_0)$	$N_{00}(x_0)\,\pi_{12}(x_0)$	$N_{00}(x_0)\,\pi_{13}(x_0)$	$N_{00}(x_0)\,\pi_{14}(x_0)$	$N_{00}(x_0)$
01	$N_{01}(x_0)\,\pi_{21}(x_0)$	$N_{01}(x_0)\,\pi_{22}(x_0)$	$N_{01}(x_0)\,\pi_{23}(x_0)$	$N_{01}(x_0)\,\pi_{24}(x_0)$	$N_{01}(x_0)$
10	$N_{10}(x_0)\,\pi_{31}(x_0)$	$N_{10}(x_0)\,\pi_{32}(x_0)$	$N_{10}(x_0)\,\pi_{33}(x_0)$	$N_{10}(x_0)\,\pi_{34}(x_0)$	$N_{10}(x_0)$
11	$N_{11}(x_0)\,\pi_{41}(x_0)$	$N_{11}(x_0)\,\pi_{42}(x_0)$	$N_{11}(x_0)\,\pi_{43}(x_0)$	$N_{11}(x_0)\,\pi_{44}(x_0)$	$N_{11}(x_0)$

$\pi_1(\mathbf{x})$ and $\rho(\mathbf{x})+\eta(\mathbf{x})$, respectively. Prior assumptions may be necessary to estimate all parameters.

Prior distributions may be considered for some parameters and the remaining parameters can be estimated based on these distributions. If certain values or small ranges of values are assumed for some parameters, then the remaining parameters can be estimated. As Section 4.4.2 indicated, attempting to fit an adequate model to describe the data is an important step in achieving estimability. Certainly, maximizing the likelihood function with varying starting conditions and examining likelihood function plots in the area surrounding the claimed maximum will be helpful in establishing parameter estimability for a particular data set. Further, if partial derivatives of the likelihood function are not prohibitive, a positive-definite Hessian matrix would support parameter estimability. Numerical derivatives could also be calculated to obtain the Hessian matrix.

Chapter 5

Effects of Ignoring Misclassification

5.1 Overview

The effect of misclassification on the parameter estimates in the case of constant inter-observation times is next considered. In the absence of covariates and misclassification predictors, the impact of misclassification on three estimators of the true process transition probabilities in the Type I counter is examined. Approximated estimators are constructed by expanding the maximum likelihood estimates around assumed small misclassification probabilities. Naive estimators based on the observed transition counts, which ignore misclassification, are compared with maximum likelihood estimators and approximated estimators. Via simulation studies, the approximated estimators prove to be reasonable for small values of the misclassification probabilities while the naive estimators are shown to be biased.

5.2 Introduction

Misclassification of states in transitional models has not widely been discussed in the literature (Satten and Longini (1996)). The observed response is considered to be correct in most studies, even where it may be subject to misclassification. The effect ignoring misclassification has on the model estimates when the process is subject to misclassification is of interest. More simply, if it is incorrectly assumed that the misclassification probabilities are both 0, then the resulting estimates will be different from the estimates which do not assume zero misclassification probabilities. The biases in the estimates will be related to the size of the misclassification probabilities involved. Quantifying the relationship between the estimates and misclassification probabilities, and hence establishing the magnitude of the biases that may result if misclassification is ignored, is an important aspect for any inferences that may be drawn.

Section 5.3 gives an approximation for the two Type I counter model transition probability estimates in terms of the misclassification probabilities when the misclassification probabilities are small. Simulations are carried out in Section 5.4 to illuminate the behaviour of the approximation and the biases resulting from incorrectly assuming zero misclassification probabilities.

5.3 Transition Probabilities as Functions of the Misclassification Probabilities

Consider the Type I counter model with exponential state sojourns and at least one nonzero misclassification probability: $v_{01} \neq 0$, $v_{10} \neq 0$ or both. Assume that the number of observations on each subject is the same, for notational convenience, and that these observations occur at constant intervals of length t. With the constant time between observations, the transition probabilities can be reparameterized as $a = P_{01}(t)$ and $b = P_{10}(t)$ with steady state probability for state 1, $\pi_1 = \dfrac{a}{a+b}$.

When the initial state is not used, the log-likelihood for this model is

$$\ell_P(a, b, v_{01}, v_{10}) = \sum_{i=1}^{I} \sum_{j=2}^{n} \ln \Pr(O_{ij} | O_i^{(j-1)}) = \sum_{i=1}^{I} \sum_{j=2}^{n} \ln \mathcal{P}_{ij}$$

where for $j = 2, 3, \ldots, n$ and $i = 1, \ldots, I$ the conditional probability of the current observation given the history of the observed process can be written as

$$\mathcal{P}_{ij} = (2O_{ij} - 1)(1 - v_{01} - v_{10})\left[\gamma_{i,j-1}(1 - a - b) + a\right] + (1 - O_{ij})(1 - v_{01}) + O_{ij} v_{01}.$$

The recursive formula for the conditional probability that the true process is in state 1 at time $t_{ij} = jt$ given the present and past observed process states, $\gamma_{ij} =$

$\Pr(\xi_{ij} = 1 \mid O_i^{(j)})$, has the form

$$\gamma_{ij} = \frac{-(1-O_{ij})\, v_{10} \left[\gamma_{i,j-1}\,(1-a-b)+a\right]}{\left[\gamma_{i,j-1}\,(1-a-b)+a\right](1-v_{01}-v_{10})-(1-v_{01})} + \frac{O_{ij}\,(1-v_{10}) \left[\gamma_{i,j-1}\,(1-a-b)+a\right]}{\left[\gamma_{i,j-1}\,(1-a-b)+a\right](1-v_{01}-v_{10})+v_{01}}$$

for $j = 2, 3, \ldots, n$ with

$$\gamma_{i1} = \frac{(1-O_{i1})\, v_{10}\, a}{(1-v_{01})\, b + v_{10}\, a} + \frac{O_{i1}\,(1-v_{10})\, a}{v_{01}\, b + (1-v_{10})\, a}.$$

How the maximum likelihood (ML) estimates $\widehat{a}$ and $\widehat{b}$ differ for different values of v_{01} and v_{10} is of primary interest. The estimators may be approximated for small values of v_{01} and v_{10}. Expanding the estimators around the point $(v_{01}, v_{10}) = (0,0)$ gives approximations $\widehat{a}_*(v_{01}, v_{10})$ and $\widehat{b}_*(v_{01}, v_{10})$,

$$\widehat{a}_*(v_{01}, v_{10}) = \widehat{a}_0 + v_{01} \left.\frac{\partial \widehat{a}}{\partial v_{01}}\right|_{(0,0)} + v_{10} \left.\frac{\partial \widehat{a}}{\partial v_{10}}\right|_{(0,0)} \approx \widehat{a}(v_{01}, v_{10}) \tag{5.1}$$

and

$$\widehat{b}_*(v_{01}, v_{10}) = \widehat{b}_0 + v_{01} \left.\frac{\partial \widehat{b}}{\partial v_{01}}\right|_{(0,0)} + v_{10} \left.\frac{\partial \widehat{b}}{\partial v_{10}}\right|_{(0,0)} \approx \widehat{b}(v_{01}, v_{10}) \tag{5.2}$$

where the partial derivatives are evaluated at $(v_{01}, v_{10}) = (0,0)$. Note that the terms $\widehat{a}_0$ and $\widehat{b}_0$ are the maximum likelihood estimators for a Markov chain model without misclassification. In the case of likelihood considered here, the first observation is not used and the ML estimates are simple functions of the observed transition counts,

$$\widehat{a}_0 = \frac{n_{01}}{n_{00} + n_{01}} \qquad \text{and} \qquad \widehat{b}_0 = \frac{n_{10}}{n_{10} + n_{11}}$$

where n_{kl} denotes the number of observed transitions from state k to state l, $k, l \in \{0, 1\}$. These estimates are referred to as the naive transition probability estimates.

The ML estimates for a and b can be found by setting the partial derivatives of ℓ_P involving these parameters to zero,

$$\begin{aligned} \frac{\partial}{\partial a}\ell_P(a, b, v_{01}, v_{10}) &= 0 \\ \frac{\partial}{\partial b}\ell_P(a, b, v_{01}, v_{10}) &= 0 \end{aligned} \tag{5.3}$$

which places a constraint on the relationships between $\{a, b\}$ and $\{v_{01}, v_{10}\}$. Use of the implicit function theorem (Buck (1956), for example), provided that suitable regularity conditions are satisfied, reveals these relationships. From this, the formulas for a and b are determined entirely in terms of the unknown probabilities v_{01} and v_{10}

$$\begin{aligned} a &= \widehat{a}(v_{01}, v_{10}) \\ b &= \widehat{b}(v_{01}, v_{10}). \end{aligned} \tag{5.4}$$

Use of (5.4) with (5.3) yields a system of algebraic constraints

$$\begin{aligned} \phi_1(\widehat{a}(v_{01}, v_{10}), \widehat{b}(v_{01}, v_{10}), v_{01}, v_{10}) &= 0 \\ \phi_2(\widehat{a}(v_{01}, v_{10}), \widehat{b}(v_{01}, v_{10}), v_{01}, v_{10}) &= 0. \end{aligned} \tag{5.5}$$

This system is satisfied for all admissible v_{01} and v_{10}. Taking the total derivatives of (5.5) with respect to v_{01} and v_{10} and applying the chain rule gives the four equations

$$\begin{aligned}
\frac{\partial\phi_1}{\partial v_{01}} + \frac{\partial\phi_1}{\partial a}\frac{\partial\widehat{a}}{\partial v_{01}} + \frac{\partial\phi_1}{\partial b}\frac{\partial\widehat{b}}{\partial v_{01}} &= 0\\
\frac{\partial\phi_1}{\partial v_{10}} + \frac{\partial\phi_1}{\partial a}\frac{\partial\widehat{a}}{\partial v_{10}} + \frac{\partial\phi_1}{\partial b}\frac{\partial\widehat{b}}{\partial v_{10}} &= 0\\
\frac{\partial\phi_2}{\partial v_{01}} + \frac{\partial\phi_2}{\partial a}\frac{\partial\widehat{a}}{\partial v_{01}} + \frac{\partial\phi_2}{\partial b}\frac{\partial\widehat{b}}{\partial v_{01}} &= 0\\
\frac{\partial\phi_2}{\partial v_{10}} + \frac{\partial\phi_2}{\partial a}\frac{\partial\widehat{a}}{\partial v_{10}} + \frac{\partial\phi_2}{\partial b}\frac{\partial\widehat{b}}{\partial v_{10}} &= 0.
\end{aligned}$$

These equations can be rewritten in matrix form as $\mathbf{AQ} = \mathbf{R}$ where

$$\mathbf{A} = \begin{pmatrix} \dfrac{\partial\widehat{a}}{\partial v_{01}} & \dfrac{\partial\widehat{b}}{\partial v_{01}} \\ \dfrac{\partial\widehat{a}}{\partial v_{10}} & \dfrac{\partial\widehat{b}}{\partial v_{10}} \end{pmatrix} \quad \mathbf{Q} = \begin{pmatrix} \dfrac{\partial\phi_1}{\partial a} & \dfrac{\partial\phi_2}{\partial a} \\ \dfrac{\partial\phi_1}{\partial b} & \dfrac{\partial\phi_2}{\partial b} \end{pmatrix} \quad \mathbf{R} = -\begin{pmatrix} \dfrac{\partial\phi_1}{\partial v_{01}} & \dfrac{\partial\phi_2}{\partial v_{01}} \\ \dfrac{\partial\phi_1}{\partial v_{10}} & \dfrac{\partial\phi_2}{\partial v_{10}} \end{pmatrix}.$$

The partial derivatives of the transition probability estimates with respect to the misclassification probabilities can be evaluated in terms of the other partial derivatives by solving for the matrix $\mathbf{A}$, $\mathbf{A} = \mathbf{RQ}^{-1}$. Hence,

$$A = \frac{1}{\varepsilon}\begin{pmatrix} -\dfrac{\partial\phi_1}{\partial v_{01}}\dfrac{\partial\phi_2}{\partial b} + \dfrac{\partial\phi_2}{\partial v_{01}}\dfrac{\partial\phi_1}{\partial b} & \dfrac{\partial\phi_1}{\partial v_{01}}\dfrac{\partial\phi_2}{\partial a} - \dfrac{\partial\phi_2}{\partial v_{01}}\dfrac{\partial\phi_1}{\partial a} \\ -\dfrac{\partial\phi_1}{\partial v_{10}}\dfrac{\partial\phi_2}{\partial b} + \dfrac{\partial\phi_2}{\partial v_{10}}\dfrac{\partial\phi_1}{\partial b} & \dfrac{\partial\phi_1}{\partial v_{10}}\dfrac{\partial\phi_2}{\partial a} - \dfrac{\partial\phi_2}{\partial v_{10}}\dfrac{\partial\phi_1}{\partial a} \end{pmatrix} \tag{5.6}$$

where $\varepsilon = \dfrac{\partial\phi_1}{\partial a}\dfrac{\partial\phi_2}{\partial b} - \dfrac{\partial\phi_1}{\partial b}\dfrac{\partial\phi_2}{\partial a}$.

The formulas for $\phi_1(a,b,v_{01},v_{10})$ and $\phi_2(a,b,v_{01},v_{10})$ are

$$\phi_1 = \sum_{i=1}^{I}\sum_{j=2}^{n}(2O_{ij}-1)(1-v_{01}-v_{10})\left(\frac{\partial\gamma_{i,j-1}}{\partial a}(1-a-b)-\gamma_{i,j-1}+1\right)\times\mathcal{P}_{ij}^{-1}$$
$$\phi_2 = \sum_{i=1}^{I}\sum_{j=2}^{n}(2O_{ij}-1)(1-v_{01}-v_{10})\left(\frac{\partial\gamma_{i,j-1}}{\partial b}(1-a-b)-\gamma_{i,j-1}\right)\times\mathcal{P}_{ij}^{-1}.$$

General formulas for the partial derivatives of ϕ_1, ϕ_2 and γ_{ij} as functions of any v_{01} and v_{10} appear in Appendix B. For the specific case when the partial derivatives are evaluated at $(v_{01},v_{10})=(0,0)$, the functions $W_0=(1-\widehat{a}_0-\widehat{b}_0)$ and $D_{ij:0}=-W_0\,O_{ij}-\widehat{a}_0$ are introduced. In the absence of misclassification, the recursion relation simplifies to the observed data, $\gamma_{ij}=O_{ij}$ for $j=1,\ldots,n$. The nonzero first order partial derivatives of γ_{i1} at $(v_{01},v_{10})=(0,0)$ are

$$\left.\frac{\partial\gamma_{i1}}{\partial v_{01}}\right|_{(0,0)} = -\frac{O_{i1}\,\widehat{b}_0}{\widehat{a}_0} \qquad \left.\frac{\partial\gamma_{i1}}{\partial v_{10}}\right|_{(0,0)} = \frac{(1-O_{i1})\,\widehat{a}_0}{\widehat{b}_0}$$

and the nonzero second order partial derivatives are

$$\left.\frac{\partial^2\gamma_{i1}}{\partial v_{01}\partial a}\right|_{(0,0)} = \frac{O_{i1}\,\widehat{b}_0}{\widehat{a}_0^2} \qquad \left.\frac{\partial^2\gamma_{i1}}{\partial v_{10}\partial a}\right|_{(0,0)} = \frac{1-O_{i1}}{\widehat{b}_0}$$
$$\left.\frac{\partial^2\gamma_{i1}}{\partial v_{01}\partial b}\right|_{(0,0)} = -\frac{O_{i1}}{\widehat{a}_0} \qquad \left.\frac{\partial^2\gamma_{i1}}{\partial v_{10}\partial b}\right|_{(0,0)} = -\frac{(1-O_{i1})\,\widehat{a}_0}{\widehat{b}_0^2}.$$

The partial derivatives of γ_{ij} at $(v_{01}, v_{10}) = (0,0)$ for $j = 2, \ldots, n$ are

$$\begin{aligned}
\left.\frac{\partial \gamma_{ij}}{\partial v_{01}}\right|_{(0,0)} &= \frac{O_{ij}\left(D_{i,j-1:0}+1\right)}{D_{i,j-1:0}} \\
\left.\frac{\partial \gamma_{ij}}{\partial v_{10}}\right|_{(0,0)} &= -\frac{\left(1-O_{ij}\right)D_{i,j-1:0}}{D_{i,j-1:0}+1} \\
\left.\frac{\partial^2 \gamma_{ij}}{\partial v_{01}\partial a}\right|_{(0,0)} &= \frac{O_{ij}\left(1-O_{i,j-1}\right)}{D^2_{i,j-1:0}} \\
\left.\frac{\partial^2 \gamma_{ij}}{\partial v_{10}\partial a}\right|_{(0,0)} &= \frac{\left(1-O_{ij}\right)\left(1-O_{i,j-1}\right)}{\left(D_{i,j-1:0}+1\right)^2} \\
\left.\frac{\partial^2 \gamma_{ij}}{\partial v_{01}\partial b}\right|_{(0,0)} &= -\frac{O_{ij}\,O_{i,j-1}}{D^2_{i,j-1:0}}
\end{aligned}$$

and

$$\left.\frac{\partial^2 \gamma_{ij}}{\partial v_{10}\partial b}\right|_{(0,0)} = \frac{\left(O_{ij}-1\right)O_{i,j-1}}{\left(D_{i,j-1:0}+1\right)^2}$$

and the partial derivatives $\frac{\partial \gamma_{ij}}{\partial a}$, $\frac{\partial \gamma_{ij}}{\partial b}$, $\frac{\partial^2 \gamma_{ij}}{\partial a \partial b}$, $\frac{\partial^2 \gamma_{ij}}{\partial a^2}$ and $\frac{\partial^2 \gamma_{ij}}{\partial b^2}$ are all zero at $(v_{01}, v_{10}) = (0,0)$.

The partial derivatives of ϕ_1 and ϕ_2 are fairly complicated. In the expressions which follow, $\mathcal{P}_{ij:0} = \left(2O_{ij}-1\right)\left(O_{i,j-1}\,W_0 + \widehat{a}_0\right) + 1 - O_{ij}$. The partial derivatives of ϕ_1 with respect to the misclassification probabilities, evaluated at $(v_{01}, v_{10}) = (0,0)$, are

$$\begin{aligned}
\left.\frac{\partial \phi_1}{\partial v_{01}}\right|_{(0,0)} = \sum_{i=1}^{I} & -\frac{(2O_{i2}-1)}{\mathcal{P}_{i2:0}}\left(1 - O_{i1} - \frac{O_{i1}\widehat{b}_0}{\widehat{a}_0^2}W_0 - \frac{O_{i1}\widehat{b}_0}{\widehat{a}_0}\right) \\
& - \frac{(2O_{i2}-1)^2}{\mathcal{P}_{i2:0}^2}(1-O_{i1})\left(D_{i1:0} - \frac{O_{i1}\widehat{b}_0}{\widehat{a}_0}W_0 + 1\right) \\
+ \sum_{i=1}^{I}\sum_{j=3}^{n} & -\frac{(2O_{ij}-1)}{\mathcal{P}_{ij:0}}\left[1 - O_{i,j-1}\left(\frac{(1-O_{i,j-2})}{D_{i,j-1:0}^2}W_0 - \frac{(D_{i,j-2:0}+1)}{D_{i,j-2:0}}\right)\right] \\
& -\frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2}(1-O_{i,j-1})\left(D_{i,j-1:0} + \frac{O_{i,j-1}\,(D_{i,j-2:0}+1)}{D_{i,j-2:0}}W_0 + 1\right)
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial \phi_1}{\partial v_{10}}\right|_{(0,0)} = \sum_{i=1}^{I} & -\frac{(2O_{i2}-1)}{\mathcal{P}_{i2:0}}\left(1 - O_{i1} - \frac{1-O_{i1}}{\widehat{b}_0}W_0 + \frac{(1-O_{i1})\widehat{a}_0}{\widehat{b}_0}\right) \\
& - \frac{(2O_{i2}-1)^2}{\mathcal{P}_{i2:0}^2}(1-O_{i1})\left(D_{1:0} + \frac{(1-O_{i1})\widehat{a}_0}{\widehat{b}_0}W_0\right) \\
+ \sum_{i=1}^{I}\sum_{j=3}^{n} & -\frac{(2O_{ij}-1)}{\mathcal{P}_{ij:0}}(1-O_{i,j-1})\left(1 - \frac{(1-O_{i,j-2})}{(D_{i,j-2:0}+1)^2}W_0 - \frac{D_{i,j-2:0}}{D_{i,j-2:0}+1}\right) \\
& - \frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2}(1-O_{i,j-1})\left(D_{i,j-1:0} - \frac{(1-O_{i,j-1})D_{i,j-2:0}}{D_{i,j-2:0}+1}W_0\right).
\end{aligned}$$

The partial derivatives of ϕ_1 with respect to the transition probabilities are quite simple when evaluated at $(v_{01}, v_{10}) = (0,0)$,

$$\left.\frac{\partial \phi_1}{\partial a}\right|_{(0,0)} = \sum_{i=1}^{I}\sum_{j=2}^{n} -\frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2}(1-O_{i,j-1})^2$$

and

$$\left.\frac{\partial \phi_1}{\partial b}\right|_{(0,0)} = \sum_{i=1}^{I}\sum_{j=2}^{n} \frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2}\left(1-O_{i,j-1}\right)O_{i,j-1}.$$

When evaluated at $(v_{01}, v_{10}) = (0,0)$, the partial derivatives of ϕ_2 with respect to the misclassification probabilities are

$$\begin{aligned}
\left.\frac{\partial \phi_2}{\partial v_{01}}\right|_{(0,0)} = & \sum_{i=1}^{I} \frac{(2O_{i2}-1)}{\mathcal{P}_{i2:0}}\left(O_{i1} - \frac{O_{i1}}{\widehat{a}_0}W_0 + \frac{O_{i1}\widehat{b}_0}{\widehat{a}_0}\right) \\
& + \frac{(2O_{i2}-1)^2}{\mathcal{P}_{i2:0}^2}O_{i1}\left(D_{i1:0} - \frac{O_{i1}\widehat{b}_0}{\widehat{a}_0}W_0 + 1\right) \\
& + \sum_{i=1}^{I}\sum_{j=3}^{n} \frac{(2O_{ij}-1)}{\mathcal{P}_{ij:0}}O_{i,j-1}\left(1 - \frac{O_{i,j-2}}{D_{i,j-2:0}^2}W_0 - \frac{D_{i,j-2:0}+1}{D_{i,j-2:0}}\right) \\
& + \frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2}O_{i,j-1}\left(D_{i,j-1:0} + \frac{O_{i,j-1}\left(D_{i,j-2:0}+1\right)}{D_{i,j-2:0}}W_0 + 1\right)
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial \phi_2}{\partial v_{10}}\right|_{(0,0)} = & \sum_{i=1}^{I} \frac{(2O_{i2}-1)}{\mathcal{P}_{i2:0}}\left(O_{i1} - \frac{(1-O_{i1})\widehat{a}_0}{\widehat{b}_0^2}W_0 - \frac{(1-O_{i1})\widehat{a}_0}{\widehat{b}_0}\right) \\
& + \frac{(2O_{i2}-1)^2}{\mathcal{P}_{i2:0}^2}O_{i1}\left(D_{i1:0} + \frac{(1-O_{i1})\widehat{a}_0}{\widehat{b}_0}W_0\right) \\
& + \sum_{i=1}^{I}\sum_{j=3}^{n} \frac{(2O_{ij}-1)}{\mathcal{P}_{ij:0}}\left(O_{i,j-1} - \frac{(1-O_{i,j-1})O_{i,j-2}}{\left(D_{i,j-1:0}+1\right)^2}W_0 + \frac{(1-O_{i,j-1})D_{i,j-2:0}}{D_{i,j-2:0}+1}\right) \\
& + \frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2}O_{i,j-1}\left(D_{i,j-1:0} - \frac{(1-O_{i,j-1})D_{i,j-2:0}}{D_{i,j-2:0}+1}W_0\right).
\end{aligned}$$

Finally, the partial derivatives of ϕ_2 with respect to the transition probabilities and evaluated at (v_{01}, v_{10}) are

$$\left.\frac{\partial \phi_2}{\partial a}\right|_{(0,0)} = \sum_{i=1}^{I}\sum_{j=2}^{n} \frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2} O_{i,j-1}(1-O_{i,j-1}) = \left.\frac{\partial \phi_1}{\partial b}\right|_{(0,0)}$$

and

$$\left.\frac{\partial \phi_2}{\partial b}\right|_{(0,0)} = \sum_{i=1}^{I}\sum_{j=2}^{n} - \frac{(2O_{ij}-1)^2}{\mathcal{P}_{ij:0}^2} O_{i,j-1}^2.$$

The above partial derivative expressions, when combined in (5.6), give a complicated formula for the dependency of the transition probability estimators on the misclassification probabilities. However, calculating these expressions for the values of the data are straightforward and easily obtained by a computer program. The programming was checked by comparing the program results with the results of test cases where the derivatives were known exactly.

The Multiple Sclerosis and parasitic infection data sets are used in an assessment of the approximations given by (5.1) and (5.2). Numerical results for the MS data under a constant inter-observation time assumption appear in Table 5.1 and Table 5.2 contains the numerical results for the parasitic infection data set. For each data set, various misclassification probabilities are assumed and the ML estimates for the transition probabilities are calculated. The approximated estimates for the transition probabilities are given as well as the error in the approximation. These results are further summarized graphically in Figure 5.1. The ML estimates are shown as well as the approximated estimates for different values of the misclassification probabilities.

The results for the Multiple Sclerosis data show that for small values of v_{01} and v_{10} the approximations seem to be close to the ML estimates. An exception is the approximated estimate of b when both misclassification probabilities are 0.1. In almost all cases, the approximations over-estimate the transition probabilities. For misclassification probabilities both greater than 0.2, the approximated estimates for a are less than 0. This result emphasizes the assumption that the approximation holds for small values of the misclassification probabilities.

The parasitic infection data results reiterate the suitability of the approximation for small misclassification probabilities. For the higher misclassification probabilities given in the table, the approximation does better than in the MS data set. The relative errors are generally less than 10% when the misclassification probabilities are both below 0.15. In particular, the error in the approximation for b when $(v_{01}, v_{10}) = (0.1, 0.1)$ is much smaller. The adequacy of the approximated estimates is likely related to the validity of the assumptions on the true underlying model. Hence, the model misspecification referred to in Section 4.4.2 may be responsible for poorer approximations in the case of the Multiple Sclerosis data set.

The impact of misclassification on the ML transition probability estimates is clearly seen from the tables. These estimates are attenuated by increasing misclassification probabilities. For fixed values of v_{01}, $\widehat{b}$ is more sensitive to changes in v_{10} than $\widehat{a}$. Conversely, for fixed values of v_{10}, $\widehat{a}$ is more sensitive to changes in v_{01} than $\widehat{b}$.

Table 5.1: Maximum likelihood estimates, approximated estimates and errors in the approximations for the transition probabilities in the Multiple Sclerosis data set. Inter-observation times are set to be 1.5 28-day units.

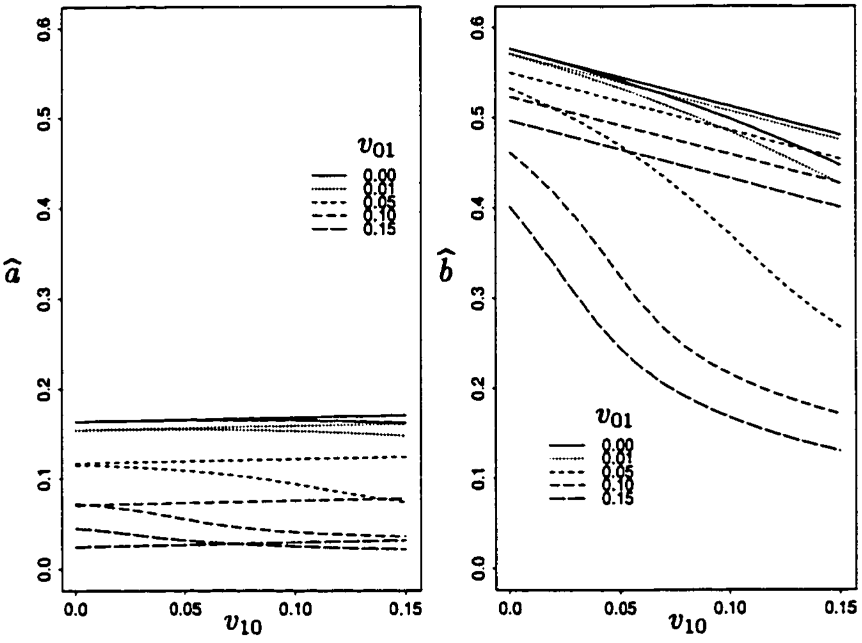
		$a = \Pr(\xi_{ij} = 1 \mid \xi_{i,j-1} = 0)$			$b = \Pr(\xi_{ij} = 0 \mid \xi_{i,j-1} = 1)$		
v_{01}	v_{10}	$\widehat{a}$	$\widehat{a}_\star$	$\widehat{a} - \widehat{a}_\star$	$\widehat{b}$	$\widehat{b}_\star$	$\widehat{b} - \widehat{b}_\star$
0.00	0.00	0.163180	0.163180	0.000000	0.576923	0.576923	0.000000
0.00	0.01	0.163650	0.163672	-0.000022	0.570382	0.570501	-0.000119
0.00	0.05	0.164998	0.165641	-0.000643	0.541606	0.544815	-0.003209
0.00	0.10	0.165005	0.168101	-0.003096	0.498731	0.512707	-0.013976
0.00	0.15	0.162354	0.170562	-0.008207	0.447351	0.480599	-0.033248
0.01	0.00	0.153823	0.153926	-0.000104	0.570985	0.571594	-0.000609
0.01	0.01	0.154110	0.154418	-0.000308	0.563912	0.565172	-0.001260
0.01	0.05	0.154568	0.156387	-0.001819	0.532470	0.539486	-0.007016
0.01	0.10	0.152938	0.158848	-0.005910	0.484659	0.507378	-0.022719
0.01	0.15	0.147869	0.161308	-0.013440	0.426521	0.475270	-0.048749
0.05	0.00	0.115029	0.116912	-0.001883	0.533062	0.550278	-0.017216
0.05	0.01	0.114285	0.117404	-0.003120	0.522131	0.543856	-0.021725
0.05	0.05	0.109004	0.119373	-0.010369	0.468858	0.518170	-0.049312
0.05	0.10	0.093787	0.121833	-0.028046	0.370822	0.486062	-0.115240
0.05	0.15	0.074041	0.124294	-0.050253	0.267697	0.453954	-0.186257
0.10	0.00	0.072138	0.070644	0.001494	0.461926	0.523633	-0.061707
0.10	0.01	0.069898	0.071136	-0.001238	0.441006	0.517211	-0.076205
0.10	0.05	0.055051	0.073105	-0.018054	0.324478	0.491525	-0.167047
0.10	0.10	0.040731	0.075565	-0.034834	0.215692	0.459417	-0.243725
0.10	0.15	0.035764	0.078026	-0.042262	0.170784	0.427309	-0.256525
0.15	0.00	0.044911	0.024376	0.020535	0.401604	0.496988	-0.095384
0.15	0.01	0.042552	0.024868	0.017683	0.371868	0.490566	-0.118698
0.15	0.05	0.031571	0.026837	0.004734	0.244385	0.464880	-0.220495
0.15	0.10	0.025095	0.029298	-0.004203	0.167454	0.432772	-0.265317
0.15	0.15	0.021778	0.031758	-0.009980	0.129860	0.400664	-0.270804
0.20	0.30	0.005158	-0.007128	-	0.024685	0.277695	-0.253010
0.30	0.20	0.000000	-0.104585	-	0.000000	0.288620	-0.288620

Table 5.2: Maximum likelihood estimates, approximated estimates and errors in the approximations for the transition probabilities in the parasitic infection data set. Inter-observation times equal 7 days.

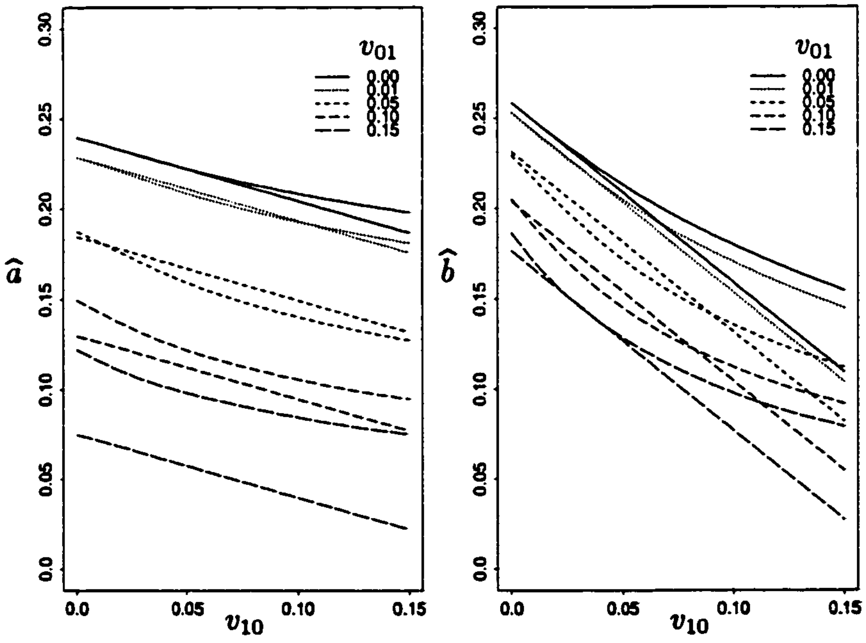
		$a = \Pr(\xi_{ij} = 1 \vert \xi_{i,j-1} = 0)$			$b = \Pr(\xi_{ij} = 0 \vert \xi_{i,j-1} = 1)$		
v_{01}	v_{10}	$\widehat{a}$	$\widehat{a}_\star$	$\widehat{a} - \widehat{a}_\star$	$\widehat{b}$	$\widehat{b}_\star$	$\widehat{b} - \widehat{b}_\star$
0.00	0.00	0.239437	0.239437	-0.000000	0.258661	0.258661	-0.000000
0.00	0.01	0.235873	0.235919	-0.000047	0.248813	0.248714	0.000099
0.00	0.05	0.221951	0.221851	0.000100	0.213245	0.208928	0.004317
0.00	0.10	0.208363	0.204265	0.004098	0.179933	0.159195	0.020738
0.00	0.15	0.198019	0.186679	0.011340	0.154744	0.109463	0.045281
0.01	0.00	0.228470	0.228466	0.000004	0.253029	0.253194	-0.000165
0.01	0.01	0.224405	0.224949	-0.000544	0.242622	0.243248	-0.000625
0.01	0.05	0.208414	0.210880	-0.002466	0.205059	0.203462	0.001597
0.01	0.10	0.192825	0.193295	-0.000469	0.170518	0.153729	0.016789
0.01	0.15	0.180999	0.175709	0.005291	0.144974	0.103996	0.040978
0.05	0.00	0.187508	0.184585	0.002922	0.229437	0.231329	-0.001892
0.05	0.01	0.181561	0.181068	0.000493	0.216573	0.221382	-0.004810
0.05	0.05	0.159202	0.167000	-0.007798	0.171808	0.181596	-0.009788
0.05	0.10	0.140230	0.149414	-0.009183	0.135748	0.131864	0.003884
0.05	0.15	0.127402	0.131828	-0.004426	0.112144	0.082131	0.030013
0.10	0.00	0.149501	0.129734	0.019766	0.205219	0.203997	0.001222
0.10	0.01	0.142818	0.126217	0.016600	0.190508	0.194050	-0.003542
0.10	0.05	0.121665	0.112148	0.009517	0.145203	0.154264	-0.009061
0.10	0.10	0.105488	0.094562	0.010926	0.112270	0.104532	0.007739
0.10	0.15	0.094642	0.076977	0.017666	0.091938	0.054799	0.037139
0.15	0.00	0.122135	0.074883	0.047252	0.186468	0.176665	0.009803
0.15	0.01	0.115600	0.071366	0.044234	0.170956	0.166719	0.004237
0.15	0.05	0.097950	0.057297	0.040653	0.128385	0.126932	0.001452
0.15	0.10	0.084416	0.039711	0.044705	0.097807	0.077200	0.020607
0.15	0.15	0.075126	0.022125	0.053001	0.079300	0.027467	0.051833
0.20	0.30	0.046449	-0.085484	-	0.041907	-0.149062	-
0.30	0.20	0.038593	-0.160014	-	0.047011	-0.104261	-

Figure 5.1: Maximum likelihood transition probability estimates as a function of the misclassification probabilities for the two data sets. The approximated estimates are the straight lines.

MS Data Set



PI Data Set



5.4 Simulation Calculations of Bias

Three biases are of interest: the bias that may result from the ignoring misclassification, the bias that may result from using the approximated estimates and the bias in the ML estimates. The first bias is particularly interesting since it quantifies the effect on estimates that may result if misclassification exists in the data but the misclassification is ignored in the modeling. The bias resulting from use of the approximated estimates is also of interest. The approximated estimates permit easier calculation than the ML estimates and readily allow the impact of differing degrees of misclassification to be seen.

The biases $\mathrm{E}(\widehat{a}_0) - a$, $\mathrm{E}(\widehat{a}_*(v_{01}, v_{10})) - a$ and $\mathrm{E}(\widehat{a}(v_{01}, v_{10})) - a$ for the 0 to 1 transition probability, as well as the biases $\mathrm{E}(\widehat{b}_0) - b$, $\mathrm{E}(\widehat{b}_*(v_{01}, v_{10})) - b$ and $\mathrm{E}(\widehat{b}(v_{01}, v_{10})) - b$ for the 1 to 0 transition probability have complicated dependencies on the misclassification probabilities. Such complicated dependencies make finding analytical expressions impossible. Simulation methods are implemented as a way to assess the magnitude of the expected bias of each estimator.

Data sets are simulated with the same number of subjects (38) and observations per subject (17) as the complete MS data set. Each misclassification probability is considered to be either 0, 0.01, 0.05, 0.1 or 0.15. The true state probability transition a was specified as 0.2, 0.4, 0.6 or 0.8. The value of b was set as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 or 0.7 subject to the restriction that $a + b < 1$. Recall from Section 2.5.1 that $a + b$ must be less than one to be consistent with a Type I model. These transition probability simulation settings were combined with all possible combinations of the misclassification probabilities for the results that follow.

The data sets were simulated in two steps. The first step simulated the true data for each subject at each observation time. The initial state of the true data was generated based on the steady state probability settings for each state. For subject i, a uniform random number on (0,1), u_{i1}, was generated, $i = 1, \ldots, 38$. If $0 < u_{i1} < \pi_1 = \frac{a}{a+b}$, then the first observation was set to 1. Conversely, if $\pi_1 \leq u_{i1} < 1$, then the first observation was set to 0. The remaining observations were generated based on the transition probability settings. For subject i at observation j ($j = 2, 3, \ldots, 17$), a uniform random number u_{ij} was generated between 0 and 1. If the previous observation was 0, then observation j was set to 1 if $0 < u_{ij} < a$ and 0 otherwise. Similarly, if the previous observation was 1, then observation j was set to 0 if $0 < u_{ij} < b$ and 1 otherwise. This data set is termed the true data set.

The second step of the data simulation involved misclassifying the true data according to the specified misclassification probabilities. A uniform random number on (0,1), u^*_{ij}, was generated for subject i and observation j, $i = 1, \ldots, 38$, $j = 1, \ldots, 17$. An observation which was 0 became misclassified as a 1 if $u^*_{ij} < v_{01}$. Similarly, an observation that was 1 became misclassified as a 0 if $u^*_{ij} < v_{10}$. The data constructed in this second step of the simulation process by misclassifying the true data set is referred to as the observed data set.

Each data set was then used to calculate the values of the partial derivatives in the matrix **A**. These values, along with the specified values for the misclassification probabilities and the naive transition probability estimates, produced approximated transition probability estimates. For each combination of simulation settings, the results from the simulated data sets were averaged to give estimates for the expected partial derivatives and the biases.

Simulations of size 50 were run with the 16 combinations of the transition probability settings. Results from four settings are shown in Tables 5.3-5.6 which start on page 105. The absolute relative biases of these results are displayed graphically in Figures 5.2-5.5. In each situation, the time between observations is set to be 1. The tables give the naive estimates, ML estimates and approximated estimates for the transition probabilities centered around the transition probability simulation settings. These values give estimates of the biases in the naive, ML, and approximated transition probability estimates. Note that these estimated biases are uncertain and have standard errors equal to the standard deviations of the 50 simulation estimates divided by $\sqrt{50}$.

Table 5.3 gives the averages and standard deviations of the simulations when both transition probabilities are set to 0.2. The average biases calculated for the ML estimates are all less than their associated standard deviations. In general, the partial likelihood estimates have relative biases less than 3%, and the biases do not show any patterns with respect to the misclassification probabilities. Their errors are mainly the result of the variability associated with the finiteness of the data sets generated from the simulation settings. The errors given for the case where both misclassification probabilities are zero illustrate this point. The ML estimates are approximately unbiased when the values of v_{01} and v_{10} are correctly specified.

The naive estimates almost always have a positive bias which increases as the misclassification probabilities increase. Except for the case without misclassification, these values are always larger than the ML estimates. The standard deviations are generally the same for the naive and ML estimates but are larger in the approximated estimates. The approximated estimates have variability associated with both the naive estimates as well as the derivative calculations which depend on the pattern of the observations. The naive estimate of a appears to be reasonable for values of the misclassification probabilities both less than 0.05. For higher misclassification probabilities the relative bias is in the 18% to 39% range when $v_{01} = 0.05$, in the 40% to 58% range when $v_{01} = 0.10$ and in the 59% to 80% range when $v_{01} = 0.15$. For these same values of v_{01} the relative bias ranges for $\widehat{b}_0$ are 8% to 67%, 15% to 76% and 20% to 80%, respectively. Clearly, $\widehat{a}_0$ is more sensitive to the value of v_{01} than $\widehat{b}_0$ and $\widehat{b}_0$ is more sensitive to the value of v_{10} than $\widehat{a}_0$. Again, the naive estimate of b appears to be reasonable for $v_{01} < 0.05$ and $v_{10} < 0.05$.

The approximated estimates perform well. Most of the estimated biases are less than their associated standard deviations. The absolute relative biases in the approximation are generally less than 10%. This figure is doubled or tripled when $v_{01} = 0.10$ and $v_{10} = 0.15$ or when $v_{01} = 0.15$ and $v_{10} \geq 0.10$. Sometimes, the approximated estimates appear to have less bias than the ML estimates. This appearance generally indicates that the average transition probabilities generating the true data sets were not exactly the same as the transition probability simulation settings. For example, when $(v_{01}, v_{10}) = (0.01, 0.05)$ the averages of the 50 simulations yielded 0.2011 for the 0-1 transition probability of the true data sets generated. The ML estimate of a is closer to 0.2011 than the approximated estimate $\widehat{a}_*$. Such situations demonstrate the variability associated with the generation of data.

Many of the same comments discussed above apply to Tables 5.4-5.6. Generally, the absolute relative bias of the ML estimates is less than 3%. The ML estimates are approximately unbiased for correct specification of the misclassification probabilities. The naive estimates are more biased and the approximated estimates do not perform as well for larger values of the misclassification probabilities. The approximated estimates are generally much closer to the simulation settings than the naive estimates. The standard deviations of the simulations, not presented here, are similar to what is shown in Table 5.3.

Within the tables for $(a, b) = (0.2, 0.4)$ and $(a, b) = (0.4, 0.2)$, the naive estimates of a are about the same in the last three blocks of settings in Table 5.4 and the last block of Table 5.5. The estimated bias is about the same for $\widehat{a}_0$ when $v_{01} = 0.15$ and v_{10} ranges from 0 to 0.15 in each of the respective tables. The absolute relative biases when $v_{01} = 0.15$ are around 60% for $(a, b) = (0.2, 0.4)$ and around 25% for $(a, b) = (0.4, 0.2)$. The absolute relative biases of $\widehat{a}_0$ when $(a, b) = (0.2, 0.4)$ for $v_{01} = 0.05$, 0.10 or 0.15 are about twice the magnitude of the corresponding absolute relative biases when $(a, b) = (0.4, 0.2)$. Hence, the naive estimate $\widehat{a}_0$ performs better when $a = 0.4$ than when $a = 0.2$. A similar statement is also true for the performance of $\widehat{b}_0$. For misclassification probabilities both greater than 0.05, the absolute relative bias ranges from 8% to 29% when $(a, b) = (0.2, 0.4)$ and from 17% to 62% when $(a, b) = (0.4, 0.2)$. For both of the transition probability settings, the approximated estimates perform about the same.

Table 5.6 displays the results when $(a, b) = (0.6, 0.3)$. In this particular table, $\widehat{a}_0$ has relative biases less than 12% which is smaller than the other tables discussed. The approximated estimates $\widehat{a}_\star$ and $\widehat{b}_\star$ generally perform better here too. When

$v_{10} = 0.05$, 0.10 or 0.15 the absolute relative biases of $\widehat{b}_0$ range from 5% to 34%. Settings such as this generated data sets which had distributions close to the stationary distribution. Convergence was a problem for some of the data sets and the details are further discussed in the next section.

A few general statements can be made as a summary of these results. The ML estimates are approximately unbiased when the misclassification probabilities are correctly specified. The naive estimates over-estimate the transition probabilities and this positive bias increases as the misclassification probabilities increase. The naive estimate for a is more sensitive to the value of v_{01} than $\widehat{b}_0$ and $\widehat{b}_0$ is more sensitive to the value of v_{10} than $\widehat{a}_0$. When both of the misclassification probabilities are less than 0.05, the naive estimates will have small bias regardless of the transition probabilities of the true process. The conclusions drawn from such estimates will likely be adequate. If one of the transition probabilities is quite large, the naive estimate of that transition probability appears to do better for slightly larger misclassification probabilities.

The approximated estimates do well and are much easier to calculate than the ML estimates. While the approximation does get worse for higher values of the misclassification probabilities, the approximated estimates do well when the misclassification probabilities are both less than 0.10 and generally when $v_{01} = 0.15$ and $v_{10} < 0.05$ or $v_{01} < 0.05$ and $v_{10} = 0.15$. The approximated estimates appear to be a reasonable choice for these values of the misclassification probabilities. For misclassification probabilities which are not too large, the approximated estimates can be calculated for a range of values of v_{01} and v_{10}. For these values, the approximated estimates are approximately unbiased for the correct misclassification probability values. If the investigator believes that the misclassification probabilities are not small, a couple of possibilities can be considered. A further approximation could be made which includes higher order partial derivatives. This approximation would then be reasonable for larger misclassification probability values. Alternatively, the ML transition probability estimates could be calculated for specified larger values of the misclassification probabilities.

Table 5.3: Averages for 50 simulations appear with standard deviations in parentheses. Simulation settings were $a = 0.2$, $b = 0.2$ and $\Delta t_{ij} = 1$ for specified misclassification probabilities.

v_{01}	v_{10}	$\widehat{a}_0 - a$	$\widehat{a} - a$	$\widehat{a}_\star - a$	$\widehat{b}_0 - b$	$\widehat{b} - b$	$\widehat{b}_\star - b$
0.00	0.00	-0.005 (0.020)	-0.005 (0.020)	-0.005 (0.020)	0.002 (0.024)	0.002 (0.024)	0.002 (0.024)
0.00	0.01	0.009 (0.029)	0.006 (0.028)	0.005 (0.029)	0.013 (0.030)	0.005 (0.030)	0.005 (0.030)
0.00	0.05	0.016 (0.026)	-0.001 (0.026)	-0.003 (0.027)	0.047 (0.023)	0.006 (0.024)	-0.002 (0.026)
0.00	0.10	0.029 (0.026)	-0.001 (0.024)	-0.002 (0.027)	0.087 (0.022)	0.005 (0.025)	-0.012 (0.029)
0.00	0.15	0.036 (0.025)	0.001 (0.027)	0.005 (0.031)	0.133 (0.027)	0.011 (0.027)	-0.009 (0.038)
0.01	0.00	0.015 (0.025)	0.007 (0.025)	0.006 (0.025)	0.007 (0.025)	0.003 (0.025)	0.003 (0.025)
0.01	0.01	0.011 (0.025)	-0.001 (0.025)	-0.002 (0.025)	0.007 (0.022)	-0.005 (0.021)	-0.006 (0.021)
0.01	0.05	0.027 (0.028)	0.002 (0.028)	0.000 (0.029)	0.046 (0.026)	0.002 (0.027)	-0.004 (0.030)
0.01	0.10	0.036 (0.024)	-0.001 (0.023)	-0.001 (0.026)	0.089 (0.030)	0.005 (0.032)	-0.010 (0.038)
0.01	0.15	0.050 (0.021)	0.005 (0.021)	0.011 (0.025)	0.123 (0.023)	-0.000 (0.024)	-0.018 (0.031)
0.05	0.00	0.038 (0.031)	-0.001 (0.031)	-0.010 (0.034)	0.017 (0.026)	0.000 (0.026)	-0.003 (0.027)
0.05	0.01	0.042 (0.031)	-0.002 (0.031)	-0.009 (0.035)	0.029 (0.025)	0.004 (0.026)	0.002 (0.028)
0.05	0.05	0.061 (0.023)	0.005 (0.022)	0.001 (0.025)	0.054 (0.031)	0.001 (0.030)	-0.003 (0.033)
0.05	0.10	0.069 (0.026)	0.002 (0.027)	0.007 (0.031)	0.101 (0.024)	0.006 (0.027)	0.004 (0.032)
0.05	0.15	0.078 (0.032)	0.002 (0.033)	0.017 (0.040)	0.135 (0.031)	-0.002 (0.036)	-0.001 (0.047)
0.10	0.00	0.081 (0.025)	0.004 (0.025)	-0.015 (0.032)	0.030 (0.025)	0.003 (0.024)	-0.001 (0.027)
0.10	0.01	0.088 (0.031)	0.004 (0.030)	-0.011 (0.038)	0.037 (0.027)	0.000 (0.027)	-0.001 (0.031)
0.10	0.05	0.098 (0.034)	0.001 (0.037)	-0.002 (0.043)	0.070 (0.029)	0.001 (0.031)	0.007 (0.035)
0.10	0.10	0.111 (0.033)	0.006 (0.035)	0.017 (0.044)	0.106 (0.027)	0.001 (0.029)	0.011 (0.038)
0.10	0.15	0.117 (0.029)	0.004 (0.037)	0.034 (0.042)	0.152 (0.028)	0.007 (0.033)	0.033 (0.040)
0.15	0.00	0.117 (0.025)	-0.006 (0.028)	-0.030 (0.035)	0.041 (0.029)	0.001 (0.027)	0.005 (0.032)
0.15	0.01	0.122 (0.029)	0.001 (0.032)	-0.018 (0.045)	0.049 (0.026)	0.005 (0.028)	0.010 (0.035)
0.15	0.05	0.142 (0.033)	0.007 (0.033)	0.008 (0.044)	0.073 (0.026)	-0.001 (0.028)	0.014 (0.034)
0.15	0.10	0.140 (0.040)	-0.001 (0.043)	0.021 (0.054)	0.126 (0.024)	0.012 (0.030)	0.042 (0.033)
0.15	0.15	0.159 (0.028)	0.008 (0.036)	0.055 (0.046)	0.160 (0.034)	0.010 (0.043)	0.058 (0.053)

Figure 5.2: Percent absolute relative bias of a and b naive estimates (NEs), approximated estimates (AEs) and ML estimates (MLEs) for simulations with $a = 0.20$ and $b = 0.20$ and specified misclassification probabilities.

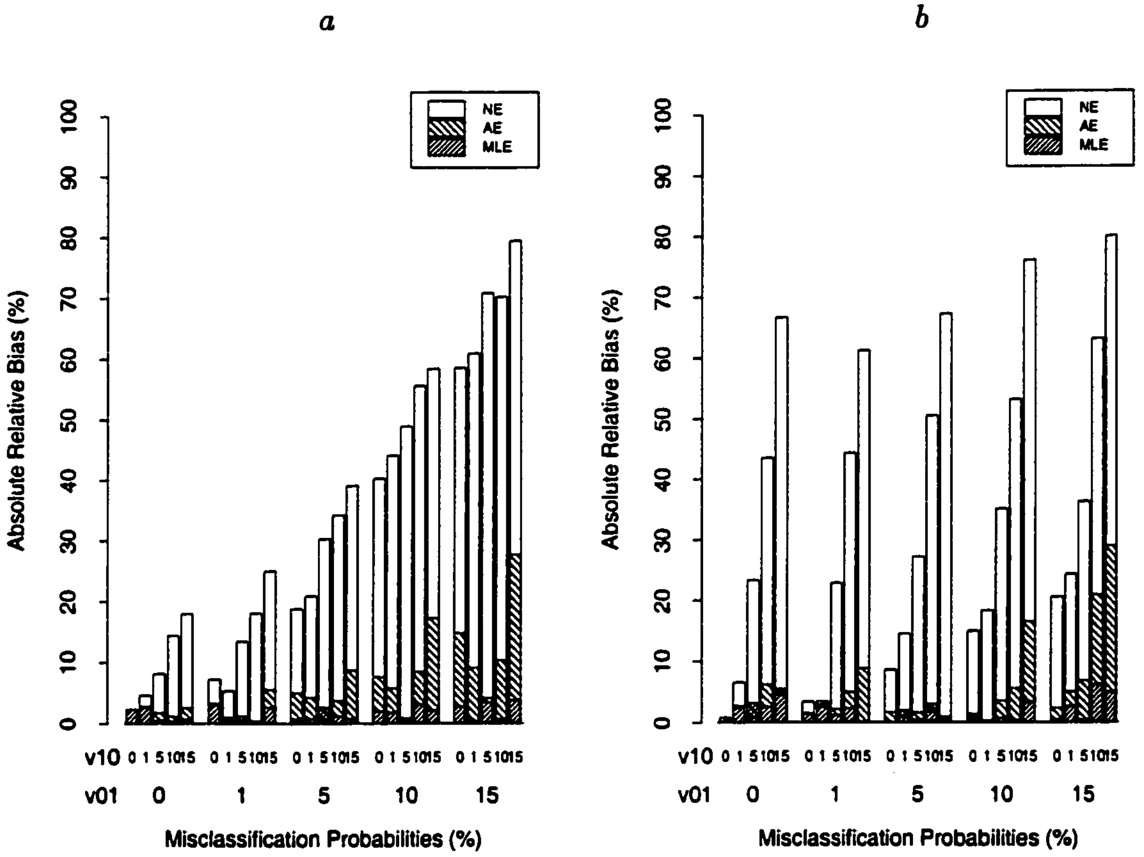


Table 5.4: Averages of 50 simulations for settings $a = 0.2$, $b = 0.4$, $\Delta t_{ij} = 1$ and specified misclassification probabilities.

v_{01}	v_{10}	$\hat{a}_0 - a$	$\hat{a} - a$	$\hat{a}_\star - a$	$\hat{b}_0 - b$	$\hat{b} - b$	$\hat{b}_\star - b$
0.00	0.00	-0.0029	-0.0029	-0.0029	0.0038	0.0038	0.0038
0.00	0.01	0.0004	0.0004	0.0004	-0.0021	-0.0082	-0.0082
0.00	0.05	0.0051	0.0060	0.0061	0.0286	-0.0009	-0.0011
0.00	0.10	-0.0014	0.0013	0.0019	0.0587	-0.0012	-0.0007
0.00	0.15	-0.0075	-0.0015	0.0000	0.0858	-0.0034	-0.0000
0.01	0.00	0.0066	-0.0014	-0.0014	0.0037	-0.0001	0.0000
0.01	0.01	0.0040	-0.0039	-0.0039	0.0154	0.0056	0.0058
0.01	0.05	0.0079	0.0011	0.0016	0.0386	0.0053	0.0060
0.01	0.10	0.0038	-0.0003	0.0010	0.0656	0.0043	0.0064
0.01	0.15	-0.0029	-0.0035	-0.0012	0.0992	0.0087	0.0147
0.05	0.00	0.0406	0.0020	0.0017	0.0158	0.0030	0.0056
0.05	0.01	0.0450	0.0057	0.0063	0.0259	0.0074	0.0110
0.05	0.05	0.0394	-0.0005	0.0021	0.0338	-0.0122	-0.0057
0.05	0.10	0.0395	0.0037	0.0090	0.0728	0.0007	0.0114
0.05	0.15	0.0338	0.0006	0.0098	0.1044	0.0015	0.0210
0.10	0.00	0.0787	0.0003	0.0024	0.0193	0.0003	0.0115
0.10	0.01	0.0822	0.0041	0.0067	0.0192	-0.0049	0.0063
0.10	0.05	0.0775	-0.0010	0.0063	0.0483	-0.0000	0.0168
0.10	0.10	0.0756	0.0019	0.0144	0.0826	0.0102	0.0333
0.10	0.15	0.0719	0.0024	0.0199	0.1148	0.0162	0.0487
0.15	0.00	0.1217	0.0024	0.0116	0.0141	-0.0039	0.0179
0.15	0.01	0.1208	0.0016	0.0130	0.0243	0.0009	0.0248
0.15	0.05	0.1205	0.0010	0.0198	0.0463	-0.0004	0.0305
0.15	0.10	0.1146	0.0002	0.0250	0.0796	0.0093	0.0471
0.15	0.15	0.1099	-0.0024	0.0315	0.1039	-0.0004	0.0533

Figure 5.3: Percent absolute relative bias of a and b NEs, AEs and MLEs for simulations with $a = 0.20$ and $b = 0.40$ and specified misclassification probabilities.

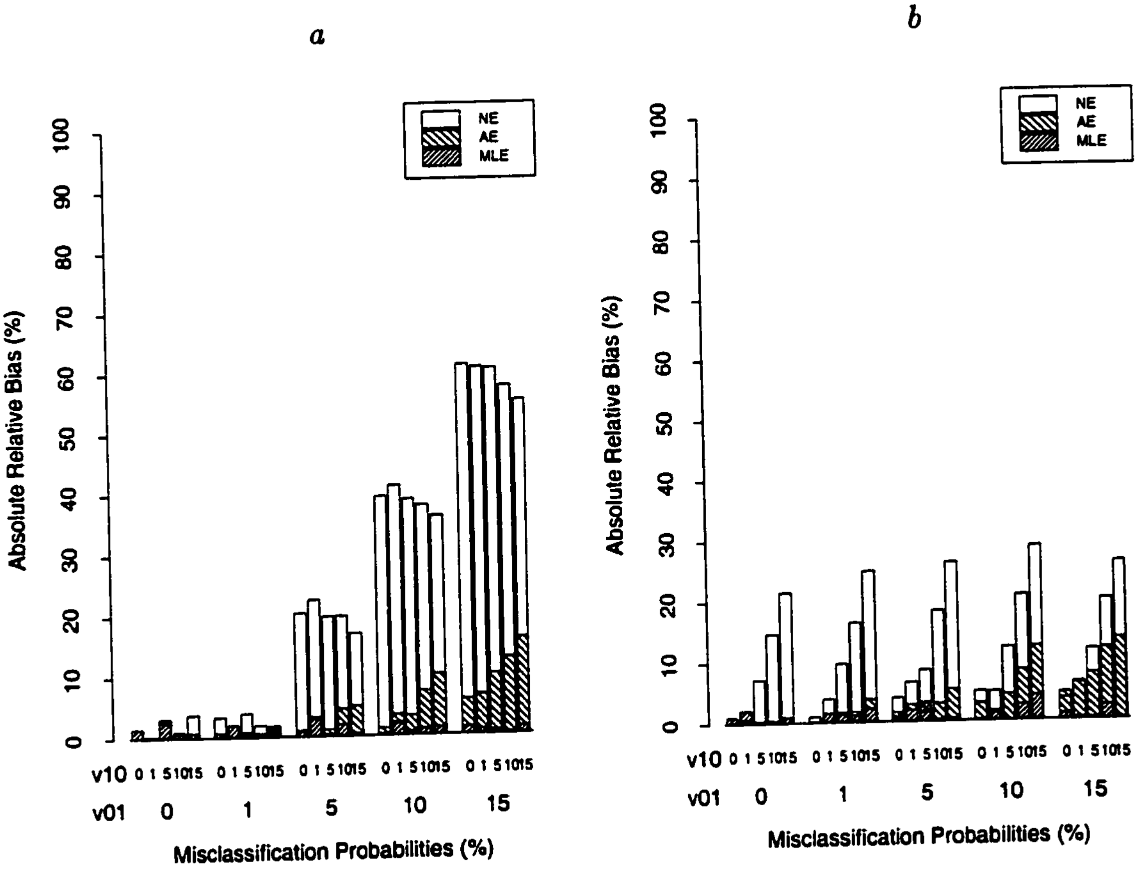
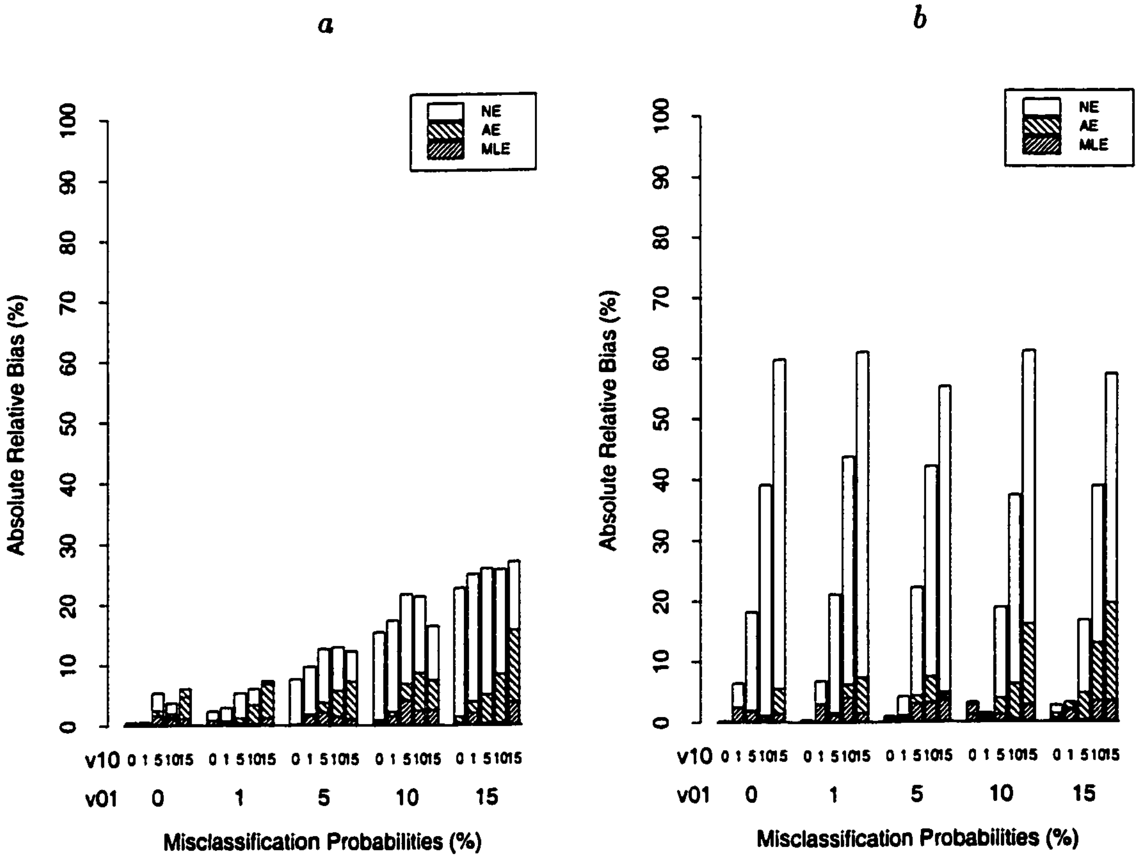


Table 5.5: Averages of 50 simulations for settings $a = 0.4$, $b = 0.2$, $\Delta t_{ij} = 1$ and specified misclassification probabilities.

v_{01}	v_{10}	$\hat{a}_0 - a$	$\hat{a} - a$	$\hat{a}_* - a$	$\hat{b}_0 - b$	$\hat{b} - b$	$\hat{b}_* - b$
0.00	0.00	0.0019	0.0019	0.0019	-0.0003	-0.0003	-0.0003
0.00	0.01	0.0011	-0.0023	-0.0022	0.0128	0.0049	0.0049
0.00	0.05	0.0215	0.0064	0.0097	0.0364	-0.0036	-0.0037
0.00	0.10	0.0146	-0.0076	0.0044	0.0782	-0.0021	0.0002
0.00	0.15	0.0186	0.0042	0.0243	0.1195	0.0026	0.0110
0.01	0.00	0.0092	0.0033	0.0033	-0.0006	-0.0005	-0.0005
0.01	0.01	0.0114	0.0022	0.0024	0.0136	0.0059	0.0059
0.01	0.05	0.0211	0.0012	0.0048	0.0422	0.0025	0.0030
0.01	0.10	0.0243	0.0008	0.0133	0.0875	0.0081	0.0124
0.01	0.15	0.0274	0.0048	0.0292	0.1220	0.0028	0.0147
0.05	0.00	0.0305	0.0003	0.0000	0.0011	0.0018	0.0019
0.05	0.01	0.0387	0.0061	0.0067	0.0083	0.0015	0.0020
0.05	0.05	0.0504	0.0085	0.0146	0.0444	0.0060	0.0086
0.05	0.10	0.0515	0.0054	0.0226	0.0842	0.0064	0.0148
0.05	0.15	0.0488	-0.0039	0.0289	0.1103	-0.0097	0.0072
0.10	0.00	0.0612	0.0023	0.0033	0.0023	0.0057	0.0064
0.10	0.01	0.0692	0.0059	0.0087	0.0020	-0.0028	-0.0012
0.10	0.05	0.0868	0.0165	0.0274	0.0378	0.0024	0.0077
0.10	0.10	0.0850	0.0094	0.0343	0.0747	-0.0009	0.0124
0.10	0.15	0.0657	-0.0099	0.0296	0.1221	0.0055	0.0321
0.15	0.00	0.0906	0.0013	0.0051	-0.0054	0.0012	0.0026
0.15	0.01	0.0998	0.0080	0.0152	0.0041	0.0037	0.0064
0.15	0.05	0.1035	0.0007	0.0199	0.0335	0.0004	0.0094
0.15	0.10	0.1027	-0.0011	0.0333	0.0777	0.0068	0.0259
0.15	0.15	0.1081	0.0149	0.0628	0.1145	0.0069	0.0391

Figure 5.4: Percent absolute relative bias of a and b NEs, AEs and MLEs for simulations with $a = 0.40$ and $b = 0.20$ and specified misclassification probabilities.



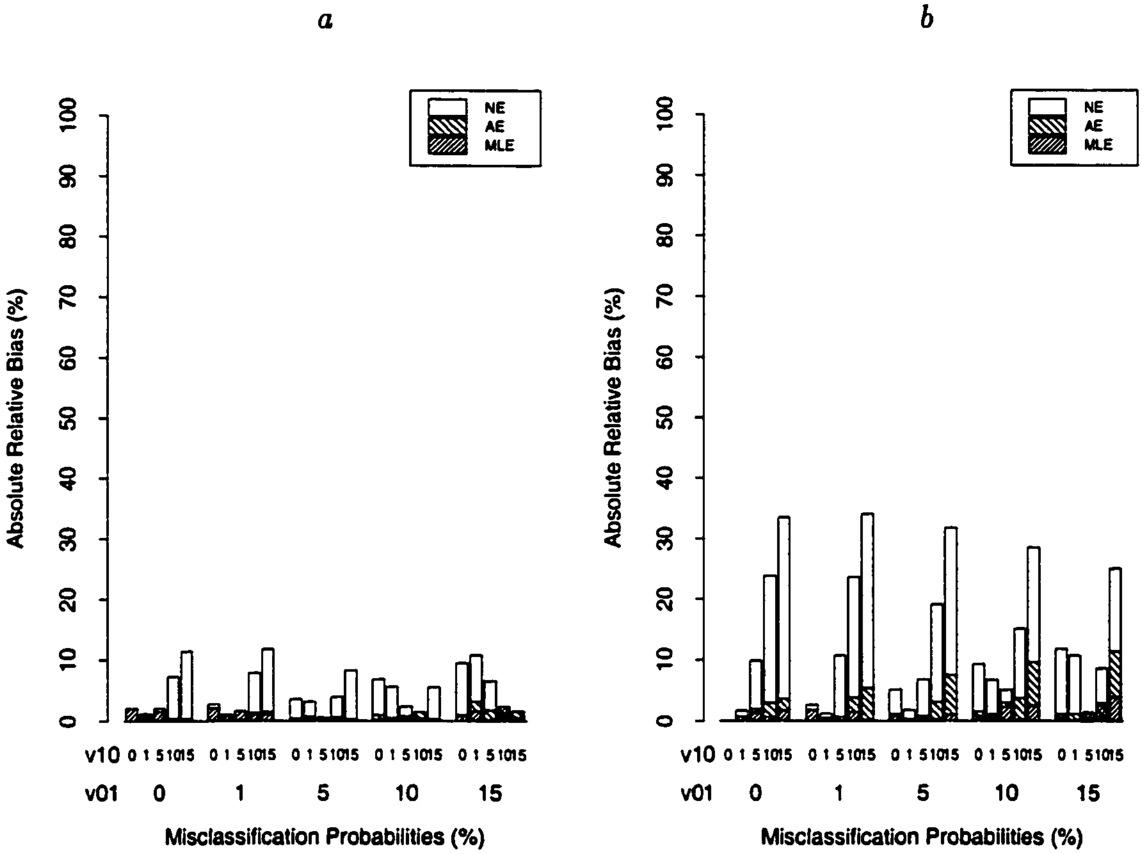
a

b

Table 5.6: Averages of 50 simulations for settings $a = 0.6$, $b = 0.3$, $\Delta t_{ij} = 1$ and specified misclassification probabilities.

v_{01}	v_{10}	$\widehat{a}_0 - a$	$\widehat{a} - a$	$\widehat{a}_\star - a$	$\widehat{b}_0 - b$	$\widehat{b} - b$	$\widehat{b}_\star - b$
0.00	0.00	0.0122	0.0120	0.0122	0.0001	0.0000	0.0001
0.00	0.01	0.0029	0.0070	0.0070	0.0050	-0.0021	-0.0020
0.00	0.05	-0.0123	0.0093	0.0094	0.0296	-0.0058	-0.0040
0.00	0.10	-0.0437	0.0021	0.0013	0.0715	0.0018	0.0088
0.00	0.15	-0.0684	0.0012	-0.0020	0.1005	-0.0050	0.0107
0.01	0.00	0.0163	0.0124	0.0125	-0.0079	-0.0054	-0.0054
0.01	0.01	0.0057	0.0052	0.0060	0.0036	-0.0013	-0.0008
0.01	0.05	-0.0096	0.0092	0.0092	0.0321	-0.0002	0.0018
0.01	0.10	-0.0473	-0.0064	-0.0075	0.0708	0.0042	0.0113
0.01	0.15	-0.0705	-0.0062	-0.0090	0.1021	-0.0003	0.0161
0.05	0.00	0.0212	0.0012	0.0022	-0.0153	-0.0028	-0.0033
0.05	0.01	0.0185	0.0030	0.0039	-0.0052	0.0007	0.0005
0.05	0.05	-0.0029	-0.0018	-0.0008	0.0203	-0.0022	0.0002
0.05	0.10	-0.0235	0.0030	0.0027	0.0573	0.0005	0.0091
0.05	0.15	-0.0497	0.0010	-0.0010	0.0952	0.0029	0.0224
0.10	0.00	0.0412	0.0012	0.0054	-0.0279	-0.0025	-0.0044
0.10	0.01	0.0337	-0.0027	0.0013	-0.0200	-0.0018	-0.0032
0.10	0.05	0.0140	-0.0049	-0.0009	0.0152	0.0068	0.0087
0.10	0.10	-0.0009	0.0022	0.0086	0.0453	-0.0004	0.0110
0.10	0.15	-0.0327	-0.0016	-0.0012	0.0856	0.0075	0.0287
0.15	0.00	0.0570	-0.0048	0.0056	-0.0355	0.0034	-0.0006
0.15	0.01	0.0645	0.0097	0.0190	-0.0321	0.0004	-0.0034
0.15	0.05	0.0391	0.0008	0.0103	-0.0012	0.0038	0.0043
0.15	0.10	0.0072	-0.0141	-0.0034	0.0257	-0.0085	0.0019
0.15	0.15	-0.0097	0.0031	0.0093	0.0751	0.0117	0.0343

Figure 5.5: Percent absolute relative bias of a and b NEs, AEs and MLEs for simulations with $a = 0.60$ and $b = 0.30$ and specified misclassification probabilities.



5.4.1 Convergence Problems

Convergence problems were seen in the estimation of the transition probabilities for some simulated data sets. The maximization algorithm maximizes the log-likelihood with respect to the parameters β_{ρ_0} and β_{η_0} which form the transition rates $\rho = \exp(\beta_{\rho_0})$ and $\eta = \exp(\beta_{\eta_0})$. In turn, these estimated rates along with the inter-observation time give the estimates of the transition probabilities. When two distinct sets of values for $\hat{\beta}_{\rho_0}$ and $\hat{\beta}_{\eta_0}$ provided the same maximum log-likelihood function, a convergence problem was detected.

These problems occurred only when the transition probabilities of the simulation settings summed to 0.8 or 0.9. The problems were more frequent when $a + b = 0.9$ and misclassification probabilities were not small. Of the 1250 data sets generated for the settings $a = 0.6$ and $b = 0.3$, 53 data sets reported convergence problems. For these situations, the true data sets were generated with estimated transition probabilities too close to the Bernoulli trials situation described in Section 4.4.1. While the sum of the simulation transition probability settings was constrained to be less than one, the random generation of the true data sets did produce some cases where the estimated transition probabilities did sum to one or more. These estimates were not consistent with the Type I counter model.

Recall from Section 2.5.1 that $(1 - a) + (1 - b) > 1$ is called the embeddability criterion (see Singer and Cohen (1980) and Kalbfleisch and Lawless (1985), for example). If the criterion is not satisfied, the transition probabilities do not correspond to any continuous-time model. Such probabilities are only possible in a discrete-time model. This result is easily seen by investigating the relationship between the transition rates and probabilities. If $\hat{a}+\hat{b} \geq 1$, then the estimates $\hat{\beta}_{\rho_0}$ and $\hat{\beta}_{\eta_0}$ lie in the complex plane. The transition rates constructed from the transition probabilities are

$$\begin{aligned} \rho = e^{\beta_{\rho_0}} &= \frac{-a}{(a+b)\,t} \log(1-a-b) \\ \eta = e^{\beta_{\eta_0}} &= \frac{-b}{(a+b)\,t} \log(1-a-b) \end{aligned}$$

where $t = 1$ in the simulations conducted here. Hence, if $a+b=1$, the log function is not defined and neither are ρ and η. If $a+b>1$, ρ and η are complex numbers. The maximization algorithm does not search the complex plane for parameter values which maximize the likelihood function and two real solutions give the largest log-likelihood function possible.

In all cases where convergence was a problem, the observed data satisfied

$$\hat{a}_0 + \hat{b}_0 = \frac{n_{01}}{n_{00}+n_{01}} + \frac{n_{10}}{n_{10}+n_{11}} \geq 1 \tag{5.7}$$

and two sets of transition probability estimates yielded the same likelihood function. Generally, if (5.7) was satisfied, then $1 \leq \hat{a}_0 + \hat{b}_0 \leq 1.05$. However, this feature did not necessarily cause the convergence problems. In several instances, the transition probabilities were estimated without problems despite the inconsistency of the point estimates from the true data with the model simulation settings.

A few cases had probability estimates from the true data which summed to slightly more than one, but with the observed data not satisfying (5.7). In these cases, convergence of estimates was not a problem. Additionally, an observed data set with property (5.7) was not sufficient to cause convergence problems. That is, a few simulations resulted in observed data which satisfied (5.7) but did not result in convergence problems. These cases numbered two in the simulations with $a = 0.6$ and $b = 0.3$.

The values of the naive estimates for the misclassified data seem to be more important determinants of convergence problems than the transition probability estimates from the true data set.

Chapter 6

Model Assessment

6.1 Overview

Model assessment techniques are considered for the models estimated previously in Chapter 3. Expected one-step transition pairs, two-step transition pairs and one-step transition triplets are calculated under a stationarity assumption and compared with observed counts. Such standard goodness-of-fit techniques for two-state Markov models prove to be inadequate in the sense that these comparisons do not detect important departures from the model. Likelihood ratio tests are shown for continuous covariate models by comparing saturated indicator models with the continuous covariate models. A simulation study suggests that a gamma distributed dead time may be more appropriate to consider for the MS data set.

6.2 Introduction

The previous chapters have discussed model specification, parameter identification and estimation as well as estimator properties. Model assessment is another important component of the model building process.

Billingsley (1961) reviews topics such as goodness-of-fit related to Markov chains. When the observations are made at discrete times, Markov models are generally assessed by comparing observed and expected transition counts under a specific model. Runs tests have been proposed by many authors (see Goodman (1958), for example) which assess the Markov property. More recently, Reeves (1993) focuses on run lengths of 0s and 1s in the two-state Markov chain. Tests for the predominance of the value one, departure from constant hazard, consistency of the equilibrium distribution, and for the geometric distribution in the presence of censoring are proposed. Further references on model assessment can be found in Lindsey (1993) and Agresti (1990).

Comparing observed and expected counts is the approach taken in Section 6.3. The observed and expected counts are compared for one-step transition pairs, two-step transition pairs and one-step transition triplets for the PI and MS data sets. In Section 6.4, MS saturated indicator models and continuous covariate models are compared using likelihood ratio tests. Gamma dead times are simulated in Section 6.5 to suggest an appropriate model for the MS data set.

6.3 Goodness-of-fit for Categorical Covariates

Even if the underlying process is Markov, the observed process will not have the Markov property when misclassification is present. However, the observed transition counts can be thought of as a reduction of the data for misclassification models even though they are not sufficient statistics. The one-step and two-step transition counts as well as the one-step transition triplets are considered for the MS single dichotomous covariate models and the model for the parasitic infection data set. Expected counts are calculated under a stationarity assumption and compared with the observed counts. For the MS data set, with misclassification predictors and covariates, these counts are stratified.

6.3.1 One-Step Transition Counts

This section presents a Pearson goodness-of-fit statistic as an indicator of model fit. The statistic requires notation for the observed and expected one-step transition counts which are stratified by misclassification predictors and covariate level. Let $N^{(i)}_{de,c_1c_2}$ be the random variable representing the number of transitions from state d to state e when subject i has misclassification predictors equal to c_1 and c_2 on subsequent visits and covariate x_i, $d, e, c_1, c_2 \in \{0, 1\}$, $i = 1, \ldots, I$. The analogous observed counts are represented by $n^{(i)}_{de,c_1c_2}$. In the MS data set for example, if subject i has active scan sequence 01010 and exacerbation sequence 11110, then $n^{(i)}_{01,11} = 2$ and $n^{(i)}_{10,11} = n^{(i)}_{10,10} = 1$. To observe state d followed by state e at subsequent visits, there are four possibilities: both states are correctly classified, both states are misclassified, state d is misclassified and state e is correctly classified, or state d is correctly classified and state e is misclassified. With these possibilities, the expected value for subject i with covariate x_i, observed process pair (d, e) and misclassification predictor pair (c_1, c_2) is

$$\begin{aligned} E(N^{(i)}_{de,c_1c_2}; x_i) = \sum_{j \in J^{(i)}_{c_1c_2}} \Big(& v_{dd}(c_1)\, v_{ee}(c_2)\, \pi_d(x_i)\, P_{de}(\Delta t_{ij}, x_i) \\ & + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, \pi_{1-d}(x_i)\, P_{1-d,1-e}(\Delta t_{ij}, x_i) \\ & + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, \pi_{1-d}(x_i)\, P_{1-d,e}(\Delta t_{ij}, x_i) \\ & + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, \pi_d(x_i)\, P_{d,1-e}(\Delta t_{ij}, x_i) \Big) \end{aligned} \tag{6.1}$$

where $J^{(i)}_{c_1c_2} = \{j \,|\, C_{i,j-1} = c_1,\ C_{ij} = c_2,\ j = 2, \ldots, n_i\}$ and $\pi_d(x_i) = \Pr(\xi_{ij} = d, x_i)$ for $d, e, c_1, c_2 \in \{0, 1\}$, $i = 1, \ldots, I$. Since $\rho(x_i) = \exp(\beta_{\rho 0} + \beta_{\rho 1} x_i)$ and $\eta(x_i) = \exp(\beta_{\eta 0} + \beta_{\eta 1} x_i)$, then $\pi_1(x_i) = \rho(x_i)/(\rho(x_i) + \eta(x_i))$ and $\pi_0(x_i) = 1 - \pi_1(x_i)$. Let $\delta_{x_i,l}$ be an indicator function taking the value 1 if $x_i = l$ and 0 if $x_i \neq l$, where l is a level of the covariate, $l \in \{x_i\}_i$. Summing over subjects, the total expected and observed number of observed process pairs (d, e) for subjects with misclassification predictor pairs (c_1, c_2) and covariate level l are

$$E(N_{de,c_1c_2,l}) = \sum_{i=1}^{I} \delta_{x_i,l}\, E(N^{(i)}_{de,c_1c_2}; x_i)$$

and $n_{de,c_1c_2,l} = \sum_{i=1}^{I} \delta_{x_i,l} n^{(i)}_{de,c_1c_2}$, respectively. The Pearson goodness-of-fit statistic

$$\chi^2_{\dagger} = \sum_{d,e,c_1,c_2,l} \frac{\left(n_{de,c_1c_2,l} - E(N_{de,c_1c_2,l})\right)^2}{E(N_{de,c_1c_2,l})}$$

can be used as an indicator of model fit. Similarly, the unstratified statistic

$$\chi^2_{\dagger\dagger} = \sum_{d,e} \frac{\left(n_{de} - E(N_{de})\right)^2}{E(N_{de})}$$

can be calculated where $E(N_{de}) = \sum_{c_1,c_2,l} E(N_{de,c_1c_2,l})$ and $n_{de} = \sum_{c_1,c_2,l} n_{de,c_1c_2,l}$. Models with continuous covariates can also be assessed in this manner. However, the dependence of the transition counts does not permit formal tests of fit based on a χ^2 distribution.

For the parasitic infection data with full likelihood estimates given in Table 3.6 and $v_{01} = 0$, the observed transition counts are seen in Table 6.1. The observed counts appear to be reasonable for the Markov assumption, as would be expected because the model has 3 parameters.

Table 6.1: Observed and expected transition counts for the parasitic infection data set under the perfect specificity model where the full likelihood estimates are $\hat{\rho} = 0.038$/day, $\hat{\eta} = 0.034$/day, $\hat{v}_{10} = 0.092$ and $v_{01} = 0$.

Parasitic Infection Transition Counts				
	00	01	10	11
Observed	378	119	112	321
Expected	372.834	116.237	116.237	323.692
			$\chi^2_{\dagger\dagger}$:	0.334

The MS observed and expected counts stratified by covariates sex and treatment as well as exacerbation status appear in Table 6.2 for the models given previously in Table 3.4. Recall that the three treatment levels are placebo (P), low dose (L) and high dose (H). No comparisons need to be conducted in these cases since the

Table 6.2: MS observed and expected transition counts for combined data and stratified by exacerbation status for the two covariates sex and treatment. Expected counts in parentheses for male (M), females (F), placebo (P), low dose (L) and high dose (H) subjects.

Exacerbation	Status	Scan Activity			
		00	01	10	11
	00	53 (52.830)	14 (14.084)	10 (14.084)	16 (12.002)
M	01	11 (14.558)	4 (2.710)	6 (4.013)	3 (2.719)
	10	15 (14.558)	3 (4.013)	3 (2.710)	3 (2.719)
	11	3 (1.944)	0 (0.377)	0 (0.377)	0 (0.301)
	00	216 (212.281)	42 (39.917)	40 (39.917)	20 (25.885)
F	01	46 (48.952)	8 (5.770)	10 (9.463)	5 (4.815)
	10	49 (48.254)	7 (9.314)	5 (5.675)	7 (4.756)
	11	7 (6.790)	0 (0.830)	1 (0.830)	1 (0.551)
					$\chi_{\dagger}^2$: 12.525
Total		400 (400.167)	78 (77.015)	75 (77.069)	55 (53.748)
					$\chi_{\dagger\dagger}^2$: 0.097

Exacerbation	Status	Scan Activity			
		00	01	10	11
	00	66 (61.398)	22 (23.754)	17 (23.754)	23 (19.093)
P	01	11 (15.781)	5 (4.842)	9 (6.004)	6 (4.374)
	10	14 (15.276)	6 (5.806)	4 (4.682)	6 (4.236)
	11	2 (1.621)	0 (0.487)	0 (0.487)	1 (0.405)
	00	92 (91.485)	18 (15.941)	15 (15.941)	7 (8.633)
L	01	22 (25.972)	5 (2.512)	6 (4.462)	2 (2.054)
	10	26 (25.983)	3 (4.452)	3 (2.502)	3 (2.064)
	11	5 (4.768)	0 (0.449)	1 (0.449)	0 (0.333)
	00	111 (111.907)	16 (15.747)	18 (15.747)	6 (7.599)
H	01	24 (21.458)	2 (1.368)	1 (2.977)	0 (1.198)
	10	24 (21.457)	1 (2.977)	1 (1.368)	1 (1.197)
	11	3 (2.557)	0 (0.158)	0 (0.158)	0 (0.127)
					$\chi_{\dagger}^2$: 22.375
Total		400 (401.220)	78 (78.371)	75 (78.432)	55 (49.976)
					$\chi_{\dagger\dagger}^2$: 0.660

observed and expected numbers are very similar in the stratified and unstratified tables. If a rough test is to be conducted, recall from Section 4.4.3 that the stratified table constructed in this manner for a covariate with two levels has around 14 degrees of freedom less the number of parameters estimated. Similarly, a 3 level covariate such as treatment has around a third more degrees of freedom associated with its table. The transition counts are modeled well for both sex and treatment models. However, the estimated mean occupancy times are unreasonable in both models except for low dose subjects in the relapsing state. Examining transition counts in this manner does not provide any insight into the fit of the model.

6.3.2 Two-Step Transition Counts

Now consider a slightly different goodness-of-fit test for such tables. Nagelkerke, Chunge and Kinoti (1990) looked at the frequency distribution of two-step transition counts. Rather than counting consecutive pairs of observations, pairs of observations are constructed two observations apart. Let $M^{(i)}_{de,c_1c_2}$ be the random variable representing the number of observations where state d is observed at time $t_{i,j-2}$ and state e is observed at time t_{ij} when subject i has misclassification predictors equal to c_1 and c_2 at these times and covariate x_i, $d, e, c_1, c_2 \in \{0, 1\}$, $i = 1, \dots, I$, $j = 3, \dots, n_i$. The analogous observed counts are denoted by $m^{(i)}_{de,c_1c_2}$ for subject i. In the example referred to earlier for a subject with active scan sequence 01010 and exacerbation sequence 11110, the observed two-step transition counts are $m^{(i)}_{00,11}$, $m^{(i)}_{00,10}$ and $m^{(i)}_{11,11}$ are each one.

The probabilities associated with these two-step transition counts are calculated in much the same way as for the one-step transition counts. The two-step transition counts sum over the possibilities for the true state occurring between the two observed states. The expected value for subject i with covariate x_i, observed process two-step pair (d, e) and misclassification predictor two-step pair (c_1, c_2) is

$$
\begin{aligned}
&E(M^{(i)}_{de,c_1c_2}; x_i) = \\
&\quad \sum_{j \in \bar{J}^{(i)}_{c_1c_2}} \Bigg(v_{dd}(c_1)\, v_{ee}(c_2)\, \pi_d(x_i) \sum_{k=0}^{1} P_{dk}(\Delta t_{i,j-1}, x_i) P_{ke}(\Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, \pi_{1-d}(x_i) \sum_{k=0}^{1} P_{1-d,k}(\Delta t_{i,j-1}, x_i) P_{k,1-e}(\Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, \pi_{1-d}(x_i) \sum_{k=0}^{1} P_{1-d,k}(\Delta t_{i,j-1}, x_i) P_{ke}(\Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, \pi_d(x_i) \sum_{k=0}^{1} P_{dk}(\Delta t_{i,j-1}, x_i) P_{k,1-e}(\Delta t_{ij}, x_i) \Bigg)
\end{aligned}
\tag{6.2}
$$

where $\bar{J}^{(i)}_{c_1c_2} = \{j \,|\, C_{i,j-2} = c_1,\, C_{ij} = c_2,\, j = 3, \ldots, n_i\}$. Summing over subjects, the total expected and observed number of observed state process two-step pairs (d, e) for subjects with misclassification predictor two-step pairs (c_1, c_2) and covariate level l are

$$E(M_{de,c_1c_2,l}) = \sum_{i=1}^{I} \delta_{x_i,l}\, E(M^{(i)}_{de,c_1c_2}; x_i)$$

and $m_{de,c_1c_2,l} = \sum_{i=1}^{I} \delta_{x_i,l} m^{(i)}_{de,c_1c_2}$, respectively. The associated Pearson goodness-of-fit statistic is denoted by

$$\chi^2_\ddagger = \sum_{d,e,c_1,c_2,l} \frac{(m_{de,c_1c_2,l} - E(M_{de,c_1c_2,l}))^2}{E(M_{de,c_1c_2,l})}.$$

Without stratification, the associated Pearson goodness-of-fit statistic is

$$\chi^2_{\ddagger\ddagger} = \sum_{d,e} \frac{(m_{de} - E(M_{de}))^2}{E(M_{de})}$$

where $E(M_{de}) = \sum_{c_1,c_2,l} E(M_{de,c_1c_2,l})$ and $m_{de} = \sum_{c_1,c_2,l} m_{de,c_1c_2,l}$.

The results for the parasitic infection data tell a slightly different story than the one-step transition counts. For the full likelihood model with only one misclassification probability, the two-step transition counts appear in Table 6.3. Nagelkerke, Chunge and Kinoti (1990) performed this goodness-of-fit test with the maximum likelihood parameter estimates obtained by excluding the initial state. For that model, the estimates were $\hat{\rho} = 0.0383$/day, $\hat{\eta} = 0.0337$/day and $\hat{v}_{10} = 0.091$ with test statistic $\chi^2_{\ddagger\ddagger} = 6.64$. The authors acknowledged that the $\chi^2_{\ddagger\ddagger}$ value was marginally significant at the 0.05 significance level and suggested that the excess of 00 observations indicated some minor heterogeneity in the state 0 to state 1 transition rate. Although the two-step transition counts are less dependent than the one-step transition counts, these counts are still dependent. Hence, 2 degrees of freedom may be approximately correct.

The observed and expected two-step counts for the covariates sex and treatment appear in Table 6.4 for the MS models given previously in Table 3.4. Again, the expected counts appear to be close to the observed counts, which indicates the

Table 6.3: Observed and expected two-step transition counts for the parasitic infection data set under the perfect specificity model where $\hat{\rho} = 0.038/\text{day}$, $\hat{\eta} = 0.034/\text{day}$, $\hat{v}_{10} = 0.092$ and $v_{01} = 0$.

Parasitic Infection Two-Step Transition Counts				
	0 0	0 1	1 0	1 1
Observed	328	143	128	273
Expected	306.649	151.920	151.920	261.510
			$\chi^2_{\ddagger\ddagger}$ on 2 df:	6.281

model is consistent with the data despite the extreme parameter estimates.

6.3.3 One-Step Transition Triplet Counts

Rather than counting consecutive pairs of observations, triples of observations can be counted. Let $N^{(i)}_{def,c_1c_2c_3}$ be the random variable representing the number of observations where state d is observed at time $t_{i,j-2}$, e is observed at time $t_{i,j-1}$ and state f is observed at time t_{ij} when subject i has misclassification predictors equal to c_1, c_2 and c_3 at these times and covariate x_i, $d, e, f, c_1, c_2, c_3 \in \{0, 1\}$, $i = 1, \ldots, I$, $j = 3, \ldots, n_i$. Similarly, $n^{(i)}_{def,c_1c_2c_3}$ denotes the observed counts for subject i with observed process (d, e, f), misclassification predictors (c_1, c_2, c_3) and covariate x_i.

The probabilities associated with these transition triplet counts are calculated in much the same way as the one-step transition counts. For each observation triplet, either all observations are classified correctly, only one observation is misclassified, only two observations are misclassified or all three observations are misclassified. The expected value for subject i with covariate x_i, observed process triplet (d, e, f)

Table 6.4: MS observed and expected two-step transition counts for combined data and stratified by exacerbation status for the two covariates sex and treatment. Expected counts in parentheses for male (M), females (F), placebo (P), low dose (L) and high dose (H) subjects.

Exacerbation Status		Scan Activity			
		0 0	0 1	1 0	1 1
	0 0	52 (51.694)	13 (13.781)	11 (13.781)	15 (11.743)
M	0 1	9 (10.919)	2 (2.032)	5 (3.010)	2 (2.039)
	1 0	12 (12.739)	5 (3.511)	2 (2.371)	2 (2.379)
	1 1	1 (3.240)	2 (0.629)	1 (0.629)	1 (0.502)
	0 0	210 (206.683)	38 (39.963)	38 (39.963)	25 (24.391)
F	0 1	39 (38.859)	6 (4.760)	7 (7.704)	3 (3.677)
	1 0	37 (36.045)	8 (7.131)	4 (4.402)	2 (3.422)
	1 1	11 (13.528)	2 (1.711)	3 (1.711)	2 (1.050)
					$\chi^2_{\ddagger}$: 13.095
Total		371 (373.707)	76 (73.518)	71 (73.571)	52 (49.203)
					$\chi^2_{\ddagger\ddagger}$: 0.352

Exacerbation Status		Scan Activity			
		0 0	0 1	1 0	1 1
	0 0	63 (60.761)	20 (23.727)	18 (23.727)	26 (18.786)
P	0 1	8 (11.677)	5 (3.624)	6 (4.486)	4 (3.213)
	1 0	10 (11.679)	7 (4.484)	5 (3.622)	1 (3.215)
	1 1	2 (3.773)	2 (1.146)	2 (1.146)	1 (0.935)
	0 0	88 (89.705)	17 (16.094)	17 (16.094)	8 (8.108)
L	0 1	23 (20.683)	2 (2.104)	2 (3.664)	1 (1.548)
	1 0	21 (20.693)	5 (3.654)	1 (2.094)	1 (1.559)
	1 1	4 (7.119)	1 (0.707)	2 (0.707)	2 (0.467)
	0 0	111 (107.311)	14 (15.271)	14 (15.271)	6 (7.148)
H	0 1	17 (17.458)	1 (1.141)	4 (2.452)	0 (0.950)
	1 0	18 (16.664)	1 (2.340)	0 (1.089)	2 (0.906)
	1 1	6 (5.957)	1 (0.378)	0 (0.378)	0 (0.287)
					$\chi^2_{\ddagger}$: 31.974
Total		371 (373.480)	76 (74.670)	71 (74.729)	52 (47.122)
					$\chi^2_{\ddagger\ddagger}$: 0.731

and misclassification triplet (c_1, c_2, c_3) is

$$\begin{aligned}
&E(N^{(i)}_{def,c_1c_2c_3}; x_i) = \\
&\sum_{j \in \bar{J}^{(i)}_{c_1c_2c_3}} \Big(v_{dd}(c_1)\, v_{ee}(c_2)\, v_{ff}(c_3)\, \pi_d(x_i)\, P_{de}(\Delta t_{i,j-1}, x_i)\, P_{ef}(\Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{ee}(c_2)\, v_{1-f,f}(c_3)\, \pi_d(x_i)\, P_{de}(\Delta t_{i,j-1}, x_i)\, P_{e,1-f}(\Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, v_{ff}(c_3)\, \pi_d(x_i)\, P_{d,1-e}(\Delta t_{i,j-1}, x_i)\, P_{1-e,f}(\Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, v_{ff}(c_3)\, \pi_{1-d}(x_i)\, P_{1-d,e}(\Delta t_{i,j-1}, x_i)\, P_{ef}(\Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, v_{1-f,f}(c_3)\, \pi_d(x_i)\, P_{d,1-e}(\Delta t_{i,j-1}, x_i)\, P_{1-e,1-f}(\Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, v_{1-f,f}(c_3)\, \pi_{1-d}(x_i)\, P_{1-d,e}(\Delta t_{i,j-1}, x_i)\, P_{e,1-f}(\Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, v_{ff}(c_3)\, \pi_{1-d}(x_i)\, P_{1-d,1-e}(\Delta t_{i,j-1}, x_i)\, P_{1-e,f}(\Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, v_{1-f,f}(c_3)\, \pi_{1-d}(x_i) \\
&\qquad\qquad\qquad \times\, P_{1-d,1-e}(\Delta t_{i,j-1}, x_i)\, P_{1-e,1-f}(\Delta t_{ij}, x_i) \Big)
\end{aligned} \tag{6.3}$$

where $\bar{J}^{(i)}_{c_1c_2c_3} = \{j \mid C_{i,j-2} = c_1,\ C_{i,j-1} = c_2,\ C_{ij} = c_3,\ j = 3, \ldots, n_i\}$. Summing over subjects, the total expected and observed number of state triplets (d, e, f) for subjects with misclassification predictor triplets (c_1, c_2, c_3) and covariate level l are

$$E(N_{def,c_1c_2c_3,l}) = \sum_{i=1}^{I} \delta_{x_i,l}\, E(N^{(i)}_{def,c_1c_2c_3}; x_i)$$

and $n_{def,c_1c_2c_3,l} = \sum_{i=1}^{I} \delta_{x_i,l}\, n^{(i)}_{def,c_1c_2c_3}$, respectively. The associated Pearson statistic is

$$\chi^2_* = \sum_{d,e,f,c_1,c_2,c_3,l} \frac{\left(n_{def,c_1c_2c_3,l} - E(N_{def,c_1c_2c_3,l})\right)^2}{E(N_{def,c_1c_2c_3,l})}.$$

Summing over misclassification predictors and the covariate levels gives

$$\chi^2_{**} = \sum_{d,e,f} \frac{\left(n_{def} - E(N_{def})\right)^2}{E(N_{def})}$$

where $E(N_{def}) = \sum_{c_1,c_2,c_3,l} E(N_{def,c_1c_2c_3,l})$ and $n_{def} = \sum_{c_1,c_2,c_3,l} n_{def,c_1c_2c_3,l}$.

The observed and expected triplet counts for the parasitic infection data appear in Table 6.5. For the very same reason that the n_{01} and n_{10} counts were not completely independent in the one-step transition counts, the one-step triplet counts n_{001} and n_{100} are not completely independent and additionally, n_{011} and n_{110} are not completely independent. Hence, Table 6.5 has fewer than 5 degrees of freedom. The Markov model appears to produce a poorer fit to the transition triplet counts than the one-step or two-step transition counts examined previously.

Table 6.5: Observed and expected transition triplet counts for the parasitic infection data set under the perfect specificity model where $\hat{\rho} = 0.038$/day, $\hat{\eta} = 0.034$/day, $\hat{v}_{10} = 0.092$ and $v_{01} = 0$.

	Parasitic Infection Transition Triplet Counts							
	000	001	010	100	011	101	110	111
Observed	285	77	43	67	66	35	61	238
Expected	277.918	71.664	28.731	71.664	80.257	37.324	80.257	224.186
							χ^2_{**}:	16.117

For the MS sex and treatment models, the observed and expected triplet counts stratified by exacerbation status triplets and covariate levels give sparse tables. The treatment model results appear in Table 6.7. The observed and expected triplet counts for the combined exacerbation status triplets and covariate levels are given in Table 6.6. The observed and expected counts are very close despite the extreme parameter estimates for the data which is not stratified by exacerbation status. For the treatment model, the number of 111 active scan triplets expected is smaller than observed. When the data are stratified by the covariate level and misclassification predictor triplets, the fit is not as nice. Similar to the situation seen with the two-step transitions, the placebo patients with 000 exacerbation triplets appear to have poorer fits for the active scan triplets than the rest of the table. Otherwise, the expected counts match well with the observed counts for both sparse cells and cells that are not so sparse. Incidentally, the data set did not have any subjects which had three consecutive exacerbations.

Table 6.6: Observed and expected transition triplet counts for the MS sex and treatment models.

Sex Model								
	Scan Activity Triplets							
	000	001	010	100	011	101	110	111
Observed	325	51	46	47	25	24	24	28
Expected	327.464	47.547	46.237	47.580	29.973	24.774	25.992	24.460
							χ^2_{**}:	1.790

Treatment Model								
	Scan Activity Triplets							
	000	001	010	100	011	101	110	111
Observed	325	51	46	47	25	24	24	28
Expected	327.393	48.568	47.542	48.609	25.952	25.012	25.979	20.944
							χ^2_{**}:	2.846

Table 6.7: Observed and expected transition triplets by exacerbation status and treatment level. Expected counts are in parentheses.

Exac'n		Scan Activity							
	Status	000	001	010	100	011	101	110	111
P	000	47(37.940)	9(10.987)	8(10.861)	7(10.987)	8(8.068)	7(7.942)	5(8.068)	11(7.147)
	001	4 (7.926)	3 (1.668)	2 (2.228)	1 (2.255)	2(1.483)	1(1.456)	5(1.646)	2(1.338)
	010	6 (9.910)	2 (2.817)	2 (2.050)	5 (2.820)	1(1.855)	4(2.021)	1(1.852)	4(1.675)
	100	7 (7.929)	3 (2.256)	1 (2.227)	2 (1.664)	4(1.643)	1(1.457)	2(1.486)	0(1.339)
	011	2 (1.267)	0 (0.260)	0 (0.256)	0 (0.354)	0(0.213)	0(0.227)	0(0.231)	1(0.192)
	101	2 (2.958)	1 (0.606)	0 (0.815)	1 (0.606)	1(0.540)	0(0.486)	1(0.540)	1(0.448)
	110	2 (1.268)	0 (0.354)	0 (0.256)	0 (0.260)	0(0.231)	0(0.227)	1(0.213)	0(0.192)
L	000	60(61.882)	10 (8.813)	12 (8.500)	10 (8.808)	2(3.816)	4(3.508)	1(3.820)	3(2.853)
	001	16(14.961)	2 (0.982)	3 (2.030)	0 (2.105)	0(0.745)	0(0.670)	1(0.904)	1(0.603)
	010	14(18.221)	3 (2.557)	2 (1.101)	5 (2.565)	2(0.908)	0(1.004)	1(0.901)	1(0.743)
	100	15(14.323)	3 (2.011)	1 (1.937)	0 (0.926)	2(0.858)	1(0.644)	0(0.718)	0(0.582)
	011	3 (3.486)	0 (0.224)	1 (0.206)	1 (0.487)	0(0.153)	0(0.150)	0(0.168)	0(0.125)
	101	4 (6.278)	1 (0.402)	0 (0.841)	0 (0.401)	0(0.305)	1(0.244)	2(0.306)	1(0.224)
	110	5 (4.187)	0 (0.581)	0 (0.246)	1 (0.268)	0(0.203)	0(0.182)	0(0.183)	0(0.151)
H	000	78(82.365)	11 (9.534)	13 (9.404)	11 (9.533)	2(3.525)	4(3.397)	3(3.527)	2(2.716)
	001	15(13.547)	1 (0.534)	0 (1.531)	2 (1.553)	0(0.451)	0(0.428)	1(0.564)	0(0.392)
	010	19(14.975)	1 (1.715)	1 (0.567)	0 (1.714)	0(0.496)	0(0.600)	0(0.497)	0(0.435)
	100	15(12.833)	0 (1.472)	0 (1.450)	0 (0.506)	1(0.534)	1(0.406)	0(0.427)	1(0.371)
	011	2 (2.296)	0 (0.088)	0 (0.084)	1 (0.261)	0(0.068)	0(0.070)	0(0.074)	0(0.059)
	101	6 (5.358)	1 (0.205)	0 (0.599)	0 (0.205)	0(0.173)	0(0.149)	0(0.173)	0(0.138)
	110	3 (2.297)	0 (0.260)	0 (0.084)	0 (0.088)	0(0.074)	0(0.070)	0(0.068)	0(0.059)

6.4 Likelihood Ratio Tests for Continuous Covariates

Model reduction can be conducted on the continuous covariates using likelihood ratio tests. Indicators can be defined for each distinct value of a covariate. The model including indicators of all levels of a covariate is referred to as the saturated model. This saturated model can be compared with the model with only the continuous covariate, the reduced model.

Saturated models were constructed for each of the standardized covariates age, duration and initial burden in the MS data set. The log-likelihood values appear for the full and reduced models in Table 6.8. In the case of age, the likelihood ratio test statistic is 7.34 on 42 $\big((23 \times 2 + 4) - (2 \times 2 + 4)\big)$ degrees of freedom. The indicator model is not required. The same conclusion is drawn for the initial burden and duration models. Hence, the reduced models for each variable perform as well as the saturated indicator models. Further likelihood ratio tests could be conducted on the models in Table 3.4.

Table 6.8: Log-likelihoods for MS models with continuous covariates and associated indicators. Each covariate value has an indicator in the saturated model.

	Covariate		
	Age	Initial Burden	Duration
number of distinct values	23	38	34
log-likelihood: saturated model	-284.552	-265.428	-271.480
log-likelihood: reduced model	-288.220	-290.782	-291.302

6.5 Simulated Gamma Dead Times

The extreme parameter estimates for the MS data set suggested that the model with exponential state sojourn time distributions may not be appropriate. The transition triplet counts of Section 6.3.3 confirmed this conclusion. Changing the sojourn time distributions is next considered. As a natural extension to the exponential distribution, a gamma distribution was examined for the dead time.

The appropriateness of a gamma dead time distribution was determined by simulation. Data sets were simulated with the same observation times as the MS data set and the simulated transition counts were compared with the actual counts obtained. The simulation was carried out in the absence of misclassification predictors and covariates. However, fixed values of the misclassification probabilities were considered.

The data were generated by first simulating the true process where exponential intervals of mean length $1/\rho$ and gamma intervals of mean length κ/ρ were constructed for the state 0 and state 1 sojourn times, respectively. Initially, an interval for state 0 was generated and subsequent intervals alternate between state 1 and state 0. The algorithm generated 100 transitions before the history for the subjects was recorded. Each subject started in state 0 at time 0. Preliminary investigations suggested mean times of 180, 210 and 240 days for state 0, $\eta = 0.1$ per day and $\kappa = 5, 6, 7, 8$ or 9 for state 1.

For each subject at each observation time, a uniform random number u_{ij}^* between 0 and 1 was generated to determine if the true state was misclassified, $i = 1, \ldots, 38$, $j = 1, \ldots, 17$. The misclassification generation follows in the same manner as described in Section 5.4. An observation which was 0 became misclassified as a 1 if $u^*_{ij} < v_{01}$ and an observation that was 1 became misclassified as a 0 if $u^*_{ij} < v_{10}$. Misclassification probabilities of 0, 0.2, 0.05 and 0.10 were considered.

Transition triplet counts for selected simulated data sets appear in Table 6.9. In the absence of covariates and misclassification predictors, these results indicate that a model where the state 0 sojourn time is exponentially distributed and the state 1 sojourn time has a gamma distribution is worth pursuing further. Nonexponential sojourn time distributions leads to the consideration of more general semi-Markov models in the next chapter.

Table 6.9: Simulated data sets with exponential and gamma distributed sojourn times for states 0 and 1, respectively, based on the MS observation times. The mean times are given in days for the two states and $\eta = 0.1/\text{day}$.

Simulated data:				Transition Triplet Counts							
$1/\rho$	κ/η	v_{01}	v_{10}	n_{000}	n_{001}	n_{010}	n_{100}	n_{011}	n_{101}	n_{110}	n_{111}
180	50	0.02	0.02	334	55	41	55	26	15	28	16
210	60	0.02	0.02	323	57	40	52	32	17	31	18
240	60	0.02	0.10	336	54	43	51	25	16	24	21
240	70	0.05	0.05	294	63	48	60	34	18	34	19
240	80	0.02	0.02	311	58	38	55	36	14	35	23
Observed data:				325	51	46	47	25	24	24	28

Chapter 7

Semi-Markov and General Markov Processes

7.1 Overview

The suggestion of a gamma distributed dead time motivates a redevelopment of the true process specification. A semi-Markov model is proposed for the true process based on joint probabilities when the process is in equilibrium. When two states are considered, the semi-Markov model developed is equivalent to a Type I counter model with exponential open time distribution and gamma dead time distribution. Via approximation, the equilibrium probabilities at three consecutive time points are used with misclassification probabilities to calculate expected frequencies. Minimizing the discrepancy between the expected and observed counts leads to parameter estimation. This approach provides a better fit for the PI data set than the Markov model with misclassification implemented previously through a likelihood approach. Models for the MS data set suggest substantial heterogeneity which is further investigated by examining simulated run length distributions.

More general Markov models are also proposed. The Type I counter is extended to a general Markov process with a small number of states. State transitions can be made from any state to any other state and misclassification probabilities are defined for each type of misclassification.

7.2 Introduction

Counter models with exponential sojourn time distributions may be too restrictive for some processes. This may be the case in the MS data set where the estimates obtained indicated the model specification was not correct. Such results motivate consideration of a different model specification for the true process. The simulated data constructed in Section 6.5 suggested that a model with a gamma distributed dead time may be appropriate.

A semi-Markov model where the two state sojourn time distributions are exponential and gamma is developed in Section 7.3. The theoretical development is described in Section 7.3.1 and the implementation and results appear in Sections 7.3.2–7.3.5. Section 7.4 generalizes the model given in Chapter 2 to include Markov processes with a small number of states and more than one misclassification predictor or covariate.

7.3 Semi-Markov Process

7.3.1 Model Development

When the waiting time in each state is not exponential, a Markov process does not adequately describe the situation. The process may follow a continuous-time semi-Markov process where the sojourn times are independently distributed given the states entered. Semi-Markov models have been considered by many authors in several fields including insurance, reliability and medicine (see McGilchrist and Hills (1991) and Cox and Miller (1965), for example).

The goal of this section is to determine joint probabilities for states at three consecutive observation times under an equilibrium assumption. The probabilities will be used to calculate expected triplet counts which will be compared with the observed counts to obtain parameter estimates. Before the joint probabilities can be determined, several conditional probabilities are defined for a two-state semi-Markov process in continuous time. The probabilities conditional on a transition at time 0 are defined as

$$f_{ab}(t)dt = \Pr(\text{stay in } a \text{ until } t, \text{ jump to } b \text{ in } [t, t+dt) \qquad\qquad \\ \qquad\qquad \mid \text{transition into } a \text{ at time } 0) \tag{7.1}$$

$$u_{ab}(t) = \Pr(\text{in } b \text{ at time } t \mid \text{transition into } a \text{ at time } 0) \tag{7.2}$$

$$u_{abk}(t,\tau) = \Pr(\text{in } b \text{ at time } t,\ k \text{ at time } t+\tau \mid \text{transition into } a \text{ at time } 0) \tag{7.3}$$

where a, b and k are states of the process, $t, \tau \geq 0$ are times (or lengths of time) and by convention $f_{aa}(t) \equiv 0$. With these definitions, note that $\sum_b u_{ab}(t) = 1$ for all t and $\sum_{b,k} u_{abk}(t, \tau) = 1$ for all t, τ. These conditional probabilities are conditional on a transition into state a at time 0.

Much of the mathematical development that follows can be extended to handle more than two states and different distributions for the state sojourn times. However, attention is restricted here to the case of two states. Suppose that state 0 has an exponential sojourn time distribution with rate ρ and that state 1, the dead time, has a gamma sojourn time distribution with scale η and shape κ. The sojourn time probability distribution functions for the two states in this case are

$$f_{01}(t) = \rho\, e^{-\rho t} \qquad f_{10}(t) = \frac{\eta\,(\eta t)^{\kappa-1}\, e^{-\eta t}}{\Gamma(\kappa)} \qquad t > 0$$

where ρ and η may depend on baseline covariates in the same manner as described in Section 2.6, $\rho, \eta > 0$. The transition probabilities associated with the semi-Markov process are defined as

$$p_{ab} = \Pr(\text{next jump is to } b \,|\, \text{transition into } a) = \int_0^\infty f_{ab}(t)\, dt.$$

For the process in equilibrium, the joint probabilities for states at two or three consecutive observation times are

$$\pi_{ab}(t) = \Pr(\text{in } a \text{ at time } t_0 \text{ and } b \text{ at time } t_0 + t) \tag{7.4}$$

$$\pi_{abk}(t, \tau) = \Pr(\text{in } a \text{ at time } t_0,\ b \text{ at time } t_0 + t \text{ and } k \text{ at time } t_0 + t + \tau) \tag{7.5}$$

and the probability of occupying state a is defined as

$$\pi_a = \Pr(\text{in } a \text{ at time } 0) = \sum_b \pi_{ab}(t) = \sum_{b,k} \pi_{abk}(t,\tau)$$

for all t and τ.

Additionally, when the process is in equilibrium, the conditional probabilities for two states given the state occupied at time 0 is defined as

$$\phi_{ab}(t)\,dt = \Pr(\text{stay in } a \text{ until } t, \text{ then jump to } b \text{ in } [t, t+dt) \mid \text{ in } a \text{ at time } 0).$$

Here, the probability is conditional on occupying state a at time 0 rather than transitioning to state a at time 0 as in (7.1). For a two state alternating renewal process, these probabilities are forward recurrence time probabilities, with densities

$$\begin{aligned}
\phi_{01}(t) &= \frac{\int_t^\infty f_{01}(w)\,dw}{\int_0^\infty w\, f_{01}(w)\,dw} = \rho e^{-\rho t} \\
\phi_{10}(t) &= \frac{\int_t^\infty f_{10}(w)\,dw}{\int_0^\infty w\, f_{10}(w)\,dw} = \frac{\eta}{\kappa}(1 - I(\kappa, \eta t))
\end{aligned}$$

where $I(\kappa, \eta t)$ is the incomplete gamma function

$$I(\kappa, \eta t) = \frac{1}{\Gamma(\kappa)} \int_0^{\eta t} w^{\kappa-1}\, e^{-w}\, dw.$$

Further, the probability of occupying a certain state is a simple function of the mean sojourn times for each state in the alternating renewal process,

$$\pi_0 = \frac{\eta}{\eta + \kappa\rho} \qquad \pi_1 = \frac{\kappa\rho}{\eta + \kappa\rho}.$$

First, the formulas for the $u_{ab}(t)$ and $u_{abk}(t,\tau)$ are required to determine the equilibrium joint probabilities. The probabilities $u_{ab}(t)$ are evaluated as

$$u_{ab}(t) = \sum_{\ell \neq a} \int_0^t f_{a\ell}(w)\, u_{\ell b}(t-w)\, dw + \delta_{ab} \sum_{\ell \neq a} \int_t^\infty f_{a\ell}(w)\, dw \tag{7.6}$$

with the initial condition $u_{ab}(0) = \delta_{ab}$ where $\delta_{ab} = 1$ if $a = b$ and 0 otherwise. To prove (7.6), note that in order to have a transition from state a so that b is occupied at time t, a transition must be made to another state at some time $w < t$ and the process must occupy b after a further time interval of length $t - w$. If $a = b$, the above description is possible or the process makes a transition into a at time 0 and occupies a for an interval of length at least t.

The development for three consecutive states follows similarly. The $u_{abk}(t,\tau)$ are

$$\begin{aligned} u_{abk}(t,\tau) = &\sum_{\ell \neq a} \int_0^t f_{a\ell}(w)\, u_{\ell bk}(t-w,\tau)\, dw \\ &+ \delta_{ab} \sum_{\ell \neq a} \int_t^{t+\tau} f_{a\ell}(w)\, u_{\ell k}(t+\tau-w)\, dw + \delta_{abk} \sum_{\ell \neq a} \int_{t+\tau}^\infty f_{a\ell}(w)\, dw \end{aligned} \tag{7.7}$$

where $\delta_{abk} = \delta_{ab}\,\delta_{bk}$. The initial conditions are $u_{abk}(0,0) = \delta_{abk}$, $u_{abk}(0,\tau) = \delta_{ab}\, u_{bk}(\tau)$ and $u_{abk}(t,0) = u_{ab}(t)\,\delta_{bk}$. To prove (7.7), note that in order to have state 0 at three consecutive observation times, then one the following possibilities must occur:

- the process makes a transition into 0 at time 0, stays in 0 for an interval of length nearly w, where w is some value in $(0, t)$; makes a transition to another state; resets the clock time to 0; and occupies state 0 at time $t - w$ and state 0 at time $t - w + \tau$;
- the process makes a transition into 0 at time 0, stays in 0 for an interval of length nearly w, where w is some value in $(t, t + \tau)$; resets the clock time to 0; and occupies state 0 at time $t + \tau - w$;
- the process makes a transition into 0 at time 0 and stays in 0 for an interval of length at least $t + \tau$.

When the process is in equilibrium, the joint probabilities for states at two and three time points are calculated as

$$\begin{aligned}
\pi_{ab}(t) &= \pi_a \sum_{\ell \neq a} \int_0^t \phi_{a\ell}(w)\, u_{\ell b}(t - w)\, dw + \delta_{ab}\, \pi_a \sum_{\ell \neq a} \int_t^\infty \phi_{a\ell}(w)\, dw \\
\pi_{abk}(t, \tau) &= \pi_a \sum_{\ell \neq a} \int_0^t \phi_{a\ell}(w)\, u_{\ell bk}(t - w, \tau)\, dw \\
&\quad + \delta_{ab}\, \pi_a \sum_{\ell \neq a} \int_t^{t+\tau} \phi_{a\ell}(w)\, u_{\ell k}(t + \tau - w)\, dw + \delta_{abk}\, \pi_a \sum_{\ell \neq a} \int_{t+\tau}^\infty \phi_{a\ell}(w)\, dw
\end{aligned} \tag{7.8}$$

with initial conditions $\pi_{ab}(0) = \delta_{ab}\pi_a$, $\pi_{abk}(0, 0) = \delta_{abk}\pi_a$, $\pi_{abk}(0, \tau) = \delta_{ab}\pi_{bk}(\tau)$ and $\pi_{abk}(t, 0) = \pi_{ab}(t)\,\delta_{bk}$. Since the joint probabilities are not conditioned on the state at time 0, (7.8) have a similar form as (7.6) and (7.7) with $\phi_{ab}(t)$ and weights π_a instead of $f_{ab}(t)$.

Evaluating the joint probabilities $\pi_{ab}(t)$ and $\pi_{abk}(t, \tau)$ is not feasible unless there are only two states and the shape parameter is 1 or 2. In these cases, Laplace transforms can be used to solve (7.6) and (7.7). For all other situations, numerical methods are required to evaluate the joint probabilities. To solve (7.6)–(7.8), an approximation scheme is described in Section 7.3.2.

These joint probabilities can be used to calculate the expected counts of transitions for categorical covariates in much the same way as (6.1) developed in Section 6.3.1. The parameters ρ and η appearing in the preceding development become $\rho(\mathbf{x}_i) = \exp(\mathbf{x}_i \boldsymbol{\beta}_\rho)$ and $\eta(\mathbf{x}_i) = \exp(\mathbf{x}_i \boldsymbol{\beta}_\eta)$. Attention is restricted to the case of one categorical covariate. The expected value for subject i with covariate x_i, observed state pair (d, e) and misclassification predictor pair (c_1, c_2) is

$$
\begin{aligned}
E(G^{(i)}_{de,c_1c_2}; x_i) = \sum_{j \in \breve{J}^{(i)}_{c_1c_2}} \Big(& v_{dd}(c_1)\, v_{ee}(c_2)\, \pi_{de}(\Delta t_{ij}, x_i) \\
& + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, \pi_{d,1-e}(\Delta t_{ij}, x_i) \\
& + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, \pi_{1-d,e}(\Delta t_{ij}, x_i) \\
& + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, \pi_{1-d,1-e}(\Delta t_{ij}, x_i) \Big)
\end{aligned} \tag{7.9}
$$

where $\breve{J}^{(i)}_{c_1c_2} = \{j \mid C_{i,j-1} = c_1, C_{ij} = c_2,\ j = 2, \ldots, n_i\}$. Summing over subjects, the total expected counts of observed state pairs (d, e) with misclassification predictor pair (c_1, c_2) and covariate level l, $l \in \{x_i\}_i$, is

$$
E(G_{de,c_1c_2,l}) = \sum_{i=1}^{I} \delta_{x_i,l}\, E(G^{(i)}_{de,c_1c_2}; x_i)
$$

and the observed count is denoted by $n_{de,c_1c_2,l}$.

Similarly, the expected counts for transition triplets can also be calculated. Let $J^{(i)}_{c_1c_2c_3} = \{j \mid C_{i,j-2} = c_1,\, C_{i,j-1} = c_2, C_{ij} = c_3,\, j = 3, \ldots, n_i\}$. The expected value for subject i with covariate x_i, observed state triplet (d, e, f) and misclassification predictor triplet (c_1, c_2, c_3) is

$$
\begin{aligned}
&E(H^{(i)}_{def,c_1c_2c_3}; x_i) \\
&\quad = \sum_{j \in J^{(i)}_{c_1c_2c_3}} \Big(v_{dd}(c_1)\, v_{ee}(c_2)\, v_{ff}(c_3)\, \pi_{def}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{ee}(c_2)\, v_{1-f,f}(c_3)\, \pi_{de,1-f}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, v_{ff}(c_3)\, \pi_{d,1-e,f}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, v_{ff}(c_3)\, \pi_{1-d,ef}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{dd}(c_1)\, v_{1-e,e}(c_2)\, v_{1-f,f}(c_3)\, \pi_{d,1-e,1-f}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{ee}(c_2)\, v_{1-f,f}(c_3)\, \pi_{1-d,e,1-f}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, v_{ff}(c_3)\, \pi_{1-d,1-e,f}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \\
&\qquad + v_{1-d,d}(c_1)\, v_{1-e,e}(c_2)\, v_{1-f,f}(c_3)\, \pi_{1-d,1-e,1-f}(\Delta t_{i,j-1}, \Delta t_{ij}, x_i) \Big).
\end{aligned}
\tag{7.10}
$$

Summing over subjects, the total observed and expected counts of observed state triplets (d, e, f), misclassification predictor triplets (c_1, c_2, c_3) with covariate level l, are $n_{def,c_1c_2c_3,l}$ and

$$E(H_{def,c_1c_2c_3,l}) = \sum_{i=1}^{I} \delta_{x_i,l}\, E(H^{(i)}_{def,c_1c_2c_3}; x_i),$$

respectively.

Considering a gamma dead time in the model makes the likelihood function intractable and an alternative method to maximum likelihood estimation is required for parameter estimation. Providing a good fit to the observed counts is a necessary condition for an appropriate model and motivates parameter estimation by minimizing a discrepancy measure that compares the observed and expected counts. Although the non-independence of the transitions prevents use of a χ^2 distribution to assess goodness-of-fit, a discrepancy function of a Pearson type form can be minimized to estimate $\boldsymbol{\Theta} = (\alpha_{01}, \alpha_{10}, \beta_\rho, \beta_\eta, \kappa)$. Let $\mathcal{O}$ and $\mathcal{E}_{\boldsymbol{\Theta}}$ represent observed and expected counts with parameter vector $\boldsymbol{\Theta}$ for a generic model, respectively. Minimizing a measure of the form

$$\sum \frac{(\mathcal{O} - \mathcal{E}_{\boldsymbol{\Theta}})^2}{\mathcal{E}_{\boldsymbol{\Theta}}} \tag{7.11}$$

with respect to $\boldsymbol{\Theta}$ leads to a biased estimating function. A better estimating function results from modifying the discrepancy measure. Consider finding $\mathcal{E}_{\boldsymbol{\Theta}}{}^{(\ell)}$ by minimizing

$$\sum \frac{(\mathcal{O} - \mathcal{E}_{\boldsymbol{\Theta}})^2}{\mathcal{E}_{\boldsymbol{\Theta}}{}^{(\ell-1)}} \tag{7.12}$$

with respect to $\mathcal{E}_{\boldsymbol{\Theta}}$. Provided that each iteration leads to an improvement in the searching algorithm, this procedure will lead to solutions of an unbiased estimating function, and (7.12) will be easier to minimize than (7.11). However, if the expectations are costly to compute or if $\boldsymbol{\Theta}$ contains several parameters, an easier function to minimize is of the form

$$\sum \frac{(\mathcal{O} - \mathcal{E}_{\boldsymbol{\Theta}})^2}{\mathcal{O}}. \tag{7.13}$$

As will be discussed further in the next section, the approximation scheme implemented to calculate the equilibrium joint probabilities requires the choice of a discrepancy measure which is easily minimized. The significant computational advantage makes (7.13) the preferred discrepancy function for the models which follow.

The minimization of the Pearson χ^2 test statistic (7.11) for parameter estimation is generally referred to as minimum χ^2 methods (see Grizzle, Starmer and Koch (1969), for example). For multinomial situations, the χ^2 distribution can be used to assess goodness-of-fit as well as more complicated hypotheses. However, in the context of the model considered here, the counts will not follow a multinomial distribution and a χ^2 distribution cannot be used to assess the model or any hypotheses.

7.3.2 Approximating the Joint Probabilities

For the data sets considered here, observations are only available at discrete time points. Hence, interest focuses on the value of the probabilities of the previous section at these discrete time points. Numerical methods can be easily implemented to solve the convolution formulas and provide estimates of the probabilities at discrete times. Deriving recursion relations that approximate $u_{ab}(t)$ and $u_{abk}(t,\tau)$ at the discrete observation times is the approach proposed.

The formulas of the previous section are recast to contain cumulative distribution functions rather than the probability density functions. This change yields more accurate numerical results. Equations (7.6) and (7.7) can be re-expressed in terms of the cumulative probability distribution for each sojourn time,

$$F_{ab}(t) = \int_0^t f_{ab}(w)\,dw. \tag{7.14}$$

Applying integration by parts to $u_{ab}(t)$ and noting $u_{ab}(0) = \delta_{ab}$ yields

$$\begin{aligned} u_{ab}(t) &= \sum_{\ell\neq a}\left(F_{a\ell}(w)\,u_{\ell b}(t-w)\big|_{w=0}^{t} + \int_0^t F_{a\ell}(w)\,\dot{u}_{\ell b}(t-w)\,dw\right) \\ &\quad + \delta_{ab}\sum_{\ell\neq a}(1 - F_{a\ell}(t)) \\ &= \sum_{\ell\neq a}\left(F_{a\ell}(t)\,(\delta_{\ell b} - \delta_{ab}) + \int_0^t F_{a\ell}(w)\,\dot{u}_{\ell b}(t-w)\,dw + \delta_{ab}\right) \end{aligned} \tag{7.15}$$

where $\dot{u}_{ab}(t) = \frac{du_{ab}(t)}{dt}$. Similarly, with a change of variables and recalling the fact that $u_{abk}(0,\tau) = \delta_{ab}\,u_{ak}(\tau)$, the formula for $u_{abk}(t,\tau)$ can be rewritten as

$$\begin{aligned} u_{abk}(t,\tau) &= \sum_{\ell\neq a}\left(F_{a\ell}(w)\,u_{\ell bk}(t-w,\tau)\big|_{w=0}^{t} + \int_0^t F_{a\ell}(w)\,\dot{u}_{\ell bk}(t-w,\tau)\,dw\right) \\ &\quad + \delta_{ab}\sum_{\ell\neq a}\left(F_{a\ell}(w)\,u_{\ell k}(t+\tau-w)\big|_{w=t}^{t+\tau} + \int_t^{t+\tau} F_{a\ell}(w)\,\dot{u}_{\ell k}(t+\tau-w)\,dw\right) \\ &\quad + \delta_{abk}\sum_{\ell\neq a}(1 - F_{a\ell}(t+\tau)) \\ &= \sum_{\ell\neq a}\left(F_{a\ell}(t)\,u_{\ell k}(\tau)\,(\delta_{\ell b} - \delta_{ab}) + \int_0^t F_{a\ell}(w)\,\dot{u}_{\ell bk}(t-w,\tau)\,dw\right) \\ &\quad + \delta_{ab}\sum_{\ell\neq a}\left(\delta_{ak} + F_{a\ell}(t+\tau)\,(\delta_{\ell k} - \delta_{ak}) + \int_0^\tau F_{a\ell}(t+w)\,\dot{u}_{\ell k}(\tau-w)\,dw\right) \end{aligned} \tag{7.16}$$

where $\dot{u}_{abk}(t,\tau) = \frac{\partial u_{abk}(t,\tau)}{\partial t}$.

Consider the interval of time $[0, T]$ where T is the maximum time between observations for the entire data set. The interval can be partitioned into n_* segments, $t_0 = 0, t_1, \ldots, t_{n_*} = T$, each of equal length $\delta t = (T - t_0)/n_*$. A given point t_r corresponds to $r\delta t + t_0$, for some positive integer r. The integrals in (7.15) and (7.16) can be approximated by Riemann sums where the leftmost endpoint is evaluated. Such method gives a first order approximation in δt.

The derivatives $\dot{u}_{ab}(t)$ and $\dot{u}_{abk}(t, \tau)$ can be approximated by first order forward differences

$$\begin{aligned}
\dot{u}_{ab}(t) &\approx \frac{u_{ab}(t + \delta t) - u_{ab}(t)}{\delta t} \\
\dot{u}_{abk}(t, \tau) &\approx \frac{u_{abk}(t + \delta t, \tau) - u_{abk}(t, \tau)}{\delta t}
\end{aligned}$$

and hence, the discretized version can be written as

$$\begin{aligned}
\dot{u}_{ab}(t_r) &\approx \frac{u_{ab}(t_{r+1}) - u_{ab}(t_r)}{\delta t} \\
\dot{u}_{abk}(t_r, t_s) &\approx \frac{u_{abk}(t_{r+1}, t_s) - u_{abk}(t_r, t_s)}{\delta t}
\end{aligned}$$

where $t_r = t$ and $\tau = t_s$ for some integers $r, s > 0$. Since the t and τ represent days in the data sets considered here, each discrete observation time can be related to one of the $t_0, t_1, \ldots, t_{n_*}$ when the associated partition is sufficiently fine.

Approximating the integrals in (7.15) with Riemann sums gives

$$\begin{aligned}
u_{ab}(t_r) &\approx \sum_{\ell\neq a}\left(\delta_{ab} + F_{a\ell}(t_r)(\delta_{\ell b} - \delta_{ab}) + \sum_{m=0}^{r-1} F_{a\ell}(t_m)\,\dot{u}_{\ell b}(t_r - t_m)\,\delta t\right) \\
&\approx \sum_{\ell\neq a}\Bigg(\delta_{ab}\,(1 - F_{a\ell}(t_r)) + \delta_{\ell b}\,(F_{a\ell}(t_r) - F_{a\ell}(t_{r-1})) \\
&\qquad + \sum_{m=0}^{r-2}(F_{a\ell}(t_{m+1}) - F_{a\ell}(t_m))\,u_{\ell b}(t_{r-1-m})\Bigg) \qquad r \geq 1
\end{aligned} \tag{7.17}$$

since $t_\ell - t_m = t_{\ell-m}$ and $F_{ab}(t_0) = F_{ab}(0) = 0$. It can be easily shown that (7.17) satisfies the initial conditions and constraints identified. Note that choosing to approximate the derivatives with the forward difference allows $u_{ab}(t_r)$ to depend only on $\{u_{\ell b}(t_s)\}_{\ell\neq a}$ for $s < r$.

Following the same discretization scheme, an approximation to $u_{abk}(t,\tau)$ is

$$\begin{aligned}
u_{abk}(t_r, t_s) &\approx \sum_{\ell\neq a}\left(F_{a\ell}(t_r)\,u_{\ell k}(t_s)\,(\delta_{\ell b} - \delta_{ab}) + \sum_{m=0}^{r-1} F_{a\ell}(t_m)\,\dot{u}_{\ell bk}(t_{r-m}, t_s)\,\delta t\right) \\
&\quad + \delta_{ab}\sum_{\ell\neq a}\left(\delta_{ak} + F_{a\ell}(t_{r+s})\,(\delta_{\ell k} - \delta_{ak}) + \sum_{m=0}^{s-1} F_{a\ell}(t_{r+m})\,\dot{u}_{\ell k}(t_{s-m})\,\delta t\right) \\
&\approx \sum_{\ell\neq a}\Bigg(\delta_{\ell b}\,(F_{a\ell}(t_r) - F_{a\ell}(t_{r-1}))\,u_{\ell k}(t_s) \\
&\qquad + \delta_{ab}(\delta_{ak}\,(1 - F_{a\ell}(t_{r+s})) + \delta_{\ell k}\,(F_{a\ell}(t_{r+s}) - F_{a\ell}(t_{r+s-1}))) \\
&\qquad + \sum_{m=0}^{r-2}(F_{a\ell}(t_{m+1}) - F_{a\ell}(t_m))\,u_{\ell bk}(t_{r-1-m}, t_s) \\
&\qquad + \delta_{ab}\sum_{m=0}^{s-2}(F_{a\ell}(t_{r+m+1}) - F_{a\ell}(t_{r+m}))\,u_{\ell k}(t_{s-1-m})\Bigg)
\end{aligned}$$

for $r \geq 1$, $s \geq 1$. All initial conditions and constraints are also satisfied by this approximation.

The values of equilibrium joint probabilities can be approximated in the same manner as $u_{ab}(t)$ and $u_{abk}(t,\tau)$. The approximations for (7.8) are

$$\begin{aligned}
\pi_{ab}(t_r) \approx \pi_a \sum_{\ell \neq a} \Bigg(& \delta_{ab}\,(1 - \Phi_{a\ell}(t_r)) + \delta_{\ell b}\,(\Phi_{a\ell}(t_r) - \Phi_{a\ell}(t_{r-1})) \\
& + \sum_{m=0}^{r-2} (\Phi_{a\ell}(t_{m+1}) - \Phi_{a\ell}(t_m))\, u_{\ell b}(t_{r-1-m}) \Bigg) \\
\pi_{abk}(t_r, t_s) \approx \pi_a \sum_{\ell \neq a} \Bigg(& \delta_{\ell b}\,(\Phi_{a\ell}(t_r) - \Phi_{a\ell}(t_{r-1}))\, u_{\ell k}(t_s) \\
& + \delta_{ab}\big(\delta_{ak}\,(1 - \Phi_{a\ell}(t_{r+s})) + \delta_{\ell k}\,(\Phi_{a\ell}(t_{r+s}) - \Phi_{a\ell}(t_{r+s-1}))\big) \\
& + \sum_{m=0}^{r-2} (\Phi_{a\ell}(t_{m+1}) - \Phi_{a\ell}(t_m))\, u_{\ell bk}(t_{r-1-m}, t_s) \\
& + \delta_{ab} \sum_{m=0}^{s-2} (\Phi_{a\ell}(t_{r+m+1}) - \Phi_{a\ell}(t_{r+m}))\, u_{\ell k}(t_{s-1-m}) \Bigg).
\end{aligned}$$

where

$$\Phi_{ab}(t) = \int_0^t \phi_{ab}(w)\, dw.$$

In the two state case considered here, the value of $\Phi_{01}(t_r)$ can be easily calculated since $\Phi_{01}(t_r) = F_{01}(t_r)$. Conversely, numerical integration methods are required to evaluate $\Phi_{10}(t_r)$.

These approximated equilibrium joint probabilities can be used in (7.13) with appropriate stratification by misclassification predictors and covariates for parameter estimation. The smaller the value of δt, the better the approximation. However, the recursion relations take more time to calculate for very small values of δt. This is particularly so for the calculation of $u_{abk}(t_r, t_s)$. Generally, $n_\star = 70$ provides accuracy of about two significant digits in a reasonable amount of time. With the estimation technique used here, this is a reasonable approximation since very small changes in the parameters do not dramatically change the measure of discrepancy between observed and expected counts. As with the likelihood models, calculations are provided by the author's C program using the linear algebra package Stuber (1996). The minimization algorithm is started at several different starting conditions and is stopped when the discrepancy measure fails to decrease by more than 10^{-6} on one iteration.

7.3.3 Parasitic Infection Fitted Model

The results of Chapter 6 suggested that an underlying Markov process was not suitable for the parasitic infection data set. Table 6.5 indicated that the observed transition triplets were not modeled well and suggested a change to the model is necessary. These findings motivate a semi-Markov model based on the equilibrium triplet probabilities for the PI data set.

The parameter estimates under a semi-Markov model with exponential open time distribution and gamma dead time distribution are given in Table 7.1. The estimates are obtained by minimizing

$$D = \sum_{d=0}^{1} \sum_{e=0}^{1} \sum_{f=0}^{1} \frac{(n_{def} - E(H_{def}))^2}{n_{def}} \tag{7.18}$$

where n_{def} and $E(H_{def})$ are the total observed and expected counts of parasitic

Table 7.1: Observed and expected transition triplet counts for the parasitic infection data under a semi-Markov model with exponential open time and gamma dead time. Drop-5 jackknife standard error estimates. ($\delta t = 0.1$, $n_* = 70$)

Parameter	Estimate (se)	Correlations			Term	Estimate	EMT (days)
$\beta_{\rho 0}$	-0.832 (0.686)				ρ	0.435	State 0 2.30
$\beta_{\eta 0}$	-4.515 (0.334)	0.226			η	0.011	State 1 2.26
$\ln(\kappa)$	-3.698 (0.703)	−0.967	−0.042		κ	0.025	
α_0^*	-2.586 (0.364)	−0.151	−0.596	−0.025	v_{10}	0.070	

	Observed and Expected Triplet Counts							
	000	001	010	100	011	101	110	111
Observed	285	77	43	67	66	35	61	238
Expected	285.190	74.552	44.046	66.252	65.771	33.050	65.456	237.683
							D:	0.547
	Equilibrium Joint Probability Estimates							
	000	001	010	100	011	101	110	111
$\pi_{def}(7,7)$	0.312	0.084	0.044	0.074	0.064	0.020	0.063	0.339
π_1	0.497							

	Observed and Expected Pair Counts under Triplet Estimates			
	00	01	10	11
Observed	378	119	112	321
Expected	383.670	117.121	105.907	323.302
	Equilibrium Joint Pair Probability Estimates			
	00	01	10	11
$\pi_{de}(7)$	0.397	0.107	0.094	0.402

detection triplets (d, e, f), $d, e, f \in \{0, 1\}$. The joint equilibrium pair and triplet probabilities are calculated based on the parameter estimates minimizing (7.18). The observed and expected pair and triplet counts are also given under the parameter estimates.

This estimation technique provides a model which more closely fits the observed triplet data than the likelihood approach. In particular, the 000 and 111 expected counts are close to the observed counts. There does not appear to be an over-representation of 000 runs with this model. The equilibrium probability estimates appear reasonable since the observed data suggests that subjects are almost equally likely to be in each state. The subjects are more likely to be in the same state for three consecutive observation times. The misclassification probability is slightly lower than the maximum likelihood estimate of 0.09. The real difference between the model specifications is seen in the ρ, η and κ estimates. The shape estimate is much smaller than the shape of 1 equivalent to an exponential distribution. Under the semi-Markov approach, the estimate of ρ is about 10 times larger and the estimate of η is about 33% of the corresponding estimates under the likelihood approach with exponential sojourn times.

In Tables 7.2 and 7.3, the estimated model in Table 7.1 is compared with models where the shape is specified and the remaining parameters are estimated. For the specified shape models, the estimated mean times increase as the shape increases. The estimate of v_{10} is fairly stable for all models and appears to be in the 7-12% range. The Markov model, $\kappa = 1$, gives estimates for ρ and η close to the maximum likelihood estimates in Table 3.6 on page 52. When the shape is allowed to

Table 7.2: Estimates for all other parameters when shape is specified as 0.5 ,1 ,2 or 5 along with estimates from Table 7.1 where $\hat{\kappa} = 0.025$. The Markov model corresponds to $\kappa = 1$ and estimated mean times (EMT) given in days.

	Estimates by Shape				
	$\hat{\kappa} = 0.025$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$	$\kappa = 5$
State 0: Uninfected					
$\beta_{\rho 0}$	-0.832	-3.058	-3.339	-3.542	-3.671
ρ	0.434	0.047	0.035	0.029	0.025
EMT	2.296	21.293	28.194	34.536	39.291
State 1: Infected					
$\beta_{\eta 0}$	-4.515	-3.757	-3.370	-2.944	-2.221
η	0.011	0.023	0.034	0.053	0.108
EMT	2.262	21.412	29.071	37.973	46.092
Misclassification					
α_0^*	-2.586	-2.662	-2.498	-2.173	-1.940
v_{10}	0.070	0.065	0.076	0.102	0.126
Statistic					
D	0.547	7.240	12.169	23.771	30.460

Table 7.3: Expected triplet counts for the PI models in Table 7.2.

		Expected Triplet Counts				
Triplet	Observed	$\hat{\kappa} = 0.025$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$	$\kappa = 5$
000	285	285.219	286.883	289.636	294.306	297.856
001	77	74.551	72.210	70.173	65.515	60.906
010	43	44.046	31.403	26.016	21.152	19.388
100	67	66.251	72.072	70.177	65.576	60.975
011	66	65.767	72.823	77.139	81.021	82.345
101	35	33.047	31.976	32.982	36.668	40.830
110	61	65.453	72.822	77.142	81.050	82.402
111	238	237.666	231.807	228.734	226.713	227.297
D	-	0.547	7.240	14.169	23.770	30.460

be estimated, the estimated mean times are small and less than the time between observations. Further examination of the expected counts for these models in Table 7.3 shows why the estimated shape is small. Generally, the expected counts are fairly similar for the models where the shape is 0.025, 0.5 or 1. The largest difference in these models appears to be the 010 frequency. This frequency is fit much better when the shape is 0.025 than when the shape is 1 or higher. This sequence seems to be influential in the estimation of the shape parameter. While this model does fit better than the Markov model, knowledge of the distribution of the statistic D is necessary to assess if this model fits significantly better than the Markov model. Simulation methods could be used to investigate the distribution and properties of D.

Using the estimated mean times in Table 7.1, simulations were conducted to see if the run length distributions of the observed PI data set could arise from the model estimates. Exponential sojourn times were generated with a mean of 2.3 days for state 0. For state 1, gamma sojourn times were generated with a mean of 2.26 days and shape 0.025. The true states were determined by comparing the observation times of the PI data set with the alternating time intervals generated for each subject. These simulated true data sets were then misclassified according to the misclassification probability settings $v_{10} = 0.07$, $v_{10} = 0.09$ and $v_{10} = 0.15$ in a similar manner as described in Section 5.4. Note that v_{01} was set to be 0 in all simulations.

One hundred true data sets were generated and misclassified 20 times for each misclassification probability setting in S-PLUS (MathSoft, Inc. (1996)). Sample

Table 7.4: Sample run length distributions for simulations† which are close to the PI data set. Data is generated based on the estimated state mean times and shape estimate in Table 7.1 with misclassification probabilities $v_{01} = 0.07$, $v_{01} = 0.09$ or $v_{01} = 0.15$. The means are given in days.

	Run Length Distribution of 0s																						
Sim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Mean
S1	52	21	14	16	16	4	10	3	4	9	2	3	1	0	2	0	1	1	0	0	0	0	29.06
S2	53	23	15	15	11	8	6	4	4	3	4	3	3	1	2	1	1	0	0	0	0	0	28.67
S3	53	24	14	11	11	9	7	8	2	5	6	1	2	2	1	0	1	0	0	0	0	1	29.37
S4	53	28	15	9	15	9	8	7	1	7	4	1	2	1	0	0	0	0	0	0	1	0	27.26
S5	56	26	14	11	15	8	8	3	1	4	2	1	2	0	1	1	1	0	1	1	0	0	26.61
S6	51	26	19	9	11	9	8	8	0	8	4	1	2	1	0	0	1	0	0	0	0	0	27.38
S7	50	25	17	17	11	10	7	1	1	4	3	0	2	0	1	1	1	0	1	1	0	0	26.81
PI	52	26	14	12	12	5	2	7	3	3	1	5	2	0	1	0	0	0	0	0	0	0	25.25

	Run Length Distribution of 1s																						
Sim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Mean
S1	61	30	11	6	8	2	3	3	3	0	1	0	0	0	1	0	0	0	0	0	0	0	17.80
S2	62	29	13	7	8	3	0	3	5	2	1	0	0	0	0	0	0	0	0	0	0	0	18.16
S3	63	28	11	12	5	2	2	2	5	0	0	0	1	0	0	0	0	0	0	0	0	0	17.37
S4	62	27	12	11	6	2	6	2	2	2	1	0	0	0	0	1	0	0	0	0	0	0	18.86
S5	57	28	12	14	4	10	3	1	2	1	0	2	1	0	0	1	0	0	0	0	0	0	20.33
S6	59	29	13	8	6	4	4	4	2	2	1	0	0	0	0	0	0	0	1	0	0	0	19.47
S7	56	27	11	13	5	3	5	0	3	2	1	1	2	1	0	0	0	0	1	0	0	0	21.48
PI	61	28	10	8	10	4	4	8	5	2	1	1	1	0	0	0	0	0	1	0	0	0	22.60

†S1, S2, S3: generated with $v_{10} = 0.15$
S4, S5: generated with $v_{10} = 0.09$
S6, S7: generated with $v_{10} = 0.07$

run length distributions for simulated data sets appear in Table 7.4. These results suggest that with the misclassification probability v_{10} in the 7-15% range, data sets can be generated with estimated mean times of 2.3 and 2.26 days for states 0 and 1, respectively, which have similar run length distributions as the PI data set. In particular, the runs of lengths 1, 2 or 3 for each of the states agree well for the simulated data sets and the PI data set. The estimates in Table 7.4 appear to be plausible in the sense that appropriate run length distributions could arise from such a model. Note that other shapes were also considered with the same state 1 mean time. Simulations with shapes 0.5, 1 and 2, conducted in the same manner as described above for $\kappa = 0.025$ failed to yield similar run length distributions as the PI data set.

7.3.4 Multiple Sclerosis Fitted Models

Unlike the PI data set, the likelihood approach for an underlying Markov model for the MS data set provided good fits to both the observed pair and triplet data despite the extreme parameter estimates. Attention is now restricted to treatment (P=placebo, L=low dose, H=high dose) models where estimation is conducted by use of the equilibrium triplet probabilities.

The full model involves parameters for each treatment level and exacerbation status which can be expressed as follows:

$$\begin{aligned}
\text{M1:} \quad & \rho(x_{Li}, x_{Hi}) = \exp(\beta_{\rho P} + \beta_{\rho L} x_{Li} + \beta_{\rho H} x_{Hi}) \\
& \eta(x_{Li}, x_{Hi}) = \exp(\beta_{\eta P} + \beta_{\eta L} x_{Li} + \beta_{\eta H} x_{Hi}), \quad \text{shape } \kappa \\
& v_{01}(c_{ij}) = \frac{e^{\alpha_0 + \alpha_1 c_{ij}}}{1 + e^{\alpha_0 + \alpha_1 c_{ij}}} \qquad v_{10}(c_{ij}) = \frac{e^{\alpha_0^* + \alpha_1^* c_{ij}}}{1 + e^{\alpha_0^* + \alpha_1^* c_{ij}}} \\
& i = 1, \ldots, 38, \; j = 1, \ldots, 17
\end{aligned}$$

where x_{Li} and x_{Hi} are indicator variables of low or high dose treatment group membership for subject i, respectively, and c_{ij} is the observed exacerbation status for subject i at visit j. In addition, some of the parameters are constrained. The shape is restricted to lie in the interval [0.01,7.00] and transitions are restricted to occur no more than once per day and be at least as frequent as once every three years. These constraints are imposed merely to facilitate timely minimization. Preliminary investigations indicated that more extreme values result in taking much time to calculate the numerical integrals and many more function evaluations for minimization yielding much larger minima than those found with other starting conditions. No constraints are applied to the misclassification probabilities other than they must lie in [0,1].

Since the exacerbation status is included in the model, the triplet counts can be stratified by both treatment level and exacerbation status triplet. However, not all cells in the table have at least one count. To alleviate division by zero, the discrepancy function is slightly altered. Consider the function

$$D_2 = \sum_{d,e,f} \sum_{c_1,c_2,c_3} \sum_{l} \frac{(n_{def,c_1c_2c_3,l} - E(H_{def,c_1c_2c_3,l}))^2}{n_{def,c_1c_2c_3,l} + 1} \tag{7.19}$$

which de-emphasizes disagreement between observed and expected counts for cells with few or no counts observed. Other alterations could be made, although the effects of different adjustments have not been definitively investigated (Grizzle, Starmer and Koch (1969)). The observed and expected counts of active scan triplets (d, e, f) and exacerbation status triplets (c_1, c_2, c_3) for treatment level l are denoted by $n_{def,c_1c_2c_3,l}$ and $E(H_{def,c_1c_2c_3,l})$, respectively, $d, e, f, c_1, c_2, c_3 \in \{0, 1\}$, $l \in \{P, L, H\}$. The parameter estimates appear in Table 7.5 and the fits appear in Tables 7.6 and 7.7. The estimates for the equilibrium joint probabilities $\pi_{def}(t, \tau)$ for $t = \tau = 42/28$ also appear in Table 7.5 along with the resulting equilibrium joint pair probabilities $\pi_{de}(t)$, where $\pi_{de}(t) = \delta_{ef}\pi_{def}(t, 0)$.

The shape estimate is small and near the boundary of acceptable shapes. The exacerbation terms appear to be insignificant and lead to the consideration of a model where misclassification probabilities are independent of exacerbation status. The $v_{01}(\cdot)$ probabilities are small while the $v_{10}(\cdot)$ probabilities are near 0.5. These values are consistent with the maximum likelihood estimates presented in Section 3.3.1. However, the estimated mean times under a semi-Markov model appear to be more reasonable for the different treatment groups. The expected counts for the active scan triplet 000 and the exacerbation status triplet 000 do not closely agree for the placebo and high dose subjects. The fit for the triplet data is similar to the fit given in Table 6.7 under the maximum likelihood estimates.

Table 7.5: Model M1 estimates found by minimizing D_2 for the MS data set with exponential open time and gamma dead time distributions with exacerbation status. Drop-1 jackknife standard error estimates, triplet and pair equilibrium probabilities for 42 day intervals. ($\delta t = 0.042$, $n_* = 58$, 28-day timescale)

Parameter	Estimate	(se)	Term	Estimate	EMT (days)
State 0: Remitting					
$\beta_{\rho P}$	−1.510	(0.388)	$\rho(0,0)$	0.221	Placebo 126
$\beta_{\rho L}$	2.444	(0.192)	$\rho(1,0)$	2.543	Low 11
$\beta_{\rho H}$	1.052	(1.289)	$\rho(0,1)$	0.632	High 44
State 1: Relapsing					
$\beta_{\eta P}$	−6.506	(0.237)	$\eta(0,0)$	0.001	Placebo 188
$\beta_{\eta L}$	4.413	(0.299)	$\eta(1,0)$	0.123	Low 2
$\beta_{\eta H}$	3.265	(1.169)	$\eta(0,1)$	0.039	High 7
$\ln(\kappa)$	−4.600	($\approx 10^{-10}$)	κ	0.010	
Misclassification					
α_0	−2.935	(0.495)	$v_{01}(0)$	0.050	
α_1	−0.853	(0.991)	$v_{01}(1)$	0.022	
α_0^*	−0.211	(0.122)	$v_{10}(0)$	0.447	
α_1^*	0.229	(0.246)	$v_{10}(1)$	0.504	

Equilibrium Joint Probability Estimates ($\pi_{def}(1.5, 1.5)$)								
	$\pi_{000}()$	$\pi_{001}()$	$\pi_{010}()$	$\pi_{100}()$	$\pi_{011}()$	$\pi_{101}()$	$\pi_{110}()$	$\pi_{111}()$
P	0.380	0.011	0.005	0.008	0.007	0.000	0.007	0.583
L	0.594	0.107	0.086	0.051	0.041	0.009	0.034	0.078
H	0.771	0.043	0.029	0.024	0.017	0.001	0.016	0.098
Equilibrium Joint Pair Probability Estimates								
	$\pi_{00}(1.5)$	$\pi_{01}(1.5)$	$\pi_{10}(1.5)$	$\pi_{11}(1.5)$	π_1			
P	0.391	0.012	0.008	0.589	0.598			
L	0.702	0.127	0.060	0.112	0.172			
H	0.815	0.046	0.025	0.114	0.140			

Table 7.6: Active scan observed and expected transition pair counts by exacerbation status and treatment level for model M1. Expected counts in parentheses.

Exacerbation Status		Scan Activity			
		00	01	10	11
P	00	66 (61.285)	22 (21.848)	17 (21.633)	23 (23.235)
	01	11 (15.655)	5 (4.479)	9 (5.842)	6 (5.024)
	10	14 (15.150)	6 (5.701)	4 (4.284)	6 (4.865)
	11	2 (1.600)	0 (0.485)	0 (0.480)	1 (0.435)
L	00	92 (96.934)	18 (17.031)	15 (12.614)	7 (5.421)
	01	22 (26.722)	5 (3.496)	6 (3.559)	2 (1.223)
	10	26 (26.643)	3 (4.740)	3 (2.400)	3 (1.217)
	11	5 (4.748)	0 (0.632)	1 (0.443)	0 (0.177)
H	00	111 (118.919)	16 (13.875)	18 (12.329)	6 (5.877)
	01	24 (22.010)	2 (1.735)	1 (2.348)	0 (0.907)
	10	24 (21.986)	1 (2.632)	1 (1.482)	1 (0.900)
	11	3 (2.530)	0 (0.206)	0 (0.177)	0 (0.088)

Next consider a model with misclassification probabilities which are independent of exacerbation status. Model M2 can be specified as

$$
\begin{aligned}
\text{M2:}\quad & \rho(x_{Li}, x_{Hi}) = \exp(\beta_{\rho P} + \beta_{\rho L} x_{Li} + \beta_{\rho H} x_{Hi}) \\
& \eta(x_{Li}, x_{Hi}) = \exp(\beta_{\eta P} + \beta_{\eta L} x_{Li} + \beta_{\eta H} x_{Hi}), \quad \text{shape } \kappa \\
& v_{01} = \frac{e^{\alpha_0}}{1 + e^{\alpha_0}} \qquad v_{10} = \frac{e^{\alpha_0^*}}{1 + e^{\alpha_0^*}} \\
& i = 1, \ldots, 38,\ j = 1, \ldots, 17.
\end{aligned}
$$

The discrepancy function used for estimation is

$$
D_3 = \sum_{d,e,f} \sum_{l} \frac{(n_{def,l} - E(H_{def,l}))^2}{n_{def,l}} \tag{7.20}
$$

Table 7.7: Active scan observed and expected transition triplet counts by exacerbation status and treatment level for model M1. Expected counts in parentheses. ($D_2 = 62.842$)

Exacerbation		Scan Activity							
	Status	000	001	010	100	011	101	110	111
P	000	47(39.730)	9 (9.106)	8 (8.945)	7(8.949)	8(8.465)	7(8.290)	5(8.454)	11(10.061)
	001	4 (8.134)	3 (1.443)	2 (1.937)	1(1.933)	2(1.476)	1(1.446)	5(1.866)	2 (1.766)
	010	6(10.165)	2 (2.461)	2 (1.764)	5(2.424)	1(1.847)	4(2.287)	1(1.843)	4 (2.209)
	100	7 (8.134)	3 (1.967)	1 (1.936)	2(1.409)	4(1.864)	1(1.445)	2(1.477)	0 (1.767)
	011	2 (1.275)	0 (0.240)	0 (0.236)	0(0.321)	0(0.197)	0(0.244)	0(0.249)	1 (0.237)
	101	2 (2.976)	1 (0.559)	0 (0.749)	1(0.548)	1(0.581)	0(0.451)	1(0.581)	1 (0.554)
	110	2 (1.275)	0 (0.325)	0 (0.235)	0(0.236)	0(0.250)	0(0.244)	1(0.197)	0 (0.237)
L	000	60(64.281)	10(10.625)	12(10.142)	10(7.713)	2(3.015)	4(2.030)	1(2.683)	3 (1.510)
	001	16(15.060)	2 (1.835)	3 (2.403)	0(1.816)	0(0.560)	0(0.375)	1(0.655)	1 (0.296)
	010	14(18.329)	3 (3.050)	2 (2.103)	5(2.238)	2(0.693)	0(0.607)	1(0.613)	1 (0.367)
	100	15(14.370)	3 (2.381)	1 (2.283)	0(1.156)	2(0.692)	1(0.346)	0(0.482)	0 (0.290)
	011	3 (3.397)	0 (0.420)	1 (0.397)	1(0.421)	0(0.103)	0(0.088)	0(0.118)	0 (0.056)
	101	4 (6.102)	1 (0.747)	0 (0.985)	0(0.504)	0(0.236)	1(0.117)	2(0.209)	1 (0.101)
	110	5 (4.068)	0 (0.680)	0 (0.472)	1(0.341)	0(0.160)	0(0.103)	0(0.109)	0 (0.068)
H	000	78(88.601)	11 (9.062)	13 (8.626)	11(7.898)	2(2.760)	4(2.218)	3(2.686)	2 (2.149)
	001	15(14.033)	1 (0.929)	0 (1.392)	2(1.272)	0(0.355)	0(0.280)	1(0.448)	0 (0.291)
	010	19(15.510)	1 (1.611)	1 (0.958)	0(1.407)	0(0.389)	0(0.416)	0(0.383)	0 (0.325)
	100	15(13.280)	0 (1.378)	0 (1.318)	0(0.727)	1(0.436)	1(0.261)	0(0.326)	1 (0.275)
	011	2 (2.290)	0 (0.155)	0 (0.145)	1(0.212)	0(0.048)	0(0.049)	0(0.059)	0 (0.041)
	101	6 (5.341)	1 (0.360)	0 (0.539)	0(0.298)	0(0.143)	0(0.086)	0(0.139)	0 (0.095)
	110	3 (2.289)	0 (0.240)	0 (0.144)	0(0.128)	0(0.061)	0(0.049)	0(0.046)	0 (0.041)

where $n_{def,l}$ and $E(H_{def,l})$ are the observed and expected counts of active scan triplets (d, e, f) for treatment level l, respectively, $d, e, f \in \{0, 1\}$, $l \in \{P, L, H\}$. Since exacerbation status does not enter the model, (7.20) does not depend on this variable. The parameter estimates appear in Table 7.8 and the fits appear in Table 7.9.

Table 7.8: Model M2 estimates found by minimizing D_3 for the MS data set with exponential open time and gamma dead time distributions. Drop-1 jackknife standard error estimates, triplet and pair equilibrium probabilities for 42 day intervals. ($\delta t = 0.042$, $n_* = 58$, 28-day timescale)

Parameter	Estimate (se)	Term	Estimate	EMT (days)	
State 0: Remitting					
$\beta_{\rho P}$	−2.252 (0.470)	$\rho(0,0)$	0.105	Placebo	266
$\beta_{\rho L}$	−0.661 (3.940)	$\rho(1,0)$	0.054	Low	516
$\beta_{\rho H}$	−1.414 (0.376)	$\rho(0,1)$	0.026	High	1095
State 1: Relapsing					
$\beta_{\eta P}$	−2.440 (5.978)	$\eta(0,0)$	0.087	Placebo	236
$\beta_{\eta L}$	0.812 (4.307)	$\eta(1,0)$	0.196	Low	105
$\beta_{\eta H}$	0.815 (0.344)	$\eta(0,1)$	0.197	High	104
$\ln(\kappa)$	−0.310 (6.244)	κ	0.733		
Misclassification					
α_0	−2.496 (0.151)	v_{01}	0.076		
α_0^*	−0.710 (0.184)	v_{10}	0.330		

	Equilibrium Joint Probability Estimates ($\pi_{def}(1.5, 1.5)$)							
	$\pi_{000}()$	$\pi_{001}()$	$\pi_{010}()$	$\pi_{100}()$	$\pi_{011}()$	$\pi_{101}()$	$\pi_{110}()$	$\pi_{111}()$
P	0.404	0.059	0.012	0.059	0.055	0.009	0.055	0.347
L	0.734	0.047	0.017	0.047	0.034	0.003	0.033	0.085
H	0.861	0.026	0.009	0.026	0.018	0.001	0.017	0.043
	Equilibrium Joint Pair Probability Estimates							
	$\pi_{00}(1.5)$	$\pi_{01}(1.5)$	$\pi_{10}(1.5)$	$\pi_{11}(1.5)$	π_1			
P	0.463	0.067	0.067	0.402	0.469			
L	0.781	0.050	0.050	0.119	0.169			
H	0.887	0.026	0.026	0.061	0.087			

Table 7.9: Active scan observed and expected transition triplet and pair counts by treatment group for model M2. Expected counts in parentheses.

	Scan Activity Triplets							
	000	001	010	100	011	101	110	111
P	70	18	13	16	16	13	15	19
	(68.366)	(18.071)	(15.077)	(18.030)	(14.573)	(11.609)	(14.601)	(19.672)
L	117	19	19	17	6	6	5	6
	(120.927)	(17.580)	(15.449)	(17.512)	(6.662)	(4.562)	(6.678)	(5.629)
H	138	14	14	14	3	5	4	3
	(136.693)	(15.301)	(14.136)	(15.264)	(3.949)	(2.799)	(3.955)	(2.904)
							D_3:	4.354
Sum over treatment level:								
	325	51	46	47	25	24	24	28
	(325.986)	(50.952)	(44.662)	(50.806)	(25.184)	(18.970)	(25.234)	(28.205)

	Scan Activity Pairs			
	00	01	10	11
P	93	33	30	36
	(92.162)	(31.665)	(31.653)	(36.520)
L	145	26	25	12
	(147.717)	(23.608)	(23.569)	(13.105)
H	162	19	20	7
	(162.107)	(19.310)	(19.287)	(7.297)

Subjects receiving the low or high doses of the drug appear to spend more time in the remitting state and less time in the relapsing state than the subjects receiving the placebo. The estimates for η are very similar for the low and high levels of the drug. The time spent remitting for high dose patients is very near the boundary of at least one transition every 3 years. The jackknife standard error estimate for $\beta_{\rho L}$ is quite large compared to the parameter estimate.

Upon examination of the estimates when one subject is dropped at a time, the estimates for $\beta_{\rho L}$ are around 2 when any of the placebo or high dose subjects are removed from the data set. Conversely, when any one of the low dose subjects is dropped, the estimates for $\beta_{\rho L}$ are around 0.77. The same situation is seen for the shape parameters estimates. If a low dose subject is dropped, the estimate for $\ln(\kappa)$ is around -2.4 while any other subject dropped yields an estimate near -4.6. Clearly, dropping any one of the low dose subjects has a large effect on these two parameters estimates and this effect leads to the relatively large jackknife standard errors. The estimate of v_{01} is fairly consistent with earlier results while the estimate of v_{10} is further away from 0.5 than with previous models. The observed and expected triplet and pair counts appearing in Table 7.9 indicate that the model provides a good fit to the data.

Note that probabilities for these models take a considerable time to compute. With $n_\star = 58$, each function evaluation requires calculation of a 58×58 recursion relation for each treatment level. With the additional computational burden of dropping 1 subject at a time for jackknife variance estimation, a model such as Model M1 with several parameters can take as much as 30 hours of computing time on a dedicated Pentium II 233 MHz computer. Several days can be spent trying to prune model parameters based on the results of larger models and the jackknife variance estimates.

Drop-5 jackknife variance estimates calculated for the parasitic infection data set are preferred to the drop-1 estimates presented for the Multiple Sclerosis data set. Certainly, drop-5 jackknife estimates reduce the computational burden significantly and were tried for the MS data set as well as drop-3 jackknife. In both of these cases, the removal of the groups of subjects resulted in large variance estimates. Models which resulted in a large reduction of the discrepancy measure did not have significant parameters. Even for models with only intercept parameters and misclassification absent, the intercept parameters were not significant. Under a drop-1 jackknife, the variance estimates were generally reasonable in the sense that an additional parameter which reduced the the discrepancy measure by a large amount had a relatively small variance estimate. Based on these observations, drop-1 jackknife variance estimation was implemented despite the increased computational burden.

Computational burden can be considerably reduced if parameter estimation is based on pairs rather than triplets. In this case, only a one-dimensional recursion relation is calculated for each treatment level at each function evaluation. When a model analogous to model M1 is fit with the pair data, the exacerbation status terms can also be dropped from the model. However, dropping such terms reduces the degrees of freedom to 6: 2 degrees of freedom for each treatment level. The pair counts should not be stratified by exacerbation status when this variable is not included in the model. Even in the model without misclassification, the number of parameters requiring estimation for the 3 treatment levels total 7. Hence, the active scan pair data cannot be used to estimate all the parameters necessary in the models considered here.

7.3.5 Multiple Sclerosis Movers Data Set

The results of the last section suggest that the shape parameter may be less than one. Gleser (1989) showed that a gamma distribution with shape parameter less than 1 can be represented as a scale mixture of exponential distributions. Hence, the estimated shapes may be explained by subject heterogeneity in the data set. Different models indicated different groups of heterogeneous subjects indicating no clear subset of the data. A natural choice is to subset the data set into "movers" and "stayers" and to look at plausible values for the movers. While other sections have considered models where the misclassification probabilities are specified and the regression parameters are estimated, this section takes the opposite approach. Misclassification probabilities values are considered based on plausible regression coefficients for the different treatment groups.

If the data are assumed to be classified correctly, only the subjects experiencing a change of state at at least one observation time can contribute information about plausible state sojourn times. For the 8 subjects whose scans are never active, any sojourn time longer than the study duration is plausible. Assuming the observation times are the transition times, mean sojourn times for the remaining 30 subjects can be used to determine plausible values. Dropping subject #s 451 (P), 424, 450

(L), 526, 541, 543 and 566 (H) yields 30 subjects termed the MS Movers data set. The results that follow are all obtained using this reduced data set.

Table 7.10 gives plausible sojourn times for each state and treatment group. These values are simply calculated by averaging the observed sojourn times for all subjects under the assumption that observation times are the transition times and ignoring the censoring of the sojourn time of the last state. With a 28-day time-scale, the results in Table 7.10 lead to regression coefficients

Table 7.10: Plausible sojourn times in days by state and treatment group for MS Movers data set.

Treatment	Remitting (0)	Relapsing (1)
Placebo (P)	168.435	88.211
Low Dose (L)	220.494	61.003
High Dose (H)	259.038	62.375

$$
\begin{aligned}
\rho(x_{Li}, x_{Hi}) &= \exp(-1.794 - 0.270\, x_{Li} - 0.431\, x_{Hi}) \\
\eta(x_{Li}, x_{Hi}) &= \exp(-1.148 + \ln(\kappa) + 0.369\, x_{Li} + 0.347\, x_{Hi})
\end{aligned}
\tag{7.21}
$$

where κ is the shape of the gamma distribution for the state 1 sojourn time.

To investigate the plausibility of these sojourn time values, simulations were conducted for different shape and misclassification probability settings. Gamma sojourn times were generated for each subject according to the model terms in (7.21) to determine subject true process histories. Using the observation times, the histories were converted to true states. The simulated true data sets were then misclassified according to the misclassification probability settings in similar manner as described in Section 5.4. Here, misclassification probabilities were allowed to be independent or dependent on the exacerbation status for each subject. Table 7.11 lists the misclassification probability settings considered. One hundred true data sets were generated and misclassified 20 times for each misclassification probability setting V1, V3, V5 and V7. Similarly, another 100 true data sets were generated and misclassified 20 times for each setting V2, V4, V6 and V8. This process was repeated for each of 3 shapes: 0.5, 1 and 2. The run length distributions for each state in the simulated data were compared with the run length distributions in the MS Movers data set. S-PLUS (MathSoft, Inc. (1996)) was used to carry out the simulation results.

Table 7.11: Misclassification probability settings for MS Movers simulations.

Code	$v_{01}(0)$	$v_{01}(1)$	$v_{10}(0)$	$v_{10}(1)$
V1	0.050	0.050	0.050	0.050
V2	0.075	0.050	0.050	0.075
V3	0.100	0.100	0.100	0.100
V4	0.120	0.100	0.100	0.120
V5	0.150	0.150	0.150	0.150
V6	0.175	0.150	0.150	0.175
V7	0.200	0.200	0.200	0.200
V8	0.220	0.200	0.200	0.220

The average run length distributions for the simulations appear in Tables 7.12 and 7.13 for the two states. Generally, the average run length distributions tend not to agree closely with the observed data for both states. The misclassification probability settings V5 and V6 provide the closest results to the run length distributions of 1s. While the run lengths of size 1 and 3 are fit well in these simulations,

Table 7.12: Average run lengths of 0s for simulated misclassified data sets for shapes 0.5, 1 and 2. Run lengths for active scans in the MS Movers data set appear in the last row.

κ	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.5	V1	17.0	11.8	9.8	7.5	6.1	5.6	4.2	3.3	2.8	2.1	1.8	1.3	1.1	0.9	0.7	0.5	1.9
	V2	19.8	13.7	10.7	8.3	7.1	5.3	4.3	3.4	2.6	2.1	1.6	1.2	1.0	0.8	0.5	0.4	1.1
	V3	24.3	15.8	12.2	9.1	7.0	5.7	4.2	3.2	2.5	1.8	1.4	1.0	0.8	0.6	0.4	0.3	0.8
	V4	26.7	17.5	12.7	9.5	7.5	5.5	4.1	3.1	2.3	1.7	1.2	0.9	0.7	0.5	0.3	0.2	0.6
	V5	31.6	19.8	14.1	10.2	7.4	5.5	3.9	2.8	2.1	1.5	1.0	0.7	0.5	0.4	0.3	0.2	0.3
	V6	34.6	21.5	14.9	10.4	7.5	5.4	3.7	2.6	1.8	1.3	0.9	0.6	0.4	0.3	0.2	0.1	0.2
	V7	38.6	23.6	15.9	10.7	7.4	5.2	3.4	2.3	1.6	1.1	0.7	0.5	0.3	0.2	0.1	0.1	0.1
	V8	41.1	24.8	16.3	10.7	7.3	4.9	3.2	2.1	1.4	1.0	0.6	0.4	0.2	0.2	0.1	0.1	0.1
1.0	V1	19.7	13.9	11.2	8.7	6.9	5.5	4.3	3.4	2.6	2.1	1.5	1.2	1.0	0.7	0.4	0.5	1.1
	V2	22.6	16.0	12.1	9.3	7.1	5.6	4.3	3.3	2.4	1.9	1.3	1.1	0.7	0.5	0.4	0.3	0.8
	V3	26.5	17.6	13.5	9.9	7.3	5.6	4.2	3.1	2.3	1.8	1.2	0.9	0.7	0.5	0.3	0.3	0.5
	V4	29.1	19.5	13.9	10.3	7.4	5.5	4.1	2.9	2.1	1.5	1.1	0.8	0.5	0.4	0.3	0.2	0.3
	V5	33.4	21.4	15.3	10.6	7.6	5.3	3.8	2.6	1.9	1.3	0.9	0.6	0.4	0.3	0.2	0.2	0.2
	V6	36.6	23.1	15.9	11.0	7.4	5.1	3.6	2.4	1.6	1.1	0.7	0.5	0.3	0.2	0.1	0.1	0.1
	V7	40.2	24.8	16.7	11.0	7.4	4.9	3.4	2.2	1.5	1.0	0.6	0.4	0.2	0.2	0.1	0.1	0.1
	V8	43.1	26.0	16.9	11.0	7.2	4.7	3.0	2.0	1.3	0.8	0.5	0.3	0.2	0.1	0.1	0.0	0.1
2.0	V1	22.2	15.8	12.8	9.5	7.4	5.6	4.3	3.2	2.4	2.0	1.4	1.1	0.8	0.6	0.4	0.3	0.8
	V2	25.8	17.7	13.5	10.0	7.8	5.6	4.3	3.1	2.1	1.7	1.2	0.9	0.6	0.5	0.3	0.2	0.5
	V3	28.6	19.4	14.6	10.4	7.7	5.6	4.0	2.9	2.1	1.6	1.1	0.7	0.6	0.4	0.3	0.2	0.4
	V4	31.7	21.0	15.1	10.6	7.8	5.5	3.9	2.7	1.8	1.3	0.9	0.6	0.5	0.3	0.2	0.1	0.2
	V5	35.2	22.8	16.1	11.1	7.6	5.4	3.6	2.4	1.7	1.2	0.8	0.5	0.4	0.2	0.2	0.1	0.2
	V6	38.5	24.4	16.6	11.2	7.6	5.1	3.4	2.2	1.4	0.9	0.6	0.4	0.3	0.2	0.1	0.1	0.1
	V7	42.0	26.0	17.1	11.1	7.4	4.9	3.1	2.0	1.3	0.9	0.5	0.3	0.2	0.1	0.1	0.1	0.1
	V8	44.6	27.2	17.3	11.0	7.1	4.5	2.9	1.8	1.1	0.7	0.4	0.3	0.2	0.1	0.1	0.0	0.0
Movers		31	21	11	7	10	4	2	1	0	2	2	1	1	3	1	1	0

Table 7.13: Average run lengths of 1s for simulated misclassified data sets for shapes 0.5, 1 and 2. Run lengths for active scans in the MS Movers data set appear in the last row.

κ	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0.5	V1	34.0	11.0	6.5	3.7	2.1	1.4	1.0	0.6	0.4	0.2	0.2	0.1	0.1	0.0	0.0	0.0	0.0
	V2	39.0	12.0	6.5	4.0	2.3	1.5	0.9	0.5	0.5	0.3	0.2	0.1	0.1	0.1	0.0	0.0	0.0
	V3	45.9	13.8	7.0	3.8	2.0	1.2	0.8	0.5	0.3	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0
	V4	49.2	14.7	7.2	4.1	2.3	1.3	0.8	0.5	0.4	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0
	V5	55.6	16.9	7.6	3.9	2.0	1.1	0.7	0.4	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	V6	58.5	18.3	8.0	4.1	2.2	1.2	0.7	0.4	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	V7	63.3	20.1	8.5	4.1	2.0	1.1	0.6	0.3	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	V8	64.8	21.4	9.1	4.2	2.2	1.1	0.6	0.3	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
1.0	V1	37.3	15.0	7.7	4.0	2.2	1.2	0.6	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V2	41.8	15.6	7.8	4.3	2.3	1.3	0.8	0.3	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	V3	48.2	17.0	8.0	4.1	2.1	1.1	0.5	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V4	51.3	17.6	8.2	4.4	2.2	1.2	0.7	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V5	57.3	19.5	8.7	4.1	2.1	1.0	0.5	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V6	59.9	20.6	9.0	4.4	2.2	1.1	0.5	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V7	64.3	22.3	9.3	4.2	2.0	0.9	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V8	65.6	23.3	9.7	4.5	2.1	1.0	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2.0	V1	40.5	17.7	8.7	4.3	1.8	0.9	0.5	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V2	44.2	19.0	9.4	4.4	2.0	0.9	0.4	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V3	50.6	19.2	9.1	4.2	1.8	0.8	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V4	53.0	20.6	9.6	4.3	1.9	0.9	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V5	58.8	21.3	9.4	4.2	1.8	0.8	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V6	60.6	22.8	10.2	4.4	2.0	0.9	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V7	65.2	23.8	10.2	4.3	1.8	0.7	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	V8	66.3	24.8	10.5	4.6	2.0	0.8	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Movers		58	11	9	6	0	0	0	0	1	0	0	0	0	0	0	0	0

the run lengths of size 2 are generally over-represented. The corresponding results for the run lengths of 0s provided higher average run counts for the smaller run lengths than those seen in the MS Movers data set. In general, the larger misclassification probabilities examined tend to shorten the run lengths for each state and all shapes considered. These findings are consistent with the results presented in Chapter 5. Higher probabilities considered for the MS data set led to smaller estimates of the transition probabilities. Smaller transition probabilities imply the possibility of longer run lengths in the true data than seen in the observed data.

The run length distributions of the simulated true data sets can also be examined before the data is misclassified. In all cases, these data sets exhibited run length distributions close to the distributions found in the MS Movers data set and generally had too few short run lengths for each state. Hence, the smaller misclassification settings V1 and V2 tend to have many fewer run lengths of size 1 for either state than the MS Movers data set. Although the plausible estimated mean times were determined under the assumption that the data is correctly classified, these plausible values do not generate appropriate run length distributions. The simulated true data set must be misclassified to produce run length distributions similar to the observed data.

In Table 7.14, individual run length distributions are given for simulated data that closely matched the observed data set. These samples agree well with the smallest three run lengths for both states and suggest that any one of the three shapes could be plausible. The simulations also indicate that misclassification probabilities within [0.05,0.25] are consistent with the plausible estimated mean times considered and may be dependent on exacerbation status.

Based on the simulation results, several models where every parameter was specified were fit using the semi-Markov approach described earlier. These models followed the specification of (7.21) for ρ and η with misclassification probabilities set to values that provided close run length distributions for shapes 0.5, 1.0 and 2.0. These models did not provide close fits. Hence, the misclassification probabilities were estimated while ρ, η and κ were specified. As seen from the complete data set exacerbation status has estimated standard errors of about the same order as the estimates. The results for the three shape models with constant misclassification probabilities appear in Table 7.15. For each of the three shapes, the models give fairly consistent estimates for the misclassification probabilities and discrepancy function. The fit appears in Table 7.16 for shape 1 and generally reflects the same patterns seen in the two other shape models. The placebo 000 triplet appears to be over-represented by the model while the 111 triplet is under-represented. Each of the other shape models from Table 7.15 also display this same result. Any of the three shapes 0.5, 1 and 2 seem to yield roughly equivalent results under the specified treatment model parameters. When a model with all possible parameters was fit, as was done with M1 in the previous section, to the MS Movers data set, the results again suggest that the shape is close to the boundary of 0.01.

Additionally, maximum likelihood estimates were also found for several different treatment models with only the MS Movers data set. Models included specified values of the treatment parameters with misclassification probability parameters estimated as well as full estimation of all misclassification probability and treatment

Table 7.14: Sample run length distributions for simulations[†] which are close to the observed MS Movers data set.

	Run Length Distribution of 0s																
Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
S1	32	21	10	10	6	8	7	5	2	1	0	1	1	0	0	0	0
S2	32	21	10	5	13	4	4	1	4	2	1	2	1	1	0	0	0
S3	32	21	10	10	6	8	7	5	2	1	0	1	1	0	0	0	0
S4	32	21	10	5	13	4	4	1	4	2	1	2	1	1	0	0	0
S5	30	22	11	7	5	10	4	2	3	2	2	0	0	1	0	0	2
S6	31	23	9	7	10	7	5	4	3	1	0	1	1	1	0	0	0
S7	32	19	11	12	5	6	6	4	2	1	0	1	0	0	1	0	0
S8	30	19	11	6	10	3	5	5	2	1	1	1	1	0	2	1	0
S9	31	21	14	10	4	5	2	6	3	0	1	1	0	2	0	0	0
Movers	31	21	11	7	10	4	2	1	0	2	2	1	1	3	1	1	0

	Run Length Distribution of 1s																
Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
S1	57	12	8	4	2	0	1	1	0	0	0	0	0	0	0	0	0
S2	57	12	8	3	1	2	0	0	1	0	0	0	0	0	0	0	0
S3	57	12	8	4	2	0	1	1	0	0	0	0	0	0	0	0	0
S4	57	12	8	3	1	2	0	0	1	0	0	0	0	0	0	0	0
S5	60	13	9	1	0	2	0	0	0	0	0	0	0	0	0	0	0
S6	58	11	9	6	0	2	0	0	0	0	0	0	0	0	0	0	0
S7	58	14	8	3	2	5	1	0	0	0	0	0	0	0	0	0	0
S8	59	13	8	2	3	0	0	0	0	0	0	0	0	0	0	0	0
S9	57	13	9	4	5	1	0	1	0	0	0	0	0	0	0	0	0
Movers	58	11	9	6	0	0	0	0	1	0	0	0	0	0	0	0	0

[†] S1, S2, S3, S4: $\kappa = 0.5$ generated with V4
S5, S6: $\kappa = 1.0$ generated with V3
S7: $\kappa = 1.0$ generated with V5
S8: $\kappa = 2.0$ generated with V3
S9: $\kappa = 2.0$ generated with V5

Table 7.15: Estimated misclassification probabilities under specified shape and treatment parameters. Standard errors in parentheses.

κ	α_0	α_0^*	v_{01}	v_{10}	D_3
0.5	-2.148 (0.038)	-0.991 (0.235)	0.105	0.271	16.444
1.0	-2.317 (0.041)	-1.208 (0.290)	0.090	0.230	16.798
2.0	-2.472 (0.061)	-1.368 (0.358)	0.078	0.203	18.299

Table 7.16: Observed and expected triplet counts by treatment group for shape 1 model from Table 7.15. Expected counts in parentheses.

	Scan Activity Triplets							
	000	001	010	100	011	101	110	111
P	55	18	13	16	16	13	15	19
	(65.742)	(19.014)	(15.818)	(18.913)	(11.734)	(8.558)	(11.749)	(13.472)
L	87	19	19	17	6	6	5	6
	(82.319)	(18.676)	(16.190)	(18.541)	(8.230)	(5.774)	(8.245)	(7.026)
H	63	14	14	14	3	5	4	3
	(62.612)	(13.235)	(11.581)	(13.152)	(5.484)	(3.844)	(5.486)	(4.606)

parameters. Generally speaking, when all parameters are estimated, a likelihood ratio test concludes that the dependence on exacerbation status can be dropped from both misclassification probabilities. When v_{01} and v_{10} are restricted to be both less than 0.5 to permit estimation, the estimates are $\hat{v}_{01} = 0.16$ and $\hat{v}_{10} \approx 0.50$ with unreasonably large estimated mean times ($> 10^5$ days) for the different treatment groups and rate terms. This model has 8 parameters and -2 log-likelihood=552.134. If the treatment parameters are specified according to Table 7.10, estimation of the misclassification probability parameters again leads to a model without exacerbation status. The misclassification probabilities are smaller and appear with observed and expected transition counts in Table 7.17.

Table 7.17: Maximum likelihood misclassification probability estimates for model with treatment rates (7.21). ($\kappa = 1$)

Parameter	Estimate (se)	Correlation	Term	Estimate
α_0	-2.216 (0.338)		v_{01}	0.098
α_0^*	-1.169 (0.497)	0.071	v_{10}	0.237
			log-likelihood:	-283.099

Treatment	Scan Activity Pairs			
	00	01	10	11
P	77 (88.426)	33 (30.258)	30 (30.258)	36 (27.058)
L	113 (106.684)	26 (26.822)	25 (26.822)	12 (15.672)
H	82 (80.360)	19 (18.657)	20 (18.657)	7 (10.326)
			χ_1^2:	7.273

In summary, the semi-Markov estimating function approach worked well for the PI data set and provided fits as least as good as the likelihood approach for the MS data set with more reasonable parameter estimates. The semi-Markov approach requires significant computation time when compared with maximum likelihood estimation but allows both misclassification probabilities to be estimated without requiring misclassification predictors or constraints on these probabilities. The models consistently showed the misclassification probabilities did not depend reliably on exacerbation status. Exacerbation status proved to be more closely linked to the v_{01} misclassification probability than to v_{10}. The exacerbation status and shape coefficients seemed to be the least stable. This lack of stability may suggest that the relationship between exacerbation status and the misclassification probabilities was misspecified for all or some subgroup of the subjects. While the remitting state is misclassified as active with a relatively small probability, the relapsing state is misclassified as inactive 2 to 3 times more frequently. There does appear to be a treatment effect with the treated subjects spending more time on average in the remitting state and less time on average in the relapsing state than the placebo subjects. Subject heterogeneity seemed to be responsible for relatively large jackknife variance estimates and this heterogeneity was still inherent in the MS Movers data set. One approach to account for the excess heterogeneity would be the use of a frailty model (Aalen (1988)). Simulations suggested misclassification was required to generate data sets with similar run length distributions as the MS Movers data set and that these misclassification probabilities are likely in the interval [0.05,0.25].

7.4 General Markov Process

Next consider a true process which has exponential sojourn time distributions for $s+1$ states, where s is small. Label the states from 0 to s and let $S = \{0, 1, \ldots, s\}$. The observed state may be misclassified. The probability of misclassification is allowed to depend on at least one predictor. Let I be the number of subjects and assume each subject has n_i observations at times $t_{i1} < t_{i2} < \cdots < t_{in_i}$, $i = 1, \ldots, I$. The time between observation j and observation $j-1$ is denoted by Δt_{ij}, $\Delta t_{ij} = t_{ij} - t_{i,j-1}$. The true state of the process for subject i at time t is $\xi_i(t)$ and at observation j is $\xi_{ij} = \xi_i(t_{ij})$, while the observed state is labeled O_{ij}, with history $O_i^{(j)} = (O_{i1}, \ldots, O_{ij})$. Let the vector of misclassification predictors for subject i at observation j be denoted by $\boldsymbol{C}_{ij} = (1, C_{1ij}, \ldots, C_{qij})$, for some positive integer q. The history of the misclassification predictors at observation t_{ij}, $C_i^{(j)}$, contains all current and past values of the misclassification predictors for subject i. The probability of being misclassified in state b at observation j when the true classification is state a is denoted by $v_{ab}(\boldsymbol{C}_{ij})$, for $a, b \in S$.

A logistic relationship is assumed between the misclassification predictors and probabilities. Denote the parameters for misclassification in state b when the true state is a by $\boldsymbol{\alpha}_{ab}$, $\boldsymbol{\alpha}_{ab}^T = (\alpha_{0ab}, \alpha_{1ab}, \ldots, \alpha_{qab})$. The misclassification probabilities can be represented as

$$v_{ab}(\boldsymbol{C}_{ij}) = \frac{\exp(\boldsymbol{C}_{ij}\boldsymbol{\alpha}_{ab})}{1+\exp(\boldsymbol{C}_{ij}\boldsymbol{\alpha}_{ab})} = \frac{\exp(\alpha_{0ab} + \cdots + \alpha_{qab}C_{qij})}{1+\exp(\alpha_{0ab} + \cdots + \alpha_{qab}C_{qij})}$$

for $a \neq b$, $a, b \in S$, and $v_{aa}(\boldsymbol{C}_{ij}) = 1 - \sum_{b=0, b\neq a}^{s} v_{ab}(\boldsymbol{C}_{ij})$.

Subject i has covariate vector $\mathbf{x}_i = (1, \mathbf{x}_{1i}, \ldots, \mathbf{x}_{pi})$ for some positive integer p. Define the steady state probabilities for subject i as $\pi_b(\mathbf{x}_i) = \Pr(\xi_{ij} = b, \mathbf{x}_i)$ for $b \in \mathcal{S}$, $i = 1, \ldots, I$. The transition probabilities for the true process are $P_{ab}(\Delta t_{ij}, \mathbf{x}_i) = \Pr(\xi_{ij} = b \,|\, \xi_{i,j-1} = a, \mathbf{x}_i)$ for subject i at observation j with covariate vector $\mathbf{x}_i$, $a, b \in \mathcal{S}$. The recursive probability is defined as $\gamma_{ij}^{(a)}(\mathbf{x}_i) = \Pr(\xi_{ij} = a \,|\, O_i^{(j)}, \mathcal{C}_i^{(j)}, \mathbf{x}_i)$ for $a \in \mathcal{S}$ and the observed state probabilities become

$$
\begin{aligned}
\Pr(O_{i1} = k \,|\, \boldsymbol{C}_{i1}, \mathbf{x}_i) &= \sum_{b=0}^{s} v_{bk}(\boldsymbol{C}_{i1}) \pi_b(\mathbf{x}_i) \\
\Pr(O_{ij} = k \,|\, O_i^{(j-1)}, \mathcal{C}_i^{(j)}, \mathbf{x}_i) &= \sum_{b=0}^{s} v_{bk}(\boldsymbol{C}_{ij}) \sum_{a=0}^{s} \gamma_{i,j-1}^{(a)}(\mathbf{x}_i) P_{ab}(\Delta t_{ij}, \mathbf{x}_i)
\end{aligned}
$$

for $k \in \mathcal{S}$, $i = 1, \ldots, I$, $j = 1, \ldots, n_i$. The quantity $\gamma_{ij}^{(a)}(\mathbf{x}_i)$ can be determined by the recursion relation for $j > 1$

$$
\gamma_{ij}^{(a)}(\mathbf{x}_i) = \frac{v_{ak}(\boldsymbol{C}_{ij}) \sum_{b=0}^{s} \gamma_{i,j-1}^{(b)}(\mathbf{x}_i) P_{ba}(\Delta t_{ij}, \mathbf{x}_i)}{\sum_{b=0}^{s} v_{bk}(\boldsymbol{C}_{ij}) \sum_{a=0}^{s} \gamma_{i,j-1}^{(a)}(\mathbf{x}_i) P_{ab}(\Delta t_{ij}, \mathbf{x}_i)} \quad \text{if } O_{ij} = k
$$

where

$$
\gamma_{i1}^{(a)}(\mathbf{x}_i) = \frac{v_{ak}(\boldsymbol{C}_{ij}) \pi_a(\mathbf{x}_i)}{\sum_{b=0}^{s} v_{bk}(\boldsymbol{C}_{ij}) \pi_b(\mathbf{x}_i)} \quad \text{if } O_{i1} = k
$$

for $a, k \in \mathcal{S}$.

For subject i, the probability intensity of going from state a to b, $\lambda_{ab}(\mathbf{x}_i)$, is defined as

$$
\lambda_{ab}(\mathbf{x}_i) = \exp(\mathbf{x}_i \boldsymbol{\beta}_{ab}) \quad a \neq b, \quad a, b \in \mathcal{S}
$$

with $\lambda_{aa}(\mathbf{x}_i) = -\sum_{b=0,\, b\neq a}^{s} \lambda_{ab}(\mathbf{x}_i)$ and $\boldsymbol{\beta}_{ab} = (\beta_{0ab}, \ldots, \beta_{pab})^T$ is the vector of regression parameters. Collecting the parameters in the vector $\boldsymbol{\Theta}$, $\boldsymbol{\Theta} = \{\boldsymbol{\alpha}_{ab}\}_{a,b} \cup \{\boldsymbol{\beta}_{ab}\}_{a,b}$ for $a \neq b$, the likelihood can be written as

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{I} \Pr(O_{i1} \,|\, C_{i1}, \mathbf{x}_i; \boldsymbol{\Theta}) \prod_{j=2}^{n_i} \Pr(O_{ij} \,|\, O_i^{(j-1)}, C_i^{(j)}, \mathbf{x}_i; \boldsymbol{\Theta}).$$

The number of parameters will be determined by the number of possible transitions as well as the number of states which are always classified correctly. If all $s+1$ states have nonzero misclassification probabilities and it is possible to move from any one of the states in $\mathcal{S}$ to any of the other states in $\mathcal{S}$, then the number of parameters is equal to $\binom{s+1}{2}\big((q+1)+(p+1)\big)$ where q is the number of misclassification predictors and p is the number of covariates.

The relationship between the transition intensities and probabilities becomes complicated as the number of states increases. Numerically, however, the relationship is simpler. Diagonalization and matrix exponentiation of the transition intensities, as was done in Section 2.5.1, can be done with standard computer software. The situation will be simplified if any states are absorbing.

A further simplification occurs if not all states can be misclassified. In the case of an absorbing state such as death, the state may always be correctly classified. Other states may similarly have misclassification probabilities equal to 0.

In general, several tests of parameters may be of interest. In the misclassification part, the dependence on misclassification predictors can be tested. It may be the case that not all of the misclassification predictors provide information about the misclassification. The effect one misclassification predictor has on two different misclassification probabilities may be the same, $\alpha_{102} = \alpha_{103}$ for example. Further, it might be reasonable to have symmetric misclassification probabilities, $v_{ab}(C_{ij}) = v_{ba}(C_{ij})$. In the covariate part, different transition intensities may have the same relationship with the covariates.

The situation where $v_{0b}(C_{ij}) = v_{1b}(C_{ij}) = \cdots = v_{sb}(C_{ij})$, for all states $b \in \mathcal{S}$, $i = 1, \ldots, I$, $j = 1, \ldots, n_i$ is degenerate. In this case, the probability of being correctly classified in state b is the same as the probability of misclassification in any of the other states. As seen in the two-state alternating case, the regression parameters $\{\beta_{ab}\}$ cannot be estimated.

Chapter 8

Discussion and Further Work

This research focused on methodology for binary longitudinal responses collected on an alternating process and subject to misclassification. The misclassification mechanism was modeled separately from the underlying progress as a logistic function of misclassification predictors available at each observation time. The true process was modeled as a continuous-time counter model with time-independent covariates. The Type I counter was the focus and the sojourn time distributions examined reduced the model to either a Markov or semi-Markov model for the true process.

With the Markov model, identifiability and estimability of the model parameters were examined in Chapter 4. Exactly two distinct sets of parameter values were shown to provide equally likely explanations of the data when a likelihood approach was used for estimation provided that each transition rate depended on the same covariates and each misclassification probability depended on the same misclassification predictors. Restricting both misclassification probabilities to be less than 0.5 removed this problem, but this solution is appropriate only if it can be assumed that both probabilities are less than 0.5. This last point was highlighted by the results of Section 4.4.2 where the restriction forced one misclassification probability for the MS data set to be 0.5 and transition rate estimates were even more extreme than previously seen in Section 3.3.1.

Parameter estimability was described as a property of both the parameterization and a particular data set. Even if the restriction proposed is appropriate and identifiability is assured, model parameters may still not be estimable. Recall from Section 4.4, that parameters may not be estimable if the likelihood function is constant in some directions of the parameters. For example, when the sampling interval is too long relative to the sojourn times of the true states, not all transition rate regression parameters and misclassification predictor parameters are estimable. The results of Section 4.4.1 suggested that for small misclassification probabilities, the observations should be made at least 3 times during an interval equal to the average length of occupancy of each state. If the data are not sampled frequently enough, some parameters will need to be specified to allow estimation of other parameters. For the properly sampled and specified model, estimability will likely not be a problem if the observed history, misclassification predictors and covariates classify subjects into more distinct groups than parameters, as was seen in Section 4.4.3.

Identifiability and estimability issues were discussed specifically for the Markov case with either constant misclassification probabilities or misclassification probabilities modeled as logistic functions of the misclassification predictor. These issues could be further investigated for a Type I counter with non-exponential sojourn times, for a Type II counter and for links other than the logistic link. These issues also need to be addressed in terms of the estimating function approach discussed in Section 7.3.

The impact of misclassification on Markov transition probability estimates was investigated in Chapter 5 through simulation studies. Three estimates of the transition probabilities were examined as functions of the specified misclassification probabilities: naive estimates calculated assuming $v_{01} = v_{10} = 0$, maximum likelihood estimates found by maximizing the likelihood for specified values of the misclassification probabilities and approximated estimates based on a linear function of naive estimates and specified misclassification probabilities. The misclassification probabilities investigated were both at most 0.15. The maximum likelihood estimates were approximately unbiased for correct specification of v_{01} and v_{10}. The naive estimates over-estimated the transition probabilities and this bias increased as the misclassification probabilities increased. Depending on the true process transition probabilities, these estimates may have relative biases of 30% or higher when misclassification probabilities are both 5% or higher. While these estimates need to be assessed in conjunction with their standard errors, conclusions drawn on such estimates when $v_{01} \geq 0.05$ and $v_{10} \geq 0.05$ will likely not be appropriate. The approximated estimates were easier to calculate than the maximum likelihood estimates and were less biased than the naive estimates. When $v_{01} \leq 0.10$ and $v_{10} \leq 0.10$ and generally when $v_{01} = 0.15$ and $v_{10} < 0.05$ or $v_{01} < 0.05$ and $v_{10} = 0.15$, the biases were small and inferences drawn with misclassification probabilities in these ranges are likely to be reasonable. The approximated estimates can be used as an exploratory tool to allow researchers to more clearly see the dependency between transition probability estimates and the misclassification probabilities.

Similar investigations of estimates could be done on the regression parameter estimates as well as the transition parameter estimates, and allowing for varying as well as fixed times between observations. Estimates could be examined when the underlying process is not first-order Markov or the misclassification probabilities are not correctly specified. Further, misclassification predictors could be included in such investigations.

Model assessment is an important aspect of the model building process. Expected counts of one-step transition pairs, two-step transition pairs and one-step transition triplets under a stationarity assumption were compared with the observed counts using traditional goodness-of-fit measures with categorical covariates. In the model assessments of Chapter 6, the parameter estimates were obtained from maximum likelihood estimates when the underlying process was modeled as a Markov process. The comparison of expected frequencies with the observed transition frequencies was not extremely useful in the sense that these comparisons did not adequately detect important and obvious departures from the model. The MS analysis in Section 3.3.1 gave extreme parameter estimates which seemed to provide adequate fits in Chapter 6. More work is required in this area. Given the dependencies of the transition counts, specific degrees of freedom for tables such as those presented in Chapter 6 are necessary for any formal model testing. Assessment tools which check the fit of run length distributions for the two states could also be incorporated. While assessing the specification of both the misclassification and true process parts of the model is difficult, progress may more easily be made by looking at assessment of the true process when the misclassification probabilities are assumed known.

Simulations in Section 6.5 suggested that a gamma dead time distribution may be appropriate for the MS data set. An estimating function approach was taken to estimate parameters for this semi-Markov model. Under equilibrium assumptions for the true process, the joint probabilities for the true states at two and three consecutive times were calculated based on approximations to convolution equations. This approach permitted tractable calculations for a gamma dead time, and other distributions can easily be considered for the open and dead times provided that the computations are feasible. The parameter estimation was conducted by minimizing the discrepancy between observed and expected frequencies. The method was computationally intensive but seemed to permit estimation of both misclassification probabilities in the absence of a misclassification predictor in the sense that solutions to minimize the discrepancy were obtained.

The semi-Markov specification yielded estimates which fit the observed data better than the Markov model for the PI data set and provided at least as good a fit in the MS data set with more reasonable estimates. The latter model suggested that the exacerbation status was not informative regarding misclassification and subject heterogeneity may have been responsible for an unstable shape estimate. Whether or not the fits are significantly better than fits of the Markov model is yet to be determined. While the discrepancy of a specific model may be assessed by simulating data according to the estimates and determining how likely a more extreme discrepancy value is obtained, further work is necessary to test whether the shape parameter is significantly different from one. Further computations, via faster simulation based estimation techniques such as Monte Carlo Markov Chain methods (eg. Hastings (1970) and Geyer (1991)), can facilitate investigation of the sensitivity of this estimating approach, the relationship between the shape and the misclassification probability estimates, and the distribution of the discrepancy measure, as well as performance of other choices to measure the discrepancy.

Several other aspects could also be investigated. The work of this thesis concentrated on continuous-time models whereas discrete-time models could also be investigated. The recursive nature of the likelihood and estimating functions guided the choice of variance estimation by the jackknife. The performance of such a variance estimator in the context of misclassified longitudinal data needs to be examined. Missing data in the misclassification predictor for the MS data set were not handled in the analyses presented. Work could be done to incorporate this type of missingness. Further work can be done to examine different design aspects as well. Guidelines for the applicability of the methods presented are needed with respect to the number of subjects, the number of observations per subject and the criteria for a potential misclassification predictor.

Several suggestions can be made for the practical application of the methodology developed in this thesis for data where the misclassification probabilities are small. An investigator planning a study involving the type of data discussed can simulate data according to likely state estimated mean times and a range of misclassification probabilities to get an idea of the type of data which may be observed. Based on these simulations, a minimum desirable sampling frequency could be determined, to be followed or exceeded as permitted by the budget. Once the investigator has obtained the data, approximated estimates can be calculated, perhaps stratified by subgroups, for specified misclassification probabilities as an exploratory tool. These estimates may suggest how sensitive the Markov transition probabilities are to changes in the degree of misclassification. Finding maximum likelihood estimates for the regression parameters with specified values of the misclassification probabilities would be the next step in assessing the effect of misclassification on estimates. All parameters can then be estimated via maximum likelihood, possibly requiring constraints on the misclassification probabilities, and the variability of these estimates can be examined. Influential subjects can be identified and the stability of estimates can be observed. Using the estimates, data sets can be simulated to check that the run length distributions seen in the simulated data sets adequately match the observed data. Further, estimates could also be obtained from the semi-Markov estimating approach either with exponential or non-exponential sojourn time distributions. These estimates could be compared to the maximum likelihood estimates for the Markov case to identify discrepancies and stable estimates.

Appendix A

Data Sets

A.1 Multiple Sclerosis (MS) Data Set

Table A.1: MS baseline covariate descriptions.

Name	Description
ID	patient identification number
age	age in years at entry
initedss	EDSS score at entry
duration	duration of disease in years
dose	0 = placebo; 1= low dose; 2 = high dose
origin	0 = Washington; 1 = BC
sex	0 = male; 1 = female
area	burden of disease in mm^2 at entry

Table A.2: MS baseline covariates.

ID	age	initedss	duration	dose	origin	sex	area
420	42	2	4.4	0	1	0	2513
421	35	1	2.8	0	1	0	567
443	27	1.5	6.9	0	1	1	1808
446	34	1.5	6.9	0	1	1	14
449	38	4	4.2	0	1	0	171
451	34	2.5	17.1	0	1	0	4829
498	31	3.5	3.8	0	1	0	4490
501	35	2	16.2	0	1	1	325
504	33	1	14.3	0	1	1	379
507	41	2.5	6.9	0	1	0	1584
522	36	2.5	4.1	0	0	0	2167
523	37	1	1.1	0	1	1	3191
539	35	2	8	0	1	1	1980
540	25	1	1.5	0	1	1	136
545	30	1	9.1	0	1	1	498
550	38	1.5	5.4	0	0	1	3179
565	37	1	13.1	0	1	1	5074
419	31	1.5	12.1	1	1	0	3228
424	29	3	4	1	1	1	42
448	51	1.5	3.4	1	1	1	2103
450	49	3.5	20	1	1	1	2611
452	39	0	6.1	1	1	1	899
499	24	1.5	9.1	1	1	1	12295
502	48	0	5.8	1	1	1	227
505	47	3	14.3	1	1	1	0
508	31	2.5	1.5	1	1	1	1302
521	35	3	10.9	1	0	1	3834
525	36	3.5	15.2	1	0	1	187
542	34	1	5.9	1	1	1	371
544	27	1.5	3.7	1	1	0	5779
547	42	2	1.8	1	0	1	236
548	47	3.5	26.1	1	0	1	297
564	26	2	3.3	1	0	1	854
568	41	3.5	18.2	1	0	1	7794

continued on next page

	baseline covariates continued from previous page						
patient	age	initedss	duration	dose	origin	sex	area
422	34	1.5	12.5	2	1	1	1648
444	44	3	18	2	1	0	560
445	47	2.5	14.9	2	1	1	2426
453	35	1	10.1	2	1	1	1673
454	42	1.5	10.1	2	1	1	5778
497	20	3.5	1.1	2	1	1	4823
500	33	1.5	10.4	2	1	1	1883
503	30	3	3.2	2	1	1	1657
506	22	3.5	9.5	2	1	1	2583
524	36	3	7.5	2	1	1	4223
526	46	2.5	9.4	2	1	0	1559
541	36	2.5	8.8	2	1	1	634
543	45	1.5	23.8	2	1	1	917
546	49	5.5	22.8	2	0	0	3118
549	34	1.5	10.5	2	0	1	3640
566	42	3	2.6	2	1	1	333

Table A.3: MS Placebo Group: Day of Scan Data.

	Scan Number																
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
420	43	85	127	169	211	253	295	337	379	421	463	505	547	589	638	673	715
421	43	85	127	169	211	253	NA	337	381	414	463	505	540	589	631	680	722
443	48	82	125	173	202	251	295	335	377	419	461	503	545	594	629	678	720
446	43	85	129	169	213	255	302	344	NA	421	463	505	554	589	631	672	722
449	38	85	127	171	213	260	304	339	381	423	458	505	549	589	638	673	730
451	45	87	127	171	211	255	297	344	381	423	465	507	549	591	633	674	724
498	45	87	127	176	213	260	297	337	388	421	463	505	549	588	638	694	722
501	45	87	129	171	220	255	297	346	NA	423	465	507	549	591	633	682	752
504	45	92	129	169	220	267	297	339	388	423	465	507	556	584	633	675	724
507	45	87	122	171	213	255	297	339	388	423	465	514	549	591	633	668	724
522	43	85	141	169	211	253	281	337	379	421	456	505	554	596	637	680	743
523	43	85	126	169	204	253	295	337	NA	421	472	512	540	598	645	673	722
539	45	87	129	185	213	260	302	337	379	421	464	505	547	582	NA	NA	NA
540	43	86	127	183	218	260	295	337	379	421	463	505	547	589	631	666	722
545	43	92	127	171	211	255	288	344	381	421	470	519	547	596	638	673	736
550	38	87	129	164	220	255	290	339	381	423	465	507	556	591	633	675	731
565	45	87	129	178	213	269	297	339	381	423	465	500	554	591	638	689	724

Table A.4: MS Low Dose Group: Day of Scan Data.

	Scan Number																
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
419	43	85	127	169	211	253	295	337	379	421	463	505	547	589	631	673	729
424	50	85	127	169	204	253	297	337	379	428	463	505	547	596	633	680	722
448	43	85	127	171	213	255	295	339	381	NA	465	507	549	591	633	NA	724
450	45	87	127	169	213	253	297	339	381	423	465	507	549	598	633	682	724
452	43	85	127	169	225	255	297	344	393	423	463	507	549	591	633	682	724
499	45	87	129	171	213	260	297	339	381	423	465	507	549	588	633	696	724
502	43	85	127	169	211	255	311	337	381	416	465	507	549	591	633	668	731
505	45	87	127	169	218	262	NA	339	381	423	465	507	549	590	640	675	724
508	45	85	128	176	225	262	295	330	379	421	463	505	561	589	631	673	722
521	45	87	129	171	213	255	311	339	381	430	472	507	577	598	640	682	731
525	50	85	127	169	211	253	295	337	379	421	463	512	547	589	631	673	729
542	50	85	127	169	218	260	302	337	379	421	463	505	547	589	631	673	722
544	44	85	142	169	211	255	316	337	379	421	463	519	570	NA	NA	NA	NA
547	38	87	129	164	213	255	311	339	381	423	465	507	549	591	633	675	731
548	45	87	129	171	213	262	297	339	381	437	465	507	549	591	633	682	724
564	50	85	127	169	197	253	302	337	371	421	463	505	561	589	638	673	729
568	43	85	127	169	211	253	288	323	379	414	484	505	547	NA	NA	NA	NA

Table A.5: MS High Dose Group: Day of Scan Data.

	Scan Number																
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
422	43	85	127	169	211	253	297	344	388	414	465	505	554	589	631	673	729
444	43	85	129	169	206	255	297	339	381	423	458	507	549	598	633	675	724
445	43	85	129	171	213	255	297	339	381	423	465	507	549	591	633	675	724
453	43	87	122	171	213	255	297	339	381	423	465	514	549	591	633	673	729
454	43	94	122	171	213	255	297	344	381	423	465	514	549	598	633	675	724
497	23	85	127	171	213	255	297	339	381	423	465	507	549	591	633	675	724
500	45	87	127	169	211	253	295	344	381	421	465	507	549	591	633	661	724
503	45	87	127	169	234	254	304	344	379	430	470	519	556	596	631	673	722
506	43	85	126	169	211	253	295	337	379	421	462	505	NA	NA	NA	NA	NA
524	45	85	136	178	213	255	290	339	381	423	458	510	549	591	633	NA	731
526	43	85	127	169	211	253	295	330	379	421	463	505	547	596	638	680	722
541	43	85	129	171	213	255	297	337	381	423	465	505	549	591	633	675	724
543	43	80	127	169	204	246	295	330	372	442	470	505	547	596	631	673	722
546	43	92	129	178	213	248	297	330	381	423	465	519	NA	NA	NA	NA	NA
549	45	87	129	171	213	255	297	338	381	423	463	507	556	597	633	675	731
566	45	94	129	171	213	255	297	346	381	430	465	507	549	591	633	675	731

Table A.6: MS Active MRI Scans: 0=inactive, 1=active.

ID	Scan Number																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
420	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
421	1	1	1	1	1	1	NA	0	0	1	0	0	0	0	0	0	0
443	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1
446	0	0	0	0	0	0	0	0	NA	0	0	0	0	0	1	0	1
449	0	0	0	1	0	1	1	1	0	1	1	1	0	1	1	0	1
451	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
498	0	0	1	1	1	1	0	0	0	0	1	1	0	1	1	1	0
501	0	0	0	0	0	0	0	1	NA	0	0	1	0	1	0	0	1
504	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
507	0	0	1	1	1	0	1	1	0	1	0	1	1	1	0	0	1
522	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
523	0	0	0	1	0	0	0	0	NA	0	0	1	1	0	0	0	0
539	1	1	1	1	1	1	1	1	0	0	1	0	0	0	NA	NA	NA
540	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
545	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
550	0	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1
565	1	0	1	1	0	1	0	0	1	0	0	1	0	1	1	0	0
419	0	0	0	1	1	1	1	0	0	1	0	1	1	1	0	1	1
424	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
448	1	0	0	0	0	0	0	1	0	NA	0	1	1	0	0	NA	0
450	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
452	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1
499	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	1	0
502	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
505	0	0	1	0	0	0	NA	0	0	0	0	0	0	0	0	0	0
508	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
521	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	1	0
525	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
542	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
544	0	1	0	1	0	0	0	1	0	1	0	0	0	NA	NA	NA	NA
547	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
548	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
564	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0

continued on next page

active scans continued from previous page

ID	Scan Number																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
568	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA
422	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
444	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
445	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	0	0
453	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
454	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
497	1	1	1	1	0	0	0	1	0	1	1	0	1	0	0	0	0
500	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0
503	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
506	0	1	1	1	0	1	0	1	0	0	0	0	NA	NA	NA	NA	NA
524	0	0	1	0	0	1	0	1	1	0	1	0	0	1	1	NA	0
526	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
541	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
543	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
546	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA	NA
549	0	0	0	1	1	1	0	1	0	0	1	0	0	0	0	0	1
566	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.7: MS Exacerbation occurred between scans: 0=no, 1=yes.

ID	Scan Number																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
420	0	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	1
421	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0
443	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	1	0
446	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
449	0	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	1
451	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
498	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
501	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
504	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
507	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
522	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0
523	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
539	1	0	0	1	0	0	0	1	0	0	0	1	0	0	NA	NA	NA
540	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
545	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0
550	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1
565	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
419	1	1	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0
424	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1
448	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
450	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
452	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0
499	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0
502	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
505	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
508	0	1	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0
521	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
525	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
542	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
544	0	0	0	0	1	0	0	0	0	0	0	0	0	NA	NA	NA	NA
547	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
548	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
564	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0

continued on next page

exacerbations continued from previous page

	Scan Number																
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
568	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA
422	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
444	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
445	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
454	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
497	1	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	1
500	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
503	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
506	1	0	0	1	1	0	0	0	0	0	0	0	NA	NA	NA	NA	NA
524	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0
526	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
541	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
543	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0
546	0	0	0	0	0	0	0	0	0	1	0	0	NA	NA	NA	NA	NA
549	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1
566	0	0	1	0	1	0	1	1	0	1	1	0	0	0	0	1	0

A.2 Parasitic Infection (PI) Data Set

Table A.8: Weekly presence (1) or absence (0) of parasitic infection.

Child	Observations
1	101100111110
2	0000100000000100111000101
3	1100000000
4	00011111111
5	1111110011111111011111111
6	0010100010110000010000000
7	111110011111
8	101011100100
9	0000000000
10	111111110110
11	01100000100000
12	000001000011000000000000
13	11111000000111011
14	1000011001
15	000000000011111111
16	1111000011111111111101011
17	11001111100001111
18	000000001111
19	0000011000000011111111
20	000001101000000100111110
21	0000000000
22	1111111111111
23	111111111100000
24	01000000000000010
25	10110011111
26	1111111011111
27	00000011100011111111110
28	0011111111101
29	10001000111
30	01000000000
31	0000011011
32	1111111101

continued on next page

continued from previous page

Child	Observations
33	01100000000001111101
34	00000000000001000
35	00000111000001
36	00000000000010000000
37	000010010001
38	00000000000000000100000010
39	011111100101001000100000111001110001000111100
40	0011110110
41	111110110011
42	10111111110111101111111111
43	1011000000001000000000110101110000011100
44	000000000000
45	0000000000001001
46	1000011011110011
47	1111111000001000000000000011
48	00001111111011
49	11111111111
50	11111111101000110110011110000 1
51	01010000000000001
52	11110100000001011111111100
53	00111111111
54	0000111111
55	0010000000010
56	010000000010
57	11111100001000
58	00011111110111111111111111111110

Appendix B

Derivative Calculations for Approximated Transition Probability Estimators

The partial derivatives required in Section 5.3 calculated by Maple (Waterloo Maple Inc. (1996)) are quite lengthy and are listed here. Recall that the the number of observations on each subject is considered to be the same for notational convenience, and the time between observations is constant. The transition probabilities are written as $a = P_{01}(t)$ and $b = P_{10}(t)$.

To simplify the partial derivatives of ϕ_1 and ϕ_2, let $W = (1 - a - b)$ and $U = (1 - v_{01} - v_{10})$ and for $j = 2, \dots, n$ let $T_{i,j-1} = \frac{\partial \gamma_{i,j-1}}{\partial a}(1 - a - b) - \gamma_{i,j-1}$, $S_{i,j-1} = \frac{\partial \gamma_{i,j-1}}{\partial b}(1 - a - b) - \gamma_{i,j-1}$ and $V_{i,j-1} = \gamma_{i,j-1}(1 - a - b) + a$. The partial derivatives of ϕ_1 become,

$$
\begin{aligned}
\frac{\partial \phi_1}{\partial v_{01}} = & \sum_{i=1}^{I}\sum_{j=2}^{n}(2O_{ij}-1)\left[-(T_{i,j-1}+1)+U\left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01}\partial a}W-\frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\right)\right]\times \mathcal{P}_{ij}^{-1}\\
& -(2O_{ij}-1)^2\,U(T_{i,j-1}+1)\left(-V_{i,j-1}+\frac{\partial \gamma_{i,j-1}}{\partial v_{01}}WU+1\right)\times \mathcal{P}_{ij}^{-2}\\
\frac{\partial \phi_1}{\partial v_{10}} = & \sum_{i=1}^{I}\sum_{j=2}^{n}(2O_{ij}-1)\left[-(T_{i,j-1}+1)+U\left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{10}\partial a}W-\frac{\partial \gamma_{i,j-1}}{\partial v_{10}}\right)\right]\times \mathcal{P}_{ij}^{-1}\\
& -(2O_{ij}-1)^2\,U(T_{i,j-1}+1)\left(-V_{i,j-1}+\frac{\partial \gamma_{i,j-1}}{\partial v_{10}}WU\right)\times \mathcal{P}_{ij}^{-2}\\
\frac{\partial \phi_1}{\partial a} = & \sum_{i=1}^{I}\sum_{j=2}^{n}(2O_{ij}-1)\,U\left(\frac{\partial^2 \gamma_{i,j-1}}{\partial a^2}W-2\frac{\partial \gamma_{i,j-1}}{\partial a}\right)\times \mathcal{P}_{ij}^{-1}\\
& -(2O_{ij}-1)^2U^2(T_{i,j-1}+1)^2\times \mathcal{P}_{ij}^{-2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \phi_1}{\partial b} = & \sum_{i=1}^{I}\sum_{j=2}^{n}(2O_{ij}-1)\,U\left(\frac{\partial^2 \gamma_{i,j-1}}{\partial a\partial b}W-\frac{\partial \gamma_{i,j-1}}{\partial a}-\frac{\partial \gamma_{i,j-1}}{\partial b}\right)\times \mathcal{P}_{ij}^{-1}\\
& -(2O_{ij}-1)^2U^2(T_{i,j-1}+1)\left(\frac{\partial \gamma_{i,j-1}}{\partial b}W-\gamma_{i,j-1}\right)\times \mathcal{P}_{ij}^{-2},
\end{aligned}
$$

where

$$
\mathcal{P}_{ij} = (2O_{ij}-1)(1-v_{01}-v_{10})\left[\gamma_{i,j-1}(1-a-b)+a\right]+(1-O_{ij})(1-v_{01})+O_{ij}v_{01}
$$

for $i = 1, \dots, I$ and $j = 2, \dots, n$. The partial derivatives of ϕ_2 become,

$$\begin{aligned}
\frac{\partial \phi_2}{\partial v_{01}} = \sum_{i=1}^{I} \sum_{j=2}^{n} & (2O_{ij} - 1) \left[-S_{i,j-1} + U \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}} \right) \right] \times \mathcal{P}_{ij}^{-1} \\
& - (2O_{ij} - 1)^2 \, U S_{i,j-1} \left(-V_{i,j-1} + \frac{\partial \gamma_{i,j-1}}{\partial v_{01}} WU + 1 \right) \times \mathcal{P}_{ij}^{-2} \\
\frac{\partial \phi_2}{\partial v_{10}} = \sum_{i=1}^{I} \sum_{j=2}^{n} & (2O_{ij} - 1) \left[-S_{i,j-1} + U \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{10} \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{10}} \right) \right] \times \mathcal{P}_{ij}^{-1} \\
& - (2O_{ij} - 1)^2 \, U S_{i,j-1} \left(-V_{i,j-1} + \frac{\partial \gamma_{i,j-1}}{\partial v_{10}} WU \right) \times \mathcal{P}_{ij}^{-2} \\
\frac{\partial \phi_2}{\partial a} = \sum_{i=1}^{I} \sum_{j=2}^{n} & (2O_{ij} - 1) \, U \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial a \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial b} - \frac{\partial \gamma_{i,j-1}}{\partial a} \right) \times \mathcal{P}_{ij}^{-1} \\
& - (2O_{ij} - 1)^2 U^2 S_{i,j-1} (T_{i,j-1} + 1) \times \mathcal{P}_{ij}^{-2}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \phi_2}{\partial b} = \sum_{i=1}^{I} \sum_{j=2}^{n} & (2O_{ij} - 1) \, U \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial b^2} W - 2 \frac{\partial \gamma_{i,j-1}}{\partial b} \right) \times \mathcal{P}_{ij}^{-1} \\
& - (2O_{ij} - 1)^2 U^2 {S_{i,j-1}}^2 \times \mathcal{P}_{ij}^{-2}.
\end{aligned}$$

The first order partial derivatives for γ_{i1} are

$$\begin{aligned}
\frac{\partial \gamma_{i1}}{\partial v_{01}} &= \frac{(1 - O_{i1}) v_{10} \, a \, b}{[(1 - v_{01}) b + v_{10} \, a]^2} - \frac{O_{i1} (1 - v_{10}) a \, b}{[v_{01} \, b + (1 - v_{10}) a]^2} \\
\frac{\partial \gamma_{i1}}{\partial v_{10}} &= - \frac{(O_{i1} - 1)(1 - v_{01}) a \, b}{[(1 - v_{01}) b + v_{10} \, a]^2} - \frac{O_{i1} v_{01} a \, b}{[v_{01} \, b + (1 - v_{10}) a]^2} \\
\frac{\partial \gamma_{i1}}{\partial a} &= \frac{-(O_{i1} - 1)(1 - v_{01}) v_{10} \, b}{[(1 - v_{01}) b + v_{10} \, a]^2} + \frac{O_{i1} v_{01} (1 - v_{10}) b}{[v_{01} \, b + (1 - v_{10}) a]^2}
\end{aligned}$$

and

$$\frac{\partial \gamma_{i1}}{\partial b} = \frac{(O_{i1}-1)(1-v_{01})v_{10}\,a}{[(1-v_{01})b+v_{10}\,a]^2} - \frac{O_{i1}v_{01}(1-v_{10})a}{[v_{01}\,b+(1-v_{10})a]^2}.$$

The second order partial derivatives for the recursion relation for $j = 1$ are

$$\begin{aligned}
\frac{\partial^2 \gamma_{i1}}{\partial v_{01}\partial a} &= -\frac{(1-O_{i1})v_{10}\,b\,[-(1-v_{01})b+v_{10}\,a]}{[(1-v_{01})b+v_{10}\,a]^3} - \frac{O_{i1}(1-v_{10})b\,[v_{01}\,b-(1-v_{10})a]}{[v_{01}\,b+(1-v_{10})a]^3} \\
\frac{\partial^2 \gamma_{i1}}{\partial v_{10}\partial a} &= \frac{(O_{i1}-1)(1-v_{01})b\,[-(1-v_{01})b+v_{10}\,a]}{[(1-v_{01})b+v_{10}\,a]^3} - \frac{O_{i1}v_{01}\,b\,[v_{01}\,b-(1-v_{10})a]}{[v_{01}\,b+(1-v_{10})a]^3} \\
\frac{\partial^2 \gamma_{i1}}{\partial a\partial b} &= -\frac{(O_{i1}-1)(1-v_{01})v_{10}\,[-(1-v_{01})b+v_{10}\,a]}{[(1-v_{01})b+v_{10}\,a]^3} \\
&\quad - \frac{O_{i1}v_{01}(1-v_{10})\,[v_{01}b-(1-v_{10})a]}{[v_{01}\,b+(1-v_{10})a]^3} \\
\frac{\partial^2 \gamma_{i1}}{\partial a^2} &= 2\,\frac{(O_{i1}-1)(1-v_{01})v_{10}{}^2 b}{[(1-v_{01})b+v_{10}\,a]^3} - 2\,\frac{O_{i1}v_{01}(1-v_{10})^2\,b}{[v_{01}\,b+(1-v_{10})\,a]^3} \\
\frac{\partial^2 \gamma_{i1}}{\partial b^2} &= -2\,\frac{(O_{i1}-1)(1-v_{01})^2 v_{10}\,a}{[(1-v_{01})b+v_{10}\,a]^3} + 2\,\frac{O_{i1}v_{01}{}^2(1-v_{10})a}{[v_{01}\,b+(1-v_{10})a]^3} \\
\frac{\partial^2 \gamma_{i1}}{\partial v_{01}\partial b} &= -\frac{(O_{i1}-1)v_{10}\,a\,[-(1-v_{01})b+v_{10}\,a]}{[(1-v_{01})b+v_{10}\,a]^3} \\
&\quad + \frac{O_{i1}(1-v_{10})a\,[v_{01}\,b-(1-v_{10})a]}{[v_{01}\,b+(1-v_{10})a]^3}
\end{aligned}$$

and

$$\frac{\partial^2 \gamma_{i1}}{\partial v_{10}\partial b} = -\frac{(O_{i1}-1)(1-v_{01})a\,[-(1-v_{01})b+v_{10}a]}{[(1-v_{01})b+v_{10}\,a]^3} + \frac{O_{i1}v_{01}a\,[v_{01}b-(1-v_{10})a]}{[v_{01}b+(1-v_{10})a]^3}.$$

The first order partial derivatives for γ_{ij}, $j = 2, \ldots, n$, are

$$\begin{aligned}
\frac{\partial \gamma_{ij}}{\partial v_{01}} =& \frac{(1-O_{ij})\, v_{10}\, \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\, W}{-V_{i,j-1}\, U + 1 - v_{01}} + \frac{O_{ij}(1-v_{10})\, \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\, W}{V_{i,j-1}\, U + v_{01}} \\
&- \frac{(1-O_{ij})\, v_{10}\, V_{i,j-1}\, (-\frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\, W\, U + \gamma_{i,j-1}\, W + a - 1)}{(-V_{i,j-1}\, U + 1 - v_{01})^2} \\
&- \frac{O_{ij}\, (1-v_{10})\, V_{i,j-1}\, (\frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\, W\, U - \gamma_{i,j-1}\, W - a + 1)}{(V_{i,j-1}\, U + v_{01})^2} \\
\frac{\partial \gamma_{ij}}{\partial v_{10}} =& \frac{(1-O_{ij})\, V_{i,j-1}}{-V_{i,j-1}\, U + 1 - v_{01}} + \frac{(1-O_{ij})\, v_{10}\, \frac{\partial \gamma_{i,j-1}}{\partial v_{10}}\, W}{-V_{i,j-1}\, U + 1 - v_{01}} \\
&- \frac{(1-O_{ij})\, v_{10}\, V_{i,j-1}(-\frac{\partial \gamma_{i,j-1}}{\partial v_{10}}\, W\, U + \gamma_{i,j-1}\, W + a)}{(-V_{i,j-1}\, U + 1 - v_{01})^2} \\
&- \frac{O_{ij}\, V_{i,j-1}}{V_{i,j-1}\, U + v_{01}} + \frac{O_{ij}\, (1-v_{10})\, \frac{\partial \gamma_{i,j-1}}{\partial v_{10}}\, W}{V_{i,j-1}\, U + v_{01}} \\
&- \frac{O_{ij}\, (1-v_{10})\, V_{i,j-1}\, (\frac{\partial \gamma_{i,j-1}}{\partial v_{10}}\, W\, U - \gamma_{i,j-1}\, W - a)}{(V_{i,j-1}\, U + v_{01})^2} \\
\frac{\partial \gamma_{ij}}{\partial a} =& \frac{(1-O_{ij})\, (1-v_{01})\, v_{10}\, (T_{i,j-1} + 1)}{(U\, V_{i,j-1} - 1 + v_{01})^2} + \frac{O_{ij}\, v_{01}\, (1-v_{10})\, (T_{i,j-1} + 1)}{(U\, V_{i,j-1} + v_{01})^2}
\end{aligned}$$

and

$$\frac{\partial \gamma_{ij}}{\partial b} = \frac{(1-O_{ij})(1-v_{01})v_{10}\, S_{i,j-1}}{(UV_{i,j-1} - 1 + v_{01})^2} + \frac{O_{ij}v_{01}(1-v_{10})S_{i,j-1}}{(UV_{i,j-1} + v_{01})^2}.$$

For γ_{ij}, $j = 2, \ldots, n$ the second order partial derivatives are

$$
\begin{aligned}
\frac{\partial^2 \gamma_{ij}}{\partial v_{01} \partial a} = & \frac{(1 - O_{ij})\, v_{10}\, (\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial a} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}})}{-U\, V_{i,j-1} + 1 - v_{01}} \\
& - \frac{(1 - O_{ij})\, v_{10}\, (T_{i,j-1} + 1)\, (-\frac{\partial \gamma_{i,j-1}}{\partial v_{01}} W\, U + V_{i,j-1} - 1)}{(-U\, V_{i,j-1} + 1 - v_{01})^2} \\
& + \frac{(1 - O_{ij})\, v_{10}\, (T_{i,j-1} + 1)\, W\, U \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}}{(-U\, V_{i,j-1} + 1 - v_{01})^2} \\
& + \frac{(1 - O_{ij})\, (W \frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial a} - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}) v_{10}\, U\, V_{i,j-1}}{(-U\, V_{i,j-1} + 1 - v_{01})^2} \\
& - \frac{(1 - O_{ij})\, v_{10}\, V_{i,j-1}\, (T_{i,j-1} + 1)}{(-U\, V_{i,j-1} + 1 - v_{01})^2} \\
& - 2 \frac{(1 - O_{ij})\, v_{10}\, V_{i,j-1}\, (T_{i,j-1} + 1)\, U(-W\, U \frac{\partial \gamma_{i,j-1}}{\partial v_{01}} + V_{i,j-1} - 1)}{(-U\, V_{i,j-1} + 1 - v_{01})^3} \\
& + \frac{O_{ij}\, (1 - v_{10})\, (\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial a} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}})}{U\, V_{i,j-1} + v_{01}} \\
& - \frac{O_{ij}\, (1 - v_{10})\, (T_{i,j-1} + 1)\, (\frac{\partial \gamma_{i,j-1}}{\partial v_{01}} W\, U - V_{i,j-1} + 1)}{(U\, V_{i,j-1} + v_{01})^2} \\
& - \frac{O_{ij}\, (1 - v_{10})\, (T_{i,j-1} + 1)\, W\, U \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}}{(U\, V_{i,j-1} + v_{01})^2} \\
& - \frac{O_{ij}\, (1 - v_{10})\, U\, V_{i,j-1} (\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial a} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}})}{(U\, V_{i,j-1} + v_{01})^2} \\
& + \frac{O_{ij}\, (1 - v_{10})\, V_{i,j-1}\, (T_{i,j-1} + 1)}{(U\, V_{i,j-1} + v_{01})^2} \\
& + 2 \frac{O_{ij}\, (1 - v_{10})\, V_{i,j-1}\, (T_{i,j-1} + 1)\, U\, (\frac{\partial \gamma_{i,j-1}}{\partial v_{01}} W\, U - V_{i,j-1} + 1)}{(U\, V_{i,j-1} + v_{01})^3}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2\gamma_{ij}}{\partial v_{10}\partial a} =& \frac{(1-O_{ij})\,(T_{i,j-1}+1)}{-U\,V_{i,j-1}+1-v_{01}} + \frac{(1-O_{ij})\,v_{10}\,(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial a}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{-U\,V_{i,j-1}+1-v_{01}} \\
&- \frac{(1-O_{ij})\,v_{10}\,(T_{i,j-1}+1)\,(-\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,W\,U + V_{i,j-1})}{(-U\,V_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})\,V_{i,j-1}\,(T_{i,j-1}+1)\,U}{(-U\,V_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})\,v_{10}\,\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,W\,(T_{i,j-1}+1)\,U}{(-U\,V_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})\,v_{10}\,V_{i,j-1}\,(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial a}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})\,U}{(-U\,V_{i,j-1}+1-v_{01})^2} \\
&- \frac{(1-O_{ij})\,v_{10}\,V_{i,j-1}\,(T_{i,j-1}+1)}{(-U\,V_{i,j-1}+1-v_{01})^2} \\
&- 2\,\frac{(1-O_{ij})\,v_{10}\,V_{i,j-1}\,(T_{i,j-1}+1)\,U\,(-\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,W\,U + V_{i,j-1})}{(-U\,V_{i,j-1}+1-v_{01})^3} \\
&- \frac{O_{ij}\,(T_{i,j-1}+1)}{U\,V_{i,j-1}+v_{01}} + \frac{O_{ij}\,(1-v_{10})\,(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial a}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{U\,V_{i,j-1}+v_{01}} \\
&- \frac{O_{ij}\,(1-v_{10})\,(T_{i,j-1}+1)\,(\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,W\,U - V_{i,j-1})}{(U\,V_{i,j-1}+v_{01})^2} \\
&+ \frac{O_{ij}\,V_{i,j-1}\,(T_{i,j-1}+1)\,U}{(U\,V_{i,j-1}+v_{01})^2} - \frac{O_{ij}\,(1-v_{10})\,W\,(T_{i,j-1}+1)\,U\,\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}}{(U\,V_{i,j-1}+v_{01})^2} \\
&- \frac{O_{ij}\,(1-v_{10})\,V_{i,j-1}U\,(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial a}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{(U\,V_{i,j-1}+v_{01})^2} \\
&+ \frac{O_{ij}\,(1-v_{10})\,(T_{i,j-1}+1)\,V_{i,j-1}}{(U\,V_{i,j-1}+v_{01})^2} \\
&+ 2\,\frac{O_{ij}\,(1-v_{10})\,V_{i,j-1}\,(T_{i,j-1}+1)\,U\,(\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,W\,U - V_{i,j-1})}{(U\,V_{i,j-1}+v_{01})^3}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2\gamma_{ij}}{\partial a\partial b} =& \frac{(1-O_{ij})v_{10}\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a\partial b}W - \frac{\partial\gamma_{i,j-1}}{\partial a} - \frac{\partial\gamma_{i,j-1}}{\partial b}\right)}{-UV_{i,j-1}+1-v_{01}} \\
&+ 2\,\frac{(1-O_{ij})v_{10}(T_{i,j-1}+1)S_{i,j-1}U}{(-UV_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})v_{10}V_{i,j-1}U\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a\partial b}W - \frac{\partial\gamma_{i,j-1}}{\partial a} - \frac{\partial\gamma_{i,j-1}}{\partial b}\right)}{(-UV_{i,j-1}+1-v_{01})^2} \\
&+ 2\,\frac{(1-O_{ij})v_{10}V_{i,j-1}(T_{i,j-1}+1)U^2S_{i,j-1}}{(-UV_{i,j-1}+1-v_{01})^3} \\
&+ \frac{O_{ij}(1-v_{10})\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a\partial b}W - \frac{\partial\gamma_{i,j-1}}{\partial a} - \frac{\partial\gamma_{i,j-1}}{\partial b}\right)}{UV_{i,j-1}+v_{01}} \\
&- 2\,\frac{O_{ij}(1-v_{10})(T_{i,j-1}+1)S_{i,j-1}U}{(UV_{i,j-1}+v_{01})^2} \\
&- \frac{O_{ij}(1-v_{10})V_{i,j-1}U\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a\partial b}W - \frac{\partial\gamma_{i,j-1}}{\partial a} - \frac{\partial\gamma_{i,j-1}}{\partial b}\right)}{(UV_{i,j-1}+v_{01})^2} \\
&+ 2\,\frac{O_{ij}(1-v_{10})(T_{i,j-1}+1)V_{i,j-1}U^2S_{i,j-1}}{(UV_{i,j-1}+v_{01})^3}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2\gamma_{ij}}{\partial a^2} =& \frac{(1-O_{ij})v_{10}\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a^2}W - 2\,\frac{\partial\gamma_{i,j-1}}{\partial a}\right)}{-UV_{i,j-1}+1-v_{01}} + 2\,\frac{(1-O_{ij})v_{10}(T_{i,j-1}+1)^2U}{(-UV_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})v_{10}V_{i,j-1}U\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a^2}W - 2\,\frac{\partial\gamma_{i,j-1}}{\partial a}\right)}{(-UV_{i,j-1}+1-v_{01})^2} \\
&+ 2\,\frac{(1-O_{ij})v_{10}V_{i,j-1}(T_{i,j-1}+1)^2U^2}{(-UV_{i,j-1}+1-v_{01})^3} \\
&+ \frac{O_{ij}(1-v_{10})\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a^2}W - 2\,\frac{\partial\gamma_{i,j-1}}{\partial a}\right)}{UV_{i,j-1}+v_{01}} - 2\,\frac{O_{ij}(1-v_{10})(T_{i,j-1}+1)^2U}{(UV_{i,j-1}+v_{01})^2} \\
&- \frac{O_{ij}(1-v_{10})V_{i,j-1}U\left(\frac{\partial^2\gamma_{i,j-1}}{\partial a^2}W - 2\,\frac{\partial\gamma_{i,j-1}}{\partial a}\right)}{(UV_{i,j-1}+v_{01})^2} \\
&+ 2\,\frac{O_{ij}(1-v_{10})V_{i,j-1}(T_{i,j-1}+1)^2U^2}{(UV_{i,j-1}+v_{01})^3}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \gamma_{ij}}{\partial b^2} = & \frac{(1-O_{ij})v_{10}\,(\frac{\partial^2 \gamma_{i,j-1}}{\partial b^2}\,W - 2\,\frac{\partial \gamma_{i,j-1}}{\partial b})}{-UV_{i,j-1}+1-v_{01}} + 2\,\frac{(1-O_{ij})v_{10}\,{S_{i,j-1}}^2 U}{(-UV_{i,j-1}+1-v_{01})^2} \\
& + \frac{(1-O_{ij})v_{10}\,V_{i,j-1}U(\frac{\partial^2 \gamma_{i,j-1}}{\partial b^2}\,W - 2\,\frac{\partial \gamma_{i,j-1}}{\partial b})}{(-UV_{i,j-1}+1-v_{01})^2} \\
& + 2\,\frac{(1-O_{ij})v_{10}\,V_{i,j-1}{S_{i,j-1}}^2 U^2}{(-UV_{i,j-1}+1-v_{01})^3} \\
& + \frac{O_{ij}(1-v_{10})(\frac{\partial^2 \gamma_{i,j-1}}{\partial b^2}\,W - 2\,\frac{\partial \gamma_{i,j-1}}{\partial b})}{UV_{i,j-1}+v_{01}} - 2\,\frac{O_{ij}(1-v_{10}){S_{i,j-1}}^2 U}{(UV_{i,j-1}+v_{01})^2} \\
& - \frac{O_{ij}(1-v_{10})V_{i,j-1}U(\frac{\partial^2 \gamma_{i,j-1}}{\partial b^2}\,W - 2\,\frac{\partial \gamma_{i,j-1}}{\partial b})}{(UV_{i,j-1}+v_{01})^2} \\
& + 2\,\frac{O_{ij}(1-v_{10})V_{i,j-1}{S_{i,j-1}}^2 U^2}{(UV_{i,j-1}+v_{01})^3}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \gamma_{ij}}{\partial v_{01} \partial b} = & \frac{(1-O_{ij}) v_{10} \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\right)}{-UV_{i,j-1} + 1 - v_{01}} \\
& - \frac{(1-O_{ij}) v_{10} S_{i,j-1} \left(-\frac{\partial \gamma_{i,j-1}}{\partial v_{01}} WU + V_{i,j-1} - 1\right)}{(-UV_{i,j-1} + 1 - v_{01})^2} \\
& + \frac{(1-O_{ij}) v_{10} W S_{i,j-1} U \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}}{(-UV_{i,j-1} + 1 - v_{01})^2} \\
& + \frac{(1-O_{ij}) v_{10} V_{i,j-1} U \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\right)}{(-UV_{i,j-1} + 1 - v_{01})^2} \\
& - \frac{(1-O_{ij}) v_{10} V_{i,j-1} S_{i,j-1}}{(-UV_{i,j-1} + 1 - v_{01})^2} \\
& - 2 \frac{(1-O_{ij}) v_{10} V_{i,j-1} S_{i,j-1} U \left(-\frac{\partial \gamma_{i,j-1}}{\partial v_{01}} WU + V_{i,j-1} - 1\right)}{(-UV_{i,j-1} - v_{01} + 1)^3} \\
& + \frac{O_{ij}(1-v_{10}) \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\right)}{UV_{i,j-1} + v_{01}} \\
& - \frac{O_{ij}(1-v_{10}) S_{i,j-1} \left(WU \frac{\partial \gamma_{i,j-1}}{\partial v_{01}} - V_{i,j-1} + 1\right)}{(UV_{i,j-1} + v_{01})^2} \\
& - \frac{O_{ij}(1-v_{10}) W S_{i,j-1} U \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}}{(UV_{i,j-1} + v_{01})^2} \\
& - \frac{O_{ij}(1-v_{10}) V_{i,j-1} U \left(\frac{\partial^2 \gamma_{i,j-1}}{\partial v_{01} \partial b} W - \frac{\partial \gamma_{i,j-1}}{\partial v_{01}}\right)}{(UV_{i,j-1} + v_{01})^2} + \frac{O_{ij}(1-v_{10}) V_{i,j-1} S_{i,j-1}}{(UV_{i,j-1} + v_{01})^2} \\
& + 2 \frac{O_{ij}(1-v_{10}) V_{i,j-1} S_{i,j-1} U \left(\frac{\partial \gamma_{i,j-1}}{\partial v_{01}} WU - V_{i,j-1} + 1\right)}{(UV_{i,j-1} + v_{01})^3}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2\gamma_{ij}}{\partial v_{10}\partial b} =& \frac{(1-O_{ij})S_{i,j-1}}{-UV_{i,j-1}-v_{01}+1} + \frac{(1-O_{ij})v_{10}\,(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial b}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{-UV_{i,j-1}+1-v_{01}} \\
&- \frac{(1-O_{ij})v_{10}\,S_{i,j-1}(-\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,WU + V_{i,j-1})}{(-UV_{i,j-1}+1-v_{01})^2} + \frac{(1-O_{ij})V_{i,j-1}S_{i,j-1}U}{(-UV_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})v_{10}\,\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,WS_{i,j-1}U}{(-UV_{i,j-1}+1-v_{01})^2} \\
&+ \frac{(1-O_{ij})v_{10}\,V_{i,j-1}U(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial b}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{(-UV_{i,j-1}+1-v_{01})^2} \\
&- \frac{(1-O_{ij})v_{10}\,V_{i,j-1}S_{i,j-1}}{(-UV_{i,j-1}+1-v_{01})^2} \\
&- 2\,\frac{(1-O_{ij})v_{10}\,V_{i,j-1}S_{i,j-1}U(-\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,WU + V_{i,j-1})}{(-UV_{i,j-1}+1-v_{01})^3} \\
&- \frac{O_{ij}S_{i,j-1}}{UV_{i,j-1}+v_{01}} + \frac{O_{ij}(1-v_{10})(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial b}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{UV_{i,j-1}+v_{01}} \\
&- \frac{O_{ij}(1-v_{10})S_{i,j-1}(\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,WU - V_{i,j-1})}{(UV_{i,j-1}+v_{01})^2} + \frac{O_{ij}V_{i,j-1}S_{i,j-1}U}{(UV_{i,j-1}+v_{01})^2} \\
&- \frac{O_{ij}(1-v_{10})\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,WS_{i,j-1}U}{(UV_{i,j-1}+v_{01})^2} \\
&- \frac{O_{ij}(1-v_{10})V_{i,j-1}U(\frac{\partial^2\gamma_{i,j-1}}{\partial v_{10}\partial b}\,W - \frac{\partial\gamma_{i,j-1}}{\partial v_{10}})}{(UV_{i,j-1}+v_{01})^2} + \frac{O_{ij}(1-v_{10})V_{i,j-1}S_{i,j-1}}{(UV_{i,j-1}+v_{01})^2} \\
&+ 2\,\frac{O_{ij}(1-v_{10})V_{i,j-1}S_{i,j-1}U(\frac{\partial\gamma_{i,j-1}}{\partial v_{10}}\,WU - V_{i,j-1})}{(UV_{i,j-1}+v_{01})^3}.
\end{aligned}
$$

References

Aalen, O. O. (1988). Dynamic description of a Markov chain with random time scale. *Mathematical Scientist*, **13**, 90–103.

Abowd, J. M. and Zellner, A. (1985). Estimating gross flows. *Journal of Business and Economic Statistics*, **3**, 254–283.

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

Albert, G. E. and Nelson, L. (1953). Contributions to the statistical theory of counter data. *Annals of Mathematical Statistics*, **24**, 9–22.

Albert, P. S., Hunsberger, S. A., and Biro, F. M. (1997). Modeling repeated measures with monotonic ordinal responses and misclassification, with application to studying maturation. *Journal of the American Statistical Association*, **92**, 1304–1311.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Bekker, P. A. (1986). Comment on identification in the linear errors in variables model. *Econometrica*, **54**, 215–217.

Bekker, P. A., Van Montfort, K., and Mooijaart, A. (1991). Regression analysis with dichotomous regressors and misclassification. *Statistica Neerlandica*, **45**, 107–119.

Biggins, J. D. and Cannings, C. (1987). Markov renewal processes, counters and repeated sequences in Markov chains. *Advances in Applied Probability*, **19**, 521–545.

Billingsley, P. (1961). Statistical methods in Markov chains. *Annals of Mathematical Statistics*, **32**, 12–40.

Buck, R. C. (1956). *Advanced Calculus.* New York: McGraw-Hill.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, **85**, 652–663.

Chen, T. T. (1989). A review of methods for misclassified categorical data in epidemiology. *Statistics in Medicine*, **8**, 1095–1106.

Chunge, R. N. (1989). *Intestinal parasites in a rural community in Kiambu District, Kenya with special reference to* Giardia lamblia. PhD thesis, University College Galway.

Cook, R. J., Ng, E. T. M., and Meade, M. O. (1999). Estimation of operating characteristics for dependent diagnostic tests based on latent Markov models. Submitted to Biometrics.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.

Cox, D. R. and Isham, V. (1980). *Point Processes.* New York: Chapman and Hall.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes.* London: Chapman and Hall.

Dvurecenskij, A. and Ososkov, G. A. (1985). On a modified counter with prolonging dead time. *Journal of Applied Probability*, **22**, 678–687.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Espeland, M. A., Platt, O. S., and Gallagher, D. (1989). Joint estimation of incidence and diagnostic error rates from irregular longitudinal data. *Journal of the American Statistical Association*, **84**, 972–979.

Geyer, C. J. (1991). Markov Chain Monte Carlo maximum likelihood. In Kerimidas, E., editor, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface.* Fairfax Station: Interface Foundation of North America.

Gleser, L. J. (1989). The gamma distribution as a mixture of exponential distributions. *The American Statistician*, **43**, 115–117.

Goodman, L. A. (1958). Simplified runs tests and likelihood ratio tests for Markov chains. *Biometrika*, **45**, 181–197.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, **25**, 489–504.

Gu, X. (1996). *Statistical Analysis of Incomplete Data arising in Biomedical Studies.* PhD thesis, University of Waterloo.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Ito, J., Amari, S.-I., and Kobayashi, K. (1992). Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Transactions on Information Theory*, **38**, 324–333.

Juang, B. H. and Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics*, **33**, 251–272.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, **80**, 863–871.

Kasagi, F., Yanagawa, T., Horibe, H., and Aoki, N. (1988). Estimating the incidence and reversion rates of hypertension under two–time follow–up data with misclassifications. In Matusita, K., editor, *Statistical Theory and Data Analysis II.* North-Holland: Elsevier Science Publishers B.V.

Kesselring, J., editor (1997). *Multiple Sclerosis*, chapter 9. Cambridge: Cambridge University Press.

Kingman, J. F. C. (1962). The imbedding problem for finite Markov chains. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **1**, 14–24.

Kuha, J. and Skinner, C. (1997). Categorical data analysis and misclassification. In Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin, editors, *Survey Measurement and Process Quality*, chapter 28. New York: Wiley.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, **33**, 1444–1452.

Le, N. D., Leroux, B. G., and Puterman, M. L. (1992). Exact likelihood evaluation in a Markov mixture model for time series of seizure counts. *Biometrics*, **48**, 317–323.

Lindsey, J. K. (1993). *Models for Repeated Measurements.* Oxford: Clarendon Press.

Lohr, K. N. and Schroeder, S. A. (1990). A strategy for quality assurance in Medicare. *New England Journal of Medicine*, **322**, 707–712.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* New York: Chapman and Hall.

MathSoft, Inc. (1996). *S-PLUS, Version 3.4 Release 1.* Seattle: Author.

Matthews, W. B., editor (1991). *McAlpine's Multiple Sclerosis.* Edinburgh: Churchill Livingstone, 2nd edition.

McGilchrist, C. A. and Hills, L. J. (1991). A semi–Markov model for ear infection. *The Australian Journal of Statistics*, **33**, 5–16.

Nagelkerke, N. J. D., Chunge, R. N., and Kinoti, S. N. (1990). Estimation of parasitic infection dynamics when detectability is imperfect. *Statistics in Medicine*, **9**, 1211–1219.

Nguyen, D. G. and Murthy, D. N. P. (1984). Cost analysis of warranty policies. *Naval Research Logistics Quarterly*, **31**, 525–541.

Oliver, R. M. (1961). A traffic counting distribution. *Operations Research*, **9**, 802–810.

Panitch, H. S., Bever, C. T., Katz, E., and Johnson, K. P. (1992). Upper respiratory tract infections trigger attacks of multiple sclerosis in patients treated with interferon–beta [abstract]. *Journal of Neuroimmunology*, **35**(suppl 1), 125.

Paty, D. W., Li, D. K. B., the UBC MS/MRI Study Group, and the IFNB Multiple Sclerosis Study Group (1993). Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology*, **43**, 662–667.

Petkau, A. J. and White, R. (1995). Longitudinal analyses for the UBC 6-weekly frequent MRI substudy of the Betaseron multiple sclerosis clinical trial. Biostatistics Research Reports 10, University of British Columbia.

Poterba, J. M. and Summers, L. H. (1986). Reporting errors and labour market dynamics. *Econometrica*, **54**, 1319–1338.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: the art of scientific computing.* Cambridge: Cambridge

University Press, 2nd edition.

Pyke, R. (1958). On renewal processes related to type I and type II counter models. *Annals of Mathematical Statistics*, **29**, 737–754.

Quenouille, M. (1949). Approximation tests of correlation in time series. *Journal of the Royal Statistical Society, Series B*, **11**, 18–84.

Reeves, G. K. (1993). Goodness-of-fit tests in two–state processes. *Biometrika*, **80**, 431–442.

Reiersol, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, **18**, 375–389.

Satten, G. A. and Longini, I. M. (1996). Markov chains with measurement error: estimating the 'true' course of a marker for the progression of Human Immunodeficiency Virus disease. *Applied Statistics*, **45**, 275–309.

Sibley, W. A., Bamford, C. R., and Clark, K. (1985). Clinical viral infections and multiple sclerosis. *Lancet*, **1**, 1313–1315.

Sibley, W. A., Bamford, C. R., Clark, K., Smith, M. S., and Laguna, J. F. (1991). A prospective study of physical trauma and multiple sclerosis. *Journal of Neurological Neurosurgery Psychiatry*, **54**, 584–589.

Singer, B. and Cohen, J. E. (1980). Estimating malaria incidence and recovery rates from panel surveys. *Mathematical Biosciences*, **49**, 273–305.

Singer, B. and Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology*, **82**, 1–54.

Singh, A. C. and Rao, J. N. K. (1995). On the adjustment of gross flow estimates for classification error with application to data from the Canadian Labour Force Survey. *Journal of the American Statistical Association*, **90**, 478–488.

Singh, S. N. (1964). A probability model for couple fertility. *Sankhya, Series B*, **21**, 89–94.

Skinner, C. J. (1992). Logistic modelling of longitudinal survey data with measurement error. In *Proceedings of Statistics Canada Symposium 92 Design and Analysis of Longitudinal Surveys.*

Smith, W. L. (1958). Renewal theory and its ramifications. *Journal of the Royal Statistical Society, Series B*, **20**, 243–284.

Srinivasan, S. K. and Rajamannar, G. (1970). Counter models and dependent renewal point processes related to neuronal firing. *Mathematical Biosciences*, **7**, 27–39.

Stienman, L. (1996). Multiple Sclerosis: A coordinated immunological attack against the Central Nervous System. *Cell*, **85**, 299–302.

Stuber, J. L. (1996). C library for linear algebra. Sofware developed for M.Math. thesis, University of Waterloo.

Takács, L. (1956). On the sequence of events, selected by a counter from a recurrent process of events. *Probability Theory and its Applications*, **1**, 81–91.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, **29**, 614.

Waterloo Maple Inc. (1996). *Maple V Release 4*. Waterloo: Author.

Wolter, K. M. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.