11-1-2016

Hidden Markov Chain Analysis: Impact of Misclassification on Effect of Covariates in Disease Progression and Regression

Haritha Polisetti

University of South Florida, haritha.polisetti@yahoo.co.in

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the Biostatistics Commons, and the Public Health Commons

Scholar Commons Citation

http://scholarcommons.usf.edu/etd/6568

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Hidden Markov Chain Analysis: Impact of Misclassification on Effect of Covariates in Disease Progression and Regression

by

Haritha Polisetti

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Public Health Department of Epidemiology & Biostatistics College of Public Health University of South Florida

Major Professor: Yougui Wu, Ph.D. Yanzing Huang, Ph.D. Amy Alman, Ph.D.

Date of Approval: November 2, 2016

Keywords: Dementia, Hidden Markov model, misclassification

Copyright © 2016, Haritha Polisetti
# TABLE OF CONTENTS

List Of Tables .......................................................................................................................... ii

List Of Figures ........................................................................................................................ iii

Abstract ..................................................................................................................................... iv

Chapter One: Introduction ....................................................................................................... 1

Chapter Two: Testing A Proportion In An Environment Of Misclassification ..................... 4
  A Brief Review ......................................................................................................................... 4
  Error Rates in a Binomial Experiment.................................................................................. 4

Chapter Three: Addressing Misclassification Using Multi State Markov Models ............... 6
  A Brief Introduction of Multi State Markov Models .............................................................. 6
  Hidden Markov Model ......................................................................................................... 9
  Multi State Models with Misclassification ......................................................................... 10
  Misclassification Probabilities using MSM ...................................................................... 12

Chapter Four: Simulation Studies ............................................................................................. 13
  Study Design ....................................................................................................................... 13
  Simulation Results ............................................................................................................ 14

Chapter Five: Illustration Using Dementia Data .................................................................... 18
  Background on Dementia .................................................................................................... 18
  Data Source ....................................................................................................................... 20
  Specifying a Multi-State Markov Model .......................................................................... 21
  Results from Fitting Multi State Model to Dementia Data ............................................... 24
  Fitting Multi State Markov Model with Misclassification .................................................. 24
  Results from Fitting Hidden Markov Model (Multi State Model with Misclassification) to Dementia Data ................................................................. 27
  Misclassification Probabilities ......................................................................................... 29

Discussion ............................................................................................................................... 30

References ............................................................................................................................... 34

Appendices ............................................................................................................................... 43
LIST OF TABLES

Table 1 Results from simulation study – scenario without covariates, comparing the mean estimates from the model addressing misclassification and the model without addressing misclassification to the true transition probabilities. ........................................ 16

Table 2 Results from simulation study – scenario with covariate, comparing the mean covariate estimates from the model addressing misclassification and the model without addressing misclassification to the true covariate coefficient. ......................... 17

Table 3 Demographics ......................................................................................................................................... 22

Table 4 Frequencies of transitions between stages of dementia ........................................................................ 23

Table 5 Possible observed states for each underlying state for this study .......................................................... 26

Table 6 Comparing results from fitting dementia data with progressive multi state model and Hidden Markov model (multi state Markov model with misclassification) .................. 28

Table 7 Misclassification probabilities .................................................................................................................. 29
LIST OF FIGURES

Figure 1: Multi state model (General form)................................................................................. 6
Figure 2: Disease progression model (General form representing k states)................................. 8
Figure 3: Depicting observed and unobserved states at three time points................................. 9
Figure 4: Dementia Progression Markov chain model ................................................................. 21
ABSTRACT

Most of the chronic diseases have a well-known natural staging system through which the disease progression is interpreted. It is well established that the transition rates from one stage of disease to other stage can be modeled by multi state Markov models. But, it is also well known that the screening systems used to diagnose disease states may subject to error some times. In this study, a simulation study is conducted to illustrate the importance of addressing for misclassification in multi-state Markov models by evaluating and comparing the estimates for the disease progression Markov model with misclassification opposed to disease progression Markov model. Results of simulation study support that models not accounting for possible misclassification leads to bias. In order to illustrate method of accounting for misclassification is illustrated using dementia data which was staged as no cognitive impairment, mild cognitive impairment and dementia and diagnosis of dementia stage is prone to error sometimes. Subjects entered the study irrespective of their state of disease and were followed for one year and their disease state at follow up visit was recorded. This data is used to illustrate that application of multi state Markov model which is an example of Hidden Markov model in accounting for misclassification which is based on an assumption that the observed (misclassified) states conditionally depend on the underlying true disease states which follow the Markov process. The misclassification probabilities for all the allowed disease transitions were also estimated. The impact of misclassification on the effect of covariates is estimated by comparing the hazard ratios estimated by fitting data with progression multi state model and by fitting data with multi
state model with misclassification which revealed that if misclassification has not been addressed
the results are biased. Results suggest that the gene APOE ε4 is significantly associated with
disease progression from mild cognitive impairment to dementia but, this effect was masked
when general multi state Markov model was used. While there is no significant relation is found
for other transitions.
CHAPTER ONE:
INTRODUCTION

Understanding a disease by its uniquely characterized progression of symptoms and pathology plays vital role in correct diagnosis and suitable treatment plan. If a disease can be diagnosed or detected at an early state, it may be more responsive to treatment. An effective method to reduce mortality due to the disease can be effectively reduced by a systematic screening of population. A detailed knowledge of the natural history of a disease is very essential in order to develop and establish a systematic screening policy. The risk of onset of disease can be used to determine the type of population and time of population to screen but to determine the intervals between successive screens should be chosen based on the risk of progression. The risk of progression may vary with current stage of disease. Multistate Markov models can be effectively used to determine the course of a disease. (Jackson et al, 2003)

These models are very useful in estimating the transition rates between each disease state and simultaneously estimate the misclassification probabilities and also to understand the effect of covariates on transitions of disease states. (Andersen and Keiding, 2002; Commenges, 1999) Examples of application of multi-state models in medicine are liver cirrhosis (Anderson et al., 1991), screening for abdominal aortic aneurysm (Jackson et al., 2011), smoking prevention (Kalbfleisch and lawless, 1985; Chen et al., 2011), psoriatic arthritis (Chen et al., 2010; Cook et al., 2004, Sutradhar and Cook, 2008), screening of breast cancer (Duffy et al., 1995; Chen et al.,
1996, 2000), chronic myelogenous leukemia (Klein et al., 1984), diabetic complications (Kosorok and Chao, 1996; Marshall and Jones, 1995; Anderson, 1988), complications after heart transplantation (Sharples, 1993; Klotz and Sharples, 1994), acquired immune deficiency syndrome and Human immunodeficiency virus infection (Longini et al., 1989; Gentleman et al., 1994; Satten and Longini, 1996; Guihenneuc-Jouyaux et al., 2000; Alioum et al., 2005), hepatitis C virus (Sweeting et al., 2010), human papillomavirus (Bureau et al., 2003; Kang and Lagakos, 2007), hepatocellular carcinoma (Kay, 1986) and bronchiolitis obliteron after lung transplantation (Jackson and Sharples, 2002).

The multi-state Markov models are characterized by Markov property which states that the distribution of forthcoming state can be determined by the current state of disease. (The msm package, version 0.6.4) More details of multi-state Markov model is described in the subsequent sections. For the multi-state Markov model to determine the course of the disease, the current stage of disease should be determined without errors. But it is well known that any screening method or diagnostic methods are prone to errors which might lead to misclassification of the disease state. It is well established that misclassification of the outcome leads to bias in the estimates. Even though, similar effects are expected to be observed with misclassification of intermittent stages of disease, there are no sufficient studies reported to support the expectation that misclassification of the intermittent stages in a multi-state model leads to bias.

In this study, illustration of application of multi-state model with misclassification in screening of dementia is performed and simulation is used to prove that misclassification of the intermittent disease states in a multi-state model when misclassification is not addressed leads to bias of estimates. In order to account for the misclassification, multi-state model with misclassification is fit to the data and the misclassification probabilities are estimated and impact
of misclassification on the effect of covariates on the transitions is estimated to fill the space in the research addressing misclassification issues in multi-state models. In this study, data from subjects at different stages of dementia (brief description of dementia and stages of dementia is given under dementia section) is used to illustrate the effect of misclassification shown by simulation study and portray the method of correcting the misclassification by fitting the data with multi state Markov model without accounting for misclassification initially and then, determine the misclassification probabilities by fitting multi state Markov model with misclassification which enables to visualize the effect of accounting for misclassification compared to not accounting for misclassification.
CHAPTER TWO:
TESTING A PROPORTION IN AN ENVIRONMENT OF MISCLASSIFICATION

A Brief Review

Before introducing the misclassification issue in a complex multi-state model, a brief introduction about how error rate is addressed to test for a proportion in a misclassification environment is depicted in this section. In a binomial experiment, to estimate a population proportion ‘p’ from a large homogenous population and through random sampling, the sample proportion is a sufficient statistic, is the maximum likelihood estimator and minimum variance unbiased estimator of the population proportion ‘p’ in the absence of misclassification. But, in reality errors happen. The diagnostic or screening methods or surveys are subject to errors and there are many reasons explaining them and errors are inevitable during data collection. Making inferences using such data can be inimical. (Bradley & Farnsworth, 2013; Rohatgi, 2003; Hogg et al., 2005)

Error Rates in a Binomial Experiment

Let ‘T’ represent the true disease state of the subject and T=0 if subject is disease free and T=1 if subject has the disease. ‘p’ represents the proportion of subjects who have the disease
and is given by the probability \( p = \Pr(T = 1) \). Let ‘O’ represent the observed disease state by the screening test and \( O=1 \) for disease and \( O=0 \) for no disease. \( r = \Pr(O = 1) \) is the probability of subject testing positive for disease. Thus, the false positive rate is given by \( r_1 = \Pr(O = 1 \mid T = 0) \) and false negative rate is given by \( r_2 = \Pr(O = 0 \mid T = 1) \). Therefore, \( r \) can be written as

\[
 r = \Pr(O = 1) \\
 = \Pr(O = 1 \mid T = 0) \Pr(T = 0) + \Pr(O = 1 \mid T = 1) \Pr(T = 0) \\
= r_1 * (1 - p) + (1 - r_2) * p \\
= p(1 - r_1 - r_2) + r_1 \\
\]

(1)

Probabilities \( p \) and \( r \) are linearly related for \( r_1 \) and \( r_2 \).

\[
1 - r_1 - r_2 > 0 \\
\]

(2)

The above equation of inequality ensures that ‘\( r \)’ increases with increasing ‘\( p \)’ and ‘\( r \)’ decreases with decreasing ‘\( p \)’ and \( r > 0 \). Given that \( r_1 \) and \( r_2 \) are error rates, if \( 1 - r_1 - r_2 > 0 \), then the identifications of disease and no disease state are interchanged so that equation (2) is satisfied. (Bradley & Farnsworth, 2013; Rohatgi, 2003; Hogg et al., 2005) For example if \( p=0.30 \), \( r_1 = 0.01 \) and \( r_2 = 0.10 \) gives \( r = 0.277 \) which is less than ‘\( p \)’ the true proportion, if \( r_1 = 0.02 \) and \( r_2 = 0.08 \) gives \( r = 0.3 \) which is equal to ‘\( p \)’ and if, \( r_1 = 0.10 \) and \( r_2 = 0.01 \) gives \( r = 0.367 \) which is greater than ‘\( p \)’. Thus, based on the error rates, the estimated proportions are biased from the true proportion accordingly.
A Brief Introduction of Multi State Markov Models

Markov chains represent a class of stochastic processes of great interest for the wide spectrum of practical applications. The course of disease is modeled often using multi state models in continuous time. A general example that illustrates multi state Markov model is shown in (Figure 1). (Jackson et al., 2003; the msm package, version 0.6.4)

Figure 1 Multi state model (General form)

The possible transitions between each disease state are represented by direction of arrow marks. The disease state \( S_i(t) \) is observed for each individual \( i \) during arbitrary times \( t \) and it
may not be same for each individual. The disease state to which the individual moves and the
time of change will be determined by the transition intensity for each pair of states m & n. The
instantaneous risk of moving from stage ‘m’ to ‘n’ can be represented by the transition intensity.
The transition intensity matrix needs to be estimated in order to fit a multi-state model to the data
and for the general multi state model in (Figure 1) the transition intensity matrix Q takes the
form as

\[
Q = \begin{pmatrix}
q_{11} & q_{12} & q_{13} & q_{14} \\
q_{21} & q_{22} & q_{23} & q_{24} \\
q_{31} & q_{32} & q_{33} & q_{34} \\
q_{41} & q_{42} & q_{43} & q_{44}
\end{pmatrix}
\]

The matrix Q represents these transition intensities whose rows sum to zero, so that the
diagonal entries are given by (equation 3). (Jackson et al., 2003; the msm package, version 0.6.4)

\[
q_{mm} = -\sum_{n \neq m} q_{mn}
\]

The disease progression model which is used in this study is different from the general
multi state model in terms of possible transitions and depicted in (figure 2), where a series of
successive states of disease ending with an absorbing stage (death) is represented.

The subject is expected to progress to adjacent stage or recover to the previous stage or
move to absorbing stage (die) at any state of disease. (Jackson et al., 2003) Though, the model
used in this study does not contain an absorbing state as subjects who died at any stage of disease
were not included into the study.
In order to calculate the likelihood for multi-state models, transition probability matrix \((P_t)\) is required. Transition probability is the probability of transition of disease from stage \(m\) at time \(c\) to stage \(n\) at time \(t + c\) and is given by (equation 4).

\[
p_{mn}(t) = \Pr(S_i(t + c) = n | S_i(c) = m) \quad (4)
\]

\[
P(t) = e^{(tQ)} \quad (5)
\]

The information regarding the time of transition from state \(m\) to \(n\) is not given and the sampling times are assumed to be non-informative. \(P(t)\) can be determined from scaled transition intensity matrix by taking matrix exponential (equation 5). (Cox and Miller, 1965)
**Hidden Markov Model**

This study involves illustration of methodology involved in addressing for misclassification using multi state Markov model with misclassification which works with the principle of Hidden Markov model (HMM). There are two process in hidden Markov model, the observed process \((S(t_i))\) and the true underlying process \((S^*(t_i))\). The true states of Markov model are not observed in a hidden Markov model (HMM). Observed states \((S(t_i))\) of HMM is expected to be governed by emission distribution conditionally on underlying true states \((S^*(t_i))\) (Figure 3). The underlying states of Markov chain are determined based on the transition intensity matrix \(Q\). In Hidden Markov models, observations were evolved based on unknown distributions, thus HMM are mixture models but based on the states in HMM the distribution involved changes with time. Hidden Markov models are the best option for studies involving population with chronic disease with definite interpretation of stages. (The msm package, version 0.6.4, Jackson et al., 2003).

![Figure 3: Depicting observed and unobserved states at three time points](image-url)
The above figure indicates that the conditional distribution of \( (S^*(t_2)) \) is determined only by the previous state \( (S^*(t_1)) \). Similarly, the conditional distribution of \( (S(t_2)) \) given all the process \( (S^*(t_1)),(S^*(t_2)),(S^*(t_3)) \) and \( (S(t_1)),(S(t_3)) \), it is determined only by the current underlying state \( (S^*(t_2)) \).

**Multi State Models with Misclassification**

Multi state model with misclassification is a type of Hidden Markov model where observed states are expected to be misclassified. Here the observed data are states, assumed to be misclassifications of the true, underlying states. The screening process or the diagnostic method used to identify disease state may subject to error at times. In such situation the true Markov disease process \( (S^*_i(t)) \) for an individual ‘i’ can only be observed through realizations \( (S_i(t)) \) and cannot be observed directly. Thus, the quality of a screening test or diagnostic test can be determined by the probability that the observed disease state and the true underlying states are equal. (The msm package, version 0.6.4)

\[
\Pr(S_i(t) = m|S^*_i(t) = m) \quad (6)
\]

The sensitivity of the test or the probability that the test is a true positive is represented by the above equation when \( m \) is a positive disease state. If \( m \) is a disease free state, then the
above equation represents specificity, the probability that the diagnostic test truly identifies disease free subjects. (The msm package, version 0.6.4)

In this study, multi-state model with misclassification is fit to the dementia data in order to simultaneously determine the misclassification probabilities and also address for misclassification in estimating the parameters. The observed disease states $S_{ij}$ for subject $i$, at observation time $t_{ij}$ will be determined conditionally on the true underlying states $S^*_{ij}$ based on misclassification matrix $E$, where $(m, n)$ entries are given by (equation 7)

$$e_{mn} = \Pr(S(t_{ij}) = n | S^*(t_{ij}) = m) \quad (7)$$

Based on the knowledge of the screening or diagnosis process some of the $e_{rs}$ might be fixed which is analogous to the entries of Q matrix. The misclassification matrix $E$ governs the observed process of the underlying states.

In this study, multi-state model is fit to dementia data to determine if addressing misclassification has impact on the effect of covariates on disease progression or regression. In order to develop a suitable multi state model that fit’s the data, the natural history of dementia and its screening methods should be understood which is briefly explained in next subsection.
Misclassification Probabilities using MSM

The unique feature about Multi State Modeling is that the misclassification probabilities for transitions from one state to another can be determined when a multi-state model for more than 2 states is defined. The methodology involved in determining the misclassification probability can be explained by the formula shown in (equation 8). In order to investigate the explanatory variables for the misclassification probability for each pair of states, logistic model can be used. (The msm package, version 0.6.4) Probability of observing state 2 (mild cognitive impairment) given the underlying true state as state1 (no cognitive impairment) or state 3 (dementia) is illustrated using (equations 10-11).

\[
\Pr(Y_{ij} = 2| S_{ij} = 1, c_{ij}; \alpha) = \Pr(Y_{ij} = 2| S_{ij} = 3, c_{ij}; \alpha) = \frac{\exp(\alpha_0 + \alpha_1 c_{ij})}{1 + \exp(\alpha_0 + \alpha_1 c_{ij})} \quad (8)
\]

If the true state is normal or dementia, then the possible misclassification rate is given by (equation 9).

\[
\int_0^1 \frac{\exp(\alpha_0 + \alpha_1 x)}{1 + \exp(\alpha_0 + \alpha_1 x)} \, dx = \frac{1}{\alpha_1} \left[ \ln(1 + e^{\alpha_0 + \alpha_1}) - \ln(1 + e^{\alpha_0}) \right] \quad (9)
\]

This is the unique feature of multi state modeling where the misclassification probabilities can be estimated which is an added advantage of MSM.
CHAPTER FOUR: SIMULATION STUDIES

Study Design

Simulation study is conducted to illustrate the importance of addressing for misclassification in multi-state Markov models by evaluating and comparing the bias in the estimates for the disease progression Markov model with misclassification opposed to disease progression Markov model.

Simulation setting includes n = 3 states (state1, state2 and state3) and are assumed to follow Markov process. The sample size or number of subjects is 500 and 1000 replications are used for both models, multi-state model with misclassification and without misclassification. Subjects are assumed to start at any state of disease among the defined 3 stages at initial visit $t_{i0} = 0$ and are observed at 12 follow up visits at equal time intervals, $t_{ij}$ where, $j = 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22$ and 24 months. All transitions were allowed for this simulation study which follows the general form of multi state Markov model shown in Figure 1. The transition intensity matrix $Q$ is supplied with values ($q_{12} = 0.1, q_{13} = 0.01, q_{21} = 0.05, q_{23} = 0.1, q_{31} = 0.02, q_{32} = 0.07$) to generate the true states of disease. The diagonal values are ignored as each row sums to 1. The observed states are conditionally dependent on the true states with misclassification probabilities. The misclassification probability matrix
considered for simulation purpose is \((e_{12} = e_{13} = e_{21} = e_{23} = e_{31} = e_{32} = 0.30)\). A prognostic covariate following binomial distribution is introduced with a proportion of 0.5 and the covariate effect on the matrix of transition intensities is considered to be same for all the transitions with a value of -0.3. Two scenarios were investigated in this study, model without covariates and model with one prognostic covariates to investigate the effect of misclassification on estimation of transition probabilities and on covariate effect on transition probabilities. Both scenarios follow the same scheme except the inclusion of covariate in which a multi-state Markov model addressing for misclassification is fit to the data and a separate multi state Markov model is fit to the data without addressing for misclassification using the same regression coefficients for the covariate in both models and the average of the estimated covariate effect on the transitions from all the simulations is calculated and were compared with the true covariate effect in order to examine the performance of both the models in order to visualize the effect of misclassification on estimates and show that if misclassification is not accounted while estimating the transition probabilities in multi-state models, it might lead to bias. A very limited research has been reported on addressing misclassification issue in multi-state models and proving this concept theoretically is complex, thus simulation study is used to show the bias in estimates if misclassification has not been addressed.

**Simulation Results**

The results from simulation study for the scenario in which covariate is not included are compared between misclassification model and the MSM model in Table 1. It is observed that
the estimated transition intensities from Multi state model fit to the simulated data without addressing for misclassification are biased when compared to the true parameter. Results from the model with covariate were shown in Table 2 and the results suggest that the estimates from the model without addressing misclassification model were biased. Whereas the results from misclassification model in both scenarios (Multi state model is fit to the simulated data addressing misclassification) showed minimal or no bias when compared to true transition probabilities and true covariate effects. In order to assess the performance, the results were compared with the model without addressing for misclassification by reporting their bias from the true parameters in Table 1 These results strongly suggest that, there is significant impact of misclassification on estimation of the transition probabilities from one state to other state. Therefore, multi-state model with misclassification is proved to give unbiased estimates when compared to the multi-state model without addressing misclassification, which strengthens the argument that if misclassification not addressed in multi-state models, leads to bias and also the methodology proposed for addressing misclassification performs better compared to MSM without addressing misclassification.
Table 1 Results from simulation study – scenario without covariates, comparing the mean estimates from the model addressing misclassification and the model without addressing misclassification to the true transition probabilities.

<table>
<thead>
<tr>
<th>Transitions</th>
<th>True transition probability</th>
<th>Model addressing misclassification mean estimate</th>
<th>Bias</th>
<th>Model not addressing misclassification mean estimate</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1 - State 1</td>
<td>0.898</td>
<td>0.899</td>
<td>0.001</td>
<td>0.439</td>
<td>-0.459</td>
</tr>
<tr>
<td>State 1 - State 2</td>
<td>0.045</td>
<td>0.043</td>
<td>-0.002</td>
<td>0.287</td>
<td>0.242</td>
</tr>
<tr>
<td>State 1 - State 3</td>
<td>0.019</td>
<td>0.02</td>
<td>0.001</td>
<td>0.295</td>
<td>0.276</td>
</tr>
<tr>
<td>State 2 - State 1</td>
<td>0.088</td>
<td>0.087</td>
<td>-0.001</td>
<td>0.302</td>
<td>0.214</td>
</tr>
<tr>
<td>State 2 - State 2</td>
<td>0.866</td>
<td>0.866</td>
<td>0.000</td>
<td>0.395</td>
<td>-0.471</td>
</tr>
<tr>
<td>State 2 - State 3</td>
<td>0.063</td>
<td>0.058</td>
<td>-0.005</td>
<td>0.306</td>
<td>0.243</td>
</tr>
<tr>
<td>State 3 - State 1</td>
<td>0.014</td>
<td>0.014</td>
<td>0.000</td>
<td>0.258</td>
<td>0.244</td>
</tr>
<tr>
<td>State 3 - State 2</td>
<td>0.089</td>
<td>0.089</td>
<td>0.000</td>
<td>0.319</td>
<td>0.23</td>
</tr>
<tr>
<td>State 3 - State 3</td>
<td>0.917</td>
<td>0.921</td>
<td>0.004</td>
<td>0.402</td>
<td>-0.515</td>
</tr>
</tbody>
</table>
Table 2 Results from simulation study – scenario with covariate, comparing the mean covariate estimates from the model addressing misclassification and the model without addressing misclassification to the true covariate coefficient.

<table>
<thead>
<tr>
<th>Transitions</th>
<th>True covariate effect (exp((\beta)))</th>
<th>Model addressing misclassification</th>
<th>Model not addressing misclassification</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1 - State 2</td>
<td>0.741</td>
<td>0.730</td>
<td>-0.011</td>
</tr>
<tr>
<td>State 1 - State 3</td>
<td>0.741</td>
<td>0.702</td>
<td>-0.039</td>
</tr>
<tr>
<td>State 2 - State 1</td>
<td>0.741</td>
<td>0.605</td>
<td>-0.136</td>
</tr>
<tr>
<td>State 2 - State 3</td>
<td>0.741</td>
<td>0.785</td>
<td>0.044</td>
</tr>
<tr>
<td>State 3 - State 1</td>
<td>0.741</td>
<td>0.772</td>
<td>0.031</td>
</tr>
<tr>
<td>State 3 - State 2</td>
<td>0.741</td>
<td>0.783</td>
<td>0.042</td>
</tr>
</tbody>
</table>
CHAPTER FIVE:
ILLUSTRATION USING DEMENTIA DATA

In this study, the proposed method of accounting for misclassification using hidden Markov models (multi state Markov model with misclassification) is illustrated using dementia data.

Background on Dementia

Although commonly used to refer to a disease state, the term “dementia” does not refer to a disease at all but rather a syndrome characterized by memory loss and impaired activities of daily living (ADLs). (American Psychiatric Association, 1994) Alzheimer’s disease is a complex neurodegenerative disease characterized by a decline in cognition, behavioral disturbance and reductions in daily functioning and independence. Alzheimer’s disease is the most common form of dementia, accounting for 60-80% of all cases in epidemiological studies. (Knopman DS, 1998)

Alzheimer’s disease is a progressive brain disorder that slowly destroys memory and thinking skills and eventually, the ability to carry out the simplest tasks. In most people with Alzheimer’s, symptoms first appear in their mid-60s. Estimates vary, but experts suggest that
more than 5 million Americans may have Alzheimer’s. Dementia is the loss of cognitive functioning—thinking, remembering, and reasoning—and behavioral abilities to such an extent that it interferes with a person’s daily life and activities. (AD fact sheet, 2015)

The classic neuropathological signs of Alzheimer’s disease are amyloid plaques and neurofibrillary tangles. Plaques consist largely of the protein fragment beta-amyloid. This fragment is produced from a "parent" molecule called amyloid precursor protein. The accumulation of neurofibrillary tangles and neuronal loss is initially observed in trans-entorhinal and entorhinal cortex (ERC), and subsequently in the hippocampus (HPC). (Braak & Braak, 1991) Atrophy in the ERC and the hippocampus on MRI scans is also predictive of future cognitive decline and conversion to AD among individuals with Mild Cognitive Impairment (MCI). (Jack et al., 1999, 2000; Rusinek, 2003)

Dementia cannot be diagnosed by a single test. It is generally diagnosed based on medical history, brain scans, physical exams, laboratory tests, characteristics changes in thinking and behavior and impact on day to day functions. Dementia can be determined at a high level of certainty but determining the exact stage of dementia is difficult because, the symptoms and brain changes can overlap. (Alzheimer’s Association)

Analyzing such data in this context is very complicated and possible challenges in screening of stages of dementia are that the subjects are observed intermittently. For example, a healthy subject at first visit might die before his follow up visit and he might have transitioned to intermittent disease state without diagnosis. Even the exact transition time may not be known in most of the situations and it is also important to know the number of transitions occurred to determine the course of disease. Thus, for this study, this data is used to illustrate the application
of multi state model with misclassification to determine the impact of misclassification on effect of covariates in disease progression or regression.

**Data Source**

Data from subjects who participated in a diagnostic study of Alzheimer’s disease is used in this study. All subjects in this study completed subject evaluation which includes full clinical history, neurological evaluation, neurophysiologic tests, MRI brain scan, verbal learning test and standard blood tests. Consensus screening was performed on all the subjects by multiple clinicians as per National Alzheimer’s coordinating center NAAC protocol. The state of disease of subjects was determined according to national Institute of neurological and communicative Disorders and Stroke (NINCDS), Alzheimer’s disease and Related Disorders Association (ADRDA) criteria for AD. Subjects with no cognitive impairment, mild cognitive impairment and dementia were included in the study. The normal subjects, participants with no cognitive impairment were determined based on the cognitive score of informant interview, where there is no decline in cognition. Participants reported with stroke or transient ischemic attack or any cerebrovascular events were excluded from the study. 802 Subjects entered the study irrespective of their state of disease at first visit. Longitudinal evaluation procedure was used where subjects were followed and reevaluated at 1-year where 441 subjects turned out for follow up visit. Each subject’s state of disease was recorded at their follow up visit.
Specifying a Multi-State Markov Model

A multi-state Markov model is fit to the data using a progressive three state disease model in which a set of states (3 stages of disease) is considered and is shown in Figure 4. It is well established that dementia is an irreversible disease which means that once if a subject is diagnosed with mild cognitive impairment, he/she is supposed to progress to dementia and recovery from dementia is not possible without surgical treatment. Thus, in this study transition from MCI to no cognitive impairment stage is considered as it is referred to natural recovery but recovery from dementia to MCI is not allowed in this study as possible transition unless it is misclassified because without surgical treatment, dementia is theoretically not possible to revert to lower stages of disease.

\[ S = [s_1, s_2, s_3] \]  \hspace{1cm} (6)

Where,

\( s_1 \) = Normal or No cognitive impairment,

\( s_2 \) = MCI-Mild Cognitive Impairment,

\( s_3 \) = Dementia.

![Figure 4: Dementia Progression Markov chain model](image-url)
If the subject is at state $S_i$ and advances to $S_j$ with a probability $P_{ij}$ (Transition probability) which is not dependent on the state in which the subject is prior to the present $S_i$ state. If the subject remains in the same state, then the transition probability is given by $P_{ii}$. A common initial state is not specified for this study because; the data used for this study was collected from a diagnostic technique development study where participants enter the study irrespective of state. The demographics of the sample collected are shown in Table 3.

Table 3: Demographics

<table>
<thead>
<tr>
<th></th>
<th>NCI</th>
<th>MCI</th>
<th>Dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>258</td>
<td>412</td>
<td>132</td>
</tr>
<tr>
<td>Age, means(std)</td>
<td>72.65 (6.61)</td>
<td>75.03 (6.29)</td>
<td>76.53 (7.25)</td>
</tr>
<tr>
<td>Female %</td>
<td>69.69</td>
<td>52.42</td>
<td>52.76</td>
</tr>
<tr>
<td>Low education %</td>
<td>7.04</td>
<td>19.89</td>
<td>17.82</td>
</tr>
<tr>
<td>Apoe 4 %</td>
<td>25.48</td>
<td>30.53</td>
<td>51.85</td>
</tr>
</tbody>
</table>

The Markov process for this study starts in one of the states mentioned in Figure 4 and moves successively from one state to the other. For example, if the subject is diagnosed with a mild cognitive impairment at his first visit, then his is screened again for his state of disease during his second visit, he might have progressed to next state of disease, reversed or recovered to the previous no disease state or might have stayed in the same state. Absorbing state is not included in the model. All this information is recorded for all the participants and used to determine the frequencies of transitions between the stages of dementia and were shown in Table 4. Frequencies were reported for the participants who showed up for the follow up visit.
It is supposed that the cognitive impairment progresses from no cognitive impairment to Mild cognitive impairment and then progress to dementia and recovery to the adjacent previous stage are considered possible. Accordingly, a plausible transition intensity matrix is developed (equation 10).

Table 4 Frequencies of transitions between stages of dementia

<table>
<thead>
<tr>
<th>Initial stage</th>
<th>Frequency of transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
</tr>
<tr>
<td>Stage 1 No Cognitive impairment</td>
<td>143</td>
</tr>
<tr>
<td>Stage 2 Mild Cognitive impairment</td>
<td>30</td>
</tr>
<tr>
<td>Stage 3 Dementia</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
Q = \begin{pmatrix}
-(q_{12}) & q_{12} & 0 \\
q_{21} & -(q_{21} + q_{23}) & q_{23} \\
0 & q_{32} & -q_{32}
\end{pmatrix}
\] (10)

It indicates progression and recovery from one stage to its adjacent stage and direct transition from no cognitive impairment to dementia or the reciprocal is not allowed as it is not possible medically in 1 year unless there is error in diagnosis. Using this transition intensity matrix, multistate model is fit to the data using msm package “R” This matrix is required to specify the allowed transitions and the transitions that are not allowed are given a value of 0. Initial values are supplied to all other possible transitions leaving the diagonal values in Q matrix
as they are the negative value of the sum of all other transitions in that row and each row sums to 1. Then, multi-state model is fit to the data by supplying the Q matrix with allowed transitions and appropriate initial values using msm package in R. (Jackson (2011); msm package R)

**Results from Fitting Multi State Model to Dementia Data**

Multi state model was fit to the data using three transition states and using apoe ε4 and low education as covariates without addressing for misclassification. Where transition probabilities and hazard ratios for disease progression and disease regression at one-year follow-up were obtained and presented in Table 6. There is no significant effect in both disease progression and regression shown by gene Apoe ε4 and low education.

According to the results from model-1 (model without accounting for misclassification), Apoe ε4 gene and low education do not show and significant effect on transition of disease from no cognitive impairment (state1) to mild cognitive impairment (state 2), mild cognitive impairment (state2) to dementia (state3) or the recovery from any state to its adjacent previous state. Therefore, claiming that positive Apoe ε4 gene and low education does not have effect on disease progression or disease regression

**Fitting Multi State Markov Model with Misclassification**

Fitting general multi state Markov model to that data might lead to biased estimates as the screening or diagnosis of dementia states are subject to error. It should be noted that though
the transition from dementia to lower states is not theoretically possible, there are subjects showing transition from dementia to MCI which is possible only in case of misclassification. Therefore, the resulting model without addressing misclassification is compared with hidden Markov model to determine the transition misclassification rates.

As stated before, for multi-state Markov model with misclassification it is assumed that the underlying true states follow Markov process with the matrix \( Q \) (transition intensity matrix) and the observed disease states are assumed to depend on the corresponding underlying true states with misclassification probability. Considering the irreversible nature of the dementia without surgical procedures and it is also known that only in rare situations recovery is possible, so the above mentioned model might be medically not realistic for majority of population. Thus, it is required to fit a multi-state Markov model with misclassification (Hidden Markov model) in order to account for misclassification. For that, the previous two intensity matrix is replaced by a one-way transition intensity matrix (equation 11) where recovery from any state is not considered.

\[
Q = \begin{pmatrix}
-q_{12} & q_{12} & 0 \\
0 & -q_{23} & q_{23} \\
0 & 0 & 0
\end{pmatrix}
\]  \hspace{1cm} (11)

It is assumed that, state 1 (no cognitive impairment could be classified as either mild cognitive impairment or no cognitive impairment (state 1 or state 2). similarly, state 2 (mild cognitive impairment could be classified as no cognitive impairment or dementia or mild
cognitive impairment (state1, state2 or state3) and likewise, dementia (state 3) could be classified as state 2 (mild cognitive impairment) or state 3. Possible observed states for each underlying true state is depicted in Table 5. Based on these possibilities; the misclassification matrix is given below (equation 12) where rows represent underlying states and columns represent observed states.

Table 5: Possible observed states for each underlying state for this study

<table>
<thead>
<tr>
<th>Underlying true state</th>
<th>Possible observed state</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>State 1</td>
</tr>
<tr>
<td></td>
<td>State 2</td>
</tr>
<tr>
<td>State 2</td>
<td>State 1</td>
</tr>
<tr>
<td></td>
<td>State 2</td>
</tr>
<tr>
<td></td>
<td>State 3</td>
</tr>
<tr>
<td>State 3</td>
<td>State 2</td>
</tr>
<tr>
<td></td>
<td>State 3</td>
</tr>
</tbody>
</table>

\[
E = \begin{pmatrix}
1 - e_{12} & e_{12} & 0 \\
e_{21} & 1 - e_{21} - e_{23} & e_{23} \\
0 & e_{32} & 1 - e_{32}
\end{pmatrix}
\]

(12)

The E matrix is defined in order to model the observed states with misclassification and the value is given as zero if there is no misclassification permitted and this is determined using Table 5 and all other misclassification probabilities \((e_{12}, e_{21}, e_{23} \text{ and } e_{32})\) were given an initial
value of 0.1. Similar to the Q matrix, the diagonal values were ignored as the rows sum to 1. A Hidden Markov model is fit to the data along with the new Q matrix and E matrix to determine the misclassification probabilities. It is also investigated whether the misclassification probabilities depend on covariates by using the ‘misc covariates’ argument in msm. (Jackson (2011); msm package R)

Results from Fitting Hidden Markov Model (Multi State Model with Misclassification) to Dementia Data

The results from model-2 (Hidden Markov model) to dementia data are contrary to that obtained from fitting multi state Markov model to dementia data for one of the disease progression transition. Though estimates for remaining transitions were in line with the estimates from model-1. Results from model-2 were shown in Table 6 which suggests that Apoe  ε4 gene is significantly associated with the disease transition from mild cognitive impairment to dementia. There is no significant effect shown by Apoe  ε4 gene for rest of the disease state transitions and also there is no significant association shown by low education on any of the disease state transitions related to neither disease progression or in the regression of the disease. The misclassification probabilities for each disease transition mentioned in the Q matrix were also estimated using this model.
Table 6 Comparing results from fitting dementia data with progressive multi state model and Hidden Markov model (multi state Markov model with misclassification)

<table>
<thead>
<tr>
<th>Progression of disease</th>
<th>Model addressing misclassification</th>
<th>Model not addressing misclassification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Apoe ε4</td>
<td>Low education</td>
</tr>
<tr>
<td></td>
<td>HR* 95% CI p-value</td>
<td>HR* 95% CI p-value</td>
</tr>
<tr>
<td>NCI-MCI</td>
<td>1.094 (0.512-1.619) 0.772</td>
<td>1.732 (1.242-1.924) 0.375</td>
</tr>
<tr>
<td>MCI-Dementia</td>
<td>1.973 (1.451-1.999) &lt;0.0001</td>
<td>1.745 (0.469-1.986) 0.651</td>
</tr>
<tr>
<td>Regression of disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI-NCI</td>
<td>0.104 (0.0001-1.967) 0.768</td>
<td>1.745 (0.469-1.987) 0.769</td>
</tr>
<tr>
<td>Dementia-MCI</td>
<td>1.35E-07 (0 - Inf) 0.6514</td>
<td>0.732 (0.101-1.724) 0.2244</td>
</tr>
</tbody>
</table>

*Notes: HR is the estimated hazard ratio; 95% CI is the 95% confidence interval.
Misclassification Probabilities

Misclassification probabilities for the defined transitions in the Q matrix were estimated and shown in Table 7. The probability of observing no disease state given the underlying true state is no disease state is found to be 0.816, which is considered as 81.6% specificity of classifying a subject with no cognitive impairment as disease state 1. The probability of misclassifying Normal state as MCI is found to be 0.183(0.091, 0.334) and the probability of misclassifying MCI as Normal is found to be 0.0157(9.25e-12, 1.0). The probability of observing MCI given MCI state is found to be 0.984(2.24e-08, 1.0) which can be considered as sensitivity of observing MCI as 98.4%. Similarly, the probability of observing dementia given dementia is found to be 0.997(1.42e-109, 1.0) which can also be considered as sensitivity of diagnosing dementia as 99.7%. The probability of misclassifying MCI as dementia is found to be 0.00012(2.88e-07, 0.049) and the probability of misclassifying Dementia as MCI is found to be 0.0028(1.12e-114, 1.0). Thus, the false positive rate for mild cognitive impairment stage is found to be 18% and false positive rate for dementia is found to be 0.012 %.

<table>
<thead>
<tr>
<th>Underlying true state</th>
<th>Observed state</th>
<th>Misclassification probability</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>State 2</td>
<td>0.1834</td>
<td>(0.0914, 0.334)</td>
</tr>
<tr>
<td>State 2</td>
<td>State 1</td>
<td>0.0157</td>
<td>(9.25e-12, 1.0)</td>
</tr>
<tr>
<td></td>
<td>State 3</td>
<td>0.00012</td>
<td>(2.88e-07, 0.049)</td>
</tr>
<tr>
<td>State 3</td>
<td>State 2</td>
<td>0.0028</td>
<td>(1.12e-114, 1.0)</td>
</tr>
</tbody>
</table>
DISCUSSION

Most of the chronic diseases have a well-known natural staging system through which the disease progression is interpreted. It is well established that the transition rates from one stage of disease to other stage can be modeled by multi state Markov models. But, it is also well known that the screening systems used to diagnose disease states may subject to error some times. In this study simulation is used to illustrate the importance of addressing for misclassification in multi-state Markov models by evaluating and comparing the estimates for the disease progression Markov model with misclassification opposed to disease progression Markov model. These results from simulation study strongly suggest that the estimates from multi state model without addressing for misclassification lead to significant bias when compared to the true parameters and suggest better performance of multi state models with misclassification. The application of hidden Markov model (multi state model with misclassification) to real data is illustrated using dementia data and impact of misclassification on effect of covariates on disease transition is evaluated.

It is well established that apoε4 is associated with Alzheimer’s disease but its association with Dementia is inconsistent. (Yi-Fang Chuang et al., 2010) There was minimal research on examining the association of apoε4 gene on transition of disease from one stage to the other. A general hidden Markov model was presented for estimating transition rates and probabilities of misclassification of stages of disease and concluded that regression of disease can be explained by misclassification. This is because even though Markov processes were well
established method of estimating rates of transition between each stage of disease, diagnosis of disease stages might be subject to error. (Jackson, Christopher H., et al, 2003) Results based on data from the Biologically Resilient Adults in Neurological Studies (BRAiNS) cohort, a longitudinal study of aging and cognition at the University of Kentucky Alzheimer’s Disease Center (UK ADC), demonstrated that well established risk factors for dementia (i.e., age, education, family history of dementia, apolipoprotein ε-4 status) were also risk factors for transitions from normal cognition to transient MCI states. (Kryscio et al, 2006, Salazar et al, 2007) In this study we examined the association of apoe ε4 in disease incidence or progression from no cognitive impairment to mild cognitive impairment or from mild cognitive impairment to dementia and also examined the association of apoe ε4 in disease regression from mild cognitive impairment to no cognitive impairment after accounting for misclassification.

To test this hypothesis, data with subjects at different stages of dementia who were followed up for 1 year was used. Multi state model with misclassification was fit to the data to test the mentioned hypothesis. This idea is supported by other studies where the impact of misclassification of age-related macular degeneration(AMD) on baseline intensity and estimated effects of age, sex on incidence, progression and regression of AMD. (Ronald E. Gangnon et al, 2014) and employing hidden Markov model allowing for misclassification is well suited to analysis of health service databases to determine the transition probabilities between two states, and of misclassification and capture bias due to the fact that the quality and accuracy of the available information are not always optimal. (Nicola Bartolomeo et al, 2011)

It is well established that apoe-4 protein levels contribute to the risk of Alzheimer’s disease (Laws, Simon M., et al, 2003) and the association between apoe-4 and vascular dementia in a large population based cohort was examined for ten years and concluded that the apoe-4
allele is associated with increased risk of vascular dementia in a dose dependent fashion. (Chuang, Y-F., et al. 2010) There are several other studies which proved association of apoe-4 with Alzheimer’s disease but there is lack of evidence through studies showing effect of apoe-4 in disease progression and regression which is addressed in this study.

In this study multi state model with misclassification is used to account for the misclassification and found that after addressing for misclassification, apoe ε4 gene explains the disease progression from mild cognitive impairment to dementia. Apoe ε4 gene is found to have significant association with the transition of disease state from mild cognitive impairment to dementia which was masked when multi state model was fit to the data without addressing for misclassification.

There is no significant association of Apoe ε4 gene found with other transitions. In this study we also estimated if there is any association of low education with disease incidence, progression or regression with and without addressing for misclassification and found there is no significant association of low education. The unique ability of multi state modeling is portrayed in a well explainable manner in this study where the methodology behind the ability of the multi-state model to determine the misclassification probabilities of the disease state even in situation where gold standard is not available. As an improvement to the general multi state modeling, multi-state model with misclassification is fit to the data in order to address misclassification issues in an effective method. The misclassification probability of diagnosing normal subjects as mild cognitive impairment subjects is 18.3% and probability of misclassifying MCI subjects as normal is 1.57% and probability of misclassifying MCI subjects as dementia patients is 0.012% and probability of misclassifying Dementia patients as subjects with MCI is 0.28%. This proves that misclassification of the disease state has occurred during diagnosis.
Following up the subjects for only 1 year is one of the limitations of this study and following up for at least 5 years is recommended in order to capture the disease transitions involving intermittent stages. While this study focused on three state models for analysis, further analysis is recommended by considering amnestic MCI and non-amnestic MCI to be more specific in terms of disease transitions and also transition from dementia to MCI can be considered if the treatment involved in recovery is included in to the model. As recovery from dementia is being made possible due to recent advancements in treatment it could be considered in future research. In simulation study including different rates of misclassification in increasing fashion like 10%, 20% and 30% is recommended to determine the effect of misclassification at different rates to study the misclassification impact in more detailed perspective.
REFERENCES


He, F. (2015). Analysis of Multi-State Models with Mismeasured Covariates or Misclassified States.


Misclassification Can Explain Most Apparent Regression of Age-Related Macular Degeneration: Results From Multistate Models With Misclassification. Misclassification Explains Most AMD Regression. *Investigative ophthalmology & visual science, 55*(3), 1780-1786.


The msm package: Multistate modeling with R, version 0.6.4.


Ying Chen and Tuan D Pham, 2013, BioMedical Engineering OnLine 2013, 12(Suppl 1):S2

http://www.biomedical-engineering-online.com/content/12/S1/S2.

APPENDICES

The R code used for simulation is shown below.

#Simulation study without covariate

```r
sim_misdf = vector("list", 1000)
test.mis = vector("list", 1000)
test_mis.msm = vector("list", 1000)
ti.mis = vector("list", 1000)
test.msm = vector("list", 1000)
ti = vector("list", 1000)

for (i in 1:1000) {

sim_misdf[[i]] <- data.frame(subject = rep(1:500, rep(13,500)), time = rep(seq(0, 24, 2), 500))

qmatrix <- rbind(c(-0.11, 0.1, 0.01 ),
    c(0.05, -0.15, 0.1 ),
    c(0.02, 0.07, -0.09))

ematrix <- rbind(c(0, 0.3, 0.3 ),
    c(0.3, 0, 0.3 ),
    c(0.3, 0.3, 0 ))

test.mis[[i]] <- simmulti.msm(sim_misdf[[i]], qmatrix, ematrix = ematrix, death = FALSE)

Q1 <- rbind(c(1,1,1),c(1,1,1),c(1,1,1))

test_mis.msm[[i]] <- msm(obs~time, subject=subject, data=test.mis[[i]], qmatrix= Q1)

ti.mis[[i]]<- test_mis.msm[[i]]$Qmatrices

write.csv(ti.mis[[i]], file=paste0('timis', i, '.csv'), row.names=FALSE)

test.msm[[i]] <- msm(state~time, subject=subject, data=test.mis[[i]], qmatrix= Q1)

ti[[i]]<- test.msm[[i]]$Qmatrices
```
write.csv(ti[i], file=paste0('newsim', i, '.csv'), row.names=FALSE)
}

# simulation study including covariate

sim_cov_truedf = vector("list", 1000)
cov_true = vector("list", 1000)
cov_true msm = vector("list", 1000)
cov_true hr = vector("list", 1000)
cov_true_hr = vector("list", 1000)
cov_mis msm = vector("list", 1000)
cov_mis hr = vector("list", 1000)

for (i in 1:1000) {
  sim_cov_truedf[i] <- data.frame(subject = rep(1:500, rep(13, 500)),
                                 time = rep(seq(0, 24, 2), 500),
                                 x=rep(rbinom(500,1,0.5),rep(13,500)))
  qmatrix <- rbind(c(-0.11, 0.1, 0.01),
                   c(0.05, -0.15, 0.1),
                   c(0.02, 0.07, -0.09))
  ematrix <- rbind(c(0, 0.3, 0.3),
                   c(0.3, 0, 0.3),
                   c(0.3, 0.3, 0))

  cov_true[i] <- simmulti.msm(sim_cov_truedf[i], qmatrix, covariates=list(x = c(-0.3,-0.3,-0.3,-0.3,-0.3,-0.3)), ematrix = ematrix)
  Q <- rbind(c(1,1,1),c(1,1,1),c(1,1,1))
  cov_true msm[i] <- msm(state~time,subject=subject,data=cov_true[i],qmatrix= Q,covariates = ~x, method='BFGS')
  cov_true hr[i] <- hazard.msm(cov_true msm[i])
  write.csv(cov_true hr[i], file=paste0('hrcov_t', i, '.csv'), row.names=FALSE)

  cov_mis msm[i] <- msm(obs~time,subject=subject,data=cov_true[i],qmatrix= Q,covariates = ~x,method='BFGS')
  cov_mis hr[i] <- hazard.msm(cov_mis msm[i])
  write.csv(cov_mis hr[i], file=paste0('hrcov_mis', i, '.csv'), row.names=FALSE)
}