

ANALYSIS OF COMPLEX LIFE-HISTORY DATA AND VARIABLE SELECTION IN
SURVIVAL ANALYSIS UNDER INTERVAL CENSORING

By

Daewoo Pak

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2018

ABSTRACT

ANALYSIS OF COMPLEX LIFE-HISTORY DATA AND VARIABLE SELECTION IN SURVIVAL ANALYSIS UNDER INTERVAL CENSORING

By

Daewoo Pak

Event-time data are routinely obtained from longitudinal investigations to evaluate the time to onset and progression of chronic diseases. Such data are commonly referred to as disease life course data and have several nontrivial complications. In many longitudinal studies, disease life course data are subject to interval censoring, within-sampling-unit clustering, and multiplicity of event states. Moreover, because medical studies usually collect a large number of hypothesized risk factors for the disease, identifying pertinent determinants of the disease life course is of interest for disease prevention and prediction.

In Chapter 1, we describe a dental caries data set from a unique longitudinal study of young low-income urban African-American children. This data set motivates the three statistical methodologies developed respectively in Chapters 2-4.

In Chapter 2, we formulate a parametric frailty Markov model coupled with a likelihood-based inference to analyze the life course data with the complications of interval censoring, within-sampling-unit clustering and three event states. We also develop a Bayesian approach to predict observational-unit-level future transition probabilities. Such probabilities have implications for precision medicine.

Albeit its ease of computations, the proposed parametric method in Chapter 2 has some limitations. An obvious limitation is the restrictive parametric model imposed on the baseline intensities. Thus, in Chapter 3, we propose a similar model but with unspecified baseline intensities and develop a penalized spline method for the model estimation. Numerical experiments demonstrate that the proposed methods perform very well in finite samples with moderate sizes.

In Chapter 4, we propose a penalized variable selection method for interval censored data under the Cox proportional hazards model. It conducts a penalized nonparametric maximum likelihood

estimation with an adaptive Lasso penalty, which can be implemented through a penalized EM algorithm. The method is proved to have the oracle property. We also extend it to left truncated and interval censored data. Our simulation studies show that the method demonstrates the oracle property in samples of modest sizes and outperforms existing approaches in terms of many operating characteristics. The practical utility of the approach is illustrated using the mouth-level dental caries data introduced in Chapter 1.

Copyright by
DAEWOO PAK
2018

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my advisor and co-advisors, Dr. Chenxi Li, Dr. David Todem and Dr. Yuehua Cui, for their unflagging support of my Ph.D study, for their incredible patience, and for their invaluable assistance. I could not have imagined having better mentors through the doctoral program.

I would also like to thank the other members of my dissertation committees, Dr. Hyokyoung Hong and Dr. Lyudmila Sakhanenko, each of whom has provided insightful comments and feedback on this dissertation. The dissertation would not be accomplished well without their helps.

Thanks to my beloved parents, Chanmo Pak and Geumhwa Jang, and my proud brother, Sewoo Pak, for their endless support during the five years that it took me to earn a Ph.D. Because of them, I was able to keep moving on and overcome any obstacle.

Thanks to my fiancée, Jin Kyoung Park, who was standing by me all the time and providing encouraging words and support with her constant love. She always let me stay positive throughout entire doctoral progress.

I also want to thank Dr. Song Yang whom I was working with at National Institute of Health. His support was very valuable when deciding the direction of further research, and the discussions with him have always inspired me in many aspects about statistical methodologies.

Thanks also goes to the graduate school and the Department of Statistics and Probability for all financial support I recieved throughout my doctoral program such as the funds for conference travels and the dissertation completion fellowship.

I am also indebted to all members in the MSU Center for Statistical Training and Consulting. I worked there as a graduate consultant for two years and had excellent opportunities for learning how to interact with people under their guidance and support, and for meeting the professors and students from many disciplines, which will be the valuable assets in my life.

Thanks to my friends, Jiwoong Kim, Metin Eroğlu, Yi-Chen Zhang and Shawn Santo. I also would like to thank my office-mates, Robin Todd and Satabdi Saha. I am so fortunate to get to

know them.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 PARAMETRIC ANALYSIS OF CORRELATED AND INTERVAL CEN- SORED EVENT-HISTORY DATA	7
2.1 Introduction	7
2.2 Statistical methodologies	9
2.2.1 Parameter estimation and inference	12
2.2.2 Prediction of future transition probabilities	12
2.3 Simulation study	14
2.3.1 Simulation setting	14
2.3.2 Simulation results	15
2.4 Application	17
2.5 Discussion	21
CHAPTER 3 SEMIPARAMETRIC ANALYSIS OF CORRELATED AND INTERVAL CENSORED EVENT-HISTORY DATA	24
3.1 Introduction	24
3.2 Statistical methodologies	27
3.2.1 A penalized likelihood estimation	28
3.2.2 Statistical inference	33
3.2.3 Prediction of future transition probabilities	33
3.3 Simulation study	35
3.3.1 Simulation setting	35
3.3.2 Simulation results	35
3.4 Application	40
3.5 Discussion	45
CHAPTER 4 VARIABLE SELECTION IN THE PROPORTIONAL HAZARDS MODEL FOR INTERVAL CENSORED DATA	48
4.1 Introduction	48
4.2 Statistical methodologies	50
4.2.1 Variable selection	51
4.2.2 Covariance matrix of the adaptive Lasso estimator	54
4.2.3 Thresholding parameter tuning	54
4.3 Asymptotic properties	55
4.4 Extension to left truncated and interval censored data	61
4.5 Simulation study	62

4.5.1	Simulation setting	62
4.5.2	Simulation results	63
4.6	Application	68
4.7	Discussion	70
APPENDICES		73
APPENDIX A	74
APPENDIX B	77
APPENDIX C	80
BIBLIOGRAPHY		82

LIST OF TABLES

Table 1.1: The classification rule for three caries states based on a two-digit number system in the International Caries Detection and Assessment System	5
Table 2.1: Simulation results for estimating the Weibull and variance parameters	15
Table 2.2: Simulation results for estimating the regression parameters β	16
Table 2.3: The DDHP data analysis using the univariate random effect model	18
Table 2.4: The Wald test statistics (p-value) for the three types of symmetry in caries transition intensity	21
Table 3.1: The simulation results for estimating β_{01}	36
Table 3.2: The simulation results for estimating β_{12}	36
Table 3.3: The proportion of the Monte Carlo iterations where the 95% prediction interval contains the true future transition probability for the last tooth of the last subject at $V_m M_m + t^*$ where $t^* = 0.1, 0.3, 0.5, 1, 1.5$	40
Table 3.4: The regression parameter estimates for the saturated model based on the DDHP data	41
Table 3.5: The Wald test statistics (p-value) for the three types of symmetry in caries transition intensity stratified by gender and transition type	43
Table 3.6: The detailed comparisons associated with the significant non-symmetries in caries transition intensity found in Table 3.5	44
Table 4.1: The variable selection percentages of our method (Ours), Zhang & Lu (2007)'s method based on mid-point imputation (ZL), and Wu & Cook (2015)'s method (WC) in the untruncated case	64
Table 4.2: Average numbers of correct and incorrect zero coefficients and mean squared errors (MSE) of the coefficient estimators for our method (Ours), Zhang & Lu (2007)'s method based on mid-point imputation (ZL), and Wu & Cook (2015)'s method (WC) in the untruncated case. In the parentheses are the ideal numbers of correct and incorrect zero coefficients	65

Table 4.3:	Simulation results on the estimation of the non-zero coefficients in the untruncated case. The oracle method is the unpenalized nonparametric maximum likelihood estimation with only the covariates whose coefficients are non-zero, i.e., X_1, X_2, X_9 and X_{10} . The standard errors and the coverage of the Normal-based confidence intervals for Wu & Cook (2015)'s method were not computed due to the bootstrap, suggested by Wu & Cook (2015) for computing the standard errors, being time-consuming	66
Table 4.4:	The candidate covariates considered in the analysis of DDHP data	69
Table 4.5:	The DDHP data analysis results. NPMLE is the coefficient estimate from the nonparametric maximum likelihood estimation, and Adaptive Lasso is the coefficient estimate from the proposed shrinkage method. Standard errors are given in the parentheses	71
Table A.1:	The regression parameter estimates for the saturated model based on the DDHP data	75
Table A.2:	The Wald test statistics (p-value) for the three types of symmetry in caries transition intensity stratified by gender and transition type	75
Table A.3:	The detailed comparisons associated with the significant non-symmetries in caries transition intensity found in Table A.2	76
Table B.1:	The variable selection percentages of our method in the truncated case	77
Table B.2:	Average numbers of correct and incorrect zero coefficients and mean squared errors (MSE) of the coefficient estimators for our method in the truncated case. In the parentheses are the ideal numbers of correct and incorrect zero coefficients	77
Table B.3:	Simulation results on the estimation of the non-zero coefficients in the truncated case. The oracle method is the unpenalized nonparametric maximum likelihood estimation with only the covariates whose coefficients are non-zero, i.e., X_1, X_2, X_9 and X_{10}	78
Table C.1:	Questionnaire items for the family-level measures (OHSE, KBU, KCOH)	80
Table C.2:	Questionnaire items for the family-level measures (OHF, SUPPORT)	81

LIST OF FIGURES

Figure 2.1: The progressive three states: state 0 for normal condition, state 1 for early caries and state 2 for advanced caries	10
Figure 2.2: Molars in a child's mouth clockwise starting from the upper-right teeth (numbers starting 5) from the dentist's view. The child's right side corresponds to the tooth chart's left side. The numbers ending 4 and 5 correspond to first and second molars respectively	16
Figure 2.3: The predicted transition probabilities from state 0 to state 1 and to state 2 for the upper right first molar (A) and from state 1 to state 2 for the upper right second molar (B) for a boy in the DDHP dataset	22
Figure 3.1: The average estimates of $h_{01}^{(0)}(t)$ and $h_{12}^{(0)}(t)$, their average 95% pointwise confidence intervals and their mean integrated squared errors (MISE) across the 1000 Monte Carlo samples	37
Figure 3.2: The comparison between the empirical standard errors and the average theoretical standard errors of the baseline intensity estimators	38
Figure 3.3: The empirical coverage probabilities of the 95% pointwise confidence intervals for the baseline intensities	39
Figure 3.4: The future transition probabilities to every possible state out of the caries state at the last visit for the upper right first molar (A) and the upper right second molar (B)	46
Figure 4.1: Normal Q-Q plots of the proposed estimators for the non-zero coefficients in the untruncated case	67
Figure B.1: Normal Q-Q plots of the proposed estimators for the non-zero coefficients in the truncated case	79

CHAPTER 1

INTRODUCTION

In medical research involving chronic diseases, it is often the practice to use a small number of predefined clinical states, which are exhaustive and mutually exclusive contemporaneously, to represent the spectrum of the disease. These classifications coupled with longitudinal investigations can be used to study the transition of subjects across the spectrum of the disease, which is paramount in understanding the life course of the condition. A good example is the joint damage in psoriatic arthritis, where researchers typically classify subject participants four states ranging from normal to severe states requiring surgery, in view of the reading of X-ray images (see, for example, Sutradhar & Cook, 2008). Many chronic diseases manifest themselves at several regions of the body, resulting in endpoints that are potentially correlated even after conditioning on observed covariates. In diabetic retinopathy, for instance, retinal lesions in the two eyes within a body develop dependently, due to the two eyes being exposed to common risk factors, including genetic variants, environmental factors, control of glucose in the blood (management of diabetes), nutritional status, etc., which may not be fully measured. Any analysis that ignores this correlation is apt to yield biased inferences for parameters describing the life-history of the disease, if multiple endpoints are recorded cross-sectionally on the same subject. Finally, in many longitudinal investigations, even in studies that aim at describing the life course of the disease, data are collected intermittently, yielding interval censored data. For such data, the exact time of the transition from one state to another is known up to being bracketed by two assessment times.

A good example of correlated and interval censored life-history data is in a life-course study of tooth decay. Tooth decay medically known as dental caries is the most prevalent chronic disease in childhood, despite being largely preventable (Selwitz et al., 2007). It is estimated that tooth decay is five times more prevalent than asthma in childhood and seven times more prevalent than hay fever (US Department of Health and Human Services, 2000; Benjamin, 2010). Its diagnosis is typically performed at the tooth-surface level where both the shape and the depth of any potential

lesions are scored on an ordinal scale representing the stages of the disease process, ranging from sound (no lesion) to lesions into the dental pulp. These detailed tooth-surface level assessments have enabled epidemiological and clinical studies to describe the intra-oral susceptibility of dental caries (Kandelman & Lewis, 1988; Batchelor & Sheiham, 2004; Zhang et al., 2011). While tooth and tooth-surface susceptibilities to cavity are well understood, the intra-oral life course of dental caries remains relatively understudied. It is well known that caries susceptibility differs between the maxilla and mandible and also across individual teeth and their surfaces, but little is known about the timing and the transition of the disease for these intra-oral characteristics. Understanding the intra-oral caries life course may provide important evidence on the timing and the transition of the disease, critical for caries prevention and treatment.

The study of intra-oral caries life course involves a longitudinal association due to the same tooth or tooth surface being observed over time. This association is of primary importance and used to model the transition across caries states. A basic statistical approach is to use a multistate survival model where the outcome of interest is the time to transition across various caries states. Due to intermittent dental assessments, however, the time to disease transition is often interval censored in many medical researches. Interval censoring in multistate survival models gives rise to a new difficulty, which does not exist in classical survival models, in that several paths are possible for transitioning from one state to another between two sequential assessments. Unlike the longitudinal association, the intra-oral association is essentially a nuisance term that needs to be accommodated to yield valid inferences for the model parameters describing all tooth-level progressions. For example, when the transition durations across the caries states are to be compared for several teeth or tooth surfaces of the same child, a correlated data model is needed to accommodate the inherent intra-oral correlation.

A unique database that provides detailed information for the study of intra-oral caries life course is the Detroit Dental Health Project (DDHP). This is a longitudinal study designed to understand the oral health of low-income African-American children (0-5 years at baseline) and their main caregivers (14+ years) residing in the city of Detroit, Michigan. Although it is well established that

low income and minority children exhibit the most adverse oral health outcomes (Vargas et al., 1998; Mouradian et al., 2000; Edelstein & Chinn, 2009), little is known about the intra-oral life course of caries in this underserved population. Among other information, this study comprises longitudinal investigations involving tooth and tooth-surface level information on caries life course which can shed light on important research questions such as: which teeth or tooth surfaces transition faster from sound to noncavitated lesions and ultimately to cavitated lesions? Another interesting intra-oral question is whether there are left-right or upper-lower symmetries in the development of tooth decay.

In addition to the intra-oral investigations, the DDHP also provides information on a relatively large number of factors (child-, household- and community-levels) that are known to be associated with caries prevalence. Because a high proportion of children (81% in the DDHP data) did not exhibit any caries activity at the first examination (Wave 1), it may of interest to determine which of these variables, albeit their associations with caries prevalence, are associated with the duration of transition from no caries to some caries. The study on the duration of caries transition or caries development is ultimately of interest in dental practice as the information gained can provide guidance on the timing of dental screening and treatment, which is especially critical for low income African American children who have fewer dental visits and fewer protective sealants than any other socio-economic and racial groups (Mouradian et al., 2000; Edelstein & Chinn, 2009).

The DDHP data contain longitudinal investigations of caries on all deciduous teeth in 1021 children during the three waves (2002-2003, 2004-2005 and 2007). The caries status was assessed at tooth-surface level and its result was coded as a two-digit number according to the International Caries Detection and Assessment System (ICDAS). The first digit identifies restorations, sealants or other conditions: 0 for surface neither restored nor sealed (sound), 1 for partial sealant, 2 for full sealant, 3 for tooth colored restoration, 4 for amalgam restoration, 5 for stainless steel crown, 6 for other various restorations (Full porcelain, PFM crown, etc.) 7 for lost restoration, 8 for temporary restoration, and 9 for surface missing due to carious or non-carious related reasons. The second digit codes the caries severity: 0 for sound surface, 1-2 for noncavitated lesions, 3-6 for cavitated

lesions with various severity (Enamel cavity, Extensive cavity, etc.), 7 for tooth missing because of caries, 8 for tooth missing for reasons other than caries, and 9 for an unerupted tooth. Since all teeth were assessed simultaneously when a child had a study visit, the examination times are the same for all teeth within a child.

The statistical modeling and intra-oral and inter-oral inferences for the DDHP data are inherently complex because the data have several theoretical complications inherent in the study of intra-oral caries life course. The ICDAS scoring system generates multiple caries states over time (up to three times in DDHP study). Moreover, intra-oral clustering of teeth or tooth-surfaces with an unusual spatial configuration poses another level of difficulty. And, finally, the exact 'failure' time to disease transition is not directly observed but it is only known up to occur between the periodic assessments, yielding interval censored survival endpoints. Such interval censoring is referred as *mixed case interval censoring* because it is generated with varying numbers of monitoring times among children, which is the most common type of interval censoring (Schick & Yu, 2000). Methodologically, there have been several works in the literature which examine related issues and complications. To our best knowledge, we are not aware of any robust statistical methodology that addresses all these complications simultaneously.

In Chapter 2 and Chapter 3, we study how to analyze intra-oral caries life course data at the tooth-level. Thus, we reformulate the DDHP tooth-surface level data to the tooth-level data by classifying all possible two-digit numbers into three states: 0 for normal condition, 1 for early caries (noncavitated), and 2 for advanced caries (cavitated), as following Table 1.1. After then, the most severe caries status of all surface of a tooth were taken to represent the caries status of the tooth. An unerupted tooth is considered as normal condition (coded 99), and a lost tooth due to caries (coded 97) is treated as being in the advanced caries state. A deciduous tooth that is missing due to replacement by a permanent tooth is assumed to be right-censored at the last visit time when the primary tooth was seen. In these two chapters, motivated by the DDHP data, we propose a multistate Markov frailty models to analyze interval censored tooth-level life-history data of caries in the deciduous dentition. Specifically, in Chapter 2, we propose a multistate frailty Markov model

Table 1.1: The classification rule for three caries states based on a two-digit number system in the International Caries Detection and Assessment System

A two-digit number system	caries state
00, 10, 20, 99	0
01, 02, 11, 12, 21, 22	1
03, 04, 05, 06, 13, 14, 15, 16, 23, 24, 25, 26, 30, 31, 32, 33, 34, 35, 36, 40, 41, 42, 43, 44, 45, 46, 50, 51, 52, 53, 54, 55, 56, 60, 61, 62, 63, 64, 65, 66, 70, 71, 72, 73, 74, 75, 76, 80, 81, 82, 83, 84, 85, 86, 97	2
96, 98	right-censored

coupled with a likelihood-based inference to analyze tooth-level life course data in caries research. In Chapter 3, the less restrictive formulation that uses spline functions to control the fitting of the baseline intensities in the multistate frailty model framework is described. In the analyses of Chapter 2 and Chapter 3, we focus on any spatial symmetry in the mouth with respect to the life course of dental caries and whether the same type of tooth has a similar decay process in male and female, using the tooth location and gender as covariates in the models. These analyses prove challenging because of nontrivial complications including intra-oral clustering, interval censoring, multiplicity of caries states, and computational complexities. Besides fitting appropriate models, the methodologies for predicting individuals' future transition probabilities are also developed. It is helpful to know the probability of a tooth transitioning to a more severe caries state at a future time point for the timing of dentist visit and disease-modifying intervention. Because there is usually heterogeneity in transition rates between different individuals that are not captured by measured covariates, the future trajectory of a subject's life-history process would be predicted more accurately if the future transition probabilities are predicted conditional on the subject's up-to-date life course data.

In Chapter 4, we introduce a new variable selection approach for the time-to-event data subject to interval censoring and left truncation. Our interest lies in identifying which among important risk factors are associated with increased risk of developing dental caries from DDHP data. Guided by the conceptual model in Fisher-Owens et al. (2007), potential variables to examine in our analysis of the DDHP data will include child-, household- and community-level characteristics. Child-level variables will include demographic variables (e.g., age) and known risk factors (e.g., weight, brushing frequency). Household-level variables will include children's oral health-related measures (e.g., oral health self-efficacy, usage of water filter/purifier on kitchen tap) and caregiver's status (e.g., education level, parenting stress score). Community-level characteristics will include the number of dentists, full-scale grocery stores and churches in predefined neighborhoods. A list of the variables of interest for the analysis will be discussed in detail in Chapter 4. Instead of tooth- or tooth-surface level caries life course data, in this chapter, we consider mouth-level caries life course data, aiming to early detection of carious lesions among the surfaces of all teeth in a mouth. Thus, we create the mouth-level data from DDHP data by taking the time intervals within which first noncavitated/cavitated lesion is observed in any surface. The proposed variable selection is conducted under the Adaptive Lasso framework, and the model estimation is through a penalized Expectation–Maximization algorithm (EM) based on the penalized nonparametric maximum likelihood approach for the baseline hazard function and the regression parameters. The consistency of the regression parameter estimator is also studied along with its sparsity.

CHAPTER 2

PARAMETRIC ANALYSIS OF CORRELATED AND INTERVAL CENSORED EVENT-HISTORY DATA

2.1 Introduction

Several survival analysis methods, which can accommodate some of the complications of the DDHP data such as interval censoring and intra-oral correlation, have been developed and applied to caries history data. For example, Tommi et al. (2000) and Komárek et al. (2005) used nonparametric and semiparametric Bayesian intensity models to analyze clustered and interval censored time-to-caries data. Komárek & Lesaffre (2008) proposed a Bayesian accelerated failure time model for clustered doubly-interval censored data for the time to caries of permanent first molars. Besides these methods tailored for caries data, there are other survival analysis methods for multiple (correlated) interval censored survival time data (see e.g., Goggins & Finkelstein, 2000; Bogaerts et al., 2002; Kim & Xue, 2002; Kor et al., 2013) that can be applied to caries research. These existing models, however, can only handle time-to-event data where the event is a binary endpoint, e.g. the incidence of caries defined according to a dichotomized version of the ICDAS scoring criterion. Although important, this approach only provides partial information on the complete time course of dental caries. To our best knowledge, only few papers (see e.g., Sutradhar & Cook, 2008; Joly et al., 2012) studied the class of multistate models for clustered life-history data under interval censoring. However, neither of Sutradhar & Cook (2008) and Joly et al. (2012) considered cluster-level or observation-unit-level covariates in their models, and they did not discuss how to predict unit-level future transition. Simulations corroborating their methods are also lacking. From a methodological viewpoint, this chapter aims at developing a statistical method that simultaneously addresses the complications of the intra-oral caries life course data and the limitations of existing works.

Intra-oral caries life course data are complex for statistical analysis. Firstly, dental caries is usually classified into multiple states according to its severity. Secondly, the caries state of

each tooth is determined at intermittent dental examinations, which leads to interval censoring of the transition time between states. Thirdly, the progressions of caries at different teeth within a mouth are correlated. To address these complexities simultaneously in the analysis, we consider a parametric estimation of multistate models for clustered interval censored data coupled with a conditionally progressive Markov model. The approach is challenging in many points. Even for uncorrelated data, non-homogeneous Markov multistate models are hard to fit with interval censored observations and general software is not readily available (Cook & Lawless, 2014). The practical novelty of the methodology relies essentially on the application of multistate models to analyze complex caries life course data. This methodological approach will advance knowledge in dental caries research by providing a more detailed information on the caries life course, both at tooth and tooth-surface levels. Additionally, the effects of potential determinants of early childhood caries transition, e.g. gender, tooth types and sites on teeth, will be easily evaluated using regression techniques through the transition intensities. And finally, the proposed model will be able to predict future caries outcomes for a tooth or tooth surface given its current disease status.

The proposed statistical methods also have general applications to modeling chronic diseases involving paired or multiple organs in cohort studies where the diseases status is measured intermittently. Besides dental caries, popular examples includes sacroiliac joint damage in psoriatic arthritis and diabetic retinopathy as discussed in Chapter 1. For these cohort studies of chronic diseases, the proposed models and the associated estimation and inferences can be used to estimate the rate of disease onset and progression, to identify demographic, genetic and environmental risk factors, and to assess therapeutic interventions; and predict future disease status at the individual or population level. Despite a wide range of applications of the proposed statistical methods, their application to dental caries is unique in view of the intral-oral clustering with an unusual spatial configuration.

The rest of the chapter is organized as follows. In Section 2.2, we introduce the three-state non-homogeneous Markov frailty model and derive the corresponding likelihood of tooth-level caries life course data. Section 2.2.1 discusses the likelihood-based estimation and inference. Section

2.2.2 presents the Bayesian prediction method. In Section 2.3, we perform a simulation study to investigate the finite sample performance of the likelihood-based inferential approach, followed by the application of the proposed methodology to the DDHP data in Section 4.6. Section 2.5 gives some concluding remarks.

2.2 Statistical methodologies

We model the caries transition profile of n specific primary teeth ($n = 20$ if all the primary teeth are modeled). The time scale is chosen to be age, so birth is the time origin. Let m be the number of children. The study visit times of the i -th child are denoted by $V_{i1}, V_{i2}, \dots, V_{iM_i}$, where M_i is the number of visits for the i -th child and can vary across children. Let $A_{ij}(t)$ be the caries state of the j -th tooth for the i -th subject ($i = 1, \dots, m; j = 1, \dots, n$) at time t ($t \geq 0$), observed only at the finite time points. Our basic model assumes that $A_{ij}(t)$ follows a conditionally progressive Markov model with the three states representing the tooth-level caries spectrum because a direct transition from state 0 to 2 is biologically impossible and the only plausible reverse transition, 1 to 0 (Ismail et al., 2011), is rare in reality (3.4% of all the transitions on primary molars in the DDHP data). Also note that any continuous intensity model for $A_{ij}(t)$ that allows reverse transitions is non-identifiable under interval censoring. Therefore, we consider tooth decay a progressive process from normal condition to early caries and finally to advanced caries as illustrated by Figure 2.1. The corresponding observed caries states of the j -th tooth are denoted by $K_{i1}^{(j)}, K_{i2}^{(j)}, \dots, K_{iM_i}^{(j)}$ respectively. Define $\mathbf{V}_i \equiv (V_{i1}, \dots, V_{iM_i})$ and $\mathbf{K}_{ij} \equiv (K_{i1}^{(j)}, \dots, K_{iM_i}^{(j)})$. Let \mathbf{X}_{ij} be the baseline covariate vector of interest for the j -th tooth of the i -th child, e.g., gender, tooth location, left-handed or right-handed, number of dentists nearby, and socioeconomic status. Define $\mathbf{Y}_{ij} \equiv (\mathbf{V}_i, \mathbf{K}_{ij}, \mathbf{X}_{ij})$ and $\mathbf{Y}_i \equiv (\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in})$. Then an interval censored tooth-level caries-history data set consists of $\mathbf{Y} \equiv (\mathbf{Y}_1, \dots, \mathbf{Y}_m)$.

Caries life course data of the teeth within a child are correlated, so are the transition times between different pairs of states in a tooth. We use a subject-level random effect on transition intensities modulated by transition types to capture these two kinds of correlation. Specifically, we

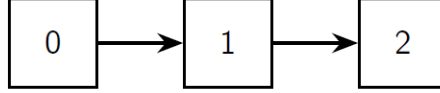


Figure 2.1: The progressive three states: state 0 for normal condition, state 1 for early caries and state 2 for advanced caries

account for the intra-oral correlation between $A_{ij}(t)$'s ($j = 1, \dots, n$) by positing a common random effect u_i in their transition intensities, and assume that

$$\begin{aligned}
 h_{ab}(t|\mathbf{X}_{ij}, u_i) &= \lim_{\Delta t \rightarrow 0} \Pr(A_{ij}(t^- + \Delta t) = b | A_{ij}(u), 0 \leq u < t, A_{ij}(t^-) = a, \mathbf{X}_{ij}, u_i) \\
 &= \lim_{\Delta t \rightarrow 0} \Pr(A_{ij}(t^- + \Delta t) = b | A_{ij}(t^-) = a, \mathbf{X}_{ij}, u_i),
 \end{aligned}$$

where $(a, b) \in \{(0, 1), (1, 2)\}$. Furthermore, we postulate the following multiplicative intensity model for $h_{01}(t|\mathbf{X}_{ij}, u_i)$ and $h_{12}(t|\mathbf{X}_{ij}, u_i)$,

$$h_{01}(t|\mathbf{X}_{ij}, u_i) = h_{01}^{(0)}(t) \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_{01} + \sigma_{01} u_i) \quad (2.1a)$$

$$h_{12}(t|\mathbf{X}_{ij}, u_i) = h_{12}^{(0)}(t) \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_{12} + \sigma_{12} u_i) \quad (2.1b)$$

where $\boldsymbol{\beta}_{01}$ and $\boldsymbol{\beta}_{12}$ are the regression coefficient vectors of \mathbf{X}_{ij} , u_i 's are assumed to be i.i.d. from the standard normal distribution, σ_{10} and σ_{12} are the standard deviations of the log frailties corresponding to the two transition types respectively, and $h_{01}^{(0)}(t)$ and $h_{12}^{(0)}(t)$ are the baseline intensity functions. In the parametric modeling framework, we can use the Weibull distribution to model the baseline intensities, namely, $h_{01}^{(0)}(t) = k_{01} r_{01} t^{r_{01}-1}$ and $h_{12}^{(0)}(t) = k_{12} r_{12} t^{r_{12}-1}$ with unknown positive parameters $(k_{01}, r_{01}, k_{12}, r_{12})$.

Denote the vector of all parameters by $\boldsymbol{\theta} \equiv (\log k_{01}, \log r_{01}, \log k_{12}, \log r_{12}, \log \sigma_{01}, \log \sigma_{12}, \boldsymbol{\beta}_{01}, \boldsymbol{\beta}_{12})$. Let $f(\cdot)$ be a generic notation of density/mass function and $f(\cdot|\cdot)$ be a generic notation of conditional density/mass function. Under the assumption that the tooth decay process is Markovian and \mathbf{Y}_i 's are independent across children, the likelihood function of the data equals, up to a multiplicative constant not dependent on $\boldsymbol{\theta}$,

$$L(\boldsymbol{\theta}; \mathbf{Y}) = \prod_{i=1}^m \int_{-\infty}^{\infty} \prod_{j=1}^n f(\mathbf{K}_{ij} | \boldsymbol{\theta}, \mathbf{V}_i, \mathbf{X}_{ij}, u_i) \phi(u_i) du_i, \quad (2.2)$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. Under the Markov assumption, $f(\mathbf{K}_{ij}|\boldsymbol{\theta}, \mathbf{V}_i, \mathbf{X}_{ij}, u_i)$ can be expressed as the product of transition probabilities, namely,

$$f(\mathbf{K}_{ij}|\boldsymbol{\theta}, \mathbf{V}_i, \mathbf{X}_{ij}, u_i) = \prod_{q=1}^{M_i} f(A_{ij}(V_{iq}) = K_{iq}^{(j)} | A_{ij}(V_{i,q-1}) = K_{i,q-1}^{(j)}, \mathbf{X}_{ij}, u_i),$$

where $V_{i0} = 0$, $K_{i0}^{(j)} = 0$ and $A_{ij}(0) = 0$ according to the caries classification scheme in Chapter 1.

Thus the log-likelihood function, $l(\boldsymbol{\theta}; \mathbf{Y})$, is simply,

$$l(\boldsymbol{\theta}; \mathbf{Y}) = \sum_{i=1}^m \log \int_{-\infty}^{\infty} \prod_{j=1}^n \left\{ \prod_{q=1}^{M_i} f(A_{ij}(V_{iq}) = K_{iq}^{(j)} | A_{ij}(V_{i,q-1}) = K_{i,q-1}^{(j)}, \mathbf{X}_{ij}, u_i) \right\} \phi(u_i) du_i. \quad (2.3)$$

To simplify the notations, we write the transition probability $f(A_{ij}(V_{iq}) = K_{iq}^{(j)} | A_{ij}(V_{i,q-1}) = K_{i,q-1}^{(j)}, \mathbf{X}_{ij}, u_i)$ on the right-hand side of (2.3) as a function $p_{K_{i,q-1}^{(j)} K_{iq}^{(j)}}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \boldsymbol{\theta})$, which can be computed using $h_{01}(t|\mathbf{X}_{ij}, u_i)$ and $h_{12}(t|\mathbf{X}_{ij}, u_i)$ as follows,

$$p_{00}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = S_{01}(V_{i,q-1}, V_{iq} | \mathbf{X}_{ij}, u_i) \quad (2.4a)$$

$$p_{01}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = \int_{V_{i,q-1}}^{V_{iq}} S_{01}(V_{i,q-1}, r | \mathbf{X}_{ij}, u_i) h_{01}(r | \mathbf{X}_{ij}, u_i) S_{12}(r, V_{iq} | \mathbf{X}_{ij}, u_i) dr \quad (2.4b)$$

$$\begin{aligned} p_{02}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) \\ &= \int_{V_{i,q-1}}^{V_{iq}} \int_{r_1}^{V_{iq}} S_{01}(V_{i,q-1}, r_1 | \mathbf{X}_{ij}, u_i) \cdot h_{01}(r_1 | \mathbf{X}_{ij}, u_i) S_{12}(r_1, r_2 | \mathbf{X}_{ij}, u_i) h_{12}(r_2 | \mathbf{X}_{ij}, u_i) dr_2 dr_1 \\ &= \int_{V_{i,q-1}}^{V_{iq}} S_{01}(V_{i,q-1}, r | \mathbf{X}_{ij}, u_i) h_{01}(r | \mathbf{X}_{ij}, u_i) \{1 - S_{12}(r, V_{iq} | \mathbf{X}_{ij}, u_i)\} dr \end{aligned} \quad (2.4c)$$

$$p_{11}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = S_{12}(V_{i,q-1}, V_{iq} | \mathbf{X}_{ij}, u_i) \quad (2.4d)$$

$$\begin{aligned}
p_{12}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \xi, \sigma) &= \int_{V_{i,q-1}}^{V_{iq}} S_{12}(V_{i,q-1}, r | \mathbf{X}_{ij}, \mathbf{z}_i) h_{12}(r | \mathbf{X}_{ij}, u_i) dr \\
&= 1 - S_{12}(V_{i,q-1}, V_{iq} | \mathbf{X}_{ij}, u_i)
\end{aligned} \tag{2.4e}$$

$$p_{22}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, u_i, \xi, \sigma) = 1 \tag{2.4f}$$

where $S_{ab}(x, y | \mathbf{X}_{ij}, u_i) \equiv \exp\{-\int_x^y h_{ab}(t | \mathbf{X}_{ij}, u_i) dt\}$. The integrals involved in p_{01} and p_{02} do not have a closed form and thus are numerically approximated using the Gauss-Jacobi quadrature.

2.2.1 Parameter estimation and inference

We maximize the log likelihood of the observed data (2.3) to get the maximum likelihood estimate (MLE) $\hat{\theta}$. The Gauss-Jacobi quadrature is used to compute the integrals over finite intervals in (2.4b, 2.4c) that do not have a closed form, and the Gauss-Hermite quadrature is used to compute the integrals with respect to the random effect in (2.3). In the Gauss-Jacobi quadrature, the number of quadrature points is determined by

$$\# \text{ of quadrature points} = \min \left(25, 50 \times \frac{\text{length}}{\text{max.length}} \right),$$

where ‘length’ is the length of the integration interval and ‘max.length’ is the maximum length among all the integration intervals. In the Gauss-Hermite quadrature, 25 quadrature points are used to get a numerical integral value. The optimization is carried out by the R function `nlminb`. The observed information matrix for θ , $I_{\text{obs}}(\hat{\theta})$, is founded by the summation of the cross products of the subject-level score function. The inference about θ is based on the asymptotic distribution of $\hat{\theta}$, approximated by $N(\theta, I_{\text{obs}}(\hat{\theta})^{-1})$.

2.2.2 Prediction of future transition probabilities

Using all the children’s caries life-history data, the transition probability of the j -th tooth in the i -th child from the last observed state is expressed by,

$$p_{K_{iM_i}^{(j)} q}(V_{iM_i}, t, \mathbf{X}_{ij}, u_i, \theta), \tag{2.5}$$

where $t > V_{iM_i}$ and $q \geq K_{iM_i}^{(j)}$. Given the data, θ and u_i are the only two unknown quantities in (2.5). So a natural estimator of (2.5) can be obtained by substituting $\hat{\theta}$ for θ and \hat{u}_i for u_i , where \hat{u}_i is the empirical Bayes estimator of u_i under the zero-one loss obtained by maximizing $f(\mathbf{Y}_i|u_i, \hat{\theta})\phi(u_i)$. However, it is difficult to construct a prediction interval for (2.5) using a frequentist approach.

We can also predict the transition probabilities using a Bayesian approach. Specifically, we assume θ has a flat prior distribution so that we can predict (2.5) by its posterior mean or median obtained from simulating (u_i, θ) according to their joint posterior distribution. The credible interval for (2.5) can also be obtained from the simulation. We derive the posterior distribution of (u_i, θ) as follows,

$$\begin{aligned}
f(u_i, \theta|\mathbf{Y}) &= \int \cdots \int f(u_1, \dots, u_m, \theta|\mathbf{Y}) du_1 \cdots du_{i-1} du_{i+1} \cdots du_m \\
&\propto \frac{1}{f(\mathbf{Y})} \int \cdots \int f(\mathbf{Y}|u_1, \dots, u_m, \theta) \prod_{j=1, j \neq i}^m \phi(u_j) du_1 \cdots du_{i-1} du_{i+1} \cdots du_m \\
&= \frac{f(\mathbf{Y}_i|u_i, \theta)\phi(u_i)}{f(\mathbf{Y})} \prod_{j=1}^m \int f(\mathbf{Y}_j|u_j, \theta)\phi(u_j) du_j \\
&= \frac{f(\mathbf{Y}_i|u_i, \theta)\phi(u_i)}{\int f(\mathbf{Y}_i|u_i, \theta)\phi(u_i) du_i} \times \frac{\prod_{j=1}^m \int f(\mathbf{Y}_j|u_j, \theta)\phi(u_j) du_j}{f(\mathbf{Y})} \\
&\propto \frac{f(\mathbf{Y}_i|u_i, \theta)\phi(u_i)}{\int f(\mathbf{Y}_i|u_i, \theta)\phi(u_i) du_i} \times f(\theta|\mathbf{Y}), \tag{2.6}
\end{aligned}$$

which implies $f(u_i|\mathbf{Y}, \theta) \propto f(\mathbf{Y}_i|u_i, \theta)\phi(u_i)$. When m is large enough, we can approximate $f(\theta|\mathbf{Y})$ by the multivariate normal distribution, $N(\hat{\theta}, I_{obs}(\hat{\theta})^{-1})$. Given the observed data of i -th child along with the child's random effect and the parameters, i.e., $(\mathbf{Y}_i, u_i, \theta)$, $f(\mathbf{Y}_i|u_i, \theta)$ can be computed using the transition probabilities (2.4a - 2.4f). Therefore, we can generate a posterior realization, (u_i^*, θ^*) , from $f(u_i, \theta|\mathbf{Y})$ in the following two steps:

- (a) generate θ^* from the multivariate normal distribution $N(\hat{\theta}, I_{obs}(\hat{\theta})^{-1})$
- (b) generate u_i^* from $f(\mathbf{Y}_i|u_i, \theta^*)\phi(u_i)$ using the Metropolis-Hastings algorithm with the burn-in period 5000 and the proposal distribution being a normal distribution with appropriate variance allowing the acceptance ratio to be between 20% and 40% in general.

From a large sample of (u_i, θ) 's simulated from $f(u_i, \theta | \mathbf{Y})$, we can get a large sample of $p_{K_{iM_i}^{(j)} q}(V_{iM_i}, t, \mathbf{X}_{ij}, u_i, \theta)$'s. Thus, the posterior mean or median of (2.5) can be approximated by the sample mean or median, and a 95% credible interval for (2.5) can be approximated by the interval between the 2.5% and 97.5% sample quantiles of the $p_{K_{iM_i}^{(j)} q}(V_{iM_i}, t, \mathbf{X}_{ij}, u_i, \theta)$'s.

2.3 Simulation study

2.3.1 Simulation setting

We performed a Monte Carlo simulation study to investigate the finite-sample performance of the likelihood-based inferential approaches for θ . Two different sets of sample sizes were considered: 1) $m = 200$ and $n = 4$, and 2) $m = 400$ and $n = 4$. With the assumption of the Weibull intensities, the vector of Weibull and variance parameters in (2.1) was chosen to be $(r_{01}, k_{01}, r_{12}, k_{12}, \sigma_{12}, \sigma_{12}) = (1.2, 0.1, 0.9, 0.2, 1, 1.2)$. After taking the logarithm of these values, we get $(\log r_{01}, \log k_{01}, \log r_{12}, \log k_{12}, \log \sigma_{01}, \log \sigma_{12}) = (0.18, -2.30, -0.11, -1.61, 0, 0.18)$. As covariates, three dummy variables X_1 , X_2 and X_3 indicating the four tooth locations and a standard normal random variable X_4 were considered. For the dummy variables, the location of the first tooth was set as the reference. The continuous covariate was chosen to be a subject-specific variable which means all teeth within a subject have the same value of X_4 . The coefficients of (X_1, X_2, X_3, X_4) were chosen to be $\beta_{01} \equiv (\beta_{01}^{(1)}, \beta_{01}^{(2)}, \beta_{01}^{(3)}, \beta_{01}^{(4)}) = (0.2, -0.1, 0.1, 0.3)$ for the 0-to-1 transition and $\beta_{12} \equiv (\beta_{12}^{(1)}, \beta_{12}^{(2)}, \beta_{12}^{(3)}, \beta_{12}^{(4)}) = (0.5, -0.2, 0.3, 0.2)$ for the 1-to-2 transition. The exact transition times from one caries state to the next were generated by the inverse cumulative distribution function method. Three visit times, V_1 , V_2 and V_3 , were generated for each subject in this manner: $V_1 \sim U(1.6, 2.4)$, $V_2 = V_1 + U(0.8, 1.2)$, and $V_3 = V_2 + 5$. The final data set in each iteration of simulation consist of visit times, caries state at each visit time, and covariates for the four teeth of every subject. The number of iterations was one thousand in the Monte Carlo simulation.

Table 2.1: Simulation results for estimating the Weibull and variance parameters

	$\log r_{01}$	$\log k_{01}$	$\log r_{12}$	$\log k_{12}$	$\log \sigma_{01}$	$\log \sigma_{12}$
True value	0.182	-2.303	-0.105	-1.609	0	0.182
<i>m=200, n = 4</i>						
Estimate [†]	0.188	-2.317	-0.083	-1.674	-0.011	0.164
Coverage [‡]	0.968	0.957	0.958	0.968	0.944	0.947
MSE	0.002	0.025	0.019	0.139	0.008	0.017
Bias	0.006	0.015	0.022	0.065	0.011	0.019
Empirical SE	0.049	0.157	0.135	0.367	0.087	0.129
Mean of SEs*	0.054	0.164	0.150	0.402	0.087	0.128
<i>m=400, n = 4</i>						
Estimate	0.183	-2.302	-0.093	-1.636	-0.017	0.157
Coverage	0.960	0.946	0.962	0.961	0.939	0.951
MSE	0.001	0.013	0.009	0.065	0.004	0.008
bias	0.000	0.000	0.012	0.027	0.017	0.026
Empirical SE of	0.035	0.115	0.092	0.254	0.062	0.087
Mean of SEs	0.038	0.113	0.103	0.275	0.061	0.088

[†] The average of the estimates across the 1000 simulations.

[‡] The empirical coverage of the 95% confidence interval based on the asymptotic distribution of $\hat{\theta}$.

* The average of the asymptotic standard errors across the 1000 simulations

2.3.2 Simulation results

Table 2.1 and Table 2.2 show the simulation results of model estimation for the two different sample sizes. As seen in Table 2.1 and Table 2.2, the bias, standard error and mean squared error of every parameter's estimator are small in both scenarios. The empirical coverage of every parameter's asymptotic confidence interval is close to the nominal level. The empirical standard error of every estimator is also close to the asymptotic standard error. When the sample size doubles, both MSE and variance of every estimator reduce about by half as expected according to the asymptotic theory. In a word, the likelihood-based inference for θ performs very well when the number of subjects is equal to or greater than 200.

Table 2.2: Simulation results for estimating the regression parameters β

	$\beta_{01}^{(1)}$	$\beta_{01}^{(2)}$	$\beta_{01}^{(3)}$	$\beta_{01}^{(4)}$	$\beta_{12}^{(1)}$	$\beta_{12}^{(2)}$	$\beta_{12}^{(3)}$	$\beta_{12}^{(4)}$
True value	0.2	-0.1	0.1	0.3	0.5	-0.2	0.3	0.2
<i>m=200, n = 4</i>								
Estimate	0.198	-0.098	0.100	0.305	0.510	-0.191	0.310	0.201
Coverage	0.953	0.963	0.954	0.928	0.972	0.954	0.965	0.938
MSE	0.017	0.017	0.018	0.008	0.035	0.040	0.036	0.014
Bias	0.002	0.002	0.000	0.005	0.010	0.009	0.010	0.001
Empirical SE	0.131	0.132	0.134	0.091	0.186	0.200	0.188	0.118
Mean of SEs	0.135	0.138	0.136	0.084	0.198	0.204	0.199	0.112
<i>m=400, n = 4</i>								
Estimate	0.200	-0.103	0.099	0.307	0.497	-0.199	0.302	0.203
Coverage	0.947	0.953	0.951	0.912	0.947	0.962	0.956	0.921
MSE	0.009	0.009	0.009	0.004	0.018	0.017	0.018	0.008
bias	0.000	0.003	0.001	0.007	0.003	0.001	0.002	0.003
Empirical SE	0.095	0.093	0.093	0.065	0.133	0.132	0.135	0.090
Mean of SEs	0.094	0.096	0.094	0.058	0.137	0.141	0.137	0.076

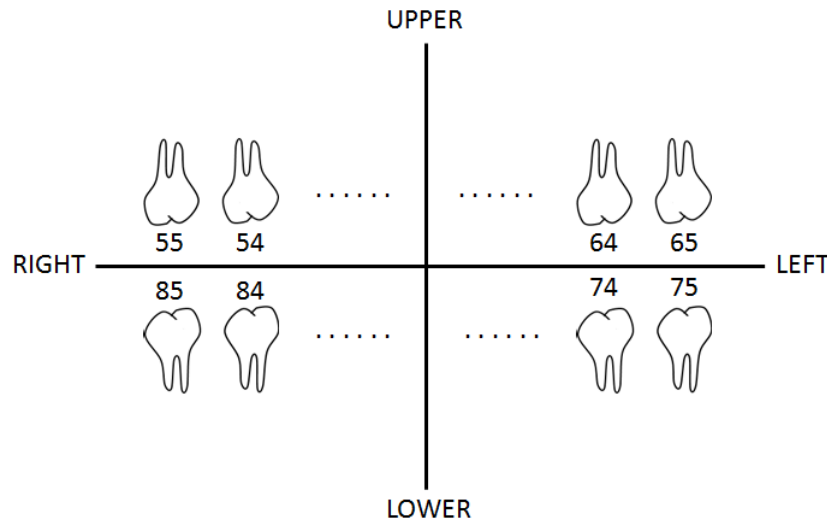


Figure 2.2: Molars in a child's mouth clockwise starting from the upper-right teeth (numbers starting 5) from the dentist's view. The child's right side corresponds to the tooth chart's left side. The numbers ending 4 and 5 correspond to first and second molars respectively

2.4 Application

In this section, we apply the three-state Markov frailty model to the Analysis of Detroit Dental Health Project (DDHP) data and demonstrate how to predict the future tooth decay based on the given information on a child. The data on the first and second molars were analyzed, resulting in eight teeth per child (see Figure 2.2). We consider fifteen dummy covariates to indicate every possible combination of tooth type (first or second molar), tooth location (upper or lower, right or left) and gender of the child in the model. The upper right first molar for male is set as the reference category. The total number of parameters is 36 because there are additionally four Weibull parameters and two random-effect scale parameters and all the parameters are specific to the transition type (0-to-1 and 1-to-2). We first fit this saturated model, and the resulting estimates and standard errors are shown in the Table 2.3.

Based on the saturated model, we try to answer the following three scientific questions:

- i. Is there an interaction effect between gender and tooth identity (type and location) on tooth decay?
- ii. Do the first and second molars have the same tooth decay profile?
- iii. Is there a upper-lower symmetry or right-left symmetry in tooth decay profile?

Every question above can be answered by testing a hypothesis of the form

$$H_0: A\theta = 0$$

formulated by choosing an appropriate matrix A with full row rank s . A test for it would be the Wald test:

$$(A\hat{\theta})^T (AI_{obs}^{-1}(\hat{\theta})A^T)^{-1} A\hat{\theta} \stackrel{H_0}{\sim} \chi_s^2. \quad (2.7)$$

To test the interaction between gender and tooth identity, the hypothesis is formulated as

$$H_0: \begin{cases} \alpha_{i,01} = \beta_{(i+1),01} - \beta_{1,01} \\ \alpha_{i,12} = \beta_{(i+1),12} - \beta_{1,12} \end{cases}, \quad i = 1, \dots, 7. \quad (2.8)$$

Table 2.3: The DDHP data analysis using the univariate random effect model

Gender	Molar Type	Quadrant	Transition	Parameter	Estimate	SE	p-value
Male	1st molar	Upper left teeth (64)	$0 \rightarrow 1$	$\alpha_{1,01}$	-0.0087	0.1219	0.9431
			$1 \rightarrow 2$	$\alpha_{1,12}$	0.1088	0.1842	0.5546
		Lower left teeth (74)	$0 \rightarrow 1$	$\alpha_{2,01}$	0.3464	0.1106	0.0017
			$1 \rightarrow 2$	$\alpha_{2,12}$	0.2495	0.1768	0.1580
		Lower right teeth (84)	$0 \rightarrow 1$	$\alpha_{3,01}$	0.3450	0.1028	0.0008
			$1 \rightarrow 2$	$\alpha_{3,12}$	0.2214	0.1657	0.1816
	2nd molar	Upper right teeth (55)	$0 \rightarrow 1$	$\alpha_{4,01}$	1.4359	0.1051	< 0.0001
			$1 \rightarrow 2$	$\alpha_{4,12}$	-0.1986	0.1709	0.2453
		Upper left teeth (65)	$0 \rightarrow 1$	$\alpha_{5,01}$	1.5080	0.1071	< 0.0001
			$1 \rightarrow 2$	$\alpha_{5,12}$	-0.2002	0.1537	0.1925
		Lower left teeth (75)	$0 \rightarrow 1$	$\alpha_{6,01}$	1.3746	0.1006	< 0.0001
			$1 \rightarrow 2$	$\alpha_{6,12}$	-0.1279	0.1585	0.4194
		Lower right teeth (85)	$0 \rightarrow 1$	$\alpha_{7,01}$	1.4169	0.1057	< 0.0001
			$1 \rightarrow 2$	$\alpha_{7,12}$	-0.1234	0.1543	0.4237
Female	1st molar	Upper right teeth (54)	$0 \rightarrow 1$	$\beta_{1,01}$	0.2178	0.1198	0.0691
			$1 \rightarrow 2$	$\beta_{1,12}$	0.0110	0.1621	0.9461
		Upper left teeth (64)	$0 \rightarrow 1$	$\beta_{2,01}$	0.3071	0.1118	0.0060
			$1 \rightarrow 2$	$\beta_{2,12}$	0.0769	0.1565	0.6230
		Lower left teeth (74)	$0 \rightarrow 1$	$\beta_{3,01}$	0.2723	0.1126	0.0156
			$1 \rightarrow 2$	$\beta_{3,12}$	0.3337	0.1518	0.0279
	2nd molar	Lower right teeth (84)	$0 \rightarrow 1$	$\beta_{4,01}$	0.3277	0.1141	0.0041
			$1 \rightarrow 2$	$\beta_{4,12}$	0.3509	0.1608	0.0291
		Upper right teeth (55)	$0 \rightarrow 1$	$\beta_{5,01}$	1.6215	0.1088	< 0.0001
			$1 \rightarrow 2$	$\beta_{5,12}$	-0.0885	0.1515	0.5590
		Upper left teeth (65)	$0 \rightarrow 1$	$\beta_{6,01}$	1.6814	0.1153	< 0.0001
			$1 \rightarrow 2$	$\beta_{6,12}$	0.0694	0.1503	0.6440
		Lower left teeth (75)	$0 \rightarrow 1$	$\beta_{7,01}$	1.8786	0.1192	< 0.0001
			$1 \rightarrow 2$	$\beta_{7,12}$	0.0509	0.1531	0.7395
	Lower right teeth (85)		$0 \rightarrow 1$	$\beta_{8,01}$	1.9004	0.1114	< 0.0001
			$1 \rightarrow 2$	$\beta_{8,12}$	-0.0679	0.1518	0.6546

According to the data analysis, this null hypothesis was rejected at a 5% significant level (p-value < 0.0001). It indicates that there is a statistically significant interaction effect between gender and tooth identity on tooth decay, so questions (ii) and (iii) would be more meaningful if asked for each gender. The subsequent null hypotheses of question (ii) are formulated as

$$H_0: \begin{cases} \alpha_{j,01} = \alpha_{(j+4),01} \\ \alpha_{4,01} = 0 \\ \alpha_{j,12} = \alpha_{(j+4),12} \\ \alpha_{4,12} = 0 \end{cases}, \quad j = 1, 2, 3, \quad (2.9)$$

for male and

$$H_0: \begin{cases} \beta_{k,01} = \beta_{(k+4),01} \\ \beta_{k,12} = \beta_{(k+4),12} \end{cases}, \quad k = 1, 2, 3, 4, \quad (2.10)$$

for female. The p-values for testing these two hypotheses are both less than 0.0001. Thus the first and second molars have different tooth decay profiles. Question (iii) asks whether the dental caries progresses in a same pattern between lower (or right) and upper (or left) teeth. The null hypotheses for the upper-lower symmetry are formulated as

$$H_0: \begin{cases} \alpha_{1,01} = \alpha_{2,01} \\ \alpha_{1,12} = \alpha_{2,12} \\ \alpha_{3,01} = \alpha_{3,12} = 0 \\ \alpha_{4,01} = \alpha_{7,01} \\ \alpha_{4,12} = \alpha_{7,12} \\ \alpha_{5,01} = \alpha_{6,01} \\ \alpha_{5,12} = \alpha_{6,12} \end{cases} \quad (2.11)$$

for male and

$$H_0: \begin{cases} \beta_{1,01} = \beta_{4,01} \\ \beta_{1,12} = \beta_{4,12} \\ \beta_{2,01} = \beta_{3,01} \\ \beta_{2,12} = \beta_{3,12} \\ \beta_{5,01} = \beta_{8,01} \\ \beta_{5,12} = \beta_{8,12} \\ \beta_{6,01} = \beta_{7,01} \\ \beta_{6,12} = \beta_{7,12} \end{cases} \quad (2.12)$$

for female, and the null hypotheses for the right-left symmetry are formulated as

$$H_0: \begin{cases} \alpha_{i,01} = \alpha_{i+1,01} \\ \alpha_{i,12} = \alpha_{i+1,12} \\ \alpha_{1,01} = \alpha_{1,12} = 0 \end{cases} \quad , \quad i = 2, 4, 6, \quad (2.13)$$

for male and

$$H_0: \begin{cases} \beta_{k,01} = \beta_{k+1,01} \\ \beta_{k,12} = \beta_{k+1,12} \end{cases} \quad , \quad k = 1, 3, 5, 7, \quad (2.14)$$

for female. The test statistics for the hypotheses (2.11), (2.12), (2.13) and (2.14) are 35.412 (p-value < 0.0001), 29.658 (p-value = 0.0002), 0.999 (p-value = 0.8415) and 2.916 (p-value = 0.9395) respectively. Therefore, right-left symmetry exists in the tooth decay profile for each gender. The hypotheses above can be expressed in simpler forms if we use more subscripts (see Section 3.4 of Chapter 3), although the current expression of the hypotheses is more intuitive.

We further performed the tests for questions (ii) and (iii) only to the transition from 0 to 1 or from 1 to 2. Table 2.4 shows the Wald test statistics and their p-values of these tests for symmetry. One can see that there is a symmetry in 1-to-2 transition intensity between the upper and lower molars in boys. Since hypotheses (2.13) and (2.14) were not rejected, it is no wonder that none of the tests for right-left symmetry in Table 2.4 shows any significant result.

Table 2.4: The Wald test statistics (p-value) for the three types of symmetry in caries transition intensity

Gender	Test for symmetry	Transition	
		0 to 1	1 to 2
Male	Upper vs Lower	28.95(< 0.0001)	4.77(0.3115)
	Right vs Left	0.52(0.9716)	0.37(0.9849)
	First molar vs Second molar	1013.65(< 0.0001)	28.65(< 0.0001)
Female	Upper vs Lower	16.58(0.0023)	12.42(0.0145)
	Right vs Left	1.08(0.8967)	1.83(0.7666)
	First molar vs Second molar	1742.93(< 0.0001)	16.45(0.0025)

To illustrate predicting the future transition probability, we chose a boy from the dataset who had only two visits (2.14 yrs and 4.23 yrs) with the following caries states: the upper right first molar and all the lower teeth in study remained sound over the two study visits; the rest of the considered teeth were sound at the first visit but were found to have early caries at the second. No tooth had advanced caries. The empirical Bayes estimator of u_i is 0.13 for the child. The posterior median and the 95 % credible interval of this boy's transition probability after his last visit were computed over the time interval (4.23 yrs, 8 yrs). A sample of 200 transition probability functions were computed based on 200 (u_i, θ) 's generated from their joint posterior distribution. The top plot of Figure 2.3 shows the probabilities of the 0-to-1 and 0-to-2 transitions for his upper right first molar, and the bottom plot shows the 1-to-2 transition probability for his upper right second molar. The solid line is the posterior median of the transition probability. The credible interval refers to the interval between the 2.5% and 97.5% sample quantiles of the transition probability. One can see that at age six, this boy's upper right first molar only has probability 0.03 of having advanced caries, but his upper right second molar has probability 0.23 to be in advanced caries state.

2.5 Discussion

We proposed a parametric three-state Markov frailty model for tooth-level caries life course data under interval censoring. The likelihood-based inference for the model was shown to perform very well in finite samples by simulation. Furthermore, we came up with a Bayesian approach to predict future transition probabilities given a child's observed caries life-history data. It is worthy to point

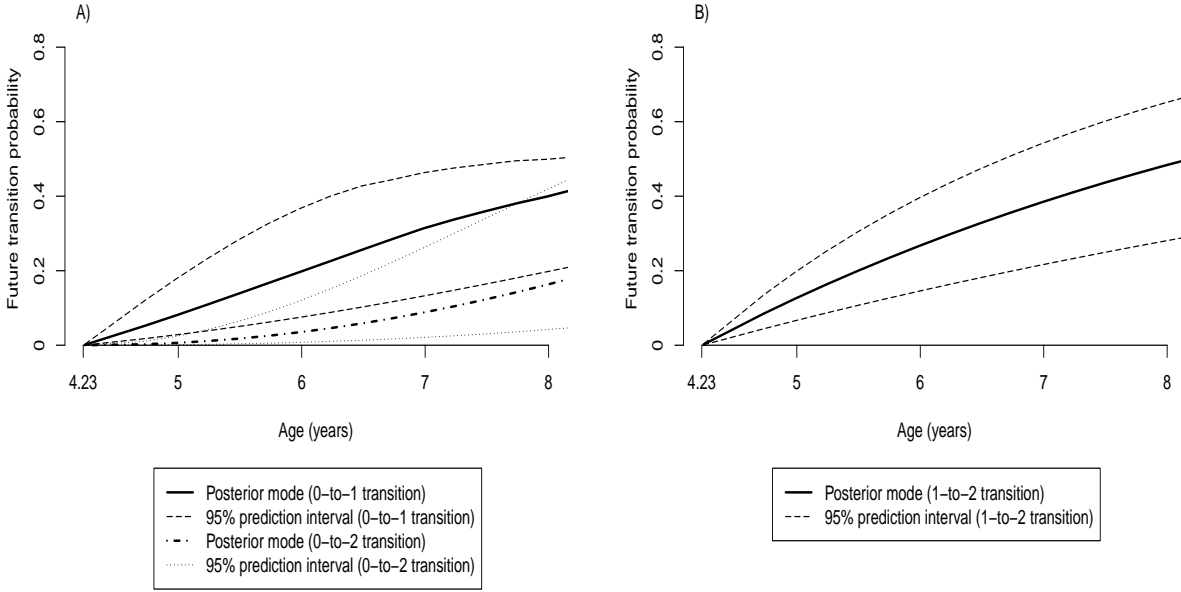


Figure 2.3: The predicted transition probabilities from state 0 to state 1 and to state 2 for the upper right first molar (A) and from state 1 to state 2 for the upper right second molar (B) for a boy in the DDHP dataset

out that the prediction approach can predict future caries development not only for individual teeth but also at the mouth level. For instance, using the Bayesian approach, one can obtain a predicted value and a prediction interval of the probability that no tooth in the mouth will develop new caries within one year after the most recent dental visit.

Information gained from applying the proposed methods to real data will provide guidance on the timing of dental screening and treatment. Our statistical methods also have general applications to modeling chronic diseases involving paired or multiple organs in cohort studies where the disease status is measured intermittently. Besides dental caries, popular examples include sacroiliac joint damage in psoriatic arthritis (Rahman et al., 1998) and diabetic retinopathy (Diabetes Control and Complications Trial Research Group, 1995).

Several future research directions are worth to pursue. Firstly, it would be appealing to model the baseline transition intensity nonparametrically, which will be discussed in Chapter 3. To achieve

this, we will use a linear spline model for the log baseline intensity since it does not induce additional numerical integrations in the likelihood calculation. Secondly, the intra-oral and the inter-transition correlations of caries life course data can be accounted for in the analysis in a different way. For instance, we can replace the shared random effect in our model by two transition-specific random effects with a bivariate normal distribution. GEE-type of analysis is also feasible. Lastly, one may relax the proportionality assumption on intensities between different teeth by forming strata (e.g., dental quadrants) and using stratum-specific baseline intensities.

CHAPTER 3

SEMIPARAMETRIC ANALYSIS OF CORRELATED AND INTERVAL CENSORED EVENT-HISTORY DATA

3.1 Introduction

As discussed in Chapter 1, in many instances, the data generated from longitudinal investigations are interval censored, in that the transition time point is only known up to a time interval. Additionally, many chronic conditions, despite being inherently continuous, are often represented by a finite number of states derived from clinical evaluations, serological tests and other medical assessment tools. Furthermore, many chronic conditions manifest themselves at various locations of the body, resulting in correlated time-to-event endpoints even after conditioning on observed covariates. In caries research, carious lesions are likely to develop on multiple teeth of a person, owing to the intra-oral association resulting from sharing common risk factors, including genetic variants and environmental factors which may not be fully measured.

In this chapter, we are interested in formulating a flexible multistate model to describe a set of multiple life-history processes observed only at intermittent time points. A data example motivating the proposed model comes from the Detroit Dental Health Project (DDHP) involving tooth-level caries assessments in primary dentition for young inner-city African-American children, which has been introduced in detail in Chapter 1. In this era of precision medicine, understanding the transition process of a tooth across the caries spectrum could aid dentists design tooth-specific prevention and treatment plans. Existing works in caries research have primarily focused on mouth-level indices such as the number of Decayed, Missing or Filled teeth or surfaces, a typical measure of caries experience in human populations (Lewsey & Thomson, 2004). Albeit important, these indices have well documented limitations regarding their inability to provide a comprehensive assessment of dental caries (Lewsey & Thomson, 2004), which is critical for personalizing preventive oral health care and treatment plans.

Analysis of correlated interval censored event-history data is nontrivial, owing to the difficulty of the theoretical argument and the computational burden of the numerical implementation. For example, the use of counting process techniques is very difficult in the framework of interval censored data, hence prohibiting the use of martingale theory. Consequently, unlike right-censored data, one has to rely on complicated empirical process arguments to investigate the asymptotic properties of the estimation procedures (Zhang & Sun, 2010). It is also well known that when multiple life-history assessments are made on the same sampling unit, any analysis involving the within-unit correlation of the data, viewed as a quantity of scientific interest or a nuisance, necessitates the use of methods for correlated survival data (Kalbfleisch & Prentice, 2002, Chapter 10). To the best of our knowledge, only few statistical papers (Sutradhar & Cook, 2008; Joly et al., 2012) have studied analysis of correlated interval censored event-history data without imposing the common modeling restriction that the life-history process is time-homogeneous, i.e. having a constant transition intensity matrix across time. Assuming a restrictive model structure is apt to yield biased analysis results due to model misspecification. The existing works and the methodological approach of Chapter 2 relied on likelihood-based inference by means of frailty models for intensity functions to account for the within-sampling-unit clustering. Albeit important, these works still have notable limitations. For example, the works of Sutradhar & Cook (2008) and Joly et al. (2012) are not amenable to analysis involving cluster-level or observation-unit level covariates. As well as the approach in Chapter 2, the works of Joly et al. (2012) also assumed that the baseline intensity functions are of Weibull type, which is apt to yield biased inferences for the covariate effects on the transition intensities. Sutradhar and Cook (2008) attempted to relax this modeling assumption by allowing the baseline intensities to be piecewise constant. A limitation, however, is that the selection of the number and locations of knots was ad hoc ignoring that a poorly chosen number of knots may introduce bias. Finally, existing works such as Sutradhar & Cook (2008) and Joly et al. (2012) often lacked numerical simulations to support the heuristic arguments being proposed.

To circumvent some of the limitations of existing works, in this chapter, we propose a semiparametric Markov model with frailties to analyze clustered interval censored tooth-level event-history

data on early childhood caries. Our model is similar in spirit to the parametric model introduced in Chapter 2, which uses frailties to tie together all life-history data from the same sampling unit. But unlike Chapter 2 which considers Weibull-type baseline intensities and shared frailties for various types of transitions, the formulation in this chapter relaxes these assumptions by using unshared frailties and approximating the unknown baseline intensities by spline functions. The estimation of the model parameters including fixed effects coefficients, spline parameters, and random effects' variance components then proceeds by maximizing a penalized log-likelihood which constrains the spline coefficients to avoid the overfitting problem. As discussed by Brumback & Rice (1998) and Cai & Betensky (2003), penalizing the spline coefficients is equivalent to treating these coefficients as random effects, and penalty parameters can be estimated as reciprocal variance components of the random effects using a restricted maximum likelihood approach. Thus, this mixed-model representation avoids the use of computationally intensive techniques such as cross-validation to select penalty parameters. More importantly, it enables the fitting algorithm and inferential approaches to be devised within the framework of hierarchical likelihood (Lee & Nelder, 1996). As a by-product of the analysis, we propose a Bayesian approach for predicting the future transition probabilities of tooth-level caries in the spirit of Section 2.2.2. Due to the inherent heterogeneity in transition rates among different individuals not captured by measured covariates, these future transition probabilities are predicted using the subject's historical data via the frailties estimates. We conduct a simulation study to evaluate the performance of the Bayesian prediction procedure.

The rest of the chapter is organized as follows. The proposed semiparametric Markov frailty model is outlined in Section 3.2, followed by a detailed description of the model estimation and associated inference, and a Bayesian approach for predicting future transition probabilities. In Section 3.3, simulation studies are conducted to evaluate the accuracy of the estimators as well as the coverage rates of the confidence intervals for the regression parameters and baseline intensities. We also investigate the coverage of the prediction intervals for future transition probabilities through simulations. The application of the proposed methods to the DDHP data is presented in Section 3.4. We give some concluding remarks in Section 3.5.

3.2 Statistical methodologies

We define $A_{ij}(t)$ as the caries state of the j -th tooth for the i -th subject ($i = 1, \dots, m; j = 1, \dots, n$) at time t ($t \geq 0$), observed only at the time points \mathbf{V}_i . The intra-oral correlation among the $A_{ij}(t)$'s ($j = 1, \dots, n$) is modeled through a bivariate frailty $\mathbf{w}_i = (w_{i,01}, w_{i,12})^T$ on transition intensities, coupled with the conditional independence assumption. To highlight the dependence on covariates and the frailty, we denote by $h_{ab}(t|\mathbf{X}_{ij}, \mathbf{w}_i)$ the transition intensity from a to b , defined as follows,

$$\begin{aligned} h_{ab}(t|\mathbf{X}_{ij}, \mathbf{w}_i) &= \lim_{\Delta t \rightarrow 0} \Pr(A_{ij}(t^- + \Delta t) = b | A_{ij}(u), 0 \leq u < t, A_{ij}(t^-) = a, \mathbf{X}_{ij}, \mathbf{w}_i) \\ &= \lim_{\Delta t \rightarrow 0} \Pr(A_{ij}(t^- + \Delta t) = b | A_{ij}(t^-) = a, \mathbf{X}_{ij}, \mathbf{w}_i), \end{aligned}$$

where $(a, b) \in \{(0, 1), (1, 2)\}$.

The following multiplicative intensity model for $h_{ab}(t|\mathbf{X}_{ij}, \mathbf{w}_i)$ is assumed

$$h_{ab}(t|\mathbf{X}_{ij}, \mathbf{w}_i) = h_{ab}^{(0)}(t) w_{i,ab} \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta}_{ab}), \quad (a, b) \in \{(0, 1), (1, 2)\},$$

where $h_{ab}^{(0)}(t)$'s are unspecified baseline transition intensities and $\log(\mathbf{w}_i) = (\log(w_{i,01}), \log(w_{i,12}))^T$ is assumed to follow a bivariate zero-mean normal distribution, that is,

$$\log(\mathbf{w}_i) \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{01}^2 & \sigma_{01}\sigma_{12}\rho \\ \sigma_{01}\sigma_{12}\rho & \sigma_{12}^2 \end{pmatrix} \right).$$

Here σ_{ab}^2 ($(a, b) \in \{(0, 1), (1, 2)\}$) reflects not only the strength of the intra-oral association of caries progressions among different teeth, but also the magnitude of heterogeneity in caries transition intensities between subjects that is not captured by observed covariates. The parameter ρ reflects the within-subject correlation between the two types of transitions, with positive ρ 's implying that the two intensities evolve similarly (for example, if the transition from 0 to 1 is fast and so is 1 to 2). The assumed bivariate frailty provides a more flexible representation of the inter-transition dependence compared to the shared frailty model in chapter 2, which constrains ρ the correlation between the log frailties of the two transitions to one.

For computational simplicity, we reparameterize $\log(\mathbf{w}_i)$ as $\log(\mathbf{w}_i) = (\sigma_{01}z_{i1}, \sigma_{12}(\rho z_{i1} + \sqrt{1 - \rho^2}z_{i2}))^T$ where z_{i1} and z_{i2} are independent standard normal random variables. Under this formulation, $h_{ab}(t|\mathbf{X}_{ij}, \mathbf{w}_i)$ are rewritten as $h_{ab}(t|\mathbf{X}_{ij}, \mathbf{z}_i)$ ($(a, b) \in \{(0, 1), (1, 2)\}$) with $\mathbf{z}_i \equiv (z_{i1}, z_{i2})^T$.

3.2.1 A penalized likelihood estimation

As described in Chapter 1, the transition times between caries states in the Detroit study are only known up to an interval between two consecutive study visits. This naturally complicates the estimation of the multistate model for tooth decay. Without interval censoring, the regression parameters and baseline intensities for different transitions can be estimated separately based on the transition-specific partial likelihood and Breslow's estimator (Kalbfleisch & Prentice, 2002, Section 8.3), even after conditioning on random effects; while under interval censoring, they have to be estimated jointly as the partial likelihood approach is not applicable. Consequently, the model estimation would require a full-likelihood-based method, which is developed in this subsection.

Unlike the model in Chapter 2 which imposes a parametric form on $h_{01}^{(0)}(\cdot)$ and $h_{12}^{(0)}(\cdot)$, we estimate these functions nonparametrically via penalized spline smoothing. Specifically, we approximate the logarithms of baseline intensity functions by linear splines,

$$\log h_{ab}^{(0)}(t) = \alpha_{0,ab} + b_{0,ab}t + \sum_{j=1}^{N_{ab}} b_{j,ab}(t - \kappa_{j,ab})_+, \quad (a, b) \in \{(0, 1), (1, 2)\}$$

where $(x)_+ = \max(0, x)$, and $\kappa_{j,01}$'s ($j = 1, \dots, N_{01}$) and $\kappa_{j,12}$'s ($j = 1, \dots, N_{12}$) are the locations of knots. The roughness of the spline functions, defined later, is penalized in the model estimation. We choose linear spline basis because it leads to closed-form expressions of cumulative baseline intensities involved in the likelihood function. The numbers of knots for these spline models increase with the sample size. In practice, a large number of knots can be used to reduce the approximation bias while not incurring the overfitting problem due to the penalization, but small N_{01} and N_{12} are preferred computationally. According to Kauermann et al. (2009), a penalized spline with dimension proportional to a small fractional power of the sample size can achieve the optimal rate of convergence (Stone, 1982). This guides our selection on the number of knots discussed later. The

knots for the a -to- b transition, $(\kappa_{1,ab}, \dots, \kappa_{N_{ab},ab})$, are equally spaced with respect to the quantiles of the unique values of $\{V_{i,l-1}, V_{il} : \text{an } a\text{-to-}b \text{ transition occurred to a tooth in } (V_{i,l-1}, V_{il}], i = 1, \dots, m, l = 1, \dots, M_i\} \cup \{V_{M_i} : \text{no } a\text{-to-}b \text{ transition occurred to a tooth by } V_{M_i}, i = 1, \dots, m\}$, where $V_{i0} = 0$.

Define $\boldsymbol{\sigma} \equiv (\sigma_{01}, \sigma_{12}, \rho)^T$. Denote the vector of all model parameters except the variance components by $\boldsymbol{\xi} \equiv (\boldsymbol{\beta}_{01}^T, \boldsymbol{\beta}_{12}^T, \alpha_{0,01}, \alpha_{0,12}, \mathbf{b}_{01}^T, \mathbf{b}_{12}^T)^T$ where $\mathbf{b}_{01} \equiv (b_{0,01}, b_{1,01}, \dots, b_{N_{01},01})^T$ and $\mathbf{b}_{12} \equiv (b_{0,12}, b_{1,12}, \dots, b_{N_{12},12})^T$. Under the assumption that the tooth decay processes are independent across subjects, the log-likelihood of the observed data \mathbf{Y} is

$$l(\mathbf{Y}|\boldsymbol{\xi}, \boldsymbol{\sigma}) = \sum_{i=1}^m \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^n f(\mathbf{Y}_{ij}|\mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) \phi(z_{i1}) \phi(z_{i2}) dz_{i1} dz_{i2},$$

where $\phi(\cdot)$ is the standard normal density. Under the conditional Markov assumption on the tooth decay process,

$$f(\mathbf{Y}_{ij}|\mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) \propto \prod_{q=1}^{M_i} Pr \left(A_{ij}(V_{iq}) = K_{iq}^{(j)} | A_{ij}(V_{i,q-1}) = K_{i,q-1}^{(j)}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma} \right), \quad (3.1)$$

where $K_{i0}^{(j)} = 0$ ($A_{ij}(0) = 0$ according to the caries classification scheme in Chapter 1). To simplify the notations, we write the transition probabilities on the right-hand side of (3.1) as $p_{K_{i,q-1}^{(j)}, K_{iq}^{(j)}}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma})$, which can be expressed in terms of the transition intensities $h_{ab}(t|\mathbf{X}_{ij}, \mathbf{z}_i)$ ($(a, b) \in \{(0, 1), (1, 2)\}$) as follows,

$$p_{00}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = S_{01}(V_{i,q-1}, V_{iq}|\mathbf{X}_{ij}, \mathbf{z}_i)$$

$$p_{01}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = \int_{V_{i,q-1}}^{V_{iq}} S_{01}(V_{i,q-1}, r|\mathbf{X}_{ij}, \mathbf{z}_i) h_{01}(r|\mathbf{X}_{ij}, \mathbf{z}_i) S_{12}(r, V_{iq}|\mathbf{X}_{ij}, \mathbf{z}_i) dr$$

$$p_{02}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = \int_{V_{i,q-1}}^{V_{iq}} S_{01}(V_{i,q-1}, r|\mathbf{X}_{ij}, \mathbf{z}_i) h_{01}(r|\mathbf{X}_{ij}, \mathbf{z}_i) \{1 - S_{12}(r, V_{iq}|\mathbf{X}_{ij}, \mathbf{z}_i)\} dr$$

$$p_{11}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = S_{12}(V_{i,q-1}, V_{iq}|\mathbf{X}_{ij}, \mathbf{z}_i)$$

$$p_{12}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = 1 - S_{12}(V_{i,q-1}, V_{iq}|\mathbf{X}_{ij}, \mathbf{z}_i)$$

$$p_{22}(V_{i,q-1}, V_{iq}, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}) = 1$$

where $S_{ab}(x, y | \mathbf{X}_{ij}, \mathbf{z}_i) \equiv \exp\{-\int_x^y h_{ab}(t | \mathbf{X}_{ij}, \mathbf{z}_i) dt\}$. The integrals involved in p_{01} and p_{02} do not have a closed form and thus are numerically approximated using the Gauss-Jacobi quadrature. In our implementation, the number of quadrature points is,

$$\max \left(5, \left\lceil 10 \times \frac{V_{iq} - V_{i,q-1}}{\max\{V_{ik} - V_{i,k-1} : \text{all } i \text{ and } k\}} \right\rceil \right).$$

We estimate (ξ, σ) by maximizing the following penalized log-likelihood,

$$l_{pen}(\mathbf{Y} | \xi, \sigma) = l(\mathbf{Y} | \xi, \sigma) - \frac{\tau_{01}}{2} \mathbf{b}_{01}^T \mathbf{b}_{01} - \frac{\tau_{12}}{2} \mathbf{b}_{12}^T \mathbf{b}_{12}, \quad (3.2)$$

where τ_{01} and τ_{12} are two penalty parameters that regularize the roughness of the two estimated baseline intensity functions. Our roughness definition is similar to $\int (\partial \log h_{ab}^{(0)}(t) / \partial t)^2 dt$ but slightly different from that in Equation (2.4) of Kauermann et al. (2009), which does not include $b_{0,ab}^2$. We tried the penalty excluding $b_{0,ab}^2$ in simulations; the results (not reported here) are worse than including $b_{0,ab}^2$. Albeit using a slightly different penalty term, we follow Assumption 4 of Kauermann et al. (2009) to set the two numbers of knots to $N_{01} = N_{12} = \max(\lceil (mn)^{1/5} \rceil, 5)$.

The values of the penalty parameters are generally determined by cross-validation techniques. However, these procedures are known to be computationally intensive, especially for penalized log-likelihoods involving intractable integrals. An alternative approach is to treat the penalized likelihood as that of a mixed-effects model and then estimate the penalty parameters as reciprocal variance components of the random effects as, e.g., in Cai & Betensky (2003). More specifically, we treat the spline parameter vectors \mathbf{b}_{01} and \mathbf{b}_{12} as random effects that follow the multivariate normal distributions below:

$$\mathbf{b}_{01} \sim N(\mathbf{0}, \theta_{01}^2 \mathbf{I}_{N_{01}+1}) \text{ and } \mathbf{b}_{12} \sim N(\mathbf{0}, \theta_{12}^2 \mathbf{I}_{N_{12}+1}),$$

where $\theta_{01}^2 = 1/\tau_{01}$, $\theta_{12}^2 = 1/\tau_{12}$, and \mathbf{I}_k denotes a $k \times k$ identity matrix ($k \in \mathbb{Z}^+$). The resulting mixed-effects model has fixed effect parameters $\zeta \equiv (\beta_{01}^T, \beta_{12}^T, \alpha_{01}, \alpha_{12})^T$, random effect parameters $\gamma \equiv (\mathbf{b}_{01}^T, \mathbf{b}_{12}^T, \mathbf{z}_1^T, \dots, \mathbf{z}_m^T)^T$, and variance component parameters $\nu \equiv (\log \sigma_{01}, \log \sigma_{12}, \psi(\rho), \log \theta_{01}, \log \theta_{12})^T$. Here, $\psi(\cdot)$ denotes the Fisher's z -transformation function.

Ideally, we would like to estimate ζ and ν by maximizing the log-likelihood of the observed data \mathbf{Y} (marginal likelihood) based on the above mixed-effects model,

$$\begin{aligned} l_{mixed}(\mathbf{Y}|\zeta, \nu) &= \log \int \int \int f(\mathbf{Y}|\mathbf{z}, \xi, \sigma) f(\mathbf{z}) f(\mathbf{b}_{01}) f(\mathbf{b}_{12}) d\mathbf{z} d\mathbf{b}_{01} d\mathbf{b}_{12} \\ &= \log \int e^{l_h(\mathbf{Y}, \gamma|\zeta, \nu)} d\gamma - (N_{01} + 1) \log \theta_{01} - (N_{12} + 1) \log \theta_{12}, \end{aligned} \quad (3.3)$$

where $\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_m^T)^T$ and

$$l_h(\mathbf{Y}, \gamma|\zeta, \nu) = \sum_{i=1}^m \left[\sum_{j=1}^n \log \{f(\mathbf{Y}_{ij}|\mathbf{z}_i, \xi, \sigma)\} - \frac{\mathbf{z}_i^T \mathbf{z}_i}{2} \right] - \frac{1}{2\theta_{01}^2} \mathbf{b}_{01}^T \mathbf{b}_{01} - \frac{1}{2\theta_{12}^2} \mathbf{b}_{12}^T \mathbf{b}_{12}. \quad (3.4)$$

However, this marginal maximum likelihood estimation for ζ and ν is numerically rather complex, due to the marginal likelihood involving a multiple integral that does not have a closed form. To circumvent this difficulty, other methods which are computationally feasible have been proposed for estimating parameters of mixed-effects models; see, e.g., Pinheiro & Bates (1995) for a summary. Here we adopt Lee & Nelder (1996)'s hierarchical likelihood approach to estimate ζ , γ and ν . This approach estimates ν by maximizing the following restricted likelihood with respect to ν ,

$$l_r(\mathbf{Y}|\nu) = \log \iint e^{l_h(\mathbf{Y}, \gamma|\zeta, \nu)} d\gamma d\zeta - (N_{01} + 1) \log \theta_{01} - (N_{12} + 1) \log \theta_{12}. \quad (3.5)$$

This restricted likelihood is constructed following Harville (1974)'s Bayesian interpretation of the restricted maximum likelihood estimator (REML) for variance components in linear mixed-effects models. Specifically, assuming that ζ and ν have a prior density that is flat relative to the likelihood function $l_{mixed}(\mathbf{Y}|\zeta, \nu)$, $l_r(\mathbf{Y}|\nu)$ is then the logarithm of the marginal posterior density for ν . Thus, the estimator for ν , denoted by $\hat{\nu}$, obtained via maximizing $l_r(\mathbf{Y}|\nu)$ is ν 's marginal posterior mode, which is the same as the Bayesian interpretation of REML for linear mixed-effects models (Harville, 1974). Given $\hat{\nu}$, the hierarchical likelihood approach estimates $\eta \equiv (\zeta^T, \gamma^T)^T$ via maximizing $l_h(\mathbf{Y}, \gamma|\zeta, \hat{\nu})$, called h -likelihood by Lee & Nelder (1996), with respect to η .

In practice, due to the lack of a closed form of the multiple integral in (3.5), we evaluate the restricted likelihood using the Laplace's approximation. This leads to the adjusted profile h -likelihood (APHL) (Lee & Nelder, 1996),

$$l_A(\mathbf{Y}|\nu) = l_h(\mathbf{Y}, \hat{\gamma}(\nu)|\hat{\zeta}(\nu), \nu) - \frac{1}{2} \log | -H_h(\hat{\eta}(\nu); \mathbf{Y}, \nu) | - (N_{01} + 1) \log \theta_{01} - (N_{12} + 1) \log \theta_{12}, \quad (3.6)$$

where $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu}) \equiv (\hat{\boldsymbol{\gamma}}(\boldsymbol{\nu}), \hat{\boldsymbol{\zeta}}(\boldsymbol{\nu}))$ maximizes $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu})$ given $\boldsymbol{\nu}$, $H_h(\hat{\boldsymbol{\eta}}(\boldsymbol{\nu}); \mathbf{Y}, \boldsymbol{\nu})$ is the Hessian matrix of $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu})$ with respect to $\boldsymbol{\eta}$ evaluated at $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu})$, and $|\cdot|$ denotes the matrix determinant. Because there is no analytical expression for $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu})$, we compute the maximum adjusted profile h -likelihood estimate (MAPHLE) $\hat{\boldsymbol{\nu}}$ for $\boldsymbol{\nu}$ and the maximum h -likelihood estimate (MHLE) $\hat{\boldsymbol{\eta}}$ for $\boldsymbol{\eta}$ by alternatively maximizing $l_A(\mathbf{Y}|\boldsymbol{\nu})$ and $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu})$ as in the following steps:

(i) Set $s = 0$, choose an initial value $\boldsymbol{\nu}^{(0)}$ for $\boldsymbol{\nu}$ and maximize $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu}^{(0)})$ w.r.t. $\boldsymbol{\eta}$ to obtain $\boldsymbol{\eta}^{(0)}$;

(ii) Maximize the following function w.r.t. $\boldsymbol{\nu}$ to obtain $\boldsymbol{\nu}^{(s+1)}$:

$$l_h(\mathbf{Y}, \boldsymbol{\gamma}^{(s)}|\boldsymbol{\zeta}^{(s)}, \boldsymbol{\nu}) - 2^{-1} \log | -H_h(\boldsymbol{\eta}^{(s)}; \mathbf{Y}, \boldsymbol{\nu}) | - (N_{01} + 1) \log \theta_{01} - (N_{12} + 1) \log \theta_{12};$$

(iii) Maximize $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu}^{(s+1)})$ w.r.t. $\boldsymbol{\eta}$ to obtain $\boldsymbol{\eta}^{(s+1)}$ and set $s = s + 1$;

(iv) If $\|(\boldsymbol{\eta}^{(s)T}, \boldsymbol{\nu}^{(s)T})^T - (\boldsymbol{\eta}^{(s-1)T}, \boldsymbol{\nu}^{(s-1)T})^T\| / \|(\boldsymbol{\eta}^{(s-1)T}, \boldsymbol{\nu}^{(s-1)T})^T\| < 0.01$, stop and set $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^{(s)}$ and $\hat{\boldsymbol{\nu}} = \boldsymbol{\nu}^{(s)}$; otherwise, go to (ii). Here, $\|\cdot\|$ denotes the Euclidean norm.

The simulation study and data analysis in this chapter were performed using the R statistical software (R 3.1.1). We used the `optim` function with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton option to maximize $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu}^{(s+1)})$ w.r.t. $\boldsymbol{\eta}$ and $l_h(\mathbf{Y}, \boldsymbol{\gamma}^{(s)}|\boldsymbol{\zeta}^{(s)}, \boldsymbol{\nu}) - 2^{-1} \log | -H_h(\boldsymbol{\eta}^{(s)}; \mathbf{Y}, \boldsymbol{\nu}) | - (N_{01} + 1) \log \theta_{01} - (N_{12} + 1) \log \theta_{12}$ w.r.t. $\boldsymbol{\nu}$. In particular, the limited-memory BFGS method was used to maximize $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu}^{(s+1)})$, since it is more suited for the maximization of a function w.r.t. a large number of arguments. We computed the gradient and the hessian matrix of $l_h(\mathbf{Y}, \boldsymbol{\gamma}|\boldsymbol{\zeta}, \boldsymbol{\nu})$ w.r.t $\boldsymbol{\eta}$ using their analytic expressions. In evaluating those expressions, we sometimes encountered the situation where the denominator of a fraction was so close to zero that the software recognized it as zero. When that happened, the denominator was replaced with 10^{-40} .

3.2.2 Statistical inference

According to Lee & Nelder (1996), the MHLE $\hat{\xi}$ is asymptotically equivalent to the marginal MLE for ξ obtained by maximizing (3.3), and both are asymptotically normal with a mean ξ and a covariance matrix that can be consistently estimated by the upper-left $(2p+2) \times (2p+2)$ sub-matrix of $-H_h^{-1}(\hat{\eta}; \mathbf{Y}, \hat{\nu})$ where $2p+2$ is the length of the vector ξ . Hence, hypothesis tests and confidence intervals for ξ can be constructed based on this asymptotic normal distribution.

In view of the random effect treatment of the penalized spline coefficients, we use a Bayesian approach to conduct inferences on $h_{01}^{(0)}(\cdot)$ and $h_{12}^{(0)}(\cdot)$. Specifically, we treat ξ and ν as random quantities with a flat joint prior density, whence $\exp\{l_h(\mathbf{Y}, \gamma|\xi, \nu)\}$ is proportional to the joint posterior density of ξ and \mathbf{z} , $f(\xi, \mathbf{z}|\mathbf{Y}, \nu)$, where $\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_m^T)^T$. A Taylor expansion similar to (5.3) in O’Sullivan (1988) can be used to show that $f(\xi|\mathbf{Y}, \hat{\nu})$ is approximately a multivariate normal distribution when m is large. Similar arguments to Appendices B, C and G in Lee & Nelder (1996) can be used to show that the normal approximation of $f(\xi|\mathbf{Y}, \hat{\nu})$ has a mean approximately equal to the MHLE $\hat{\xi}$ and a covariance matrix approximately equal to the upper-left $(2p+4+N_{01}+N_{12}) \times (2p+4+N_{01}+N_{12})$ block of $-H_h^{-1}(\hat{\eta}; \mathbf{Y}, \hat{\nu})$, which enables one to construct pointwise Bayesian confidence intervals, a.k.a. credible intervals, for $h_{01}^{(0)}(t)$ and $h_{12}^{(0)}(t)$. The use of the conditional posterior $f(\xi|\mathbf{Y}, \hat{\nu})$ is justified because ξ and ν are expected to be asymptotically orthogonal (Ha et al., 2016, Section 4.1).

3.2.3 Prediction of future transition probabilities

Understanding the process by which a tooth transitions from being sound to being decayed is of great interest for primary and secondary prevention in dental care. An important step toward reaching this goal is to predict the tooth-level probabilities of future caries state transitions. The future transition probability is defined here as the conditional probability of a tooth being in a specific caries state at a future time point given the individual’s up-to-date caries life-history data. In terms of the previous notations, the future transition probability of the j -th tooth of the i -th

subject is

$$p_{K_{iM_i}^{(j)} q}(V_{iM_i}, t, \mathbf{X}_{ij}, \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma}), \quad (3.7)$$

where $t > V_{iM_i}$ and $q \geq K_{iM_i}^{(j)}$. Note that V_{iM_i} is the last dental exam time for the i -th child and $K_{iM_i}^{(j)}$ is the dental caries state of the j -th tooth of the i -th child at V_{iM_i} . A natural point predictor for (3.7) is $p_{K_{iM_i}^{(j)} q}(V_{iM_i}, t, \mathbf{X}_{ij}, \hat{\mathbf{z}}_i, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\sigma}})$, which is the empirical Bayes estimator of (3.7) under the zero-one loss. It is, however, not straightforward to construct a $1 - \alpha$ prediction interval for the future transition probability, so we resort to Bayesian methods again. Treating $\boldsymbol{\xi}$ and $\boldsymbol{\nu}$ as random quantities with a flat joint prior density, we can obtain credible intervals for (3.7) as its prediction intervals through simulating $(\boldsymbol{\xi}, \mathbf{z}_i)$ from their joint posterior density $f(\boldsymbol{\xi}, \mathbf{z}_i | \mathbf{Y}, \hat{\boldsymbol{\nu}})$. The use of this conditional posterior is justified because (3.7) depends on $(\mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\sigma})$ through $(\boldsymbol{\xi}, \mathbf{w}_i)$, which is expected to be asymptotically orthogonal to $\boldsymbol{\nu}$ (Ha et al., 2016, Section 4.1). A posterior realization of $(\boldsymbol{\xi}^*, \mathbf{z}_i^*)$ is generated as follows:

(a) Generate $\boldsymbol{\xi}^*$ from $f(\boldsymbol{\xi} | \mathbf{Y}, \hat{\boldsymbol{\nu}}) \approx N(A\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}}), -AH_h^{-1}(\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}}); \mathbf{Y}, \hat{\boldsymbol{\nu}})A^T)$, where

$A = (\mathbf{I}_{2p+4+N_{01}+N_{12}}, \mathbf{0}_{(2p+4+N_{01}+N_{12}) \times 2m})$ and $\mathbf{0}_{(2p+4+N_{01}+N_{12}) \times 2m}$ is a $(2p + 4 + N_{01} + N_{12}) \times 2m$ zero matrix.

(b) Generate \mathbf{z}_i^* from $f(\mathbf{z}_i | \mathbf{Y}, \boldsymbol{\xi}^*, \hat{\boldsymbol{\nu}}) \propto f(\mathbf{Y}_i | \mathbf{z}_i, \boldsymbol{\xi}^*, \hat{\boldsymbol{\nu}}) \phi(z_{i1}) \phi(z_{i2})$ using the Metropolis-Hastings algorithm which employs a standard bivariate normal distribution as the proposal distribution and has a burn-in period of 5000.

We repeat steps (a) and (b) until we obtain enough pairs of $(\boldsymbol{\xi}^*, \mathbf{z}_i^*)$, each of which results in a $p_{K_{iM_i}^{(j)} q}(V_{iM_i}, t, \mathbf{X}_{ij}, \mathbf{z}_i^*, \boldsymbol{\xi}^*, \hat{\boldsymbol{\sigma}})$, a realization of (3.7) from its posterior distribution. The credible interval formed by the $\alpha/2$ -th and $(1 - \alpha/2)$ -th sample quantiles of the posterior sample of (3.7) is then used as a $1 - \alpha$ prediction interval for (3.7). The posterior sample mean and median are two alternative point predictors for (3.7) to the empirical Bayes estimator.

3.3 Simulation study

3.3.1 Simulation setting

A Monte Carlo simulation study was performed to evaluate the finite-sample performance of the estimation and inference methods for the regression parameters and the baseline intensity functions, and the prediction method for the future transition probability. The simulation study was conducted for two sample sizes $m = 100$ and $m = 200$, and for $n = 8$ tooth-level processes. Weibull intensities, $h_{01}^{(0)}(t) = 0.12t^{1.2-1}$ and $h_{12}^{(0)}(t) = 0.18t^{0.9-1}$ were used as the baseline intensities. Seven dummy variables X_1 through X_7 were generated to indicate the eight tooth sites, with the first tooth serving as the reference. Additionally, one continuous tooth-specific covariate X_8 was generated from a standard normal distribution. The coefficients of (X_1, \dots, X_8) were chosen to be $(\beta_{01}^{(1)}, \beta_{01}^{(2)}, \beta_{01}^{(3)}, \beta_{01}^{(4)}, \beta_{01}^{(5)}, \beta_{01}^{(6)}, \beta_{01}^{(7)}, \beta_{01}^{(8)}) = (0.2, -0.1, 0.1, 0.3, 0.2, -0.3, 0.1, 0.2)$ for the 0-to-1 transition and $(\beta_{12}^{(1)}, \beta_{12}^{(2)}, \beta_{12}^{(3)}, \beta_{12}^{(4)}, \beta_{12}^{(5)}, \beta_{12}^{(6)}, \beta_{12}^{(7)}, \beta_{12}^{(8)}) = (0.5, -0.2, 0.3, 0.2, 0.2, -0.3, 0.2, -0.1)$ for the 1-to-2 transition. We set σ_{01}, σ_{12} and ρ as 1, 0.8 and 0.8, respectively. Three dental exam times were simulated as follows: $V_1 \sim U(1.6, 2.4)$, $V_2 = V_1 + U(1.6, 2.4)$, and $V_3 = V_2 + U(3.2, 4.8)$. And finally, 1000 Monte Carlo samples were generated for each of the two sample sizes.

3.3.2 Simulation results

Tables 3.1 and 3.2 show the simulation results for estimating $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_{01}^T, \boldsymbol{\beta}_{12}^T)^T$. The mean estimates are close to the true parameter values. The empirical coverage probabilities of the 95% confidence intervals based on the asymptotic normal distribution of $\hat{\boldsymbol{\beta}}$ are centered around the nominal level. The standard deviation of the estimates (empirical standard error) for each component in $\boldsymbol{\beta}$ is almost the same as the mean of the estimated asymptotic standard errors. The mean squared errors (MSEs) and the empirical variances decrease roughly by half as the number of subjects doubles, which is consistent with the asymptotic theory.

Figure 3.1 shows that the average baseline intensity estimates are reasonably close to the true baseline intensities over the interval $[1.6, 9.6]$, with the endpoints being the lower and the upper

Table 3.1: The simulation results for estimating β_{01}

(0-to-1 transition)	$\beta_{01}^{(1)}$	$\beta_{01}^{(2)}$	$\beta_{01}^{(3)}$	$\beta_{01}^{(4)}$	$\beta_{01}^{(5)}$	$\beta_{01}^{(6)}$	$\beta_{01}^{(7)}$	$\beta_{01}^{(8)}$
Truth	0.2	-0.1	0.1	0.3	0.2	-0.3	0.1	0.2
<i>m=100</i>								
Estimate	0.201	-0.106	0.096	0.301	0.193	-0.304	0.098	0.194
Coverage	0.955	0.940	0.954	0.956	0.954	0.953	0.958	0.961
MSE	0.033	0.035	0.032	0.031	0.032	0.035	0.032	0.012
Bias	0.001	0.006	0.004	0.001	0.007	0.004	0.002	0.006
Empirical SE	0.181	0.186	0.179	0.177	0.178	0.186	0.178	0.109
Mean of SEs	0.181	0.185	0.182	0.180	0.181	0.188	0.182	0.113
<i>m=200</i>								
Estimate	0.199	-0.100	0.105	0.303	0.200	-0.299	0.097	0.195
Coverage	0.950	0.955	0.959	0.954	0.954	0.943	0.964	0.956
MSE	0.016	0.016	0.016	0.016	0.015	0.017	0.015	0.006
Bias	0.001	0.000	0.005	0.003	0.000	0.001	0.003	0.005
Empirical SE	0.128	0.125	0.125	0.127	0.124	0.132	0.124	0.076
Mean of SEs	0.128	0.130	0.128	0.127	0.128	0.133	0.129	0.079

Table 3.2: The simulation results for estimating β_{12}

(1-to-2 transition)	$\beta_{12}^{(1)}$	$\beta_{12}^{(2)}$	$\beta_{12}^{(3)}$	$\beta_{12}^{(4)}$	$\beta_{12}^{(5)}$	$\beta_{12}^{(6)}$	$\beta_{12}^{(7)}$	$\beta_{12}^{(8)}$
Truth	0.5	-0.2	0.3	0.2	0.2	-0.3	0.2	-0.1
<i>m=100</i>								
Estimate	0.509	-0.191	0.308	0.222	0.205	-0.281	0.211	-0.104
Coverage	0.966	0.961	0.938	0.960	0.958	0.943	0.947	0.942
MSE	0.053	0.062	0.060	0.054	0.056	0.074	0.060	0.012
Bias	0.009	0.009	0.008	0.022	0.005	0.019	0.011	0.000
Empirical SE	0.230	0.249	0.245	0.231	0.237	0.272	0.245	0.108
Mean of SEs	0.242	0.259	0.245	0.241	0.244	0.269	0.246	0.105
<i>m=200</i>								
Estimate	0.484	-0.197	0.291	0.189	0.197	-0.298	0.188	-0.099
Coverage	0.959	0.948	0.961	0.954	0.954	0.953	0.950	0.948
MSE	0.027	0.033	0.028	0.028	0.028	0.035	0.030	0.005
Bias	0.016	0.003	0.009	0.011	0.003	0.002	0.012	0.001
Empirical SE	0.164	0.181	0.166	0.167	0.166	0.188	0.174	0.072
Mean of SEs	0.170	0.182	0.172	0.170	0.171	0.189	0.173	0.073

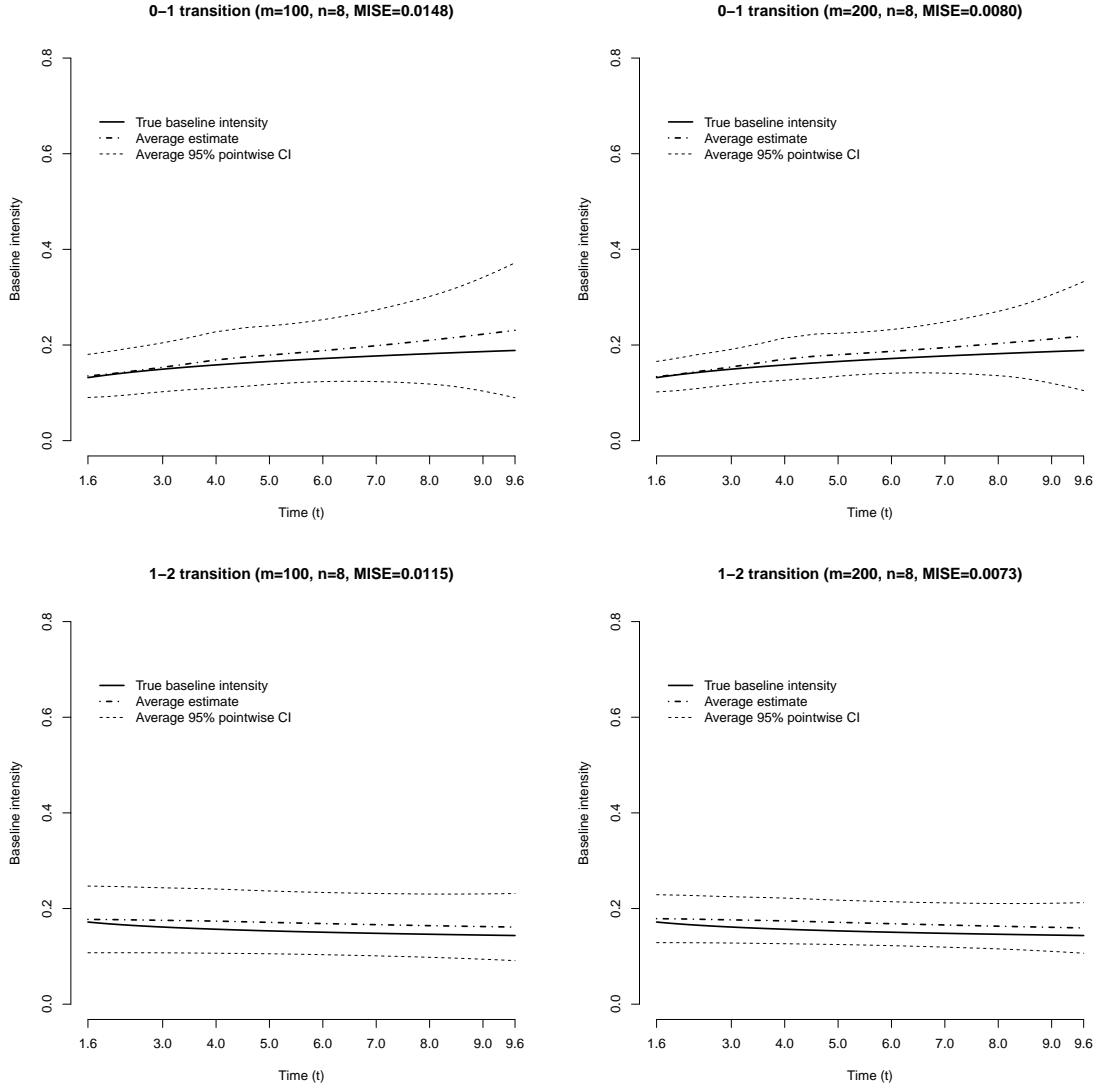


Figure 3.1: The average estimates of $h_{01}^{(0)}(t)$ and $h_{12}^{(0)}(t)$, their average 95% pointwise confidence intervals and their mean integrated squared errors (MISE) across the 1000 Monte Carlo samples

bounds of V_1 and V_3 respectively. The mean integrated squared errors (MISE), $E[\int_{1.6}^{9.6} \{\hat{h}_{ab}^{(0)}(t) - h_{ab}^{(0)}(t)\}^2 dt]$ ($ab = 01, 12$), decrease with the sample size m .

The average 95% pointwise confidence intervals for the baseline intensities are of proper widths and contain the true values at all the time points. The posterior standard deviations of the baseline intensities in the calculation of the Bayesian confidence intervals were obtained by the delta method, and they are treated as the theoretical standard errors of the baseline intensity estimators. The pointwise averages of these theoretical standard errors over the 1000 Monte Carlo

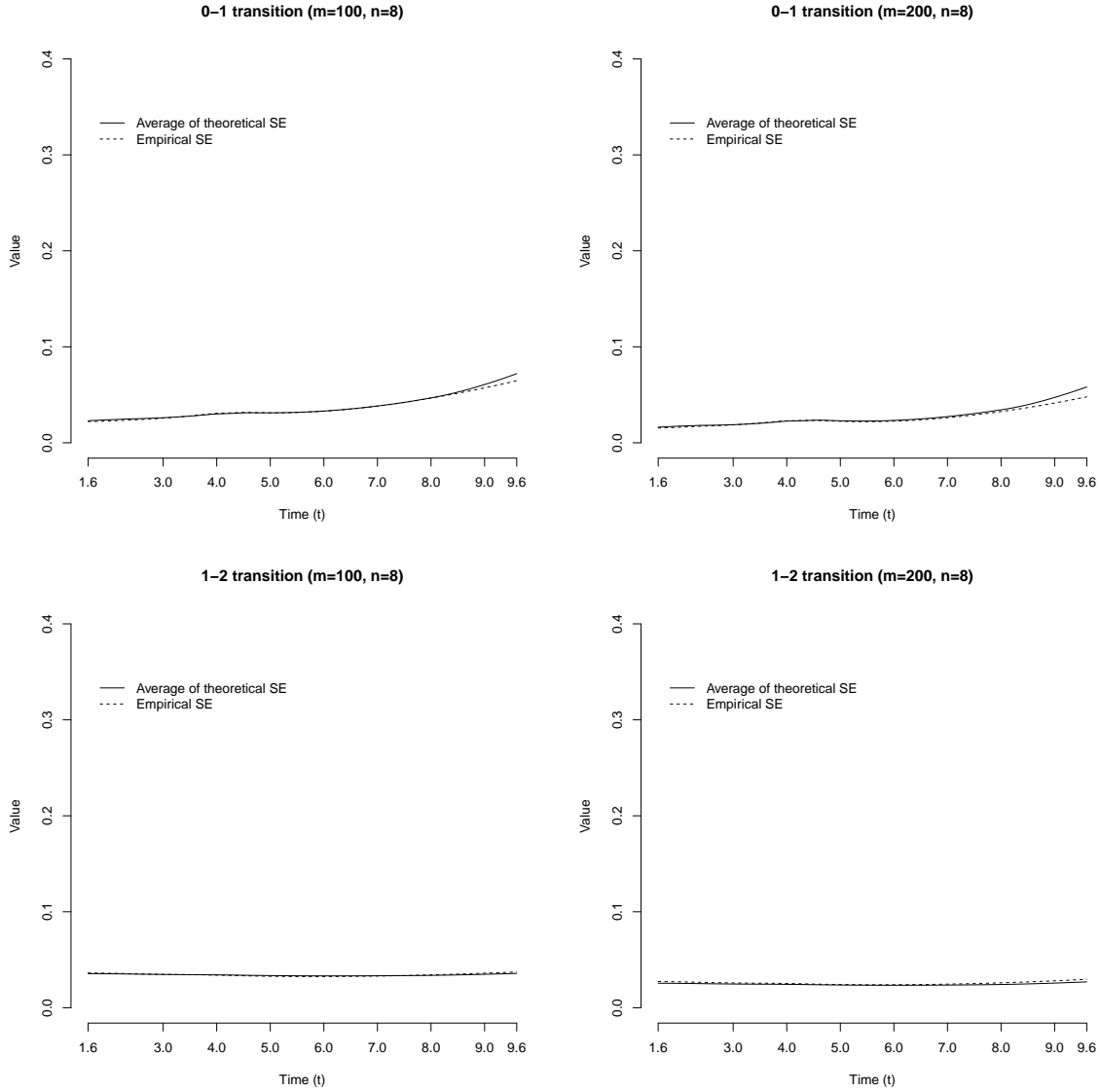


Figure 3.2: The comparison between the empirical standard errors and the average theoretical standard errors of the baseline intensity estimators

samples are shown to be close to the empirical standard errors of the baseline intensity estimators in Figure 3.2, except for $\hat{h}_{01}^{(0)}(t)$ near the upper boundary. In addition, both standard errors reduce as the sample size increases.

Figure 3.3 displays the empirical coverage probabilities of the 95% pointwise confidence intervals for the baseline intensities. Most of the coverage probabilities are scattered around 95% except when the time point is near the upper boundary, which shows that the Bayesian confidence interval

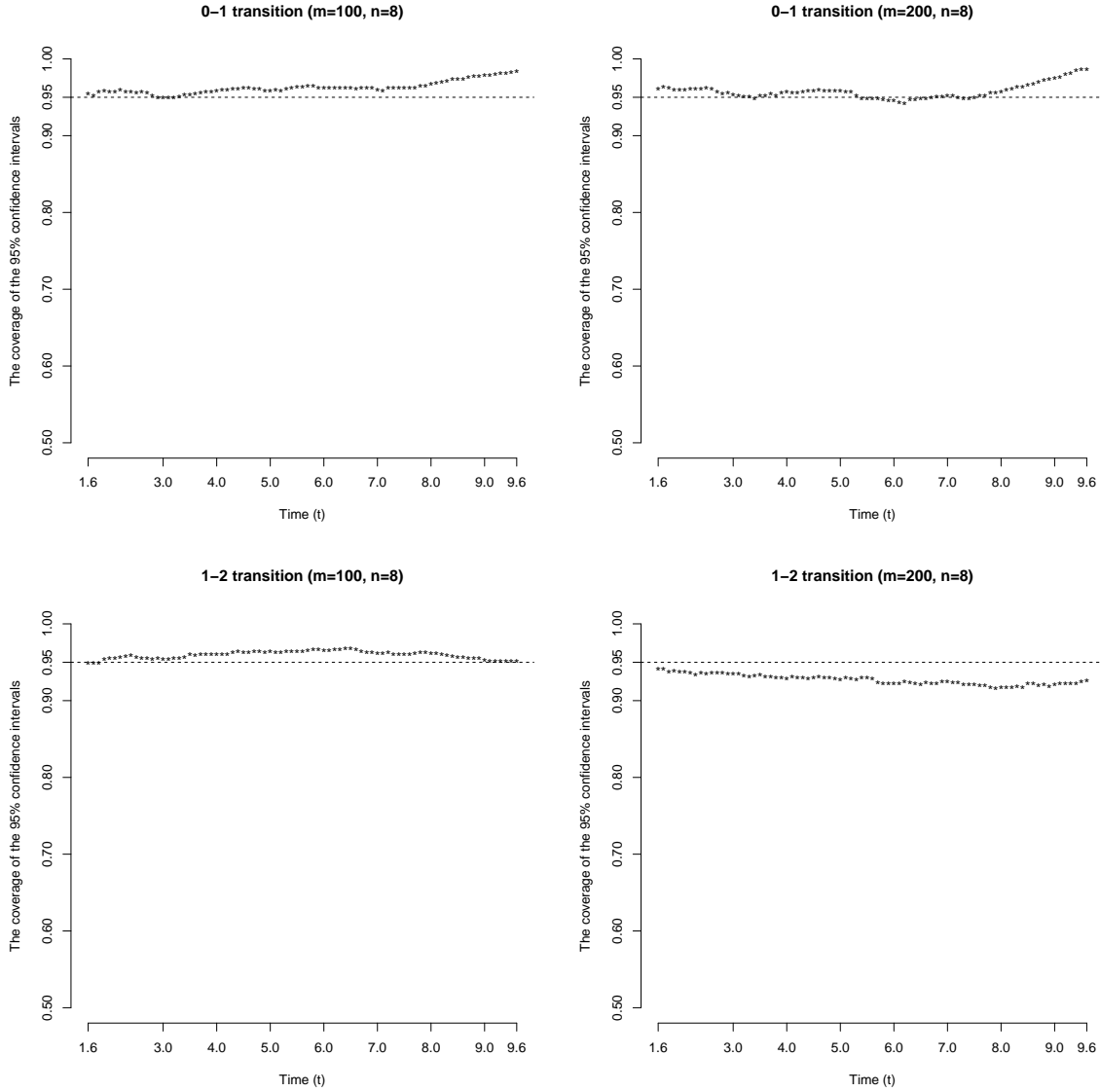


Figure 3.3: The empirical coverage probabilities of the 95% pointwise confidence intervals for the baseline intensities

has a good frequentist performance.

To evaluate the accuracy of the prediction procedure for the future transition probability described in Section 3.2.3, we picked the last tooth (the 8-th) of the last (the m -th) subject in each Monte Carlo sample and checked whether the 95% prediction interval contains the true probability of that tooth remaining in the last observed caries state at several future time points. We considered five future time points, which were obtained by adding 0.1, 0.3, 0.5, 1 and 1.5 to the last visit time

Table 3.3: The proportion of the Monte Carlo iterations where the 95% prediction interval contains the true future transition probability for the last tooth of the last subject at $V_{mM_m} + t^*$ where $t^* = 0.1, 0.3, 0.5, 1, 1.5$

m	$V_{mM_m} + t^*$				
	$t^* = 0.1$	0.3	0.5	1	1.5
100	0.9445	0.9428	0.9428	0.9428	0.9376
200	0.9381	0.9381	0.9434	0.9434	0.9416

respectively. Five hundred posterior realizations of (ξ^*, \mathbf{z}_i^*) were used to compute the prediction intervals. Table 3.3 shows the coverage rates of the 95% prediction intervals at the future time points. They are all close to 0.95. The Monte Carlo iterations where the m -th subject has an advanced caries state at the last visit were not considered in the calculation of the coverage rates because that is an absorbing state. The number of removed iterations was 423 for $m = 100$ and 435 for $m = 200$.

Overall, the estimation and inference methods for regression parameters and baseline intensities as well as the prediction procedure for the future transition probability all perform reasonably well when the number of subjects is no less than 100 and the number of teeth considered is at least eight.

3.4 Application

We applied the proposed methods to the analysis of the DDHP data. This analysis was restricted to caries data on first and second molars of each quadrant, resulting in $n = 8$ teeth per child being studied. One subject was removed from the analysis data set because that subject's birth date was missing, leading to $m = 1020$ including 487 boys and 533 girls. The caries state of a tooth may be missing at a study visit if some tooth surface is not visible; the tooth is missing for reasons other than caries; or the child simply missed the visit. In these instances, the observation was removed from the analysis data set, assuming a missing at random mechanism. The remaining observations are then used to study the intra-oral distribution of dental caries by contrasting for example the transition intensities for the first molar vs. second molar, the upper molar (maxilla) vs. lower (mandible) molar, and right quadrant vs. left quadrant. More importantly, they are used to predict tooth-level future transition probabilities for caries formation.

Table 3.4: The regression parameter estimates for the saturated model based on the DDHP data

Gender	Molar Type	Quadrant	Transition	Parameter	Estimate	SE	p-value
Male	1st molar	Upper right tooth (54)	0 \rightarrow 1	$\beta_{1,1,1,01}$	The reference category		
			1 \rightarrow 2	$\beta_{1,1,1,12}$			
		Upper left tooth (64)	0 \rightarrow 1	$\beta_{1,1,2,01}$	-0.0200	0.1147	0.8614
			1 \rightarrow 2	$\beta_{1,1,2,12}$	0.1208	0.1681	0.4724
		Lower left tooth (74)	0 \rightarrow 1	$\beta_{1,1,3,01}$	0.3410	0.1118	0.0023
			1 \rightarrow 2	$\beta_{1,1,3,12}$	0.2174	0.1591	0.1719
		Lower right tooth (84)	0 \rightarrow 1	$\beta_{1,1,4,01}$	0.3580	0.1125	0.0015
			1 \rightarrow 2	$\beta_{1,1,4,12}$	0.2139	0.1581	0.1762
	2nd molar	Upper right tooth (55)	0 \rightarrow 1	$\beta_{1,2,1,01}$	1.4652	0.1075	< 0.0001
			1 \rightarrow 2	$\beta_{1,2,1,12}$	-0.3856	0.1504	0.0104
		Upper left tooth (65)	0 \rightarrow 1	$\beta_{1,2,2,01}$	1.5330	0.1081	< 0.0001
			1 \rightarrow 2	$\beta_{1,2,2,12}$	-0.3883	0.1498	0.0095
		Lower left tooth (75)	0 \rightarrow 1	$\beta_{1,2,3,01}$	1.4061	0.1078	< 0.0001
			1 \rightarrow 2	$\beta_{1,2,3,12}$	-0.3307	0.1517	0.0293
		Lower right tooth (85)	0 \rightarrow 1	$\beta_{1,2,4,01}$	1.4426	0.1076	< 0.0001
			1 \rightarrow 2	$\beta_{1,2,4,12}$	-0.3115	0.1496	0.0373
Female	1st molar	Upper right tooth (54)	0 \rightarrow 1	$\beta_{2,1,1,01}$	0.1274	0.1617	0.4306
			1 \rightarrow 2	$\beta_{2,1,1,12}$	0.0096	0.1624	0.9530
		Upper left tooth (64)	0 \rightarrow 1	$\beta_{2,1,2,01}$	0.2324	0.1608	0.1484
			1 \rightarrow 2	$\beta_{2,1,2,12}$	0.0483	0.1582	0.7599
		Lower left tooth (74)	0 \rightarrow 1	$\beta_{2,1,3,01}$	0.1805	0.1612	0.2628
			1 \rightarrow 2	$\beta_{2,1,3,12}$	0.2544	0.1584	0.1083
		Lower right tooth (84)	0 \rightarrow 1	$\beta_{2,1,4,01}$	0.2326	0.1606	0.1477
			1 \rightarrow 2	$\beta_{2,1,4,12}$	0.2647	0.1581	0.0941
	2nd molar	Upper right tooth (55)	0 \rightarrow 1	$\beta_{2,2,1,01}$	1.5512	0.1569	< 0.0001
			1 \rightarrow 2	$\beta_{2,2,1,12}$	-0.3185	0.1467	0.0300
		Upper left tooth (65)	0 \rightarrow 1	$\beta_{2,2,2,01}$	1.6083	0.1573	< 0.0001
			1 \rightarrow 2	$\beta_{2,2,2,12}$	-0.2099	0.1457	0.1496
		Lower left tooth (75)	0 \rightarrow 1	$\beta_{2,2,3,01}$	1.8307	0.1579	< 0.0001
			1 \rightarrow 2	$\beta_{2,2,3,12}$	-0.2834	0.1454	0.0512
		Lower right tooth (85)	0 \rightarrow 1	$\beta_{2,2,4,01}$	1.8725	0.1580	< 0.0001
			1 \rightarrow 2	$\beta_{2,2,4,12}$	-0.3777	0.1458	0.0096

Because each child has a total of eight primary molars, we considered a saturated model with fifteen dummy covariate variables indicating the combination of tooth identity and gender. The upper right first molar for boys was set as the reference category. Seven knots were used for approximating $\log h_{01}^{(0)}(t)$ and $\log h_{12}^{(0)}(t)$ by linear splines. The total number of parameters in the model is 51 including the spline parameters, the regression coefficients and the variance components of the log frailties for both types of transitions.

The regression coefficient estimates and their standard errors for the saturated model are shown in Table 3.4. The subscripts of β represent gender (1= Male, 2= Female), molar type (1=First molar, 2=Second molar), quadrant (1=Upper right, 2=Upper left, 3=Lower left, 4=Lower right) and transition type (01=0-to-1 transition, 12=1-to-2 transition) in order. For instance, $\beta_{1,1,1,01}$ and $\beta_{1,1,1,12}$ are the coefficients for 0-1 and 1-2 transitions of the upper right first molar in a boy, the reference category, respectively, thus $\beta_{1,1,1,01} = \beta_{1,1,1,12} = 0$. It is of scientific interest to compare tooth decay profiles between first and second molars, between upper and lower molars, and between right and left molars. These symmetry evaluations need to be stratified by gender if there exists an interaction effect between gender and tooth identity (type and quadrant) on the tooth decay profile. All these questions can be answered by testing a hypothesis of the form $H_0 : G\beta = \mathbf{0}$, where G is an appropriate matrix with full row rank, denoted by r , and $\mathbf{0}$ is a zero vector of length r . Based on the asymptotic normality of $\hat{\beta}$, the corresponding Wald test is,

$$(G\hat{\beta})^T (GV\hat{ar}(\hat{\beta})G^T)^{-1} G\hat{\beta} \overset{H_0}{\rightsquigarrow} \chi_r^2,$$

where $V\hat{ar}(\hat{\beta})$ is the sub-matrix of $-H_h^{-1}(\hat{\eta}; \mathbf{Y}, \hat{\nu})$ corresponding to β .

The null hypothesis for testing the interaction between gender and tooth identity can be formulated as

$$H_0 : \beta_{1,t,l,ab} - \beta_{1,1,1,ab} = \beta_{2,t,l,ab} - \beta_{2,1,1,ab} \text{ for all } t, l, ab.$$

This null hypothesis was rejected at 5% significant level (p-value = 0.0016) based on the data. Hence we evaluate whether there are any intra-oral symmetries with respect to the tooth decay profile for each gender. The null hypothesis of comparing the tooth decay profile between first and second molars for each gender is,

$$H_0 : \beta_{g,1,l,ab} = \beta_{g,2,l,ab} \text{ for all } l, ab,$$

where g takes values 1 or 2. These two null hypotheses were rejected (both p-values less than 0.0001). The null hypothesis of comparing the tooth decay profile between upper and lower molars for each gender is,

$$H_0 : \beta_{g,t,1,ab} = \beta_{g,t,4,ab} \text{ } \beta_{g,t,2,ab} = \beta_{g,t,3,ab} \text{ for all } t, ab,$$

Table 3.5: The Wald test statistics (p-value) for the three types of symmetry in caries transition intensity stratified by gender and transition type

Gender	Test for symmetry	Transition	
		0 to 1	1 to 2
Male	Upper vs Lower	21.87(0.0002)	2.74(0.6020)
	Right vs Left	0.71(0.9498)	0.54(0.9695)
	First molar vs Second molar	538.75(< 0.0001)	48.35(< 0.0001)
Female	Upper vs Lower	18.96(0.0008)	5.75(0.2180)
	Right vs Left	1.86(0.7617)	1.66(0.7973)
	First molar vs Second molar	775.38(< 0.0001)	51.22(0.0004)

where g takes values 1 or 2. This hypothesis was rejected for each gender (p-value = 0.0011 for boys and p-value= 0.0020 for girls). Lastly, the right-left symmetry in tooth decay for each gender can be tested using the null hypothesis:

$$H_0 : \beta_{g,t,1,ab} = \beta_{g,t,2,ab}, \beta_{g,t,3,ab} = \beta_{g,t,4,ab} \text{ for all } t, ab, \quad (3.8)$$

where g takes values 1 or 2. The corresponding test statistics are 1.256 (p-value = 0.9961) for boys and 3.690 (p-value= 0.8840) for girls, which indicates that the right-left symmetry in tooth decay exists for both genders.

We further performed the tests for the three types of symmetries in caries transition intensity stratified by gender and transition type. The Wald test statistics and the p-values for these tests are shown in Table 3.5. Since the null hypotheses (3.8) were not rejected, it is expected that none of the tests comparing the right and left molars in Table 3.5 was rejected. The detailed comparisons associated with the significant non-symmetries found in Table 3.5 (at 5% level) were conducted in order to find out where the difference is. The results of these specific tests can be found in Table 3.6. The upper-lower non-symmetry in 0-to-1 transition intensity for boys is due to lower first molars transitioning faster from state 0 to 1 than upper first molars in boys. The first vs. second difference in 0-to-1 transition intensity for boys is because of the first molar transitioning slower from 0 to 1 than the second molar in every quadrant of a boy's mouth. In contrast, the first molar transitions faster from state 1 to 2 than the second molar in every quadrant of a boy's mouth, which explains the first vs. second non-symmetry in 1-to-2 transition intensity for boys. The upper-lower

Table 3.6: The detailed comparisons associated with the significant non-symmetries in caries transition intensity found in Table 3.5

Gender	Transition	Teeth Compared		Wald Test Statistic (p-value)
Male	0 → 1	Upper right first	Lower right first	-3.181 (0.0007)
		Upper left first	Lower left first	-3.200 (0.0007)
		Upper right second	Lower right second	0.236 (0.4068)
		Upper left second	Lower left second	1.323 (0.0928)
	0 → 1	Upper right first	Upper right second	-13.625 (< 0.0001)
		Upper left first	Upper left second	-14.212 (< 0.0001)
		Lower right first	Lower right second	-10.382 (< 0.0001)
		Lower left first	Lower left second	-10.261 (< 0.0001)
	1 → 2	Upper right first	Upper right second	2.563 (0.0052)
		Upper left first	Upper left second	3.443 (0.0003)
		Lower right first	Lower right second	3.864 (0.0001)
		Lower left first	Lower left second	3.934 (< 0.0001)
Female	0 → 1	Upper right first	Lower right first	-0.988 (0.1592)
		Upper left first	Lower left first	0.494 (0.3015)
		Upper right second	Lower right second	-3.489 (0.0002)
		Upper left second	Lower left second	-2.423 (0.0077)
	0 → 1	Upper right first	Upper right second	-14.055 (< 0.0001)
		Upper left first	Upper left second	-13.681 (< 0.0001)
		Lower right first	Lower right second	-16.200 (< 0.0001)
		Lower left first	Lower left second	-16.168 (< 0.0001)
	1 → 2	Upper right first	Upper right second	2.424 (0.0077)
		Upper left first	Upper left second	2.000 (0.0227)
		Lower right first	Lower right second	4.966 (< 0.0001)
		Lower left first	Lower left second	4.167 (< 0.0001)

non-symmetry in 0-to-1 transition intensity for girls is due to lower second molars transitioning faster from state 0 to 1 than upper second molars in girls. The first vs. second difference in 0-to-1 transition intensity for girls is because of the first molar transitioning slower from 0 to 1 than the second molar in every quadrant of a girl's mouth, but it is the opposite for comparing the first molars with the second in 1-to-2 transition intensity for girls: the first molar transitions faster from 1 to 2 than the second in every quadrant of a girl's mouth.

We predicted future transition probabilities for two molars of a boy in the data set who had only two study visits at 2.15 and 4.25 years of age respectively. The boy is the same child who was used for the data analysis in Chapter 2. The molars of interest were his upper right first and second

molars. His upper right second molar was sound at the first visit but was in the early caries state at the second visit, whereas his upper right first molar remained sound over the two visits. For each molar, the empirical Bayes estimates and the 95% prediction intervals for the probabilities of all possible transitions after his last visit were computed up to age 8 using 500 posterior realizations of (ξ^*, \mathbf{z}_i^*) . The left-side plot in Figure 3.4 shows the probabilities of the 0-to-0, 0-to-1 and 0-to-2 transitions for his upper right first molar, and the right-side plot shows the probabilities of the 1-to-1 and 1-to-2 transitions for his upper right second molar. The probabilities for the molars to stay in the current state get smaller as the child ages, and those for the molars to develop caries get greater as time goes by. Dentists can decide the proper timing of future dental exams for this child based on these probabilities.

The correlation parameter ρ in the distribution of the bivariate frailty could be impacted by covariates. To examine the heterogeneity in ρ and assess its impact on the analysis results, we fit our model with ρ being gender-specific to the Detroit data. The analysis results are given in Appendix A and they are very similar to the results obtained from assuming a homogeneous ρ , although the likelihood ratio test for the heterogeneity based on the adjusted profile h -likelihood (3.6) showed that gender has a statistically significant effect on ρ (p-value = 0.0343).

3.5 Discussion

We have proposed a semiparametric three-state Markov frailty model for tooth-level caries life course data subject to interval censoring. Estimation of the model proceeds by maximizing a penalized likelihood, imposing a constraint on the coefficients of the linear splines that approximate the unknown baseline intensities. The linear spline approximation is computationally convenient as it avoids certain numerical integrations with respect time appearing in the log-likelihood function, the score function and the hessian matrix. The penalized likelihood resembles the joint likelihood of a mixed model, treating the coefficients of the spline as random effects. Using the mixed-model representation avoids the classical problem of penalty parameter tuning. Because the marginal maximum likelihood estimation for the mixed-effects model involves high-dimensional intractable

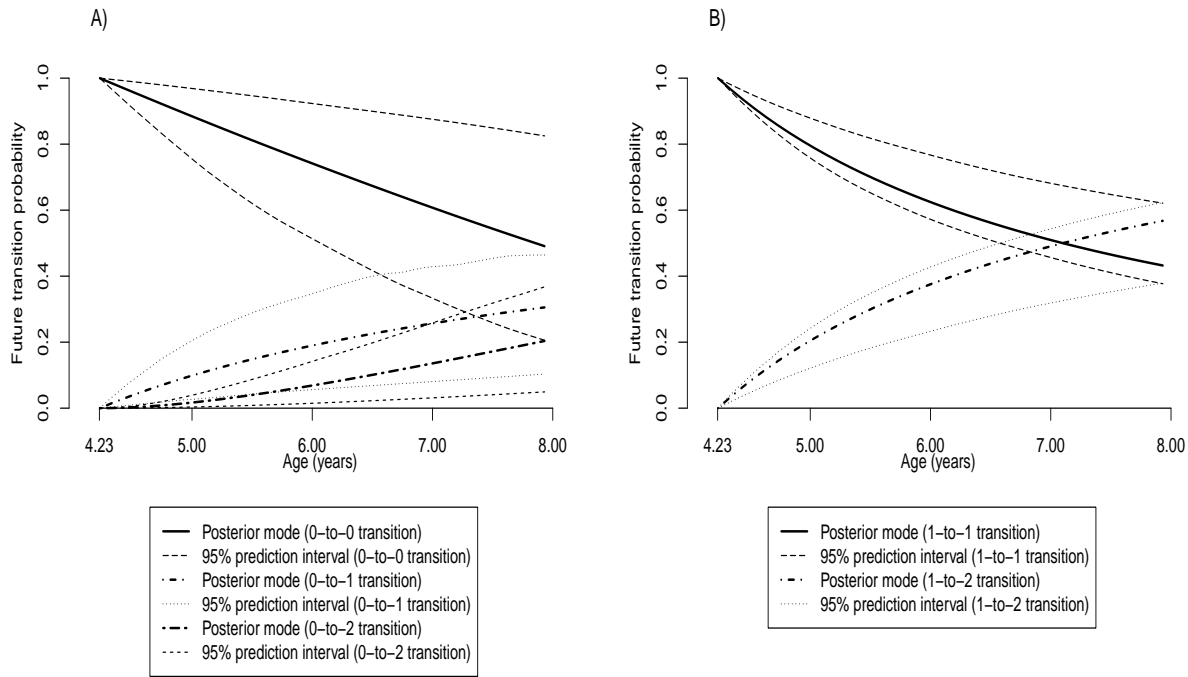


Figure 3.4: The future transition probabilities to every possible state out of the caries state at the last visit for the upper right first molar (A) and the upper right second molar (B)

integrals, we adopted the hierarchical likelihood approach of Lee and Nelder (1996) for model fitting and inference. In addition, we developed a Bayesian approach, as a by-product of the model estimation, to predict future caries transition probabilities at the tooth level. Tooth-level caries transition risks are useful for prevention and treatment planning in dental care.

The simulation study demonstrated that the estimation and inference methods for the regression parameters and baseline intensities have a good performance. The Bayesian prediction interval for the future transition probability is also shown to have a good frequentist property. The methodology is applied to the DDHP data to describe the intra-oral life-history of caries, with an emphasis on spatial symmetry evaluation. Our statistical methods are generally applicable to cohort studies and medical records involving chronic diseases in which the disease progression is assessed at various locations of the body and at intermittent time points.

The model has some limitations. For example, to account for the intra-oral clustering, we assume that the intensities of different teeth share a bivariate frailty. This restriction can be relaxed

by allowing quadrant-specific frailties. Also, the conditional Markov assumption might not be realistic and could be replaced by a conditional semi-Markov assumption. These extensions of the model merit future research.

CHAPTER 4

VARIABLE SELECTION IN THE PROPORTIONAL HAZARDS MODEL FOR INTERVAL CENSORED DATA

4.1 Introduction

Epidemiologic/clinical studies usually collect a large number of variables such as subjects' biomarkers, genotypes, demographic characteristics, and environmental risk factors to study their relationships with the outcome. Building a regression model including all those variables is often undesirable because it has low prediction accuracy and is hard to interpret. Therefore, variable selection is an important topic in regression modeling. There have been many statistical methods on variable selection. Among them, the traditional method of best subset selection is being generally used in many applications, but it is known to be unstable as resulting in poor predictive performance (Breiman, 1996). Another popular class of methods, which is more stable, is variable selection via regularization, also known as penalized variable selection. This class of methods can simultaneously determine which variables to put in a regression model and estimate their effects on the outcome. Popular penalized variable selection methods include Lasso (Tibshirani, 1996), the smoothy-clipped absolute deviation (SCAD; Fan & Li, 2001) and adaptive Lasso (Zou, 2006). The latter two methods enjoy the so-called oracle property: when the sample size is large, they build the regression model as if they knew which variables are important and which are not.

Most penalized variable selection methods were proposed for linear or generalized linear models, where the outcome is fully observed. There are some methods for survival analysis. For example, Lasso, SCAD and adaptive Lasso have been extended to the Cox model for right censored data by Tibshirani (1997), Fan & Li (2002) and Zhang & Lu (2007) respectively. However, to the best of our knowledge, there is only one published method (Wu & Cook, 2015) on penalized variable selection for interval censored data. Wu & Cook (2015) assumes a proportional hazards model with a piecewise constant baseline hazard function, and performs the variable selection through

penalized likelihood estimation with a LASSO, SCAD or adaptive LASSO penalty. They did not prove any asymptotic property of the method and used an ad hoc approach to specify the number and location of break points for the piecewise constant baseline hazard function.

In this chapter, we propose a penalized variable selection method for the Cox proportional hazards model of interval censored data and investigate the performance of the proposed model in realistic samples, in terms of both variable selection and estimation consistency. The model is semiparametric because its baseline hazard function is completely unspecified. The nonparametric maximum likelihood estimator for the baseline hazard function is considered with the support set which was introduced by Alioum & Commenges (1996). The variable selection is performed via a penalized nonparametric maximum likelihood estimation with an adaptive Lasso penalty. In the model estimation procedure, we first obtain the unpenalized nonparametric maximum likelihood estimators for the regression parameters as following the EM algorithm in Zeng et al. (2016) with some modifications, and use these consistent estimates in the penalized likelihood estimation procedure with the adaptive LASSO penalty. We prove that our method has the oracle property and give an approach to estimate the covariance matrix of the penalized estimator. In addition, we extend the method to left truncated and interval censored data.

To evaluate model performance, we compared our proposed method to the existing methods introduced by Zhang & Lu (2007) and Wu & Cook (2015) under interval censoring. Because Zhang and Lu's method is designed for right censored data, the evaluation on their method is conducted using mid-point imputations of the censoring intervals, which is a common practice in many studies.

The application of the proposed model in this chapter is performed with the mouth-level DDHP data described in Chapter 1. We consider the event that a child has at least one noncavitated/cavitated lesion in the surfaces of all teeth. Unlike Chapter 2 and Chapter 3, we have only one censored interval for each child. In addition, we use the child-level, family-level or community-level influence variables as covariates of the model to determine social and behavioral factors that play a critical role in developing dental caries in deciduous dentition.

In Section 4.2, we introduce a new variable selection approach via adaptive Lasso estimation for interval censored time-to-event data. Followed by Section 4.2, the asymptotic property of the adaptive Lasso estimator from the proposed model is shown in Section 4.3. The extension of the proposed model to and interval censored data is demonstrated in Section 4.4. We also conduct the simulation study in Section 4.5, as comparing the proposed model with the existing variable selection approaches. The simulation study with left truncated and interval censored data is also described in this section. The practical utility of the proposed variable selection approach is illustrated with the mouth-level DDHP data in Section 4.6. The concluding remarks are given in Section 4.7.

4.2 Statistical methodologies

We consider a random sample of n independent subjects. Let T_i and \mathbf{X}_i respectively denote the time to event of interest and a p -dimensional vector of covariates for subject i ($i = 1, \dots, n$). We study how to select significant covariates among \mathbf{X}_i for the time-to-event outcome T_i in the situation where T_i is subject to mixed case interval censoring (Schick & Yu, 2000). Denote the sequence of inspection times for subject i by $\vec{V}_i = (V_{i1}, \dots, V_{iK_i})^T$. Define $\Delta_{ik} = I(V_{i,k-1} < T_i \leq V_{ik})$ ($k = 1, \dots, K_i$) with $V_{i0} = 0$, and $\vec{\Delta}_i = (\Delta_{i1}, \dots, \Delta_{iK_i})^T$. Then the observed data consist of $\{\mathcal{O}_i = (\vec{V}_i, \vec{\Delta}_i, \mathbf{X}_i)\}_{i=1}^n$.

We assume that the inspection process is independent of the time to event given the covariates, i.e. $T_i \perp (\vec{V}_i, K_i) | \mathbf{X}_i$. We also assume that the conditional distribution of T_i given \mathbf{X}_i satisfies the Cox proportional hazards model, i.e.

$$h(t|\mathbf{X}_i) = h(t) \exp(\boldsymbol{\beta}^T \mathbf{X}_i), \quad (4.1)$$

where $h(t|\mathbf{X}_i) \equiv \lim_{\Delta t \rightarrow 0+} \text{pr}(t \leq T_i < t + \Delta t | T_i \geq t, \mathbf{X}_i) / \Delta t$, $h(t)$ is an unspecified baseline hazard function, and $\boldsymbol{\beta}$ is a vector of unknown regression parameters.

4.2.1 Variable selection

We conduct the variable selection via an adaptive Lasso estimation. Let L_i and R_i denote respectively the last inspection time before T_i and the first inspection time after T_i . Set $L_i = 0$ if T_i is smaller than V_{i1} and $R_i = \infty$ if T_i is larger than V_{iK_i} , i.e. T_i is right censored. Under the Cox model (4.1) and the mixed case interval censoring, the logarithm of the observed-data likelihood is

$$l_n(\boldsymbol{\beta}, H) = \sum_{i=1}^n \log \left[\exp\{-H(L_i) \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\} - \exp\{-H(R_i) \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\} \right], \quad (4.2)$$

where $H(t) = \int_0^t h(s)ds$ and the convention, $\exp\{-H(\infty) \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\} = 0$, is used. The adaptive Lasso estimation is

$$\max_{\boldsymbol{\beta}, H} \left\{ l_n(\boldsymbol{\beta}, H) - n\theta \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j| \right\}, \quad (4.3)$$

where $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\beta}_1, \dots, \tilde{\beta}_d)^T$ is the unpenalized nonparametric maximum likelihood estimator for $\boldsymbol{\beta}$, which can be obtained using the EM algorithm in Zeng et al. (2016), θ is a thresholding parameter whose selection is discussed later, and the maximization with respect to H is over the space of nondecreasing nonnegative functions. We denote the estimator from (4.3) by $(\hat{\boldsymbol{\beta}}, \hat{H})$.

Note that the penalty in (4.3) does not involve H . Thus the support set over which $\hat{H}(\cdot)$ increases is the same as that of the unpenalized nonparametric maximum likelihood estimator for H , which was characterized by Alioum & Commenges (1996). Specifically, the estimator \hat{H} increases only on so-called *maximal intersections*: intervals of the form $(l, u]$ where $l \in \{L_i\}_{i=1}^n$, $u \in \{R_i\}_{i=1}^n$ and there is no L_i or R_i in (l, u) . Additionally, \hat{H} is indifferent to how it increases on the maximal intersections, as only the overall jump sizes over $(l, u]$'s ($u < \infty$), $\hat{H}(u) - \hat{H}(l)$, affect the penalized likelihood in (4.3). Write the maximal intersections with a finite upper endpoint as $(l_1, u_1], \dots, (l_m, u_m]$. According to the characterizations of \hat{H} , we can assume, just for the purpose of computing \hat{H} , that H is flat on $\overline{\cup_{j=1}^m (l_j, u_j]}$. Define $h_k = H(u_k) - H(l_k)$ ($k = 1, \dots, m$). Then

the log likelihood $l_n(\boldsymbol{\beta}, H)$ can be written as $l_n(\boldsymbol{\beta}, \mathbf{h})$, where $\mathbf{h} = (h_1, \dots, h_m)^T$, and

$$\begin{aligned} l_n(\boldsymbol{\beta}, \mathbf{h}) &= \sum_{i=1}^n \log \left[\exp\left\{-\sum_{u_k \leq L_i} h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\right\} - I(R_i < \infty) \exp\left\{-\sum_{u_k \leq R_i} h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\right\} \right] \\ &= \sum_{i=1}^n \log \left(\exp\left\{-\sum_{u_k \leq L_i} h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\right\} \left[1 - \exp\left\{-\sum_{L_i < u_k \leq R_i} h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\right\} \right]^{I(R_i < \infty)} \right). \end{aligned}$$

Direct maximization of $l_n(\boldsymbol{\beta}, \mathbf{h}) - n\theta \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j|$ is challenging because of no closed-form expression for the maximizer $\hat{\mathbf{h}}$, whose dimension increases with the sample size. In the spirit of Zeng et al. (2016), we propose an EM algorithm for the adaptive Lasso estimation as follows. Let W_{ik} ($i = 1, \dots, n; k = 1, \dots, m$) be independent Poisson random variables with means $h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)$. Define $A_i = \sum_{u_k \leq L_i} W_{ik}$ and $B_i = I(R_i < \infty) \sum_{L_i < u_k \leq R_i} W_{ik}$. Consider $\tilde{\mathcal{O}} \equiv \{(L_i, R_i, \mathbf{X}_i, A_i = 0) : 1 \leq i \leq n \text{ and } I(R_i) = \infty\} \cup \{(L_i, R_i, \mathbf{X}_i, A_i = 0, B_i > 0) : 1 \leq i \leq n \text{ and } I(R_i) < \infty\}$ as an observed data set, where $A_i = 0$ means that A_i is observed to be zero and $B_i > 0$ means that B_i is observed to be positive. The log likelihood of $\tilde{\mathcal{O}}$ has the form

$$\begin{aligned} &\sum_{i=1}^n \log \left[\left\{ \prod_{u_k \leq L_i} \text{pr}(W_{ik} = 0) \right\} \left\{ 1 - \text{pr} \left(\sum_{L_i < u_k \leq R_i} W_{ik} = 0 \right) \right\}^{I(R_i < \infty)} \right] \\ &= \sum_{i=1}^n \log \left(\exp\left\{-\sum_{u_k \leq L_i} h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\right\} \left[1 - \exp\left\{-\sum_{L_i < u_k \leq R_i} h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\right\} \right]^{I(R_i < \infty)} \right), \end{aligned}$$

which is the same as $l_n(\boldsymbol{\beta}, \mathbf{h})$. Therefore, we can maximize $l_n(\boldsymbol{\beta}, \mathbf{h}) - n\theta \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j|$ via an EM algorithm by treating $\{W_{ik} : i = 1, \dots, n \text{ and } u_k \leq R_i^*\}$, where $R_i^* = L_i I(R_i = \infty) + R_i I(R_i < \infty)$, as the complete data corresponding to $\tilde{\mathcal{O}}$.

The complete-data log likelihood is

$$\sum_{i=1}^n \sum_{k=1}^m I(u_k \leq R_i^*) \left[W_{ik} \log\{h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i)\} - h_k \exp(\boldsymbol{\beta}^T \mathbf{X}_i) - \log W_{ik}! \right]. \quad (4.4)$$

At the E-step, we compute, $\hat{E}(W_{ik})$'s, the conditional means of W_{ik} 's given the observed data $\tilde{\mathcal{O}}$ and the current parameter updates $(\boldsymbol{\beta}^{(s)}, \mathbf{h}^{(s)})$ ($s = 0, 1, \dots$). For $u_k \leq L_i$, $\hat{E}(W_{ik}) = 0$ since $A_i = 0$.

For $L_i < u_k \leq R_i$ with $R_i < \infty$,

$$\begin{aligned}\hat{E}(W_{ik}) &= E(W_{ik} | L_i, R_i, \mathbf{X}_i, A_i = 0, B_i > 0) \\ &= E(W_{ik} | \sum_{L_i < u_j \leq R_i} W_{ij} > 0) \\ &= \frac{h_k^{(s)} \exp(\boldsymbol{\beta}^{(s)T} \mathbf{X}_i)}{1 - \exp\{-\sum_{L_i < u_j \leq R_i} h_j^{(s)} \exp(\boldsymbol{\beta}^{(s)T} \mathbf{X}_i)\}}.\end{aligned}$$

At the M-step, we first maximize the expected complete-data log likelihood with respect to \mathbf{h} conditioning on $\boldsymbol{\beta}$. The maximizer has an analytical expression:

$$h_k^{(s+1)}(\boldsymbol{\beta}) \equiv \frac{\sum_{i=1}^n I(u_k \leq R_i^*) \hat{E}(W_{ik})}{\sum_{i=1}^n I(u_k \leq R_i^*) \exp(\boldsymbol{\beta}^T \mathbf{X}_i)} \quad (k = 1, \dots, m).$$

Plugging $h_k = h_k^{(s+1)}(\boldsymbol{\beta})$ into the conditional expectation of (4.4), we update $\boldsymbol{\beta}$ by maximizing

$$Q(\boldsymbol{\beta}, \mathbf{h}^{(s+1)}(\boldsymbol{\beta}) | \boldsymbol{\beta}^{(s)}, \mathbf{h}^{(s)}) - n\theta \sum_{j=1}^d |\beta_j| / |\tilde{\beta}_j|,$$

where

$$Q(\boldsymbol{\beta}, \mathbf{h}^{(s+1)}(\boldsymbol{\beta}) | \boldsymbol{\beta}^{(s)}, \mathbf{h}^{(s)}) = \sum_{i=1}^n \sum_{k=1}^m I(u_k \leq R_i^*) \hat{E}(W_{ik}) \left[-\log \left\{ \sum_{j=1}^n I(u_k \leq R_j^*) \exp(\boldsymbol{\beta}^T \mathbf{X}_j) \right\} + \boldsymbol{\beta}^T \mathbf{X}_i \right]$$

and $\mathbf{h}^{(s+1)}(\boldsymbol{\beta}) = (h_1^{(s+1)}(\boldsymbol{\beta}), \dots, h_m^{(s+1)}(\boldsymbol{\beta}))^T$. To perform this maximization, we approximate $-Q(\boldsymbol{\beta}, \mathbf{h}^{(s+1)}(\boldsymbol{\beta}) | \boldsymbol{\beta}^{(s)}, \mathbf{h}^{(s)})$ by a second-order Taylor expansion around $\boldsymbol{\beta}^{(s)}$. It can be written in a quadratic form $2^{-1}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})$, where \mathbf{Z} is from the Cholesky decomposition of $\nabla^2 Q(\boldsymbol{\beta}^{(s)}) \equiv -\partial^2 Q / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T |_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(s)}}$, that is $\nabla^2 Q(\boldsymbol{\beta}^{(s)}) = \mathbf{Z}^T \mathbf{Z}$, and $\mathbf{Y} = (\mathbf{Z}^T)^{-1} \{\nabla^2 Q(\boldsymbol{\beta}^{(s)}) \boldsymbol{\beta}^{(s)} - \nabla Q(\boldsymbol{\beta}^{(s)})\}$ with $\nabla Q(\boldsymbol{\beta}^{(s)}) = -\partial Q / \partial \boldsymbol{\beta} |_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(s)}}$. Then we minimize

$$\frac{1}{2}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}) + n\theta \sum_{j=1}^d |\beta_j| / |\tilde{\beta}_j| \quad (4.5)$$

to obtain $\boldsymbol{\beta}^{(s+1)}$, using the modified shooting algorithm in Zhang & Lu (2007). The corresponding update for \mathbf{h} is $\mathbf{h}^{(s+1)} = \mathbf{h}^{(s+1)}(\boldsymbol{\beta}^{(s+1)})$.

The EM algorithm stops if the maximum of the absolute differences between the estimates at two successive iterations is smaller than, say 10^{-3} . We choose the initial parameter values

$(\boldsymbol{\beta}^{(0)}, \mathbf{h}^{(0)})$ to be the unpenalized nonparametric maximum likelihood estimator $(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{h}})$, which can be obtained from the same EM algorithm as the above except that the objective function (4.5) becomes $2^{-1}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})$.

4.2.2 Covariance matrix of the adaptive Lasso estimator

Note that the estimator $\hat{\boldsymbol{\beta}}$ obtained from (4.3) is equivalent to the solution of the following penalized profile likelihood estimation,

$$\max_{\boldsymbol{\beta}} \left\{ l_{pn}(\boldsymbol{\beta}) - n\theta \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j| \right\}, \quad (4.6)$$

where $l_{pn}(\boldsymbol{\beta}) = \sup_H l_n(\boldsymbol{\beta}, H)$. Adapting the standard error derivation in Section 4.1 of Lu & Zhang (2007) to the Adaptive Lasso estimation (4.6), the covariance matrix of $\hat{\boldsymbol{\beta}}$ can be estimated by a sandwich formula:

$$\{\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}}) + n\theta A(\hat{\boldsymbol{\beta}})\}^{-1} \Sigma(\hat{\boldsymbol{\beta}}) \{\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}}) + n\theta A(\hat{\boldsymbol{\beta}})\}^{-1}, \quad (4.7)$$

where $\nabla^2 l_{pn}(\boldsymbol{\beta})$ is the negative hessian of $l_{pn}(\boldsymbol{\beta})$, $\Sigma(\hat{\boldsymbol{\beta}}) = \{\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}}) + n\theta D(\hat{\boldsymbol{\beta}})\} \{\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}})\}^{-1} \{\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}}) + n\theta D(\hat{\boldsymbol{\beta}})\}$, $A(\boldsymbol{\beta}) = \text{diag}(1/\beta_1^2, \dots, 1/\beta_d^2)$ and $D(\boldsymbol{\beta}) = \text{diag}(I(\beta_1 \neq 0)/\beta_1^2, \dots, I(\beta_d \neq 0)/\beta_d^2)$. Here we take the convention $0/0 = 0$ and set $1/0$ to a very large number, say 10^{10} .

Since $l_{pn}(\boldsymbol{\beta})$ does not have a closed form, we calculate $\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}})$ using a second-order numerical difference as in Murphy & Van Der Vaart (1999), that is,

$$(\nabla^2 l_{pn}(\hat{\boldsymbol{\beta}}))_{i,j} \approx -\frac{l_{pn}(\hat{\boldsymbol{\beta}} + \delta_n e_i + \delta_n e_j) - l_{pn}(\hat{\boldsymbol{\beta}} + \delta_n e_i) - l_{pn}(\hat{\boldsymbol{\beta}} + \delta_n e_j) + l_{pn}(\hat{\boldsymbol{\beta}})}{\delta_n^2}, \quad (4.8)$$

where e_i is the i th unit vector in \mathbb{R}^d and $\delta_n = O_p(n^{-1/2})$. The value of $l_{pn}(\boldsymbol{\beta})$ can be evaluated using the EM algorithm in Section 4.2.1 with $\boldsymbol{\beta}$ held fixed.

4.2.3 Thresholding parameter tuning

Like all the variable selection methods via regularization, the performance of our adaptive Lasso estimator $\hat{\boldsymbol{\beta}}_\theta$ depends critically on the choice of the thresholding parameter θ . To select it, we

minimize the following Bayesian information criterion (BIC),

$$\text{BIC}(\theta) = -2l_{pn}(\hat{\boldsymbol{\beta}}_\theta) + |\alpha_\theta| \log(n), \quad (4.9)$$

where $\alpha_\theta = \{j : \hat{\boldsymbol{\beta}}_{\theta,j} \neq 0\}$ is the active set identified by the adaptive Lasso estimation with thresholding parameter θ , and $|\alpha_\theta|$ is its size. For generalized linear models with a fixed number of covariates, the adaptive Lasso with the thresholding parameter selected using BIC identifies the true model consistently (Zhang et al., 2010; Hui et al., 2015). This motivates us to use BIC to select the thresholding parameter in our variable selection method for interval censored data.

The grid points of θ is constructed by following (Simon et al., 2011, Section 2.3). We first find the smallest tuning parameter θ which all coefficient estimates are zero through a rough search, and denote it as θ_{max} because it will be used as the largest value among the possible tuning parameters. After then, the minimum value for the tuning parameter is calculated from θ_{max} by $\theta_{min} = \epsilon \theta_{max}$ where $\epsilon = 0.0001$. In our simulation study and data analysis, we use one hundred grid points for θ between θ_{min} and θ_{max} (including the maximum and the minimum) so as to ignore a value of θ very near 0 and to distribute them reasonable. Thus, as suggested by Simon et al. (2011), the j -th grid point is chosen by $\theta_l = \theta_{max}(\theta_{min}/\theta_{max})^{l/100}$, $l = 1, \dots, 100$.

4.3 Asymptotic properties

We study the asymptotic properties of the adaptive Lasso estimator $\hat{\boldsymbol{\beta}}$ obtained from maximizing the penalized likelihood,

$$W_n(\boldsymbol{\beta}, H) = l_n(\boldsymbol{\beta}, H) - n\theta_n \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j|, \quad (4.10)$$

with respect to $\boldsymbol{\beta}$ and H . Denote the true value of $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}^T, \boldsymbol{\beta}_{20}^T)^T$, where $\boldsymbol{\beta}_{10}$ denotes the vector of all q non-zero components ($1 \leq q \leq d$) and $\boldsymbol{\beta}_{20}$ the vector of zero components. Write $\hat{\boldsymbol{\beta}}$ as $(\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T)^T$ accordingly.

We assume the following regularity conditions:

(C1) The true value $\boldsymbol{\beta}_0$ belongs to the interior of a known compact set \mathcal{B} . The union of the supports of (V_1, \dots, V_K) is a finite interval $[\zeta, \tau]$, where $0 < \zeta < \tau < \infty$. The true value of

$H(\cdot)$, denoted by $H_0(\cdot)$, is strictly increasing and continuously differentiable on $[\zeta, \tau]$, and $0 < H_0(\zeta) < H_0(\tau) < \infty$.

(C2) The covariate vector \mathbf{X} is bounded almost surely.

(C3) The covariance matrix of \mathbf{X} is positive definite.

(C4) The number of inspection times, K , is positive almost surely, and $E(K) < \infty$. Additionally, $\text{pr}(V_{j+1} - V_j \geq \eta | \mathbf{X}, K) = 1$ ($j = 1, \dots, K-1$) for some positive constant η . Furthermore, the conditional densities of (V_j, V_{j+1}) given (\mathbf{X}, K) , denoted by $f_j(s, t | \mathbf{X}, K)$ ($j = 1, \dots, K-1$), have continuous second-order derivatives with respect to s and t when $t - s > \eta$ and are continuously differentiable with respect to \mathbf{X} .

Condition (C1) is the regularity condition 1 assumed in Zeng et al. (2016). (C2) and (C3) are the special cases of their regularity conditions 2 and 3 respectively. (C4) is almost the same as their regularity condition 4 except that we do not assume $\text{pr}(V_K = \tau | \mathbf{X}, K)$ to be greater than a positive constant since this assumption is too restrictive and unnecessary for proving the asymptotic properties in Zeng et al. (2016) as discussed by Zeng et al. (2017). Note that the regularity condition 5 in Zeng et al. (2016) automatically holds for the Cox proportional hazards model. Conditions (C1)–(C4) ensure the root- n consistency of the unpenalized maximum likelihood estimator $\tilde{\boldsymbol{\beta}}$ (Zeng et al., 2016), which is required for the penalty term in (4.10) to be an adaptive Lasso penalty as defined in Zou (2006). These conditions also ensure that the log profile likelihood $l_{pn}(\boldsymbol{\beta})$ has a quadratic expansion around $\boldsymbol{\beta}_0$ (see Remark A1 in Zeng et al., 2016). Our proofs of the asymptotic properties of $\hat{\boldsymbol{\beta}}$ relies on it.

The following asymptotic results are obtained under the above regularity conditions.

Theorem 4.3.1 *If $\sqrt{n}\theta_n = O_p(1)$, then $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$.*

Proof First note that the estimator $\hat{\boldsymbol{\beta}}$ obtained from maximizing (4.10) is equivalent to the solution

of maximizing the following penalized profile likelihood,

$$Q_n(\boldsymbol{\beta}) = l_{pn}(\boldsymbol{\beta}) - n\theta_n \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j|. \quad (4.11)$$

Under the regularity conditions (C1)–(C4), Theorem 1 of Murphy & van der Vaart (2000) is applicable (see Remark A1 in Zeng et al., 2016). Hence, for any random sequence $\boldsymbol{\beta}_n^* \xrightarrow{P} \boldsymbol{\beta}_0$, the log profile likelihood $l_{pn}(\boldsymbol{\beta}_n^*)$ has the following quadratic expansion around $\boldsymbol{\beta}_0$:

$$l_{pn}(\boldsymbol{\beta}_n^*) = l_{pn}(\boldsymbol{\beta}_0) + (\boldsymbol{\beta}_n^* - \boldsymbol{\beta}_0)^T \sum_{i=1}^n \tilde{S}_0(O_i) - \frac{1}{2} n (\boldsymbol{\beta}_n^* - \boldsymbol{\beta}_0)^T \tilde{I}_0 (\boldsymbol{\beta}_n^* - \boldsymbol{\beta}_0) + o_{P_{\boldsymbol{\beta}_0, H_0}}(\sqrt{n} \|\boldsymbol{\beta}_n^* - \boldsymbol{\beta}_0\| + 1)^2, \quad (4.12)$$

where \tilde{S}_0 is the efficient score function for $\boldsymbol{\beta}$ as given implicitly in Zeng et al. (2016) and \tilde{I}_0 is its covariance matrix, the efficient Fisher information matrix.

According to Proposition 3.1 and the discussion in Section 4.4 of Huang & Wellner (1997), $l_n(\boldsymbol{\beta}, \mathbf{h})$ is concave in $(\boldsymbol{\beta}, \mathbf{h})$. Following the proof of Theorem 1 in Zeng et al. (2016), it can be shown that $\hat{H}(\tau)$ is bounded almost surely when n is large. Thus, without loss of generality, we may restrict the parameter \mathbf{h} to a compact space. Then $l_{pn}(\boldsymbol{\beta}) = \sup_{\mathbf{h}} l_n(\boldsymbol{\beta}, \mathbf{h})$ is concave in $\boldsymbol{\beta}$. This together with the concavity of $-n\theta_n \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j|$ leads to that $Q_n(\boldsymbol{\beta})$ is concave. Hence, to show the root- n consistency of $\hat{\boldsymbol{\beta}}$, it is sufficient to show that, for any $\epsilon > 0$, there exists a large constant C such that

$$\liminf_{n \rightarrow \infty} \Pr \left\{ \sup_{\|\mathbf{u}\|=C} Q_n(\boldsymbol{\beta}_0 + n^{-1/2} \mathbf{u}) < Q_n(\boldsymbol{\beta}_0) \right\} \geq 1 - \epsilon. \quad (4.13)$$

This implies that when n is large, there is a local maximizer of $Q_n(\boldsymbol{\beta})$ in the interior of the ball $\{\boldsymbol{\beta}_0 + n^{-1/2} \mathbf{u} : \|\mathbf{u}\| \leq C\}$ with probability at least $1 - \epsilon$. By the concavity of $Q_n(\boldsymbol{\beta})$, the local maximizer must be $\hat{\boldsymbol{\beta}}$ and thus $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$.

By the quadratic expansion (4.12), we have, when n is large,

$$\begin{aligned} & \frac{1}{n} \{l_{pn}(\boldsymbol{\beta}_0 + n^{-1/2} \mathbf{u}) - l_{pn}(\boldsymbol{\beta}_0)\} \\ &= \frac{1}{n} \mathbf{u}^T n^{-1/2} \sum_{i=1}^n \tilde{S}_0(O_i) - \frac{1}{2n} \mathbf{u}^T \tilde{I}_0 \mathbf{u} + o_{P_{\boldsymbol{\beta}_0, H_0}}(n^{-1/2} \|\mathbf{u}\| + n^{-1/2})^2 \\ &= \frac{1}{n} O_p(1) \sum_{j=1}^d |u_j| - \frac{1}{2n} \mathbf{u}^T \tilde{I}_0 \mathbf{u} + o_{P_{\boldsymbol{\beta}_0, H_0}} \left\{ \frac{(\|\mathbf{u}\| + 1)^2}{n} \right\}, \end{aligned}$$

where the second equality holds because $n^{-1/2} \sum_{i=1}^n \tilde{S}_0(O_i) = O_p(1)$. It follows that

$$\begin{aligned}
& \frac{1}{n} \{Q_n(\boldsymbol{\beta}_0 + n^{-1/2} \mathbf{u}) - Q_n(\boldsymbol{\beta}_0)\} \\
&= \frac{1}{n} \{l_{pn}(\boldsymbol{\beta}_0 + n^{-1/2} \mathbf{u}) - l_{pn}(\boldsymbol{\beta}_0)\} - \theta_n \sum_{j=1}^d \frac{|\beta_{j0} + n^{-1/2} u_j| - |\beta_{j0}|}{|\tilde{\beta}_j|} \\
&\leq \frac{1}{n} \{l_{pn}(\boldsymbol{\beta}_0 + n^{-1/2} \mathbf{u}) - l_{pn}(\boldsymbol{\beta}_0)\} - \theta_n \sum_{j=1}^q \frac{|\beta_{j0} + n^{-1/2} u_j| - |\beta_{j0}|}{|\tilde{\beta}_j|} \\
&\leq \frac{1}{n} O_p(1) \sum_{j=1}^d |u_j| - \frac{1}{2n} \mathbf{u}^T \tilde{I}_0 \mathbf{u} + o_{P_{\boldsymbol{\beta}_0, H_0}} \left\{ \frac{(\|\mathbf{u}\| + 1)^2}{n} \right\} + n^{-1/2} \theta_n \sum_{j=1}^q \frac{|u_j|}{|\tilde{\beta}_j|}. \tag{4.14}
\end{aligned}$$

Since the unpenalized maximum likelihood estimator $\tilde{\boldsymbol{\beta}}$ satisfies $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$, we have, for $1 \leq j \leq q$,

$$\frac{1}{|\tilde{\beta}_j|} = \frac{1}{|\beta_{j0}|} - \frac{\text{sign}(\beta_{j0})}{\beta_{j0}^2} (\tilde{\beta}_j - \beta_{j0}) + o_p(|\tilde{\beta}_j - \beta_{j0}|) = \frac{1}{|\beta_{j0}|} + \frac{O_p(1)}{\sqrt{n}}.$$

Then, under the condition $\sqrt{n}\theta_n = O_p(1)$, we have

$$n^{-1/2} \theta_n \sum_{j=1}^q \frac{|u_j|}{|\tilde{\beta}_j|} = n^{-1/2} \theta_n \sum_{j=1}^q \left(\frac{|u_j|}{|\beta_{j0}|} + \frac{|u_j|}{\sqrt{n}} O_p(1) \right) \leq C n^{-1/2} \theta_n O_p(1) = C n^{-1} O_p(1)$$

Therefore, in (4.14), the first and fourth terms are of the order Cn^{-1} , the third term is of a order smaller than $C^2 n^{-1}$, and the second term is of the order $C^2 n^{-1}$ because \tilde{I}_0 is positive definite (see the proof in Zeng et al., 2016). If C is sufficiently large, the second term dominates the others in (4.14). Hence, (4.13) holds, which completes the proof.

Theorem 4.3.2 *If $\sqrt{n}\theta_n \rightarrow 0$ and $n\theta_n \rightarrow \infty$, then $\hat{\boldsymbol{\beta}}$ has the following properties:*

(i) $\lim_{n \rightarrow \infty} \text{pr}(\hat{\boldsymbol{\beta}}_2 = \mathbf{0}) = 1$;

(ii) $\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \rightsquigarrow N(\mathbf{0}, \tilde{I}_{10}^{-1})$, where \tilde{I}_{10} is the upper-left $q \times q$ sub-matrix of the efficient Fisher information matrix for $\boldsymbol{\beta}$, denoted by \tilde{I}_0 , as given implicitly in Zeng et al. (2016).

Proof (i) Since $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$ according to Theorem 4.3.1, to prove $\lim_{n \rightarrow \infty} \text{pr}(\hat{\boldsymbol{\beta}}_2 = \mathbf{0}) = 1$, it suffices to show that for any sequence $\boldsymbol{\beta}_1$ such that $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{10}\| = O_p(n^{-1/2})$ and for any

positive constant C ,

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ Q_n(\boldsymbol{\beta}_1, \mathbf{0}) = \max_{\|\boldsymbol{\beta}_2\| \leq Cn^{-1/2}} Q_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \right\} = 1.$$

We show this by proving that for any such sequence $\boldsymbol{\beta}_1$ and any positive constant C , $\partial Q_n(\boldsymbol{\beta})/\partial \beta_j$ and β_j have opposite signs for any $\beta_j \in (-Cn^{-1/2}, 0) \cup (0, Cn^{-1/2})$ ($j = q+1, \dots, d$) with probability tending to 1.

For any $\boldsymbol{\beta}_1$ satisfying $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{10}\| = O_p(n^{-1/2})$ and any $\boldsymbol{\beta}_2$ such that $\|\boldsymbol{\beta}_2\| \leq Cn^{-1/2}$, $l_{pn}\{\boldsymbol{\beta}\}$ has the quadratic expansion (4.12) when n is large. It is then easy to see that $\partial l_{pn}\{\boldsymbol{\beta}\}/\partial \beta_j = O_p(n^{1/2})$ ($j = q+1, \dots, d$). Hence, for $\beta_j \in (-Cn^{-1/2}, 0) \cup (0, Cn^{-1/2})$ ($j = q+1, \dots, d$),

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial l_{pn}(\boldsymbol{\beta})}{\partial \beta_j} - n\theta_n \frac{\text{sign}(\beta_j)}{|\tilde{\beta}_j|} = O_p(n^{1/2}) - (n\theta_n)n^{1/2} \frac{\text{sign}(\beta_j)}{|n^{1/2}\tilde{\beta}_j|}.$$

Since $n^{1/2}(\tilde{\beta}_j - 0) = O_p(1)$, it follows that

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \beta_j} = n^{1/2} \left\{ O_p(1) - n\theta_n \frac{\text{sign}(\beta_j)}{|O_p(1)|} \right\}. \quad (4.15)$$

Because $n\theta_n \rightarrow \infty$, the sign of β_j determines the sign of $\partial Q_n(\boldsymbol{\beta})/\partial \beta_j$ in (4.15) when n is large, and they have opposite signs.

(ii) According to (i), $\lim_{n \rightarrow \infty} \text{pr}(\hat{\boldsymbol{\beta}}_2 = \mathbf{0}) = 1$. Thus we just need to derive the asymptotic representation of $\hat{\boldsymbol{\beta}}_1$ in the event $\{\hat{\boldsymbol{\beta}}_2 = \mathbf{0}\}$. Set $\hat{h} = \sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10})$ and $\Delta_{1n} = n^{-1/2} \sum_{i=1}^n \tilde{S}_{10}(O_i)$, where \tilde{S}_{10} is the sub-vector of \tilde{S}_0 corresponding to $\boldsymbol{\beta}_1$. For any random sample in the event $\{\hat{\boldsymbol{\beta}}_2 = \mathbf{0}\}$, we have $Q_n(\hat{\boldsymbol{\beta}}_1, \mathbf{0}) \geq Q_n(\boldsymbol{\beta}_0 + n^{-1/2}(\tilde{I}_{10}^{-1}\Delta_{1n}, \mathbf{0}))$. Note that $\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}\| = O_p(n^{-1/2})$ according to Theorem 4.3.1 and $\tilde{I}_{10}^{-1}\Delta_{1n} = O_p(1)$. By (4.12), we have

$$\begin{aligned} Q_n(\hat{\boldsymbol{\beta}}_1, \mathbf{0}) &= l_{pn}(\boldsymbol{\beta}_0) + \hat{h}^T \Delta_{1n} - \frac{1}{2} \hat{h}^T \tilde{I}_{10} \hat{h} + o_p(\|\hat{h}\| + 1)^2 - n\theta_n \sum_{j=1}^q \frac{|\hat{\beta}_j|}{|\tilde{\beta}_j|} \\ &= l_{pn}(\boldsymbol{\beta}_0) + \hat{h}^T \Delta_{1n} - \frac{1}{2} \hat{h}^T \tilde{I}_{10} \hat{h} + o_p(1) - n\theta_n \sum_{j=1}^q \frac{|\hat{\beta}_j|}{|\tilde{\beta}_j|} \end{aligned}$$

and

$$\begin{aligned}
& Q_n(\boldsymbol{\beta}_0 + n^{-1/2}(\tilde{I}_{10}^{-1}\Delta_{1n}, \mathbf{0})) \\
&= l_{pn}(\boldsymbol{\beta}_0) + \Delta_{1n}^T \tilde{I}_{10}^{-1} \Delta_{1n} - \frac{1}{2} \Delta_{1n}^T \tilde{I}_{10}^{-1} \Delta_{1n} + o_p(\|\tilde{I}_{10}^{-1} \Delta_{1n}\| + 1)^2 - n\theta_n \sum_{j=1}^q \frac{|\beta_{j0} + n^{-1/2}(\tilde{I}_{10}^{-1} \Delta_{1n})_j|}{|\tilde{\beta}_j|} \\
&= l_{pn}(\boldsymbol{\beta}_0) + \frac{1}{2} \Delta_{1n}^T \tilde{I}_{10}^{-1} \Delta_{1n} + o_p(1) - n\theta_n \sum_{j=1}^q \frac{|\beta_{j0} + n^{-1/2}(\tilde{I}_{10}^{-1} \Delta_{1n})_j|}{|\tilde{\beta}_j|}.
\end{aligned}$$

Since $Q_n(\hat{\boldsymbol{\beta}}_1, \mathbf{0}) - Q_n(\boldsymbol{\beta}_0 + n^{-1/2}(\tilde{I}_{10}^{-1}\Delta_{1n}, \mathbf{0})) \geq 0$, it follows that

$$\hat{h}^T \Delta_{1n} - \frac{1}{2} \hat{h}^T \tilde{I}_{10} \hat{h} - \frac{1}{2} \Delta_{1n}^T \tilde{I}_{10}^{-1} \Delta_{1n} - n\theta_n \sum_{j=1}^q \frac{|\hat{\beta}_j| - |\beta_{j0} + n^{-1/2}(\tilde{I}_{10}^{-1} \Delta_{1n})_j|}{|\tilde{\beta}_j|} \geq o_p(1) \quad (4.16)$$

The left side of (4.16) is equal to

$$\begin{aligned}
& -\frac{1}{2}(\hat{h} - \tilde{I}_{10}^{-1} \Delta_{1n})^T \tilde{I}_{10}(\hat{h} - \tilde{I}_{10}^{-1} \Delta_{1n}) - n\theta_n \sum_{j=1}^q \frac{|\hat{\beta}_j| - |\beta_{j0} + n^{-1/2}(\tilde{I}_{10}^{-1} \Delta_{1n})_j|}{|\tilde{\beta}_j|} \\
& \leq -c\|\hat{h} - \tilde{I}_{10}^{-1} \Delta_{1n}\|^2 + \sqrt{n}\theta_n \sum_{j=1}^q \frac{|\sqrt{n}(\hat{\beta}_j - \beta_{j0}) - (\tilde{I}_{10}^{-1} \Delta_{1n})_j|}{|\tilde{\beta}_j|}
\end{aligned}$$

for some positive constant c since \tilde{I}_{10} is positive definite. It follows that

$$\|\hat{h} - \tilde{I}_{10}^{-1} \Delta_{1n}\|^2 \leq \frac{1}{c} \sqrt{n}\theta_n \sum_{j=1}^q \frac{|\sqrt{n}(\hat{\beta}_j - \beta_{j0}) - (\tilde{I}_{10}^{-1} \Delta_{1n})_j|}{|\tilde{\beta}_j|} + o_p(1).$$

Note that $|\tilde{\beta}_j| \xrightarrow{p} |\beta_{j0}| > 0$ for $j = 1, \dots, q$. Thus, under the condition $\sqrt{n}\theta_n \rightarrow 0$, $\|\hat{h} - \tilde{I}_{10}^{-1} \Delta_{1n}\| = o_p(1)$, and then $\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) = \tilde{I}_{10}^{-1} \Delta_{1n} + o_p(1) \rightsquigarrow N(\mathbf{0}, \tilde{I}_{10}^{-1})$.

The consistency and sparsity of $\hat{\boldsymbol{\beta}}$ shown in Theorem 4.3.1 and Theorem 4.3.2(i) respectively imply that the adaptive Lasso estimator enjoys the selection consistency property, that is, $\lim_{n \rightarrow \infty} \text{pr}(\mathcal{A}_n = \{1, \dots, q\}) = 1$ with $\mathcal{A}_n = \{j : \hat{\beta}_j \neq 0\}$. Theorem 4.3.2(ii) implies that the adaptive Lasso estimator for the non-zero regression parameters is semiparametrically efficient as if the unimportant covariates were known. It is more efficient than the unpenalized maximum likelihood estimator $\tilde{\boldsymbol{\beta}}_1$, whose covariance matrix is $(\tilde{I}_0^{-1})_{11}$, the leading $q \times q$ sub-matrix of \tilde{I}_0^{-1} .

4.4 Extension to left truncated and interval censored data

In this section, we extend the variable selection method to left truncated and interval censored data. Such data have one additional variable V_{i0} ($i = 1, \dots, n$), the time to entering the study, to the observed data described in Section 1. Left truncation means that the random sample comes from the subpopulation of subjects whose time to event is greater than his/her time to study entry, also called left truncation time. To avoid confusion in understanding the sampling plan, we define T^* , V_0^* and \mathbf{X}^* to be the time to event, left truncation time and covariate vector of a subject from the target population, respectively. Then $(T_i, V_{i0}, \mathbf{X}_i)$ ($i = 1, \dots, n$) are a random sample from the subpopulation of $(T^*, V_0^*, \mathbf{X}^*)$ with $T^* > V_0^*$. To describe the assumption below on the inspection process, we introduce a positive random variable U_k ($k = 1, \dots, K$) to represent the time from study entry to the k -th inspection of a sampled subject, where K denotes the random total number of inspections. Hence $U_{ik} = V_{ik} - V_{i0}$ ($i = 1, \dots, n; k = 1, \dots, K_i$). Define $\vec{U} = (U_1, \dots, U_K)^T$. We assume that T^* is independent of (V_0^*, K, \vec{U}) given \mathbf{X}^* and that the joint distribution of $(V_0^*, K, \vec{U}, \mathbf{X}^*)$ does not involve β and H .

Under the above assumptions, the log likelihood of left truncated and interval censored data is, up to an additive constant free of (β, H) ,

$$l_n^{(T)}(\beta, H) = \sum_{i=1}^n \log \left[\exp\{-(H(L_i) - H(V_{i0})) \exp(\beta^T \mathbf{X}_i)\} - \exp\{-(H(R_i) - H(V_{i0})) \exp(\beta^T \mathbf{X}_i)\} \right], \quad (4.17)$$

where the superscript (T) means truncation. We perform the variable selection through the following adaptive Lasso estimation,

$$\max_{\beta, H} \left\{ l_n^{(T)}(\beta, H) - n\theta \sum_{j=1}^d |\beta_j|/|\tilde{\beta}_j| \right\}, \quad (4.18)$$

where $\tilde{\beta} \equiv (\tilde{\beta}_1, \dots, \tilde{\beta}_d)^T$ is the unpenalized nonparametric maximum likelihood estimator for β , which can be obtained using the EM algorithm below except no penalty being involved, θ is a thresholding parameter whose selection can follow the method in Section 4.2.3, and the maximization with respect to H is over the space of nondecreasing nonnegative functions. We

denote the estimator from (4.18) by $(\hat{\beta}, \hat{H})$.

Similar to the case of interval censored data, \hat{H} has the same characterizations as the unpenalized nonparametric maximum likelihood estimator for H , which were given in Alioum & Commenges (1996). Specifically, the estimator \hat{H} increases only on intervals of the form $(l, u]$ where $l \in \{L_i\}_{i=1}^n$, $u \in \{R_i, V_{i0}\}_{i=1}^n$ and there is no L_i, R_i or V_{i0} in (l, u) . Additionally, \hat{H} is indifferent to how it increases on those intervals, as only the overall jump sizes over $(l, u]$'s with $l > 0$ and $u < \infty$, $\hat{H}(u) - \hat{H}(l)$, affect the penalized likelihood in (4.18). Write the intervals $(l, u]$'s with $l > 0$ and $u < \infty$ as $(l_1, u_1], \dots, (l_m, u_m]$. According to the characterizations of \hat{H} , we can assume, just for the purpose of computing \hat{H} , that H is flat on $\overline{\cup_{j=1}^m (l_j, u_j]}$. Define $h_k = H(u_k) - H(l_k)$ ($k = 1, \dots, m$). Then the log likelihood $l_n^{(T)}(\beta, H)$ can be written as $l_n^{(T)}(\beta, \mathbf{h})$, where $\mathbf{h} = (h_1, \dots, h_m)^T$, and

$$\begin{aligned} l_n^{(T)}(\beta, \mathbf{h}) &= \sum_{i=1}^n \log \left[\exp\left\{-\sum_{V_{i0} < u_k \leq L_i} h_k \exp(\beta^T \mathbf{X}_i)\right\} - I(R_i < \infty) \exp\left\{-\sum_{V_{i0} < u_k \leq R_i} h_k \exp(\beta^T \mathbf{X}_i)\right\} \right] \\ &= \sum_{i=1}^n \log \left(\exp\left\{-\sum_{V_{i0} < u_k \leq L_i} h_k \exp(\beta^T \mathbf{X}_i)\right\} \left[1 - \exp\left\{-\sum_{L_i < u_k \leq R_i} h_k \exp(\beta^T \mathbf{X}_i)\right\} \right]^{I(R_i < \infty)} \right). \end{aligned}$$

In view of the similarity between $l_n^{(T)}(\beta, \mathbf{h})$ and $l_n(\beta, \mathbf{h})$, the adaptive Lasso estimation (4.18) can be performed using the same EM algorithm as for the case of interval censored data except that the indices $u_k \leq L_i$ are replaced by $V_{i0} < u_k \leq L_i$ throughout the algorithm.

4.5 Simulation study

4.5.1 Simulation setting

To check the finite sample performance of our variable selection methods, we conducted numerical experiments in two cases, one without left truncation and one with it. In the former, we also compare the performance of our method with Wu & Cook (2015)'s and Zhang & Lu (2007)'s, for the latter of which the event times were obtained using mid-point imputations. We used the R codes developed by the authors to implement Wu & Cook (2015)'s and Zhang & Lu (2007)'s methods.

The simulation scenario is the following. For every subject of a sample, a vector of ten covariates is generated from a multivariate normal distribution, $\mathbf{X}_i \stackrel{i.i.d.}{\sim} MVN_{10}(0, \Sigma)$ where Σ is

the covariance matrix whose ij -th element is $\Sigma_{ij} = 0.5^{|i-j|}$. We set $\beta_j = 0.5$ for $j = 1, 2, 9$, and 10 , and $\beta_j = 0$ for $j = 3, \dots, 8$ in the Cox model, $h(t|\mathbf{X}_i) = h(t) \exp(\boldsymbol{\beta}^T \mathbf{X}_i)$. We use the Weibull hazard as the baseline hazard function, $h(t) = \kappa \eta (\eta t)^{\kappa-1}$ with $\kappa = 1.5$ and $\eta = 0.02$, which renders $P(T_i < 10 | \mathbf{X}_i = 0) \approx 0.95$. The number of planned inspections is three, but we allow subjects to miss each of the second and third planned inspections with a 5% chance so that the actual number of inspections could vary across subjects. In the untruncated case, the inspection times V_1, V_2 and V_3 are generated from $V_1 \sim U(3.2, 4.8)$, $V_2 = V_1 + U(1.5, 2.5)$, and $V_3 = V_2 + U(1.5, 2.5)$. In the truncated case, we generate the left truncation time V_0 from $V_0 = 2.5 + U(0, 4)$, and the inspection times are obtained from $V_1 = V_0 + U(3.2, 4.8)$, $V_2 = V_1 + U(1.5, 2.5)$ and $V_3 = V_2 + U(1.5, 2.5)$. The proportions of subjects being right-censored are 24.2% in the untruncated case and 29.7% in the truncated case. We considered two sample sizes, 200 and 400. One thousand Monte Carlo samples were generated for each sample size. However, for 247 Monte Carlo samples of size 200 and 210 samples of size 400, Wu & Cook (2015)'s method failed to converge for a wide range of thresholding parameter values. We removed these Monte Carlo samples from computing the simulation performance measurements for Wu & Cook (2015)'s method.

4.5.2 Simulation results

Table 4.1 gives the variable selection percentages of our method, Zhang & Lu (2007)'s method based on mid-point imputation, and Wu & Cook (2015)'s method in the untruncated case. Our method's selection percentages of significant covariates are almost one and higher than the other two methods. Our method's selection percentages of non-significant covariates are less than 8% for $n = 200$ and less than 5% for $n = 400$, slightly worse than Zhang & Lu (2007)'s but better than Wu & Cook (2015)'s. All the three methods' variable selection performances improved as the sample size increased. Table 4.2 shows the average numbers of correct and incorrect zero coefficients as well as the mean squared error of the coefficient estimator, $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \Sigma (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, where Σ is the population covariance matrix of the covariates. Our method outperforms the other two in terms of the average numbers of incorrect zero coefficients and the mean squared error. Zhang &

Table 4.1: The variable selection percentages of our method (Ours), Zhang & Lu (2007)’s method based on mid-point imputation (ZL), and Wu & Cook (2015)’s method (WC) in the untruncated case

n	Method	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
200	Ours	0.997	0.993	0.067	0.067	0.069	0.060	0.072	0.066	0.992	0.993
	ZL	0.974	0.963	0.038	0.019	0.032	0.030	0.035	0.035	0.967	0.971
	WC	0.991	0.992	0.121	0.102	0.112	0.086	0.106	0.130	0.987	0.983
400	Ours	1.000	1.000	0.037	0.026	0.042	0.040	0.047	0.033	1.000	1.000
	ZL	0.999	0.998	0.012	0.009	0.008	0.009	0.012	0.008	1.000	0.999
	WC	0.996	0.995	0.113	0.086	0.085	0.090	0.106	0.109	0.996	0.998

Lu (2007)’s method has a relatively large mean squared error because of the estimation bias for the non-zero coefficients caused by the mid-point imputation, as shown in Table 4.3. Our method’s average numbers of correct zero coefficients are only worse than Zhang & Lu (2007)’s, but are still reasonably good. Table 4.3 gives the mean estimate, the empirical standard error of the estimator, the mean of the standard error estimates, and the coverage of the Normal-based confidence interval for the non-zero coefficients. For the variance estimation of our method and the oracle method, the latter of which applies Zeng et al. (2016)’s approach to the true model, we set $\delta_n = 5n^{-1/2}$, the same as in Zeng et al. (2016). Concurring with the theory, our method performs like the oracle method as the sample size increases. Even in small samples, our variance formula (4.7) is rather accurate, and our estimators for the non-zero coefficients showed normality, as reflected by the coverage of the Normal-based confidence intervals and the Normal Q-Q plots in Figure 4.1. Wu & Cook (2015)’s method performs well in terms of bias and empirical variance, but it has convergence issues as mentioned earlier. Moreover, the standard errors and the coverage of the Normal-based confidence intervals for Wu & Cook (2015)’s method were not computed, because the bootstrap, suggested by Wu & Cook (2015) for computing the standard errors, was very time-consuming in the simulations and unstable in getting moderate standard errors. Figure 4.1, the Normal Q-Q plots of our regularized estimators for the non-zero coefficients, shows that the estimators’ distributions are very close to Normal.

The simulation results for the truncated case are available in Appendix B. In the truncated case,

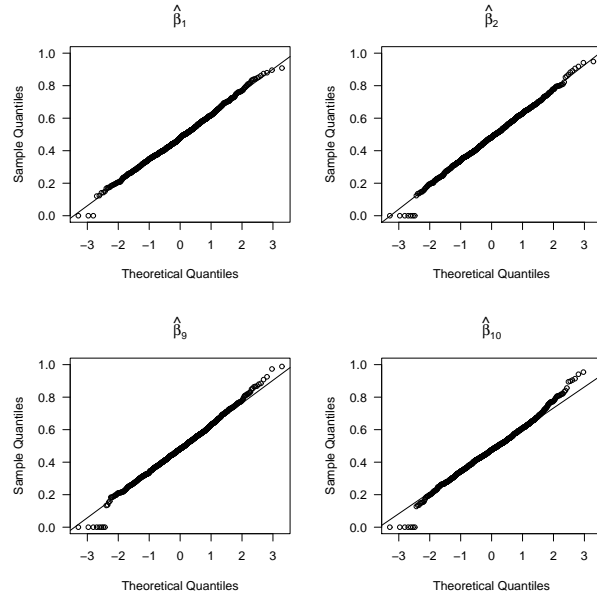
Table 4.2: Average numbers of correct and incorrect zero coefficients and mean squared errors (MSE) of the coefficient estimators for our method (Ours), Zhang & Lu (2007)’s method based on mid-point imputation (ZL), and Wu & Cook (2015)’s method (WC) in the untruncated case. In the parentheses are the ideal numbers of correct and incorrect zero coefficients

n		Correct (6)	Incorrect (0)	MSE
200	Ours	5.599	0.025	0.083
	ZL	5.811	0.125	0.297
	WC	5.343	0.048	0.098
400	Ours	5.775	0	0.034
	ZL	5.942	0.004	0.229
	WC	5.411	0.015	0.043

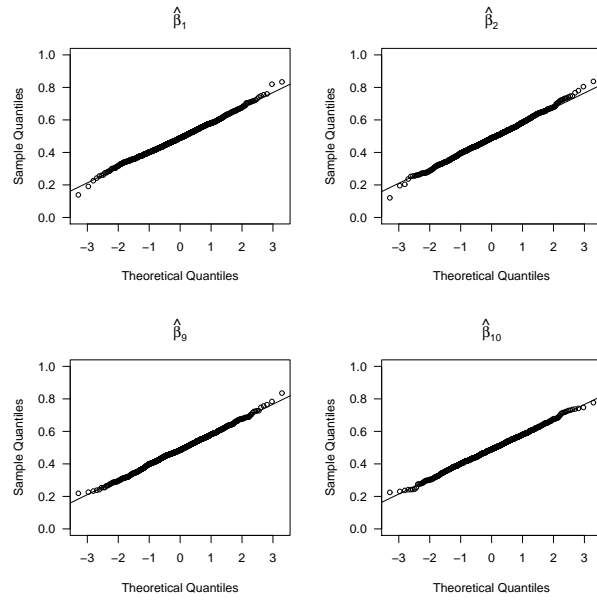
our adaptive Lasso method also performed very well in terms of variable selection percentages (Table B.1), average numbers of correct and incorrect zero coefficients (Table B.2), and estimation accuracy of the non-zero coefficients (Table B.3). For the variance estimation of our method and the oracle method, we again set $\delta_n = 5n^{-1/2}$. Figure B.1 shows that the distributions of our regularized estimators for the non-zero coefficients are very close to Normal.

Table 4.3: Simulation results on the estimation of the non-zero coefficients in the untruncated case. The oracle method is the unpenalized nonparametric maximum likelihood estimation with only the covariates whose coefficients are non-zero, i.e., X_1, X_2, X_9 and X_{10} . The standard errors and the coverage of the Normal-based confidence intervals for Wu & Cook (2015)'s method were not computed due to the bootstrap, suggested by Wu & Cook (2015) for computing the standard errors, being time-consuming

Method	$n = 200$				$n = 400$			
	Estimate	Coverage	Empirical SE	Mean of SEs	Estimate	Coverage	Empirical SE	Mean of SEs
<i>Oracle</i>								
$\beta_1 = 0.5$	0.531	0.974	0.123	0.137	0.518	0.963	0.085	0.090
$\beta_2 = 0.5$	0.535	0.972	0.124	0.136	0.516	0.961	0.090	0.089
$\beta_9 = 0.5$	0.536	0.972	0.127	0.137	0.517	0.953	0.087	0.089
$\beta_{10} = 0.5$	0.525	0.969	0.124	0.136	0.517	0.961	0.084	0.090
<i>Our method</i>								
β_1	0.481	0.939	0.140	0.139	0.491	0.952	0.091	0.091
β_2	0.483	0.948	0.147	0.150	0.488	0.948	0.098	0.098
β_9	0.484	0.955	0.148	0.149	0.488	0.947	0.094	0.098
β_{10}	0.476	0.936	0.142	0.138	0.490	0.937	0.091	0.091
<i>Zhang & Lu (2007)'s</i>								
β_1	0.298	0.470	0.133	0.097	0.318	0.299	0.089	0.068
β_2	0.296	0.493	0.134	0.097	0.314	0.280	0.090	0.068
β_9	0.300	0.485	0.135	0.097	0.314	0.285	0.090	0.068
β_{10}	0.295	0.469	0.132	0.097	0.315	0.285	0.086	0.068
<i>Wu & Cook (2015)'s</i>								
β_1	0.514	-	0.132	-	0.507	-	0.094	-
β_2	0.514	-	0.139	-	0.504	-	0.101	-
β_9	0.515	-	0.150	-	0.504	-	0.102	-
β_{10}	0.512	-	0.146	-	0.505	-	0.091	-



(a) $n = 200$



(b) $n = 400$

Figure 4.1: Normal Q-Q plots of the proposed estimators for the non-zero coefficients in the untruncated case

4.6 Application

We apply the proposed method to the data of Detroit Dental Health Project (DDHP). The study collected a broad array of hypothesized determinants of oral health and made tooth-surface-level caries assessments on the participants over three waves from 2003 to 2007. In the analysis, we consider the event of interest to be that a child has at least one non-cavitated or cavitated lesion in the primary dentition. We use age as the time scale and call the time to event age to caries. According to the study design, each child had 1-3 inspection times depending on whether he or she missed study visits. So the age to caries is either left censored, interval censored or right censored. The number of children in the analyzed data set is 1020, because one child does not have age-at-inspection information. We study the effects of child-, family- and community-level factors on mouth-level primary dental caries development. A list of variables considered in the analysis are in Table 4.4. All of them except WATER were picked out according to Ismail et al. (2009). These variables' values at Wave I were used in the analysis, assuming that the time-dependent variables in the list did not change much during the follow-up, which is reasonable to some extent owing to the dichotomization of many of the time-dependent variables.

The included child-level factors are gender (0 for female, 1 for male), weight-for-age percentiles, frequency of brushing during the preceding week (0 for < 7 , 1 for ≥ 7), and frequencies of wiping teeth, cleaning teeth with water (0 for never or rarely, 1 for sometimes or usually) and whether a child participates in WIC or Head Start (0 for no, 1 for yes). The weight-for-age percentiles are obtained using the 2000 Centers for Disease Control and Prevention growth charts (see Kuczmarski, et al., 2000). Eleven family-level factors are taken into consideration in the model, inclusive of five children's oral health-related measures (OHSE, KBU, KCOH, OHF, KITCHEN) and six caregiver's status-related measures (CESD, PARENTSTRESS, SUPPORT, EDU, EMPLOY, RELIGION). Some family-level measures (OHSE, KBU, KCOH) are constructed by averaging the responses to the questionnaire. SUPPORT was created based on availability of social support in five areas: 1 for support from all five areas and 0 for otherwise (Ismail et al., 2009). The items of the questionnaire for each measure (if applicable) are shown in Appendix C (see Table C.1 and

Table 4.4: The candidate covariates considered in the analysis of DDHP data

Variable Name	Variable Description
<i>Child-level</i>	
GENDER	Gender of child
WEIGHT	Weight-for-age percentile
BRUSHRATE	Brushing frequency during the preceding week (0 for < 7; 1 for ≥ 7)
WIPE	Frequency of wiping teeth of the child (0=never or rarely; 1=sometimes or usually)
WATER	Frequency of cleaning teeth of the child with water (0=never or rarely; 1=sometimes or usually)
WIC	Participating in WIC (0=no; 1=yes)
HEADSTART	Participating in Head Start (0=no; 1=yes)
<i>Family-level</i>	
OHSE	Caregiver's score of perception of self-efficacy related to brushing the child's teeth regularly
KBU	Caregiver's score of knowledge of bottle use
KCOH	Caregiver's score of knowledge of children's oral hygiene
OHF	Caregiver's belief in oral health fatalism (0 = neutral, disagree, or strongly disagree; 1 = agree or strongly agree)
KITCHEN	Water filter/purifier on kitchen tap (0=no; 1=yes)
CESD	Caregiver's depressive symptoms (0=absence; 1=presence)
PARENTSTRESS	Caregiver's parenting stress score
SUPPORT	Social support received by the caregiver (0=low; 1=high)
EDU	Caregiver's education attainment (0 = less than high school, 1 = high school diploma or more)
EMPLOYMENT	Caregiver's full-time employment status (0 = no, 1 = yes)
RELIGIOUSNESS	Caregiver's frequency of attending religious services (1 for less than three times a month; 0 for at least once a week)
<i>Community-level</i>	
DENTIST	Number of dentists in the neighborhood
GROCER	Number of grocery stores in the neighborhood
CHURCH	Number of churches in the neighborhood

Table C.2). The depression scale (CESD) is dichotomized additionally (0 for < 23 , 1 for ≥ 23) to indicate the presence of symptoms by following Siefert et al. (2007). The rest of the factors are also dichotomized: oral health-related fatalism (0 for neutral or disagree, 1 for agree), the level of caregiver's education (0 for less than high school, 1 for high school diploma or more), Having water filter/purifier on kitchen tap (0 for no, 1 for yes), the employment status (0 for no, 1 for yes), and frequency of attending religious services (0 for less than three times a month, 1 for at least once a week). Lastly, the numbers of dentists, grocery stores and churches in the neighborhood are included as community-level factors, based on geocoding (Tellez et al., 2006).

The data analysis results are shown in Table 4.5. Six covariates (BRUSHRATE, WIPE, HEAD-START, KBU, PARENTSTRESS and EDU) were selected in the Cox model by the proposed adaptive Lasso approach. All of their effect directions except WIPE's and PARENTSTRESS's make common sense. A possible explanation about the positive effect of WIPE could be that the wiping cloth was dirty so that wiping teeth accelerated the caries development. A possible explanation about the negative effect of PARENTSTRESS could be that spending more efforts taking care of the child's oral health increased the caregiver's parenting stress. As can be seen from Table 4.5, the 5%-level Wald tests for each covariate based on the unpenalized nonparametric maximum likelihood estimation would pick the same set of significant variables as the adaptive Lasso. But this test-based variable selection approach suffers from the multiple-testing issue, especially in this case of 21 candidate covariates.

4.7 Discussion

In this chapter, we have proposed a new adaptive Lasso approach for left truncated and interval censored data, based on nonparametric maximum likelihood estimation. The simulation study shows that the proposed model led to better and reasonable performance in parameter estimation and variable selection, compared to the other two models. The data application of the proposed model was performed with the mouth-level DDHP data, focusing on the effects of individual-, family- and community-level factors on the development of noncavitated/cavitated lesions in a

Table 4.5: The DDHP data analysis results. NPMLE is the coefficient estimate from the nonparametric maximum likelihood estimation, and Adaptive Lasso is the coefficient estimate from the proposed shrinkage method. Standard errors are given in the parentheses

Variable	NPMLE	Adaptive Lasso
<i>Child-level</i>		
GENDER	-0.039 (0.040)	0 (-)
WEIGHT	-0.038 (0.039)	0 (-)
BRUSHRATE	-0.114 (0.043)	-0.092 (0.087)
WIPE	0.160 (0.041)	0.138 (0.086)
WATER	-0.002 (0.041)	0 (-)
WIC	0.033 (0.04)	0 (-)
HEADSTART	-0.112 (0.042)	-0.087 (0.107)
<i>Family-level</i>		
OHSE	-0.032 (0.042)	0 (-)
KBU	-0.098 (0.044)	-0.060 (0.043)
KCOH	0.058 (0.045)	0 (-)
OHF	0.040 (0.041)	0 (-)
KITCHEN	0.015 (0.040)	0 (-)
CESD	-0.003 (0.041)	0 (-)
PARENTSTRESS	-0.118 (0.044)	-0.094 (0.060)
SUPPORT	-0.023 (0.042)	0 (-)
EDU	-0.094 (0.042)	-0.080 (0.085)
EMPLOYMENT	-0.023 (0.040)	0 (-)
RELIGIOUSNESS	-0.026 (0.045)	0 (-)
<i>Community-level</i>		
DENTIST	0.054 (0.051)	0 (-)
GROCER	0.040 (0.052)	0 (-)
CHURCH	-0.077 (0.052)	0 (-)

child's caries experience trajectory.

We consider the support set of Alioum & Commenges (1996) that has positive mass of NPMLE of H when constructing the log-likelihood. This support set is also applicable for the left truncated and interval censored data, which fact allows us to extend the proposed approach to the truncated case with a simple modification of the estimation procedure. In the truncated case, the corresponding covariance matrix can also be estimated using the profile likelihood method (Murphy & van der Vaart, 2000). These approaches, to the best of our knowledge, are new in the literature. They performed well in finite samples, as seen in Table B.3 of Appendix B.

The proposed variable selection method can be readily extended to interval censored data with time-dependent covariates whose trajectories are fully observed, e.g., marital status and parity. The asymptotic properties for the extension can be also easily derived based on this chapter and Zeng et al. (2016), which considered fully-observed time-dependent covariates. A more interesting and challenging extension is to the high-dimensional setting, i.e., d is comparable to n or even much larger than n . The modification of the EM algorithm by replacing the penalty term with a Lasso penalty will be applicable. However, it will loose the consistency of the estimators, as being generally discussed in much of the Lasso approaches.

APPENDICES

APPENDIX A

We regress the correlation parameter ρ in the distribution of the bivariate frailty on gender after the Fisher z-transformation, namely,

$$\frac{1}{2} \log \frac{1 + \rho}{1 - \rho} = \beta_0 + \beta_1 \text{gender}.$$

Table A.1 shows the model estimation results based on the heterogeneity in ρ . The further results are available in Table A.2 and Table A.3.

Table A.1: The regression parameter estimates for the saturated model based on the DDHP data

Gender	Molar Type	Quadrant	Transition	Parameter	Estimate	SE	p-value
Male	1st molar	Upper right tooth (54)	0 → 1	$\beta_{1,1,1,01}$	The reference category		
			1 → 2	$\beta_{1,1,1,12}$			
		Upper left tooth (64)	0 → 1	$\beta_{1,1,2,01}$	-0.0216	0.1147	0.8506
			1 → 2	$\beta_{1,1,2,12}$	0.1192	0.168	0.4781
		Lower left tooth (74)	0 → 1	$\beta_{1,1,3,01}$	0.3386	0.1118	0.0025
			1 → 2	$\beta_{1,1,3,12}$	0.2177	0.159	0.1709
		Lower right tooth (84)	0 → 1	$\beta_{1,1,4,01}$	0.3556	0.1125	0.0016
			1 → 2	$\beta_{1,1,4,12}$	0.2131	0.158	0.1775
	2nd molar	Upper right tooth (55)	0 → 1	$\beta_{1,2,1,01}$	1.4635	0.1075	< 0.0001
			1 → 2	$\beta_{1,2,1,12}$	-0.3788	0.1504	0.0118
		Upper left tooth (65)	0 → 1	$\beta_{1,2,2,01}$	1.5315	0.1081	< 0.0001
			1 → 2	$\beta_{1,2,2,12}$	-0.3816	0.1497	0.0108
		Lower left tooth (75)	0 → 1	$\beta_{1,2,3,01}$	1.4042	0.1078	< 0.0001
			1 → 2	$\beta_{1,2,3,12}$	-0.3242	0.1516	0.0326
		Lower right tooth (85)	0 → 1	$\beta_{1,2,4,01}$	1.4407	0.1076	< 0.0001
			1 → 2	$\beta_{1,2,4,12}$	-0.3054	0.1495	0.0411
Female	1st molar	Upper right tooth (54)	0 → 1	$\beta_{2,1,1,01}$	0.1258	0.1616	0.4361
			1 → 2	$\beta_{2,1,1,12}$	0.0293	0.1623	0.8567
		Upper left tooth (64)	0 → 1	$\beta_{2,1,2,01}$	0.2306	0.1607	0.1512
			1 → 2	$\beta_{2,1,2,12}$	0.0680	0.1581	0.6672
		Lower left tooth (74)	0 → 1	$\beta_{2,1,3,01}$	0.1788	0.1611	0.2670
			1 → 2	$\beta_{2,1,3,12}$	0.2740	0.1583	0.0835
		Lower right tooth (84)	0 → 1	$\beta_{2,1,4,01}$	0.2308	0.1605	0.1504
			1 → 2	$\beta_{2,1,4,12}$	0.2839	0.158	0.0723
	2nd molar	Upper right tooth (55)	0 → 1	$\beta_{2,2,1,01}$	1.5512	0.1568	< 0.0001
			1 → 2	$\beta_{2,2,1,12}$	-0.2991	0.1466	0.0413
		Upper left tooth (65)	0 → 1	$\beta_{2,2,2,01}$	1.6082	0.1571	< 0.0001
			1 → 2	$\beta_{2,2,2,12}$	-0.1906	0.1456	0.1904
		Lower left tooth (75)	0 → 1	$\beta_{2,2,3,01}$	1.8309	0.1575	< 0.0001
			1 → 2	$\beta_{2,2,3,12}$	-0.2635	0.1453	0.0697
		Lower right tooth (85)	0 → 1	$\beta_{2,2,4,01}$	1.8726	0.1575	< 0.0001
			1 → 2	$\beta_{2,2,4,12}$	-0.3576	0.1457	0.0141

Table A.2: The Wald test statistics (p-value) for the three types of symmetry in caries transition intensity stratified by gender and transition type

Gender	Test for symmetry	Transition	
		0 to 1	1 to 2
Male	Upper vs Lower	21.71(0.0002)	2.74(0.6025)
	Right vs Left	0.72(0.9490)	0.53(0.9709)
	First molar vs Second molar	538.41(< 0.0001)	47.08(< 0.0001)
Female	Upper vs Lower	19.05(0.0008)	5.73(0.2202)
	Right vs Left	1.85(0.7625)	1.66(0.7976)
	First molar vs Second molar	779.47(< 0.0001)	51.20(0.0004)

Table A.3: The detailed comparisons associated with the significant non-symmetries in caries transition intensity found in Table A.2

Gender	Transition	Teeth Compared		Wald Test Statistic (p-value)
Male	0 → 1	Upper right first	Lower right first	-3.160 (0.0008)
		Upper left first	Lower left first	-3.193 (0.0007)
		Upper right second	Lower right second	0.238 (0.4060)
		Upper left second	Lower left second	1.327 (0.0923)
	0 → 1	Upper right first	Upper right second	-13.611 (< 0.0001)
		Upper left first	Upper left second	-14.208 (< 0.0001)
		Lower right first	Lower right second	-10.381 (< 0.0001)
		Lower left first	Lower left second	-10.261 (< 0.0001)
	1 → 2	Upper right first	Upper right second	2.519 (0.0059)
		Upper left first	Upper left second	3.388 (0.0004)
		Lower right first	Lower right second	3.815 (0.0001)
		Lower left first	Lower left second	3.891 (< 0.0001)
Female	0 → 1	Upper right first	Lower right first	-0.986 (0.1597)
		Upper left first	Lower left first	0.494 (0.3016)
		Upper right second	Lower right second	-3.495 (0.0002)
		Upper left second	Lower left second	-2.427 (0.0076)
	0 → 1	Upper right first	Upper right second	-14.066 (< 0.0001)
		Upper left first	Upper left second	-13.693 (< 0.0001)
		Lower right first	Lower right second	-16.235 (< 0.0001)
		Lower left first	Lower left second	-16.201 (< 0.0001)
	1 → 2	Upper right first	Upper right second	2.428 (0.0076)
		Upper left first	Upper left second	2.003 (0.0226)
		Lower right first	Lower right second	4.961 (< 0.0001)
		Lower left first	Lower left second	4.165 (< 0.0001)

APPENDIX B

Table B.1: The variable selection percentages of our method in the truncated case

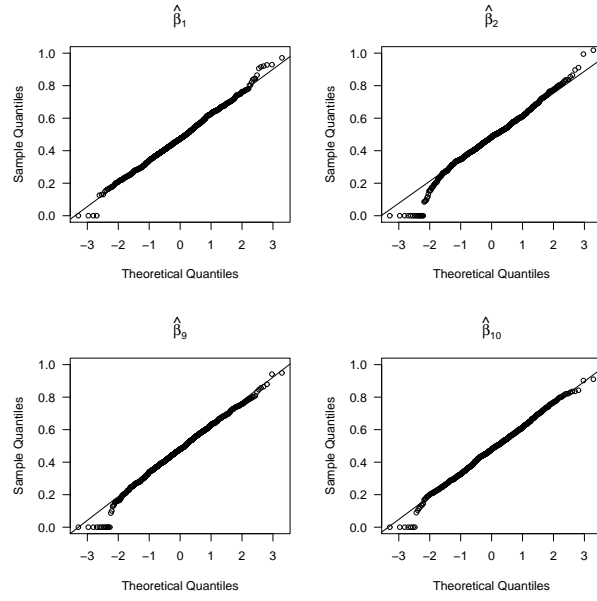
n	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
200	0.996	0.986	0.063	0.062	0.064	0.070	0.075	0.087	0.988	0.993
400	1	1	0.031	0.039	0.035	0.033	0.031	0.042	1	1

Table B.2: Average numbers of correct and incorrect zero coefficients and mean squared errors (MSE) of the coefficient estimators for our method in the truncated case. In the parentheses are the ideal numbers of correct and incorrect zero coefficients

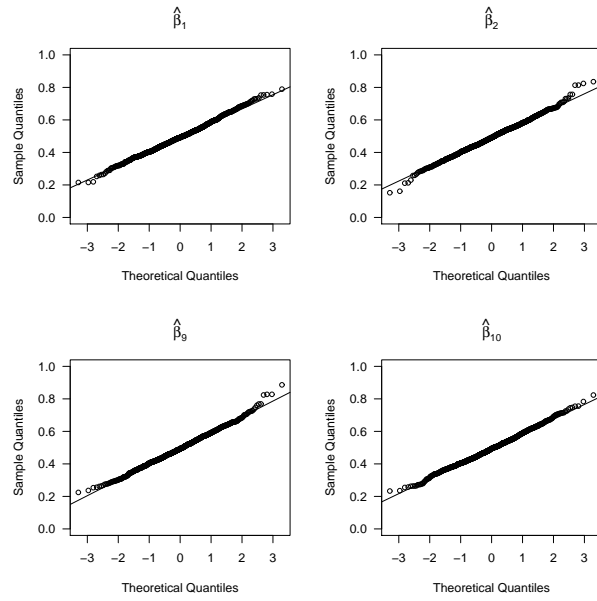
n	Correct (6)	Incorrect (0)	MSE
200	5.579	0.037	0.0877
400	5.789	0	0.0347

Table B.3: Simulation results on the estimation of the non-zero coefficients in the truncated case. The oracle method is the unpenalized nonparametric maximum likelihood estimation with only the covariates whose coefficients are non-zero, i.e., X_1, X_2, X_9 and X_{10}

Method	$n = 200$				$n = 400$			
	Estimate	Coverage	Empirical SE	Mean of SEs	Estimate	Coverage	Empirical SE	Mean of SEs
<i>Oracle</i>								
$\beta_1 = 0.5$	0.542	0.968	0.128	0.134	0.542	0.946	0.086	0.089
$\beta_2 = 0.5$	0.546	0.955	0.129	0.134	0.546	0.948	0.086	0.089
$\beta_9 = 0.5$	0.545	0.958	0.129	0.134	0.545	0.948	0.088	0.089
$\beta_{10} = 0.5$	0.539	0.960	0.129	0.133	0.539	0.937	0.088	0.089
<i>Our method</i>								
$\beta_1 = 0.5$	0.479	0.934	0.143	0.137	0.495	0.944	0.091	0.091
$\beta_2 = 0.5$	0.480	0.942	0.148	0.147	0.492	0.958	0.093	0.098
$\beta_9 = 0.5$	0.478	0.945	0.150	0.147	0.496	0.953	0.096	0.098
$\beta_{10} = 0.5$	0.475	0.931	0.145	0.136	0.495	0.942	0.092	0.091



(a) $n = 200$



(b) $n = 400$

Figure B.1: Normal Q-Q plots of the proposed estimators for the non-zero coefficients in the truncated case

APPENDIX C

Table C.1: Questionnaire items for the family-level measures (OHSE, KBU, KCOH)

Questionnaire items
<p><i>Oral health self-efficacy (OHSE): 1 = not at all confident, 4 = very confident</i></p> <p>(Q1 - Q9) How confident are you that the child's teeth will get brushed when:</p> <ul style="list-style-type: none"> - you are stressed - you are depressed - you are anxious - you are too busy - you are tired - you are worried - child is crying - child does not stay still - child do not feel like brushing right now
<p><i>Knowledge of bottle use (KBU): 1 = strongly agree, 4 = strongly disagree</i></p> <p>(Q1 - Q3) Putting a baby to bed with a bottle helps the child:</p> <ul style="list-style-type: none"> - to be better fed - sleep better - to gain weight and grow <p>Q4. There is nothing wrong with putting a baby to bed with a bottle</p>
<p><i>Knowledge of children's oral hygiene (KCOH): 1 = strongly agree, 5 = strongly disagree</i></p> <p>Q1. Cavities in baby teeth don't matter since they fall out anyway</p> <p>Q2. Keeping baby teeth clean is not very important; after all, they fall out</p> <p>Q3. There is not much I can do to stop the child from developing dental cavities</p> <p>Q4. There is not much I can do to help the child have healthy teeth</p> <p>Q5. Children don't need to brush every day until they get their permanent teeth</p> <p>Q6. Children don't really need their own toothbrush until all their teeth are in</p>

Table C.2: Questionnaire items for the family-level measures (OHF, SUPPORT)

Questionnaire items
<p><i>Oral health-related fatalism (OHF): 1 = strongly disagree, 5 = strongly agree</i></p> <p>Q1. Most children eventually develop dental cavities</p>
<p><i>Social support (SUPPORT): 1 = yes, 0 = no</i></p> <p>(Q1 - Q5). Is there someone you could count on to:</p> <ul style="list-style-type: none"> - run errand for you if you needed them to - lend you some money if you really needed it in a time of financial crisis - give you encouragement and reassurance if you really needed it - watch your (child/children) for you if you needed them to - lend you a car or give you a ride if you needed them

BIBLIOGRAPHY

BIBLIOGRAPHY

- Alioum, Ahmadou & Daniel Commenges. 1996. A proportional hazards model for arbitrarily censored and truncated data. *Biometrics* 52(2). 512–524.
- Batchelor, Paul A & Aubrey Sheiham. 2004. Grouping of tooth surfaces by susceptibility to caries: a study in 5–16 year-old children. *BMC Oral Health* 4(1). 2.
- Benjamin, Regina M. 2010. Oral health: the silent epidemic. *Public health reports* 125(2). 158–159.
- Bogaerts, Kris, Roos Leroy, Emmanuel Lesaffre & Dominique Declerck. 2002. Modelling tooth emergence data based on multivariate interval-censored data. *Statistics in medicine* 21(24). 3775–3787.
- Breiman, Leo. 1996. Heuristics of instability and stabilization in model selection. *Ann. Statist.* 24(6). 2350–2383.
- Brumback, Babette A & John A Rice. 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* 93(443). 961–976.
- Cai, Tianxi & Rebecca A Betensky. 2003. Hazard regression for interval-censored data with penalized spline. *Biometrics* 59(3). 570–579.
- Cook, Richard J & Jerald F Lawless. 2014. Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences* 6(1). 127–161.
- Diabetes Control and Complications Trial Research Group. 1995. Progression of retinopathy with intensive versus conventional treatment in the diabetes control and complications trial. *Ophthalmology* 102. 647–661.
- Edelstein, Burton L. & Courtney H. Chinn. 2009. Update on disparities in oral health and access to dental care for America's children. *Academic Pediatrics* 9(6). 415 – 419.
- Fan, Jianqing & Runze Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96(456). 1348–1360.
- Fan, Jianqing & Runze Li. 2002. Variable selection for cox's proportional hazards model and frailty model. *Ann. Statist.* 30(1). 74–99.
- Fisher-Owens, Susan A, Stuart A Gansky, Larry J Platt, Jane A Weintraub, Mah-J Soobader, Matthew D Bramlett & Paul W Newacheck. 2007. Influences on children's oral health: A conceptual model. *Pediatrics* 120(3). e510–e520.
- Goggins, William B & Dianne M Finkelstein. 2000. A proportional hazards model for multivariate interval-censored failure time data. *Biometrics* 56(3). 940–943.

- Ha, Il Do, Florin Vaida & Youngjo Lee. 2016. Interval estimation of random effects in proportional hazards models with frailties. *Statistical Methods in Medical Research* 25(2). 936–953.
- Harville, David A. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika* 383–385.
- Huang, Jian & Jon A Wellner. 1997. Interval censored survival data: a review of recent progress. In *Proceedings of the first seattle symposium in biostatistics*, 123–169. Springer.
- Hui, Francis K. C., David I. Warton & Scott D. Foster. 2015. Tuning parameter selection for the adaptive lasso using eric. *Journal of the American Statistical Association* 110(509). 262–269.
- Ismail, A.I., W. Sohn, S. Lim & J.M. Willem. 2009. Predictors of dental caries progression in primary teeth. *Journal of Dental Research* 88(3). 270–275.
- Ismail, Amid I., Lim Sungwoo & Sohn Woosung. 2011. A transition scoring system of caries increment with adjustment of reversals in longitudinal study: evaluation using primary tooth surface data. *Community Dentistry and Oral Epidemiology* 39(1). 61–68.
- Joly, Pierre, Thomas A. Gerds, Vibeke Qvist, Daniel Commenges & Niels Keiding. 2012. Estimating survival of dental fillings on the basis of interval-censored data and multi-state models. *Statistics in Medicine* 31. 1139–1149.
- Kalbfleisch, John D & Ross L Prentice. 2002. *The statistical analysis of failure time data*. John Wiley & Sons.
- Kandelman, D. P. & D. W. Lewis. 1988. Pit and fissure sealants. *Preventive Dental Services second edition*. Ottawa: Minister of Supply and Services Canada.[Cat. No. H 39-4/1988E] 13–31.
- Kauermann, Göran, Tatyana Krivobokova & Ludwig Fahrmeir. 2009. Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71(2). 487–503.
- Kim, Mimi Y & Xiaonan Xue. 2002. The analysis of multivariate interval-censored survival data. *Statistics in Medicine* 21(23). 3715–3726.
- Komárek, Arnost & Emmanuel Lesaffre. 2008. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association* 103(482). 523–533.
- Komárek, Arnost, Emmanuel Lesaffre, Tommi T Härkänen, Dominique Declerck & Jorma I. Virtanen. 2005. A bayesian analysis of multivariate doubly-interval-censored dental data. *Biostatistics* 6 1. 145–55.
- Kor, Chew-Teng, Kuang-Fu Cheng & Yi-Hau Chen. 2013. A method for analyzing clustered interval-censored data based on cox’s model. *Statistics in medicine* 32(5). 822–832.
- Kuczmarski, Robert J. 2000. Cdc growth charts; united states .

- Lee, Youngjo & John A Nelder. 1996. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)* 619–678.
- Lewsey, J. D. & W. M. Thomson. 2004. The utility of the zero-inflated poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal dmf data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology* 32(3). 183–189.
- Lu, Wenbin & Hao H. Zhang. 2007. Variable selection for proportional odds model. *Statistics in Medicine* 26(20). 3771–3781.
- Mouradian, WE, E Wehr & JJ Crall. 2000. Disparities in children's oral health and access to dental care. *JAMA* 284(20). 2625–2631.
- Murphy, S. A. & A. W. van der Vaart. 2000. On profile likelihood. *Journal of the American Statistical Association* 95(450). 449–465.
- Murphy, Susan A. & Aad W. Van Der Vaart. 1999. Observed information in semi-parametric models. *Bernoulli* 5(3). 381–412.
- O'Sullivan, F. 1988. Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing* 9(2). 363–379.
- Pinheiro, José C & Douglas M Bates. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics* 4(1). 12–35.
- Rahman, P, D D Gladman, R J Cook, Y Zhou, G Young & D Salonen. 1998. Radiological assessment in psoriatic arthritis. *Rheumatology* 37. 760–765.
- Schick, Anton & Qiqing Yu. 2000. Consistency of the gmle with mixed case interval-censored data. *Scandinavian Journal of Statistics* 27(1). 45–55.
- Selwitz, Robert H, Amid I Ismail & Nigel B Pitts. 2007. Dental caries. *The Lancet* 369(9555). 51–59.
- Siefert, Kristine, David R Williams, Tracy L Finlayson, Jorge Delva & Amid I Ismail. 2007. Modifiable risk and protective factors for depressive symptoms in low-income african american mothers. *American Journal of Orthopsychiatry* 77(1). 113–123.
- Simon, Noah, Jerome Friedman, Trevor Hastie & Rob Tibshirani. 2011. Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of statistical software* 39(5). 1.
- Stone, Charles J. 1982. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10(4). 1040–1053.
- Sutradhar, Rinku & Richard J Cook. 2008. Analysis of interval-censored data from clustered multistate processes: application to joint damage in psoriatic arthritis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57. 553–566.

- Tellez, Marisol, Woosung Sohn, Brian A Burt & Amid I Ismail. 2006. Assessment of the relationship between neighborhood characteristics and dental caries severity among low-income african-americans: a multilevel approach. *Journal of public health dentistry* 66(1). 30–36.
- Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1). 267–288.
- Tibshirani, robert. 1997. The lasso method for variable selection in the cox model. *Statistics in Medicine* 16(4). 385–395.
- Tommi, Harkanen, Virtanen Jorma I. & Arjas Elja. 2000. Caries on permanent teeth: A non-parametric bayesian analysis. *Scandinavian Journal of Statistics* 27(4). 577–588.
- US Department of Health and Human Services. 2000. National Institute of Dental and Craniofacial Research. *Oral Health in America: A Report of the Surgeon General. Washington, DC: May* .
- Vargas, Clemencia M., James J. Crall & Donald A. Schneider. 1998. Sociodemographic distribution of pediatric dental caries: NHANES III, 1988-1994. *The Journal of the American Dental Association* 129. 1229–1238.
- Wu, Ying & Richard J. Cook. 2015. Penalized regression for interval-censored times of disease progression: Selection of hla markers in psoriatic arthritis. *Biometrics* 71(3). 782–791.
- Zeng, Donglin, Fei Gao & D. Y. Lin. 2017. Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. *Biometrika* 104(3). 505–525.
- Zeng, Donglin, Lu Mao & D. Y. Lin. 2016. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* 103(2). 253–271.
- Zhang, Hao Helen & Wenbin Lu. 2007. Adaptive lasso for cox’s proportional hazards model. *Biometrika* 94(3). 691–703.
- Zhang, Yang, D. Todem, K. Kim & Emmanuel Lesaffre. 2011. Bayesian latent variable models for spatially correlated tooth-level binary data in caries research. *Statistical modelling* 11(1). 25–47.
- Zhang, Yiyun, Runze Li & Chih-Ling Tsai. 2010. Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association* 105(489). 312–323.
- Zhang, Zhigang & Jianguo Sun. 2010. Interval censoring. *Statistical Methods in Medical Research* 19(1). 53–70. PMID: 19654168.
- Zou, Hui. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101(476). 1418–1429.