RICE UNIVERSITY

Statistical Issues in Breast Cancer Screening and Clustered Survival Data Analysis

by

Xiuyu Cong

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Dr. Yu Shen, Advisor
Associate Professor of Biotatistics

Dr. Guosheng Yin, Co-advisor
Assistant Professor of Biostatistics

Dr. Javier Rojo, Committee Chair
Professor of Statistics

Dr. Dennis D. Cox
Professor of Statistics

Dr. Enrique V. Barrera
Professor of Mechanical Engineering and
Material Sciences

Houston, Texas

December, 2005

Abstract

Statistical Issues in Breast Cancer Screening and Clustered Survival Data Analysis

by

Xiuyu Cong

This dissertation addresses certain statistical issues in two biomedical fields, namely, modeling breast cancer screening program and correlated survival data analysis.

For the breast cancer screening project, this study investigates statistical approaches to quantitatively describing the age effect on screening sensitivity and sojourn time distribution. Such an investigation is directly motivated by the need to understand the inherent relationships between age and these important quantities. Age effect is incorporated through generalized linear models under a progressive disease modeling framework. Parameter estimates are obtained by maximizing conditional likelihood functions. Among a set of potential models, the Akeike's information criterion and likelihood ratio test are used in model selection and inferences. Extensive simulation studies show that the estimators have reasonable accuracy and the model

selection criterion works well. The proposed methods are illustrated using data from two large breast cancer screening trials.

For correlated survival data analysis, an interesting yet often ignored problem is considered, that is when cluster sizes may be informative to the outcome of interest, based on a within-cluster resampling approach and a weighted marginal model. Large sample properties for the within-cluster resampling estimators are derived under the Cox proportional hazards model, including the consistency and asymptotic normality of the regression coefficient estimators and the weak convergence property of the estimated baseline cumulative hazard function. The weighted marginal model is constructed by incorporating the inverse of cluster sizes as weights in the estimating equations. Simulation studies are conducted to assess and compare the finite-sample behaviors of the estimators and the proposed methods are applied to a dental data example as an illustration.

Acknowledgements

There would be no dissertation for me without all the help from all sources during the past four and a half years. I am obligated to thank them before presenting my work.

I would like to start by thanking my committee members, Drs Yu Shen, Guosheng Yin, Javier Rojo, Dennis Cox, and Enrique Barrera. My special thanks go to my thesis advisor, Dr. Yu Shen, associate professor of biostatistics at the University of Texas MD Anderson Cancer Center. Her instructions and guidance has always been instrumental. Her mothering nature made me feel like a little girl being taken care of. I would also like to thank my co-advisor, Dr. Guosheng Yin, assistant professor of biostatistics at MD Anderson Cancer Center, whose encouragement and criticism has made me more serious about conducting research. I also thank all my committee members for spending their valuable time sitting on the committee and overseeing my progress. Without their help, I would not go this far.

Secondly, I would like to thank the joint biostatistics program offered between Rice University and MD Anderson Cancer Center. I thank the faculties and staffs from both sides for leading me into this exciting world of biostatistics and for assisting me in all aspects. I would also like to thank my fellow graduate students within the department for being a wonderful team. In particular, my officemates and friends, Garrett Fox and Talithia Daniel, Meichun Ding, for all those precious sharings in

personal, spiritual, and academic issues. Because of all of them, graduate school has been one of the best parts of my life.

Thirdly, I would like to acknowledge the generous financial support from Dr. Shen's grant, the National Institute of Health Grants R01 CA79466, and Dr. Yin's grant from MD Anderson Cancer Center. I would also like to thank Drs. AB Miller, CF Spiekerman and DY Lin for the data source, Drs. JP Nolan, M Zelen for their helpful and constructive suggestions, and the editor of Statistics in Medicine and two anonymous referees for their comments on the breast cancer screening paper.

Fourthly, I would like to thank my dearest grandma, parents, brother and husband for their always being there to support and comfort me.

Last but not least, I thank God for His non-stopping love and provisions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis intends to address certain important statistical issues in two biomedical research areas, namely, breast cancer screening (Chapter 2) and multivariate correlated survival data analysis (Chapter 3). Since the two areas and the issues considered are not closely related, I would like to give separate introductions in what follows.

1.1 Breast Cancer Screening

1.1.1 Statement of Problem

In evaluating breast cancer screening programs, it is very important to estimate screening sensitivity and the distribution of sojourn time during the early stage. Better understanding of these two can better help in assessing the effectiveness and quality of existing screening trials and facilitate public health policy makers in rec-

ommending optimal screening strategies. Most existing studies of screening trials consider sensitivity as an unknown constant and sojourn times follow a distribution with constant parameters. Evidence from clinical studies shows that women's age are related to screening sensitivity and sojourn time distribution. However, such evidence is mostly from subset data analysis by stratifying data into age groups. Little has been done to mathematically model such relationships. Due to the sparseness of screening data, loss of efficiency is a big concern in subset analysis, therefore it is interesting and necessary to use structural models to estimate age-specific sensitivity and sojourn time distribution.

1.1.2 Methods

Based on the progressive model framework of Shen and Zelen (1999), I incorporate the age effect on sensitivity and sojourn time distribution through generalized linear models. Three models are proposed.

- *Age-specific sensitivity only (Model 1).* In this model, age effect is only on sensitivity through a generalized linear model, e.g. logistic model. Sojourn time distribution does not depend on age.

- *Age-specific sojourn time distribution (Model 2).* In this model, age effect is on the mean of the sojourn time distribution through a simple linear model. Sensitivity is an unknown constant.

- *Age-specific sensitivity and sojourn time distribution (Model 3).* In this model, both sensitivity and sojourn time distribution depend on age.

Simulation studies are conducted to compare the performance of these models comparing with the model without age effect (Model 0). Finally, the proposed models are applied to two large breast cancer screening trials.

1.2 Clustered Survival Data Analysis

1.2.1 Statement of Problem

There are two interesting and important issues in analyzing clustered data: one is the correlation among individuals within each cluster and the other is the possible informativeness of cluster size. The former is well studied in both survival and non-survival data analysis. The latter has been ignored and only brought to researchers' notice very recently. For analyzing categorical and uncensored continuous data, Hoffman, Sen, and Weinberg (2001) proposed a within cluster resampling model and Williamson, Datta, and Satten (2003) proposed a weighted estimating equation approach to account for informative cluster sizes. Other researchers have elaborated such models, however theory only applies for categorical and uncensored continuous data. For survival data with censoring, the problem is more complicated and challenging due to censoring and the infinite dimensionality of baseline hazard function. In Chapter 3, I propose to extend both methods to survival data analysis based on

the Cox proportional hazards model framework.

1.2.2 Methods

- *Within Cluster Resampling (WCR).* Randomly sample one individual from each cluster, apply the usual Cox model to the resulting dataset to estimate the regression coefficients, estimate the cumulative baseline hazard using the Breslow-Aalen estimator. Repeating a large number of times, the WCR estimators can be constructed as the mean from the resamplings. Large sample properties are derived.

- *Weighted Marginal Model (WMM).* Inverse of cluster sizes are incorporated as weights into the score function based on Cox marginal model under working independence assumptions.

Both models adequately account for the effect of cluster size when it is informative to survival probability. Simulation studies are conducted to assess small sample properties. The proposed methods are applied to a dental study and compared with the analysis result from unweighted marginal model.

Chapter 2

Breast Cancer Screening

This chapter proposes mathematical models for age-specific estimation of screening sensitivity and sojourn time distribution under the progressive disease modeling framework of Shen and Zelen (1999), and applies these models to two breast cancer screening trials, namely, the New York Health Insurance Plan (HIP) and the Canadian National Breast Screening Studies (CNBSS). A paper based on this chapter has been published in Statistics in Medicine (Cong, Shen, and Miller, 2005). Section 2.1 provides background and literature review. Section 2.2 states model assumptions and notations, and describes models for age-specific estimation of sensitivity and mean sojourn time. Section 2.3 presents the simulation studies. In Section 2.4 the results of fitting these models to the two breast cancer screening trials are reported, while some issues related to model fitting and inferences are discussed in section 2.5.

2.1 Background

Breast cancer is the second most frequently diagnosed non-skin cancer and the second leading cause of cancer deaths among women in North America. According to the Cancer Figures and Facts (2003), an estimated 211,300 invasive breast cancer cases and 39,800 deaths were expected to have occurred among women in the United States during 2003. Breast cancer incidence has continued to increase since the 1980s, however, mortality has declined by 1.4% per year during 1989-1995 and by 3.2% since then. The decline of breast cancer mortality over the last twenty years is probably due to improvement in treatment for breast cancer, especially adjuvant therapy for stage II disease, and regular early detection programs (Jatoi and Miller, 2003).

The ultimate goal for an early detection program is to detect cancer in its early stage, which is often defined as the preclinical stage where cancer exists with no clinical symptoms, so that patients may obtain more effective treatments. To assess the efficacy of a screening program, in addition to mortality reduction, two key quantities, screening sensitivity and sojourn time during the preclinical stage, play important roles. Screening sensitivity is the conditional probability of classifying a person as in the preclinical stage given that person actually has preclinical disease. Sensitivity measures the accuracy of a screening program in detecting the preclinical disease. Sojourn time is the length of time the disease remains in the preclinical stage and hence the maximum duration that a screening program can advance the diagnosis. Knowledge of these two quantities is critical in assessing the effectiveness

and quality of screening programs and facilitating public health policy makers in recommending optimal breast cancer screening strategies. Hence, many researches in the area of cancer screening have focused on the estimation of these two quantities with various modeling strategies (Zelen and Feinleib, 1969; Walter and Day, 1983; Day and Walter, 1984; Peer et al., 1996; Etzioni and Shen, 1997; Straatman et al., 1997; Shen and Zelen, 1999).

Most of the previous studies assume that the screening sensitivity is constant and sojourn times follow a parametric distribution with a constant mean. On the other hand, evidence from many breast cancer screening studies has suggested that the age at detection of breast cancer contributes to the variation of screening sensitivity and the mean sojourn time. In particular, recent studies have shown that the sensitivity of mammography is higher in older women (Peer et al., 1996; Miller et al., 1992a,b; Shen and Zelen, 2001). For example, from the Nijmegen study, the sensitivity of mammography was estimated to be 64%, 85%, and 80% within the following age groups: younger than 50, 50-69, and older than 69 years of age, respectively (Peer et al., 1996). The increase in sensitivity of mammography with age may be explained by the fact that breast tissue is primarily fatty and less dense for older women but more dense in younger women (Kerlikowske et al., 1996a,b). Notably, the age at detection has also been found to be positively associated with the duration of the preclinical phase of breast cancer. Several studies have shown that younger women tend to have shorter sojourn time due to faster tumor growth (Shen and Zelen, 1999;

Tabar et al., 2000). For example, using the CNBSS data, the estimated mean sojourn time was found to be longer for the 50-59 age group relative to the 40-49 age group: that is 3.8 years versus 2.1 years based on the results from Shen and Zelen (1999). Similar results are seen in the reports of the Swedish Two-County trial data by Tabar et al. (2000). They reported the mean sojourn times for women by age group as follows: 2.4 years (ages 40 to 49); 3.7 years (50 to 59); 4.2 years (60 to 69) and 4 years (70 to 79).

The observation of differences in the screening sensitivity and preclinical duration among various age cohorts raises a challenge for the design and conduct of screening trials, as too long a screening interval will result in missed diagnosis of cancer in the younger cohorts, and too short intervals will result in an unnecessarily larger number of screening examinations. Joint estimation of the sensitivity of the screening program and the preclinical duration of the disease is preferred for a complete evaluation of screening clinical trials. Therefore, it is of great interest to estimate the sojourn time and screening sensitivity by age and to quantify the association between them. However, little has been done to model such relationships and to quantitatively explore how the screening sensitivity and sojourn time vary across age groups based on all available data besides subset analyses.

2.2 Proposed Models

2.2.1 Assumptions, Notations and Previous Work

Consider a cohort of initially asymptomatic individuals, who are enrolled in a screening trial for early detection of a chronic disease. Assume that the disease progresses through the following three states: a disease-free state or a state where disease cannot be detected by current screening techniques, denoted as S_0; a preclinical state where disease is detectable by the screening examinations yet shows no clinical symptoms, S_p; and a clinical state where symptoms are developed, S_c. Sojourn time in S_p is defined as the duration from the time at onset of S_p to the time at onset of S_c. The three-state stochastic model was originally used to describe the natural history of a chronic disease by Zelen and Feinleib (1969).

Let $q(t)$ be the probability density function of the sojourn time in S_p, and $Q(t) = \int_t^\infty q(x)dx$, the survival function. Let θ be the parameter vector associated with the sojourn time distribution $Q(t)$. Define $w(t)dt$ to be the probability of transition from S_0 to S_p within the time interval $(t, t+dt)$, and $I(t)dt$ the probability of transition from S_p to S_c within $(t, t+dt)$. The following relationship among $w(t)$, $I(t)$ and $q(t)$ holds,

$$I(t) = \int_0^t w(\nu)q(t-\nu)d\nu.$$

Let $t_0 < t_1 < ... < t_{k-1} (< T)$ represent k ordered screening times, which can refer to chronological time or age, where T is the follow-up time past the time of the last

screening examination. Define the ith screening interval as $[t_{i-1}, t_i)$ for $i = 1, 2, ..., k$, where $t_k = T$ and the length of the ith interval is $\delta_i = t_i - t_{i-1}$. Screening examinations occur on the left-hand boundary time of each screening interval. Conditioning on the screening history, we have trinomial samples within each screening interval. The three exclusive possible outcomes are (i) being screening detected at t_{i-1}, (ii) clinically diagnosed to have the disease between the two scheduled screening examinations (t_{i-1}, t_i) (also called interval cases) and (iii) not diagnosed with either mode in the screening interval. The last category consists of individuals without the disease of interest plus diseased women with false negative tests.

Let D_i be the probability of an individual being detected to have preclinical disease at the ith examination given at time t_{i-1}, and I_i be the probability that an interval case occurs in the ith interval. Zelen (1993) derived the functional forms of D_i and I_i as the following, which are functions of the screening sensitivity, β, transition rate $w(.)$ and sojourn time distribution.

$$D_i = \beta \left\{ \sum_{j=0}^{i-1} (1-\beta)^{i-j-1} \int_{t_{j-1}}^{t_j} w(\nu) Q(t_{i-1} - \nu) d\nu \right\} \tag{2.1}$$

and

$$\begin{aligned} I_i &= \int_{t_{i-1}}^{t_i} \left\{ \sum_{j=0}^{i-1} (1-\beta)^{i-j} \int_{t_{j-1}}^{t_j} w(\nu) q(t-\nu) d\nu + \int_{t_{i-1}}^{t} w(\nu) q(t-\nu) d\nu \right\} dt \\ &= \sum_{j=0}^{i-1} (1-\beta)^{i-j} \int_{t_{j-1}}^{t_j} w(\nu) \left[Q(t_{i-1} - \nu) - Q(t_i - \nu) \right] d\nu + \\ &\quad \int_{t_{i-1}}^{t} w(\nu) \left[1 - Q(t_i - \nu) \right] d\nu. \end{aligned} \tag{2.2}$$

Let us first consider an important special case, the stable disease model. Under

the stable disease model assumption, the preclinical and clinical incidence rates in a given population remain as a constant over time, and the corresponding transition probability is also a constant. Thus, we have $w(t) \equiv w$. Under the stable disease model, the above probabilities can be simplified as,

$$D_i = \beta w \mu(\theta) \left\{ 1 - \beta \sum_{j=0}^{i-1} (1-\beta)^{i-j-2} Q_0(t_{i-1} - t_{j-1}) \right\} \tag{2.3}$$

and

$$I_i = w \left\{ \delta_i - \mu(\theta)\beta \sum_{j=0}^{i-1} (1-\beta)^{i-j-1} \left[Q_0(t_{i-1} - t_j) - Q_0(t_i - t_j) \right] \right\}, \tag{2.4}$$

where $Q_0(t) = \mu(\theta)^{-1} \int_t^\infty Q(\nu) d\nu$, and $\mu(\theta) = \int_0^\infty Q(x) dx$ is the mean sojourn time in S_p.

When the age effect is not a concern in the modeling of screening sensitivity and sojourn time, Shen and Zelen (1999) described both full and conditional likelihood methods to estimate the parameters of interest, utilizing periodic screening data. The full likelihood function based on k screening intervals is proportional to

$$\prod_{i=1}^{k} [D_i]^{s_i} [I_i]^{r_i} [1 - D_i - I_i]^{n_i - s_i - r_i},$$

where n_i is the total number of participants at time t_{i-1}; s_i is the number of individuals who were screening detected in S_p at t_{i-1}; and r_i is the number of interval cases in the ith interval. An alternative approach is the conditional likelihood method. Conditioning on the total number of diagnosed cases, $m_i = s_i + r_i$, in the ith interval, the likelihood of an individual being screen-detected in $[t_{i-1}, t_i)$ is proportional to

$$L_i(\beta, \theta | m_i) = \frac{D_i^{s_i} I_i^{m_i - s_i}}{[D_i + I_i]^{m_i}}.$$

The overall conditional likelihood function is obtained by multiplying the likelihood contribution from each of the k intervals. With the observed data for each screening interval, $\{(s_i, r_i), i = 1, 2, ..., k\}$, the conditional log likelihood function follows

$$l(\beta, \theta | m_1, ..., m_k) = \sum_{i=1}^{k} \{s_i \log D_i + r_i \log I_i - (s_i + r_i) \log(D_i + I_i)\}.$$

For simplicity, I refer to this model without the age effect as Model 0.

As shown in previous studies, the full and conditional maximum likelihood methods both lead to reasonable estimates of sensitivity and mean sojourn time. The full likelihood method usually yields estimates with smaller variations compared with the conditional likelihood method (Shen and Zelen, 1999). Nevertheless, the conditional likelihood method will be particularly useful when the age effect is incorporated into the modeling in the following sections. For most screening trials, we have well-documented data regarding the age at detection among the diseased cases, whereas the same information is sometimes missing among the asymptomatic or healthy participants at the end of follow-up. The lack of age information among $(n_i - r_i - s_i)$ subjects in the ith screening interval makes it difficult to calculate the age-specific component, $(1 - D_i - I_i)$, in the full likelihood function. Hence, the conditional likelihood that only focuses on diagnosed cases is preferred for modeling the age-specific screening sensitivity and sojourn time distribution in the following models.

2.2.2 Model 1: Age-Specific Sensitivity Model

Although we can explore the age effect on screening sensitivity and sojourn time distribution by fitting data from different age cohorts with Model 0, a major concern is the loss of efficiency in the estimation procedure with limited sample sizes in such subset analyses. It is hence of great interest to use a structural model to study the associations between age and sensitivity, and between age and sojourn time distribution.

A generalized linear model is used to model the association between the screening sensitivity and the age at screening examinations. Let x denote the age at a screening examination, and $\beta(x)$ define the age-specific screening sensitivity. The model can be formed as

$$g\{\beta(x)\} = a + b(x - x_0),$$

where x_0 is the minimum age of the cohort at study entry, and g is a link function. Since $\beta(x)$ has to be in the range of $[0, 1]$ as a probability measure, we naturally consider a simple logit link function, from which we obtain

$$\beta(x) = \frac{\exp[a + b(x - x_0)]}{1 + \exp[a + b(x - x_0)]}. \tag{2.5}$$

With logit link and $b > 0$, the sensitivity function $\beta(x)$ monotonically increases to 1 as age x increases. As $b \to 0$, the curve reduces to a straight horizontal line, in which case the sensitivity is independent of age. This formula also provides a straightforward interpretation for parameter b in terms of the odds ratio. When

$b > 0$, the odds of being screening detected increase multiplicatively by e^b for every unit increase in age. It may also be plausible to use other link functions, e.g. the identity link, to describe this association. However, the estimate of β can be out of $[0, 1]$ range without imposing complex constraints for parameters a and b.

Let us first derive the age-specific probabilities under the stable disease model, assuming that the sojourn time distribution is independent of age. Let $D_i(\beta(x), \theta)$ be the probability of being detected at the ith screening examination at age x. If a subject is diagnosed to have the disease at her first screening examination at age x, she must have had the preclinical disease before her first examination and remain in S_p at age x. By incorporating age-specific screening sensitivity into (2.1),

$$D_1(\beta(x), \theta) = w\mu(\theta)\beta(x).$$

For $i > 1$, the probability of an individual being detected at the ith screening examination can be expressed as a sum of three components:

$$\begin{aligned} D_i(\beta(x), \theta) &= w\mu(\theta)\beta(x)\Big\{ \prod_{j=0}^{i-2}(1 - \beta(x - t_{i-1} + t_j))Q_0(t_{i-1} - t_0) + \\ &\quad \sum_{j=1}^{i-2} \prod_{l=j}^{i-j-1} (1 - \beta(x - t_{i-1} + t_l)) \left[Q_0(t_{i-1} - t_j) - Q_0(t_{i-1} - t_{j-1})\right] + \\ &\quad 1 - Q_0(t_{i-1} - t_{i-2})\Big\}. \end{aligned} \tag{2.6}$$

The first component describes the probability that an individual is screening detected to have the disease at her ith examination, while she entered S_p before t_0. Thus, she must have failed to be detected at $(i - 1)$ previous examinations given at times $t_0, ..., t_{i-2}$. The second term summarizes the possibilities that the individual has been

missed on $(i-j-1)$ previous examinations if she had her preclinical disease onset in the jth interval, where $1 \leq j < i-1$. The last term in (2.6) corresponds to the possibility that the individual had her preclinical disease onset after her last negative screening examination at t_{i-2}, and was detected at t_{i-1}. Given age x at t_{i-1} and the length of each screening interval, we can obtain the age at each previous screening examination, and the age-specific sensitivity according to (2.5), which is incorporated into (2.6).

Similarly, let $I_i(\beta(x),\theta)$ be the probability of an individual having her disease clinically manifested in the ith screening interval at age x,

$$\begin{aligned} I_i(\beta(x),\theta) &= w\mu(\theta)\left\{\prod_{j=0}^{i-1}(1-\beta(x-t_{i-1}+t_j))\left[Q_0(t_{i-1}-t_0)-Q_0(t_i-t_0)\right]+\right. \\ &\quad \sum_{j=1}^{i-1}\prod_{l=j}^{i-j}(1-\beta(x-t_{i-1}+t_l))[Q_0(t_{i-1}-t_j)-Q_0(t_i-t_j)- \\ &\quad Q_0(t_{i-1}-t_{j-1})+Q_0(t_i-t_{j-1})]+ \\ &\quad \left.\frac{\delta_i}{\mu(\theta)}-1+Q_0(t_i-t_{i-1})\right\}. \end{aligned} \tag{2.7}$$

The above formula also summarizes the possibilities of an individual (i) entering S_p before t_0 but having all her previous i screening examinations negative (false); (ii) entering S_p during $[t_{j-1},t_j)$ and failing to be detected at $(i-j)$ screening examinations, where $1 \leq j \leq i-1$; and (iii) entering S_p and being clinically manifested in the same interval (t_{i-1},t_i). Again, age-specific sensitivity at each missed screening examination is used here.

Within each screening interval $[t_{i-1},t_i)$, $i=1,...,k$, we observe data $(s_i,\mathbf{x_i^s},r_i,\mathbf{x_i^r})$,

where $\mathbf{x_i^s}$ consists of the ages of s_i screen-detected individuals at the ith examination, i.e. $\mathbf{x_i^s} = (x_{ij}^s, j = 1, 2, ..., s_i)$; and $\mathbf{x_i^r}$ contains the ages at clinical diagnosis of r_i interval cases, i.e. $\mathbf{x_i^r} = (x_{ij}^r, j = 1, 2, ..., r_i)$. With the observed data, the conditional log likelihood function is of the form

$$l(a, b, \theta | m_i, \mathbf{x_i^s}, \mathbf{x_i^r}) = \sum_{i=1}^{k} \left\{ \sum_{j=1}^{s_i} \log D_i(\beta(x_{ij}^s), \theta) + \sum_{j=1}^{r_i} \log I_i(\beta(x_{ij}^r), \theta) - \sum_{l=s,r} \sum_{j=1}^{l} \log \left[D_i(\beta(x_{ij}^l), \theta) + I_i(\beta(x_{ij}^l), \theta) \right] \right\}.$$

This conditional log likelihood function can then be maximized to obtain the corresponding estimates within the space for parameters a, b, θ using a nonlinear optimization method proposed by Byrd et al. (1995). This quasi-Newton method allows the parameters to be estimated with constraints so that each of the parameters can be in a meaningful range. For example, the mean sojourn time $\mu(\theta)$ must be positive. This algorithm is readily implemented in statistics softwares such as R.

2.2.3 Model 2: Age-Specific Sojourn Time Distribution

This section studies the age effect on sojourn time distribution. Previous breast cancer screening trials have also indicated that the duration of breast cancer in S_p may be age dependent. However, few studies have explicitly explored the association between sojourn time and the age at diagnosis. Because the mean of sojourn time is often an important summary statistic in understanding the natural history of the disease, we consider a generalized linear model to relate the age at detection with the

mean sojourn time,

$$\mu(x) = \mu_0 + \xi(x - x_0), \tag{2.8}$$

where μ_0 stands for the mean sojourn time at age x_0, x is age at detection. Here, age at detection includes the possibilities of age at screening detection and age at clinical diagnosis.

It is concerned that the sojourn time distribution may not be uniquely determined by its mean. When sojourn time is exponentially distributed, its survival function, $Q(\nu|x)$, is determined by its mean $\mu(x)$ through $Q(\nu|x) = \exp\{-\nu/\mu(x)\}$. However, such a relationship may not hold in general. In this case, we can make certain transformation on the relevant parameters to allow the mean of sojourn times change with age. For instance, if sojourn times follow a Gamma distribution, $\Gamma(\alpha, \lambda)$, the mean and variance can be expressed by α and λ through $\mu = \alpha\lambda$ and $\sigma^2 = \alpha\lambda^2$, respectively. By modeling the age effect on the mean, $\mu(x)$, only, the corresponding parameters α and λ are functions of age x, $\alpha(x) = \mu(x)^2/\sigma^2$, and $\lambda(x) = \sigma^2/\mu(x)$.

In a simple linear model, the coefficient $\xi > 0$ indicates that the mean sojourn time increases as age increases and vice versa. Furthermore, ξ can be explained as the change in the mean sojourn time for a unit change in age. Given the age at detection x, $Q(t|\mu(x)) = \int_t^\infty q(\nu|\mu(x))d\nu$, where $q(t|\mu(x))$ is the age-specific probability density function of the sojourn time in S_p, $\mu(x)$ is used here to indicate the dependence of the sojourn time distribution on the age at detection through the mean. Moreover, $Q_0(t|\mu(x)) = \frac{1}{\mu(x)} \int_t^\infty Q(\nu|\mu(x))d\nu$.

Accordingly, the age-specific probabilities of being screening detected at the ith examination and being an interval case within the ith interval are denoted by $D_i(\beta, \mu(x))$ and $I_i(\beta, \mu(x))$, respectively. Assuming a constant screening sensitivity and letting sojourn time distribution depend on age at detection, we can obtain,

$$D_i(\beta, \mu(x)) = \beta \left\{ \sum_{j=0}^{i-1} (1-\beta)^{i-j-1} \int_{t_{j-1}}^{t_j} w(\nu) Q(t_{i-1} - \nu | \mu(x)) d\nu \right\} \tag{2.9}$$

and

$$\begin{aligned} I_i(\beta, \mu(x)) &= \sum_{j=0}^{i-1} (1-\beta)^{i-j} \int_{t_{j-1}}^{t_j} w(\nu) \left[Q(t_{i-1} - \nu | \mu(x)) - Q(t_i - \nu | \mu(x)) \right] d\nu \\ &\quad + \int_{t_{i-1}}^{t} w(\nu) \left[1 - Q(t_i - \nu | \mu(x)) \right] d\nu. \end{aligned} \tag{2.10}$$

When the sojourn time distribution is age-dependent, the stable disease model generally does not hold. Therefore, we use the nonstable disease model instead. We consider a simple case where the transition rate $w(t)$ is constant. This can be a reasonable assumption when the screening horizon is short compared with a participant's age. The following two probabilities can be readily obtained from (2.9) and (2.10), with a similar decomposition as in section 2.2.2:

$$\begin{aligned} D_i(\beta, \mu(x)) &= w\mu(x)\beta \Big\{ (1-\beta)^{i-1} Q_0(t_{i-1} - t_0 | \mu(x)) + \\ &\quad \sum_{j=1}^{i-1} (1-\beta)^{i-j-1} \left[Q_0(t_{i-1} - t_j | \mu(x)) - Q_0(t_{i-1} - t_{j-1} | \mu(x)) \right] \Big\} \end{aligned} \tag{2.11}$$

and

$$
\begin{aligned}
I_i(\beta,\mu(x)) \;=\; & w\mu(x)\Big\{(1-\beta)^i\left[Q_0(t_{i-1}-t_0|\mu(x))-Q_0(t_i-t_0|\mu(x))\right]+\\
& \sum_{j=1}^{i-1}(1-\beta)^{i-j}\left[Q_0(t_{i-1}-t_j|\mu(x))-Q_0(t_i-t_j|\mu(x))-\right.\\
& Q_0(t_{i-1}-t_{j-1}|\mu(x))+Q_0(t_i-t_{j-1}|\mu(x))]+\\
& \left.\frac{\delta_i}{\mu(x)}-1+Q_0(t_i-t_{i-1}|\mu(x))\right\}. \qquad (2.12)
\end{aligned}
$$

Similarly, the conditional log likelihood function for this age-specific sojourn time model is of the following form,

$$
\begin{aligned}
l(\beta,\mu_0,\xi|m_i,\mathbf{x_i^s},\mathbf{x_i^r}) \;=\; & \sum_{i=1}^{k}\left\{\sum_{j=1}^{s_i}\log D_i(\beta,\mu(x_{ij}^s))+\sum_{j=1}^{r_i}\log I_i(\beta,\mu(x_{ij}^r))-\right.\\
& \left.\sum_{l=s,r}\sum_{j=1}^{l}\log\left[D_i(\beta,\mu(x_{ij}^l))+I_i(\beta,\mu(x_{ij}^l))\right]\right\},
\end{aligned}
$$

which can be maximized to obtain the corresponding estimates for β, μ_0 and ξ. The constant transition rate w can be canceled out in the conditional likelihood.

One reviewer of the paper pointed out that the age at detection (either screening detected or clinically diagnosed) may not be an appropriate anchor to model the sojourn time distribution. In fact, it is more appropriate to use the age of preclinical onset to model the age effect on the mean sojourn time, since the age of preclinical onset is not affected by the mode of detection. By replacing $Q(t_i-\nu|\mu(x))$ with $Q(t_i-\nu|\mu(\nu))$ in (2.9) and (2.10), we can model the association between sojourn time distribution and the age at onset of preclinical disease. However, the computation of the likelihood is more complicated, since some asymptotic numerical integration can

be required. In the following simulation section, simulation studies will be conducted to compare the estimated parameters between using the age at onset of preclinical disease and the age at detection in modeling the mean sojourn time, while generating data assuming the mean sojourn time depends on age at onset.

2.2.4 Model 3: Age-specific Sensitivity and Sojourn Time Distribution

For completeness, let both sensitivity and mean sojourn time dependent on age at detection as specified in (2.5) and (2.8), respectively. Correspondingly, the probability of being screening detected at age x at time t_{i-1} and that of an interval case at age x between $[t_{i-1}, t_i)$ are given as follows:

$$\begin{aligned} D_i(\beta(x), \mu(x)) &= w\mu(x)\beta(x)\Big\{\prod_{j=0}^{i-2}(1-\beta(x-t_{i-1}+t_j))Q_0(t_{i-1}-t_0|\mu(x)) + \\ &\quad \sum_{j=1}^{i-2}\prod_{l=j}^{i-j-1}(1-\beta(x-t_{i-1}+t_l))\left[Q_0(t_{i-1}-t_j|\mu(x))-\right. \\ &\quad \left. Q_0(t_{i-1}-t_{j-1}|\mu(x))\right] + 1 - Q_0(t_{i-1}-t_{i-2}|\mu(x))\Big\}, \end{aligned} \tag{2.13}$$

and

$$
\begin{aligned}
I_i(\beta(x),\mu(x)) \;=\; & w\mu(x)\left\{\prod_{j=0}^{i-1}(1-\beta(x-t_{i-1}+t_j))[Q_0(t_{i-1}-t_0|\mu(x))- \right. \\
& Q_0(t_i-t_0|\mu(x))] + \sum_{j=1}^{i-1}\prod_{l=j}^{i-j}(1-\beta(x-t_{i-1}+t_l))[Q_0(t_{i-1}-t_j|\mu(x))- \\
& Q_0(t_i-t_j|\mu(x)) - Q_0(t_{i-1}-t_{j-1}|\mu(x)) + Q_0(t_i-t_{j-1}|\mu(x))] + \\
& \left.\frac{\delta_i}{\mu(x)} - 1 + Q_0(t_i-t_{i-1}|\mu(x))\right\}. \qquad (2.14)
\end{aligned}
$$

The conditional log likelihood function can be written similarly as in the previous two models substituting these two probabilities, and maximized to obtain estimates for a, b, μ_0 and ξ simultaneously.

2.2.5 Model Inferences

Under usual regularity conditions for the likelihood function, the asymptotic properties of the conditional maximum likelihood estimators can be inferred (Serfling, 1980). When the models are correctly specified, the conditional likelihood equations yield consistent estimates of the true parameters. The analytic formulae for the variance estimators of the parameters are complicated, though they are not impossible to derive. Instead, the bootstrap variance estimates can be used, and the corresponding inference for individual parameters can be made by the Wald test.

To assess whether the proposed models are adequate to describe the relationships under study, certain hypothesis tests can be carried out. Note that Model 0 is nested within Models 1, 2 and 3. Hence the likelihood ratio test based on the conditional

likelihood can be used for model checking. For example, to evaluate the age effect on screening sensitivity in Model 1, we can test the null hypothesis, $H_0 : b = 0$, using the likelihood ratio test. The test statistic will be $-2\log(LR)$, where LR is an abbreviation for likelihood ratio, which is asymptotically χ^2_1 distributed (Mukerjee, 1992). To test the age dependence of the mean sojourn time, we perform a likelihood ratio test between Model 0 and Model 2 with the null hypothesis, $H_0 : \xi = 0$.

When it is of interest to compare the model fittings between Model 1 and Model 2, the likelihood ratio test is not applicable, since Model 1 and Model 2 are not nested. Instead, Akeike's information criterion (AIC) can be used for model selection among a set of nonnested models (Burnham and Anderson, 2002). The AIC criterion is defined as

$$AIC = -2\log L + 2K,$$

where L is the likelihood and K is the number of parameters included in the model. AIC incorporates both the log likelihood and a penalty for the number of parameters. In model selection, a model with the smallest AIC value is often preferred.

2.3 Simulation Studies

A series of simulation studies are conducted to evaluate the proposed models in terms of estimation accuracy and ability to capture the relationship between age at screening examinations and screening sensitivity, or between age at detection and mean sojourn time. Data are generated from four possible scenarios.

- *Simulation I.* Generate data according to Model 0, where neither screening sensitivity nor sojourn time distribution is age-dependent. Let the sensitivity be equal to 0.85, and sojourn times generated from an exponential distribution with mean 2.5 years.

- *Simulation II.* Generate data according to Model 0, where screening sensitivity is age-dependent and mean sojourn time is constant. Let $a = 0.8$, $b = 1.5$, which resulted in an average sensitivity of 0.80, and sojourn times simulated the same way as in Simulation I.

- *Simulation III.* Generate data according to Model 2, where mean sojourn time depends on age and screening sensitivity is constant. Let $\mu_0 = 1.5$, $\xi = 0.6$, so that the average sojourn time was 2.60 years, and sensitivity equal to 0.85.

- *Simulation IV.* Generate data according to Model 3, where both screening sensitivity and mean sojourn time are age-dependent. Let $a = 0.8$, $b = 1.5$, $\mu_0 = 1.5$ and $\xi = 0.6$.

For each scenario, 1000 simulations were repeated, each consisting of 50,000 screening individuals at the start. The individuals' ages at entry were generated uniformly from 40 to 65 years old. For each simulation cohort, four equally spaced screening intervals were used, each of which had a length of one year.

In simulation IV, data were generated according to Model 3, then all four models were fitted to the simulated datasets. For reasons explained in section 2.3.4, in simu-

lations I-III, only Models 0, 1 and 2 were fitted to the simulated data. To determine which model provided the best fit for each simulated data, the one with the smallest AIC value was selected as the best model.

2.3.1 Simulation I

Table 2.1 shows the results for simulation I. When data were generated from Model 0 such that both sensitivity and mean sojourn time were independent of age, the estimates for sensitivity and mean sojourn time using Model 0, the correct model, were closest to the true with smallest bias and variation. Moreover, the correct model had the highest probability of being selected as the best fit, 928 out of 1000 simulations as shown in the last column. Fitting the same data using Model 1 resulted in large standard deviation for parameter estimates, in particular, $\hat{b}$, indicating that sensitivity is unlikely to be age-dependent. Similarly, for the fit with Model 2, the mean sojourn time is not age dependent. Model fitting results are consistent with the data structure.

2.3.2 Simulation II

Table 2.2 shows the results for simulation II. When data were generated from Model 1, the estimates of parameters a and b using Model 1 were close to the true values with small standard deviation. The resulting average sensitivity and mean sojourn time were also close to the true values. Model 0 underestimated the mean sojourn time

Table 2.1: Simulation I: data generated from Model 0, fitted with Models 0, 1, 2

True Model	Model 0 $b = 0$ $\xi = 0$	Model 1 $b \neq 0$ $\xi = 0$	Model 2 $b = 0$ $\xi \neq 0$	Power
Model 0				
$a =$		4.41(5.46)		
$b =$		2.20(5.64)		92.8%
$\mu_0 =$			2.79(.81)	
$\xi =$			.12(.06)	
$\beta = .85$	**.84**(.10)	.94(.11)	.80(.12)	
$\mu = 2.5$	**2.64**(.66)	7.19(6.32)	2.98(.84)	

Each entry is mean (standard deviation) from 1000 simulations.
Power is the proportion of selecting the correct model.

in this case. The worst fit was by Model 3 considering the large standard deviations. Similarly as in simulation I, Model 1, the correct model, was selected as the best fit 883 times out of the 1000 simulations.

It is interesting to know how the age-specific detection probabilities change versus age at detection. As an example, Figure 2.1 shows the relationship between age-specific probabilities of screening and clinical detection (2.6, 2.7) based on Model 1 in simulation II. It is shown that the probability of screening detection increases with age at detection at all four examinations, where the first exam has the highest probability, the last exam the lowest. For interval detection, the first interval has the highest probability of being interval detected, the last interval the lowest. The interval detection probability appears to be decreasing with age at detection because when

Table 2.2: Simulation II: data generated from Model 1, fitted with Models 0, 1, 2

True Model	Model 0 $b = 0$ $\xi = 0$	Model 1 $b \neq 0$ $\xi = 0$	Model 2 $b = 0$ $\xi \neq 0$	Power
Model 1				
$a = -.8$		$-.\mathbf{68}(.29)$		
$b = 1.5$		$\mathbf{1.67}(.43)$		88.3%
$\mu_0 =$			1.97(1.93)	
$\xi =$			1.45(1.21)	
average $\beta = .\mathbf{80}$	.82(.14)	$.\mathbf{82}(.06)$	.74(.33)	
$\mu = \mathbf{2.5}$	2.45(.93)	$\mathbf{2.56}(.46)$	4.70(4.12)	

Each entry is mean (standard deviation) from 1000 simulations.
Power is the proportion of selecting the correct model.

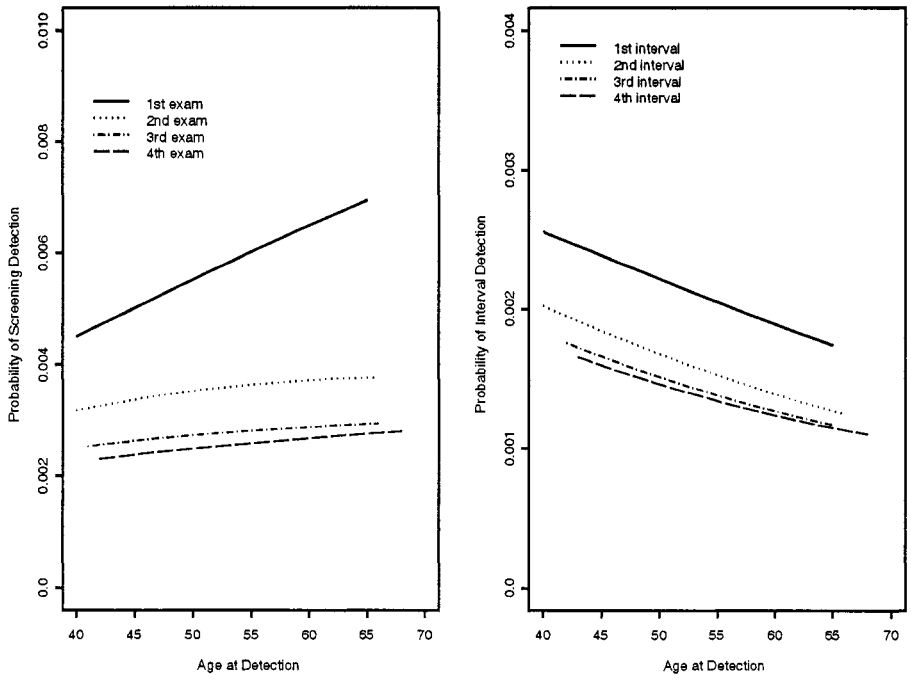
assuming a stable disease and with strong positive age effect on screening sensitivity, more women are detected by mammography when they age.

2.3.3 Simulation III

Table 2.3 shows the results for simulation III. When data were generated from Model 2, in fitting the data, Model 2 provided the best estimates of mean sojourn time and sensitivity. The correct model had the highest empirical probability of being selected as the best model, even though the proportion was rather low, only 664 out of 1000 simulations.

In response to the reviewer's comment, additional simulation studies for Model 2 were conducted to investigate the difference between using the age at onset of S_p and

Figure 2.1: Probability of detection versus age at detection



the age at detection to model the mean sojourn time, while the true mean sojourn time depends on age at onset. Data were generated by letting the mean sojourn time depend on the age at onset of S_p. Then, Model 2 was fitted to the generated datasets by assuming the mean sojourn time depending on the age at onset, and by assuming the mean sojourn time depending on the age at detection, respectively. The results are summarized in the second and third columns of Table 2.4. In fact, with sojourn time distribution similar to the one under investigation as in breast cancer, the estimators do not show substantial difference between these two fittings, though both are somewhat biased. Estimators from fitting the model with the age at onset

Table 2.3: Simulation III: data generated from Model 2, fitted with Models 0, 1, 2

True Model	Model 0 $b = 0$ $\xi = 0$	Model 1 $b \neq 0$ $\xi = 0$	Model 2 $b = 0$ $\xi \neq 0$	Power
Model 2				
$a =$		1.49(1.73)		
$b =$		1.23(1.79)		66.4%
$\mu_0 = 1.5$			2.01(.38)	
$\xi = .6$			.38(.22)	
$\beta = .85$	.92(.07)	.90(.07)	**.87**(.12)	
average $\mu = 2.6$	2.31(.42)	2.48(.44)	**2.68**(.81)	

Each entry is mean (standard deviation) from 1000 simulations.
Power is the proportion of selecting the correct model.

of S_p (correctly) have smaller variations than those using the age at detection.

2.3.4 Simulation IV

Table 2.5 shows the results for simulation IV. When data were simulated using Model 3, the results of fitting Models 1-3 failed to show significant parameter estimates, b or ξ. All three models resulted in similar estimated average sensitivity and mean sojourn times. This indicates that the model including age effect on both sensitivity and mean sojourn time does not gain much improvement over Model 1 or 2 when only sensitivity or sojourn time depends on age. For this reason and to save computation time, Model 3 is omitted in simulations I-III and in the subsequent data application.

Table 2.4: Simulation Results: data generated with mean sojourn time depending on age at onset.

True values	Model 2 Age at onset*	Model 2 Age at detection†
$\mu_0 = 1.5$	1.94(.38)	2.00(.58)
$\xi = .6$	.35(.19)	.37(.21)
$\beta = .85$	**.88**(.10)	**.87**(.12)
average $\mu = 2.6$	**2.58**(.54)	**2.68**(.79)

Each entry is mean(standard deviation) from simulations
* Fitting Model 2 assuming dependence on age at onset
† Fitting Model 2 assuming dependence on age at detection

Table 2.5: Simulation IV: data generated from Model 3, fitted with all four models

True Model	Model 0 $b = 0$ $\xi = 0$	Model 1 $b \neq 0$ $\xi = 0$	Model 2 $b = 0$ $\xi \neq 0$	Model 3 $b \neq 0$ $\xi \neq 0$
Model 3				
$a = -.8$		$-1.13(.23)$		$-.\mathbf{32}(.42)$
$b = 1.5$		1.80(.31)		**1.34**(.54)
$\mu_0 = 1.5$			.59(.15)	**1.61**(.47)
$\xi = .6$			.94(.09)	**.57**(.29)
average $\beta = .80$	.96(.05)	.80(.03)	.88(.07)	**.82**(.07)
average $\mu = 2.62$	1.75(.23)	2.75(.24)	2.35(.26)	**2.68**(.46)

Each entry is mean(standard deviation) from 1000 simulations

2.4 Application to Two Breast Cancer Screening Trials

As motivated by the breast cancer screening trials, we illustrate the proposed models and estimation methods using data from two large screening trials: HIP study and CNBSS. The three proposed models were fitted to both the HIP trial and the CNBSS. Based on previous findings, the exponential distribution is the best-fitting distribution for the sojourn times among several parametric families that have been explored for early detection of breast cancer (Zelen and Feinleib, 1969; Walter and Day, 1983). Thus, an exponential model for sojourn time distribution is assumed. Model 2 was fitted to both datasets with age at detection, assuming a constant transition rate, for computational simplicity.

2.4.1 Data Description

The HIP study was carried out in the 1960s and was the first randomized breast cancer screening trial to evaluate screening mammography. The study enrolled approximately 62,000 asymptomatic women, 40 to 64 years old at entry. The participants were evenly randomized to a screening or control group. Women in the screening group were offered an initial and three additional annual screening examinations. No screening examination was offered to the control group. Women in this group followed their usual practice in obtaining medical care. In this analysis, only data from

the screening group for the first five years of follow-up were used. The numbers of participants, screening detected at each exam and interval cases among each interval are summarized in Table 2.6.

Table 2.6: HIP: total participants, screening detected and interval cases.

	Screening Exam			
	First	Second	Third	Fourth
No. of total participants	20,166	15,936	13,679	11,971
No. of screening detected	55	32	18	27
No. of interval cases	13	8	11	13

The CNBSS investigators conducted two randomized controlled screening trials by the age of entry. In the first trial about 50,340 women aged 40 to 49 years at entry were randomized to either a study or control group. Women in the study group had an initial and then four annual screening examinations. The second trial enrolled women aged 50-59 years, with 19,711 in the study group and 19,694 in the control group. The same screening schedule and screening modality were given to women in the study group as in the first trial. Since we are primarily interested in age-specific sensitivity and sojourn time, to obtain a wider age range, we combined the data from the two screening groups from the CNBSS with the first five years of follow-up in this analysis. The numbers of participants, screening detected at each exam and interval cases among each interval are summarized in Table 2.7.

Table 2.7: CNBSS: total participants, screening detected and interval cases.

	Screening Exam			
	First	Second	Third	Fourth
No. of total participants	44,925	40,093	39,413	39,032
No. of screening detected	238	100	85	105
No. of interval cases	36	30	17	26

2.4.2 Analysis Results

For the purpose of model selection, the AIC value for each model is calculated and shown in Table 2.8 for HIP and 2.9 for CNBSS. In addition, the likelihood ratio test statistic, $-2\log(LR)$, with its p-value in the parentheses, is estimated between the nested models. For the HIP study, the AIC value for Model 1 is the smallest, though the differences among the three models are not substantial. It seems to indicate an indifference between the models under comparison, especially between Model 1 and Model 2 (Burnham and Anderson, 2002). For the nested models (Model 0 and Model 1), the likelihood ratio test indicates a marginal statistical significance of Model 1 over Model 0 (p=0.07), while the effect of age on the sojourn time distribution is even weaker. For the CNBSS, the difference in the AIC values between Model 1 and Model 0 is 5.72, which shows significant evidence of Model 1 being favored over Model 0. On the other hand, the age effect on the sojourn time distribution is relatively weak, similar to the results from the HIP study. In addition, the likelihood ratio

Table 2.8: Parameter estimates from the proposed models for HIP

	Model 0 $b = 0$ $\xi = 0$	Model 1 $b \neq 0$ $\xi = 0$	Model 2 $b = 0$ $\xi \neq 0$
$-\log L$	98.86	97.23	97.36
estimates	$\hat{\beta} = .75(.15)*$ $\hat{\mu} = 2.13(.59)$	$\hat{a} = -.13(1.44)$ $\hat{b} = .47(.29)$ $\hat{\mu} = 2.57(.46)$	$\hat{\beta} = .59(.15)$ $\hat{\mu}_0 = 2.01(.65)$ $\hat{\xi} = .52(.32)$
$-2\log(LR)$		3.26(.07) **	3.0(.08)
AIC	201.72	**200.46**	200.72

* parameter estimates from data (bootstrap standard error).
** $-2\log(LR)$ value (p-value).

statistic for testing Model 1 versus Model 0 also leads to a statistically significant age effect on screening sensitivity; whereas the age effect is marginally significant for the mean sojourn time. The Wald test results from the conditional likelihood are less stable than those from the likelihood ratio tests, as found previously (Shen et al., 2001). Thus, the inference here does not rely on the Wald test.

As shown in Tables 2.8 and 2.9, the coefficient estimates $\hat{b}$ are positive in both trials, indicating that screening sensitivity increases with age at screening examinations. Figure 2.2 shows the relationship between screening sensitivity and age in both studies based on the parameter estimates. The overall sensitivity of the CNBSS in the observed range of age is generally higher than that of the HIP study, and the sensitivity increases more rapidly with age in the CNBSS.

Table 2.9: Parameter estimates from the proposed models for CNBSS

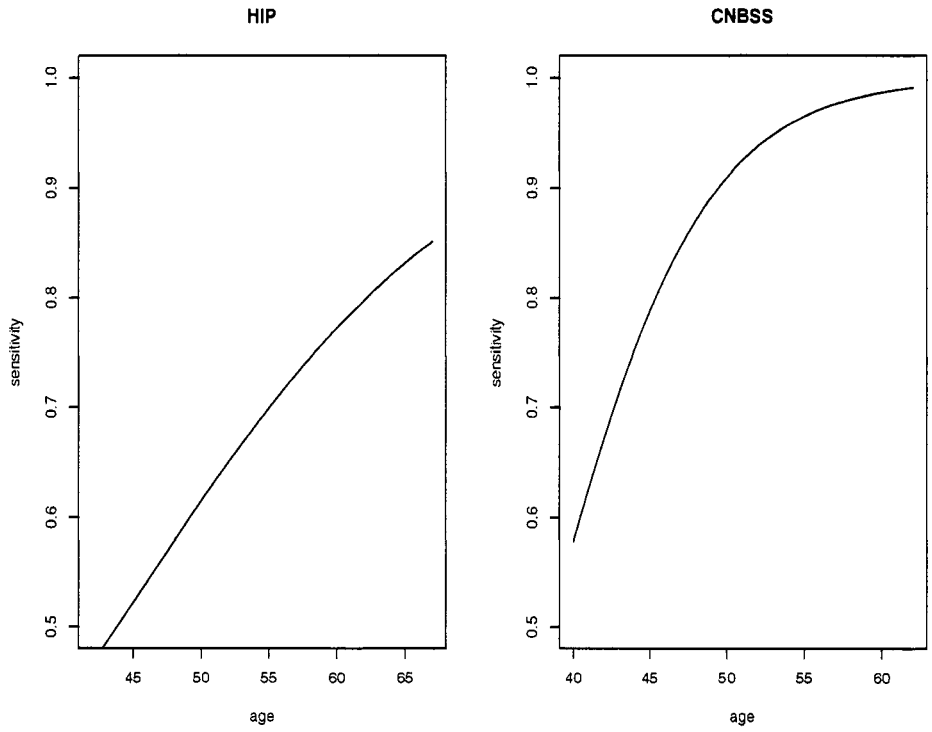
	Model 0 $b = 0$ $\xi = 0$	Model 1 $b \neq 0$ $\xi = 0$	Model 2 $b = 0$ $\xi \neq 0$
$-\log L$	289.53	285.67	287.77
estimates	$\hat{\beta} = .82(.10)*$ $\hat{\mu} = 2.68(.48)$	$\hat{a} = .31(.2.78)$ $\hat{b} = 1.07(13.26)$ $\hat{\mu} = 2.42(1.22)$	$\hat{\beta} = .80(.20)$ $\hat{\mu}_0 = 1.70(.85)$ $\hat{\xi} = .11(.07)$
$-2\log(LR)$		7.72(.005) **	3.52(.06)
AIC	583.06	**577.34**	581.54

* parameter estimates from data (bootstrap standard error).
** $-2\log(LR)$ value (p-value).

Based on the estimated $\hat{a}$ and $\hat{b}$ in Model 1, the screening sensitivity for the HIP study varies from .47 to .85 for the observed age range, with an average of .69. The mean sojourn time is estimated to be 2.57 years with a bootstrap SE of .46. The average sensitivity and mean sojourn time are comparable with those estimated by Shen and Zelen (2001), which are the estimates under Model 0, and those by Day and Walter (1984) of .82 for sensitivity and 1.7 years for the mean sojourn time. For the CNBSS, sensitivity changes from .58 to .99 with an average being .90 and the mean sojourn time is 2.42 years with a bootstrap SE of 1.22 under Model 1. These are also comparable with the parameters estimated in Shen and Zelen (2001).

When fitting Model 2 with the mean sojourn time specified as in (2.8), we obtained positive estimates for $\hat{\xi}$ in both studies, which indicates an increasing trend for the mean sojourn time with age at detection, though such a relationship is not sta-

Figure 2.2: Sensitivity versus age at screening examinations



tistically significant. The estimates lead to an average mean sojourn time of 3.04 and 2.9 years, respectively, for the HIP study and the CNBSS. The corresponding screening sensitivity is .59 (SE, .15) and .80 (SE, .20) for the HIP trial and the CNBSS, respectively.

2.5 Discussion

Models are proposed to evaluate the age effect on screening sensitivity and sojourn time distribution in breast cancer screening trials. The results suggested, to a rea-

sonable degree, that the screening sensitivity increases with the age at screening examinations. Previous studies noted this tendency as described in the subset analyses by stratifying women into different age groups (Peer et al., 1996). Parmigiani, on the other hand, proposed some parametric models to relate age to screening sensitivity, as well as to sojourn times (Parmigiani, 2002). However, the estimates of parameters based on his models were loosely determined using the subset analysis results of Peer et al. (1996).

Naturally, we can jointly model the age dependencies on both the sojourn time distribution and screening sensitivity (Model 3). Simulation study IV showed that the model including age effect on both screening sensitivity and mean sojourn time does not gain much improvement over Model 1 or Model 2. Neither the AIC selection procedure nor the likelihood ratio test led to significant differences. In other words, when either the screening sensitivity or mean sojourn time is age dependent, introducing age dependency on the second quantity does not significantly improve the model fitting. A plausible explanation is that screening sensitivity and mean sojourn time are strongly correlated. When the screening sensitivity is higher, smaller tumors become detectable so that the sojourn time in S_p becomes longer. This correlation is most likely to weaken the identifiability of the age effect on both screening sensitivity and mean sojourn time in the same model.

Even though it is more appropriate to use the age at onset of preclinical disease to model the sojourn time distribution, the difference in model fitting would not be

substantial for breast cancer than using age at detection according to our simulation studies. However, the difference could be significant for cancers with much longer sojourn times, such as prostate cancer.

A simple logit model for screening sensitivity and a linear model for mean sojourn time are suggested in analyzing the two screening trials, the general methodology to estimate the unknown parameters remains to be the same with any alternative parametric models. As suggested in some studies, as women age, the increase in the mean sojourn time may slow down (Tabar et al., 2000). In this case, a three-parameter logistic curve for the effect of age may be more appropriate to capture this trend, $\mu(x) = \frac{a}{1+be^{c(x-x_0)}}$. However, it is worth noting that, compared with the simple linear model, more parameters need to be estimated for this flexible model, which may yield less efficient estimators.

Chapter 3

Informative Cluster Size in Clustered Survival Data Analysis

In this chapter, I generalize two methods for analyzing clustered survival data with informative cluster sizes. One is the within cluster resampling method of Hoffman et al. (2001) based on the Cox proportional hazards model framework. Asymptotic theories are developed for the regression coefficient estimators, so are the weak convergence property of the estimated cumulative baseline hazard function, and the corresponding consistent variance-covariance estimators. An alternative is to generalize the standard marginal model (MM) of Lee, Wei, and Amato (1992) by incorporating the inverse of cluster sizes as weights into the score function to account for informative cluster sizes.

The rest of this chapter is organized as follows. Section 3.1 provides a motivating

example and literature review. Section 3.2 introduces the within-cluster resampling method and weighted marginal models under the Cox proportional hazard model. The large sample properties for the estimators of the regression coefficients and the cumulative baseline hazard function are derived for the within-cluster resampling method. Section 3.3 conducts simulation studies to assess the finite sample properties of the proposed methods. Section 3.5 applies the two proposed methods to the dental study and compare the results with those from the standard marginal models.

3.1 Motivating Example and Literature Review

3.1.1 Motivating Example

Correlated survival data often arise in biomedical research settings. For example, in randomized multi-center clinical trials, patients are recruited and grouped by study centers. In dental or family disease studies, all teeth of the same person or all members of the same family are naturally clustered together. Observations within the same cluster are likely to be correlated, where the correlation needs to be accounted for in statistical estimation and inference.

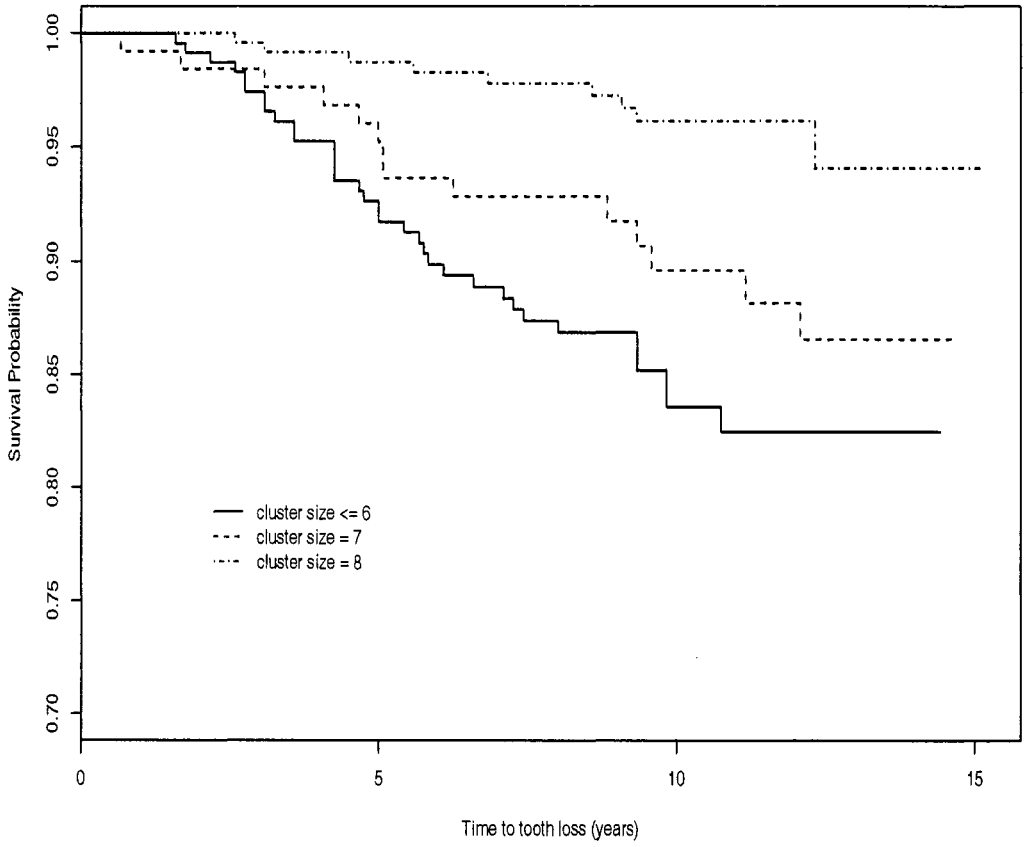
Another interesting problem in clustered/correlated survival data, which is often ignored, is the possible informativeness of cluster sizes. Cluster size is informative when the outcome of interest among individuals in a cluster is associated with the size of that cluster. For example, in a toxicology study assessing the effect of mother

mice being exposed to certain toxicant, mothers that are particularly susceptible to the toxicant may produce more offspring with birth defects that have lower survival probabilities and meanwhile may experience more fetal resorptions, hence reducing the litter size. In this case, pups of a smaller litter tend to have shorter survival, thus the cluster size is informative to the effect of toxicant exposure on offspring survival. Another example is found in a dental study, where the effects of behavioral factors such as cigarette smoking and hygiene status, may predict tooth survival for patients with chronic periodontitis (McGuire and Nunn, 1996). The outcome of interest was the time to tooth loss from initiation of therapy. Shown in Figure 3.1 are the Kaplan-Meier survival curves for molar teeth stratified by the cluster sizes (the number of molars of a patient ≤ 6, =7 or =8). We can see that patients with more teeth have higher molar survival probability. As a result, cluster size is informative to tooth survival.

3.1.2 Literature Review

Marginal models have been proposed and widely used for analyzing clustered survival data. Under the proportional hazards model framework, Lee et al. (1992) proposed a multiplicative intensity model where they estimated the regression coefficients assuming independence among observations and provided a "sandwich" form covariance matrix estimator that takes account of the intra-cluster correlation. Similarly, in analyzing multivariate failure times where individuals experience multiple failure

Figure 3.1: Kaplan-Meier curves for the survival of molars in the dental study



types, for example, heart disease and cancer, or repeated failures of the same type such as multiple tumor recurrence, Wei, Lin, and Weissfeld (1989) proposed using the marginal Cox proportional hazards model (Cox, 1972, 1975) for each failure type without specifying any dependence structure among the distinct failure types while the covariance matrix was estimated jointly across all failure types to adjust for the correlation. Spiekerman and Lin (1998) and Clegg, Cai, and Sen (1999) presented a more general marginal regression model for multivariate failure time data which included Wei et al. (1989) and Lee et al. (1992) as special cases. Alternative marginal models include the accelerated failure time model, additive hazard model and linear transformation models. Yin and Cai (2004) considered marginal additive hazards models for multivariate survival data. Lin and Wei (1992) proposed a semiparametric accelerated failure time model where the covariates are linearly related to the logarithm of the survival times for the multivariate case. Lee, Wei, and Ying (1993) applied a log-linear regression model based on a population-averaged approach to highly stratified failure time data. The class of semiparametric linear transformation models includes the proportional hazards and the proportional odds models as special cases. In these models, the covariates are linearly related to an unknown transformation of the failure time with a fully specified error distribution. Chen and Wei (1997) proposed an approach similar to that of Wei et al. (1989) for multivariate failure time data by modeling each marginal failure time with a linear transformation model. Cai, Wei, and Wilcox (2000) combined the idea of the generalized estimating

equation method (Liang and Zeger, 1986) and the linear transformation models for correlated survival data.

However, all the aforementioned methods do not explicitly consider the possible informativeness of cluster sizes. When cluster size is informative, the standard marginal models tend to overweight the large clusters since each individual observation contributes equally in the likelihood function and therefore produce biased estimates. Hoffman et al. (2001) proposed a within-cluster resampling (WCR) procedure, where one observation is randomly sampled from each cluster. The observations in the resampled dataset are thus independent and standard methods can be readily applied. By resampling the observed data with replacement many times, the final estimator can be obtained through averaging over the estimates from the resampled data. Williamson et al. (2003) proposed a modified generalized estimating equation method, where the estimating equation is inversely weighted by cluster sizes. Both methods can adequately account for informative cluster sizes by weighting clusters equally and produce valid inferences for the parameters of interest. Follmann, Proschan, and Leifer (2003) discussed the asymptotic theories and broad applications of the WCR approach under a different name, multiple outputation. More recently, Benhin, Rao, and Scott (2005) gave a thorough and in-depth discussion on a mean estimating equation approach that is in essence analogous to the work of Williamson et al. (2003).

The theoretical development of the within cluster resampling and weighted esti-

mating equation is well established for uncensored categorical and continuous data. For correlated survival data, the problem is much more complicated and challenging due to the existence of censoring and the infinite dimensionality of the unknown hazard function.

3.2 Proposed Methods

3.2.1 Notations

Let $i = 1, \ldots, m$ index the clusters which are assumed to be independent of each other, and $j = 1, \ldots, n_i$ denote the individuals within the ith cluster. Let T_{ij} and C_{ij} be the failure and censoring times for the jth individual in the ith cluster, respectively. Let $\mathbf{Z}_{ij}(t)$ denote a p-vector of possibly time-dependent covariates, where $t \in [0, \tau]$ for some finite constant $\tau > 0$. Assume that T_{ij} is conditionally independent of C_{ij} given $\mathbf{Z}_{ij}(t)$. The observed times are $X_{ij} = \min(T_{ij}, C_{ij})$, with the censoring indicator $\Delta_{ij} = I(X_{ij} = T_{ij})$, where $I(\cdot)$ is the indicator function.

If cluster sizes are ignorable (non-informative to survival), the usual marginal proportional hazards model (Lee et al., 1992) is defined as

$$\lambda(t|\mathbf{Z}_{ij}(t)) = \lambda_0(t)\exp(\boldsymbol{\beta}_0'\mathbf{Z}_{ij}(t)), \tag{3.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and $\boldsymbol{\beta}_0$ is the regression coefficient vector. Define the counting process $N_{ij}(t) = I(X_{ij} \leq t, \Delta_{ij} = 1)$, the at-risk

process $Y_{ij}(t) = I(X_{ij} \geq t)$, and

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_{ij}(u)) \lambda_0(u) du.$$

Note that $M_{ij}(t)$ is a local square-integrable martingale with respect to the marginal filtration $\mathcal{F}_{ij}(t) = \sigma\{N_{ij}(u), Y_{ij}(u), \mathbf{Z}_{ij}(u) : 0 \leq u \leq t\}$. However, due to the within-cluster dependence, $M_{ij}(t)$ is not a martingale with respect to the joint filtration generated by the history of all the failure, censoring and covariate information up to time t. Regression based on (3.1) would yield consistent estimators and valid inference. However, when cluster sizes are informative, the model estimators may be asymptotically biased by a direct application of (3.1). The causes for cluster sizes being informative can be complicated and usually unknown, since some latent variables may implicitly affect the baseline hazard for each cluster and/or the covariates.

3.2.2 The Within-Cluster Resampling Method

Randomly sample, with replacement, one individual from each of the m clusters. The bth resampled dataset denoted by $\{X_i^b, \Delta_i^b, \mathbf{Z}_i^b(t); i = 1, \ldots, m, t \in [0, \tau]\}$, consists of m independent observations, which can be analyzed using the Cox proportional hazards model for independent failure time data. For $b = 1, \ldots, B$, we introduce the following necessary notation:

$$\mathbf{S}_b^{(k)}(\boldsymbol{\beta}, t) = m^{-1} \sum_{i=1}^{m} Y_i^b(t) \{\mathbf{Z}_i^b(t)\}^{\otimes k} \exp(\boldsymbol{\beta}' \mathbf{Z}_i^b(t)),$$

$$\mathbf{s}^{(k)}(\boldsymbol{\beta}, t) = E\{\mathbf{S}_b^{(k)}(\boldsymbol{\beta}, t)\}, \quad \mathbf{e}(\boldsymbol{\beta}, t) = \mathbf{s}^{(1)}(\boldsymbol{\beta}, t) / s^{(0)}(\boldsymbol{\beta}, t),$$

$$\mathbf{V}_b(\boldsymbol{\beta}, t) = \mathbf{S}_b^{(2)}(\boldsymbol{\beta}, t)/S_b^{(0)}(\boldsymbol{\beta}, t) - \{\mathbf{S}_b^{(1)}(\boldsymbol{\beta}, t)/S_b^{(0)}(\boldsymbol{\beta}, t)\}^{\otimes 2},$$

where $\mathbf{a}^{\otimes k} = \mathbf{1}, \mathbf{a}, \mathbf{a}\mathbf{a}'$, for $k = 0, 1, 2$.

For the bth resampled data, the partial likelihood function is given by

$$L_b(\boldsymbol{\beta}) = \prod_{i=1}^{m} \left\{ \frac{\exp(\boldsymbol{\beta}' \mathbf{Z}_i^b(X_i^b))}{m S_b^{(0)}(\boldsymbol{\beta}, X_i^b)} \right\}^{\Delta_i^b}, \tag{3.2}$$

and accordingly, the score function is

$$\mathbf{U}_b(\boldsymbol{\beta}) = \sum_{i=1}^{m} \int_0^{\tau} \left\{ \mathbf{Z}_i^b(t) - \frac{\mathbf{S}_b^{(1)}(\boldsymbol{\beta}, t)}{S_b^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i^b(t). \tag{3.3}$$

Solving $\mathbf{U}_b(\boldsymbol{\beta}) = 0$, one can obtain a consistent estimator for $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}}_b$. Analogously, the baseline cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ can be estimated by the Breslow-Aalen estimator, which for the bth resampled dataset is given by

$$\hat{\Lambda}_0^b(t, \hat{\boldsymbol{\beta}}_b) = \sum_{i=1}^{m} \int_0^t \frac{dN_i^b(u)}{\sum_{j=1}^{m} Y_j^b(u) \exp(\hat{\boldsymbol{\beta}}_b' \mathbf{Z}_j^b(u))}.$$

Repeating this procedure a large number of times, the WCR estimator for $\boldsymbol{\beta}$ is constructed as the average of the B resample-based estimates,

$$\bar{\boldsymbol{\beta}}_{\text{wcr}} = \frac{1}{B} \sum_{b=1}^{B} \hat{\boldsymbol{\beta}}_b, \tag{3.4}$$

and similarly, the WCR estimator for $\Lambda_0(t)$ is

$$\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) = \frac{1}{B} \sum_{b=1}^{B} \hat{\Lambda}_0^b(t, \hat{\boldsymbol{\beta}}_b), \tag{3.5}$$

where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_B)$.

An attractive feature of the WCR method is that the estimates can be obtained by maximizing the standard partial likelihood function for independent data without specifying any correlation matrix. Simply by averaging over the estimates from

the resampled data, we can obtain a consistent estimators for $\boldsymbol{\beta}$, and the variance-covariance matrix of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ in a relatively straightforward fashion, as shown in the next section.

3.2.3 Inference Procedures

Under certain regularity conditions (Anderson and Gill, 1982) and (Fleming and Harrington, 1990, pp.289-290)), for each resampled dataset, $\hat{\boldsymbol{\beta}}_b$ is consistent and asymptotically normal. To prove the asymptotic normality of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$, the central limit theorem (CLT) cannot be directly applied since $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ is the average of B identically distributed but dependent maximum partial likelihood estimators. Following similar arguments as in (Hoffman et al., 2001), $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ can be rewritten as as the sum of m independent cluster-specific terms so that the multivariate CLT can be applied. The asymptotic normality is stated in the following theorem, for which the proof is outlined in the Appendix.

Theorem 3.2.1. *Under regularity conditions, as the number of clusters $m \to \infty$, the within-cluster resampling estimator $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ is asymptotically normal, that is*

$$\sqrt{m}(\bar{\boldsymbol{\beta}}_{\text{wcr}} - \boldsymbol{\beta}_0) \to N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

in distribution, where $\boldsymbol{\Sigma}$ is a finite and positive definite matrix.

A consistent estimator for $\boldsymbol{\Sigma}$ is given as

$$\hat{\boldsymbol{\Sigma}} = \frac{m}{B}\left\{\sum_{b=1}^{B}\hat{\boldsymbol{\Sigma}}_b - (B-1)\hat{\boldsymbol{\Omega}}\right\}, \tag{3.6}$$

where $\hat{\boldsymbol{\Sigma}}_b$ is the estimated variance-covariance matrix for $\hat{\boldsymbol{\beta}}_b$ given by

$$\hat{\boldsymbol{\Sigma}}_b = \left\{ \sum_{i=1}^{m} \int_0^{\tau} \mathbf{V}_b(\hat{\boldsymbol{\beta}}_b, t) dN_i^b(t) \right\}^{-1}, \tag{3.7}$$

and $\hat{\boldsymbol{\Omega}}$ is the estimated covariance matrix among the B resample-based estimates $\hat{\boldsymbol{\beta}}_b$,

$$\hat{\boldsymbol{\Omega}} = (B-1)^{-1} \sum_{b=1}^{B} (\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})'.$$

The consistency of $\hat{\boldsymbol{\Sigma}}$ is given in Theorem 2, with a sketched proof in the Appendix.

Theorem 3.2.2. *Under the same regularity conditions, $\hat{\boldsymbol{\Sigma}}$ is consistent for $\boldsymbol{\Sigma}$.*

Let $W(t) = \sqrt{m}(\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) - \Lambda_0(t))$, $t \in [0, \tau]$, and let $\mathcal{W}(t)$ be a zero-mean Gaussian process with a finite covariance function between $\mathcal{W}(t_1)$ and $\mathcal{W}(t_2)$.

Theorem 3.2.3. *The random process $W(t)$ converges weakly to $\mathcal{W}(t)$ for $t \in [0, \tau]$.*

Similar to equation (3.6), the covariance function of $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$, between time t_1 and t_2, can be estimated by

$$\frac{m}{B} \left[\sum_{b=1}^{B} \text{Cov}\{\hat{\Lambda}_0^b(t_1, \hat{\boldsymbol{\beta}}_b), \hat{\Lambda}_0^b(t_2, \hat{\boldsymbol{\beta}}_b)\} - (B-1) \sum_{b=1}^{B} \{\hat{\Lambda}_0^b(t_1, \hat{\boldsymbol{\beta}}_b) - \bar{\Lambda}_0(t_1, \hat{\boldsymbol{\beta}})\}\{\hat{\Lambda}_0^b(t_2, \hat{\boldsymbol{\beta}}_b) - \bar{\Lambda}_0(t_2, \hat{\boldsymbol{\beta}})\} \right], \tag{3.8}$$

where $\text{Cov}\{\hat{\Lambda}_0^b(t_1, \hat{\boldsymbol{\beta}}_b), \hat{\Lambda}_0^b(t_2, \hat{\boldsymbol{\beta}}_b)\}$ can be estimated by

$$\int_0^{\min(t_1, t_2)} \frac{d\hat{\Lambda}_0^b(u)}{S_b^{(0)}(\hat{\boldsymbol{\beta}}_b, u)} + \mathbf{H}'(\hat{\boldsymbol{\beta}}_b, t_1) \hat{\boldsymbol{\Sigma}}_b \mathbf{H}(\hat{\boldsymbol{\beta}}_b, t_2),$$

and

$$\mathbf{H}(\hat{\boldsymbol{\beta}}_b, t) = -\sum_{i=1}^{m} \int_0^t \frac{\mathbf{S}_b^{(1)}(\hat{\boldsymbol{\beta}}_b, u)}{m\{S_b^{(0)}(\hat{\boldsymbol{\beta}}_b, u)\}^2} dN_i^b(u).$$

3.2.4 Weighted Marginal Model

To avoid the intensive computation of WCR, consider extending the work of Williamson et al. (2003) to the clustered survival data with informative cluster sizes. Consider the following weighted score function

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{m} \frac{1}{n_i} \sum_{j=1}^{n_i} \int_0^{\tau} \left\{ \mathbf{Z}_{ij}(t) - \frac{\sum_{k=1}^{m} \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(t) \mathbf{Z}_{kl}(t) \exp(\boldsymbol{\beta}' \mathbf{Z}_{kl}(t))}{\sum_{k=1}^{m} \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(t) \exp(\boldsymbol{\beta}' \mathbf{Z}_{kl}(t))} \right\} dN_{ij}(t). \tag{3.9}$$

By setting $\mathbf{U}(\boldsymbol{\beta}) = 0$, one can obtain the estimator from the weighted marginal model, denoted by $\hat{\boldsymbol{\beta}}_{\text{wmm}}$. To account for the informative cluster sizes, the contribution of each individual to the overall estimating equation is inversely weighted by the corresponding cluster size. Similar to the WCR approach, this weighting scheme eliminates the overweight for larger clusters as opposed to the standard unweighted marginal models. Asymptotic properties of this estimator can be justified following arguments similar to those in Cai and Prentice (1997) by letting the weights equal the inverse of each cluster size.

The covariance matrix of $\sqrt{m}(\hat{\boldsymbol{\beta}}_{\text{wmm}} - \boldsymbol{\beta}_0)$ can be consistently estimated by the following robust "sandwich" form estimator:

$$\boldsymbol{\Gamma} = \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_{\text{wmm}}) \mathbf{V} \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_{\text{wmm}}), \tag{3.10}$$

where $\mathbf{I}(\hat{\boldsymbol{\beta}}_{\text{wmm}})$ is the information matrix with $\boldsymbol{\beta}_0$ replaced by $\hat{\boldsymbol{\beta}}_{\text{wmm}}$,

$$\mathbf{V} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n_i^2} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} \hat{\mathbf{U}}_{ij}(\tau) \hat{\mathbf{U}}'_{ik}(\tau),$$

and

$$\hat{\mathbf{U}}_{ij}(\tau) = \int_0^\tau \left\{ \mathbf{Z}_{ij}(t) - \frac{\sum_{k=1}^m \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(t)\mathbf{Z}_{kl}(t) \exp(\hat{\boldsymbol{\beta}}'_{\text{wmm}}\mathbf{Z}_{kl}(t))}{\sum_{k=1}^m \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(t) \exp(\hat{\boldsymbol{\beta}}'_{\text{wmm}}\mathbf{Z}_{kl}(t))} \right\} d\hat{M}_{ij}(t).$$

The baseline cumulative hazard function at time t can be estimated using the weighted Breslow-Aalen estimator,

$$\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\text{wmm}}) = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \int_0^t \frac{dN_{ij}(u)}{\sum_{k=1}^m \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(u) \exp(\hat{\boldsymbol{\beta}}'_{\text{wmm}}\mathbf{Z}_{kl}(u))}.$$

The covariance function can be estimated as

$$\text{Cov}(\hat{\Lambda}_0(t_1, \hat{\boldsymbol{\beta}}_{\text{wmm}}), \hat{\Lambda}_0(t_2, \hat{\boldsymbol{\beta}}_{\text{wmm}})) = \sum_{i=1}^m \psi_i(t_1)\psi_i(t_2), \tag{3.11}$$

where

$$\psi_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} \int_0^t \frac{d\hat{M}_{ij}(u)}{\sum_{k=1}^m \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(u) \exp(\hat{\boldsymbol{\beta}}'_{\text{wmm}}\mathbf{Z}_{kl}(u))} - \mathbf{H}'(t)\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_{\text{wmm}})\hat{\mathbf{U}}_i(t),$$

$$\hat{\mathbf{U}}_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{\mathbf{U}}_{ij}(t),$$

and

$$\mathbf{H}(t) = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \int_0^t \frac{\sum_{k=1}^m \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(u)\mathbf{Z}_{kl}(u) \exp(\hat{\boldsymbol{\beta}}'_{\text{wmm}}\mathbf{Z}_{kl}(u))}{\{\sum_{k=1}^m \frac{1}{n_k} \sum_{l=1}^{n_k} Y_{kl}(u) \exp(\hat{\boldsymbol{\beta}}'_{\text{wmm}}\mathbf{Z}_{kl}(u))\}^2} dN_{ij}(u).$$

3.3 Simulation Studies

Simulation studies were conducted to assess the performance of the proposed methods. Different scenarios were considered by varying within-cluster correlations, censoring rates, and most importantly, informative and non-informative cluster sizes on covariate effects. Two sets of simulations were conducted: one with datasets of clusters

$m = 100$ and the other with $m = 200$. For each set of simulation, I simulated 1,000 datasets and analyzed each dataset using the WCR method with $B = 2,000$ resamples, the WMM and the MM methods under the working independence assumption.

3.3.1 Simulation Procedure

Correlated failure times were simulated using the Cox PH model with positive stable frailty (Hougaard, 1986),

$$\lambda(t_{ij}|\mathbf{Z}_{ij}, w_i) = \lambda_{0i} w_i \exp(\boldsymbol{\gamma}_0' \mathbf{Z}_{ij})$$

where w_i follows a positive stable distribution with parameter α, $\alpha \in (0, 1)$.

- *Positive stable frailty* w_i. The family of stable distributions is usually characterized with four parameters $\alpha, \eta, \sigma, \delta$, where $\alpha \in (0, 2]$ is the index of stability, $\eta \in [-1, 1]$ the skewness parameter, $\sigma > 0$ the scale parameter and $\delta \in (-\infty, +\infty)$ the location parameter (Nolan, 2006). When $\alpha \in (0, 1), \eta = 1, \delta = 0$, the distribution has a support on the positive half of the real line and is called the positive stable distribution. In this simulation, I fixed the scale parameter $\sigma = 1$ for identifiability purpose and the positive stable distribution has essentially one parameter α. A positive stable variable can be simulated using the following representation (Chambers et al., 1976; Nolan, 2006),

$$W = (a(\theta)/\xi)^{(1-\alpha)/\alpha}$$

where θ and ξ are independent, θ is uniform on $(0, \pi)$, ξ is exponential with mean one, and

$$a(\theta) = \frac{(\sin(1-\alpha)\theta)(\sin \alpha\theta)^{\alpha/(1-\alpha)}}{(\sin \theta)^{1/(1-\alpha)}}.$$

The parameter α represents the degree of correlation between cluster members, with $\alpha \to 0$ giving maximal positive dependence, and $\alpha \to 1$ corresponding to the independent case. Let α equal 0.5 and 0.75, corresponding to the within-cluster correlation $\rho = 0.3, 0.15$, respectively. For ease of exposition, I considered a constant baseline hazard, $\lambda_{0i}(t) = 1$. After integrating out the frailty, the true marginal regression parameters are $\boldsymbol{\beta}_0 = \alpha\boldsymbol{\gamma}_0$ and the true baseline cumulative hazard at time t is $\Lambda_0(t) = t^{\alpha}$.

- *Cluster size* n_i. The size of each cluster is either randomly generated, the non-informative case, or depends on the frailty value. To simulate informative cluster sizes, I let the size of each cluster depend on the value of the generated frailty such that

$$n_i = 2 + k/10, \text{if } q_k \leq w_i < q_{k+10}, k = 0, 10, 20, \ldots, 90.$$

where q_k is the kth percentile of the frailty distribution. As a result, the cluster sizes vary from 2 to 11, depending on which percentile range the frailty values fall. For the non-informative cases, the cluster sizes are randomly taken from $\{2, \ldots, 11\}$ no matter what the frailty values are.

- *Correlated failure times* T_{ij}. For cluster i, given the generated the frailty w_i,

the survival times of the individuals within this cluster, $T_{ij}, j = 1, \ldots, n_i$, can be independently generated from the above model.

- *Censoring times* C_{ij}. The censoring times were generated independently from a uniform distribution $U(0, a)$, where the value of a can be selected to achieve desired censoring rates.

- *Covariates* $\mathbf{Z}_{ij}$. Two covariates were included in the simulation: one was a binary variable, Z_1, taking the value of 0 or 1 with an equal probability of 0.5, which may represent the treatment or control group; the other was a continuous variable, Z_2, independently generated from a uniform $U(0, 1)$ distribution.

3.3.2 Simulation Results

For each simulated dataset, I obtained the point estimates and standard errors of the regression coefficients using all three methods, $\bar{\boldsymbol{\beta}}_{\text{wcr}}$, $\hat{\boldsymbol{\beta}}_{\text{wmm}}$ and $\hat{\boldsymbol{\beta}}_{\text{mm}}$. I also calculated the sample standard deviation over 1,000 simulations, the mean standard error and the 95% confidence interval coverage rate for each estimated coefficient.

3.3.2.1 Informative cluster size

This section shows the simulation results when cluster sizes are informative. As shown in Table 3.1, when cluster sizes are informative, the point estimates of the regression coefficients using the WCR and WMM methods are approximately unbiased and for which the 95% confidence interval coverage rates are close to the nominal value,

whereas the MM estimates are substantially biased. The variation of the parameter estimates decreases when the number of clusters increases, and increases when the censoring rate increases. The sample standard deviations (SD) are close to the mean standard errors (SE) for the WCR and WMM methods, which suggests that our variance estimators (3.6) and (3.10) provide good estimates for the variability of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ and $\hat{\boldsymbol{\beta}}_{\text{wmm}}$.

Figure 3.2 shows the quantile-quantile (Q-Q) plots for the estimated coefficients after being standardized versus a standard normal distribution. All six Q-Q plots appear to lie closely on the identity line, which indicate that the parameter estimators follow approximately normal distributions. The deviation from the identity line (the solid line) in the bottom two plots shows the bias in the MM estimator.

For the cumulative baseline hazard, we computed the point-wise estimates and standard errors for some selected time points, which were the 20th, 40th, 60th and 80th percentiles of the underlying true failure time distribution. The results were shown in Tables 3.2 and 3.3. The WCR estimator, $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$, and the WMM estimator $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\text{wmm}})$ provide point-wise estimates that are very close to the true cumulative baseline hazard with satisfactory 95% confidence interval coverage rate. The MM estimator is significantly biased with poor coverage rate. The WCR and MM methods yielded similar point-wise estimates for the cumulative baseline hazard, however, the WCR estimator has uniformly better 95% confidence interval coverage rates. The similarity between the columns SD and SE suggests that (3.8) and (3.11) serve as

good estimators for the variance of $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$ and $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\text{wmm}})$, respectively.

Figure 3.3 shows the Q-Q plots for the point-wise estimates of the cumulative baseline hazard at two time points, i.e. $t = 0.04$ and $t = 0.25$. All plots appear as straight lines, indicating asymptotic normality. However, on the ones for the MM estimates, the points are substantially deviant from the solid identity line, which shows the biasedness of the MM estimator similarly as for the regression estimates.

3.3.2.2 Non-informative cluster size

Similar simulation studies are conducted for non-informative cluster sizes. Results are shown in a similar fashion as in the previous section. Table 3.4 shows the results for the regression parameters. One can see that when cluster sizes are not informative, all point estimates are approximately unbiased and the coverage rates of all three methods are reasonably close to the nominal level. The variation of the parameter estimates decreases when the number of clusters increases, and increases when the censoring rate increases. In Figure 3.4, all six Q-Q plots appear as straight line and stay close to the identity line.

For the cumulative baseline hazard, as shown in Tables 3.5 and 3.6, all three models perform comparably. As an example, in Figure 3.5 the Q-Q plots at the two time points are similar for all three models. The points align closely along with the identity line except at the tails.

Table 3.1: Simulation results for $\boldsymbol{\beta}$ with informative cluster size

			WCR				WMM				MM			
	True	Cen%	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
							$m = 200$							
	0.5	25	0.500	0.090	0.089	94.0	0.499	0.089	0.090	94.7	0.466	0.067	0.068	92.1
		50	0.498	0.105	0.099	92.1	0.495	0.104	0.101	94.0	0.470	0.085	0.083	92.4
	0.75	25	0.757	0.088	0.083	93.6	0.753	0.088	0.084	93.8	0.710	0.070	0.067	88.9
		50	0.756	0.092	0.087	93.0	0.751	0.091	0.088	94.6	0.710	0.074	0.071	90.1
β_1							$m = 300$							
	0.5	25	0.501	0.074	0.072	93.6	0.500	0.073	0.072	93.9	0.464	0.053	0.053	88.7
		50	0.501	0.076	0.073	93.8	0.500	0.076	0.074	94.4	0.467	0.056	0.055	89.5
	0.75	25	0.751	0.069	0.068	94.8	0.748	0.068	0.068	95.8	0.705	0.051	0.054	87.8
		50	0.751	0.072	0.072	94.7	0.748	0.072	0.073	95.2	0.706	0.056	0.058	89.3
							$m = 200$							
	0.5	25	0.501	0.154	0.150	93.5	0.499	0.153	0.152	95.0	0.466	0.111	0.113	94.4
		50	0.508	0.175	0.165	92.3	0.506	0.173	0.170	94.7	0.479	0.140	0.137	93.7
	0.75	25	0.766	0.144	0.139	93.6	0.760	0.143	0.141	94.3	0.716	0.107	0.109	94.0
		50	0.764	0.152	0.146	94.1	0.759	0.150	0.150	95.4	0.715	0.117	0.117	94.4
β_2							$m = 300$							
	0.5	25	0.504	0.120	0.121	95.4	0.502	0.120	0.121	95.8	0.466	0.086	0.087	92.9
		50	0.505	0.124	0.123	94.7	0.504	0.124	0.124	96.2	0.469	0.092	0.092	94.0
	0.75	25	0.753	0.121	0.114	92.9	0.749	0.121	0.116	94.2	0.704	0.090	0.089	91.1
		50	0.753	0.125	0.120	93.8	0.749	0.124	0.122	95.2	0.705	0.096	0.096	93.1

Est: estimate; SD: standard deviation; SE: standard error; CR: 95% confidence interval coverage rate in percentage; True: true value; Cen%: censoring rate.

Figure 3.2: Q-Q plots for β with informative cluster sizes: 1,000 simulations with $m = 200$, $\alpha = 0.5$ and 50% censoring rate.

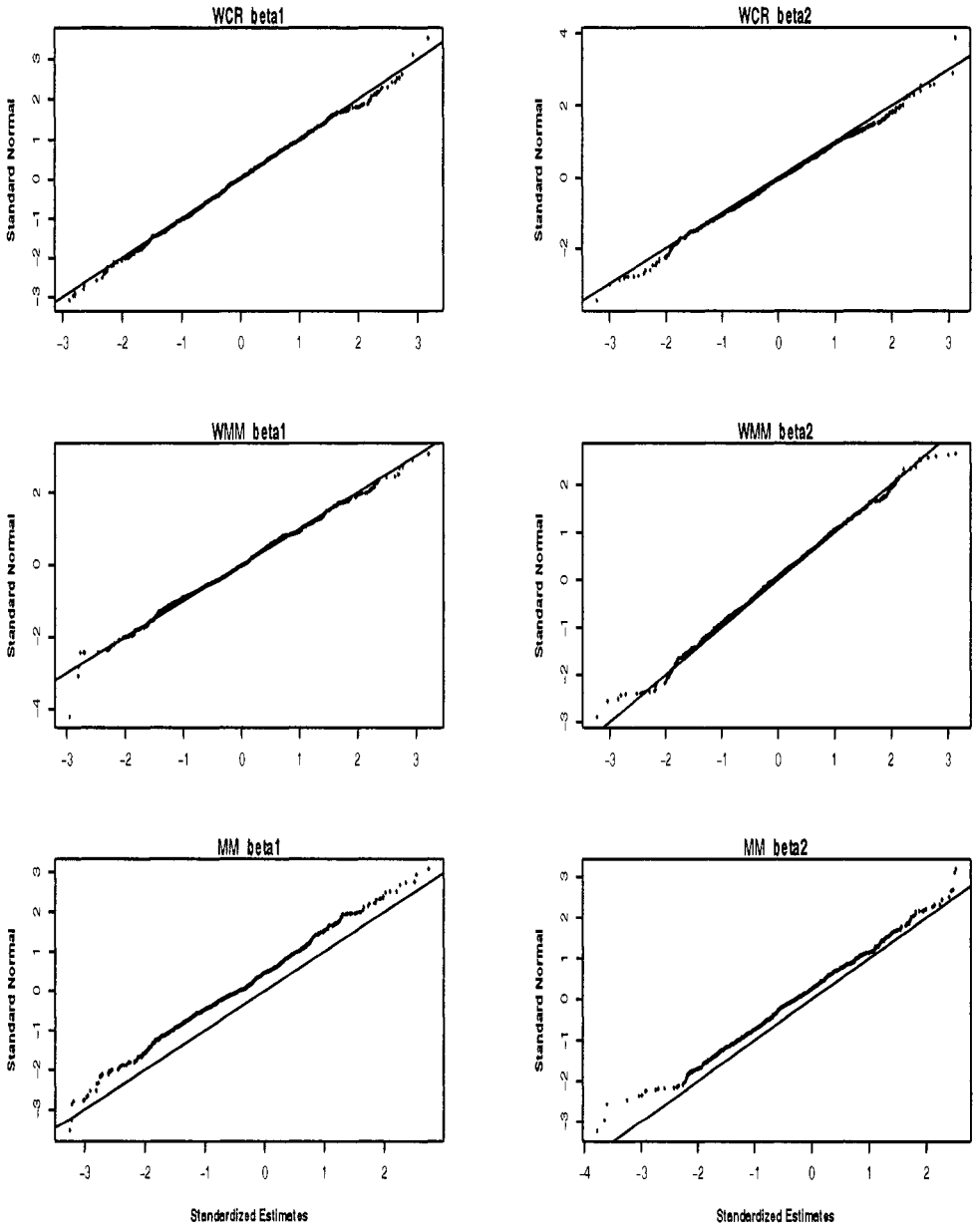


Table 3.2: Simulation results for $\Lambda_0(t)$ with informative cluster sizes I: $\alpha = 0.5$.

			WCR				WMM				MM			
Cen%	t	$\Lambda_0(t)$	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
							$m = 200$							
	0.04	0.200	0.202	0.030	0.031	95.8	0.201	0.030	0.029	94.2	0.333	0.045	0.042	9.2
	0.15	0.387	0.392	0.051	0.051	94.7	0.389	0.050	0.047	92.6	0.619	0.068	0.061	2.1
	0.25	0.500	0.505	0.061	0.063	95.1	0.502	0.061	0.058	93.2	0.784	0.078	0.071	1.2
	0.50	0.707	0.716	0.082	0.084	95.7	0.710	0.081	0.076	93.0	1.079	0.098	0.088	0.5
25%							$m = 300$							
	0.04	0.200	0.201	0.024	0.025	96.3	0.200	0.024	0.024	95.2	0.331	0.035	0.034	0.8
	0.15	0.387	0.389	0.040	0.041	96.1	0.388	0.039	0.039	94.7	0.616	0.052	0.050	0
	0.25	0.500	0.503	0.049	0.051	96.3	0.501	0.049	0.047	94.8	0.782	0.061	0.058	0
	0.50	0.707	0.712	0.066	0.068	96.1	0.708	0.065	0.062	94.6	1.076	0.077	0.072	0
							$m = 200$							
	0.04	0.200	0.203	0.031	0.031	95.7	0.202	0.031	0.030	94.5	0.332	0.046	0.043	10.5
	0.15	0.387	0.392	0.052	0.052	95.1	0.390	0.051	0.048	93.5	0.618	0.069	0.063	2.6
	0.25	0.500	0.507	0.063	0.064	95.7	0.503	0.062	0.059	93.4	0.782	0.080	0.074	1.7
	0.50	0.707	0.719	0.085	0.087	95.8	0.711	0.084	0.079	94.5	1.077	0.101	0.092	0.8
50%							$m = 300$							
	0.04	0.200	0.201	0.024	0.025	96.2	0.200	0.024	0.024	95.7	0.330	0.035	0.035	1.6
	0.15	0.387	0.389	0.040	0.042	95.7	0.387	0.040	0.039	94.7	0.614	0.053	0.051	0.2
	0.25	0.500	0.503	0.050	0.052	95.6	0.500	0.050	0.048	93.7	0.780	0.063	0.060	0.1
	0.50	0.707	0.713	0.068	0.070	96.1	0.708	0.068	0.064	94.4	1.074	0.081	0.075	0.1

Est: estimate; SD: standard deviation; SE: standard error; CR: 95% confidence interval coverage rate in percentage; Cen%: censoring rate.

Table 3.3: Simulation results for $\Lambda_0(t)$ with informative cluster sizes II: $\alpha = 0.75$.

			WCR				WMM				MM			
Cen%	t	$\Lambda_0(t)$	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
							$m = 200$							
	0.04	0.089	0.090	0.014	0.014	93.5	0.089	0.014	0.014	93.2	0.134	0.020	0.020	38.0
	0.15	0.241	0.242	0.032	0.030	93.4	0.241	0.032	0.030	93.0	0.343	0.039	0.038	20.7
	0.25	0.354	0.356	0.044	0.041	93.9	0.354	0.044	0.041	93.4	0.492	0.052	0.049	17.2
	0.50	0.595	0.600	0.069	0.065	92.7	0.596	0.069	0.062	92.0	1.803	0.077	0.069	13.8
25							$m = 300$							
	0.04	0.089	0.090	0.011	0.011	95.1	0.090	0.011	0.011	95.4	0.135	0.016	0.017	15.1
	0.15	0.241	0.243	0.025	0.025	95.5	0.243	0.025	0.025	95.3	0.346	0.031	0.032	4.4
	0.25	0.354	0.357	0.035	0.034	94.2	0.356	0.034	0.034	94.7	0.496	0.041	0.040	3.2
	0.50	0.595	0.601	0.055	0.053	93.8	0.598	0.055	0.051	93.3	0.808	0.061	0.056	2.4
							$m = 200$							
	0.04	0.089	0.090	0.015	0.015	93.7	0.090	0.015	0.014	93.4	0.134	0.021	0.021	41.0
	0.15	0.241	0.243	0.033	0.032	94.3	0.242	0.032	0.031	93.6	0.343	0.041	0.040	24.7
	0.25	0.354	0.358	0.045	0.044	93.0	0.355	0.045	0.043	92.8	0.492	0.054	0.052	20.3
	0.50	0.595	0.603	0.073	0.069	94.2	0.597	0.072	0.066	92.5	1.803	0.080	0.073	18.0
50							$m = 300$							
	0.04	0.089	0.090	0.012	0.012	95.2	0.090	0.012	0.012	95.2	0.135	0.017	0.017	19.1
	0.15	0.241	0.244	0.026	0.026	95.8	0.243	0.026	0.026	95.6	0.346	0.032	0.033	6.5
	0.25	0.354	0.358	0.036	0.036	95.5	0.356	0.036	0.035	95.8	0.496	0.043	0.042	5.4
	0.50	0.595	0.602	0.058	0.056	94.6	0.598	0.058	0.054	93.3	0.807	0.065	0.060	4.2

Est: estimate; SD: standard deviation; SE: standard error; CR: 95% confidence interval coverage rate in percentage; Cen%: censoring rate.

Figure 3.3: Q-Q plots for $\Lambda_0(t)$ with informative cluster sizes: 1,000 simulations with $m = 200$, $\alpha = 0.5$ and 50% censoring rate.

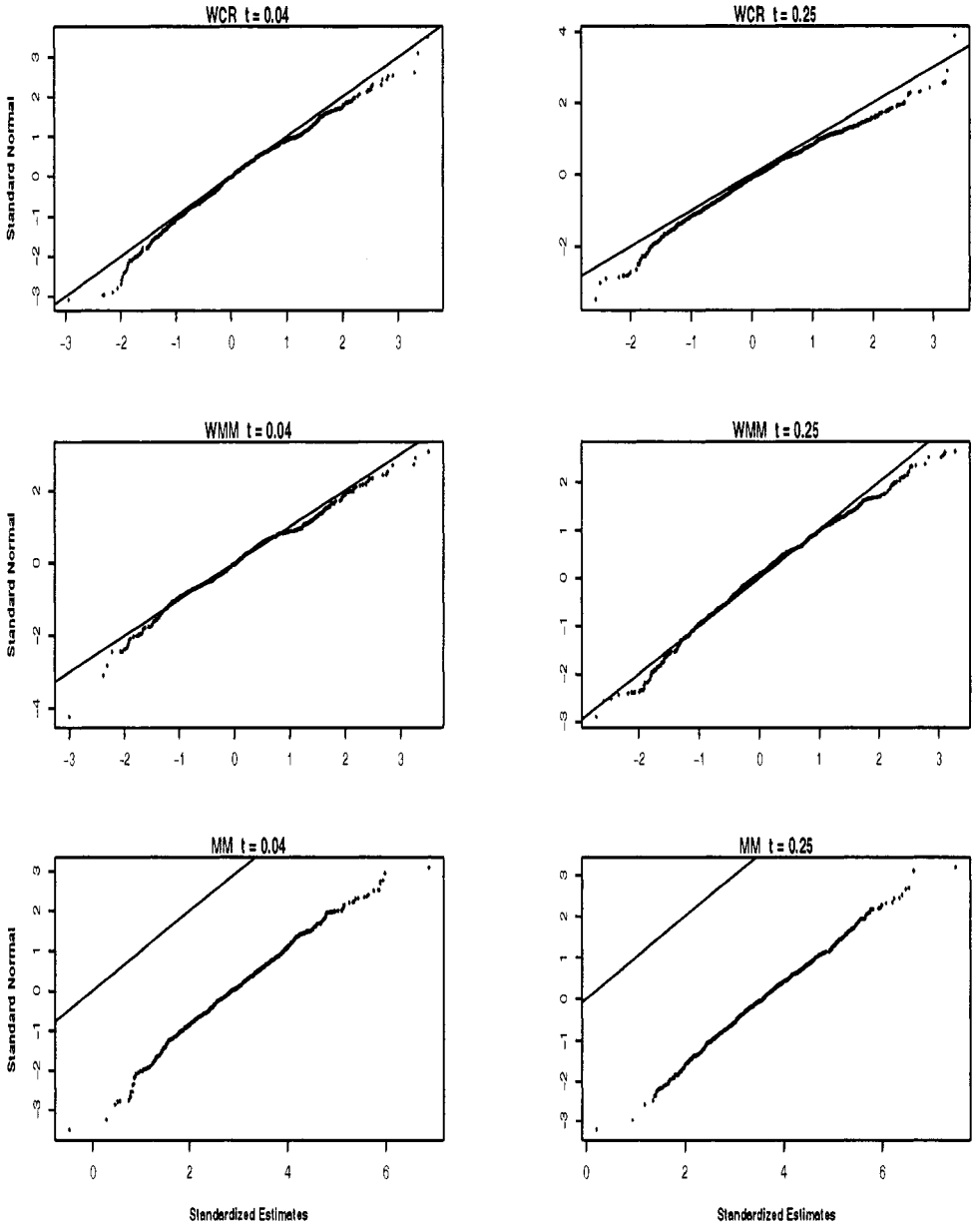


Table 3.4: Simulation results for $\boldsymbol{\beta}$ with non-informative cluster size.

			WCR				WMM				MM			
	True	Cen%	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
							$m = 200$							
	0.5	25	0.502	0.076	0.075	95.2	0.499	0.076	0.076	95.6	0.500	0.068	0.068	94.9
		50	0.506	0.087	0.082	92.9	0.503	0.086	0.083	93.9	0.503	0.076	0.074	93.8
	0.75	25	0.754	0.075	0.074	94.0	0.748	0.074	0.076	95.5	0.750	0.068	0.068	95.0
		50	0.759	0.088	0.082	92.5	0.754	0.087	0.084	93.6	0.754	0.075	0.075	95.3
β_1							$m = 300$							
	0.5	25	0.505	0.062	0.062	94.9	0.503	0.062	0.063	95.2	0.504	0.055	0.056	95.5
		50	0.506	0.068	0.067	93.6	0.505	0.067	0.068	94.7	0.505	0.060	0.061	95.7
	0.75	25	0.759	0.063	0.061	94.2	0.754	0.062	0.062	95.0	0.754	0.056	0.056	95.3
		50	0.758	0.069	0.067	93.4	0.755	0.069	0.068	95.4	0.754	0.061	0.061	95.1
							$m = 200$							
	0.5	25	0.503	0.133	0.125	92.2	0.500	0.131	0.128	94.7	0.500	0.114	0.113	94.8
		50	0.511	0.142	0.137	93.7	0.508	0.140	0.139	95.6	0.509	0.121	0.123	95.5
	0.75	25	0.755	0.127	0.123	92.6	0.749	0.124	0.126	95.3	0.749	0.110	0.112	94.4
		50	0.763	0.141	0.137	93.0	0.757	0.139	0.139	95.1	0.758	0.120	0.123	96.1
β_2							$m = 300$							
	0.5	25	0.501	0.107	0.104	93.9	0.499	0.106	0.105	95.0	0.503	0.091	0.093	95.7
		50	0.499	0.115	0.112	94.3	0.496	0.114	0.114	94.9	0.500	0.099	0.100	94.9
	0.75	25	0.752	0.107	0.102	91.9	0.748	0.106	0.103	93.3	0.752	0.092	0.091	94.7
		50	0.750	0.118	0.112	93.1	0.746	0.117	0.114	94.0	0.751	0.102	0.100	94.6

Est: estimate; SD: standard deviation; SE: standard error; CR: 95% confidence interval coverage rate in percentage; True: true value; Cen%: censoring rate.

Figure 3.4: Q-Q plots for $\boldsymbol{\beta}$ with non-informative cluster sizes: 1,000 simulations with $m = 200$, $\alpha = 0.5$ and 50% censoring rate.

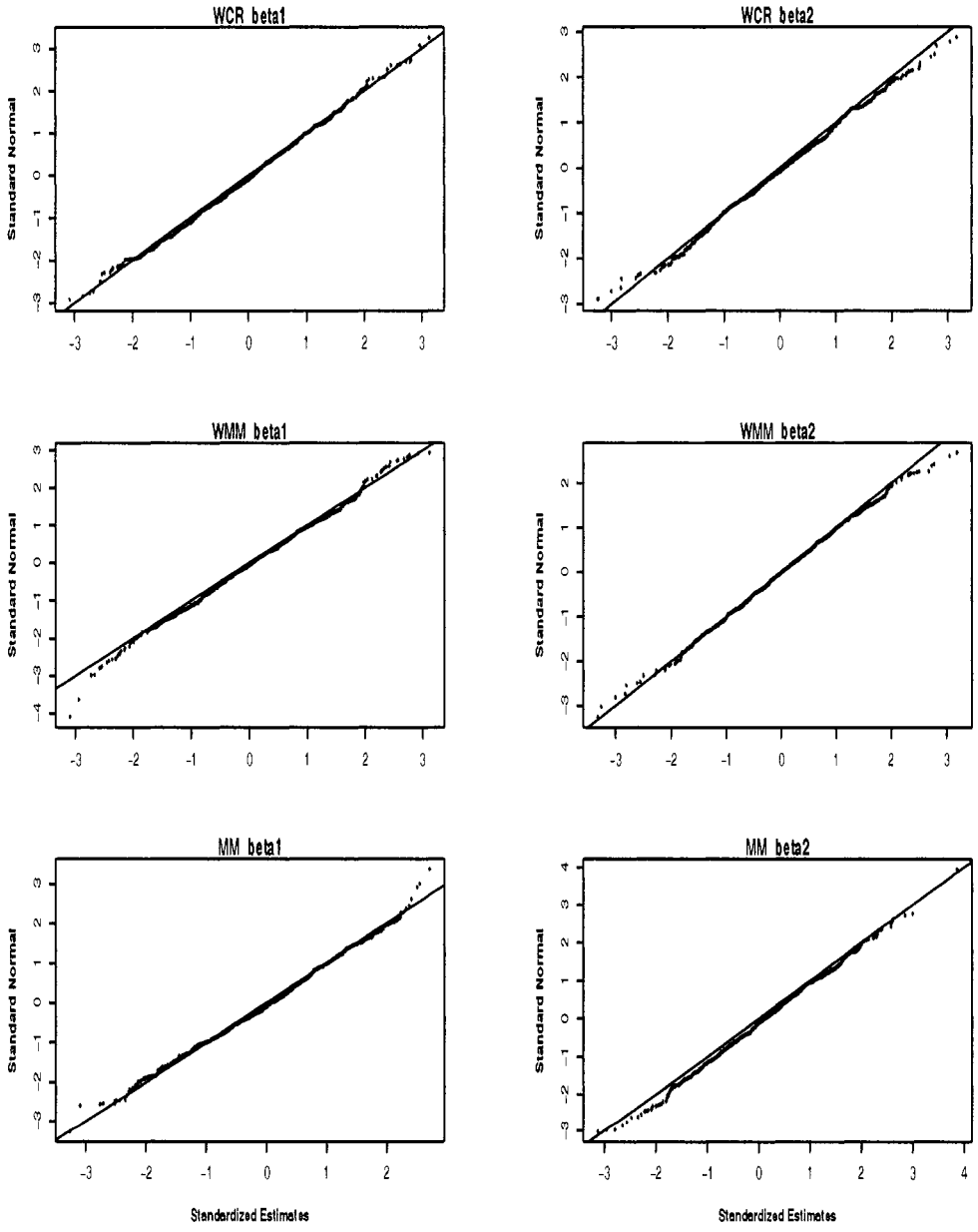


Table 3.5: Simulation results for $\Lambda_0(t)$ with non-informative cluster sizes I: $\alpha = 0.5$.

			WCR				WMM				MM			
Cen%	t	$\Lambda_0(t)$	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
							$m = 200$							
	0.04	0.200	0.202	0.028	0.030	95.1	0.201	0.028	0.027	93.2	0.201	0.029	0.028	94.0
	0.15	0.387	0.393	0.047	0.049	96.0	0.390	0.046	0.044	93.9	0.390	0.046	0.043	93.1
	0.25	0.500	0.507	0.057	0.060	95.3	0.504	0.056	0.053	93.8	0.504	0.056	0.051	92.0
	0.50	0.707	0.719	0.077	0.079	95.1	0.713	0.076	0.069	93.1	0.713	0.074	0.065	91.6
25							$m = 300$							
	0.04	0.200	0.201	0.024	0.024	95.0	0.201	0.024	0.022	93.5	0.200	0.024	0.023	93.8
	0.15	0.387	0.390	0.040	0.040	95.2	0.388	0.040	0.036	93.1	0.387	0.039	0.035	91.9
	0.25	0.500	0.504	0.049	0.048	94.5	0.501	0.048	0.043	92.3	0.500	0.046	0.042	91.3
	0.50	0.707	0.713	0.064	0.064	95.0	0.709	0.064	0.056	92.9	0.707	0.060	0.053	91.3
							$m = 200$							
	0.04	0.200	0.203	0.031	0.031	94.8	0.202	0.031	0.028	92.6	0.202	0.031	0.029	92.9
	0.15	0.387	0.393	0.052	0.051	93.8	0.390	0.051	0.046	90.7	0.390	0.050	0.045	92.6
	0.25	0.500	0.508	0.064	0.063	94.0	0.503	0.063	0.056	91.1	0.503	0.059	0.054	93.0
	0.50	0.707	0.718	0.084	0.084	94.2	0.710	0.083	0.074	90.6	0.710	0.078	0.069	92.5
50							$m = 300$							
	0.04	0.200	0.201	0.025	0.025	95.2	0.201	0.025	0.023	94.3	0.200	0.024	0.023	93.6
	0.15	0.387	0.390	0.042	0.041	95.0	0.389	0.041	0.038	92.6	0.388	0.040	0.037	92.2
	0.25	0.500	0.505	0.051	0.051	94.7	0.502	0.050	0.046	92.3	0.500	0.048	0.044	92.1
	0.50	0.707	0.716	0.067	0.068	94.6	0.710	0.067	0.060	92.6	0.708	0.063	0.057	91.3

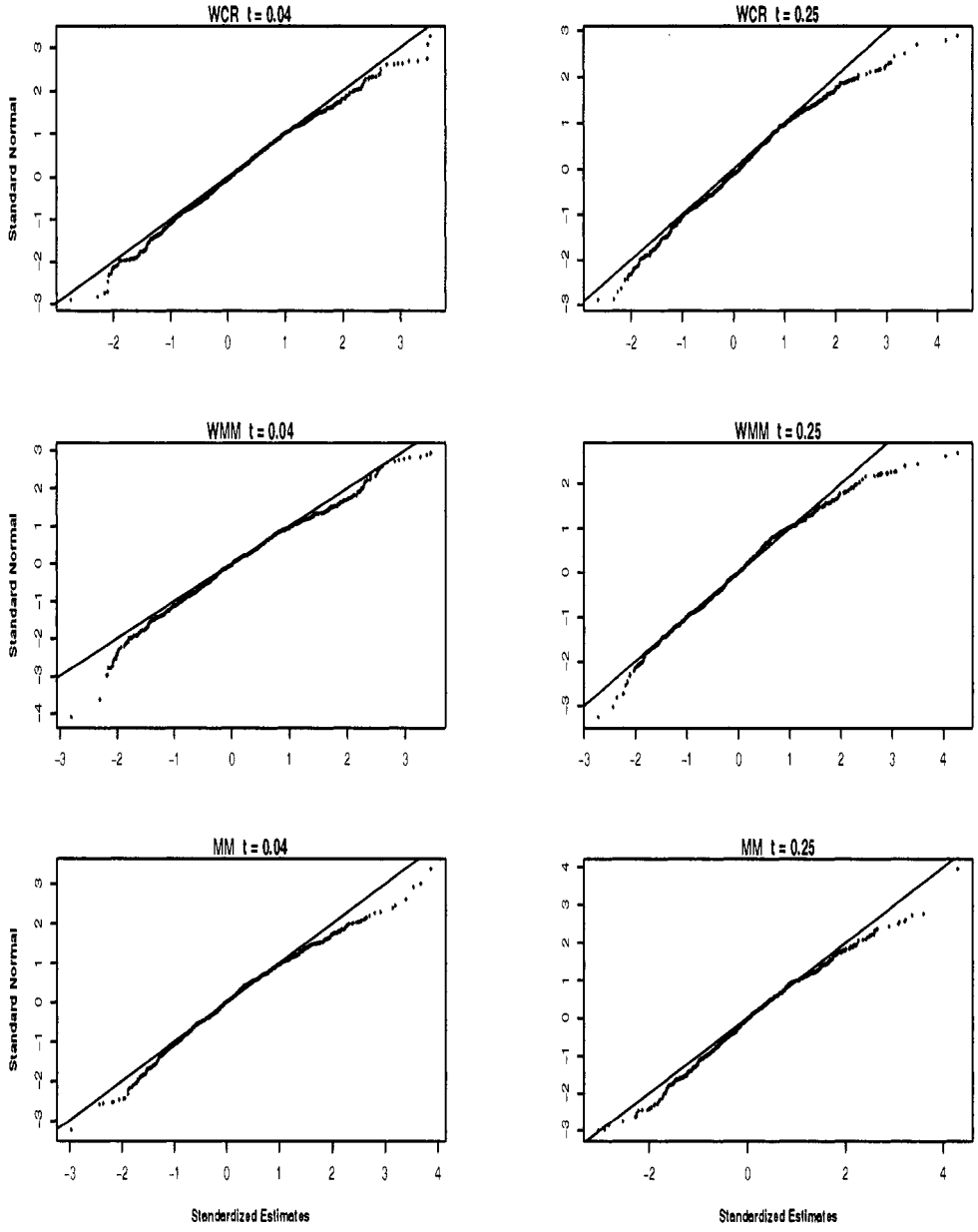
Est: estimate; SD: standard deviation; SE: standard error; CR: 95% confidence interval coverage rate in percentage; Cen%: censoring rate.

Table 3.6: Simulation results for $\Lambda_0(t)$ with non-informative cluster sizes II: $\alpha = 0.75$.

			WCR				WMM				MM			
Cen%	t	$\Lambda_0(t)$	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
							$m = 200$							
	0.04	0.089	0.091	0.013	0.014	94.6	0.090	0.013	0.013	94.2	0.090	0.013	0.013	94.6
	0.15	0.241	0.244	0.028	0.030	96.3	0.243	0.028	0.028	95.7	0.243	0.027	0.027	95.0
	0.25	0.354	0.358	0.038	0.040	95.6	0.356	0.038	0.038	95.1	0.356	0.036	0.037	94.7
	0.50	0.595	0.602	0.060	0.062	95.0	0.598	0.059	0.057	94.0	0.598	0.057	0.054	93.0
25							$m = 300$							
	0.04	0.089	0.090	0.011	0.011	94.7	0.090	0.011	0.011	94.5	0.089	0.011	0.011	93.9
	0.15	0.241	0.242	0.024	0.024	94.7	0.241	0.024	0.023	94.6	0.241	0.023	0.022	93.9
	0.25	0.354	0.355	0.033	0.033	94.1	0.353	0.033	0.031	93.2	0.353	0.031	0.030	93.7
	0.50	0.595	0.599	0.051	0.050	94.8	0.596	0.051	0.047	92.9	0.594	0.047	0.044	93.3
							$m = 200$							
	0.04	0.089	0.090	0.015	0.015	94.4	0.090	0.014	0.014	93.8	0.090	0.014	0.014	95.5
	0.15	0.241	0.244	0.032	0.031	94.0	0.242	0.032	0.030	93.4	0.242	0.030	0.029	95.0
	0.25	0.354	0.357	0.044	0.043	94.3	0.354	0.043	0.041	93.4	0.354	0.039	0.039	94.7
	0.50	0.595	0.602	0.069	0.068	94.0	0.596	0.068	0.062	92.5	0.594	0.061	0.058	93.5
50							$m = 300$							
	0.04	0.089	0.090	0.012	0.012	95.2	0.090	0.012	0.011	94.4	0.089	0.011	0.011	93.9
	0.15	0.241	0.243	0.026	0.025	94.8	0.242	0.026	0.025	94.2	0.241	0.024	0.024	94.0
	0.25	0.354	0.356	0.035	0.035	93.8	0.353	0.035	0.034	93.7	0.353	0.033	0.032	93.9
	0.50	0.595	0.601	0.055	0.055	94.2	0.597	0.055	0.051	93.2	0.595	0.050	0.048	93.9

Est: estimate; SD: standard deviation; SE: standard error; CR: 95% confidence interval coverage rate in percentage; Cen%: censoring rate.

Figure 3.5: Q-Q plots for $\Lambda_0(t)$ with non-informative cluster sizes: 1,000 simulations with $m = 200$, $\alpha = 0.5$ and 50% censoring rate.



3.4 Data Application

I applied the WCR (with $B = 10,000$), WMM and MM methods to the motivating example, the dental study in Spiekerman and Lin (1998). The original study was conducted by McGuire and Nunn (1996) to assess the effect of some commonly measured clinical factors in predicting tooth survival. The dataset consists of 100 consecutive patients from Dr. McGuire's appointment book. All of these patients had been diagnosed with moderate to severe chronic adult periodontitis and had received at least five years of maintenance care. For this analysis, I considered the effect on tooth survival for two covariates, age and smoking status (0 = smoker, and 1 otherwise). The failure time for each tooth is defined as the time to tooth loss measured from the initiation of active periodontal therapy.

In this analysis, only the upper and lower molars were considered (a normal person should have a maximum of eight molars). There were 96 patients with both upper and lower molars, i.e. the number of clusters was 96. The cluster size ranged from 1 to 8. The total number of teeth was 598 with 58 observed failures. As illustrated in section 3.1.1, the cluster size might be informative to tooth survival. Patients who have more teeth tend to have higher tooth survival probability; more precisely, larger cluster sizes indicate better survival.

The analysis results are summarized in Table 4, where I compare the estimates of the regression coefficients using the WCR and WMM methods with those using the standard MM. The point estimate for the effect of smoking status is similar using

Table 3.7: Estimates of the regression coefficients for the dental study.

	WCR		WMM		MM	
Covariate	Estimate	SE	Estimate	SE	Estimate	SE
Smoker	-0.501	0.304	-0.500	0.342	-0.558	0.340
Age	1.877	0.984	1.721	0.743	1.759	0.724

WCR and WMM but quite different from that obtained by the MM method. This might be due to the informativeness of cluster sizes. The hazard ratio of tooth loss for cigarette smoking is 1.647 (WCR and WMM) and 1.747 (MM) with overlapping 95% confidence intervals. Smoking is a very important factor that hastens tooth loss. The estimates of the age effect from the three methods are consistent, indicating that older patients would lose their tooth sooner than younger ones.

3.5 Discussion

When cluster sizes are informative to the covariate effect on the outcome of interest, the estimated regression coefficients could be substantially biased when using the standard marginal model approach. In contrast, both the within-cluster resampling and weighted marginal model methods provide valid estimates. Simulation studies have shown that the estimates under the within-cluster resampling and weighted marginal model methods are approximately unbiased with reasonable 95% confidence

interval coverage. On the other hand, when cluster sizes are informative such that the hazard ratio increases as cluster sizes get bigger, any estimation procedure that overweights the larger clusters will overestimate the covariate effect. This is precisely shown in our simulation studies where the estimates by the standard marginal model are positively biased.

The merit, among others, of the within-cluster resampling method is that it is unnecessary to make assumptions of the underlying correlation structure. By resampling one observation out of each cluster, the estimation problem is reduced to an independent case, hence standard methods for independent data can be readily applied.

The within-cluster resampling method is computationally intensive due to the resampling scheme. The variance estimators defined in (3.6) and (3.8) involve the subtraction of two terms, thus it is possible to obtain negative estimates. However, this is very rare in the simulations. In particular, the occurrence frequency is less than one out of a thousand when the number of resampling is large. Another potential problem in applying the within-cluster resampling method is that the estimator from a single resampling dataset might be unstable under heavy censoring since each resample only consists of partial information from the original data.

Chapter 4

Conclusions and Future Work

In summary, I have considered age-specific estimation of screening sensitivity and sojourn time distribution for breast cancer screening trials, and marginal analysis of multivariate survival data when cluster sizes are informative.

4.1 Breast Cancer Screening

I generalized the model framework of Shen and Zelen (1999) to incorporate age information in estimating screening sensitivity and sojourn time distribution. Three models were proposed: the first one considered screening sensitivity as age-dependent through a generalized linear model while keeping sojourn time with a constant mean; the second one had mean sojourn time depend on age through a simple linear model and sensitivity as a constant; the third one considered both screening sensitivity and mean sojourn time as age-dependent, which is essentially a combination of the other

two models.

Through simulation studies, I found that these models have reasonable performance. They captured the structure correctly in simulated data in that when data were generated with certain age dependence, and analyzed with all models, the model indicating that same age dependence has the best fit and provides best estimates with smallest variations for the parameters of interest. Applying the proposed models to two breast cancer screening trials, I found that in both trials, screening sensitivity is positively related to age at screening examinations, and mean sojourn time may also depend on age but less significantly.

One of the goals for better estimation of sensitivity and sojourn time distribution is to provide better insights for public policy makers in making decisions for optimal breast cancer screening strategies. The findings that age is related to sensitivity and sojourn time can be incorporated into decision analytical models, cost-effective analysis and meta-analysis models for breast cancer screening trials.

The models are proposed for, but should not be limited to breast cancer screening. I believe with careful consideration of the natural history of other chronic disease, the models can be applied to other settings.

4.2 Clustered Survival Data Analysis

I have considered two marginal approaches for analyzing clustered survival data when cluster sizes are informative to survival probability. The first approach extends the

work of Hoffman et al. (2001), provides necessary theoretical development for application of the within cluster resampling method to survival data with independent censoring. Asymptotic properties were developed for both regression coefficients and baseline cumulative hazards. The second approach combines the idea of standard marginal model (Lee et al., 1992) and the weighted estimating equation approach (Williamson et al., 2003) by incorporating inverse cluster sizes as weights into the score function. Robust variance estimators were provided for both regression coefficients and baseline cumulative hazards.

Simulation studies confirmed the consistency and asymptotic normality of the estimators for both methods. I also found that when cluster sizes are informative to survival probabilities, estimates using standard marginal models ignoring cluster sizes tend to be substantially biased, whereas the two proposed models provides consistent estimates. On the other hand, when cluster sizes are not informative, the proposed methods perform comparably with the standard marginal models.

Both proposed models are based on Cox proportional hazards model for estimation. When nonproportionality is a concern, it would be of interest to explore other model frameworks, for example, additive hazards model, accelerate failure time model, and linear transformation model. In fact, appeared in a recent issue of Lifetime Data Analysis, Lu (2005) discussed a similar weighted model under the linear transformation model framework. However, the author did not provide simulation or data application results for his model.

I have discussed in Chapter 3 that within cluster resampling estimates may be unstable under heavy censoring due to the fact that resampled datasets only contain partial information of the original data. There is another issue with smaller number of clusters, m, within the original dataset. When there are a small number of clusters, the number of resamplings, B, can achieve an upper limit. Follmann et al. (2003) discussed this problem in a non-survival analysis setting. It will be interesting to explore this problem for better usage of within cluster resampling in multivariate survival data analysis.

Appendix A

Asymptotic Properties of WCR

A.1 Asymptotic Normality of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$

Expanding the score function of the bth resample, $\mathbf{U}_b(\boldsymbol{\beta})$, defined in (3.3), in a Taylor series around $\boldsymbol{\beta}_0$, we get

$$\mathbf{U}_b(\boldsymbol{\beta}_0) = \mathbf{J}_b(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0), \tag{A.1}$$

where $\boldsymbol{\beta}^*$ lies on the segment between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}_b$, which can be rewritten as

$$\sqrt{m}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{m}}\mathbf{U}_b(\boldsymbol{\beta}_0)m\mathbf{J}_b^{-1}(\boldsymbol{\beta}^*).$$

Averaging over $b = 1, \ldots, B$ resamples, we obtain

$$\sqrt{m}\left(\bar{\boldsymbol{\beta}}_{\text{wcr}} - \boldsymbol{\beta}_0\right) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{\sqrt{m}}\mathbf{U}_b(\boldsymbol{\beta}_0)m\mathbf{J}_b^{-1}(\boldsymbol{\beta}^*).$$

Anderson and Gill Anderson and Gill (1982) showed that $m^{-1}\mathbf{J}_b(\boldsymbol{\beta}^*)$ converges to a positive definite matrix in probability. Hence, it remains to show that $B^{-1}m^{-1/2}\sum_{b=1}^{B}\mathbf{U}_b(\boldsymbol{\beta}_0)$

converges in distribution to a normal distribution as $m \to \infty$. Interchanging the order of summation, we rewrite

$$\begin{aligned}
\tfrac{1}{\sqrt{m}B}\sum_{b=1}^{B}\mathbf{U}_b(\boldsymbol{\beta}_0) &= \frac{1}{\sqrt{m}B}\sum_{b=1}^{B}\sum_{i=1}^{m}\int_0^\tau\left\{\mathbf{Z}_i^b(t)-\frac{\mathbf{S}_b^{(1)}(\boldsymbol{\beta}_0,t)}{S_b^{(0)}(\boldsymbol{\beta}_0,t)}\right\}dM_i^b(t)\\
&= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\frac{1}{B}\sum_{b=1}^{B}\int_0^\tau\left\{\mathbf{Z}_i^b(t)-\mathbf{e}(\boldsymbol{\beta}_0,t)\right\}dM_i^b(t)+o_p(1)\\
&= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\mathbf{U}_i(\boldsymbol{\beta}_0)+o_p(1)
\end{aligned}$$

where $\mathbf{U}_i(\boldsymbol{\beta}_0)=\frac{1}{B}\sum_{b=1}^{B}\int_0^\tau\left\{\mathbf{Z}_i^b(t)-\mathbf{e}(\boldsymbol{\beta}_0,t)\right\}dM_i^b(t)$. Since $\mathbf{U}_i(\boldsymbol{\beta}_0)$ are independent with zero mean and finite variance, by the multivariate CLT Sen and Singer (1993), $m^{-1/2}\sum_{b=1}^{B}\mathbf{U}_b(\boldsymbol{\beta}_0)$ is asymptotically normal with mean zero and some positive definite covariance matrix.

Finally, by Slutsky's theorem, $\sqrt{m}(\bar{\boldsymbol{\beta}}_{\text{wcr}}-\boldsymbol{\beta}_0)\to N_p(\mathbf{0},\boldsymbol{\Sigma})$ in distribution as $m\to\infty$, where $\boldsymbol{\Sigma}$ is positive definite, for which a consistent estimator is provided. The proof of its consistency follows.

A.2 Consistency of $\hat{\boldsymbol{\Sigma}}$

The consistency of $\hat{\boldsymbol{\Sigma}}$, defined in (3.8), follows similarly as in Appendix 2 of Hoffman et al. (2001). First, we write

$$\text{Var}(\sqrt{m}\hat{\boldsymbol{\beta}}_b)=E\{\text{Var}(\sqrt{m}\hat{\boldsymbol{\beta}}_b|\text{data})\}+\text{Var}\{E(\sqrt{m}\hat{\boldsymbol{\beta}}_b|\text{data})\},$$

where the expectations on the right-hand side are over the resampling distribution for $\hat{\boldsymbol{\beta}}_b$ given the data. Note that $E(\hat{\boldsymbol{\beta}}_b|\text{data}) = \bar{\boldsymbol{\beta}}_{\text{wcr}}$, thus

$$\text{Var}(\sqrt{m}\bar{\boldsymbol{\beta}}_{\text{wcr}}) = m\text{Var}(\hat{\boldsymbol{\beta}}_b) - mE\{\text{Var}(\hat{\boldsymbol{\beta}}_b|\text{data})\}. \tag{A.2}$$

For the first term on the right-hand side of (A.2), from each resample we obtain a consistent estimator of $\text{Var}(\hat{\boldsymbol{\beta}}_b)$, denoted as $\hat{\boldsymbol{\Sigma}}_b$, given in equation (3.7). By averaging over the B resamples, the resulting estimator is also consistent. The second term in (A.2) can be estimated as the covariance matrix among the B resample-based estimate $\hat{\boldsymbol{\beta}}_b$, that is

$$\hat{\boldsymbol{\Omega}} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})'.$$

Since $E\{\text{Var}(\sqrt{m}\hat{\boldsymbol{\beta}}_b|\text{data})\} = mE(\hat{\boldsymbol{\Omega}})$, to show the consistency of $\hat{\boldsymbol{\Sigma}}$, we need to show that $m\hat{\boldsymbol{\Omega}} - mE(\hat{\boldsymbol{\Omega}}) \to 0$ in probability as $m \to \infty$. To prove this, we can show that $m^2\text{Var}(\hat{\boldsymbol{\Omega}}) \to 0$, Since the B resamples are identically distributed although not independent, and all the pairwise covariance between resamples are also identical, we have

$$\begin{aligned} m^2\text{Var}(\hat{\boldsymbol{\Omega}}) &= \frac{B}{(B-1)^2}\text{Var}\{\sqrt{m}(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})\sqrt{m}(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})'\} \\ &\quad + \frac{m^2 B}{B-1}\text{Cov}\{(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})', (\hat{\boldsymbol{\beta}}_{b^*} - \bar{\boldsymbol{\beta}}_{\text{wcr}})(\hat{\boldsymbol{\beta}}_{b^*} - \bar{\boldsymbol{\beta}}_{\text{wcr}})'\}, \end{aligned} \tag{A.3}$$

The first term of (A.3) is approximately zero for large B, the second term of (A.3) goes to zero as $m \to \infty$ by a similar argument as in Hoffman et al. (2001). This completes the proof for consistency of $\hat{\boldsymbol{\Sigma}}$.

A.3 Weak Convergence Properties of $\bar{\Lambda}_0(t)$

By adding and subtracting a common term of $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$ evaluated at the true parameter values, denoted as $\bar{\Lambda}_0(t, \boldsymbol{\beta}_0)$, we rewrite

$$\begin{aligned} W(t) &= \sqrt{m}\left\{\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) - \Lambda_0(t)\right\} \\ &= \sqrt{m}\left\{\bar{\Lambda}_0(t, \boldsymbol{\beta}_0) - \Lambda_0(t)\right\} + \sqrt{m}\left\{\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) - \bar{\Lambda}_0(t, \boldsymbol{\beta}_0)\right\}. \end{aligned} \tag{A.4}$$

The first term on the right-hand side of (A.4) can be written as

$$\begin{aligned} \sqrt{m}\left\{\bar{\Lambda}_0(t, \boldsymbol{\beta}_0) - \Lambda_0(t)\right\} &= \sqrt{m}\left[\frac{1}{mB}\sum_{b=1}^{B}\sum_{i=1}^{m}\int_0^t \frac{dN_i^b(u)}{S_b^{(0)}(\boldsymbol{\beta}_0, u)} - \int_0^t d\Lambda_0(u)\right] \\ &= \sqrt{m}\left[\frac{1}{mB}\sum_{b=1}^{B}\sum_{i=1}^{m}\int_0^t \left\{\frac{dN_i^b(u)}{S_b^{(0)}(\boldsymbol{\beta}_0, u)} - d\Lambda_0(u)\right\}\right] \\ &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\frac{1}{B}\sum_{b=1}^{B}\int_0^t \frac{dM_i^b(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} + o_p(1). \end{aligned}$$

The second term of (A.4) can be written in the following form

$$\sqrt{m}\left\{\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) - \bar{\Lambda}_0(t, \boldsymbol{\beta}_0)\right\} = \sqrt{m}\frac{1}{B}\sum_{b=1}^{B}\left\{\hat{\Lambda}_0^b(t; \hat{\boldsymbol{\beta}}_b) - \hat{\Lambda}_0^b(t; \boldsymbol{\beta}_0)\right\}.$$

By the Taylor series expansion,

$$\begin{aligned} \hat{\Lambda}_0^b(t; \hat{\boldsymbol{\beta}}_b) - \hat{\Lambda}_0^b(t; \boldsymbol{\beta}_0) &= -(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0)' \int_0^t \frac{\mathbf{S}_b^{(1)}(\boldsymbol{\beta}^*, u)}{S_b^{(0)}(\boldsymbol{\beta}^*, u)} d\hat{\Lambda}_0^b(u; \boldsymbol{\beta}^*) \\ &= -(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0)' \int_0^t \mathbf{e}(\boldsymbol{\beta}_0, u) d\Lambda_0(u) + o_p(1). \end{aligned}$$

Therefore,

$$\sqrt{m}\left\{\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) - \bar{\Lambda}_0(t, \boldsymbol{\beta}_0)\right\} = -\sqrt{m}(\bar{\boldsymbol{\beta}}_{\text{wcr}} - \boldsymbol{\beta}_0)' \int_0^t \mathbf{e}(\boldsymbol{\beta}_0, u) d\Lambda_0(u) + o_p(1).$$

It then follows that

$$W(t) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left\{ \frac{1}{B} \sum_{b=1}^{B} \int_0^t \frac{dM_i^b(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} - m(\bar{\boldsymbol{\beta}}_{\text{wcr}} - \boldsymbol{\beta}_0)' \int_0^t \mathbf{e}(\boldsymbol{\beta}_0, u) d\Lambda_0(u) \right\} + o_p(1) \tag{A.5}$$

and since this is a sum of m independent terms with expected mean zero and finite variance, by the multivariate CLT Sen and Singer (1993), the finite-dimensional distribution of $W(t)$ converges weakly to a zero-mean Gaussian process, denoted as $\mathcal{W}(t)$, with finite covariance function.

We then need to show the tightness of $W(t)$. Define the two terms of (A.5) as $Q_1(t) = B^{-1} \sum_{b=1}^{B} m^{-1/2} \sum_{i=1}^{m} \int_0^t s^{(0)}(\boldsymbol{\beta}_0, u)^{-1} dM_i^b(u)$ and $Q_2(t) = -\sqrt{m}(\bar{\boldsymbol{\beta}}_{\text{wcr}} - \boldsymbol{\beta}_0)' \int_0^t \mathbf{e}(\boldsymbol{\beta}_0, u) d\Lambda_0(u)$. The tightness of $W(t)$ follows the tightness of $Q_1(t)$ and $Q_2(t)$. Under the regularity conditions, $m^{-1/2} \sum_{i=1}^{m} \int_0^t s^{(0)}(\boldsymbol{\beta}_0, u)^{-1} dM_i^b(u)$ is a square-integrable martingale with respect to the joint filtration

$$\mathcal{F}^b(t) = \vee_{i=1}^{m} \sigma \left\{ N_i^b(u), Y_i^b(u), \mathbf{Z}_i^b(u) : 0 \leq u \leq t \right\}.$$

Thus, $Q_1(t)$ is tight by Theorem 10.2 of Pollard (1990). Under the regularity conditions, the tightness of $Q_2(t)$ follows from Theorem 1. This completes the proof for Theorem 3.

Bibliography

P. K. Anderson and R. D. Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10:1100–1120, 1982.

E. Benhin, N. K. Rao, and A. J. Scott. Mean estimating equation approach to analysing cluster-correlated data with nonignorable cluster sizes. *Biometrika*, 92: 435–450, 2005.

K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference a Practical Information-Theoretic Approach.* Springer, second edition, 2002.

R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, 16:1190–1208, 1995.

J. Cai and R. L. Prentice. Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis*, 3:197–213, 1997.

T. Cai, L. J. Wei, and M. Wilcox. Semiparametric regression analysis for clustered failure time data. *Biometrika*, 87:867–878, 2000.

J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71:340–344, 1976.

L. Chen and L. J. Wei. Analysis of multivariate survival times with non-proportional hazards models. In D. Y. Lin and T. R. Fleming, editors, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, New York, 1997. Springer-Verlag.

L. X. Clegg, J. Cai, and P. K. Sen. A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics*, 55:805–812, 1999.

X. J. Cong, Y. Shen, and A. B. Miller. Estimation of age-specific sensitivity and sojourn time in breast cancer screening studies. *Statistics in Medicine*, 24:3123–3138, 2005.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

D. R. Cox. Partial likelihood. *Biometrika*, 62:169–276, 1975.

N. E. Day and S. D. Walter. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics*, 40:1–14, 1984.

R. Etzioni and Y. Shen. Estimating asymptomatic duration in cancer: the AIDS connection. *Statistics in Medicine*, 16:627–644, 1997.

Cancer Figures and Facts. *http://www.cancer.org/downloads/STT/CAFF2003PWSecured.pdf*, November 2003.

T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis.* John Wileys & Sons, Inc., New York, 1990.

D. Follmann, M. Proschan, and E. Leifer. Multiple outputation: inference for complex clustered data by averaging analyses from independent data. *Biometrics*, 59:420–429, 2003.

E. B. Hoffman, P. K. Sen, and C. R. Weinberg. Within-cluster resampling. *Biometrika*, 88:1121–1134, 2001.

P. Hougaard. A class of multivariate failure time distribution. *Biometrika*, 73:671–678, 1986.

I. Jatoi and A. B. Miller. Why is breast-cancer mortality declining? *Lancet Oncology*, 4:251–254, 2003.

K. Kerlikowske, D. Crady, J. Barclay, E. A. Sickles, and V. Ernster. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *The Journal of the American Medical Association*, 276:33–38, 1996a.

K. Kerlikowske, D. Grady, J. Barclay, E. A. Sickles, and V. Ernster. Likelihood ratios for modern screening mammography: risk of breast cancer based on age and mammographic interpretation. *The Journal of the American Medical Association*, 276:39–43, 1996b.

E. W. Lee, L. J. Wei, and D. A. Amato. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In J. P. Klein and P. K. Goel, editors, *Survival Analysis: State of the Arts*, Dordrecht, 1992. Kluwer Academic Publishers.

E. W. Lee, L. J. Wei, and Z. Ying. Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association*, 88:557–565, 1993.

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

J. S. Lin and L. J. Wei. Linear regression analysis for multivariate failure time observations. *Journal of American Statistical Association*, 87:1091–1097, 1992.

W. Lu. Marginal regression of multivariate event times based on linear transformation models. *Lifetime Data Analysis*, 11:389–404, 2005.

M. K. McGuire and M. E. Nunn. Prognosis versus actual outcome III: The effects of clinical parameters in accurately predicting tooth survival. *Journal of Periodontology*, 67:666–674, 1996.

A. B. Miller, C. J. Baines, T. To, and C. Wall. Canadian national breast screening study: 1. breast cancer detection and death rates among women aged 40 to 49 years. *Canadian Medical Association Journal*, 147:1459–1476, 1992a.

A. B. Miller, C. J. Baines, T. To, and C. Wall. Canadian national breast screening study: 2. breast cancer detection and death rates among women aged 50 to 59 years. *Canadian Medical Association Journal*, 147:1477–88, 1992b.

R. Mukerjee. Comparison between the conditional likelihood ratio and the usual likelihood ratio tests. *Journal of Royal Statistical Society, Series B*, 54:189–194, 1992.

J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data.* Birkhäuser, Boston, 2006. In progress, Chapter 1 online at cademic2.american.edu/~jpnolan.

G. Parmigiani. *Modeling in medical decision making: a Bayesian approach.* John Wiley, New York, 2002.

P. G. M. Peer, A. L. M. Verbeek, H. Straatman, J. H. C. L. Hendriks, and R. Holland. Age-specific sensitivity of mammographic screening for breast cancer. *Breast Cancer Research and Treatment*, 38:153–160, 1996.

D. Pollard. *Empirical Processes: Theory and Applications.* Institute of Mathematical Statistics, Hayward, CA, 1990.

P. K. Sen and J. M. Singer. *Large Sample Methods in Statistics.* Chapman & Hall, New York, 1993.

R. J. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley, New York, 1980.

Y. Shen, D. Wu, and M. Zelen. Testing the independence of two diagnostic tests. *Biometrics*, 57:1009–1017, 2001.

Y. Shen and M. Zelen. Parametric estimation procedures for screening programmes: stable and nonstable disease models for multimodality case finding. *Biometrika*, 86:503–515, 1999.

Y. Shen and M. Zelen. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *Journal of Clinical Oncology*, 19:3490–3499, 2001.

C. F. Spiekerman and D. Y. Lin. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93:1164–1175, 1998.

H. Straatman, P. G. M. Peer, and A. L. M. Verbeek. Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group. *Biometrics*, 53:217–229, 1997.

L. Tabar, B. Vitak, and H. H. Chen. The Swedish Two-County trial twenty years later.

updated mortality results and new insights from long-term follow-up. *Radiologic Clinics of North America*, 38:625–651, 2000.

S. D. Walter and N. E. Day. Estimation of the duration of a preclinical disease state using screening data. *American Journal of Epidemiology*, 118:856–886, 1983.

L. J. Wei, D. Y. Lin, and L. Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84:1065–1073, 1989.

J. M. Williamson, S. Datta, and G. A. Satten. Marginal analyses of clustered data when cluster size is informative. *Biometrics*, 59:36–42, 2003.

G. Yin and J. Cai. Additive hazards model with multivariate failure time data. *Biometrika*, 91:801–818, 2004.

M. Zelen. Optimal scheduling of examinations for the early detection of disease. *Biometrika*, 80:279–293, 1993.

M. Zelen and M. Feinleib. On the theory of screening for chronic diseases. *Biometrika*, 56:601–614, 1969.