Robust Methods for Interval-Censored Life History Data

by

David Tolusso

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

Waterloo, Ontario, Canada, 2008

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

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

Abstract

Interval censoring arises frequently in life history data, as individuals are often only observed at a sequence of assessment times. This leads to a situation where we do not know when an event of interest occurs, only that it occurred somewhere between two assessment times. Here, the focus will be on methods of estimation for recurrent event data, current status data, and multistate data, subject to interval censoring.

With recurrent event data, the focus is often on estimating the rate and mean functions. Nonparametric estimates are readily available, but are not smooth. Methods based on local likelihood and the assumption of a Poisson process are developed to obtain smooth estimates of the rate and mean functions without specifying a parametric form. Covariates and extra-Poisson variation are accommodated by using a pseudo-profile local likelihood. The methods are assessed by simulations and applied to a number of datasets, including data from a psoriatic arthritis clinic.

Current status data is an extreme form of interval censoring that occurs when each individual is observed at only one assessment time. If current status data arise in clusters, this must be taken into account in order to obtain valid conclusions. Copulas offer a convenient framework for modelling the association separately from the margins. Estimating equations are developed for estimating marginal parameters as well as association parameters. Efficiency and robustness to the choice of copula are examined for first and second order estimating equations. The methods are applied to data from an orthopedic surgery study as well as data on joint damage in psoriatic arthritis.

Multistate models can be used to characterize the progression of a disease as individuals move through different states. Considerable attention is given to a three-state model to characterize the development of a back condition known as spondylitis in psoriatic arthritis, along with the associated risk of mortality. Robust estimates of the state occupancy probabilities are derived based on a difference in distribution functions of the entry times. A five-state model which differentiates between left-side and right-side spondylitis is also considered, which allows us to characterize what effect spondylitis on one side of the body has on the development of spondylitis on the other side. Covariate effects are considered through multiplicative time homogeneous Markov models. The robust state occupancy probabilities are also applied to data on CMV infection in patients with HIV.

Acknowledgements

I wish to thank a number of people who have helped me either directly or indirectly create this thesis.

First and foremost, I must express my sincere gratitude to my supervisor, Dr. Richard Cook. He provided me with interesting and challenging problems as well as valuable insight and advice whenever difficulties were encountered. This thesis would not have been possible without him.

I also would like to thank the members of my committee for their time reading my thesis. Their helpful comments and suggestions greatly improved this work. A big thank you goes to Dr. Joel Dubin and Dr. Jerry Lawless from the Department of Statistics and Actuarial Science. I also wish to thank the External Examiner, Dr. John Braun, from the Department of Statistical and Actuarial Sciences at the University of Western Ontario. My appreciation also goes to the Internal-External Examiner, Dr. Paul Stolee (Department of Health Studies and Gerontology) for reading my thesis even though statistics is not his area of expertise.

I wish to acknowledge the faculty, students and staff of the Department of Statistics and Actuarial Science. In particular I wish to mention the graduate secretary, Mary Lou Dufton, as well as the students with whom I had the pleasure of sharing office space over the years, Adam Metzler, So-Yeun Kim, Theo Koulis, Joonghee Huh and Changkee Lee.

My thanks also go to Dr. Dafna Gladman and Dr. Vinod Chandran for the use of the data from the University of Toronto Psoriatic Arthritis Clinic. I also thank Dr. Ted Warkentin from McMaster University regarding the data from the orthopedic studies. I would also like to thank Ker-Ai Lee for her help with the data and computing issues.

I am grateful for the financial support I received from the Natural Sciences and Engineering Research Council of Canada, the University of Waterloo, the Faculty of Mathematics, and the Department of Statistics and Actuarial Science, and Dr. Richard Cook.

Contents

List of Tables **x**

List of Figures **xiii**

1 Introduction **1**

1.1 Overview 1

1.2 Analysis of Failure Time Data 2

1.2.1 Analysis of Right-Censored Failure Time Data 2

1.2.2 Analysis of Interval-Censored Failure Time Data 10

1.2.3 Analysis of Truncated Failure Time Data 20

1.3 Analyses Based on Multistate Models 23

1.3.1 Intensity Functions for Multistate Models 23

1.3.2 Analysis of Right-Censored Multistate Data 27

1.3.3 Analysis of Interval-Censored Multistate Data 27

1.4 Analysis of Recurrent Events 31

1.4.1 Analysis of Right-Censored Recurrent Events 31

1.4.2 Analysis of Interval-Censored Recurrent Events 32

1.5 Outline of Thesis 35

1.5.1 Chapter 2: Local Likelihood methods for Interval-Censored Recurrent Event Data 35

1.5.2 Chapter 3: Estimating Equations for Clustered Current Status Data 36

1.5.3 Chapter 4: Multistate Models With Interval Censoring . . . 37

2 Local Likelihood for Interval-Censored Recurrent Event Data 39

2.1 Introduction and Overview 39

2.2 Estimation of the Mean Function for Right-Censored Count Data . 40

2.2.1 The One Sample Problem 40

2.2.2 Regression Models 41

2.3 Estimation of the Mean Function for Interval-Censored Recurrent Event Data 44

2.3.1 The One Sample Problem 44

2.3.2 Regression Models 45

2.4 Simulation Studies 47

2.5 Applications 55

2.5.1 Data on Incidence of Superficial Bladder Tumors 55

2.5.2 Rat Tumorgenicity Data 61

2.5.3 Nuclear Plant Reliability Study 61

2.5.4 Counts of Damaged Joints in Psoriatic Arthritis 62

2.6 Summary 66

3 Estimating Equations for Clustered Current Status Data 68

3.1 Introduction and Overview 68

3.2 Notation 69

3.3 Methods of Estimation and Inference 71

3.3.1 Generalized Estimating Equations 71

3.3.2 Relative Efficiency of GEE1 vs. GEE2 75

3.3.3 Robustness of GEE1 and GEE2 to Copula Misspecification . 77

3.3.4 Computational Notes for the Clayton Copula 80

3.4 Simulation Studies 81

3.4.1 Simulating Clustered Current Status Data via Copulas . . . 81

3.4.2 Assessing the Empirical Performance of Estimators 82

3.5 Applications 83

3.5.1 Analysis of Seroconversion in Orthopedic Surgery 83
3.5.2 Joint Damage at Clinic Entry in Psoriatic Arthritis 89
3.6 Summary 89

4 Multistate Models With Interval Censored Data: Applications 93
4.1 Introduction and Overview 93
4.2 "Pepe" Estimation of Prevalence Functions 94
4.2.1 Estimation under Piecewise Constant Models with Order Restrictions 95
4.3 Applications 98
4.3.1 Prevalence and Regression Analysis for Spondylitis in PsA . 98
4.3.2 Spondylitis Defined by Unilateral Involvement 110
4.3.3 A 4-State Model for Bivariate Interval-Censored Data 115
4.4 Summary 121

5 Future Work 127
5.1 Summary 127
5.2 Methods for Recurrent Events and Death 128
5.2.1 Methods Based on Markov Models 129
5.2.2 Pepe Estimation 132
5.3 Clustered Current Status Data 135
5.3.1 Second Order GEE for Regression on Association Parameter 135
5.3.2 Copula Goodness-of-Fit 135
5.3.3 Random Effects Approach 136

APPENDICES 140

A Full Simulation Results From Chapter 3 141

References 156

List of Tables

2.1 Integrated mean squared error for the local likelihood methods based on Poisson models, Pepe approach and piecewise constant models under consideration based on 1000 simulated datasets of 500 subjects. 49

2.2 Integrated absolute bias for the local likelihood methods based on Poisson models, Pepe approach and piecewise constant models under consideration based on 1000 simulated datasets of 500 subjects. . . 50

2.3 Empirical bias and coverage probability of local likelihood and piecewise constant estimates of the regression coefficient under Poisson and Mixed Poisson models based on 1000 simulated datasets of 500 subjects. 54

2.4 Kolmogorov-Smirnov test statistics (KS) and p-values for assessing the validity of the χ^2_1 approximation to the profile likelihood ratio statistic based on 1000 simulated datasets of 500 subjects with mean function $(\theta t)^\gamma$. 55

2.5 Local likelihood and Cox estimates and 95% confidence intervals for the effect of treatment in the bladder cancer study. 58

2.6 Local likelihood and Cox estimates and 95% confidence intervals for the effect of treatment in the rat tumorgenicity study of Gail et al. (1980). 62

2.7 Local likelihood estimates of the regression parameters in the psoriatic arthritis study. 66

3.1 Asymptotic relative efficiencies of estimators under working independence and GEE1 relative to GEE2. 78

3.2 Asymptotic bias in parameter estimators from assuming a Clayton copula when the true copula is a Gumbel copula under GEE1 and GEE2 analyses. 79

3.3 Empirical bias ($\times 10^2$) and empirical (ESE) and average standard errors (SE) for $\widehat{\beta}$ ($\rho = 0.4$ and $\beta = \log 0.80$). 84

3.4 Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution. 85

3.5 Estimates and 95% confidence intervals for proportion testing positive 10 days after surgery in the orthopedic surgery study. 86

3.6 Regression coefficient estimates for 3 piece and Weibull models in the orthopedic surgery study. 88

3.7 Estimates and 95% confidence intervals for proportion of damaged joints 5 years after diagnosis based on the University of Toronto Psoriatic Arthritis Clinic data. 91

3.8 Estimates of regression coefficients obtained from fitting the five piece piecewise constant and Weibull regression models to data from the University of Toronto Psoriatic Arthritis Clinic. 92

4.1 Frequency distributions of the continuous covariates used to predict spondylitis in psoriatic arthritis. 102

4.2 Summaries of the binary covariates used to predict spondylitis in psoriatic arthritis. 103

4.3 Univariate effects on transitions using spondylitis Definition 3. . . . 107

4.4 Multivariate effects on transitions using spondylitis Definition 3. . . 108

4.5 Univariate effects on transitions using spondylitis Definition 5. . . . 111

4.6 Multivariate effects on transitions using spondylitis Definition 5. . . 112

4.7 Univariate effects on transitions in the model of Figure 4.5. 116

4.8 Multivariate effects on transitions in the model of Figure 4.5. 117

4.9 Estimates of model parameters from fitting the Markov model in Figure 4.5 to the psoriatic arthritis data. 118

A.1 Asymptotic relative efficiencies compared to GEE2 ($n = 2$) 142

A.2 Asymptotic relative efficiencies compared to GEE2 ($n = 5$) 143

A.3 Asymptotic relative efficiencies compared to GEE2 ($n = 10$) 144

A.4 Bias due to misspecification of the copula ($n = 2$). 145

A.5 Bias due to misspecification of the copula ($n = 5$). 146

A.6 Bias due to misspecification of the copula ($n = 10$). 147

A.7 Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient. 148

A.8 Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient (continued). 149

A.9 Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient (continued). 150

A.10 Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient (continued). 151

A.11 Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution. 152

A.12 Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution (continued). 153

A.13 Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution (continued). 154

A.14 Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution (continued). 155

List of Figures

1.1 Six possibilities for the censoring interval relative to the cut-points in a piecewise model. 13

1.2 Example of how the intervals of support for the NPMLE are determined in the case of interval-censored failure time data. 15

1.3 2-State survival model. 25

1.4 Progressive 3-State model. 25

1.5 Illness-Death model. 25

1.6 Competing Risks model. 26

1.7 Recurrent events model. 31

2.1 Recurrent events model. 39

2.2 Empirical bias of the Poisson local likelihood and Pepe local likelihood estimates as well as piecewise constant estimates (3 piece–PW3, 6 piece–PW6) of the mean function for interval-censored recurrent event data. 51

2.3 Empirical bias of the Poisson local likelihood estimate as well as piecewise constant estimates (3 piece–PW3, 6 piece–PW6) of the rate function for interval-censored recurrent event data 52

2.4 QQ-plots comparing the simulated quantiles of the profile likelihood ratio statistic with the theoretical quantiles of the χ^2_1 distribution based on 1000 simulated datasets of 500 subjects with Poisson mean function $(\theta t)^\gamma$. 56

2.5 QQ-plots comparing the simulated quantiles of the profile likelihood ratio statistic with the theoretical quantiles of the χ^2_1 distribution based on 1000 simulated datasets of 500 subjects with mixed-Poisson mean function $(\theta t)^\gamma$ and $\phi = 0.25$. 57

2.6 Timelines of 24 thiotepa patients from the bladder tumor study of Byar (1980). 59

2.7 Plot of the total number of tumors detected against time on study in the bladder tumor data of Byar (1980). 60

2.8 Estimates of the mean number of losses of feedwater flow obtained nonparametrically and using local likelihood. 63

2.9 Timeline diagram of joint damage for a sample of seven patients from the University of Toronto psoriatic arthritis clinic. 64

2.10 Poisson local likelihood and piecewise constant estimates of the mean function with three (PW3) and six (PW6) pieces based on the joint damage from the psoriatic arthritis data. 65

3.1 Correlation between Y_1 and Y_2 as a function of C_1 and C_2. 72

3.2 Proportion of subjects testing positive under a three piece piecewise constant and Weibull model using GEE and GEE2 in the orthopedic surgery study. 87

3.3 Proportion of damaged joints under a 5 piece and Weibull model. . 90

4.1 Three state diagram for onset of spondylitis in patients with psoriatic arthritis. 98

4.2 Timeline diagrams for sample of patients indicating the states occupied and censoring intervals for onset of spondylitis. 100

4.3 Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and time homogeneous Markov estimates of remaining alive and having spondylitis over 15 years from clinic entry, based on Definition 3. 105

4.4 Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and time homogeneous Markov estimates of remaining alive and having spondylitis over 15 years from clinic entry, based on Definition 5. 109

4.5 Five state diagram for onset of left and right side spondylitis in patients with psoriatic arthritis. 110

4.6 Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of state occupancy probabilities in the side-specific analysis of spondylitis in psoriatic arthritis. 114

4.7 Four state diagram for bivariate interval-censored failure time data. 118

4.8 Censoring intervals for the times to shedding in blood and urine for a sample of patients in the CMV study. 120

4.9 Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of having shedding in the blood only over 18 months from study entry in the CMV data. 122

4.10 Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of having shedding in the urine only over 18 months from study entry in the CMV data. 123

4.11 Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of having both blood and urine shedding over 18 months from study entry in the CMV data. 124

4.12 Bivariate nonparametric, univariate piecewise constant, univariate local likelihood and multistate Markov estimates of the distributions of time to shedding in the blood and shedding in the urine. 125

5.1 Kaplan-Meier estimates of survival for data from the study of breast cancer patients with bone metastases (Hortobagyi et al. 1996). . . . 129

5.2 Kaplan-Meier estimates of time to first detected lesion or death for data from the study of breast cancer patients with bone metastases (Hortobagyi et al. 1996). 130

5.3 Sample of patient timelines from the study of Hortobagyi et al. (1996).131

5.4 Multistate model for recurrent events with a terminating event. . . 132

5.5 Alternative model for recurrent events with a terminating event appropriate for Pepe estimation of the mean function. 133

5.6 Competing risks model. 133

Chapter 1

Introduction

1.1 Overview

In many chronic disease processes, it is natural to model the disease course with multistate stochastic models. An example of this is the simple case where a healthy individual may develop a disease and subsequently die. In this case, the states can be defined as "healthy", "diseased" and "dead" and transitions occur between these states as the process evolves over time. A complication in studies of this setting is that it is often not possible to observe individuals continuously over the period of interest. Instead, the status of individuals is known only at a sequence of assessment times. At these times it can be determined what state an individual is in, and if a transition is known to have occurred, an interval over which the transition occurred may be known. The transition times are said to be interval-censored. Models for counting processes can also be formulated based on a multistate model. Here the state space is the set of non-negative integers representing the cumulative number of recurrent events an individual has experienced. These models are progressive in the sense that transitions are only possible in one direction.

The goal of this thesis is to develop statistical methods for settings where the process of interest can be characterized through a multistate or recurrent event model but the transition or event times are subject to interval censoring. In particular, focus will be on estimation of prevalence functions and transition intensities for multistate models and transition rates and mean functions for recurrent event models. We restrict attention to progressive models and give special attention to several common multistate models.

The remainder of this chapter is organized as follows. In Sections 1.2, 1.3 and 1.4

we review methods for the analysis of failure time, multistate, and recurrent event data respectively and consider both the analysis of right-censored and interval-censored data. Methods to be used in later chapters are introduced here, but readers familiar with this material can proceed to Section 1.5 where the topics of particular interest are discussed briefly. Details on the specific topics are given in Chapters 2 to 4 and plans for future research are given in Chapter 5.

1.2 Analysis of Failure Time Data

In many settings interest lies in the time until a certain event occurs. Often, this is death and so the time under study is referred to as a lifetime or failure time. Let $T > 0$ be a random variable representing the time under study, and t its realized value. Usually, it is assumed that T has a continuous distribution with density $f(t)$ and cdf $F(t) = P(T \leq t)$. The *survivor function* is

$$S(t) = P(T > t) = \int_t^\infty f(u)\, du = 1 - F(t)$$

and the *hazard function* is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

The hazard function expresses the instantaneous conditional probability of failing at t given that an individual survived up to t. A related function of interest is the *cumulative hazard function*,

$$H(t) = \int_0^t h(u)\, du$$

and it can be shown that $S(t) = \exp(-H(t))$ in the continuous case and more generally using the product integral (see Andersen et al. 1993, Chap. 2),

$$S(t) = \prod_{(0,t)} (1 - dH(u)).$$

1.2.1 Analysis of Right-Censored Failure Time Data

In many situations failure times are not known exactly, but are only known to have fallen in a particular interval. In such cases they are said to be *censored*. Suppose in a medical study a subject is followed from some time origin and does

not experience the event under study. In this case all that can be determined is that the event occurred at some point after the end of the period of observation. This is an example of *right censoring*, where the censoring time corresponds to the end of the study. More generally, a failure time would be called right-censored when it is known to occur after some given time, but the time itself is unknown.

A common way of describing right-censored data is to define random variables T_i and δ_i for the ith subject, $i = 1, \ldots, n$ where $\delta_i = I(\text{subject } i\text{'s lifetime is known})$ and T_i is the lifetime if $\delta_i = 1$ or the censoring time if $\delta_i = 0$ for subject i. The likelihood function can be constructed based on these random variables for simple censoring schemes, however a much more general approach will be adopted here that encompasses many censoring schemes. Let $N_i(t)$ be a counting process that counts failures for subject i, $i = 1, \ldots, n$. In this case, a subject can only fail once, say at time t_i, so $N_i(t) = 0$ for $t < t_i$ and $N_i(t) = 1$ for $t \geq t_i$. Similarly, a counting process $C_i(t)$ can be defined that counts when individual i is censored. A third process, $Y_i(t)$, is the at-risk process which takes the value 1 if individual i is at risk of failure at time t and 0 otherwise. Also define $dN_i(t) = I(\text{individual } i \text{ failed in } [t, t + dt))$ and $dC_i(t) = I(\text{individual } i \text{ censored in } [t, t + dt))$. Define $D(t) = \{i : dN_i(t) = 1\}$ and $C(t) = \{i : dC_i(t) = 1\}$. The likelihood can be constructed as a product of conditional probabilities of the form $P(D(t), C(t)|\mathcal{H}(t))$ where $\mathcal{H}(t) = \{N_i(u), 0 \leq u < t; Y_i(u), 0 \leq u \leq t, i = 1, \ldots, n\}$ is the history of the study to time t. Using the product integral, the likelihood is

$$L = \prod_{(0,\infty)} P(D(t), C(t)|\mathcal{H}(t))$$

which can be broken down to

$$L = \prod_{(0,\infty)} P(D(t)|\mathcal{H}(t)) \prod_{(0,\infty)} P(C(t)|D(t), \mathcal{H}(t)).$$

In order to further simplify the likelihood, assumptions must be made about the censoring mechanism. Standard assumptions (Lawless 2003; Kalbfleisch and Prentice 2002) are that for individuals at risk at time t, the failure mechanisms act independently and that

$$P(dN_i(t) = 1|\mathcal{H}(t)) = Y_i(t)h(t).$$

This leads to

$$P(D(t)|\mathcal{H}(t)) = \prod_{i=1}^{n} h(t)^{dN_i(t)}(1 - h(t))^{Y_i(t)(1-dN_i(t))}$$

and if $P(C(t)|D(t), \mathcal{H}(t))$ does not depend on the parameters of interest, these terms can be dropped from the likelihood leaving

$$L = \prod_{i=1}^{n} \prod_{(0,\infty)} h(t)^{dN_i(t)} (1 - h(t))^{Y_i(t)(1-dN_i(t))}$$

which simplifies to

$$L = \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

The assumptions above define the class of censoring mechanisms known as *independent censoring* since it amounts to a conditional independence between the failure mechanism and the censoring mechanism. This means that a subject who has not failed and not been censored by time t has the same probability of failure at time t had there been no censoring. Special cases of independent censoring include both type 1 and type 2 censoring as well as independent random censoring.

When $P(C(t)|D(t), \mathcal{H}(t))$ does not depend on the parameters of interest it is termed *noninformative censoring*. In cases where the terms $P(C(t)|D(t), \mathcal{H}(t))$ do depend on the parameters of interest (*i.e.* we have informative censoring) they can still be dropped and L can be viewed as a partial likelihood. Inferences remain valid, although possibly at a loss of efficiency.

Parametric Models

When the survival function takes a parametric form, $S(t; \theta)$, dependent on a vector of parameters θ, the likelihood is as above,

$$L(\theta) = \prod_{t=1}^{n} f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}.$$

Standard likelihood based inference can be used to estimate θ and parametric regression models may also be considered (see Lawless 2003, Chaps. 4, 5 & 6).

Piecewise-Constant Models

A special case of parametric models are those where the hazard function is piecewise constant. This means there is a partition $0 = a_0 < a_1 < a_2 < \ldots < a_m$ where the

hazard in $(a_{j-1}, a_j]$ is a constant λ_j, $j = 1, \ldots, m$. In this case,

$$\begin{aligned} H(t;\lambda) &= \sum_{j=1}^{m} \lambda_j w_j(t) \\ S(t;\lambda) &= \exp(-H(t)) \\ f(t;\lambda) &= \prod_{j=1}^{m} \lambda_j^{I(a_{j-1}<t\le a_j)} \exp(-H(t)) \end{aligned}$$

where $w_j(t) = \int_{a_{j-1}}^{a_j} I(u \le t)\, du$. The likelihood becomes

$$L(\lambda) = \prod_{i=1}^{n} \prod_{j=1}^{m} \lambda_j^{I(a_{j-1}<t_i\le a_j)\delta_i} \exp\left(- \sum_{j=1}^{m} \lambda_j w_j(t_i) \right).$$

This likelihood is straightforward to maximize, since the log-likelihood can be written as

$$\ell(\lambda) = \sum_{j=1}^{m} \left(d_j \log \lambda_j - v_j \lambda_j \right)$$

where $d_j = \sum_{i=1}^{n} I(a_{j-1} < t_i \le a_j)\delta_i$ and $v_j = \sum_{i=1}^{n} w_j(t_i)$. It follows that $\widehat{\lambda}_j = d_j/v_j$, $j = 1, \ldots, m$.

Nonparametric Estimation

If a nonparametric approach is adopted, the derivation of the nonparametric MLE of the survivor function is as follows (see Kalbfleisch and Prentice 2002, Chap. 1). Let $t_1, \ldots, t_k$ be the ordered observed survival times. Let r_j be the number of individuals who have a failure time greater or equal to t_j and let d_j be the number of individuals with a failure time equal to t_j. Let the number of individuals with censoring times lying in $[t_j, t_{j+1})$, $j = 0, \ldots, k$ be c_j, where $t_0 = 0$ and $t_{k+1} = \infty$. Let the censoring times be denoted by t_{jr}, $r = 1, \ldots, c_j$, $j = 0, \ldots, k$. Then an observed time contributes $S(t_j) - S(t_j^+)$ and a censored time contributes $S(t_{jr}^+)$ to the likelihood, hence the likelihood becomes

$$L = \prod_{j=0}^{k} \left[\left(S(t_j) - S(t_j^+) \right)^{d_j} \prod_{r=1}^{c_j} S(t_{jr}^+) \right].$$

From this, we can see the nonparametric MLE, $\widehat{S}(t)$, must be discontinuous at each t_j. Also, since $S(t)$ is a non-increasing function, the $S(t_{jr}^+)$ terms are

maximized by setting $S(t_{jr}^+) = S(t_j^+)$. In other words, $\widehat{S}(t)$ has jumps at $t_1, \ldots, t_k$ and is constant everywhere else. This means that we can write

$$\widehat{S}(t_j) = \prod_{r<j}(1 - \widehat{h}_r)$$
$$\widehat{S}(t_j^+) = \prod_{r\leq j}(1 - \widehat{h}_r)$$

and the log-likelihood becomes

$$\begin{aligned}\ell &= \sum_{j=1}^{k}\left[d_j \log(\widehat{h}_j) + d_j \sum_{r<j}\log(1-\widehat{h}_r) + c_j \sum_{r\leq j}\log(1-\widehat{h}_r)\right] \\ &= \sum_{j=1}^{k}\Big[d_j \log(\widehat{h}_j) + (r_j - d_j)\log(1-\widehat{h}_j)\Big]\end{aligned}$$

This can easily be maximized and $\widehat{h}_j = d_j/r_j$, which gives the Kaplan-Meier estimator,

$$\widehat{S}(t) = \prod_{j|t_j<t}(1 - d_j/r_j).$$

A nonparametric MLE of the cumulative hazard function, $H(t)$ can be obtained (see Lawless 2003, Chap. 3) from the likelihood function

$$L = \prod_{(0,\infty)} dH(t)^{dN_\cdot(t)}(1 - dH(t))^{Y_\cdot(t) - dN_\cdot(t)}$$

where $dN_\cdot(t) = \sum_{i=1}^n dN_i(t)$ and $Y_\cdot(t) = \sum_{i=1}^n Y_i(t)$. Maximizing this likelihood corresponds to setting $d\widehat{H}(t) = dN_\cdot(t)/Y_\cdot(t)$ so that

$$\widehat{H}(t) = \int_0^t \frac{dN_\cdot(u)}{Y_\cdot(u)} = \sum_{j|t_j<t}\frac{d_j}{r_j}.$$

This is the Nelson-Aalen estimate of the cumulative hazard function. The Kaplan-Meier estimate can be derived using the relationship

$$\widehat{S}(t) = \prod_{(0,t)}(1 - d\widehat{H}(u)).$$

Kernel Smoothing and Local Likelihood

The Kaplan-Meier and Nelson-Aalen estimates both produce step-functions. Sometimes it is desirable to have a smooth nonparametric estimates. A smooth estimate

of the density function can be obtained by kernel smoothing the Kaplan-Meier estimate. Similarly, a smooth estimate of the hazard function can be obtained by kernel smoothing the Nelson-Aalen estimate. Let $K_b(u) = K(u/b)/b$ where $K(u)$ is the kernel function. Typical examples of kernel functions are the rectangular kernel, $K(u) = 1/2$ for $|u| < 1$, the Epanechnikov kernel, $K(u) = 3/4(1 - u^2)$ for $|u| < 1$ and the Gaussian kernel, $K(u) = 1/\sqrt{2\pi}\exp(-u^2/2)$. Each of these kernels have properties that may suggest the use of one in a particular application, however in general the bandwidth is what will frequently have the greatest effect on the resulting estimate.

The kernel density estimates of the density and hazard are (Ramlau-Hansen 1983)

$$\widehat{f}(t) = \int_{-\infty}^{\infty} K_b(u - t)\, d\widehat{F}(u)$$

and

$$\widehat{h}(t) = \int_{0}^{\infty} K_b(u - t)\, d\widehat{H}(u)$$

where $\widehat{F}(u) = 1 - \widehat{S}(u)$, $\widehat{S}(u)$ is the Kaplan-Meier estimate and $\widehat{H}(u)$ is the Nelson-Aalen estimate. The downside to these estimates is they can be imprecise in the tails leading to bias (see Lawless 2003, Chap. 3).

The kernel density estimate can also be motivated by local likelihood arguments. For uncensored data, the local log-likelihood is (Loader 1996; Hjort and Jones 1996)

$$\ell(f, t) = \sum_{i=1}^{n} K_b(T_i - t) \log f(T_i) - n \int_{-\infty}^{\infty} K_b(u - t) f(u)\, du.$$

Hjort and Jones (1996) provide a number of justifications for this likelihood, the simplest of which being that the derivative of the first term does not have expectation 0 so the second term is added as a correction. The density function around the point t can be approximated by a parametric function. Frequently, a polynomial log density, $\log f(s) \approx \alpha_{0t} + \alpha_{1t}(s - t) + \cdots + \alpha_{pt}(s - t)^p$ for s near t is used. If the log of the density is approximated by a constant, *i.e.* $\log f(s) = \alpha_{0t}$ for s near t then the log-likelihood can be written as

$$\ell(f, t) = \sum_{i=1}^{n} K_b(T_i - t)\alpha_{0t} - n \exp(\alpha_{0t}).$$

Differentiating with respect to α_{0t} and setting to zero gives

$$\widehat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} K_b(T_i - t) = \int_{-\infty}^{\infty} K_b(u - t)\, d\widehat{F}(u).$$

Loader (1996) and Hjort and Jones (1996) also discuss reducing the bias of the kernel density estimators by approximating $\log f(u)$ for u near t by a polynomial.

Local likelihood estimation of the hazard function is discussed in Betensky et al. (1999) who deal with right-censored data. A similar approach is taken where the local log-likelihood function is taken to be

$$\ell(h,t) = \sum_{i=1}^{n} \left\{ \delta_i K_b(T_i - t) \log h(T_i) - \int_0^{T_i} h(u) K_b(u-t)\, du \right\}$$

where δ_i is 1 (0) if subject i's failure time is observed (censored). This local likelihood is constructed by weighting each individual's contributions according to its distance from the point t where the weight is assigned according to the chosen kernel. This leads to a local likelihood of the form

$$L(h,t) = \prod_{i=1}^{n} \prod_{(0,\infty)} (h(u)\, du)^{dN_i(u) K_b(u-t)} (1 - h(u)\, du)^{Y_i(u) K_b(u-t)(1 - dN_i(u))}$$

which reduces to

$$\begin{aligned} L(h,t) &= \prod_{i=1}^{n} h(T_i)^{\delta_i K_b(T_i - t)} \exp\left\{ - \int_0^{\infty} Y_i(u) K_b(u-t) h(u)\, du \right\} \\ &= \prod_{i=1}^{n} h(T_i)^{\delta_i K_b(T_i - t)} \exp\left\{ - \int_0^{T_i} K_b(u-t) h(u)\, du \right\}. \end{aligned}$$

If the kernel function is taken to be the rectangular kernel then the local log-likelihood at t is a restriction of the ordinary log-likelihood to the region $[t-b, t+b]$. For kernel functions of other forms, it amounts to assigning more weight to information closer to the time t.

The hazard function around the point t can be approximated by a parametric function, typically, $\log(h(s)) \approx \alpha_{0t} + \alpha_{1t}(s-t) + \cdots + \alpha_{pt}(s-t)^p$ for s near t leading to

$$\begin{aligned} \ell(h,t) = \sum_{i=1}^{n} \{ & \delta_i K_b(T_i - t) \left\{ \alpha_{0t} + \cdots + \alpha_{pt}(s-t)^p \right\} \\ & - \int_0^{T_i} \exp\left\{ \alpha_{0t} + \cdots + \alpha_{pt}(s-t)^p \right\} K_b(u-t)\, du \Big\} \end{aligned}$$

The estimate of $h(t)$ is then $e^{\widehat{\alpha}_{0t}}$ where $\widehat{\alpha}_{0t}, \ldots, \widehat{\alpha}_{pt}$ are the solutions to the local

score equations

$$\sum_{i=1}^{n} \Big\{ \delta_i K_b(T_i - t)(T_i - t)^d \\ - \int_0^{T_i} (u-t)^d \exp\left\{\alpha_{0t} + \cdots + \alpha_{pt}(u-t)^p\right\} K_b(u-t)\, du \Big\} = 0$$

for $d = 0, 1, \ldots, p$.

If the log of the hazard function around the point t is approximated by a constant, the resulting score equation has a closed form solution and

$$\widehat{h}(t) = \frac{\sum_{i=1}^{n} \delta_i K_b(T_i - t)}{\sum_{i=1}^{n} \int_0^{T_i} K_b(u-t)\, du}.$$

For the local likelihood methods, the bandwidth, kernel and degree of the polynomial must be specified. Betensky et al. (1999) suggest using a nearest-neighbour bandwidth which ensures a percentage of the data are always within the fitting window. Hjort and Jones (1996) and Loader (1999) discuss cross-validation techniques. For example, a least squares cross-validation for estimating the density chooses the bandwidth b such that it minimizes

$$\int_{-\infty}^{\infty} \widehat{f}^2(u; b)\, du - \frac{2}{n} \sum_{i=1}^{n} \widehat{f}_{-i}(T_i; b)$$

where $\widehat{f}_{-i}(t; b)$ is the estimate of f obtained by deleting the ith observation. An alternative criterion is the likelihood cross validation given by

$$\sum_{i=1}^{n} \log \widehat{f}_{-i}(T_i; b) - n \left(\int_{-\infty}^{\infty} \widehat{f}(u; b)\, du - 1 \right). \tag{1.1}$$

The second term is included as a kind of penalty for density estimates that do not integrate to 1. The likelihood cross validation criterion may be used without the penalty term. For example, the implementation in Loader (1999) does not include this term by default. Loader (1999) also notes that fixed bandwidths may perform poorly in the tails of the distribution, hence the motivation for nearest-neighbours bandwidths. Using polynomials of order 1 or 2 can help reduce bias at the boundaries.

1.2.2 Analysis of Interval-Censored Failure Time Data

Another form of incomplete data arises when subjects are periodically observed. In this situation we can only determine that the event of interest occurred between two time points. The event time is then only known to fall in an interval of time and this type of data is known as *interval-censored data*, (Sun 2006). If each subject is only observed once, then we can only observe whether or not the individual has experienced the event. Such data are referred to as case I interval-censored data, or *current status data*. The general situation of many assessments per individual are sometimes called case II interval-censored data.

It should also be noted that as described by Lindsey (1998), all observations on continuous time variables are interval censored. An event time is recorded to the nearest day, month and so on. One question is when does this censoring have an impact? If events occur on average once every year but we can only record to the nearest day, then the censoring intervals are quite small with respect to the time unit, and we are not losing too much by ignoring the interval censoring. Lindsey (1998) suggests that if less than an average of 0.2 events occur per unit of time per individual then interval censoring does not greatly affect the conclusions.

Suppose n individuals are interval-censored with censoring interval $(l_i, r_i]$ for individual i, $i = 1, 2, \ldots, n$. The likelihood under an independent censoring mechanism (Sun 2006) in this case is

$$L = \prod_{i=1}^{n} [S(l_i) - S(r_i)]$$

Parametric Models

If a parametric form, $S(t; \theta)$, is assumed then the likelihood becomes

$$L(\theta) = \prod_{i=1}^{n} [S(l_i; \theta) - S(r_i; \theta)]$$

and standard likelihood inference can be applied (*e.g.*, see Lawless 2003, Chapter 4). Kalbfleisch and Prentice (2002, Chapter 3) describes inference procedures for fitting accelerated failure time regression models to interval-censored data. Sun (2006, Chaps. 2 & 6) also discusses parametric models under interval censoring.

An EM Algorithm for Piecewise-Constant Models

Consider now the special case of the piecewise-constant model with interval censoring (Lindsey and Ryan 1998). In order to construct the likelihood function, individuals can be grouped into three classes. Let $\mathcal{O}$ be individuals with known failure times, $\mathcal{C}_R$ be individuals with right-censored failure times, and $\mathcal{C}_I$ be individuals with interval-censored failure times. Also, let t_i denote the survival (censoring) time for individuals with known (right-censored) failure times, and let (l_i, r_i) be the censoring interval for individuals with interval-censored failure times. It follows that the likelihood is given by

$$\begin{aligned} L(\lambda) = &\prod_{i\in\mathcal{O}} \left[\prod_{j=1}^{m} \lambda_j^{I(a_{j-1}<t_i\leq a_j)} \exp\left(-\sum_{j=1}^{m} \lambda_j w_j(t_i) \right) \right] \\ &\times \prod_{i\in\mathcal{C}_R} \left[\exp\left(-\sum_{j=1}^{m} \lambda_j w_j(t_i) \right) \right] \\ &\times \prod_{i\in\mathcal{C}_I} \left[\exp\left(-\sum_{j=1}^{m} \lambda_j w_j(l_i) \right) - \exp\left(-\sum_{j=1}^{m} \lambda_j w_j(r_i) \right) \right]. \end{aligned}$$

Numerical methods can be used to maximize this likelihood, however the likelihood function is messy. This suggests the use of an EM algorithm since it involves rewriting the likelihood as something mathematically simpler.

Let the complete data likelihood be the likelihood for known and right-censored observations only,

$$L_C(\lambda) = \prod_{i=1}^{n} \prod_{j=1}^{m} \lambda_j^{I(a_{j-1}<t_i\leq a_j)\delta_i} \exp\left(-\sum_{j=1}^{m} \lambda_j w_j(t_i) \right)$$

where $\delta_i = 1$ if individual i has a known failure time and $\delta_i = 0$ if individual i has a right-censored failure time. Note that for the purposes of the EM algorithm, exact and right-censored failure times are considered "complete" while interval-censored observations are considered "incomplete" exact observations which will be estimated at each E-step.

The complete data log-likelihood has a particularly simple form,

$$\ell_C(\lambda) = \sum_{j=1}^{m} \left(d_j \log \lambda_j - v_j \lambda_j \right)$$

where $d_j = \sum_{i=1}^{n} I(a_{j-1} < t_i \leq a_j)\delta_i$ and $v_j = \sum_{i=1}^{n} w_j(t_i)$. Note that a right-censored observation does not contribute anything to d_j.

The E-Step

At the rth iteration of the EM algorithm, we require $E(\ell_C(\lambda)|\text{data}, \widehat{\lambda}^{(r-1)})$ for the E-step, hence the required expectations are,

$$\begin{aligned} E(d_j|\text{data}, \widehat{\lambda}^{(r-1)}) &= \sum_{i\in\mathcal{O}} I(a_{j-1} < t_i \leq a_j) \\ &\quad + \sum_{i\in\mathcal{C}_I} P(a_{j-1} < t_i \leq a_j | l_i < t_i < r_i, \widehat{\lambda}^{(r-1)}) \\ E(v_j|\text{data}, \widehat{\lambda}^{(r-1)}) &= \sum_{i\in\mathcal{O}} w_j(t_i) + \sum_{i\in\mathcal{C}_R} w_j(t_i) \\ &\quad + \sum_{i\in\mathcal{C}_I} E(w_j(t_i)|l_i < t_i < r_i, \widehat{\lambda}^{(r-1)}) \end{aligned}$$

so for each interval-censored observation, the probability of failing in $(a_{j-1}, a_j]$ and the expected time at risk in $(a_{j-1}, a_j]$ must be computed. Note that for each interval-censored observation and given interval $(a_{j-1}, a_j]$, there are six possibilities as shown in Figure 1.1. For each case, the probability of failing and expected time at risk can be estimated.

Expressions for the probabilities are as follows where we suppress the dependence on λ

Case (A) or (B)

$$P(a_{j-1} < t_i \leq a_j | l_i < t_i \leq r_i) = 0$$

Case (C)

$$P(a_{j-1} < t_i \leq a_j | l_i < t_i \leq r_i) = 1$$

Case (D)

$$P(a_{j-1} < t_i \leq a_j | l_i < t_i \leq r_i) = \frac{S(a_{j-1}) - S(r_i)}{S(l_i) - S(r_i)}$$

Case (E)

$$P(a_{j-1} < t_i \leq a_j | l_i < t_i \leq r_i) = \frac{S(l_i) - S(a_j)}{S(l_i) - S(r_i)}$$

Case (F)

$$P(a_{j-1} < t_i \leq a_j | l_i < t_i \leq r_i) = \frac{S(a_{j-1}) - S(a_j)}{S(l_i) - S(r_i)}$$

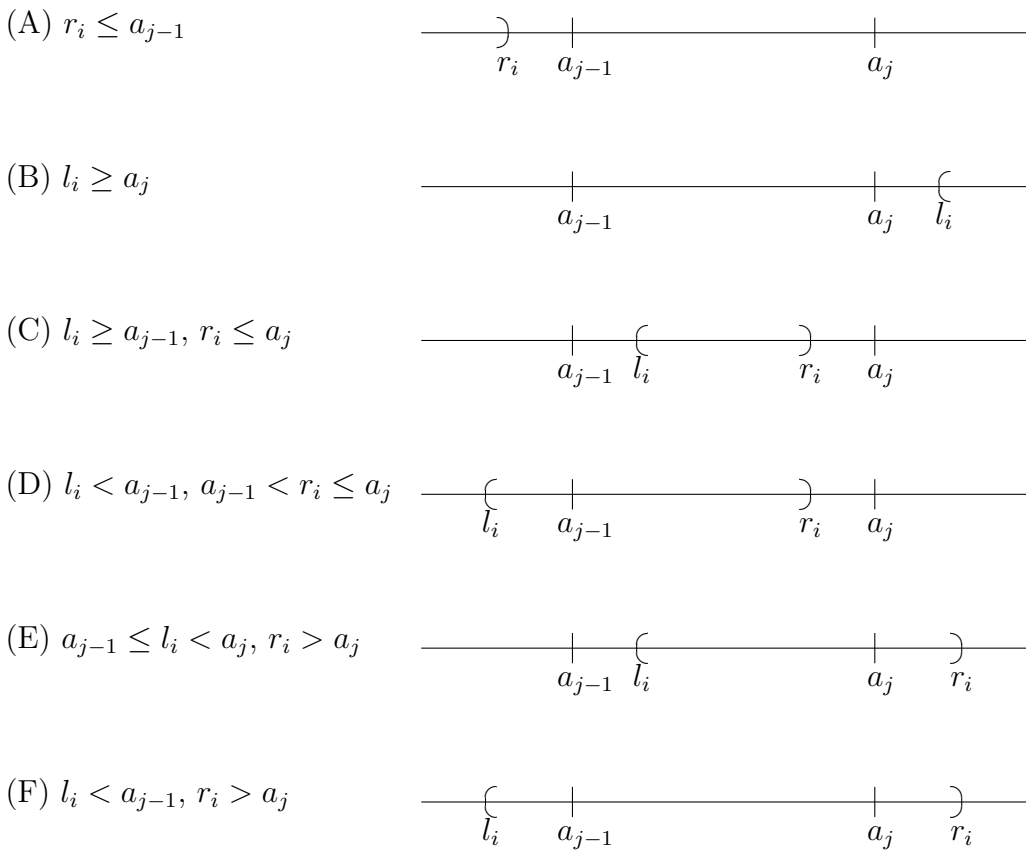

Figure 1.1: Six possibilities for the censoring interval relative to the cut-points in a piecewise model.

while expressions for the expected times at risk are

Case (A)

$$E(w_j(t_i)|l_i < t_i \leq r_i) = 0$$

Case (B)

$$E(w_j(t_i)|l_i < t_i \leq r_i) = a_j - a_{j-1}$$

Case (C)

$$E(w_j(t_i)|l_i < t_i \leq r_i) = l_i + \frac{1}{\lambda_j} + \frac{r_i - l_i}{1 - e^{\lambda_j(r_i - l_i)}} - a_{j-1}$$

Case (D)

$$E(w_j(t_i)|l_i < t_i \leq r_i) = \frac{S(a_{j-1}) - S(r_i)}{S(l_i) - S(r_i)} \left(\frac{1}{\lambda_j} + \frac{R_i - a_{j-1}}{1 - e^{\lambda_j(r_i - a_{j-1})}} \right)$$

Case (E)

$$\begin{aligned} E(w_j(t_i)|l_i < t_i \leq r_i) = & \frac{S(l_i) - S(a_j)}{S(l_i) - S(r_i)} \left(l_i + \frac{1}{\lambda_j} + \frac{a_j - l_i}{1 - e^{\lambda_j(a_j - l_i)}} - a_{j-1} \right) \\ & + (a_j - a_{j-1}) \sum_{k=j+1}^{m} P(a_{k-1} < t_i \leq a_k | l_i < t_i \leq r_i) \end{aligned}$$

Case (F)

$$\begin{aligned} E(w_j(t_i)|l_i < t_i \leq r_i) = & \frac{S(a_{j-1}) - S(a_j)}{S(l_i) - S(r_i)} \left(\frac{1}{\lambda_j} + \frac{a_j - a_{j-1}}{1 - e^{\lambda_j(a_j - a_{j-1})}} \right) \\ & + (a_j - a_{j-1}) \sum_{k=j+1}^{m} P(a_{k-1} < t_i \leq a_k | l_i < t_i \leq r_i). \end{aligned}$$

It should be noted that if some λ_j's are zero and this results in $S(l_i) = S(r_i)$ then $P(a_{j-1} < t_i \leq a_j | l_i < t_i \leq r_i) = 0$ and $E(w_j(t_i)|l_i < t_i \leq r_i) = 0$ since in this case $\{l_i < t_i \leq r_i\}$ is an impossible event.

The M-Step

The M-step is easy, due to the simple form of the complete data log-likelihood. For each j,

$$\widehat{\lambda}_j^{(r)} = \frac{E(d_j|\text{data}, \widehat{\lambda}^{(r-1)})}{E(v_j|\text{data}, \widehat{\lambda}^{(r-1)})}.$$

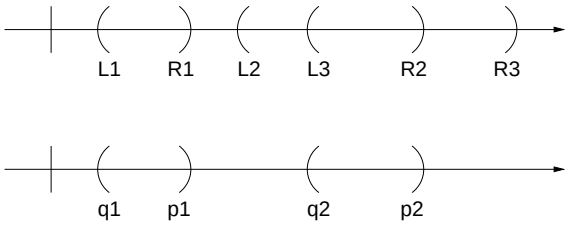

Figure 1.2: Example of how the intervals of support for the NPMLE are determined in the case of interval-censored failure time data.

The iterations are terminated when convergence is achieved, typically based on the requirement that the difference in estimates on successive iterations drops below a specified threshold (*e.g.* 10^{-6}). Alternatively, convergence can be declared by calculating the log-likelihood at each iteration and stopping when an iteration produces a change in the log-likelihood less than a specified tolerance.

Nonparametric Estimation

For a nonparametric approach to interval censoring, the most common method is due to Turnbull (1976). He developed an algorithm for obtaining the nonparametric MLE of the cumulative distribution function.

The observed data are $(l_i, r_i]$, $i = 1, \ldots, n$. Peto (1973) constructed a set that is crucial to determining the NPLME. Let $[q_1, p_1], \ldots, [q_m, p_m]$, $q_1 \leq p_1 < \ldots < q_m \leq p_m$ such that the q_j's are chosen from the l_i's, the p_j's are chosen from the r_i's and no $[q_j, p_j]$ contains any other l_i's or r_i's. This can be done by arranging the endpoints of the observed intervals on a timeline as in Figure 1.2. An interval $[q_j, p_j]$ occurs when a left endpoint is immediately followed by a right endpoint. Let $C = \bigcup_{j=1}^{m} [q_j, p_j]$. The set C is sometimes referred to as "innermost intervals".

The likelihood is

$$L(S) = \prod_{i=1}^{n} [S(l_i) - S(r_i)]$$

Peto (1973) noted that for $\widehat{S}(t)$ to be the NPMLE it must satisfy two conditions

1. $\widehat{S}(t)$ cannot decrease outside the set C.
2. The likelihood does not depend on the behaviour of $\widehat{S}(t)$ within the set C.

Hence, $\widehat{S}(t)$ can be parameterized as a distribution such that it is flat outside C, undefined within C and has jumps of size $s_j = \widehat{S}(q_j) - \widehat{S}(p_j)$ over the intervals

$[q_j, p_j]$. Note that $\sum_{j=1}^m s_j = 1$. The likelihood can now be reduced to

$$L(s) = \prod_{i=1}^n \sum_{j=1}^m \alpha_{ij} s_j$$

where $\alpha_{ij} = I([q_j, p_j] \subseteq (l_i, r_i))$. This likelihood is simple to maximize if there are very few s_j's, however in general it would be difficult using standard calculus techniques. Peto (1973) developed a Newton-Raphson-type algorithm with many boundary and step-size considerations, while Turnbull (1976) developed an easier to implement self-consistency algorithm to obtain the MLE's. Turnbull (1976) also noted at the end of his paper that this algorithm can be viewed as an EM algorithm. The latter approach will be adopted here. Other algorithms are described in Section 3.4 of Sun (2006).

For the purposes of the EM algorithm, the data are "complete" when it is known which interval $[q_j, p_j]$ contains t_i. Let $j_{(i)}$ be the index of the interval containing t_i. The complete data likelihood is

$$\begin{aligned} L_C(s) &= \prod_{i=1}^n P(t_i \in [q_{j_{(i)}}, p_{j_{(i)}}]; s) \\ &= \prod_{i=1}^n \prod_{j=1}^m s_j^{I(t_i \in [q_j, p_j])} \end{aligned}$$

with log-likelihood

$$\ell_C(s) = \sum_{i=1}^n \sum_{j=1}^m I(t_i \in [q_j, p_j]) \log s_j$$

At the E-step of the rth iteration, $P(t_i \in [q_j, p_j] | t_i \in (l_i, r_i), s^{(r-1)})$ is required. This is easy to compute as

$$P(t_i \in [q_j, p_j] | t_i \in (l_i, r_i), s^{(r-1)}) = \frac{\alpha_{ij} s_j^{(r-1)}}{\sum_{k=1}^m \alpha_{ik} s_k^{(r-1)}} = \mu_{ij}(s^{(r-1)})$$

For the M-step, the constraint $\sum_{j=1}^m s_j = 1$ must be imposed. Let

$$g(s^{(r)}, \lambda) = \sum_{i=1}^n \sum_{j=1}^m \mu_{ij}(s^{(r-1)}) \log s_j^{(r)} + \lambda \left(1 - \sum_{j=1}^m s_j^{(r)} \right)$$

Then,

$$\frac{\partial g}{\partial s_j^{(r)}} = \sum_{i=1}^n \frac{\mu_{ij}(s^{(r-1)})}{s_j^{(r)}} - \lambda$$

which implies

$$s_j^{(r)} = \frac{1}{\lambda} \sum_{i=1}^{n} \mu_{ij}(s^{(r-1)})$$

Keeping in mind that $\sum_{j=1}^{m} s_j^{(r)} = 1$ we get $\lambda = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij}(s^{(r-1)}) = n$, so an iteration of the EM algorithm (or Turnbull's self-consistency algorithm) is given by

$$s_j^{(r)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_{ij} s_j^{(r-1)}}{\sum_{k=1}^{m} \alpha_{ik} s_k^{(r-1)}}, \quad j = 1, \ldots, m$$

The algorithm is continued until changes in L are negligible. The NPMLE of S then becomes

$$\widehat{S}(t) = \begin{cases} 1 & \text{if } t < q_1 \\ 1 - s_1 - \cdots - s_j & \text{if } p_j < t < q_{j+1} \\ 0 & \text{if } t > p_m \\ \text{undefined} & \text{if } t \in C \end{cases}$$

It should also be noted that this algorithm was also obtained by Betensky (2000) as a generalization of the redistribution of mass algorithm (Dinse 1985; Efron 1967). In addition, Gentleman and Geyer (1994) provide conditions for verifying that the algorithm converges to the maximum and that this maximum is unique. Finkelstein (1986) generalizes the method to allow covariates using a proportional hazards model. A full likelihood approach is taken, requiring estimation of the baseline survivor function. Alternatives which do not require estimation of the baseline survivor function are considered in Satten (1996) and Goggins et al. (1998). These approaches require considerable computational effort as they rely on Gibbs sampling and Monte Carlo EM respectively. Finkelstein et al. (2002) consider the situation where the censoring mechanism is not independent.

The resulting NPMLE of the survivor function can be kernel smoothed according to

$$\widehat{f}(t) = E_{\widehat{S}(t)}\{K_b(T - t)\}$$

and then integrated to get a smooth estimate of the distribution function, however this expectation depends on how the probability masses are distributed within the set C. Braun et al. (2005) show how assigning mass in different ways lead to different estimates, hence there is no unique kernel smoothed estimate. Li et al. (1997) propose an iterative algorithm which can be viewed as iteratively taking the conditional expectation of the empirical distribution function, where at the jth

iteration,

$$\widetilde{F}_j(t) = E_{\widetilde{F}_{j-1}(t)}\left[\frac{1}{n}\sum_{i=1}^{n} I(T_i \le t)\middle| I_1, \ldots, I_n\right].$$

This leads to an estimator that coincides with Turnbull's algorithm outside the set C, and interpolates within C, however, the interpolation within C is different depending on the initial distribution, $\widetilde{F}_0(t)$. Braun et al. (2005) propose a density estimate which smooths at each iteration,

$$\widetilde{f}_j(t) = \frac{1}{n}\sum_{i=1}^{n} E_{\widetilde{f}_{j-1}(t)}\{K_b(T_i - t)|I_i\}.$$

The same estimate is obtained regardless of initial value. This estimator can also be derived based on local likelihood arguments, however in the context of failure time data it is more natural to work with the hazard function as opposed to the density.

Betensky et al. (1999) describe local likelihood methodology for estimating the hazard function in the presence of interval censoring. The complete data local log-likelihood is

$$\ell(h,t) = \sum_{i=1}^{n}\left\{K_b(T_i - t)\log h(T_i) - \int_0^{T_i} K_b(u-t)h(u)\,du\right\}.$$

The log-hazard can be approximated near t by a polynomial or simply by letting $\log h(u) = \alpha_{0t}$ for u near t. If $\widehat{\alpha}_{0t}$ is an estimate of this, then the estimate of the hazard at t is then $\widehat{h}(t) = \exp(\widehat{\alpha}_{0t})$.

An EM type algorithm can be developed to obtain such an estimate. Upon plugging in the locally constant approximation of the log-hazard and taking conditional expectations of $\ell(h,t)$ above we get

$$E\left[\ell(h,t)|I_i\right] = \sum_{i=1}^{n}\left\{\alpha_{0t}E\left[K_b(T_i - t)|I_i\right] - e^{\alpha_{0t}}E\left[\int_0^{T_i} K_b(u-t)\,du\middle| I_i\right]\right\}.$$

Differentiating with respect to α_{0t} and setting to zero yields the fixed point equation

$$\widehat{h}(t) = \frac{\displaystyle\sum_{i=1}^{n} E\left[K_b(T_i - t)|I_i\right]}{\displaystyle\sum_{i=1}^{n} E\left[\int_0^{T_i} K_b(u-t)\,du\middle| I_i\right]}$$

where the conditional expectations depend on the unknown hazard function.

In order to solve the fixed point equation, the conditional expectations are evaluated using the trapezoid rule for numerical integration. If we let $t_1, t_2, \ldots, t_M$ be a sequence of equally spaced points with $t_k - t_{k-1} = \Delta$ and $f_k = f(t_k) = h(t_k)\exp(-\int_0^{t_k} h(u)\,du) \approx h(t_k)\exp(-\Delta\sum_{i=1}^{k} h(t_i))$ then for any function g

$$E\left[g(T)\right] = \frac{\sum\limits_{l:t_l \in I_i} g(t_l) f_l \Delta}{\sum\limits_{l:t_l \in I_i} f_l \Delta} = \frac{\sum\limits_{l:t_l \in I_i} g(t_l) f_l}{\sum\limits_{l:t_l \in I_i} f_l}$$

and an EM algorithm can be constructed. Betensky et al. (1999) discusses further issues such as the use of linear and quadratic approximations to the log-hazard and standard error calculations. The choice of the smoothing parameter is also addressed. Betensky et al. (1999) also propose 400 to be a safe choice for the number of grid points. Extensions to incorporate covariates in a proportional hazards regression model are explored in Betensky et al. (2002).

Automatic bandwidth selection becomes more difficult in the interval censored case. Methods for right-censored data may carry over to the interval censored case (Betensky et al. 1999). Such methods may be computationally intensive, hence Bebchuck and Betensky (2001) chose a bandwidth by visual inspection. They also note that nearest-neighbours bandwidths perform better in hazard estimation than for density estimation. If the density function is of interest, local likelihood can be used as described in Braun et al. (2005). They also propose a cross-validation procedure based on the intervals of support for the NPMLE (innermost intervals). Sun (2006) describes other ways of smoothing the NPMLE as well as a penalized likelihood approach incorporating splines.

Bivariate Failure Time Data

Consider a study where individuals have two failure times, T_{i1} and T_{i2}, $i = 1, \ldots, n$. We would like to estimate the bivariate joint distribution function is given by $F(t_1, t_2) = P(T_{i1} \leq t_1, T_{i2} \leq t_2)$. When individuals are only seen at periodic assessment times, we observe $U_i = (L_{i1}, R_{i1}] \times (L_{i2}, R_{i2}]$, $i = 1, \ldots, n$, where $(L_{i1}, R_{i1}]$ and $(L_{i2}, R_{i2}]$ are the univariate censoring intervals for T_{i1} and T_{i2} respectively. Hence, for each individual we have a rectangle within which their failure times may have occurred.

In order to obtain the NPMLE of $F(t_1, t_2)$, we use the likelihood (Sun 2006,

Chapter 7)

$$L(F) = \prod_{i=1}^{n} \{F(R_{i1}, R_{i2}) - F(R_{i1}, L_{i2}) - F(L_{i1}, R_{i2}) + F(L_{i1}, L_{i2})\}.$$

In the univariate case, the probability masses were concentrated on the innermost intervals given by the set C. Betensky and Finkelstein (1999) and Gentleman and Vandal (2002) show that a similar situation occurs in the bivariate setting with the masses being concentrated on a region made up of intersections of the observation rectangles. Let this region be denoted by $H = \{H_j, j = 1 \ldots, m\}$, where a rectangle H_j is the bivariate analogue of the univariate innermost interval. Algorithms for determining H are given in Betensky and Finkelstein (1999), Gentleman and Vandal (2002) as well as Bogaerts and Lesaffre (2004). Sun (2006, Chapter 7) discusses the differences between the three approaches.

Defining s_j to be the probability mass on H_j and $\alpha_{ij} = I(H_j \subseteq U_i)$, the likelihood becomes

$$L(s) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} s_j$$

where $s = (s_1, \ldots, s_m)'$. Maximization of this likelihood subject to $\sum_{j=1}^{m} s_j = 1$ can be carried out using one of the algorithms for univariate failure time data, such as the method described previously (Turnbull 1976) or one of the methods in Section 3.4 of Sun (2006).

1.2.3 Analysis of Truncated Failure Time Data

In some situations an individual is only included in a study if their survival time lies within a certain interval. In this case we say the individual's failure time is *truncated*, and the respective interval is called the *truncation interval*.

A special case of truncation is when the interval is of the form (u, ∞). This is referred to as *left truncation*. This situation can arise in a prospective study when an individual is selected for inclusion if they are event-free at some time u after the beginning of a process of interest. In this case, the individual's failure time must therefore lie in the interval (u, ∞), and they are said to have a left truncation time u.

Klein and Moeschberger (1997) provides an example where the ages of death for individuals living in a retirement centre are of interest. Since the individuals

must reach a certain age before they can be admitted, the ages of death are left-truncated. Individuals who die before they are eligible to enter the centre are not included in the data set.

Another special case is when the interval is of the form $(0, v)$. This is known as *right truncation*. This can arise in a retrospective study where an individual is selected such that their failure time must be less than some time v.

Lawless (2003) describes a study where the time of interest was the time between HIV infection and AIDS diagnosis. The subjects were selected in 1987 and consisted of individuals who had been diagnosed with AIDS prior to July 1, 1986. The time of HIV infection was retrospectively determined, so that the time of interest was right-truncated, with the right-truncation time given by the time between HIV infection and July 1, 1986.

Data can be subject to both censoring and truncation at the same time. The most general case is when each individual is interval-censored and truncated with censoring interval $A_i = (l_i, r_i)$ and truncating set $B_i = (u_i, v_i)$. Turnbull's algorithm can be used in this case in very much the same manner as before.

The likelihood for this case becomes

$$\begin{aligned} L &= \prod_{i=1}^{n} P(t_i \in A_i | t_i \in B_i) \\ &= \prod_{i=1}^{n} \frac{S(l_i) - S(r_i)}{S(u_i) - S(v_i)} \end{aligned}$$

Frydman (1994) and Alioum and Commenges (1996) both describe how to construct a set Q in the presence of both censoring and truncation, which is analogous to the set C in the case of interval-censored data only. Let

$$\begin{aligned} \mathcal{L} &= \{l_i; i = 1, \ldots, n\} \cup \{v_i; i = 1, \ldots, n\} \cup \{0\} \\ \mathcal{R} &= \{r_i; i = 1, \ldots, n\} \cup \{u_i; i = 1, \ldots, n\} \cup \{\infty\} \end{aligned}$$

Let $[q_1, p_1], \ldots, [q_m, p_m]$, $q_1 \leq p_1 < \ldots < q_m \leq p_m$ such that the q_j's are chosen from $\mathcal{L}$, the p_j's are chosen from $\mathcal{R}$ and no $[q_j, p_j]$ contains any other elements of $\mathcal{L}$ or $\mathcal{R}$. Let $Q = \bigcup_{j=1}^{m} [q_j, p_j]$. This set can be decomposed into three parts. Let C be the union of intervals $[q_j, p_j]$ covered by at least one censoring interval, W be the union of intervals $[q_j, p_j]$ covered by at least one truncating set but not covered by any censoring interval, and D be the union of intervals $[q_j, p_j]$ not covered by any truncating set. Note that $D = \bigcap_{i=1}^{n} B_i^{\complement}$. Then $Q = C \cup W \cup D$.

Alioum and Commenges (1996) provided results similar to those of Peto (1973) for characterizing the NPMLE. $\widehat{S}(t)$ cannot decrease outside the set $C \cup D$. Furthermore, the likelihood does not depend on $P(T \in D)$ since the data provides no information about the region D. This means all that can be estimated is $P(T > t | T \notin D)$. If $D = \emptyset$ (*i.e.* if at least one observation is not truncated) then $S(t)$ can be estimated.

Based on these results, the problem becomes that of estimating $S(t)$ ($P(T \in D)$ is known) or $P(T > t | T \notin D)$ ($P(T \in D)$ is unknown). In either case, the set C is the only part of Q that affects the remainder of the estimation procedure. For simplicity, assume that $P(T \in D) = 0$, so we are considering estimating $S(t)$. The last fact required to parameterize the problem is to note that the likelihood does not depend on the behaviour of $\widehat{S}(t)$ within the set C.

Letting $s_j = \widehat{S}(q_j) - \widehat{S}(p_j)$, the likelihood becomes

$$L(s) = \prod_{i=1}^{n} \frac{\sum_{j=1}^{m} \alpha_{ij} s_j}{\sum_{j=1}^{m} \beta_{ij} s_j}$$

where $\alpha_{ij} = I([q_j, p_j] \subseteq A_i)$ and $\beta_{ij} = I([q_j, p_j] \subseteq B_i)$.

Again, this likelihood is difficult to maximize directly, so an EM approach will be used. In order to construct the complete data likelihood, the effects of the truncation must be taken into account. Each subject observed under this truncation scheme can be viewed as the only subject observed among a group in which the remaining members were not observed due to truncation, *i.e.* their times lie in $B_i^{\complement}$. Let the (unknown) number of subjects in the group corresponding to subject i be G_i. Turnbull referred to these G_i subjects not observed due to truncation as subject i's "ghosts". The complete data likelihood is given by

$$\begin{aligned} L_C &= \prod_{i=1}^{n} P(t_i \in [q_{j_{(i)}}, p_{j_{(i)}}]) \prod_{g=1}^{G_i} P(t_{ig} \in [q_{j_{(ig)}}, p_{j_{(ig)}}]) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{m} s_j^{I(t_i \in [q_j, p_j]) + G_i I(t_{ig} \in [q_j, p_j])} \end{aligned}$$

with log-likelihood

$$\ell_C = \sum_{i=1}^{n} \sum_{j=1}^{m} [I(t_i \in [q_j, p_j]) + G_i I(t_{ig} \in [q_j, p_j])] \log s_j$$

The required expectations are

$$P(t_i \in [q_j, p_j] | t_i \in A_i, s^{(r-1)}) = \frac{\alpha_{ij} s_j^{(r-1)}}{\sum_{k=1}^{m} \alpha_{ik} s_k^{(r-1)}} = \mu_{ij}(s^{(r-1)})$$

and

$$E(G_i I(t_{ig} \in [q_j, p_j]) | t_i \in B_i, s^{(r-1)})$$

which we can write as

$$E(G_i P(t_{ig} \in [q_j, p_j] | G_i, t_{ig} \in B_i^{\complement}, s^{(r-1)}) | t_i \in B_i, s^{(r-1)})$$

or

$$\begin{aligned}
E\left(G_i \left. \frac{(1-\beta_{ij}) s_j^{(r-1)}}{\sum_{k=1}^{m} (1-\beta_{ik}) s_k^{(r-1)}} \right| t_i \in B_i, s^{(r-1)} \right) &= \frac{P(t_i \in B_i^{\complement} | s^{(r-1)})}{P(t_i \in B_i | s^{(r-1)})} \frac{(1-\beta_{ij}) s_j^{(r-1)}}{\sum_{k=1}^{m} (1-\beta_{ik}) s_k^{(r-1)}} \\
&= \frac{\sum_{k=1}^{m} (1-\beta_{ik}) s_k^{(r-1)}}{\sum_{k=1}^{m} \beta_{ik} s_k^{(r-1)}} \frac{(1-\beta_{ij}) s_j^{(r-1)}}{\sum_{k=1}^{m} (1-\beta_{ik}) s_k^{(r-1)}} \\
&= \frac{(1-\beta_{ij}) s_j^{(r-1)}}{\sum_{k=1}^{m} \beta_{ik} s_k^{(r-1)}} \\
&= \nu_{ij}(s^{(r-1)})
\end{aligned}$$

The M-step is very similar to the interval-censored case, leading to an iteration having the form

$$s_j^{(r)} = \frac{\sum_{i=1}^{n} (\mu_{ij}(s^{(r-1)}) + \nu_{ij}(s^{(r-1)}))}{\sum_{i=1}^{n} \sum_{j=1}^{m} (\mu_{ij}(s^{(r-1)}) + \nu_{ij}(s^{(r-1)}))}.$$

1.3 Analyses Based on Multistate Models

1.3.1 Intensity Functions for Multistate Models

If the different conditions an individual can experience can be expressed in terms of several well-defined states, then multistate models are useful for modelling the

course of a disease. Individuals begin initially in a given state and make transitions to other states as time progresses.

Multistate data is often expressed using counting process notation. Suppose there are J states. Let $N_{ijk}(t)$ count the number of $j \to k$ transitions subject i experienced over $(0,t]$, $j \neq k$, $j,k = 1,\ldots,J$. Let $\Delta N_{ijk}(t) = N_{ijk}((t+\Delta t)^-) - N_{ijk}(t^-)$. Define the state occupancy variable $Z_i(t) = j$ which means subject i is in state j at time t and let $Y_{ij}(t) = I(Z_i(t^-) = j)$. The history of this process can be defined as $\mathcal{H}_i(t) = \{N_i(u), 0 \leq u < t; X_i(u), 0 \leq u \leq t\}$ where $N_i(t)$ is the vector of counting processes and $X_i(t)$ is a vector of covariates.

We can now define the *intensity function*,

$$\lambda_{ijk}(t|\mathcal{H}_i(t)) = \lim_{\Delta t \to 0} \frac{P(N_{ijk}((t+\Delta t)^-) - N_{ijk}(t^-) = 1|\mathcal{H}_i(t), Z_i(t^-) = j)}{\Delta t}$$

which is the instantaneous probability of individual i making a $j \to k$ transition at time t given the history over $[0,t)$ and that individual i is in state j just before t.

If the observed process is subject to right censoring, then we define $Y_i(t) = I(t \leq C_i)$, which is an indicator that subject i is under observation at time t and let $N^c_{ijk}(t) = \int_0^t Y_i(u)\, dN_{ijk}(u)$ denote the observed counts process. The history can now be expanded to be $\mathcal{H}_i(t) = \{N_i(u), Y_i(u), 0 \leq u < t; X_i(u), 0 \leq u \leq t\}$. The intensity function for the observed process in the presence of right censoring becomes

$$\lambda^c_{ijk}(t|\mathcal{H}_i(t)) = \lim_{\Delta t \to 0} \frac{P(N^c_{ijk}((t+\Delta t)^-) - N^c_{ijk}(t^-) = 1|\mathcal{H}_i(t), Z_i(t^-) = j)}{\Delta t}$$

and under independent censoring we have $\lambda^c_{ijk}(t|\mathcal{H}_i(t)) = Y_i(t)\lambda_{ijk}(t|\mathcal{H}_i(t))$.

The intensity functions can be readily used to construct the likelihood function when no interval censoring or truncation is present. Consider individual i observed over $(0,\tau]$. Let $t_{i1},\ldots,t_{im_i}$ be the times when transitions occurred and let $t_{i0} = 0$ and $t_{im_{i+1}} = \tau$. Let j_{ik} denote the state occupied at t_{ik}, $k = 0,\ldots,m_i$. Then, individual i's contribution to the likelihood is

$$L_i = \prod_{k=1}^{m_i} \lambda_{ij_{i,k-1}j_{ik}}(t_{ik}|\mathcal{H}(t_{ik})) \times \prod_{k=1}^{m_i+1} \exp\left[-\sum_{j=1}^{J} I(j \neq j_{i,k-1}) \int_{t_{i,k-1}}^{t_{ik}} \lambda_{ij_{i,k-1}j}(u|\mathcal{H}(u))\, du\right]$$

and $L = \prod_{i=1}^n L_i$, (Lawless 2003).

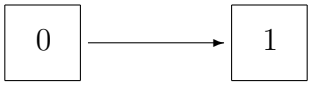

Figure 1.3: 2-State survival model.

Survival data can be thought of as the simplest example of a multistate model. In this case there are 2 states, the first representing no event, the second indicating the individual has experienced the event. This model is illustrated in Figure 1.3.

This model is easily extended to the progressive three state model, where there are three states (say 0, 1 and 2) and individuals in state 0 can only make transitions to state 1 and likewise individuals in state 1 can only make transitions to state 2, as in Figure 1.4. For example, an individual may be in state 0 if they are healthy, move to state 1 when they develop a disease and move to state 2 when they develop a complication caused by the disease. In this example it would be assumed that individuals can only develop the complication if they have the disease.

Figure 1.4: Progressive 3-State model.

Another 3-state model is the illness-death model, where individuals in state 0 can make transitions to either state 1 or 2, while individuals in state 1 can only make transitions to state 2, see Figure 1.5. The name of this model is derived from the fact that often state 0 represents healthy, state 1 represents illness and state 2 represents death. Individuals do not necessarily have the illness when they die, hence transitions from state 0 to state 2 are permitted.

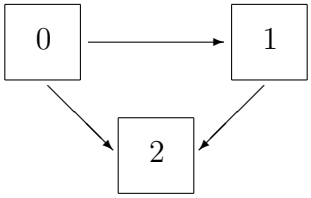

Figure 1.5: Illness-Death model.

Multistate models can also be used in the problem of competing risks. In this situation individuals begin in state 0 and can experience one and only one of k

different events. Figure 1.6 shows the state transition diagram for the problem of $k = 3$ competing risks. An example for this kind of model would be to have state 0 represent alive, state 1 represent death by heart attack, state 2 represent death by cancer and state 3 represent death by car accident.

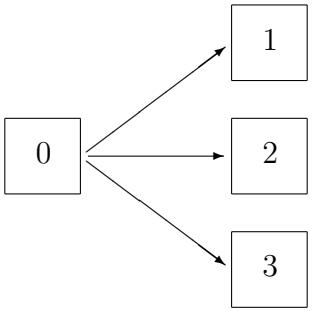

Figure 1.6: Competing Risks model.

These are just a few of the models that can be examined in a multistate framework. Many other models for all kinds of problems can be easily defined (see, for example, Hougaard 1999).

A frequent assumption is that the process does not depend on what has happened previously and all that matters is the current state and time. This is the Markov assumption, which means that given the process is in state j at time t^-, the intensity of a $j \to k$ transition is independent of $\mathcal{H}(t)$. In this case we can write $\lambda_{ijk}(t|\mathcal{H}_i(t)) = \rho_{ijk}(t)$. *Markov models* are useful in situations where the time since the beginning of the process is of importance. Such situations include cases when aging effects may be present.

If it is further assumed that $\rho_{ijk}(t) = \rho_{ijk}$ for all j and k, then the model is a time homogeneous Markov model. This assumption basically reduces the process to a parametric model with a single parameter describing the rate for each type of transition.

An alternative assumption to the Markov assumption is to assume the process only depends on the current state and the amount of time spent in that state. This is the case with semi-Markov models in which, given the process is in state j and has been in state j for a length of time s, the intensity of a $j \to k$ transition is independent of $\mathcal{H}(t)$. In this case we can write $\lambda_{ijk}(t|\mathcal{H}_i(t)) = h_{ijk}(s)$. Semi-Markov models are useful in situations where a change of state leads to a fundamental change in the process, since this essentially causes a change in the time origin. This includes situations where the duration of a condition is of primary interest.

If $h_{ijk}(s) = \rho_{ijk}$ for all j and k then we have a time homogeneous Semi-Markov model which is identical to the time homogeneous Markov model.

Frequently, one wishes to know about the probability of being in a certain state at a given time. This information can be summarized by the *prevalence functions*, defined by $p_{0j}(t) = P(Z(t) = j|Z(0) = 0)$ for $j = 1, \ldots, J$.

1.3.2 Analysis of Right-Censored Multistate Data

When data are right-censored, a common approach to modelling is based on a Markov assumption. We wish to estimate the transition probability matrix $P(s,t)$ which has (j,k) entry given by $P(Z(t) = k|Z(s) = j)$. Under a Markov assumption, there is a simple relationship between $P(s,t)$ and the intensity functions, $\rho_{jk}(t)$. Let $Q(t)$ be the matrix with (j,k) entry given by $\rho_{jk}(t)$, $j \neq k$ and $\rho_{jj}(t) = -\sum_{k \neq j} \rho_{jk}(t)$. Then $P(s,t)$ can be expressed as the product integral,

$$\prod_{(s,t]} \{I + Q(u)\, du\}$$

hence estimates of $P(s,t)$ can be constructed from estimates of $Q(t)$ using this relationship. If the transition intensities are estimated using the Nelson-Aalen estimates,

$$\widehat{\rho}_{jk}(t) = \frac{\sum_{i=1}^{n} Y_{ij}(t) dN_{ijk}(t)}{\sum_{i=1}^{n} Y_{ij}(t)}$$

then this gives the Aalen-Johansen estimate (Andersen et al. 1993, Chap. 4) of $P(s,t)$. Estimates of the prevalence functions can be obtained from the first row of $P(s,t)$. Even though this estimator was derived under a Markov assumption, Datta and Satten (2001) show that the estimated prevalence functions are consistent even when the underlying process is not Markov under independent censoring.

1.3.3 Analysis of Interval-Censored Multistate Data

Interval censoring in multistate models often occurs when subjects are examined at inspection times. The state a subject occupies is only known at the inspection times. As in the case for right censoring, the inspection process must satisfy conditions in order to construct the likelihood function (Lawless and Zhan 1998; Grüger et al. 1991).

Define the inspection times for a subject to be $b_1, \ldots, b_m$ and define the histories $\mathcal{H}_j = \{b_1, \ldots, b_j, Z(b_1), \ldots, Z(b_j)\}$ and $\mathcal{H}_{j-} = \{b_1, \ldots, b_j, Z(b_1), \ldots, Z(b_{j-1})\}$. Then the likelihood is

$$L = P(\mathcal{H}_m) = P(\mathcal{H}_0) \prod_{j=1}^{m} P(Z(b_j)|b_j, \mathcal{H}_{j-1}) \prod_{j=1}^{m} P(b_j|\mathcal{H}_{j-1})$$

The independent inspection process condition can now be defined as

$$P(Z(b_j)|b_j, \mathcal{H}_{j-1}) = P(Z(b_j)|Z(b_1), \ldots, Z(b_{j-1}))$$

i.e. the state occupied at b_j cannot depend on the inspection process. The non-informative condition is simply $P(b_j|\mathcal{H}_{j-1})$ does not depend on the parameters of interest. When both these conditions are satisfied, the likelihood (conditional on $\mathcal{H}_0$) simplifies to

$$L = P(Z(b_1), \ldots, Z(b_m)).$$

Examples of particular inspection processes are given in Grüger et al. (1991). If an inspection scheme is given by pre-specified (non-random) visit times then it is an independent inspection process. The same can be said if the inspection times are from a random process independent of the multistate process. If a subject is being treated by a doctor and the doctor schedules the next inspection time based on the current state the subject is in, this type of inspection process still satisfies the independence condition. However, if a subject schedules an inspection based on how they feel, an inspection process of this type violates the independence condition since the inspection time may be dependent on the (possibly updated) state the subject is observed to be in.

When the data are interval-censored, Markov models are most often used. Kalbfleisch and Lawless (1985) describe the time-homogeneous Markov model. In this case, $\rho_{jk}(t) = \rho_{jk}$ and $P(t) = P(s, s+t) = P(0, t)$. It is assumed that $\rho_{jk} = \rho_{jk}(\theta)$ depends on a vector of parameters.

Assume that individuals are observed at times $t_1, \ldots, t_m$. Define n_{jkl} to be the number of individuals in state j at t_{l-1} and in state k at t_l. Then the likelihood becomes

$$\begin{aligned} L(\theta) &= \prod_{l=1}^{m} \prod_{j=1}^{J} \prod_{k=1}^{J} p_{jk}(t_{l-1}, t_l)^{n_{jkl}} \\ &= \prod_{l=1}^{m} \prod_{j=1}^{J} \prod_{k=1}^{J} p_{jk}(w_l)^{n_{jkl}} \end{aligned}$$

where $w_l = t_l - t_{l-1}$. The log-likelihood function is

$$\ell(\theta) = \sum_{l=1}^{m} \sum_{j=1}^{J} \sum_{k=1}^{J} n_{jkl} \log p_{jk}(w_l)$$

Kalbfleisch and Lawless (1985) describe a quasi-Newton procedure for maximizing the log-likelihood. They show that the expected information matrix can be computed using only first derivatives of the p_{jk}'s. The uth entry of the score vector is

$$U_u(\theta) = \sum_{l=1}^{m} \sum_{j=1}^{J} \sum_{k=1}^{J} n_{jkl} \frac{\partial p_{jk}(w_l)/\partial \theta_u}{p_{jk}(w_l)} \tag{1.2}$$

The (u, v)th entry of the observed information is given by

$$\sum_{l=1}^{m} \sum_{j=1}^{J} \sum_{k=1}^{J} n_{jkl} \left\{ \frac{\partial p_{jk}(w_l)/\partial \theta_u \partial p_{jk}(w_l)/\partial \theta_v}{p_{jk}^2(w_l)} - \frac{\partial^2 p_{jk}(w_l)/\partial \theta_u \partial \theta_v}{p_{jk}(w_l)} \right\}.$$

The expectation of these terms can be taken by first conditioning on $N_j(t_{l-1}) = \sum_{k=1}^{J} n_{jkl}$. Then since

$$E\{n_{jkl} | N_j(t_{l-1})\} = p_{jk}(w_l) N_j(t_{l-1})$$

and

$$\sum_{k=1}^{J} \partial^2 p_{jk}(w_l)/\partial \theta_u \partial \theta_v = 0$$

the (u, v)th entry of the expected information is given by

$$M_{uv}(\theta) = \sum_{l=1}^{m} \sum_{j=1}^{J} \sum_{k=1}^{J} \frac{E\{N_j(t_{l-1})\}}{p_{jk}(w_l)} \frac{\partial p_{jk}(w_l)}{\partial \theta_u} \frac{\partial p_{jk}(w_l)}{\partial \theta_v} \tag{1.3}$$

which can be estimated by replacing $E\{N_j(t_{l-1})\}$ by the observed $N_j(t_{l-1})$. Let $U(\theta)$ be the score vector obtained by summing terms (1.2) over all m subjects in the sample, and $M(\theta)$ the corresponding estimated expected information matrix from (1.3). One step of the Fisher-scoring algorithm is performed by computing $\theta^{(r)} = \theta^{(r-1)} + M^{-1}(\theta^{(r-1)}) U(\theta^{(r-1)})$. Details on the computation of the derivatives of the p_{jk} can be found in Kalbfleisch and Lawless (1985).

The estimate of $P(t)$ is found by noting that if $Q = ADA^{-1}$ is the eigenvalue-eigenvector decomposition of Q then

$$P(t) = A e^{D(t)} A^{-1}$$

where $e^{D(t)} = \text{diag}(e^{d_1 t}, \ldots, e^{d_J t})$.

Numerous extensions to this method are provided in Kalbfleisch and Lawless (1985, 1989). If a non-homogeneous Markov model is assumed such that $Q(t) = Q_0 g(t;\lambda)$, then the method applies with a slight modification. Here, Q_0 is a fixed intensity matrix with unknown entries, and $g(t;\lambda)$ is a function of time that depends on an unknown parameter λ. If we let $s = \int_0^t g(u;\lambda)\,du$ and $Z_0(s) = Z(t)$, then $Z_0(s)$ is a time-homogeneous Markov process with intensity matrix Q_0. Hence, for a given λ the parameters of Q_0 can be estimated, and λ can be varied to find the value which maximizes the log-likelihood.

Another way of dealing with non-homogeneity is to assume $Q(t)$ changes at specified time points, but remains constant between time points. In this case the method is applied separately to each interval where $Q(t)$ is constant.

Kalbfleisch and Lawless (1989) also provide a test of time homogeneity. The basic idea is to assume $Q(t) = Q + H\gamma t$, where H is a matrix specifying which entries of $Q(t)$ may be non-homogeneous. A score test of $\gamma = 0$ can be used to determine if time homogeneity appears reasonable or not. Gentleman et al. (1994) generalizes these methods by allowing piecewise-constant transition intensities.

When a non-homogeneous Markov model is of interest, nonparametric procedures similar to Turnbull's method can be used. Frydman (1992) considers a progressive three state model where $0 \to 1$ transitions can be interval-censored and $1 \to 2$ transitions can be right-censored. The functions of interest are F_1, the distribution function of the time until a $0 \to 1$ transition and Λ_2, the cumulative intensity of $1 \to 2$ transitions.

As in Turnbull's algorithm, a set C can be constructed which characterizes the NPLME. In this case, F_1 is flat outside C and undefined within C while Λ_2 increases only at failure times. Again, this basically reduces the problem to a parametric one where the jumps in F_1 and Λ_2 are the parameters to be estimated. Frydman (1992) provides a self-consistent algorithm to obtain the NPMLE. Using a similar idea, Frydman (1995) shows how to obtain NPMLE's of the cumulative intensities of an illness-death model. An extension to allow left-truncated data is also provided. An overview of issues and methods for the analysis of interval-censored multistate data can be found in Commenges (2002).

The competing risks model with both interval censoring and truncation has been examined by Hudgens et al. (2001). If T is the failure time and there are J possible failure types, then define $I_j(t) = P(T \le t, \text{failure of type } j),\ j = 1 \ldots, J$. Note that $\sum_{j=1}^{J} I_j(t) = 1 - S(t)$. Straightforward modification of Turnbull's algorithm

leads to the NPMLE's of each $I_j(t)$. A complication in this case is that each $\widehat{I}_j(t)$ is undefined on a set C_j, so that $\widehat{S}(t) = 1 - \sum_{j=1}^{J} I_j(t)$ is undefined on $\cup_{j=1}^{J} C_j$. If $\widetilde{S}(t)$ is the estimator of $S(t)$ ignoring failure type then it is undefined on a set C. It is possible that $C \subset \cup_{j=1}^{J} C_j$ meaning that $\widehat{S}(t)$ is undefined on a larger region than $\widetilde{S}(t)$.

Hudgens et al. (2001) work around this problem by defining what they call a pseudolikelihood method. The pseudolikelihood estimates (PLE's) are obtained by maximizing the likelihood subject to $\sum_{j=1}^{J} \widetilde{I}_j(t) = 1 - \widetilde{S}(t)$. The resulting $\widetilde{I}_j(t)$'s are all undefined within the same set C and constant outside C. The maximization procedure is again a simple extension of Turnbull's algorithm.

1.4 Analysis of Recurrent Events

A special case of multistate model which has received considerable attention is the recurrent events setting. In this case, events of interest occur repeatedly in individual subjects. This can be modeled in a multistate framework by defining states according to the number of events a subject has experienced, as in Figure 1.7.

Figure 1.7: Recurrent events model.

In such a model, there is a different intensity function $\lambda_k(t|\mathcal{H}(t))$ for transitions out of each state $k = 0, 1, \ldots$. A common assumption is to use *Poisson processes*, where the intensity functions are Markovian and identical for each k, i.e. $\lambda_k(t|\mathcal{H}(t)) = \rho(t)$, $k = 0, 1, \ldots$.

With Poisson processes, the notation can be simplified by only defining counting processes $N_i(t)$ which count the number of events individual i has experienced at time t. The mean number of events, $E\{N_i(t)\}$, can be expressed simply as

$$\mu(t) = E\{N_i(t)\} = \int_0^t \rho(u)\, du.$$

1.4.1 Analysis of Right-Censored Recurrent Events

Suppose individuals are each followed for a period of time τ_i. Let n_i be the number of events individual i experienced, which occurred at times $t_{i1}, t_{i2}, \ldots, t_{in_i}$. Then

the likelihood contribution for individual i becomes

$$L_i = \left\{ \prod_{j=1}^{n_i} \rho(t_{ij}) \right\} e^{-\mu(\tau_i)}.$$

If a parametric model is assumed, then the usual likelihood methods can be used to obtain estimates and standard errors of the model parameters.

If a nonparametric approach is adopted, the mean function can be estimated by the Nelson-Aalen estimate,

$$\widehat{\mu}(t) = \sum_{h:t_h \leq t} \frac{dN_{\cdot}(t_h)}{Y_{\cdot}(t_h)}$$

where $t_1, \ldots, t_H$ are the distinct event times from all individuals, $dN_{\cdot}(t)$ gives the number of events that occurred at time t and $Y_{\cdot}(t)$ gives the number of individuals at risk of events at time t.

If the intensity functions are not identical for all transitions but still Markov, i.e. $\lambda_k(t|\mathcal{H}(t)) = \rho_k(t)$, then another nonparametric estimate of the mean function can be constructed. If the transition probability matrix is estimated using the Aalen-Johansen estimator, then an estimate of the mean function is

$$\widehat{\mu}(t) = \sum_{k=0}^{\infty} k \widehat{p}_{0k}(t)$$

where $\widehat{p}_{0k}(t)$ is the Aalen-Johansen estimate of the prevalence function for having experienced k events.

1.4.2 Analysis of Interval-Censored Recurrent Events

Suppose individual i is only seen at times $b_{i1}, b_{i2}, \ldots, b_{iJ_i}$ and the data consists of counts, n_{ij}, of the number of events in each interval $(b_{i,j-1}, b_{ij}]$, $j = 1, \ldots, J_i$. Under a Poisson model with parametric mean function $\mu(t;\theta)$ then the likelihood function is (Sun 2006)

$$L(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{J_i} \exp\left\{-(\mu(b_{ij};\theta) - \mu(b_{i,j-1};\theta))\right\} (\mu(b_{ij};\theta) - \mu(b_{i,j-1};\theta))^{n_{ij}}.$$

Maximum likelihood estimates are readily available.

A nonparametric approach was introduced in Sun and Kalbfleisch (1993) and further described in Sun and Kalbfleisch (1995). If we let $s_1, s_2, \ldots, s_m$ denote the

distinct observation times, then a nonparametric estimate of $\mu(t)$ is identifiable only at these points. Define $\mu_j = \mu(s_j)$ for $j = 1, \ldots, m$. Note that the μ_j must be nondecreasing. Also define w_j to be the number of individuals observed at s_j and $\bar{n}_j$ to be the average number of events observed at s_j. Consider the simplest case of current status data, where each subject is observed once, and a count of the number of events that have occurred is recorded. A Poisson assumption then leads to the log-likelihood

$$\ell = \sum_{j=1}^{m} w_j \left\{ \bar{n}_j \log \mu_j - \mu_j \right\}.$$

Maximizing this likelihood subject to $\mu_1 \leq \cdots \leq \mu_m$ is equivalent to minimizing

$$\sum_{j=1}^{m} w_j \left\{ \bar{n}_j - \mu_j \right\}^2$$

subject to this constraint. This is the same as carrying out isotonic regression of $\bar{n}_1, \ldots, \bar{n}_m$ with weights $w_1, \ldots, w_m$ (see Barlow et al. 1972).

Sun and Kalbfleisch (1995) recommend using this method in the general case as well, although it is not the MLE. Wellner and Zhang (2000) show how this estimator can be derived using a "pseudo-likelihood" function which ignores the dependency between $N_i(s_j)$ and $N_i(s_k)$ for $j \neq k$. Wellner and Zhang (2000) also show how to construct the nonparametric MLE of the mean function under a Poisson process, as well as establish the consistency of both the MLE and pseudo MLE even if the underlying process is not Poisson.

Thall and Lachin (1988) propose a simpler estimate of the rate function $\lambda(t) = \mu'(t)$. Their idea is to estimate the rate function for each individual as

$$\widehat{\lambda}_i(t) = \sum_{j=1}^{J_i} \frac{n_{ij}}{b_{ij} - b_{i,j-1}} I(t \in (b_{i,j-1}, b_{ij}])$$

and estimate the rate function by the mean of these rate functions,

$$\widehat{\lambda}(t) = \frac{\displaystyle\sum_{i=1}^{n} \widehat{\lambda}_i(t)}{\displaystyle\sum_{i=1}^{n} \sum_{j=1}^{J_i} I(t \in (b_{i,j-1}, b_{ij}])}.$$

Lawless and Zhan (1998) consider a piecewise-constant rate function under a mixed Poisson assumption. They consider models where conditional on a random effect u_i, the rate function is

$$\lambda_i(t | \mathcal{H}_i(t), u_i) = u_i \rho_0(t) e^{x_i' \beta}$$

where $\rho_0(t)$ is the baseline piecewise-constant rate function, x_i is a covariate vector and β is a parameter vector. The random effect distribution is taken to be gamma with mean 1 and variance ϕ.

The parameters may be estimated by an EM algorithm. The complete data log-likelihood can be written as the sum of two terms, $\ell_C = \ell_1(\phi) + \ell_2(\rho, \beta)$ where

$$\ell_1 = -n\left\{\log\Gamma(\phi^{-1}) + \phi^{-1}\log\phi\right\} + \sum_{i=1}^{n}\phi^{-1}(\log u_i - u_i)$$

$$\ell_2 = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\sum_{k=1}^{r} n_{ijk}\log\mu_{ijk} - \sum_{i=1}^{m} u_i\mu_i$$

where n_{ijk} is the unobserved number of events individual i experienced in the interval $A_k \cap B_{ij}$ and $A_k = a_k - a_{k-1}$ (where the a's denote breakpoints for the piecewise-constant rate function) and $B_{ij} = b_{ij} - b_{i,j-1}$ (where the b's denote assessment times). Also, $\mu_{ijk} = \rho_k w_k(i,j)e^{x_i'\beta}$ where ρ_k is the constant rate over A_k and $w_k(i,j)$ gives the time individual i spent in $A_k \cap B_{ij}$.

For the E-step, expectations of n_{ijk}, u_i and $\log u_i$ are required. For details, see Lawless and Zhan (1998). The M-step is straightforward, as after obtaining the expectations, the expected log-likelihood is straightforward to maximize. Lawless and Zhan (1998) also discuss variance estimation for both the mixed Poisson case as well as robust methods which do not require a Poisson assumption.

Thall (1988) considers likelihood based analyses of interval count data with parametric rate functions approximations of the mean function. Empirical Bayes methods are suggested for inference about the random effects. The need to make parametric assumptions about the baseline rate was relaxed in Staniswalis et al. (1997) who described profile likelihood methods (Severini and Wong 1992) based on smoothing techniques for profiling out the baseline mean to permit estimation of the regression coefficients. For inferences about the baseline mean, further smoothing with the imposition of monotonicity is required. Interest here lies in the estimation of the mean function in one sample problems and the baseline mean in multiplicative rate function regression models.

An estimating equations approach is taken in Sun and Wei (2000), while other regression models are considered in Chapter 9 of Sun (2006). Chen et al. (2005) considers multiple types of events.

1.5 Outline of Thesis

1.5.1 Chapter 2: Local Likelihood methods for Interval-Censored Recurrent Event Data

When we have interval-censored recurrent event data, all that is known is counts of the occurrences of events within specified intervals. Data of this type often occurs in studies where subjects are observed only at inspection times, so the exact timing of events is usually unknown, but the number of events since the last inspection time can be determined. Data of this type are sometimes referred to as interval-grouped recurrent events or panel count data (Lawless and Zhan 1998).

Lawless and Zhan (1998) looked at a study where patients develop superficial bladder tumours which are observed only at clinic visits. Thall (1988) and Thall and Lachin (1988) look at a clinical trial where the recurrent event is an occurrence of nausea in patients with gall stones. Another example is given in Sun and Kalbfleisch (1995) where loss of feedwater flow at a nuclear plant is the recurrent event of interest. These data are based on current status observations.

An estimate of the mean function can already be obtained based on the methods of estimating prevalence functions by

$$\widehat{\mu}(t) = \sum_{j=1}^{J} j\widehat{p}_j(t).$$

where $\widehat{p}_j(t)$ is an estimate of the prevalence function for state j, $j = 1, \ldots, J$. Alternative estimates can also be derived by applying the local likelihood methods under a working Poisson or mixed Poisson model.

OBJECTIVE: Local Likelihood Mean Function Estimation

A primary objective for many analyses of interval-censored recurrent event data is estimation of the mean function. The mean function represents a relatively easy quantity to interpret and a natural basis for the evaluation of proposed methods. Local likelihood methods (Betensky et al. 1999, 2002) will be developed for estimation of the rate function, which in turn leads to an estimate of the mean function. The methods will be developed first under a Poisson assumption, then generalized to allow extra-Poisson variation. Covariates will also be considered using multiplicative rate functions of the form $\rho_i(t) = \rho_0(t)e^{x_i'\beta}$.

The performance of established methods for estimating mean functions (Lawless and Zhan 1998) as well as new methods will be investigated in terms of robustness and efficiency. Non-Markov processes will be considered by using renewal processes for the gap time distributions. The comparisons will be on the basis of bias and efficiency for estimation of mean functions over time.

The methods will be applied to data from the University of Toronto Psoriatic Arthritis Clinic, where we will consider a damaged joint as an event in a recurrent event process.

1.5.2 Chapter 3: Estimating Equations for Clustered Current Status Data

An extreme form of interval censoring occurs when individuals are observed at only one time point. At such an assessment, all that can be ascertained is whether or not the event of interest has occurred, hence the name *current status data*. If subjects are grouped into clusters, say according to centres in a clinical trial or by family for example, then subjects within a cluster may not have independent responses. It is of interest to estimate the survival distribution as well as the effects of covariates in multiplicative models.

OBJECTIVE: Estimate survival function parameters, covariate effects and association parameters with clustered current status data

Current status data can be viewed as being similar to binary data since at an individual's assessment we have a binary response of whether the event has occurred or not. In fact, current status data under a Weibull model can be viewed as a binary GLM with complementary log-log link. This similarity to binary data allows for the use of generalized estimating equations (Liang and Zeger 1986) techniques. Typically with binary GEE's the association structure is modelled by parameterizing the association between the binary responses. However with current status data, the association can be imposed on the failure times themselves. This approach will be adopted with the dependence between failure times induced by a copula. The methods will be developed under a working independence model, which corrects naive variance estimates to account for the dependence, and for first and second order GEE's which make use of the copula based dependence structure.

Performance will be assessed using both Weibull and piecewise constant hazard functions. In addition, the asymptotic efficiency of the first order GEE will be

compared with that of the second order GEE, since for large cluster sizes, GEE2 can be computationally intensive.

The methods will be applied to a series of studies of patients undergoing orthopedic surgery. Individuals were tested for the presence of antibodies that indicate a reaction to a drug administered for recovery after surgery. Clusters are defined according to the centre an individual is seen. The methods will also be applied to the University of Toronto Psoriatic Arthritis Clinic Data, where we restrict attention to the joints of the hands. Each individual represents a cluster of hand joints.

1.5.3 Chapter 4: Multistate Models With Interval Censoring

With interval-censored multistate data, the intervals in which transitions occurred and the states occupied at the assessment times are all that is known. This is usually as a result of studies where subjects are only observed at inspection times, such as the Toronto Psoriatic Arthritis Clinic. Other examples of such data can be found in Frydman (1992) which looked at a study where individuals received blood transfusions and could fall into one of three states (non-infected, HIV positive, AIDS) in a progressive three state model. In another setting, Frydman (1995) uses an illness-death model to look at a study of Danish diabetics who are either alive without complications, alive with complications or dead. A different kind of multistate model was considered in Kalbfleisch and Lawless (1985), which looked at children and smoking. The states were defined as "never smoked" (state 0), "currently smokes" (state 1) and "quit smoking" (state 2). In this model, once a subject leaves state 0 they will never return, but it is possible for subjects to go back and forth between states 1 and 2. Another example involving an AIDS study is Gentleman et al. (1994) where subjects can be in a number of intermediary states before entering a state representing AIDS.

OBJECTIVE 1: "Pepe" Estimation of Prevalence Functions

A primary objective here is to develop estimators of the prevalence function, or state occupancy probability, $P(Z(t) = k|Z(0) = 0)$ for multistate models that are robust (*i.e.* applicable for a variety of underlying models). The approach will be to develop estimates based on the fact that $P(Z(t) = k|Z(0) = 0) = P(T_{k+1} > t) - P(T_k > t)$ where T_k is the time of entry to state k, following Pepe et al. (1991) who considered this approach for right-censored data. Estimation of the survivor functions can

be accomplished using a variety of methods including *i)* standard nonparametric estimation (Turnbull 1976), *ii)* piecewise-constant models (Lindsey and Ryan 1998) or, *iii)* Local EM estimation (Betensky et al. 1999). The advantages of the second and third approaches include the fact that the resulting estimates are defined on the positive real line.

OBJECTIVE 2: Estimate Covariate Effects in Markov Models

Patients with psoriatic arthritis are also at risk of a back condition known as spondylitis. A multistate model can be formulated where individuals move among states defined according to whether or not they have spondylitis, and whether or not they are still alive. A number of covariates may affect transitions between these states, so it is of interest to determine which covariates affect spondylitis and which affect mortality.

Applications

The Psoriatic Arthritis Clinic data will be used to illustrate the methods. The state occupancy probabilities will be estimated in a 3-state model with states alive with PsA, alive with spondylitis, and dead. A 5-state model will also be considered which separates the spondylitis state according to whether the left, right, or both sides of the body are affected. Markov models with multiplicative covariate effects will be fit to assess the effects of covariates on the transition intensities.

Methods for estimation of state occupancy probabilities will also be applied to the bivariate interval-censored data on viral shedding in HIV patients with CMV infection, described in Betensky and Finkelstein (1999) and Goggins and Finkelstein (2000).

Chapter 2

Local Likelihood for Interval-Censored Recurrent Event Data

2.1 Introduction and Overview

Figure 2.1: Recurrent events model.

We now focus on the simple class of multistate models often used to represent the state space for recurrent event data; see Figure 2.1. We focus on the problem of estimating the mean function, $\mu(t)$, of a point process, but also consider use of multiplicative models, in which interest lies in both the baseline mean function and covariate effects.

As noted in Section 1.4, Sun and Kalbfleisch (1995) discuss nonparametric estimation of the mean function with interval-censored counts. In particular, they note that a nonparametric estimate of the mean function is defined only at the observation times. They propose an estimator defined such that the mean function at the jth distinct inspection time, $\widehat{\mu}_j$ is the mean number of events observed at that inspection time, and if the $\widehat{\mu}_j$'s are not nondecreasing then the estimates are made so using ideas of isotonic regression. For Poisson current status data this gives the NPMLE, but Wellner and Zhang (2000) show how the estimator can be

viewed as a "pseudo-likelihood" estimator more generally and prove its consistency in the general case of interval-censored data and without the Poisson assumption. Lawless and Zhan (1998) describe an EM algorithm for estimating covariate effects in a mixed-Poisson recurrent event process.

The state occupancy probability, or prevalence function gives the probability being in a given state at time t, so in the context of recurrent event data the prevalence function for state k gives the probability of having experienced exactly k events by time t. This leads to a natural estimate of the mean function given by

$$\widehat{\mu}(t) = \sum_{j=1}^{J} j\widehat{p}_j(t)$$

where $\widehat{p}_j(t)$ is the prevalence function estimate for state j, $j = 1, \ldots, J$.

2.2 Estimation of the Mean Function for Right-Censored Count Data

2.2.1 The One Sample Problem

Suppose events are generated according to a Poisson process with rate function

$$\lambda_i(t|\mathcal{H}_i(t)) = \rho(t).$$

Let $T_{i1}, \ldots, T_{in_i}$ denote the times of the n_i events experienced by individual i over the period of observation from 0 to τ_i. The likelihood contribution from individual i is

$$L_i = \left\{ \prod_{j=1}^{n_i} \rho(T_{ij}) \right\} \exp\left(- \int_0^{\tau_i} \rho(u)\, du \right). \tag{2.1}$$

Let the rate function around t be approximated by a function dependent on a parameter vector $\alpha_t = (\alpha_{0t}, \ldots, \alpha_{pt})'$. For example, one could adopt $\log \rho(u) = \alpha_{0t} + \alpha_{1t}(u-t) + \cdots + \alpha_{pt}(u-t)^p$, and then the local log-likelihood at t becomes

$$\ell(\alpha_t; t) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} [K_b(T_{ij} - t) \log \rho(T_{ij}; \alpha_t)] - \int_0^{\tau_i} K_b(u-t)\rho(u; \alpha_t)\, du \right\} \tag{2.2}$$

where $K_b(u) = K(u/b)/b$ and $K(u)$ is a kernel function. The kernel function is used to weight the contributions to the local log-likelihood function at t. Observations

closer to t will be given more weight according to the kernel function while those further away will be given less weight, or even no weight at all. The kernel causes the local log-likelihood to be affected only by observations within b units of t.

Estimates of α_t are obtained by solving the $p+1$ estimating equations, $U_{tp}(\alpha_t) = 0$, where U_{tp} is obtained by differentiating $\ell(\alpha_t; t)$ with respect to the pth component of α_t.

Variance estimates can be obtained in a similar manner as in Betensky et al. (1999) using the theory of estimating functions. If the rate function is being estimated at M equally spaced grid points, $t_1, \ldots, t_M$, then let U_i be the vector of contributions from individual i to the set of $M(p+1)$ estimating functions and let α be an $M(p+1) \times 1$ vector consisting of all the elements of α_g, $g = 1, \ldots, M$. Let $U = \sum_{i=1}^{n} U_i$. The covariance matrix of $\widehat{\alpha}$ is estimated by

$$\left\{ \left(\frac{\partial U}{\partial \alpha'} \right)^{-1} \right\}' \left(\sum_{i=1}^{n} U_i U_i' \right) \left(\frac{\partial U}{\partial \alpha'} \right)^{-1}$$

evaluated at $\widehat{\alpha}$. For example, under a locally constant approximation,

$$U_{ig} = \sum_{j=1}^{n_i} K_b(T_{ij} - t_g) - e^{\alpha_g} \int_0^{\tau_i} K_b(u - t_g)\, du$$

and the g estimating functions are orthogonal, meaning $-\partial U / \partial \alpha'$ is a diagonal matrix with entries

$$e^{\alpha_g} \sum_{i=1}^{n} \int_0^{\tau_i} K_b(u - t_g)\, du, \quad g = 1, \ldots, M.$$

2.2.2 Regression Models

With covariates, multiplicative models of the form $\rho_i(t) = \rho_0(t) \exp(x_i' \beta)$ are useful. Estimation of β using local likelihood can be carried out using profile likelihood. The idea is to use local likelihood to estimate ρ_0 for a fixed value of β and plug this estimate into the log-likelihood

$$\ell(\rho_0(\cdot), \beta) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} \left[\log \widehat{\rho}_0(T_{ij}; \beta) + x_i' \beta \right] - e^{x_i' \beta} \widehat{\mu}_i(\tau_i; \beta) \right\}.$$

Estimates of ρ_0 and μ_0 are obtained by maximizing the local log-likelihood

$$\ell(\alpha_t; t) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} K_b(T_{ij} - t) \left[\log \rho(T_{ij}; \alpha_t) + x_i' \beta \right] \right.$$
$$\left. - e^{x_i' \beta} \int_0^{\tau_i} K_b(u - t) \rho(u; \alpha_t)\, du \right\} \quad (2.3)$$

for α_t. The profile log-likelihood, $\ell(\widehat{\rho}_0(\cdot;\beta),\beta_0)$ is then maximized to obtain $\widehat{\beta}$. In the one-parameter case, confidence intervals for β can be constructed by finding the β that satisfy $2[\ell(\widehat{\rho}_0(\cdot;\widehat{\beta}),\widehat{\beta}) - \ell(\widehat{\rho}_0(\cdot;\beta),\beta)] \leq \chi^2_{1,\alpha}$. In principle this can be extended to the multiparameter case.

In order to justify the use of the χ^2 distribution, consider the following heuristic argument. We will restrict attention to the case of a locally constant approximation to the log of the rate function. Higher order polynomials may be handled similarly. The estimate $\widehat{\rho}_0(t;\beta)$ is given by $e^{\widehat{\alpha}_{0t}}$. The derivative of (2.3) with respect to α_{0t} is given by

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_{0t}} &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} K_b(T_{ij}-t) - e^{\alpha_{0t}} \sum_{i=1}^{n}\int_0^{\tau_i} K_b(u-t)e^{x_i'\beta}\,du \\ &= \sum_{i=1}^{n}\int_0^{\tau_i} K_b(u-t)\,dN_i(u) - e^{\alpha_{0t}} \sum_{i=1}^{n}\int_0^{\tau_i} K_b(u-t)e^{x_i'\beta}\,du.\end{aligned}$$

Let $\widetilde{\alpha}_{0t}$ denote the solution to $E\{\partial\ell/\partial\alpha_{0t}\} = 0$. We will take the expectation of individual i's terms first given τ_i. We have

$$E\left\{\int_0^{\tau_i} K_b(u-t)\,dN_i(u)\,\middle|\,\tau_i\right\} = \int_0^{\tau_i} K_b(u-t)\rho_0(u)e^{x_i'\beta}\,du. \tag{2.4}$$

For $\tau_i > t$, and b sufficiently small, we can write this as

$$\int_{-1}^{1} K(u)\rho_0(t+bu)e^{x_i'\beta}\,du \to \rho_0(t)e^{x_i'\beta} \qquad \text{as } b \to 0.$$

Similarly, for $\tau_i < t$, (2.4) goes to 0 as $b \to 0$ and for $\tau_i = t$, (2.4) goes to $0.5\rho_0(t)e^{x_i'\beta}$ as $b \to 0$. It also follows that

$$\int_0^{\tau_i} K_b(u-t)e^{x_i'\beta}\,du$$

converges to $e^{x_i'\beta}$, 0, and $0.5e^{x_i'\beta}$ for $\tau_i > t$, $\tau_i < t$ and $\tau_i = t$ respectively, as $b \to 0$. Hence as $b \to 0$,

$$e^{\widetilde{\alpha}_{0t}} = \frac{\sum_{i=1}^{n} W_i\rho_0(t)}{\sum_{i=1}^{n} W_i} = \rho_0(t)$$

where $W_i = e^{x_i'\beta}\{P(\tau_i > t) + 0.5P(\tau_i = t)\}$. In the language of Severini and Wong (1992), the profile local likelihood estimate of $\rho_0(t)$ given β consistently estimates the same "least favourable curve" as the standard semiparametric approach (Cook

and Lawless 2007, Chapter 3), hence the profile likelihood has the usual asymptotic properties.

The Poisson model is restrictive in the sense that $\text{var}(N_i(t)) = \mu_i(t)$. If this is not reasonable, a simple alternative is to consider a variance function of the form $\text{var}(N_i(t)) = \nu\mu_i(t)$. Such a model is often used when there is over-dispersion in a Poisson GLM (McCullagh and Nelder 1989). In the Poisson GLM case, the log-likelihood is multiplied by ν^{-1} so the resulting estimate of β is unchanged but the variability of β is inflated by ν. The parameter ν can be estimated by the method of moments,

$$\widehat{\nu} = \frac{1}{n}\sum_{i=1}^{n}\frac{(n_i - \widehat{\mu}_i(\tau_i))^2}{\widehat{\mu}_i(\tau_i)}. \tag{2.5}$$

The profile likelihood interval then becomes the set of β which satisfy

$$2[\ell(\widehat{\rho}_0(\cdot;\widehat{\beta}),\widehat{\beta}) - \ell(\widehat{\rho}_0(\cdot;\beta),\beta)] \leq \widehat{\nu}\chi^2_{1,\alpha}.$$

Another form of variance function which is frequently used arises from a mixed-Poisson process formulation and is given by $\text{var}(N_i(t)) = \mu_i(t) + \phi\mu_i^2(t)$. In this situation we suggest an estimate of $\rho_0(t)$ for given β using the Poisson model and an estimate of ϕ is then obtained by solving

$$\sum_{i=1}^{n}\left\{\frac{(n_i - \widehat{\mu}_i)^2 - \widehat{\mu}_i(1+\phi\widehat{\mu}_i)}{(1+\phi\widehat{\mu}_i)^2}\right\} = 0$$

where $\widehat{\mu}_i = \widehat{\mu}(\tau_i;\beta)$ (Dean 1991). Under the assumption that the random effect follows a gamma distribution, an estimate of β is obtained by maximizing the profile likelihood

$$\begin{aligned}\ell(\beta,\widehat{\rho}_0(\cdot;\beta),\widehat{\phi}(\beta)) = \sum_{i=1}^{n}\Bigg\{&\sum_{j=1}^{n_i}\log\widehat{\rho}_0(T_{ij}) + n_i(\log\widehat{\phi} + x_i'\beta) \\ &-(n_i + \widehat{\phi}^{-1})\log(1+\widehat{\phi}\widehat{\mu}_i(\tau_i)) + \log\Gamma(n_i+\widehat{\phi}^{-1}) - \log\Gamma(\widehat{\phi}^{-1})\Bigg\}. \end{aligned}\tag{2.6}$$

Confidence intervals are obtained by finding the β that satisfy $2[\ell(\widehat{\beta},\widehat{\rho}_0(\cdot;\widehat{\beta}),\widehat{\phi}(\widehat{\beta})) - \ell(\beta,\widehat{\rho}_0(\cdot;\beta),\widehat{\phi}(\beta))] \leq \chi^2_{1,\alpha}$.

2.3 Estimation of the Mean Function for Interval-Censored Recurrent Event Data

2.3.1 The One Sample Problem

Suppose individual i is only seen at inspection times $b_{i1}, \ldots, b_{im_i}$. Here we assume the inspection process satisfies the conditions in Grüger et al. (1991).

We consider the problem of developing a local likelihood EM algorithm to estimate the rate under a Poisson model in a one-sample problem. This would be in a similar spirit to the EM approach of Lawless and Zhan (1998), but one would get a smooth estimate of the rate $\rho(t)$, in the spirit of Betensky et al. (1999).

Let $T_{i1}, \ldots, T_{in_i}$ denote the times of the n_i events experienced by individual i over the period of observation from 0 to τ_i. Let $(L_{ij}, R_{ij}]$ be the censoring interval for T_{ij}. These censoring intervals are created as a result of a continuous time recurrent event process only being observed at periodic assessment times.

Under a Poisson process the complete data likelihood contribution from individual i is given by (2.1). Let the baseline rate function around t be approximated by a function dependent on a parameter vector $\alpha_t = (\alpha_{0t}, \ldots, \alpha_{pt})'$ so the local log-likelihood at t is given by (2.2).

The parameters can now be estimated by an EM algorithm which incorporates smoothing. A grid of equally spaced points at which the baseline hazard will be estimated, $t_1, \ldots, t_M$, must be defined. Let $\widehat{\rho}$ denote the estimate of ρ at each of the grid points, so the gth element of $\widehat{\rho}$ depends on $\widehat{\alpha}_{t_g}$. Define

$$Q^{(t)}(\alpha_t; \widehat{\rho}^{(r-1)}) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} \left[E[K_b(T_{ij} - t) | \widehat{\rho}^{(r-1)}, L_{ij}, R_{ij}] \log \rho(T_{ij}; \alpha_t) \right] - \int_0^{\tau_i} K_b(u - t) \rho(u; \alpha_t)\, du \right\}$$

where parameters with a superscript $(r-1)$ indicate the parameter estimate at the $(r-1)$th iteration.

The expectations involving the T_{ij} must be evaluated numerically, for example using the trapezoid rule. To obtain the conditional density of the T_{ij} given the data, it is useful to transform the times by defining $S_{ij} = \mu(T_{ij})$. The S_{ij} then follow a homogeneous Poisson process with rate 1. Due to the independent increments property of Poisson processes, the conditional density of S_{ij}, given it occurred in the

interval $(\mu(b_{i,k-1}), \mu(b_{ik})]$, does not depend on any times outside that interval. The times within that interval are uniformly distributed over $(\mu(b_{i,k-1}), \mu(b_{ik})]$, hence the density of T_{ij} given it occurred in $(b_{i,k-1}, b_{ik}]$ has the form $\rho(t_{ij})/(\mu(b_{ik}) - \mu(b_{i,k-1}))$, $b_{i,k-1} < t_{ij} \leq b_{ik}$. The resulting expectation is then given by

$$\begin{aligned} E[K_b(T_{ij} - t)|\widehat{\rho}^{(r-1)}, L_{ij}, R_{ij}] &= \sum_{g: L_{ij} \leq t_g \leq R_{ij}} K_b(t_g - t) \frac{\widehat{\rho}^{(r-1)}(t_g)}{\widehat{\mu}^{(r-1)}(R_{ij}) - \widehat{\mu}^{(r-1)}(L_{ij})} \\ &= \frac{\sum_{g: L_{ij} \leq t_g \leq R_{ij}} K_b(t_g - t) \widehat{\rho}^{(r-1)}(t_g)}{\sum_{g: L_{ij} \leq t_g \leq R_{ij}} \widehat{\rho}^{(r-1)}(t_g)}. \end{aligned}$$

The M-step involves maximizing $Q^{(t)}(\alpha_t; \widehat{\rho}^{(r-1)})$ to obtain $\alpha_t^{(r)}$ at each grid point using, for example, Newton-Raphson. The EM algorithm is continued until the difference between parameter estimates obtained in successive steps becomes negligible.

Variance estimates for $\widehat{\alpha}$ can be obtained as described in the previous section. If we let $\widehat{\rho}$ and $\widehat{\mu}$ denote the vector of estimates of $\rho(t)$ and $\mu(t)$ respectively at each of the grid points then $\widehat{\mu} = \Delta W \widehat{\rho}$, where the jth row of W is comprised of the weights used in the numerical integration to obtain $\widehat{\mu}(t_j)$ and Δ is the grid size. Variance estimates for $\widehat{\mu}$ can be obtained by the delta method, giving

$$\operatorname{var}(\widehat{\mu}) = \Delta^2 \left(W \operatorname{diag} \widehat{\rho}\right) \operatorname{var}(\widehat{\alpha}) \left(W \operatorname{diag} \widehat{\rho}\right)'.$$

Alternatively, resampling techniques such as the bootstrap may also be considered. Extensions to deal with mixed-Poisson processes are straightforward.

2.3.2 Regression Models

When there are covariates of interest, multiplicative models can be used. Estimation of $\rho_0(\cdot)$ and β can be done using an approach analogous to the profile likelihood as described for the right-censored case of Section 2.2.2. With interval censoring, the profile log-likelihood is

$$\ell(\rho_0(\cdot), \beta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ n_{ij} (\log \mu_{ij} + x_i' \beta) - e^{x_i' \beta} \mu_{ij} \right\},$$

where n_{ij} is the number of events in $(b_{i,j-1}, b_{ij}]$ and $\mu_{ij} = \int_{b_{i,j-1}}^{b_{ij}} \rho_0(u)\, du$.

In order to obtain $\widehat{\rho}(\cdot;\beta)$ an EM algorithm must be used. Here we employ an EM algorithm to estimate $\rho(\cdot)$ for fixed β as follows.

The complete data likelihood contribution from individual i is

$$L_i = \left\{ \prod_{j=1}^{n_i} \rho_0(T_{ij}) \exp(x_i'\beta) \right\} \exp\left(- \int_0^{\tau_i} \rho_0(u) \exp(x_i'\beta)\, du \right)$$

and the log-likelihood is,

$$\ell_C(\rho_0, \beta) = \sum_{i=1}^{n} \left[\sum_{j=1}^{n_i} [\log \rho_0(T_{ij}) + x_i'\beta] - e^{x_i'\beta} \int_0^{\tau_i} \rho_0(u)\, du \right].$$

Again, we let the baseline rate function around t be approximated by a function dependent on a parameter vector α_t (*e.g.* $\log \rho_0(u) = \alpha_{0t} + \alpha_{1t}(u-t) + \cdots + \alpha_{pt}(u-t)^p$), then the local log likelihood at t becomes

$$\ell_C(\alpha_t, \beta) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} K_b(T_{ij} - t)[\log \rho_0(T_{ij}; \alpha_t) + x_i'\beta] \right.$$
$$\left. - e^{x_i'\beta} \int_0^{\tau_i} K_b(u-t)\rho_0(u;\alpha_t)\, du \right\}$$

We let $\widehat{\rho}_0$ denote the estimate of ρ_0 at each of the grid points, so the gth element of $\widehat{\rho}_0$ depends on $\widehat{\alpha}_{t_g}$. Furthermore we let

$$Q^{(t)}(\alpha_t, \beta; \widehat{\rho}_0^{(r-1)}, \widehat{\beta}^{(r-1)}) =$$
$$\sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} E[K_b(T_{ij} - t) | \widehat{\rho}_0^{(r-1)}, \widehat{\beta}^{(r-1)}, L_{ij}, R_{ij}][\log \rho_0(T_{ij}; \alpha_t) + x_i'\beta] \right.$$
$$\left. - e^{x_i'\beta} \int_0^{\tau_i} K_b(u-t)\rho_0(u;\alpha_t)\, du \right\}$$

where parameters with a superscript $(r-1)$ indicate the parameter estimates at the $(r-1)$th iteration. Again, the expectations involving the T_{ij} must be evaluated numerically.

The M-step is to maximize $Q^{(t)}(\alpha_t, \beta; \widehat{\rho}_0^{(r-1)})$ to obtain $\alpha_t^{(r)}$ at each grid point. The EM algorithm is continued until the difference between successive parameter estimates becomes negligible.

The "profile" log-likelihood, $\ell(\widehat{\rho}_0(\cdot;\beta), \beta)$ is then maximized to obtain $\widehat{\beta}$. In the one-parameter case, confidence intervals for β can be constructed by finding the β that satisfy $2[\ell(\widehat{\rho}_0(\cdot;\widehat{\beta}), \widehat{\beta}) - \ell(\widehat{\rho}_0(\cdot;\beta), \beta)] \leq \chi^2_{1,\alpha}$. A justification of the

distribution of the profile likelihood ratio statistic is much more difficult in the interval-censored case, and its validity is examined via simulations in Section 2.4. The difficulty arises since there is no explicit expression for the least favourable curve (Staniswalis et al. 1997). Resampling techniques such as the bootstrap may also be considered. Variance functions of the form $\text{var}(N_i(t)) = \nu\mu_i(t)$ can be handled in a similar manner as for right-censored data. Estimates are computed as for the Poisson case and the estimate of ν then is

$$\widehat{\nu} = \frac{1}{n}\sum_{i=1}^{n}\frac{(n_i - \widehat{\mu}_i(\tau_i))^2}{\widehat{\mu}_i(\tau_i)}. \tag{2.7}$$

For variances of the mixed-Poisson form $\text{var}(N_i(t)) = \mu_i(t) + \phi\mu_i^2(t)$ the "profile" likelihood is calculated as follows. First, a Poisson model is fitted using the EM algorithm to estimate $\rho_0(t)$ for fixed β, $\widehat{\rho}_0(\cdot;\beta)$. Secondly, a moment estimate of ϕ, is calculated by solving

$$\sum_{i=1}^{n}\left\{\frac{(n_i - \widehat{\mu}_i)^2 - \widehat{\mu}_i(1+\phi\widehat{\mu}_i)}{(1+\phi\widehat{\mu}_i)^2}\right\} = 0$$

where $\widehat{\mu}_i = \int_0^{\tau_i}\widehat{\rho}_0(u;\beta)\,du$. These estimates are then used to compute the profile negative binomial likelihood,

$$\ell(\beta, \widehat{\rho}_0(\cdot;\beta), \widehat{\phi}(\beta)) = \sum_{i=1}^{n}\left\{\sum_{i=1}^{m_i} n_{ij}(\log\mu_{ij} + x_i'\beta) + n_i\log\phi + \log\Gamma(n_i + \phi^{-1}) \right.$$
$$\left. - \log\Gamma(\phi^{-1}) - (n_i + \phi^{-1})\log(1 + \phi e^{x_i'\beta}\mu_{ij})\right\}.$$

The profile likelihood is then maximized to obtain $\widehat{\beta}$ and confidence intervals are obtained by finding the β that satisfy $2[\ell(\widehat{\beta}, \widehat{\rho}_0(\cdot;\widehat{\beta}), \widehat{\phi}(\widehat{\beta})) - \ell(\beta, \widehat{\rho}_0(\cdot;\beta), \widehat{\phi}(\beta))] \leq \chi^2_{1,\alpha}$.

2.4 Simulation Studies

Simulation studies were conducted to assess the performance of the Poisson local likelihood estimator. One thousand samples of 500 subjects were generated over the interval $(0, 1]$ such that the mean number of events experienced by an individual was 4. The underlying processes were Poisson, mixed-Poisson and renewal processes. For the Poisson processes, the mean function took the form $(\theta t)^\gamma$. The values of γ were taken to be 1, 0.75 and 1.2 in order to examine the effects of a trend in the intensity. The same form was used as the baseline mean function conditional on

the random effect in the mixed-Poisson case, while the random effect was taken to be gamma distributed with mean 1 and variance 0.25. The interarrival distribution for the renewal process was taken to be gamma with shape parameter 2. The scale parameters in all cases were chosen such that $\mu(1) = 4$. The assessment times were generated according to a homogeneous Poisson process with mean number of visits equal to 5 or 10.

For comparison, the Poisson local likelihood estimator was compared with estimates obtained using a piecewise constant rate function with 3 or 6 pieces. In the case of mean function estimation, comparisons can also be made with a simple estimator of the mean function based on state occupancy probabilities in a multistate model. If we consider a recurrent event process as a multistate model with states defined by the number of events an individual has experienced, as in Figure 2.1, then estimates of the state occupancy probabilities can be obtained as in Pepe et al. (1991). If $Z(t)$ indicates the state an individual is in at time t then $P(Z(t) = k) = P(T_{k+1} > t) - P(T_k > t)$ where T_k denotes the time of the kth event and $T_0 = 0$. The marginal survivor functions can be estimated separately using local likelihood as in Betensky et al. (1999) and an estimate of the state occupancy probabilities can be obtained by taking the appropriate difference. An estimate of the mean function is obtained by $\widehat{\mu}(t) = \sum_{j=1}^{J} j\widehat{p}_j(t)$ where $\widehat{p}_j(t)$ is the prevalence function estimate for state j, $j = 1, \ldots, J$.

The estimators were examined in terms of bias and MSE. A locally constant polynomial, Epanechnikov kernel and 20% "nearest neighbours" bandwidth (Loader 1999) were used in all cases. The results of the simulations are displayed in Tables 2.1 and 2.2. Figures 2.2 and 2.3 graphically display the performance of the estimators.

Examination of Tables 2.1 and 2.2 shows that the Poisson-based local likelihood estimator outperforms the robust Pepe estimator in all situations. It can also be seen that the MSE and bias of the local likelihood estimator are close to that of the piecewise constant estimators. The local likelihood estimator tends to perform slightly better than the 3-piece model, while the 6-piece model tends to perform slightly better than the local likelihood estimator. Figure 2.2 displays the bias of the estimators. In all cases, the bias of the Pepe estimator is largest. All methods suffer early bias when the model is not time homogeneous. The piecewise constant methods have large bias early on, which gets smaller with time, while the local likelihood estimator has smaller initial bias, but there is a slight amount of bias that persists over time. Figure 2.3 displays the performance of the local likelihood estimator versus the piecewise constant estimators of the rate function.

Table 2.1: Integrated mean squared error for the local likelihood methods based on Poisson models, Pepe approach and piecewise constant models under consideration based on 1000 simulated datasets of 500 subjects.

		Integrated MSE							
		5 Visits				10 Visits			
		LL–Poisson	LL–Pepe	PW3	PW6	LL–Poisson	LL–Pepe	PW3	PW6
Poisson	$\gamma = 1$	0.0048	0.0083	0.0044	0.0049	0.0042	0.0110	0.0039	0.0042
	$\gamma = 0.75$	0.0062	0.0086	0.0090	0.0058	0.0060	0.0132	0.0095	0.0057
	$\gamma = 1.2$	0.0045	0.0084	0.0048	0.0047	0.0040	0.0101	0.0043	0.0039
Mixed Poisson	$\gamma = 1$	0.0068	0.0108	0.0065	0.0070	0.0069	0.0156	0.0066	0.0069
	$\gamma = 0.75$	0.0101	0.0134	0.0128	0.0096	0.0094	0.0218	0.0126	0.0087
	$\gamma = 1.2$	0.0069	0.0108	0.0070	0.0070	0.0065	0.0144	0.0068	0.0064
Renewal		0.0033	0.0074	0.0045	0.0030	0.0034	0.0078	0.0044	0.0027

Table 2.2: Integrated absolute bias for the local likelihood methods based on Poisson models, Pepe approach and piecewise constant models under consideration based on 1000 simulated datasets of 500 subjects.

		Integrated Absolute Bias							
		5 Visits				10 Visits			
		LL–Poisson	LL–Pepe	PW3	PW6	LL–Poisson	LL–Pepe	PW3	PW6
Poisson	$\gamma = 1$	0.0115	0.0460	0.0011	0.0009	0.0103	0.0690	0.0022	0.0022
	$\gamma = 0.75$	0.0343	0.0534	0.0413	0.0137	0.0301	0.0844	0.0411	0.0156
	$\gamma = 1.2$	0.0080	0.0467	0.0165	0.0050	0.0107	0.0649	0.0169	0.0046
Mixed Poisson	$\gamma = 1$	0.0120	0.0502	0.0009	0.0010	0.0101	0.0784	0.0044	0.0043
	$\gamma = 0.75$	0.0341	0.0617	0.0413	0.0138	0.0388	0.1095	0.0418	0.0148
	$\gamma = 1.2$	0.0102	0.0494	0.0168	0.0056	0.0102	0.0740	0.0168	0.0045
Renewal		0.0222	0.0530	0.0300	0.0105	0.0308	0.0629	0.0283	0.0089

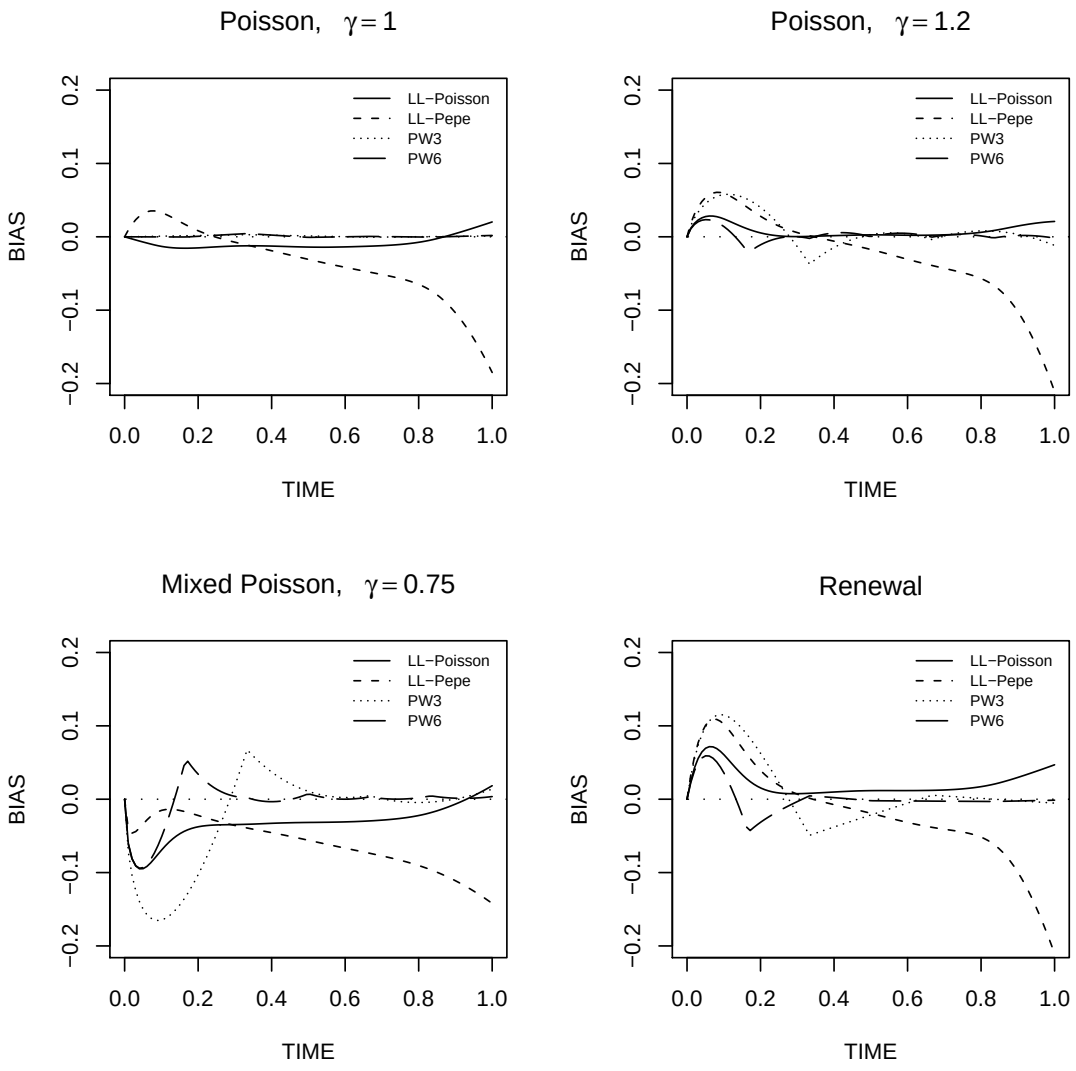

Figure 2.2: Empirical bias of the Poisson local likelihood and Pepe local likelihood estimates as well as piecewise constant estimates (3 piece–PW3, 6 piece–PW6) of the mean function for interval-censored recurrent event data; visits were Poisson distributed with a mean of 5 visits.

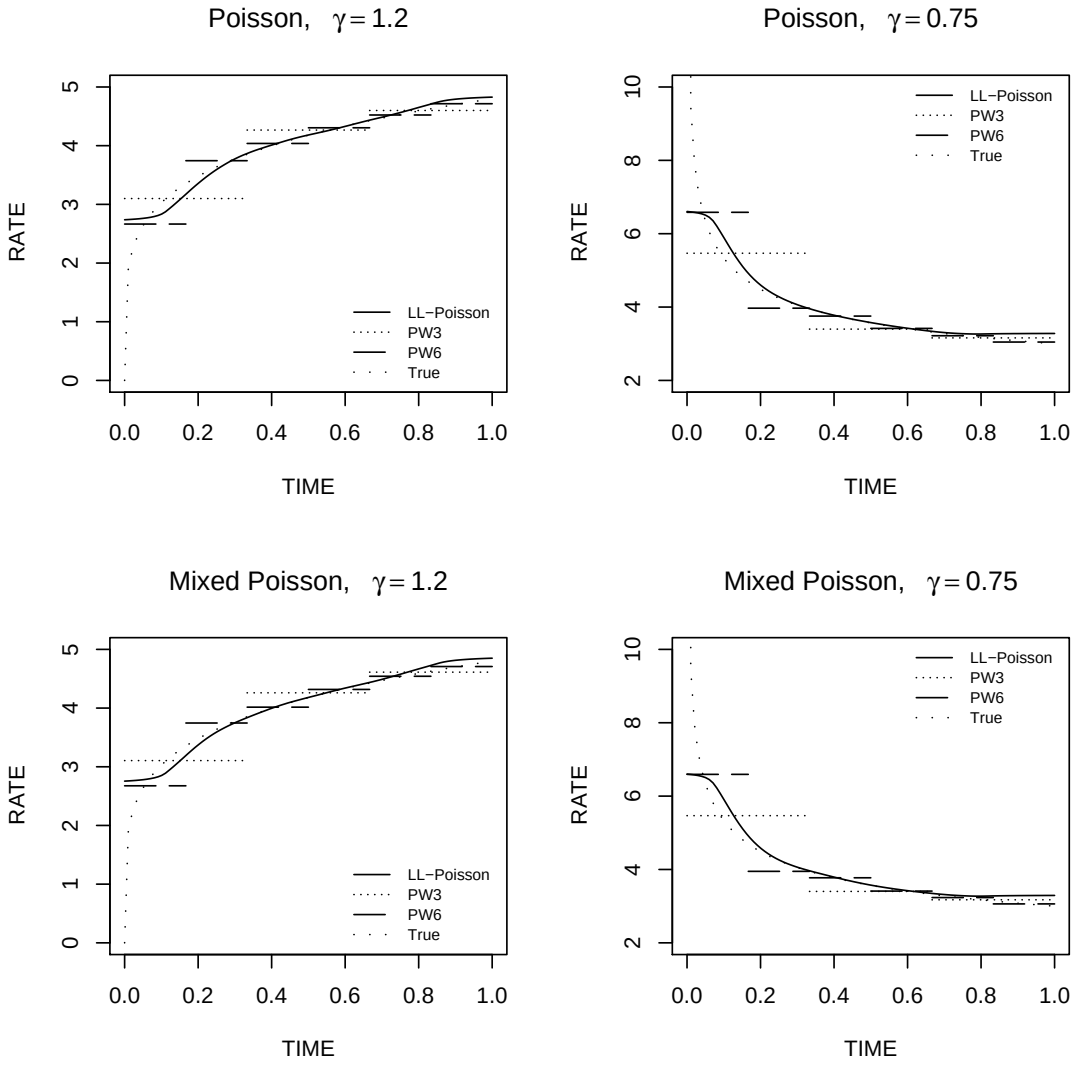

Figure 2.3: Empirical bias of the Poisson local likelihood estimate as well as piecewise constant estimates (3 piece–PW3, 6 piece–PW6) of the rate function for interval-censored recurrent event data based on 1000 simulated datasets of 500 subjects; visits were Poisson distributed with a mean of 5 visits.

The local likelihood estimate tracks the true rate function very closely except very early on near the boundary at zero. It is clear from this figure that the local likelihood method provides a smooth estimate as compared with the piecewise constant estimates.

Simulations were also conducted to assess the performance of the estimators of a regression coefficient. This was done for the right-censored case. The true value of the regression coefficient was set to $\beta = \log 0.75 = -0.28768$ with covariate values taking the values 1 and 0 with equal probabilities. Baseline rate functions considered were the same as for the previous case. Poisson and mixed-Poisson models were used to generate the data with an average of two or four events over the interval $(0, 1]$. For the local likelihood methods, the extra-Poisson procedure involving estimation of ν was used when the data were generated according to a mixed-Poisson process. The results are displayed in Table 2.3. For comparison, results obtained using a piecewise constant rate function (with four or ten pieces) and robust variance estimates are also displayed. The biases are all small relative to their standard errors. The coverage probabilities are all close to the nominal level of 0.95.

The validity of the chi-square approximation to the distribution of the profile likelihood ratio statistic was investigated for the interval-censored case. One thousand datasets composed of 500 subjects were simulated as before. Poisson local likelihood methods with nearest-neighbours bandwidths of 0.2 and 0.6 were used when the data follow a Poisson process, and the method based on (2.7) was used when the data follow a mixed-Poisson process. The mean function at the end of follow-up was taken to be 4 and the random effect was taken to be gamma with mean 1 and variance $\phi = 0.25$. This corresponds to a doubling of the variance relative to a Poisson process by the end of follow-up. Figures 2.4 and 2.5 show QQ-plots of the simulated quantiles of the profile likelihood ratio statistic versus the theoretical quantiles of the χ^2_1 distribution along with 95% pointwise confidence intervals. The QQ-plots were obtained using the `qq.plot` function in the R library `car` (Fox 2007). There appears to be good agreement between the empirical and chi-square (1 df) quantiles for most configurations since the empirical quantiles are within the "confidence envelope" (Fox 1997) in all cases except the Poisson case with $\gamma = 1.2$, $b = 0.2$ and the mixed-Poisson case with $\gamma = 1.2$, $b = 0.6$. An omnibus Kolmogorov-Smirnov test of the null hypothesis that the profile likelihood ratio statistic

$$2[\ell(\widehat{\rho}_0(\cdot;\widehat{\beta}),\widehat{\beta}) - \ell(\widehat{\rho}_0(\cdot;\beta),\beta)]$$

Table 2.3: Empirical bias and coverage probability of local likelihood and piecewise constant estimates of the regression coefficient under Poisson and Mixed Poisson models based on 1000 simulated datasets of 500 subjects.

			Local Likelihood		PW4		PW10	
Model	$\mu_0(1)$	γ	Bias (ESE)	CP	Bias (ESE)	CP	Bias (ESE)	CP
Poisson	2	0.75	−0.001 (0.071)	0.961	−0.001 (0.071)	0.962	−0.001 (0.071)	0.960
		1.0	0.001 (0.070)	0.954	0.001 (0.070)	0.946	0.001 (0.070)	0.946
		1.2	−0.000 (0.075)	0.944	−0.000 (0.075)	0.947	−0.000 (0.075)	0.946
	4	0.75	−0.000 (0.050)	0.950	−0.000 (0.050)	0.950	−0.000 (0.050)	0.949
		1.0	−0.001 (0.048)	0.968	−0.001 (0.048)	0.962	−0.001 (0.048)	0.960
		1.2	−0.003 (0.051)	0.950	−0.003 (0.051)	0.946	−0.003 (0.051)	0.946
Mixed Poisson	2	0.75	0.000 (0.084)	0.953	−0.001 (0.084)	0.954	−0.001 (0.084)	0.954
		1.0	0.003 (0.086)	0.942	0.003 (0.086)	0.943	0.003 (0.086)	0.943
		1.2	−0.001 (0.087)	0.948	−0.000 (0.087)	0.945	−0.000 (0.087)	0.945
	4	0.75	0.001 (0.070)	0.936	−0.000 (0.071)	0.932	−0.000 (0.071)	0.932
		1.0	−0.001 (0.067)	0.960	−0.001 (0.067)	0.957	−0.001 (0.067)	0.958
		1.2	0.003 (0.069)	0.950	0.004 (0.070)	0.943	0.004 (0.070)	0.943

Table 2.4: Kolmogorov-Smirnov test statistics (KS) and p-values for assessing the validity of the χ^2_1 approximation to the profile likelihood ratio statistic based on 1000 simulated datasets of 500 subjects with mean function $(\theta t)^\gamma$.

		Poisson		Mixed-Poisson	
γ	b	KS	p-value	KS	p-value
0.75	0.2	0.029	0.390	0.034	0.187
	0.6	0.019	0.872	0.028	0.410
1.00	0.2	0.020	0.818	0.015	0.972
	0.6	0.014	0.991	0.032	0.263
1.20	0.2	0.038	0.106	0.034	0.207
	0.6	0.016	0.952	0.033	0.217

for the Poisson case and

$$2[\ell(\widehat{\rho}_0(\cdot;\widehat{\beta}),\widehat{\beta}) - \ell(\widehat{\rho}_0(\cdot;\beta),\beta)]/\widehat{\nu}$$

for the mixed-Poisson case follows a χ^2_1 distribution is given in Table 2.4. There is insufficient evidence to reject this null hypothesis for any of the parameter configurations considered here. The validity of this distributional assumption is further substantiated empirically by the good empirical coverage probability of the 95% profile likelihood ratio confidence intervals. In practical applications it may be worthwhile to use bootstrap methods to corroborate profile likelihood intervals; further analytic work will be required to ascertain whether conditions like those of Staniswalis et al. (1997) can be specified to provide rigorous broader justification for this distributional approximation.

2.5 Applications

2.5.1 Data on Incidence of Superficial Bladder Tumors

Byar (1980) describes a randomized clinical trial of patients who experienced superficial bladder tumors. This dataset has also been studied in Lawless and Zhan (1998), Wellner and Zhang (2000) and Kalbfleisch and Prentice (2002). Patients were randomly assigned either pyridoxine pills, thiotepa (a chemotherapeutic agent) or a placebo. Each time a patient was seen, any tumors present were counted and

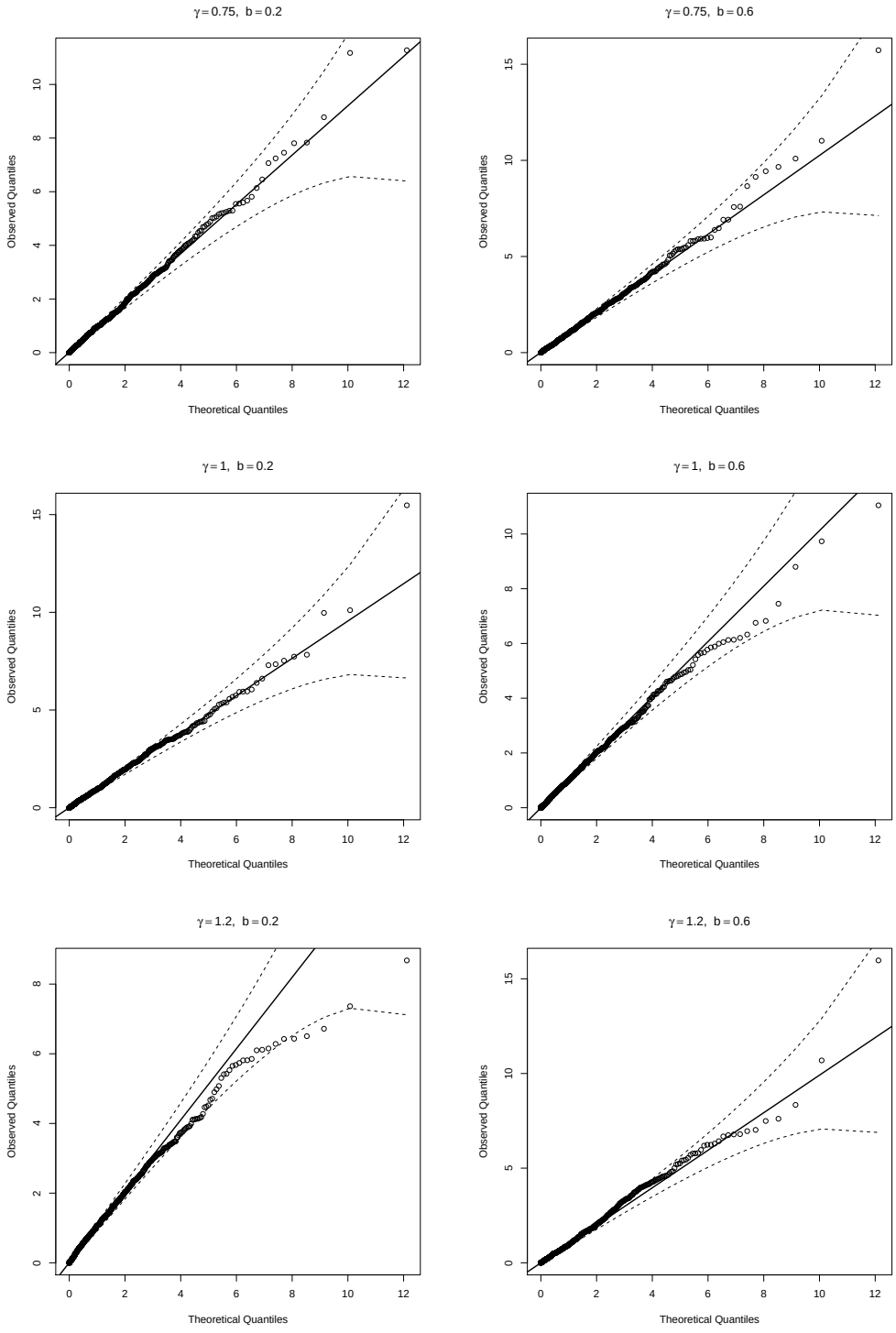

Figure 2.4: QQ-plots comparing the simulated quantiles of the profile likelihood ratio statistic with the theoretical quantiles of the χ^2_1 distribution based on 1000 simulated datasets of 500 subjects with Poisson mean function $(\theta t)^\gamma$.

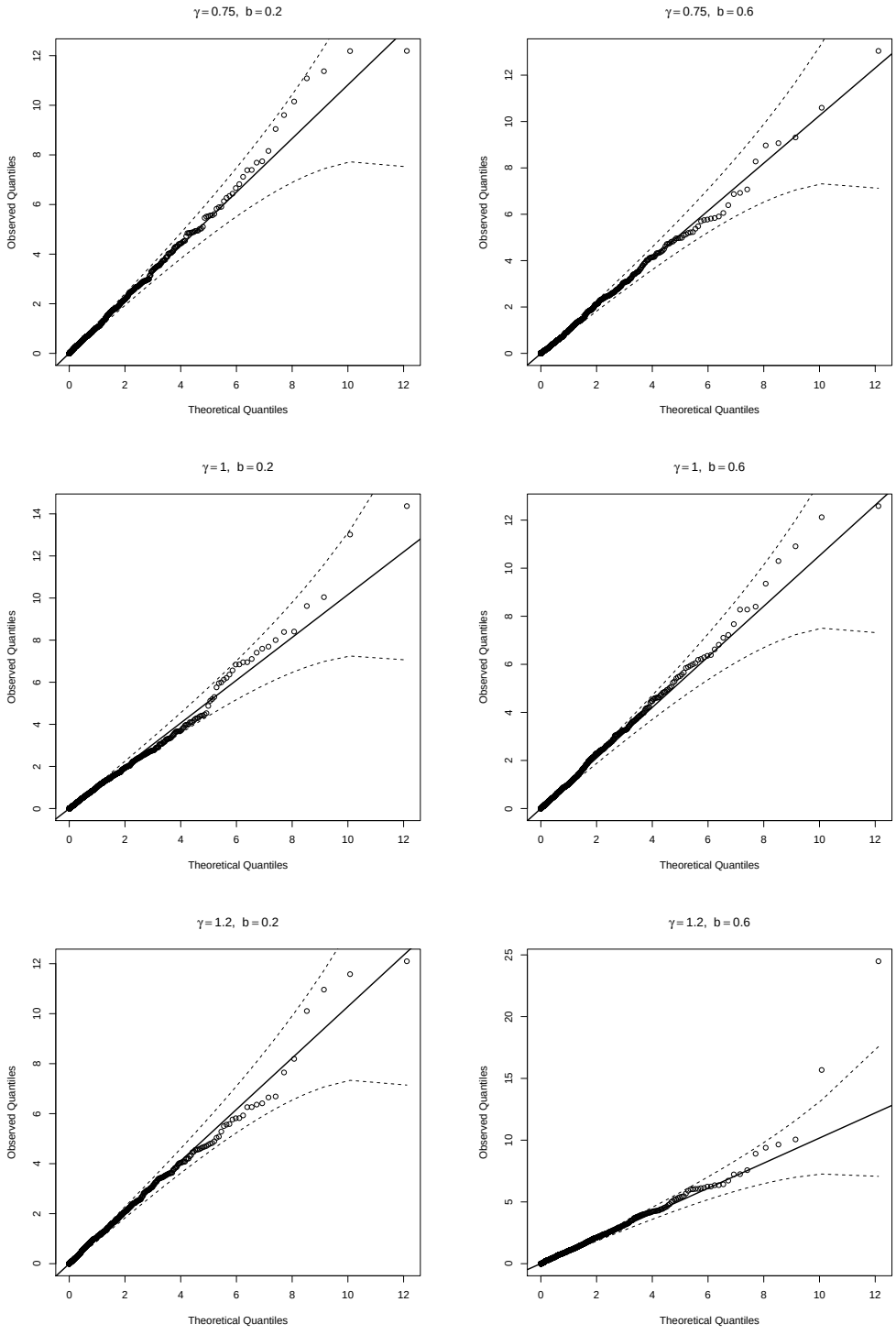

Figure 2.5: QQ-plots comparing the simulated quantiles of the profile likelihood ratio statistic with the theoretical quantiles of the χ^2_1 distribution based on 1000 simulated datasets of 500 subjects with mixed-Poisson mean function $(\theta t)^\gamma$ and $\phi = 0.25$.

Table 2.5: Local likelihood and Cox estimates and 95% confidence intervals for the effect of treatment in the bladder cancer study.

	Local Likelihood		Andersen-Gill	
Poisson	-0.392	$(-0.760, -0.037)$	-0.381	$(-0.743, -0.019)$
Extra-Poisson / Robust	-0.392	$(-0.939, 0.126)$	-0.381	$(-0.954, 0.192)$
Mixed-Poisson	-0.367	$(-0.927, 0.290)$	-0.302	$(-0.900, 0.296)$

removed. Figure 2.6 shows the number of tumors counted at inspection times for a sample of 24 patients.

As in Lawless and Zhan (1998) and Kalbfleisch and Prentice (2002) only the placebo and thiotepa groups will be considered in the analysis. A plot like that in in Lawless and Zhan (1998) of the total number of tumors against the total time on study is given in Figure 2.7.

Here, we will consider estimation of the mean functions separately for the treatment and placebo groups using the methods of Section 2.3. Hence, 85 patients were included in the study, of which 38 were assigned the treatment, thiotepa. Information on the initial number of tumors as well as the size of the largest initial tumor was also recorded for each patient.

The results of using local likelihood with the Epanechnikov kernel $K(u) = 3/4(1-u^2)$, $|u| \leq 1$, locally constant polynomial and 0.2 nearest-neighbours bandwidth lead to the results displayed in Table 2.5 for the effect of treatment. We consider the event times as being exact by assuming events occur at the inspection time they are observed. Other nearest-neighbours bandwidths were tried, ranging from 0.1 to 0.6, however the estimates did not vary greatly. For example, the Poisson local likelihood estimate of the treatment effect ranged between -0.406 and -0.377. Results based on a tricube kernel $K(u) = 70/81(1-|u|^3)^3$, $|u| \leq 1$ with the same bandwidth ranged between -0.405 and -0.375. Hence, the bandwidth had a small effect on the results but not enough to drastically change the conclusions. The choice of kernel had an even smaller impact, as expected.

For the sake of comparison, the results of fitting a semiparametric Andersen-Gill regression model including the effect of treatment are also displayed, along with 95% confidence intervals based on a Poisson variance (Andersen and Gill 1982), robust variance (Lawless and Nadeau 1995) and mixed-Poisson variance (Lawless 1987). The local likelihood estimates and confidence intervals are close to those

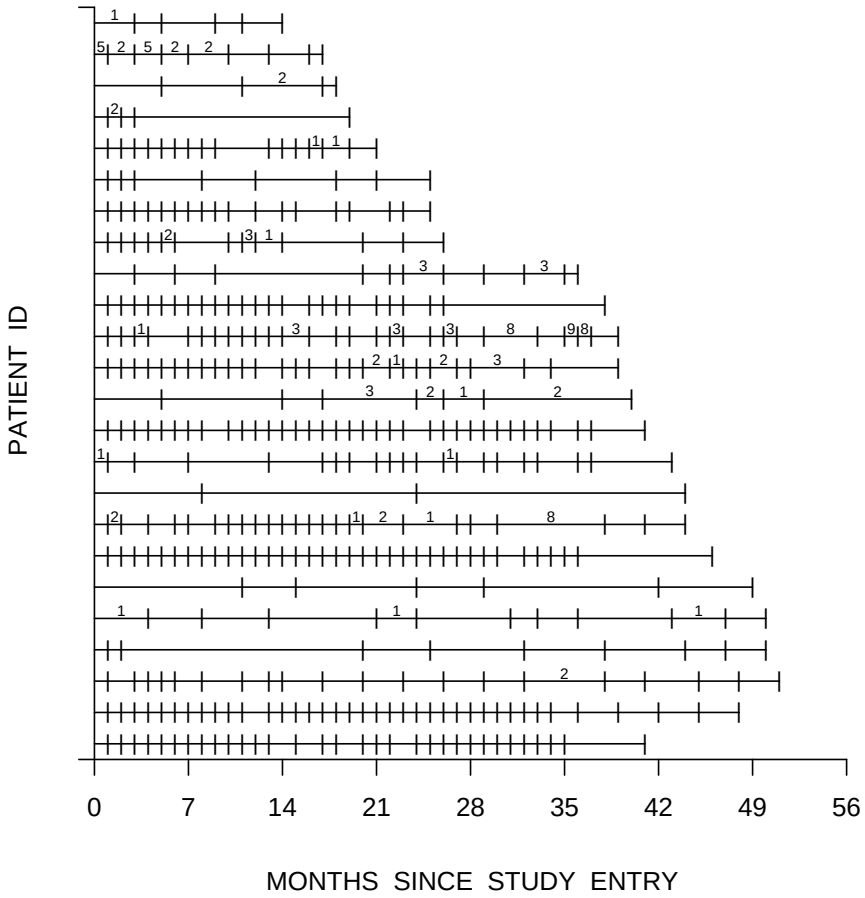

Figure 2.6: Timelines of 24 thiotepa patients from the bladder tumor study of Byar (1980). For a given patient, the horizontal line indicates the time under study while the vertical lines represent assessment times. The numbers between visit marks indicate the number of tumors that developed during that time interval (if greater than 0).

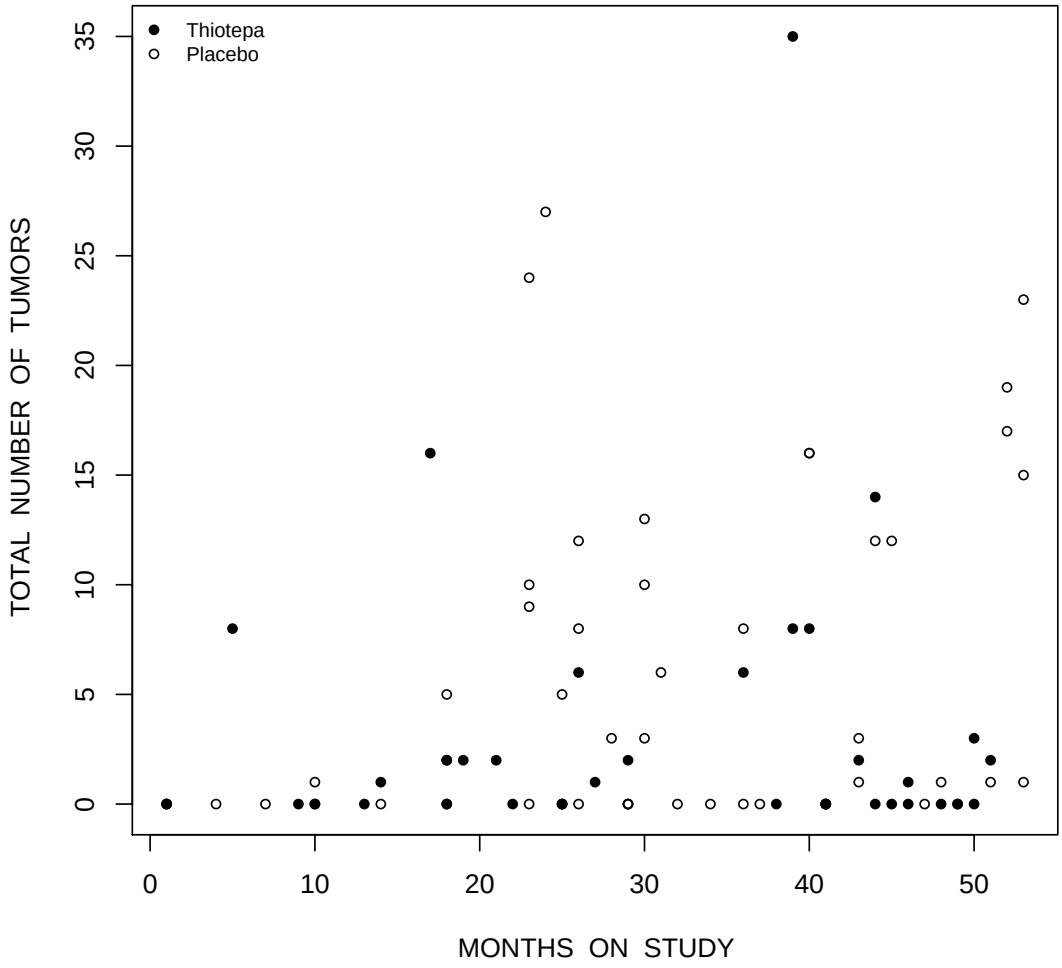

Figure 2.7: Plot of the total number of tumors detected against time on study in the bladder tumor data of Byar (1980).

obtained by a semiparametric PH model except for the mixed-Poisson case where the estimates differ, but the confidence intervals are quite similar.

2.5.2 Rat Tumorgenicity Data

Gail et al. (1980) discusses a randomized experiment involving 48 female rats. Rats were exposed to a carcinogen and further conditioned for 60 days prior to randomization to receive either a treatment or control. A followup period of 122 days began after randomization, during which they were examined every few days for the development of new tumours. The data are given in Gail et al. (1980) and in Cook and Lawless (2007) where the data are reported with the times from the beginning of the period of examination instead of from the time of exposure to the carcinogen.

Local likelihood with a locally constant polynomial, Epanechnikov kernel and 0.2 nearest-neighbours bandwidth lead to the results displayed in the first two columns of Table 2.6 for the effect of treatment with the event times right-censored. Varying the bandwidth produced the same estimates to three decimal places. The first row of Table 2.6 contains estimates obtained via local likelihood with a Poisson model, the second row reports the same point estimates with an adjusted variance estimate obtained by multiplying the Poisson variance estimate by $\widehat{\nu}$ given by (2.5), and the third row reports estimates obtained by maximizing (2.6) under the negative binomial model. Also provided are the estimates from fitting an Andersen-Gill model with Poisson variance estimates in the top row (Andersen and Gill 1982), robust variance estimates in the second row (Lawless and Nadeau 1995), and a negative binomial model in the third row (Lawless 1987). The local likelihood estimates and confidence intervals are generally close to those obtained by a semiparametric analysis except for the mixed-Poisson case where the point estimates differ slightly; the confidence intervals are quite similar however. All analyses suggest a strong and highly significant reduction in the rate of tumours among treated rats with over a 50% reduction in the event rate.

2.5.3 Nuclear Plant Reliability Study

Sun and Kalbfleisch (1995) consider an example where 30 nuclear plants are observed for loss of feedwater flow. Each plant was observed once and the number of losses of feedwater flow were recorded. The inspection times ranged from 1 to 15 years (with quantiles of 2.0, 3.5 and 5.0 years). Figure 2.8 contains the same

Table 2.6: Local likelihood and Cox estimates and 95% confidence intervals for the effect of treatment in the rat tumorgenicity study of Gail et al. (1980).

	Local Likelihood		Andersen-Gill	
Poisson	-0.823	$(-1.127, -0.531)$	-0.848	$(-1.146, -0.550)$
Extra-Poisson / Robust	-0.823	$(-1.255, -0.416)$	-0.848	$(-1.262, -0.434)$
Mixed-Poisson	-0.820	$(-1.233, -0.397)$	-0.861	$(-1.300, -0.424)$

nonparametric estimate of the mean function as Figure 1 in Sun and Kalbfleisch (1995) with local likelihood estimates of the mean function superimposed. The nonparametric estimate of the mean function suggested by Sun and Kalbfleisch (1995) is given by the filled circles while the clear circles indicate the average number of losses of feedwater flow per nuclear plant at each time point. Both local likelihood estimates are based on the Epanechnikov kernel with 100 grid points and a locally constant polynomial. The dashed line corresponds to a bandwidth of 1.0 while the solid line corresponds to a bandwidth of 0.33 chosen by cross validation. This bandwidth was obtained by finding the b which maximizes the likelihood criterion given by (1.1) without the penalty term which in this case is given by

$$\sum_{i=1}^{n} \{n_i \log \widehat{\mu}_{-i}(C_i; b) - \widehat{\mu}_{-i}(C_i; b)\}$$

where $\widehat{\mu}_{-i}(t; b)$ is the local likelihood estimate with bandwidth b and C_i is the time the ith plant was observed. Other kernels were tried, but the resulting estimates provided little visual difference.

The local likelihood estimates tracks the nonparametric one quite closely while offering smooth estimates of the mean function. The plot suggests that the rate of occurrence of loss of feedwater flow begins to decrease at about 4 years and increases after about 7 years.

2.5.4 Counts of Damaged Joints in Psoriatic Arthritis

The event of interest here is the development of damage in joints and interest lies in both rate of occurrence of damage and the expected cumulative number of damaged joints over time. Moreover, identification of important covariate effects is also of interest to help characterize risk among patients present in the clinic.

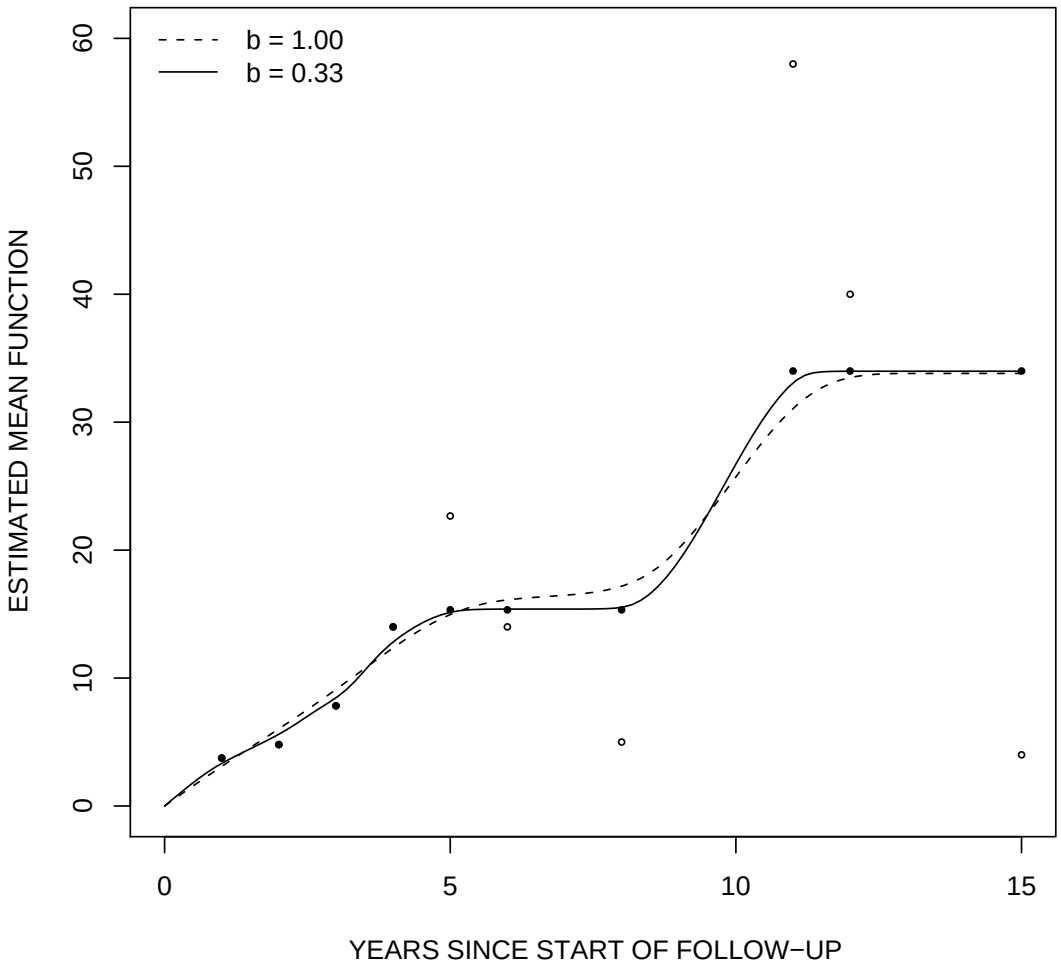

Figure 2.8: Estimates of the mean number of losses of feedwater flow obtained non-parametrically (filled circles) and using local likelihood with bandwidth 1.0 (dashed) and 0.33 (solid); clear circles indicate the mean number of losses of feedwater flow across plants observed at that time point.

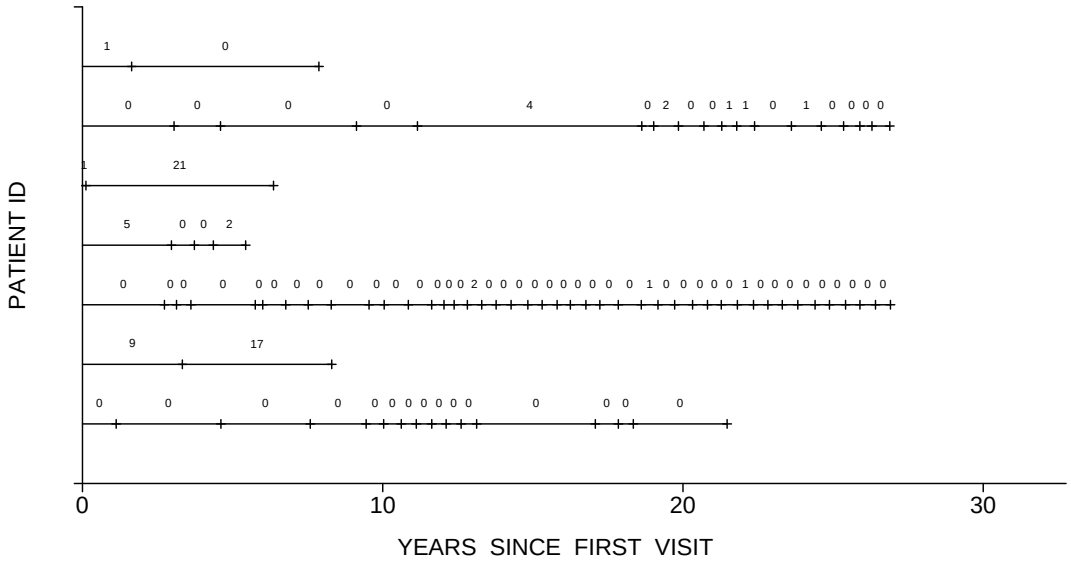

Figure 2.9: Timeline diagram of joint damage for a sample of seven patients from the University of Toronto psoriatic arthritis clinic; tick marks indicate assessments and numbers indicate the number of newly damaged joints.

While assessments are scheduled annually, in reality there is considerable variability in the frequency and timing of clinic visits. Figure 2.9 shows the times of assessments for a sample of patients along with counts of the number of damaged joints occurring between assessments. As can be seen, some patients adhere to a regular schedule of clinic visits and others attend less frequently and more irregularly.

Estimates of the expected cumulative number of damaged joints over time are given in Figure 2.10 based on Poisson models via the local likelihood approach as well as based on piecewise constant models of Lawless and Zhan (1998) with 3 and 6 pieces. The local likelihood estimate, based on an Epanechnikov kernel, a bandwidth of 3 and a locally constant function agrees very well with the piecewise estimates over the majority of the followup. There is a slight divergence in these estimates towards the end of followup due to boundary problems commonly observed in local likelihood methods (Loader 1999). Other kernels and bandwidths may be applied, however due to the apparent linear nature of the mean function, there is little difference in the estimates.

Table 2.7 contains the results of fitting the mixed Poisson model for estimation

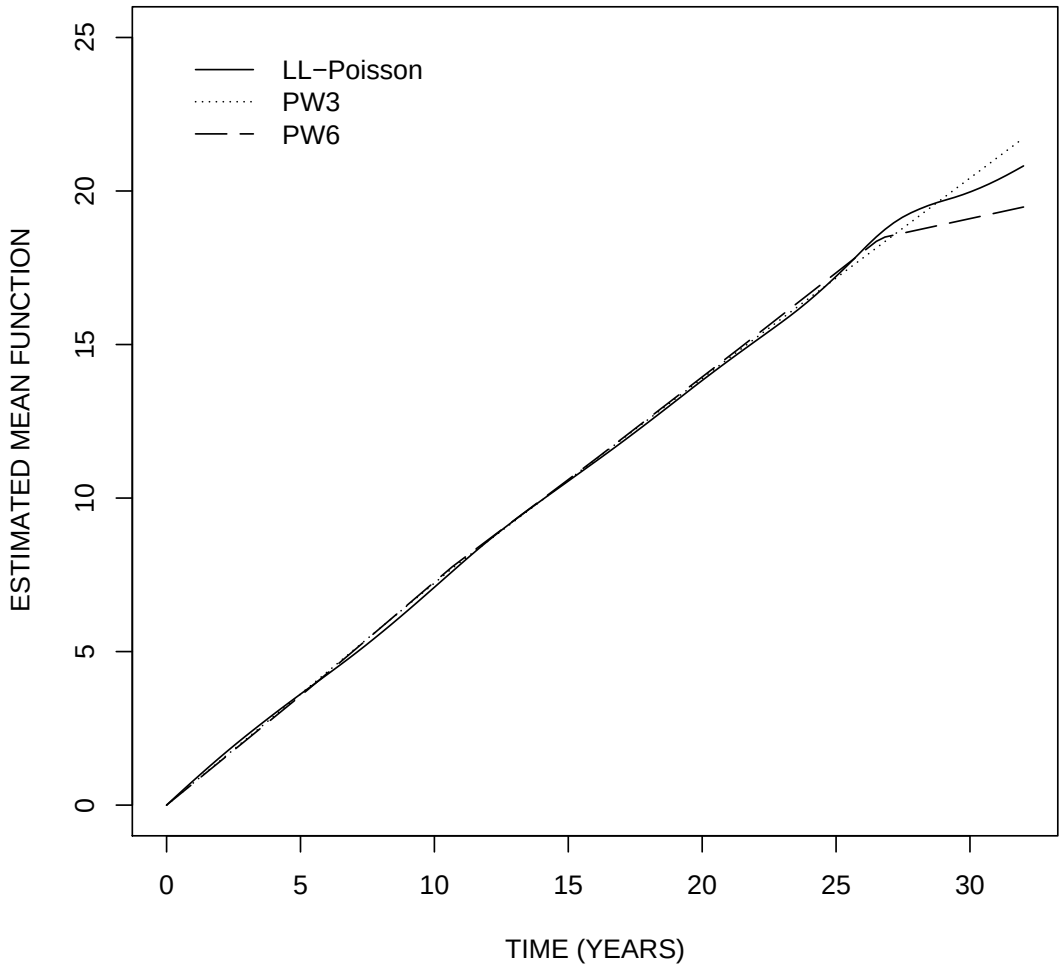

Figure 2.10: Poisson local likelihood and piecewise constant estimates of the mean function with three (PW3) and six (PW6) pieces based on the joint damage from the psoriatic arthritis data.

Table 2.7: Local likelihood estimates of the regression parameters in the psoriatic arthritis study.

Covariate	Estimate	Jackknife SE	Jackknife 95% CI
Duration of Arthritis (years)	0.026	0.005	$(-0.015, -0.037)$
Sex (male vs female)	-0.034	0.133	$(-0.295, 0.227)$
Family History of PsA (yes vs no)	-0.346	0.252	$(-0.839, 0.148)$

and profile likelihood intervals for a multivariate model with covariates duration of disease (years), sex (1=male, 0=female) and family history of disease(1=yes,0=no). Varying the bandwidth from 0.5 to 10 years produced little change in the estimates. The estimate of the effect due to the duration of arthritis did not change at all, and the largest difference was for the effect due to sex, which ranged from -0.034 ($b = 10$) to -0.031 ($b = 0.5$). The findings show that the rate of damaged joints increases with each additional year of disease (RR=1.026; 95% CI $(1.016, 1.037)$); there is no significant effect of sex or family history of disease.

2.6 Summary

In this chapter, interval-censored recurrent event data was considered. Main objectives were smooth estimation of the mean function, as existing methods lack smoothness and may not be everywhere defined. Local likelihood methods were used to obtain smooth estimates of both the mean and rate function. Bandwidth selection was mainly done by visual inspection, although leave-one-out cross validation was considered for the nuclear reliability example. The choice of the bandwidth is always a difficult issue when working with smoothing techniques. The simulations and examples considered the use of the Epanechnikov kernel with a constant local approximation to the log of the rate function. Other kernels may be used, although changing kernels appears to have little impact on the results. The use of higher order polynomials in the local approximation may also be considered, although the computations become much more complicated. In the case of survival data, Hjort and Jones (1996) note that using a Gaussian kernel simplifies the calculations required with a higher-order local polynomial and Braun et al. (2005) show that closed form expressions for the iterations of the local EM algorithm can be obtained with polynomials of order 1 or 2. Boundary bias was also observed in the simulations. Hjort and Jones (1996) examined reducing the bias at the boundary

for survival data by increasing the order of the local polynomial, and Loader (1999) notes that doing so comes at the cost of an increase in variability. Methods for covariates were also considered along with relaxing the at times restrictive form of the variance function of a Poisson process. Simulation studies demonstrated the good performance of the local likelihood methods. The results of this work can also be found in Tolusso and Cook (2008a).

The methods of this chapter are robust in the sense that the simulations demonstrated the consistency of the Poisson based local likelihood estimates even when the underlying process was not Poisson. In addition, the Pepe estimate is constructed without ever making reference to form of the underlying event process.

Chapter 3

Estimating Equations for Clustered Current Status Data

3.1 Introduction and Overview

Suppose we have a study where each subject can experience a particular event, and it is the time to the event that is of interest. Suppose that the time of the event is not observed exactly, rather each subject has a specified follow-up time at which it is determined whether or not the event has occurred. Such data are termed type I interval censored or current status data and is an extreme form of interval censoring.

With current status data, the likelihood is straightforward to construct and a parametric estimate of the cumulative distribution function is readily available. Nonparametric estimates may be obtained using isotonic regression methods such as the min-max formula of Barlow et al. (1972) or the pooled-adjacent violators algorithm of Ayer et al. (1955). Regression models may also be considered (Sun 2006, Chapter 5), and in some instances (for example proportional hazards with Weibull baseline hazard) may be fit using existing software for generalized linear models. A weakly parametric approach may also be taken where the baseline hazard is assumed to be piecewise constant (Zhan 1999). A semiparametric approach is taken by Shiboski (1998) using generalized additive models methodology.

Current status data may be extended to the case where for each subject there are two event times of interest, which may be dependent. Wang and Ding (2000) consider this case with the focus being on estimating the association between the two failure times. They propose the use of a copula to model the association. Two-

stage semiparametric estimation of the association follows along the lines of Shih and Louis (1995) for right-censored data. Jewell et al. (2005) consider nonparametric estimation of the joint distribution of bivariate current status. They use an EM algorithm to estimate the identifiable parts of the joint distribution under univariate monitoring times. They also provide a test of independence and goodness-of-fit of a copula. Chapter 7 of Sun and Kalbfleisch (1995) also discusses the analysis of bivariate interval-censored data using Copula models.

Another generalization of univariate current status data is the case where subjects may be grouped into clusters. Individuals within a cluster have potentially dependent event times. This situation will be considered. For fully parametric models, the data is similar to binary data. Hence methods of generalized estimating equations (Liang and Zeger 1986) may be used to obtain estimates of the model parameters when faced with possible dependence within clusters. Here, a copula may be used to model the association between event times within a cluster, which in turn induces an association structure which may be used to define working covariance matrices.

The organization of the remainder is as follows. Section 3.2 describes the notation and the general approach to be taken. Section 3.3 develops the methodology. Simulation studies are discussed in Section 3.4. Simulations were carried out to evaluate the performance of piecewise constant models for estimation of regression coefficients as well as quantiles and survival probabilities of the failure time distribution. The methods are illustrated by example in Section 3.5.

3.2 Notation

Suppose there are m clusters of individuals with n_i observations per cluster, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$. Let S_{ij} denote the event time for the jth individual in cluster i, and $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ denote a $p \times 1$ covariate vector. We assume a proportional hazards formulation to examine covariate effects and hence assume S_{ij} has a marginal survivor function $\mathcal{F}(s|x_{ij};\theta) = [\mathcal{F}_0(s;\alpha)]^{\exp(x_{ij}'\beta)}$ where $\mathcal{F}_0(s;\alpha)$ is a baseline survivor function indexed by α, β is a $p \times 1$ vector of regression coefficients, and $\theta = (\alpha', \beta')'$. Let C_{ij} denote the inspection time for individual j in cluster i. For each individual, in addition to the covariate x_{ij} we observe the indicator $Y_{ij} = I(S_{ij} \leq C_{ij})$ which indicates if the event has occurred before the inspection time. The data are therefore binary where $p_{ij} = P(Y_{ij} = 1|C_{ij}, x_{ij};\theta) = 1 - \mathcal{F}(C_{ij}|x_{ij};\theta)$. If $\Lambda_0(s;\alpha)$ is the cumulative baseline hazard function, a Weibull

model is given by $\Lambda(s;\alpha) = (\lambda s)^\gamma$ where $\alpha = (\lambda, \gamma)'$, and a piecewise constant hazard model is given by $\Lambda(s;\alpha) = \sum_{k=1}^K \lambda_k w_k(s)$, where $\alpha_k = \log \lambda_k$, $k = 1, \ldots, K$, and $\alpha = (\alpha_1, \ldots, \alpha_K)'$.

It is often desirable to formulate covariate effects on marginal features of a multivariate distribution, which is one of the reasons that generalized estimating equations have seen such widespread use in the analysis of clustered data with marginal distributions in the exponential family (Liang and Zeger 1986), marginal methods based on rate and mean functions are popular for recurrent event analysis (Lawless and Nadeau 1995), and marginal methods are popular for multivariate failure time data (Wei et al. 1989). These, however, are all examples of methods of analysis. Copula functions (Joe 1997) provide a convenient framework for constructing joint distributions with specified margins, and we consider their use here.

A copula function in n dimensions is a multivariate distribution on the unit hypercube $[0,1]^n$, with uniform margins. Consider the general family of Archimedean copulas (Genest and MacKay 1986),

$$H_n(u) = \Phi^{-1}(\Phi(u_1) + \cdots + \Phi(u_n))$$

where $\Phi(v)$ is a function known as the generator. Omitting the subscript i for convenience if we have a cluster of size n, the multivariate failure time distribution of all times in this cluster are generated by treating $\mathcal{F}(s_{ij}; x_{ij}; \theta)$ as a uniform random variable, and obtaining

$$P(S_{i1} > s_{i1}, \ldots, S_{in} > s_{in} | x_i; \theta) = H_n(\mathcal{F}(s_{i1}|x_{i1};\theta), \ldots, \mathcal{F}(s_{in}|x_{in};\theta))$$

The Clayton copula is obtained by using the generator $\Phi(v) = v^{-\phi} - 1$ which gives $\Phi^{-1}(v) = (v+1)^{-1/\phi}$ and

$$H_n(u;\phi) = (u_1^{-\phi} + \cdots + u_n^{-\phi} - n + 1)^{-1/\phi}. \tag{3.1}$$

The resulting joint survivor function $P(S_{i1} > s_{i1}, \ldots, S_{in} > s_{in} | x_i; \theta, \phi)$ is

$$(\mathcal{F}(s_{i1}|x_{i1};\theta)^{-\phi} + \cdots + \mathcal{F}(s_{in}|x_{in};\theta)^{-\phi} - n + 1)^{-1/\phi}. \tag{3.2}$$

Note that we can marginalize over s_k and obtain

$$P(S_1 > s_1, \ldots, S_{k-1} > s_{k-1}, S_{k+1} > s_{k+1}, \ldots, S_n > s_n | x_i; \theta, \phi)$$

simply by inserting $s_k = 0$ into (3.2). We can then obtain

$$\begin{aligned} P(S_1 > s_1, \ldots, S_{k-1} > s_{k-1}, S_k < s_k, S_{k+1} > s_{k+1}, \ldots, S_n > s_n | x_i; \theta, \phi) \\ = P(S_1 > s_1, \ldots, S_{k-1} > s_{k-1}, S_{k+1} > s_{k+1}, \ldots, S_n > s_n | x_i; \theta, \phi) \\ - P(S_1 > s_1, \ldots, S_n > s_n | x_i; \theta, \phi). \end{aligned}$$

General expression for the current status data from such a cluster are similarly obtained and likelihood functions may be constructed based on these expressions.

Maximization of the resulting likelihood is challenging, however, and we consider instead the use of generalized estimating functions for estimation and inference. Estimation can then proceed under the working independence assumption with an appropriate sandwich variance estimate, or methods of generalized estimating equations (Liang and Zeger 1986) may be used. One significant difference between previous GEE methods for binary data and the current setup is that previous binary data methods parameterize the association between Y_{ij} and Y_{ik} using, for example odds ratios (Liang et al. 1992) or correlations (Prentice 1988). In the current setting, the association between S_{ij} and S_{ik} is parameterized via the copula function, which induces an association structure between the elements of $Y_i = (Y_{i1}, \ldots, Y_{in})'$ unlike any of those considered before for clustered binary data. Figure 3.1 shows the Pearson correlation of Y_1 and Y_2 as a function of the corresponding assessment times C_1 and C_2 under a Weibull model with shape parameter γ and Clayton copula with association parameter ϕ.

3.3 Methods of Estimation and Inference

3.3.1 Generalized Estimating Equations

Let $p_{ij}(\theta) = P(Y_{ij} = 1|C_{ij}, X_{ij}; \theta) = 1 - \mathcal{F}(C_{ij}|X_{ij}; \theta)$ denote the probability individual j in cluster i tests positive given C_{ij} and X_{ij}. In standard use of generalized estimating equations for binary data one would choose among the standard link functions for binary data to specify a generalized linear model. In the context of a proportional hazards model for the underlying failure time distribution, the complementary log-log link will give a linear model with a common linear predictor, but additional parameters will need to be estimated for the baseline hazard function. Moreover, we would typically let $R_i(\rho)$ denote a working correlation structure for the observations in cluster i, and obtain $V_i(\theta, \rho) = A_i^{1/2} R_i(\rho) A_i^{1/2}$ as the working covariance structure where $A_i = \text{diag}\{\partial p_{i1}/\partial\theta_1, \ldots, \partial p_{in_i}/\partial\theta_r\}$. It is well known that consistent estimates for θ are obtained even if the working correlation structure is incorrect (Liang and Zeger 1986), and that maximum efficiency is obtained the closer this correlation structure is to the true structure. While the responses are binary in the present setting, it is inappropriate to use any of the standard correlation structures for binary data because of the underlying joint distribution

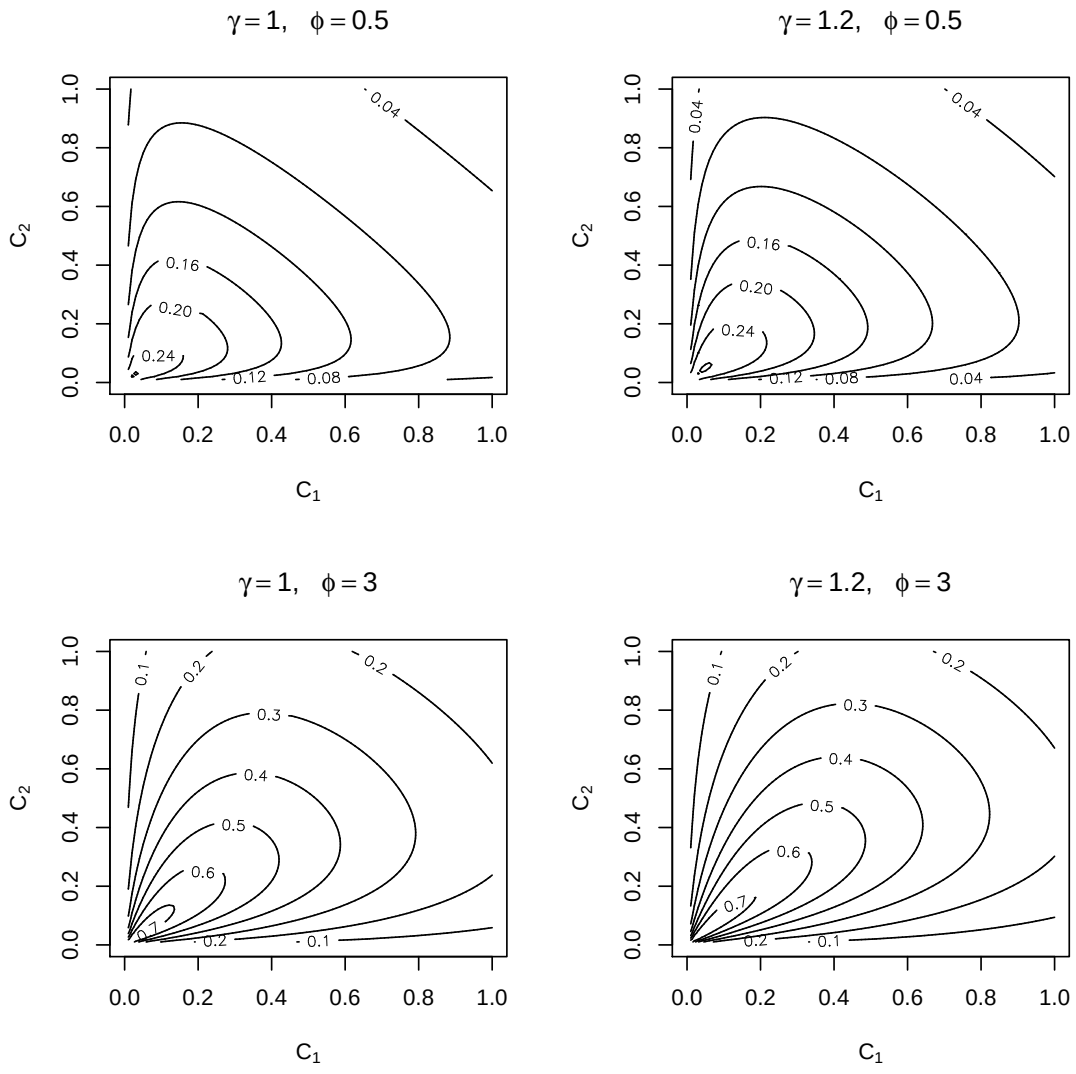

Figure 3.1: Correlation between Y_1 and Y_2 as a function of C_1 and C_2.

of the failure times. We therefore consider covariance matrices obtained under the copula formulation of the previous section.

Recall $\theta = (\alpha', \beta')'$, let $\kappa = \log \phi$ and let $\psi = (\theta', \kappa)'$. Consider the first order estimating equations for θ,

$$U_\theta(\psi) = \sum_{i=1}^{m} D_i'(\theta) V_i^{-1}(\psi)(Y_i - p_i(\theta)) \tag{3.3}$$

where $Y_i = (Y_{i1}, \ldots, Y_{in_i})'$, $p_i(\theta) = (p_{i1}(\theta), \ldots, p_{in_i}(\theta))'$, $D_i(\theta) = \partial p_i(\theta)/\partial\theta'$ and $V_i(\psi)$ is an $n_i \times n_i$ covariance matrix with $V_{ijj}(\psi) = p_{ij}(1 - p_{ij})$ in the diagonals and off diagonals given by

$$V_{ijk}(\psi) = \mathrm{cov}(Y_{ij}, Y_{ik}|X_i; \psi) = P(T_{ij} \le C_{ij}, T_{ik} \le C_{ik}|X_i; \psi) - p_{ij}(\theta)p_{ik}(\theta)$$

for $j \neq k$ with $P(T_{ij} \le C_{ij}, T_{ik} \le C_{ik}|X)i; \psi)$ determined by the copula.

A second set of estimating equations is required to estimate κ for given θ. This can be constructed using the pairwise combinations of the elements of Y_i as in Prentice (1988) and Zhao and Prentice (1990). Here, and in most of what follows, we suppress the dependence on C_{ij} in the notation. Let $Z_i = (Y_{i1}Y_{i2}, \ldots, Y_{i,n_i-1}Y_{in_i})'$ be a column vector of all $r_i = n_i(n_i - 1)/2$ pairwise products and $\eta_i(\psi) = E[Z_i|X_i]$ the $r_i \times 1$ vector of conditional expectations of the elements of Z_i, which are again determined by the copula through

$$\begin{aligned} E[Z_{ij}|X_i] &= P(S_{ij} < C_{ij}, S_{ik} < C_{ik}|X_i; \psi) \\ &= 1 - \mathcal{F}(C_{ij}|X_{ij}; \theta) - \mathcal{F}(C_{ik}|X_{ik}; \theta) + P(S_{ij} \ge C_{ij}, S_{ik} \ge C_{ik}; \psi) \ . \end{aligned}$$

The estimating equation for κ is then

$$U_\kappa(\psi) = \sum_{i=1}^{m} G_i'(\psi) W_i^{-1}(\psi)(Z_i - \eta_i(\psi)) = 0 \ , \tag{3.4}$$

where $G_i(\psi) = \partial\eta_i/\partial\kappa$ and $W(\psi) = \mathrm{diag}\{\eta_{i\ell}(1 - \eta_{i\ell}), \ell = 1, \ldots, r_i\}$ and zeros elsewhere. Let $U(\psi) = (U_\theta'(\psi), U_\kappa(\theta))'$, and let $U_i(\psi)$ denote the contribution to $U(\psi)$ from cluster i, $i = 1, \ldots, m$. Let $\widehat{\psi}$ denote the solution to $U(\psi) = 0$. Then the asymptotic variance of $\sqrt{m}(\widehat{\psi} - \psi)$ has the sandwich form,

$$\mathcal{A}^{-1}(\psi)\mathcal{B}(\psi)\{\mathcal{A}^{-1}(\psi)\}'$$

where $\mathcal{A}(\psi) = E(\partial U_i(\psi)/\partial\psi)$ and $\mathcal{B}(\psi) = E(U_i(\psi)U_i'(\psi))$ are the functions to which

$$A(\psi) = \frac{1}{m}\sum_{i=1}^{m} \begin{pmatrix} D_i'V_i^{-1}D_i & 0 \\ G_i'W_i^{-1}Q_i & G_i'W_i^{-1}G_i \end{pmatrix}$$

and

$$B(\psi) = \frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} D_i'V_i^{-1}(Y_i-p_i)(Y_i-p_i)'V_i^{-1}D_i & D_i'V_i^{-1}(Y_i-p_i)(Z_i-\eta_i)'W_i^{-1}G_i \\ G_iW_i^{-1}(Z_i-\eta_i)(Y_i-p_i)'V_i^{-1}D_i & G_i'W_i^{-1}(Z_i-\eta_i)(Z_i-\eta_i)'W_i^{-1}G_i \end{pmatrix}$$

converge in probability, where $Q_i(\psi) = \partial\eta_i(\psi)/\partial\theta$ is a $r_i \times p$ matrix of derivatives, and we suppress the dependence on θ or κ on the right hand sides. A consistent estimate of the variance of $\sqrt{m}(\widehat{\psi}-\psi)$ is then given by $A(\widehat{\psi})^{-1}B(\widehat{\psi})[A(\widehat{\psi})^{-1}]'$.

The solution to the estimating equations (3.3) and (3.4) can be obtained using Fisher scoring. If $\widehat{\psi}^{(t-1)}$ denotes the estimate of ψ at the $(t-1)$st iteration, then the iterations take the form

$$\widehat{\psi}^{(t)} = \widehat{\psi}^{(t-1)} + \left(mA(\widehat{\psi}^{(t-1)})\right)^{-1} U(\widehat{\psi}^{(t-1)})$$

and iterations proceed until the difference between successive estimates decreases to a level below a specified tolerance.

The described procedure is robust to misspecification of the dependence structure in the sense that only $\mathcal{F}(s|X_{ij};\theta)$ need be correct to obtain consistent estimators for θ since the estimating equations for θ and ψ are constructed separately. Also note that since the top-right block of $\mathcal{A}$ is 0, the variance of θ is not affected by the choice of W_i, hence the rationale for using a simple diagonal matrix for W_i, instead of the optimal choice, $\mathrm{cov}(Z_i|X_i)$ (McCullagh and Nelder 1989).

As an alternative approach, a joint estimating equation (GEE2) can be constructed as suggested by Prentice (1988), Zhao and Prentice (1990) and Liang et al. (1992). This can improve efficiency by exploiting information about the parameters of the marginal distribution in the second moments, but requires correct specification of the association structure in order to obtain consistent estimates of ψ. The GEE2 equations are obtained by setting the estimating functions

$$\begin{aligned}\bar{U}(\psi) &= \sum_{i=1}^{m}\begin{pmatrix} \dfrac{\partial(p_i',\eta_i')'}{\partial\psi'} \end{pmatrix}'\begin{pmatrix} V_i & H_i \\ H_i' & W_i \end{pmatrix}^{-1}\begin{pmatrix} Y_i-p_i(\theta) \\ Z_i-\eta_i(\psi) \end{pmatrix} \\ &= \sum_{i=1}^{m}\begin{pmatrix} \partial p_i/\partial\theta' & \partial p_i/\partial\kappa \\ \partial\eta_i/\partial\theta' & \partial\eta_i/\partial\kappa \end{pmatrix}'\begin{pmatrix} V_i & H_i \\ H_i' & W_i \end{pmatrix}^{-1}\begin{pmatrix} Y_i-p_i(\theta) \\ Z_i-\eta_i(\psi) \end{pmatrix}\end{aligned} \tag{3.5}$$

equal to zero, where $\partial p_i/\partial\kappa = 0$ and $H_i = \mathrm{cov}(Y_i, Z_i'|X_i)$. Obtaining the solutions to the GEE2 equations (3.5) is computationally more intensive than solving (3.3) and (3.4) since the $(n_i+r_i)\times(n_i+r_i)$ covariance matrix (the second matrix in (3.5)) must be inverted.

If $\bar{U}_i(\psi) = \sum_{i=1}^m \bar{U}_i(\psi)$, then let $\bar{\psi}$ denote the solution. Then $\sqrt{m}(\bar{\psi} - \psi)$ has asymptotic variance $\bar{\mathcal{A}}^{-1}(\psi)\bar{\mathcal{B}}(\psi)[\bar{\mathcal{A}}^{-1}(\psi)]^{-1}$, where $\bar{\mathcal{A}}(\psi) = E(\partial \bar{U}_i(\psi)/\partial\psi)$ and $\bar{\mathcal{B}}(\psi) = E(\bar{U}_i(\psi)\bar{U}_i'(\psi))$ which are the limiting functions (in probability) of

$$\bar{A}(\psi) = \frac{1}{m}\sum_{i=1}^m \begin{pmatrix} \frac{\partial p_i}{\partial \theta'} & \frac{\partial p_i}{\partial \kappa} \\ \frac{\partial \eta_i}{\partial \theta'} & \frac{\partial \eta_i}{\partial \kappa} \end{pmatrix}' \begin{pmatrix} V_i & H_i \\ H_i' & W_i \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial p_i}{\partial \theta'} & \frac{\partial p_i}{\partial \kappa} \\ \frac{\partial \eta_i}{\partial \theta'} & \frac{\partial \eta_i}{\partial \kappa} \end{pmatrix}$$

and

$$\bar{B}(\psi) = \frac{1}{m}\sum_{i=1}^m \bar{U}_i(\psi)\bar{U}_i'(\psi)$$

Estimates of $\bar{A}(\psi)$ and $\bar{B}(\psi)$ are given by inserting $\bar{\psi}$ into these expressions, and an estimate of $\text{var}(\sqrt{m}(\bar{\psi} - \psi))$ is $\bar{A}^{-1}(\bar{\psi})\bar{B}(\bar{\psi})\{\bar{A}^{-1}(\bar{\psi})\}'$.

3.3.2 Relative Efficiency of GEE1 vs. GEE2

It is known that the working independence and GEE1 methods are not fully efficient. Liang and Zeger (1986), Liang et al. (1992) and Carey et al. (1993) have shown that for estimation of parameters other than the association parameters, GEE1 methods are nearly as efficient as GEE2 methods, and GEE2 methods are nearly fully efficient. A study of whether or not this holds for the case of clustered current status data with a Weibull baseline hazard is considered. In the GEE1 case, $\text{var}(\sqrt{m}(\widehat{\psi} - \psi)) = \mathcal{A}^{-1}(\psi)\mathcal{B}(\psi)\{\mathcal{A}^{-1}(\psi)\}'$, where

$$\mathcal{A}(\psi) = E\left\{ \begin{pmatrix} D_i' V_i^{-1} D_i & 0 \\ G_i' W_i^{-1} Q_i & G_i' W_i^{-1} G_i \end{pmatrix} \right\}$$

and

$$\mathcal{B}(\psi) = E\left\{ \begin{pmatrix} D_i' V_i^{-1} D_i & D_i' V_i^{-1} H_i W_i^{-1} G_i \\ G_i' W_i^{-1} H_i' V_i^{-1} D_i & G_i' W_i^{-1} G_i \end{pmatrix} \right\}$$

where, unlike the usual case of clustered binary data, here the expectation is taken with respect to both the covariate distribution and the inspection time distribution. We let $G(C_{ij}|X_{ij}) = G(C_{ij})$ denote the distribution function for the inspection times and suppose it has mean μ and variance σ^2.

In the case of GEE2, $\text{var}(\sqrt{m}(\bar{\psi} - \psi)) = \bar{\mathcal{A}}^{-1}(\psi)\bar{\mathcal{B}}(\bar{\psi})\{\bar{\mathcal{A}}^{-1}(\psi)\}'$ where

$$\bar{\mathcal{A}}(\psi) = E\left\{ \begin{pmatrix} \frac{\partial p_i}{\partial \theta'} & \frac{\partial p_i}{\partial \kappa} \\ \frac{\partial \eta_i}{\partial \theta'} & \frac{\partial \eta_i}{\partial \kappa} \end{pmatrix}' \begin{pmatrix} V_i & H_i \\ H_i' & W_i \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial p_i}{\partial \theta'} & \frac{\partial p_i}{\partial \kappa} \\ \frac{\partial \eta_i}{\partial \theta'} & \frac{\partial \eta_i}{\partial \kappa} \end{pmatrix} \right\}$$

and

$$\bar{\mathcal{B}}(\bar{\psi}) = E\left\{ \begin{pmatrix} \frac{\partial p_i}{\partial \theta'} & \frac{\partial p_i}{\partial \kappa} \\ \frac{\partial \eta_i}{\partial \theta'} & \frac{\partial \eta_i}{\partial \kappa} \end{pmatrix}' \begin{pmatrix} V_i & H_i \\ H_i' & W_i \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial p_i}{\partial \theta'} & \frac{\partial p_i}{\partial \kappa} \\ \frac{\partial \eta_i}{\partial \theta'} & \frac{\partial \eta_i}{\partial \kappa} \end{pmatrix} \right\}.$$

For the asymptotic variance of GEE2 methods we require the entries of W_i and H_i. Let $p_i(j,k,h,l)$ denote the elements of p_i with subscripts among the unique elements of (j,k,h,l). Then entries of W_i are of the form $H(p_i(j,k,h,l);\phi) - H(p_i(j,k);\phi)H(p_i(h,l);\phi)$ for $j \neq k$ and $h \neq l$, and entries of $H_i = \text{cov}(Y_i, Z_i|X_i)$ are given by $H(p_i(j,h,l);\phi) - p_{ij}H(p_i(h,l);\phi)$ for $h \neq l$.

The asymptotic relative efficiencies can be evaluated for given inspection time and covariate distributions by evaluating the matrices $\mathcal{A}(\psi)$, $\mathcal{B}(\psi)$, $\bar{\mathcal{A}}(\psi)$ and $\bar{\mathcal{B}}(\psi)$. To evaluate the requisite expectations and study this further, we make the following distributional assumptions. Suppose subjects have their assessments in the interval $(0,\tau]$. Let the underlying marginal distribution of the time be Weibull with shape parameter γ and rate parameter λ. For a given γ, λ is chosen such that $P(S > \tau) = p$. We assume a Clayton copula (3.1) with association parameter ϕ giving a joint distribution as in (3.2). If we let C^* be gamma distributed with mean μ and variance σ^2, we take the inspection time to be $C = \min(C^*, \tau)$. For given σ^2, γ and p, μ is chosen such that the probability an individual tests positive is $P(T < C) = \rho$. In the two-sample case, the covariate X_{ij} is generated as a binary random variable with $P(X_{ij} = 1) = 0.5$. The parameters p and ρ are chosen such that $P(T > \tau|x = 0) = p$ and $P(T < C|x = 0) = \rho$.

The asymptotic relative efficiencies are obtained by evaluating the expectations using Monte Carlo methods based on 100,000 Monte Carlo samples. The relative efficiencies of working independence and GEE1 estimators versus GEE2 estimators are presented in Table 3.1 for a variety of parameter configurations. Specifically, we set $\rho = 0.4, 0.6$, $\beta = \log 0.8, 0$, $\gamma = 1, 1.2$, $\sigma^2 = 0.75, 1$, $p = 0.05$ and $\tau = 1$. The association parameter was chosen to give Kendall's τ of 0.2 and 0.6, leading to values of ϕ of 0.5 and 3 respectively.

The results suggest that the estimates of the parameters of the marginal distribution are quite efficiently estimated under working independence assumptions or GEE1 when the association is weak, but there can be substantial losses in efficiency when the association within clusters is stronger (e.g. $\phi = 3.0$); in practise such large values are unlikely to be realized but we explore them here to assess the rate of loss of information. For any degree of within cluster association, the larger the cluster sizes the greater the efficiency loss with WI or GEE1 analyses. The greatest loss in

efficiency with GEE1 analysis is seen for the association parameter κ, where even for relatively low degrees of association the loss of efficiency can be as great as 74%.

3.3.3 Robustness of GEE1 and GEE2 to Copula Misspecification

While there can be efficiency gains, GEE2 methods may not provide consistent estimates of the parameters if the association structure is misspecified. This bias can be investigated by examining the expectation of equation (3.5) given the underlying correct model following White (1982); this approach has been used to investigate misspecified semiparametric models in Rotnitzky and Wypij (1994) and Cook et al. (2004). For this analysis, we assume the true association structure is governed by a Gumbel copula (Joe 1997), an Archimedean copula with generator $\Phi(v) = (-\log v)^{\phi}$, but adopt the Clayton copula for the specification of the moments in the estimating equations. When taking the expectation of (3.5), note that $E[Y_i|X_i] = p_i(\theta)$ even under this misspecification since the mean structure is unchanged. However, $E[Z_i|X_i]$ is affected, and this can be obtained from the Gumbel copula by evaluating

$$\begin{aligned} E[Z_{ij}|X_i, C_i] &= P(T_{ij_1} \leq C_{ij_1}, T_{ij_2} \leq C_{ij_2}|X_i) \\ &= \exp\left[-\left\{(-\log p_{ij_1}(\theta))^{\phi} + (-\log p_{ij_2}(\theta))^{\phi}\right\}^{1/\phi}\right] \end{aligned}$$

Again, Monte Carlo methods can be used to evaluate the expected estimating equations, and find the limit to which $\widehat{\psi}$ converges. Table 3.2 contains the results of these calculations where we find zero bias for the parameters of the marginal distributions for GEE1 and negligible biases for those from GEE2. The influence of the misspecification of the higher moments is seen to be greatest for larger cluster sizes and again stronger associations. The biases in the estimates of Kendall's tau are considerable under both GEE1 and GEE2.

Table 3.1: Asymptotic relative efficiencies of estimators under working independence and GEE1 relative to GEE2.

				Working Independence			GEE1			
α	σ^2	ϕ	n	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\kappa}$
1.0	0.75	0.5	2	0.990	0.982	0.984	0.999	0.996	0.998	0.418
1.0	0.75	0.5	5	0.972	0.948	0.952	0.996	0.991	0.993	0.348
1.0	0.75	0.5	10	0.958	0.913	0.922	0.994	0.990	0.987	0.274
1.0	0.75	3.0	2	0.934	0.930	0.884	0.997	0.994	0.992	0.108
1.0	0.75	3.0	5	0.870	0.848	0.722	0.991	0.988	0.977	0.083
1.0	0.75	3.0	10	0.863	0.807	0.622	0.985	0.989	0.962	0.076
1.0	1.00	0.5	2	0.990	0.983	0.985	0.999	0.996	0.998	0.410
1.0	1.00	0.5	5	0.973	0.948	0.955	0.997	0.991	0.993	0.336
1.0	1.00	0.5	10	0.958	0.910	0.927	0.995	0.990	0.988	0.262
1.0	1.00	3.0	2	0.937	0.933	0.893	0.997	0.995	0.993	0.105
1.0	1.00	3.0	5	0.874	0.848	0.736	0.991	0.989	0.978	0.081
1.0	1.00	3.0	10	0.864	0.803	0.635	0.986	0.989	0.964	0.074
1.2	0.75	0.5	2	0.990	0.983	0.985	0.999	0.996	0.998	0.414
1.2	0.75	0.5	5	0.972	0.950	0.954	0.997	0.991	0.993	0.342
1.2	0.75	0.5	10	0.959	0.919	0.926	0.994	0.991	0.988	0.276
1.2	0.75	3.0	2	0.936	0.933	0.891	0.997	0.995	0.993	0.105
1.2	0.75	3.0	5	0.872	0.853	0.732	0.991	0.989	0.978	0.081
1.2	0.75	3.0	10	0.864	0.816	0.633	0.986	0.989	0.963	0.074
1.2	1.00	0.5	2	0.991	0.984	0.986	0.999	0.997	0.998	0.401
1.2	1.00	0.5	5	0.973	0.951	0.957	0.997	0.992	0.994	0.332
1.2	1.00	0.5	10	0.959	0.917	0.930	0.995	0.991	0.989	0.266
1.2	1.00	3.0	2	0.940	0.935	0.899	0.997	0.995	0.993	0.103
1.2	1.00	3.0	5	0.875	0.853	0.745	0.992	0.989	0.979	0.080
1.2	1.00	3.0	10	0.865	0.813	0.647	0.987	0.989	0.965	0.072

Table 3.2: Asymptotic bias in parameter estimators from assuming a Clayton copula when the true copula is a Gumbel copula under GEE1 and GEE2 analyses.

				GEE1				GEE2			
α	σ^2	$\tau^\dagger$	n	$\lambda^* - \lambda$	$\alpha^* - \alpha$	$\beta^* - \beta$	τ^*	$\lambda^* - \lambda$	$\alpha^* - \alpha$	$\beta^* - \beta$	τ^*
1.0	0.75	0.2	2	0	0	0	0.103	0.001	-0.003	0.001	0.139
1.0	0.75	0.2	5	0	0	0	0.102	0.003	-0.010	0.001	0.143
1.0	0.75	0.2	10	0	0	0	0.102	0.000	-0.017	0.002	0.149
1.0	0.75	0.6	2	0	0	0	0.541	0.000	-0.012	0.001	0.503
1.0	0.75	0.6	5	0	0	0	0.585	-0.040	-0.032	0.006	0.508
1.0	0.75	0.6	10	0	0	0	0.563	-0.122	-0.028	0.002	0.493
1.0	1.00	0.2	2	0	0	0	0.104	0.002	-0.004	0.000	0.145
1.0	1.00	0.2	5	0	0	0	0.099	0.006	-0.011	0.002	0.147
1.0	1.00	0.2	10	0	0	0	0.100	0.004	-0.017	0.001	0.151
1.0	1.00	0.6	2	0	0	0	0.541	0.001	-0.013	0.002	0.501
1.0	1.00	0.6	5	0	0	0	0.582	-0.028	-0.031	0.001	0.511
1.0	1.00	0.6	10	0	0	0	0.562	-0.117	-0.028	0.002	0.498
1.2	0.75	0.2	2	0	0	0	0.101	0.001	-0.004	0.001	0.141
1.2	0.75	0.2	5	0	0	0	0.105	0.003	-0.012	0.001	0.147
1.2	0.75	0.2	10	0	0	0	0.103	0.002	-0.020	0.002	0.152
1.2	0.75	0.6	2	0	0	0	0.563	0.001	-0.016	0.001	0.517
1.2	0.75	0.6	5	0	0	0	0.540	-0.024	-0.033	0.003	0.502
1.2	0.75	0.6	10	0	0	0	0.565	-0.076	-0.035	0.001	0.499
1.2	1.00	0.2	2	0	0	0	0.103	0.003	-0.004	0.000	0.147
1.2	1.00	0.2	5	0	0	0	0.100	0.005	-0.012	0.001	0.147
1.2	1.00	0.2	10	0	0	0	0.101	0.004	-0.020	0.002	0.152
1.2	1.00	0.6	2	0	0	0	0.530	0.005	-0.016	-0.001	0.510
1.2	1.00	0.6	5	0	0	0	0.587	-0.018	-0.036	0.003	0.514
1.2	1.00	0.6	10	0	0	0	0.581	-0.072	-0.035	-0.001	0.504

3.3.4 Computational Notes for the Clayton Copula

Use of the GEE methods require derivatives of the copula function in order to obtain the matrices G_i and Q_i. For the Clayton copula, elements of G_i are obtained from

$$\begin{aligned}\frac{\partial H(p_{ij},p_{ik};\phi)}{\partial\psi} &= \frac{\partial H(p_{ij},p_{ik};\phi)}{\partial\phi}\frac{\partial\phi}{\partial\psi}\\ &= \eta_{il}\left\{\frac{1}{\phi^2}\log(p_{ij}^{-\phi}+p_{ik}^{-\phi}-1)\right.\\ &\qquad\left.+\frac{1}{\phi}\frac{1}{(p_{ij}^{-\phi}+p_{ik}^{-\phi}-1)}\left(p_{ij}^{-\phi}\log p_{ij}+p_{ik}^{-\phi}\log p_{ik}\right)\right\}\phi\\ &= \eta_{il}\left\{-\log\eta_{il}+\eta_{il}^{\phi}\left(p_{ij}^{-\phi}\log p_{ij}+p_{ik}^{-\phi}\log p_{ik}\right)\right\}\end{aligned}$$

where $\eta_{il}=H(p_{ij},p_{ik};\phi)$. Elements of Q_i can be obtained by noting that

$$\begin{aligned}\frac{\partial H(p_{ij},p_{ik};\phi)}{\partial\theta_h} &= -\frac{1}{\phi}(p_{ij}^{-\phi}+p_{ik}^{-\phi}-1)^{-1/\phi-1}\left\{-\phi p_{ij}^{-\phi-1}\frac{\partial p_{ij}}{\partial\theta_h}-\phi p_{ik}^{-\phi-1}\frac{\partial p_{ik}}{\partial\theta_h}\right\}\\ &= \eta_{il}^{1+\phi}\left\{p_{ij}^{-\phi-1}\frac{\partial p_{ij}}{\partial\theta_h}+p_{ik}^{-\phi-1}\frac{\partial p_{ik}}{\partial\theta_h}\right\}.\end{aligned}$$

Some computational difficulties can arise when ϕ becomes either too small or too large. Computation of η_i, G_i and Q_i become problematic. A way around this is to replace those computations with the corresponding limit as ϕ goes to zero or infinity as required.

Consider first when ϕ goes to zero. For η_{il}, it is convenient to work with $\log\eta_{il}$, since we obtain a $0/0$ which can be evaluated using l'Hôpital's rule.

$$\begin{aligned}\lim_{\phi\to0}\log\eta_{il} &= \lim_{\phi\to0}-\frac{1}{\phi}\log(p_{ij}^{-\phi}+p_{ik}^{-\phi}-1)\\ &= \lim_{\phi\to0}\frac{p_{ij}^{-\phi}\log p_{ij}+p_{ik}^{-\phi}\log p_{ik}}{p_{ij}^{-\phi}+p_{ik}^{-\phi}-1}\\ &= \log p_{ij}+\log p_{ik}\end{aligned}$$

hence $\lim_{\phi\to0}\eta_{il}=p_{ij}p_{ik}$. We also have that

$$\lim_{\phi\to0}\frac{\partial\eta_{il}}{\partial\psi}=\lim_{\phi\to0}\eta_{il}\left\{-\log\eta_{il}+\eta_{il}^{\phi}\left(p_{ij}^{-\phi}\log p_{ij}+p_{ik}^{-\phi}\log p_{ik}\right)\right\}=0$$

As for the elements of Q_i,

$$\lim_{\phi\to0}\frac{\partial\eta_{il}}{\partial\theta_h}=\frac{\partial p_{ij}}{\partial\theta_h}p_{ik}+p_{ij}\frac{\partial p_{ik}}{\partial\theta_h}$$

When ϕ tends to infinity, the limit of η_{il} can be obtained by noting that $p_{ij}^{-\phi} + p_{ik}^{-\phi} - 1$ is dominated by $\min(p_{ij}, p_{ik})^{-\phi}$ and hence $\lim_{\phi\to\infty} \eta_{il} = \min(p_{ij}, p_{ik})$. Similarly for G_i, $p_{ij}^{-\phi} \log p_{ij} + p_{ik}^{-\phi} \log p_{ik}$ is dominated by $\min(p_{ij}, p_{ik})^{-\phi} \log(\min(p_{ij}, p_{ik}))$ leading to $\lim_{\phi\to\infty} G_i = 0$. The elements of Q_i are obtained from $\lim_{\phi\to\infty} \partial\eta_{il}/\partial\theta_h = \partial p_{ij}/\partial\theta_h I(p_{ij} \le p_{ik}) + \partial p_{ik}/\partial\theta_h I(p_{ij} > p_{ik})$.

3.4 Simulation Studies

3.4.1 Simulating Clustered Current Status Data via Copulas

Consider the general family of Archimedean copulas,

$$H_n(u) = \Phi^{-1}(\Phi(u_1) + \cdots + \Phi(u_n))$$

where $\Phi(v)$ is known as the generator. The Clayton copula is obtained by using the generator $\Phi(v) = v^{-\phi} - 1$. Hence, $\Phi^{-1}(v) = (v+1)^{-1/\phi}$ and

$$H_n(u;\phi) = (u_1^{-\phi} + \cdots + u_n^{-\phi} - n + 1)^{-1/\phi}.$$

With a two-dimensional Clayton copula, $H_2(u;\phi) = (u_1^{-\phi} + u_2^{-\phi} - 1)^{-1/\phi}$. Since U_1 has a uniform marginal distribution, the distribution of U_2 given $U_1 = u_1$ is $\partial H/\partial u_1$, *i.e.*

$$P(U_2 \le u_2 | U_1 = u_1) = (u_1^{-\phi} + u_2^{-\phi} - 1)^{-1/\phi-1} u_1^{-\phi-1}.$$

The pair (u_1, u_2) can be generated as follows. Generate variables u_1 and z as independent uniform random variables over $(0, 1]$. The variable u_2 can be obtained by solving $P(U_2 \le u_2 | U_1 = u_1) = z$ for u_2, *i.e.*

$$u_2 = \left[\left(z^{-\phi/(\phi+1)} - 1\right) u_1^{-\phi} + 1\right]^{-1/\phi}.$$

In higher dimensions, to obtain the distribution of U_n given the values of $U_1, \ldots, U_{n-1}$, H_n must be differentiated with respect to $u_1, \ldots, u_{n-1}$. The joint density of $u_1, \ldots, u_{n-1}$ can be found by noting that $H_{n-1}(u_1, \ldots, u_{n-1}; \phi)$ gives the joint distribution of $U_1, \ldots, U_{n-1}$, so differentiating H_{n-1} with respect to $u_1, \ldots, u_{n-1}$ gives the required joint density. The distribution of U_n given the previous values $U_1, \ldots, U_{n-1}$ is obtained as

$$P(U_n \le u_n | U_1 = u_1, \ldots, U_{n-1} = u_{n-1}) = \frac{\partial^{n-1} H_n}{\partial u_1 \cdots \partial u_{n-1}} \Big/ \frac{\partial^{n-1} H_{n-1}}{\partial u_1 \cdots \partial u_{n-1}}$$

and with the Clayton copula,

$$P(U_n \leq u_n | U_1 = u_1, \ldots, U_{n-1} = u_{n-1}) \\ = \left(\frac{u_1^{-\phi} + \cdots + u_n^{-\phi} - n + 1}{u_1^{-\phi} + \cdots + u_{n-1}^{-\phi} - (n-1) + 1} \right)^{-1/\phi - (n-1)}.$$

The required random variables can be generated recursively as follows. Generate u_1 as a uniform $(0, 1]$ random variable. The kth variable can be obtained by solving

$$\left(\frac{u_1^{-\phi} + \cdots + u_k^{-\phi} - k + 1}{u_1^{-\phi} + \cdots + u_{k-1}^{-\phi} - (k-1) + 1} \right)^{-1/\phi - (k-1)} = z$$

for u_k where z is another independent uniform $(0, 1]$ random variable. This gives

$$u_k = \left[\left(z^{-\theta/(1+(k-1)\theta)} - 1 \right) \left(u_1^{-\phi} + \cdots + u_{k-1}^{-\phi} - (k-1) + 1 \right) + 1 \right]^{-1/\theta}$$

3.4.2 Assessing the Empirical Performance of Estimators

Here we evaluate the empirical performance of the estimators obtained from the working independence assumption, and methods based on GEE1. We consider the same parameter configurations as given in Section 3.3.2 and suppose there are 1000 clusters of size 5. We assume a Clayton copula with association parameter ϕ to generate the joint distribution of the event times. The parameter values considered were $\rho = 0.4, 0.6$, $\beta = \log 0.8, 0$, $\gamma = 1, 1.2$, $\phi = 0.5, 3.0$, $\sigma^2 = 0.75, 1$, $p = 0.05$, and $\tau = 1$. For a given γ, λ is chosen such that $P(T > 1) = p$. Analyses were carried out under the assumption of an exponential and Weibull (correct) marginal event time distributions, as well as under piecewise constant models with 3 and 5 pieces defined by equally spaced cut-points over $(0, 1]$. The empirical bias ($\times 10^2$), empirical standard errors and average estimated standard errors are displayed in Tables 3.3 and 3.4 for $\rho = 0.4$ and $\beta = \log 0.80$. Full simulation results can be found in the appendix.

The empirical bias of β in Table 3.3 is generally quite small but is largest for the exponential model when there is a trend ($\gamma = 1.2$) in the hazard function; The piecewise constant specification yields estimators with smaller bias in these settings. There is close agreement between the empirical (ESE) and average estimated standard errors (SE) throughout the table, and in settings where the association within clusters is large, the efficiency gains from the GEE1 analysis are apparent. In Table 3.4 we report on estimators of the probabilty $S_{ij} > 0.50$ for individuals on the

control treatment, as well as the associated median. The biases here can be more substantial; typically the bias under the exponential model is greater than that of the piecewise constant estimates if $\gamma = 1.2$, although this is not true for estimation of the medians. Again, however, the biases under the Weibull model are negligible and the efficiency gains from GEE1 versus a working independence assumption are clear, but more modest.

3.5 Applications

3.5.1 Analysis of Seroconversion in Orthopedic Surgery

Patients undergoing orthopedic surgery for hip or knee replacement are at risk of developing thrombosis and experiencing the associated complications, including death. As a result, orthopedic patients are routinely administered anticoagulants such as low molecular weight heparin (LMWH). In a small fraction of patients LMWH can induce a serological response and the resulting platelet-activating antibodies put patients at risk of thrombocytopenia (Warkentin et al. 2005). We consider data from recent international orthopedic studies providing data on the serological response to LMWH (Bauer et al. 2001; Turpie et al. 2002; Lassen et al. 2002; Eriksson et al. 2001), and consider the objective of identifying which factors are associated with seroconversion.

Patients are antibody negative at the time of surgery, and receive the injection of LMWH within 4-12 hours of surgery. Following surgery patients recover in hospital for 3-10 days and provide a blood sample at the time of discharge. These blood samples are then tested for the presence of antibodies. The time of seroconversion is therefore subject to type I interval censoring and the resulting data are current status data. Due to regional variations in race, socioeconomic status, surgical technique, etc., it we consider centers as defining clusters of individuals and take this into account in our analyses. There were 340 centers altogether and the numbers of subjects per center ranged from 1 to 63 (first, second and third quartiles were 3, 7, and 12); the total number of subjects included in this analysis was 3150.

The marginal methods of Section 3.3 were fit to this data based on a working independence assumption as well as GEE1 and GEE2 under a Clayton copula. Models were fit with a piecewise exponential baseline hazard function, with break points at 3.333 and 6.667 days, and with a Weibull baseline hazard. The estimated cumulative distribution functions giving the probability of seroconversion are given

Table 3.3: Empirical bias ($\times 10^2$) and empirical (ESE) and average standard errors (SE) for $\widehat{\beta}$ ($\rho = 0.4$ and $\beta = \log 0.80$).

				Working Indep.			GEE1		
α	σ^2	ϕ	Model	BIAS	ESE	SE	BIAS	ESE	SE
1.0	0.75	0.5	Exp	−0.065	0.052	0.053	−0.138	0.051	0.052
1.0	0.75	0.5	3 Piece	−0.105	0.053	0.053	−0.156	0.052	0.052
1.0	0.75	0.5	5 Piece	−0.112	0.053	0.053	−0.157	0.052	0.052
1.0	0.75	0.5	Weibull	−0.099	0.052	0.053	−0.153	0.052	0.052
1.0	0.75	3.0	Exp	−0.018	0.054	0.053	−0.061	0.047	0.045
1.0	0.75	3.0	3 Piece	−0.065	0.054	0.053	−0.083	0.047	0.046
1.0	0.75	3.0	5 Piece	−0.088	0.054	0.053	−0.123	0.047	0.046
1.0	0.75	3.0	Weibull	−0.050	0.054	0.053	−0.090	0.047	0.046
1.0	1.00	0.5	Exp	−0.294	0.052	0.053	−0.262	0.051	0.052
1.0	1.00	0.5	3 Piece	−0.326	0.052	0.053	−0.295	0.052	0.052
1.0	1.00	0.5	5 Piece	−0.353	0.053	0.053	−0.331	0.052	0.052
1.0	1.00	0.5	Weibull	−0.323	0.052	0.053	−0.295	0.052	0.052
1.0	1.00	3.0	Exp	0.107	0.055	0.053	0.153	0.048	0.046
1.0	1.00	3.0	3 Piece	0.054	0.055	0.053	0.046	0.048	0.046
1.0	1.00	3.0	5 Piece	0.045	0.055	0.053	0.028	0.048	0.047
1.0	1.00	3.0	Weibull	0.065	0.055	0.053	0.063	0.048	0.047
1.2	0.75	0.5	Exp	2.168	0.047	0.046	2.222	0.047	0.046
1.2	0.75	0.5	3 Piece	0.032	0.052	0.053	0.058	0.051	0.052
1.2	0.75	0.5	5 Piece	−0.006	0.052	0.053	−0.010	0.051	0.052
1.2	0.75	0.5	Weibull	−0.008	0.052	0.053	−0.012	0.051	0.052
1.2	0.75	3.0	Exp	2.118	0.050	0.047	2.370	0.045	0.043
1.2	0.75	3.0	3 Piece	−0.034	0.055	0.053	0.035	0.048	0.046
1.2	0.75	3.0	5 Piece	−0.069	0.055	0.053	−0.092	0.048	0.046
1.2	0.75	3.0	Weibull	−0.053	0.055	0.053	−0.109	0.048	0.046
1.2	1.00	0.5	Exp	2.172	0.048	0.047	2.210	0.047	0.047
1.2	1.00	0.5	3 Piece	0.012	0.052	0.053	0.044	0.052	0.053
1.2	1.00	0.5	5 Piece	−0.020	0.052	0.054	−0.022	0.052	0.053
1.2	1.00	0.5	Weibull	−0.015	0.053	0.054	−0.020	0.052	0.053
1.2	1.00	3.0	Exp	1.944	0.048	0.047	2.218	0.044	0.043
1.2	1.00	3.0	3 Piece	−0.285	0.053	0.054	−0.085	0.046	0.047
1.2	1.00	3.0	5 Piece	−0.311	0.053	0.054	−0.238	0.047	0.047
1.2	1.00	3.0	Weibull	−0.315	0.053	0.054	−0.253	0.047	0.047

Table 3.4: Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution.

				$P(S_{ij} > 0.5\|x_{ij} = 0)$				Median			
				WI		GEE1		WI		GEE1	
α	σ^2	ϕ	model	BIAS	ESE	BIAS	ESE	BIAS	ESE	BIAS	ESE
1.0	0.75	0.5	Exp	−0.017	0.013	−0.006	0.013	−0.002	0.009	0.005	0.009
1.0	0.75	0.5	3 Piece	−0.044	0.013	−0.046	0.013	0.047	0.012	0.047	0.012
1.0	0.75	0.5	5 Piece	−0.090	0.018	−0.084	0.018	0.015	0.014	0.011	0.013
1.0	0.75	0.5	Weibull	−0.031	0.013	−0.034	0.013	−0.016	0.009	−0.018	0.009
1.0	0.75	3.0	Exp	−0.007	0.015	−0.005	0.014	0.007	0.010	0.007	0.010
1.0	0.75	3.0	3 Piece	−0.027	0.015	−0.005	0.014	0.113	0.014	0.105	0.013
1.0	0.75	3.0	5 Piece	0.033	0.019	0.053	0.018	0.001	0.016	0.015	0.014
1.0	0.75	3.0	Weibull	−0.025	0.015	−0.011	0.014	−0.012	0.010	−0.004	0.010
1.0	1.00	0.5	Exp	−0.044	0.013	−0.018	0.013	−0.021	0.009	−0.004	0.009
1.0	1.00	0.5	3 Piece	−0.064	0.013	−0.059	0.013	0.011	0.012	0.018	0.012
1.0	1.00	0.5	5 Piece	−0.120	0.017	−0.112	0.017	−0.055	0.015	−0.047	0.014
1.0	1.00	0.5	Weibull	−0.050	0.013	−0.044	0.013	−0.030	0.009	−0.025	0.009
1.0	1.00	3.0	Exp	0.104	0.015	0.117	0.014	0.084	0.011	0.091	0.010
1.0	1.00	3.0	3 Piece	0.080	0.016	0.099	0.014	0.173	0.015	0.162	0.013
1.0	1.00	3.0	5 Piece	0.069	0.019	0.082	0.018	0.094	0.017	0.110	0.015
1.0	1.00	3.0	Weibull	0.087	0.015	0.108	0.014	0.063	0.010	0.077	0.010
1.2	0.75	0.5	Exp	1.063	0.013	0.728	0.013	1.581	0.010	1.325	0.010
1.2	0.75	0.5	3 Piece	0.381	0.014	0.374	0.014	3.958	0.015	3.948	0.014
1.2	0.75	0.5	5 Piece	−0.446	0.016	−0.451	0.016	3.963	0.014	3.967	0.014
1.2	0.75	0.5	Weibull	−0.007	0.014	−0.003	0.014	−0.016	0.008	−0.016	0.008
1.2	0.75	3.0	Exp	1.078	0.014	−0.012	0.013	1.595	0.011	0.766	0.010
1.2	0.75	3.0	3 Piece	0.454	0.017	0.458	0.015	4.079	0.017	4.022	0.015
1.2	0.75	3.0	5 Piece	−0.441	0.018	−0.384	0.017	4.064	0.017	4.072	0.014
1.2	0.75	3.0	Weibull	0.010	0.015	0.032	0.014	−0.005	0.009	0.009	0.009
1.2	1.00	0.5	Exp	0.837	0.013	0.553	0.013	1.409	0.010	1.192	0.010
1.2	1.00	0.5	3 Piece	0.777	0.016	0.766	0.016	4.051	0.016	4.035	0.016
1.2	1.00	0.5	5 Piece	−0.452	0.017	−0.451	0.017	4.141	0.015	4.150	0.015
1.2	1.00	0.5	Weibull	0.006	0.014	0.016	0.014	0.000	0.009	0.002	0.009
1.2	1.00	3.0	Exp	0.779	0.015	−0.210	0.014	1.368	0.011	0.618	0.011
1.2	1.00	3.0	3 Piece	0.686	0.017	0.689	0.016	3.984	0.017	3.918	0.016
1.2	1.00	3.0	5 Piece	−0.550	0.019	−0.471	0.018	4.076	0.016	4.081	0.015
1.2	1.00	3.0	Weibull	−0.059	0.016	−0.018	0.015	−0.048	0.010	−0.018	0.009

Table 3.5: Estimates and 95% confidence intervals for proportion testing positive 10 days after surgery in the orthopedic surgery study.

Model		Estimate	95% CI
Working Independence	3 Piece	0.043	$(0.024, 0.061)$
	Weibull	0.052	$(0.037, 0.074)$
GEE	3 Piece	0.043	$(0.024, 0.062)$
	Weibull	0.053	$(0.038, 0.074)$
GEE2	3 Piece	0.045	$(0.026, 0.064)$
	Weibull	0.055	$(0.039, 0.076)$

in Figure 3.2 for the three piece and Weibull models under GEE1 and GEE2, along with the nonparametric estimate of the cumulative distribution function using the pooled-adjacent violators algorithm (Sun 2006). The estimate based on the working independence assumption is almost indistinguishable from the GEE1 estimate and hence is not plotted. Estimates and 95% confidence intervals for the proportion of patients testing positive 10 days after surgery are shown in Table 3.5. The parametric estimates agree quite closely with the nonparametric estimates and the fits from a particular parametric model are quite similar under the two estimation schemes.

Regression models were considered with covariates sex (male versus female), an indicator of whether the surgery was for hip or knee replacement, and an indicator of whether there was an injection prior to surgery (yes versus no). The estimated regression coefficients are displayed in Table 3.6. Also provided are the estimates of κ under GEE1 and GEE2 analyses which give corresponding small estimates of ϕ (0.01 for both GEE1 analyses and 0.05 for both GEE2 analyses) suggesting a weak association within centers in the seroconversion times. The confidence intervals for Kendall's tau are (0.010,0.053) and (0.010,0.055) for the GEE2 analyses under piecewise constant and Weibull hazard functions respectively. Despite this weak association, there is some evidence of increased efficiency from the GEE2 analysis compared to the working independence or GEE1 analyses with slightly smaller standard errors observed for GEE2 analyses for the sex and type of surgery covariates.

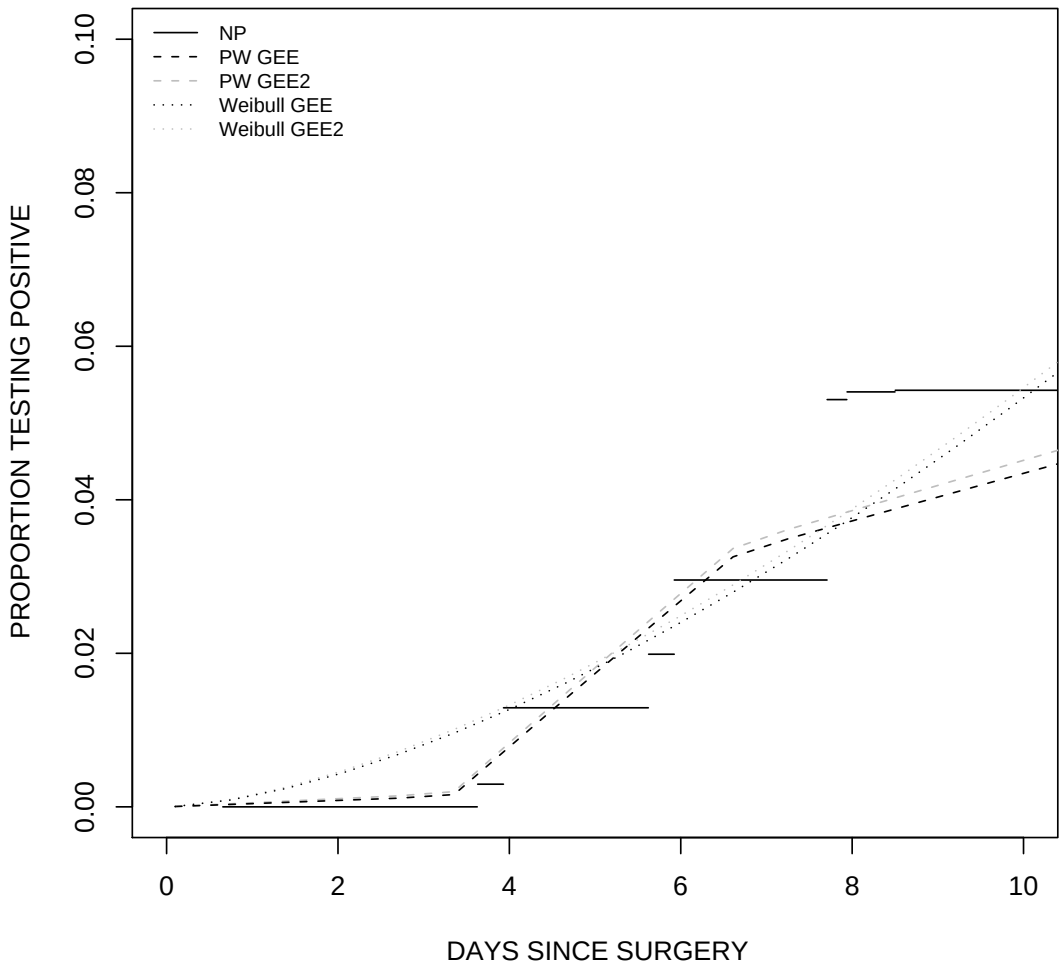

Figure 3.2: Proportion of subjects testing positive under a three piece piecewise constant and Weibull model using GEE and GEE2 in the orthopedic surgery study; the nonparametric estimate is also plotted.

Table 3.6: Regression coefficient estimates for 3 piece and Weibull models in the orthopedic surgery study.

		3 Piece			Weibull		
		Est	SE	p	Est	SE	p
Injection Prior to Surgery	WI	0.223	0.258	0.386	0.266	0.261	0.309
	GEE1	0.236	0.256	0.356	0.286	0.259	0.270
	GEE2	0.285	0.258	0.269	0.329	0.260	0.205
Hip vs Knee Surgery	WI	-1.213	0.347	< 0.001	-1.140	0.348	0.001
	GEE1	-1.224	0.344	< 0.001	-1.153	0.345	< 0.001
	GEE2	-1.221	0.340	< 0.001	-1.132	0.344	< 0.001
Male vs Female	WI	-0.261	0.252	0.299	-0.242	0.252	0.337
	GEE1	-0.259	0.251	0.301	-0.238	0.251	0.342
	GEE2	-0.246	0.244	0.314	-0.225	0.245	0.358
κ	GEE1	-5.099	3.543	—	-4.608	2.028	—
	GEE2	-3.087	0.460	—	-3.005	0.438	—

3.5.2 Joint Damage at Clinic Entry in Psoriatic Arthritis

Upon first entry to the clinic patients undergo a detailed clinical and radiological assessment with each of 44 joints of the 64 joint examined and graded according to the severity of damage using the modified Steinbrocker scale (Rahman et al. 1998). Here we consider joints to be damaged if they have a score of 1 or higher, corresponding to the presence of soft tissue swelling, surface erosions, joint space narrowing, disorganization or need for surgery. Due to genetic and environmental factors, progression rates in joint within the same patient tend to more similar than progression rates in joints from different patients, and so there is a need to account for clustering within patients. Interest lies in features of and covariate effects on the marginal distribution of the time to joint damage in this patient population, suggesting the use of the methods of Section 3.3. Here we restrict attention to the joints of the hands to ensure they are a comparable group.

Figure 3.3 displays estimates of the cumulative distribution of the time to damage based on the nonparametric MLE (Sun 2006), and five piece exponential and Weibull hazards estimated using GEE1 and GEE2; again the five estimates appear quite compatible. The cut points for the 5 piece model were chosen according to the quintiles of the time from diagnosis to first clinic visit, and were 1.1, 3.0, 6.5 and 12.1 years. Under the 5 piece model the estimates of Kendall's tau were 0.236 (95% CI: 0.195, 0.282) and 0.247 (95% CI: 0.204, 0.297) for the GEE1 and GEE2 analyses respectively.

Table 3.8 displays estimates of the parameters in the proportional hazards models with covariates of interest being sex (male versus female), race (caucasian versus other), family history of psoriasis (yes versus no), family history of psoriatic arthritis (yes versus no), and age at diagnosis. GEE2 results are at times quite different. This is a consequence of the dependence of GEE2 estimates on the correlation structure and as a result the reliability of the GEE2 estimates is questionable. In this example, it may be best to rely on the working independence and GEE1 estimates pending further investigation into the validity of the chosen copula.

3.6 Summary

Here we have considered current status data where the failure times arise in clusters. This can be viewed as dependent binary data, where the dependence arises due to the clustering. The association was modelled using a copula which allowed for

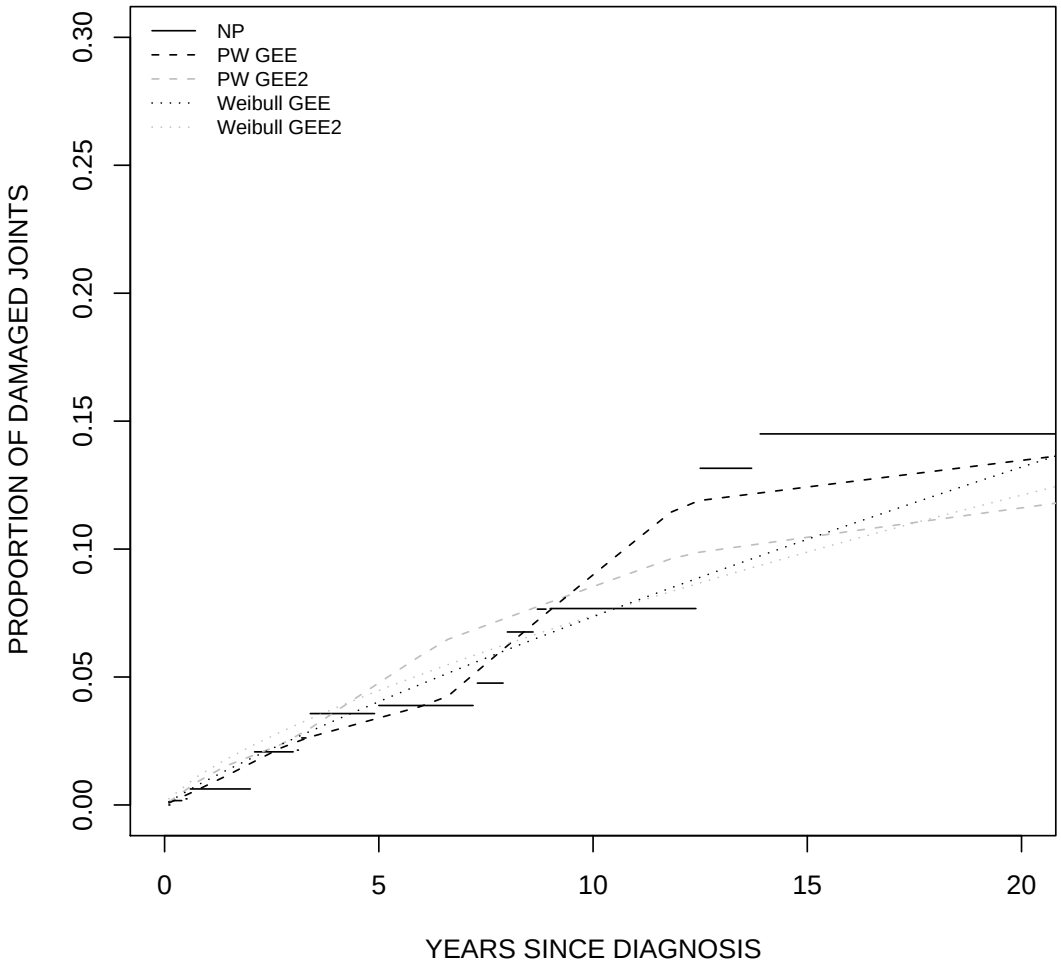

Figure 3.3: Proportion of damaged joints under a 5 piece and Weibull model; the nonparametric estimate is also plotted.

Table 3.7: Estimates and 95% confidence intervals for proportion of damaged joints 5 years after diagnosis based on the University of Toronto Psoriatic Arthritis Clinic data.

Model		Est	95% CI
Working Independence	5 Piece	0.034	(0.021,0.047)
	Weibull	0.040	(0.033,0.049)
GEE	5 Piece	0.034	(0.021,0.047)
	Weibull	0.040	(0.033,0.049)
GEE2	5 Piece	0.048	(0.033,0.062)
	Weibull	0.045	(0.037,0.054)

parameterizing the association between the event times themselves. Covariates effects were included using a proportional hazards form, although this work carries over to the accelerated failure time model as well by assuming

$$\mathcal{F}(s, |x_{ij}; \theta) = \mathcal{F}_0 \left(\frac{s - \alpha - x'_{ij}\beta}{\sigma} \right)$$

where $\mathcal{F}_0(s)$ is a specified survivor function, and $\theta = (\alpha, \sigma, \beta')'$. The derivatives of p_{ij} may be obtained and the methods applied in the same manner as described for the proportional hazards setting. In much of the work, the Clayton copula was used, however this copula need not be used in every setting. Depending on the nature of the association, it may be preferable to choose one copula over another. Both GEE1 and GEE2 approaches were considered. A study of the asymptotic relative efficiency suggested the GEE1 approach has high efficiency relative to the GEE2 approach for estimating marginal parameters, but may be inefficient for estimating the association parameter. The GEE2 approach may lead to biased estimates as a result of mis-specifying the dependence structure. This is illustrated in the psoriatic arthritis example. These results may be found in Tolusso and Cook (2008c).

First order GEE's can often be used to obtain robust estimators. In this particular case, the GEE1 approach leads to robust estimation of the marginal parameters regardless of the association structure. Consistent estimates are obtained as a result of the separate construction of the estimating equations. Variance estimates that remain valid regardless of the underlying association can be obtained using the usual sandwich forms.

Table 3.8: Estimates of regression coefficients obtained from fitting the five piece piecewise constant and Weibull regression models to data from the University of Toronto Psoriatic Arthritis Clinic.

		5 Piece			Weibull		
		Est	SE	p	Est	SE	p
Male vs Female	WI	0.284	0.191	0.136	0.304	0.193	0.116
	GEE	0.275	0.191	0.151	0.292	0.194	0.132
	GEE2	0.089	0.156	0.568	0.096	0.158	0.542
Caucasian vs Other	WI	0.380	0.374	0.310	0.356	0.363	0.327
	GEE	0.385	0.374	0.305	0.359	0.364	0.323
	GEE2	−0.001	0.372	0.998	−0.003	0.371	0.993
Family History of Psoriasis	WI	−0.109	0.234	0.642	−0.117	0.234	0.618
	GEE	−0.108	0.235	0.646	−0.116	0.235	0.620
	GEE2	−0.338	0.183	0.066	−0.336	0.183	0.066
Family History of PsA	WI	0.148	0.338	0.661	0.164	0.336	0.624
	GEE	0.148	0.339	0.662	0.167	0.336	0.619
	GEE2	0.290	0.252	0.248	0.292	0.251	0.244
$30 <$ Age ≤ 40 vs Age ≤ 30	WI	0.202	0.267	0.450	0.256	0.256	0.317
	GEE	0.204	0.268	0.447	0.261	0.256	0.308
	GEE2	0.325	0.213	0.126	0.334	0.207	0.106
Age > 40 vs Age ≤ 30	WI	0.672	0.246	0.006	0.730	0.231	0.002
	GEE	0.660	0.247	0.008	0.721	0.232	0.002
	GEE2	0.826	0.206	< 0.001	0.845	0.199	< 0.001

Chapter 4

Multistate Models With Interval Censored Data: Applications

4.1 Introduction and Overview

When using multistate models to characterize the course of a disease process, it is often of interest to know the probability of being in a particular state. This may be obtained from state occupancy probability or prevalence functions. We consider here robust estimation of the state occupancy probability functions from multistate models based on interval-censored failure time data.

There are two approaches that will be considered. The first is to follow the idea of Pepe et al. (1991) where it is noted that in the context of a progressive model such as the one in Figure 2.1, $P(Z(t) = k) = P(T_{k+1} > t) - P(T_k > t)$, where T_k is the time to entry of state k. An estimate of the prevalence function can therefore be obtained by taking the difference of two estimates of survival functions, and the problem is reduced to estimating a survivor function under interval censoring. The survivor function estimators under consideration could be *i)* nonparametric MLE due to Turnbull (1976), *ii)* piecewise constant hazard due to Lindsey and Ryan (1998) and *iii)* local likelihood as described in Betensky et al. (1999). Since the NPMLE of a survivor function under interval censoring can be undefined (see Section 1.2.2), it is of primary interest to see how models with piecewise constant hazards and the local likelihood methods perform. Care must also be taken to ensure that the resulting prevalence function estimate is non-negative.

The second approach considered is to assume a Markov model holds. Estimation then proceeds by assuming time homogeneous, or piecewise constant transition

intensities. Again this is because the nonparametric estimates resulting from a Markov assumption are not always defined everywhere (Frydman 1992, 1995).

This work is partially motivated by a desire to obtain robust estimates of state occupancy probabilities to characterize the proportion of patients diagnosed with psoriatic arthritis in a particular disease state. Specifically we wish to examine the proportion of patients who develop back disease following entry to the University of Toronto Psoriatic Arthritis Clinic. Given the clinical motivation for these analyses, we also consider fitting multiplicative intensity based Markov models to identify and characterize the effect of important risk factors.

4.2 "Pepe" Estimation of Prevalence Functions

With right-censored data, Pepe et al. (1991) estimated state occupancy probabilities based on estimates of the marginal distributions of the entry times. For example, in the progressive three-state model of Figure 1.4, if we define T_1 and T_2 to be the time of entry to states 1 and 2 respectively, then $P(Z(t) = 0) = P(T_1 > t)$ and $P(Z(t) = 2) = P(T_2 \leq t)$, hence $P(Z(t) = 1) = 1 - P(T_1 > t) - P(T_2 \leq t)$, or $P(Z(t) = 1) = P(T_1 \leq t) - P(T_2 \leq t)$. This also could be obtained intuitively by thinking of the probability of being in state 1 as being the probability of having left state 0 ($P(T_1 \leq t)$) minus the probability of having entered state 2 ($P(T_2 \leq t)$). This suggests estimates of the prevalence functions can be obtained by plugging in the appropriate estimates of the distribution functions for T_1 and T_2. This approach may be extended to progressive models with an arbitrary number of states.

For other models, such as the illness-death model of Figure 1.5, this approach can still be applied by carefully choosing which marginal distribution functions are estimated. In the illness-death model, define S_0 to be the time spent in state 0, and T_2 to be the time of entry to state 2. Then $P(Z(t) = 0) = P(S_0 > t)$ and $P(Z(t) = 2) = P(T_2 \leq t)$, hence $P(Z(t) = 1) = P(S_0 \leq t) - P(T_2 \leq t)$, or intuitively, the probability of having left state 0 minus the probability of having entered state 2. This is very similar to the progressive three-state model, with the difference being that in the progressive three-state model the time spent in state 0 is the same as the time of entry to state 1, while in the illness-death model they are not. Other models may be handled in a similar manner.

The marginal distribution functions may be estimated in a variety of ways. We will consider the use of the methods of Section 1.2.2, namely the nonparametric estimate of Turnbull (1976), piecewise constant hazards (Lindsey and Ryan 1998),

and local likelihood (Betensky et al. 1999; Braun et al. 2005), and refer to the resulting methods as "Pepe-Turnbull", "Pepe-PW", and "Pepe-Local-Likelihood", respectively. The marginal distribution functions must be estimated taking care to ensure the order restrictions are satisfied and hence avoiding negative prevalence function estimates, however this situation has rarely been found to occur, and did not occur in the applications to follow. Nevertheless, Section 4.2.1 considers this problem for the piecewise constant model.

4.2.1 Estimation under Piecewise Constant Models with Order Restrictions

Nonparametric estimation of a single survival curve under interval censoring produces an estimator that is not defined everywhere, so here estimation is considered using marginal piecewise constant hazard models. Consider now estimation of two ordered survival curves in the presence of interval censoring. Let $S_1(t)$ and $S_2(t)$ be survival functions satisfying $S_1(t) \leq S_2(t)$ for all t. Suppose we have interval-censored observations from each distribution. The two survival functions can be estimated using a piecewise constant model via the EM algorithm.

Let $\ell_C = \ell_{C1} + \ell_{C2}$ where ℓ_{Ch} is the complete data log-likelihood as in Section 1.2.2 for estimation of $S_h(t)$ under piecewise constant hazards, $h = 1, 2$. This amounts to assuming the two distributions are independent. In reality this assumption may not be correct, but it is used for the purposes of robustness. Let $a_1, \ldots, a_m$ be the common cut-points for both survival functions. Let $\lambda_h = (\lambda_{h1}, \ldots, \lambda_{hm})$ be the vector of rates where λ_{hj} denotes the rate over $(a_{j-1}, a_j]$ for $S_h(t)$, $h = 1, 2$, $j = 1, \ldots, m$.

At the E-step of the rth iteration, we require

$$E(\ell_C|\text{data}, \widehat{\lambda}^{(r-1)}) = E(\ell_{C1}|\text{data}, \widehat{\lambda}_1^{(r-1)}) + E(\ell_{C2}|\text{data}, \widehat{\lambda}_2^{(r-1)}).$$

At the M-step, the maximization must ensure the constraints are satisfied. The nonparametric version of this problem has been discussed in the right-censored case by Dykstra (1982) and Præstgaard and Huang (1996). If we let d_{hj} and r_{hj} be, respectively, the expected number of deaths and expected total time spent in the jth interval for distribution h, then the expected log-likelihood for distribution h can be written as,

$$E(\ell_{Ch}|\text{data}, \widehat{\lambda}^{(r-1)}) = \sum_{j=1}^{m} (d_{hj} \log \lambda_{hj} - r_{hj}\lambda_{hj})$$

where all expectations are conditional on the observed data and parameter estimates at the previous iteration, obtained as in Section 1.2.2.

The problem is now reduced to maximizing $E(\ell_C|\text{data}, \widehat{\lambda}^{(r-1)})$ subject to $S_1(t) \leq S_2(t)$. In fact, the restriction can be simplified since under a piecewise constant model it is equivalent to

$$\sum_{j=1}^{m} \lambda_{2j} w_j(a_l) \leq \sum_{j=1}^{m} \lambda_{1j} w_j(a_l) \quad \text{for all } l = 1, \ldots, m$$

where again, $w_j(t) = \int_{a_{j-1}}^{a_j} I(u \leq t)\, du$. This is now a nonlinear optimization problem with linear inequality constraints. In order to solve such a problem, the solution must satisfy the Kuhn-Tucker (KT) conditions (Chiang 1984, Chap. 21). The KT conditions essentially are a set of equations and inequalities relating to the derivatives of the Lagrangian of the objective function.

In this case, the objective function is

$$\ell = \ell_1 + \ell_2 + \sum_{l=1}^{m} \alpha_l \left[\sum_{j=1}^{m} \lambda_{1j} w_j(a_l) - \sum_{j=1}^{m} \lambda_{2j} w_j(a_l) \right]$$

where for ease of notation, $\ell_h = E(\ell_{Ch}|\text{data}, \widehat{\lambda}^{(r-1)})$, $h = 1, 2$, and $\alpha_1, \ldots, \alpha_m$ are Lagrange multipliers. The derivatives of ℓ are

$$\begin{aligned}
\frac{\partial \ell}{\partial \lambda_{1j}} &= \frac{d_{1j}}{\lambda_{1j}} - r_{1j} + (a_j - a_{j-1}) \sum_{l=j}^{m} \alpha_l \\
\frac{\partial \ell}{\partial \lambda_{2j}} &= \frac{d_{2j}}{\lambda_{2j}} - r_{2j} - (a_j - a_{j-1}) \sum_{l=j}^{m} \alpha_l \\
\frac{\partial \ell}{\partial \alpha_l} &= \sum_{j=1}^{l} \lambda_{1j}(a_j - a_{j-1}) - \sum_{j=1}^{l} \lambda_{2j}(a_j - a_{j-1})
\end{aligned}$$

so the resulting KT conditions are

$$\begin{array}{lll}
\dfrac{\partial \ell}{\partial \lambda_{1j}} \leq 0 & \lambda_{1j} \geq 0 & \lambda_{1j} \dfrac{\partial \ell}{\partial \lambda_{1j}} = 0 \\
\dfrac{\partial \ell}{\partial \lambda_{2j}} \leq 0 & \lambda_{2j} \geq 0 & \lambda_{2j} \dfrac{\partial \ell}{\partial \lambda_{2j}} = 0 \\
\dfrac{\partial \ell}{\partial \alpha_l} \geq 0 & \alpha_l \geq 0 & \alpha_l \dfrac{\partial \ell}{\partial \alpha_l} = 0
\end{array}$$

From the KT conditions, it is clear that when $d_{1j} > 0$,

$$\widehat{\lambda}_{1j} = \frac{d_{1j}}{r_{1j} - (a_j - a_{j-1}) \sum_{l=j}^{m} \alpha_l}$$

and if $d_{1j} = 0$ then $\widehat{\lambda}_{1j}$ must be 0 so the above form for $\widehat{\lambda}_{1j}$ holds for $d_{1j} \geq 0$. Similarly, it can be shown that

$$\widehat{\lambda}_{2j} = \frac{d_{2j}}{r_{2j} + (a_j - a_{j-1}) \sum_{l=j}^{m} \alpha_l}$$

To solve the problem, the values of the Lagrange multipliers must be determined.

The algorithm used by Dykstra (1982) for the nonparametric case can be modified for use in this case. It is useful to reparameterize the Lagrange multipliers and solve for $\gamma_j = \sum_{l=j}^{m} \alpha_l$. The modified algorithm is as follows.

1. Find the index j' such that $\gamma_{j'}$ is the largest solution to

$$\sum_{j=1}^{m} \frac{d_{1j}[w_j(a_{j'}) - w_j(0)]}{r_{1j} - (a_j - a_{j-1})\gamma_{j'}} = \sum_{j=1}^{m} \frac{d_{2j}[w_j(a_{j'}) - w_j(0)]}{r_{2j} + (a_j - a_{j-1})\gamma_{j'}}$$

   If there is more than one index satisfying the above condition then select j' so that $a_{j'}$ is largest.

2. Set $\gamma_1 = \gamma_2 = \cdots = \gamma_{j'}$

3. Find the index j'' such that $\gamma_{j''}$ is the largest solution to

$$\sum_{j=1}^{m} \frac{d_{1j}[w_j(a_{j''}) - w_j(a_{j'})]}{r_{1j} - (a_j - a_{j-1})\gamma_{j''}} = \sum_{j=1}^{m} \frac{d_{2j}[w_j(a_{j''}) - w_j(a_{j'})]}{r_{2j} + (a_j - a_{j-1})\gamma_{j''}}$$

   If there is more than one index satisfying the above condition then select j'' so that $a_{j''}$ is largest.

4. Set $\gamma_{j'+1} = \gamma_{j'+2} = \cdots = \gamma_{j''}$

5. Continue in a similar manner. If at some point no positive γ exists, then the remaining γ terms are set to 0.

It suffices to show that the third row of the KT conditions are satisfied. When j' is found, it implies that $\partial\ell/\partial\alpha_{j'} = 0$, $\alpha_{j'} \neq 0$ and $\alpha_1 = \alpha_2 = \cdots = \alpha_{j'-1} = 0$. Similar conclusions can be reached regarding the remaining indices.

The only piece of the KT conditions not yet satisfied is $\partial\ell/\partial\alpha_l \geq 0$. Define $g_j(\gamma)$ to be $\partial\ell/\partial\alpha_j$ evaluated at $\gamma_1 = \gamma_2 = \cdots = \gamma_j = \gamma$. Suppose there exists a j^* such that $a_{j^*} < a_{j'}$ and $g_{j^*}(\gamma_{j'}) < 0$. Then since $g_j(\gamma)$ increases with γ it implies there exists $\gamma_{j^*} > \gamma_{j'}$ such that $g_{j^*}(\gamma_{j^*}) = 0$. But if $\gamma_{j'}$ is found using the algorithm,

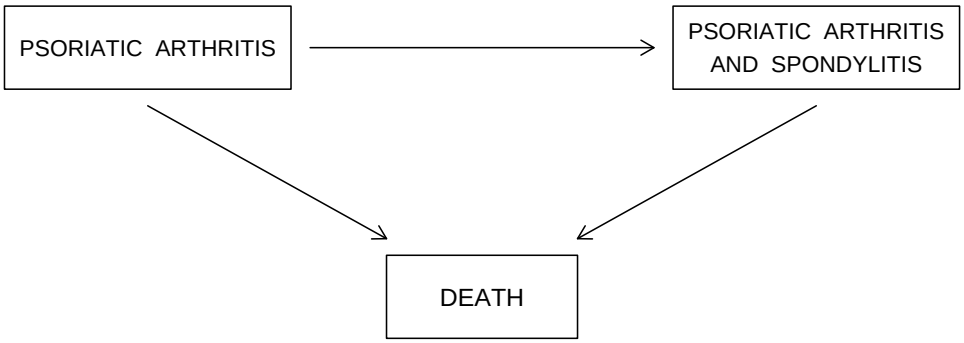

Figure 4.1: Three state diagram for onset of spondylitis in patients with psoriatic arthritis.

then no such γ_{j^*} can exist and we arrive at a contradiction, hence the last part of the KT conditions is satisfied.

Note that if the survival functions are estimated without imposing the constraints and the resulting survival curves satisfy the constraints, this is equivalent to $\alpha_l = 0$ for all l. Hence, the above algorithm need only be used if the estimates of $S_1(t)$ and $S_2(t)$ assuming independence do not satisfy the KT conditions.

4.3 Applications

4.3.1 Prevalence and Regression Analysis for Spondylitis in PsA

Figure 4.1 is the three state diagram we fit to characterize the onset of back involvement in patients with psoriatic arthritis. We label the psoriatic arthritis state (with no spondylitis) as state 0, psoriatic arthritis with spondylitis state as state 1 and the death state as state 2. Figure 4.2 contains timeline diagrams which indicate the state occupied by a selection of individuals over time. The length of the line represents the time from clinic entry to last contact or death. The solid lines correspond to periods in which patients were in state 0 and the dashed lines correspond to them being in state 1. The solid circles at the end of the lines indicate the times of deaths and the open circles indicate that the patient's survival time is right-censored. The breaks in the lines correspond to periods in which the state

occupied is unknown. The left endpoint of such intervals is the last assessment at which there was no back involvement and the right endpoint is the time of the first assessment that spondylitis was detected.

The sacroiliac joints are graded on a scale from 0 to 4 with 0 being normal, 1 being equivocal, 2 being abnormal with erosions or sclerosis, 3 being the presence of more than one of erosions, sclerosis, widening, narrowing or partial ankylosis, and 4 being total ankylosis. Patients are said to have limitation of movement if they have a reduced range of chest expansion, back movement, or neck mobility.

There are six definitions of spondylitis based on various combinations of radiological and clinical assessments, which were defined by researchers at the University of Toronto Psoriatic Arthritis Clinic.

Defintion 1

- Both left and right sacroiliac joints graded 2 or higher, or one sacroiliac joint graded 3 or higher

Definition 2

- One sacroiliac joint graded 2 or higher
- Pain and stiffness in the neck or back

Definition 3

- Radiologic evidence as in Definition 1
- Pain and stiffness in the neck or back

Definition 4

- Radiologic evidence as in Definition 1
- Pain and stiffness in the neck or back
- Limitation of movement

Definition 5

- Radiologic evidence as in Definition 1

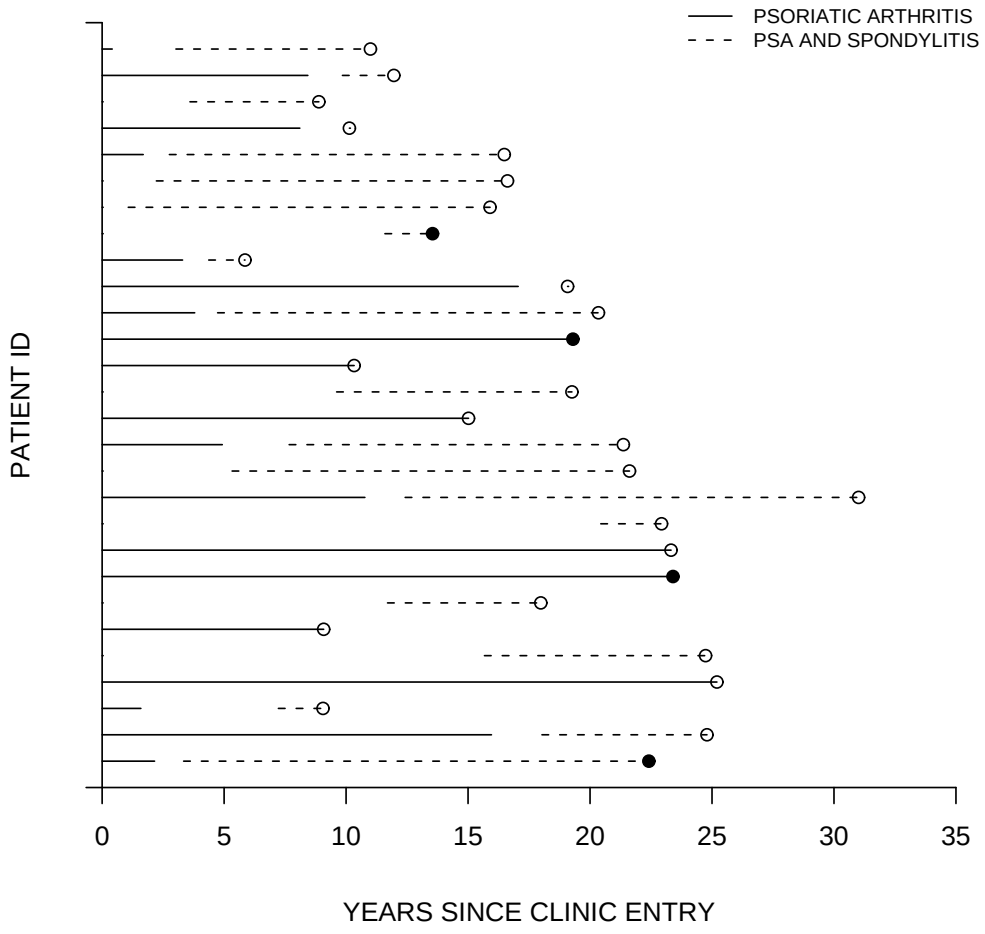

Figure 4.2: Timeline diagrams for sample of patients indicating the states occupied and censoring intervals for onset of spondylitis.

- Pain and stiffness in the neck or back, or limitation of movement

Definition 6

- One sacroiliac joint graded 1 or higher
- Pain and stiffness in the neck or back, or limitation of movement

The definitions provide varying amounts of inclusivity, with Definition 4 being the least inclusive, meaning that patients who meet Definition 4 will meet all other definitions as well. Here we consider the results from analyses using two of the six definitions for illustration.

We aim to examine the proportion of patients that develop spondylitis over time, the effect of spondylitis on mortality, and risk factors for spondylitis based on the definitions above. The data are comprised of the interval and right-censored transition times as in Figure 4.2, and baseline characteristics.

Spondylitis Definition 3: Prevalence and Risk Factors

Two hundred and fifty subjects had complete covariate information and 32 patients had spondylitis at clinic entry. Eleven patients died (2 with path $0 \to 1 \to 2$, 8 with path $0 \to 2$, and 1 with path $1 \to 2$). The followup in the clinic ranged from 0.50 years to 31.0 years, and the mean followup was 8.3 years (S.D. 6.9) with quartiles 2.7, 6.0, and 11.5 years. The covariates used to predict spondylitis and mortality are displayed in Tables 4.1 and 4.2 along with summaries of their distributions.

Figure 4.3 shows three prevalence estimates for state 1, representing the proportion of psoriatic arthritis patients alive with spondylitis, using the Pepe-Turnbull, Pepe-Local-Likelihood and Pepe-Piecewise methods, along with the state occupancy probabilities obtained from fitting a time-homogeneous Markov model. The local likelihood methods were employed with a nearest neighbours bandwidth of 0.1, Epanechnikov kernel $K(u) = 0.75(1-u^2)$, $|u| < 1$, locally constant approximation, and 400 grid points. The piecewise constant methods were fit with six equally spaced pieces with cut-points at 2.5, 5.0, 7.5, 10.0, 12.5, and 15.0 years. Bootstrap standard errors were obtained by generating 500 resampled datasets, and the 95% confidence intervals are given by the 0.025 and 0.975 quantiles. The prevalence for state 1 need not be a monotonic function since subjects must develop spondylitis to enter state 1, but leave state 1 when they die. The four estimators track each other quite well however, and we conclude that after 10 years of followup we can expect

Table 4.1: Frequency distributions of the continuous covariates used to predict spondylitis in psoriatic arthritis.

Covariate		mean	sd	min	1st quartile	median	3rd quartile	max
Age	Years	43.3	13.0	15.9	33.3	43.3	53.2	74.8
Arthritis Duration	Years	7.1	7.8	0.0	1.4	4.3	9.8	39.1
Psoriasis Duration	Years	14.6	12.5	0.0	4.9	10.8	21.6	69.8
# Actively inflammed joints	-	10.9	9.3	0.0	4.0	9.0	16.0	48.0
# Effused joints	-	3.6	4.3	0.0	0.0	2.0	6.0	21.0
# Clinically damaged joints	-	4.6	10.0	0.0	0.0	0.0	4.0	53.0
# Radiologically damaged joints	-	5.6	8.7	0.0	0.0	2.0	6.0	41.0
ESR	mm/h	27.4	21.6	0.0	10.0	23.5	39.8	105.0

Table 4.2: Summaries of the binary covariates used to predict spondylitis in psoriatic arthritis.

Covariate	# Yes	%
Male	144	57.6
Caucasian	242	96.8
Family history of psoriasis	98	39.2
Family history of psoriatic arthritis	23	9.2
Dactylitis	97	38.8
Nail involvement	195	78.0
Periostitis	48	19.2
Enthesitis	60	24.0
Spurs	103	41.2
Prior NSAID use	123	49.2
Prior DMARD use	88	35.2
B27	38	15.2
Smoker	54	21.6
Hypertension	30	12.0

close to 20% of PsA patients to develop spondylitis (LL - EST=0.172, bootstrap SE=0.037, 95% bootstrap CI 0.119:0.265; PW - EST=0.180, bootstrap SE=0.039, 95% bootstrap CI 0.093:0.247).

For the regression analyses, the R package `msm` (Jackson 2007) was used to estimate the baseline transition rates and covariate effects. The intensity for the onset of spondylitis was taken to be $\lambda_{01}(t)e^{x'_{01}\beta_{01}}$ where $\lambda_{01}(t)$ is the baseline intensity while x_{01} and β_{01} are the covariate vector and regression coefficient vector respectively. It was assumed that the covariates and regression coefficients for transitions into state 2 were the same, and the effect of spondylitis on mortality was assumed to be multiplicative. This leads to the intensity for death without spondylitis as being $\lambda_{02}(t)e^{x'_{02}\beta_{02}}$ and the intensity for death with spondylitis as $\lambda_{02}(t)e^{x'_{02}\beta_{02}+\gamma}$ where $\lambda_{02}(t)$ is the baseline intensity for mortality and x_{02} and β_{02} are the covariate and regression coefficient vectors, respectively, for mortality. The γ term is the multiplicative effect on mortality due to having spondylitis. Table 4.3 displays the estimates of the regression coefficients for the $0 \rightarrow 1$ transition as well as the coefficient for transitions to state 2 from univariate analyses. Also listed are the p-values from a likelihood ratio test of the null hypothesis $H_0 : \beta_{02} = \beta_{12}$. The effects on mortality of race, periostitis, and enthesitis had large standard errors due to the fact that only caucasians were observed to die, and no one with periostitis or enthesitis was observed to die. Also, only two non-caucasians developed spondylitis, causing the large standard error for the effect of race on spondylitis.

Data were somewhat sparse and it was difficult obtaining convergence in a full multivariate model with all prognostic variables. We therefore used a somewhat ad hoc approach to fitting a multivariate model: the univariate effects which were significant at the 0.25 level were included in a multivariate model. The model was then reduced by backwards elimination until the remaining effects were significant at the 0.05 level. At that point, all other effects were tested for re-entry to the model (no effects were added at this stage). Estimates of the final model regression coefficients are shown in Table 4.4.

Only the number of radiologically damaged joints and ESR were associated with an increased risk of spondylitis in the multivariate analysis. Each additional radiologically damaged joint increases the risk of spondylitis by 5% (RR=1.05, 95% CI 1.01:1.09) while an increase of 1 mm/h in ESR increases spondylitis risk by 2% (RR=1.02, 95% CI 1.00:1.03), controlling for other factors. In the mortality regression models, age and smoking status were associated with increased risk of death. Specifically, for each additional year of age at presentation there was a 10% increased risk of death (RR=1.10, 95% CI 1.04: 1.17) and smokers had a

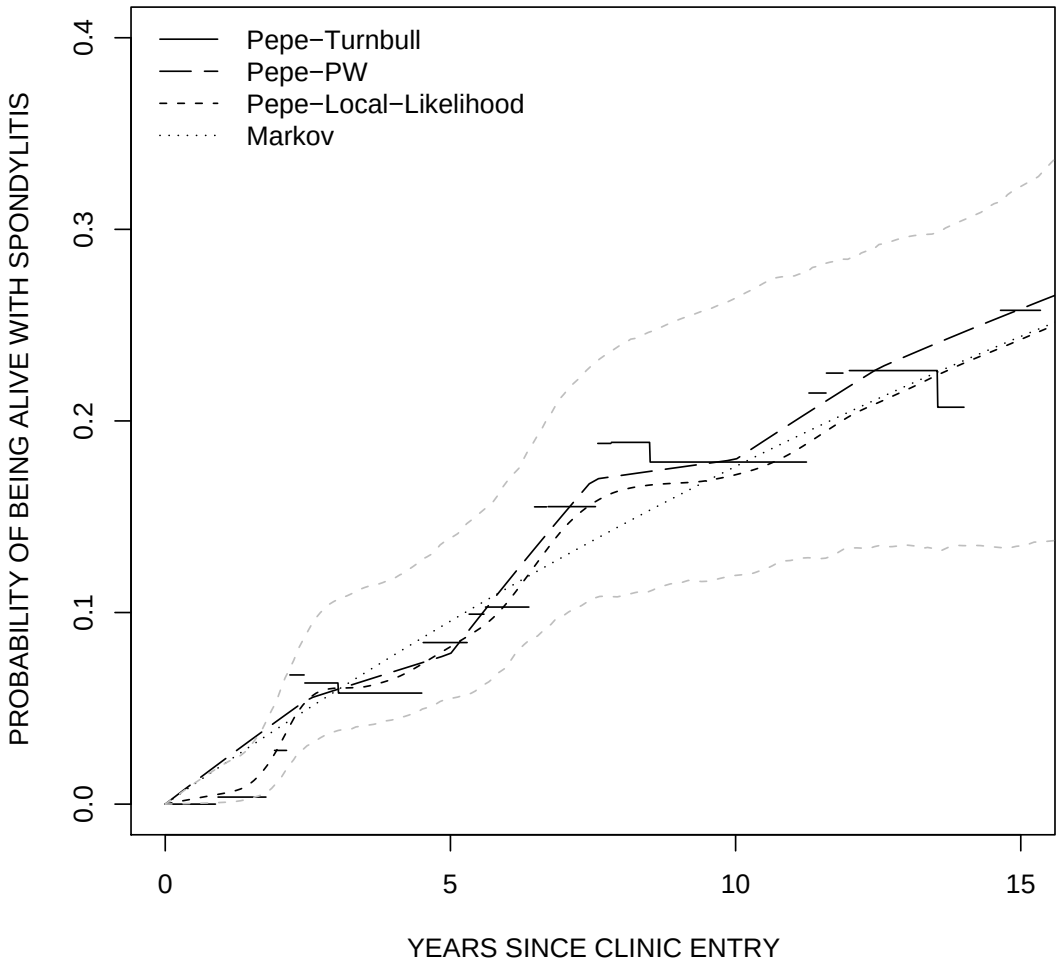

Figure 4.3: Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and time homogeneous Markov estimates of remaining alive and having spondylitis over 15 years from clinic entry, based on Definition 3; Pepe-Local-Likelihood 95% bootstrap confidence intervals are also plotted.

roughly five-fold increased risk of death (RR=4.77, 95% CI 1.36: 16.69). The tests of homogeneity both fail to suggest that the effects of these risk factors differ for subjects with and without spondylitis. Finally, the effect of spondylitis on the mortality rate was not significant with $\widehat{\gamma} = 1.126$ (RR=3.08, 95% CI 0.66:14.32, $p = 0.15$), so there is little evidence that spondylitis affects mortality.

Spondylitis Definition 5: Prevalence and Risk Factors

The same analyses were performed for each definition of spondylitis, but here we report the results for Definition 5. Based on this definition, 206 subjects had complete covariate information and 43 patients had spondylitis at clinic entry. Eleven patients died (2 with path $0 \to 1 \to 2$, 6 with path $0 \to 2$, and 3 with path $1 \to 2$).

Figure 4.4 shows the Pepe-Turnbull, Pepe-Local-Likelihood, Pepe-Piecewise, and Markov estimates of the prevalence function for state 1. Again, we conclude that after 10 years of followup we can expect up to 15% of PsA patients to develop spondylitis (LL - EST=0.146, bootstrap SE=0.041, 95% bootstrap CI 0.089:0.244; PW - EST=0.151, bootstrap SE=0.043, 95% bootstrap CI 0.058:0.222).

The regression model was built in the same manner as for the previous definition of spondylitis. Table 4.5 gives the univariate estimates while the final model estimates are given in Table 4.6. Again, the effects on mortality of race, periostitis, and enthesitis had large standard errors due to the fact that only caucasians were observed to die, and no one with periostitis or enthesitis was observed to die. Based, on definition 5, no non-caucasians developed spondylitis, causing the large standard error for the effect of race on spondylitis.

Nail involvement, periostitis, number of effused joints, number of radiologically damaged joints and ESR were associated with increased risk of spondylitis. Those with nail involvement have 7 times the risk of developing spondylitis (RR=7.07, 95% CI 1.56:32.04) and periostitis leads to almost 6 times the risk (RR=5.67, 95% CI 2.11:15.27), controlling for other factors. Each effused joint reduces risk by 17% (RR=0.83, 95% CI 0.73:0.95) while each radiologically damaged joint increases risk by 11% (RR=1.11, 95% CI 1.05:1.17). An increase of 1 mm/h in ESR leads to a 3% increase in risk (RR=1.03, 95% CI 1.01:1.05).

Factors affecting mortality include age and smoking status. For each additional year in age at presentation there was a 10% increase in risk of death (RR=1.10, 95% CI 1.03:1.16) while smokers had almost four times the risk of death (RR=3.73, 95% CI 1.07:12.92) compared with non-smokers. Again, the tests of homogeneity

Table 4.3: Univariate effects on transitions using spondylitis Definition 3.

		Effect on spondylitis			Effect on mortality			
Covariate		Est	SE	p	Est	SE	p	$p^\dagger$
Gender	Male vs Female	0.169	0.349	0.628	0.294	0.639	0.646	0.366
Race	Caucasian vs Other	6.195	39.150	0.874	2.655	15.276	0.862	0.933
Family history of psoriasis	yes vs. no	-0.275	0.361	0.446	-0.592	0.680	0.384	0.101
Family history of psoriatic arthritis	yes vs. no	0.011	0.597	0.986	0.185	1.059	0.861	0.268
Dactylitis	yes vs. no	0.031	0.352	0.931	0.023	0.635	0.971	0.933
Nail involvement	yes vs. no	0.573	0.486	0.238	0.157	0.785	0.841	0.176
Periostitis	yes vs. no	0.265	0.534	0.620	-7.563	43.528	0.862	1.000
Enthesitis	yes vs. no	0.465	0.426	0.276	-10.771	157.323	0.945	1.000
Spurs	yes vs. no	0.198	0.355	0.578	-0.772	0.783	0.324	0.107
B27	yes vs. no	0.546	0.406	0.178	0.093	0.783	0.906	0.134
Age	Years	0.008	0.014	0.559	0.079	0.027	0.004	0.669
Arthritis Duration	Years	0.012	0.024	0.618	-0.031	0.052	0.551	0.849
Psoriasis Duration	Years	-0.004	0.016	0.796	0.014	0.025	0.574	0.132
# Actively inflammed joints	-	0.004	0.018	0.808	0.038	0.026	0.145	0.652
# Effused joints	-	-0.044	0.050	0.376	0.021	0.073	0.775	0.005
# Clinically damaged joints	-	0.026	0.018	0.131	0.001	0.027	0.986	0.260
# Radiologically damaged joints	-	0.050	0.020	0.014	-0.007	0.032	0.832	0.627
ESR	mm/h	0.017	0.007	0.010	0.003	0.013	0.848	0.949
Prior NSAID use	yes vs. no	-0.308	0.351	0.380	-0.146	0.606	0.809	0.016
Prior DMARD use	yes vs. no	0.146	0.377	0.698	0.351	0.627	0.575	0.036
Smoker	yes vs. no	-0.180	0.425	0.671	1.004	0.606	0.098	0.685
Hypertension	yes vs. no	0.424	0.533	0.427	1.497	0.677	0.027	0.144

† p-value for test of common covariate effect on $0 \rightarrow 2$ and $1 \rightarrow 2$ transitions

Table 4.4: Multivariate effects on transitions using spondylitis Definition 3.

		Effect on spondylitis			Effect on mortality			
Covariate		Est	SE	p	Est	SE	p	$p^\dagger$
# Radiologically damaged joints	-	0.050	0.021	0.015	—	—	—	—
ESR	mm/h	0.017	0.007	0.010	—	—	—	—
Age	Years	—	—	—	0.096	0.030	0.002	0.696
Smoker	yes vs. no	—	—	—	1.562	0.639	0.015	0.811

† p-value for test of common covariate effect on $0 \to 2$ and $1 \to 2$ transitions

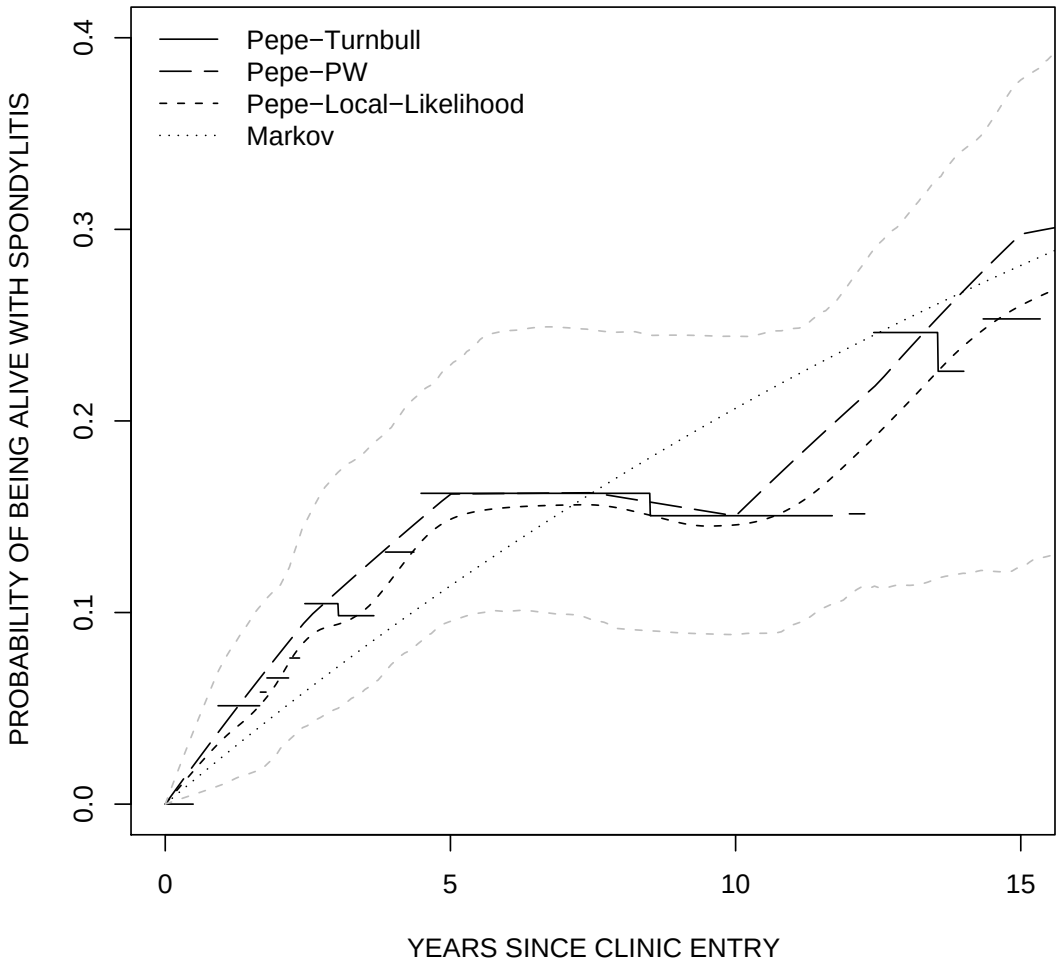

Figure 4.4: Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and time homogeneous Markov estimates of remaining alive and having spondylitis over 15 years from clinic entry, based on Definition 5; Pepe-Local-Likelihood 95% bootstrap confidence intervals are also plotted.

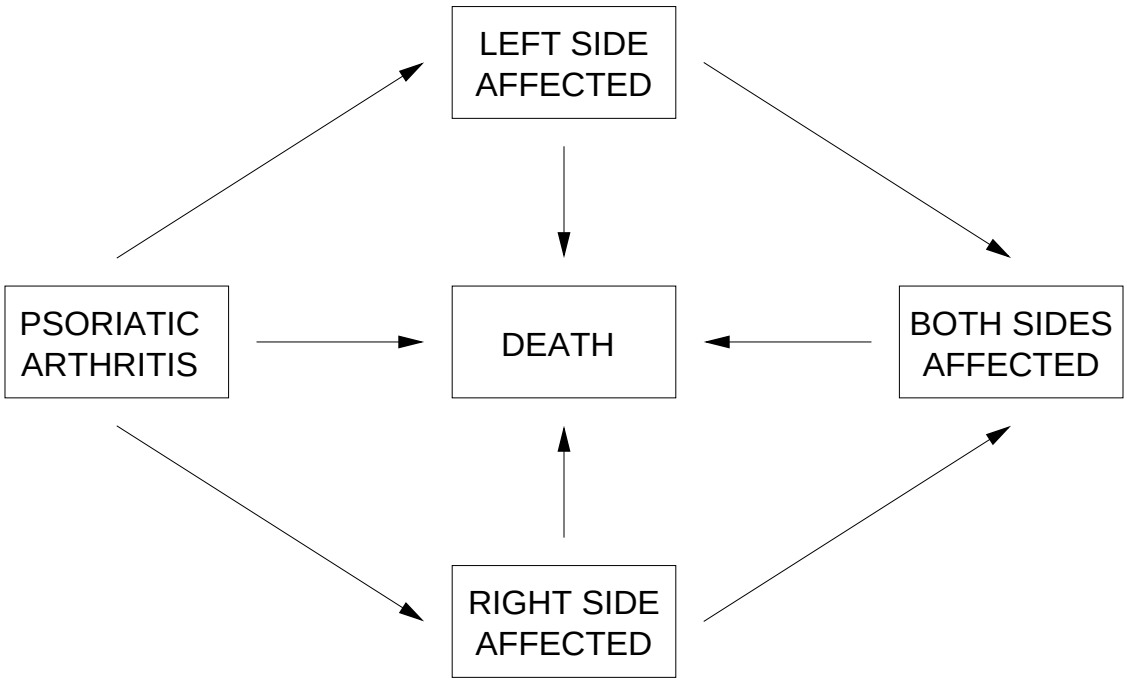

Figure 4.5: Five state diagram for onset of left and right side spondylitis in patients with psoriatic arthritis.

fail to suggest these effects are different for subjects with and without spondylitis. The effect of spondylitis on the mortality rate was not significant with $\widehat{\gamma} = 0.772$ (RR=2.16, 95% CI 0.57:8.24, $p = 0.26$). The mortality analysis results in the same findings regardless of the choice of definition of spondylitis.

4.3.2 Spondylitis Defined by Unilateral Involvement

If interest lies in the onset of spondylitis on a specific side of the body then the 5-state model of Figure 4.5 may be used. Here we define spondylitis as being present on a given side of the body if the corresponding sacroiliac joint is grade 3 or higher. A time-homogeneous Markov model may be fit where λ_{0L} and λ_{0R} are the intensities of left-back and right-back involvement respectively among individuals with no prior back involvement. Let λ_{LR} (λ_{RL}) denote the transition intensity for the onset of back involvement on the right (left) side given involvement on the left (right). Finally, λ_{0D}, λ_{LD}, λ_{RD}, and λ_{LRD} are the transition rates for death among individuals with no back involvement, left, right, or bilateral back involvement. The effect of already having left-back involvement on the intensity of right-back involvement is given by γ_L under the model $\lambda_{LR} = \lambda_{0R}e^{\gamma_L}$, and similarly γ_R is

Table 4.5: Univariate effects on transitions using spondylitis Definition 5.

		Effect on spondylitis			Effect on mortality			
Covariate		Est	SE	p	Est	SE	p	$p^\dagger$
Gender	Male vs. Female	0.453	0.385	0.239	0.248	0.636	0.697	0.148
Race	Caucasian vs. Other	5.374	48.006	0.911	4.985	61.084	0.935	1.000
Family history of psoriasis	yes vs. no	-0.091	0.379	0.811	-0.483	0.686	0.481	0.197
Family history of Psoriatic Arthritis	yes vs. no	0.739	0.480	0.123	0.174	1.055	0.869	0.352
Dactylitis	yes vs. no	-0.052	0.382	0.891	0.052	0.634	0.934	0.188
Nail involvement	yes vs. no	1.527	0.733	0.037	0.077	0.790	0.922	0.057
Periostitis	yes vs. no	1.183	0.466	0.011	-9.182	97.247	0.925	1.000
Enthesitis	yes vs. no	0.615	0.461	0.182	-9.780	94.460	0.918	1.000
Spurs	yes vs. no	0.528	0.383	0.169	-0.765	0.785	0.329	0.430
B27	yes vs. no	0.122	0.540	0.821	0.139	0.787	0.859	0.527
Age	Years	0.012	0.015	0.435	0.082	0.028	0.003	0.549
Arthritis Duration	Years	0.019	0.030	0.518	-0.031	0.054	0.561	0.721
Psoriasis Duration	Years	-0.005	0.019	0.793	0.016	0.026	0.546	0.483
# Actively inflammed joints	-	0.018	0.018	0.334	0.040	0.028	0.156	0.357
# Effused joints	-	-0.085	0.064	0.180	0.024	0.073	0.743	0.007
# Clinically damaged joints	-	0.018	0.024	0.450	-0.001	0.027	0.962	0.434
# Radiologically damaged joints	-	0.062	0.024	0.010	-0.015	0.034	0.667	0.625
ESR	mm/h	0.013	0.007	0.085	-0.001	0.013	0.951	0.520
Prior NSAID use	yes vs. no	0.318	0.384	0.407	-0.210	0.606	0.729	0.625
Prior DMARD use	yes vs. no	0.129	0.430	0.764	0.418	0.629	0.506	0.273
Smoker	yes vs. no	-0.090	0.436	0.837	0.934	0.606	0.123	0.105
Hypertension	yes vs. no	1.083	0.495	0.029	1.612	0.678	0.017	0.613

† p-value for test of common covariate effect on $0 \rightarrow 2$ and $1 \rightarrow 2$ transitions

Table 4.6: Multivariate effects on transitions using spondylitis Definition 5.

		Effect on spondylitis			Effect on mortality			
Covariate		Est	SE	p	Est	SE	p	$p^\dagger$
Nail involvement	yes vs. no	1.956	0.771	0.011	—	—	—	—
Periostitis	yes vs. no	1.735	0.505	< 0.001	—	—	—	—
# Effused joints	-	-0.182	0.069	0.008	—	—	—	—
# Radiologically damaged joints	-	0.105	0.029	< 0.001	—	—	—	—
ESR	mm/h	0.030	0.009	0.001	—	—	—	—
Age	Years	—	—	—	0.092	0.030	0.002	0.468
Smoker	yes vs. no	—	—	—	1.315	0.634	0.038	0.192

† p-value for test of common covariate effect on $0 \rightarrow 2$ and $1 \rightarrow 2$ transitions

the effect of already having right-back involvement on the intensity of left-back involvement through $\lambda_{RL} = \lambda_{0L}e^{\gamma_R}$. The baseline mortality rate is λ_{0D}, which is modulated by δ_1 when an individual has spondylitis on one side of the body (*i.e.* $\lambda_{LD} = \lambda_{RD} = \lambda_{0D}e^{\delta_1}$) and δ_2 when both sides are affected (*i.e.* $\lambda_{LRD} = \lambda_{0D}e^{\delta_2}$).

Robust estimates of the prevalence functions may be obtained as in Section 4.2. Here, the focus is on the prevalences of unilateral or bilateral spondylitis, so robust prevalence estimates may be based on the estimates of the distribution functions of the time spent in the initial state, S_0, the time to death, T_D, and the time of bilateral involvement or death, $\min(T_D, T_B)$. Defining state 0 to be the initial state, state 1 to be left side affected, state 2 to be right side affected, state 3 to be both sides affected, and state 4 to be death, we have,

$$\begin{aligned} P(Z(t) = 0) &= P(S_0 > t), \\ P(Z(t) = 1 \text{ or } Z(t) = 2) &= P(S_0 \leq t) - P(\min(T_D, T_B) \leq t), \\ P(Z(t) = 3) &= P(\min(T_D, T_B) \leq t) - P(T_D \leq t), \\ P(Z(t) = 4) &= P(T_D \leq t). \end{aligned}$$

Plots of the prevalences of unilateral (either left or right side affected) and bilateral (both sides affected) spondylitis are given in Figure 4.6. These suggest that after 10 years, roughly 5% of PsA patients develop unilateral spondylitis (LL - EST=0.053, bootstrap SE=0.026, 95% bootstrap CI 0.003:0.108; PW - EST=0.050, bootstrap SE=0.028, 95% bootstrap CI 0.000:0.108). As for bilateral spondylitis, the lower panel of Figure 4.6 suggests about 10% of PsA patients develop bilateral spondylitis after 10 years (LL - EST=0.097, bootstrap SE=0.036, 95% bootstrap CI 0.044:0.181; PW - EST=0.109, bootstrap SE=0.040, 95% bootstrap CI 0.015:0.168). It is interesting to compare this with the results of Section 4.3.1, where it was suggested that after 10 years, roughly 15% of individuals developed spondylitis, based on Definition 5. Hence, there appears to be roughly twice as many bilateral cases of spondylitis as there are univariate cases, after 10 years.

Regression models may also be fit using `msm` (Jackson 2007). Here, we take the effect of covariates to be the same on transitions leading to further development of spondylitis ($0 \to 1$, $0 \to 2$, $1 \to 3$ and $2 \to 3$), and the effect on the transitions leading to death ($0 \to 4$, $1 \to 4$, $2 \to 4$ and $3 \to 4$) to be the same. The model can be built as before, with the univariate estimates in Table 4.7. With this model, continuous covariates were difficult to fit, so the continuous variables were divided into categorical variables, using the median a guideline in choosing the cut-point.

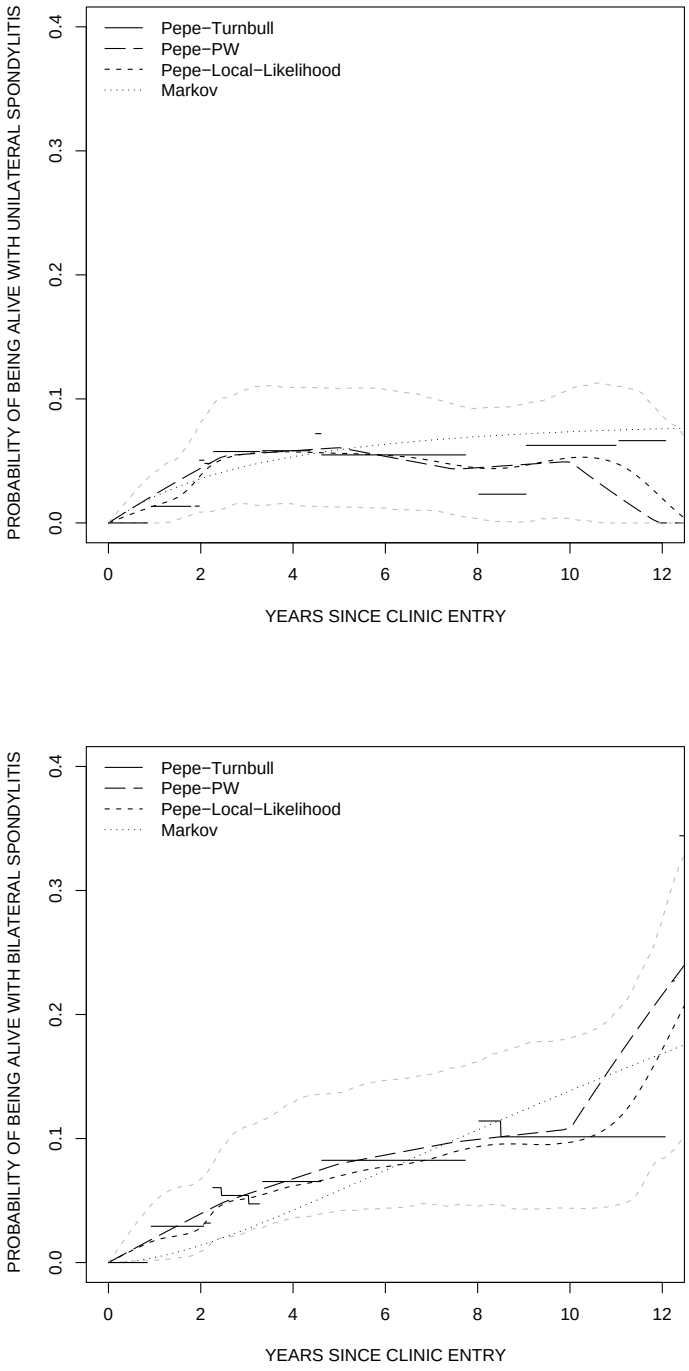

Figure 4.6: Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of state occupancy probabilities for being alive with spondylitis on one side of the body (top) and alive with spondylitis on both sides of the body (bottom); Pepe-Local-Likelihood 95% bootstrap confidence intervals are also plotted.

The final model is given in Table 4.8. We see that gender, family history of psoriasis, and age affect spondylitis development. Males have more than twice the risk (RR=2.36, 95% CI 1.27:4.37) of developing spondylitis compared with females, holding other factors constant. Having a family history of psoriasis reduces the risk of spondylitis by about 50% (RR=0.45, 95% CI 0.24:0.86) while those over 40 years old have about half the risk (RR=0.50, 95% CI 0.27:0.91) compared with those 40 years old or younger. Only age and smoking status affect mortality, with those over 40 years old having 13 times the risk of dying (RR=13.19, 95% CI 1.64:106.12). The large standard error may be attributed to the fact that only one individual died in the age group 40 years or less. Smoking leads to almost 4 times the risk of death (RR=3.77, 95% CI 1.11:12.74).

The results of fitting a Markov model are shown in Table 4.9. The estimates of γ_L and γ_R both suggest that having spondylitis on one side of the body increases the risk of spondylitis on the other side. The estimates of parameters relating to mortality are not significant, suggesting that the onset of spondylitis does not affect mortality. Looking at the final model estimates in Table 4.9, we see that developing spondylitis on the left side of the body leads to a very elevated risk of bilateral involvement. If spondylitis occurs first on the right side of the body, then there is no significant effect on the intensity of bilateral involvement. It also appears that neither unilateral nor bilateral spondylitis has a significant effect on mortality.

4.3.3 A 4-State Model for Bivariate Interval-Censored Data

Cook et al. (2008) consider the use of a four-state model to fit the joint distribution of bivariate interval-censored failure time data. They described an EM algorithm for fitting a piecewise constant multiplicative intensity Markov model, and obtained estimates of state occupancy probabilities based on piecewise constant models. Here we consider the same problem, but discuss Pepe estimation of the state occupancy probabilities.

Let T_j denote the time of the event of type j. For the model of Figure 4.7

Table 4.7: Univariate effects on transitions in the model of Figure 4.5.

		Effect on spondylitis			Effect on mortality		
Covariate		Est	SE	p	Est	SE	p
Gender	Male vs Female	0.726	0.298	0.015	0.337	0.640	0.598
Race	Caucasian vs Other	-0.516	1.159	0.656	6.267	104.434	0.952
Family history of psoriasis	yes vs. no	-0.757	0.319	0.018	-0.628	0.681	0.356
Family history of psoriatic arthritis	yes vs. no	-1.033	0.512	0.044	0.055	1.067	0.959
Dactylitis	yes vs. no	0.140	0.312	0.654	-0.019	0.636	0.976
Nail involvement	yes vs. no	-0.177	0.328	0.590	0.292	0.785	0.710
Periostitis	yes vs. no	0.540	0.488	0.269	-8.652	78.428	0.912
Enthesitis	yes vs. no	0.344	0.374	0.358	-8.408	47.314	0.859
Spurs	yes vs. no	-0.367	0.338	0.278	-0.841	0.790	0.287
B27	yes vs. no	0.147	0.342	0.667	0.083	0.789	0.916
Age $>$ 40 years	yes vs. no	-0.419	0.293	0.153	2.423	1.064	0.023
Arthritis Duration $>$ 3.5 years	yes vs. no	0.131	0.302	0.664	-0.094	0.611	0.878
Psoriasis Duration $>$ 10 years	yes vs. no	-0.973	0.324	0.003	0.853	0.680	0.210
# Actively inflammed joints $\geq$ 10	yes vs. no	-0.353	0.285	0.215	1.026	0.678	0.130
# Effused joints $\geq$ 3	yes vs. no	-0.425	0.298	0.154	-0.460	0.686	0.503
# Clinically damaged joints $\geq$ 1	yes vs. no	-0.192	0.291	0.509	0.117	0.616	0.850
# Radiologically damaged joints $\geq$ 3	yes vs. no	-0.023	0.312	0.943	-0.084	0.663	0.899
ESR $>$ 25 mm/h	yes vs. no	0.326	0.284	0.251	0.889	0.705	0.207
Prior NSAID use	yes vs. no	-0.397	0.287	0.167	-0.216	0.630	0.731
Prior DMARD use	yes vs. no	0.344	0.309	0.266	0.445	0.638	0.486
Smoker	yes vs. no	-0.209	0.364	0.566	1.038	0.616	0.092
Hypertension	yes vs. no	-8.359	37.016	0.821	1.990	0.780	0.011

Table 4.8: Multivariate effects on transitions in the model of Figure 4.5.

		Effect on spondylitis			Effect on mortality		
Covariate		Est	SE	p	Est	SE	p
Gender	Male vs Female	0.857	0.315	0.007	—	—	—
Family history of psoriasis	yes vs. no	-0.794	0.327	0.015	—	—	—
Age $>$ 40 years	yes vs. no	-0.703	0.312	0.024	2.579	1.064	0.015
Smoker	yes vs. no	—	—	—	1.326	0.622	0.033

Table 4.9: Estimates of model parameters from fitting the Markov model in Figure 4.5 to the psoriatic arthritis data.

	No Covariates			Final Model		
	EST	SE	p-value	EST	SE	p-value
λ_{0L}	0.018	0.004	—	0.021	0.007	—
λ_{0R}	0.008	0.003	—	0.009	0.004	—
γ_L	4.181	0.457	< 0.001	4.453	0.494	< 0.001
γ_R	1.084	0.650	0.095	0.558	0.633	0.379
λ_{0D}	0.005	0.002	—	0.001	0.001	—
δ_1	1.041	1.262	0.409	1.811	1.219	0.137
δ_2	0.571	0.851	0.503	0.107	0.863	0.901

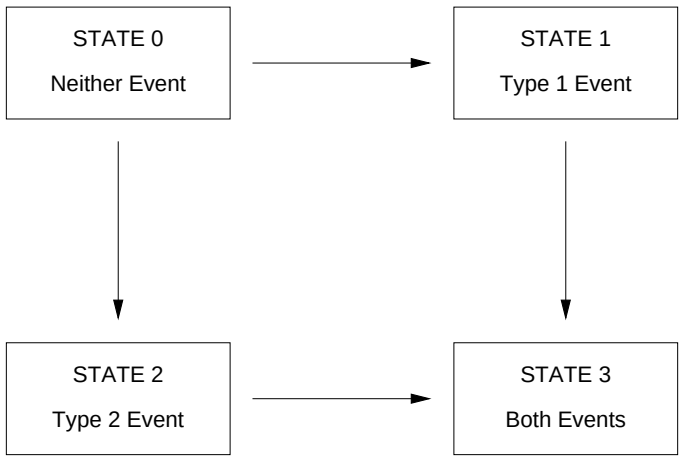

Figure 4.7: Four state diagram for bivariate interval-censored failure time data.

expressions for the state occupancy probabilities are

$$\begin{aligned} P(Z(t)=0|Z(0)=0) &= P(\min(T_1,T_2)>t) \\ P(Z(t)=1|Z(0)=0) &= P(T_2>t) - P(Z(t)=0|Z(0)=0) \\ P(Z(t)=2|Z(0)=0) &= P(T_1>t) - P(Z(t)=0|Z(0)=0) \\ P(Z(t)=3|Z(0)=0) &= P(T_2\leq t) - P(Z(t)=2|Z(0)=0) \\ &= P(T_1\leq t) - P(Z(t)=1|Z(0)=0). \end{aligned} \tag{4.1}$$

The intuition behind these equations is again quite simple. Consider equation (4.1). The event $\{T_2 > t\}$ indicates that the subject has not experienced the event of type 2, and hence is in either state 0 or 1. The occupancy probability for state 1 is obtained by subtracting the occupancy probability of state 0 from the probability of $\{T_2 > t\}$. All other expressions are obtained similarly.

The expressions given were constructed by beginning with $P(Z(t) = 0|Z(0) = 0)$ and working forward, but we could just as easily started with $P(Z(t) = 3|Z(0) = 0)$ and worked backwards. In this case, $P(Z(t) = 3|Z(0) = 0) = P(\max(T_1, T_2) \leq t)$ and similar expressions can be derived. It seems preferable to begin at state 0 since in practice, $\max(T_1, T_2)$ will be subject to a larger degree of right censoring than $\min(T_1, T_2)$.

The marginal distributions of T_1, T_2 and $\min(T_1, T_2)$ can be estimated in a number of ways. Possibilities include nonparametric (Turnbull 1976), piecewise-constant hazards (Lindsey and Ryan 1998) or local likelihood (Betensky et al. 1999).

State Occupancy Probabilities for CMV Shedding

We now consider the bivariate interval-censored data discussed in Betensky and Finkelstein (1999) and Goggins and Finkelstein (2000). HIV-infected individuals are susceptible to the opportunistic infection CMV. Once infected, the virus may be shed in the blood and urine. We consider data from 232 patients who were followed at a clinic and tested for the presence of the CMV virus. Urine tests were scheduled roughly every 4 weeks, while blood tests were administered every 12 weeks. Figure 4.8 shows a plot of the censoring intervals for a sample of patients in the study. Each patient has a corresponding rectangle within which the time point where shedding in the blood and urine may be. This data was analyzed by Cook et al. (2008) in the context of a multistate Markov model, and covariate effects were considered based on multiplicative models. Here, we focus on estimation of the state

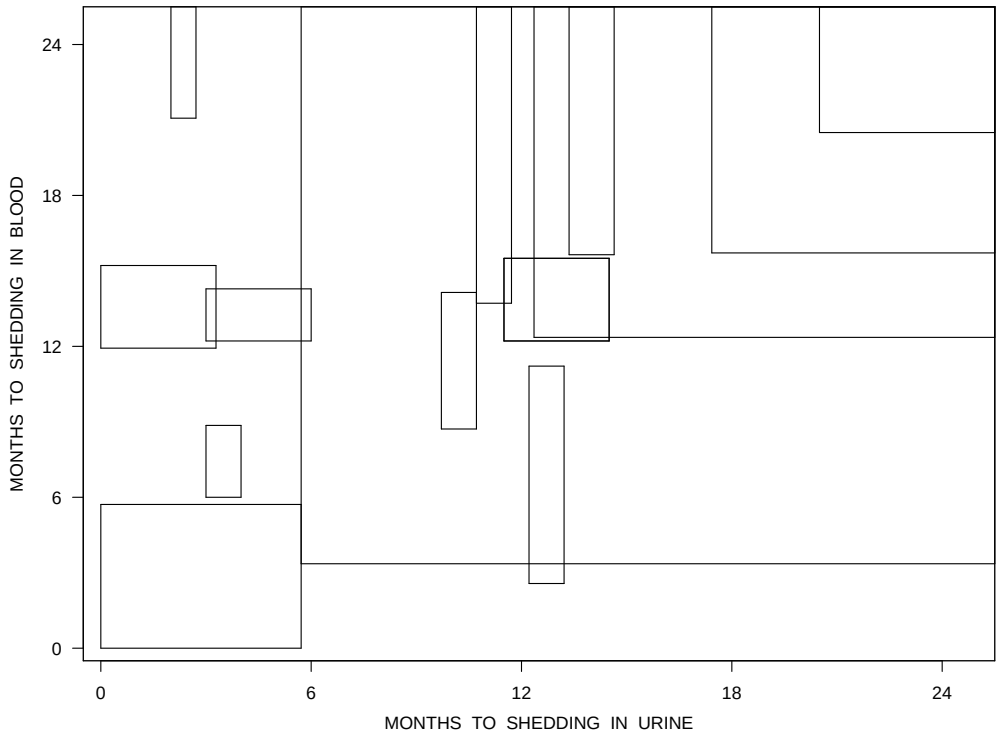

Figure 4.8: Censoring intervals for the times to shedding in blood and urine for a sample of patients in the CMV study.

occupancy probabilities using the Pepe methods. For the Pepe-Local-Likelihood method, a fixed bandwidth of 0.5 months was used, along with a locally constant approximation and Epanechnikov kernel. A 6-piece Pepe-Piecewise method was used with cut-points at 3, 6, 9, 12, 15 and 18 months. The resulting estimates were compared with the state occupancy probabilities obtained from the 3-piece multistate Markov model considered in Cook et al. (2008)

Figure 4.9 displays the state occupancy probability of having only blood shedding. About 2% of patients have only blood shedding 12 months after study entry (LL - EST=0.025, bootstrap SE=0.008, 95% bootstrap CI 0.011:0.042; PW - EST=0.034, bootstrap SE=0.008, 95% bootstrap CI 0.018:0.052). Figure 4.10 shows the state occupancy probability of having urine shedding only. Almost half the patients have only urine shedding 12 months after study entry (LL - EST=0.443, bootstrap SE=0.041, 95% bootstrap CI 0.367:0.529; PW - EST=0.516, bootstrap SE=0.043, 95% bootstrap CI 0.429:0.604). The state occupancy probability of hav-

ing both blood and urine shedding is shown in Figure 4.11. Roughly 10% of patients have both blood and urine shedding 12 months after study entry (LL - EST=0.108, bootstrap SE=0.026, 95% bootstrap CI 0.062:0.160; PW - EST=0.114, bootstrap SE=0.029, 95% bootstrap CI 0.064:0.174). It appears that the presence of CMV in the urine is fairly common, while having CMV in the blood alone is much less likely to be observed.

The estimates of the distribution functions for time to shedding in the blood or shedding in the urine may be plotted along with the estimates obtained from the bivariate nonparametric estimate described in Section 1.2.2. Figure 4.12 shows the bivariate nonparametric estimates along with univariate local likelihood and piecewise constant estimates. Also plotted is the 3-piece multistate Markov estimate from Cook et al. (2008). The bivariate estimate was obtained using the R package `Icens` (Gentleman and Vandal 2008). We see that for shedding in the blood, the estimates are all very similar. For shedding in the urine the estimates are more spread out, with the largest gap being between the two univariate estimates.

4.4 Summary

This chapter looked at several applications of multistate methods involving interval-censored observations. Estimates of the state occupancy probabilities were obtained from marginal distributions of the state entry times. This was done for 3, 4 and 5-state models relating to both spondylitis in psoriatic arthritis patients and CMV in HIV patients. The issue of bandwidth selection arises again when using the local likelihood based Pepe estimator. It would be of interest to develop a method of choosing a an overall bandwidth that can be used for each CDF required to obtain the estimated state occupancy probabilities, and compare this with choosing an individual bandwidth for each CDF. Regression models were also considered to identify factors which affect spondylitis development and mortality in psoriatic arthritis patients. The regression models were considered with time-homogeneous transition intensities. Piecewise constant intensities could be applied in this setting as well. It would also be of interest to see if using piecewise constant intensities alters the conclusions, although this would be at the expense of greater computational burden. These results are also discussed in Tolusso and Cook (2008b), Chandran et al. (2008a) and Chandran et al. (2008b).

The Pepe methods are robust in the sense that they do not require the Markov assumption. Regardless of the underlying multistate process, we can obtain esti-

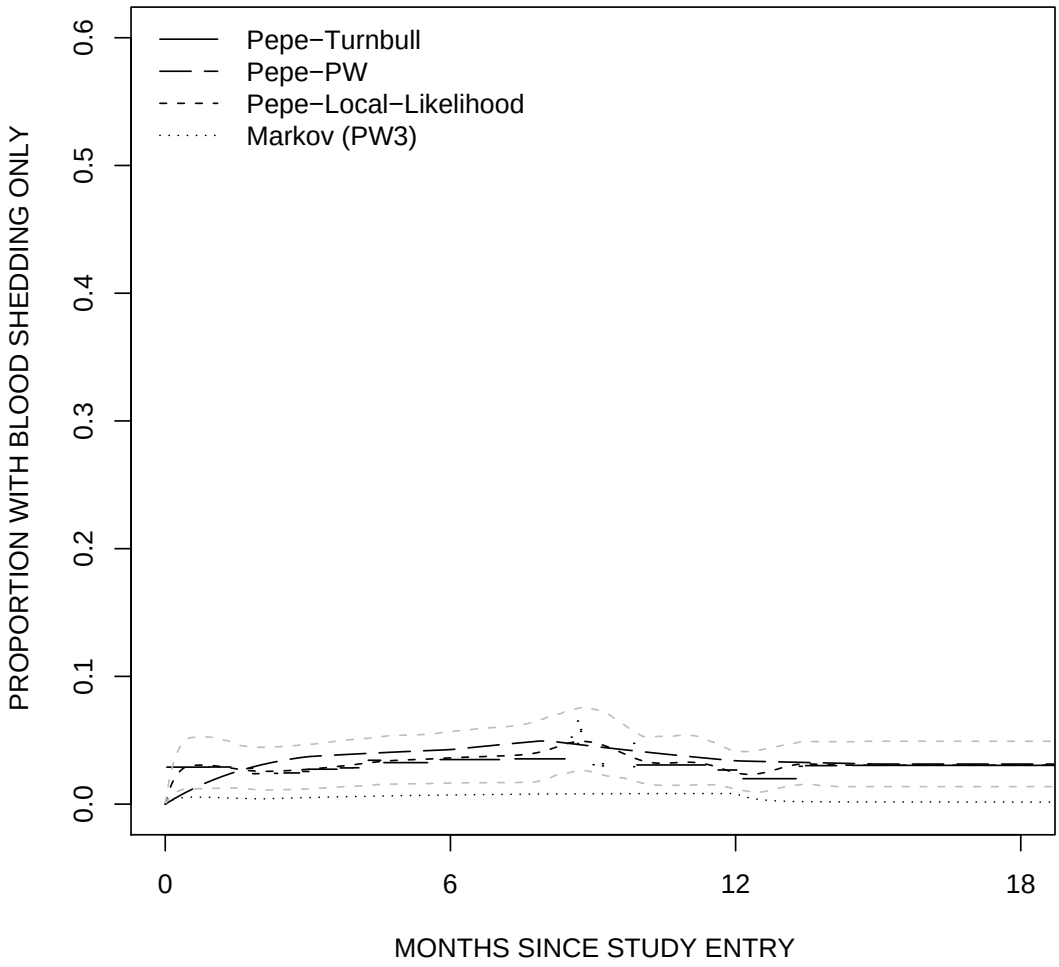

Figure 4.9: Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of having shedding in the blood only over 18 months from study entry in the CMV data; Pepe-Local-Likelihood 95% bootstrap confidence intervals are also plotted.

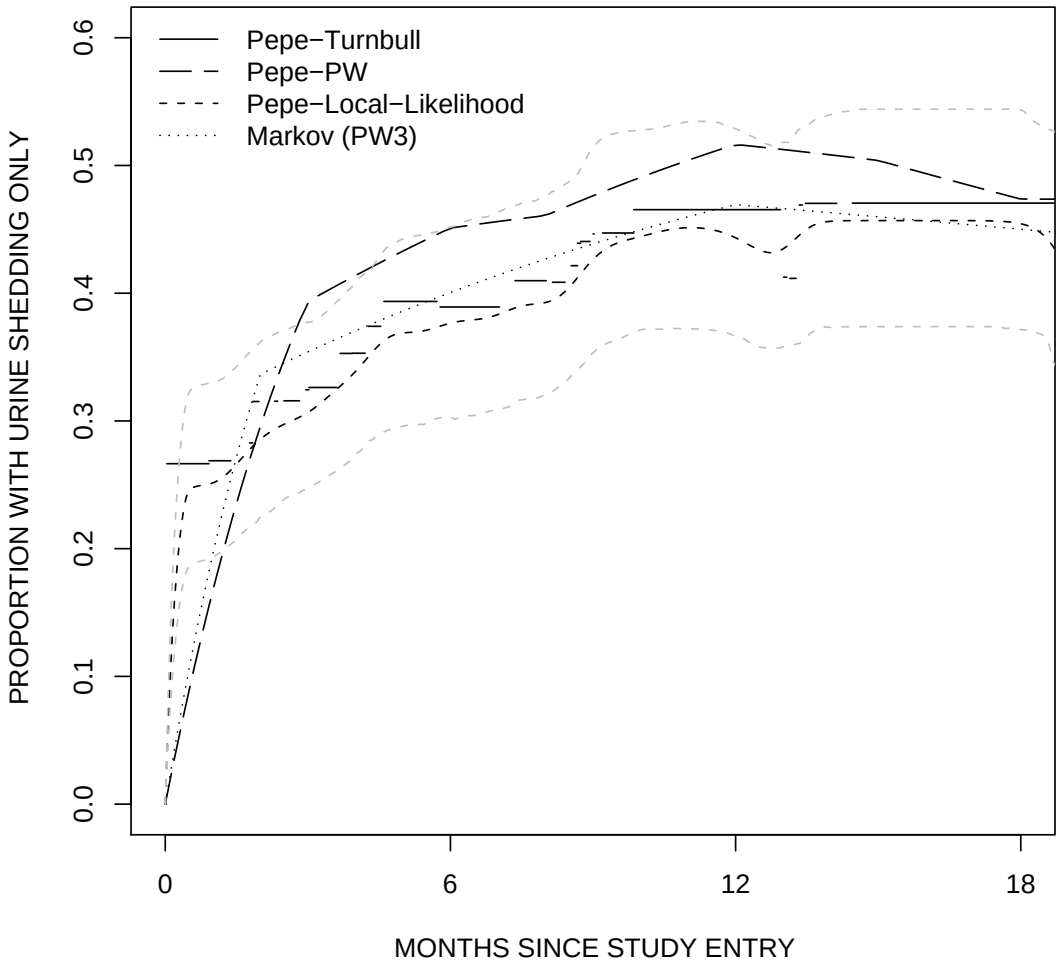

Figure 4.10: Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of having shedding in the urine only over 18 months from study entry in the CMV data; Pepe-Local-Likelihood 95% bootstrap confidence intervals are also plotted.

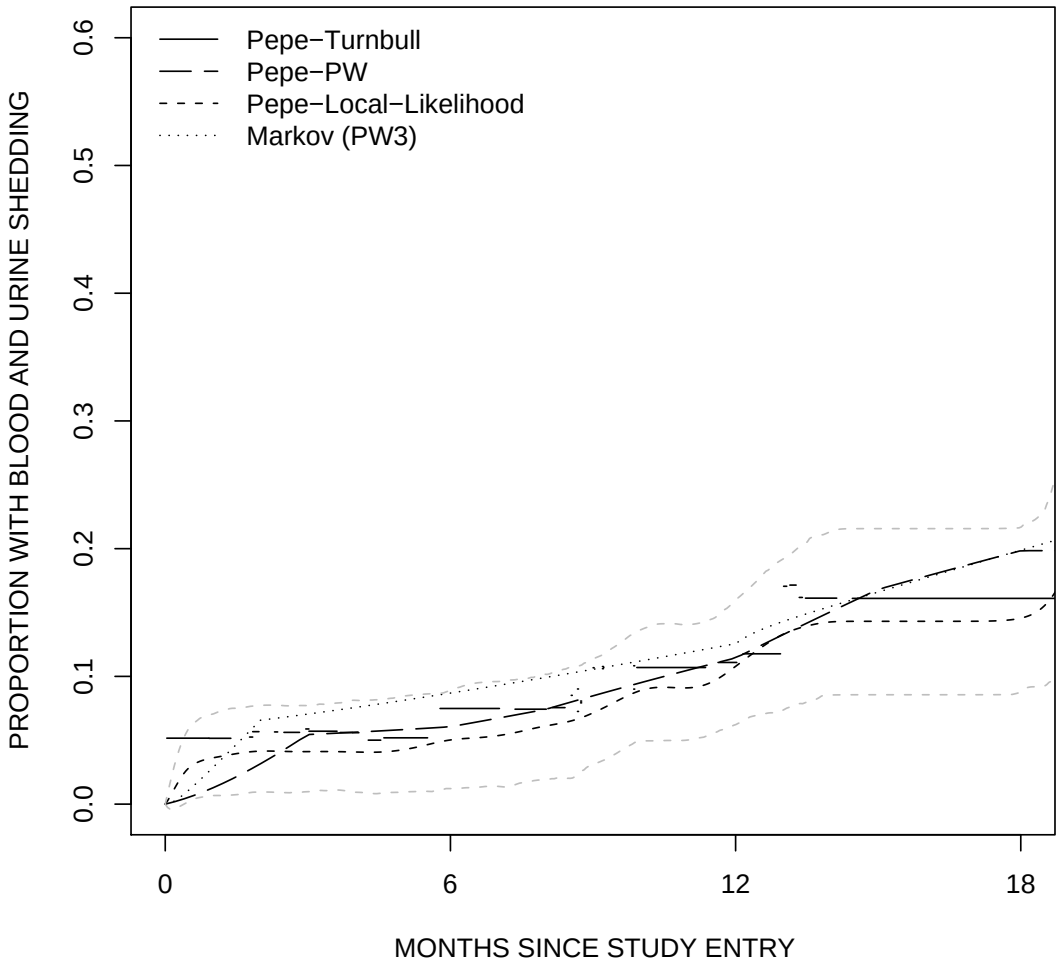

Figure 4.11: Pepe-Turnbull, Pepe-PW, Pepe-Local-Likelihood, and Markov estimates of having both blood and urine shedding over 18 months from study entry in the CMV data; Pepe-Local-Likelihood 95% bootstrap confidence intervals are also plotted.

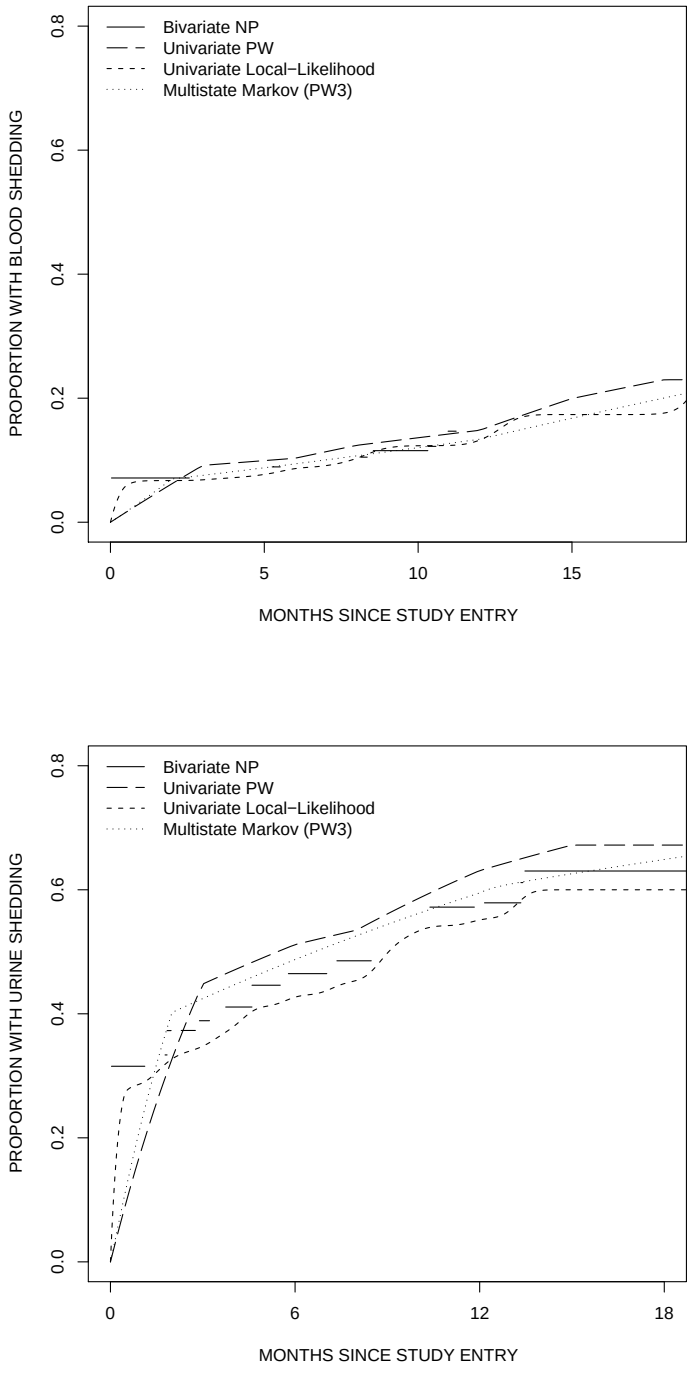

Figure 4.12: Bivariate nonparametric, univariate piecewise constant, univariate local likelihood and multistate Markov estimates of the distributions of time to shedding in the blood and shedding in the urine.

mates of the state occupancy probabilities using the Pepe approach as long as we use consistent methods to estimate the cumulative distribution functions.

Chapter 5

Future Work

5.1 Summary

Chapter 2 considered recurrent event data subject to interval censoring. Drawbacks of existing methods were the undefined regions and lack of smoothness of nonparametric estimates of the mean function (Sun and Kalbfleisch 1995; Wellner and Zhang 2000). Other techniques required selection of pieces and cut-points (Lawless and Zhan 1998) or the imposition of monotonicity and smoothing on a previously smoothed estimate (Staniswalis et al. 1997). Smooth estimates of the mean and rate function were developed based on local likelihood methods (Loader 1999; Betensky et al. 1999). Regression models were also considered using a profile likelihood (Severini and Wong 1992; Staniswalis et al. 1997). Simulation studies showed the local likelihood estimate performed as well as the piecewise-constant methods in terms of bias and mean squared error. These findings have been summarized in Tolusso and Cook (2008a).

Current status data was considered in Chapter 3 where it was assumed the failure times were possibly dependent. Current status data can be viewed as binary data, however existing methods for dependent binary data (Prentice 1988; Zhao and Prentice 1990; Liang et al. 1992) do not model the association in a manner appropriate for failure time data. A copula approach was taken to model the dependence structure with parameter estimates obtained by GEE methods (Liang and Zeger 1986). Both first and second order GEE's were considered. The bias and asymptotic relative efficiencies of the methods were evaluated as well as the performance of piecewise constant baseline hazards. Tolusso and Cook (2008c) discusses these results.

Chapter 4 detailed the application of multistate methods to a number of problems. In particular, robust estimates of state occupancy probabilities were obtained from marginal distributions using the approach of Pepe et al. (1991). The marginal distributions were estimated nonparametrically (Turnbull 1976), using piecewise constant hazards (Lindsey and Ryan 1998) and local likelihood (Betensky et al. 1999). Much of the focus was on the development of spondylitis in patients with psoriatic arthritis. Three-state models were considered as well as a five-state model which distinguished between unilateral and bilateral spondylitis. Regression models were considered to identify which factors affect development of spondylitis as well as mortality. These findings were summarized in Tolusso and Cook (2008b) as well as Chandran et al. (2008a) and Chandran et al. (2008b).

There still remains open problems related to this work. The remainder of this chapter will be devoted to topics to be explored in future research.

5.2 Methods for Recurrent Events and Death

In many women with advanced breast cancer, bone metastases often occur. The bone destruction that occurs in these lesions leads to increased pain, immobility, and deterioration in quality of life. Hortobagyi et al. (1996, 1998) describes a study of women with breast cancer that has metastasized to bone. The study examines the effect of a treatment known as pamidronate disodium. There were 380 patients in the study, of which 185 were randomly assigned the treatment, while the remaining patients received a placebo. Patients were followed for 24 months, or until they died or were lost to follow-up. Figure 5.1 shows plots of the Kaplan-Meier estimates of the survival functions for both the treatment and placebo groups. Figure 5.2 shows Kaplan-Meier estimates of the survival functions for the time until the first lesion or death, whichever occurs first. As this plot is being used to get initial impressions of the data, the interval censoring was not taken into account, and the time of the first lesion was taken to be the assessment time at which it was discovered. Figure 5.3 shows sample data patterns for some of the individuals in the study. Visits are indicated by vertical bars and the numbers indicated how many lesions were detected between visits. Solid lines represent when the individual was being examined for bone lesions, while the dashed line represents the time after their last assessment, but before their time of death. The time after the last assessment but before death must be treated carefully, since it is unknown how many lesions occur in this time period.

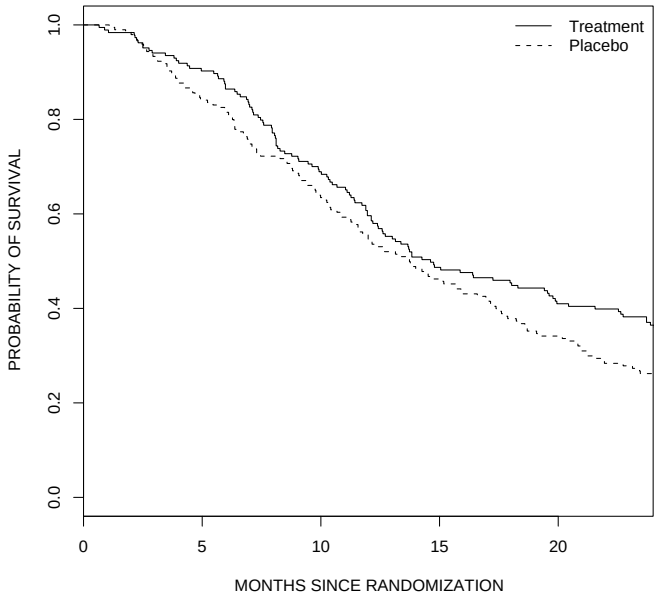

Figure 5.1: Kaplan-Meier estimates of survival for data from the study of breast cancer patients with bone metastases (Hortobagyi et al. 1996).

One aim is to estimate the marginal mean function, $E[N(t)]$.

5.2.1 Methods Based on Markov Models

Here we consider the problem of estimation and inference about rate functions for recurrent events in the presence of a terminal event. Figure 5.4 contains a general multi-state diagram for a Markov process which can characterize the occurrence of events in the presence of risk for death. The prevalence function represents the probability being in a given state at time t, so for recurrent events the prevalence function for state k gives the probability of having experienced exactly k events by time t. This leads to a natural estimate of the mean function given by

$$\widehat{\mu}(t) = \sum_{j=1}^{J} j\widehat{p}_j(t) \tag{5.1}$$

where $\widehat{p}_j(t)$ is the prevalence function for state j, $j = 1, \ldots, J$, and J is chosen to be large enough to capture all events in a particular dataset.

Special models one could consider include $\rho_k(t) = \rho(t)$, in which case there is a common rate of event occurrence, and $\mu_k(t) = \mu(t)\exp(\beta N(t^-))$, in which case there

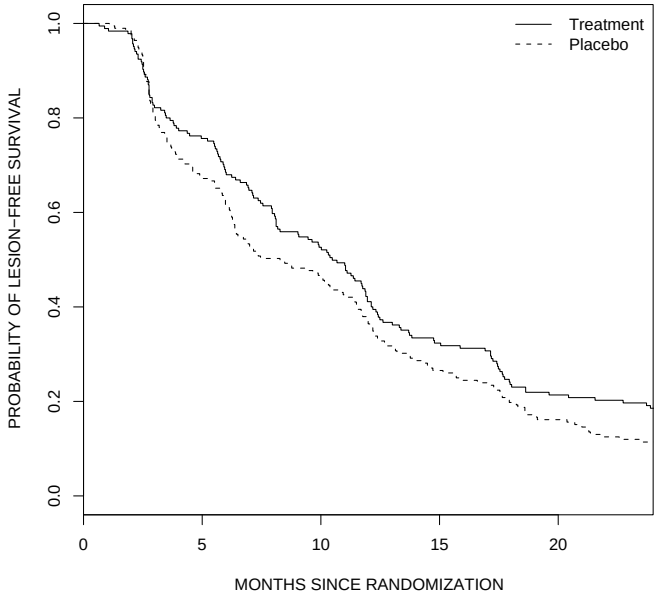

Figure 5.2: Kaplan-Meier estimates of time to first detected lesion or death for data from the study of breast cancer patients with bone metastases (Hortobagyi et al. 1996).

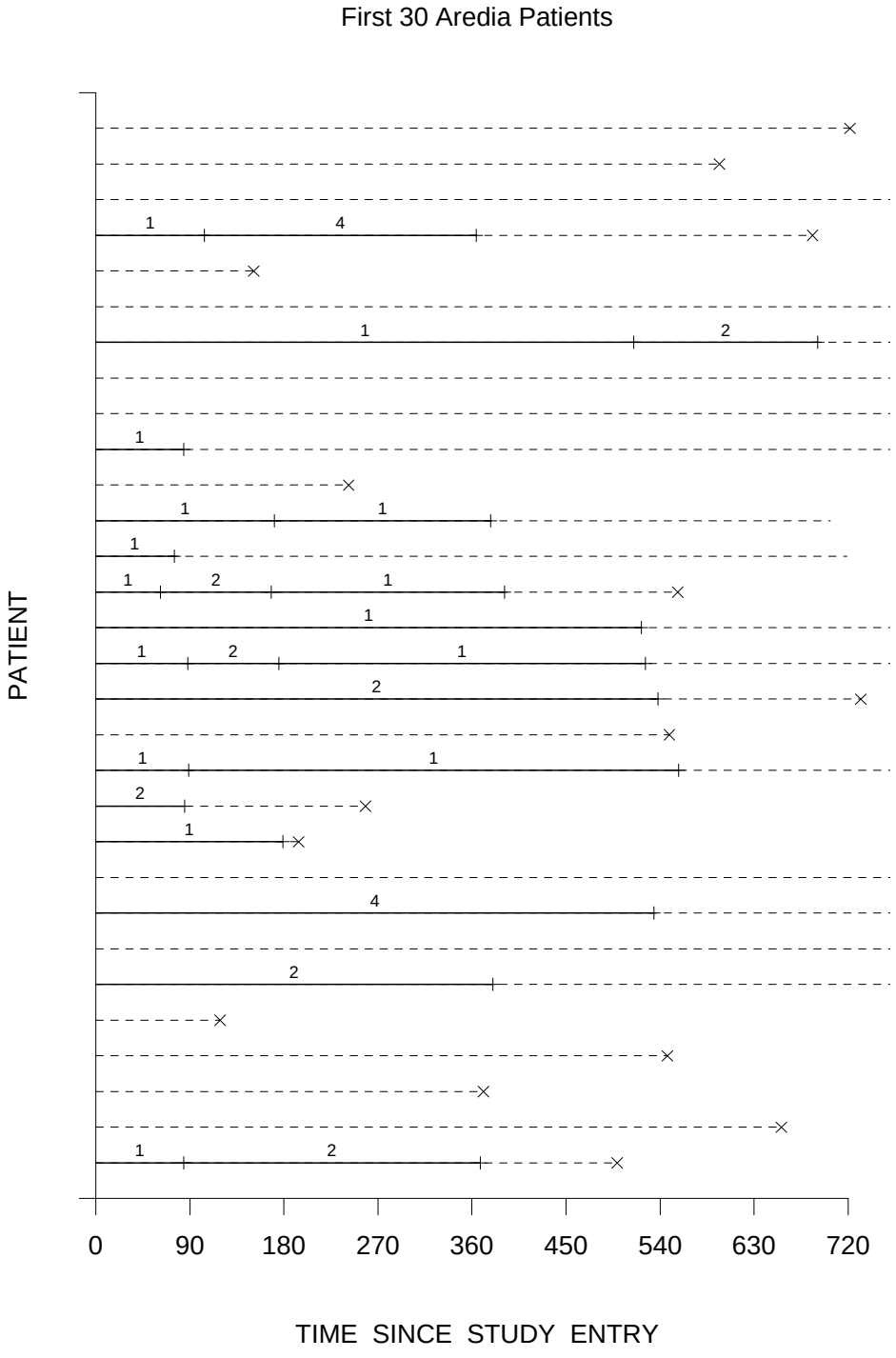

Figure 5.3: Sample of patient timelines from the study of Hortobagyi et al. (1996); numbers indicate number of lesions detected between visits indicated by vertical bars.

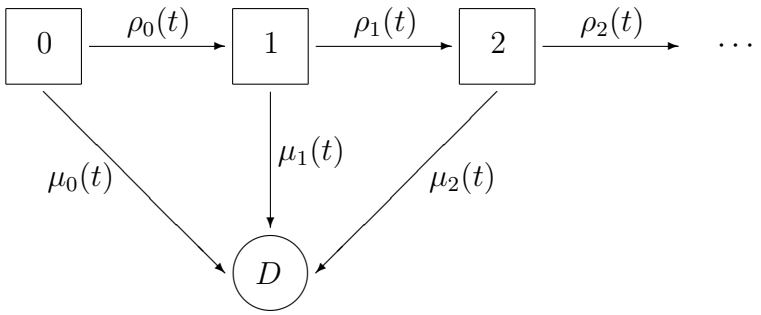

Figure 5.4: Multistate model for recurrent events with a terminating event.

is a common baseline mortality rate modulated by the cumulative number of events. A local likelihood EM algorithm will be used to estimate these parameters and the associated mean functions in both the one sample problem and in multiplicative models. This will involve adapting the methods of Section 2.3 to deal with the terminating event.

A complication that arises is that individuals may die or be lost to followup between observation times. The number of events that occur between the last assessment time, b_{im}, and the time of death/loss of followup, τ_i, is most likely unknown, however it is known the individual was alive between those times. Possible ways of dealing with this include disregarding the information that the individual survived from b_{im} to τ_i, or preferably, treating the number of events individual i experienced during that time as unknown, and estimating it in the E-step of the EM algorithm. The latter is preferred since the former may violate the assumptions on the inspection process discussed in Section 1.3.3.

5.2.2 Pepe Estimation

In addition, for the one sample case, estimates of the mean function can also be based on the Pepe approach. In this case, prevalence function estimates are based on a difference in cumulative incidence functions (CIF's),

$$P(N(t) = k) = P(Z(t) = E_k \text{ or } D_k) = CIF_{E_k}(t) - CIF_{E_{k+1}}(t) \qquad (5.2)$$

where E_k represents alive with k events and D_k represents died with k events, as in Figure 5.5.

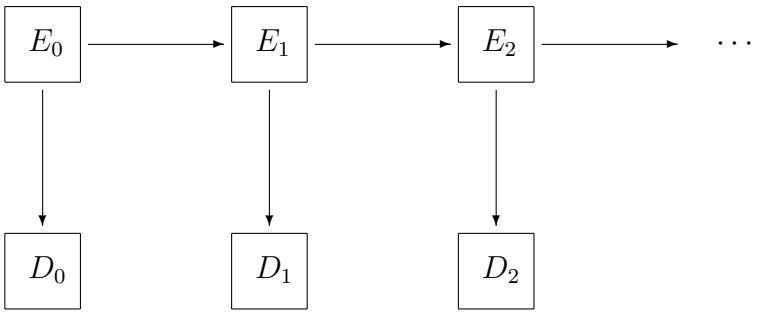

Figure 5.5: Alternative model for recurrent events with a terminating event appropriate for Pepe estimation of the mean function.

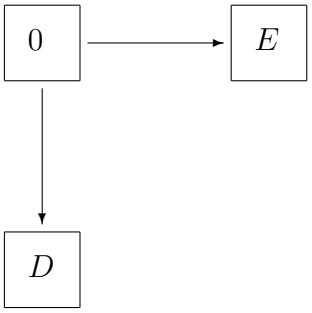

Figure 5.6: Competing risks model.

With the Pepe approach, the problem has essentially been reduced to obtaining an estimate of the CIF's in the competing risks problem. Consider Figure 5.6 which illustrates a competing risks problem where individuals initially begin in state 0 and can move either to state E (representing the event of interest) or state D (representing death). The CIF for transitions to state E is of interest, where transitions to state E may be interval censored and transitions to state D may be right censored.

A local likelihood EM algorithm will be used to estimate the CIF. Let the intensity of transitions to states E and D be denoted $\lambda_E(t)$ and $\lambda_D(t)$ respectively. For individual i, let

$$\delta_{Ei} = I(\text{the observed transition is to state } E)$$
$$\delta_{Di} = I(\text{the observed transition is to state } D)$$

and let T_i be the time of the transition (if either $\delta_{Ei} = 1$ or $\delta_{Di} = 1$) or the right censoring time (if $\delta_{Ei} = \delta_{Di} = 0$). The complete data log-likelihood is

$$\ell_C = \sum_{i=1}^{n} \left[\delta_{Ei} \log \lambda_E(T_i) + \delta_{Di} \log \lambda_D(T_i) - \int_0^{T_i} \lambda_E(u)\, du - \int_0^{T_i} \lambda_D(u)\, du \right]$$

with corresponding local likelihood

$$\begin{aligned} \ell_C(\lambda_E, \lambda_D, t) = \sum_{i=1}^{n} \Bigg[& \delta_{Ei} K_b(T_i - t) \log \lambda_E(T_i) + \delta_{Di} K_b(T_i - t) \log \lambda_D(T_i) \\ & - \int_0^{T_i} K_b(u - t) \lambda_E(u)\, du - \int_0^{T_i} K_b(u - t) \lambda_D(u)\, du \Bigg]. \end{aligned}$$

The log of the intensities can be approximated by polynomials as in previous local likelihood methods. If a transition to state E is interval censored, the expectation of terms involving T_i must be taken with respect to the density of T_i given that $T_i \in (L_i, R_i]$ and the transition was to state E. This density can be obtained from the density of T_i given the transition was to state E, given by (see Kalbfleisch and Prentice 2002, Chap. 8)

$$f_E(t) = \frac{\lambda_E(t) \exp\{-\Lambda_E(t) - \Lambda_D(t)\}}{P(E)}$$

where

$$\begin{aligned} \Lambda_E(t) &= \int_0^t \lambda_E(u)\, du \\ \Lambda_D(t) &= \int_0^t \lambda_D(u)\, du \\ P(E) &= \int_0^\infty \lambda_E(u) \exp\{-\Lambda_E(u) - \Lambda_D(u)\}\, du. \end{aligned}$$

The EM algorithm then proceeds as other local EM algorithms (see Sections 1.2.2 and 2.3). The resulting estimates of the intensities are denoted $\widehat{\lambda}_E(t)$ and $\widehat{\lambda}_D(t)$, and the local likelihood estimate of the CIF for transitions to E is

$$\widehat{CIF}_E(t) = \int_0^t \widehat{\lambda}_E(t) \exp(-\widehat{\Lambda}_E(t) - \widehat{\Lambda}_D(t)).$$

The mean function may then be estimated by applying (5.2) and (5.1).

5.3 Clustered Current Status Data

5.3.1 Second Order GEE for Regression on Association Parameter

As mentioned in Chapter 3, second order estimating functions have received a considerable amount of attention in the analysis of clustered data. It represents an appealing way to increase efficiency in estimates of regression parameters in the marginal means, and an approach to understanding how covariates might influence the nature of the dependence within clusters. Prentice (1988) and Zhao and Prentice (1990) discuss this in the context of correlated binary data.

Interestingly, there has been little work in extending this regression approach to the analysis of clustered failure time data. The concluding remarks of Wang (2003) comment on this approach, and mention how statistical inference under such a model is challenging in the case of right-censored failure time data. However, the current status setting has close connections to the binary analyses of Prentice (1988) and Zhao and Prentice (1990), so fitting such a regression model on the association is feasible.

Kendall's τ is a parameter that has an appealing interpretation regardless of the chosen copula, which arises in the joint distribution specified by the copula function. It would be of interest to formulate models of the form

$$\log\left(\frac{1+\tau}{1-\tau}\right) = z'\psi$$

and fit these using GEE2 to learn about factors that influence the strength of association within clusters.

Such methods could also be used in the context of multivariate current status data. This problem would be somewhat different than the clustered data problem, in that the dimension J would be fixed and typically quite small, and the marginal distributions would differ depending on the endpoint of particular interest. One could, however, still model association parameters to learn which covariates most heavily influenced the dependence between processes.

5.3.2 Copula Goodness-of-Fit

One issue regarding the second order GEE approach is that the dependence structure must be correct in order to obtain consistent estimates of not only the associ-

ation parameter, but the marginal parameters as well. It would then be of interest to develop methods for assessing the goodness-of-fit of a chosen copula.

Genest and Rivest (1993) consider the problem of selecting a copula from the class of Archimedean copulas for bivariate data. Their approach is based on the function $K(v) = P(H(U_1, U_2) \leq v)$ where U_1 and U_2 have uniform marginal distributions and H is a bivariate copula. They construct an estimate of K which can be used to aid in the selection of a copula.

Wang and Wells (2000) extend this approach by providing an estimate of K which may be used with right-censored data. They also provide a goodness-of-fit statistic to assist with the choice of copula. Genest et al. (2006) further extend this work by defining Cramér-von Mises and Kolmogorov-Smirnov goodness-of-fit statistics, however only for complete data. They also extend the approach to copulas of arbitrary dimension.

It would be of interest to further extend this work to the current status setting. It may be possible to obtain an estimate of K using methods for current status data, although dealing with the potential undefined regions may be a challenge. Such goodness-of-fit statistics would be useful to determine if a copula adequately models the association structure and provides little bias in the marginal parameters.

5.3.3 Random Effects Approach

An alternative to the marginal approach of Chapter 3 is to assume that subjects within a cluster share a common random effect. The random effect serves to capture homogeneity within a cluster and induce heterogeneity between clusters. With failure time data, we can use the model

$$\mathcal{F}(s|u_i, x_{ij}; \theta) = \exp(-u_i \Lambda_0(s; \alpha) e^{x'_{ij}\beta})$$

where the random effect u_i is distributed over the positive real line.

Direct Maximization

Consider the situation where we have a treatment indicator x_{ij}. The likelihood can be constructed in a similar fashion as in Cook (1999). Let $P(Y_{ij}|C_{ij}, x_{ij}, u_i) = \left\{1 - \exp(-u_i \Lambda(C_{ij}) e^{x_{ij}\beta})\right\}^{Y_{ij}} \left\{\exp(-u_i \Lambda(C_{ij}) e^{x_{ij}\beta})\right\}^{1-Y_{ij}}$, where u_i is a gamma random variable with mean 1 and variance ϕ. Let $C_{i(k)}$ denote the kth unique inspec-

tion time for cluster i. Define

$$Y^1_{i(k)} = \sum_{j:C_{ij}=C_{i(k)}} Y_{ij}x_{ij}, \qquad Y^0_{i(k)} = \sum_{j:C_{ij}=C_{i(k)}} Y_{ij}(1-x_{ij}),$$
$$r^1_{i(k)} = \sum_{j:C_{ij}=C_{i(k)}} x_{ij}, \qquad r^0_{i(k)} = \sum_{j:C_{ij}=C_{i(k)}} (1-x_{ij}).$$

Then, the conditional joint probability distribution of $Y^0_{i(k)}$ and $Y^1_{i(k)}$ given $C_{i(k)}$, $\{x_{ij} : C_{ij} = C_{i(k)}\}$, $r^1_{i(k)}$, $r^0_{i(k)}$ and u_i is given by

$$\left(1-e^{-u_i\Lambda(C_{i(k)})}\right)^{Y^0_{i(k)}}\left(e^{-u_i\Lambda(C_{i(k)})}\right)^{r^0_{i(k)}-Y^0_{i(k)}}\left(1-e^{-u_i\Lambda(C_{i(k)})e^{\beta}}\right)^{Y^1_{i(k)}}\left(e^{-u_i\Lambda(C_{i(k)})e^{\beta}}\right)^{r^1_{i(k)}-Y^1_{i(k)}}.$$

Binomial expansion of the first and third terms leads to

$$\left\{\sum_{l_0=0}^{Y^0_{i(k)}}\binom{Y^0_{i(k)}}{l_0}(-1)^{l_0}\left(e^{-u_i\Lambda(C_{i(k)})}\right)^{r^0_{i(k)}-Y^0_{i(k)}+l_0}\right\}$$
$$\times\left\{\sum_{l_1=0}^{Y^1_{i(k)}}\binom{Y^1_{i(k)}}{l_1}(-1)^{l_1}\left(e^{-u_i\Lambda(C_{i(k)})e^{\beta}}\right)^{r^1_{i(k)}-Y^1_{i(k)}+l_1}\right\}$$

which can be rewritten as

$$\sum_{l_0=0}^{Y^0_{i(k)}}\sum_{l_1=0}^{Y^1_{i(k)}}\binom{Y^0_{i(k)}}{l_0}\binom{Y^1_{i(k)}}{l_1}(-1)^{l_0+l_1}\left(e^{-u_i\Lambda(C_{i(k)})}\right)^{r^0_{i(k)}-Y^0_{i(k)}+l_0}\left(e^{-u_i\Lambda(C_{i(k)})e^{\beta}}\right)^{r^1_{i(k)}-Y^1_{i(k)}+l_1}$$

If we define

$$a_{ik}(l_k) = \binom{Y^0_{i(k)}}{l_0}\binom{Y^1_{i(k)}}{l_1}(-1)^{l_0+l_1}$$
$$b_{ik}(l_k) = r^0_{i(k)} - Y^0_{i(k)} + l_{k0}$$
$$d_{ik}(l_k) = r^1_{i(k)} - Y^1_{i(k)} + l_{k1}$$

then the likelihood contribution from cluster i,

$$\int_0^\infty \prod_{k=1}^{K_i} P(Y^0_{i(k)}, Y^1_{i(k)}|C_{i(k)}, \{x_{ij} : C_{ij} = C_{i(k)}\}, r^1_{i(k)}, r^0_{i(k)}, u_i)g(u_i)\,du_i$$

becomes

$$L_i(\lambda,\beta,\phi) = \sum_{l_1\in D_{i1}}\cdots\sum_{l_{k_i}\in D_{ik_i}} \frac{\prod_{i=1}^{k_i} a_{ik}(l_k)}{\left[1+\phi\left\{\sum_{k=1}^{K_i}\Lambda(C_{ik})(b_{ik}(l_k)+e^{\beta}d_{ik}(l_k))\right\}\right]^{\phi^{-1}}}$$

where $D_{ik} = \left\{ l_k : 0 \leq l_{k0} \leq Y^0_{i(k)}, 0 \leq l_{k1} \leq Y^1_{i(k)} \right\}$. This follows from the fact that

$$\prod_{k=1}^{K_i} P(Y^0_{i(k)}, Y^1_{i(k)} | C_{i(k)}, \{x_{ij} : C_{ij} = C_{i(k)}\}, r^1_{i(k)}, r^0_{i(k)}, u_i)$$

can be rewritten as

$$\sum_{l_1 \in D_{i1}} \cdots \sum_{l_{k_i} \in D_{ik_i}} \left[\prod_{i=1}^{k_i} a_{ik}(l_k) \exp\left\{ -(b_{ik}(l_k) u_i \Lambda(C_{i(k)}) + d_{ik}(l_k) u_i \Lambda(C_{i(k)}) e^{\beta}) \right\} \right].$$

EM Algorithm

Direct maximization of the observed data likelihood can be cumbersome. An alternative is to use the EM algorithm. However, as can be seen from the direct approach, the distribution of the u_i's given the observed data is quite difficult to obtain. In this situation it is more reasonable to define the complete data as being the unobserved times of the events in addition to the latent random effects. The complete data likelihood is

$$\begin{aligned} L_c &= \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} u_i \lambda(t_{ij}) e^{x'_{ij}\beta} \exp\left(-u_i \Lambda(t_{ij}) e^{x'_{ij}\beta} \right) \right\} \frac{u_i^{\phi^{-1}-1} \exp(-\phi^{-1} u_i)}{\Gamma(\phi^{-1}) \phi^{\phi^{-1}}} \\ &= \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) e^{x'_{ij}\beta} \right\} \frac{u_i^{\phi^{-1}+n_i-1}}{\Gamma(\phi^{-1}) \phi^{\phi^{-1}}} \exp\left\{ -u_i \left(\phi^{-1} + \sum_{j=1}^{n_i} \Lambda(t_{ij}) e^{x'_{ij}\beta} \right) \right\}. \end{aligned}$$

From this, it can be seen that u_i given $t_{i1}, \ldots, t_{in_i}$ follows a gamma distribution with rate $\phi^{-1} + \sum_{j=1}^{n_i} \Lambda(t_{ij}) e^{x'_{ij}\beta}$ and shape $\phi^{-1} + n_i$. Taking the logarithm of L_c and disregarding terms involving data only, we obtain the complete-data log-likelihood, $\ell_c(\lambda, \beta, \phi) = \ell_{c1}(\lambda, \beta) + \ell_{c2}(\phi)$ where

$$\begin{aligned} \ell_{c1}(\lambda, \beta) &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \log \lambda(t_{ij}) + x'_{ij}\beta - u_i \Lambda(t_{ij}) e^{x'_{ij}\beta} \right\} \\ \ell_{c2}(\phi) &= -m \left\{ \log \Gamma(\phi^{-1}) + \phi^{-1} \log(\phi) \right\} + \sum_{i=1}^{m} \phi^{-1} (\log u_i - u_i). \end{aligned}$$

If we assume a piecewise constant hazard function, then substituting the corresponding piecewise-constant forms for $\lambda(t)$ and $\Lambda(t)$ we get

$$\ell_{c1}(\lambda, \beta) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \sum_{k=1}^{K} \left[d_{ijk} \log \lambda_k - u_i \lambda_k w_k(t_{ij}) e^{x'_{ij}\beta} \right] + x'_{ij}\beta \right\}.$$

In the E-step, we take the expectation of the complete-data log-likelihood given the observed data and parameter estimates from the previous iteration. The expectations can be taken in two stages. First, we take the expectation given the t_{ij}'s and observed data. In this stage we require the expectations of u_i and $\log u_i$ which are given by

$$E\left[u_i|\{t_{ij}\},\{C_{ij}\},\{Y_{ij}\},\theta\right] = \frac{\phi^{-1}+n_i}{\phi^{-1}+\sum_{j=1}^{n_i}\Lambda(t_{ij})e^{x'_{ij}\beta}}$$

$$E\left[\log u_i|\{t_{ij}\},\{C_{ij}\},\{Y_{ij}\},\theta\right] = \Psi(\phi^{-1}+n_i) - \log\left\{\phi^{-1}+\sum_{j=1}^{n_i}\Lambda(t_{ij})e^{x'_{ij}\beta}\right\}$$

where $\Psi(u) = d\log\Gamma(u)/du$.

The second stage of the E-step involves computing expectations given the observed data alone. We now require the expectations of $d_{ijk} = I(a_{k-1} < t_{ij} \leq a_k)$, $u_i w_k(t_{ij})$, u_i and $\log u_i$. The distribution of the t_{ij}'s given the observed data is difficult to work out, so a Monte Carlo approach can be taken. Given estimates from the previous iteration $\widehat{\lambda}_k^{(r)}$, $k = 1,\ldots,K$, $\widehat{\beta}^{(r)}$, $\widehat{\phi}^{(r)}$ we generate B replications of t_{ij} as follows. Generate u_i^* to be gamma with mean 1 and variance $\widehat{\phi}^{(r)}$. If $Y_{ij} = 1$ then t_{ij} solves

$$\frac{\exp(-u_i^* e^{x'_{ij}\widehat{\beta}^{(r)}}\widehat{\Lambda}(t))}{1-\exp(-u_i^* e^{x'_{ij}\widehat{\beta}^{(r)}}\widehat{\Lambda}(C_{ij}))} = z$$

where z is a uniform $(0,1]$ random variable. This implies that

$$t_{ij} = \widehat{\Lambda}^{-1}\left\{-\log\left[z(1-\exp(-u_i^* e^{x'_{ij}\widehat{\beta}^{(r)}}\widehat{\Lambda}(C_{ij})))\right]/u_i^* e^{x'_{ij}\widehat{\beta}^{(r)}}\right\}.$$

Similarly, if $Y_{ij} = 0$ then

$$t_{ij} = \widehat{\Lambda}^{-1}\left\{-\log\left[z\exp(-u_i^* e^{x'_{ij}\widehat{\beta}^{(r)}}\widehat{\Lambda}(C_{ij}))\right]/u_i^* e^{x'_{ij}\widehat{\beta}^{(r)}}\right\}.$$

The M-step in this case does not have a closed form, so numerical methods such as Newton-Raphson must be used to maximize the expected log-likelihood. Iteration continues until differences in successive parameter estimates fall below a specified tolerance.

Clearly, both the direct maximization and EM approaches require difficult computations. The direct approach results in a likelihood that is not easy to evaluate while the EM approach requires Monte Carlo in the E-step and a numerical solution in the M-step.

APPENDICES

Appendix A

Full Simulation Results From Chapter 3

Presented here are the full results of the simulation studies from Chapter 3. The complete results of the relative efficiency study are in Tables A.1 through A.3. The asymptotic relative efficiencies are obtained by evaluating the expectations in Section 3.3.2 using Monte Carlo methods based on 100,000 Monte Carlo samples. The full results of the robustness to the choice of copula are given in Tables A.4 through A.6. Here, the true association was taken to be Gumbel, with estimation according to the Clayton copula. Tables A.7 through A.14 contain the full simulation results corresponding to Section 3.4. In all cases, the parameters considered were $\rho = 0.4, 0.6$, $\beta = \log 0.8, 0$, $\gamma = 1, 1.2$, $\sigma^2 = 0.75, 1$, $p = 0.05$ and $\tau = 1$.

Table A.1: Asymptotic relative efficiencies compared to GEE2 ($n = 2$)

						$\widehat{\lambda}$		$\widehat{\alpha}$		$\widehat{\beta}$		$\widehat{\psi}$
n	α	σ^2	ρ	ϕ	β	WI	GEE1	WI	GEE1	WI	GEE1	GEE1
2	1.0	0.75	0.4	0.5	$\log 0.8$	0.990	0.999	0.982	0.996	0.984	0.998	0.418
2	1.0	0.75	0.4	0.5	0	0.990	0.999	0.984	0.996	0.985	0.998	0.414
2	1.0	0.75	0.4	3.0	$\log 0.8$	0.934	0.997	0.930	0.994	0.884	0.992	0.107
2	1.0	0.75	0.4	3.0	0	0.937	0.997	0.936	0.995	0.890	0.993	0.105
2	1.0	0.75	0.6	0.5	$\log 0.8$	0.991	0.999	0.983	0.997	0.985	0.998	0.432
2	1.0	0.75	0.6	0.5	0	0.991	0.999	0.984	0.997	0.987	0.998	0.409
2	1.0	0.75	0.6	3.0	$\log 0.8$	0.932	0.996	0.915	0.992	0.875	0.991	0.141
2	1.0	0.75	0.6	3.0	0	0.938	0.996	0.924	0.994	0.885	0.991	0.135
2	1.0	1.00	0.4	0.5	$\log 0.8$	0.990	0.999	0.983	0.996	0.985	0.998	0.410
2	1.0	1.00	0.4	0.5	0	0.991	0.999	0.984	0.996	0.986	0.998	0.403
2	1.0	1.00	0.4	3.0	$\log 0.8$	0.937	0.997	0.933	0.995	0.893	0.993	0.104
2	1.0	1.00	0.4	3.0	0	0.940	0.997	0.938	0.995	0.898	0.993	0.102
2	1.0	1.00	0.6	0.5	$\log 0.8$	0.991	0.999	0.983	0.997	0.986	0.998	0.415
2	1.0	1.00	0.6	0.5	0	0.992	0.999	0.985	0.997	0.988	0.998	0.394
2	1.0	1.00	0.6	3.0	$\log 0.8$	0.935	0.996	0.919	0.993	0.884	0.991	0.136
2	1.0	1.00	0.6	3.0	0	0.941	0.997	0.927	0.994	0.893	0.992	0.129
2	1.2	0.75	0.4	0.5	$\log 0.8$	0.990	0.999	0.983	0.996	0.985	0.998	0.414
2	1.2	0.75	0.4	0.5	0	0.991	0.999	0.984	0.997	0.986	0.998	0.403
2	1.2	0.75	0.4	3.0	$\log 0.8$	0.936	0.997	0.933	0.995	0.891	0.993	0.104
2	1.2	0.75	0.4	3.0	0	0.938	0.997	0.938	0.995	0.895	0.993	0.103
2	1.2	0.75	0.6	0.5	$\log 0.8$	0.991	0.999	0.983	0.997	0.986	0.998	0.422
2	1.2	0.75	0.6	0.5	0	0.991	0.999	0.985	0.997	0.987	0.998	0.401
2	1.2	0.75	0.6	3.0	$\log 0.8$	0.935	0.996	0.918	0.993	0.882	0.991	0.136
2	1.2	0.75	0.6	3.0	0	0.940	0.997	0.927	0.994	0.891	0.992	0.129
2	1.2	1.00	0.4	0.5	$\log 0.8$	0.991	0.999	0.984	0.997	0.986	0.998	0.400
2	1.2	1.00	0.4	0.5	0	0.991	0.999	0.985	0.997	0.987	0.998	0.390
2	1.2	1.00	0.4	3.0	$\log 0.8$	0.940	0.997	0.935	0.995	0.899	0.993	0.103
2	1.2	1.00	0.4	3.0	0	0.943	0.997	0.940	0.996	0.904	0.994	0.099
2	1.2	1.00	0.6	0.5	$\log 0.8$	0.991	0.999	0.983	0.997	0.987	0.998	0.404
2	1.2	1.00	0.6	0.5	0	0.992	0.999	0.985	0.997	0.988	0.998	0.387
2	1.2	1.00	0.6	3.0	$\log 0.8$	0.937	0.996	0.921	0.993	0.890	0.992	0.131
2	1.2	1.00	0.6	3.0	0	0.942	0.997	0.930	0.994	0.898	0.992	0.125

Table A.2: Asymptotic relative efficiencies compared to GEE2 ($n = 5$)

						$\widehat{\lambda}$		$\widehat{\alpha}$		$\widehat{\beta}$		$\widehat{\psi}$
n	α	σ^2	ρ	ϕ	β	WI	GEE1	WI	GEE1	WI	GEE1	GEE1
5	1.0	0.75	0.4	0.5	$\log 0.8$	0.972	0.996	0.948	0.991	0.952	0.993	0.347
5	1.0	0.75	0.4	0.5	0	0.973	0.997	0.951	0.991	0.954	0.993	0.339
5	1.0	0.75	0.4	3.0	$\log 0.8$	0.870	0.991	0.848	0.988	0.722	0.977	0.083
5	1.0	0.75	0.4	3.0	0	0.874	0.991	0.856	0.990	0.729	0.978	0.080
5	1.0	0.75	0.6	0.5	$\log 0.8$	0.975	0.997	0.952	0.992	0.956	0.994	0.375
5	1.0	0.75	0.6	0.5	0	0.976	0.997	0.957	0.993	0.959	0.994	0.351
5	1.0	0.75	0.6	3.0	$\log 0.8$	0.883	0.990	0.842	0.987	0.719	0.973	0.111
5	1.0	0.75	0.6	3.0	0	0.889	0.991	0.853	0.989	0.730	0.974	0.103
5	1.0	1.00	0.4	0.5	$\log 0.8$	0.973	0.997	0.948	0.991	0.955	0.993	0.336
5	1.0	1.00	0.4	0.5	0	0.973	0.997	0.951	0.992	0.956	0.994	0.329
5	1.0	1.00	0.4	3.0	$\log 0.8$	0.874	0.991	0.848	0.989	0.736	0.978	0.081
5	1.0	1.00	0.4	3.0	0	0.877	0.992	0.856	0.990	0.741	0.979	0.079
5	1.0	1.00	0.6	0.5	$\log 0.8$	0.975	0.997	0.953	0.992	0.958	0.994	0.360
5	1.0	1.00	0.6	0.5	0	0.977	0.997	0.957	0.993	0.961	0.995	0.339
5	1.0	1.00	0.6	3.0	$\log 0.8$	0.885	0.990	0.843	0.987	0.732	0.974	0.106
5	1.0	1.00	0.6	3.0	0	0.891	0.991	0.855	0.990	0.743	0.975	0.099
5	1.2	0.75	0.4	0.5	$\log 0.8$	0.972	0.997	0.950	0.991	0.954	0.993	0.342
5	1.2	0.75	0.4	0.5	0	0.973	0.997	0.953	0.992	0.956	0.993	0.333
5	1.2	0.75	0.4	3.0	$\log 0.8$	0.872	0.991	0.853	0.989	0.732	0.978	0.080
5	1.2	0.75	0.4	3.0	0	0.875	0.991	0.861	0.991	0.737	0.978	0.078
5	1.2	0.75	0.6	0.5	$\log 0.8$	0.975	0.997	0.953	0.992	0.958	0.994	0.363
5	1.2	0.75	0.6	0.5	0	0.977	0.997	0.957	0.993	0.960	0.995	0.343
5	1.2	0.75	0.6	3.0	$\log 0.8$	0.884	0.990	0.844	0.988	0.729	0.974	0.107
5	1.2	0.75	0.6	3.0	0	0.890	0.991	0.855	0.990	0.739	0.975	0.099
5	1.2	1.00	0.4	0.5	$\log 0.8$	0.973	0.997	0.951	0.992	0.957	0.994	0.331
5	1.2	1.00	0.4	0.5	0	0.974	0.997	0.954	0.992	0.959	0.994	0.324
5	1.2	1.00	0.4	3.0	$\log 0.8$	0.875	0.992	0.853	0.989	0.745	0.979	0.079
5	1.2	1.00	0.4	3.0	0	0.879	0.992	0.862	0.991	0.753	0.980	0.077
5	1.2	1.00	0.6	0.5	$\log 0.8$	0.975	0.997	0.954	0.993	0.960	0.994	0.349
5	1.2	1.00	0.6	0.5	0	0.977	0.997	0.958	0.994	0.963	0.995	0.331
5	1.2	1.00	0.6	3.0	$\log 0.8$	0.886	0.991	0.845	0.988	0.743	0.976	0.102
5	1.2	1.00	0.6	3.0	0	0.893	0.992	0.857	0.990	0.754	0.977	0.094

Table A.3: Asymptotic relative efficiencies compared to GEE2 ($n = 10$)

						$\widehat{\lambda}$		$\widehat{\alpha}$		$\widehat{\beta}$		$\widehat{\psi}$
n	α	σ^2	ρ	ϕ	β	WI	GEE1	WI	GEE1	WI	GEE1	GEE1
10	1.0	0.75	0.4	0.5	$\log 0.8$	0.958	0.994	0.913	0.990	0.922	0.987	0.274
10	1.0	0.75	0.4	0.5	0	0.959	0.995	0.916	0.991	0.925	0.988	0.264
10	1.0	0.75	0.4	3.0	$\log 0.8$	0.863	0.985	0.807	0.989	0.622	0.962	0.076
10	1.0	0.75	0.4	3.0	0	0.865	0.985	0.814	0.991	0.624	0.962	0.074
10	1.0	0.75	0.6	0.5	$\log 0.8$	0.965	0.995	0.933	0.992	0.930	0.989	0.331
10	1.0	0.75	0.6	0.5	0	0.967	0.996	0.937	0.993	0.934	0.990	0.309
10	1.0	0.75	0.6	3.0	$\log 0.8$	0.886	0.984	0.831	0.989	0.633	0.957	0.106
10	1.0	0.75	0.6	3.0	0	0.891	0.985	0.840	0.991	0.641	0.958	0.098
10	1.0	1.00	0.4	0.5	$\log 0.8$	0.958	0.995	0.910	0.990	0.927	0.988	0.262
10	1.0	1.00	0.4	0.5	0	0.959	0.995	0.914	0.991	0.929	0.988	0.254
10	1.0	1.00	0.4	3.0	$\log 0.8$	0.864	0.986	0.803	0.989	0.635	0.964	0.074
10	1.0	1.00	0.4	3.0	0	0.867	0.986	0.810	0.991	0.639	0.964	0.071
10	1.0	1.00	0.6	0.5	$\log 0.8$	0.965	0.996	0.932	0.992	0.933	0.990	0.317
10	1.0	1.00	0.6	0.5	0	0.967	0.996	0.937	0.993	0.937	0.990	0.295
10	1.0	1.00	0.6	3.0	$\log 0.8$	0.886	0.985	0.830	0.989	0.645	0.959	0.100
10	1.0	1.00	0.6	3.0	0	0.891	0.986	0.839	0.991	0.653	0.960	0.092
10	1.2	0.75	0.4	0.5	$\log 0.8$	0.959	0.994	0.919	0.991	0.926	0.988	0.276
10	1.2	0.75	0.4	0.5	0	0.960	0.995	0.922	0.992	0.928	0.988	0.267
10	1.2	0.75	0.4	3.0	$\log 0.8$	0.864	0.986	0.816	0.989	0.633	0.963	0.074
10	1.2	0.75	0.4	3.0	0	0.866	0.986	0.823	0.991	0.635	0.963	0.071
10	1.2	0.75	0.6	0.5	$\log 0.8$	0.965	0.996	0.933	0.992	0.932	0.990	0.321
10	1.2	0.75	0.6	0.5	0	0.967	0.996	0.938	0.993	0.936	0.990	0.299
10	1.2	0.75	0.6	3.0	$\log 0.8$	0.886	0.985	0.831	0.989	0.643	0.959	0.101
10	1.2	0.75	0.6	3.0	0	0.891	0.986	0.840	0.991	0.650	0.959	0.093
10	1.2	1.00	0.4	0.5	$\log 0.8$	0.959	0.995	0.917	0.991	0.930	0.989	0.266
10	1.2	1.00	0.4	0.5	0	0.960	0.995	0.921	0.992	0.932	0.989	0.257
10	1.2	1.00	0.4	3.0	$\log 0.8$	0.865	0.987	0.813	0.989	0.647	0.965	0.072
10	1.2	1.00	0.4	3.0	0	0.868	0.987	0.821	0.991	0.650	0.965	0.068
10	1.2	1.00	0.6	0.5	$\log 0.8$	0.965	0.996	0.933	0.992	0.936	0.990	0.306
10	1.2	1.00	0.6	0.5	0	0.967	0.996	0.937	0.993	0.939	0.991	0.287
10	1.2	1.00	0.6	3.0	$\log 0.8$	0.886	0.986	0.829	0.989	0.656	0.961	0.096
10	1.2	1.00	0.6	3.0	0	0.891	0.987	0.838	0.991	0.664	0.962	0.088

Table A.4: Bias due to misspecification of the copula ($n = 2$).

						GEE1				GEE2			
n	α	σ^2	ρ	$\tau^\dagger$	β	$\widehat{\lambda}-\lambda$	$\widehat{\alpha}-\alpha$	$\widehat{\beta}-\beta$	$\widehat{\tau}^\ddagger$	$\widehat{\lambda}-\lambda$	$\widehat{\alpha}-\alpha$	$\widehat{\beta}-\beta$	$\widehat{\tau}^\ddagger$
2	1.0	0.75	0.4	0.2	-0.223	0	0	0	0.103	0.001	-0.003	0.001	0.139
2	1.0	0.75	0.4	0.2	0.000	0	0	0	0.105	0.001	-0.003	0.000	0.143
2	1.0	0.75	0.4	0.6	-0.223	0	0	0	0.541	0.000	-0.012	0.001	0.503
2	1.0	0.75	0.4	0.6	0.000	0	0	0	0.553	0.001	-0.011	-0.001	0.517
2	1.0	0.75	0.6	0.2	-0.223	0	0	0	0.145	0.007	-0.006	0.001	0.198
2	1.0	0.75	0.6	0.2	0.000	0	0	0	0.142	0.006	-0.005	0.000	0.198
2	1.0	0.75	0.6	0.6	-0.223	0	0	0	0.751	0.013	-0.020	0.002	0.633
2	1.0	0.75	0.6	0.6	0.000	0	0	0	0.784	0.010	-0.017	-0.001	0.634
2	1.0	1.00	0.4	0.2	-0.223	0	0	0	0.104	0.002	-0.004	0.000	0.145
2	1.0	1.00	0.4	0.2	0.000	0	0	0	0.102	0.002	-0.003	0.000	0.143
2	1.0	1.00	0.4	0.6	-0.223	0	0	0	0.541	0.001	-0.013	0.002	0.501
2	1.0	1.00	0.4	0.6	0.000	0	0	0	0.612	0.000	-0.011	0.000	0.511
2	1.0	1.00	0.6	0.2	-0.223	0	0	0	0.146	0.007	-0.006	0.001	0.205
2	1.0	1.00	0.6	0.2	0.000	0	0	0	0.141	0.008	-0.005	0.000	0.206
2	1.0	1.00	0.6	0.6	-0.223	0	0	0	0.773	0.008	-0.019	0.004	0.638
2	1.0	1.00	0.6	0.6	0.000	0	0	0	0.904	0.010	-0.019	0.001	0.643
2	1.2	0.75	0.4	0.2	-0.223	0	0	0	0.101	0.001	-0.004	0.001	0.141
2	1.2	0.75	0.4	0.2	0.000	0	0	0	0.101	0.002	-0.004	0.000	0.144
2	1.2	0.75	0.4	0.6	-0.223	0	0	0	0.563	0.001	-0.016	0.001	0.517
2	1.2	0.75	0.4	0.6	0.000	0	0	0	0.532	-0.002	-0.015	0.001	0.511
2	1.2	0.75	0.6	0.2	-0.223	0	0	0	0.143	0.004	-0.006	0.001	0.195
2	1.2	0.75	0.6	0.2	0.000	0	0	0	0.139	0.005	-0.006	0.000	0.200
2	1.2	0.75	0.6	0.6	-0.223	0	0	0	0.794	0.011	-0.024	0.000	0.629
2	1.2	0.75	0.6	0.6	0.000	0	0	0	0.909	0.007	-0.021	-0.001	0.651
2	1.2	1.00	0.4	0.2	-0.223	0	0	0	0.103	0.003	-0.004	0.000	0.147
2	1.2	1.00	0.4	0.2	0.000	0	0	0	0.101	0.001	-0.003	0.000	0.144
2	1.2	1.00	0.4	0.6	-0.223	0	0	0	0.530	0.005	-0.016	-0.001	0.510
2	1.2	1.00	0.4	0.6	0.000	0	0	0	0.511	0.001	-0.014	0.001	0.502
2	1.2	1.00	0.6	0.2	-0.223	0	0	0	0.146	0.006	-0.007	0.000	0.205
2	1.2	1.00	0.6	0.2	0.000	0	0	0	0.142	0.006	-0.006	0.000	0.208
2	1.2	1.00	0.6	0.6	-0.223	0	0	0	0.886	0.008	-0.023	0.004	0.635
2	1.2	1.00	0.6	0.6	0.000	0	0	0	0.916	0.009	-0.022	0.000	0.629

† Kendall's τ under the Gumbel copula

‡ Estimate of Kendall's τ under the Clayton copula

Table A.5: Bias due to misspecification of the copula ($n = 5$).

						GEE1				GEE2			
n	α	σ^2	ρ	$\tau^\dagger$	β	$\widehat{\lambda}-\lambda$	$\widehat{\alpha}-\alpha$	$\widehat{\beta}-\beta$	$\widehat{\tau}^\ddagger$	$\widehat{\lambda}-\lambda$	$\widehat{\alpha}-\alpha$	$\widehat{\beta}-\beta$	$\widehat{\tau}^\ddagger$
5	1.0	0.75	0.4	0.2	-0.223	0	0	0	0.102	0.003	-0.010	0.001	0.143
5	1.0	0.75	0.4	0.2	0.000	0	0	0	0.103	0.004	-0.010	0.000	0.145
5	1.0	0.75	0.4	0.6	-0.223	0	0	0	0.585	-0.040	-0.032	0.006	0.508
5	1.0	0.75	0.4	0.6	0.000	0	0	0	0.572	-0.038	-0.028	-0.001	0.517
5	1.0	0.75	0.6	0.2	-0.223	0	0	0	0.148	0.013	-0.015	0.002	0.197
5	1.0	0.75	0.6	0.2	0.000	0	0	0	0.151	0.014	-0.014	0.000	0.206
5	1.0	0.75	0.6	0.6	-0.223	0	0	0	0.772	-0.050	-0.045	0.002	0.606
5	1.0	0.75	0.6	0.6	0.000	0	0	0	0.764	-0.047	-0.032	0.000	0.609
5	1.0	1.00	0.4	0.2	-0.223	0	0	0	0.099	0.006	-0.011	0.002	0.147
5	1.0	1.00	0.4	0.2	0.000	0	0	0	0.098	0.005	-0.010	0.000	0.144
5	1.0	1.00	0.4	0.6	-0.223	0	0	0	0.582	-0.028	-0.031	0.001	0.511
5	1.0	1.00	0.4	0.6	0.000	0	0	0	0.556	-0.032	-0.027	0.001	0.507
5	1.0	1.00	0.6	0.2	-0.223	0	0	0	0.142	0.016	-0.016	0.002	0.199
5	1.0	1.00	0.6	0.2	0.000	0	0	0	0.142	0.014	-0.014	0.000	0.200
5	1.0	1.00	0.6	0.6	-0.223	0	0	0	0.811	-0.045	-0.036	0.002	0.607
5	1.0	1.00	0.6	0.6	0.000	0	0	0	0.915	-0.044	-0.033	0.003	0.615
5	1.2	0.75	0.4	0.2	-0.223	0	0	0	0.105	0.003	-0.012	0.001	0.147
5	1.2	0.75	0.4	0.2	0.000	0	0	0	0.102	0.004	-0.011	0.000	0.146
5	1.2	0.75	0.4	0.6	-0.223	0	0	0	0.540	-0.024	-0.033	0.002	0.502
5	1.2	0.75	0.4	0.6	0.000	0	0	0	0.542	-0.019	-0.031	-0.002	0.505
5	1.2	0.75	0.6	0.2	-0.223	0	0	0	0.143	0.011	-0.019	0.002	0.200
5	1.2	0.75	0.6	0.2	0.000	0	0	0	0.142	0.010	-0.017	0.001	0.200
5	1.2	0.75	0.6	0.6	-0.223	0	0	0	0.830	-0.034	-0.043	0.004	0.610
5	1.2	0.75	0.6	0.6	0.000	0	0	0	0.861	-0.032	-0.038	0.002	0.615
5	1.2	1.00	0.4	0.2	-0.223	0	0	0	0.100	0.005	-0.012	0.001	0.147
5	1.2	1.00	0.4	0.2	0.000	0	0	0	0.100	0.004	-0.011	0.000	0.147
5	1.2	1.00	0.4	0.6	-0.223	0	0	0	0.587	-0.018	-0.036	0.003	0.514
5	1.2	1.00	0.4	0.6	0.000	0	0	0	0.568	-0.016	-0.033	0.000	0.518
5	1.2	1.00	0.6	0.2	-0.223	0	0	0	0.140	0.013	-0.019	0.001	0.200
5	1.2	1.00	0.6	0.2	0.000	0	0	0	0.140	0.011	-0.016	0.000	0.201
5	1.2	1.00	0.6	0.6	-0.223	0	0	0	0.865	-0.022	-0.046	-0.001	0.614
5	1.2	1.00	0.6	0.6	0.000	0	0	0	0.921	-0.022	-0.040	-0.001	0.616

† Kendall's τ under the Gumbel copula

‡ Estimate of Kendall's τ under the Clayton copula

Table A.6: Bias due to misspecification of the copula ($n = 10$).

						GEE1				GEE2			
n	α	σ^2	ρ	$\tau^\dagger$	β	$\widehat{\lambda}-\lambda$	$\widehat{\alpha}-\alpha$	$\widehat{\beta}-\beta$	$\widehat{\tau}^\ddagger$	$\widehat{\lambda}-\lambda$	$\widehat{\alpha}-\alpha$	$\widehat{\beta}-\beta$	$\widehat{\tau}^\ddagger$
10	1.0	0.75	0.4	0.2	-0.223	0	0	0	0.102	0.000	-0.017	0.002	0.149
10	1.0	0.75	0.4	0.2	0.000	0	0	0	0.102	0.001	-0.016	0.000	0.152
10	1.0	0.75	0.4	0.6	-0.223	0	0	0	0.563	-0.122	-0.028	0.002	0.493
10	1.0	0.75	0.4	0.6	0.000	0	0	0	0.549	-0.120	-0.023	0.001	0.494
10	1.0	0.75	0.6	0.2	-0.223	0	0	0	0.145	0.009	-0.021	0.002	0.197
10	1.0	0.75	0.6	0.2	0.000	0	0	0	0.150	0.010	-0.019	-0.001	0.205
10	1.0	0.75	0.6	0.6	-0.223	0	0	0	0.764	-0.178	-0.022	0.001	0.576
10	1.0	0.75	0.6	0.6	0.000	0	0	0	0.827	-0.168	-0.017	0.000	0.583
10	1.0	1.00	0.4	0.2	-0.223	0	0	0	0.100	0.004	-0.017	0.001	0.151
10	1.0	1.00	0.4	0.2	0.000	0	0	0	0.098	0.004	-0.016	-0.000	0.151
10	1.0	1.00	0.4	0.6	-0.223	0	0	0	0.562	-0.117	-0.028	0.002	0.498
10	1.0	1.00	0.4	0.6	0.000	0	0	0	0.552	-0.112	-0.024	-0.001	0.496
10	1.0	1.00	0.6	0.2	-0.223	0	0	0	0.143	0.013	-0.021	0.002	0.199
10	1.0	1.00	0.6	0.2	0.000	0	0	0	0.144	0.011	-0.019	-0.000	0.203
10	1.0	1.00	0.6	0.6	-0.223	0	0	0	0.816	-0.164	-0.024	-0.002	0.578
10	1.0	1.00	0.6	0.6	0.000	0	0	0	0.898	-0.154	-0.018	-0.003	0.584
10	1.2	0.75	0.4	0.2	-0.223	0	0	0	0.103	0.002	-0.020	0.002	0.152
10	1.2	0.75	0.4	0.2	0.000	0	0	0	0.102	0.001	-0.018	0.001	0.151
10	1.2	0.75	0.4	0.6	-0.223	0	0	0	0.565	-0.076	-0.034	0.001	0.499
10	1.2	0.75	0.4	0.6	0.000	0	0	0	0.558	-0.078	-0.027	0.000	0.501
10	1.2	0.75	0.6	0.2	-0.223	0	0	0	0.145	0.008	-0.025	0.002	0.200
10	1.2	0.75	0.6	0.2	0.000	0	0	0	0.141	0.007	-0.022	0.001	0.199
10	1.2	0.75	0.6	0.6	-0.223	0	0	0	0.819	-0.120	-0.028	0.002	0.578
10	1.2	0.75	0.6	0.6	0.000	0	0	0	0.892	-0.113	-0.020	0.000	0.585
10	1.2	1.00	0.4	0.2	-0.223	0	0	0	0.101	0.004	-0.020	0.002	0.152
10	1.2	1.00	0.4	0.2	0.000	0	0	0	0.100	0.003	-0.018	0.000	0.151
10	1.2	1.00	0.4	0.6	-0.223	0	0	0	0.581	-0.072	-0.035	-0.001	0.504
10	1.2	1.00	0.4	0.6	0.000	0	0	0	0.542	-0.070	-0.028	-0.002	0.497
10	1.2	1.00	0.6	0.2	-0.223	0	0	0	0.139	0.010	-0.026	0.002	0.199
10	1.2	1.00	0.6	0.2	0.000	0	0	0	0.139	0.009	-0.023	-0.000	0.202
10	1.2	1.00	0.6	0.6	-0.223	0	0	0	0.862	-0.113	-0.029	-0.001	0.583
10	1.2	1.00	0.6	0.6	0.000	0	0	0	0.921	-0.104	-0.023	-0.002	0.583

† Kendall's τ under the Gumbel copula

‡ Estimate of Kendall's τ under the Clayton copula

Table A.7: Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient.

						Working Independence			GEE		
α	σ^2	ρ	β	ϕ	model	bias	ese	se	bias	ese	se
1.0	0.75	0.4	0.5	-0.223	Exponential	-0.065	0.052	0.053	-0.138	0.051	0.052
1.0	0.75	0.4	0.5	-0.223	3 Piece	-0.105	0.053	0.053	-0.156	0.052	0.052
1.0	0.75	0.4	0.5	-0.223	5 Piece	-0.112	0.053	0.053	-0.157	0.052	0.052
1.0	0.75	0.4	0.5	-0.223	Weibull	-0.099	0.052	0.053	-0.153	0.052	0.052
1.0	0.75	0.4	0.5	0.000	Exponential	0.404	0.053	0.053	0.318	0.052	0.052
1.0	0.75	0.4	0.5	0.000	3 Piece	0.402	0.053	0.053	0.323	0.052	0.052
1.0	0.75	0.4	0.5	0.000	5 Piece	0.410	0.053	0.053	0.332	0.052	0.052
1.0	0.75	0.4	0.5	0.000	Weibull	0.405	0.053	0.053	0.326	0.052	0.052
1.0	0.75	0.4	3.0	-0.223	Exponential	-0.018	0.054	0.053	-0.061	0.047	0.045
1.0	0.75	0.4	3.0	-0.223	3 Piece	-0.065	0.054	0.053	-0.083	0.047	0.046
1.0	0.75	0.4	3.0	-0.223	5 Piece	-0.088	0.054	0.053	-0.123	0.047	0.046
1.0	0.75	0.4	3.0	-0.223	Weibull	-0.050	0.054	0.053	-0.090	0.047	0.046
1.0	0.75	0.4	3.0	0.000	Exponential	-0.172	0.052	0.053	-0.017	0.046	0.046
1.0	0.75	0.4	3.0	0.000	3 Piece	-0.174	0.052	0.053	0.016	0.046	0.046
1.0	0.75	0.4	3.0	0.000	5 Piece	-0.162	0.052	0.053	0.025	0.046	0.046
1.0	0.75	0.4	3.0	0.000	Weibull	-0.173	0.052	0.053	0.033	0.046	0.046
1.0	0.75	0.6	0.5	-0.223	Exponential	-0.064	0.044	0.043	-0.087	0.043	0.043
1.0	0.75	0.6	0.5	-0.223	3 Piece	-0.110	0.045	0.044	-0.137	0.043	0.043
1.0	0.75	0.6	0.5	-0.223	5 Piece	-0.117	0.045	0.044	-0.140	0.043	0.043
1.0	0.75	0.6	0.5	-0.223	Weibull	-0.096	0.045	0.044	-0.121	0.043	0.043
1.0	0.75	0.6	0.5	0.000	Exponential	-0.043	0.042	0.044	-0.049	0.041	0.043
1.0	0.75	0.6	0.5	0.000	3 Piece	-0.042	0.042	0.044	-0.037	0.041	0.043
1.0	0.75	0.6	0.5	0.000	5 Piece	-0.039	0.042	0.044	-0.035	0.041	0.043
1.0	0.75	0.6	0.5	0.000	Weibull	-0.043	0.042	0.044	-0.039	0.041	0.043
1.0	0.75	0.6	3.0	-0.223	Exponential	0.129	0.045	0.043	0.001	0.038	0.037
1.0	0.75	0.6	3.0	-0.223	3 Piece	0.114	0.045	0.044	0.025	0.039	0.038
1.0	0.75	0.6	3.0	-0.223	5 Piece	0.102	0.046	0.044	0.011	0.039	0.038
1.0	0.75	0.6	3.0	-0.223	Weibull	0.119	0.045	0.044	0.024	0.039	0.038
1.0	0.75	0.6	3.0	0.000	Exponential	0.082	0.043	0.044	0.082	0.038	0.038
1.0	0.75	0.6	3.0	0.000	3 Piece	0.084	0.043	0.044	0.076	0.038	0.038
1.0	0.75	0.6	3.0	0.000	5 Piece	0.088	0.043	0.044	0.077	0.038	0.038
1.0	0.75	0.6	3.0	0.000	Weibull	0.083	0.043	0.044	0.078	0.038	0.038

Table A.8: Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient (continued).

						Working Independence			GEE		
α	σ^2	ρ	β	ϕ	model	bias	ese	se	bias	ese	se
1.0	1.00	0.4	0.5	-0.223	Exponential	-0.294	0.052	0.053	-0.262	0.051	0.052
1.0	1.00	0.4	0.5	-0.223	3 Piece	-0.326	0.052	0.053	-0.295	0.052	0.052
1.0	1.00	0.4	0.5	-0.223	5 Piece	-0.353	0.053	0.053	-0.331	0.052	0.052
1.0	1.00	0.4	0.5	-0.223	Weibull	-0.323	0.052	0.053	-0.295	0.052	0.052
1.0	1.00	0.4	0.5	0.000	Exponential	0.012	0.050	0.053	0.025	0.049	0.052
1.0	1.00	0.4	0.5	0.000	3 Piece	0.017	0.050	0.053	0.021	0.049	0.052
1.0	1.00	0.4	0.5	0.000	5 Piece	0.019	0.050	0.053	0.024	0.049	0.052
1.0	1.00	0.4	0.5	0.000	Weibull	0.013	0.050	0.053	0.014	0.049	0.052
1.0	1.00	0.4	3.0	-0.223	Exponential	0.107	0.055	0.053	0.153	0.048	0.046
1.0	1.00	0.4	3.0	-0.223	3 Piece	0.054	0.055	0.053	0.046	0.048	0.047
1.0	1.00	0.4	3.0	-0.223	5 Piece	0.045	0.055	0.053	0.028	0.048	0.047
1.0	1.00	0.4	3.0	-0.223	Weibull	0.065	0.055	0.053	0.063	0.048	0.047
1.0	1.00	0.4	3.0	0.000	Exponential	0.170	0.053	0.053	0.053	0.047	0.046
1.0	1.00	0.4	3.0	0.000	3 Piece	0.164	0.053	0.053	0.024	0.047	0.046
1.0	1.00	0.4	3.0	0.000	5 Piece	0.153	0.053	0.053	0.031	0.047	0.046
1.0	1.00	0.4	3.0	0.000	Weibull	0.170	0.053	0.053	0.029	0.046	0.046
1.0	1.00	0.6	0.5	-0.223	Exponential	-0.042	0.045	0.044	-0.088	0.045	0.043
1.0	1.00	0.6	0.5	-0.223	3 Piece	-0.070	0.045	0.044	-0.116	0.045	0.043
1.0	1.00	0.6	0.5	-0.223	5 Piece	-0.072	0.045	0.044	-0.111	0.045	0.043
1.0	1.00	0.6	0.5	-0.223	Weibull	-0.056	0.045	0.044	-0.097	0.045	0.043
1.0	1.00	0.6	0.5	0.000	Exponential	-0.065	0.044	0.044	-0.040	0.044	0.043
1.0	1.00	0.6	0.5	0.000	3 Piece	-0.071	0.045	0.044	-0.039	0.044	0.043
1.0	1.00	0.6	0.5	0.000	5 Piece	-0.071	0.045	0.044	-0.040	0.044	0.043
1.0	1.00	0.6	0.5	0.000	Weibull	-0.065	0.045	0.044	-0.034	0.044	0.043
1.0	1.00	0.6	3.0	-0.223	Exponential	0.159	0.043	0.044	0.042	0.038	0.038
1.0	1.00	0.6	3.0	-0.223	3 Piece	0.139	0.044	0.044	0.015	0.039	0.038
1.0	1.00	0.6	3.0	-0.223	5 Piece	0.134	0.044	0.044	0.016	0.039	0.038
1.0	1.00	0.6	3.0	-0.223	Weibull	0.153	0.043	0.044	0.033	0.038	0.038
1.0	1.00	0.6	3.0	0.000	Exponential	-0.059	0.047	0.044	-0.068	0.040	0.039
1.0	1.00	0.6	3.0	0.000	3 Piece	-0.061	0.047	0.044	-0.075	0.040	0.039
1.0	1.00	0.6	3.0	0.000	5 Piece	-0.058	0.047	0.044	-0.076	0.040	0.039
1.0	1.00	0.6	3.0	0.000	Weibull	-0.061	0.047	0.044	-0.075	0.040	0.039

Table A.9: Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient (continued).

						Working Independence			GEE		
α	σ^2	ρ	β	ϕ	model	bias	ese	se	bias	ese	se
1.2	0.75	0.4	0.5	-0.223	Exponential	2.168	0.047	0.046	2.222	0.047	0.046
1.2	0.75	0.4	0.5	-0.223	3 Piece	0.032	0.052	0.053	0.058	0.051	0.052
1.2	0.75	0.4	0.5	-0.223	5 Piece	-0.006	0.052	0.053	-0.010	0.051	0.052
1.2	0.75	0.4	0.5	-0.223	Weibull	-0.008	0.052	0.053	-0.012	0.051	0.052
1.2	0.75	0.4	0.5	0.000	Exponential	0.074	0.048	0.046	0.089	0.048	0.046
1.2	0.75	0.4	0.5	0.000	3 Piece	0.079	0.054	0.053	0.086	0.053	0.052
1.2	0.75	0.4	0.5	0.000	5 Piece	0.078	0.054	0.053	0.089	0.053	0.052
1.2	0.75	0.4	0.5	0.000	Weibull	0.079	0.054	0.053	0.086	0.053	0.052
1.2	0.75	0.4	3.0	-0.223	Exponential	2.118	0.050	0.047	2.370	0.045	0.043
1.2	0.75	0.4	3.0	-0.223	3 Piece	-0.034	0.055	0.053	0.035	0.048	0.046
1.2	0.75	0.4	3.0	-0.223	5 Piece	-0.069	0.055	0.053	-0.092	0.048	0.046
1.2	0.75	0.4	3.0	-0.223	Weibull	-0.053	0.055	0.053	-0.109	0.048	0.046
1.2	0.75	0.4	3.0	0.000	Exponential	0.242	0.051	0.046	0.203	0.046	0.043
1.2	0.75	0.4	3.0	0.000	3 Piece	0.288	0.056	0.053	0.163	0.048	0.046
1.2	0.75	0.4	3.0	0.000	5 Piece	0.276	0.056	0.053	0.198	0.049	0.046
1.2	0.75	0.4	3.0	0.000	Weibull	0.288	0.056	0.053	0.167	0.048	0.046
1.2	0.75	0.6	0.5	-0.223	Exponential	1.880	0.041	0.039	2.037	0.039	0.039
1.2	0.75	0.6	0.5	-0.223	3 Piece	0.120	0.045	0.044	0.168	0.044	0.043
1.2	0.75	0.6	0.5	-0.223	5 Piece	0.096	0.045	0.044	0.089	0.044	0.043
1.2	0.75	0.6	0.5	-0.223	Weibull	0.099	0.045	0.044	0.085	0.044	0.043
1.2	0.75	0.6	0.5	0.000	Exponential	0.018	0.040	0.039	0.042	0.041	0.039
1.2	0.75	0.6	0.5	0.000	3 Piece	0.051	0.044	0.044	0.030	0.043	0.043
1.2	0.75	0.6	0.5	0.000	5 Piece	0.046	0.044	0.044	0.023	0.043	0.043
1.2	0.75	0.6	0.5	0.000	Weibull	0.046	0.044	0.044	0.025	0.043	0.043
1.2	0.75	0.6	3.0	-0.223	Exponential	1.629	0.042	0.039	1.979	0.037	0.036
1.2	0.75	0.6	3.0	-0.223	3 Piece	-0.134	0.045	0.044	0.112	0.038	0.038
1.2	0.75	0.6	3.0	-0.223	5 Piece	-0.160	0.045	0.044	-0.147	0.039	0.038
1.2	0.75	0.6	3.0	-0.223	Weibull	-0.145	0.045	0.044	-0.144	0.039	0.038
1.2	0.75	0.6	3.0	0.000	Exponential	-0.087	0.039	0.039	-0.055	0.036	0.036
1.2	0.75	0.6	3.0	0.000	3 Piece	-0.113	0.043	0.044	-0.029	0.037	0.038
1.2	0.75	0.6	3.0	0.000	5 Piece	-0.114	0.043	0.044	-0.028	0.037	0.039
1.2	0.75	0.6	3.0	0.000	Weibull	-0.116	0.043	0.044	-0.038	0.037	0.039

Table A.10: Empirical bias ($\times 10^2$) and standard errors for WI and GEE estimators of treatment coefficient (continued).

						Working Independence			GEE		
α	σ^2	ρ	β	ϕ	model	bias	ese	se	bias	ese	se
1.2	1.00	0.4	0.5	-0.223	Exponential	2.172	0.048	0.047	2.210	0.047	0.047
1.2	1.00	0.4	0.5	-0.223	3 Piece	0.012	0.052	0.053	0.044	0.052	0.053
1.2	1.00	0.4	0.5	-0.223	5 Piece	-0.020	0.052	0.054	-0.022	0.052	0.053
1.2	1.00	0.4	0.5	-0.223	Weibull	-0.015	0.053	0.054	-0.020	0.052	0.053
1.2	1.00	0.4	0.5	0.000	Exponential	0.013	0.047	0.046	-0.020	0.047	0.046
1.2	1.00	0.4	0.5	0.000	3 Piece	-0.024	0.052	0.053	-0.011	0.051	0.053
1.2	1.00	0.4	0.5	0.000	5 Piece	-0.064	0.052	0.054	-0.055	0.051	0.053
1.2	1.00	0.4	0.5	0.000	Weibull	-0.019	0.052	0.054	-0.005	0.051	0.053
1.2	1.00	0.4	3.0	-0.223	Exponential	1.944	0.048	0.047	2.218	0.044	0.043
1.2	1.00	0.4	3.0	-0.223	3 Piece	-0.285	0.053	0.054	-0.085	0.046	0.047
1.2	1.00	0.4	3.0	-0.223	5 Piece	-0.311	0.053	0.054	-0.238	0.047	0.047
1.2	1.00	0.4	3.0	-0.223	Weibull	-0.315	0.053	0.054	-0.253	0.047	0.047
1.2	1.00	0.4	3.0	0.000	Exponential	-0.258	0.048	0.046	-0.256	0.044	0.043
1.2	1.00	0.4	3.0	0.000	3 Piece	-0.294	0.053	0.053	-0.167	0.048	0.047
1.2	1.00	0.4	3.0	0.000	5 Piece	-0.277	0.053	0.054	-0.192	0.048	0.047
1.2	1.00	0.4	3.0	0.000	Weibull	-0.291	0.053	0.054	-0.179	0.048	0.047
1.2	1.00	0.6	0.5	-0.223	Exponential	1.482	0.040	0.040	1.528	0.039	0.040
1.2	1.00	0.6	0.5	-0.223	3 Piece	-0.339	0.044	0.044	-0.275	0.043	0.043
1.2	1.00	0.6	0.5	-0.223	5 Piece	-0.363	0.044	0.044	-0.357	0.043	0.044
1.2	1.00	0.6	0.5	-0.223	Weibull	-0.363	0.044	0.044	-0.362	0.043	0.044
1.2	1.00	0.6	0.5	0.000	Exponential	0.003	0.041	0.039	-0.123	0.040	0.039
1.2	1.00	0.6	0.5	0.000	3 Piece	-0.002	0.045	0.044	-0.015	0.045	0.044
1.2	1.00	0.6	0.5	0.000	5 Piece	-0.004	0.045	0.045	-0.014	0.045	0.044
1.2	1.00	0.6	0.5	0.000	Weibull	-0.003	0.045	0.045	-0.015	0.045	0.044
1.2	1.00	0.6	3.0	-0.223	Exponential	1.840	0.040	0.040	2.160	0.036	0.037
1.2	1.00	0.6	3.0	-0.223	3 Piece	0.036	0.044	0.044	0.300	0.038	0.038
1.2	1.00	0.6	3.0	-0.223	5 Piece	0.003	0.044	0.044	0.066	0.038	0.039
1.2	1.00	0.6	3.0	-0.223	Weibull	0.023	0.044	0.044	0.076	0.038	0.039
1.2	1.00	0.6	3.0	0.000	Exponential	0.063	0.041	0.039	0.044	0.037	0.037
1.2	1.00	0.6	3.0	0.000	3 Piece	0.078	0.045	0.044	0.075	0.039	0.039
1.2	1.00	0.6	3.0	0.000	5 Piece	0.081	0.045	0.044	0.075	0.040	0.039
1.2	1.00	0.6	3.0	0.000	Weibull	0.082	0.045	0.044	0.085	0.040	0.039

Table A.11: Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution.

						$P(S_{ij} > 0.5\|x_{ij} = 0)$				Median			
						WI		GEE1		WI		GEE1	
α	σ^2	ρ	ϕ	β	model	BIAS	ESE	BIAS	ESE	BIAS	ESE	BIAS	ESE
1.0	0.75	0.4	0.5	-0.223	Exponential	-0.017	0.013	-0.002	0.009	-0.006	0.013	0.005	0.009
1.0	0.75	0.4	0.5	-0.223	3 Piece	-0.044	0.013	0.047	0.012	-0.046	0.013	0.047	0.012
1.0	0.75	0.4	0.5	-0.223	5 Piece	-0.090	0.018	0.015	0.014	-0.084	0.018	0.011	0.013
1.0	0.75	0.4	0.5	-0.223	Weibull	-0.031	0.013	-0.016	0.009	-0.034	0.013	-0.018	0.009
1.0	0.75	0.4	0.5	0.000	Exponential	0.041	0.014	0.038	0.010	0.054	0.014	0.047	0.009
1.0	0.75	0.4	0.5	0.000	3 Piece	0.036	0.014	0.115	0.013	0.020	0.014	0.105	0.013
1.0	0.75	0.4	0.5	0.000	5 Piece	0.049	0.019	0.034	0.014	0.037	0.018	0.028	0.014
1.0	0.75	0.4	0.5	0.000	Weibull	0.038	0.014	0.031	0.010	0.026	0.014	0.024	0.009
1.0	0.75	0.4	3.0	-0.223	Exponential	-0.007	0.015	0.007	0.010	-0.005	0.014	0.007	0.010
1.0	0.75	0.4	3.0	-0.223	3 Piece	-0.027	0.015	0.113	0.014	-0.005	0.014	0.105	0.013
1.0	0.75	0.4	3.0	-0.223	5 Piece	0.033	0.019	0.001	0.016	0.053	0.018	0.015	0.014
1.0	0.75	0.4	3.0	-0.223	Weibull	-0.025	0.015	-0.012	0.010	-0.011	0.014	-0.004	0.010
1.0	0.75	0.4	3.0	0.000	Exponential	-0.062	0.015	-0.032	0.010	0.013	0.014	0.019	0.010
1.0	0.75	0.4	3.0	0.000	3 Piece	-0.104	0.015	-0.033	0.014	-0.018	0.014	0.052	0.013
1.0	0.75	0.4	3.0	0.000	5 Piece	-0.042	0.019	-0.135	0.016	0.001	0.017	-0.009	0.014
1.0	0.75	0.4	3.0	0.000	Weibull	-0.074	0.015	-0.050	0.010	0.004	0.014	0.004	0.010
1.0	0.75	0.6	0.5	-0.223	Exponential	-0.003	0.011	0.004	0.008	0.009	0.011	0.013	0.008
1.0	0.75	0.6	0.5	-0.223	3 Piece	0.002	0.015	0.010	0.011	0.002	0.015	0.037	0.011
1.0	0.75	0.6	0.5	-0.223	5 Piece	0.024	0.017	0.048	0.013	0.023	0.017	0.062	0.012
1.0	0.75	0.6	0.5	-0.223	Weibull	-0.010	0.011	-0.007	0.008	-0.006	0.011	-0.003	0.008
1.0	0.75	0.6	0.5	0.000	Exponential	-0.004	0.011	0.004	0.008	0.018	0.011	0.018	0.007
1.0	0.75	0.6	0.5	0.000	3 Piece	0.001	0.014	-0.005	0.011	0.008	0.014	0.013	0.010
1.0	0.75	0.6	0.5	0.000	5 Piece	0.026	0.017	0.038	0.011	0.025	0.017	0.051	0.011
1.0	0.75	0.6	0.5	0.000	Weibull	-0.007	0.011	-0.006	0.007	-0.001	0.011	-0.002	0.007
1.0	0.75	0.6	3.0	-0.223	Exponential	0.055	0.012	0.046	0.009	0.042	0.011	0.035	0.008
1.0	0.75	0.6	3.0	-0.223	3 Piece	0.089	0.016	-0.020	0.013	0.080	0.014	-0.026	0.012
1.0	0.75	0.6	3.0	-0.223	5 Piece	0.092	0.018	0.059	0.014	0.086	0.016	0.037	0.012
1.0	0.75	0.6	3.0	-0.223	Weibull	0.057	0.012	0.034	0.008	0.048	0.011	0.029	0.008
1.0	0.75	0.6	3.0	0.000	Exponential	-0.023	0.012	-0.009	0.008	-0.013	0.011	-0.003	0.008
1.0	0.75	0.6	3.0	0.000	3 Piece	0.019	0.015	-0.045	0.013	0.052	0.014	-0.011	0.012
1.0	0.75	0.6	3.0	0.000	5 Piece	0.005	0.017	0.035	0.014	0.034	0.016	0.056	0.013
1.0	0.75	0.6	3.0	0.000	Weibull	-0.024	0.012	-0.022	0.008	-0.002	0.011	-0.006	0.008

Table A.12: Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution (continued).

						$P(S_{ij} > 0.5\|x_{ij} = 0)$				Median			
						WI		GEE1		WI		GEE1	
α	σ^2	ρ	ϕ	β	model	BIAS	ESE	BIAS	ESE	BIAS	ESE	BIAS	ESE
1.0	1.00	0.4	0.5	-0.223	Exponential	-0.044	0.013	-0.021	0.009	-0.018	0.013	-0.004	0.009
1.0	1.00	0.4	0.5	-0.223	3 Piece	-0.064	0.013	0.011	0.012	-0.059	0.013	0.018	0.012
1.0	1.00	0.4	0.5	-0.223	5 Piece	-0.120	0.017	-0.055	0.015	-0.112	0.017	-0.047	0.014
1.0	1.00	0.4	0.5	-0.223	Weibull	-0.050	0.013	-0.030	0.009	-0.044	0.013	-0.025	0.009
1.0	1.00	0.4	0.5	0.000	Exponential	-0.028	0.014	-0.010	0.009	0.007	0.013	0.014	0.009
1.0	1.00	0.4	0.5	0.000	3 Piece	-0.051	0.014	0.006	0.013	-0.041	0.014	0.020	0.013
1.0	1.00	0.4	0.5	0.000	5 Piece	-0.058	0.018	-0.066	0.015	-0.053	0.018	-0.038	0.014
1.0	1.00	0.4	0.5	0.000	Weibull	-0.028	0.013	-0.017	0.009	-0.020	0.013	-0.011	0.009
1.0	1.00	0.4	3.0	-0.223	Exponential	0.104	0.015	0.084	0.011	0.117	0.014	0.091	0.010
1.0	1.00	0.4	3.0	-0.223	3 Piece	0.080	0.016	0.173	0.015	0.099	0.014	0.162	0.013
1.0	1.00	0.4	3.0	-0.223	5 Piece	0.069	0.019	0.094	0.017	0.082	0.018	0.110	0.015
1.0	1.00	0.4	3.0	-0.223	Weibull	0.087	0.015	0.063	0.010	0.108	0.014	0.077	0.010
1.0	1.00	0.4	3.0	0.000	Exponential	0.060	0.015	0.053	0.010	0.059	0.014	0.051	0.010
1.0	1.00	0.4	3.0	0.000	3 Piece	0.030	0.015	0.110	0.014	0.035	0.014	0.104	0.013
1.0	1.00	0.4	3.0	0.000	5 Piece	0.027	0.019	0.034	0.016	0.038	0.018	0.054	0.014
1.0	1.00	0.4	3.0	0.000	Weibull	0.049	0.015	0.035	0.010	0.049	0.014	0.036	0.010
1.0	1.00	0.6	0.5	-0.223	Exponential	-0.034	0.011	-0.018	0.007	-0.018	0.011	-0.007	0.007
1.0	1.00	0.6	0.5	-0.223	3 Piece	-0.022	0.014	-0.038	0.010	-0.015	0.014	-0.016	0.010
1.0	1.00	0.6	0.5	-0.223	5 Piece	0.004	0.017	0.024	0.012	0.011	0.017	0.035	0.012
1.0	1.00	0.6	0.5	-0.223	Weibull	-0.036	0.011	-0.026	0.007	-0.035	0.011	-0.025	0.007
1.0	1.00	0.6	0.5	0.000	Exponential	-0.013	0.011	-0.003	0.008	0.015	0.011	0.016	0.008
1.0	1.00	0.6	0.5	0.000	3 Piece	-0.060	0.014	-0.022	0.010	-0.048	0.014	0.010	0.010
1.0	1.00	0.6	0.5	0.000	5 Piece	-0.072	0.017	0.092	0.012	-0.063	0.017	0.110	0.012
1.0	1.00	0.6	0.5	0.000	Weibull	-0.014	0.011	-0.011	0.007	-0.002	0.011	-0.002	0.008
1.0	1.00	0.6	3.0	-0.223	Exponential	0.021	0.013	0.023	0.009	0.016	0.012	0.019	0.008
1.0	1.00	0.6	3.0	-0.223	3 Piece	0.056	0.016	-0.042	0.013	0.057	0.015	-0.008	0.012
1.0	1.00	0.6	3.0	-0.223	5 Piece	0.092	0.019	0.051	0.015	0.081	0.018	0.078	0.013
1.0	1.00	0.6	3.0	-0.223	Weibull	0.025	0.013	0.011	0.009	0.024	0.012	0.012	0.008
1.0	1.00	0.6	3.0	0.000	Exponential	-0.045	0.013	-0.022	0.009	-0.057	0.012	-0.032	0.008
1.0	1.00	0.6	3.0	0.000	3 Piece	-0.032	0.016	-0.001	0.013	-0.001	0.015	-0.006	0.012
1.0	1.00	0.6	3.0	0.000	5 Piece	-0.003	0.019	0.087	0.015	0.025	0.018	0.067	0.013
1.0	1.00	0.6	3.0	0.000	Weibull	-0.050	0.013	-0.035	0.009	-0.044	0.012	-0.032	0.008

Table A.13: Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution (continued).

						$P(S_{ij} > 0.5\|x_{ij} = 0)$				Median			
						WI		GEE1		WI		GEE1	
α	σ^2	ρ	ϕ	β	model	BIAS	ESE	BIAS	ESE	BIAS	ESE	BIAS	ESE
1.2	0.75	0.4	0.5	-0.223	Exponential	1.063	0.013	1.581	0.010	0.728	0.013	1.325	0.010
1.2	0.75	0.4	0.5	-0.223	3 Piece	0.381	0.014	3.958	0.015	0.374	0.014	3.948	0.014
1.2	0.75	0.4	0.5	-0.223	5 Piece	-0.446	0.016	3.963	0.014	-0.451	0.016	3.967	0.014
1.2	0.75	0.4	0.5	-0.223	Weibull	-0.007	0.014	-0.016	0.008	-0.003	0.014	-0.016	0.008
1.2	0.75	0.4	0.5	0.000	Exponential	1.036	0.013	1.561	0.010	0.729	0.013	1.326	0.010
1.2	0.75	0.4	0.5	0.000	3 Piece	0.422	0.015	4.031	0.015	0.416	0.015	4.018	0.015
1.2	0.75	0.4	0.5	0.000	5 Piece	-0.397	0.018	4.008	0.014	-0.391	0.017	4.011	0.014
1.2	0.75	0.4	0.5	0.000	Weibull	-0.021	0.014	-0.012	0.009	-0.010	0.014	-0.008	0.009
1.2	0.75	0.4	3.0	-0.223	Exponential	1.078	0.014	1.595	0.011	-0.012	0.013	0.766	0.010
1.2	0.75	0.4	3.0	-0.223	3 Piece	0.454	0.017	4.079	0.017	0.458	0.015	4.022	0.015
1.2	0.75	0.4	3.0	-0.223	5 Piece	-0.441	0.018	4.064	0.017	-0.384	0.017	4.072	0.014
1.2	0.75	0.4	3.0	-0.223	Weibull	0.010	0.015	-0.005	0.009	0.032	0.014	0.009	0.009
1.2	0.75	0.4	3.0	0.000	Exponential	1.092	0.015	1.607	0.011	-0.008	0.014	0.769	0.010
1.2	0.75	0.4	3.0	0.000	3 Piece	0.479	0.017	4.074	0.017	0.468	0.016	4.019	0.016
1.2	0.75	0.4	3.0	0.000	5 Piece	-0.393	0.019	4.033	0.016	-0.362	0.017	4.064	0.014
1.2	0.75	0.4	3.0	0.000	Weibull	0.048	0.016	0.028	0.010	0.059	0.015	0.036	0.009
1.2	0.75	0.6	0.5	-0.223	Exponential	-0.147	0.011	0.661	0.008	-0.283	0.011	0.558	0.008
1.2	0.75	0.6	0.5	-0.223	3 Piece	-0.398	0.015	4.616	0.010	-0.455	0.015	4.503	0.010
1.2	0.75	0.6	0.5	-0.223	5 Piece	-0.117	0.017	3.727	0.015	-0.098	0.017	3.725	0.014
1.2	0.75	0.6	0.5	-0.223	Weibull	0.028	0.012	0.007	0.007	0.034	0.012	0.009	0.007
1.2	0.75	0.6	0.5	0.000	Exponential	-0.183	0.011	0.633	0.008	-0.432	0.012	0.448	0.009
1.2	0.75	0.6	0.5	0.000	3 Piece	-0.390	0.015	4.624	0.009	-0.470	0.015	4.509	0.009
1.2	0.75	0.6	0.5	0.000	5 Piece	-0.090	0.017	3.712	0.014	-0.103	0.017	3.711	0.014
1.2	0.75	0.6	0.5	0.000	Weibull	-0.005	0.012	-0.021	0.007	-0.001	0.012	-0.021	0.007
1.2	0.75	0.6	3.0	-0.223	Exponential	-0.172	0.013	0.645	0.010	-1.069	0.012	-0.023	0.009
1.2	0.75	0.6	3.0	-0.223	3 Piece	-0.348	0.016	4.547	0.011	-0.279	0.015	4.299	0.011
1.2	0.75	0.6	3.0	-0.223	5 Piece	-0.053	0.018	3.638	0.016	-0.003	0.017	3.640	0.014
1.2	0.75	0.6	3.0	-0.223	Weibull	-0.005	0.014	-0.001	0.008	0.026	0.013	0.015	0.008
1.2	0.75	0.6	3.0	0.000	Exponential	-0.176	0.012	0.641	0.009	-1.054	0.012	-0.012	0.009
1.2	0.75	0.6	3.0	0.000	3 Piece	-0.377	0.016	4.580	0.011	-0.358	0.015	4.298	0.010
1.2	0.75	0.6	3.0	0.000	5 Piece	-0.094	0.018	3.668	0.015	-0.067	0.017	3.652	0.013
1.2	0.75	0.6	3.0	0.000	Weibull	-0.009	0.014	-0.010	0.008	0.006	0.013	0.001	0.008

Table A.14: Empirical bias ($\times 10^2$) and empirical standard errors for $P(S_{ij} > 0.5|x_{ij} = 0)$ and the median of the baseline distribution (continued).

						$P(S_{ij} > 0.5\|x_{ij} = 0)$				Median			
						WI		GEE1		WI		GEE1	
α	σ^2	ρ	ϕ	β	model	BIAS	ESE	BIAS	ESE	BIAS	ESE	BIAS	ESE
1.2	1.00	0.4	0.5	-0.223	Exponential	0.837	0.013	1.409	0.010	0.553	0.013	1.192	0.010
1.2	1.00	0.4	0.5	-0.223	3 Piece	0.777	0.016	4.051	0.016	0.766	0.016	4.035	0.016
1.2	1.00	0.4	0.5	-0.223	5 Piece	-0.452	0.017	4.141	0.015	-0.451	0.017	4.150	0.015
1.2	1.00	0.4	0.5	-0.223	Weibull	0.006	0.014	0.000	0.009	0.016	0.014	0.002	0.009
1.2	1.00	0.4	0.5	0.000	Exponential	0.784	0.012	1.367	0.010	0.497	0.013	1.149	0.010
1.2	1.00	0.4	0.5	0.000	3 Piece	0.733	0.015	4.028	0.015	0.730	0.015	4.022	0.015
1.2	1.00	0.4	0.5	0.000	5 Piece	-0.458	0.017	4.085	0.015	-0.445	0.017	4.102	0.014
1.2	1.00	0.4	0.5	0.000	Weibull	-0.057	0.014	-0.042	0.008	-0.039	0.014	-0.034	0.008
1.2	1.00	0.4	3.0	-0.223	Exponential	0.779	0.015	1.368	0.011	-0.210	0.014	0.618	0.011
1.2	1.00	0.4	3.0	-0.223	3 Piece	0.686	0.017	3.984	0.017	0.689	0.016	3.918	0.016
1.2	1.00	0.4	3.0	-0.223	5 Piece	-0.550	0.019	4.076	0.016	-0.471	0.018	4.081	0.015
1.2	1.00	0.4	3.0	-0.223	Weibull	-0.059	0.016	-0.048	0.010	-0.018	0.015	-0.018	0.009
1.2	1.00	0.4	3.0	0.000	Exponential	0.740	0.014	1.337	0.011	-0.260	0.014	0.580	0.010
1.2	1.00	0.4	3.0	0.000	3 Piece	0.650	0.017	3.971	0.017	0.673	0.016	3.947	0.016
1.2	1.00	0.4	3.0	0.000	5 Piece	-0.537	0.018	4.037	0.016	-0.464	0.017	4.079	0.014
1.2	1.00	0.4	3.0	0.000	Weibull	-0.102	0.015	-0.076	0.009	-0.044	0.015	-0.041	0.009
1.2	1.00	0.6	0.5	-0.223	Exponential	-0.384	0.011	0.483	0.008	-0.585	0.010	0.332	0.008
1.2	1.00	0.6	0.5	-0.223	3 Piece	-0.496	0.014	4.488	0.009	-0.563	0.014	4.381	0.009
1.2	1.00	0.6	0.5	-0.223	5 Piece	-0.159	0.016	3.699	0.015	-0.168	0.016	3.712	0.015
1.2	1.00	0.6	0.5	-0.223	Weibull	-0.048	0.012	-0.037	0.007	-0.040	0.012	-0.034	0.007
1.2	1.00	0.6	0.5	0.000	Exponential	-0.359	0.011	0.501	0.008	-0.581	0.011	0.336	0.008
1.2	1.00	0.6	0.5	0.000	3 Piece	-0.429	0.014	4.526	0.009	-0.498	0.014	4.413	0.009
1.2	1.00	0.6	0.5	0.000	5 Piece	-0.043	0.017	3.757	0.015	-0.043	0.016	3.769	0.015
1.2	1.00	0.6	0.5	0.000	Weibull	-0.023	0.012	-0.022	0.007	-0.015	0.012	-0.019	0.007
1.2	1.00	0.6	3.0	-0.223	Exponential	-0.383	0.012	0.485	0.009	-1.221	0.011	-0.136	0.008
1.2	1.00	0.6	3.0	-0.223	3 Piece	-0.478	0.015	4.474	0.011	-0.453	0.014	4.226	0.010
1.2	1.00	0.6	3.0	-0.223	5 Piece	-0.116	0.018	3.733	0.017	-0.103	0.016	3.744	0.015
1.2	1.00	0.6	3.0	-0.223	Weibull	-0.049	0.013	-0.035	0.008	-0.021	0.013	-0.017	0.007
1.2	1.00	0.6	3.0	0.000	Exponential	-0.283	0.012	0.560	0.009	-1.105	0.012	-0.049	0.009
1.2	1.00	0.6	3.0	0.000	3 Piece	-0.403	0.016	4.602	0.011	-0.364	0.015	4.331	0.011
1.2	1.00	0.6	3.0	0.000	5 Piece	-0.023	0.018	3.823	0.017	0.029	0.017	3.823	0.015
1.2	1.00	0.6	3.0	0.000	Weibull	0.072	0.014	0.018	0.008	0.099	0.013	0.035	0.008

References

Alioum, A. and Commenges, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics* 52 512–524.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 10 1100–1120.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and Silverman, E. (1955). An emprirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* 26 641–647.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. John Wiley & Sons.

Bauer, K. A., Eriksson, B. I., Lassen, M. R. and Turpie, A. G. G. (2001). Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after elective major knee surgery. *New England Journal of Medicine* 345 1305–1310.

Bebchuck, J. D. and Betensky, R. A. (2001). Local likelihood analysis of survival data with censored intermediate events. *Journal of the American Statistical Association* 96 449–457.

Betensky, R. A. (2000). Redistribution algorithms for censored data. *Statistics & Probability Letters* 46 385–389.

Betensky, R. A. and Finkelstein, D. M. (1999). A nonparametric maximum likelihood estimator for bivariate interval censored data. *Statistics in Medicine* 18 3089–3100.

Betensky, R. A., Lindsey, J. C., Ryan, L. M. and Wand, M. P. (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics* 55 238–245.

Betensky, R. A., Lindsey, J. C., Ryan, L. M. and Wand, M. P. (2002). A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine* 21 263–275.

Bogaerts, K. and Lesaffre, E. (2004). A new, fast algorithm to find the regions of possible support for bivariate interval-censored data. *Journal of Computational and Graphical Statistics* 13 330–340.

Braun, J., Duchesne, T. and Stafford, J. E. (2005). Local likelihood density estimation for interval censored data. *The Canadian Journal of Statistics* 33 39–60.

Byar, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent state I bladder tumors: Comparisons of placebo, pyridoxine, and topical thiotepa. In M. Pavone-Macaluso, P. H. Smith and F. Edsmyr, eds., *Bladder Tumors and Other Topics in Urological Oncology*. New York: Plenum, 363–370.

Carey, V., Zeger, S. L. and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* 80 517–526.

Chandran, V., Tolusso, D., Schentag, C. T., Cook, R. J. and Gladman, D. D. (2008a). Defining psoriatic spondylitis. ACR Abstract.

Chandran, V., Tolusso, D., Schentag, C. T., Cook, R. J. and Gladman, D. D. (2008b). Risk factors for the development of spondylitis in patients with psoriatic arthritis. ACR Abstract.

Chen, B. E., Cook, R. J., Lawless, J. F. and Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine* 24 671–691.

Chiang, A. C. (1984). *Fundamental Methods of Mathematical Economics*. McGraw Hill, Inc.

Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research* 11 167–182.

Cook, R. J. (1999). A mixed model for two-state Markov processes under panel observation. *Biometrics* 55 915–920.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Springer.

Cook, R. J., Zeng, L. and Lee, K.-A. (2008). A multistate model for bivariate interval-censored failure time data. *Biometrics* To appear.

Cook, R. J., Zeng, L. and Yi, G. Y. (2004). Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* 60 820–828.

Datta, S. and Satten, G. A. (2001). Validity of the Aalen-Johansen estimators of state occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Statistics & Probability Letters* 55 403–411.

Dean, C. B. (1991). Estimating equations for mixed-Poisson models. In V. P. Godambe, ed., *Estimating Functions*. Oxford University Press, 35–46.

Dinse, G. E. (1985). An alternative to Efron's redistribution-of-mass construction of the Kaplan-Meier estimator. *The American Statistician* 39 299–300.

Dykstra, R. L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association* 77 621–628.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the fifth Berkeley Symposium, Vol. 4*. Berkeley, California: University of California Press, 831–853.

Eriksson, B. I., Bauer, K. A., Lassen, M. R. and Turpie, A. G. G. (2001). Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after hip-replacement surgery. *New England Journal of Medicine* 345 1298–1304.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* 42 845–854.

Finkelstein, D. M., Goggins, W. B. and Schoenfeld, D. A. (2002). Analysis of failure time data with dependent interval censoring. *Biometrics* 58 298–304.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Sage.

Fox, J. (2007). *car: Companion to Applied Regression*. R package version 1.2-2.

Frydman, H. (1992). A nonparametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *Journal of the Royal Statistical Society, Series B* 54 853–866.

Frydman, H. (1994). A note on nonparametric estimation of the distribution function from interval-censored and truncated observations. *Journal of the Royal Statistical Society, Series B* 56 71–74.

Frydman, H. (1995). Nonparametric estimation of a Markov 'illness-death' process from interval-censored observations, with application to diabetes survival data. *Biometrika* 82 773–789.

Gail, M., Santner, T. and Brown, C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* 36 255–266.

Genest, C. and MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician* 40 280–283.

Genest, C., Quessy, J.-F. and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* 33 337–366.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88 1034–1043.

Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika* 81 618–623.

Gentleman, R. and Vandal, A. (2008). *Icens: NPMLE for censored and trucated data*. R package version 1.2.0.

Gentleman, R. and Vandal, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *The Canadian Journal of Statistics* 30 557–571.

Gentleman, R. C., Lawless, J. F., Lindsey, J. C. and Yan, P. (1994). Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine* 13 805–821.

Goggins, W. B. and Finkelstein, D. M. (2000). A proportional hazards model for multivariate interval-censored failure time data. *Biometrics* 56 940–943.

Goggins, W. B., Finkelstein, D. M., Schoenfeld, D. A. and Zaslavsky, A. M. (1998). A Markov chain monte carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model. *Biometrics* 54 1498–1507.

Grüger, J., Kay, R. and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics* 47 595–605.

Hjort, N. L. and Jones, M. C. (1996). Locally parametric nonparametric density estimation. *The Annals of Statistics* 24 1619–1647.

Hortobagyi, G. N., Theriault, R. L., Lipton, A., Porter, L., Blaney, D., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., Knight, R. D., Heffernan, M., Mellars, K. and Reitsma, D. J. (1998). Long-term prevention of skeletal complications of metastatic breast cancer with Pamidronate. *Journal of Clinical Oncology* 16 2038–2044.

Hortobagyi, G. N., Theriault, R. L., Porter, L., Blaney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., Knight, R. D., Heffernan, M. and Reitsma, D. J. (1996). Efficacy of Pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *The New England Journal of Medicine* 335 1785–1791.

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis* 5 239–264.

Hudgens, M. G., Satten, G. A. and Longini, I. M. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics* 57 74–80.

Jackson, C. (2007). *msm: Multi-state Markov and hidden Markov models in continuous time*. R package version 0.7.2.

Jewell, N. P., van der Laan, M. and Lei, X. (2005). Bivariate current status data with univariate monitoring times. *Biometrika* 92 847–862.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall, New York.

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

Kalbfleisch, J. D. and Lawless, J. F. (1989). Some statistical methods for panel life history data. In *Proceedings of the Statistics Canada Symposium on Analysis of Data in Time*. Ottawa, Ontario: Statistics Canada, 185–192.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag.

Lassen, M. R., Bauer, K. A., Eriksson, B. I. and Turpie, A. G. G. (2002). Postoperative fondaparinux versus preoperative enoxaparin for prevention of venous thromboembolism in elective hip-replacement surgery: a randomized double-blind comparison. *Lancet* 359 1715–1720.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15 209–225.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons.

Lawless, J. F. and Nadeau, J. C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* 37 158–168.

Lawless, J. F. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *The Canadian Journal of Statistics* 26 549–565.

Li, L., Watkins, T. and Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics* 24 531–542.

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

Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B* 54 3–40.

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics. methods for interval-censored data. *Statistics in Medicine* 17 219–238.

Lindsey, J. K. (1998). Counts and times to events. *Statistics in Medicine* 17 1745–1751.

Loader, C. (1999). *Local Regression and Likelihood*. Springer.

Loader, C. R. (1996). Local likelihood density estimation. *The Annals of Statistics* 24 1602–1618.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall.

Pepe, M. S., Longton, G. and Thornquist, M. (1991). A qualifier Q for the survival function to describe the prevalence of a transient condition. *Statistics in Medicine* 10 413–421.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics* 22 86–91.

Præstgaard, J. T. and Huang, J. (1996). Asymptotic theory for nonparametric estimation of survival curves under order restrictions. *The Annals of Statistics* 24 1679–1716.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* 44 1033–1048.

Rahman, P., Gladman, D. D., Cook, R. J., Zhou, Y., Young, G. and Salonen, D. (1998). Radiologic assessment in psoriatic arthritis. *British Journal of Rheumatology* 37 760–765.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics* 11 453–466.

Rotnitzky, A. and Wypij, D. (1994). A note on the bias of estimators with missing data. *Biometrics* 50 1163–1170.

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika* 83 355–370.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics* 20 1768–1802.

Shiboski, S. C. (1998). Generalized additive models for current status data. *Lifetime Data Analysis* 4 29–50.

Shih, J. H. and Louis, T. A. (1995). Inference on the association parameter in copula models for bivariate survival data. *Biometrics* 51 1384–1399.

Staniswalis, J. G., Thall, P. F. and Salch, J. (1997). Semiparametric regression analysis for recurrent event interval counts. *Biometrics* 53 1334–1353.

Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer.

Sun, J. and Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association* 88 1449–1454.

Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* 5 279–290.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society, Series B* 62 293–302.

Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* 44 197–209.

Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association* 83 339–347.

Tolusso, D. and Cook, R. J. (2008a). Local likelihood for right and interval-censored recurrent event data. Working paper, University of Waterloo.

Tolusso, D. and Cook, R. J. (2008b). Multistate analysis of interval censored onset times of spondylitis in psoriatic arthritis using parametric, nonparametric and local likelihood methods. Submitted to *Communications in Statistics – Theory and Methods*.

Tolusso, D. and Cook, R. J. (2008c). Second order estimating equations for the analysis of clustered current status data. Working paper, University of Waterloo.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B* 38 290–295.

Turpie, A. G. G., Bauer, K. A., Eriksson, B. I. and Lassen, M. R. (2002). Postoperative fondaparinux versus postoperative enoxaparin for prevention of venous thromboembolism after elective hip-replacement surgery: a randomized double-blind trial. *Lancet* 359 1721–1726.

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society, Series B* 65 257–273.

Wang, W. and Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika* 87 879–893.

Wang, W. and Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association* 95 62–72.

Warkentin, T. E., Cook, R. J., Marder, V. J., Sheppard, J.-A. I., Moore, J. C., Eriksson, B. I., Greinacher, A. and Kelton, J. G. (2005). Anti-platelet factor 4/heparin antibodies in orthopedic surgery patients receiving prophylaxis with fondaparinux or enoxaparin. *Blood* 106 3791–3796.

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

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics* 28 779–814.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50 1–25.

Zhan, M. (1999). *Analysis of Incomplete Event History Data*. Ph.D. thesis, University of Waterloo.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* 77 642–648.