Analysis of Recurrent Event Data

by

Robert F. Balshaw

B.Sc. University of Manitoba, 1986

M.Sc. University of Manitoba, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Mathematics and Statistics

SIMON FRASER UNIVERSITY
November 1997

APPROVAL

Name: Robert F. Balshaw

Degree: Doctor of Philosophy

Title of thesis: Analysis of Recurrent Event Data

Examining Committee: Dr. A. Lachlan
Chair

Dr. C.B. Dean
Senior Supervisor

Dr. R. Routledge
Department of Mathematics and Statistics

Dr. R. Lockhart
Department of Mathematics and Statistics

Dr. C. Schwarz
Department of Mathematics and Statistics

Dr. B. Leroux
Department of Statistics, University of Washington

Date Approved: November 27, 1997

Abstract

Inference for point processes is most efficient if the event times for each individual are available. Sometimes, the study design is such that only count data are collected, consisting of the number of events or recurrences for each individual over the entire follow-up period or between multiple follow-up times. This thesis discusses the loss in efficiency of an analysis of such count data versus an analysis of the actual event-times. One particular case is exemplified, that in which the purpose of the experiment or trial is to compare the effects of treatments, and the loss in efficiency in the estimator of the treatment effect is computed.

The specific point process considered here is the non-homogeneous Poisson process, with a proportional intensity model for the treatment effects. Random effects models are also considered, with estimation via a quasi-likelihood approach. The quasi-likelihood analysis proposed here is an extension of such techniques for the homogeneous Poisson process. The resulting estimating equations for the parameters in the random effects models are simple and intuitive. The results show that for many usual situations, treatment effects are efficiently estimated using aggregated data; however, when only end-of-follow-up counts are collected, the underlying intensity function is not. Multiple follow-up count data is shown to recover much of the information lost by end-of-follow-up counts.

The efficiency of the quasi-likelihood estimators is shown to be high relative to specific likelihood alternatives. Tests and diagnostic procedures for checking model assumptions are presented.

The quasi-likelihood estimators developed here require the assumption of a parametric form for the intensity function. This thesis also develops a nonparametric approach to the estimation of the intensity function. Combined with quasi-likelihood estimators for covariates, this provides a simple method for the analysis of recurrent event data, requiring less stringent assumptions than traditional methods.

We examine the small sample behaviour of these procedures with simulation studies. The studies show that for the situations we consider, the methods work well and display adequate small sample characteristics. Analyses of illustrative examples demonstrate the application of the procedures.

Dedication

For MaryAnn and Michael.

Acknowledgements

I would like to thank both NSERC and Simon Fraser University for their financial support. Also, a big thank you to all the staff, students, and faculty in the Department who helped in so many different ways.

Thanks to Crawford McNair for the use of his Cherry Bark Tortix data.

A special thank you to the members of my examining committee.

Finally, I would like to express my sincere gratitude to my supervisor, Dr. Charmaine Dean, for her guidance and support.

Contents

Abstract iii
Dedication v
Acknowledgements vi
List of Tables x
List of Figures xii
1 Introduction 1
2 Background 5
2.1 The Structure of Recurrent Event Data 5
2.2 Illustrations 6
2.2.1 Air Conditioner Data 6
2.2.2 Bladder Cancer Data 6
2.2.3 Cherry Bark Tortrix Data 7
2.3 Models for Recurrent Event Data 8
2.4 Poisson Process Models 8
2.5 Proportional Intensity Models 10
2.6 Frailty Models for Heterogeneity 10
2.7 Other Stochastic Process Models 11
2.7.1 Renewal Process Models 11
2.7.2 Discrete State Markov Process Models 12
2.8 Estimation 12
2.8.1 Likelihood Estimation 13
2.8.2 Quasi-Likelihood Estimation and Other Moment Methods . . 13
2.8.3 Robust Variance Estimation 14
3 Event-Times vs. End-of-Follow-Up Counts 16

3.1 Introduction 16
3.2 Efficiency Comparisons 17
3.3 Derivations for Poisson Process Model 25
3.3.1 Case 1: Scalar α 25
3.3.2 Case 2: Vector $\boldsymbol{\alpha}$ 27
3.4 Models with random effects 29
3.5 Derivations for Quasi-Likelihood Estimators 35
3.5.1 Asymptotic Variance of $\tilde{\boldsymbol{\theta}}_{\mathrm{t}}$ and $\tilde{\boldsymbol{\theta}}_{\mathrm{c}}$ 35
3.5.2 Asymptotic Relative Efficiency of $\tilde{\boldsymbol{\theta}}_{\mathrm{c}}$ 36
3.6 Illustration 37
3.7 Discussion 44
4 The Analysis of Recurrent Event Panel Data 45
4.1 Introduction 45
4.2 Likelihood and Quasi-Likelihood Estimators 46
4.2.1 Estimation Based on the Event-times 47
4.2.2 Estimation Based on Panel Data 52
4.2.3 The Efficiency of Quasi-Likelihood Relative to Likelihood Estimation 54
4.2.4 Small Sample Characteristics of Likelihood and Quasi-Likelihood Estimators 56
4.3 Efficiency of Panel Data vs. Count Data 68
4.3.1 Asymptotic Comparison of Panel Data vs. Count Data . . . 68
4.3.2 Small Sample Comparison of Panel Data vs. Count Data . . 75
4.4 Tests and Diagnostics for Model-Checking 76
4.4.1 Testing for Overdispersion 76
4.4.2 Testing for Non-Homogeneity in the Poisson Process 78
4.4.3 Diagnostic Plot for Checking the Baseline Intensity 79
4.5 Illustration 79
5 Semiparametric Analysis of Panel Data 85
5.1 Introduction 85
5.2 A Semiparametric Model 86
5.3 Test for Parametric Baseline 91
5.4 Efficiency of Semiparametric Estimators 92

5.5 Illustrations 96
5.5.1 Analysis of the Bladder Cancer Data 96
5.5.2 Analysis of the Cherry Bark Tortrix Data 99
5.6 Small Sample Characteristics 106
6 Conclusion 126
6.1 Summary 126
6.2 Future Work 128
6.2.1 Further Investigation of Nonparametric Baseline Intensity Models 128
6.2.2 Prediction and Generalizations of the Variance Structure . . 129
6.2.3 Relationship to Generalized Estimating Equations Methods . 129
6.2.4 Two-State Model for Recurrent Events 130

List of Tables

3.1 Asymptotic Relative Efficiencies of the Estimators $\tilde{\beta}_{1c}$ and $\tilde{\alpha}_c$. 23
3.2 Asymptotic Relative Efficiencies of $\hat{\beta}_{1c}$ for Three Cases with Increasing Values of $(1 + e^{\beta_2} + e^{\beta_3})$. 24
3.3 Distribution of the Number of Recurrences Observed for the Subjects in each of the Three Treatment Groups in the Bladder Cancer Study. 38
3.4 Some Summary Statistics Relating to the Follow-Up Times of Subjects in the Three Treatment Groups for the Bladder Cancer Study. 38
3.5 Parameters Estimates from the Poisson Point Process Likelihood Analysis of the Bladder Cancer Data. 41
3.6 Parameters Estimates from the Poisson-Gamma Mixture Likelihood and the Quasi-Likelihood Analysis of the Bladder Cancer Data. 42
3.7 Asymptotic Relative Efficiencies of the Analysis of Count Data. 43

4.1 Asymptotic Relative Efficiency of Quasi-Likelihood Estimator $\tilde{\tau}$. 55
4.2 Proportion of Simulated Datasets Yielding Estimates which Failed to Converge. 57
4.3 Simulation Results for Estimation of $\boldsymbol{\theta}$ Based on End-of-Follow-Up Count Data. 58
4.4 Simulation Results for Estimation of $\boldsymbol{\theta}$ Based on Panel Data. 63
4.5 Simulation Results for Estimation of $\boldsymbol{\theta}$ Based on Event-Time Data. 63
4.6 Accuracy and Precision of Variance Estimators. 65
4.7 Observed Coverage of Nominal 95% Large-Sample Confidence Intervals. . . . 66
4.8 Ratio of Estimated Variances to Sampling Variance for End-of-Follow-Up Count Data. 66
4.9 Estimated ARE($\tilde{\theta}$) from Simulation Study. 67
4.10 Design Combinations Examined in Numerical Study. 71
4.11 Balanced and Imbalanced Designs Considered in the Numerical Study. . . . 72

4.12 Small Sample Efficiency of the Analysis of End-of-Follow-Up Counts and Panel Data. 76
4.13 Number of Subjects at Risk in the Modified Bladder Cancer Dataset. 81
4.14 Results of Likelihood and Quasi-Likelihood Analyses of the Bladder Cancer Recurrent Event Data. 81

5.1 Factors Considered in ARE Calculations for Semiparametric Estimator β_{2S}. . 93
5.2 Relative Efficiency of the Semiparametric Estimator $\tilde{\beta}_{2S}$ for $\tau = 0$. 94
5.3 Asymptotic Relative Efficiency of the Semiparametric Estimator $\tilde{\beta}_{2S}$ for $\tau = 0.8$. 95
5.4 Parameter Estimates for the Bladder Cancer Data. 97
5.5 Estimates of the Baseline Cumulative Mean Functions for the Bladder Cancer Data. 99
5.6 Parameter Estimates for Cherry Bark Tortrix Data. 103
5.7 Simulation Results for Semiparametric Estimation of $\boldsymbol{\theta}$. 117
5.8 Accuracy and Precision of Variance Estimators for ϕ_2 and ϕ_5. 118
5.9 Accuracy and Precision of Variance Estimators for β_2 and τ. 119
5.10 Observed Coverage Probabilities for Asymptotic Normal Confidence Intervals. 121
5.11 Observed Rejection Rates for Testing Specific Intensity Models When the Null Hypothesis is True. 122
5.12 Observed Rejection Rate for Testing $H_0 : \alpha = 1$ When the Null Hypothesis is False. 124

List of Figures

2.1 An Example of Recurrent Event Data — Bladder Cancer Data 7

3.1 Asymptotic Relative Efficiency of $\hat{\beta}_{1c}$. 24
3.2 Cumulative Mean Functions for the Three Treatment Groups in the Bladder Cancer Study. 39
3.3 Boxplot of End-of-Follow-Up Times By Treatment Group. 40

4.1 Asymptotic Relative Efficiency of Quasi-Likelihood Estimator $\tilde{\tau}$. 55
4.2 Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of β_1. 59
4.3 Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of β_2. 60
4.4 Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of α. 61
4.5 Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of τ. . 62
4.6 Efficiency Gains for Increased Number of Follow-Up Times, Balanced Dropout. 73
4.7 Efficiency Gains for Increased Number of Follow-Up Times, Imbalanced Dropout. 74
4.8 Estimated Cumulative Mean Functions for the Bladder Cancer Data 82
4.9 A Diagnostic Plot for Checking the Baseline Intensity Function. 83

5.1 Estimated Baseline Cumulative Mean Functions For the Bladder Cancer Data. 98
5.2 Estimates of the Cumulative Mean Functions of the Bladder Cancer Data. . . 100
5.3 Estimated Cumulative Intensities for Cherry Bark Tortrix Data. 102
5.4 Estimated Baseline Intensity Function for Cherry Bark Tortrix Data. 104
5.5 Estimated Cumulative Baseline Intensity Function for the Cherry Bark Tortrix Data. 105
5.6 Normal Probability Plot for the Standardized End-of-Follow-Up Residuals from the Cherry Bark Tortrix Data. 107

5.7 Sampling Distribution of the Weibull Intensity Function Model Estimators, Model-Based Standardization. 109
5.8 Sampling Distribution of the Weibull Intensity Function Model Estimators, Robust Standardization. 110
5.9 Sampling Distribution of the Semiparametric Model Estimators, Model-Based Standardization. 111
5.10 Sampling Distribution of the Semiparametric Model Estimators, Robust Standardization. 112
5.11 Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Weibull Intensity Model, Model-Based Standardization. 113
5.12 Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Weibull Intensity Model, Robust Standardization. 114
5.13 Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Semiparametric Model, Model-Based Standardization. 115
5.14 Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Semiparametric Model, Robust Standardization. 116
5.15 Effect of Baseline Intensity Parameter α on the Sampling Distribution of Test Statistics for the Weibull Intensity Model for Groups of Size 24. 123
5.16 Effect of Group Size on the Sampling Distribution of Test Statistics for the Weibull Intensity Model for $\alpha = 0.7$. 125

Chapter 1

Introduction

Many statistical experiments are designed to study the recurrence rate of non-fatal events. The data from these experiments, which we will refer to as recurrent event data, can be viewed as a more general form of survival data. In a classical survival study, the event of interest is fatal; that is, each subject can experience at most a single event. However, in studies of recurrent events, the event is non-fatal and each subject can experience multiple events. Whereas models for survival data depend on the specification of a hazard function, i.e., the instantaneous probability of death, models for recurrent events depend on the intensity function, the instantaneous probability of a recurrence. However, while there are several well developed methods for the analysis of survival data, methods for recurrent event data are still under development. This thesis examines some models and methods for the analysis of recurrent event data.

Like survival data, recurrent event data can arise in many contexts: medical, social, mechanical, etc. (Abu-Libdeh, Turnbull, and Clark 1990; Bacchetti 1990; Lawless 1995). However, for simplicity, the terminology of the biological context has been adopted here. For example, the word subject is used here in a general sense, and could refer as easily to automobile brakes or computer operating systems as to patients in a clinical trial.

Recurrent event data can be collected to investigate a variety of questions, such as: Does the recurrence rate appear to change over time? Are there characteristics associated with the subjects which affect the frequency of recurrence? Is there more variability than expected among the subjects?

This thesis proposes new methodology for the analysis of recurrent event data. The methodology is quite general in that it can handle many of the complications which can

arise in this type of analysis. For example, the methods are specifically designed to handle overdispersion, as well as unequal follow-up periods due to non-informative left or right censoring. The methods can be adapted to cope with discontinuous follow-up and time-dependent covariates that are independent of the event-history.

One of these potential complications, overdispersion, receives particular attention in the thesis. Overdispersion is commonly observed in the analysis of count data (*c.f.* Cox 1983)., and it is natural to anticipate its appearance in recurrent event data as well. In recurrent event data, overdispersion results in more subject-to-subject variability than would be reasonable under the proposed model. The inclusion of a subject-specific random effect is a natural method to model this excess variation. However, likelihood-based inference would then require the specification of a distribution for the random effects. An alternative is the use of quasi-likelihood estimation, which requires only low order moment assumptions. The efficiency of quasi-likelihood inference is examined for the models developed in later chapters. We also compare robust and model-based variance estimators.

In Chapter 2 we attempt to locate the work in this thesis within the broad context of the analysis of longitudinal data. We begin by examining the general structure of recurrent event data and describing three illustrative studies of recurrent events. Next, we survey some of the models and estimation methods currently used for the analysis of this type of data.

In some studies of recurrent events, it is possible to record the exact event-times for each recurrence. In other studies, the event-times are not available, perhaps because the examination process is too expensive or invasive, or because the events occur too rapidly to record the exact times. In these circumstances, we could record event counts instead, choosing either one count per subject, recorded at the end of the follow-up period, or a series of counts for each subject, collected at intermediate follow-up times. The relative efficiency these two forms of count data is examined in detail in Chapters 3 and 4.

Chapter 3 describes a nonhomogeneous Poisson process model for recurrent event data. A proportional intensity function is adopted for incorporating the effects of covariates, and a random subject-specific effect is introduced to account for overdispersion. We examine the relative efficiency of studies utilizing event-time data versus end-of-follow-up count data, and establish that count data is highly efficient for the estimation of covariate effects under very reasonable design constraints. However, count data is shown to have low efficiency for parameters associated with the underlying intensity function. Quasi-likelihood estimators,

which require only low order moment assumptions, are proposed as a more robust alternative to likelihood estimators. The efficiency and robustness qualities of the quasi-likelihood estimators for the analysis of "simple" count data (e.g., McCullagh and Nelder 1989; Breslow 1990) make them a reasonable choice for these models. Asymptotic results are illustrated with a numerical study, and a bladder cancer recurrence data set is used to illustrate the techniques.

Chapter 4 develops methods for the analysis of recurrent event data in the form of counts collected at multiple follow-up times per subject. This is sometimes referred to as panel data. Panel data are examined as a means of recovering much of the information that was lost by the use of end-of-follow-up count data in Chapter 3. Like end-of-follow-up count data, panel data is shown to be highly efficient for the estimation of covariate effects. However, where end-of-follow-up count data was not efficient for the estimation of the overall intensity, panel data is shown to be reasonably efficient with as few as two or three follow-up-times in the numerical studies we conducted. This chapter also describes a quasi-score test for overdispersion and diagnostics for the appropriateness of the intensity and variance models used. A small numerical study examines the asymptotic characteristics of the procedures and a simulation study examines their small-sample behaviour, including the performance of the model-based and robust variance estimators. The bladder cancer data set examined in Chapter 3 is revisited and a panel data analysis is compared to analyses based on event-times and end-of-follow-up counts.

In Chapter 5, a nonparametric approach to the modeling of the intensity is described. This is a particularly exciting development as it permits removal of the parametric assumptions associated with the intensity function. Combined with the quasi-likelihood methods of Chapter 4, which eliminate the parametric assumptions for the subject-specific random effect, the result is a reasonably robust semiparametric model for recurrent events collected as panel data. Score-type tests for the fit of parametric intensity models are developed. The asymptotic and small-sample behaviour of the estimators and tests are examined through numerical and simulation studies, and the techniques are illustrated by a reanalysis of the bladder cancer data. The analysis of a second data set, from an experiment examining a new method for the control of the Cherry Bark Tortrix moth, highlights the usefulness of the semiparametric approach.

The thesis concludes with a brief summary and discussion of several areas which will be pursued in later work. For example, one topic for future research is the extension of

semiparametric methods of Chapter 5, including development of smoothing techniques for the nonparametric intensity estimator. Another area of interest is the prediction of future events based on current event history. For some applications, there is less interest in testing of covariate effects and more in the prediction of final event-counts based on intermediate information. Generalizations of the variance structures used in this thesis also provide an opportunity for future work, and perhaps connect with generalized linear mixed models and generalized estimating equations methods. Another direction to pursue is the development of methods for recurrent events which are not point events — i.e., for events which have positive duration. Applications of such techniques may include many disease processes, such as Multiple Sclerosis. Here the event of interest is a prolonged period of exacerbation of symptoms. For example, with M.S., an exacerbation can last for more than six weeks. It should be possible to modify the methods in this thesis to address a two-state model for such processes.

Chapter 2

Background

2.1 The Structure of Recurrent Event Data

We shall consider recurrent event data arising where a reasonably large number of subjects is monitored for the recurrence of a non-fatal event over a reasonably short period of time. Let (T_{i0}, T_{ie}) denote the *follow-up period* for subject i, $i = 1, 2, \ldots, M$. Then recurrent event data can be recorded in several ways. *Event-time data* records the exact event times for each subject. *Panel data* records the number of recurrences between the intermediate follow-up times $T_{i1} < T_{i2} < \ldots < T_{is_i}$ where all $T_{ij} \in (T_{i0}, T_{ie}]$. End-of-follow-up count data, referred to here as *count data* or where ambiguous, *end-of-follow-up count data*, records the total number of events occurring during the follow-up period.

In many studies, each subject has a unique start-of-follow-up time T_{i0}. However, in this thesis we will assume that the time index can be redefined so that $T_{i0} = 0$ for all subjects, $i = 1, \ldots, M$. We also assume that the intermediate follow-up times are the same for all subjects, i.e., $T_{ij} = T_j$ for all $i = 1, \ldots M$, and $j = 1, \ldots, s$. These assumptions reflect the reality of many recurrent event studies. For example, many clinical trials consider the application of the treatment to occur at time $t = 0$ and each subject is then examined according to the same schedule. However, we do not assume common end-of-follow-up times, and in this way accommodate unequal follow-up periods due to subjects being lost to follow-up.

Many studies of recurrent events focus on determining the effects of covariates. Covariates can be either fixed or time-dependent; however, only fixed covariates are considered in this thesis. While it is possible to include time-dependent covariates, care must be taken

when covariate changes are associated with the history of the process (*cf.* Kalbfleisch and Prentice 1980).

A typical event-time recurrent event dataset is displayed in Figure 2.1. Here we show the follow-up period for each subject and the event-times and the subjects are grouped according to treatment. The data are described in Section 2.2.2, but at present it is sufficient to note the characteristics described above: multiple events per subject, common start-of-follow-up times, unequal follow-up periods and the presence of a covariate (treatment).

2.2 Illustrations

In this section we describe three illustrations which demonstrate the broad applicability of the techniques developed in the thesis.

2.2.1 Air Conditioner Data

Proschan (1963) provides data consisting of inter-event times, measured in hours of operation between successive failures of the air conditioning equipment aboard a fleet of 13 Boeing 720 aircraft. The data have been discussed by a number of authors. As there are no recorded covariates, the focus of analyses for this data is on describing the pattern of recurrent failures, detecting time trends, and determining if there is variability between the aircraft. An important question for this data set is whether there is evidence of a decreasing failure rate (i.e., a decreasing intensity function).

2.2.2 Bladder Cancer Data

In the early 1970s, the Veterans' Administration Co-Operative Urological Research Group conducted a randomized clinical trial investigating the effectiveness of three treatments on the frequency of bladder cancer recurrence. Each of the 118 subjects had superficial bladder tumours when they entered the trial. The tumours were removed and the subjects were assigned to one of three treatment arms: placebo pills, pyridoxine (Vitamin B_6) pills, or periodic instillation of thiotepa into the bladder. Each patient was examined at frequently repeated follow-up visits; any tumours found were removed and the treatment was continued. The data record the times of each recurrence and the number and size of any tumours found. These data were originally examined in Byar et al. (1977). They have been analyzed in

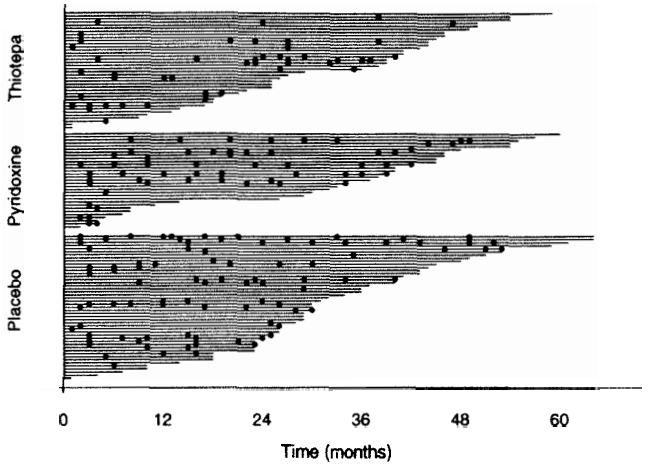

Figure 2.1: An Example of Recurrent Event Data — Bladder Cancer Data. The horizontal lines represent the follow-up period for each subject and the points mark recurrence times. Figure adapted from Byar et al. (1977, Figure 1).

a number of other articles, including Byar (1980) and Davis and Wei (1988). A simple graphical summary of the event-times is given in Figure 2.1.

2.2.3 Cherry Bark Tortrix Data

The Cherry Bark Tortrix moth (*Enarmonia formosana*) infests ornamental cherry trees on boulevards throughout the Lower Mainland. Because the trees are located in residential neighbourhoods, insecticide controls were deemed inappropriate. An experiment was designed to test an alternative pest-control strategy which would disrupt mating by disorienting mate-seeking males and preventing them from finding females. This scent-based mating disruption strategy used pheromone dispensers to emit a cloud of artificial female pheromone sufficiently dense to disorient any mate-seeking male. The artificial pheromone chosen for the experiment was known to be competitive with virgin female moths at attracting males to the traps. For the experiment, twenty cherry trees were chosen along a residential street in New Westminster, and a chemical dispenser was installed on each tree. Half of the dispensers, randomly chosen, were filled with the pheromone; the remaining ten trees were used as a control group. Traps were placed in similar locations in each tree and

the bait in each trap was impregnated with the pheromone. Over the course of the summer of 1995, the trees were visited 19 times, approximately once per week, and the number of male moths per trap was recorded. Once every three weeks, the baits were refreshed.

These data were originally examined in McNair, Gries, and Gries (1997) where the strong treatment effect was described. The data are re-analyzed in Chapter 5 as an example of an experiment where the semiparametric models we present are particularly suitable.

2.3 Models for Recurrent Event Data

Methodology for recurrent event data has been examined in many references, including Cox and Isham (1980) and Ross (1983). In this section, we review the models which are used in later chapters. We begin with the the nonhomogeneous Poisson process (NHPP) model as a basic framework for the analysis of recurrent events and then incorporate covariates through the use of the proportional intensity model. Subject-to-subject heterogeneity is accommodated through frailty models.

The models are described below for the event-time data. End-of-follow-up count and panel data models are described in Chapters 3, 4, and 5.

2.4 Poisson Process Models

A Poisson process is a stochastic process associated with events occurring randomly in time. Let $N(a,b)$ be the number of events occurring during the interval (a,b), let $N(t)$ correspond to $N(0,t)$ and let $\lambda(t)$ be a left continuous function such that

$$\int_0^t \lambda(u)\,du = \Lambda(t) < \infty \quad \text{for all } t > 0.$$

Then, $\{N(t)\}_{t=0}^{\infty}$ is a Poisson process with intensity function $\lambda(t)$ and cumulative intensity $\Lambda(t)$ if and only if

1. $N(0) = 0$;
2. $\Pr\{N(t) - N(t-h) = 1 | \mathcal{N}(t-h)\} = \lambda(t)h + o(h)$
3. $\Pr\{N(t) - N(t-h) > 1 | \mathcal{N}(t-h)\} = o(h)$

for small h and all $t > 0$. The history of the process, $\mathcal{N}(t)$, is the record of all events to time t, $\mathcal{N}(t) = \{N(u) : 0 \leq u \leq t\}$, and $o(h)$ represents a quantity that decreases to 0 more rapidly than h as h tends to 0, i.e., $\lim_{h \to 0} h^{-1} o(h) = 0$. If the intensity function is constant with respect to time, $\lambda(t) = \lambda$, then the Poisson process is a homogeneous Poisson process. Otherwise, the process is nonhomogeneous in time. In other words, a nonhomogeneous Poisson process (NHPP) has an intensity function that is time dependent; a homogeneous Poisson process (HPP) has constant intensity.

The Poisson process has a number of useful characteristics:

- The process is memoryless. Events in non-overlapping time intervals are completely independent.
- If (a, b) and (c, d) are non-overlapping intervals, then $N(a, b)$ and $N(c, d)$ are independent Poisson random variables with means $\int_a^b \lambda(u)\ du$ and $\int_c^d \lambda(u)\ du$.
- If the process is homogeneous, then the inter-event times are independent exponential random variables with rate λ, or mean $1/\lambda$.
- For $t > 0$, $N(t)$ is a Poisson random variable with mean $\Lambda(t)$. The cumulative intensity function $\Lambda(t)$ is sometimes referred to as the cumulative mean function for this reason.

Two distributional results that will be useful later are stated here. First, we observe that conditional on the number of events during the period $(0, T_E]$, the distribution of event times within that period is that of order statistics. That is, given $N(T_E) = n$, the probability density function (p.d.f.) of the event-times $T_1, T_2, \ldots, T_n$ is given by

$$f(t_1, t_2, \ldots, t_n | n) = n! \prod_{j=1}^{n} \frac{\lambda(t_j)}{\Lambda(T_E)^n} \quad \text{for } 0 < t_1 < t_2 < \ldots < t_n \leq T_E,$$

which is the distribution of the order statistics from a sample of size n from a distribution with density $\lambda(t)/\Lambda(T_E)$, $0 \leq t \leq T_E$. Second, the likelihood based on the n event times $0 \leq t_1 < t_2 < \ldots < t_n \leq T_E$ occurring during the fixed interval $[0, T_E]$ is

$$\begin{aligned} \mathcal{L} &= \left[\prod_{j=1}^{n} \lambda(t_j) \exp\left(- \int_{t_{j-1}}^{t_j} \lambda(u)\ du \right) \right] \exp\left(- \int_{t_n}^{T_E} \lambda(u)\, du \right) \\ &= \left(\prod_{j=1}^{n} \lambda(t_j) \right) \exp\left(- \int_0^{T_E} \lambda(u)\, du \right) \end{aligned}$$

References for the Poisson Process are numerous. See for example, Cox and Isham (1980),Karr (1986),Lewis (1972),and Snyder (1975). Cox and Lewis (1966, Chapter 2) discuss methods based on counts of the number of events in fixed time intervals. Ross (1983, Chapter 2) and Lawless (1982, Chapter 10) provide general overviews of the Poisson process.

2.5 Proportional Intensity Models

In a study of recurrent events with covariates, the assumption of a proportional intensity model means that any two subjects with covariate values $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have intensity functions that are proportional to one another. That is, the ratio $\lambda(t|\boldsymbol{x}_1)/\lambda(t|\boldsymbol{x}_2)$ does not depend on time. This implies that the intensity function can be written $\lambda(t|\boldsymbol{x}) = \lambda_0(t)g(\boldsymbol{x})$, where $\lambda_0(\boldsymbol{x})$ is called the baseline intensity and corresponds to a subject with $g(\boldsymbol{x}) = 1$. One very common and useful version of this model has $g(\boldsymbol{x}) = \exp(\boldsymbol{x}'\boldsymbol{\beta})$. In this model the covariates have a multiplicative effect on the intensity function. This is the same method for the inclusion of covariate effects as used in proportional hazards survival analysis models. The proportional intensity model is discussed in many standard references, including Kalbfleisch and Prentice (1980) and Lawless (1982).

2.6 Frailty Models for Heterogeneity

Heterogeneity, a phenomenon frequently encountered in the analysis of count data, occurs when the variance exhibited by the observations is larger than that predicted by the model; for example, when the data are modeled as Poisson with mean $\hat{\Lambda(t)}$ and the sample variance of the data is found to be much larger than $\hat{\Lambda}(t)$. Cox (1982) describes the general phenomenon and points out that analyses conducted which ignore overdispersion will underestimate the standard errors of parameters and result in tests with inflated type I error rates.

In an analysis of recurrent events, one method to account for overdispersion is to allow each subject to have a unique Poisson intensity, i.e., $\lambda_i(t) = \nu_i\lambda(t)$. However, this introduces a new parameter ν_i for each subject. An alternative is to treat these parameters as unobservable realizations from a mixing distribution $p(\nu)$. In this case, the conditional distribution of the events is given by the intensity function $\lambda_i(t)$; that is, given the value ν_i

the events follow a Poisson process.

Based on the event-times for subject i, whose n_i observations occured at times t_{ij}, where $0 < t_{i1} < t_{i2} < \ldots < t_{in_i} \leq T_{ie}$, the likelihood is

$$\begin{aligned} \mathcal{L}_i &= \int_0^\infty \left(\prod_{l=1}^{n_i} \lambda(t_{il}; \nu_i, \boldsymbol{x}_i) \right) \exp\left[-\Lambda(T_{ie}; \nu_i, \boldsymbol{x}_i) \right] \, d\nu_i \\ &= \left\{ \prod_{l=1}^{n_i} \frac{\lambda_0(t_{il})}{\Lambda_0(t_{ie})} \right\} \int_0^\infty \left[\nu_i \Lambda_0(T_{ie}) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \right]^{n_i} \exp\left[-\nu_i \Lambda_0(T_{ie}) \right] \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) p(\nu_i) \, d\nu_i. \end{aligned}$$

For example, if the conditional distribution of y_i given ν_i is Poisson with mean μ_i, and the subject-specific effects ν_i are *i.i.d.* gamma random variables with mean 1 and variance τ, the marginal distribution of y_i is negative binomial with mean μ_i and variance $\mu_i(1 + \tau\mu_i)$. This scenario is described fully in Lawless (1987a).

2.7 Other Stochastic Process Models

There are many other models for stochastic processes, but two at least deserve mention here because of their potential for describing the behaviour of recurrent events. The first is the renewal process, of which the Poisson process is a special case. The renewal process is a natural candidate model for processes where the inter-event times are thought to be *i.i.d.* The second is the discrete state Markov Process. Also known as a Markov chain in continuous time, these models may be appropriate for processes which pass from state-to-state and where interest centers on the sojourn times (i.e., the times spent between shifts from state-to-state). This is a rich class of models with broad applicability, but is described only briefly here as part of the overall context of the models adopted in the thesis.

2.7.1 Renewal Process Models

A homogeneous Poisson process is a counting process with independent exponentially distributed inter-event times. A generalization of this is the renewal process, a counting process with independent and identically distributed inter-event times from an arbitrary distribution. In renewal models, interest naturally centers on the inter-event times, denoted $\{Z_i\}_{i=1}^\infty$, all $Z_i > 0$. An example which could be modeled as a renewal process would be the lifetimes of light bulbs replaced immediately upon failure, assuming that the light bulb lifetimes are

i.i.d. A delayed renewal process is a counting process where the first event time has a different distribution than the remaining events. For example, computer components may have a different lifetime distribution when first installed and after being replaced for the first time.

One useful characteristic of the renewal process is that the intensity function for the event process depends only on the time since the most recent event. That is,

$$\begin{aligned}\lambda(t;\mathcal{N}(t)) &= \lim_{h\to 0}\frac{1}{h}\Pr[N(t)-N(t-h)=1|\mathcal{N}(t-h)] \\ &= \lambda_0(t-T_{N(t)})\end{aligned}$$

where $T_{N(t)}$ is the time of the latest event and $\lambda_0(z) = f(z)/[1-F(z)]$ is the hazard function for the inter-event distribution with p.d.f. $f(z)$.

2.7.2 Discrete State Markov Process Models

One form of the Markov process which is useful for modeling recurrent events is the continuous-time Markov chain, a discrete state process in continuous time. Let $Y(t)$ represent the state a process is in at time t. The states are labeled 0, 1, 2, ...; the probability that the process will go from state i to state j during the interval $(s, s+t)$ is represented as $P_{ij}(s, s+t) = \Pr[Y(t+s) = j|Y(s) = i]$, where $s \geq 0$ and $t > 0$. If the P_{ij} do not depend on s, then the process is homogeneous in time. The transition probabilities must satisfy the Markov assumption (Feller 1968, Section 17.9). A Markov process has the property that the conditional distribution of $Y(t)$, given the event history of $Y(t)$, depends only on the most recent value of $Y(t)$. Markov processes display what could be called first-order dependence. Markov processes have been used to model changes in a subject's behaviour or disease state (e.g., Kalbfleisch and Lawless 1985; Bartholomew 1983).

A semi-Markov process is one that passes among states $1, 2, \ldots$ according to a Markov chain with transition probability matrix $\mathbf{P} = (p_{ij})$; p_{ij} is the probability that the system will enter state j next from state i. Also, the sojourn time, the time spent in state i before passing to state j, is a random variable with a p.d.f. $f_{ij}(t)$.

2.8 Estimation

This section reviews estimation methods used in later chapters. First is likelihood estimation, where by specifying a full distributional model, estimates are found that maximize

the likelihood for that model given the observed data. Quasi-likelihood and other moment methods find estimates as the roots of estimating equations.

2.8.1 Likelihood Estimation

Adopting the NHPP model with proportional intensity function for incorporating covariates and the frailty model for accommodating overdispersion, the likelihood function for the event-time data is

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{M} \left\{ \prod_{j=1}^{n_i} \frac{\lambda_0(t_{ij})}{\Lambda_0(t_{iE})} \right\} \int_0^\infty \left[\nu_i \Lambda_0(T_{iE}) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})\right]^{n_i} \exp\left[-\nu_i \Lambda_0(T_i)\right] \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) p(\nu_i)\, d\nu_i.$$

The likelihoods for the end-of-follow-up count data and panel data are given in Chapters 3 and 4. Likelihood inference for event-time data is explored extensively in Lawless (1987b) and Thall (1988), with additional results given by Lawless (1987a) for the mixed-Poisson model with subject-specific effects distributed as gamma random variables.

Strategies for finding the maximum likelihood estimates include the usual iterative procedures (quasi-Newton or Fisher Scoring algorithms). Models which include a random subject-specific effect are most easily handled by maximizing the profile likelihood for the overdispersion parameter τ, defined as in Section 2.6. Negative values of $\hat{\tau}$ are possible, but for the random-effects model proposed have no sensible interpretation and are thus set to zero.

2.8.2 Quasi-Likelihood Estimation and Other Moment Methods

Recently, moment-based estimation methods have become popular for the analysis of "simple" count data, (e.g., McCullagh and Nelder 1989, Chapter 9). The use of only first and second moment assumptions to construct estimating functions in the context of generalized linear models for count data is attributed to Wedderburn (1974). We describe the use of these techniques for the analysis of event-time, panel, and count data in later chapters.

The usual quasi-likelihood estimating functions can be used for the intensity function parameters (McCullagh and Nelder 1989). Several estimating functions have been suggested for τ, including

$$\sum_{i=1}^{M} \frac{(n_i - \mu_i)^2}{\mu_i(1 + \tau\mu_i)} - (M - p) \tag{2.1}$$

where n_i is the count of the number of events for subject i and has expectation $\mu_i = \mu_i(\theta)$, M is the number of subjects and p is the dimension of the intensity function parameter vector (Breslow 1984). The resulting equation sets the Pearson statistic equal to its degrees of freedom. Using the deviance statistic as the first term of this function yields a similar estimator (McCullagh and Nelder 1989). Davidian and Carroll (1987) suggest a similar estimating function for τ,

$$\sum_{i=1}^{M} \frac{(n_i - \mu_i)^2 - \mu_i(1 + \tau\mu_i)}{(1 + \tau\mu_i)^2}.$$

The function compares the observed sample variance and the model-based (in this case, negative binomial) variance and effectively corresponds to the likelihood estimating function were the subject-specific random-effects to be from a normal distribution. We adopt this estimating equation for the quasi-likelihood approach for reasons discussed later, and consider the efficiency and bias of the resulting estimators in Chapter 3. Lawless (1987a) also discusses efficiency and robustness of the estimators based on (2.1).

2.8.3 Robust Variance Estimation

For inference based on quasi-likelihood, let $\boldsymbol{g}$ be an unbiased estimating function for the parameter $\boldsymbol{\theta}$. That is, assume that $\mathrm{E}(\boldsymbol{g}) = \mathbf{0}$. Then, the quasi-likelihood estimates, $\tilde{\boldsymbol{\theta}}$, are obtained as the roots of $\boldsymbol{g}(\boldsymbol{\theta})$. The results of Huber (1967), Inagaki (1973), and White (1982) provide that under suitable regularity conditions, similar to those for maximum likelihood asymptotic theory, $\sqrt{M}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal, with mean 0 and asymptotic variance given by

$$\mathrm{E}\left(-\lim_{M\to\infty}\frac{1}{M}\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\theta}^T}\right)^{-1}\mathrm{E}\left(\lim_{M\to\infty}\frac{1}{M}\boldsymbol{g}\boldsymbol{g}^T\right)\mathrm{E}\left(-\lim_{M\to\infty}\frac{1}{M}\frac{\partial \boldsymbol{g}^T}{\partial \boldsymbol{\theta}}\right)^{-1} \tag{2.2}$$

evaluated at $\tilde{\boldsymbol{\theta}}$. For the quasi-likelihood results, the regularity conditions require the matrix given by (2.2) to be positive definite. A robust estimator of the variance of $\tilde{\boldsymbol{\theta}}$ is,

$$\mathrm{E}\left(-\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\theta}^T}\right)^{-1}\left(\boldsymbol{g}\boldsymbol{g}^T\right)\mathrm{E}\left(-\frac{\partial \boldsymbol{g}^T}{\partial \boldsymbol{\theta}}\right)^{-1}$$

This is the so-called "sandwich" variance estimator (e.g., Liang and Zeger 1986; Zeger et al. 1988). The robust variance estimator has been used extensively, for example in the papers Lawless (1987b), Lawless and Nadeau (1995), Moore and Tsiatis (1991).

In the following chapters, we will examine the likelihood and proposed quasi-likelihood estimators for a nonhomogeneous Poisson process model. Covariate effects will be incorporated via a proportional intensity model, and overdispersion will be accounted for via a subject-specific random effect. Chapter 3 compares the efficiency of analyses based on end-of-follow-up count data to those based on event-time data and Chapter 4 examines the efficiency of analyses of panel data. Chapter 5 describes a nonparametric intensity model and examines the efficiency of analyses based on this model for panel data.

Chapter 3

Analysis of Event-Time Data vs. End-of-Follow-Up Counts

3.1 Introduction

Inference for point processes is most efficient when the exact times of occurrence of events are available. Sometimes, however, the event-times are unavailable and only the number of events that have occurred in a fixed interval of time is recorded. For example, this type of data arises in medical studies where an event is recognized only when a subject is examined, or when events occur too frequently for exact times of occurrence to be noticed (*cf.* Stukel 1993).

In this chapter, we consider a study where every subject in the study gives rise to a non-homogeneous Poisson process. The aim of the study is to compare processes for individuals with different covariate values, e.g., undergoing different treatments. We use the proportional intensity function for the inclusion of covariate effects. The proportional intensity model has been discussed by many authors, and surveys of related articles are provided in Andersen and Gill (1982) and Lawless (1987b). Its usage is common in part because it is a tractable and flexible model. Subject-to-subject heterogeneity not explained by the regression model is incorporated via random effects models and estimators are found using quasi-likelihood estimating functions.

There is some loss of efficiency in the estimation of treatment effects when analyzing only the end-of-follow-up counts instead of the event-times. This chapter derives this loss for

non-homogeneous Poisson processes in Section 3.2, and for Poisson processes with random effects (*cf.* Lawless 1987b) in Section 3.4. The random effects models allow for extra-Poisson variation in the distributions of the counts. Quasi-likelihood inference for these models is also discussed in Section 3.4. The quasi-likelihood analysis proposed here is an extension of such techniques for the homogeneous Poisson process. The methods result in estimating equations for the parameters in the random effects model that are simple and intuitive. Section 3.6 considers an example concerning the recurrences of tumors in patients with bladder cancer. Our results show that for many usual situations, covariate effects can be estimated with high efficiency based on end-of-follow-up counts, though this is not the case for the parameters of the underlying intensity function of the process. The chapter closes with a brief discussion of the results.

3.2 Efficiency comparisons for the non-homogeneous Poisson process

Consider a comparison of k treatments, where m_r individuals are given treatment r, and let $M = \sum_{r=1}^{k} m_r$ denote the total sample size of the study. Each individual is monitored for events, and observations on the i-th of the M individuals consists of $Y_i(t)$, the number of events observed up to time t, $t \in (0, T_{ie}]$, so individual i is followed for a time T_{ie}. We assume that T_{ie} is independent of the counting process $Y_i(t)$. Inference is possible in the absence of such independence, but would require further assumptions regarding the joint distribution of the end-of-follow-up times, T_{ie}, and the counting process. Note that this would permit more general withdrawal schemes such as censoring at the r-th failure, as is common in reliability studies (*cf.* Lawless and Nadeau 1995).

Define the end-of-follow-up counts to be $n_i = Y_i(T_{ie})$, and let $\{\omega_{il}, l = 1, \ldots, n_i\}$ be the event-times for the i-th individual. Let $\boldsymbol{x}_i$ be a $k \times 1$ covariate vector for the i-th individual with $x_{i1} = 1$, and

$$x_{ir} = \begin{cases} 1 & \text{if individual } i \text{ received treatment } r, \\ 0 & \text{otherwise} \end{cases}$$

$r = 2, \ldots, k$. The counting process $Y_i(t)$ is modeled as a Poisson process with rate function

$$\lambda_i(t) = \lambda_0(t; \boldsymbol{\alpha}) e^{\boldsymbol{x}_i' \boldsymbol{\beta}}$$

where λ_0 is a twice differentiable baseline intensity function, depending on $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ is a vector of regression parameters. We have parameterized the treatment effects such that treatment 1 is the "reference treatment." That is, β_1 determines the level of the intensity function $\lambda_i(t)$ for treatment group 1 and β_r measures the effect of treatment r relative to treatment 1 ($r = 2, \ldots, k$). The parameter $\boldsymbol{\alpha}$ determines the overall "shape" of the intensity function $\lambda_0(t;\boldsymbol{\alpha})$. If a constant intensity function were assumed, $\lambda_0(t;\boldsymbol{\alpha})$ would be identically one and $\boldsymbol{\alpha}$ would not be required. Writing $\Lambda_0(T_{ie};\boldsymbol{\alpha}) = \int_0^{T_{ie}} \lambda_0(u;\boldsymbol{\alpha})\, du$ for the cumulative baseline intensity function, then $\mu_i = E(n_i) = e^{\boldsymbol{x}_i'\boldsymbol{\beta}}\Lambda_0(T_{ie})$.

The likelihood function for $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$, based on the event-time data, is

$$L_t(\boldsymbol{\gamma}) = \exp\left(\sum_{i=1}^{M} n_i \boldsymbol{x}_i'\boldsymbol{\beta}\right)\left\{\prod_{i=1}^{M}\prod_{l=1}^{n_i}\lambda_0(\omega_{il};\boldsymbol{\alpha})\right\}\exp\left(-\sum_{i=1}^{M}\mu_i\right). \tag{3.1}$$

Based on the count data, consisting only of the $n_i, i = 1, \ldots M$, the likelihood is the simple Poisson kernel:

$$\begin{aligned} L_c(\boldsymbol{\gamma}) &= \prod_{i=1}^{M} \mu_i^{n_i} e^{-\mu_i} \\ &= \exp\left(\sum_{i=1}^{M} n_i \boldsymbol{x}_i'\boldsymbol{\beta}\right)\left\{\prod_{i=1}^{M}\Lambda_0(T_{ie};\boldsymbol{\alpha})^{n_i}\right\}\exp\left(-\sum_{i=1}^{M}\mu_i\right). \end{aligned} \tag{3.2}$$

As pointed out by Lawless (1987b), there is a factorization of the event-time data likelihood (3.1) of the form

$$L_t(\boldsymbol{\gamma}) = L_\alpha(\boldsymbol{\alpha})L_c(\boldsymbol{\gamma}), \tag{3.3}$$

where

$$L_\alpha(\boldsymbol{\alpha}) = \prod_{i=1}^{M}\prod_{l=1}^{n_i}\frac{\lambda_0(\omega_{il};\boldsymbol{\alpha})}{\Lambda_0(T_{ie};\boldsymbol{\alpha})}. \tag{3.4}$$

The factorization of $\lambda_i(t)$ into two components, one a function of $\boldsymbol{\alpha}$, the other a function of $\boldsymbol{\beta}$, is the key to the factorization of the likelihood in (3.2) above. This factorization of the likelihood will be exploited in computing the asymptotic efficiency of $\widehat{\boldsymbol{\beta}}_c$, the estimator of $\boldsymbol{\beta}$ derived from L_c, with respect to $\widehat{\boldsymbol{\beta}}_t$, the estimator of $\boldsymbol{\beta}$ derived from L_t. The asymptotic efficiency of the corresponding estimator of $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\alpha}}_c$ from L_c with respect to $\widehat{\boldsymbol{\alpha}}_t$ from L_t, will also be computed. Note that if $\boldsymbol{\alpha}$ were known, the likelihood kernel from (3.1) would equal that from (3.2), and analysis of the counts would be fully efficient for inference concerning $\boldsymbol{\beta}$.

The maximum likelihood estimate of $\boldsymbol{\beta}$, based on either L_c or L_t, satisfies

$$D_\beta' V^{-1}(\boldsymbol{N} - \boldsymbol{\mu}) = \mathbf{0}, \tag{3.5}$$

where $D_\beta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}'$, $V = \text{diag}(\mu_i; i = 1, \ldots, M)$, and $\boldsymbol{N}$ and $\boldsymbol{\mu}$ are vectors of counts and their expected values, respectively. This simplifies to

$$X'(\boldsymbol{N} - \boldsymbol{\mu}) = \mathbf{0}, \tag{3.6}$$

where X is the covariate matrix having ir-th entry x_{ir}.

Let G_r index the set of individuals who received treatment r, $G_r = \{$ all $i \ni$ individual i received treatment $r\}$, $r = 1, \ldots, k$. To simplify notation, let $[a]_{r+}$ denote $\sum_{i \in G_r} a_i$, so $[n]_{1+} = \sum_{i \in G_1} n_i$, the total number of events observed for all individuals receiving treatment 1; let $[ab]_{r+}$ denote $\sum_{i \in G_r} a_i b_i$, etc., and for grand totals, let $[a]_+$ denote $\sum_{i=1}^M a_i$, $[ab]_+$ denote $\sum_{i=1}^M a_i b_i$, etc. Then,

$$e^{\hat{\beta}_{1t}} = \frac{[n]_{1+}}{[\Lambda_0(T; \hat{\boldsymbol{\alpha}}_t)]_{1+}} = \frac{\sum_{i \in G_1} n_i}{\sum_{i \in G_1} \Lambda_0(T_i; \hat{\boldsymbol{\alpha}}_t)}, \tag{3.7}$$

and

$$e^{\hat{\beta}_{rt}} = \frac{[n]_{r+}}{e^{\hat{\beta}_{1t}}[\Lambda_0(T; \hat{\boldsymbol{\alpha}}_t)]_{r+}} = \frac{\sum_{i \in G_r} n_i}{e^{\hat{\beta}_{1t}} \sum_{i \in G_r} \Lambda_0(T_i; \hat{\boldsymbol{\alpha}}_t)}, \tag{3.8}$$

$r = 2, \ldots, k$. Similarly, $e^{\hat{\beta}_{1c}} = [n]_{1+}/[\Lambda_0(T; \hat{\boldsymbol{\alpha}}_c)]_{1+}$, and $e^{\hat{\beta}_{rc}} = [n]_{r+}/(e^{\hat{\beta}_{1c}}[\Lambda_0(T; \hat{\boldsymbol{\alpha}}_c)]_{r+})$. An important result holds when the follow-up times for individuals on treatment 1 are the same as for those on treatment $r \geq 2$, $\{T_{ie}, i \in G_1\} = \{T_{ie}, i \in G_r\}$; Here, $\{T_{ie}, i \in G_r\}$ has length m_r, so replicate times are retained in the set; e.g., while $\{1, 1, 2\} = \{1, 2\}$ in traditional set notation, here $\{1, 1, 2\}$ is not so reduced. When this condition is true, the estimates of β_r from the analyses of event-time and panel data are identical,

$$e^{\hat{\beta}_{rt}} = e^{\hat{\beta}_{rc}} = \frac{[n]_{r+}}{[n]_{1+}},$$

so analysis of the counts is fully efficient for estimation of the treatment effects β_r, $r = 2, \ldots, k$.

Based on L_c, the information matrix is

$$\mathcal{I}_c(\gamma) = \begin{pmatrix} X'VX & X'VZ \\ Z'VX & Z'VZ \end{pmatrix}, \tag{3.9}$$

where V is defined as previously in (3.5), and conceptually here is playing the role of the inverse of the generalized linear model iterated weight matrix for log-linear Poisson regression (McCullagh and Nelder 1989, Sec. 2.5); the matrix Z has ir-th entry

$$Z_{ir} = \frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha_t}, \quad t = 1, \ldots, \dim(\boldsymbol{\alpha}).$$

The information matrix based on the event-time data is

$$I_t(\gamma) = \begin{pmatrix} X'VX & X'VZ \\ Z'VX & Z'VZ + H \end{pmatrix},$$

where

$$H = \sum_{i=1}^{M} \mathrm{E}\left(\sum_{l=1}^{n_i} - \frac{\partial^2 \log[\lambda_0(\omega_{il}; \boldsymbol{\alpha}) / \Lambda_0(T_{ie}; \boldsymbol{\alpha})]}{\partial \boldsymbol{\alpha}' \partial \boldsymbol{\alpha}} \right). \tag{3.10}$$

The following results provide the asymptotic relative efficiencies of the estimators derived from the end-of-follow-up count data, relative to those derived from the event-time data. For simplicity, results are presented for scalar α. The derivation of these results is provided in the following section, including a proof of the corresponding results for vector $\boldsymbol{\alpha}$.

Result 1 *The asymptotic relative efficiency of* $\hat{\alpha}_c$ *relative to* $\hat{\alpha}_t$ *is*

$$\mathrm{ARE}(\hat{\alpha}_c) = 1 - \frac{H}{E + H}, \tag{3.11}$$

where

$$E = \sum_{r=1}^{k} \sum_{i \in G_r} \mu_i \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha} - \frac{\left[\Lambda_0 \frac{\partial \log \Lambda_0}{\partial \alpha} \right]_{r+}}{[\Lambda_0]_{r+}} \right)^2, \tag{3.12}$$

is a corrected sum of squares.

Result 2 *The asymptotic relative efficiency of* $\hat{\beta}_{1c}$ *relative to* $\hat{\beta}_{1t}$ *is*

$$\mathrm{ARE}(\hat{\beta}_{1c}) = 1 - \left\{ \frac{l_1}{[\mu]_{1+}^{-1} E + l_1} \right\} \frac{H}{E + H}, \tag{3.13}$$

where

$$l_1 = \left(\frac{[\partial \Lambda_0 / \partial \alpha]_{1+}}{[\Lambda_0]_{1+}} \right)^2.$$

Result 3 *The asymptotic relative efficiency of* $\hat{\beta}_{rc}$ *relative to* $\hat{\beta}_{rt}$, $r = 2, \ldots, k$, *is*

$$\mathrm{ARE}(\hat{\beta}_{rc}) = 1 - \left\{ \frac{l_r}{([\mu]_{r+}^{-1} + [\mu]_{1+}^{-1})E + l_r} \right\} \frac{H}{E+H}, \tag{3.14}$$

where

$$l_r = \left(\frac{[\partial \Lambda_0/\partial \alpha]_{r+}}{[\Lambda_0]_{r+}} - \frac{[\partial \Lambda_0/\partial \alpha]_{1+}}{[\Lambda_0]_{1+}} \right)^2. \tag{3.15}$$

The first observation is that $\mathrm{ARE}(\hat{\beta}_{rc}) > \mathrm{ARE}(\hat{\alpha}_c)$, $r = 2, \ldots, k$. That is, an analysis of counts rather than event-time data loses more information about the shape of the baseline intensity function than about the differences between the treatments. As many experiments are conducted more to examine treatment differences than the baseline intensity, loss of information about the baseline intensity poses less of a drawback than loss of information about the treatments.

Next, we point out that if H in (3.10) can be written as

$$H = \sum_{i=1}^{M} \mu_i f(T_{ie}; \alpha), \tag{3.16}$$

where $f(T_{ie}; \alpha)$ is a function of the end-of-follow-up times, T_{ie}, and α only, then further observations regarding the AREs can be made. This form for H is possible when $\lambda_0(t; \alpha)$ takes common parametric forms, including the Weibull ($\alpha t^{\alpha-1}$) and exponential ($\exp(\alpha t)$). The Weibull is commonly used as the intensity function of count processes. Bain (1978), Seeber (1989), Lawless (1987b), for example, illustrate and discuss its use.

When (3.16) holds, e^{β_1} can be factored out of both E and H leaving remainders which are not functions of β_1. Using (3.11) and (3.13) we then obtain that $\mathrm{ARE}(\hat{\alpha}_c)$ and $\mathrm{ARE}(\hat{\beta}_{1c})$ do not depend on β_1. If, in addition, $\{T_{ie}, i \in G_1\} = \ldots = \{T_{ie}, i \in G_k\}$ then the βs enter E and H through a factor, $(e^{\beta_1} + \sum_{r=2}^{k} e^{\beta_1+\beta_r})$; in this situation, from (3.11) we have that $\mathrm{ARE}(\hat{\alpha}_c)$ does not depend on $\boldsymbol{\beta}$, and the representation in (3.13) shows that $\mathrm{ARE}(\hat{\beta}_{1c})$ increases with increasing values of $\sum_{r=2}^{k} e^{\beta_r}$.

If the follow-up times for all the individuals in the study are the same, $T_{ie} = T_e$, $\forall i$, then $E = 0$, and $\mathrm{ARE}(\hat{\alpha}_c) = \mathrm{ARE}(\hat{\beta}_{1c}) = 0$. That is, when the T_{ie}s are all equal, the parameters α and β_1 cannot be simultaneously estimated. We need information on the total number of events observed over at least two distinct follow-up times in order to estimate these two parameters. However, the larger the range of the Ts, (i.e. the larger the variability of $\partial \log \Lambda_0/\partial \alpha$), the larger E (3.12), and hence the larger the AREs of $\tilde{\alpha}$, $\hat{\beta}_{1c}, \ldots, \hat{\beta}_{kc}$ will be.

The most important results relate to the ARE($\hat{\beta}_{rc}$), $r = 2, \ldots, k$, and these, indeed, are the quantities of greatest interest. If (3.16) holds, then ARE($\hat{\beta}_{rc}$) does not depend on β_1. More importantly, if $\{T_{ie}, i \in G_r\} = \{T_{ie}, i \in G_1\}$, then $l_r = 0$, so ARE($\hat{\beta}_{rc}$) = 1, and Section 3.3.1 shows that the asymptotic covariance of the estimators of β_r and α equals 0, $r = 2, \ldots, k$. Recall, from (3.2), that in this situation the estimators from the full analysis are identical to those from the aggregated analysis.

There are other conditions which lead to ARE($\hat{\beta}_{rc}$) being unity. The most inclusive, of course, is that l_r be 0. This results when $\{T_{ie}, i \in G_r\} = \{T_{ie}, i \in G_1\}$; however, the cardinality of G_r need not equal that of G_1 in order that $l_r = 0$. When we refer to a *balance in the follow-up times over the treatment groups* we mean conditions leading to $l_r = 0, r = 2, \ldots, k$. For example, if the end-of-follow-up times for individuals under treatment 1 were $\{T_{1e}, \ldots, T_{m_1 e}\}$, and there were $(p \times m_1)$ individuals receiving treatment r whose observation times were $\{T_{1e}, \ldots, T_{m_1 e}\}$, replicated p times, then l_r would equal 0, and ARE($\hat{\beta}_{rc}$) would equal 1. Hence, while the variability of $\partial \log \Lambda_0 / \partial \alpha$, as measured by E, is important for high efficiencies, we also need the weighted means of $\partial \log \Lambda_0 / \partial \alpha$ for treatment groups 1 and r to be similar; l_r, (3.15), measures the squared difference of these weighted means

$$l_r = \left(\frac{\left[\Lambda_0 \frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{r+}}{[\Lambda_0]_{r+}} - \frac{\left[\Lambda_0 \frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{1+}}{[\Lambda_0]_{1+}} \right)^2 .$$

When $l_r \neq 0$ and if (3.16) holds, then the ARE($\hat{\beta}_{rc}$) becomes large as β_r increases in absolute magnitude. To see this notice that when (3.16) holds, $H/(E + H)$ is bounded as $|\beta_r|$ increases, with all other parameters fixed. Since

$$([\mu]_{r+}^{-1} + [\mu]_{1+}^{-1})E = (e^{-\beta_r}[\Lambda_0]_{r+}^{-1} + [\Lambda_0]_{1+}^{-1})(c_1 + c_2 e^{\beta_2} + \cdots + c_j e^{\beta_r} + \cdots + c_k e^{\beta_k}),$$

where the c_rs are functions of the T_{ie}s and α, then as $|\beta_r| \to \infty, ([\mu]_{r+}^{-1} + [\mu]_{1+}^{-1})E \to \infty$ and from (3.14) we have that ARE($\hat{\beta}_{rc}$) increases as $|\beta_r|$ increases. In the case of imbalanced follow-up times, the larger the treatment effect, the more efficiently we can estimate it.

If $\boldsymbol{\alpha}$ is a vector parameter, similar observations to those given in the paragraphs above can be made, and this is discussed further in the next section.

Table 3.1 and Figure 3.1 display asymptotic relative efficiencies for the situation where there are two groups to be compared (i.e., $k = 2$), and with $m_1 = m_2$ individuals per group; the follow-up times for members of each group are 10 - 1's and 10 - 2's intended to mimic

α	ARE($\hat{\beta}_{1c}$)	ARE($\hat{\alpha}_c$)
0.5	0.598	0.028
1.0	0.548	0.096
1.5	0.515	0.173
2.0	0.490	0.235
2.5	0.466	0.277
3.0	0.439	0.299

Table 3.1: Asymptotic Relative Efficiencies of the Estimators $\tilde{\beta}_{1c}$ and $\tilde{\alpha}_c$. The study was conducted with $\beta_2 = 0$, $k = 2$ groups, $m_1 = m_2 = 20$ individuals per groups; the follow-up times are identical for the two groups.

a 2-year study where 1/2 of the members of each of the control and treatment groups are recruited only half-way through the study. The intensity function is modeled as Weibull, $\lambda_0(t;\alpha) = \alpha t^{\alpha-1}$. The ARE($\tilde{\beta}_2$) = 1 here since the follow-up times for the members of the first group are the same as for the second. Because (3.16) holds and the follow-up times are the same in the two groups ARE($\hat{\alpha}_c$) does not depend on β in Table 3.1. The table gives ARE($\tilde{\beta}_1$) and ARE($\hat{\alpha}_c$), for selected α values, $0.5 \leq \alpha \leq 3$ and with $\beta_2 = 0$. These efficiency values are all low, especially for $\tilde{\alpha}$.

Figure 3.1 displays ARE($\hat{\beta}_{1c}$), with the treatment effect β_2 varying between -3 and $+3$ for three cases: (a) $\alpha = 0.5$, (b) $\alpha = 1$, and (c) $\alpha = 2$. The efficiency of $\hat{\beta}_{1c}$ increases as α decreases and β_2 increases. Recall that it does not depend on β_1 here.

Table 3.2 displays asymptotic relative efficiencies corresponding to designs with $k = 3$, $m_1 = m_2 = m_3 = 20$ with the follow-up times for each group being 10 - 1's and 10 - 2's, and for three values of β. The relative efficiencies of $\tilde{\alpha}$ are identical to those obtained in Table 3.1. Recall here that ARE($\hat{\beta}_{2c}$) = ARE($\hat{\beta}_{3c}$) = 1. In Table 3.2, the value of $(1 + e^{\beta_2} + e^{\beta_3})$ increases from column 1, Case (a), to column 3, Case (c), as do the ARE()s.

To examine the effect of lack of balance in follow-up times on the asymptotic relative efficiency of the treatment estimator, we considered several imbalanced designs. We present results for one fairly extreme situation below, as an example, and a summary statement follows. In the example, there are two groups of 20 subjects, with the end-of-follow-up times for the individuals in group 1 being fifteen 1's and five 2's; in group 2, 5 of the subjects were followed up to time 1 and 15 to time 2. The Weibull intensity function is used, $\lambda_0(t;\alpha) = \alpha t^{\alpha-1}$. The study used the same values for α and β_2 as in Table 3.1. For

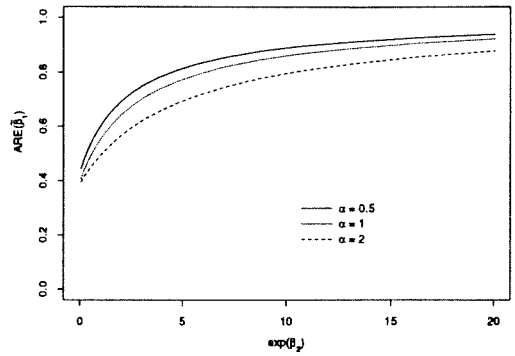

Figure 3.1: Asymptotic Relative Efficiency of $\hat{\beta}_{1c}$. The study was conducted with $-3 \leq \beta_2 \leq 3$, and for $\alpha = 0.5, 1, 2$; $k = 2$ groups, $m_1 = m_2 = 20$ subjects per group; the follow-up times are identical for the two groups.

	Case (a)	Case (b)	Case (c)
α	$\beta_2 = -1, \beta_3 = -2$	$\beta_2 = -1, \beta_3 = 1$	$\beta_2 = -2, \beta_3 = 2$
0.5	0.529	0.750	0.862
1.0	0.484	0.703	0.828
1.5	0.460	0.662	0.794
2.0	0.444	0.622	0.756
2.5	0.429	0.580	0.712
3.0	0.410	0.536	0.661

Table 3.2: Asymptotic Relative Efficiencies of $\hat{\beta}_{1c}$ for Three Cases with Increasing Values of $(1 + e^{\beta_2} + e^{\beta_3})$. For each case, $k = 3$ groups, $m_1 = m_2 = m_3 = 20$ individuals per group; follow-up times identical for the three groups.

$\alpha = 0.5$, the ARE($\hat{\beta}_{2c}$) ranges from 0.76 for $\beta_2 = 0$ to about 0.95 for $\beta_2 = -3$ or 3. For $\alpha = 3$, the ARE($\hat{\beta}_{2c}$) ranges from 0.92 for $\beta_2 = 0$ to about 0.98 for $\beta_2 = -3$ or 3. Using this set of follow-up times, the minimum value of ARE($\hat{\beta}_{2c}$) occurs when $\alpha = \beta_2 = 0$, and is 0.75. In summary, it requires quite extreme imbalance for the ARE($\hat{\beta}_{2c}$) to drop to approximately 75%.

We also considered complete separation in the follow-up times between the groups. This is not a scenario often encountered in practice, and inference concerning treatment effects and other parameters using data arising from such a design would be quite suspect. However, we suppose group 1 has 10 subjects with $T_{ie} = 0.5$ and 10 with $T_{ie} = 1$ while group 2 has 10 subjects with $T_{ie} = 1.5$ and 10 with $T_{ie} = 2$. For β_2 set to zero, the ARE($\hat{\beta}_{2c}$) ranges from 0.265 for $\alpha = 0.5$ to 0.443 for $\alpha = 3$. Thus, it is possible to see quite poor efficiency for the estimation of treatment effects, but it requires complete separation of the sets of follow-up times among the treatment groups — a scenario which might cast doubt on the assumption that the follow-up times are independent of the counting process. We will return to a discussion of censoring in the context of the illustration in Section 3.6.

3.3 Derivation of Efficiency Results in Poisson Process Model

3.3.1 Case 1: Scalar α

The inverse of the information matrix $\mathcal{I}_c$ (3.9) is

$$\mathcal{I}_c^{-1} = \begin{pmatrix} \mathcal{I}_c^{11} & \mathcal{I}_c^{12} \\ \mathcal{I}_c^{21} & \mathcal{I}_c^{22} \end{pmatrix} \tag{3.17}$$

where

$$\begin{aligned} \mathcal{I}_c^{11} &= (X'VX)^{-1} + (X'VX)^{-1}X'VZ\,\mathcal{I}_c^{22}\,Z'VX(X'VX)^{-1}, \\ \mathcal{I}_c^{12} &= (\mathcal{I}_c^{21})' = -(X'VX)^{-1}X'VZ\,\mathcal{I}_c^{22}, \\ \mathcal{I}_c^{22} &= E^{-1} = (Z'VZ - Z'VX(X'VX)^{-1}X'VZ)^{-1}. \end{aligned}$$

Here,

$$(X'VX)^{-1} = \begin{pmatrix} [\mu]_{1+}^{-1} & -\mathbf{1}'[\mu]_{1+}^{-1} \\ -\mathbf{1}[\mu]_{1+}^{-1} & \mathbf{1}\mathbf{1}'[\mu]_{1+}^{-1} + \mathrm{diag}([\mu]_{r+}^{-1}; r = 2, \ldots, k) \end{pmatrix},$$

where $\mathbf{1}$ is a $(k-1)\times 1$ vector of ones. For α a scalar quantity,

$$(X'VX)^{-1}X'VZ = \begin{pmatrix} \frac{\left[\mu\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{1+}}{[\mu]_{1+}} \\ \frac{\left[\mu\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{2+}}{[\mu]_{2+}} - \frac{\left[\mu\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{1+}}{[\mu]_{1+}} \\ \vdots \\ \frac{\left[\mu\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{k+}}{[\mu]_{k+}} - \frac{\left[\mu\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{1+}}{[\mu]_{1+}} \end{pmatrix} \tag{3.18}$$

$$= \left(l_1^{1/2}, l_2^{1/2}, \ldots, l_k^{1/2}\right)', \tag{3.19}$$

where

$$l_1 = \left(\frac{\left[\frac{\partial \Lambda_0}{\partial \alpha}\right]_{1+}}{[\Lambda_0]_{1+}}\right)^2,$$

and

$$l_r = \left(\frac{\left[\frac{\partial \Lambda_0}{\partial \alpha}\right]_{r+}}{[\Lambda_0]_{r+}} - \frac{\left[\frac{\partial \Lambda_0}{\partial \alpha}\right]_{1+}}{[\Lambda_0]_{1+}}\right)^2, \quad r = 2, \ldots, k,$$

as in Section 3.2. Also,

$$E = \left[\mu\left(\frac{\partial \log \Lambda_0}{\partial \alpha}\right)^2\right]_+ - \sum_{r=1}^{k} \frac{\left[\mu\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{r+}^2}{[\mu]_{r+}}$$

$$= \sum_{r=1}^{k}\sum_{i\in G_r} \mu_i \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha} - \frac{\left[\Lambda_0\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{r+}}{[\Lambda_0]_{j+}}\right)^2.$$

Hence the asymptotic variances of the estimators $\hat{\beta}_{rc}$ and $\tilde{\alpha}_c$ are

$$\begin{aligned} \mathrm{AsVar}(\hat{\beta}_{1c}) &= [\mu]_{1+}^{-1} + l_1 E^{-1}, \\ \mathrm{AsVar}(\hat{\beta}_{rc}) &= ([\mu]_{r+}^{-1} + [\mu]_{1+}^{-1}) + l_r E^{-1}, \quad j = 2, \ldots, k, \\ \mathrm{AsVar}(\hat{\alpha}_c) &= E^{-1}. \end{aligned} \tag{3.20}$$

Note that throughout the thesis, the notation $\mathrm{AsVar}(\hat{\theta})$ is used to indicate that

$$\sqrt{M}(\hat{\theta} - \theta)$$

has asymptotic variance given by

$$\lim_{M\to\infty} M\,\mathrm{AsVar}(\hat{\theta}).$$

The inverse of $\mathcal{I}_t$, the information matrix based on the event-time data, can be similarly computed to obtain the asymptotic variance of $\hat{\beta}_{rt}$, $r = 1, \ldots, k$, and $(\hat{\alpha}_t)$. These are calculated using identical formulae as above, except E is replaced with

$$E_t = Z'VZ + H - Z'VZ(X'VX)^{-1}X'VZ = E + H.$$

Thus

$$\text{ARE}(\hat{\alpha}_c) = \frac{E}{E_t} = 1 - \frac{H}{E+H},$$

and

$$\begin{aligned}
\text{ARE}(\hat{\beta}_{1c}) &= \frac{[\mu]_{1+}^{-1} + l_1 E_t^{-1}}{[\mu]_{1+}^{-1} + l_1 E^{-1}} \\
&= \frac{([\mu]_{1+}^{-1} E_t + l_1) E_t^{-1}}{([\mu]_{1+}^{-1} E + l_1) E^{-1}} \\
&= \frac{([\mu]_{1+}^{-1} E + l_1) E + HE[\mu]_{1+}^{-1}}{([\mu]_{1+}^{-1} E + l_1)(E+H)} \\
&= 1 - \left(\frac{l_1}{[\mu]_{1+}^{-1} E + l_1}\right)\left(\frac{H}{E+H}\right).
\end{aligned}$$

$\text{ARE}(\hat{\beta}_{rc})$, $r = 2, \ldots, k$ is simplified in a similar manner as above.

When $l_r = 0$, from (3.17) and (3.18) we have that the asymptotic covariance of (β_j, α) equals 0, $r = 2, \ldots, k$. This would occur if, for example, $\{T_{ie}, i \in G_1\} = \{T_{ie}, i \in G_r\}$, $r = 2, \ldots, k$.

3.3.2 Case 2: Vector α

If $\boldsymbol{\alpha}$ is vector valued, with $\dim(\boldsymbol{\alpha}) = a$, then

$$X'VZ = \begin{pmatrix} \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{1+} \\ \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{2+} \\ \vdots \\ \left[\mu (\frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{k+} \end{pmatrix} \tag{3.21}$$

and

$$(X'VX)^{-1}X'VZ = \begin{pmatrix} \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{1+} / [\mu]_{1+} \\ \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{2+} / [\mu]_{2+} - \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{1+} / [\mu]_{1+} \\ \vdots \\ \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{k+} / [\mu]_{k+} - \left[\mu \frac{\partial \log \Lambda_0}{\partial \boldsymbol{\alpha}'}\right]_{1+} / [\mu]_{1+} \end{pmatrix}. \tag{3.22}$$

The elements of $(X'VX)^{-1}X'VZ$ do not depend on the β_rs; they are functions of the T_{ie}s and $\boldsymbol{\alpha}$.

For vector $\boldsymbol{\alpha}$, the quantity H (3.10) is a matrix of dimension $a \times a$. The asymptotic relative efficiency of the r-th element of $\hat{\boldsymbol{\alpha}}_c$ is $\mathrm{ARE}(\hat{\alpha}_{rc})) = (E_t^{-1})_{(r,r)}/(E^{-1})_{(r,r)} = ((I + E^{-1}H)E^{-1})_{(r,r)}/(E^{-1})_{(r,r)}$. How different this is from 1 is determined by $E^{-1}H$, and we shall look at this quantity more closely.

First note that E has the same structure as in (3.12), and its (c,d) element can be shown to equal

$$E_{(c,d)} = \sum_{r=1}^{k} \sum_{i \in G_r} \mu_i \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha_c} - \frac{\left[\frac{\partial \Lambda_0}{\partial \alpha_c}\right]_{r+}}{[\Lambda_0]_{r+}} \right) \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha_d} - \frac{\left[\frac{\partial \Lambda_0}{\partial \alpha_d}\right]_{r+}}{[\Lambda_0]_{r+}} \right). \tag{3.23}$$

If $T_{ie} = T_e$, $\forall i$, then E will be a matrix of zeros and only one of $\alpha_r, r = 1, \ldots, a, \beta_1$ can be estimated. If H has (c,d) element which can be written as

$$H_{c,d} = \sum_{i=1}^{M} \mu_i f(T_{ie}; \boldsymbol{\alpha}), \tag{3.24}$$

as in (3.16), then the only term involving β_1 in H and E is e^{β_1}, and occurs as a factor in each element of both of these matrices, so $\mathrm{ARE}(\hat{\alpha}_{rc})$ does not depend on β_1. If, in addition, $\{T_{ie}, i \in G_1\} = \cdots = \{T_{ie}, i \in G_k\}$, then the only term involving $\beta_1, \beta_2, \ldots, \beta_k$ in E and H is $(e^{\beta_1} + \sum_{r=2}^{k} e^{\beta_1 + \beta_r})$ which also occurs as a factor in each element of these matrices, so $\mathrm{ARE}(\hat{\alpha}_{rc})$ does not depend on $\boldsymbol{\beta}$, $r = 1, \ldots, k$. From (3.17) we recall that

$$\mathrm{AsVar}(\hat{\boldsymbol{\beta}}_c) = (X'VX)^{-1} + (X'VX)^{-1}X'VZE^{-1}Z'VX(X'VX)^{-1}$$

and

$$\mathrm{AsVar}(\hat{\boldsymbol{\beta}}_t) = (X'VX)^{-1} + (X'VX)^{-1}X'VZ(E+H)^{-1}Z'VX(X'VX)^{-1}.$$

Recall also that $(X'VX)^{-1}X'VZ$, (3.18), does not depend on the βs, and

$$(X'VX)^{-1} = e^{-\beta_1} f_1(\beta_2, \ldots, \beta_k, \boldsymbol{\alpha}, T_{ie}, i = 1, \ldots, M),$$

where $f_1(\cdot)$ is a matrix function, so the representation above signifies that $e^{-\beta_1}$ factors out of $(X'VX)^{-1}$ leaving a remainder term which is not a function of β_1. The matrix E^{-1} has a similar factorization as (3.23) and $(H + E)^{-1}$ also does when (3.24) holds, so $\mathrm{ARE}(\hat{\beta}_{ic})$

does not depend on β_1 in this situation, $r = 1, \ldots, k$. Finally, since (i) $(X'VX)^{-1}X'VZ$ does not depend on the βs; and (ii) because when $\{T_{ie}, i \in G_1\} = \ldots = \{T_i, i \in G_k\}$

$$E = \left(e^{\beta_1} + \sum_{r=2}^{k} e^{\beta_1+\beta_r}\right) f_2(T_{ie}, i = 1, \ldots, M, \boldsymbol{\alpha}),$$

and

$$H = \left(e^{\beta_1} + \sum_{r=2}^{k} e^{\beta_1+\beta_r}\right) f_3(T_{ie}, i = 1, \ldots, M, \boldsymbol{\alpha}),$$

where $f_2(\cdot)$ and $f_3(\cdot)$ are matrix functions; then, in this scenario, ARE$(\hat{\beta}_{1c})$ increases with increasing values of $\sum_{r=2}^{k} e^{\beta_r}$.

When the r-th row in the matrix $(X'VX)^{-1}X'VZ$ (3.18) is a vector of zeros, $r = 2, \ldots, k$ then, from (3.17), the asymptotic covariances of both $(\hat{\beta}_{rc}, \hat{\boldsymbol{\alpha}}_c)$ and $(\hat{\beta}_{rt}, \hat{\boldsymbol{\alpha}}_t)$ will be zero vectors and $(\mathcal{I}_c^{11})_{(r,r)} = (\mathcal{I}_t^{11})_{(r,r)} = ((X'VX)^{-1})_{(r,r)}$; hence, the estimator of the r-th treatment effect, $\hat{\beta}_{rc}$, will be fully efficient, $r = 2, \ldots, k$. This would occur, for example, when $\{T_{ie}, i \in G_r\} = \{T_i, i \in G_1\}$.

3.4 Models with random effects

Since it is common for counts to display extra-Poisson variation (Breslow 1984), it is natural to also consider models which deal with this phenomenon, commonly termed overdispersion. To incorporate overdispersion, we suppose a model with an individual-specific random effect ν_i, so that the intensity function corresponding to the i-th individual is

$$\lambda_i(t) = \nu_i \lambda_0(t; \boldsymbol{\alpha}) e^{\mathbf{x}_i'\boldsymbol{\beta}},$$

where the ν_is have some specified distribution. This is a mixture model, and if $\boldsymbol{\beta}$ includes an intercept, we may take the mean of the ν_is to be 1, without loss of generality; we set the variance of the ν_is to be τ. Then the mean of the counts is the same as previously, μ_i, but their variance is inflated to $\mu_i(1 + \tau\mu_i)$.

For a likelihood analysis, the distribution of the random effects, $p(\nu; \tau)$, must be specified. A flexible choice is the gamma distribution. In this case, the distribution of the counts is negative binomial. Other forms for the distribution of the counts are also tractable, for example, the Poisson-log-normal (Hinde 1982) or, the Poisson-inverse Gaussian (Dean, Lawless, and Willmot 1989). Because the gamma is often a reasonable choice (Lawless

1987a), we will outline analysis under this assumption. Simple modifications are required for likelihood analyses under the inverse Gaussian or log-normal mixtures.

The likelihood based on the event-time data factorizes as before (Lawless 1987b):

$$L_{\text{ot}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = L_{\alpha}(\boldsymbol{\alpha}) L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau), \tag{3.25}$$

where $L_{\alpha}(\boldsymbol{\alpha})$ is given in (3.4), and

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \prod_{i=1}^{M} \int_0^{\infty} (\nu_i \mu_i)^{n_i} e^{-\nu_i \mu_i} (n_i!)^{-1} p(\nu_i)\ d\nu_i,$$

is the likelihood for a mixed Poisson model. If ν_i is gamma distributed, $L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ is the negative binomial likelihood:

$$L_{\text{NB}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \prod_{i=1}^{M} \frac{\Gamma(n_i + \tau^{-1})}{(n_i!)\Gamma(\tau^{-1})} \left(\frac{\tau \mu_i}{1 + \tau \mu_i}\right)^{n_i} \left(\frac{1}{1 + \tau \mu_i}\right)^{\tau^{-1}}.$$

Because of the factorization in (3.25), the likelihood equations for $\boldsymbol{\beta}$ and τ arise as derivatives of the logarithm of L equated to zero. The likelihood equation for $\boldsymbol{\alpha}$ based on the event-time data is

$$\boldsymbol{g}_{\alpha t} = \sum_{i=1}^{M} \sum_{l=1}^{n_i} \left\{ \frac{\partial}{\partial \boldsymbol{\alpha}} \log \frac{\lambda_0(\omega_{ij}; \boldsymbol{\alpha})}{\Lambda_0(T_i; \boldsymbol{\alpha})} \right\} + \frac{\partial \log L}{\partial \boldsymbol{\alpha}} = \mathbf{0}. \tag{3.26}$$

For inference using the gamma mixture, note that Lawless (1987a) gives first and second derivatives of the logarithm of L_{NB} with respect to the parameters.

Likelihood inference based on the end-of-follow-up counts uses derivatives of the logarithm of L with respect to the parameters; the equations for $\boldsymbol{\beta}$ and τ are the same, but that for $\boldsymbol{\alpha}$ consists of only the second term in (3.26) above, equated to zero.

Lately, increasing importance has been given to quasi-likelihood or moment methods of estimation, and it is this approach that will be pursued here. The procedure suggested is simply to replace the contribution from L in the estimating equations for the parameters by suitable estimating functions. We use quasi-likelihood (Wedderburn 1974) for estimation of $\boldsymbol{\beta}$, and for estimation of $\boldsymbol{\alpha}$ when only the counts are available. The quasi-likelihood estimating equation for $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$ is

$$D_{\gamma}' V_{\text{o}}^{-1} (\boldsymbol{N} - \boldsymbol{\mu}) = \mathbf{0} \tag{3.27}$$

where $D_{\gamma} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\gamma}'$ and $V_{\text{o}} = \text{diag}(\mu_i(1 + \tau \mu_i); i = 1, \ldots, M)$.

This reduces to a set of equations

$$g_\tau = \sum_{i=1}^{M} \frac{(n_i - \mu_i)x_{i\tau}}{(1 + \tau\mu_i)} = 0, \quad \tau = 1, \ldots, k, \tag{3.28}$$

for estimating $\boldsymbol{\beta}$, and

$$\boldsymbol{g}_{\alpha c} = \sum_{i=1}^{M} \frac{(n_i - \mu_i)}{(1 + \tau\mu_i)} \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \boldsymbol{\alpha}} \right) = \mathbf{0},$$

for estimating $\boldsymbol{\alpha}$. Note that (3.27) reduces to (3.5) when $\tau = 0$.

There are many important reasons for the widespread use of the quasi-likelihood approach. For generalized linear models with a full likelihood, these are the maximum likelihood equations. Hence (3.28) is the maximum likelihood equation for $\boldsymbol{\beta}$ using either event-time or count data when the random effects are assumed to be gamma distributed. However, (3.5) is also applicable more generally. The asymptotic variance of the estimate of $\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}$, is independent of the choice of the estimating function for τ, and depends only on the first two moments of the distribution of the counts. The estimator of the asymptotic variance is consistent as long as the estimator of τ is consistent. Usual estimating equations for τ require first and second moment assumptions to hold for consistency of the estimator. This imparts a robust quality to the quasi-likelihood estimator. In contrast, for consistency of the maximum likelihood estimate of τ we require further distributional assumptions to be correct.

The quasi-likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are fully efficient with respect to maximum likelihood analysis of the counts under a negative binomial distribution (Lawless 1987a), and have high efficiency, for example, under a Poisson-inverse Gaussian distribution (Dean, Lawless, and Willmot 1989). Simulation studies have been conducted to investigate the performance of $\tilde{\boldsymbol{\beta}}$ in small samples; they support the unbiasedness and efficiency of this estimator (Nelder and Lee 1992).

Usual choices for the estimating equation for τ are as follows: equating the Pearson statistic to its degrees of freedom (Breslow 1984; Williams 1982), optimal quadratic estimation (Crowder 1987; Godambe and Thompson 1989), pseudo-likelihood estimation (Davidian and Carroll 1987) and extended quasi-likelihood (Nelder and Pregibon 1987). Davidian and Carroll (1987) derived the pseudo-likelihood equation as the maximum likelihood equation when residuals are normally distributed. Extended quasi-likelihood can be obtained as the non-normalized saddle-point approximation for exponential families as discussed by

Barndorff-Nielsen and Cox (1979). Davidian and Carroll (1987) have pointed out problems with the extended quasi-likelihood approach, including inconsistency of the estimates. Despite this drawback, extended quasi-likelihood performs as well as pseudo-likelihood in simulation studies, for the analysis of count data (Nelder and Lee 1992). Optimal quadratic estimation is optimal in the sense that the resulting estimator has minimal variance among estimators derived from quadratic estimating equations. It gives rise to equations for $\boldsymbol{\beta}$ and τ which require the specification of third and fourth moments of the distribution, which is an obvious disadvantage. However, if we presume these are as for the normal distribution, the equations reduce to quasi-likelihood for $\boldsymbol{\beta}$, and the pseudo-likelihood estimating equation for τ. Because of these unbiasedness and optimality properties, the pseudo-likelihood estimator is used here. The pseudo-likelihood estimating equation for τ is

$$g_\tau = \sum_{i=1}^{M} \frac{(n_i - \mu_i)^2 - (1 - h_i)\mu_i(1 + \tau\mu_i)}{(1 + \tau\mu_i)^2} = 0,$$

where h_i is the i-th element of the diagonal of the "hat matrix", $W^{1/2}X'(X'WX)^{-1}X'W^{1/2}$ and represents a small sample correction, and $W = \mathrm{diag}(\mu_i; i = 1, \ldots, M)$.

With the event-time data a suitable estimating equation for $\boldsymbol{\alpha}$ is obtained by combining $\partial \log L_\alpha / \partial \boldsymbol{\alpha}$ with quasi-likelihood estimation, yielding

$$\boldsymbol{g}_{\alpha t} = \sum_{i=1}^{M}\sum_{l=1}^{n_i} \left(\frac{\partial}{\partial \boldsymbol{\alpha}} \log \frac{\lambda_0(\omega_{il}; \boldsymbol{\alpha})}{\Lambda_0(T_{ie})} \right) + \sum_{i=1}^{M} \frac{(n_i - \mu_i)}{(1 + \tau\mu_i)} \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \boldsymbol{\alpha}} \right) = \mathbf{0}. \tag{3.29}$$

Hence, we solve $\boldsymbol{g}_c = (g_r, r = 1, \ldots, k, \boldsymbol{g}'_{\alpha c}, g_\tau)' = \mathbf{0}'$ when only count data are available, and $\boldsymbol{g}_t = (g_r, r = 1, \ldots, k, \boldsymbol{g}'_{\alpha t}, g_\tau)' = \mathbf{0}'$ with the event-time data. The quasi-likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \tau)'$ obtained by solving $\boldsymbol{g}_t = \mathbf{0}$ is denoted $\tilde{\boldsymbol{\theta}}_t$. Under standard conditions for the application of asymptotic results for estimating equations (Inagaki 1973; White 1982), $\sqrt{M}(\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta})$ is asymptotically normal with asymptotic covariance

$$\mathrm{E}\left(- \lim_{M\to\infty} \frac{1}{M} \frac{\partial \boldsymbol{g}_t}{\partial \boldsymbol{\theta}'} \right)^{-1} \mathrm{E}\left(\lim_{M\to\infty} \frac{1}{M} \boldsymbol{g}_t \boldsymbol{g}'_t \right) \mathrm{E}\left(- \lim_{M\to\infty} \frac{1}{M} \frac{\partial \boldsymbol{g}'_t}{\partial \boldsymbol{\theta}} \right)^{-1}. \tag{3.30}$$

It turns out that because of three sets of identities: (a) $\mathrm{E}(-\partial g_r/\partial\tau) = 0$, $r = 1, \ldots, k$, (*cf.* Moore 1986; Lawless 1987b), (b) $\mathrm{E}(-\partial \boldsymbol{g}_{\alpha t}/\partial\tau) = \mathbf{0}$; and, (c) the $\{k+\dim(\boldsymbol{\alpha})\}\times\{k+\dim(\boldsymbol{\alpha})\}$ upper left submatrix of $\mathrm{E}(\boldsymbol{g}_t \boldsymbol{g}'_t)$ is the same as the corresponding submatrix of $\mathrm{E}(-\partial \boldsymbol{g}_t/\partial\boldsymbol{\theta})$, the asymptotic variance of $\sqrt{M}(\tilde{\gamma}_t - \gamma)$, $\gamma' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$, from the overdispersed analysis is

$(\lim_{M\to\infty} \frac{1}{M}\mathcal{I}_{ot})^{-1}$ where

$$\mathcal{I}_{ot} = \begin{pmatrix} X'WV_o^{-1}WX & X'WV_o^{-1}WZ \\ Z'WV_o^{-1}WX & Z'WV_o^{-1}WZ + H \end{pmatrix}, \tag{3.31}$$

$W = \text{diag}(\mu_i; i = 1, \ldots, M)$, and H (3.10) is as defined in Section 3.2.

In a similar manner, the asymptotic variance of $\sqrt{M}(\tilde{\gamma}_c - \gamma)$, where $\tilde{\gamma}_c$ is the estimator from the overdispersed count data analysis, is found to be $(\lim_{M\to\infty} \frac{1}{M}\mathcal{I}_{oc})^{-1}$, where $\mathcal{I}_{oc}$ is the same as $\mathcal{I}_{ot}$ except for the $\dim(\boldsymbol{\alpha})\times\dim(\boldsymbol{\alpha})$ lower right submatrix, which is $Z'WV_o^{-1}WZ$.

The following results give the asymptotic relative efficiency of $\tilde{\gamma}_c$ relative to $\tilde{\gamma}_t$, for scalar α. These results are derived for scalar α in Section 3.5, beginning with the expressions for the asymptotic variance of $\tilde{\gamma}_c$ and $\tilde{\gamma}_t$. Comments on extensions for a vector-valued $\boldsymbol{\alpha}$ are also provided.

Result 4 *The asymptotic relative efficiency of $\tilde{\alpha}_c$ relative to $\tilde{\alpha}_t$ from the overdispersed Poisson analysis is*

$$\text{ARE}(\tilde{\alpha}_c) = 1 - \frac{H}{E_o + H},$$

where

$$E_o = \sum_{r=1}^{k}\sum_{i\in G_r} \frac{\mu_i}{1+\tau\mu_i}\left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial\alpha} - \frac{\left[\frac{\mu}{1+\tau\mu}\frac{\partial \log \Lambda_0}{\partial\alpha}\right]_{r+}}{\left[\frac{\mu}{1+\tau\mu}\right]_{r+}}\right)^2.$$

Result 5 *The asymptotic relative efficiency of $\tilde{\beta}_{1c}$ relative to $\tilde{\beta}_{1t}$ from the overdispersed analysis is*

$$\text{ARE}(\tilde{\beta}_{1c}) = 1 - \left\{\frac{\ell_1}{\left[\frac{\mu}{1+\tau\mu}\right]_{1+}^{-1} E_o + \ell_1}\right\}\left(\frac{H}{E_o + H}\right),$$

where

$$\ell_1 = \left(\frac{\left[\frac{\mu}{1+\tau\mu}\frac{\partial \log \Lambda_0}{\partial\alpha}\right]_{1+}}{\left[\frac{\mu}{1+\tau\mu}\right]_{1+}}\right)^2.$$

Result 6 *The asymptotic relative efficiency of $\tilde{\beta}_{rc}$ relative to $\tilde{\beta}_{rt}$, $r = 2, \ldots, k$, from the overdispersed analysis is*

$$\text{ARE}(\tilde{\beta}_{rc}) = 1 - \left\{\frac{\ell_r}{\left(\left[\frac{\mu}{1+\tau\mu}\right]_{r+}^{-1} + \left[\frac{\mu}{1+\tau\mu}\right]_{1+}^{-1}\right) E_o + l_{or}}\right\}\left(\frac{H}{E_o + H}\right),$$

where

$$\ell_r = \left(\frac{\left[\frac{\mu}{1+\tau\mu}\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{r+}}{\left[\frac{\mu}{1+\tau\mu}\right]_{r+}} - \frac{\left[\frac{\mu}{1+\tau\mu}\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{1+}}{\left[\frac{\mu}{1+\tau\mu}\right]_{1+}} \right)^2 .$$

With the overdispersed analysis, some important observations concerning the asymptotic relative efficiency of the estimator of the treatment effects can still be made.

First, we note that $\text{ARE}(\tilde{\beta}_{rc}) > \text{ARE}(\tilde{\alpha}_c)$, for $r = 1, \ldots, k$. As in Section 3.2, if $T_{ie} = T_e$, $\forall i$, then $E_o = 0$, and $\text{ARE}(\tilde{\alpha}_c) = \text{ARE}(\tilde{\beta}_{1c}) = 0$; the parameters α and β_1 cannot be simultaneously estimated.

The important result carried over from Section 3.2 to the present scenario of handling overdispersed data, defines full efficiency for the treatment estimators. When (a) there is *balance in the follow-up times* for the subjects under treatments 1 and r, as described in Section 3.2; and, (b) $\tau\beta_r = 0$, then $\text{ARE}(\tilde{\beta}_{rc}) = 1, r = 2, \ldots, k$. We demonstrate below and in Section 3.6 that even when (b) above does not hold, as long as the follow-up times are not extremely imbalanced over the treatment groups, the treatment estimators from the count data analysis retain very high efficiency.

To examine the effect of overdispersion on the asymptotic relative efficiency of the treatment estimator, we will again consider the experiments described at the end of Section 3.2 for the simple Poisson model. Recall that the balanced experiment compared two treatment groups made up of 10 subjects with $T_{ie} = 1$ and 10 at $T_{ie} = 2$; the experiment showing fairly extreme imbalance has the two groups made up as follows: the first group with 15 subjects at $T_{ie} = 1$ and 5 at $T_{ie} = 2$, and the second group with 5 subjects at $T_{ie} = 1$ and 15 at $T_{ie} = 2$. As before, the Weibull baseline intensity was used. The numerical studies allowed β_2 to range from -3 to +3, α from 0.5 to 3, and τ from 0.1 to 2.

In the balanced experiment, the effect of overdispersion is minimal. The lowest $\text{ARE}(\tilde{\beta}_{2c})$ observed is over 95%, and this occurs only for the largest values of τ and α and the smallest β_2.

In the imbalanced experiment, the comments made for the Poisson model still hold with reasonable accuracy, provided the amount of overdispersion is small. For example, for $\tau = 0.1$, the $\text{ARE}(\tilde{\beta}_{2c})$ is always greater than 70% and greater than 85% for much of the region examined. However, as the overdispersion increases, the size of the region over which the $\text{ARE}(\tilde{\beta}_{2c}) > 0.85$ decreases. In general, the relationship between the $\text{ARE}(\tilde{\beta}_{2c})$ and the other parameters is not as easily described as in Section 3.2. However, the effect

of overdispersion is to lower the efficiency of $\tilde{\beta}_2$ from that observed for the Poisson model, though it takes quite extreme combinations of parameters to see efficiencies as low as 70%.

3.5 Derivations of Asymptotic Variances and Asymptotic Relative Efficiencies of Quasi-Likelihood Estimators

3.5.1 Asymptotic Variance of $\tilde{\boldsymbol{\theta}}_t$ and $\tilde{\boldsymbol{\theta}}_c$

The asymptotic covariance of $\sqrt{M}(\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta})$, where $\tilde{\boldsymbol{\theta}}_t$ is the solution to $\boldsymbol{g}_t = \mathbf{0}$, is obtained from (3.30). Finite sample variance estimates are obtained by substituting $\tilde{\boldsymbol{\theta}}_t$ for $\boldsymbol{\theta}$ and omitting the expressions $\lim_{M\to\infty} \frac{1}{M}$. Note that

$$\mathrm{E}\left(-\frac{\partial \boldsymbol{g}_t}{\partial \boldsymbol{\theta}'}\right) = \begin{pmatrix} \mathcal{I}_{ot} & \mathbf{0} \\ \boldsymbol{b}' & b_0 \end{pmatrix},$$

and

$$\mathrm{E}\left(\boldsymbol{g}_t \boldsymbol{g}_t'\right) = \begin{pmatrix} \mathcal{I}_{ot} & \boldsymbol{c} \\ \boldsymbol{c}' & c_0 \end{pmatrix}.$$

Here $\mathcal{I}_{ot}$ is given in (3.31), $\mathbf{0}, \boldsymbol{b}$, and $\boldsymbol{c}$ are $(k+a)\times 1$ vectors, $a = \dim(\boldsymbol{\alpha})$, $\mathbf{0}$ is a vector of zeros, and $\boldsymbol{b}$ and $\boldsymbol{c}$ have elements

$$\begin{aligned}
b_r &= \sum_{i=1}^{M} \frac{1+2\tau\mu_i}{(1+\tau\mu_i)} x_{ir}, \quad r = 1,\ldots,k, \\
b_{k+s} &= \sum_{i=1}^{M} \frac{1+2\tau\mu_i}{1+\tau\mu_i} \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha_s}\right), \quad s = 1,\ldots,a, \\
c_r &= \sum_{i=1}^{M} \frac{\gamma_{3i}}{(1+\tau\mu_i)^3} x_{ir}, \quad \gamma_{3i} = \mathrm{E}\left((Y_i-\mu_i)^3\right), \quad r = 1,\ldots,k, \\
c_{k+s} &= \sum_{i=1}^{M} \frac{\gamma_{3i}}{(1+\tau\mu_i)^3} \left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha_s}\right), \quad s = 1,\ldots,a,
\end{aligned}$$

and

$$\begin{aligned}
b_0 &= \sum_{i=1}^{M} \frac{\mu_i^2}{(1+\tau\mu_i)^2}, \\
c_0 &= \sum_{i=1}^{M} \frac{\gamma_{4i} - \mu_i^2(1+\tau\mu_i)^2}{(1+\tau\mu_i)^4}, \quad \gamma_{4i} = \mathrm{E}\left((Y_i-\mu_i)^4\right).
\end{aligned}$$

The asymptotic variance of $\tilde{\boldsymbol{\theta}}_t$ is then estimated as

$$\begin{pmatrix} \mathcal{I}_{ot}^{-1} & \frac{1}{b_0}\mathcal{I}_{ot}^{-1}(\boldsymbol{c}-\boldsymbol{b}) \\ \frac{1}{b_0}(\boldsymbol{c}-\boldsymbol{b})'\mathcal{I}_{ot}^{-1} & \frac{1}{b_0^2}(c_0 - 2\boldsymbol{c}'\mathcal{I}_{ot}^{-1}\boldsymbol{b} + \boldsymbol{b}'\mathcal{I}_{ot}^{-1}\boldsymbol{b}) \end{pmatrix},$$

replacing $\boldsymbol{\theta}$ by $\tilde{\boldsymbol{\theta}}_t$.

The asymptotic variance of $\tilde{\boldsymbol{\theta}}_c$ is estimated as above, replacing $\mathcal{I}_{ot}$ with $\mathcal{I}_{oc}$ and $\boldsymbol{\theta}$ with $\tilde{\boldsymbol{\theta}}_c$.

If we assume $\gamma_{3i} = \mu_i(1+\tau\mu_i)(1+2\tau\mu_i)$ and $\gamma_{4i} = 6\tau\mu_i^2(1+\tau\mu_i)^2 + \mu_i(1+\tau\mu_i)(1+3\mu_i+3\tau\mu_i^2)$ as for the negative binomial distribution, then $\boldsymbol{c} = \boldsymbol{b}$. In this case, the estimators of $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$ and τ from either the event-time data or the end-of-follow-up count data are asymptotically independent. For a longer-tailed alternative we may use third and fourth moments as for the Poisson-inverse Gaussian mixture, $\gamma_{3i} = \mu_i(1+\tau\mu_i)(1+2\tau\mu_i)+\tau^2\mu_i^3$ and $\gamma_{4i} = 7\tau\mu_i^2(1+\tau\mu_i)^2 + \mu_i(1+4\tau^2\mu_i^2+8\tau^3\mu_i^3)$. Note that only mean and variance assumptions are required for consistency of the asymptotic variance of $\tilde{\gamma}_t$ or $\tilde{\gamma}_c$.

3.5.2 Asymptotic Relative Efficiency of $\tilde{\boldsymbol{\theta}}_c$

The inverse of the information matrix $\mathcal{I}_{oc}$ is

$$\mathcal{I}_{oc}^{-1} = \begin{pmatrix} I_{oc}^{11} & I_{oc}^{12} \\ I_{oc}^{21} & I_{oc}^{22} \end{pmatrix}$$

where

$$\begin{aligned}
\mathcal{I}_{oc}^{11} &= (X'WV_o^{-1}WX)^{-1} \\
&\quad + (X'WV_o^{-1}WX)^{-1}X'WV_o^{-1}WZ\mathcal{I}_{oc}^{22}Z'WV_o^{-1}WX(X'WV_o^{-1}WX)^{-1}, \\
\mathcal{I}_{oc}^{12} &= \mathcal{I}_{oc}^{21} = -(X'WV_o^{-1}WX)^{-1}X'WV_o^{-1}WZ\mathcal{I}_{oc}^{22}, \\
\mathcal{I}_{oc}^{22} &= E_o^{-1} = (Z'WV_o^{-1}WZ - Z'WV_o^{-1}WX(X'WV_o^{-1}WX)^{-1}X'WV_o^{-1}WZ)^{-1}.
\end{aligned}$$

Let $v_r = [\mu/(1+\tau\mu)]_{r+}, r = 1, \ldots, k$, and note

$$(X'WV_o^{-1}WX)^{-1} = \begin{pmatrix} v_1^{-1} & -\mathbf{1}'v_1^{-1} \\ -\mathbf{1}v_1^{-1} & \mathbf{1}\mathbf{1}'v_1^{-1} + \mathrm{diag}(v_r^{-1}; r = 2, \ldots, k) \end{pmatrix},$$

where $\mathbf{1}$ is a $(k-1)\times 1$ vector of ones. For α a scalar quantity, and letting

$$q_r = \left[\frac{\mu}{1+\tau\mu}\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{r+},$$

we have

$$X'WV_o^{-1}WZ = ([q]_+, q_2, \ldots, q_k),$$

and

$$(X'WV_o^{-1}WX)^{-1}X'WV_o^{-1}WZ = (\ell_1^{1/2}, \ldots, \ell_k^{1/2}).$$

Here, as in Section 3.4, $\ell_1 = (q_1/v_1)^2$, and $\ell_r = \{(q_r/v_r) - \ell_1^{1/2}\}^2$, $r = 2, \ldots, k$. Then,

$$\begin{aligned} E_o &= \left[\frac{\mu}{1+\tau\mu}\left(\frac{\partial \log \Lambda_0}{\partial \alpha}\right)^2\right]_+ - \sum_{r=1}^{k}\frac{q_r^2}{v_r} \\ &= \sum_{r=1}^{k}\sum_{i\in G_r}\frac{\mu_i}{1+\tau\mu_i}\left(\frac{\partial \log \Lambda_0(T_{ie})}{\partial \alpha} - \frac{\left[\frac{\mu}{1+\tau\mu}\frac{\partial \log \Lambda_0}{\partial \alpha}\right]_{r+}}{\left[\frac{\mu}{1+\tau\mu}\right]_{r+}}\right)^2. \end{aligned}$$

Hence, following from Section 3.3, $\mathrm{AsVar}(\tilde{\beta}_{1c})$, $\mathrm{AsVar}(\tilde{\beta}_{rc})$, and $\mathrm{AsVar}(\tilde{\alpha}_c)$ are the same as in (3.20), except $[\mu]_{r+}$ is replaced with $[\mu/(1+\tau\mu)]_{r+}$, $r = 1, \ldots, k$, E is replaced with E_o, and l_r becomes ℓ_r. The inverse of $\mathcal{I}_{ot}$, the information matrix based on the analysis using the event-time data with overdispersion incorporated, is similarly computed. We obtain identical formulae for the asymptotic variances of the parameter estimators as in (3.20), except $[\mu]_{r+}$ is replaced with $[\mu/(1+\tau\mu)]_{r+}$, l_r becomes ℓ_r and E is replaced with $E_o + H$. The asymptotic relative efficiencies are computed as in Section 3.3.

With overdispersion incorporated into the analysis, we need $\ell_r = 0$ in order that $\mathrm{ARE}(\beta_{rc}) = 1$. This occurs when (a) $l_r = 0$, i.e., there is a *balance in the follow-up times* for individuals under treatments 1 and r; and, (b) $\beta_r = 0$. Condition (a) will hold, for example, when $\{T_{ie}, i \in G_r\} = \{T_{ie}, i \in G_1\}$. For $\boldsymbol{\alpha}$ a vector-valued quantity, similar algebraic manipulations as in Section 3.3 will lead to the requirement of the same conditions (a) and (b) above, in order that the treatment estimator based on the count data be fully efficient.

3.6 Illustration

The trial conducted by the Veterans Administrative Co-operative Urological Research Group (Byar et al. 1977) on the recurrence of bladder cancer provides a scenario for illustrating these comparisons. This randomized clinical trial studied the effect of three treatments on the frequency of recurrence of bladder cancer. The treatments were placebo pills, pyridoxine pills and periodic instillation of thiotepa into the bladder. The data appears in Andrews

	No. of Recurrences										Total No. of
Treatment	0	1	2	3	4	5	6	7	8	9	Patients
1. Placebo	18	10	4	6	2	4	1	0	1	1	47
2. Pyridoxine	16	5	4	0	0	2	0	0	2	2	31
3. Thiotepa	20	8	3	2	2	2	0	1	0	0	38

Table 3.3: Distribution of the Number of Recurrences Observed for the Subjects in each of the Three Treatment Groups in the Bladder Cancer Study.

	Follow-Up Times in Months					
Treatment	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
1. Placebo	1	23.00	30.0	32.51	43.0	64
2. Pyridoxine	2	12.50	37.0	32.03	45.5	60
3. Thiotepa	1	18.25	32.5	31.13	44.0	59

Table 3.4: Some Summary Statistics Relating to the Follow-Up Times of Subjects in the Three Treatment Groups for the Bladder Cancer Study.

and Herzberg (1985). All 116 patients had superficial bladder cancer when they entered the study. The tumors were removed and the patients were randomly assigned to one of the three treatments.

Table 3.3 shows the distribution of the number of recurrences observed for the patients in each of the three groups. There were 47 patients in the first group (Placebo), 31 in the second (Pyridoxine Treatment) and 38 in the third group (Thiotepa Treatment). The plot of the Cumulative Mean Functions (CMFs) is given in Figure 3.2. It would appear that the placebo and pyridoxine treatments have similar recurrence patterns and CMFs, while the thiotepa treatment appears to have been effective at reducing the number of recurrences. However, the variability associated with the CMFs (not shown) is substantial and makes it clear that it will be difficult to distinguish among the treatment groups. A likelihood ratio test for group specific α's had a p-value of 0.95. So, it seems that the proportional intensity assumption is reasonable for these data, but we cannot expect to detect a difference between the Placebo and Pyridoxine groups.

Because the distribution of the follow-up times over the three groups plays an important role in determining the asymptotic relative efficiencies of the estimators based on the aggregated data, summary statistics for the follow-up time distributions are presented here.

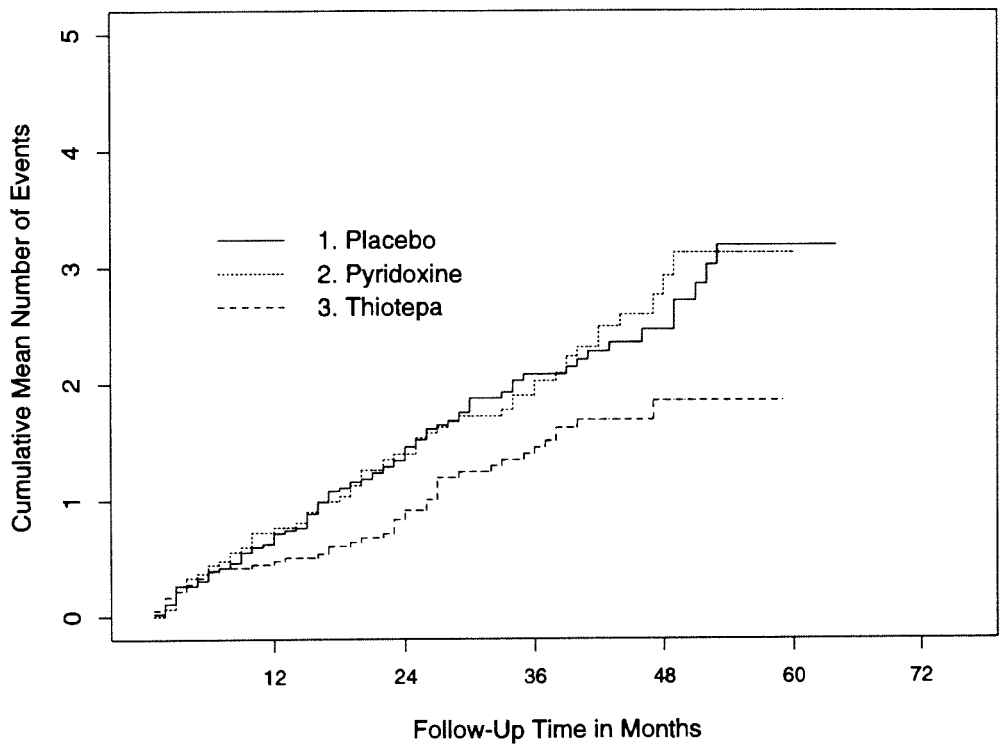

Figure 3.2: Cumulative Mean Functions for the Three Treatment Groups in the Bladder Cancer Study.

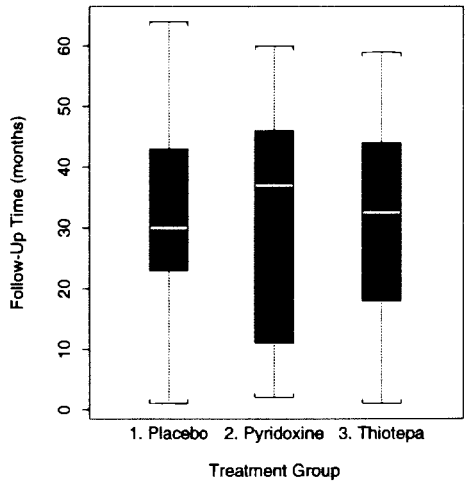

Figure 3.3: Boxplot of End-of-Follow-Up Times By Treatment Group. The end-of-follow-up times T_{ie} are shown for the three treatment groups of the bladder cancer recurrence study. The groups appear to be reasonably balanced with respect to follow-up times.

The follow-up times for the patients ranged from about one month to five years for each of the three groups. Table 3.4 gives summary statistics for the follow-up times in three groups and Figure 3.3 shows a boxplot of the follow-up times by group. In general, the follow-up-time distributions appear to be similar for the three groups. For example, the overall mean follow-up time was 32 months, and the means of the three groups are 32.5, 32.0, and 31.1 months. Because the distributions are similar, though not identical, the efficiencies of the estimators of treatment effects should be close to 100%.

The distribution of follow-up times is important for reasons beyond efficiency concerns. The methods proposed have assumed complete independence of follow-up times and recurrence times. When this assumption is not reasonable, these methods may not be appropriate. Consider, for example, a study in which subjects on the placebo experienced higher recurrence rates, and subjects who experienced higher recurrence rates were more likely to withdraw early. Under these conditions, parameter estimates based on these methods could be seriously biased.

Estimates of the parameters obtained by fitting the model using the Poisson process likelihood are given in Table 3.5. Table 3.6 shows results from fitting the overdispersed

	Event-Time Data		Count Data	
Parameter	Estimate	Std. Error	Estimate	Std. Error
β_1	-2.852	0.2615	-2.025	0.5238
β_2	0.008	0.1704	0.020	0.1705
β_3	-0.403	0.1836	-0.407	0.1837
α	0.996	0.0662	0.765	0.1440

Table 3.5: Parameters Estimates from the Poisson Point Process Likelihood Analysis of the Bladder Cancer Data. Estimates have been calculated using both the event-time data and the end-of-follow-up count data. Standard errors have been calculated from the model-based variance estimators. Note that likelihood and quasi-likelihood methods yield the same estimators for the Poisson model.

Poisson process using both likelihood and quasi-likelihood methods. A Weibull baseline intensity function $\lambda_0(t) = \alpha t^{\alpha-1}$ was used for all analyses. The standard errors obtained from the simple Poisson model appear to be underestimated, as expected in the presence of overdispersion. Examining the estimates and their standard errors, there appears to be little evidence against the hypotheses that $\beta_2 = 0$ or $\beta_3 = 0$; there is strong evidence that $\tau \neq 0$, and $\alpha = 1$ would appear to be a reasonable hypothesis.

It is perhaps surprising that the standard errors for the quasi-likelihood estimates in Table 3.6 are smaller than those for the likelihood estimates. However, for this data set, the likelihood estimate of the overdispersion parameter τ is larger than the quasi-likelihood estimate. The effect of this apparent increase in overdispersion is to inflate the standard errors for the likelihood estimates. With identical parameter estimates, the true variances of the likelihood estimates would, of course, be smaller than those of the quasi-likelihood estimates, though the standard errors may not be.

Relative efficiencies can be computed for this study by comparing the variances of the estimates from the aggregated and full analyses. In addition, we compute asymptotic relative efficiencies for several combinations of $\boldsymbol{\beta}$, α and τ around their quasi-likelihood estimates and using the follow-up times from the study. Some asymptotic relative efficiencies are displayed in Table 3.7. Looking at efficiency values for all combinations of the parameters shows that they do not vary sharply; the efficiencies of $\tilde{\beta}_{2c}$ and $\tilde{\beta}_{3c}$ are close to 1, while that of $\tilde{\beta}_{1c}$ and $\tilde{\alpha}_c$ are quite low.

To reiterate, recording and analyzing only the end-of-follow-up counts would lead to an imprecise estimate of the Weibull baseline intensity parameter and the mean recurrence

Likelihood Analysis

	Event-Time Data		Count Data	
Parameter	Estimate	Std. Error	Estimate	Std. Error
β_1	-2.955	0.3179	-2.004	0.7457
β_2	0.132	0.3318	0.095	0.3258
β_3	-0.282	0.3228	-0.317	0.3183
α	1.019	0.0693	0.746	0.2088
τ	1.351	0.3181	1.319	0.3148

Quasi-Likelihood Analysis

	Event-Time Data		Count Data	
Parameter	Estimate	Std. Error	Estimate	Std. Error
β_1	-2.935	0.3016	-1.993	0.6929
β_2	0.104	0.2908	0.083	0.2905
β_3	-0.301	0.2857	-0.328	0.2863
α	1.014	0.0688	0.745	0.1937
τ	0.909	0.2858	0.934	0.2923

Table 3.6: Parameters Estimates from the Poisson-Gamma Mixture Likelihood and the Quasi-Likelihood Analysis of the Bladder Cancer Data. Estimates have been calculated using both the event-time data and the end-of-follow-up count data. Standard errors have been calculated from the model-based variance estimators.

β	τ	α	ARE($\tilde{\beta}_{1c}$)	ARE($\tilde{\beta}_{2c}$)	ARE($\tilde{\beta}_{3c}$)	ARE($\tilde{\alpha}_c$)
(-3, 0.1, -0.3)	0.50	0.75	0.155	1.000	1.000	0.109
		1.00	0.161	0.999	1.000	0.120
		1.25	0.150	0.999	1.000	0.109
	1.00	0.75	0.148	1.000	1.000	0.096
		1.00	0.147	1.000	1.000	0.098
		1.25	0.132	1.000	1.000	0.081
	2.00	0.75	0.140	1.000	1.000	0.079
		1.00	0.134	1.000	1.000	0.074
		1.25	0.120	0.999	1.000	0.057
(-3, 1.0, -1.0)	0.50	0.75	0.166	0.999	1.000	0.107
		1.00	0.161	0.999	1.000	0.111
		1.25	0.144	0.998	0.999	0.096
	1.00	0.75	0.153	0.997	1.000	0.089
		1.00	0.143	0.996	1.000	0.087
		1.25	0.127	0.995	0.999	0.069
	2.00	0.75	0.139	0.995	1.000	0.069
		1.00	0.129	0.994	1.000	0.062
		1.25	0.115	0.993	0.999	0.046
(-3, 3.0, -3.0)	0.50	0.75	0.174	0.983	1.000	0.051
		1.00	0.131	0.973	1.000	0.044
		1.25	0.103	0.969	0.999	0.032
	1.00	0.75	0.137	0.976	1.000	0.033
		1.00	0.108	0.970	0.999	0.027
		1.25	0.091	0.970	0.997	0.019
	2.00	0.75	0.112	0.974	1.000	0.020
		1.00	0.095	0.972	0.998	0.016
		1.25	0.086	0.974	0.995	0.011

Table 3.7: Asymptotic Relative Efficiencies of the Analysis of Count Data. The combinations of parameter values chosen are similar to those observed in the quasi-likelihood analysis of the bladder cancer study.

rate, but gives estimates of the treatment effects which are almost as precise as those based on the event-times.

3.7 Discussion

This chapter has used quasi-likelihood based estimating functions for the analysis of random effects Poisson process models, a natural extension of quasi-likelihood methodology, in view of the factorization of the likelihood in (3.25). By using these quasi-likelihood equations to obtain parameter estimates and standard errors, instead of a maximum likelihood approach, we need not make full distributional assumptions concerning the random effects, but only first and second moment assumptions. To provide increased robustness against misspecification of the variance form, we could also employ the so-called "sandwich" or covariance estimator (e.g., Liang and Zeger 1986; Liang, Zeger, and Qaqish 1992). This replaces the model-based middle term, $E\{\boldsymbol{g}\boldsymbol{g}'\}|_{\tilde{\boldsymbol{\theta}}}$, by a robust estimator, $\boldsymbol{g}\boldsymbol{g}'$, in the finite sample estimate of the variance in (3.30). However, there is a corresponding loss of efficiency through the use of the sandwich estimator, and it seems that fairly large samples are required, else the asymptotic variance tends to be underestimated Breslow (1990).

The focus of the chapter has been a consideration of the efficiency loss in the analysis of end-of-follow-up counts of the number of events instead of actual event-times for random effects Poisson processes during a specified observation period. We have shown that such loss can be substantial for estimation of the baseline intensity function, which is not surprising. There are situations, however, where the estimators of treatment effects based on the aggregated data analysis will have high efficiency. Precise conditions for such high efficiency have been detailed. Loosely speaking, the follow-up times for subject in the treatment groups must be reasonably similar over the groups. For example, no one treatment group should contain only the smallest or largest follow-up times. We have noted, in addition, that the treatment estimates obtained from the event-time and count data analyses will be identical when the follow-up times are the same for the treatment groups, and if there is no overdispersion.

A straightforward extension to this work would study the gains in asymptotic relative efficiency of the baseline intensity function in the analysis of *panel* data, i.e. when counts are recorded at multiple follow-up times. We present these results in Chapter 4.

Chapter 4

The Analysis of Recurrent Event Panel Data

4.1 Introduction

In the previous chapter we examined a simple variation of the standard quasi-likelihood methods, adapted for inference for nonhomogeneous Poisson processes. We focussed on examining the efficiency of analyses based on end-of-follow-up counts relative to analyses based on event-time data and showed that under reasonable conditions, which will be reviewed in this chapter, covariate effects can be estimated efficiently based on end-of-follow-up counts; however, end-of-follow-up counts are not efficient for the estimation of other model parameters.

In this chapter, we consider a natural way to recover information that is lost when only end-of-follow-up counts are recorded, without resorting to collecting event-time data. We compare the efficiency of the analysis of *panel data* to the efficiency of the analysis of end-of-follow-up count data. Assuming M subjects,

Panel Data consists of the set $\{n_{ij} : i = 1, \ldots, M, j = 1, \ldots, s_i\}$; n_{ij} is the number of events occurring for subject i, in period j. Subjects are examined periodically, at times $T_1, T_2, \ldots, T_s$. Period j is the interval $(T_{j-1}, T_j]$. Drop-outs are considered to happen immediately following the subject's last observed follow-up-time, T_{s_i}. Estimators based on panel data are subscripted p.

End-of-follow-up Count Data consists of the set of end-of-follow-up counts, $\{n_{i+} : i =$

$1, \ldots, M\}$; n_{i+} is the total number of events to have occurred for subject i between times 0 and T_{s_i}; $n_{i+} = \sum_{i=1}^{s_i} n_{ij}$. Estimators based on count data are subscripted c.

Event-Time Data consists of the set of event times, $\{\omega_{ijl} : i = 1, \ldots, M, j = 1, \ldots, s_i, l = 1, \ldots, n_{ij}\}$; ω_{ijl} is the time elapsed from $t = 0$ for the l-th event occurring for subject i, in period j (i.e., occurring in $(T_{j-1}, T_j]$). Subjects are examined continuously from time T_0 until time T_s. Estimators based on event-time data are subscripted t.

In Section 4.2 we first review the quasi-likelihood estimator and consider its performance relative to the likelihood estimator. The efficiency of the quasi-likelihood analysis of panel data versus end-of-follow-up count data is discussed in Section 4.3. We consider both asymptotic comparisons and small sample investigations through a simulation study. Score-type tests for overdispersion and nonhomogeneity are discussed in Section 4.4 along with diagnostic checks for the baseline intensity function. In Section 4.5 we illustrate our methods by analyzing data from a clinical trial examining bladder cancer recurrence rates.

4.2 Likelihood and Quasi-Likelihood Estimators

Let $Y_i(t)$ represent the counting process for the i-th individual. For example, $Y_i(t)$ could be a count of the number of failures of a computer system, or the number of recurrences of a medical condition (e.g., migraine headaches or seizures). In these examples, t represents the time since the i-th subject was enrolled in the study. This could represent the time from some common starting point — e.g., $t = 0$ could correspond to 9:00 a.m. on the day that the reliability study was started — or it could represent a subject-specific starting time — e.g., $t = 0$ corresponds to the time the subject was first treated for migraines. We assume that $Y_i(t)$ has an intensity function of the proportional intensity form:

$$\lambda_i(t) = \nu_i \lambda_0(t; \boldsymbol{\alpha}) e^{\boldsymbol{x}_i' \boldsymbol{\beta}}, \tag{4.1}$$

where $\lambda_0(\cdot)$ is a twice differentiable baseline intensity function of t, depending on $\boldsymbol{\alpha}$; $\boldsymbol{x}_i$ is the covariate vector for the i-th individual; $\boldsymbol{\beta}$ is the vector of associated regression parameters; and ν_i represents the random effect associated with subject i. By including ν_i, we explicitly permit the overdispersion that is often evident in the analysis of count data. As discussed in the previous chapter, we assume the random effects are *i.i.d.* with probability density function $p(\nu; \tau)$; here $\text{Var}(\nu) = \tau$ reflects the extent of the overdispersion. If $\boldsymbol{\beta}$ contains a

general mean term, for example when $x_{i1} = 1$ for all i, then we may set $\mathrm{E}(\nu) \equiv 1$ without loss of generality. Thus, given ν_i, $Y_i(t)$ is modeled as a Poisson process with intensity $\lambda_i(t)$, independently for each subject.

In the first subsection below, likelihood and quasi-likelihood analyses based on the event-time data are compared. The second subsection gives corresponding analyses for the panel data.

4.2.1 Estimation Based on the Event-times

We briefly review likelihood estimation for the event-time data. The notation here is the same as in Chapter 3, except that the presence of the intermediate follow-up times T_1, T_2, T_{s_i} necessitates the addition of a subscript j to indicate events occurring during interval j, $(T_{j-1}, T_j]$.

Conditionally on the random effects $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M)'$, the likelihood based on the event-time data is

$$L_{\mathrm{t}}(\boldsymbol{\beta}, \boldsymbol{\alpha} \mid \boldsymbol{\nu}) = \prod_{i=1}^{M} \prod_{j=1}^{s_i} \left\{ \prod_{l=1}^{n_{ij}} \lambda_i(\omega_{ijl}) \exp\left(-\int_{w_{ij,l-1}}^{\omega_{ijl}} \lambda_i(u)\, du\right) \right\} \times \exp\left(-\int_{w_{ijn_{ij}}}^{T_{s_i}} \lambda_i(u)\, du\right). \tag{4.2}$$

To handle missing data, for example data from the interval $(T_{j-1}, T_j]$ for subject i, we simply omit the contributions from that interval in the product term above. However, as in the previous chapter, any censoring or other incidence of missing data is hereafter assumed to be independent of the counting process. The likelihood (4.2) simplifies to

$$L_{\mathrm{t}}(\boldsymbol{\beta}, \boldsymbol{\alpha} \mid \boldsymbol{\nu}) = \exp\left(\sum_{i=1}^{M} \sum_{j=1}^{s_i} n_{ij} \boldsymbol{x}_i' \boldsymbol{\beta}\right) \left\{ \prod_{i=1}^{M} \prod_{j=1}^{s_i} \prod_{l=1}^{n_{ij}} \nu_i \lambda_0(\omega_{ijl}; \boldsymbol{\alpha}) \right\} \times \exp\left(-\sum_{i=1}^{M} \nu_i e^{\boldsymbol{x}_i' \boldsymbol{\beta}} \Lambda_0(T_{s_i})\right), \tag{4.3}$$

where $\Lambda_0(T_j) = \Lambda_0(T_j; \boldsymbol{\alpha}) = \int_0^{T_j} \lambda_0(u; \boldsymbol{\alpha})\, du$ is the cumulative baseline intensity function.

After integrating over the density $p(\nu)$, the (unconditional) likelihood becomes

$$L_{\mathrm{t}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \int_{\boldsymbol{\nu}} L_{\mathrm{t}}(\boldsymbol{\beta}, \boldsymbol{\alpha} | \boldsymbol{\nu}) p(\boldsymbol{\nu})\, d\boldsymbol{\nu}. \tag{4.4}$$

Let $\mu_{ij} = E(n_{ij})$, $\mu_{ij} = \{\Lambda_0(T_j) - \Lambda_0(T_{j-1})\}e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$, and $\mu_{i+} = \sum_{j=1}^{s_i} \mu_{ij}$, so $\mu_{i+} = \Lambda_0(T_{s_i})e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$; $\mu_{i+} = E(n_{i+})$. Then, we can write the likelihood (4.4) as a product of two factors,

$$L_t(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = L_\alpha(\boldsymbol{\alpha}) L_{\mathrm{MP}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau), \tag{4.5}$$

where

$$L_\alpha(\boldsymbol{\alpha}) = \prod_{i=1}^{M} \prod_{j=1}^{s_i} \prod_{l=1}^{n_{ij}} \frac{\lambda_0(\omega_{ijl}; \boldsymbol{\alpha})}{\Lambda_0(T_{s_i})}, \tag{4.6}$$

and

$$L_{\mathrm{MP}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \prod_{i=1}^{M} \int_0^\infty (\nu_i \mu_{i+})^{n_{i+}} e^{-\nu \mu_{i+}} p(\nu_i; \boldsymbol{\alpha})\, d\nu_i. \tag{4.7}$$

The first term is the probability kernel from the conditional density of the event times, given n_{i+} events occurred, $i = 1, \ldots, M$; the second term is the mixed Poisson likelihood kernel for the probability that n_{i+} events occur in $(0, T_{s_i}]$, $i = 1, \ldots, M$. This factorization is key to the construction of estimating functions which follows.

A commonly used parametric form for the intensity of counting processes is the Weibull (Lawless 1987b; Bain 1978). In this case $\lambda_0(w) = \alpha w^{\alpha-1}$ and (4.6) becomes

$$L_{\alpha \mathrm{W}}(\alpha) = \prod_{i=1}^{M} \prod_{j=1}^{s_i} \prod_{l=1}^{n_{ij}} \frac{\alpha \omega_{ijl}^{\alpha-1}}{T_{s_i}^{\alpha}} \tag{4.8}$$

where the subscript W denotes Weibull. If we assume a gamma distribution for the random effects, ν_i, then the probability function for n_{i+} is negative binomial and L_{MP} becomes

$$L_{\mathrm{NB}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \frac{\Gamma(n_{i+} + \tau^{-1})}{n_{i+}! \Gamma(\tau^{-1})} \left(\frac{\tau \mu_{i+}}{1 + \tau \mu_{i+}}\right)^{n_{i+}} \left(\frac{1}{1 + \tau \mu_{i+}}\right)^{\tau^{-1}}; \tag{4.9}$$

where the subscript NB replaces MP to signify that this particular mixed Poisson model is the negative binomial. In the development which follows, we derive results for the special case of the Weibull baseline intensity and negative binomial mixture. These forms have a history of common usage in counting processes (Crow 1974; Lee and Lee 1978; Lawless 1987a). Likelihood estimates using other parametric forms can be obtained using simple modifications to the procedures discussed below.

First and expected second derivatives of the logarithm of the likelihood L_t (4.5) with respect to the parameters $\boldsymbol{\beta}$ and τ equal the corresponding derivatives of L_{MP}. These derivatives are given in Lawless (1987a) when L_{MP} equals the negative binomial likelihood (4.9) and in Dean, Lawless, and Willmot (1989) when L_{MP} equals the Poisson-inverse Gaussian

mixture. For the special case of a Weibull baseline and negative binomial mixture, the first derivative of the logarithm of L_t with respect to α is

$$\frac{\partial \log L_t}{\partial \alpha} = \sum_{i=1}^{M} \sum_{j=1}^{s_i} \sum_{l=1}^{n_{ij}} \left(\log \omega_{ijl} + \frac{1}{\alpha} - \log T_{s_i} \right) + \frac{\partial \log L_{NB}}{\partial \alpha}. \tag{4.10}$$

The second term in this expression is the derivative of the logarithm of the negative binomial likelihood

$$\frac{\partial \log L_{NB}}{\partial \alpha} = \sum_{i=1}^{M} \frac{(n_{i+} - \mu_{i+}) \log T_{s_i}}{1 + \tau \mu_{i+}} = D_\alpha' V_o^{-1} (\boldsymbol{n}_+ - \boldsymbol{\mu}_+) \tag{4.11}$$

where $\boldsymbol{n}_+ = (n_{1+}, \ldots, n_{M+})'$ is the vector of end-of-follow-up counts; $\boldsymbol{\mu}_+ = (\mu_{1+}, \ldots, \mu_{M+})'$ is the corresponding vector of expected end-of-follow-up counts; $V_o = \text{diag}(\mu_{i+}(1+\tau\mu_{i+}); i = 1, \ldots, M)$ is the diagonal matrix of variances of the n_{i+}s; and, $D_\alpha = \partial \boldsymbol{\mu}_+ / \partial \alpha$. The expected second derivative of the log-likelihood with respect to α is

$$\text{E}\left(-\frac{\partial^2 \log L_t}{\partial \alpha^2} \right) = \sum_{i=1}^{M} \frac{\mu_{i+}}{\alpha^2} + \text{E}\left(-\frac{\partial^2 \log L_{NB}}{\partial \alpha^2} \right), \tag{4.12}$$

where

$$\text{E}\left(-\frac{\partial^2 \log L_{NB}}{\partial \alpha^2} \right) = \sum_{i=1}^{M} \frac{\mu_{i+} (\log T_{s_i})^2}{1 + \tau \mu_{i+}}. \tag{4.13}$$

As described in Sec 3.4, quasi-likelihood and other moment methods have become popular for inference concerning count data. Unlike likelihood methods, which require a full distributional specification, quasi-likelihood requires only first and second moment assumptions for n_{i+} in order to specify consistent estimating functions. As was the case in the previous chapter, the quasi-likelihood estimating function for $\boldsymbol{\beta}$ is the same as the maximum likelihood estimating function under a negative binomial likelihood. However, the quasi-likelihood estimating function is more robust because only first and second moment assumptions are sufficient to ensure that $\tilde{\boldsymbol{\beta}}$ and its variance can be consistently estimated; using the likelihood estimating functions requires the specification of full distributional assumptions. This is described in greater detail in Section 3.5, as well as later in this section when we examine the variance of the quasi-likelihood estimators.

The quasi-likelihood equations for the event-time data were detailed in Chapter 3, and we briefly review them here. Continuing to work under the assumption of a Weibull baseline intensity function, the estimating functions are

$$g_\beta = D_\beta' V_o^{-1} (\boldsymbol{n}_+ - \boldsymbol{\mu}_+), \tag{4.14}$$

$$g_{\alpha t} = \sum_{i=1}^{M}\sum_{j=1}^{s_i}\sum_{l=1}^{n_{ij}}\left(\log\omega_{ijl} + \frac{1}{\alpha} - \log T_{s_i}\right) + D_\alpha' V_o^{-1}(\boldsymbol{n}_+ - \boldsymbol{\mu}_+), \tag{4.15}$$

$$g_\tau = \sum_{i=1}^{M}\frac{(n_{i+}-\mu_{i+})^2 - \mu_{i+}(1+\tau\mu_{i+})}{(1+\tau\mu_{i+})^2}, \tag{4.16}$$

where $D_\beta = \partial\boldsymbol{\mu}_{i+}/\partial\boldsymbol{\beta}'$ and $V_o = \mathrm{diag}(\mu_{i+}(1+\tau\mu_{i+}); i = 1,\ldots,M)$. We obtain the quasi-likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha, \tau)'$ by solving $\boldsymbol{g}_t = \boldsymbol{0}$, where $\boldsymbol{g}_t' = (\boldsymbol{g}_\beta', g_\alpha, g_\tau)$. The estimator will be written $\tilde{\boldsymbol{\theta}}_t$, with components $\tilde{\boldsymbol{\beta}}_t, \tilde{\alpha}_t$, and $\tilde{\tau}_t$; the subscript t indicates an estimator based on the event-time data. The estimating function for τ (4.16) is a bias corrected form of the pseudo-likelihood estimating equation suggested by Davidian and Carroll (1987). Note that it has been recommended that a small sample correction be used (e.g., Breslow 1984; Dean 1994). The estimating equation (4.16) used in practice is then

$$g_\tau = \sum_{i=1}^{M}\frac{(n_{i+}-\mu_{i+})^2 - (1-h_i)\,\mu_{i+}(1+\tau\mu_{i+})}{(1+\tau\mu_{i+})^2},$$

where h_i is the i-th diagonal element of the leverage matrix, $W^{1/2}X'(X'WX)^{-1}XW^{1/2}$, with $W = \mathrm{diag}(\mu_i; i = 1,\ldots,M)$. This correction has no effect on the asymptotic properties of the estimator, which will be developed directly from (4.16).

The estimating functions retain the structure suggested by the factored likelihood, (4.5). However, we replace any contributions from L_{MP} by the quasi-likelihood estimating functions for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

Results of Inagaki (1973) and White (1982) show that under suitable regularity conditions, $\sqrt{M}(\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta})$ is asymptotically normal as $M \to \infty$, with asymptotic variance given by

$$\mathrm{E}\left(-\lim_{M\to\infty}\frac{1}{M}\frac{\partial\boldsymbol{g}_t}{\partial\boldsymbol{\theta}'}\right)^{-1}\mathrm{E}\left(\lim_{M\to\infty}\frac{1}{M}\boldsymbol{g}_t\boldsymbol{g}_t'\right)\mathrm{E}\left(-\lim_{M\to\infty}\frac{1}{M}\frac{\partial\boldsymbol{g}_t'}{\partial\boldsymbol{\theta}}\right)^{-1}. \tag{4.17}$$

The regularity conditions are similar to those for maximum likelihood asymptotics and require that the matrix (4.17) tend to a positive definite limit.

Let

$$A = \mathrm{E}\left(-\frac{\partial\boldsymbol{g}_t}{\partial\boldsymbol{\theta}'}\right) = \begin{pmatrix} A_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{21} & a_{22} \end{pmatrix}$$

where A has been partitioned according to the dimensions of $\gamma = (\boldsymbol{\beta}', \alpha)'$ and τ. The subscript t has been suppressed for ease of exposition. Denote the inverse of A as

$$A^{-1} = \begin{pmatrix} A^{11} & \boldsymbol{a}^{12} \\ \boldsymbol{a}^{21} & a^{22} \end{pmatrix}.$$

Let

$$B = \mathrm{E}\,(\boldsymbol{g}_t \boldsymbol{g}_t') = \begin{pmatrix} B_{11} & \boldsymbol{b}_{12} \\ \boldsymbol{b}_{21} & b_{22} \end{pmatrix},$$

where the matrices A^{-1} and B have been partitioned in the same way as A. Assuming that the mean has been correctly specified (i.e., assuming that $\mathrm{E}\,(n_{i+}) = \mu_{i+}$) then $\boldsymbol{a}_{12} = \mathbf{0}$ and $A_{11} = B_{11}$. Then the inverse of A can be shown to be

$$A^{-1} = \left(\begin{array}{c|c} A_{11}^{-1} & \mathbf{0} \\ \hline -a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1} & a_{22}^{-1} \end{array} \right).$$

As a result, when the mean and variance of the marginal counts, n_{i+}, have been correctly specified, the asymptotic variance, (4.17), simplifies to

$$\lim_{M\to\infty} M \left(\begin{array}{c|c} A_{11}^{-1} & (-a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1})' + A_{11}^{-1}\boldsymbol{b}_{12}a_{22}^{-1} \\ \hline -a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1} + a_{22}^{-1}\boldsymbol{b}_{21}(A_{11}^{-1})' & \begin{array}{c} -a_{22}^{-1}\boldsymbol{a}_{21}(-a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1})' + a_{22}^{-1}\boldsymbol{b}_{21}(-a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1})' \\ + - a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1}\boldsymbol{b}_{12}a_{22}^{-1} + a_{22}^{-1}b_{22}a_{22}^{-1} \end{array} \end{array} \right).$$

Assuming that the third moments of the distribution of the counts, n_{i+}, match those of a negative binomial, then $\boldsymbol{a}_{21} = \boldsymbol{b}_{21} = \boldsymbol{b}_{12}'$ and this leads to further simplification of (4.17),

$$\lim_{M\to\infty} M \begin{pmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0}' & -a_{22}^{-1}\boldsymbol{a}_{21}A_{11}^{-1}\boldsymbol{a}_{21}'a_{22}^{-1} + a_{22}^{-1}b_{22}a_{22}^{-1} \end{pmatrix}.$$

Assuming, finally, that fourth moments of the counts, n_{i+}, match those of a negative binomial random variable, then $a_{22} = b_{22}$, and the lower-right element of the matrix becomes $a_{22}^{-2}(a_{22} - \boldsymbol{a}_{21}A_{11}^{-1}\boldsymbol{a}_{21}')$. So, using the estimating equations $\boldsymbol{g}_\beta, g_{\alpha t}$, and g_τ and assuming that the first through fourth moments of the distribution of counts matches that of a negative binomial distribution, the asymptotic variance of $\sqrt{M}(\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta})$ is given by $\lim_{M\to\infty} M\,\mathrm{AsVar}(\tilde{\theta}_t)$ where

$$\mathrm{AsVar}(\tilde{\theta}_t) = \begin{pmatrix} A_t^{-1} & \mathbf{0} \\ \mathbf{0}' & v_\tau \end{pmatrix} \tag{4.18}$$

with

$$v_\tau = a_{22}^{-2}(a_{22} - \boldsymbol{a}_{21}A_{11}^{-1}\boldsymbol{a}_{21}'). \tag{4.19}$$

The expression A_{11} has been written here as A_t to indicate that this term is calculated based on the event-time data. Note that A_t is also the Fisher information for the likelihood estimator of $\boldsymbol{\gamma}_t' = (\boldsymbol{\beta}_t', \alpha_t)$, and also note, from the derivation above, that only first and

second moment assumptions are required for determining the asymptotic variance of $\tilde{\gamma}_t$. Here,

$$A_t = \mathrm{E}\left(\frac{-\partial \boldsymbol{g}_{\gamma t}}{\partial \gamma'}\right) = \begin{pmatrix} X'WV_o^{-1}WX & X'WV_o^{-1}WZ \\ Z'WV_o^{-1}WX & Z'WV_o^{-1}WZ + H_t \end{pmatrix}$$

where $W = \mathrm{diag}(\mu_{i+}; i = 1, \ldots M), V_o = \mathrm{diag}(\mu_{i+}(1 + \tau\mu_{i+}); i = 1, \ldots, M)$, and the matrix Z has ij-th entry

$$z_{ij} = \frac{\partial}{\partial \alpha}\Lambda_0(T_{ie}), \quad i = 1, \ldots, M.$$

In the lower right corner of this matrix A_t is the expression for the Fisher information for α from the end-of-follow-up count data, (4.13) plus the additional information available from the event-time data,

$$H_t = \sum_{i=1}^{M} \mathrm{E}\left(\sum_{j=1}^{s}\sum_{l=1}^{n_{ij}} -\frac{\partial^2 \log[\lambda(w_{ijl}; \alpha)/\Lambda(T_{s_i})]}{\partial \alpha^2}\right)$$

For the Weibull intensity function, $H_t = -\sum_{i=1}^{M} \mu_{i+}/\alpha^2$. (For further details regarding the derivation of these asymptotic variances, refer to Section 3.5.)

A finite sample variance estimate for $\tilde{\boldsymbol{\theta}}_t$ is obtained removing the limit arguments and M^{-1}, substituting $\tilde{\boldsymbol{\theta}}_t$ for $\boldsymbol{\theta}$ in $\mathrm{AsVar}(\tilde{\boldsymbol{\theta}}_t)$, (4.18); this would be a model-based estimator of variance. Replacing the middle term in (4.17), $\mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t')$, by $\boldsymbol{g}_t\boldsymbol{g}_t'$ evaluated at $\tilde{\boldsymbol{\theta}}_t$ results in the so-called "sandwich" variance estimate (Liang and Zeger 1986). The sandwich estimator provides an empirical estimate of the variance, and hence it is robust to misspecification of the assumed variance of the n_{i+}'s for purposes of estimating the variance of γ_t. The sandwich estimator may underestimate the asymptotic variance in small samples (Breslow 1990); however, because we consider scenarios involving fairly large numbers of individuals followed for relatively short periods of time, the use of this variance estimator is likely to be appropriate here. We evaluate its performance in Section 4.2.4 for situations which typify this type of study.

4.2.2 Estimation Based on Panel Data

Based on panel data, consisting of n_{ij}'s, the likelihood contribution from individual i is the product of the probability of observing n_{i+} events and the probability that n_{i1} events occurred in the first follow-up period, n_{i2} in the second, etc. The likelihood is therefore

$$L_p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = \prod_{i=1}^{M}\left\{\binom{n_{i+}}{n_{i1}\cdots n_{is_i}}\prod_{j=1}^{s_i}\left(\frac{\mu_{ij}}{\mu_{i+}}\right)^{n_{ij}}\right\} L_{\mathrm{MP}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) \qquad (4.20)$$

where the subscript p signifies panel data and $L_{\mathrm{MP}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ is given in (4.7).

First and second derivatives of L_{p} with respect to $\boldsymbol{\beta}$ and τ are the same as for L_{t}, (4.5). Assuming a Weibull baseline intensity and gamma distributed random effects, the derivative and expected information with respect to α are

$$\frac{\partial \log L_{\mathrm{p}}}{\partial \alpha} = \sum_{i=1}^{M} \sum_{j=1}^{s_i} n_{ij} \left\{ \frac{T_j^{\alpha} \log T_j - T_{j-1}^{\alpha} \log T_{j-1}}{T_j^{\alpha} - T_{j-1}^{\alpha}} - \log T_{s_i} \right\} + \frac{\partial \log L_{\mathrm{NB}}}{\partial \alpha} \tag{4.21}$$

and

$$\begin{aligned} \mathrm{E}\left(-\frac{\partial^2 \log L_{\mathrm{p}}}{\partial^2 \alpha}\right) = & \sum_{i=1}^{M} \sum_{j=1}^{s_i} \mu_{ij} \left\{ \left(\frac{T_j^{\alpha} \log T_j - T_{j-1}^{\alpha} \log T_{j-1}}{T_j^{\alpha} - T_{j-1}^{\alpha}} \right)^2 \right. \\ & \left. - \frac{T_j^{\alpha} (\log T_j)^2 - T_{j-1}^{\alpha} (\log T_{j-1})^2}{T_j^{\alpha} - T_{j-1}^{\alpha}} \right\} \\ & + \mathrm{E}\left(-\frac{\partial^2 \log L_{\mathrm{NB}}}{\partial^2 \alpha}\right) \end{aligned}$$

where $\partial \log L_{\mathrm{NB}}/\partial \alpha$ and $\mathrm{E}\,(-\partial^2 \log L_{\mathrm{NB}}/\partial \alpha^2)$ are given in (4.11) and (4.13).

Quasi-likelihood estimation follows in a straightforward manner from Section 4.2.1. We replace the contribution from L_{MP} in the estimating functions by quasi-likelihood and pseudo-likelihood estimating functions. The estimating equation becomes $\boldsymbol{g}_{\mathrm{p}} = \mathbf{0}$, $\boldsymbol{g}_{\mathrm{p}}' = (\boldsymbol{g}_{\beta}', g_{\alpha p}, g_{\tau})$, where $\boldsymbol{g}_{\beta}$ and g_{τ} are given in (4.14) and (4.16); and, we use for $g_{\alpha p}$ the likelihood estimating function $\partial \log L_{\mathrm{p}}/\partial \alpha$ given in (4.21). The resulting estimator is denoted $\tilde{\boldsymbol{\theta}}_{\mathrm{p}}$, and $\sqrt{M}(\tilde{\boldsymbol{\theta}}_{\mathrm{p}} - \boldsymbol{\theta})$ is asymptotically normally distributed with asymptotic variance (4.17), replacing $\boldsymbol{g}_{\mathrm{t}}$ with $\boldsymbol{g}_{\mathrm{p}}$. Conditions for this asymptotic property are given in Inagaki (1973) and White (1982). Finite sample variance estimates are obtained in the same manner as for $\tilde{\boldsymbol{\theta}}_{\mathrm{t}}$, again using the sandwich variance estimator if desired.

The asymptotic variance of $\sqrt{M}(\tilde{\boldsymbol{\theta}}_{\mathrm{p}} - \boldsymbol{\theta})$ based on a Weibull intensity and gamma distributed random effects has the same form as for $\sqrt{M}(\tilde{\boldsymbol{\theta}}_{\mathrm{t}} - \boldsymbol{\theta})$ with A_{t} replaced by A_{p}; A_{p} is identical to A_{t} except that H_{t} is replaced with

$$H_{\mathrm{p}} = \sum_{i=1}^{M} \sum_{j=1}^{s_i} \mu_{ij} \left\{ \left(\frac{T_j^{\alpha} \log T_j - T_{j-1}^{\alpha} \log T_{j-1}}{T_j^{\alpha} - T_{j-1}^{\alpha}} \right)^2 - \frac{T_j^{\alpha} (\log T_j)^2 - T_{j-1}^{\alpha} (\log T_{j-1})^2}{T_j^{\alpha} - T_{j-1}^{\alpha}} \right\} \tag{4.22}$$

This term characterizes the information available in the panel data which was not available in the end-of-follow-up count data, in the same way that H_{t} characterizes the extra information available in the event-time data.

4.2.3 The Efficiency of Quasi-Likelihood Relative to Likelihood Estimation

In the previous section, we observed that the quasi-likelihood estimating functions $\boldsymbol{g}_\beta$ and g_α are the same as the likelihood estimating functions for the negative binomial model. Only the functions for τ differ. Also, the asymptotic variance matrices are the same, except for the asymptotic variances for the estimators of τ. As a result, we conclude that when the data are generated by the negative binomial model, the quasi-likelihood estimators of $\boldsymbol{\beta}$ and α are fully efficient relative to the likelihood estimators. However, because the quasi-likelihood estimating function for τ is not the negative binomial likelihood estimating function, the quasi-likelihood estimator $\tilde{\tau}$ has a larger asymptotic variance than the likelihood estimator $\hat{\tau}$.

Let

$$\text{ARE}(\tilde{\tau}) = \frac{\text{AsVar}(\hat{\tau})}{\text{AsVar}(\tilde{\tau})}$$

be the asymptotic relative efficiency of the quasi-likelihood estimator of τ. To explore conditions when $\text{ARE}(\tilde{\tau})$ will be high we have conducted a small numerical study. Several scenarios were examined, each with two treatment groups of equal size, but with a variety of parameter settings. Note that the asymptotic variance of the likelihood estimator, $\hat{\tau}$, does not depend on whether we examine event-time, panel, or count data; the same is true for the quasi-likelihood estimator, $\tilde{\tau}$. Thus, these results apply to all three follow-up structures. Parameter values used in this study were chosen to be similar to the estimates obtained in Section 4.5 in the analysis of the Bladder Cancer data.

Table 4.1 and Figure 4.1 summarize the study. Comparing scenarios (a.1) and (a.2), we see the small influence of the treatment effect β_2 on the relative efficiency of $\tilde{\tau}$. Comparing scenario (a.1) with (b) and (c) reveals the more substantial effect of large changes in the cumulative means due to varying values of α. However, it is clear that the general pattern is for the $\text{ARE}(\tilde{\tau})$ to decrease as the cumulative means and/or the level of overdispersion increase. In other words, the quasi-likelihood estimate of τ is more efficient when $\tau\mu_{i+}$ is small.

				Group Means		ARE($\tilde{\tau}$)		
Scenario	β_1	β_2	α	μ_1	μ_2	$\tau = 0.2$	$\tau = 0.8$	$\tau = 1.5$
a.1	-2	0.1	0.7	2.701	2.985	0.9282	0.6773	0.5228
a.2	-2	0.5	0.7	2.701	4.453	0.9145	0.6503	0.4969
b	-2	0.1	1.0	9.744	10.769	0.8617	0.5533	0.4029
c	-2	0.1	1.3	35.151	38.848	0.8126	0.4895	0.3423

Table 4.1: Asymptotic Relative Efficiency of Quasi-Likelihood Estimator $\tilde{\tau}$. The efficiency of the quasi-likelihood estimator $\tilde{\tau}$ is calculated relative to the likelihood estimator $\hat{\tau}$ for three values of τ and for different values of β and α, with equal sized treatment groups. Also refer to Figure 4.1.

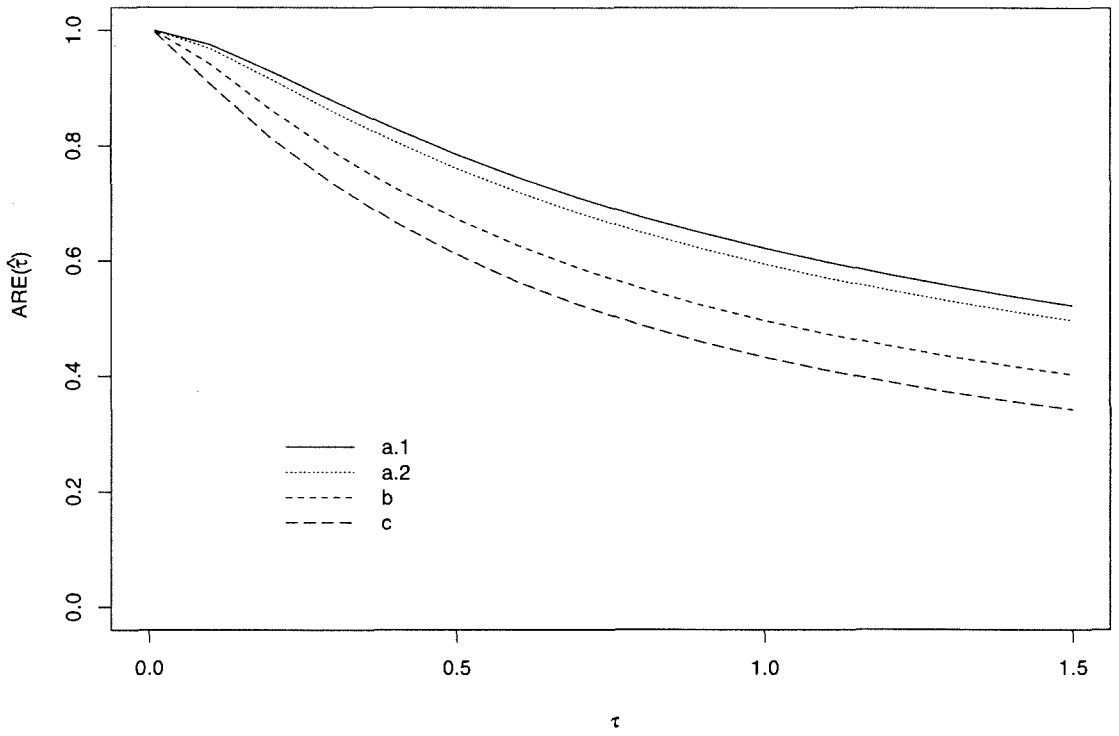

Figure 4.1: Asymptotic Relative Efficiency of Quasi-Likelihood Estimator $\tilde{\tau}$. See Table 4.1 for details.

4.2.4 Small Sample Characteristics of Likelihood and Quasi-Likelihood Estimators

A simulation study was undertaken to compare the performance of the likelihood and quasi-likelihood estimators, and how sample sizes and the type of recurrent event data affect this comparison.

The simulation was designed to mimic a two-group clinical trial, similar to the bladder cancer trial examined in Section 4.5. Event-time data were generated according to the NHPP model with intensity function $\lambda_i(t) = \nu_i \lambda_0(t;\alpha) e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$, as described in Section 4.2. Parameters were chosen to be similar to the estimates observed for the end-of-follow-up count data analysis of the bladder cancer data, namely $\beta_1 = -2, \beta_2 = 0.1, \alpha = 0.7$ and $\tau = 1.5$. The total follow-up time was 72 months, but to simulate the presence of dropouts, 25% of the subjects were eliminated after months 18, 36, and 54. The effect of sample size was examined by repeating the study twice with sample sizes of 48 and 96 subjects in each group. A total of 5000 replications of the simulation were conducted for each of the sample sizes. Panel data were obtained as if the counts were recorded at four follow-up times, months 18, 36, 54, and 72. End-of-follow-up counts were obtained as if the counts were recorded only once, at the end of their follow-up period.

Both likelihood and quasi-likelihood estimates were obtained from each dataset in all three forms: event-time, panel, and end-of-follow-up count data. That is, a total of six analyses was conducted for every simulated dataset. Model-based and robust variance estimates were calculated for the quasi-likelihood estimators. The usual inverses of the expected and observed information matrices were used in the analagous roles for the likelihood estimators.

As a robust (or "empirical") variance estimator of the maximum likelihood estimator, the usual inverse of the observed information matrix was used.

We summarize the following small sample characteristics:

(a) The accuracy of the normal approximation for the standardized estimators; where noteworthy, we also comment on the coverage properties of the approximate 95% confidence intervals (i.e., $\hat{\theta} \pm 1.96 \times \text{s.e.}(\hat{\theta})$), the relative bias of the estimators, and the accuracy of the two variance estimators. The standard errors based on the robust variances are used for standardization, but the results were found to be very similar if model-based variances were used.

Estimation Method	Group Size	Type of Data Event-time	Panel	Count
Likelihood	48	0	0	0.0046
	96	0	0	0
Quasi-likelihood	48	0.0010	0.0010	0.0072
	96	0	0	0

Table 4.2: Proportion of Simulated Datasets Yielding Estimates which Failed to Converge. A total of 5000 datasets were simulated for each sample size.

(b) The observed ratio of model-based variance estimates for the likelihood versus quasi-likelihood estimators — that is, the "observed" relative efficiency of the quasi-likelihood estimators.

The simulation study described above was intended to compare the behaviour of the likelihood and quasi-likelihood estimators at various sample sizes for a single combination of model parameters. Because the small sample behaviour of quasi-likelihood estimators has been studied less thoroughly than likelihood estimators, a more general simulation study of the quasi-likelihood estimator was also undertaken. This broader study considered all combinations of the parameters $\alpha = \{0.7, 1, 1.3\}$ and $\tau = \{0.8, 1.5\}$ for the same experimental design used in the comparative study. The parameter values were chosen to cover the range of estimates obtained in the various analyses of the bladder cancer data. For all combinations of α and τ in this more general study, the quasi-likelihood estimator performs in the same way as it does in the more restrictive study. Therefore only the comparative study is summarized below.

Table 4.2 summarizes the convergence rates of the iterative root-finding routine used for both likelihood and quasi-likelihood estimation. The rates are broken down by method of analysis and type of recurrent event data. Overall, convergence rates were high, with fewer than 0.8% failing to converge. With the random effects following a gamma distribution, it is necessary to approximate an infinite series in the calculation of the expected information for the likelihood estimate of τ (Lawless 1987a, eqn. (2.8)). Because of this, the likelihood method was more time consuming during this simulation. Determining when the series had converged to a reasonable approximation required careful consideration of the relative size of the contribution of each term in the series. However, the time required for this calculation

		End-of-Follow-Up Count Data Estimation Method			
Parameter	Group	Likelihood		Quasi-likelihood	
(true value)	Size	Mean	(Std. Err.)	Mean	(Std. Err.)
β_1	48	-2.0993	(1.1146)	-2.1012	(1.1127)
(-2)	96	-2.0365	(0.7771)	-2.0365	(0.7772)
β_2	48	0.1039	(0.2968)	0.1042	(0.2968)
(0.1)	96	0.0976	(0.2090)	0.0976	(0.2090)
α	48	0.7185	(0.2917)	0.7191	(0.2915)
(0.7)	96	0.7063	(0.2050)	0.7063	(0.2049)
τ	48	1.4324	(0.3562)	1.4383	(0.4232)
(1.5)	96	1.4729	(0.2493)	1.4670	(0.3133)

Table 4.3: Simulation Results for Estimation of $\boldsymbol{\theta}$ Based on End-of-Follow-Up Count Data. The standard errors (Std. Err.) have been calculated as the square roots of the simulated variances of the estimators.

should not be a factor for routine use.

The comparison of the likelihood and quasi-likelihood estimators did not appear to depend on whether the event-time, panel, and end-of-follow-up count data were examined, so unless otherwise noted, summaries are provided here for the analyses of the end-of-follow-up count data only. Also, since Section 4.3 focuses on the efficiency of the panel and end-of-follow-up count data analyses, we will postpone a direct comparison of the panel and count data estimators until then.

Figures 4.2 through 4.5 show normal probability plots for the standardized likelihood and quasi-likelihood estimators. It is clear that the normal approximation is adequate for both likelihood and quasi-likelihood estimators of β_1, β_2, and α, but inadequate for estimators of τ whose distributions are negatively skewed. The normal approximation appears to be slightly better for the likelihood estimators than for the quasi-likelihood estimators, though less so for larger samples ($m_i = 96$).

Tables 4.3 through 4.5 show the means and standard errors of the estimators of $\boldsymbol{\theta}$. The standard errors used are the square roots of the sampling variances of the estimators in the simulation. In general, the relative bias for the estimators is quite low, less than 3% for all

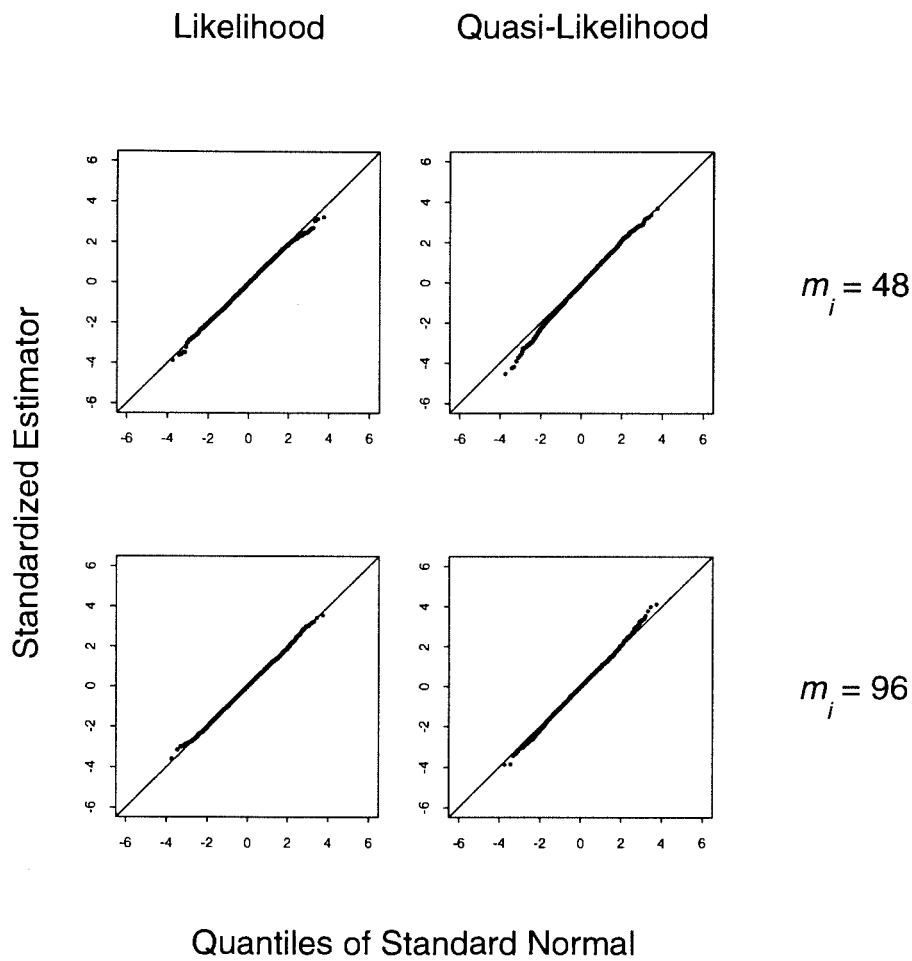

Figure 4.2: Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of β_1. A total of 5000 datasets were simulated according to the negative-binomial mixture with a Weibull intensity function with parameters $\beta_1 = -2, \beta_2 = 0.1, \alpha = 0.7, \tau = 1.5$. The standardized estimates are the estimates divided by their robust standard errors. The reference line corresponds to the standard normal distribution.

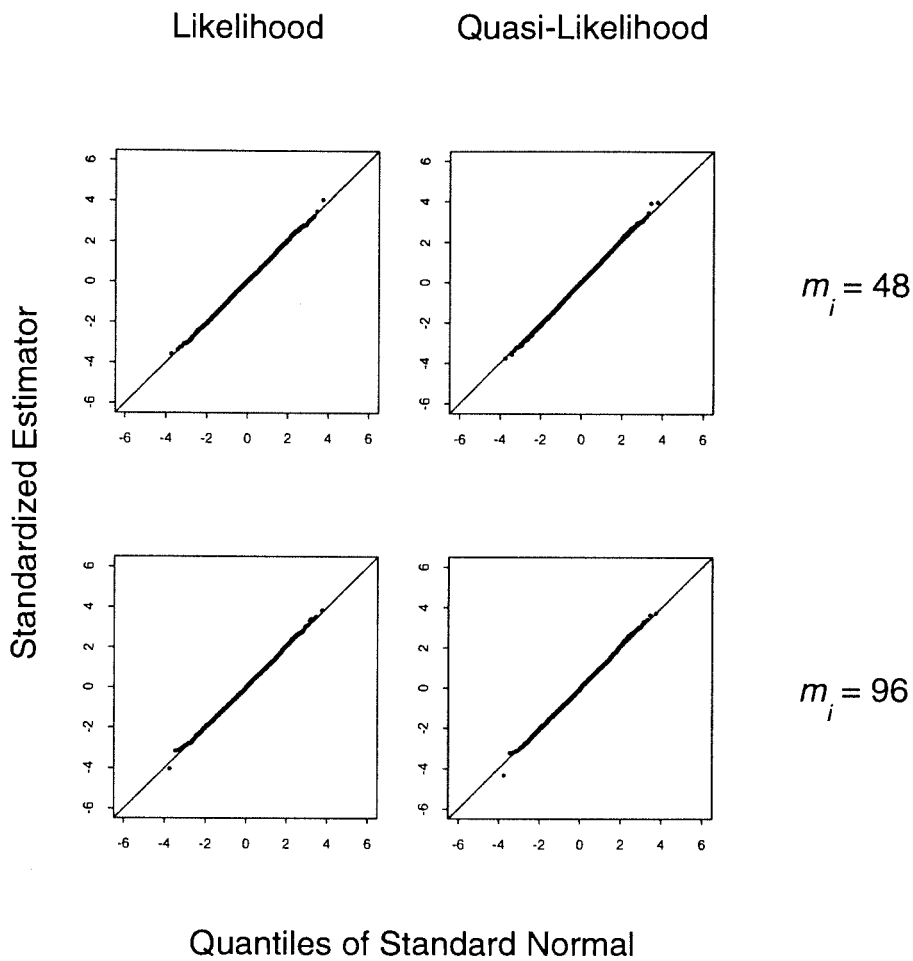

Figure 4.3: Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of β_2. A total of 5000 datasets were simulated according to the negative-binomial mixture with a Weibull intensity function with parameters $\beta_1 = -2, \beta_2 = 0.1, \alpha = 0.7, \tau = 1.5$. The standardized estimates are the estimates divided by their robust standard errors. The reference line corresponds to the standard normal distribution.

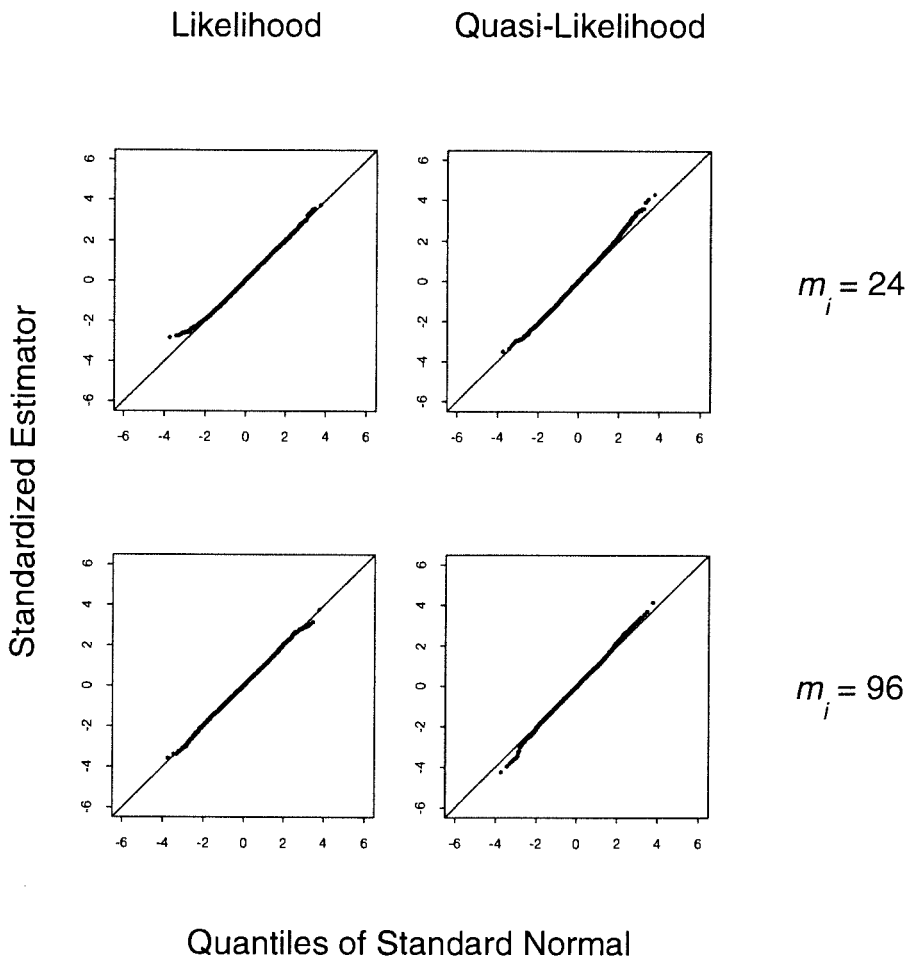

Figure 4.4: Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of α. The estimators were calculated from the end-of-follow-up counts. A total of 5000 datasets were simulated according to the negative-binomial mixture with a Weibull intensity function with parameters $\beta_1 = -2, \beta_2 = 0.1, \alpha = 0.7, \tau = 1.5$. The standardized estimates are the estimates divided by their robust standard errors. The reference line corresponds to the standard normal distribution.

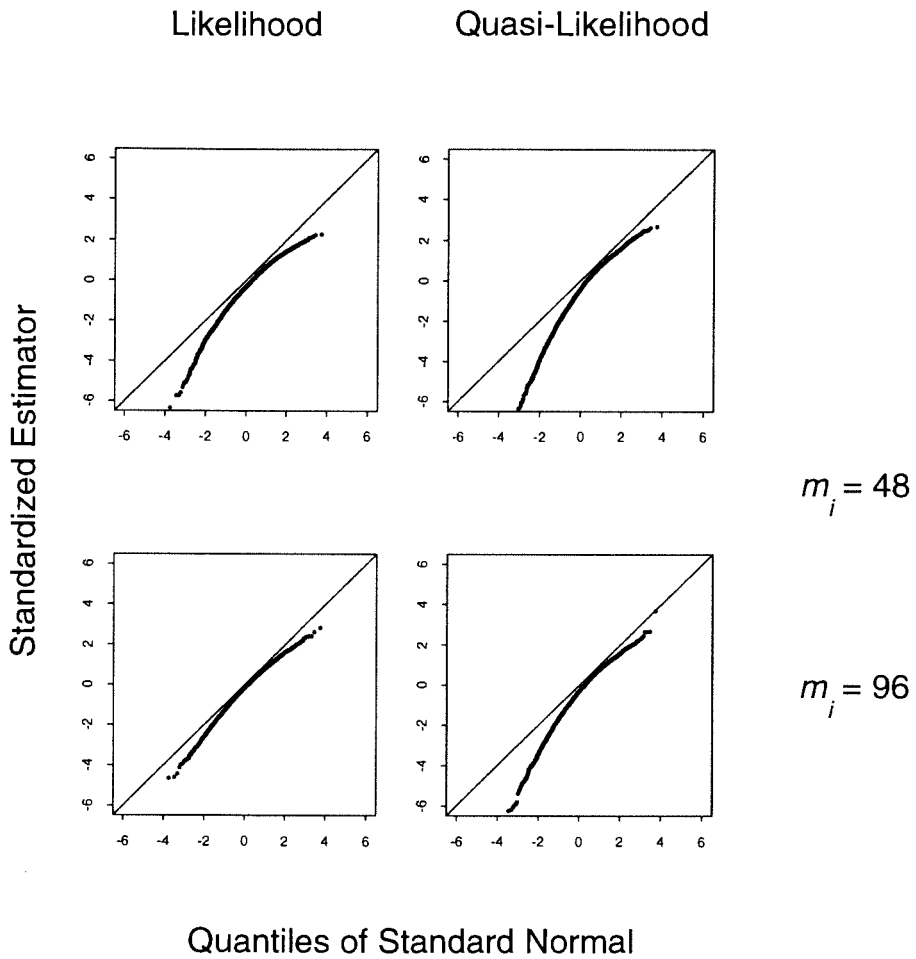

Figure 4.5: Simulated Distributions of Likelihood and Quasi-Likelihood Estimators of τ. A total of 5000 datasets were simulated according to the negative-binomial mixture with a Weibull intensity function with parameters $\beta_1 = -2, \beta_2 = 0.1, \alpha = 0.7, \tau = 1.5$. The standardized estimates are the estimates divided by their robust standard errors. The reference line corresponds to the standard normal distribution.

		Panel Data Estimation Method			
Parameter	Group	Likelihood		Quasi-likelihood	
(true value)	Size	Mean	(Std. Err.)	Mean	(Std. Err.)
β_1	48	-2.0271	(0.3419)	-2.0266	(0.3419)
(-2)	96	-2.0103	(0.2391)	-2.0101	(0.2391)
β_2	48	0.1035	(0.2954)	0.1036	(0.2953)
(0.1)	96	0.0977	(0.2084)	0.0977	(0.2084)
α	48	0.7018	(0.0714)	0.7017	(0.0715)
(0.7)	96	0.7006	(0.0510)	0.7005	(0.0510)
τ	48	1.4583	(0.3608)	1.4805	(0.4438)
(1.5)	96	1.4859	(0.2504)	1.4889	(0.3192)

Table 4.4: Simulation Results for Estimation of $\boldsymbol{\theta}$ Based on Panel Data. The standard errors (Std. Err.) have been calculated as the square roots of the simulated variances of the estimators.

		Event-Time Data Estimation Method			
Parameter	Group	Likelihood		Quasi-likelihood	
(true value)	Size	Mean	(Std. Err.)	Mean	(Std. Err.)
β_1	48	-2.0344	(0.2832)	-2.0340	(0.2831)
(-2)	96	-2.0160	(0.1966)	-2.0160	(0.1967)
β_2	48	0.1034	(0.2953)	0.1034	(0.2953)
(0.1)	96	0.0977	(0.2084)	0.0977	(0.2084)
α	48	0.7038	(0.0502)	0.7037	(0.0503)
(0.7)	96	0.7022	(0.0348)	0.7021	(0.0348)
τ	48	1.4593	(0.3609)	1.4812	(0.4425)
(1.5)	96	1.4863	(0.2505)	1.4894	(0.3192)

Table 4.5: Simulation Results for Estimation of $\boldsymbol{\theta}$ Based on Event-Time Data. The standard errors (Std. Err.) have been calculated as the square roots of the simulated variances of the estimators.

estimators except those for τ. Both the likelihood $\hat{\tau}$ and quasi-likelihood $\tilde{\tau}$ underestimated τ by up to 10%. The bias of all estimators decreases with increased sample sizes.

Table 4.6 summarizes the accuracy and precision of the various variance estimators. The standard deviations shown on this table are the square roots of the sampling variance of the model-based and robust variance estimates. The simulated variance is simply the sampling variance of the estimators of $\boldsymbol{\theta}$ in the simulation study. The standard deviations reported in the table are the square roots of the sampling variances of the model-based and robust variance estimates. The standard deviations of the robust variance estimates are generally larger than the standard deviations of the model-based variance estimates. This is the cost of making less stringent modeling assumptions. Only for the quasi-likelihood estimator of τ is this reversed, but this can be explained by the tendency of $\tilde{\tau}$ to underestimate the true value of τ. For larger samples, the bias in $\tilde{\tau}$ has been overcome, and the robust variance is again more variable than the model-based variance.

Asymptotic normal confidence interval coverage probabilities are presented in Table 4.7. The confidence intervals were calculated as the estimate plus or minus 1.96 times the standard error of the estimate. The standard errors were based on the robust variance estimates. In general, both likelihood and quasi-likelihood yielded estimators with good coverage probabilities for the larger samples ($m_i = 96$), ranging between 94% and 96% for all estimators except $\hat{\tau}$ and $\tilde{\tau}$. The likelihood estimator, $\hat{\tau}$, performed much better than the quasi-likelihood, $\tilde{\tau}$; for the larger sample size ($m_i = 96$), the likelihood interval achieved coverage of 93% compared to only 88% for the quasi-likelihood interval. With the smaller sample size ($m_i = 48$), the intervals based on the quasi-likelihood estimators appeared to have poorer coverage, approximately 93% for parameters other than τ. The corresponding likelihood intervals were more reliable here, with coverage approximately 95%.

Table 4.8 compares the accuracy of the estimators of the variance-covariance matrices. Overall, the likelihood variance estimators were more accurate than the quasi-likelihood estimators. This is not surprising given that the data were simulated according to the likelihood model. Also, the model-based variance estimators were more accurate than the robust variance estimators, though for the likelihood estimators there was very little advantage to the model-based estimators for large samples. The variance estimators for $\tilde{\tau}$ were the least accurate. Even for the larger sample size, the model-based variance of the quasi-likelihood estimator $\tilde{\tau}$ overestimated the sampling variance by about 14% and the robust variance estimator underestimated by about 16%. In contrast, for the same sample size, the model-based

			Estimation Method			
	Group	Variance	Likelihood		Quasi-likelihood	
Parameter	Size	Type	Mean	(Std. Dev.)	Mean	(Std. Dev.)
β_1	48	Simulated	1.2424		1.2381	
		Model-based	1.2098	(0.2258)	1.2128	(0.2583)
		Robust	1.2358	(0.2625)	1.1284	(0.4066)
	96	Simulated	0.6039		0.6040	
		Model-based	0.6077	(0.0766)	0.6061	(0.0927)
		Robust	0.6140	(0.0867)	0.5865	(0.1563)
β_2	48	Simulated	0.0881		0.0881	
		Model-based	0.0838	(0.0157)	0.0841	(0.0183)
		Robust	0.0848	(0.0161)	0.0806	(0.0176)
	96	Simulated	0.0437		0.0437	
		Model-based	0.0424	(0.0054)	0.0423	(0.0067)
		Robust	0.0427	(0.0055)	0.0415	(0.0066)
α	48	Simulated	0.0852		0.0850	
		Model-based	0.0832	(0.0154)	0.0834	(0.0178)
		Robust	0.0850	(0.0179)	0.0780	(0.0270)
	96	Simulated	0.0420		0.0420	
		Model-based	0.0419	(0.0053)	0.0418	(0.0064)
		Robust	0.0423	(0.0060)	0.0404	(0.0103)
τ	48	Simulated	0.1269		0.1791	
		Model-based	0.1221	(0.0547)	0.2293	(0.1665)
		Robust	0.1233	(0.0556)	0.1322	(0.1384)
	96	Simulated	0.0622		0.0981	
		Model-based	0.0615	(0.0184)	0.1116	(0.0561)
		Robust	0.0619	(0.0187)	0.0820	(0.0824)

Table 4.6: Accuracy and Precision of Variance Estimators. These results are based on the analysis of end-of-follow-up count data. Similar results were observed for the analysis of panel and event-time data. Simulated variance is the sampling variance of the estimator; standard deviations (Std. Dev.) are the square roots of the sampling variances of the model-based and robust variance estimates.

Data Type	Estimation Method	Group Size	Parameter β_1	β_2	α	τ
Count	Quasi-likelihood	48	0.927	0.936	0.928	0.833
		96	0.938	0.941	0.937	0.875
	Likelihood	48	0.953	0.942	0.954	0.903
		96	0.953	0.946	0.949	0.930
Panel	Quasi-likelihood	48	0.938	0.938	0.937	0.852
		96	0.948	0.943	0.944	0.884
	Likelihood	48	0.956	0.941	0.955	0.913
		96	0.954	0.947	0.950	0.936
Event-time	Quasi-likelihood	48	0.936	0.937	0.934	0.852
		96	0.945	0.943	0.946	0.884
	Likelihood	48	0.949	0.941	0.951	0.915
		96	0.953	0.947	0.955	0.936

Table 4.7: Observed Coverage of Nominal 95% Large-Sample Confidence Intervals. The confidence intervals have been calculated only for the estimates which converged using standard errors based on the robust variance estimates.

	Ratio of Robust Variance to Sampling Variance							
	Quasi-Likelihood Estimators				Likelihood Estimators			
Group Size	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\alpha}$	$\tilde{\tau}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\tau}$
48	0.911	0.915	0.918	0.738	0.995	0.962	0.998	0.972
96	0.971	0.949	0.962	0.835	1.017	0.976	1.008	0.996

	Ratio of Model-Based Variance to Sampling Variance							
	Quasi-Likelihood Estimators				Likelihood Estimators			
Group Size	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\alpha}$	$\tilde{\tau}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\tau}$
48	0.980	0.954	0.982	1.281	0.974	0.952	0.976	0.962
96	1.003	0.969	0.994	1.137	1.006	0.972	0.997	0.990

Table 4.8: Ratio of Estimated Variances to Sampling Variance for End-of-Follow-Up Count Data. Similar results were observed for the analysis of panel and event-time data. Empirical Variance and Model-Based Variance are the means of the variance estimates calculated in the simulation study. Sampling Variance is the observed sampling variance of the parameter estimates. Similar results were observed for the panel and event-time data.

Observed ARE($\tilde{\theta}$) from End-of-Follow-Up Count Data

Group Size	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\alpha}$	$\tilde{\tau}$
End-of-Follow-Up Count Data				
48	1.0035	0.9998	1.0032	0.7086
96	0.9998	1.0001	0.9997	0.6334
Theoretical	1	1	1	0.5654
Panel Data				
48	0.9996	1.0004	0.9981	0.6609
96	0.9997	1.0001	0.9994	0.6154
Theoretical	1	1	1	0.5654
Event-Time Data				
48	1.0002	1.0005	0.9990	0.6652
96	0.9997	1.0001	0.9996	0.6157
Theoretical	1	1	1	0.5654

Table 4.9: Estimated ARE($\tilde{\theta}$) from Simulation Study. The efficiency of the quasi-likelihood estimator relative to the likelihood estimator is calculated as the ratio of the simulated variance of $\hat{\theta}$ to the simulated variance of $\tilde{\theta}$.

and robust variance of the likelihood estimator $\hat{\tau}$ were within 1% of the sampling variance.

Table 4.9 displays the estimated relative efficiencies for the quasi-likelihood estimators. By estimated relative efficiency, we mean the ratio of the sampling variances of the likelihood estimator to the sampling variance of the quasi-likelihood estimator. The table shows that the estimated ARE($\tilde{\theta}$) closely approximates the theoretical ARE($\tilde{\theta}$).

In summary, the findings of our simulation study of the small sample behaviour of the estimators suggest that the quasi-likelihood estimators have no serious small-sample drawbacks compared to the likelihood estimators for inference regarding parameters other than τ. However, since the quasi-likelihood methods require fewer distributional assumptions than the likelihood estimators, we will restrict our attention to the quasi-likelihood estimators in later sections where we investigate the efficiency of panel data relative to event-time data.

4.3 Efficiency of Panel Data vs. Count Data

One of the key findings of the previous chapter was that even though end-of-follow-up count data can be very efficient for the estimation of covariate effects, the same is not true for the other parameters. Unfortunately, the intensity parameters α and β_1 (from the Weibull intensity model described earlier) will be poorly estimated, with relative efficiencies of 10% and below for some scenarios. We now examine how well panel data recovers this lost information. We find that the panel data estimates of the intensity parameters become much more efficient for even a small number of follow-up times.

We first conduct a large-sample comparison by computing the ARE($\tilde{\beta}_{r\mathrm{p}}$), the asymptotic efficiency of the estimators of β_r based on panel data, $\tilde{\beta}_{r\mathrm{p}}$, relative to the estimators based on event-time data, $\tilde{\beta}_{r\mathrm{t}}$. We then compare the ARE($\tilde{\beta}_{r\mathrm{c}}$) to the ARE($\tilde{\beta}_{r\mathrm{p}}$). We also consider the effect of dropouts on the efficiency of the estimators, i.e., subjects or systems that are lost to follow-up before the final follow-up time T_s. Dropouts are an almost inevitable complication of clinical trials and other recurrent event studies, and we show that the efficiency improvements due to panel data are relatively unaffected by dropouts.

We then consider the small-sample results by returning to the simulation study described in Section 4.2.4. and comparing the observed sampling variances of the count and panel data estimators.

Note that since the quasi-likelihood estimators of the covariate effects $\boldsymbol{\beta}$ are asymptotically 100% efficient relative to their likelihood counterparts (*cf.* Section 4.2.3), we will examine only the quasi-likelihood estimators in this section.

4.3.1 Asymptotic Comparison of Panel Data vs. Count Data

As might be expected given the similarity in the estimating equations for count and panel data analyses, our findings here for panel data are similar to those for count data in the previous chapter in that under certain conditions covariate effects representing treatments can be estimated with 100% efficiency. To examine these conditions, we restrict ourselves to the same situation as in Chapter 3, where the covariates are indicator variables for the treatments. That is, $x_{i1} = 1$ for all subjects, and

$$x_{ir} = \begin{cases} 1 & \text{if individual } i \text{ received treatment } r, \\ 0 & \text{otherwise} \end{cases}$$

for $r = 2, \ldots, k$. As in the previous chapter, we adopt the following notation. Let G_r index the set of individuals who received treatment r, $G_r = \{$ all $i \ni$ individual i received treatment $r\}$, $r = 1, \ldots, k$. To simplify notation, let $[a]_{r+}$ denote $\sum_{i \in G_r} a_i$, so $[n]_{1+} = \sum_{i \in G_1} n_i$, the total number of events observed for all individuals receiving treatment 1; let $[ab]_{r+}$ denote $\sum_{i \in G_r} a_i b_i$, etc., and for grand totals, let $[a]_+$ denote $\sum_{i=1}^{M} a_i$, $[ab]_+$ denote $\sum_{i=1}^{M} a_i b_i$, etc. Also, all occurrences of μ in this notation represent marginal means, $\mu_{i+} = \sum_{j=1}^{s_i} \mu_{ij}$.

Using this notation, the following result gives the asymptotic relative efficiency of $\tilde{\beta}_{\mathrm{rp}}$.

Result 7 *The asymptotic relative efficiency of $\tilde{\beta}_{\mathrm{rp}}$ relative to $\tilde{\beta}_{\mathrm{rt}}$, $r = 2, \ldots, k$, is*

$$\mathrm{ARE}(\tilde{\beta}_{\mathrm{rp}}) = 1 - \left\{ \frac{\ell_{\mathrm{p}r}}{\left(\left[\frac{\mu}{1+\tau\mu} \right]_{r+}^{-1} + \left[\frac{\mu}{1+\tau\mu} \right]_{1+}^{-1} \right) E_{\mathrm{o}} + l_{\mathrm{p}r}} \right\} \left(\frac{H_{\mathrm{p}}}{E_{\mathrm{o}} + H_{\mathrm{p}}} \right),$$

where

$$\ell_{\mathrm{p}r} = \left\{ \frac{\left[\frac{\mu}{1+\tau\mu} \frac{\partial \log \Lambda(T_s)}{\partial \alpha} \right]_{r+}}{\left[\frac{\mu}{1+\tau\mu} \right]_{r+}} - \frac{\left[\frac{\mu}{1+\tau\mu} \frac{\partial \log \Lambda(T_{s_i})}{\partial \alpha} \right]_{r+}}{\left[\frac{\mu}{1+\tau\mu} \right]_{1+}} \right\}^2, \tag{4.23}$$

and

$$E_{\mathrm{p}} = \sum_{r=1}^{k} \sum_{i \in G_r} \frac{\mu_{i+}}{1 + \tau\mu_{i+}} \left(\frac{\partial \log \Lambda_0(T_{s_i})}{\partial \alpha} - \frac{\left[\frac{\mu}{1+\tau\mu} \frac{\partial \log \Lambda_0}{\partial \alpha} \right]_{r+}}{\left[\frac{\mu}{1+\tau\mu} \right]_{r+}} \right)^2.$$

The term H_{p} was defined in (4.22).

This is obtained from the asymptotic variance (4.18) following algebraic manipulations similar to those in Section 3.5.2. The same manipulations yield the asymptotic relative efficiencies of $\tilde{\alpha}_{\mathrm{p}}$ and $\tilde{\beta}_{1\mathrm{p}}$. These results are of the same form as those for the count data (Results 5 and 6), with H_{p} replacing the count data term H. Hence we have omitted the results for these estimators.

Using equation (4.23), we deduce certain observations regarding the efficiency of analyses based on panel data, paralleling those in Section 3.4 for the analysis of end-of-follow-up counts. For the analysis of panel data in the situation where (a) there is balance in the follow-up times for the subjects in treatments 1 and r, and (b) $\tau\beta_r = 0$, then the estimates of the treatment effect, β_r, based on the end-of-follow-up counts and the panel data are 100% efficient. As with end-of-follow-up count data, treatment groups 1 and r are balanced in the follow-up times when the distributions of final follow-up times, T_{s_i}, is the same in the two groups. More precisely, group 1 and group r are balanced with respect to follow-up

times if $\ell_{\mathrm{pr}} = 0$. Note that this condition applies equally to the event-time, panel, and count data; we also note that the estimators of τ based on both the end-of-follow-up counts and panel data are fully efficient with respect to the estimator based on the event-time data. However, the use of the information in the term corresponding to the multinomial distribution of the n_{ij} conditionally on n_{i+} results in a substantial improvement in efficiency for inference about the intensity parameters α and β_1.

Information from the intermediate follow-up times $T_1, \ldots, T_{s_i-1}$ enters ℓ_{pr} only through the marginal means, μ_{i+}. So, if a subject misses a follow-up time, say T_j in a clinical trial for instance, then provided that any events occurring during that follow-up time will be recorded as having occurred during the interval $(T_{j-1}, T_{j+1}]$, the estimators of the treatment effects, $\beta_r, r = 2, \ldots, k$ will be fully efficient. The end-of-follow-up count data could be viewed as an extreme case of this type of missing data, with every subject missing all the intermediate follow-up times and reporting only a single count at the end-of-follow-up.

To illustrate these findings, we present the results of a small numerical study. The study had two goals: first, to quantify the gains in relative efficiency due to the collection of panel data vs. end-of-follow-up counts; and, second, to assess the impact of dropouts on the efficiency of the estimators of α and β_1 obtained from analyses of the panel data. The designs we examined were for two treatment groups of 60 subjects each. We considered designs with one, two, or four follow-up times. Follow-ups occurred at $T_4 = 8$ (one follow-up), or $\{T_2 = 4, T_4 = 8\}$ (two follow-ups), or $\{T_1 = 2, T_2 = 4, T_3 = 6, T_4 = 8\}$ (four follow-ups).

To investigate the effects of dropout, four different dropout rates were considered: no dropout, 10%, 30%, and 60% dropout. The dropout rates describe the total proportion of subjects who did not complete the full follow-up period. The levels of dropout rate were chosen to reflect a reasonably broad range, from very little dropout to more than half the subjects dropping out before the final follow-up time. Table 4.10 shows the number of subjects examined at each follow-up time, for each combination of follow-up times and dropout rates. The parameter values chosen for the intensity function were $\alpha = 0.7, \beta_1 = 0$, and $\beta_2 = 0.3$. These values were chosen to give give marginal means μ_{i+} ranging from 4 to 20. The overdispersion parameter, τ, varies from 0 to 0.5. We first examine the balanced designs, where the two groups have the same dropout patterns and the same number of follow-up times. We consider the imbalanced designs separately, below. Table 4.11 shows the balanced and imbalanced designs examined. Note that we do not consider designs that

Dropout Rate	Number of Subjects Examined											
	One Follow-up				Two Follow-ups				Four Follow-ups			
	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4
none	–	–	–	60	–	60	–	60	60	60	60	60
10%	2	2	2	54	–	60	–	54	60	58	56	54
30%	6	6	6	42	–	60	–	42	60	54	48	42
60%	12	12	12	24	–	60	–	24	60	48	36	24

Table 4.10: Design Combinations Examined in Numerical Study. Dropout rate refers to the proportion of subjects who do not complete the follow-up period. The entries represent the number of subjects examined at each follow-up time.

involved different numbers of follow-up times in groups 1 and 2. The single follow-up designs investigate the effect of changes in the lengths of the follow-up periods; this more closely mimics a single follow-up study designed with staggered enrollment dates and a common end-of-follow-up date.

For the balanced designs, Figure 4.6 displays the changes in asymptotic relative efficiency as the number of follow-up times increases. The plots also show the effect of τ and dropout rates on the efficiencies. The efficiency of both $\tilde{\alpha}$ and $\tilde{\beta}_1$ increases dramatically with as few as two follow-ups, and the gains continue as the number of follow-ups increases from two to four. That is, panel data with as few as two follow-up times has much higher efficiency than the end-of-follow-up count data. Overdispersion appears to have little effect on efficiency, especially for $\tilde{\alpha}$. Dropouts reduce efficiency slightly for the panel data estimators, but increased dropout rates of course will actually improve the efficiency of the end-of-follow-up estimators. With no dropouts and only one follow-up, all counts are collected at exactly the same time T_s, and the parameters α and β_1 are totally confounded. Plots are not provided for β_2 and τ because the efficiency of β_2 does not drop below 99.9% for this numerical study, and both the count data and panel data estimators of τ are fully efficient relative to the event-time estimator.

Figure 4.7 shows the same efficiency calculations for the imbalanced designs examined in the study. In comparison with the balanced designs summarized in Figure 4.6, the imbalanced designs show a very similar pattern of change. The efficiency of the panel data estimators with more follow-up times are much higher than the end-of-follow-up count data, especially for the parameters α and β_1. The most notable difference from the results for the

Group 1		Group 2	
Follow-Up Times	Dropout Rate	Follow-Up Times	Dropout Rate
Balanced Combinations Examined			
T_4	none	T_4	none
T_4	10%	T_4	10%
T_4	30%	T_4	30%
T_4	60%	T_4	60%
T_2, T_4	none	T_2, T_4	none
T_2, T_4	10%	T_2, T_4	10%
T_2, T_4	30%	T_2, T_4	30%
T_2, T_4	60%	T_2, T_4	60%
T_1, T_2, T_3, T_4	none	T_1, T_2, T_3, T_4	none
T_1, T_2, T_3, T_4	10%	T_1, T_2, T_3, T_4	10%
T_1, T_2, T_3, T_4	30%	T_1, T_2, T_3, T_4	30%
T_1, T_2, T_3, T_4	60%	T_1, T_2, T_3, T_4	60%
Imbalanced Combinations Examined			
T_4	10%	T_4	60%
T_4	60%	T_4	10%
T_2, T_4	10%	T_2, T_4	60%
T_2, T_4	60%	T_2, T_4	10%
T_1, T_2, T_3, T_4	10%	T_1, T_2, T_3, T_4	60%
T_1, T_2, T_3, T_4	60%	T_1, T_2, T_3, T_4	10%

Table 4.11: Balanced and Imbalanced Designs Considered in the Numerical Study. The dropout rate refers to the percentage of subjects who leave the study before the final follow-up time.

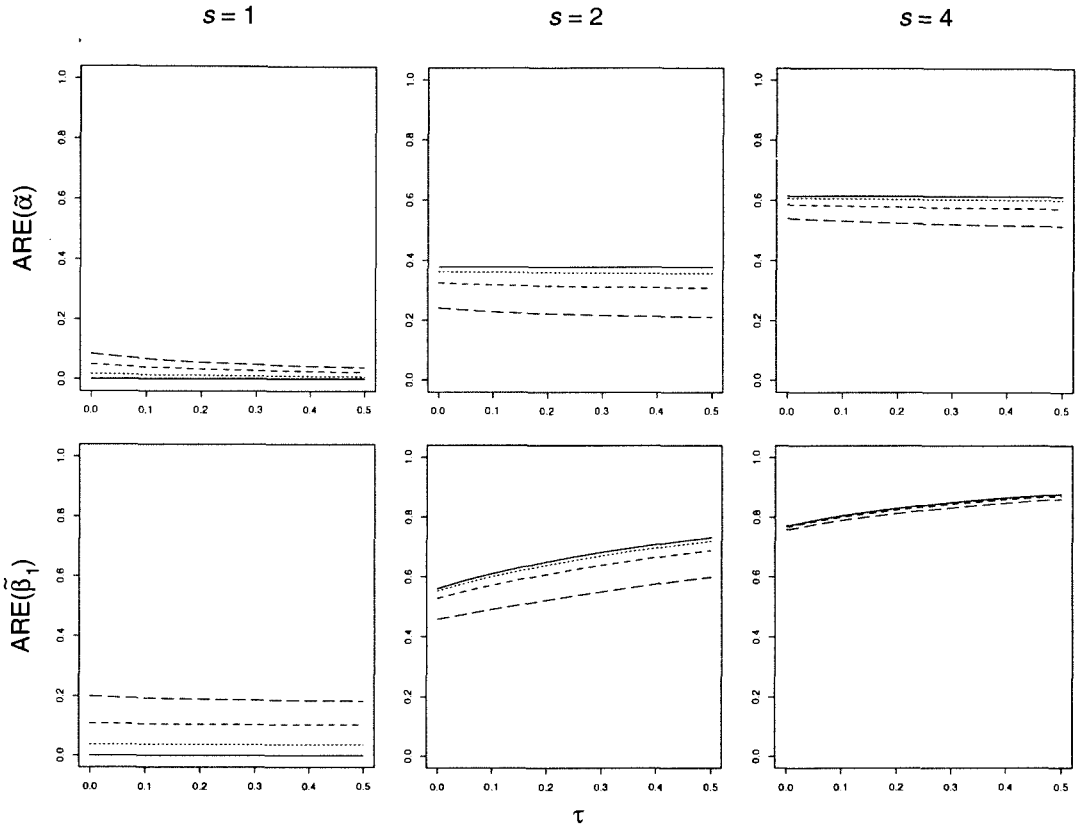

Figure 4.6: Efficiency Gains for Increased Number of Follow-Up Times, Balanced Dropout. The rows show the changes in the ARE as the number of follow-up times increases from $s = 1$ (end-of-follow-up count data) to $s = 2$ and $s = 4$ (panel data). Dropout rates considered were: no dropout(solid line), 10% dropouts (dotted line), 30% dropouts (dashed line), and 60% dropouts (long dashed line). The dropout rates are the same in both groups, occurring uniformly after each follow-up time.

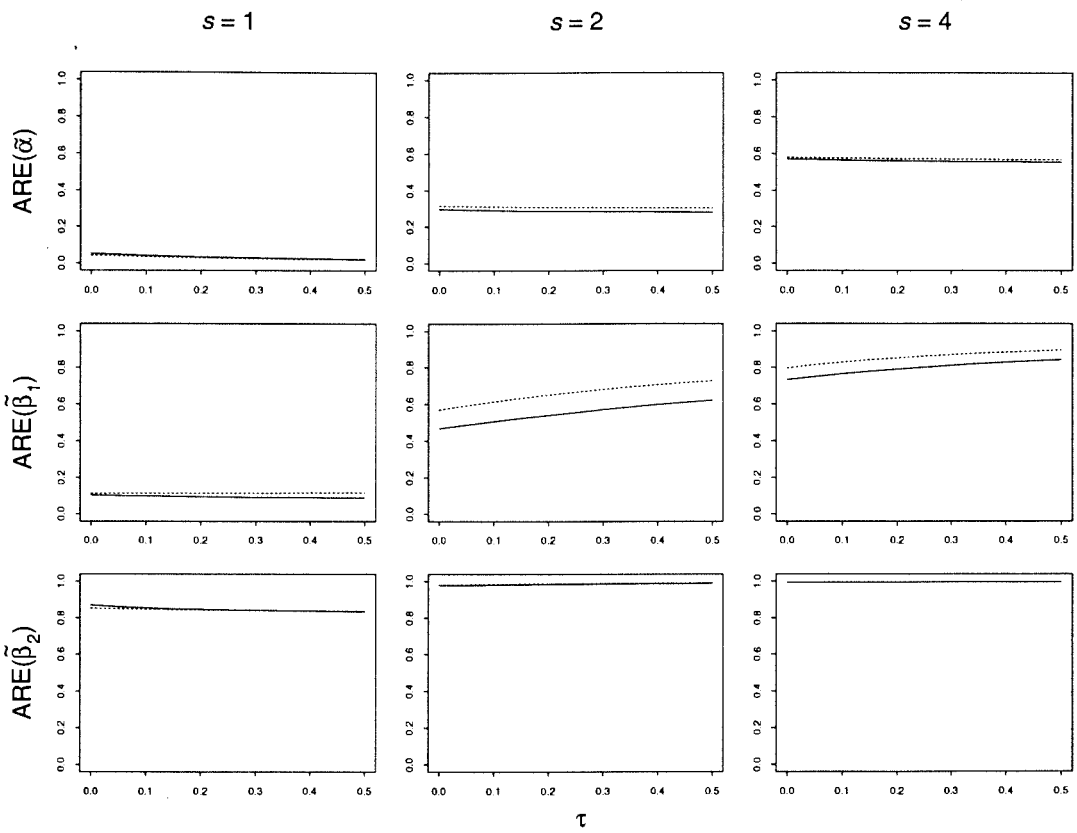

Figure 4.7: Efficiency Gains for Increased Number of Follow-Up Times, Imbalanced Dropout. The rows show the changes in the ARE as the number of follow-up times increases from $s = 1$ (end-of-follow-up count data) to $s = 2$ and $s = 4$ (panel data). The effect of the imbalanced dropout rates is minimal, though the overall dropout rate has much the same effect as in the balanced case (solid line: 10% dropout in group 1 and 60% dropout in group 2; dashed line: 60% dropout in group 1 and 10% dropout in group 2.

balanced designs is that the efficiency of the treatment effect estimator, $\tilde{\beta}_2$, can be as low as 80-85% for the end-of-follow-up counts, whereas for the balanced designs in the study, the efficiency of $\tilde{\beta}_2$ was never less than 99%. Given that the degree of imbalance considered here was quite extreme, with 10% dropout in one group and 60% dropout in the other, these results suggest that for a very broad range of practical situations, inference based on panel data will be highly efficient for a treatment effect and will be much more efficient for inference regarding α and β_1 than inference based on end-of-follow-up counts. While efficiencies will be even lower when there is greater imbalance, here we refrain from calculating efficiency for designs which would bring into question the assumption of independent censoring.

4.3.2 Small Sample Comparison of Panel Data vs. Count Data

Returning to the simulation study described in Section 4.2.4, we now examine the small sample efficiency of the estimators based on the end-of-follow-up counts and the panel data. Recall that the simulation study was conducted to examine the small sample properties of the estimators and their variances for a clinical trial similar to that seen in the bladder cancer data. The design was for two equal-sized groups of subjects, each studied for a total follow-up period of 72 months. Twenty-five percent of the subjects in each group dropped out after months 18, 36, and 54. The data were simulated as event-time data from a NHPP with Weibull intensity function, and subject-specific random effects were generated as independent gamma random variables with mean 1 and variance τ. Parameter values ($\beta_1 = -2, \beta_2 = 0.1, \alpha = 0.7$, and $\tau = 1.5$) were chosen to be similar to the estimates obtained from the bladder cancer data. End-of-follow-up counts and panel data were constructed from the event-time data in the usual way, by counting the number of events occurring for each subject during the total length of follow-up (for end-of-follow-up counts) or during the follow-up periods between 0, 18, 36, 54, and 72 months (for panel data).

In Table 4.12, we summarize the observed relative efficiency of the count and panel data estimators. We compute the observed relative efficiency here as the ratio of the observed sampling variance of the estimator based on the count or panel data to the sampling variance of the corresponding estimator based on the event-times. We note that sample size does not significantly affect the efficiency values — they are very similar to the theoretical values for groups of size 48. The efficiency of the estimators for β_1 and α behave as predicted by the asymptotic results in the previous section; that is, though still low, they show a large improvement in the efficiency of β_1 and α when based on the panel data with four

	End-of-Follow-Up Counts				Panel Data			
Group Size	$\tilde{\beta}_{1c}$	$\tilde{\beta}_{2c}$	$\tilde{\alpha}_c$	$\tilde{\tau}_c$	$\tilde{\beta}_{1p}$	$\tilde{\beta}_{2p}$	$\tilde{\alpha}_p$	$\tilde{\tau}_p$
48	0.065	0.990	0.028	1.093	0.686	1.000	0.490	1.094
96	0.064	0.993	0.028	1.038	0.677	1.000	0.462	1.000

Table 4.12: Small Sample Efficiency of the Analysis of End-of-Follow-Up Counts and Panel Data. Efficiency is measured as the ratio of the observed sample variance of the estimator to the observed sample variance of the event-time estimator.

follow-up times rather than on the end-of-follow-up count data. The estimators of β_2 and τ also behave as predicted, achieving very high efficiencies regardless of the type of recurrent event data analyzed.

4.4 Tests and Diagnostics for Model-Checking

4.4.1 Testing for Overdispersion

Since the Poisson process model offers simplicity for inference concerning the model parameters relative to the random effects model, we develop here a score-type test for overdispersion. The test is based on the estimating function g_τ, (4.16), and is akin to likelihood score testing procedures (Breslow 1990).

Let $\gamma' = (\beta', \alpha')$, and let $\boldsymbol{g}_\gamma$ denote the set of estimating functions for γ; i.e., $\boldsymbol{g}'_\gamma = (\boldsymbol{g}'_\beta, \boldsymbol{g}_\alpha)$, where the components $\boldsymbol{g}_\beta$ and $\boldsymbol{g}_\alpha$ are the quasi-likelihood estimating functions based on the event-time data assuming a Weibull intensity, (4.14) and (4.15). Note that the Weibull intensity is used here for consistency with earlier sections, but we later point out that the results are more generally applicable. We first obtain an estimator of γ under $H_{0\tau} : \tau = 0$ by solving $\boldsymbol{g}_\gamma = \boldsymbol{0}$, and denote the solution $\tilde{\gamma}^*$, the superscript "*" indicating that the quantity has been calculated under $H_{0\tau} : \tau = 0$. The test statistic is then $\tilde{\boldsymbol{g}}^*_\tau$,

$$\tilde{g}^*_\tau = \sum_{i=1}^{M} \{(n_{i+} - \tilde{\mu}^*_{i+})^2 - \tilde{\mu}^*_{i+}\},$$

where $\tilde{\mu}^*_{i+} = T_{s_i}^{\tilde{\alpha}^*} \exp(\boldsymbol{x}'_i \tilde{\boldsymbol{\beta}}^*)$, and $\tilde{\alpha}^*$ and $\tilde{\boldsymbol{\beta}}^*$ are the components of $\tilde{\gamma}^*$.

The asymptotic variance of $\tilde{\boldsymbol{g}}^*_\tau$ is estimated by

$$\text{AsVar}(\tilde{\boldsymbol{g}}^*_\tau) = \boldsymbol{b}_{22} - \boldsymbol{a}_{21} A_{11}^{-1} \boldsymbol{b}_{12} - \boldsymbol{b}_{21} A_{11}^{-1} \boldsymbol{a}'_{21} + \boldsymbol{b}_{21} A_{11}^{-1} \boldsymbol{b}_{11} A_{11}^{-1} \boldsymbol{a}'_{21}, \tag{4.24}$$

where the A's are expectations of the negative of the derivatives of the estimating functions and the B's are the variances and covariances of the estimating functions: let $\boldsymbol{g}' = (\boldsymbol{g}'_\gamma, g_\tau)$, then, as before, $A = \mathrm{E}(-\partial \boldsymbol{g}/\partial \boldsymbol{\theta})$, $B = \mathrm{E}(\boldsymbol{g}\boldsymbol{g}')$, partitioned as

$$A = \begin{pmatrix} A_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{21} & a_{22} \end{pmatrix} \text{ and } B = \begin{pmatrix} B_{11} & \boldsymbol{b}_{12} \\ \boldsymbol{b}_{21} & b_{22} \end{pmatrix},$$

A_{11} and B_{11} being square matrices with row dimension equal to that of γ. If the components of $\boldsymbol{g}$ are (4.15) – (4.16), and assuming third and fourth moments of the mixing distribution to be the same as for the gamma distribution, then (4.24) simplifies considerably under $H_{0\tau} : \tau = 0$. We have

$$A^* = \begin{pmatrix} X'VX & X'VZ & \mathbf{0} \\ Z'VX & Z'VZ + \frac{\mu_{++}}{\alpha^2} & 0 \\ \mathbf{1}'VX & \mathbf{1}'VZ & \mathbf{1}'V^2\mathbf{1} \end{pmatrix},$$

and

$$B^* = \begin{pmatrix} X'VX & X'VZ & X'V\mathbf{1} \\ Z'VX & Z'VZ + \frac{\mu_{++}}{\alpha^2} & Z'V\mathbf{1} \\ \mathbf{1}'VX & \mathbf{1}'VZ & \mathbf{1}'V\mathbf{1} + 2\mathbf{1}'V^2\mathbf{1} \end{pmatrix};$$

here, as before, $V = \mathrm{diag}(\mu_{i+}; i = 1, \ldots, M)$, $Z = \mathrm{diag}(\partial/\partial\alpha \log \Lambda_0(T_{s_i}); i = 1, \ldots, M) = \mathrm{diag}(\log T_{s_i})$, for the Weibull baseline intensity, and $\mathbf{1}$ is the $M \times 1$ unit vector. Hence, $\mathrm{AsVar}(\tilde{g}^*_\tau)$ under $H_{0\tau}$ simplifies to

$$\boldsymbol{b}^*_{22} - \boldsymbol{a}^*_{21} A^{*-1}_{11} \boldsymbol{b}^*_{12}. \tag{4.25}$$

which simplifies to $2\sum_{i=1}^{M} \tilde{\mu}^{*2}_{i+}$. Thus, the standardized version of the test statistic is

$$S_1 = \sum_{i=1}^{M} \frac{(n_{i+} - \tilde{\mu}^*_{i+})^2 - \tilde{\mu}^*_{i+}}{\sqrt{2\sum_{i=1}^{M} \tilde{\mu}^{*2}_{i+}}}. \tag{4.26}$$

Under the usual regularity conditions (Inagaki 1973; White 1982) and assuming that the mean and variance have been correctly modeled, the statistic S_1 is asymptotically standard normal.

There are some points worth highlighting regarding the construction of S_1. First, it is possible to replace B in (4.24) by an empirical variance estimator, say the sandwich estimator, so that third and fourth moment assumptions need not be made. However, the

simplification gained by using these third and fourth order moment assumptions make S_1 rather easy to compute. Second, a key requirement for the simplification is that X contain a column of $1's$. In this case, the first column of A_{11}^* is B_{12}^*. Thus $A_{11}^{*-1}B_{12}^*$ is a vector with 1 in the first entry and zeros elsewhere, and $A_{21}^*A_{11}^{*-1}B_{12}^*$ becomes $\mathbf{1}'V\mathbf{1}$ under $H_{0\tau}$. Hence (4.25) becomes $2\sum_{i=1}^M \tilde{\mu}_{i+}^{*2}$ under $H_{0\tau}$. Third, recall that (4.26) was constructed using an estimating function for α based on the event-time data (4.15); if, instead, we were to use the estimating function for α based on the panel data, $g_{\alpha p}$, then similar computations as performed above would show that the form of the test statistic obtained is identical to (4.26); the only difference would be that the estimate of μ_{i+}^* would be based on solving $(\boldsymbol{g}_\beta^{*\prime}, g_{\alpha p}^*)' = \mathbf{0}$ under $H_{0\tau}$. The form of the test statistic is the same for intensities other than the Weibull or for $\boldsymbol{\alpha}$ vector-valued. We would, of course, be required to modify the estimating function for $\boldsymbol{\alpha}$ to account for a different baseline intensity. Similar test statistics have been evaluated in previous studies and were shown to benefit from a small sample correction (Breslow 1990; Dean 1992; Dean and Lawless 1989). Applying a small sample correction to the numerator of S_1 yields an adjusted statistic with numerator

$$\sum_{i=1}^{M}\{(n_{i+} - \tilde{\mu}_{i+}^*)^2 - (1 - h_i^*)\tilde{\mu}_{i+}^*\},$$

where h_i^* is the leverage estimated under $H_{0\tau}$. Other types of test statistics for overdispersion could be considered, e.g., adaptations of the results from Spinelli (1994). This test is also very similar to those derived by Dean (1992) using likelihood methods.

4.4.2 Testing for Non-Homogeneity in the Poisson Process

To test whether the data may be adequately modeled by a homogeneous Poisson process we test that $\lambda_0(t)$ is constant. With a Weibull baseline intensity function, this is equivalent to testing $H_{0\alpha}$: $\alpha = 1$. A test for $H_{0\alpha}$ is derived below, but similar tests can be constructed when other parametric forms for the baseline intensity λ_0 are used. The procedure discussed in Section 4.4.1 yields a score-type test for $H_{0\alpha}$. Based on the event-time data, the test statistic is

$$\tilde{g}_\alpha^\dagger = \sum_{i=1}^{M}\left\{\sum_{j=1}^{s_i}\sum_{l=1}^{n_{ij}}\left(1 + \log\frac{\omega_{ijl}}{T_{s_i}}\right) + \frac{(n_{i+} - \tilde{\mu}_{i+}^\dagger)\log T_{s_i}}{1 + \tilde{\tau}^\dagger\tilde{\mu}_{i+}^\dagger}\right\}. \tag{4.27}$$

Here $\tilde{\mu}_{i+}^\dagger = T_{s_i}\exp(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}^\dagger)$ and $(\tilde{\boldsymbol{\beta}}^\dagger, \tilde{\tau}^\dagger)$ are estimates of $\boldsymbol{\beta}$ and τ obtained under $H_{0\alpha}$ by solving $\boldsymbol{g}_\beta = \mathbf{0}$ and $g_\tau = 0$, (4.14) and (4.16), with $\alpha = 1$. We standardize the test statistic

in a similar manner as in Section 4.4.1. Under $H_{0\alpha}$,

$$\text{AsVar}(\tilde{g}_\alpha^\dagger) = Z'WV_o^{-1}WZ + \mu_{++} - Z'WV_o^{-1}WX(X'WV_o^{-1}WX)^{-1}X'WV_o^{-1}WZ,$$

where, as usual, $W = \text{diag}(\mu_{i+}; i = 1, \ldots, M)$ and $V_o = \text{diag}(\mu_{i+}(1 + \tau\mu_{i+}); i = 1, \ldots, M)$. Under certain regularity conditions (Inagaki 1973; White 1982), and assuming correct specification of the mean, the standardized test statistic is asymptotically distributed as standard normal.

4.4.3 Diagnostic Plot for Checking the Baseline Intensity

The factorizations of the likelihood given in (4.5) and (4.20) suggest that diagnostics for the model may be considered in two stages: first, based on the end-of-follow-up counts n_{i+}; second, based on either the conditional distribution of the event-times given n_{i+}, $i = 1, \ldots, M$, or the multinomial conditional distribution of the panel counts, n_{ij}, $j = 1, \ldots, s_i$, given n_{i+}, $i = 1, \ldots, M$. The choice in the second stage reflects whether we have observed the event times or the panel data.

Residuals based on n_{i+} are the standard residuals for generalized linear modeling of count data (McCullagh and Nelder 1989, Sec. 12.5). The residuals based on the second stage convey the essential information concerning the parametric form of the baseline intensity function (cf. Section 4.2). For the construction of these residuals we note that conditional on n_{i+}, $r_{ijl} = \Lambda_0(\omega_{ijl})/\Lambda_0(T_{s_i})$ are distributed as order statistics from $U(0,1)$ (see, for example, Ross 1983, Chapter 2). For the Weibull model $r_{ijl} = (\omega_{ijl}/T_{s_i})^\alpha$ and in practice, we substitute an estimate of α; the estimates of r_{ijl} would then be approximately distributed as uniform order statistics. In Section 4.5 we illustrate the use of these r_{ijl}s for preparing Q-Q plots to assess how well the Weibull intensity fits the bladder cancer recurrence data. We will also test the fit of the random effects non-homogeneous Poisson process model using the tests developed in Sections 4.4.1 and 4.4.2.

4.5 Illustration

In Chapter 3 we examined the bladder cancer data collected during a clinical trial conducted by the Veterans Administrative Co-operative Urological Research Group. Recall that the trial compares the efficacy of three treatments for recurring bladder cancer: placebo pills, pyridoxine pills, and periodic instillations of thiotepa into the bladder. The data were

originally presented in Byar et al. (1977) and also appear in Andrews and Herzberg (1985). We return to this dataset to illustrate the analysis of panel data and the use of the diagnostic measures presented in Section 4.4.

A total of 116 subjects diagnosed with bladder cancer were enrolled in the study. The tumors were surgically removed and the subjects were then randomly assigned to one of the three treatment groups: (1) Placebo; (2) Pyridoxine; and, (3) Thiotepa. The data consist of the bladder cancer recurrence times for each of the subjects.

As conducted, the study resulted in event-time data rather than panel data. The subjects were monitored continuously and could drop out at arbitrary times. In Chapter 3 these data were used to compare the efficiency of the end-of-follow-up counts and the event-time data. In this chapter, we impose an artificial panel data structure to allow comparison of the event-time, panel and count data efficiencies. The dataset is modified to simulate a five year panel study with yearly follow-up exams. We treat as lost to dropout any observation which would not have been available based on yearly follow-ups. The subjects' drop-out times have been truncated to the next lowest multiple of 12 months, and any information following the last observed follow-up time has been omitted. For example, suppose a subject had a recurrence at 27 months before dropping out of the trial at month 30. His final observed follow-up would be at 24 months and the recurrence at 27 months would not be observed in the modified dataset. As a result of these modifications, 22 subjects were eliminated because they did not complete a full year. End-of-follow-up count data was created from this modified dataset by aggregating over the observed yearly follow-ups for each subject. We shall refer to this as the modified bladder cancer dataset and, except where noted, all analyses have been based on this modified dataset. Note that this means the analyses of this end-of-follow-up and event-time data in this chapter will not match those in Chapter 3.

Table 4.13 displays the number of subjects at risk at each follow-up time. The treatment groups appear to have similar dropout rates. In Chapter 3, we observed that the same was true of the unmodified drop-out times. This provides informal reassurance that the censoring process (i.e., end-of-follow-up time, T_{ie}) is likely independent of the event process.

Table 4.14 presents the results from the analyses of the modified data and Figure 4.8 displays quasi-likelihood estimates of the cumulative intensity functions based on the three quasi-likelihood analyses of the modified data (the likelihood analyses of the panel data results in a plots that are visually indistinguishable from the quasi-likelihood versions and are not presented in the plots). The Weibull intensity model is used for all analyses. Comparing

	Follow-Up Times in Months					
Treatment Group	$T_0 = 0$	$T_1 = 12$	$T_2 = 24$	$T_3 = 36$	$T_4 = 48$	$T_5 = 60$
1. Placebo	47	42	34	19	9	3
(dropout rate)		(0.106)	(0.277)	(0.596)	(0.809)	(0.936)
2. Pyridoxine	31	23	22	16	6	1
(dropout rate)		(0.258)	(0.290)	(0.484)	(0.806)	(0.968)
3. Thiotepa	38	33	25	19	6	0
(dropout rate)		(0.132)	(0.342)	(0.500)	(0.842)	(1)

Table 4.13: Number of Subjects at Risk in the Modified Bladder Cancer Dataset. The table gives the number of subjects at risk at time T_j. The dropout rate is cumulative from time T_0.

Likelihood Analyses: Estimate (Std. Err.)

	End-of-Follow-Up Count Data		Panel Data		Event-Time Data	
β_1	-1.959	(1.148)	-2.604	(0.390)	-2.866	(0.349)
β_2	-0.048	(0.362)	-0.075	(0.362)	-0.087	(0.363)
β_3	-0.296	(0.346)	-0.279	(0.348)	-0.272	(0.350)
α	0.736	(0.329)	0.925	(0.094)	1.002	(0.079)
τ	1.312	(0.357)	1.317	(0.358)	1.325	(0.360)

Quasi-likelihood Analyses: Estimate (Std. Err.)

	End-of-Follow-Up Count Data		Panel Data		Event-Time Data	
β_1	-1.975	(1.099)	-2.596	(0.342)	-2.860	(0.256)
β_2	-0.049	(0.341)	-0.074	(0.332)	-0.084	(0.331)
β_3	-0.305	(0.312)	-0.293	(0.310)	-0.287	(0.312)
α	0.742	(0.331)	0.923	(0.089)	1.000	(0.064)
τ	1.039	(0.247)	0.989	(0.245)	0.995	(0.252)

Table 4.14: Results of Likelihood and Quasi-Likelihood Analyses of the Bladder Cancer Recurrent Event Data. Standard errors (Std. Err.) are based on the robust variances. The data have been modified to fit the panel data structure.

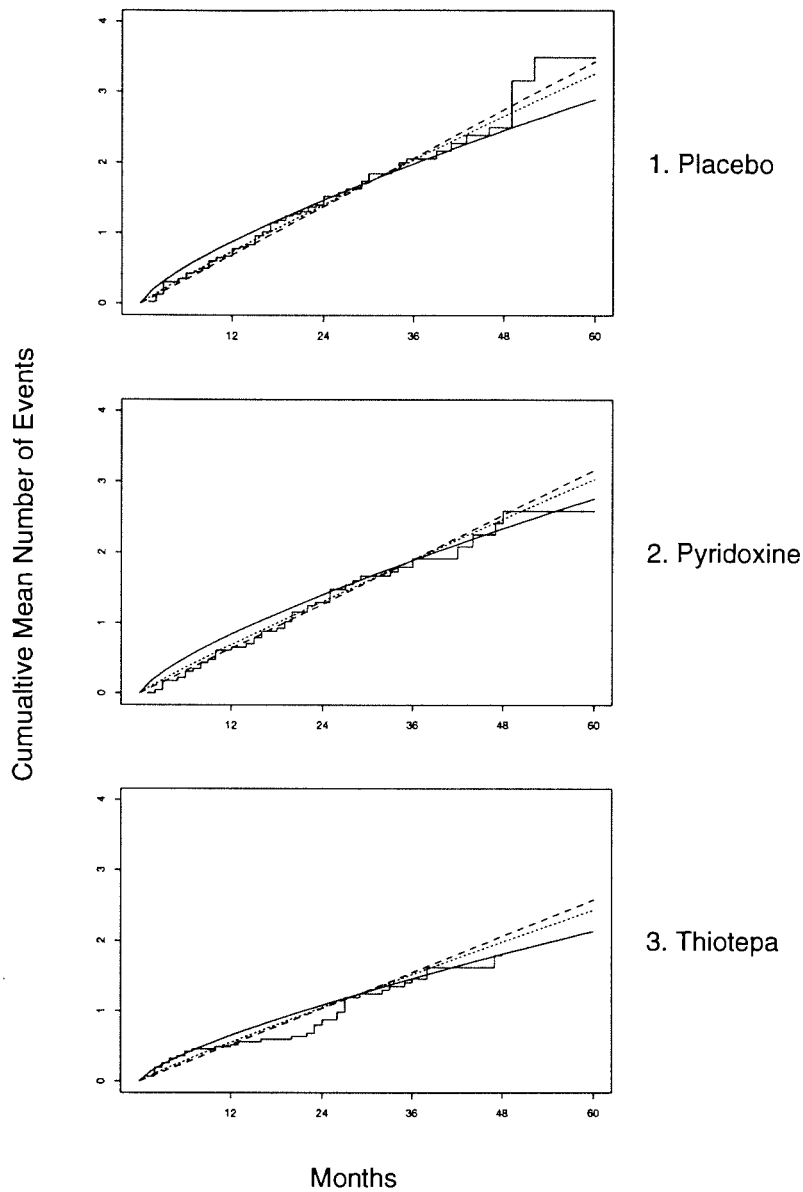

Figure 4.8: Estimated Cumulative Mean Functions for the Bladder Cancer Data. For each group, the smooth curves show the estimates of the cumulative mean function $t^{\tilde{\alpha}} \exp(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}})$ based on the three types of recurrent event data (count: solid line; panel: dotted line; event-time: dashed line). The observed cumulative mean function for the modified event-time data is shown as a step function.

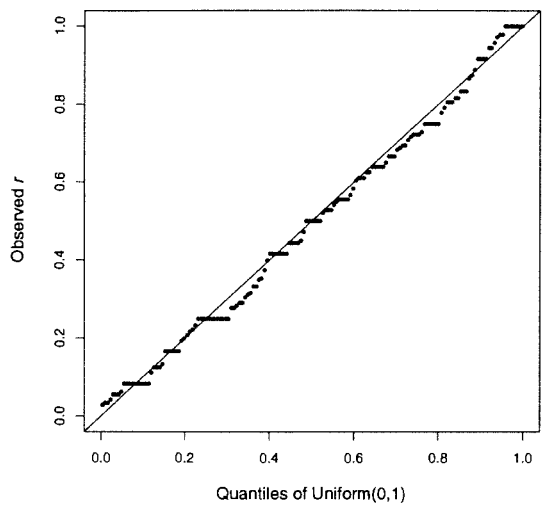

Figure 4.9: A Diagnostic Plot for Checking the Baseline Intensity Function. The observed values of $r_{ijl} = (\omega_{ijl}/T_{s_i})^{\tilde{\alpha}_i}$ are plotted against the quantiles of the Uniform(0,1) distribution.

the estimates based on the count, panel, and event-time data in Table 4.14 reveals that the estimates of the intensity parameters β_1 and α show considerable variation depending on the type of data analyzed. As expected, the data-type appears to have less effect on the estimators of treatment effects, β_2, β_3, and the overdispersion parameter, τ. Also, Figure 4.8 reveals that the estimated cumulative mean functions based on the three data structures are quite similar, especially for the period before 48 months. After this time, the cumulative dropout rate is approximately 80%, and estimates would be less reliable in any case.

Figure 4.9 demonstrates the use of the residual r_{ijl} for checking the goodness-of-fit of the baseline intensity function. The figure was created by plotting values of the residuals $r_{ijl} = (\omega_{ijl}/T_{s_i})^{\tilde{\alpha}_i}$ against the quantiles of a uniform distribution. The plot shows little evidence of serious lack-of-fit, though the "ledges" induced by the imposition of the panel data structure make the plot slightly more difficult to interpret. The overall impression is that the Weibull intensity model is very reasonable.

Testing for overdispersion (see Section 4.4.1) using the panel data results in a value of $S_1 = 14.45$. Based on the event-time and count data, the statistic takes on the values 12.71 and 14.53. There is strong evidence that these data demonstrate overdispersion relative to

the Poisson model.

Testing for non-homogeneity of the Poisson process based on the Weibull intensity model corresponds to testing the hypothesis $H_0 : \alpha = 1$ (see Section 4.4.2). Using the event-time data, the standardized test statistic is equal to -0.01 (p-value = 0.504 based on the normal approximation). In other words, the data suggest that a homogeneous Poisson process would be a reasonable model for this data. In the next chapter, we describe a nonparametric intensity model for recurrent event data recorded as panel data.

Chapter 5

Semiparametric Analysis of Panel Recurrent Event Data

5.1 Introduction

In a study of recurrent events, panel data is a record of the number of events occurring for each subject between periodic follow-up times. This is in contrast with event-time data, where the exact recurrence times are recorded, and end-of-follow-up count data, where a count is recorded only at the end of the follow-up period. In Chapter 4, we examined the efficiency of panel data versus event-time data and concluded that panel data can be highly efficient for inference regarding covariate effects under reasonable conditions; we also showed that panel data can be quite efficient for inference regarding changes in the Poisson process intensity function over time, unlike end-of-follow-up count data. When event-time data is too invasive or expensive to collect, or when it is not otherwise feasible to record the event times accurately, panel data is thus an attractive alternative.

In Chapter 4, we explored the use of quasi-likelihood estimators for the analysis of panel recurrent event data. Quasi-likelihood methods were adopted specifically to reduce dependence on assumptions regarding the distribution of the subject-specific random effect. In this chapter, we develop a nonparametric estimator for the baseline intensity function, making the methods even robust to misspecification of the baseline intensity model.

The flexibility of the Weibull intensity function, relative to the constant intensity exponential model, has made it a popular choice in reliability and survival studies. However, it

will not be appropriate in all cases. Other parametric intensity functions are possible, but to avoid requiring a parametric specification of the baseline intensity function, we propose a nonparametric estimator of the baseline intensity function. Our estimator is related to the Nelson-Aalen estimator used by Lawless and Nadeau (1995); however, where Lawless and Nadeau use likelihood methods based on Poisson models and account for overdispersion only through the use of robust variance estimates, we have explicitly accounted for subject-to-subject variation in the estimating functions. Used in combination with the quasi-likelihood estimators developed in the previous chapter, the techniques provide a robust semiparametric method for the analysis of recurrent event data.

In Section 5.2 we review the quasi-likelihood estimator for overdispersed Poisson processes recorded as panel data and present the nonparametric intensity function estimator. Section 5.3 proposes quasi-score tests for the fit of specific parametric intensity models. Section 5.4 calculates the relative efficiency of the semiparametric estimator of the treatment effect, including a small numerical study. Section 5.5 illustrates the use of these techniques with analyses of two illustrative examples: the bladder cancer data and the Cherry Bark Tortrix data. The bladder cancer data have been examined in Chapters 3 and 4 under the assumption of a Weibull intensity model. We find that the semiparametric analysis conducted in this chapter yields very similar results. The Cherry Bark Tortrix data, on the other hand, yield very different estimates of the cumulative mean functions under the Weibull and semiparametric analyses, highlighting the usefulness of the nonparametric intensity function. The chapter concludes in Section 5.6 with an examination of the small sample characteristics of the estimators and our proposed test statistics via simulation.

5.2 A Semiparametric Model for Overdispersed Panel Data

In Section 4.2.2 of the previous chapter we presented quasi-likelihood estimating functions for inference with overdispersed panel data under the assumption of a parametric model for the baseline intensity function, λ_0. In this section we will adopt a nonparametric baseline intensity model to replace this parametric baseline function. Other than this substitution, the same intensity model still applies, i.e., covariate effects enter the model via proportional intensity assumptions and subject heterogeneity is accommodated via a random subject-specific frailty. Quasi-likelihood estimates are found using a new estimating equation for the nonparametric intensity and the same quasi-likelihood estimating functions for β and

τ as in the previous chapter. We begin by briefly reviewing the panel data structure and intensity model to motivate construction of the nonparametric baseline function.

A panel data experiment involves M subjects with covariate vectors $\boldsymbol{x}_i$. Without loss of generality, we can assume that the subjects are enrolled in the study at time $T_0 = 0$. At follow-up-times $T_1, T_2, \ldots, T_s$, each subject is examined and the number of events occurring since the previous follow-up is recorded as n_{ij}, $i = 1, \ldots, M$ and $j = 1, \ldots, s$. The counts n_{ij} are realizations of the counting process $N_i(t)$ which has an intensity function of the proportional intensity form

$$\lambda_i(t) = \nu_i \lambda_0(t) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

where ν_i is a subject specific random effect, $\lambda_0(t)$ is a baseline intensity function, and $\boldsymbol{\beta}$ is a k dimensional vector of covariate effects. The subject-specific effect ν_i is included to explicitly model overdispersion of the counts. We will assume that the mixing distribution associated with ν has mean one and variance τ. Conditionally on the subject-specific effects ν_i, the counting processes $N_i(t)$ are independent, nonhomogeneous Poisson processes with intensity functions $\lambda_i(t)$.

Because observations are collected only at the follow-up times, $\{T_1, T_2, \ldots, T_s\}$, the analysis of panel data requires estimates of the intensity function only at the follow-up times. This suggests that we consider the intensity as step function; this leads to a convenient parameterization,

$$e^{\phi_j}(T_j - T_{j-1}) = \Lambda_0(T_j) - \Lambda_0(T_{j-1}) \quad j = 1, \ldots, s, \tag{5.1}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$ is the cumulative baseline intensity function, and so $\Lambda(T_0) = 0$. Thus e^{ϕ_j} represents the *mean intensity* during the interval $(T_{j-1}, T_j]$. In the parametric model,

$$\mu_{ij} = \mathrm{E}\,(n_{ij}) = \int_{T_{j-1}}^{T_j} \lambda_0(u)\, du \exp(\boldsymbol{x}_i' \boldsymbol{\beta}).$$

Under (5.1), this becomes

$$\mu_{ij} = \mathrm{E}\,(n_{ij}) = e^{\phi_j}(T_j - T_{j-1}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}). \tag{5.2}$$

Using this parameterization, the baseline cumulative intensity function $\Lambda_0(T_j)$ is also the cumulative mean function for those subjects with $\boldsymbol{x}_i = \mathbf{0}$; the common intensity parameter β_1 used in the parametric model has been absorbed in the parameter $\boldsymbol{\phi}$. Thus, for a k group design only k-1 indicator variables are required in the covariate vector $\boldsymbol{x}_i$. For consistency

of notation, we henceforth assume that $x_{i1} = 0$, effectively eliminating the parameter β_1 from the semiparametric formulation. As before, we concentrate on the k-group problem, where

$$x_{ir} = \begin{cases} 1 & \text{if individual } i \text{ received treatment } r, \\ 0 & \text{otherwise} \end{cases}$$

for $r = 2, \ldots, k$. Treatment group 1 still serves as the reference group, but now the mean function for group 1 is given directly by the cumulative intensity function.

Using this semiparametric intensity model, the likelihood is

$$L_{\mathrm{S}}(\boldsymbol{\beta}, \boldsymbol{\phi}, \tau) = \prod_{i=1}^{M} \left\{ \binom{n_{i+}}{n_{i1} \cdots n_{is_i}} \prod_{j=1}^{s_i} \left(\frac{\mu_{ij}}{\mu_{i+}} \right)^{n_{ij}} \right\} L_{\mathrm{MP}}(\boldsymbol{\beta}, \boldsymbol{\phi}, \tau)$$

where the subscript S signifies the semiparametric intensity model for panel data and $L_{\mathrm{MP}}(\boldsymbol{\beta}, \boldsymbol{\phi}, \tau)$ is the mixed-Poisson likelihood as given in (4.7) where the parameter $\boldsymbol{\alpha}$ is replaced by $\boldsymbol{\phi}$.

Assuming a gamma distribution for the subject-specific effects ν_i, the first and second derivatives of L_{S} with respect to $\boldsymbol{\beta}$ and τ equal the corresponding derivatives of L_{t}, equation (4.5), and are given in Lawless (1987a). The derivatives and expected information with respect to $\boldsymbol{\phi}$ are

$$\begin{aligned} \frac{\partial \log L_{\mathrm{S}}}{\partial \phi_h} &= \sum_{i=1}^{M} \sum_{j=1}^{s_i} n_{ij} \frac{\partial}{\partial \phi_h} \log \frac{\mu_{ij}}{\mu i+} + \frac{\partial \log L_{\mathrm{NB}}}{\partial \phi_h} \\ &= \sum_{i=1}^{M} \left(n_{ih} - \frac{n_{i+}}{\mu_{i+}} \mu_{ih} \right) + \sum_{i=1}^{M} \sum_{j=1}^{s_i} \frac{n_{i+} - \mu_{i+}}{\mu_{i+}(1 + \tau \mu_{i+})} \mu_{ih} \end{aligned}$$

with second derivatives

$$\begin{aligned} \mathrm{E}\left(-\frac{\partial^2 \log L_{\mathrm{S}}}{\partial \phi_h \partial \phi_{h'}} \right) &= \sum_{i=1}^{M} \left\{ \mu_{ih} \operatorname{Ind}(h = h') - \frac{\tau \mu_{ih} \mu_{ih'}}{1 + \tau \mu_{i+}} \right\} \\ \mathrm{E}\left(-\frac{\partial^2 \log L_{\mathrm{S}}}{\partial \phi_h \partial \beta_r} \right) &= \sum_{i=1}^{M} \frac{\mu_{ih}}{1 + \tau \mu_{i+}} x_{ir} \\ \mathrm{E}\left(-\frac{\partial^2 \log L_{\mathrm{S}}}{\partial \phi_h \partial \tau} \right) &= 0 \end{aligned}$$

where $\operatorname{Ind}(A)$ is the indicator function for the event A.

As was described in earlier chapters, quasi-likelihood methods are becoming popular for the analysis of count data displaying overdispersion. Quasi-likelihood methods require only

low order moment assumptions rather than the full distributional assumptions of likelihood methods, and, in Chapter 4, were shown to be highly efficient for the analysis of panel data. Consequently, we concentrate our attention on quasi-likelihood estimators from this point onwards.

As the structure of the intensity function has not changed with respect to $\boldsymbol{\beta}$, and neither has our overdispersion model, we will continue to use the same estimating functions for $\boldsymbol{\beta}$ and τ that were developed in earlier chapters,

$$\begin{aligned} g_{\beta_r} &= \sum_{i=1}^{M} \frac{n_{i+} - \mu_{i+}}{1 + \tau\mu_{i+}} x_{ij} \\ g_{\tau} &= \sum_{i=1}^{M} \frac{(n_{i+} - \mu_{i+})^2 - (1 - h_i)\,\mu_{i+}(1 + \tau\mu_{i+})}{(1 + \tau\mu_{i+})^2}. \end{aligned}$$

For $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_s)'$ we adopt the likelihood estimating functions, without adopting the full likelihood model;

$$\begin{aligned} g_{\phi_j} &= \sum_{i=1}^{M} \left(n_{ij} - \frac{n_{i+}}{\mu_{i+}} \mu_{ij} \right) + \sum_{i=1}^{M} \frac{n_{i+} - \mu_{i+}}{\mu_{i+}(1 + \tau\mu_{i+})} \mu_{ij} \\ &= \sum_{i=1}^{M} \frac{(1 + \tau\mu_{i+})n_{ij} - (1 + \tau n_{i+})\mu_{ij}}{1 + \tau\mu_{i+}}, \quad j = 1, \ldots, s. \end{aligned}$$

The second form of this function makes clear that the estimator is found by comparing the observed and expected counts at each follow-up time, and that the comparison takes into account the amount of overdispersion in the data. The estimating functions, $g_{\phi_j}, j = 1, 2, \ldots, s$ are unbiased provided that the final follow-up times T_{s_i} are independent of the event process. Let $\boldsymbol{g}_{\phi}$ be the vector of estimating functions for the parameter $\boldsymbol{\phi}$, and let $\boldsymbol{g}_S = (\boldsymbol{g}'_{\beta}, \boldsymbol{g}'_{\phi}, g_{\tau})'$ be the semiparametric estimating function for $\boldsymbol{\theta}_S = (\boldsymbol{\beta}', \boldsymbol{\phi}', \tau)'$. Then the semiparametric estimator, $\tilde{\boldsymbol{\theta}}_S$, is obtained as the solution to the estimating equation $\boldsymbol{g}_S = \mathbf{0}$.

In the simple case when there is no dropout, i.e., all subjects have the same final follow-up-time, T_s, the estimators of β_r and ϕ_j are

$$\begin{aligned} e^{\tilde{\phi}_j} &= \frac{n_{+j}}{\sum_{i=1}^{M} \exp(\boldsymbol{x}'_i \tilde{\boldsymbol{\beta}})} \\ e^{\tilde{\beta}_r} &= \frac{[n_{i+}]_{r+}}{m_r \tilde{\Lambda}_0(T_s)} \end{aligned}$$

where $n_{+j} = \sum_{i=1}^{M} n_{ij}$ and $[n_{i+}]_{r+} = \sum n_{i+}$ for the subjects in Group r, for $r = 2, \ldots, k$ and $j = 1, 2, \ldots, s$. Note that these estimators do not depend upon τ.

Subject to the standard regularity assumptions for asymptotic results in estimating functions (Inagaki 1973; White 1982), the asymptotic distribution of $\sqrt{M}(\tilde{\boldsymbol{\theta}}_S - \boldsymbol{\theta}_S)$ will be normal with variance

$$A^{-1}B(A^{-1})', \tag{5.3}$$

where

$$A = \lim_{M\to\infty} \frac{1}{M} \mathrm{E}\left(-\frac{\partial \boldsymbol{g}_S}{\partial \boldsymbol{\theta}_S}\right)$$

and

$$B = \lim_{M\to\infty} \frac{1}{M} \mathrm{E}\left(\boldsymbol{g}_S \boldsymbol{g}_S'\right).$$

The regularity conditions essentially specify that the matrix (5.3) is positive definite. Note that these results require the number of follow-up times to be fixed as M increases.

The asymptotic variance of $\tilde{\boldsymbol{\theta}}_S$ has the same form as the variance matrix for $\tilde{\boldsymbol{\theta}}_t$ under the parametric intensity in Section 4.2.1, with the parameter $\boldsymbol{\phi}$ entering in place of the parameter $\boldsymbol{\alpha}$. Thus the assumptions required for consistent estimation of the variance of $\tilde{\gamma}_S = (\tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\phi}}')$ are only that the mean and variance of the counts have been correctly specified. If the third and fourth moments of the distribution of subject-specific effects ν_i are the same as those of a gamma random variable, the estimates of $(\boldsymbol{\beta}', \boldsymbol{\phi}')'$ are asymptotically independent of the estimator of τ. That is, assuming that the first to fourth order moments of ν_i are the same as a gamma random variable with mean 1 and variance τ, then the asymptotic variance of $\tilde{\boldsymbol{\theta}}_S$ is

$$\mathrm{AsVar}(\tilde{\theta}_t) = \begin{pmatrix} A_S^{-1} & \mathbf{0} \\ \mathbf{0}' & v_\tau \end{pmatrix} \tag{5.4}$$

where A_S is $\mathrm{E}\left(-\partial \boldsymbol{g}_{\gamma S}/\partial \gamma_S\right)$, writing $\boldsymbol{g}_{\gamma S} = (\boldsymbol{g}_\beta, \boldsymbol{g}_\phi')'$, and v_τ, (4.19), is the asymptotic variance of $\tilde{\tau}$, unchanged from earlier chapters. Note that the asymptotic variance of the quasi-likelihood estimator of the treatment effect β_r is the same as for the semiparametric, likelihood estimator. Thus, the quasi-likelihood estimator is fully efficient relative to the likelihood estimator, as was demonstrated for the parametric intensity models of the previous two chapters.

A finite sample estimator of this variance can be obtained in the usual ways by finding consistent estimators for the matrices A and B. Writing $\tilde{A} = \mathrm{E}(-\partial \boldsymbol{g}_S/\partial \boldsymbol{\theta}_S)\,|_{\tilde{\theta}}$ and $\tilde{B} = \mathrm{E}\left(\boldsymbol{g}_S \boldsymbol{g}_S'\right)|_{\tilde{\theta}}$ where the tilde indicates a finite sample estimator of A or B, evaluated at $\tilde{\boldsymbol{\theta}}$. Using these estimators leads to the model-based variance estimator

$$\tilde{A}^{-1}\tilde{B}(\tilde{A}^{-1})';$$

using the observed $\boldsymbol{g}_S\boldsymbol{g}_S'$ as an empirical estimate of B results in the robust variance estimator

$$\tilde{A}^{-1}(\boldsymbol{g}_S\boldsymbol{g}_S')(\tilde{A}^{-1})'.$$

5.3 Test for a Specific Parametric Baseline Intensity Model

It will often be of interest to test whether the data can be accurately modeled by a specific parametric intensity function. We will construct a quasi-score test using the same basic principles as used in Section 4.4.1. The test is based on the estimating function $\boldsymbol{g}_\phi$, (5.3), and is similar to likelihood score testing procedures (Breslow 1990). For consistency with earlier sections, the test will be constructed for the Weibull intensity function, but we later point out that the technique is more generally applicable.

The null hypothesis to be tested is that the baseline intensity function is Weibull, H_{0W} : $\Lambda_0(t) = t^\alpha$. Using the nonparametric baseline for $\Lambda_0(t)$, this can be rewritten

$$H_{0W} : e^{\phi_j}(T_j - T_{j-1}) = e^{\beta_1}(T_j^\alpha - T_{j-1}^\alpha), \quad j = 1, 2, \ldots, s,$$

where s is the number of follow-up periods and $T_0 = 0$. We first obtain estimates $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}', \tilde{\alpha}', \tilde{\tau})$ under the null hypothesis, as described in Section 4.2.2. The event counts are estimated under H_{0W} as $\tilde{\mu}_{ij}^* = e^{\tilde{\beta}_1}(T_j^{\tilde{\alpha}} - T_{j-1}^{\tilde{\alpha}})$ where the superscript "*" indicates estimates under H_{0W}. The fit of the parametric model is then evaluated by calculating the value of the estimating function for ϕ,

$$\begin{aligned} g_{\phi_j}^* &= \sum_{i=1}^{M}\left(n_{ij} - \frac{n_{i+}}{\tilde{\mu}_{i+}^*}\tilde{\mu}_{ij}^*\right) + \sum_{i=1}^{M}\frac{n_{i+} - \tilde{\mu}_{i+}^*}{\tilde{\mu}_{i+}^*(1+\tilde{\tau}^*\tilde{\mu}_{i+}^*)}\tilde{\mu}_{ij}^* \\ &= \sum_{i=1}^{M}\left\{n_{ij} - n_{i+}\frac{T_j^{\tilde{\alpha}} - T_{j-1}^{\tilde{\alpha}}}{T_{s_i}^{\tilde{\alpha}}}\right\} \end{aligned}$$

for $j = 1, 2, \ldots, s$. The asymptotic variance for the estimating function $\boldsymbol{g}_\phi^*$ is

$$V_\phi = B_{22} - A_{21}A_{11}^{-1}B_{12} - B_{21}A_{11}^{-1}A_{12} + A_{21}A_{11}^{-1}B_{11}A_{11}^{-1}A_{12}$$

where the matrices $A = \mathrm{E}(-\partial\boldsymbol{g}_S/\partial\boldsymbol{\theta}_S)$ and $B = \mathrm{E}(\boldsymbol{g}_S'\boldsymbol{g}_S)$ have been partitioned appropriately into submatrices corresponding to the vectors $(\boldsymbol{\beta}', \tau)'$ and ϕ. An estimator of V_ϕ is available via either of the usual methods: a model based estimator is found by evaluating A and B at the estimates found under H_{0W}; a robust estimator is found by substituting $(\boldsymbol{g}_S'\boldsymbol{g}_S)$, the empirical estimator of B. The test statistic for H_{0W} is then

$$\boldsymbol{g}_\phi^{*\prime}\tilde{V}_\phi^{-1}\boldsymbol{g}_\phi^*$$

which we will compare to a χ^2 random variable with $s-2$ d.f.

To further illustrate the construction of this type of test, consider a test of the fit of the exponential intensity model (i.e., that the Poisson processes are homogeneous). This test would be conducted by finding estimates of $\boldsymbol{\beta}$ and τ for the Weibull model with α fixed at 1. Then the fit is evaluated by calculating $\boldsymbol{g}_{\phi}^{*}$ at these estimates. The resulting test statistic will then have $s-1$ d.f. Other intensity functions would require similar modifications to the test procedure.

These tests will be demonstrated in Section 5.5, and their small sample performance will be examined in Section 5.6.

5.4 Efficiency of Semiparametric vs. Parametric Estimators

By replacing the parametric baseline intensity model, which uses the parameters β_1 and α, with the nonparametric model, which uses s parameters, $\phi_1,\ldots,\phi_s$, the analysis has been made less model dependent. However, this freedom may come at the expense of information otherwise available for the estimation of the covariate effects. Since the primary goal of many experiments is to estimate treatment effects, we will now examine the efficiency of the semiparametric estimator of the treatment effect β_j. For this purpose, we define

$$\mathrm{ARE}(\tilde{\beta}_{r\mathrm{S}}) = \frac{\mathrm{AsVar}(\tilde{\beta}_{r\mathrm{W}})}{\mathrm{AsVar}(\tilde{\beta}_{r\mathrm{S}})}, \quad r = 2,\ldots,k, \tag{5.5}$$

where $\mathrm{AsVar}(\tilde{\beta}_{r\mathrm{W}})$ is the variance of the quasi-likelihood estimator of the treatment effect based on a Weibull intensity model as described in Section 4.2.2.

As described in Section 5.2, if the final follow-up-times, T_{s_i} are the same for all subjects, then the parametric and semiparametric estimates of β_r are the same, and the semiparametric estimators have 100% efficiency. Also, when there are only two follow-up times, requiring ϕ_1 and ϕ_2, the semiparametric specification of the intensity function is a simple reparameterization of the parametric intensity function in terms of β_1 and α, so the semiparametric estimators of β_r will have 100% efficiency. Finally, since we showed in Section 3.3 that the variance of $\boldsymbol{\theta}$ has the same form for a baseline intensity function parameterized by a scalar α as for an arbitrary vector $\boldsymbol{\alpha}$, we note that the asymptotic variance results of Section 3.3 can be applied for the nonparametric baseline intensity.

We calculated this efficiency for a variety of two group designs, for which the true baseline intensity function was Weibull. The study considered combinations of the following factors:

Factor	Levels
Group Size	30 or 60
Follow-Up Times	2 follow-up times: {4, 8};
	4 follow-up times: {2, 4, 6, 8}; or
	8 follow-up times: {1,2,3,4,5,6,7,8}
Dropouts	no dropouts, or 50% dropout at time 4
β_1	-2, 0
β_2	-1.5, 0.5, 0, 0.5, 1.5
α	0.7, 1.0, 1.3
τ	0, 0.8, 1.5

Table 5.1: Factors Considered in ARE Calculations for Semiparametric Estimator β_{2S}.

group size, number of follow-up times, dropouts, and values of the parameters β and τ. Table 5.1 shows the factors levels used in the construction of the designs studied. The study did not consider designs with unequal numbers of follow-up times in the two groups, but did allow certain imbalanced designs, i.e., designs with different dropout rates and sample sizes.

Tables 5.2 and 5.3 present the asymptotic relative efficiency values for designs constructed with parameter values: $\beta_1 = 0$, $\alpha = 1$, and $\tau = 0$ and 0.8. The effect of changes in the parameters β_1 and α were small, and the following comments apply across values of these two parameters, except as noted below.

In summary, the efficiency of the semiparametric estimator, $\tilde{\beta}_{2S}$, is greater than 75% for the two-group panel designs we examined. When the groups were balanced with respect to follow-up times (i.e., when the groups were the same size with the same dropouts, or when the groups were of different sizes with no dropouts), the semiparametric estimator of the treatment effect was 100% efficient. The lowest efficiency observed was 0.7526 when group 1 was large with no dropouts and group 2 was small with 50% dropouts, and the treatment effect was $\beta_2 = 1.5$. In general, efficiencies were lower when the group associated with β_2 (i.e., the group designated as the "treatment" group) was the smaller of the two groups. The absolute sample sizes do not affect the efficiency figures, only relative sample sizes mattered; e.g., the efficiencies for group sizes (30, 30) are the same as for the situation with group sizes (60,60). Increasing the number of follow-ups had only very small effect on efficiency for inference regarding the treatment effect, β_2.

# of Follow-Up Times	Group Size		Dropout		Asymptotic Relative Efficiency of $\tilde{\beta}_{2S}$ Treatment Effect β_2				
s	m_1	m_2	Group 1	Group 2	−1.5	−0.5	0	0.5	1.5
2	any	any	any	any	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	30	none	50%	0.9951	0.9909	0.9893	0.9887	0.9913
4	30	30	50%	none	0.9913	0.9887	0.9893	0.9909	0.9951
4	30	30	50%	50%	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	60	none	50%	0.9985	0.9981	0.9983	0.9985	0.9992
4	30	60	50%	none	0.9890	0.9898	0.9917	0.9939	0.9972
4	30	60	50%	50%	0.9963	0.9963	0.9969	0.9976	0.9989
4	60	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	60	30	none	50%	0.9815	0.9528	0.9270	0.8904	0.7847
4	60	30	50%	none	0.9942	0.9901	0.9889	0.9889	0.9922
4	60	30	50%	50%	0.9846	0.9612	0.9406	0.9124	0.8363
4	60	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	60	60	none	50%	0.9951	0.9909	0.9893	0.9887	0.9913
4	60	60	50%	none	0.9913	0.9887	0.9893	0.9909	0.9951
4	60	60	50%	50%	1.0000	1.0000	1.0000	1.0000	1.0000
8	30	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	30	30	none	50%	0.9937	0.9885	0.9865	0.9858	0.9891
8	30	30	50%	none	0.9891	0.9858	0.9865	0.9885	0.9937
8	30	30	50%	50%	1.0000	1.0000	1.0000	1.0000	1.0000
8	30	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	30	60	none	50%	0.9981	0.9976	0.9978	0.9981	0.9990
8	30	60	50%	none	0.9862	0.9871	0.9895	0.9922	0.9964
8	30	60	50%	50%	0.9954	0.9953	0.9960	0.9970	0.9986
8	60	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	60	30	none	50%	0.9764	0.9408	0.9096	0.8669	0.7526
8	60	30	50%	none	0.9928	0.9876	0.9860	0.9860	0.9901
8	60	30	50%	50%	0.9811	0.9532	0.9293	0.8976	0.8181
8	60	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	60	60	none	50%	0.9937	0.9885	0.9865	0.9858	0.9891
8	60	60	50%	none	0.9891	0.9858	0.9865	0.9885	0.9937
8	60	60	50%	50%	1.0000	1.0000	1.0000	1.0000	1.0000

Table 5.2: Relative Efficiency of the Semiparametric Estimator $\tilde{\beta}_{2S}$ for $\tau = 0$. Efficiency is measured relative to the quasi-likelihood estimator based on the Weibull intensity, under the Weibull intensity model. The panel studies considered here have $s = 2$, 4, or 8 follow-ups during the same period of time, with either no dropouts or 50% dropouts at the middle follow-up time. Parameter values chosen are $\beta_1 = 0, \alpha = 1, \tau = 0$. The ARE($\tilde{\beta}_{2S}$) is 100% for any study with $s = 2$ follow-up times.

Asymptotic Relative Efficiency of $\tilde{\beta}_{2S}$

# of Follow Up Times	Group Size		Dropout		Treatment Effect β_2				
s	m_1	m_2	Group 1	Group 2	−1.5	−0.5	0	0.5	1.5
2	any	any	any	any	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	30	none	50%	0.9965	0.9947	0.9942	0.9943	0.9958
4	30	30	50%	none	0.9908	0.9925	0.9942	0.9959	0.9982
4	30	30	50%	50%	0.9998	1.0000	1.0000	1.0000	1.0000
4	30	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	30	60	none	50%	0.9991	0.9989	0.9990	0.9991	0.9995
4	30	60	50%	none	0.9901	0.9938	0.9956	0.9971	0.9988
4	30	60	50%	50%	0.9960	0.9976	0.9984	0.9990	0.9996
4	60	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	60	30	none	50%	0.9875	0.9777	0.9719	0.9654	0.9502
4	60	30	50%	none	0.9931	0.9929	0.9939	0.9952	0.9977
4	60	30	50%	50%	0.9905	0.9834	0.9794	0.9753	0.9669
4	60	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
4	60	60	none	50%	0.9965	0.9947	0.9942	0.9943	0.9958
4	60	60	50%	none	0.9908	0.9925	0.9942	0.9959	0.9982
4	60	60	50%	50%	0.9998	1.0000	1.0000	1.0000	1.0000
8	30	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	30	30	none	50%	0.9965	0.9954	0.9953	0.9956	0.9970
8	30	30	50%	none	0.9914	0.9937	0.9953	0.9967	0.9986
8	30	30	50%	50%	0.9999	1.0000	1.0000	1.0000	1.0000
8	30	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	30	60	none	50%	0.9991	0.9990	0.9991	0.9993	0.9996
8	30	60	50%	none	0.9911	0.9947	0.9964	0.9976	0.9991
8	30	60	50%	50%	0.9965	0.9981	0.9987	0.9992	0.9997
8	60	30	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	60	30	none	50%	0.9880	0.9817	0.9785	0.9751	0.9676
8	60	30	50%	none	0.9934	0.9940	0.9950	0.9962	0.9982
8	60	30	50%	50%	0.9916	0.9874	0.9855	0.9836	0.9802
8	60	60	none	none	1.0000	1.0000	1.0000	1.0000	1.0000
8	60	60	none	50%	0.9965	0.9954	0.9953	0.9956	0.9970
8	60	60	50%	none	0.9914	0.9937	0.9953	0.9967	0.9986
8	60	60	50%	50%	0.9999	1.0000	1.0000	1.0000	1.0000

Table 5.3: Asymptotic Relative Efficiency of the Semiparametric Estimator $\tilde{\beta}_{2S}$ for $\tau = 0.8$. Asymptotic Efficiency is measured relative to the quasi-likelihood estimator based on the Weibull intensity, under the Weibull intensity model. The panel studies considered here have $s = 2$, 4, or 8 follow-ups during the same period of time, with either no dropouts or 50% dropouts at the middle follow-up time. Parameter values chosen are $\beta_1 = 0, \alpha = 1, \tau = 0.8$. The ARE($\tilde{\beta}_{2S}$) is 100% for any study with $s = 2$ follow-up times.

Other than the treatment effect, β_2, the parameters β_1, α, and τ had little effect on the results in this study. More specifically, the value of β_1 had no effect on the relative efficiency of the treatment effect estimator. Allowing for overdispersion, i.e., $\tau > 0$, increased efficiencies somewhat for imbalanced designs, but had no effect when the designs were balanced. When $\alpha < 1$ the effect of imbalance is similarly reduced; conversely, setting $\alpha > 1$ increased the effect of imbalance. The effect of β_2 on the efficiency is most clearly seen by comparing the results for two designs with the same number of follow-ups and group sizes, but with opposite dropout patterns, e.g., the third and fourth lines of Table 5.2. Here we can see that when the group with 50% dropouts is the group with the higher intensity, the efficiency of $\tilde{\beta}_{2S}$ will be lower than when subjects drop out from the lower intensity treatment group.

Overall, $\tilde{\beta}_{2S}$ is more efficient when the expected numbers of recurrences observed for the treatment group increases. So, factors reducing the relative size of the expected number of recurrences to be observed, such as dropouts, designation of treatment group, size of treatment effect, and the relative sizes of the groups, can all reduce the efficiency of the semiparametric estimator of the treatment effect in this study.

An alternative way to assess the efficiency of the semiparametric estimator is to compare the standard errors of the predicted cumulative means for the follow-up times in the design. For example, in the bladder cancer data, the standard error of the estimated cumulative mean for the placebo group (i.e., the cumulative baseline intensity) at 36 months is 0.342 based on the semiparametric analysis, only slightly wider than the standard error of 0.332 for the parametric (Weibull) estimate. These comparisons will be examined again in Section 5.5.1.

5.5 Illustrations

5.5.1 Analysis of the Bladder Cancer Data

To demonstrate the use of the nonparametric baseline intensity model, we return again to the bladder cancer data examined in Chapters 3 and 4. Recall that the data were collected as event-time data from a clinical trial designed to compare the efficacy of three treatments for recurring bladder cancer. The subjects, all of whom had been treated for bladder cancer, were randomly assigned to one of three treatment groups: (1) Placebo, (2) Pyridoxine, and (3) Thiotepa. In total, 116 subjects were followed for just over five years, recording the times of recurrence of bladder cancer tumours. The data have been modified to fit a panel data

structure corresponding to yearly follow-up times. The nonparametric intensity model will be fit to this modified data, and contrasted with the analysis based on the Weibull intensity model obtained in Chapter 4. These two analyses will be referred to as the Nonparametric Intensity Analysis (or "Semiparametric Analysis") and the Weibull Intensity Analysis (or "Parametric Analysis"). Summary statistics for the modified data have been presented in Sections 3.6 and 4.5.

The parametric and semiparametric analyses conducted as described in Sections 4.2.2 and 5.2 result in the estimates displayed in Table 5.4. There is little difference between the

Intensity Model		Estimate	Model-Based Std. Err.	Robust Std. Err.
Weibull	β_2	-0.073	(0.330)	(0.332)
	β_3	-0.293	(0.315)	(0.310)
	β_1	-2.596	(0.376)	(0.342)
	α	0.923	(0.093)	(0.089)
	τ	-0.989	(0.338)	(0.245)
Nonparametric	β_2	-0.081	(0.325)	(0.315)
	β_3	-0.302	(0.310)	(0.309)
	ϕ_1	-2.790	(0.218)	(0.180)
	ϕ_2	-2.874	(0.234)	(0.206)
	ϕ_3	-2.932	(0.267)	(0.217)
	ϕ_4	-2.958	(0.362)	(0.284)
	ϕ_5	-3.212	(0.789)	(0.696)
	τ	0.939	(0.338)	(0.236)

Table 5.4: Parameter Estimates for the Bladder Cancer Data. Both parametric (Weibull) and nonparametric intensity models have been fit. The standard errors were calculated from both the model-based and robust variance estimates. The data have been modified to fit the yearly follow-up panel data structure.

two analyses with respect to the estimate of the treatment effects, β_2 and β_3, as well as for the overdispersion parameter, τ. This suggests that the two intensity models fit the data equally well.

The high efficiency of the semiparametric analysis when compared to the parametric can be seen in Figure 5.1 and Table 5.5. These display the parametric and semiparametric approximate 95% pointwise confidence intervals for the cumulative baseline mean function.

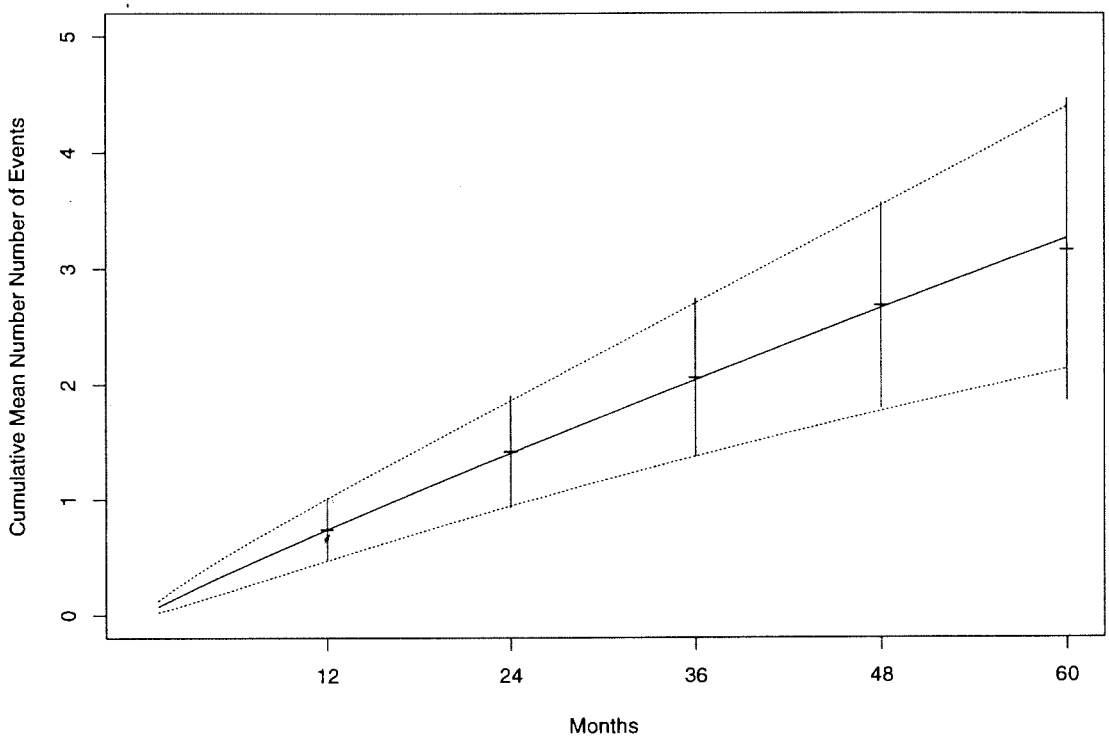

Figure 5.1: Estimated Baseline Cumulative Mean Functions For the Bladder Cancer Recurrent Event Data. Confidence limits shown are ± 2 standard errors for the estimated cumulative mean at time T_j. See also Table 5.5. The smooth curve and the dotted lines give the estimated cumulative mean function and approximate 95% pointwise confidence interval for the Weibull baseline intensity. The vertical lines and their crossbars show the approximate 95% pointwise confidence intervals and the point estimate for the cumulative mean number of events at the yearly follow-up times. Standard errors are based on robust variances.

The high efficiency of the semiparametric estimator is evident in the similarity of the widths of the confidence intervals at the first few follow-up times. The semiparametric intervals become wider for the last few follow-up times, but this is due to decreased information available because of dropouts. The mean at time T_j has been estimated by $T_j^{\tilde{\alpha}}$ for the Weibull analysis and by $\sum_{h=1}^{j} e^{\tilde{\phi}_j}$ for the semiparametric analysis. The usual asymptotic normal theory intervals were used, with variances found by delta-method approximations based on the robust covariance estimates of the original parameter vectors. Similar results are found when the model-based covariance estimates are used.

Further evidence that the Weibull model is reasonable can be seen in the similarity of the fitted means from the analyses based on the Weibull and nonparametric intensity models shown in Figure 5.2 and Table 5.5. Used in this manner, the semiparametric estimates of

Estimated Cumulative Means (Standard Errors)

Month	Weibull		Semiparametric	
12	0.738	(0.134)	0.737	(0.132)
24	1.400	(0.229)	1.415	(0.242)
36	2.035	(0.332)	2.054	(0.342)
48	2.654	(0.445)	2.678	(0.444)
60	3.262	(0.566)	3.161	(0.653)

Table 5.5: Estimates of the Baseline Cumulative Mean Functions for the Bladder Cancer Data. Standard errors (in parentheses) are calculated from the robust variances.

the cumulative means provide a useful diagnostic tool, which in this case supports the use of the proposed parametric intensity model. In contrast, the analysis of the Cherry Bark Tortrix data in the next subsection will emphasize the flexibility of the semiparametric approach for data which do not fit any of the usual parametric intensity models.

5.5.2 Analysis of the Cherry Bark Tortrix Data

This experiment was conducted to test the effectiveness of pheromone-based mating disruption of the Cherry Bark Tortrix moth (*Enarmonia fomosana*). The pheromone used was known to be competitive with caged virgin females in attracting males into traps. It was expected that males would become disoriented by the release of female scent, and would then be unable to locate and mate with any female moth in the tree.

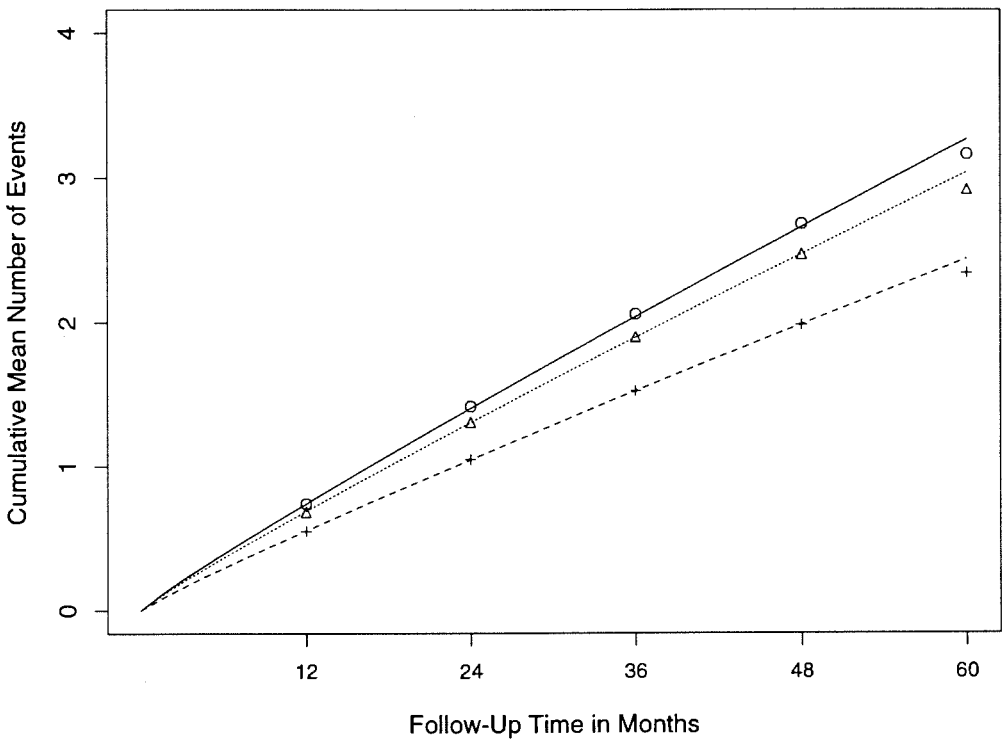

Figure 5.2: Estimates of the Cumulative Mean Functions of the Bladder Cancer Data. The smooth curves represent estimates based on the Weibull intensity model. The symbols represent the semiparametric estimates (solid line and ○: placebo; dotted line and △: pyridoxine; dashed line and +: thiotepa).

Twenty cherry trees were fitted with pheromone dispensers. Ten of the trees were selected at random and their dispensers were loaded with the pheromone; the remaining ten trees were used as a control. One trap was placed in a similar location in each of the twenty trees and baited with the female pheromone. Approximately once per week the traps were emptied of males, for a total of 19 follow-ups in 18 weeks. Once every three weeks, the baits were refreshed. It was hoped that the number of males caught in traps on the treatment trees would be lower than caught in traps on the control trees, suggesting that the treatment was effective at confusing the mate-seeking males.

Figure 5.3 displays the observed and estimated cumulative mean functions for this experiment. The treatment has substantially reduced the number of males caught, from a mean catch per tree of 141.2 in the control trees to 5.1 in the treatment trees.

While it is interesting to observe the effectiveness of the treatment, our primary interest is in comparing the semiparametric and fully parametric analyses. The biology of the species suggests that the number of males will vary over time, suggesting the use of a nonhomogeneous Poisson model rather than a homogeneous process. We consider the use of the Weibull intensity model for our parametric approach.

The parametric and semiparametric estimates of the treatment effect are shown in Table 5.6. Quasi-likelihood estimating functions were used for both analyses. Estimates of the parameters that appear in both of the models, β_2 and τ, are very similar; of course, the estimates of β_2 must be identical when there are no dropouts (see Section 5.4). Standard errors are also similar.

It is possible to apply the test procedures for the fits of specific intensity models to these data (Section 5.3). The test statistics for both the null hypothesis that the data follow a Weibull intensity model and the null hypothesis that the data follow an exponential intensity model result are strongly rejected, with observed significance $p < 0.0001$. The test statistics had 17 and 18 d.f., respectively. Testing for overdispersion, as described in Section 4.4.1, yields a test statistic with observed significance $p < 0.0001$.

Figures 5.4 and 5.5 display the estimated baseline and cumulative baseline intensity functions from the parametric and semiparametric analyses. The Weibull parameter α is estimated to be approximately 1, suggesting nearly constant intensity. However, it is clear that the intensity is changing over time. It is also clear that the Weibull model is not sufficiently flexible to adequately describe the changes observed in the means. The results from the semiparametric model suggests that the male moths present themselves in

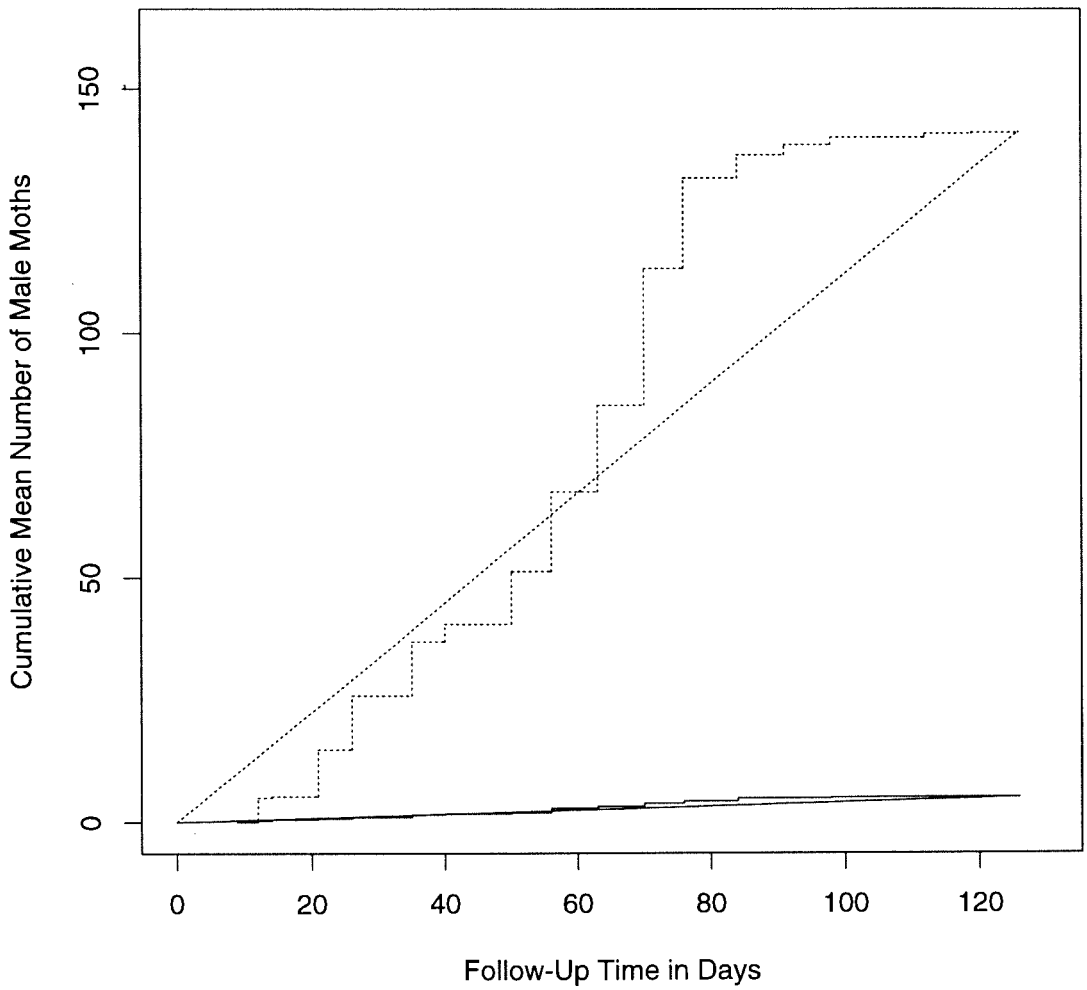

Figure 5.3: Estimated Cumulative Intensities for Cherry Bark Tortrix Data. The smooth curves represent the estimated cumulative intensity functions based on the Weibull baseline intensity function; the step functions represent the semiparametric estimates. The treatment group (solid lines) has a dramatically lower cumulative intensity function than the control group (dotted lines).

Intensity Model		Estimate	Model-Based Std. Err.	Robust Std. Err.
Weibull	β_1	-3.204	(0.315)	(0.260)
	β_2	3.321	(0.381)	(0.343)
	α	0.999	(0.027)	(0.031)
	τ	0.623	(0.285)	(0.209)
Nonparametric	β_2	3.321	(0.364)	(0.343)
	ϕ_1	-6.470	(0.570)	(0.675)
	ϕ_2	-2.866	(0.309)	(0.240)
	ϕ_3	-4.743	(0.524)	(0.613)
	ϕ_4	-3.030	(0.292)	(0.238)
	ϕ_5	-2.524	(0.289)	(0.234)
	ϕ_6	-3.111	(0.289)	(0.252)
	ϕ_7	-3.631	(0.318)	(0.366)
	ϕ_8	-3.279	(0.290)	(0.257)
	ϕ_9	-2.303	(0.284)	(0.265)
	ϕ_{10}	-2.406	(0.284)	(0.250)
	ϕ_{11}	-1.949	(0.280)	(0.231)
	ϕ_{12}	-2.209	(0.283)	(0.215)
	ϕ_{13}	-3.749	(0.306)	(0.269)
	ϕ_{14}	-4.560	(0.350)	(0.306)
	ϕ_{15}	-4.832	(0.371)	(0.262)
	ϕ_{16}	-7.605	(1.037)	(1.207)
	ϕ_{17}	-5.525	(0.447)	(0.456)
	ϕ_{18}	-6.912	(0.758)	(0.623)
	ϕ_{19}	-6.912	(0.758)	(0.604)
	τ	0.559	(0.288)	(0.194)

Table 5.6: Parameter Estimates for Cherry Bark Tortrix Data. Both parametric (Weibull) and nonparametric intensity models have been fit. The standard errors were calculated from the model-based or robust variance estimates.

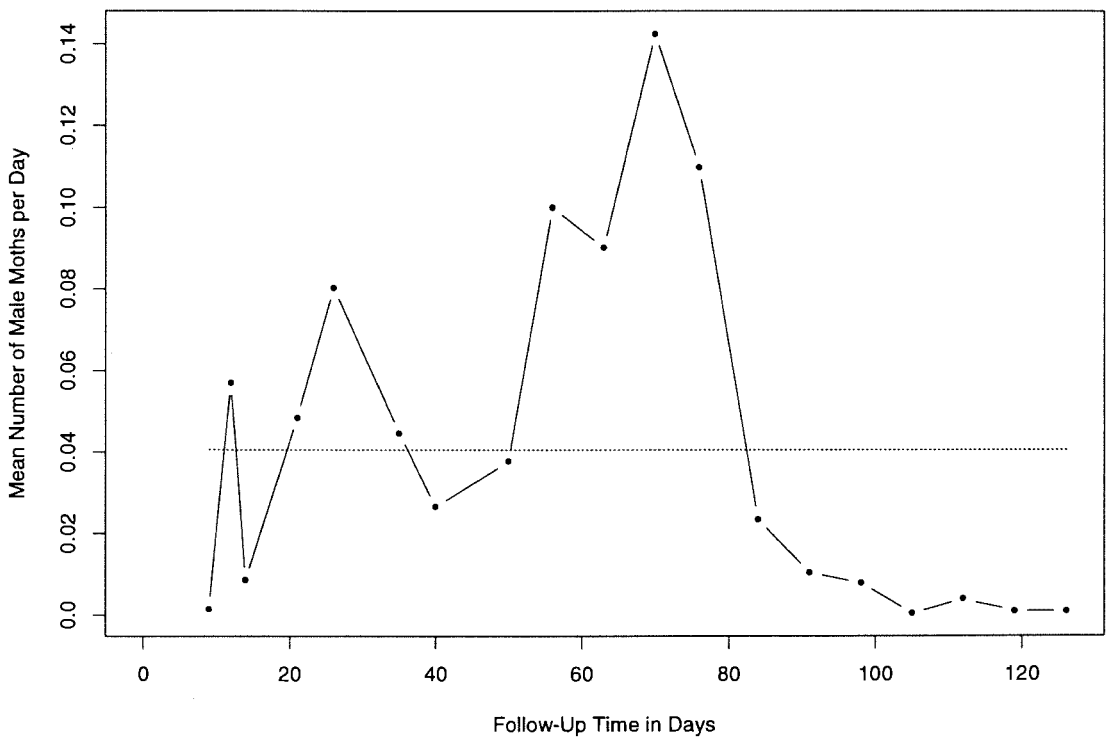

Figure 5.4: Estimated Baseline Intensity Function for Cherry Bark Tortrix Data. The semiparametric estimates (•) are $e^{\tilde{\phi}_j}(T_j - T_{j-1})$ and the parametric estimates (dotted line) are $e^{\tilde{\beta}_1}(T_j^{\tilde{\alpha}} - T_{j-1}^{\tilde{\alpha}})$. The estimates have been standardized by dividing by the time between follow-ups to give daily figures. The semiparametric estimates have been connected to enhance visibility.

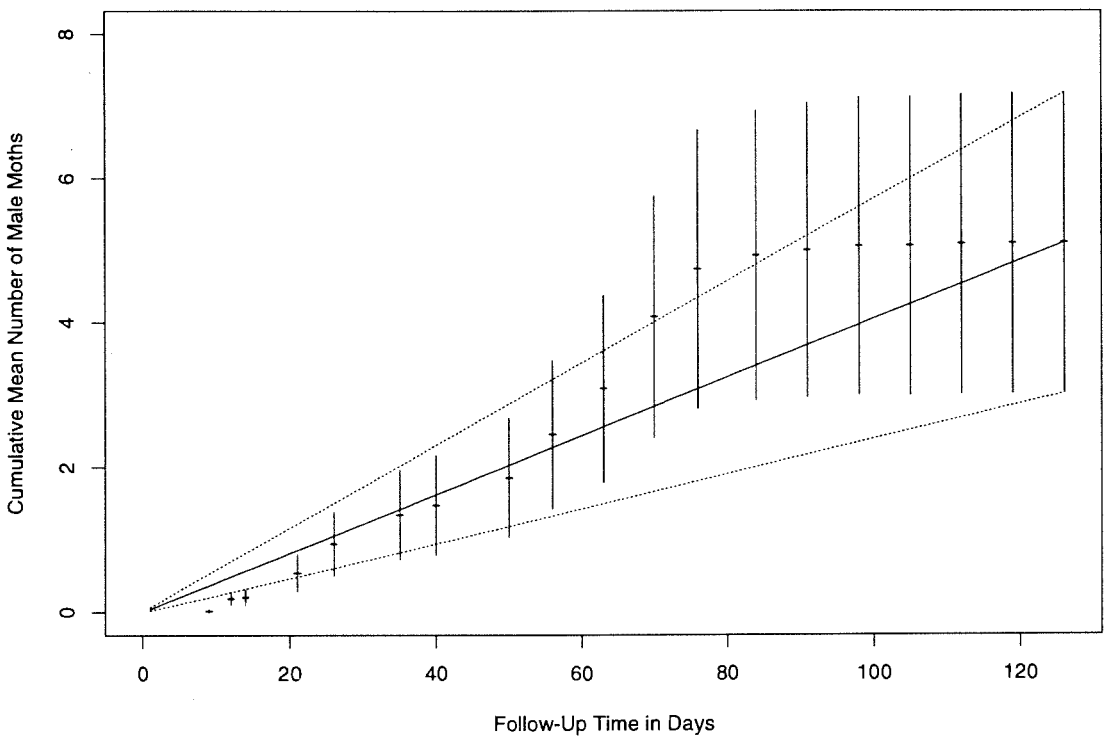

Figure 5.5: Estimated Cumulative Baseline Intensity Function for the Cherry Bark Tortrix Data. The Weibull intensity model estimate (solid line) is plotted with approximate 95% pointwise confidence intervals (dotted lines). The nonparametric intensity model estimates are shown (horizontal marks) at the center of their approximate 95% pointwise confidence intervals (vertical lines). Standard errors are based on robust variances.

two waves; this information was an unexpected bonus for the entomologist involved, and apparently corresponds with the biology of this species. While we did not test this notion in any formal sense, the information was not available from the Weibull intensity model analysis and demonstrates the extra flexibility of the semiparametric approach.

Figure 5.6 displays the residuals, calculated from the end-of-follow-up counts,

$$\frac{n_{i+} - \tilde{\mu}_{i+}}{\sqrt{\tilde{\mu}_{i+}(1 + \tilde{\tau}\tilde{\mu}_{i+})}}$$

as suggested in McCullagh and Nelder (1989, Section 12.5). The residuals appear to have a reasonably normal distribution with little difference in the behaviour of the two groups. This suggests that the variance model $\mu_{i+}(1 + \tau\mu_{i+})$ provides an adequate fit to the data.

5.6 Small Sample Characteristics of the Semiparametric Estimators

To examine the small sample behaviour of the semiparametric estimators and the test statistics developed in the previous sections, we have conducted a small simulation study. Data were generated according to a Weibull intensity model, with parameter values chosen to be similar to those observed in the Bladder Cancer data analysis in Section 5.5 and those used in the previous simulation study in Section 4.2.4. Subject-specific random effects ν_i were generated from a gamma distribution with mean 1 and variance $\tau = 0.8$. For each subject, event times were generated according to the intensity function

$$\lambda_i(t) = \nu_i \exp(-\beta_1 + \beta_2 x_i)\lambda_0(t; \alpha),$$

where $\boldsymbol{\beta} = (-2, 0.1)'$, $\lambda_0(t; \alpha) = \alpha t^{\alpha-1}$ is the Weibull baseline intensity function, and $x_i = 1$ for subjects in group 2, 0 otherwise. Three values of α used were: $\alpha = 0.7$ (decreasing intensity), $\alpha = 1$ (constant intensity), and $\alpha = 1.3$ (increasing intensity). The event times were modified by the imposition of a panel data structure (as described in the previous chapter) with follow-ups at 12, 24, 36, 48, 60, and 72 months. A total of 2000 replications were conducted for each combination of α ($\alpha = 0.7, 1.0, 1.3$) and group size ($m_i = 24, 48, 96$).

The quasi-likelihood estimating equations from Section 4.2.2 were used to find estimates for the Weibull intensity model analysis and the equations from Section 5.2 for the semiparametric model. For both sets of estimates, model-based and robust variance estimates

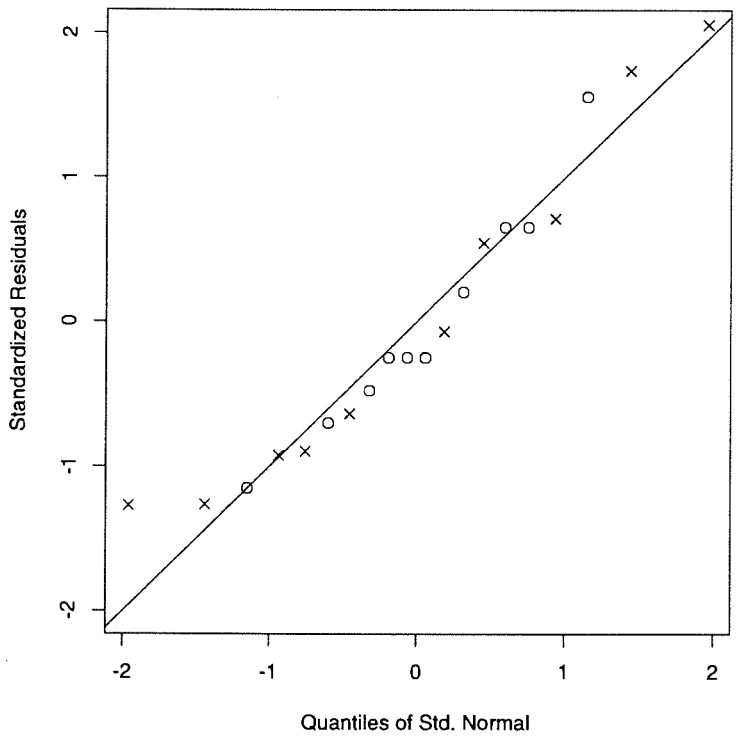

Figure 5.6: Normal Probability Plot for the Standardized End-of-Follow-Up Residuals from the Cherry Bark Tortrix Data. The treatment group is represented by o and the control group by $\times$. The reference line has a slope of 1 and intercept of 0.

were calculated. Similarly, both the model-based and robust forms of the test statistics for the Weibull and constant intensity models were calculated.

The basic Newton-Raphson algorithm used to find the roots of the estimating functions had little difficulty converging on estimates, averaging between 5 and 15 iterations for the Weibull and semiparametric models. However, the constant intensity exponential model (with α set equal to 1) was more difficult to fit when the data were generated using $\alpha = 1.3$; approximately 3% of these analyses failed to converge. Decreasing the step size of the Newton-Raphson algorithm allowed convergence for many of the problematic datasets.

Figures 5.7 through 5.10 show normal probability plots comparing the distributions of the standardized estimators for the Weibull and semiparametric models to the standard normal for the small group size simulation (i.e., $m_i = 24$). The estimators have been standardized by subtracting the true parameter value and dividing by either the model-based or the robust standard error. By standardizing the estimators in this way, the QQ-plot becomes a diagnostic for the accuracy of the mean and variance of the estimators, as well as the shape of the distribution. Because all ϕ parameter estimators behaved similarly, the distributions of only ϕ_2 and ϕ_5 are shown here. Our focus here is to compare the parametric and semiparametric estimators. However, there appears to be little difference in small sample behaviour between the two types of estimators. Overall, both the parametric and semiparametric estimators appear to have near-normal distributions for group sizes as small as 24, except for the overdispersion parameter τ. Because the distributions of the estimators of τ appear to be seriously non-normal for groups of size 24, we have also presented how the sampling distribution of the estimator of τ changes as the group size becomes larger. Figures 5.11, 5.12, 5.13, and 5.14 show normal probability plots of the quasi-likelihood estimator, $\tilde{\tau}$, for group sizes 24, 48, and 96 subjects per group. Two points are apparent. First, the parametric and semiparametric estimators of τ behave similarly – neither estimator appears to approach normality quickly. Second, the distribution of $\tilde{\tau}$ approaches normality more quickly when standardized by dividing by the model-based standard errors than when standardized using the robust standard errors, at least in this situation where the assumed variance model is the correct one. This may be due in part to the extra variability observed in the robust variance estimates, which we describe below (see Tables 5.8 and 5.9).

Table 5.7 summarizes the means and standard errors of the estimators of $\boldsymbol{\theta}_S$. The standard errors used are the square roots of the sampling variances of the estimators in

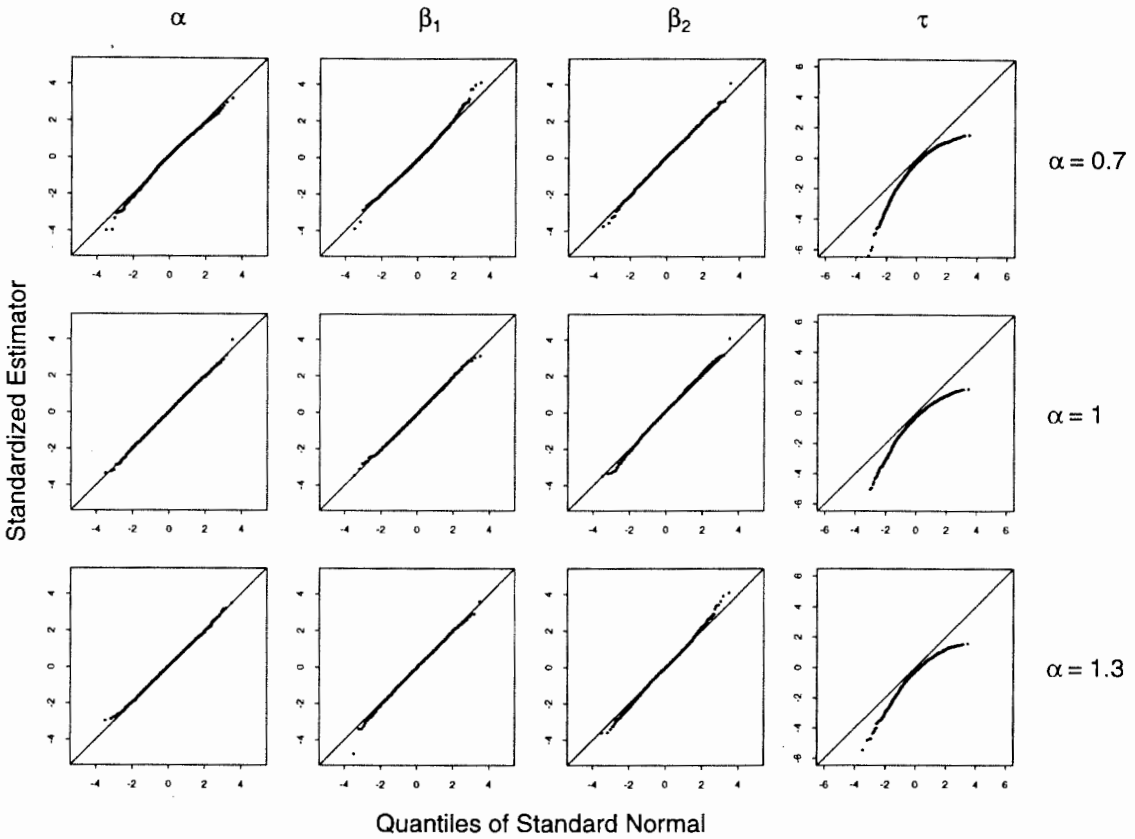

Figure 5.7: Sampling Distribution of the Weibull Intensity Function Model Estimators, Model-Based Standardization. Each panel plots standardized parameter estimates from 2000 simulated data sets against quantiles of the standard normal distribution. Note different scale for τ. The data sets consisted of two groups of 24 subjects and six follow-up times. Each column corresponds to a different parameter; each row shows the results of simulations using a different value of α.

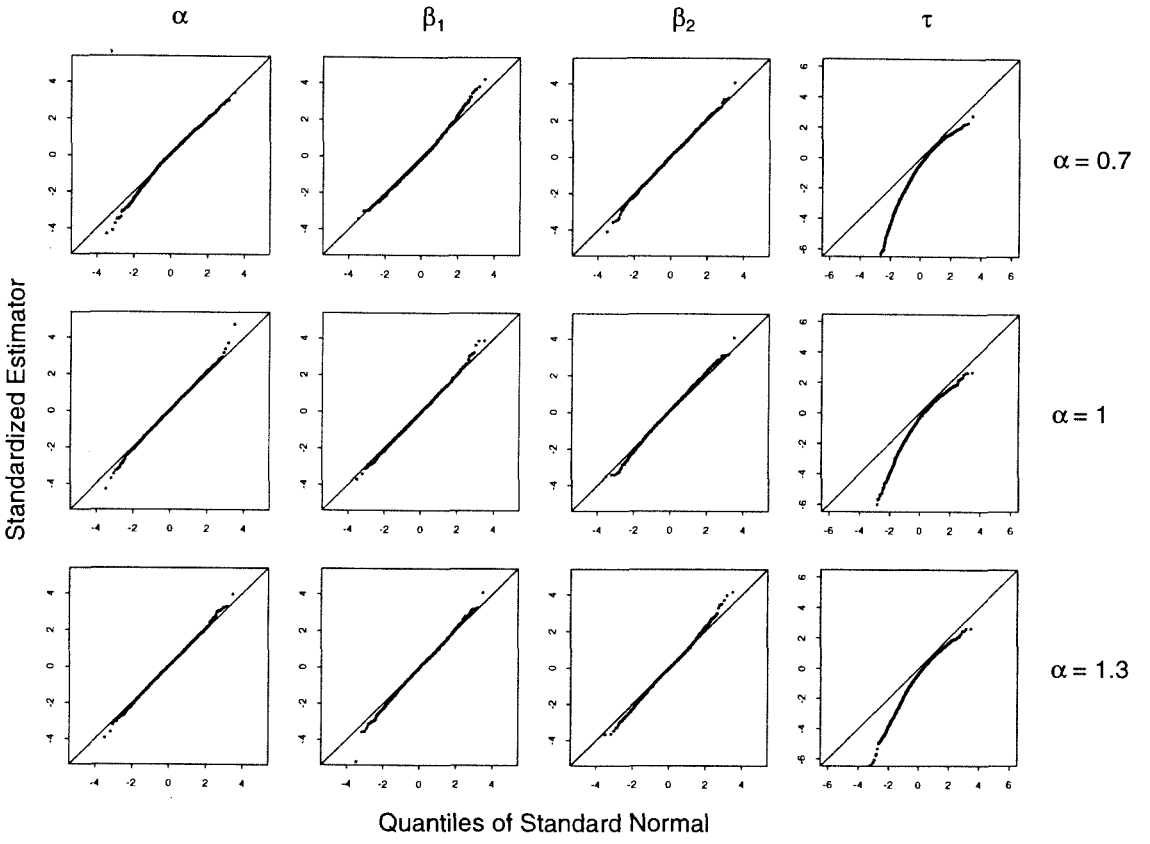

Figure 5.8: Sampling Distribution of the Weibull Intensity Function Model Estimators, Robust Standardization. Each panel plots standardized parameter estimates from 2000 simulated data sets against quantiles of the standard normal distribution. Note different scale for τ. The data sets consisted of two groups of 24 subjects and six follow-up times. Each column corresponds to a different parameter; each row shows the results of simulations using a different value of α.

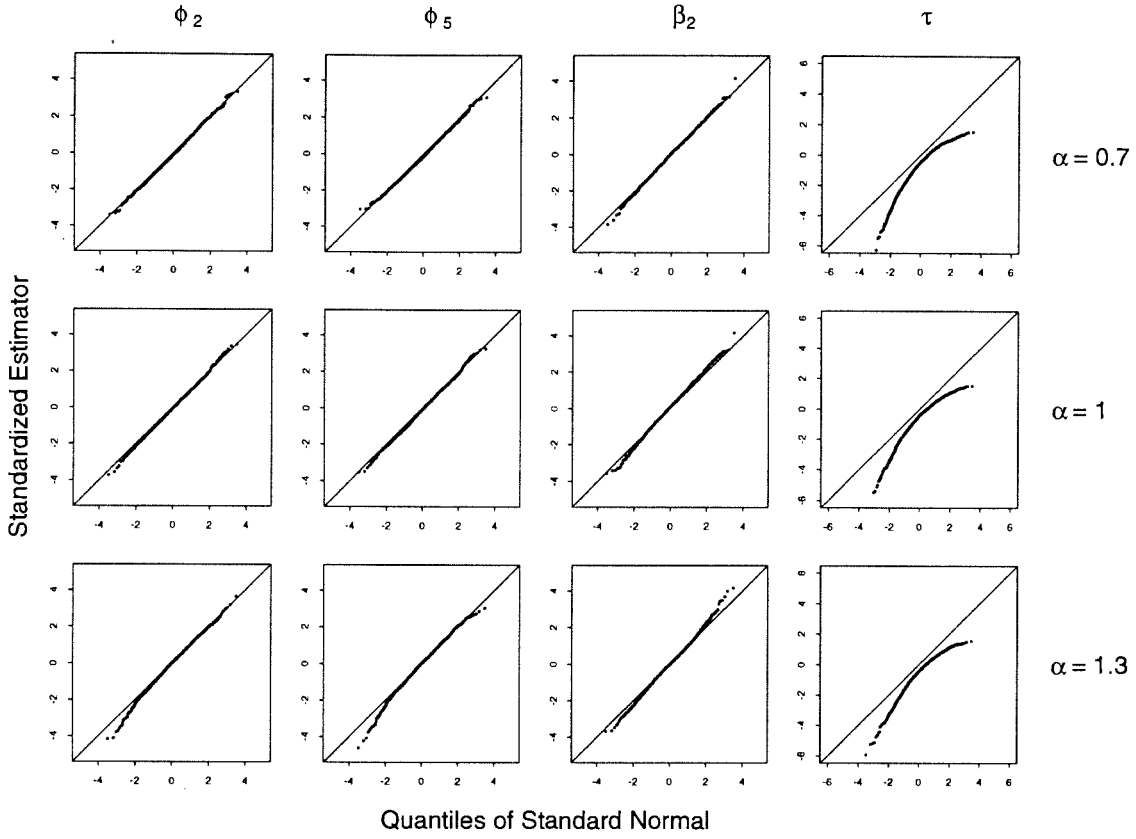

Figure 5.9: Sampling Distribution of the Semiparametric Model Estimators, Model-Based Standardization. Each panel plots standardized parameter estimates from 2000 simulated data sets against quantiles of the standard normal distribution. Note different scale for τ. The data sets consisted of two groups of 24 subjects and six follow-up times. Each column corresponds to a different parameter; each row shows the results of simulations using a different value of α.

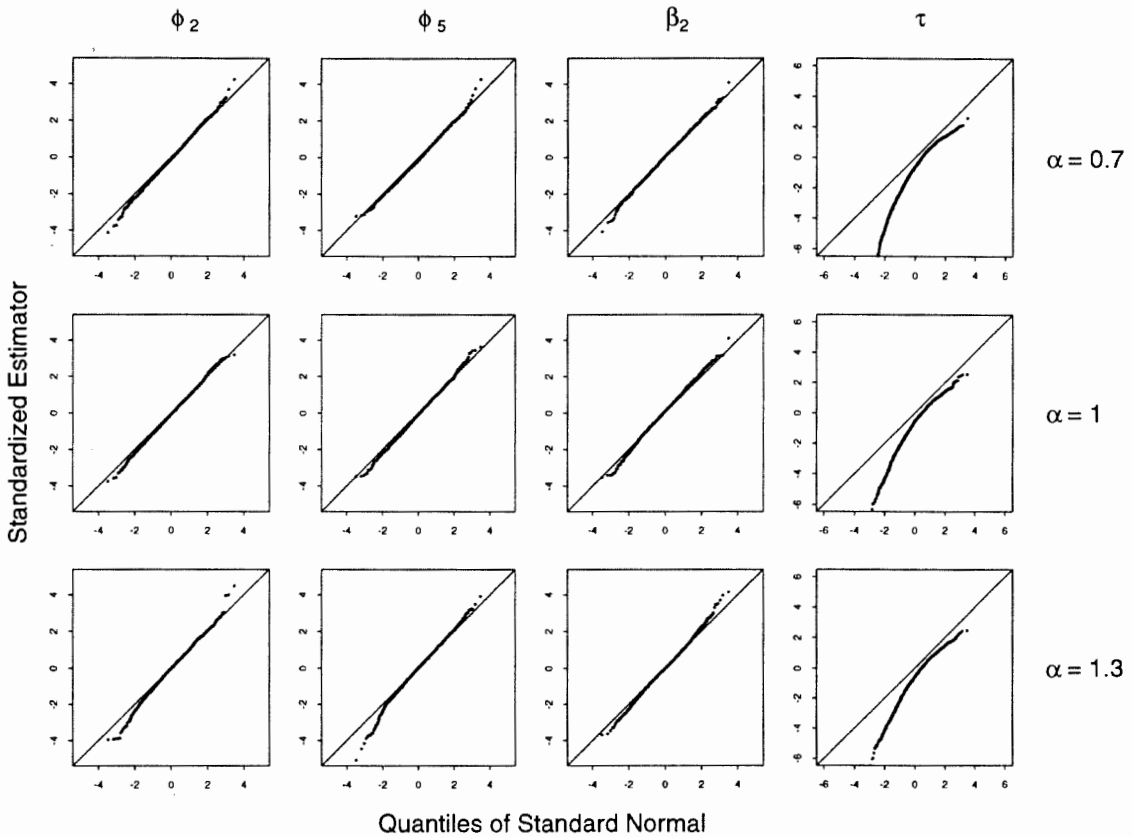

Figure 5.10: Sampling Distribution of the Semiparametric Model Estimators, Robust Standardization. Each panel plots standardized parameter estimates from 2000 simulated data sets against quantiles of the standard normal distribution. Note different scale for τ. The data sets consisted of two groups of 24 subjects and six follow-up times. Each column corresponds to a different parameter; each row shows the results of simulations using a different value of α.

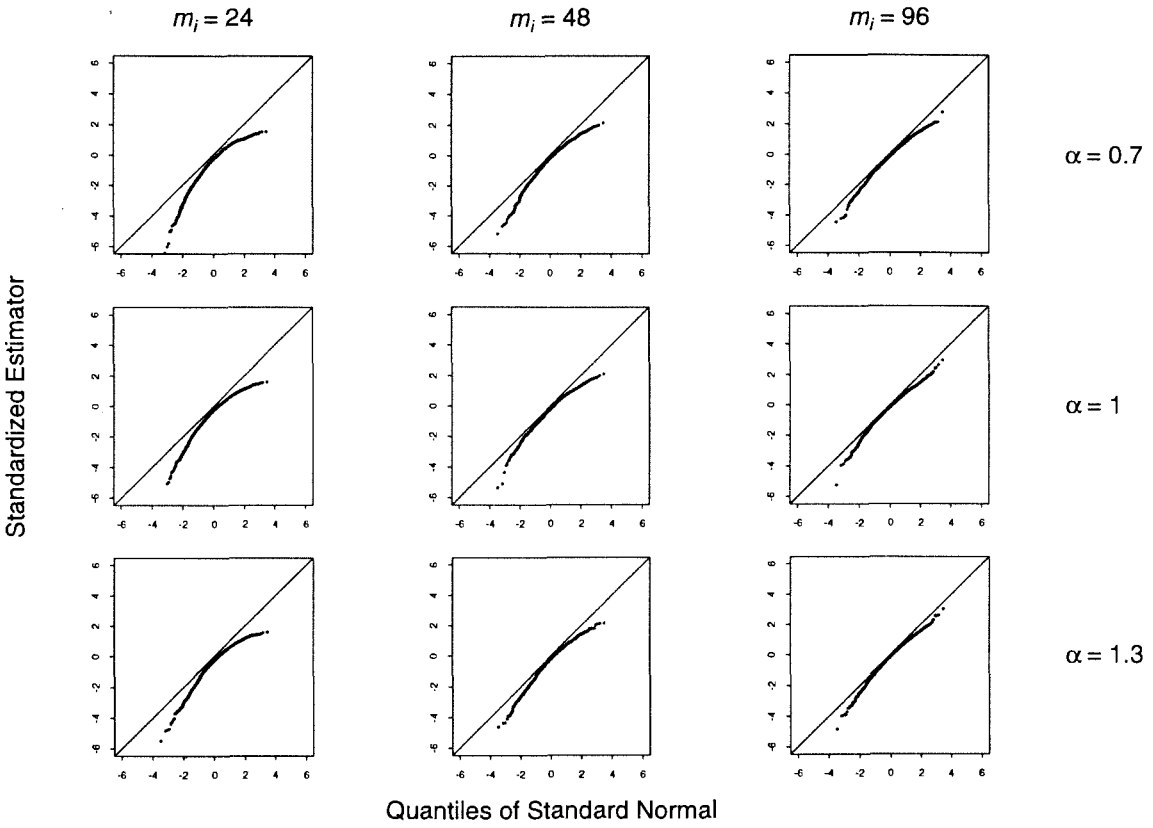

Figure 5.11: Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Weibull Intensity Model, Model-Based Standardization. Each panel plots standardized values of $\tilde{\tau}$ against quantiles of the standard normal distribution. The group size (m_i) appears above each column.

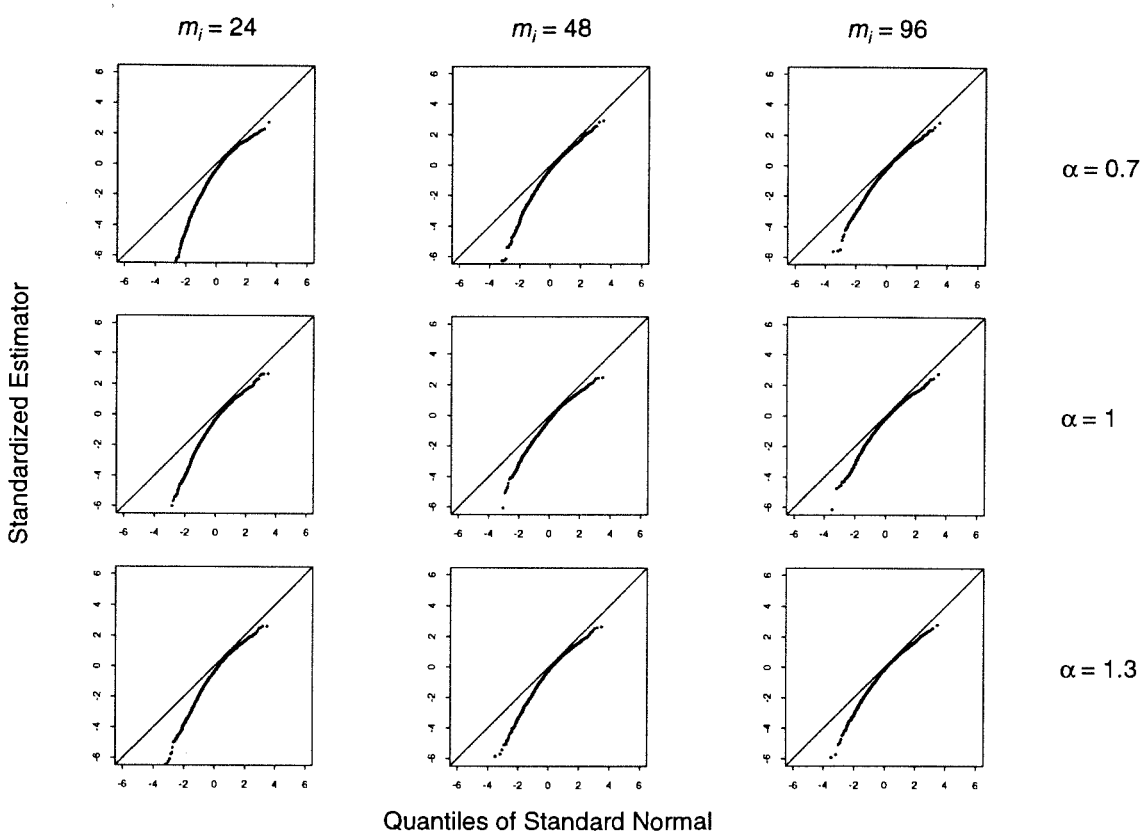

Figure 5.12: Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Weibull Intensity Model, Robust Standardization. Each panel plots standardized values of $\tilde{\tau}$ against quantiles of the standard normal distribution. The group size (m_i) appears above each column.

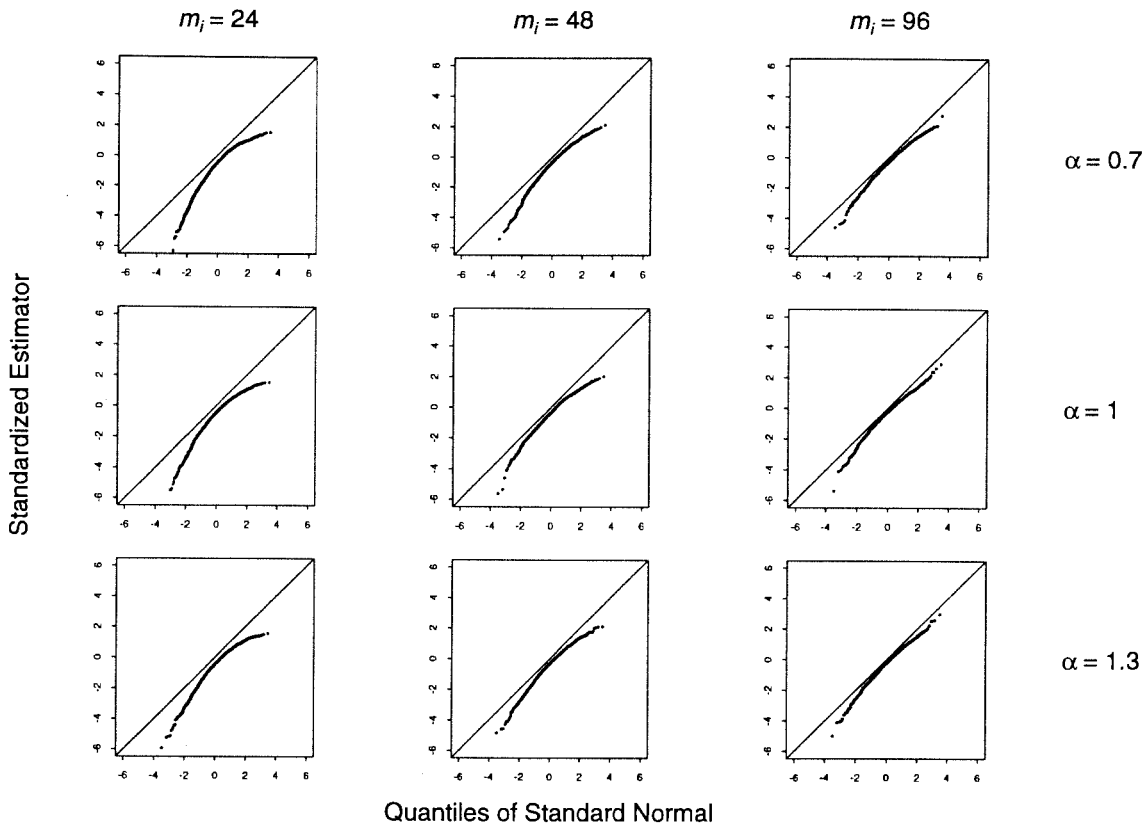

Figure 5.13: Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Semiparametric Model, Model-Based Standardization. Each panel plots standardized values of $\tilde{\tau}$ against quantiles of the standard normal distribution. The group size (m_i) appears above each column.

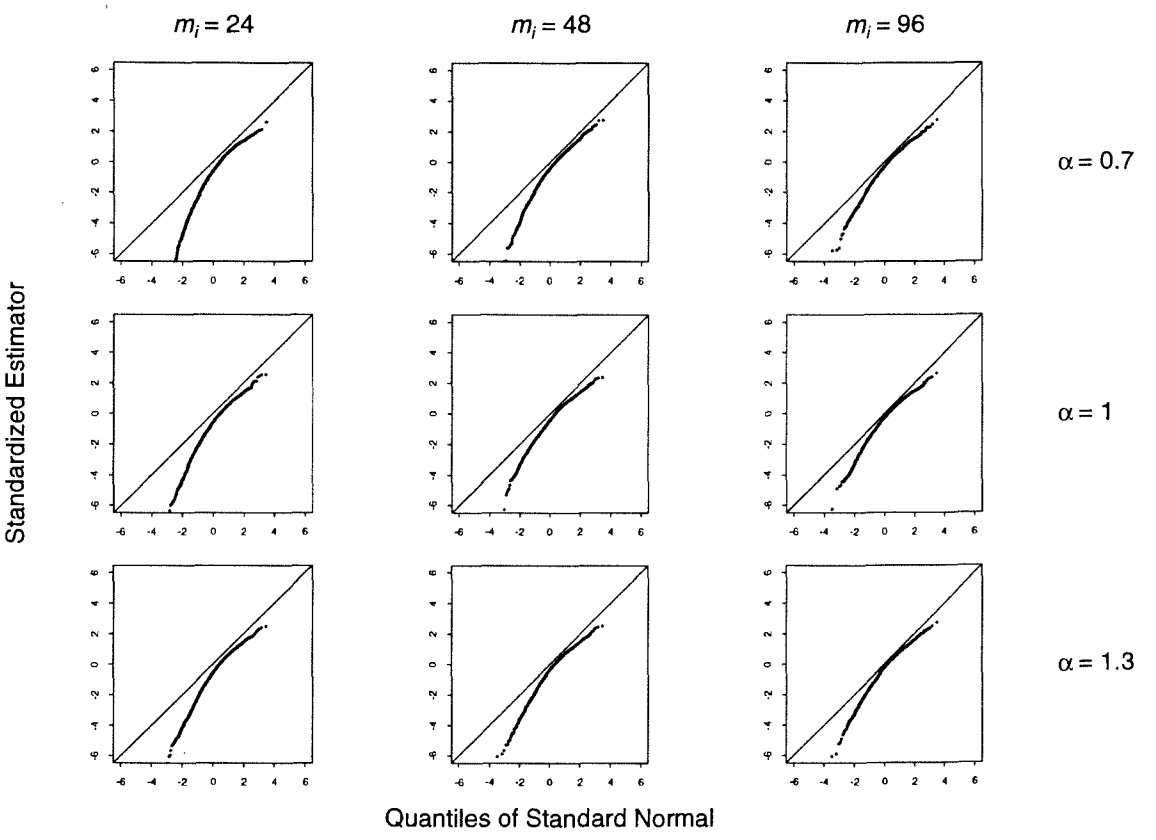

Figure 5.14: Effect of Group Size on Sampling Distribution of $\tilde{\tau}$ from Semiparametric Model, Robust Standardization. Each panel plots standardized values of $\tilde{\tau}$ against quantiles of the standard normal distribution. The group size (m_i) appears above each column.

	Group	$\alpha = 0.7$		$\alpha = 1.0$		$\alpha = 1.3$	
	Size	Mean	(Std. Err.)	Mean	(Std. Err.)	Mean	(Std. Err.)
ϕ_2	true	-3.2163		-2		-0.8745	
	24	-3.2640	(0.2669)	-2.0250	(0.2180)	-0.8872	(0.1957)
	48	-3.2405	(0.2085)	-2.0071	(0.1544)	-0.8860	(0.1395)
	96	-3.2255	(0.1416)	-2.0004	(0.1075)	-0.8782	(0.1011)
ϕ_5	true	-3.5526		-2		-0.5414	
	24	-3.6145	(0.3274)	-2.0278	(0.2189)	-0.5495	(0.1934)
	48	-3.5766	(0.2255)	-2.0100	(0.1519)	-0.5513	(0.1373)
	96	-3.5663	(0.1633)	-2.0036	(0.1067)	-0.5454	(0.0979)
β_2	true	0.1		0.1		0.1	
	24	0.1048	(0.3094)	0.1159	(0.2782)	0.0896	(0.2666)
	48	0.1034	(0.2246)	0.1000	(0.1924)	0.1022	(0.1831)
	96	0.1002	(0.1599)	0.0980	(0.1356)	0.1017	(0.1296)
τ	true	0.8		0.8		0.8	
	24	0.7122	(0.2786)	0.7315	(0.2087)	0.7381	(0.1957)
	48	0.7560	(0.3562)	0.7658	(0.1561)	0.7692	(0.1510)
	96	0.7759	(0.1512)	0.7818	(0.1168)	0.7852	(0.1119)

Table 5.7: Simulation Results for Semiparametric Estimation of $\boldsymbol{\theta}$. The standard errors (Std. Err.) have been calculated as the square roots of the simulated variances of the estimators. Parameters ϕ_1, ϕ_3, ϕ_4 and ϕ_6 have been omitted as they show results similar to ϕ_2 and ϕ_5. The simulation involved 2000 replications of each combination of group size and α. Data were generated using the Weibull baseline intensity function, $\lambda_0(t) = \alpha t^{\alpha-1}$, with $\boldsymbol{\beta} = (-2, 0.1)$ and $\tau = 0.8$.

	Group Size	Variance Type	$\alpha = 0.7$ Mean	(Std. Dev.)	$\alpha = 1.0$ Mean	(Std. Dev.)	$\alpha = 1.3$ Mean	(Std. Dev)
ϕ_2	24	Simulated	0.0905		0.0475		0.0382	
		Model-based	0.0824	(0.0176)	0.0453	(0.0091)	0.0354	(0.0082)
		Robust	0.0804	(0.0274)	0.0451	(0.0151)	0.0350	(0.0122)
	48	Simulated	0.0435		0.0239		0.0195	
		Model-based	0.0412	(0.0061)	0.0233	(0.0033)	0.0183	(0.0032)
		Robust	0.0409	(0.0104)	0.0233	(0.0055)	0.0184	(0.0049)
	96	Simulated	0.0201		0.0116		0.0951	
		Model-based	0.0206	(0.0021)	0.0118	(0.0013)	0.0093	(0.0012)
		Robust	0.0205	(0.0037)	0.0118	(0.0021)	0.0093	(0.0017)
ϕ_5	24	Simulated	0.1072		0.0479		0.0374	
		Model-based	0.1015	(0.0249)	0.0454	(0.0091)	0.0343	(0.0082)
		Robust	0.1004	(0.0389)	0.0448	(0.0146)	0.0341	(0.0120)
	48	Simulated	0.0508		0.0231		0.0188	
		Model-based	0.0498	(0.0080)	0.0233	(0.0033)	0.0178	(0.0032)
		Robust	0.0493	(0.0122)	0.0232	(0.0055)	0.0178	(0.0047)
	96	Simulated	0.0267		0.0114		0.0096	
		Model-based	0.0248	(0.0027)	0.0118	(0.0013)	0.0090	(0.0012)
		Robust	0.0247	(0.0045)	0.0117	(0.0021)	0.0090	(0.0018)

Table 5.8: Accuracy and Precision of Variance Estimators for ϕ_2 and ϕ_5. The simulated variance is the sample variance of the estimator during the simulation. The standard deviations (Std. Dev.) are the square roots of the sample variance of the model-based and robust variance estimates. The simulation summarized in this table was conducted 2000 times with $\boldsymbol{\beta} = (-2, 0.1)', \tau = 0.8$ and α as indicated.

the simulation. Other than τ, all parameters are well estimated, even for small samples. Defining relative bias = $(\text{Mean}(\tilde{\theta}) - \theta)/\theta$, the parameters ϕ_2 and ϕ_5 have a relative bias of less than 2% even for the smallest group size ($m_i = 24$) and less than than 0.5% for the largest group size ($m_i = 96$). The relative bias in the treatment effect was between -10% and 5% for the smallest group size and -2% and 2% for the largest group size ($m_i = 96$). The overdispersion parameter exhibits relative bias ranging from -11% to -8% for the smallest groups and from -3% to -2% for the large groups. The bias decreases for the larger values of α and larger samples.

Tables 5.8 and 5.9 summarize the accuracy and precision of the various variance estimators. The simulated variance is the sampling variance of the estimators of $\boldsymbol{\theta}$ in the

	Group	Variance	$\alpha = 0.7$		$\alpha = 1.0$		$\alpha = 1.3$	
	Size	Type	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
β_2	24	Simulated	0.0957		0.0774		0.0711	
		Model-based	0.0903	(0.0238)	0.0694	(0.0175)	0.0638	(0.0163)
		Robust	0.0904	(0.0239)	0.0694	(0.0175)	0.0638	(0.0163)
	48	Simulated	0.0504		0.0370		0.0335	
		Model-based	0.0465	(0.0091)	0.0361	(0.0065)	0.0332	(0.0063)
		Robust	0.0466	(0.0091)	0.0361	(0.0065)	0.0332	(0.0063)
	96	Simulated	0.0256		0.0184		0.0168	
		Model-based	0.0236	(0.0032)	0.0183	(0.0024)	0.0169	(0.0023)
		Robust	0.0236	(0.0032)	0.0183	(0.0024)	0.0169	(0.0023)
τ	24	Simulated	0.0776		0.0434		0.0383	
		Model-based	0.0966	(0.0764)	0.0564	(0.0401)	0.0480	(0.0334)
		Robust	0.0599	(0.0642)	0.0355	(0.0346)	0.0305	(0.0293)
	48	Simulated	0.0459		0.0244		0.0228	
		Model-based	0.0495	(0.0292)	0.0294	(0.0143)	0.0251	(0.0129)
		Robust	0.0374	(0.0344)	0.0231	(0.0252)	0.0198	(0.0216)
	96	Simulated	0.0229		0.0136		0.0125	
		Model-based	0.0247	(0.0095)	0.0149	(0.0053)	0.0127	(0.0048)
		Robust	0.0217	(0.0270)	0.0128	(0.0104)	0.0113	(0.0121)

Table 5.9: Accuracy and Precision of Variance Estimators for β_2 and τ. The simulated variance is the sample variance of the estimator during the simulation. The standard deviations (Std. Dev.) are the square roots of the sample variance of the model-based and robust variance estimates. The simulation summarized in this table was conducted 2000 times with $\boldsymbol{\beta} = (-2, 0.1)', \tau = 0.8$ and α as indicated.

simulation study. The standard deviations shown on this table are the square roots of the sampling variance of the model-based and robust variance estimates and thus summarize the dispersion of the variance estimators themselves. In general, the robust variances tend to underestimate the sampling variance, though less so for larger group sizes, and are more variable than the model-based estimates for the scenarios examined here. This is particularly noticeable for variance estimation for the overdispersion parameter τ, where the large group size robust-variance estimates have standard deviations two to three times as large as the model-based estimates. This extra variability may explain the slower convergence to its asymptotic distribution observed in the behaviour of the estimator $\tilde{\tau}$ described earlier.

Table 5.10 summarizes the observed coverage properties of the asymptotic normal confidence intervals for the parameters. The coverage probabilities for $\tilde{\beta}_1$ are very reasonable even for $m_i = 24$, ranging from 93% to 95%. Similar rates hold for the other parameters except for τ, where the model-based intervals covered τ as little as 90% of the time for the Weibull model and 87% for the semiparametric. The robust intervals performed even worse, with coverages of τ as low as 79% and 83%, respectively.

The above results point out the need for special care in the use of asymptotic methods for inference in the relatively rare occasion when τ is of greatest interest.

The test statistics examined in the simulation were the model-based and robust versions of the tests for constant baseline intensity and for Weibull baseline intensity. Under the null hypotheses, the test statistics should have had approximately χ^2 distributions. In general, the approximations were reasonable, improving for larger group sizes. For example, the observed rejection rates were between 4% and 6% for the test of the Weibull fit based on the model-based test statistic at a nominal type I error rate of 5%, though the robust test statistic appears to have an inflated type I error rate. Table 5.11 summarizes the observed rejection rate of the tests for various nominal levels. Also, as a very preliminary investigation of the power of these tests, Table 5.12 displays the observed rejection rates for the test of the fit of the homogeneous intensity function when the true baseline intensity function was $\lambda(t) = t^{\alpha}$, for $\alpha = 0.7$ and 1.3. The rejection rates are modest for groups of size 24 and for $\alpha = 0.7$. For larger group sizes or for $\alpha = 1.3$, the test shows reasonable power.

Figure 5.15 displays QQ-plots of the sampling distributions of the test statistic for Weibull fit vs. the quantiles of the χ^2 distribution with 4 d.f. The model-based and robust versions of the test statistic are both examined at the three levels of α. These plots show the distributions for the small group size ($m_i = 24$) and 2000 simulated data sets. The right

	Group	Weibull Model				Semiparametric Model			
α	Size	α	β_1	β_2	τ	ϕ_2	ϕ_5	β_2	τ
				Model-Based Variance Estimator					
0.7	24	0.946	0.948	0.942	0.901	0.938	0.944	0.954	0.868
	48	0.944	0.946	0.942	0.932	0.940	0.944	0.955	0.910
	96	0.944	0.947	0.942	0.940	0.940	0.951	0.940	0.930
1.0	24	0.951	0.948	0.936	0.917	0.930	0.942	0.944	0.892
	48	0.954	0.954	0.946	0.950	0.942	0.946	0.952	0.930
	96	0.945	0.952	0.949	0.946	0.948	0.953	0.952	0.936
1.3	24	0.955	0.946	0.929	0.909	0.922	0.937	0.930	0.873
	48	0.948	0.950	0.948	0.929	0.944	0.937	0.939	0.912
	96	0.944	0.948	0.948	0.942	0.946	0.948	0.944	0.936
				Robust Variance Estimator					
0.7	24	0.931	0.936	0.938	0.826	0.938	0.924	0.942	0.790
	48	0.942	0.938	0.940	0.860	0.940	0.942	0.949	0.843
	96	0.942	0.940	0.942	0.898	0.942	0.949	0.938	0.890
1.0	24	0.935	0.936	0.931	0.839	0.931	0.932	0.930	0.805
	48	0.946	0.954	0.942	0.890	0.942	0.942	0.949	0.872
	96	0.940	0.947	0.948	0.906	0.948	0.948	0.948	0.894
1.3	24	0.942	0.930	0.922	0.838	0.922	0.930	0.932	0.808
	48	0.944	0.940	0.944	0.866	0.944	0.932	0.936	0.848
	96	0.942	0.940	0.947	0.899	0.947	0.948	0.940	0.888

Table 5.10: Observed Coverage Probabilities for Asymptotic Normal Confidence Intervals. Results are based on the analyses of 2000 simulated datasets and so are accurate to within 0.01, 19 times in 20. Confidence intervals were calculated using either the model-based or robust variance estimates. The simulations were conducted with $\beta = (-2, 0.1)$, $\tau = 0.8$ and α as indicated in the table.

		Nominal Size of Test					
	Group	0.01		0.05		0.10	
α	Size	model-based	robust	model-based	robust	model-based	robust
		Testing H_0: Weibull Intensity Function					
0.7	24	0.009	0.018	0.043	0.077	0.101	0.157
	48	0.013	0.017	0.053	0.067	0.109	0.138
	96	0.013	0.018	0.060	0.068	0.108	0.122
1.0	24	0.010	0.013	0.048	0.061	0.092	0.128
	48	0.011	0.011	0.058	0.069	0.111	0.126
	96	0.009	0.009	0.043	0.050	0.105	0.118
1.3	24	0.011	0.011	0.043	0.065	0.097	0.132
	48	0.009	0.011	0.053	0.062	0.102	0.114
	96	0.011	0.013	0.057	0.058	0.105	0.114
		Testing H_0: Constant Intensity Function					
1.0	24	0.009	0.009	0.048	0.055	0.095	0.117
	48	0.013	0.010	0.056	0.054	0.109	0.114
	96	0.009	0.010	0.050	0.061	0.104	0.114

Table 5.11: Observed Rejection Rates for Testing Specific Intensity Models When the Null Hypothesis is True. The rejection rates are based on 2000 simulated data sets. The estimates for the nominal 5% level are accurate to within approximately one percentage point, 19 times in 20.

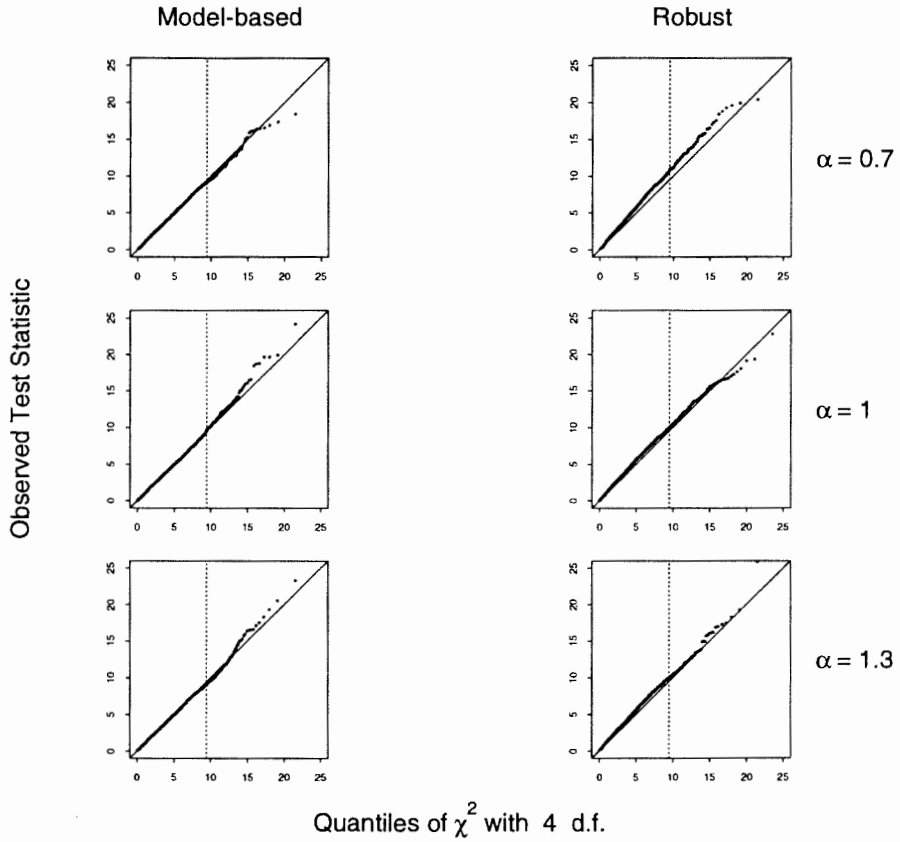

Figure 5.15: Effect of Baseline Intensity Parameter α on the Sampling Distribution of Test Statistics for the Weibull Intensity Model for Groups of Size 24. Each panel plots the observed values of the test statistic from 2000 simulated data sets against quantiles of the χ^2 distribution with 4 d.f. The data sets were simulated according to a Weibull intensity function; a different value of α were used in each row. The left column displays values of the model-based test statistic and the right shows the robust test statistic. The dashed line shows the 95th percentile of the χ^2_4 distribution.

		Nominal Size of Test					
	Group	0.01		0.05		0.10	
α	Size	model-based	robust	model-based	robust	model-based	robust
0.7	24	0.647	0.347	0.822	0.668	0.888	0.801
	48	0.967	0.873	0.991	0.969	0.996	0.990
	96	1.000	0.999	1.000	1.000	1.000	1.000
1.3	24	0.999	1.000	1.000	1.000	1.000	1.000
	48	1.000	1.000	1.000	1.000	1.000	1.000
	96	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.12: Observed Rejection Rate for Testing $H_0 : \alpha = 1$ When the Null Hypothesis is False. The data were generated with a baseline intensity function $\lambda_0(t) = t^\alpha$. Observed rates are based on 2000 simulated analyses and are accurate to within approximately 0.022, 19 times in 20.

hand panels suggest that the robust test statistic has slightly greater variance than the χ^2 distribution with 4 d.f.; this effect is most noticeable for $\alpha = 0.7$. The effect of group size on this tendency is examined in Figure 5.16. These plots show the distribution for group sizes $m_i = 24, 48$, and 96 and suggest that the tendency to underestimate the variance of the test statistic decreases with increasing group size.

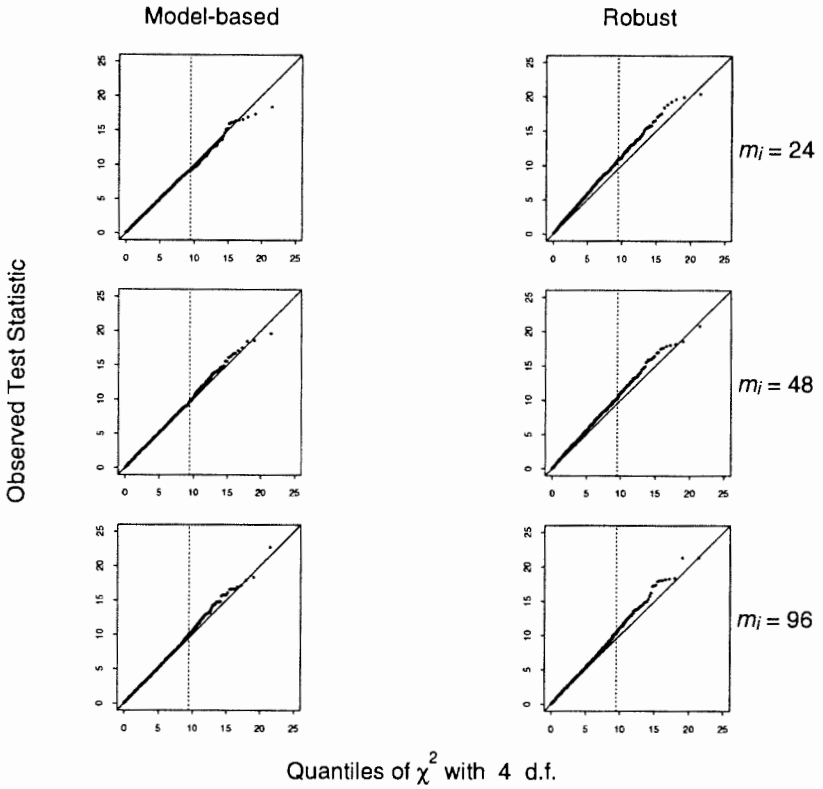

Figure 5.16: Effect of Group Size on the Sampling Distribution of Test Statistics for the Weibull Intensity Model for $\alpha = 0.7$. Each panel plots the observed values of the test statistic from 2000 simulated data sets against quantiles of the χ^2 distribution with 4 d.f. The data sets were simulated according to a Weibull intensity function; a different group size was used in each row. The left column displays values of the model-based test statistic and the right shows the robust test statistic. The dashed line shows the 95th percentile of the χ^2_4 distribution.

Chapter 6

Conclusion

6.1 Summary

In broad terms, this thesis has concentrated on two themes: evaluation of the efficiency of analyses based on count and panel data relative to the analysis of event-time data, and the development of quasi-likelihood and semiparametric methods for the analysis of recurrent event data. The first has practical implications for the design and analysis of studies of recurrent events, and some basic guidelines were given suggesting when each type of data structure would be appropriate. The second has the general effect of making analyses less model dependent, requiring fewer restrictive distributional assumptions and less rigid assumptions regarding the baseline intensity function.

In Chapter 3, methods were developed for the quasi-likelihood analysis of end-of-follow-up counts and event-times in studies of recurrent events. The model proposed was a non-homogeneous Poisson process with subject-specific random effects to account for overdispersion. Covariate effects were incorporated through a multiplicative intensity function. Quasi-likelihood estimators for the effects of covariates and for the parameters of the baseline intensity function were developed and shown to be consistent under first moment assumptions; second order moment assumptions were sufficient to obtain consistent variance estimates as well. Analyses of the count data were shown to achieve high efficiency relative to analyses based on the event-times for inference regarding covariate effects under very reasonable conditions, but had poor efficiency for estimation of parameters describing the baseline intensity.

Chapter 4 generalized the methods and results of Chapter 3 to cover the analysis of

panel data, where observations are taken at multiple follow-up times. Under conditions similar to those for end-of-follow-up count data, the panel data analyses were shown to have high efficiency for the estimation of covariate effects with respect to analyses of event-times. Estimation regarding the intensity parameters based on panel data was demonstrated to recover quite a bit of the information lost by the analysis of end-of-follow-up count data with a few follow-up times. However, we note that efficiencies of the estimator of the baseline intensity function parameters may still not be sufficiently high. Overall, however, panel data thus present an effective and possibly lower cost alternative to the analysis of event-time data.

Also Chapter 4, quasi-score test procedures were developed based on likelihood score procedures (Breslow 1990); the tests for overdispersion and homogeneity of the Poisson process were found to have reasonable small sample properties.

In a comparison of the estimation methods, the quasi-likelihood estimators of the intensity parameters and covariate effects were found to be 100% efficient relative to the likelihood estimators, regardless of the data structure considered. However, the pseudo-likelihood estimator used for the overdispersion parameter was shown to have reduced efficiency when the amount of overdispersion is high, again, regardless of the data structure.

Chapter 5 presented a semiparametric method for the analysis of panel count data, making use of a nonparametric baseline intensity estimator. Test statistics for specific parametric models were also developed, both for inference and as a diagnostic tool. In a small numerical study, the semiparametric estimators of covariate effects achieved very high efficiency relative to estimators from a parametric model. The semiparametric estimators and tests appear to show reasonable small-sample behaviour. However, the estimators of τ appear to approach asymptotic behaviour very slowly, and care must be used for inference regarding this parameter.

Taken together, the techniques developed in this thesis present a coherent framework for the analysis of recurrent event data. Analyses based on count or panel data will possess high efficiency for the estimation of treatment effects for reasonably balanced multi-group experiments. Further work is planned to investigate the efficiency of panel and count data estimators for more general covariate effects. Also, the performance of the semiparametric estimators will be examined under more general parametric intensity models. However, we expect that the results will prove similar to those observed here under the Weibull baseline model because the Weibull model is often a reasonable approximation. This approximation

probably will be especially good when follow-up times are relatively short, as they are in many clinical trials. The methods developed in the thesis are robust to misspecification of the baseline intensity function and require only low-order moment assumptions. Diagnostic procedures have been demonstrated that enable the appropriateness of the model to be assessed. The utility of less expensive and/or less invasive end-of-follow-up counts and panel data have been evaluated, showing that for many of the most common inferential purposes these aggregated counts are very efficient alternatives to event-time data.

In the following section, the thesis concludes with a brief description of questions to be addressed in future work.

6.2 Future Work

6.2.1 Further Investigation of Nonparametric Baseline Intensity Models

The nonparametric baseline intensity function developed in this thesis was proposed as a means of removing the need for specification of a parametric intensity function. The resulting baseline intensity estimates are completely flexible, but it is clear that it should be possible to develop a method of smoothing these estimates in some reasonably nonparametric way. Several possibilities suggest themselves: simple nonparametric smoothers such as moving averages; low order polynomial models; smoothing splines. These smoothing techniques could be applied either to the the baseline intensity estimates or to the cumulative baseline estimates.

Two technical issues also deserve to be addressed. First, alternative specifications of the nonparametric intensity could be examined. In particular, the model could be written with the ϕ_j terms constrained so that $\sum_{j=1}^{s} \phi_j = 0$. This would require a new parameter ϕ_0, corresponding to the Weibull parameter β_1. This would mean that the parameters $\phi_1, \ldots, \phi_s$ would parameterize only the "shape" of the intensity model, making their interpretation the same regardless of the overall intensity level of the process. Second, the power of the quasi-score test for specific parametric intensity model alternatives should be examined in more detail, specifically its power to detect departures from the constant intensity and other low dimensional null hypotheses.

6.2.2 Prediction and Generalizations of the Variance Structure

This thesis has concentrated on the problem of parameter estimation and testing. It would be worthwhile to consider the problem of prediction for an individual subject. Thall (1988) adopted an empirical Bayes methodology, with much the same likelihood as discussed in Chapter 4, except generalized for arbitrary follow-up patterns. Thall obtained a shrinkage estimator of the subject specific effect, ν_i. The use of this type of estimator may present a method for predicting time to next recurrence or total number of recurrences given a partial event history.

The variance structure examined in this thesis is largely a consequence of the use of a subject-specific effect, ν_i with a mean of 1 and variance τ. Two variants on this simple model include: unequal variances for ν_i according to some function of covariates, $\boldsymbol{z}_i$; and possible hierarchical modeling of subject-to-subject variability, perhaps along the lines of Lee and Nelder (1996).

6.2.3 Relationship to Generalized Estimating Equations Methods

In a series of articles (e.g. Liang and Zeger 1986; Zeger and Liang 1986; Zeger, Liang, and Albert 1988; Liang, Zeger, and Qaqish 1992), Liang and Zeger have demonstrated the use of a general method for the analysis of clustered and longitudinal data, known as Generalized Estimating Equations (GEE). The method develops estimators as the solution to a set of quasi-likelihood estimating equations. However, the method does not require specification of an explicit model for the within-subject correlation structure. Instead, a "working correlation matrix" is proposed and used for obtaining parameter estimates. Then the robust variance estimator described above (Section 2.8.3) is used to ensure that estimated variances are reasonable.

The methods were developed for the analysis of clustered and longitudinal data for which generalized linear models would otherwise be appropriate. However, there does not appear to be a corresponding GEE model for the event-time data, only for the panel and end-of-follow-up count data structures. It would be interesting to develop the relationship between the semiparametric methods from Chapter 5 and the GEE methods.

Stukel (1993) compares the GEE and full parametric likelihood estimators (Thall and Vail 1990) for the homogeneous and nonhomogeneous Poisson process for data structures similar to the bladder cancer data examined in this thesis. Stukel concludes that the two

methods produce similar results except when the intensity function model is misspecified. In this case the GEE variance estimates were more reliable. This suggests that the semiparametric estimators developed here may be a desirable complement to the GEE methodology. An examination of the relationship beween the GEE and semiparametric models presented in this thesis would characterize the advantages and disadvantages of both methods, providing guidance regarding when each method should be preferred.

6.2.4 Two-State Model for Recurrent Events

When the event of interest is not accurately modeled as a point event, the methods developed in this thesis will not be appropriate. Consider, for example, a disease process in which the subject is in one of two states, State 0 ("remission") and State 1 ("exacerbation"). The phenomenon of interest in such an experiment might be the total time spent in State 1, or the average time between exacerbations (i.e., average time between transitions from State 0 to State 1). If the average sojourn time in State 1 is short relative to the overall follow-up period, and the total number of exacerbations is small, then the process would be approximately Poisson. Under these conditions, the methods developed in this thesis may be approximately correct. However, as the proportion of time spent in remission (State 0) decreases, the accuracy of this approximation will decrease. It would be useful to develop a diagnostic tool that would enable this type of behaviour to be detected when the nonhomogeneous Poisson process (NHPP) model is used. As well, guidelines outlining the point where the NHPP model is no longer appropriate should be developed.

As an alternative, a two state model could be specified with a baseline intensity function describing the instantaneous rate of transition from State 0 to State 1, $\lambda_0^{01}(t)$, and a second intensity function describing the instantaneous rate of transition from State 1 to State 0, $\lambda_0^{10}(t)$. The NHPP model used in this thesis would be a special case of this model with $\lambda_0^{01}(t) = \lambda_0(t)$ and $\lambda_0^{10}(0)$ defined as infinite. Building from the models used in this thesis, the intensity function for each subject could then be specified for the subjects in State 0 as

$$\lambda_i^{01}(t) = \nu_i \, \lambda_0^{01}(t) \, \exp(\boldsymbol{x}_i'\boldsymbol{\beta}),$$

and for the subjects in State 1

$$\lambda_i^{10}(t) = \eta_i \, \lambda_0^{10}(t) \, \exp(\boldsymbol{z}_i'\boldsymbol{\gamma})$$

where the covariate vectors $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ may or may not contain the same covariates and ν_i and η_i are subject-specific random effects. The distributions of ν_i and η_i would be specified to have means of 1, with variances τ_0 and τ_1, and covariance ρ. A natural restriction to consider in this model formulation would be that the set of covariates are the same for both transitions from State 0 to 1 and from State 1 to 0, i.e., $\boldsymbol{x}_i = \boldsymbol{z}_i$. It would likely be of interest to examine the relationship between the parameter estimates β_j and γ_j. Also, specifying that the subject specific effects were the same for both functions, $\nu_i = \eta_i$, would simplify the analyses, and would correspond to a model in which each subject has a natural tendency to switch from one state to another (*cf.* Gönül and Srinivasan (1993) who examine diaper purchasing behaviour). Alternatively, the sojourn times in State 1 could be assumed to follow an exponential distribution, unaffected by any covariates or subject-specific effects. This would be a much simpler model, but triggers the following question: How reasonable is it to ignore the duration of the sojourn in State 1 and model the data according to the methods presented in this thesis? Many of the same strategies followed in this thesis could be used to investigate these questions, starting with the development of quasi-likelihood estimating functions for the various model parameters.

Bibliography

Abu-Libdeh, H., B. W. Turnbull, and L. C. Clark (1990). Analysis of multi-type recurrent events in longitudinal studies: Application to a skin cancer prevention trial. *Biometrics 46*, 1017–1034.

Andersen, P. and R. Gill (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics 10*, 1100–1120.

Andrews, D. and A. Herzberg (1985). *Data.* New York: Springer-Verlag.

Bacchetti, P. (1990). Estimating the incubation period of aids by comparing population infection and diagnosis patterns. *Journal of the American Statistical Association 85*, 1002–1008.

Bain, L. (1978). *Statistical Analysis of Probability and Life-Testing Models: Theory and Methods*, Volume 24. New York: Marcel Dekker.

Barndorff-Nielsen, O. and D. Cox (1979). Edgeworth and saddle-point approximations with statistical applications. *Journal of the Royal Statistical Society, Series B 41*, 279–299.

Bartholomew, D. (1983). Some recent developments in social statistics. *International Statistical Review 51*, 1–9.

Breslow, N. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics 33*, 38–44.

Breslow, N. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association 85*, 565–571.

Byar, D. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: Comparison of placebo, pyridoxine, and topical thiotepa. In

M. Pavone-Macaluso, P. Smith, and F. Edsmyr (Eds.), *Bladder Tumors and Other Topics in Urological Oncology*, pp. 363–370. New York: Plenum.

Byar, D., C. Blackard, and The V.A. Co-operative Urological Research Group (1977). Comparisons of placebo, pyridoxine and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology 10*, 556–561.

Cox, D. (1982). The statistical analysis of dependencies in point processes. In P. Lewis (Ed.), *Stochastic Point Processes*, pp. 55–66. New York: Wiley.

Cox, D. (1983). Some remarks on overdispersion. *Biometrika 70*, 269–274.

Cox, D. and V. Isham (1980). *Point Processes*. London: Chapman & Hall.

Cox, D. and P. Lewis (1966). *The Statistical Analysis of Series of Events*. London: Methuen.

Crow, L. (1974). Reliability analysis for complex, repairable systems. In F. Proschan and R. Serfling (Eds.), *Reliability and Biometry*, pp. 379–410. Philadelphi: SIAM.

Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika 74*, 591–597.

Davidian, M. and R. Carroll (1987). Variance function estimation. *Journal of the American Statistical Association 82*, 1079–1091.

Davis, C. S. and L. Wei (1988). Nonparametric methods for analyzing incomplete nondecreasing repeated measurements. *Biometrics 44*, 1005–1018.

Dean, C. (1994). Modified pseudo-likelihood estimator of the overdispersion parameter in Poisson mixture models. *Journal of Applied Statistics 21*, 523–532.

Dean, C. and J. Lawless (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association 84*, 467–472.

Dean, C., J. Lawless, and G. Willmot (1989). A mixed Poisson-inverse Gaussian regression model. *Canadian Journal of Statistics 17*, 171–182.

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association 87*, 451–457.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (Third ed.), Volume 1. New York: Wiley.

Godambe, V. and M. Thompson (1989). An extension of quasi-likeihood estimation. *Journal of Statistical Planning and Inference 22*, 137–152.

Gönül, F. and K. Srinivasan (1993). Consumer purchase behavior in a frequently bought product category: Estimation issues and managerial insights from a hazard function model with heterogeneity. *Journal of the American Statistical Association 88*, 1219–1227.

Hinde, J. (1982). Compound Poisson regression models. In R. Gilchrist (Ed.), *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, Berlin, pp. 109–121. Springer-Verlag.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, 221–233.

Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimatior. *Annals of the Institute of Statistical Mathematics 25*, 1–26.

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

Kalbfleisch, J. and R. Prentice (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Karr, A. F. (1986). *Point Processes and Their Statistical Inference.* New York: Marcel Dekker.

Lawless, J. (1982). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Lawless, J. (1995). The analysis of recurrent events for multiple subjects. *Applied Statistics 44*, 487–498.

Lawless, J. and C. Nadeau (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics 37*, 158–168.

Lawless, J. F. (1987a). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics 15*, 209–225.

Lawless, J. F. (1987b). Regression methods for Poisson process data. *Journal of the American Statistical Association 82*, 808–815.

Lee, L. and S. K. Lee (1978). Some results on inference for the weibull proecess. *Technometrics 20*, 41–45.

Lee, Y. and J. Nelder (1996). Hierachical generalized linear models. *Journal of the Royal Statistical Society, Series B 58*, 619 - 678.

Lewis, P. A. (Ed.) (1972). *Stochastic Point Process: Statistical Analysis, Theory and Applications.* New York: Wiley–Interscience.

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

Liang, K.-Y., S. L. Zeger, and B. Qaqish (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B 54*, 3–40.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models* (Second ed.). University Press, Cambridge: Chapman & Hall.

McNair, C., C. Gries, and R. Gries (1997). Towards pheromone-based mating disruption of the cherry bark tortrix, *Enarmonia formosana* (scopoli) (lepidoptera: Tortricidae), on ornamental cherry trees. Submitted to *The Canadian Entomologist.*

Moore, D. (1986). Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika 73*, 583–588.

Moore, D. and A. Tsiatis (1991). Robust estimation of the variance in moment methods for extra-binomial and extra-Poisson variation. *Biometrics 47*, 383–401.

Nelder, J. and Y. Lee (1992). Likelihood, quasi-likelihood and pseudolikelihood: Some comparisons. *Journal of the Royal Statistical Society, Series B 54*, 273–284.

Nelder, J. and D. Pregibon (1987). An extended quasi-likelihood function. *Biometrika 74*, 221–232.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics 5*, 375–383.

Ross, S. M. (1983). *Stochastic Processes.* New York: Wiley.

Seeber, G. (1989). On the regression analysis of tumour recurrence rates. *Statistics in Medicince 8*, 1363–1369.

Snyder, D. L. (1975). *Random Point Processes.* New York: Wiley.

Spinelli, J. (1994). *Cramer-von-Mises Statistics for Discrete Distributions.* Ph. D. thesis, Simon Fraser University.

Stukel, T. (1993). Comparison of methods for the analysis of longitudinal interval count data. *Statistics in Medicine 12*, 1339–1351.

Thall, Peter, F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics 44*, 197–209.

Thall, P. F. and S. C. Vail (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics 46*, 657–671.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models and the gauss-newton algorithm. *Biometrika 61*, 439–447.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

Williams, D. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics 31*, 144–148.

Zeger, S. and K. Liang (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics 42*, 121–130.

Zeger, S. L., K.-Y. Liang, and P. S. Albert (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics 44*, 1049–1060.