Parametric Models for Optimal Treatment Schedule Finding in Adaptive Early-Phase Clinical Trials

by

Changying Liu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2007

Doctoral Committee:

Assistant Professor Thomas M. Braun, Chair
Professor Jeremy M. G. Taylor
Associate Professor Debashis Ghosh
Associate Professor John E. Levine

© Changying Angela Liu 2007
All Rights Reserved

To my husband Bolang, my daughters Helena and Tiffany, and my parents

ACKNOWLEDGEMENTS

First, I give my sincere thanks to my advisor, Dr. Thomas M. Braun. For over three years, he has provided me with profound insight, constant encouragement and generous support in all aspect of my life. I also would like to thank my dissertation committee members, Dr. Jeremy M. G. Taylor, Dr. Debashis Ghosh and Dr. John E. Levine for their valuable advice and comments.

I am also very grateful to the faculties in the department for their valuable advice and support. Dr. Michael Boehnke and Dr. Morton B. Brown gave me constructive suggestions on the research direction. Dr. Jack D. Kalbfleisch, Dr. Roderick J.A. Little and the department provided funding support when I needed. Dr. Timothy D. Johnson and Dr. Sebastian Zöllner were very helpful on statistical computing issues in my research.

I am also very thankful for my fellow students in the department to share their knowledge and have helpful discussions.

Finally, my deepest appreciation goes to my husband, my daughters and my parents for their dedication and sacrifice, their love and encouragement.

TABLE OF CONTENTS

DEDICATION . . . . . **ii**

ACKNOWLEDGEMENTS . . . . . **iii**

LIST OF FIGURES . . . . . **vi**

LIST OF TABLES . . . . . **vii**

CHAPTER

I. Introduction 1

1.1 Review of Phase I Clinical Trial Designs 1
1.2 Review of Cure Rate Modeling Techniques 6
1.2.1 Mixture Cure Models 7
1.2.2 Non-mixture Cure Models 10
1.3 Review of Methodology for Change-Point Problems and Boundary Parameter Problems 12
1.4 The Purpose of This Paper and Its Structure 15

II. A Mixture Cure Model for Optimal Treatment Schedule Finding 17

2.1 Motivation 17
2.2 Single Administration Setting 20
2.2.1 Notation and Model Specification 20
2.2.2 Likelihood Function and Estimation 23
2.3 Multiple Schedule Setting 27
2.3.1 Notation and Model Specification 27
2.3.2 Likelihood Function and Estimation 29
2.4 Trial Conduct 30
2.5 Estimation and Schedule Finding by Bayesian Approach 33
2.5.1 Priors Based on Historical Data 33
2.5.2 Elicited Priors 35
2.5.3 Calibrating the Prior Distribution for Parameters of Interest . . . 36
2.6 Application to KGF trial 37
2.6.1 Study Setup 37
2.6.2 Study Result & Conclusion 40

III. A Triangular Hazard Model for Optimal Treatment Schedule Finding . . 50

3.1 Motivation 50
3.2 Single Administration Setting without Censoring 50
3.2.1 Notation and Statistical Model 50
3.2.2 Derivation of MLEs for a Sample without Censoring 52

3.2.3 Consistency and Limiting Distributions of Constrained MLEs for a Sample without Censoring . . . 55
3.3 Single Administration Setting with Censoring . . . 58
3.3.1 Derivation of MLEs for a Censored Sample . . . 58
3.3.2 Consistency and Limiting Distributions of Constrained MLEs for a Censored Sample . . . 61
3.4 Numerical Studies . . . 62
3.5 Estimation and Schedule Finding by Maximum Likelihood Method . . . 65
3.5.1 Likelihood Function and Estimation in Original Parametric Form 65
3.5.2 Likelihood Function and Estimation in Reparameterized Form . . 68
3.6 Estimation and Schedule Finding by Bayesian Approach . . . 71
3.7 Application to KGF trial . . . 72
3.7.1 Study Result and Conclusion . . . 74

IV. A Non-mixture Cure Model for Optimal Treatment Schedule Finding . . 81

4.1 Introduction . . . 81
4.2 Model Specification . . . 82
4.3 Estimation and Schedule Finding by Maximum Likelihood Method . . . 86
4.4 Estimation and Schedule Finding by Bayesian Approach . . . 86
4.4.1 Priors Based on Historical Data . . . 87
4.4.2 Elicited Priors . . . 89
4.4.3 Calibrating the Prior Distributions of Parameters of Interest . . . 90
4.5 Application to KGF trial . . . 92
4.5.1 Study Result & Conclusion . . . 94

V. Summary and Future Research . . . 101

5.1 Summary . . . 101
5.2 Future Research . . . 104

APPENDICES . . . **109**

BIBLIOGRAPHY . . . **118**

LIST OF FIGURES

Figure

2.1 A triangular hazard function for a single administration of an agent 18

2.2 A sectional Weibull hazard function for a single administration of an agent 19

3.1 Function $\widehat{r}(\theta_2, \theta_3)$ as a function of θ_3 at given θ_2 54

4.1 A Weibull pdf as the hazard function for a single administration of an agent 85

LIST OF TABLES

Table

2.1 **Contingency table showing distribution of subjects.** 24

2.2 **True parameter values of the mixture cure model for simulation studies** 39

2.3 **Estimated parameter values of the mixture cure model by *Maximum Likelihood* for setting *without treatment schedule reassignment*. Each entry is the estimated parameter value (standard deviation).** 41

2.4 **Performance of the mixture cure model with 30 patients by *Maximum Likelihood* for setting *without treatment schedule reassignment*. Each entry is the percentage of schedule selection, with the percentage of patients assigned to that schedule on 2nd line within each scenario.** . . 42

2.5 **Estimated parameter values of the mixture cure model by *Maximum Likelihood* for setting *with treatment schedule reassignment*. Each entry is the estimated parameter value (standard deviation).** 43

2.6 **Performance of the mixture cure model with 30 patients by *Maximum Likelihood* for setting *with treatment schedule reassignment*. Each entry is the percentage of schedule selection, with the percentage of patients assigned to that schedule on 2nd line within each scenario.** 44

2.7 **Estimated parameter values of the mixture cure model by *Bayesian approach* for setting *with treatment schedule reassignment*. Each entry is the estimated parameter value (standard deviation).** 45

2.8 **Performance of the mixture cure model with 30 patients by *Bayesian approach* for setting *with treatment schedule reassignment*. Each entry on the 1st row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on 2nd line within each scenario.** 46

2.9 **Average number of observed toxicities (out of 30 subjects) in the simulated trials using the mixture cure model** 47

2.10 **Average number of subjects (out of 30 subjects) had treatment schedule reassignment in the simulated trials using the mixture cure model** . . . 48

3.1 **Parameter estimates and their inference with 1000 replications in which $\theta_1 = 4$, $\theta_2 = 0.05$ and $\theta_3 = 18$** 63

3.2 **Parameter estimates and their inference with 1000 replications in which $\theta_1 = 3$, $\theta_2 = 0.004$ and $\theta_3 = 18$** 64

3.3 **True parameter values of the triangular hazard model for simulation studies where $\theta_1 = 0.3$ and $\theta_3 = 1.8$** 73

3.4 **Estimated parameter values of the triangular hazard model. Each entry is the estimated parameter value (standard deviation).** 75

3.5 **Performance of the triangular model with 30 patients by *Maximum Likelihood*. Each entry on the 1st row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on 2nd line within each scenario.** 76

3.6 **Performance of the triangular model with 30 patients by *Bayesian Approach*. Each entry on the 1st row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on 2nd line within each scenario.** 77

3.7 **Average number of observed toxicities (out of 30 subjects) in the simulated trials using the triangular hazard model.** 78

3.8 **Average number of subjects (out of 30 subjects) had reassignment in the simulated trials using the triangular hazard model.** 79

4.1 **True parameter values of the non-mixture cure model for simulation studies** 93

4.2 **Estimated parameter values of the non-mixture cure model. Each entry is the estimated parameter value (standard deviation).** 95

4.3 **Performance of the non-mixture cure model with 30 patients by *maximum likelihood method*. Each entry on the $1st$ row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on $2nd$ line within each scenario.** . 96

4.4 **Performance of the non-mixture cure model with 30 patients by *Bayesian approach*. Each entry on the $1st$ row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on $2nd$ line within each scenario.** 97

4.5 **Average number of observed toxicities (out of 30 subjects) in the simulated trials using the non-mixture cure model.** 99

4.6 **Average number of subjects (out of 30 subjects) had reassignment in the simulated trials using the *non-mixture cure model*.** 100

CHAPTER I

Introduction

1.1 Review of Phase I Clinical Trial Designs

Conventional Phase I clinical trials are designed to determine the maximum dose of a new therapeutic agent that leads to toxicity in an "acceptable" proportion of patients. This dose is identified as the maximum tolerated dose (MTD), based on the implicit assumption of a monotonically increasing relationship between dose and probability of toxicity. Historically, Phase I trials have been designed with simple algorithmic approaches, four (A, B, C and D) of which are compared in Storer (1989). Design A is a traditional "3+3" design in which a cohort of three subjects is treated with the same dose. Escalation occurs if no toxicities are observed, and the study is terminated if at least two subjects in the cohort experience toxicity, with the next lowest dose identified as the MTD. If one subject in the cohort experiences toxicity, an additional three subjects are treated at the same dose level. If none of those additional three patients experiences toxicity, escalation occurs; otherwise, the trial stops, with the next lowest dose identified as the MTD. A major limitation of this design is that de-escalation is never an option and the study terminates once two toxicities are observed in a cohort of three or six subjects. Limitations of the "3+3" design have been extensively studied in Korn et al. (1994), Ahn (1998) and

Garrett-Mayer (2006).

Designs B, C, and D are referred to as "up-and-down" designs because they allow for de-escalation. In Design B, a single subject is treated at a dose level d_j. The next subject is treated at dose level d_{j-1} if the previous subject experiences toxicity or at dose d_{j+1} if the previous subject does not experience toxicity. If consecutive de-escalations happen, then the trial stops, and the dose level at the second de-escalation of the consecutive de-escalations is taken as the MTD. If a study fails to stop early and reaches its planned sample size, then the dose given to the final subject or the next lower dose is usually taken as the estimate of MTD depending on the actual amount and degree of toxicity observed in the last group of evaluated subjects.

Design C is a variant of Design B, in which two consecutive nontoxic responses must be obtained before escalation occurs, whereas de-escalation occurs whenever a toxic response is seen. The rule of identifying the MTD is the same as that of Design B. Design D is a modified version of Design A, in which de-escalation occurs if more than one subject in a cohort has toxicity. If a single patient has toxicity, then the next cohort of subjects is treated at the same dose level.

Storer also proposed a pair of two-stage designs, BC and BD, that combine single-stage designs. The first stage follows design B until the first toxic response is observed. From the point at which the next subject is entered at the next lower dose level, the second stage design is implemented with fixed sample size. Storer demonstrated via simulation that the two-stage designs perform better than the single-stage designs by comparing the expected fractions of subjects that would be treated at dose levels above the prespecified threshold percentage of toxicity. Storer also found that the designs performed well as long as the MTD is not chosen to be too extreme of a

percentile (low extreme). However, the designs considered by Storer can have very poor operating characteristics when starting at a dose far below the MTD. Other limitations of those designs are that subjects who enroll first in the trial are most likely assigned to doses with sub-optimal efficacy and the MTD selected at the end of the study has no general interpretation as an estimate of the dose yielding a specified toxicity rate.

In order to overcome the limitations of algorithmic Phase I trial designs, O'Quigley et al. (1990) proposed the Continual Reassessment Method (CRM). CRM is based upon an assumed model for the association of dose and probability of toxicity and uses Bayesian methods to adaptively assign a dose to each subject. O'Quigley and Shen (1996) proposed that maximum-likelihood methods could also be used with the CRM. However, because maximum-likelihood methods fail to be useful until a toxicity is observed, the authors suggested a combination of the traditional Bayesian CRM and their proposed maximum-likelihood CRM. At the beginning of the trial, a Bayesian framework is useful because it allows prior information to be incorporated into the study design in the absence of toxicity. However, when the sample size becomes "large", eg. greater than 12 (as suggested by O'Quigley and Shen, 1996), the numerical integrals or Monte Carlo methods necessary in the Bayesian design become computationally intensive. When at least 12 subjects are enrolled and at least a toxicity is also observed, it is at this point that maximum likelihood methods can then replace the Bayesian methods. As the sample size increases, O'Quigley and Shen (1996) show that the recommended dose level of this hybrid CRM approach converges to the true MTD. There are other competing designs to the CRM, including efficient dose escalation with overdose control (EWOC) (Babb et al., 1998), a curve-free method (CFM) (Gasparini and Eisele, 2000) and the biased coin up-and-down

design (BCD) (Stylianou and Flournoy, 2002). Rosenberger and Haines (2002) gave a comprehensive review of the currently available Phase I trial designs.

Even though the proposed CRM approaches are superior to the classical Phase I design schemes, there are difficulties associated with the use of the CRM. Most notably, each subject (or small group of subjects) must be followed completely for toxicity before the next subject or group is enrolled. Although O'Quigley et al. (1990) proposed a solution to this problem by using the data from fully followed subjects only, this approach is inefficient because it does not fully utilize the information available at the time of evaluation. These difficulties may result in trials of impractically long duration when a trial is designed to evaluate late-onset effects of a new therapeutic agent and fully utilize the information available at each time of evaluation, thereby requiring a long follow-up period for each subject.

Cheung and Chappell (2000) proposed a Time-to-Event Continual Reassessment Method (TITE-CRM) to overcome the difficulties associated with the use of the CRM. The CRM assumes a parametric cumulative distribution function (CDF) $F(d, \beta)$ to describe the relationship between the dose d and toxicity. The TITE-CRM extends the CRM by considering a weighted dose-response model $G(w, d, \beta) = wF(d, \beta)$ that monotonically increase in w with constraints $G(0, d, \beta) = 0$ and $G(1, d, \beta) = F(d, \beta)$ for all $d, \beta, 0 \leq w \leq 1$. The weight w is a function of time-to-event of subjects. Thus the TITE-CRM incorporates the subject's time in a study into the model, and fully utilize each subject's information up to each evaluation time more efficiently than the CRM. As a result, the TITE-CRM allows subjects to be enrolled whenever they are available and evaluate the long-term toxicity more naturally, hence significantly shortening a study's duration without delaying the accrual. Furthermore, the TITE-CRM approach can be generalized to any Phase I design that

involves likelihood-based estimation of a MTD and provides an alternative method to classical design schemes as practitioners desire.

However, the Phase I studies using the TITE-CRM method just like any conventional dose-finding Phase I studies are inadequate for trials in which the agent is administered repeatedly over time and evaluation of long-term cumulative effects is important because the TITE-CRM method like any conventional approaches bases dose-finding on one, initial administration or course of therapy. To overcome this shortcoming, Braun et al. (2003) presented a modified version of the TITE-CRM based on a maximum tolerated cumulative dose (MTCD). Their setting was a bone marrow transplant trial that planned to determine how many weeks of recombinant human keratinocyte growth factor (KGF) could be administered while keeping toxicity rates below a desired threshold. Each subject was enrolled on the best estimate of the MTCD; each time a previously enrolled subject completed his follow-up, the estimate of the MTCD was updated and assignments for all currently enrolled subjects were modified based on whether they were assigned the current MTCD. However, this approach still considered each schedule as a "dose", as a result, the subjects who received an incomplete schedule were only evaluated up to the point of their last fully completed schedule. Furthermore, due to a period of follow-up after a schedule was completed, the "doses" overlapped, leading to some ambiguity as to which "dose" contributed to a late-onset toxicity.

To avoid these limitations, Braun et al. (2005) constructed a new paradigm for Phase I trial designs that allows for the evaluation and comparison of several treatment schedules, each consisting of a sequence of administration times. This method uses time to toxicity as the outcome instead of a binary indicator of toxicity, with the total hazard of toxicity modeled as the sum of a sequence of hazards, each associ-

ated with one administration. The goal of this design is to determine the maximum tolerated schedule (MTS) instead of a traditional MTD. Subject accrual, Bayesian estimation and outcome adaptive decision-making are done in a sequential fashion as in classical Phase I trial designs.

As an alternative to the method of Braun et al. (2005), we first propose a parametric mixture cure model for determining the MTS using adaptive designs in early-phase clinical trials in Chapter II. We discuss the rationale on how to choose our sectional Weibull hazard model to allow for the non-monotonic toxicity change pattern, we then develop an EM algorithm to obtain maximum likelihood estimators (MLEs) of parameters of interest in this mixture cure model. Based on a pre-specified maximum tolerated toxicity level, we define a decision rule on how to determine the MTS for the next subject entering a trial. Later on, we also implement the proposed mixture cure model by a Bayesian approach. Via simulations, we demonstrate the performance of the proposed mixture cure model by both estimation methods.

1.2 Review of Cure Rate Modeling Techniques

The cure rate models have become a very useful tool in biomedical research areas, such as cancer research and AIDS clinical trials. Cure rate modeling is also a rapidly developing research area in statistical modeling techniques, statistical inference and real data applications. The survival analysis invoking the concept of cure rate enhances our ability to interpret models in a more meaningful way with more flexibility in modeling.

A cure model is applicable when it is believed that the survivor function for a time-to-event random variable plateaus to a non-zero constant and does not decay to zero. Such a model is applicable in survival data settings when the empirical

survival curve tends to plateau at a value c ($0 < c < 1$) as time increases. Using a cure model, we assume the subject population is a mixture of two groups: susceptible subjects who will eventually experience the event of interest and cured subjects who will never experience the event of interest. The proportion of cured subjects in the population is called the cure rate. Currently, there are two different approaches to cure rate modeling: mixture cure models and non-mixture cure models.

1.2.1 Mixture Cure Models

Mixture cure models have been a popular method in parametric analyzing survival data with cured subjects for decades, starting with the two-component mixture parametric model of Boag (1949). In this mixture model, one component represents the survival time distribution for the subjects who experienced the event and the other component is a degenerate distribution allowing for infinite survival times of cured subjects. Let T denote a random survival time with population survival function $S_p(t)$, B denote a binary random variable taking values 1 and 0 with probability p (event rate) and $1-p$ (cure rate), respectively, where $1-p = Pr(T = \infty)$. If we let $S(t) = Pr(T \geq t|T < \infty)$ denote the latent survival distribution for the susceptible group, the population survival function $S_p(t)$ can be represented as

$$S_p(t \mid \theta) \quad = \quad E[S(t \mid \theta)^B] = 1 - p + pS(t \mid \theta), \tag{1.1}$$

which is a mixture of the survival function of susceptibles $S(t)$ and cure rate p. Different parametric distributions have been used to model the conditional survival function $S(t)$, including exponential and Weibull distributions (Berkson and Gage, 1952; Farewell, 1977a) . Nonparametric choices for $S(t)$ have also been considered in the literature (Taylor, 1995; Kuk and Chen, 1992; Sy and Taylor, 2000 and Peng and Dear, 2000).

The effects of time independent covariates on the event rate p and the survival distribution $S(t)$ can be modeled seperately. Let (T_i, C_i, Z_i) be a vector of observations in which Z_i is a vector of time independent covariates, T_i is the length of follow-up time and C_i is a censoring indicator. Let B_i indicate the cure status for each subject such that $B_i = 1$ for susceptible subjects who will eventually have a event and $B_i = 0$ for cured subjects. Note that if $C_i = 1$, then $B_i = 1$; however, if $C_i = 0$, then B_i is unknown (latent). The effects of time independent covariates on the event rate p is usually modeled by a logistic regression where

$$P(B_i = 1 \mid \boldsymbol{\beta},\ Z_i) = \frac{\exp(\boldsymbol{\beta}' Z_i)}{1 + \exp(\boldsymbol{\beta}' Z_i)} \tag{1.2}$$

and $\boldsymbol{\beta}$ is vector of parameters.

Regarding the survival function for susceptibles, Boag (1949) assumed a lognormal distribution and used maximum likelihood methods to estimate the proportion of cured subjects and regression coefficients. Berkson and Gage (1952) used a similar two-component mixture model of an exponential distribution and a degenerate distribution to allow for cure rate. Using the method of least squares, they fit their mixture model to a data set on stomach cancer from the Mayo Clinic. Various mixture-based cure models have been considered, specifically in the area of mixture parametric cure models; see also Farewell (1977b), Farewell (1977a). Maller and Zhou (1996) provided a comprehensive treatment to the topic of cure models, especially on various parametric failure time regression models, and they also studied extensively one-sample nonparametric failure time models. Tsodikov et al. (2003) provided a useful summary on nonparametric work for a homogenous sample by Maller and Zhou (1996).

Recently, more research work has been focused on the nonparametric or semiparametric failure time cure models. Taylor (1995) assumed a model with a logistic

probability model for the cure rate and a unspecified failure time process for the failure times of the susceptibles, which was estimated using Kaplan-Meier method. Kuk and Chen (1992), Sy and Taylor (2000) and Peng and Dear (2000) considered semi-parametric Cox proportional hazards models for the failure time process. Li and Taylor (2002) used a semi-parametric accelerated failure time model for the failure times. However, because the sample sizes in Phase I trials are relatively small and insufficient for nonparametric methods performing well, we will not consider nonparametric methods in our work.

Despite its popularity and advantage, a number of problems are associated with the mixture cure model approach. One of the problems is identifiability of parameters in the proposed models, first discussed by Farewell (1986). This problem arises when there is little information in the data about the tail of the survival distribution, so that a long-tailed survival curve could mimic the effect of a nonzero probability of cured subjects. We would have great difficulty in distinguishing the models with high event rates and long tails of survival functions from low event rates and short tails of survival functions. Li et al. (2001) showed that the mixture cure model with a general model for the failure time process is identifiable if a parametric model such as (1.2) for the event rate is assumed. They also considered other important special cases of mixture cure models and non-mixture cure models, establishing conditions for identifiability. Our proposed mixture sectional Weibull hazard model is identifiable according to their results.

A second problem is that a test for the presence of cured subjects or not is a nonstandard inference problem in the sense that it is testing at the boundary of the parameter space. Maller and Zhou (1992) proposed a nonparametric test for the null hypothesis no cured subject present in the population while Ghitany et al. (1994)

proposed a likelihood ratio test for the presence of cured subjects when assuming an exponential distribution for the failure times of susceptibles, which is a parametric test for the null hypothesis. Vu et al. (1998) extended the likelihood ratio test to a generalized exponential family. However, the limiting distribution of the likelihood ratio test under mild regularity conditions is not a standard chi-square distribution, but instead is a 50-50 mixture of a point mass at 0 and a chi-square distribution with 1 degree of freedom.

1.2.2 Non-mixture Cure Models

Non-mixture cure models are an alternative approach in survival analysis to account for the cured subjects. In these models (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen et al., 1999), the probability of cure is incorporated into the model by assuming a bounded cumulative hazard (BCH), $H_p(t)$. In this model, $H_p(t) \leq C$, where C is the finite limit of $H_p(t)$ as $t \to \infty$. As a result, $S_p(t) = \exp(-H_p(t))$ is approaching $\exp(-C)$ and does not decay to zero as $t \to \infty$. Tsodikov et al. (2003) provided a comprehensive review of these BCH modeling techniques in cure rate estimation and associated statistical problems. We summarize the main ideas here.

Let T be the survival time with corresponding population survival function $S_p(t)$. The bounded cumulative hazard model is given by

$$S_p(t) \;=\; \exp(-\theta F(t)),\ \theta > 0 \tag{1.3}$$

where $F(t)$ is a CDF of some nonnegative random variable such that $F(0) = 0$. The cure rate is $S_p(\infty) \;=\; \exp(-\theta)$. Tsodikov et al. (2003) used a series of formulas demonstrating the relationship between a mixture cure model and a non-mixture cure model. They also showed that when using nonparametric estimation meth-

ods, both models can be used equivalently to estimate the cure fraction. However, this equivalence of estimating cure fractions will vanish if the survival function for the susceptible is parametrically specified because of the confounding effects by the parameter θ on both the cure fraction and the survival function for the susceptibles.

Tsodikov et al. (2003) has summarized the three distinct advantages of a non-mixture cure model over a mixture cure model from their thorough review of modeling techniques and estimation methods associated with both models: First, a comprehensive class of nonlinear transformation models (NTM) can be constructed under the non-mixture cure model framework to incorporate complex covariate effects in the regression. The traditional proportional hazards model (PH) is a special case of this rich class of NTM models. However, the mixture cure model does not have the PH property for the population hazard function. Therefore, the non-mixture cure model can be a great tool for studying and testing departures from the PH assumption. Second, in some biomedical applications, a much more biologically meaningful interpretation of the results for the data analysis can be presented by the non-mixture cure model. Third, when developing maximum likelihood or Bayesian estimation procedures, a naturally technical structure is provided by the non-mixture cure model.

Regarding the parametric mixture cure models, existing research has used hazard functions that are either monotonically increasing or decreasing or piecewise constants. These hazard-based parametric models are too restrictive in our setting because they are not flexible enough to entertain situations where the hazard is non-monotonic. Therefore, it is desirable to obtain hazard-based models that allow for a hazard function with changing trends (increase-decrease or decrease-increase) while retain the simple structure of a parametric model. Shao and Zhou (2004) developed

a new parametric mixture cure model using a three-parameter Burr XII distribution for the analysis of survival data with cured subjects. The Weibull distribution is a special case of the Burr XII distribution; thus, the proposed mixture cure model by Shao and Zhou (2004) includes the Weibull and exponential mixture cure models as special cases. Shao and Zhou (2004) demonstrated the proposed mixture cure model fit the given data substantially better than the existing parametric models. Contrary to using the Burr XII hazard in our setting, we consider a triangular hazard model in Chapter III. In summary, a triangular hazard model consists of two piecewise linear functions as a hazard function for susceptibles. Refer readers to Chapter III for details. The estimation of the triangular parameters involves the estimation of a change-point and a boundary point. We develop an algorithm to derive the MLEs of the parameters and demonstrate the consistency and limiting distributions of the MLEs.

1.3 Review of Methodology for Change-Point Problems and Boundary Parameter Problems

The problem of testing for, detecting, and locating a change in the distribution in a sequence of random variables has been examined by both parametric and non-parametric methods. Using a parametric approach, Hinkley (1970) was the first to study the maximum likelihood estimator (MLE) of a unknown change-point in a sequence of time-ordered observations and recognized the key role of the extremum of a two-sided random walk in the limiting distribution of the MLE of a change-point. Hinkley also pioneered the work on inference for change-point parameters including the bias of parameter estimates and confidence regions for change-points. Hinkley (1972) later demonstrated that the asymptotic distribution of the MLE of a change-point is unaffected by the estimation of nuisance parameters.

Jandhyala and Fotopoulos (1999) extended the work by Hinkley and provided a more accurate approximation to the asymptotic distribution of the MLE of a change-point in a sequence of time-ordered observations. They also presented a computationally more efficient and easier algorithm than that of Hinkley for deriving the bounds of the asymptotic distribution of the MLE that could be used to derive the confidence interval for a change-point. Although there are published nonparametric methods for estimation of and inference for change-point problems (Dumbgen, 1994; Yao and Huang, 1994; Eubank and Speckman, 1994), we will not consider nonparametric approaches in our research due to the fact that the sample sizes in Phase I trials are relatively small and insufficient for nonparametric methods performing well.

Detection and location of a change-point in a hazard function under random censoring has been addressed by parametric, nonparametric and semiparametric methods. But existing parametric methods focus on a piecewise constant hazard function for a sample without censoring i.e. the hazard function $\lambda(t)$ of a failure time variable T can be written as

$$\lambda(t|\boldsymbol{\theta}) = \sum_{l=1}^{m} \lambda_l I(\tau_{l-1} \leq t < \tau_l),$$

where $I()$ denotes the set indicator function. The change-points $0 = \tau_0 < \tau_1 < .... < \tau_m$ and the constant hazard rates $\lambda_l > 0, l = 1, ..., m$ are unknown. Estimation of the change-points is a nonregular problem in the sense that the probability density function (PDF) is discontinuous at the unknown change-points. The published results for the limiting distributions of the change-point estimates exist only for $m = 2$ i.e. one change-point. For one change-point (τ) case, Nguyen et al. (1984) constructed a stochastic process $X_n(t), t \geq 0$ for which $X_n(\tau)$ converges to 0, and derived a consistent estimator $\widehat{\tau}$ that satisfies $X_n(\tau) = 0$. However, no asymptotic distribution

was derived. Nguyen et al. (1984) called this estimator a pseudo-maximum likelihood estimator because exact maximum likelihood method was not applicable due to the unboundedness of the likelihood function. Yao (1986) found a constrained MLE $\widehat{\tau}$ that is consistent, and showed that $n(\widehat{\tau}-\tau)$ converges in distribution to two independent random walks that are functions of the unknown parameters τ, λ_1 and λ_2.

The one-change-point exponential survival model can be embedded in a much broader family of densities with an unknown discontinuity point considered by Chernoff and Rubin (1956). In Chapter III, we use their results to prove the limiting distribution of the MLE for the unknown change-point in our triangular hazard model (See Chapter II for definition of triangular hazard function).

In our triangular hazard model, we also encounter the problem of estimating a boundary parameter, in which the upper bound on the support of the hazard function is an unknown parameter. To our knowledge, estimation methods in literature exist only for the setting in which a parameter θ places a lower bound on the support of a PDF and defines a family of PDFs with one unknown location parameter $f(t;\theta) = f_0(t-\theta)$ $(\theta < t < +\infty)$ under the restriction that $f_0(t) \to \alpha c t^{\alpha-1}$ as $t \downarrow 0$. Woodroofe (1972) stated results on asymptotic normality of the MLE when a PDF is differentiable and the MLE is consistent, and stated that for $\alpha > 2$, the Fisher information is finite and the MLE has the same asymptotic properties as regular MLEs under regularity conditions as stated in Cam (1970). For $\alpha = 2$, the MLE is also asymptotically normal (Woodroofe, 1972) and efficient (Weiss and Wolfowitz, 1973), but the convergence rate is $O\{(n\log(n))^{\frac{1}{2}}\}$ instead of the usual $O(n^{\frac{1}{2}})$. For $1 < \alpha < 2$, the MLE has a non-normal limiting distribution with the convergence rate $O(n^{\frac{1}{\alpha}})$ (Woodroofe, 1974). Smith (1985) extended these results to distributions

with nuisance parameters ϕ. The MLEs of ϕ and θ are not only consistent and asymptotically normally distributed but also asymptotically independent. We use the results of Smith (1985) to prove the limiting distribution of the MLE for the boundary parameter in our triangular hazard model.

1.4 The Purpose of This Paper and Its Structure

This dissertation develops several classes of parametric models in optimal treatment schedule finding for Phase I clinical trials using sequential designs. Our goal is to compare the performance of different models by same estimation methods and the performance of different estimation methods under the same model assumptions. The discussion is based on multiple nested treatment schedules, each schedule contains multiple administrations. It is organized as follows. In Chapter I, Introduction, we review the literature pertaining to the research in this dissertation.

In Chapter II, we propose a mixture cure model with sectional Weibull hazard to evaluate a fixed number of nested treatment schedules to determine the MTS. In this mixture cure model, we model the event rate using a logistic regression model and model the conditional hazard function for the susceptible using a combination of two Weibull distributions to account for the non-monotonic nature of the hazard of toxicity. We use both maximum likelihood and Bayesian methods to estimate the parameters of interest. We then compare the performance of the modified maximum likelihood method to that of the Bayesian approach via simulation studies.

In Chapter III, we develop a maximum likelihood procedure to derive the MLEs of unknown parameters in a triangular hazard model for a single administration and prove the asymptotic properties of the MLEs. Then, we extend the results from the single administration setting to the multiple administration (treatment schedule)

setting. We develop an additive triangular hazard model to determine the MTS via maximum likelihood method. We also compare the performance of the triangular model by the maximum likelihood method to the results by the Bayesian approach via simulations.

In Chapter IV, we propose a non-mixture cure model for optimal treatment schedule finding in early-phase clinical trials. We use both maximum likelihood and Bayesian approaches to estimate the unknown parameters and determine MTS. In the simulation studies of Chapters II-IV, subject accrual, data monitoring and outcome-adaptive decision-making are done sequentially through the study.

In Chapter V, we compare the performance of different models by same estimation methods and the performance of different estimation methods under the same model assumptions. We also provide practical recommendations based on pros and cons of the proposed models. We conclude Chapter V with a list of selected future research areas.

CHAPTER II

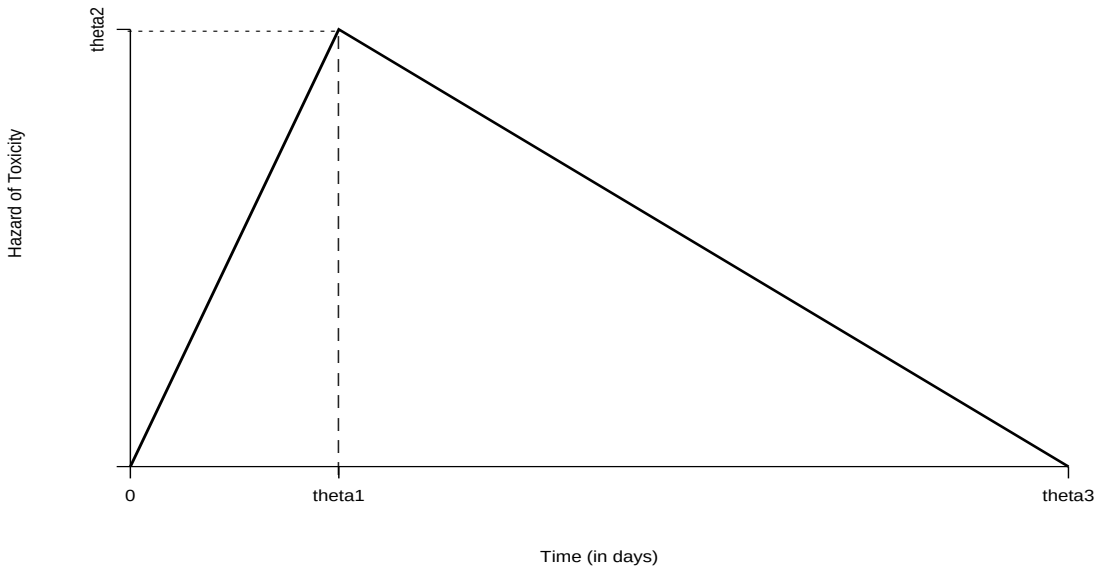
A Mixture Cure Model for Optimal Treatment Schedule Finding

2.1 Motivation

Our motivating example is that described in Braun et al. (2005). Specifically, in a Phase I trial of allogeneic bone marrow transplant (BMT) recipents, the investigators were interested in how long recombinant human keratinocyte growth factor (KGF) could be administered as prophylaxis for graft-versus-host disease (GVHD). During the study, each patient received 60mg/kg of KGF on each of the 2 days prior to BMT, and on the day of BMT. After 4 days of rest with no KGF, the patient received KGF for 3 more days. Therefore, KGF was administered using the 10-day schedule (3-days-on/4-days-off/3-days-on), which is denoted by (3+, 4-, 3+). Toxicity was monitored for 28 days, motivated by the assumption that any adverse effect due to a single adminstration of KGF is certain to occur within 18 days. Although one course of KGF using the (3+, 4-, 3+) schedule is proved to be safe, the investigators believed that this may not be sufficient prophylaxis for GVHD, which may take up to roughly 100 days after BMT to develop. For safety concerns, the investigators wished to evaluate multiple courses of KGF with 4 days of rest between consecutive courses, and follow up subjects for 100 days.

Braun et al. (2005) proposed a new method for the motivating example that took an existing MTD and sought to determine how long that dose could be administered to subjects without causing unacceptable cumulative toxicity. This method used the subject's time to toxicity as the outcome, with the hazard of toxicity modeled as the sum of a sequence of hazards, each associated with one administration. The hazard of toxicity attributed to a single administration was modeled by a triangular function.(See Figure 2.1)

Figure 2.1. A triangular hazard function for a single administration of an agent

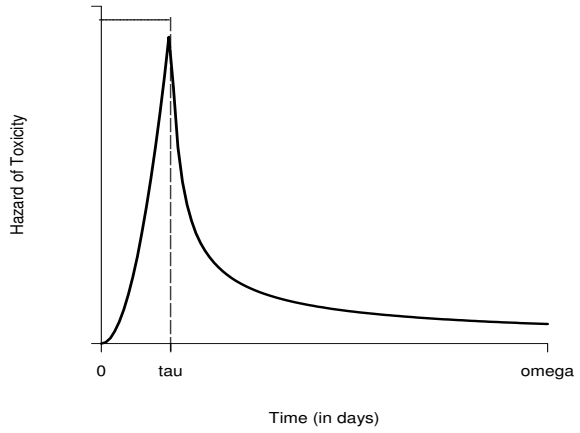


However, the underlying survival distribution specified in this model was improper in the sense that the cumulative distribution function (CDF) $F(t) \nrightarrow 1$ as $t \to +\infty$. As a result, the survival function $S(t) \to c > 0$ as $t \to +\infty$, in which this survival fraction c is referred to as the cure rate. As an alternative, we propose to model the

cure fraction explicitly and adopt a mixture cure model for the latent survival time.

Furthermore, the survival function $S_0(t)$ for susceptible subjects should approach 0 as $t \to +\infty$. So we propose to model the hazard of toxicity as a combination of two Weibull distributions to account for the changing pattern of the hazard (increase to maximum then decrease) during a trial. (See Figure 2.2)

Figure 2.2. A sectional Weibull hazard function for a single administration of an agent



In this chapter, we first consider a mixture cure model for the single administration setting. We then extend the mixture cure model to the setting of multiple schedules with the goal of estimating the probability of toxicity occurring within a pre-specified follow-up period in a trial, allowing for administrative censoring of partially followed subjects. We use both maximum likelihood and Bayesian approaches to estimate parameters of interest in maximum tolerated schedule (MTS) finding in this chapter. Then we compare the simulation results using the maximum likelihood method to

those using the Bayesian approach.

2.2 Single Administration Setting

2.2.1 Notation and Model Specification

Let t^{cur} denote any given time from the beginning of a trial when the evaluation of the data is conducted and n^{cur} be the number of subjects enrolled in the trial at time t^{cur}. Let $T_i^*, i = 1, 2, \ldots, n$, denote the true, possibly unobserved toxicity time for subject i and t_i^{ari} be the entry time for subject i. Since the evaluation is administrative, the assumption of the independence between T^* and $t^{cur} - t_i^{ari}$ are valid. At evaluation time t^{cur}, the amount of time that subject i has been observed is denoted by T_i where $T_i = \min(T_i^*, t^{cur} - t_i^{ari})$ and the indicator of whether or not a subject is observed with a toxicity prior to time t^{cur} is denoted by C_i where

$$C_i = \begin{cases} 1 \;\; ; \;\; T_i^* \leq t^{cur} - t_i^{ari}, \\ 0 \;\; ; \;\; T_i^* > t^{cur} - t_i^{ari}. \end{cases}$$

Let B_i indicate whether or not a subject i would eventually have a toxicity in a trial with infinite follow-up i.e.

$$B_i = \begin{cases} 1 \;\; ; \;\; \text{subjects who will have a toxicity} \\ 0 \;\; ; \;\; \text{subjects who will not have a toxicity} \end{cases}$$

Thus, $B_i = 1$ for subjects for whom $C_i = 1$ as well as a portion of subjects for whom $C_i = 0$. In other words, the value of B_i is latent for censored subjects. We say subjects are susceptible for having a toxicity if $B_i = 1$ and denote p as the probability that $B_i = 1$. We call p the event rate and $(1 - p)$ the cure rate.

Assume the conditional CDF for susceptible i is $P(T_i \leq t | B_i = 1) = F_0(t)$ and for subjects who are non susceptible at time t is $P(T_i \leq t | B_i = 0) = 0$. Then, the marginal CDF of the time to toxicity T_i for subject i is $F(t) = P(T_i \leq t) \;=\; pF_0(t)$.

Note that this marginal CDF $F(t)$ never reaches 1 and has an asymptote at the event rate p.

Let $h_0(t)$ denote the hazard function at time t attributed to a single administration for susceptibles. In our setting, we feel it is plausible that as the body metabolizes the study drug, the hazard of toxicity of the agent increases, reaches a maximum, and then diminishes as the study drug is cleared from the body. There are few parametric hazard functions satisfying this pattern, as most parametric hazard functions assume a monotonic pattern. There exist a class of parametric lifetime hazard functions that allow for non-monotonicity, known as a Burr XII distribution (Shao and Zhou, 2004). However, the functional form of Burr XII is quite complicated, making estimation of its parameters difficult.

As a simpler approach, we assume the hazard of a susceptible is a combination of two Weibull distributions, one with an increasing hazard and the other with a decreasing hazard, with a change-point at time $t = \tau$. A natural choice would be the shift hazard model as follows

$$h_1(t) = \begin{cases} \alpha_1 \lambda_1^{\alpha_1} t^{\alpha_1 - 1} & ; \quad 0 \leq t < \tau, \alpha_1 \geq 1, \lambda_1 > 0 \\ \alpha_2 \lambda_2^{\alpha_2} (t - \tau)^{\alpha_2 - 1} & ; \quad t \geq \tau, 0 < \alpha_2 \leq 1, \lambda_2 > 0. \end{cases}$$

However, when $\alpha_2 < 1$, this hazard would approach infinity as t approaches τ from the right. Such a result is unappealing in our setting as we would like our hazard to be bounded. Furthermore, the corresponding PDF of this shift hazard model would also be unbounded. As we plan to use maximum likelihood to estimate the parameters of interest, the unboundness of the PDF at τ would cause problems for estimation. Therefore, the shift hazard model is not considered further.

In order for our approach to create a bounded hazard function, we instead try a combination of two truncated Weibull hazards. This truncated Weibull hazard

model has the following form

$$h_2(t) = \begin{cases} \alpha_1 \lambda_1^{\alpha_1} t^{\alpha_1 - 1} & ; \quad 0 \leq t < \tau, \alpha_1 \geq 1, \lambda_1 > 0 \\ \alpha_2 \lambda_2^{\alpha_2} t^{\alpha_2 - 1} & ; \quad t \geq \tau, 0 < \alpha_2 \leq 1, \lambda_2 > 0. \end{cases}$$

However, the CDFs and PDFs corresponding to the two functional forms of $h_2(t)$ are equal at the change-point τ if and only if $\alpha_1 = \alpha_2$, $\lambda_1 = \lambda_2$, which violates our model requirement of a non-monotonic hazard function, i.e. $\alpha_1 \neq \alpha_2$.

Therefore, we have chosen to use a sectional model involving two Weibull distributions, where the first part before change-point τ is a two-parameter (α_1, λ_1) Weibull distribution and the second part after τ is a three-parameter $(\alpha_2, \lambda_2, t_\tau)$ Weibull distribution where t_τ is a shift parameter.

This leads to a hazard function (HF) attributed to one administration as

$$h_0(t) = \begin{cases} \alpha_1 \lambda_1^{\alpha_1} t^{\alpha_1 - 1} & ; \quad 0 \leq t < \tau, \alpha_1 \geq 1, \lambda_1 > 0 \\ \alpha_2 \lambda_2^{\alpha_2} (t - t_\tau)^{\alpha_2 - 1} & ; \quad t \geq \tau, 0 < \alpha_2 \leq 1, \lambda_2 > 0 \end{cases} \tag{2.1}$$

the corresponding survival function as

$$S_0(t) = \begin{cases} \exp[-(\lambda_1 t)^{\alpha_1}] & ; \quad 0 \leq t < \tau, \alpha_1 \geq 1, \lambda_1 > 0 \\ \exp\{-[\lambda_2 (t - t_\tau)]^{\alpha_2}\} & ; \quad t \geq \tau, 0 < \alpha_2 \leq 1, \lambda_2 > 0 \end{cases} \tag{2.2}$$

and the corresponding probability density function (PDF) as

$$f_0(t) = \begin{cases} \alpha_1 \lambda_1^{\alpha_1} t^{\alpha_1 - 1} \exp[-\lambda_1^{\alpha_1} t^{\alpha_1}] & ; \quad 0 \leq t < \tau, \alpha_1 \geq 1, \lambda_1 > 0 \\ \alpha_2 \lambda_2^{\alpha_2} (t - t_\tau)^{\alpha_2 - 1} \exp[-\lambda_2^{\alpha_2} (t - t_\tau)^{\alpha_2}] & ; \quad t \geq \tau, 0 < \alpha_2 \leq 1, \lambda_2 > 0. \end{cases}$$

Therefore, the proposed sectional hazard model is characterized by six parameters $(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \tau, t_\tau)$. In order to reduce number of unknown parameters and simplify the likelihood function for the part when $t > \tau$, we place constraints on both CDFs and PDFs that they take a single value at τ. This requirement implies that the two functional forms of both hazard function $h_0(t)$ and corresponding survival function

$S_0(t)$ are equal at τ, thus, the six parameters ($\alpha_1, \lambda_1, \alpha_2, \lambda_2, \tau, t_\tau$) are constrained to satisfy the following equations:

$$\exp[-(\lambda_1\tau)^{\alpha_1}] = exp[-(\lambda_2(\tau - t_\tau))^{\alpha_2}] \quad (2.3)$$

$$\alpha_1(\lambda_1)^{\alpha_1}(\tau)^{\alpha_1-1} = \alpha_2(\lambda_2)^{\alpha_2}(\tau - t_\tau)^{\alpha_2-1}. \quad (2.4)$$

Simplifying the two constraint equalities, we can write τ and t_τ as functions of $\alpha_1, \lambda_1, \alpha_2, \lambda_2$

$$\tau = [\lambda_2^{\alpha_2}/\lambda_1^{\alpha_1}(\alpha_2/\alpha_1)^{\alpha_2}]^{1/(\alpha_1-\alpha_2)}, \quad (2.5)$$

$$t_\tau = [(\alpha_1 - \alpha_2)\tau]/\alpha_1. \quad (2.6)$$

Thus, the two parameters τ and t_τ are not directly estimated.

2.2.2 Likelihood Function and Estimation

If we denote $\boldsymbol{\phi} = (\boldsymbol{\theta}, p)$ where $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \alpha_1, \alpha_2)$, then the likelihood function for $\boldsymbol{\phi}$ is given by

$$\begin{aligned} L_n(\boldsymbol{\phi} \mid \boldsymbol{T},\ \boldsymbol{C}) &= \prod_{i=1}^{n}(f(t_i))^{c_i}(1 - F(t_i))^{1-c_i} \\ &= \prod_{i=1}^{n}[pf_0(t_i)]^{c_i}[1 - p + pS_0(t_i)]^{(1-c_i)}, \end{aligned} \quad (2.7)$$

in which $(\boldsymbol{T},\ \boldsymbol{C}) = \{(t_i, c_i), i = 1, ..., n\}$.

We first use maximum likelihood to estimate the parameters of interest. During the estimation process, the component $1-p+pS_0(t_i)$ will appear in the denominator of the score equations whenever $c_i = 0$. Therefore, censored observations complicate computation of the MLEs.

To simplify the computation, we rewrite the likelihood using partially complete censored observations. Each censored subject who will eventually have an event contributes $P(B_i = 1)P(T_i > t_i|B_i = 1) = pS_0(t_i)$ to the overall likelihood while

the censored subject who will not have an event contributes $P(B_i = 0) = 1 - p$ to the overall likelihood. Therefore, the likelihood function is changed to the modified likelihood function (MLF)

$$L_c(\boldsymbol{\phi} \mid \boldsymbol{T}, \boldsymbol{C}, \boldsymbol{B}) \;=\; \prod_{i=1}^{n} (p f_0(t_i))^{c_i} [(1-p)^{(1-B_i)} (p S_0(t_i))^{B_i}]^{(1-c_i)} \tag{2.8}$$

where B_i is the latent variable for cure status of each censored subject and ($\boldsymbol{T}$, $\boldsymbol{C}$, $\boldsymbol{B}$) $= \{(t_i, c_i, B_i), i = 1, ..., n\}$.

Table 2.1. **Contingency table showing distribution of subjects.**

		Censoring Indicator $c_i = 1$	Censoring Indicator $c_i = 0$	
Observed Times	$t_i \le \tau$	n_{11}	n_{12}	$n_{1.}$
	$t_i > \tau$	n_{21}	n_{22}	$n_{2.}$
		$n_{.1}$	$n_{.2}$	n

Table 2.1 displays the distribution of subjects based on whether they are censored and whether their follow-up is before the change-point, thus dividing the total sample of n subjects into 4 groups. The cells contain the frequency counts of subjects in each group. Let $T_1 \le T_2 \le ... \le T_{n_1.} \le \tau < T_{n_1.+1} \le ... \le T_n$ be the **ordered** observed times in a study from a set of n subjects. We assume the change point τ lies in $[T_{n_1.}, T_{n_1.+1})$. Let the subscript i index the ordered observed times where $i = 1$ indexes the earliest observed time and $i = n$ indexes the latest observed time.

The log MLF is written as

$$\ell_c(\boldsymbol{\theta}) \;=\; \ell_p(p) + \ell_s(\lambda_1, \lambda_2, \alpha_1, \alpha_2), \tag{2.9}$$

where ℓ_p stands for the part related to the event rate p and ℓ_s stands for the part

related to the survival function of a susceptible and

$$\ell_p = \log(p)\sum_{i=1}^{n}[c_i + B_i(1-c_i)] + \log(1-p)\sum_{i=1}^{n}(1-B_i)(1-c_i), \quad (2.10)$$

$$\begin{aligned}\ell_s = &\sum_{i=1}^{n_{1.}}[\log(\alpha_1) + \alpha_1\log(\lambda_1) + (\alpha_1-1)\log(t_i)]c_i + \\ &\sum_{i=1}^{n_{1.}}\{[c_i + B_i(1-c_i)][-(\lambda_1 t_i)^{\alpha_1}]\} + \\ &\sum_{i=n_{1.}+1}^{n}[\log(\alpha_2) + \alpha_2\log(\lambda_2) + (\alpha_2-1)\log(t_i - t_\tau)]c_i + \\ &\sum_{i=n_{1.}+1}^{n}\{[c_i + B_i(1-c_i)][-(\lambda_2(t_i-t_\tau))^{\alpha_2}]\}.\end{aligned} \quad (2.11)$$

The expectation (E) step of the algorithm (Larson and Dinse, 1985) involves creating a set of "pseudo-data" in which the uncensored observations are left intact and the unit mass associated with each fully censored observation is fractionated and assigned to the group of susceptible with a partially complete pseudo-observation. The fractional mass assigned to each pseudo-observation g_i is the conditional probability that the subject will eventually have a toxicity given that no toxicity has occurred by time t:

$$P(B_i = 1|T_i > t) = pS_0(t_i)/(1-p+pS_0(t_i)).$$

The g_i, which will be used in the maximization step, is calculated as the expectation of B_i given the current estimates of $\alpha_1, \lambda_1, \alpha_2, \lambda_2, \tau, t_\tau$:

$$g_i = E(B_i|\widehat{p}, \widehat{\alpha_1}, \widehat{\lambda_1}, \widehat{\alpha_2}, \widehat{\lambda_2}, \widehat{\tau}, \widehat{t_\tau}, t_i, c_i),$$

The maximization (M) step of the algorithm involves calculating the parameter values that maximize the log MLF of the pseudo-data, i.e. replacing B_i by g_i in log MLF (2.10) and (2.11). Separate optimization procedures are used to find the values of $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$ that maximize l_s under constraints (2.5) and (2.6) and the value of p that maximizes l_p.

The EM algorithm is an iterative procedure that begins by choosing initial estimates of $(\alpha_1, \lambda_1, \alpha_2, \lambda_2, p)$, such as those obtained by ignoring the censored observations. At each subsequent iteration, the algorithm's E-step treats the current estimates of $(\alpha_1, \lambda_1, \alpha_2, \lambda_2, p)$ as known in order to update each estimate of g_i, and then the M-step treats the current g_i values as known and updates the estimates of $(\alpha_1, \lambda_1, \alpha_2, \lambda_2, p)$. The convergence criteria can be based on relative changes in the parameter estimates or the log likelihood values over successive iterations.

We use the following method to select initial values for the iteration procedure. Let $X = log(T)$, $u = -log(\lambda_1)$, and $b = {\alpha_1}^{-1}$. Then X has an extreme-value distribution, i.e. the p.d.f. before change-point τ is

$$f_0(x; u, b) = \frac{1}{b} e^{\frac{x-u}{b}} \exp(-e^{\frac{x-u}{b}}); -\infty < x < log(\tau), 0 < b \leq 1, -\infty < u < \infty.$$

We can derive initial guess estimates of u and b by plotting the product-limit survivor function $\widehat{S}(x)$ against x based on observed data. If $\log[-\log \widehat{S}(x)]$ is plotted against x for $x < log(\tau)$ (τ can be chosen as the median observed time point) , then u and b can be estimated from the intercept and slope of a straight line that the plot should approximate if the extreme value model is appropriate. Denote these estimates as u_1 for the intercept and b_1 for the slope, leading to the initial values for $(\widehat{\alpha_1}, \widehat{\lambda_1})$ as $({b_1}^{-1}, \exp(-u_1))$. The initial values for $(\widehat{\alpha_2}, \widehat{\lambda_2})$ could be estimated in a similar fashion since the constraints are placed on the survival function $S_0(t)$. Since the threshold of the probability of toxicity which is defined in next section is known, we use that as the initial estimate for $\widehat{p}$.

2.3 Multiple Schedule Setting

2.3.1 Notation and Model Specification

We assume that k treatment schedules, $s^{(1)}, ..., s^{(k)}$, are investigated in a trial where $s^{(j)}=(s_1, s_2, ..., s_{m_j})$ and that the jth schedule has a total of m_j administrations. Furthermore, $s^{(j)}$ is nested in $s^{(j+1)}$ for each $j = 1, ..., k-1$, so that the duration of a treatment schedule increases with j and $m_1 < m_2 < ... < m_k$. In our motivating example, one course of the (3+, 4-, 3+) schedule corresponds to $s^{(1)} = (1, 2, 3, 8, 9, 10)$, two courses corresponds to $s^{(2)} = (1, 2, 3, 8, 9, 10, 15, 16, 17, 22, 23, 24) = (s^{(1)}, s^{(1)} + 14)$, and so on, with BMT at day 3 in any case.

Let $s_i = \{s_{i,1}, ..., s_{i,m}\}$ denote the consecutive times at which the ith subject receives an administration, where $s_{i,1}$ coincides with subject's study start time. Let m_i denote the number of administrations received by subject i at interim study time t. Although m_j administrations are planned for schedule $s^{(j)}$, at time t it may be the case that $m_i < m_j$ either due to administrative censoring or because subject i had toxicity at time $s_{i,m_i} \leq t$ and thus received no further administrations. Let ω denote the fixed maximum length of follow-up for a trial. ω should be chosen by the medical investigators for clinical reasons, but must be large enough to accommodate the longest schedule, s_k. In our motivating example, $\omega = 100$ days. A fixed target probability p_ω is elicited from the physicians and is defined as the threshold probability of toxicity at the given follow-up time ω.

Conditional Hazard Model

Let $h_0(t|\boldsymbol{\theta})$ be the hazard function attributed to a single administration for the subjects in susceptible group, where $\boldsymbol{\theta}$ is the parameter vector $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$. See equation (2.1). We define the total hazard of toxicity at time t for a subject treated

with schedule $s^{(k)}$ to be

$$\lambda_k(t|\boldsymbol{\theta}, s^{(k)}) = \sum_{l=1}^{m_k} h_0(t - s_l|\boldsymbol{\theta}),$$

with $h_0(u|\boldsymbol{\theta}) = 0$ if $u < 0$ and $h_0(t|\boldsymbol{\theta})$ defined in equation (2.1). We assume that the form of $h_0(.)$ does not change with successive administrations, although this assumption can be relaxed. The cumulative hazard function (CHF) up to t for a subject treated with schedule $s^{(k)}$ is

$$\Lambda_k(t|\boldsymbol{\theta}, s^{(k)}) = \sum_{l=1}^{m_k} H_0(t - s_l|\boldsymbol{\theta}),$$

where

$$H_0(t - s_l|\boldsymbol{\theta}) = \int_0^t h_0(u - s_l|\boldsymbol{\theta})du.$$

The PDF at t for a subject treated with schedule $s^{(k)}$ is

$$f_k(t|\boldsymbol{\theta}, s^{(k)}) = \lambda_k(t|\boldsymbol{\theta}, s^{(k)}) \exp[-\Lambda_k(t|\boldsymbol{\theta}, s^{(k)})]$$

and the survival function up to t is

$$S_k(t|\boldsymbol{\theta}, s^{(k)}) = \exp[-\Lambda_k(t|\boldsymbol{\theta}, s^{(k)})] = \exp[-\sum_{l=1}^{m_k} H_0(t - s_l|\boldsymbol{\theta})].$$

Event Rate Model

Let (T_{ik}, C_{ik}, B_{ik}) denote the follow-up time, censoring indicator and cure status of the ith subject assigned to schedule $s^{(k)}$, respectively, $i = 1, ..., n_k$ where n_k is the number of observations in schedule $s^{(k)}$. Let p_k denote the event rate for subjects assigned to schedule $s^{(k)}$.

Assume the probability of an individual B_{ik} in a susceptible group is modeled by a logistic regression such that

$$logit(p_k) = \beta_0 + \beta_1 k \tag{2.12}$$

for $k = 1, ..., K$ where k indexes the administration schedules. We can also model the event rate as a function of the number of administrations in a given treatment schedule. In this chapter, we have modeled the event rate as the function of the number of treatment schedule in the simulation studies.

2.3.2 Likelihood Function and Estimation

Following the similar discussion as in subsection 2.2.2, let $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \alpha_1, \alpha_2)$, $\boldsymbol{\beta} = (\beta_0, \beta_1)$, and $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{\beta})$, then the modified likelihood function (MLF) for $\boldsymbol{\phi}$ under multiple schedules is given by

$$\begin{aligned} & L_n(\boldsymbol{\phi} \mid \boldsymbol{T}, \boldsymbol{C}, \boldsymbol{B}) \qquad (2.13) \\ = & \prod_{k=1}^{K} \prod_{i=1}^{n_k} (p_k f_k(t_{ik}))^{c_{ik}} [(1-p_k)^{(1-B_{ik})} (p_k S_k(t_{ik}))^{B_{ik}}]^{(1-c_{ik})} \end{aligned}$$

where $(\boldsymbol{T}, \boldsymbol{C}, \boldsymbol{B}) = \{(t_{ik}, c_{ik}, B_{ik}), i = 1, ..., n_k; k = 1, ..., K\}$, and the log MLF is given in (2.9) with the exception that ℓ_p and ℓ_s are changed to

$$\begin{aligned} \ell_p &= \sum_{k=1}^{K} \{\log(p_k) \sum_{i=1}^{n_k} [c_{ik} + B_{ik}(1 - c_{ik})] \qquad (2.14) \\ &\quad + \log(1-p_k) \sum_{i=1}^{n_k} (1 - B_{ik})(1 - c_{ik})\}, \\ \ell_s &= \sum_{k=1}^{K} \sum_{i=1}^{n_k} [c_{ik} \log(\lambda_k(t_{ik}))] \qquad (2.15) \\ &\quad - \sum_{k=1}^{K} \sum_{i=1}^{n_k} [c_i + B_i(1 - c_i)] \Lambda_k(t_{ik}). \end{aligned}$$

The EM estimation procedure for $\boldsymbol{\phi}$ is similar to that in subsection 2.2.2, with some modifications as follows: The fractional mass assigned to each pseudo-observation g_{ik} changes to

$$P(B_{ik} = 1 | T_{ik} > t) = \frac{p_k S_k(t_{ik})}{1 - p_k + p_k S_k(t_{ik})}.$$

and the formula for g_{ik} changes from definition of g_i accordingly.

2.4 Trial Conduct

We assume N be the maximum number of subjects enrolled in a trial with each subject assigned a treatment schedule upon arrival. The first subject is assigned the shortest schedule, $s^{(1)}$. Each subject is followed for up to ω days, with treatment terminated if a toxicity is observed. Given a threshold p_ω for the cumulative probability of toxicity $F(\omega|\boldsymbol{\phi}, s^{(k)})$ by time ω for schedule $s^{(k)}$, we compute the estimated probability of toxicity by time ω, $\widehat{F} = F(\omega|\widehat{\boldsymbol{\phi}}, s^{(k)})$ for each schedule $k,\ k = 1, ..., K$. The best schedule is defined as that having $\widehat{F}$ closest to p_ω and i.e. that minimizing $|\widehat{F} - p_\omega|$. This criterion, as a function of treatment schedule, is analogous to the CRM criterion (O'Quigley et al., 1990) as a function of dose. Using this criterion, the best schedule is identified using the currently available trial data and is assigned to the next accrued subject. At the end of the trial, the MTS is defined as the best schedule based upon the complete data of all N subjects.

To get the trial underway using maximum likelihood approach, we must have heterogeneity among the responses. In order to ensure the existence of meaningful MLEs, we will not consider using the likelihood approach until we see toxicities occur. Furthermore, because we enroll subjects sequentially, very little information is available at the beginning of a trial. In order to use the maximum likelihood in a trial, we not only need to observe a toxicity occurs but also need to have a sufficient sample size. We could use a Bayesian approach coupled with proposed sectional Weibull model or any forms of standard Up-and-Down scheme (Storer, 1989) to obtain a sufficient sample of subjects with at least one toxicity response. A simple and reasonable way is to enroll subjects in cohort of size $m \geq 1$ at a time, starting at the shortest schedule. We escalate to the next schedule if no toxicity is observed

in the previously accrued cohort of subjects and switch to the likelihood approach once a toxicity occurs and enough number of subjects are enrolled in the study. In our application, we found via trial-and-error, $15 - 20$ subjects is a sufficient number for using a maximum likelihood method. Regarding the cohort size m, it is set to $1 - 3$ in most applications although the larger values of m are possible. However, an important practical consideration is that larger values of m will slow the speed of schedule escalation so that the trial will take longer before enrolling subjects on longer schedules.

In contrast, a Bayesian approach does not require a toxicity to occur and a trial can start right away with first available subject enrolling the shortest schedule. The proposed model coupled with a Bayesian estimation method can be used to determine the schedule assignment for next available subject. Such evaluation/enrollment procedures continue until all N subjects are enrolled in the trial. The safety precautions described in next few paragraphs are applicable to any trials no matter which estimation method is used.

Some practical constraints are implemented in our design to minimize the risk of giving a subject an overly toxic schedule. First, our algorithm determines an updated MTS each time a new subject enrolls in the trial and assigns the newest subject to that updated MTS with the constraint of only incremental schedule escalation allowed. For example, if our algorithm recommends to escalate the schedule, we escalate only to the next longest schedule, regardless of the actual schedule selected by our algorithm. We do not put any constraint on schedule deescalation.

Our trial design also implement the constraint to retroactively modify the schedule of a currently enrolled subject if: (1) the currently estimated MTS is different from the schedule assigned to the currently enrolled subject and (2) the subject has not

yet received all of the initially assigned schedule. For example, suppose five subjects have been enrolled: two on the first schedule, two on the second schedule, and one on the third schedule. If the sixth subject is enrolled and assigned to the third schedule and a week later the fifth subject experiences toxicity, the algorithm can reduce the sixth subject's assignment to the second schedule if the proposed method determines such a modification is necessary. The number of subjects impacted by such potential reassignments depends upon the rate of accrual. During rapid accrual, a potentially larger number of subjects may be reassigned to a different schedule because they will have been enrolled before full observation of previously enrolled subjects is completed. During slow accrual, however, the number of reassignments will be minimal because previously enrolled subjects will have been monitored for the full observation period.

The TITEr-CRM design in Braun et al. (2003) is the first study design to allow for dose reassignment, regardless of how many subjects actually have their dose reassigned. Our approach is an extension of their method in multiple schedule scenario. In our case, each new schedule of the experimental agent is given over weeks rather than hours or days, this approach can incorporate new information regarding the safety profile of the agent in time to modify a schedule assignment during the treatment of an individual patient. This practical consideration is not discussed in other Phase I trials (e.g. Cheung and Chappell, 2000; Braun et al., 2005).

Note that all future planned treatment for a subject is stopped once a toxicity occurs. Thus, a toxicity is assigned to the schedule last administered when the toxicity occurred. For example, if a subject is assigned to schedule 3, but experiences toxicity while receiving schedule 2, the toxicity is assigned to schedule 2. Furthermore, a toxicity is prescribed to the originally assigned schedule if a subject has fully received his or her originally assigned schedule and then experiences toxicity during

the posttreatment follow-up period.

2.5 Estimation and Schedule Finding by Bayesian Approach

As a competing alternative to the maximum likelihood method in optimal schedule finding using the proposed mixture cure model, we use a Bayesian approach in this section. Our schedule-finding algorithm begins with independent informative priors for αs, λs and βs. The informative priors may be obtained either based on historical data from previous single administration studies or by elicitation from the investigators. Since the posterior distributions can not be solved analytically under the assumed model, Markov chain Monte Carlo (MCMC) technique is used to compute the posterior quantities. Specifically, a Metropolis-Hastings algorithm ((Robert and Casella, 1999; Gelman et al., 2004) is used. We experimented with different starting values and were convinced that the chains converged and covered the entire posterior distribution using multiple sequences and plots. We eliminated a total of 1000 iterations as burn-in and then generated additional 3000 samples for summarization.

2.5.1 Priors Based on Historical Data

If dose-toxicity data for a single administration are available from previous studies, these data may be used to obtain the priors in the multiple schedule trials. Denote the time of ith subject in the historical trial and the toxicity indicator of this subject by (T_{hi}, C_{hi}) and $D_h = \{(T_{hi}, C_{hi}), i = 1, 2, ...n_h\}$, where n_h is the number of subjects in the historical trial. Then the likelihood function of the available historical data is

$$\begin{aligned} L_h(\boldsymbol{\phi} \mid D_h) &= \prod_{i=1}^{n_h} (f(t_{hi}))^{c_{hi}} (1 - F(t_{hi}))^{1-c_{hi}} \\ &= \prod_{i=1}^{n} [p f_0(t_{hi})]^{c_{hi}} [1 - p + p S_0(t_{hi})]^{(1-c_{hi})}, \end{aligned}$$

If we assume a vague prior on all parameters of interest before the historical data are observed, then the posterior of $\boldsymbol{\phi}$ given the historical data is then $f(\boldsymbol{\phi} \mid D_h) \propto L_h(\boldsymbol{\phi} \mid D_h)$ and the prior used at the start of the schedule-finding trial is $f(\boldsymbol{\phi} \mid D_h)$. The informative priors based on historical data in cure model setting have been studied extensively by Chen et al. (1999). We refer readers to their work for more details.

An alternative way of using the available historical data to define priors is to estimate the parameters of interest using either Bayesian or maximum likelihood methods. With maximum likelihood methods, we can compute estimates of each parameter, as well as variance estimates of those parameter estimates. We then set the mean of the prior distribution at the parameter estimate and the variance of the prior distribution at the variance estimate. With Bayesian methods, we first select a prior distribution for the parameters and then combine it with the historical data to derive posterior mean and variance for each parameter. We then set the mean of the prior distribution at the posterior mean and the variance of the prior distribution at the posterior variance. We let $\widehat{\mu}_*$ and $\widehat{\sigma}^2_*$ denote the respective mean and variance of the prior distribution for a parameter of interest derived from the historical data as described earlier where $*$ stands for each of parameters α_1, $\lambda_1, \lambda_2, \beta_0$, and β_1.

We then select specific functional forms for the prior distributions. Because $\alpha_1 > 1$, we assume $\alpha_1 - 1$ has a gamma distribution with parameters (c_1, d_1) such that α_1 has mean c_1/d_1 and variance c_1/d_1^2. Given the prior mean and variance of α_1 that were derived from historical data, We set $c_1/d_1 + 1 = \widehat{\mu}_\alpha$ $c_1/d_1^2 = \widehat{\sigma}^2_\alpha$, then find $c_1 = (\widehat{\mu}_{\alpha_1} - 1)^2/\widehat{\sigma}^2_{\alpha_1}$ and $d_1 = (\widehat{\mu}_{\alpha_1} - 1)/\widehat{\sigma}^2_{\alpha_1}$. Since $0 < \alpha_2 < 1$, we assume $1/\alpha_2 - 1$ follows a gamma distribution with parameters (c_2, d_2) and we use the same approach as that used with α_1 to solve for the actual values of c_2 and d_2.

Since both λ_1 and λ_2 are positive, to simplify the prior selection, we reparameterize equation (2.1), letting $\gamma_i = \log(\lambda_i^{\alpha_i})$ $i = 1, 2$. As γ_i may be any real number, we assume γ_i follows a normal distribution $N(\mu_{\gamma_i}, \sigma_{\gamma_i})$. We then set μ_{γ_i} at the parameter estimates and σ_{γ_i} at the variance estimates.

Furthermore, if dose-toxicity data for a single course consisting of multiple administrations are also available from previous studies, then these data can also be used to obtain the priors on $\boldsymbol{\beta}$ in addition to the $\alpha_i, \lambda_i, i = 1, 2$ in the multiple schedule trials. Following similar arguments as above, we assume $\beta_i, i = 0, 1$ follows a normal distribution with mean as $\widehat{\mu}_{\beta_i}$ and variance as $\widehat{\sigma}^2_{\beta_i}$ that are derived from the historical data.

When the individual subject data from trials of the single administration are available but no data available for a single course, the source for the informative priors on the parameters of interest can be a mixture of historical data and elicitation from the investigators. For example, in our assumed model, the priors for $\alpha_i, \gamma_i, i = 1, 2$ may be from historical data while the priors on $\beta_i, i = 0, 1$ are elicited from investigators.

2.5.2 Elicited Priors

When individual subject data from trials of the single administration or a single course are not available, informative priors must be elicited from the investigators. This may be done in various ways, with the particular elicitation method tailored to the clinical setting and investigators' level of technical expertise. We employed the following method in our simulation trials.

With regard to the cure fraction parameters $\boldsymbol{\beta}$, we ask the investigators to specify an *a priori* value, P_k, for the cumulative probability of toxicity for schedule $k, k = 1, 2, \ldots, K$. Based upon the simple linear regression model $E\{logit(P_k)\} = b_0 + b_1 k$,

we use ordinary least squares to find estimates of b_0 and b_1; we let μ_{β_0} equal the estimate of b_0 and μ_{β_1} equal the estimate of b_1.

With regard to the hazard shape parameters $\boldsymbol{\theta}$, we do not have a specific elicitation strategy. We do rely on the historic data on single dose. For simulation purpose, we have set mean values of the prior distributions close to true values.

2.5.3 Calibrating the Prior Distribution for Parameters of Interest

In order for the data to dominate the prior distribution, sensitivity analysis of priors on parameters of interest is essential. As a result, those initially estimated hyperparameters still need fine tuning for priors to work in conjunction with the data to allow the schedule-finding algorithm provide a safe and reliable design.

Recall the priors on α_i $(i = 1, 2)$ follow the gamma distribution with parameters (c_i, d_i). We set $c_i = a\widehat{c_i}$ and $d_i = a\widehat{d_i}$. The tuning constant a scales the values of (c_i, d_i) and modulates the variability of $f(\alpha_i)$. In addition, the priors on γ_i $(i = 1, 2)$ follow the normal distribution $N(\mu_{\gamma_i}, \sigma_{\gamma_i})$. We set $\mu_{\gamma_i} = \widehat{\mu}_{\gamma_i}$, $\sigma_{\gamma_i} = b\widehat{\sigma}_{\gamma_i}$ and b is the tuning constant. Similarly, the priors on $\beta_i, i = 0, 1$ follow the normal distribution $N(\mu_{\beta_i}, \sigma_{\beta_i})$. We set $\mu_{\beta_i} = \widehat{\beta}_{i0}$, $\sigma_{\beta_i} = d\widehat{\sigma}(\beta_{i0})$ and d is a tuning constant used to modulate the variances of $\beta_i, i = 0, 1$. By simulating the toxicity times of a small number of subjects, we can compare the prior means for parameter vector $\boldsymbol{\phi}$ to their respective posterior values and evaluate the effects of a small amount of data on prior $f(\boldsymbol{\phi})$.

The prior variances can not be made arbitrarily large, as is often done with Bayesian analysis of large data sets. In any small sample size Phase I trials using adaptive designs, very few data are available, especially at the beginning of a trial. If there is substantial prior information over too broad range, then it can not be overcome by a small amount of data. In our application, undoubtedly, large prior

variances would severely hinder the algorithm's ability to assign best schedule during the trial and select a MTS at the end.

Another consideration of priors' effect is to examine the predictive probability of toxicity $F(\omega|\boldsymbol{\phi}, s^{(j)})$ for each schedule $s^{(j)}$ to determine whether the prior may produce pathological behavior by placing too much of the probability mass of $F(\omega|\boldsymbol{\phi}, s^{(j)})$ near 0 or 1 because the consequent distribution of $F(\omega|\boldsymbol{\phi}, s^{(j)})$ determine which schedule to be identified as MTS.

We have tried different starting values for the Markov chains in the Bayesian estimation procedures. We also used misspecified priors such that the prior means are different from the true parameters of interest. The simulation results changed slightly, but overall, the final conclusions were relatively unchanged. Thus the proposed model is insensitive to the misspecified priors and different starting values as long as the priors are informative at the beginning of a trial but do not dominate the data at later points in the trial.

2.6 Application to KGF trial

In this section, we investigate the performance of the proposed mixture cure model in MTS finding via simulation studies using both Bayesian and maximum likelihood methods. All results are produced in SAS.

2.6.1 Study Setup

In the motivating example, the investigators wished to study $k = 6$ treatment schedules corresponding to 2, 4, 6, 8, 10, 12 weeks of therapy. The maximum period to monitor toxicity was specified to be $\omega = 100$ days because aGVHD occurs during the first 100 days after transplant. Per the adaptive design, a schedule is specified for each subject and the therapy is discontinued if a subject experiences toxicity before

100 days. The goal is to determine how long a subject can receive the therapy while controlling the probability of toxicity within a pre-specified follow-up period ω to be less than or equal to the threshold value p_ω.

We studied the design with a maximum sample size of 30 patients, which is feasible in Phase I trials but also sufficient to determine the MTS with reasonable accuracy demonstrated in our simulations. In each simulation, the subject interarrival times were assumed to be uniformly distributed from 12 to 16 days. When the maximum likelihood method was used, at the beginning of the trial, the traditional up-and-down scheme (Storer, 1989) was implemented. More specifically, three subjects were assigned to the shortest schedule. If there were no toxicities among the three, then we escalated to the next longer schedule and assigned additional 3 subjects to this schedule. As soon as we observed a toxicity and total number of enrolled subjects over 15, we switched to the likelihood approach proposed in this chapter and thereafter included one subject at a time. We considered 6 therapy schedules, $s^{(1)}, ..., s^{(6)}$, in which $s^{(k)}$ did not have natural units and $s^{(k)} = \{s_{lk}, l = 1, ..., m_k\}$ for $k = 1, ..., 6$. When the Bayesian approach was used, one subject was assigned to shortest schedule and the trial started right away.

We examined the design's performance in nine scenarios using the criterion specified in trial conduct Section 2.4. In the first six scenarios, schedule $s^{(j)}$ is optimal under the jth scenario for $j = 1, ..., 6$. In scenario 7, the MTS was located between schedule 2 and 3, while in scenario 8, the target schedule (MTS) lay between schedule 3 and 4 but closer to schedule 3. If an existing available schedule is close to the target schedule, then the proposed method should tend to allocate subjects to that schedule. If the target schedule lies midway between two available schedules, then the method tend to allocate subjects to both schedules.

Furthermore, we examined the design's performance under model misspecification case in scenario 9, where schedule 3 was the MTS, but the data was not simulated from the sectional Weibull model. We assumed the toxicity occurred uniformly over the interval $[10 + 14(j-1), 10 + 14j]$ under schedule $s^{(j)}$.

Except in scenario 9, the actual times to toxicity were simulated assuming the parameter values shown in Table 2.2 under a sectional Weibull model. Table 2.2 also contains the actual probabilities of toxicity by day 100 for each schedule and the threshold probabilities of toxicity for all scenarios. The convergence criteria for parameter estimates were based on relative changes in estimated parameter values and log likelihood values.

Table 2.2. **True parameter values of the mixture cure model for simulation studies**

Scenario	α_1	λ_1	α_2	λ_2	β_0	β_1	True Toxicity Prob of Schedule 1	2	3	4	5	6	Threshold Prob of Toxicity
1	3	0.5	0.3	5	-1.92	0.53	0.20	0.30	0.42	0.55	0.68	0.78	0.2
2	3	0.5	0.3	5	-3.00	0.81	0.10	0.20	0.36	0.56	0.74	0.86	0.2
3	3	0.5	0.3	5	-3.00	0.53	0.07	0.12	0.20	0.30	0.42	0.56	0.2
4	3	0.5	0.3	5	-2.17	0.44	0.15	0.21	0.30	0.40	0.51	0.62	0.4
5	3	0.5	0.3	5	-2.61	0.44	0.10	0.15	0.22	0.30	0.40	0.51	0.4
6	3	0.5	0.3	5	-3.06	0.44	0.07	0.10	0.15	0.22	0.30	0.40	0.4
7	3	0.5	0.3	5	-3.35	0.98	0.09	0.20	0.40	0.64	0.82	0.92	0.3
8	3	0.5	0.3	5	-5.54	1.38	0.02	0.06	0.20	0.50	0.80	0.94	0.3
9	na	na	na	na	-3.00	0.53	0.07	0.12	0.20	0.30	0.42	0.56	0.2

We conducted the simulation studies in two different settings. In first setting, the simulation trials were designed without the constraint of reassigning the treatment schedule to subjects who have not complete the assigned treatment but are not on the updated MTS during the course of a trial. We ran the simulations for scenarios 1-8. The results for the first setting are summarized in Tables 2.3 and 2.4. In the

second setting, the simulation trials were designed with reassigning the treatment schedule to subjects who satisfied the conditions in the Section 2.4. Simulations were conducted for all 9 scenarios. The results for this setting are given in Tables 2.5 and 2.6.

Regarding Bayesian approach, the simulation trials were only conducted under the setting with treatment schedule reassignment. The simulation results are displayed in Tables 2.7 and 2.8. With regard to the prior distributions for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, we use the elicited values from the investigators as in Table 2.2. For example, the investigators supply the values $P_1 = 0.07$, $P_2 = 0.10$, $P_3 = 0.15$, $P_4 = 0.22$, $P_5 = 0.30$, $P_6 = 0.40$, corresponding to the scenario 6 row in Table 2.2. Thus, they believe the longest schedule, $\boldsymbol{s}^{(6)}$, is optimal, a belief that leads to a misspecified prior for the first five scenarios. Similarly, we can use any other row of probabilities of toxicity to specify the priors for the $\boldsymbol{\beta}$ as long as the row we choose is the investigators' belief. From these elicited values, we used the methods described in Section 2.5.2 to estimate the mean hyperparameter values $\mu_{\beta_0} = -3.0$ and $\mu_{\beta_1} = 0.45$. We set variance hyperparameter values close to two times of the corresponding mean values so that $\sigma_{\beta_0} = 6$ and $\sigma_{\beta_1} = 1$. For $\boldsymbol{\theta}$, we set the mean values close to the true values so that $\mu_{\alpha_1} = 2.9$, $\mu_{\alpha_2} = 0.4$, $\mu_{\gamma_1} = -2.0$, $\mu_{\gamma_2} = 0.4$ and set the variances close to 1 so that $\sigma_{\alpha_1} = 1$, $\sigma_{\alpha_2} = 1$, $\sigma_{\gamma_1} = 1$, $\sigma_{\gamma_2} = 1$. After a thorough sensitivity analysis, we determined that the tuning parameters $a = 1.5$, $b = 0.33$ and $d = 1/3$ to allow the data to have dominate influence on the posterior of the parameters of interest.

2.6.2 Study Result & Conclusion

Tables 2.3 and 2.4 display the simulation results for the setting without treatment schedule reassignment. Table 2.3 displays the estimated parameter values and corresponding standard deviations under eight scenarios. Table 2.4 summarizes the

Table 2.3. **Estimated parameter values of the mixture cure model by *Maximum Likelihood* for setting *without treatment schedule reassignment*. Each entry is the estimated parameter value (standard deviation).**

Scenario	Estimated Value of α_1	λ_1	α_2	λ_2	β_0	β_1
1	3.48 (1.79)	0.52 (0.125)	0.42 (0.214)	5.98 (1.78)	-1.89 (0.97)	0.63 (0.41)
2	3.72 (1.87)	0.49 (0.131)	0.38 (0.189)	5.95 (1.96)	-3.47 (1.77)	0.98 (0.54)
3	3.78 (1.96)	0.43 (0.133)	0.37 (0.172)	5.75 (1.78)	-3.99 (1.94)	0.73 (0.42)
4	3.96 (1.97)	0.49 (0.146)	0.35 (0.174)	5.89 (1.83)	-2.81 (1.74)	0.64 (0.37)
5	3.57 (1.64)	0.48 (0.135)	0.38 (0.183)	5.37 (1.92)	-3.13 (1.76)	0.56 (0.31)
6	3.89 (1.76)	0.45 (0.128)	0.36 (0.184)	5.74 (1.94)	-3.33 (1.84)	0.63 (0.46)
7	3.51 (1.83)	0.49 (0.142)	0.41 (0.185)	5.68 (2.01)	-4.03 (1.92)	1.12 (0.66)
8	3.47 (1.94)	0.43 (0.152)	0.35 (0.186)	6.04 (2.12)	-6.03 (1.97)	1.56 (0.97)

frequency of each schedule selected as MTS. The first row of each scenario in Table 2.4 contains the percentages of simulations in which each schedule was identified as the MTS while the second row of each scenario contains the mean percentages of subjects assigned to each schedule among the simulations. Note that the percentages in each row may not add up to exactly 100% due to rounding. The parameter estimates in Table 2.3 are reasonably close to the true values and provide confidence for us to interpret the results in Table 2.4.

The probabilities of toxicity for neighboring schedules in some of the scenarios are very close to each other, making it more difficult to identify the target schedule. Despite this, our maximum likelihood approach recommended the correct schedule at least 30% of the time in the scenarios we investigated, and within one level above or below the true MTS $70\% - 90\%$ of the time for scenarios $1 - 6$. Specifically, the algorithm identified the MTS within one schedule of the true MTS in 96%, 92%, 83%, 80%, 76% and 73% of the simulations in scenarios $1 - 6$, respectively. We also note that the algorithm tends to misidentify the MTS at shorter schedules more

Table 2.4. **Performance of the mixture cure model with 30 patients by *Maximum Likelihood* for setting *without treatment schedule reassignment*. Each entry is the percentage of schedule selection, with the percentage of patients assigned to that schedule on 2nd line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	64.9	32.4	2.7	0.0	0	0
	51.4	30.6	9.1	6.5	2.4	0
2	35.5	43.7	12.6	8.2	0	0
	39.6	40.1	10.3	4.8	2.7	2.5
3	12.9	27.5	35.4	20.5	2.8	0.9
	13.5	25.1	31.6	20.4	6.2	3.2
4	2.8	14.1	26.4	34.5	16.7	5.5
	7.9	13.7	20.4	29.5	19.3	9.2
5	2.1	8.7	13.8	20.1	34.8	20.5
	4.3	14.4	17.1	18.7	26.2	19.3
6	0.5	1.5	7.4	17.7	26.2	46.7
	3.4	7.4	11.8	15.2	20.8	41.4
7	15.7	34.8	30.4	13.3	5.8	0.
	20.4	31.8	27.2	12.3	7.1	1.2
8	7.5	19.3	33.1	25.2	12.6	2.3
	13.3	20.2	29.9	21.8	10.7	4.1

often than at longer schedules, indicating that our constraints in trial conduct were implemented correctly in the algorithm of selecting the MTS and reflects our desire to promote patient safety.

Our algorithm performed best in scenarios 1 and 6 when the optimal schedule existed at either the lowest or highest schedule. Furthermore, 51% and 42% of subjects were assigned to the true MTS in scenarios 1 and 6, respectively. Although fewer subjects were assigned to the true MTS in the other scenarios, we still found that no more than approximately 20% of subjects were assigned to a schedule more than 1 schedule above the true MTS in scenarios 1-6.

In scenario 7, the target schedule was located between schedules 2 and 3. The

algorithm still performed well, selecting either schedule 2 or schedule 3 in 65% of simulations. In scenario 8, the true MTS lay between schedules 3 and 4 but was closer to schedule 3. Our method picked schedule 3 in 33% of simulations and chose schedule 4 in 25% of simulations, for a total of 58%. In summary, our algorithm performed well even in the scenarios the true MTS does not exist among the available schedules but lies between two given schedules.

Table 2.5. **Estimated parameter values of the mixture cure model by *Maximum Likelihood* for setting *with treatment schedule reassignment*. Each entry is the estimated parameter value (standard deviation).**

Scenario	Estimated Value of α_1	λ_1	α_2	λ_2	β_0	β_1
1	3.97 (1.62)	0.48 (0.121)	0.39 (0.201)	5.70 (1.92)	-2.16 (1.02)	0.69 (0.41)
2	3.55 (1.42)	0.45 (0.145)	0.35 (0.193)	6.33 (2.12)	-3.32 (1.11)	1.01 (0.52)
3	3.75 (1.45)	0.46 (0.131)	0.35 (0.187)	5.73 (1.87)	-3.36 (1.21)	0.76 (0.37)
4	3.86 (1.59)	0.39 (0.136)	0.36 (0.182)	6.01 (2.15)	-2.49 (1.36)	0.72 (0.43)
5	3.61 (1.47)	0.43 (0.138)	0.34 (0.191)	5.21 (1.62)	-3.25 (1.65)	0.67 (0.41)
6	3.31 (1.45)	0.39 (0.118)	0.39 (0.187)	5.74 (1.85)	-3.71 (1.87)	0.71 (0.45)
7	3.58 (1.73)	0.48 (0.147)	0.38 (0.186)	5.85 (1.97)	-4.37 (2.23)	1.23 (0.63)
8	4.09 (1.93)	0.47 (0.129)	0.39 (0.193)	5.37 (1.74)	-6.35 (2.32)	2.04 (0.92)

We now review the results for the setting with reassignment of treatment schedule. Table 2.5 displays the estimated parameter values and corresponding standard errors under eight scenarios. Table 2.6 summaries the frequency of each schedule selected as MTS and the average percentages of subjects assigned to each schedule among the simulations. Table 2.5 shows similar results as Table 2.3 while Table 2.6 displays the results similar to those in Table 2.4. We note that the bias for the estimated value of α_1 is about 20% in scenarios 1 and 8 in Table 2.5. Despite the apparent bias, our algorithm still recommended the right schedule as MTS over 60% of simulations in Table 2.6. We suspect that the bias in estimating α_1 has little impact on the

results seen in Table 2.6 because the event rate is captured in p. Furthermore, as the follow-up time gets longer, the survival function for susceptibles does not contribute to the overall event rate because $F_0(t)$ is close to 1. As a result, as long as the biases of the estimated $\boldsymbol{\beta}$ are reasonably small, the estimated probability of toxicity for the correct treatment schedule should be close to the threshold.

Table 2.6. **Performance of the mixture cure model with 30 patients by *Maximum Likelihood* for setting *with treatment schedule reassignment*. Each entry is the percentage of schedule selection, with the percentage of patients assigned to that schedule on 2nd line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	65.8	33.0	1.1	0.0	0	0
	53.7	32.5	10.2	2.9	0.6	0
2	30.5	49.7	14.6	3.1	1.7	0.3
	31.3	36.1	19.2	8.8	3.4	1.2
3	6.6	22.2	38.2	24.3	6.6	2.1
	12.3	27.2	29.8	19.2	8.8	2.7
4	4.2	16.4	23.2	36.2	16.1	3.9
	5.7	12.2	25.7	31.6	19.4	5.4
5	3.3	10.4	15.1	17.6	37.2	16.4
	5.8	8.9	11.9	26.6	30.4	17.0
6	3.6	4.3	7.2	11.9	24.3	48.7
	6.4	9.4	12.9	14.0	16.7	40.6
7	14.4	40.7	31.3	13.6	0.	0.
	12.9	37.8	30.6	11.9	5.3	1.5
8	4.7	19.6	35.3	24.4	12.3	3.7
	7.4	16.8	28.2	29.6	12.8	5.2
9	10.8	29.5	32.4	21.6	5.0	0.7
	13.2	27.8	28.7	18.3	8.4	3.6

Scenario 9 of Table 2.6 displays the performance of our algorithm under model misspecification, where schedule 3 was the true MTS but the actual times to toxicity did not follow the assumed additive sectional Weibull model. The results follow the same pattern as that in scenario 3, where the true toxicity distribution matches that

of the assumed model. In scenario 9, a lower frequency of identifying schedule 3 as the MTS than that in scenario 3 is found, as is expected.

Our proposed likelihood approach coupled with sectional Weibull model worked fairly well in the cases where the toxicity of the treatment schedules are moderate. If the toxicity levels are very small for the shorter schedules or most of the schedules or too high for the longer schedules, then the likelihood approach may not be a good choice because the true values of those parameters may be close to the boundary of the parameter space.

Table 2.7. **Estimated parameter values of the mixture cure model by *Bayesian approach* for setting *with treatment schedule reassignment*. Each entry is the estimated parameter value (standard deviation).**

Scenario	Estimated Value of α_1	λ_1	α_2	λ_2	β_0	β_1
1	3.21 (0.534)	0.37 (0.080)	0.39 (0.069)	3.73 (1.539)	-2.21 (0.509)	0.58 (0.145)
2	2.87 (0.542)	0.49 (0.112)	0.38 (0.067)	4.80 (1.055)	-3.34 (0.413)	0.86 (0.092)
3	2.88 (0.455)	0.44 (0.141)	0.39 (0.077)	4.53 (1.269)	-3.44 (0.478)	0.63 (0.126)
4	3.02 (0.305)	0.41 (0.071)	0.39 (0.093)	4.45 (1.596)	-2.41 (0.349)	0.58 (0.085)
5	2.79 (0.445)	0.45 (0.143)	0.38 (0.088)	4.62 (1.453)	-3.13 (0.365)	0.54 (0.122)
6	2.88 (0.658)	0.48 (0.219)	0.38 (0.059)	4.51 (0.704)	-3.56 (0.421)	0.46 (0.123)
7	2.72 (0.632)	0.39 (0.198)	0.37 (0.067)	4.34 (1.233)	-3.83 (0.465)	1.05 (0.126)
8	3.07 (0.767)	0.43 (0.286)	0.39 (0.116)	4.45 (1.451)	-5.49 (0.896)	1.51 (0.186)

The simulation results by Bayesian approach are displayed in Tables 2.7 and 2.8. The estimated parameter values and corresponding standard deviations for all scenarios are listed in Table 2.7 while the frequency of each schedule chosen as MTS are summarized in Table 2.8. Within each scenario, each entry on the first row in Table 2.8 contains the percentage of simulations in which the given schedule is identified as the MTS while the entry on the second row contains the average percentage of subjects assigned to the specified schedule during a trial.

We note that the parameter estimates in Table 2.7 have smaller biases in most scenarios compared to those estimates in both Tables 2.3 and 2.5 by the maximum likelihood method. Furthermore, the standard error estimates in Table 2.7 are smaller than those in both Tables 2.3 and 2.5 among all scenarios. These results may be due to the small sample size in our Phase I simulation trials and the informative priors used in the Bayesian approach. Most importantly, the fact that the parameter estimates are reasonably close to the true parameter values gives us confidence to interpret the results in Table 2.8.

Table 2.8. **Performance of the mixture cure model with 30 patients by *Bayesian approach* for setting *with treatment schedule reassignment*. Each entry on the 1st row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on 2nd line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	85.8	11.6	2.6	0	0	0
	64.9	18.5	12.6	2.0	1.3	0.7
2	30.6	52.3	11.8	5.2	0.1	0.
	25.0	40.1	20.0	12.4	3.3	1.2
3	6.3	19.7	51.8	16.6	5.6	0
	11.3	31.1	39.8	10.4	4.1	3.2
4	0	7.3	31.9	41.8	15.3	3.7
	5.0	14.4	23.6	39.0	14.7	3.3
5	2.1	10.4	17.4	20.3	40.3	19.5
	5.9	8.5	15.3	17.2	34.2	18.9
6	0.	1.8	6.2	13.8	19.8	58.4
	4.1	6.3	7.9	14.4	21.5	45.8
7	9.4	41.9	32.7	12.1	3.9	0.
	11.6	35.0	31.4	14.1	7.9	0.
8	3.2	11.3	42.8	30.3	9.8	2.6
	5.8	17.2	35.9	25.9	10.7	4.5
9	5.8	29.1	41.1	16.1	5.8	2.1
	9.8	26.3	32.7	18.9	7.2	5.1

Most important information we can take away from Tables 2.6 and 2.8 is that

there is no marked difference between the results as final recommendation is concern by two estimation methods. In addition, there is no big difference as far as final recommendation is concern no matter whether the assumed model is right or not. However, in scenario 1, the Bayesian approach has a much higher percentage (15% more) of identifying the optimal schedule 1 as the MTS than the maximum likelihood method. In scenarios 2-6, the Bayesian approach had a moderately higher percentage ($3\% - 10\%$ more) of correctly selecting the optimal schedule (schedule j for scenario j as true MTS) as the MTS than maximum likelihood method. We can also say, in scenarios 2-5, both Bayesian and maximum likelihood methods have similar percentages of choosing the true optimal schedule as MTS. Furthermore, the pattern of the results in scenario 7 is similar as that in scenarios 2-3. The results in scenario 8 follow the same pattern as scenarios 3-4.

The results from scenario 9 in both Tables 2.6 and 2.8 demonstrate the effects of the model misspecification on the MTS identification. We note that the results in scenario 9 is similar to those in scenario 3, although scenario 3 has a higher percentage of selecting schedule 3 as the MTS. In summary, both maximum likelihood and Bayesian approaches work well under model misspecification cases even though more subjects are assigned to the true optimal schedule when the model is correctly specified.

Table 2.9. **Average number of observed toxicities (out of 30 subjects) in the simulated trials using the mixture cure model**

Estimation Method	Schedule Reassignment	Scenario 1	2	3	4	5	6	7	8
MLE	No	7.72	6.23	5.65	12.13	10.84	9.32	8.29	7.15
MLE	Yes	7.51	6.13	5.57	11.56	10.65	9.67	7.48	6.27
Bayesian	Yes	6.80	6.34	5.69	11.83	10.49	9.72	7.85	6.47

An additional comparison of the safety profiles of the trial implementation with or without treatment schedule reassignment and by different estimation methods is found in Table 2.9, which compares the two ways of implementing the trials with regard to the average number of subjects out of 30 that experienced toxicity in each scenario. Overall, the rate of toxicities is similar in both implementations (with or without schedule reassignment), although the rate of toxicities is marginally higher in the trial without treatment schedule reassignment than the one with it. We also note that the numbers are very close for both estimation methods. There is no clear pattern on which method had more observed toxicities among all scenarios. Note that the average number of observed toxicities were not decreasing from scenario 1 to scenario 6 but were higher in scenarios 4-6 than in scenarios 1-3 because the probability of toxicity $p_\omega = 0.2$ for scenarios 1 to 3 while that $p_\omega = 0.4$ for scenarios 4 to 6.

Table 2.10. **Average number of subjects (out of 30 subjects) had treatment schedule reassignment in the simulated trials using the mixture cure model**

Estimation Method	Scenario								
	1	2	3	4	5	6	7	8	9
MLE	0.89	1.86	3.79	5.05	6.59	7.46	2.17	4.31	4.28
Bayesian	0.43	1.39	4.46	5.73	6.97	7.63	3.08	6.23	3.58

An interesting information regarding the average number of subjects out of 30 who had treatment schedule reassignment is listed in Table 2.10. Overall, the numbers of subjects who went through treatment reassignment are higher in the scenarios where the longer schedules were the true MTS. The numbers are comparable for both methods with no apparent trend across all scenarios. For example, in scenario 3, four subjects experienced treatment reassignment when maximum likelihood used while five subjects had schedule reassignment when Bayesian approach used. On

the contrary, in scenario 9, four subjects experienced treatment reassignment when maximum likelihood method used but only three subjects had schedule reassignment when Bayesian approach used.

In this chapter, we proposed a mixture cure model with sectional Weibull distributions to evaluate a fixed number of nested treatment schedules to determine the MTS, in which we modeled the event rate by a logistic regression and modeled the conditional hazard function for the susceptible with a combination of two Weibull distributions to account for the non-monotonic nature of the hazard of toxicity. We used both maximum likelihood and Bayesian approaches to estimate parameters of interest. We performed simulation studies to investigate the performance of our proposed model in identifying MTS and found that our proposed model performed well in the scenarios we investigated.

CHAPTER III

A Triangular Hazard Model for Optimal Treatment Schedule Finding

3.1 Motivation

Recall the motivating example in Chapter II. In this setting, Braun et al. (2005) proposed a new Phase I trial design with parameter estimation based upon Bayesian methods. In this chapter, we will explore parameter estimation of the triangular hazard model proposed in Braun et al. (2005) via a maximum likelihood method. We will also derive the large sample properties of the MLEs when all subjects receive a single adminstration. Then we will compare the performance of the optimal treatment schedule finding by the maximum likelihood to that by the Bayesian approach under the setting with treatment schedule reassignment.

3.2 Single Administration Setting without Censoring

3.2.1 Notation and Statistical Model

Let $T_i, C_i, i = 1, 2, \ldots, n$, be the same notation as in section 2.2.1 for subject i. We define the distribution for T_i through its hazard function as proposed by Braun

et al. (2005)

$$h_0(t \mid \boldsymbol{\theta}) = \begin{cases} \frac{\theta_2 t}{\theta_1} & ; \quad 0 < t \leq \theta_1, \theta_1 > 0, \theta_2 > 0 \\ \frac{\theta_2(\theta_3 - t)}{\theta_3 - \theta_1} & ; \quad \theta_1 < t \leq \theta_3, \theta_1 < \theta_3, \\ 0 & ; \quad t > \theta_3, or, t \leq 0. \end{cases} \tag{3.1}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. We can also write the hazard function as $h_0(t \mid \boldsymbol{\theta}) = \frac{\theta_2 t}{\theta_1} I_{(0,\theta_1]} + \frac{\theta_2(\theta_3 - t)}{\theta_3 - \theta_1} I_{(\theta_1,\theta_3]}$, in which $I_{(a,b]}$ is the indicator that t lies in the open interval $(a, b]$. We use a closed bracket to indicate if either endpoint should be included in the interval. Thus, θ_1 denotes the time at which the hazard reaches its maximum, θ_2, and θ_3 denotes the time when the hazard vanishes to zero. We refer to θ_1 as the change-point of $h_0(t \mid \boldsymbol{\theta})$ and θ_3 as the duration of $h_0(t \mid \boldsymbol{\theta})$. See Figure 2.1 in Chapter II for a plot of the triangular hazard function. Note that the constraint imposed on the parameters of the triangular hazard model is $0 < \theta_1 < \theta_3 < +\infty$.

Then the corresponding cumulative hazard function (CHF) for T_i is

$$H_0(t \mid \boldsymbol{\theta}) = \begin{cases} 0 & ; \quad t \leq 0, \\ \frac{\theta_2 t^2}{2\theta_1} & ; \quad 0 < t \leq \theta_1, \theta_1 > 0, \theta_2 > 0, \\ \frac{\theta_2 \theta_3}{2} - \frac{\theta_2(\theta_3 - t)^2}{2(\theta_3 - \theta_1)} & ; \quad \theta_1 < t \leq \theta_3, \theta_1 \leq \theta_3, \\ \frac{\theta_2 \theta_3}{2} & ; \quad t > \theta_3 \end{cases} \tag{3.2}$$

and the corresponding survival function and PDF are

$$\begin{aligned} S_0(t \mid \boldsymbol{\theta}) &= \exp[-H_0(t \mid \boldsymbol{\theta})] \\ f_0(t \mid \boldsymbol{\theta}) &= h_0(t \mid \boldsymbol{\theta}) \ \exp[-H_0(t \mid \boldsymbol{\theta})] \end{aligned}$$

respectively.

Note that the CDF $F_0(t \mid \boldsymbol{\theta}) = 1 - \exp(\frac{-\theta_2 \theta_3}{2})$ for $t > \theta_3$. $F_0(t \mid \boldsymbol{\theta})$ is not a standard CDF in the sense that $F_0(t \mid \boldsymbol{\theta}) \nrightarrow 1$ as $t \to +\infty$. Because a standard CDF is needed for proving that the score equations for estimating the θs are unbiased, we define

$K(\theta_2, \theta_3)$ be the normalization factor $1 - \exp(-\frac{\theta_2\theta_3}{2})$. Then a standard CDF can be defined as $F(t \mid \boldsymbol{\theta}) = F_0(t \mid \boldsymbol{\theta})/K(\theta_2, \theta_3)$ so that $F(t \mid \boldsymbol{\theta}) = 1$ for $t > \theta_3$. Thus, the standardized PDF $f(t \mid \boldsymbol{\theta}) = f_0(t \mid \boldsymbol{\theta})/K(\theta_2, \theta_3)$ is what we choose to work with.

3.2.2 Derivation of MLEs for a Sample without Censoring

We first consider a setting in which no censoring occurs, i.e. $T_i = T_i^* \leq \theta_3$. Since the derivation of the MLE of $\boldsymbol{\theta}$ will be based upon the **ordered** observed times $T_1 \leq T_2 \leq ... \leq T_n$, we modify the notation so that i indexes the ordered observed times where $i = 1$ indexes the earliest observed time and $i = n$ indexes the latest observed time. Thus, the likelihood function is given by

$$L_n(\boldsymbol{\theta} \mid \boldsymbol{T}) = \prod_{i=1}^{n} h_0(T_i \mid \boldsymbol{\theta}) \exp[-H_0(T_i \mid \boldsymbol{\theta})]/K(\theta_2, \theta_3) \tag{3.3}$$

in which $\boldsymbol{T} = (T_1, ..., T_n)$. And the corresponding log likelihood function is written as

$$\ell_n(\boldsymbol{\theta} \mid \boldsymbol{T}) = Q_1(\theta_1, \theta_3) - \theta_2 Q_2(\theta_1, \theta_3) + Q_3(\theta_2, \theta_3), \tag{3.4}$$

where

$$\begin{aligned}
Q_1(\theta_1, \theta_3) &= \sum_{i=1}^{n} \log\left[\frac{T_i}{\theta_1} I_{[0,\theta_1]} + \frac{\theta_3 - T_i}{\theta_3 - \theta_1} I_{(\theta_1,\theta_3]}\right] \\
Q_2(\theta_1, \theta_3) &= \frac{1}{2}\sum_{i=1}^{n} \left\{\frac{T_i^2}{\theta_1} I_{[0,\theta_1]} + \left[\theta_3 - \frac{(\theta_3 - T_i)^2}{\theta_3 - \theta_1}\right] I_{(\theta_1,\theta_3]}\right\} \\
Q_3(\theta_2, \theta_3) &= n\log(\theta_2) - n\log(K(\theta_2, \theta_3)).
\end{aligned}$$

Because θ_1 is the change-point of the PDF $f(t, \boldsymbol{\theta})$, we can not use the score equation to derive the MLE of θ_1. So we consider an alternative approach to derive the MLE of θ_1. In the following theorem, we summarize the result about the MLE of θ_1.

Theorem 1. For given θ_2 and θ_3 which satisfy $\theta_2 < min(\frac{1}{\theta_1},\ \frac{1}{\theta_3\ -\ \theta_1})$, the log likelihood function (3.4) attains its maximum at one of the order statistics $(T_1, ..., T_n)$.

See the Appendix for a proof. The specific order statistic assigned to MLE $\widehat{\theta_1}$ will depend upon the values of the other two parameters. Specifically, let

$$\widehat{r}(\theta_2, \theta_3) = \arg \max_{r \in \{1,...,n\}} \{\ell_n(T_r, \theta_2, \theta_3)\}, \tag{3.5}$$

then $\widehat{r}(\theta_2, \theta_3)$ is a function of θ_2 and θ_3 that can only take the discrete values in [1, n], thus the MLE of θ_1 is

$$\widehat{\theta_1} = T_{\widehat{r}(\theta_2,\theta_3)} \tag{3.6}$$

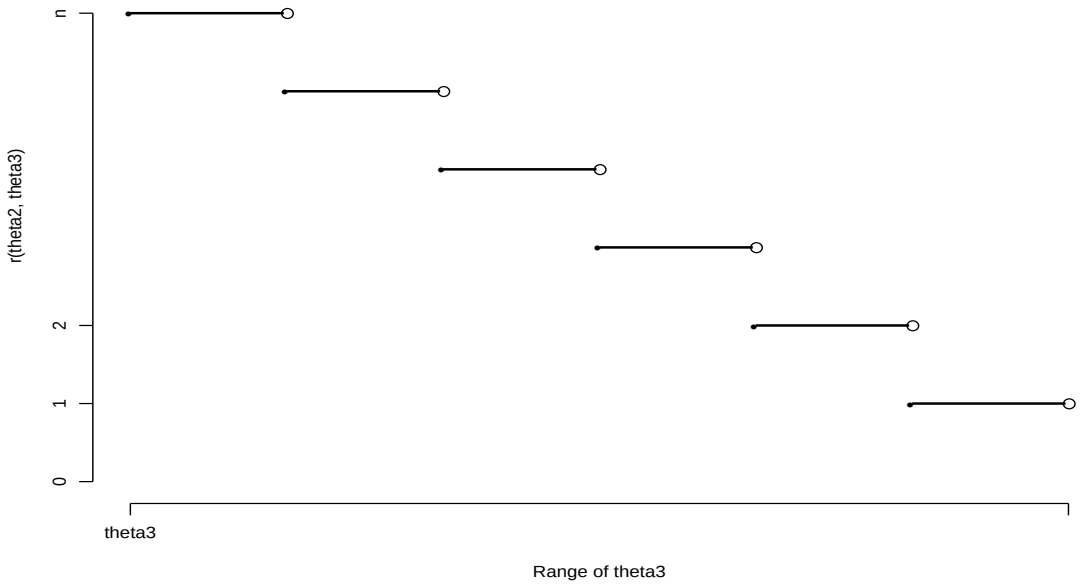
See Figure 3.1 for a pictorial representation of $\widehat{r}(\theta_2, \theta_3)$. The properties of $\widehat{r}(\theta_2, \theta_3)$ are summarized in the following theorems.

Theorem 2. For given θ_2, $\widehat{r}(\theta_2, \theta_3)$ has following properties:

1. *$\widehat{r}(\theta_2, \theta_3)$ is a non-increasing function of θ_3*
2. *$\lim_{\theta_3 \to \infty} \widehat{r}(\theta_2, \theta_3) = 1$*
3. *$\lim_{\theta_3 \to T_n} \widehat{r}(\theta_2, \theta_3) = n$*
4. *$\widehat{r}(\theta_2, \theta_3)$ is a step function having $(n-1)$ discontinuity points at $\theta_{3,(n-1)} < \theta_{3,(n-2)} < ... < \theta_{3,1}$ in $(T_n, +\infty)$.*

$$\widehat{r}(\theta_2, \theta_3) \quad = \quad \begin{cases} n & ;\ \ \theta_3 \in (T_n, \theta_{3,(n-1)}) \\ n-1 & ;\ \ \theta_3 \in [\theta_{3,(n-1)}, \theta_{3,(n-2)}) \\ ... & \\ r & ;\ \ \theta_3 \in [\theta_{3,r}, \theta_{3,(r-1)}) \\ ... & \\ 1 & ;\ \ \theta_3 \in [\theta_{3,1}, +\infty). \end{cases} \tag{3.7}$$

Figure 3.1. Function $\widehat{r}(\theta_2, \theta_3)$ as a function of θ_3 at given θ_2



See the appendix for a proof. Recall from the definition of the triangular hazard model that θ_3 must be finite and greater than θ_1. From Theorem 2, we have the following results: 1) if $\theta_3 = +\infty$, then $\theta_1 = T_1$ 2) if $\theta_3 = T_n$, then $\theta_1 = T_n$. As these two results violate our model assumptions, we put constraints around θ_1 and θ_3 such that $T_1 < \theta_1 < T_n < \theta_3 < +\infty$. To be more precise, $T_2 \leq \theta_1 \leq T_{n-1}$, $\theta_{3,(n-2)} \leq \theta_3 \leq \theta_{3,2}$ where $\theta_{3,(n-2)}$ and $\theta_{3,2}$ are defined in Theorem 2.

Similar properties can also be derived for $\widehat{r}(\theta_2, \theta_3)$ as a function of θ_2 for given θ_3 with the exception that

we are able to derive a closed form solution for $\theta_{2,r}$ as:

$$\begin{aligned}\theta_{2,r} &= \log[(\frac{T_r}{T_{r+1}})^r(\frac{\theta_3 - T_r}{\theta_3 - T_{r+1}})^{n-r}](-\frac{\sum_{i=1}^{r} T_i^2}{2T_r} + \frac{\sum_{i=1}^{r+1} T_i^2}{2T_{r+1}} \\ &\quad -\frac{\theta_3}{2} + \frac{\sum_{i=r+1}^{n}(\theta_3 - T_i)^2}{2(\theta_3 - T_r)} - \frac{\sum_{i=r+2}^{n}(\theta_3 - T_i)^2}{2(\theta_3 - T_{r+1})})^{-1}\end{aligned} \tag{3.8}$$

where $r = 1, ..., n-1$. In summary, the function $\widehat{r}(\theta_2, \theta_3)$ can be viewed as a bivariate step function which has a countable number of discontinuity points. Except those discontinuity points, the log likelihood function $\ell_n(T_r, \theta_2, \theta_3)$ is differentiable with respect to θ_2 and θ_3.

We have discussed the estimation of θ_1 for given θ_2 and θ_3. Now, we discuss the estimation of θ_2 and θ_3 for given θ_1. To simplify the notation, we replace θ_1 by the estimator T_r where r is a function of θ_2 and θ_3 and we drop the 'hat' from $\widehat{r}(\theta_2, \theta_3)$. The log likelihood function of θ_2 and θ_3 for given θ_1 is defined in equations (3.4), (A.2) and (A.3).

Let

$$\widehat{\theta_2}(\theta_3) = \arg \max_{\theta_{2,n-2} \le \theta_2 \le \theta_{2,2}} \{\ell_n(T_r, \theta_2, \theta_3)\}. \tag{3.9}$$

and

$$\widehat{\theta_3} = \arg \max_{\theta_{3,n-2} \le \theta_3 \le \theta_{3,2}} \{\ell_n(T_r, \widehat{\theta_2}(\theta_3), \theta_3)\}. \tag{3.10}$$

Under the constraints $T_2 \le \theta_1 \le T_{n-1}$, $\theta_{2,(n-2)} \le \theta_2 \le \theta_{2,2}$ and $\theta_{3,(n-2)} \le \theta_3 \le \theta_{3,2}$, using results from Theorem 1 and the score equations of θ_2 and θ_3, we evaluate the log likelihood function (3.4) iteratively and yield MLEs as follows: $\widehat{\theta_3}$, $\widehat{\theta_2} = \widehat{\theta_2}(\widehat{\theta_3}), \widehat{\theta_1} = T_{r(\widehat{\theta_2},\ \widehat{\theta_3})}$, where $\widehat{\theta_2}(.)$ and $\widehat{r}(.,.)$ are defined in equations (3.9) and (3.5) respectively.

3.2.3 Consistency and Limiting Distributions of Constrained MLEs for a Sample without Censoring

We consider the density function $f(t \mid \boldsymbol{\theta})$ as defined in subsection 3.2.1 with parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and constraints $\{T_2 \le \theta_1 \le T_{n-1},\ \theta_{2,(n-2)} \le \theta_2 \le \theta_{2,2},\ \theta_{3,(n-2)} \le$

$\theta_3 \leq \theta_{3,2}\}$ where $\theta_{2,(n-2)}, \theta_{2,2}$ as defined in (3.8) and $\theta_{3,(n-2)}, \theta_{3,2}$ as defined in (3.7). Before stating our main results in this section, we define more notation for later use. First, let $\boldsymbol{\theta}_0 = (\theta_{1(0)}, \theta_{2(0)}, \theta_{3(0)})$ denote the true value of $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{1(n)}, \widehat{\theta}_{2(n)}, \widehat{\theta}_{3(n)})$ denote the MLEs of $\boldsymbol{\theta}$ for sample size n. Second, let $\widehat{\theta}_{i:j}$ be the MLE of θ_i given θ_j at the true value of $\theta_{j(0)}$ when $i \neq j$. We also state one condition and two lemmas.

Condition 1: Assume θ_3 is known. For any $\theta_1 \neq \theta_{1(0)} \in [0, \theta_3]$, there exists a $\delta(\theta_1, \theta_{1(0)}) > 0$ such that

$$E_{\theta_{1(0)}}[sup\{\log f(t, \theta_1) - \log f(t, \theta_{1(0)}) : |\theta_1 - \theta_{1(0)}| \leq \delta(\theta_1, \theta_{1(0)})\}] \tag{3.11}$$

is less than 0.

Lemma 1: $\widehat{\theta}_{1:3}$ *is consistent under the given constraints and condition 1.*

Lemma 2: Let $X_1, ..., X_n$ *be i.i.d. with PDF* $g(x)$ *satisfying* $g(x) = 0$ *for* $x < 0$ *or* $x > \theta_3$ *and* $g(\theta_1) = \alpha > 0$ *where* $0 < \theta_1 < \theta_3$ *and* $var(X_i) = 1$. *Let* $X_{(i)}$ *denote the order statistics in relation to the change-point such that* $X_{(-r)} \leq X_{-(r-1)} \leq ... \leq X_{(-1)} \leq X_{(0)} = \theta_1 \leq X_{(1)} \leq ... \leq X_{(n-r)}$. *Then for any positive integer* $m \leq n - r$, $\{n(X_{(1)} - \theta_1) > 0, n(X_{(2)} - X_{(1)}), ..., n(X_{(m)} - X_{(m-1)})\}$ *converges in distribution to* $Y_1, ..., Y_m$ *where* $Y_1, ..., Y_m$ *are i.i.d and* Y_i*s are exponential with* $E(Y_i) = \alpha$. *For any positive integer* $m \leq r$, $\{n(X_{(-m)} - X_{(-m+1)}), ..., n(X_{(-2)} - X_{(-1)}), n(X_{(-1)} - \theta_1) < 0\}$ *converges in distribution to* $-Y_{-m}, ..., -Y_{-1}$ *where* $Y_{-m}, ..., Y_{-1}$ *are i.i.d. and* Y_{-i} *are exponential with* $E(Y_{-i}) = \alpha$.

We define the following quantities that will be used to define the variances of limiting distributions.

Definition 1: Part A) For θ_2, let

$$m_{22}(\boldsymbol{\theta}) \;=\; \mathbf{E}\{(\frac{\partial}{\partial \theta_2}) \log f(T|\boldsymbol{\theta})(\frac{\partial}{\partial \theta_2}) \log f(T|\boldsymbol{\theta})\} \tag{3.12}$$

where $\mathbf{E}_\theta$ denotes expectation with respect to $f(.|\theta)$.

Part B) For θ_3, we can not define m_{33} in the same way as m_{22} because θ_3 is the upper boundary of the support for the given PDF. Thus, we adopt the definition of m_{33} in Smith (1985) and define m_{33} as

$$m_{33}(\boldsymbol{\theta}) \;=\; \frac{1}{2} lim_{t\to\theta_3} \frac{f(t|\boldsymbol{\theta})}{\theta_3 - t} = \frac{\theta_2(1 - K(\theta_2,\theta_3))}{2(\theta_3-\theta_1)K(\theta_2,\theta_3)} \tag{3.13}$$

where

$$K(\theta_2,\theta_3) \;=\; 1 - \exp(-\frac{\theta_2\theta_3}{2})$$

Definition 2: Let $\{Y_n\}$ be a sequence of random variables and $\{r_n\}$ be a sequence of positive constants. If

$$lim_{a\to+\infty} lim_{n\to+\infty}\; sup\; pr(|Y_n| > ar_n) = 0,$$

then we say $Y_n \leq_p r_n$.

Definition 3: Let $\{X_n\}$ and $\{Y_n\}$ be sequences of i.i.d. random variables which are exponential with hazard rate 1. Without loss of generality, we assume $\theta_1 \leq \theta_3 - \theta_1$, let $\omega_1 = \theta_1 \log(\frac{\theta_3-\theta_1}{\theta_1})/(\theta_3 - 2\theta_1)$ and $\omega_2 = (\theta_3-\theta_1)\log(\frac{\theta_3-\theta_1}{\theta_1})/(\theta_3-2\theta_1)$. Then we define

$$S_1(\omega_1) \;= \textstyle\sum_{i=1}^{J_1}(X_i - \omega_1),\; S_2(\omega_2) \;= \displaystyle\sum_{i=1}^{J_2}(Y_i - \omega_2)$$

where J_1 is that value of j $(1 \leq j \leq n)$ that minimizes $\sum_{i=1}^{j}(X_i - \omega_1)$ and J_2 is that value of j that maximize $\sum_{i=1}^{j}(Y_i - \omega_2)$. Furthermore, let

$$\psi = \frac{\theta_2(\theta_3 - 2\theta_1)}{\theta_1(\theta_3-\theta_1)K(\theta_2,\theta_3)\log(\frac{\theta_3-\theta_1}{\theta_1})} \exp\,(-\frac{\theta_2\theta_1}{2}),$$

we define

$$Z_1(\omega_1,\; \omega_2,\; \psi) = \begin{cases} \psi^{-1}{\omega_1}^{-1}\sum_{i=1}^{J1} X_i & ;\;\; {\omega_1}^{-1}S_1 \leq -{\omega_2}^{-1}S_2 - 1 \\ -\psi^{-1}{\omega_2}^{-1}\sum_{i=1}^{J2} Y_i & ;\;\; otherwise. \end{cases} \tag{3.14}$$

We are in a position to state our main results about the asymptotic distributions of the MLEs for $\boldsymbol{\theta}$.

Theorem 4. Assume all constraints specified at the beginning of subsection 3.2.3 are satisfied. The values $\{m_{ii}, i = 2, 3\}$ are defined in Definition 1. Then

1. *there exist MLEs $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{1(n)}, \widehat{\theta}_{2(n)}, \widehat{\theta}_{3(n)})$ subject to the above constraints such that*

$$\widehat{\theta}_{1(n)} - \theta_{1(0)} \leq_p n^{-1}, \tag{3.15}$$

$$\widehat{\theta}_{2(n)} - \theta_{2(0)} \leq_p n^{-\frac{1}{2}}, \tag{3.16}$$

$$\widehat{\theta}_{3(n)} - \theta_{3(0)} \leq_p (n \log n)^{-\frac{1}{2}}. \tag{3.17}$$

In other words, the MLEs are consistent. Moreover,

$$\widehat{\theta}_{1(n)} - \widehat{\theta}_{1:3} = o_p(n^{-1}), \tag{3.18}$$

$$\widehat{\theta}_{2(n)} - \widehat{\theta}_{2:1} \leq_p (n \log n)^{-\frac{1}{2}}, \tag{3.19}$$

$$\widehat{\theta}_{3(n)} - \widehat{\theta}_{3:1} \leq_p n^{-\frac{1}{2}} (\log n)^{-1} \tag{3.20}$$

2. *$\{n(\widehat{\theta}_{1(n)} - \theta_{1(0)}), n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)}), (n \log n)^{\frac{1}{2}}(\widehat{\theta}_{3(n)} - \theta_{3(0)})\}$ converges in distribution to a random vector (Z_1, Z_2, Z_3) where (Z_1, Z_2, Z_3) are independent, Z_1 is defined in Definition 3, (Z_2, Z_3) are normal random variables with common mean 0 and respective variances m_{22}^{-1} and m_{33}^{-1} where $\{m_{ii}\}$ are defined in Definition 1.*

3.3 Single Administration Setting with Censoring

3.3.1 Derivation of MLEs for a Censored Sample

We extend our results from a sample without censoring to data with independent right censoring. Since the administration censoring in our motivating example is independent of survival time, the assumption of independent right censoring is valid

in Phase I clinical trials. Let $T_1 \leq T_2 \leq ... \leq T_r \leq \theta_1 < T_{r+1} \leq ... \leq T_n \leq ... \leq T_N \leq \theta_3$ be the order statistics of the uncensored and the censored observed times from N subjects with common hazard function $h(t)$ as defined in equation (3.1). Let T_n index the largest uncensored time among all the ordered time points, the change-point θ_1 lies in $[T_r, T_{r+1})$ and T_{N+1} is $+\infty$. In addition, θ_1 must be less than T_N in order for θ_1 and θ_3 to be identifiable. We use the same notation as those in section 3.2.2 to be consistent.

Then, the likelihood function for $\boldsymbol{\theta}$ is given by

$$\begin{aligned} L_N(\boldsymbol{\theta}|\boldsymbol{T}, \boldsymbol{C}) &= \prod_{i=1}^{N} [f(T_i)]^{C_i} [S(T_i)]^{1-C_i} \\ &= \frac{1}{K^N} \prod_{i=1}^{n} [h_0(T_i)]^{C_i} [S_0(T_i)]^{C_i} [K - 1 + S_0(T_i)]^{1-C_i} \\ &\quad \prod_{i=n+1}^{N} [K - 1 + S_0(T_i)]^{1-C_i}, \end{aligned} \tag{3.21}$$

in which $\boldsymbol{T} = (T_1, ..., T_N)$, $\boldsymbol{C} = (C_1, ..., C_N)$, and $K = K(\theta_2, \theta_3)$. From the assumptions, we have $C_n = 1$, $C_k = 0$ for any $k > n$. And the log-likelihood function of the proposed model can be written as

$$\begin{aligned} \ell_k(\boldsymbol{\theta} \mid \boldsymbol{T},\ \boldsymbol{C}) &= \sum_{i=1}^{n} C_i (\log(h_0(T_i|\boldsymbol{\theta})) - \sum_{i=1}^{n} C_i H_0(T_i|\boldsymbol{\theta}) \\ &\quad -k \log(K(\theta_2, \theta_3)) - \sum_{i=1}^{k} (1 - C_i) \log(\exp(-H_0(T_i|\boldsymbol{\theta})) \\ &\quad - \exp(-\frac{\theta_2 \theta_3}{2})) \\ &= Q_1(\theta_1, \theta_3) - \theta_2 Q_2(\theta_1, \theta_3) + Q_3(\theta_2, \theta_3) + Q_4(\theta_1, \theta_2, \theta_3), \end{aligned} \tag{3.22}$$

in which

$$
\begin{aligned}
Q_1(\theta_1,\theta_3) &= \sum_{i=1}^{n} C_i \log\left[\frac{T_i}{\theta_1} I_{[0,\theta_1]} + \frac{\theta_3 - T_i}{\theta_3 - \theta_1} I_{(\theta_1,\theta_3]}\right], \\
Q_2(\theta_1,\theta_3) &= \frac{1}{2}\sum_{i=1}^{n} C_i \left\{\frac{T_i^2}{\theta_1} I_{[0,\theta_1]} + \left[\theta_3 - \frac{(\theta_3 - T_i)^2}{\theta_3 - \theta_1}\right] I_{(\theta_1,\theta_3]}\right\}, \\
Q_3(\theta_2,\theta_3) &= \log(\theta_2)\sum_{i=1}^{n} C_i - k\log(K(\theta_2,\theta_3)), \\
Q_4(\theta_1,\theta_2,\theta_3) &= \sum_{i=1}^{k}(1-C_i)\log[\exp(-\frac{\theta_2}{2}\left\{\frac{T_i^2}{\theta_1} I_{[0,\theta_1]} + \left[\theta_3 - \frac{(\theta_3 - T_i)^2}{\theta_3 - \theta_1}\right] I_{(\theta_1,\theta_3]}\right\}) \\
&\quad - \exp(-\frac{\theta_2\theta_3}{2})].
\end{aligned}
$$

Theorems 1 & 2 for a sample without censoring still hold for a sample of independently right censored data with some modifications, such as replacing n by N in Theorems 1 & 2 where the likelihood function is based on (3.21) instead of (3.3). Similar properties still hold for $\widehat{r}(\theta_2,\theta_3)$ as a function of θ_2 for given θ_3.

We assume that the location of θ_3 in $[0, T_{N+1}]$ is unknown where T_{N+1} is $+\infty$ and θ_3 is greater than the largest observed time point T_N. We have identified the constraints in subsection 3.2.2. The constraints are $T_1 < \theta_1 < T_N$, $0 < \theta_2 < \infty$ and $T_N < \theta_3 < \infty$. More precisely, $T_2 \le \theta_1 \le T_{N-1}$, $\theta_{2,(N-2)} \le \theta_2 \le \theta_{2,2}$ and $\theta_{3,(N-2)} \le \theta_3 \le \theta_{3,2}$, where $\theta_{3,(N-2)}$ and $\theta_{3,2}$, $\theta_{2,(N-2)}$ and $\theta_{2,2}$ are derived similarly as in subsection 3.2.2. Barring the points of discontinuity of the function $\widehat{r}(\theta_2,\theta_3)$, the log likelihood function $\ell_N(\boldsymbol{\theta})$ is differentiable with respect to θ_2 and θ_3. We summarize the maximum likelihood estimation procedure as follows: Using results from Theorem 1 and the score equations of θ_2 and θ_3, evaluate log likelihood function (3.21) iteratively and yield MLEs as follows:

$$
\widehat{\theta}_3,\ \widehat{\theta}_2 = \widehat{\theta}_2(\widehat{\theta}_3),\ \widehat{\theta}_1 = T_{r(\widehat{\theta}_2,\ \widehat{\theta}_3)},
$$

where $\widehat{\theta}_2(.)$ and $\widehat{r}(.,.)$ are defined in equations (3.9) and (3.5) respectively.

3.3.2 Consistency and Limiting Distributions of Constrained MLEs for a Censored Sample

We consider the density function $f(t \mid \boldsymbol{\theta})$ and survival function $S(t \mid \boldsymbol{\theta})$ as defined in subsection 3.2.1 with parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and constraints $\{T_2 \leq \theta_1 \leq T_{N-1}$, $\theta_{2,(N-2)} \leq \theta_2 \leq \theta_{2,2}$, $\theta_{3,(N-2)} \leq \theta_3 \leq \theta_{3,2}\}$ where $\theta_{3,(N-2)}$ and $\theta_{3,2}$, $\theta_{2,(N-2)}$ and $\theta_{2,2}$ are derived similarly as in subsection 3.2.2.

Since the administration censoring in our motivating example is independent of survival time, the independent right censoring assumption is valid. Similar definitions and lemmas as those in subsection 3.2.3 hold under the assumption of independent right censoring with some modifications, such as replacing (3.11) by

$$E_{\theta_{1(0)}}[sup\{\log f^c(t,\theta_1) - \log f^c(t,\theta_{1(0)}) + \log S^{1-c}(t,\theta_1) - \log S^{1-c}(t,\theta_{1(0)}) : \quad (3.23)$$
$$|\theta_1 - \theta_{1(0)}| \leq \delta(\theta_1, \theta_{1(0)})\}]$$

in Condition 1, replacing 3.12 by

$$m_{22}(\boldsymbol{\theta}) \;=\; \mathbf{E}\{(\frac{\partial}{\partial\theta_2}) \log f^c(T|\boldsymbol{\theta}) S^{1-c}(T|\boldsymbol{\theta}) (\frac{\partial}{\partial\theta_2}) \log f^c(T|\boldsymbol{\theta}) S^{1-c}(T|\boldsymbol{\theta})\} \quad (3.24)$$

in Definition 1 where c is the censor indicator associated with time T. Furthermore, Similar results as Theorem 4 in subsection 3.2.3 holds for censored samples with some modifications, such as replacing n by N in Theorem 4.

The rationale is as follows: First, from the estimation procedure, we have $\widehat{\theta_1} = T_r$ where T_r is an order statistic from sequence $\{T_i\}_{i=1}^N$. The arguments to derive two independent random walks still apply. Second, extending the asymptotic results of $\widehat{\theta_2}$ and $\widehat{\theta_3}$ from a sample without censoring to censored data follows similar argument as extending the regular MLE asymptotic properties to censored samples.

3.4 Numerical Studies

In this section, we investigate the finite sample behavior of the constrained MLEs for the parameters of interest in the triangular hazard model via simulation studies. All results were produced in SAS.

We generated 1000 random samples from each of the following two triangular hazard models: (1) $\theta_1 = 4$, $\theta_2 = 0.05$ and $\theta_3 = 18$, (2) $\theta_1 = 3$, $\theta_2 = 0.004$ and $\theta_3 = 18$. Uniform censoring times were generated in the interval $(0, U)$ with U selected to give expected censoring proportion of 0 (no censoring) and 20% (censoring), respectively. For both censoring proportions, samples of $n = 30$, $n = 50$, $n = 100$ and $n = 200$ (where n is the sample size) were generated. The convergence criteria were based on relative changes of estimated parameter values and log likelihood values. All the computations of $\widehat{\boldsymbol{\theta}}$ were subject to the specified constraints in subsection 3.2.3 and subsection 3.3.2.

The means of the MLEs and the coverage probability of 95% confidence intervals for the MLEs $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ in 1000 replications were computed. The 95% confidence intervals for the MLEs $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ were calculated based on Theorem 4. Note that no asymptotic variance estimator is available for the MLE of θ_1. Therefore, no confidence interval is derived for $\widehat{\theta}_1$.

The simulation results pertaining to the evaluation of $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ are presented in Table 3.1, and Table 3.2. We list the mean of MLEs, the mean of asymptotic standard error (SE), the empirical (sample) SE and the coverage probability of 95% confidence interval for different sample sizes and censoring proportions.

In Table 3.1, the data were simulated under $\theta_1 = 4$, $\theta_2 = 0.05$ and $\theta_3 = 18$. The overall performance of the proposed MLEs for the given triangular hazard model is

Table 3.1. **Parameter estimates and their inference with 1000 replications in which $\theta_1 = 4$, $\theta_2 = 0.05$ and $\theta_3 = 18$**

Sample size (n)	Parameter Censoring proportion	Mean of MLE	Mean of Asymptotic SE	Empirical (Sample) SE	Coverage Prob. of 95% CI (in percent)
30	0%	3.634	NA	0.734	NA
		0.0458	0.0196	0.0192	87.2
		17.695	1.696	1.217	88.7
	20%	4.418	NA	0.843	NA
		0.0561	0.0212	0.0216	85.1
		17.432	1.599	1.412	86.8
50	0%	4.282	NA	0.623	NA
		0.0472	0.0183	0.0186	89.1
		18.349	1.275	1.114	90.6
	20%	3.675	NA	0.747	NA
		0.0551	0.0201	0.0203	88.8
		17.572	1.214	1.108	89.6
100	0%	3.884	NA	0.575	NA
		0.0523	0.0161	0.0151	91.3
		18.326	0.996	1.012	92.1
	20%	4.211	NA	0.667	NA
		0.0533	0.0181	0.0171	90.9
		18.426	1.204	1.214	91.4
200	0%	4.003	NA	0.463	NA
		0.05083	0.0156	0.0145	92.1
		18.231	0.861	0.866	93.6
	20%	3.943	NA	0.576	NA
		0.0516	0.0162	0.0152	91.5
		17.682	0.989	1.003	92.4

reasonably well regardless of the sample size and censoring proportion. The bias of the mean of MLEs decreases as the sample size increases. The coverage probability of 95% confidence interval is increasing as the sample size increases. When sample sizes are 100, 200, the biases are small, the estimated standard errors agree well with the sample standard errors, and the coverage probabilities are accurate. For all sample sizes, the biases and coverage probability of CIs are of same magnitudes. The mean of asymptotic SEs based on Theorem 4 are reasonably close to those empirical sample SEs based on 1000 Monte Carlo runs. The difference between these two SEs does not change very much as the sample size changes.

Table 3.2. **Parameter estimates and their inference with 1000 replications in which $\theta_1 = 3$, $\theta_2 = 0.004$ and $\theta_3 = 18$**

Sample size (n)	Parameter Censoring proportion	Parameter Name	Mean of MLE	Mean of Asymptotic SE	Empirical (Sample) SE	Coverage Prob. of 95% CI (in percent)
30	0%	θ_1	2.791	NA	0.771	NA
		θ_2	0.0171	0.0176	0.0164	98.2
		θ_3	17.728	1.762	1.749	91.2
	20%	θ_1	3.305	NA	0.878	NA
		θ_2	0.0179	0.0186	0.0178	98.1
		θ_3	17.438	1.886	1.857	90.8
50	0%	θ_1	2.812	NA	0.739	NA
		θ_2	0.0121	0.0163	0.0153	98.3
		θ_3	18.079	1.684	1.676	92.2
	20%	θ_1	3.204	NA	0.778	NA
		θ_2	0.0124	0.0172	0.0159	98.2
		θ_3	17.887	1.754	1.746	91.8
100	0%	θ_1	3.149	NA	0.646	NA
		θ_2	0.0098	0.0151	0.0143	98.6
		θ_3	17.892	1.592	1.588	93.3
	20%	θ_1	2.845	NA	0.686	NA
		θ_2	0.0102	0.0161	0.0149	98.5
		θ_3	18.068	1.666	1.654	93.1
200	0%	θ_1	2.938	NA	0.589	NA
		θ_2	0.0083	0.0128	0.0122	99.4
		θ_3	18.031	1.458	1.445	94.2
	20%	θ_1	3.082	NA	0.0676	NA
		θ_2	0.0094	0.0145	0.0139	99.1
		θ_3	18.055	1.565	1.552	93.4

In Table 3.2, the data were simulated under $\theta_1 = 3$, $\theta_2 = 0.004$ and $\theta_3 = 18$. The mean of MLEs performed reasonably well except for θ_2. When true θ_2 value was close to 0 - the boundary of the parameter support, the relative bias of the mean MLE was much larger than the one when the true value is further away from 0. However, the coverage probability of 95% confidence interval of $\widehat{\theta}_2$ is very high compared to the case when true θ_2 is further away from 0 because the relative larger variance when the true value is close to 0.

3.5 Estimation and Schedule Finding by Maximum Likelihood Method

3.5.1 Likelihood Function and Estimation in Original Parametric Form

We first consider the estimation procedures for one schedule with multiple administrations without censoring. We demonstrate whether the new MLE estimation procedure for a single administration setting developed in early part of this chapter can be extended to multiple administration setting.

The likelihood function of $\boldsymbol{\theta}$ for given $\boldsymbol{T} = \{t_i, i = 1, ..., n\}$ and schedule $s = \{s_l, l = 1, ..., m\}$ is

$$L_n(\boldsymbol{\theta} \mid \boldsymbol{T}, s) = \prod_{i=1}^{n} \lambda(t_i|\boldsymbol{\theta}, s) \exp[-\Lambda(t_i|\boldsymbol{\theta}, s)] \tag{3.25}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. Then, the log likelihood function is in the following form

$$\ell_n(\boldsymbol{\theta}) = Q_1(\theta_1, \theta_3) - \theta_2 Q_2(\theta_1, \theta_3) + Q_3(\theta_2, \theta_3) \tag{3.26}$$

where

$$\begin{aligned} Q_1(\theta_1, \theta_3) &= \sum_{i=1}^{n} \log[\sum_{l=1}^{m} h(t_i - s_l|\boldsymbol{\theta})] - \sum_{i=1}^{n} \log(\theta_2), \\ Q_2(\theta_1, \theta_3) &= \sum_{i=1}^{n} \sum_{l=1}^{m} H(t_i - s_l|\boldsymbol{\theta})/\theta_2, \\ Q_3(\theta_2, \theta_3) &= \sum_{i=1}^{n} \log(\theta_2), \end{aligned}$$

$$h(t_i - s_l|\boldsymbol{\theta}) = \theta_2 \frac{(t_i - s_l)}{\theta_1} I(0, \theta_1] + \theta_2 \frac{(\theta_3 - t_i + s_l)}{\theta_3 - \theta_1} I(\theta_1, \theta_3] \tag{3.27}$$

$$H(t_i - s_l|\boldsymbol{\theta}) = \theta_2 \frac{(t_i - s_l)^2}{2\theta_1} I(0, \theta_1] + \theta_2 [\frac{\theta_3}{2} - \frac{(\theta_3 - t_i + s_l)^2}{2(\theta_3 - \theta_1)}] I(\theta_1, \theta_3]. \tag{3.28}$$

where $I(0, \theta_1]$ is an indicator function of $t_i - s_l$ in the interval $(0, \theta_1]$ and $I(\theta_1, \theta_3]$ is an indicator function of $\theta_3 - t_i - s_l$ in the interval $(\theta_1, \theta_3]$.

Recall our finding earlier in this chapter that the log likelihood function for θ_1 with θ_2 and θ_3 fixed attains its maximum at one of the order statistics $(T_1, ..., T_n)$ of the observed $\{t_i, i = 1, ..., n\}$ where T_1 is the earliest observed time and T_n is the latest observed time. Specifically,

$$\widehat{\theta_1} = T_{\widehat{r}(\theta_2,\theta_3)}$$

where $\widehat{r}(\theta_2, \theta_3)$ is a function of θ_2 and θ_3 that can only take the discrete values in [1, n] and

$$\widehat{r}(\theta_2, \theta_3) = \arg \underset{r\in\{1,...,n\}}{max} \{\ell_n(T_r, \theta_2, \theta_3)\}$$

In order to extend this result to multiple administration setting, we need to demonstrate either the first order derivative of the log likelihood function (3.26) about θ_1 is greater than 0 (or less than 0) consistently within each interval $[T_r, T_{r+1}]$ $(1 \leq r < n)$ or the second order derivative of the log likelihood function (3.26) about θ_1 is greater than 0 consistently within each $[T_r, T_{r+1}]$. From the log likelihood function (3.26), the first derivative about θ_1 is

$$\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_1}Q_1(\theta_1, \theta_3) - \theta_2\frac{\partial}{\partial\theta_1}Q_2(\theta_1, \theta_3). \tag{3.29}$$

where

$$\frac{\partial}{\partial\theta_1}Q_1(\theta_1, \theta_3) = \sum_{i=1}^{n}\frac{1}{\sum_{l=1}^{m}h(t_i - s_l)}\frac{\partial}{\partial\theta_1}h(t_i - s_l) \tag{3.30}$$

$$\frac{\partial}{\partial\theta_1}Q_2(\theta_1, \theta_3) = \sum_{i=1}^{n}\sum_{l=1}^{m}\frac{\partial}{\partial\theta_1}H(t_i - s_l). \tag{3.31}$$

Similarly, the second derivative about θ_1 is

$$\frac{\partial^2}{\partial\theta_1^2}\ell_n(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_1^2}Q_1(\theta_1, \theta_3) - \theta_2\frac{\partial^2}{\partial\theta_1^2}Q_2(\theta_1, \theta_3). \tag{3.32}$$

where

$$\frac{\partial^2}{\partial\theta_1^2}Q_1(\theta_1,\theta_3) = \sum_{i=1}^{n}\frac{-1}{\sum_{l=1}^{m}h^2(t_i-s_l)}[\frac{\partial}{\partial\theta_1}h(t_i-s_l)]^2 + \sum_{i=1}^{n}\frac{1}{\sum_{l=1}^{m}h(t_i-s_l)}\frac{\partial^2}{\partial\theta_1^2}h(t_i-s_l) \tag{3.33}$$

$$\frac{\partial^2}{\partial\theta_1^2}Q_2(\theta_1,\theta_3) = \sum_{i=1}^{n}\sum_{l=1}^{m}\frac{\partial^2}{\partial\theta_1^2}H(t_i-s_l). \tag{3.34}$$

We first show $\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta})$ is not consistently greater > 0 (or < 0) in $[T_r, T_{r+1}]$. Let $A_{i1} = \{t_i - s_l \mid t_i - s_l \in (0,\theta_1], l = 1,...,m_i\}$, $A_{i2} = \{t_i - s_l \mid t_i - s_l \in (\theta_1,\theta_3], l = 1,...,m_i\}$ and $A_{i3} = \{t_i - s_l \mid t_i - s_l \in (\theta_3,+\infty), l = 1,...,m_i\}$. From definition of the triangular hazard function $h(t|\boldsymbol{\theta})$, the observed time points in A_{i3} do not contribute to the likelihood function through hazard but do so through cumulative hazard function. However, the cumulative hazard function of the time points in A_{i3} is a function of θ_2 and θ_3. So they do not contribute to the derivatives of the log likelihood function about θ_1. Substituting the hazard function (3.27), cumulative hazard function (3.28) and their derivatives into the first order derivative (3.29), we have the following results

$$\frac{\partial}{\partial\theta_1}Q_1(\theta_1,\theta_3) = \sum_{i=1}^{n}\frac{\sum_{A_{i1}}-\frac{t_i-s_l}{\theta_1^2}+\sum_{A_{i2}}\frac{\theta_3-t_i+s_l}{(\theta_3-\theta_1)^2}}{\sum_{A_{i1}}\frac{t_i-s_l}{\theta_1}+\sum_{A_{i2}}\frac{\theta_3-t_i+s_l}{(\theta_3-\theta_1)}} \tag{3.35}$$

$$\frac{\partial}{\partial\theta_1}Q_2(\theta_1,\theta_3) = -\sum_{i=1}^{n}\{\sum_{A_{i1}}\frac{(t_i-s_l)^2}{2\theta_1^2}+\sum_{A_{i2}}\frac{(\theta_3-t_i+s_l)^2}{2(\theta_3-\theta_1)^2}\}. \tag{3.36}$$

Since the multiple administrations are taken consecutively, the hazard functions overlap. The term $\sum_{A_{i1}}\frac{t_i-s_l}{\theta_1}+\sum_{A_{i2}}\frac{\theta_3-t_i+s_l}{(\theta_3-\theta_1)}$ in the first order derivative is in the denominator, which complicates the calculation. Even with the constraints as those defined in Chapter III, the negative part in $\frac{\partial}{\partial\theta_1}Q_1(\theta_1,\theta_3)$ makes it possible for the first order derivative to be less than 0 for some θ_1 while greater than 0 for other θ_1 in $[T_r, T_{r-1}]$.

Similarly, we can show that $\frac{\partial^2}{\partial\theta_1^2}\ell_n(\boldsymbol{\theta})$ is not consistently greater than 0 in $[T_r, T_{r+1}]$. In addition to the existing observed time points, we also have the shifted time points $\{t_i - s_l, l = 1, ..., m_i; i = 1, 2, ..., n\}$. For estimation purposes, all the observed time points and shifted time points are incorporated into the likelihood function. However, we can not prove that the MLE of θ_1 can be any one of the observed time points or shifted time points under the identifiability constraints $T_1 < \theta_1 < T_n < \theta_3$. Therefore, the findings in early part of this chapter can not be directly extended to the setting of multiple administrations. In next section, we will explore the option of reparameterizing the original triangular hazard model and its application in multiple schedule setting.

3.5.2 Likelihood Function and Estimation in Reparameterized Form

Since the proposed MLEs for a single administration setting in last section can not be extended to a multiple schedule setting, we use the original triangular hazard model proposed by Braun et al. (2005) for estimation procedures in a multiple schedule setting in this section. For easier computing implementation, we reparameterize the triangular hazard function as

$$h(t \mid \boldsymbol{\beta}) = \begin{cases} \beta_1 t & ; \quad 0 \le t \le \frac{\beta_2\theta_3}{\beta_1+\beta_2}, \beta_1 > 0 \\ \beta_2(\theta_3 - t) & ; \quad \frac{\beta_2\theta_3}{\beta_1+\beta_2} < t \le \theta_3, \beta_2 > 0, \\ 0 & ; \quad t > \theta_3, or\ t < 0. \end{cases} \tag{3.37}$$

in which $\boldsymbol{\beta} = (\beta_1, \beta_2, \theta_3)$. We can also write the hazard function as $h_0(t \mid \boldsymbol{\beta}) = \beta_1 t I_{(0,\theta_1]} + \beta_2(\theta_3 - t)I_{(\theta_1,\theta_3]}$, in which $I_{(a,b]}$ is the indicator that t lies in the interval $(a, b]$. By setting the intersection of the two lines in the new hazard function (3.37) as the change point in the old hazard function (3.1), we can express θ_1 and θ_2 in

terms of β_1, β_2 and θ_3 as

$$\theta_1 = \frac{\beta_2}{\beta_1 + \beta_2}\theta_3 \tag{3.38}$$

$$\theta_2 = \frac{\beta_2}{\beta_1 + \beta_2}\beta_1\theta_3 \tag{3.39}$$

The new hazard function not only has a simpler mathematical structure but also leads to a nice interpretation for the change point. The change point can be considered as a proportion of the total duration of hazard. This proportion is expressed as the second slope over the sum of the two slopes (we have defined the slopes as positive values only in the new hazard function). If the hazard goes up quickly and decreases slowly, then the change point should be closer to 0 than θ_3. Otherwise, if the hazard goes up slowly and decreases quickly, then the change point should be closer to θ_3 than 0. When we discuss the prior information for the change-point with medical investigators, we can consider the change-point as the proportion of total duration. Through this proportion, we can derive the relationship between the two slopes of the triangular hazards. Note that we do not use the reparameterized hazard form in Bayesian estimation procedures. They are only used in maximum likelihood estimation procedures. We do not use the normalized CDF and PDF in reparameterized form either.

Then the new cumulative hazard function (CHF) for T_i is

$$H(t \mid \boldsymbol{\beta}) = \begin{cases} 0 & ; \quad t < 0, \\ \frac{1}{2}\beta_1 t^2 & ; \quad 0 \le t \le \frac{\beta_2\theta_3}{\beta_1+\beta_2}, \beta_1 > 0 \\ \frac{\beta_1\beta_2\theta_3^2}{2(\beta_1+\beta_2)} - \frac{1}{2}\beta_2(\theta_3 - t)^2 & ; \quad \frac{\beta_2\theta_3}{\beta_1+\beta_2} < t \le \theta_3, \\ \frac{\beta_1\beta_2\theta_3^2}{2(\beta_1+\beta_2)} & ; \quad t > \theta_3 \end{cases}$$

For the multiple administration setting, the likelihood function (3.25) and log likeli-

hood function (3.26) still hold with the modifications to $h(t_i - s_l)$ and $H(t_i - s_l)$ as follows

$$\begin{aligned} & h(t_i - s_l|\boldsymbol{\beta}) \qquad (3.40) \\ = \;& \beta_1(t_i - s_l)I(0, \frac{\beta_2\theta_3}{\beta_1+\beta_2}] + \beta_2(\theta_3 - t_i + s_l)I(\frac{\beta_2\theta_3}{\beta_1+\beta_2}, \theta_3] \end{aligned}$$

$$\begin{aligned} & H(t_i - s_l|\boldsymbol{\beta}) \qquad (3.41) \\ = \;& \beta_1\frac{(t_i - s_l)^2}{2}I(0, \frac{\beta_2\theta_3}{\beta_1+\beta_2}] + [\frac{\beta_1\beta_2\theta_3^2}{2(\beta_1+\beta_2)} - \frac{1}{2}\beta_2(\theta_3 - t_i + s_l)^2]I(\frac{\beta_2\theta_3}{\beta_1+\beta_2}, \theta_3]. \end{aligned}$$

Therefore, we can estimate $(\beta_1, \beta_2, \theta_3)$ using maximum likelihood with (3.38) and (3.39) as constraints. So the MLEs of $(\beta_1, \beta_2, \theta_3)$ are constrained MLEs in the reparameterized form.

Following the notation in subsection 3.5, we extend the likelihood function to include censored observations. Let (t_{ik}, c_{ik}) denote the ith observations in schedule $s^{(k)}$ and n_k denote the number of observations in schedule $s^{(k)}$. Then the likelihood function is

$$L_n(\boldsymbol{\beta} \mid \boldsymbol{T}, \boldsymbol{C}) \;=\; \prod_{k=1}^{K}\prod_{i=1}^{n_k} f_k(t_{ik})^{c_{ik}} S_k(t_{ik})^{(1-c_{ik})} \qquad (3.42)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \theta_3)$, $\boldsymbol{T} = \{t_{ik}, i = 1, ..., n_k, k = 1, ..., K\}$, $\boldsymbol{C} = \{c_{ik}, i = 1, ..., n_k, k = 1, ..., K\}$. And the log likelihood function is

$$\begin{aligned} & \ell_n(\boldsymbol{\beta}) \qquad (3.43) \\ = \;& \sum_{k=1}^{K}\sum_{i=1}^{n_k}[c_{ik}\log(\sum_{l=1}^{m_k} h(t_{ik} - s_l)) - \sum_{l=1}^{m_k} H(t_{ik} - s_l)] \end{aligned}$$

where $h(t_{ik} - s_l))$ and $H(t_{ik} - s_l))$ are defined in (3.40) and (3.41). The maximum likelihood method can be used to estimate the unknown parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \theta_3)$ with constraints (3.38) and (3.39).

3.6 Estimation and Schedule Finding by Bayesian Approach

The basic structure is described in Braun et al. (2005), where the detailed methodology and algorithm can be found. We summarize the main ideas here before proceeding to simulation studies. Note that we use the original parameterization i.e. $\boldsymbol{\theta}$ proposed by Braun et al. (2005) in the Bayesian estimation procedure.

Let D denote the data available at evaluation time t_{cur}, then the likelihood function is (3.42), denoting the prior by $p(\boldsymbol{\theta})$, the posterior of $\boldsymbol{\theta}$ is

$$g(\boldsymbol{\theta}|D) = \frac{L_n(\boldsymbol{\theta} \mid \boldsymbol{T}, \boldsymbol{C})p(\boldsymbol{\theta})}{\int L_n(\boldsymbol{\theta} \mid \boldsymbol{T}, \boldsymbol{C})p(\boldsymbol{\theta})d\theta} \tag{3.44}$$

Because the integral in the posterior can not be obtained analytically under our assumed model, the posterior quantities are computed via Markov chain Monte Carlo (MCMC) methods. Specifically, a Metropolis-Hastings algorithm (Robert and Casella, 1999; Gelman et al., 2004) is used. We experiment with different starting values and are convinced that the chains converge and cover the entire posterior distribution using multiple sequences and plots. We eliminate a total of 1000 iterations as burn-in and then generate additional 3000 samples for summarization.

The priors ($\boldsymbol{\theta}$) are chosen in the form of $p_3(\theta_3)p_1(\theta_1|\theta_3)p_2(\theta_2)$ because $\theta_1 < \theta_3$. Both $p_3(\theta_3)$ and $p_1(\theta_1|\theta_3)$ follow generalized Beta distribution. $p_2(\theta_2)$ follows Gamma distribution. Tuning parameters k_1, k_2 and k_3 are specified to calibrate the prior distribution so that the data dominate the posterior distribution. The same notations are used as those in Braun et al. (2005). For prior elicitation, refer Braun et al. (2005) for detailed description.

In the following section, we investigated the finite sample behavior of the constrained MLEs for the parameters of interest in the reparameterized triangular hazard model and the performance of the proposed model in MTS finding via simulation

studies. We also implement the simulation study designed in Braun et al. (2005) with treatment schedule reassignment. The purpose of these simulation studies is to compare the results by maximum likelihood method to those by Bayesian approach. All results are produced in SAS.

3.7 Application to KGF trial

In order to compare the performance of the triangular hazard model using maximum likelihood method and Bayesian approach, we use the same true parameter values for both methods. Note that β_1 and β_2 were only estimated directly by maximum likelihood method while θ_1 and θ_2 were only estimated directly by Bayesian approach. θ_3 was directly estimated by both MLE and Bayesian procedures.

We note that θ_2 true values were very small (< 0.001) in the simulation study of Braun et al. (2005). From the numerical studies in early part of this chapter, the maximum likelihood method had trouble in estimating θ_2 when the true values are less than 0.004 because these small values are too close to the boundary of the support 0. When the true parameter values are very close to the boundary of the parameter space, the MLEs are not stable. It may induce larger bias (Hall and Wang, 1999). Therefore, we rescale the time such that 1 unit time in Braun et al. (2005) as 0.1 unit time in our simulation studies. For example, in Braun et al. (2005), $\theta_3 = 18$ and $\theta_1 = 3$. We set $\theta_3 = 1.8$ and $\theta_1 = 0.3$ so that we can have θ_2 value reasonably far away from 0 (the boundary of the support for θ_2). Then we can derive reasonable estimate of θ_2 using maximum likelihood method and keep the threshold for the probability of toxicity in our simulation study to be the same as that in Braun et al. (2005). As a result, reasonable comparisons of the two estimation methods can be carried out in the simulation result section. Based on equations (3.38) and (3.39),

we derive the true $\beta_i, i = 1, 2$ values for simulation. The true parameter values are shown in Table 3.3, which also contains the actual day 100 ($\omega = 10$ in simulations) probability of toxicity for each schedule under all scenarios. For trial implementation, refer to Section 2.4 for details and safety concerns.

Table 3.3. **True parameter values of the triangular hazard model for simulation studies where $\theta_1 = 0.3$ and $\theta_3 = 1.8$**

Scenario	θ_2	β_1	β_2	True Toxicity Prob of Schedule 1	2	3	4	5	6	Threshold Prob of Toxicity
1	0.08	0.27	0.06	0.20	0.36	0.48	0.59	0.67	0.74	0.2
2	0.04	0.14	0.03	0.11	0.20	0.28	0.36	0.43	0.49	0.2
3	0.03	0.09	0.02	0.07	0.14	0.20	0.26	0.31	0.36	0.2
4	0.05	0.16	0.03	0.12	0.23	0.32	0.40	0.47	0.54	0.4
5	0.04	0.13	0.03	0.10	0.18	0.26	0.33	0.40	0.46	0.4
6	0.03	0.11	0.02	0.08	0.16	0.23	0.29	0.35	0.40	0.4
7	0.05	0.18	0.04	0.13	0.25	0.35	0.43	0.51	0.58	0.3
8	0.04	0.14	0.03	0.11	0.20	0.28	0.36	0.43	0.49	0.3
9	na	na	na	0.07	0.14	0.20	0.26	0.31	0.36	0.2

Recall the motivating example in Chapter II in which the investigators wished to study $K = 6$ schedules corresponding to 2, 4, 6, 8, 10 and 12 weeks of therapy. We considered 6 therapy schedules in our simulation studies, $s^{(1)}, ..., s^{(6)}$, in which $s^{(k)}$ did not have natural units and $s^{(k)} = \{s_{lk}, l = 1, ..., m_k\}$ for $k = 1, ..., 6$.

We studied the design with a maximum sample size of 30 patients, which is feasible in Phase I trials but also sufficient to determine the MTS with reasonable accuracy demonstrated in our simulations. In each simulation, the subject interarrival times were assumed to be uniformly distributed from 12 to 16 days.

We examined the design's performance in nine scenarios using the criterion specified in section 2.4. In the first six scenarios, schedule $s^{(j)}$ was true optimal schedule under the jth scenario for $j = 1, ..., 6$. In scenario 7, the true MTS was located be-

tween schedule 2 and 3, while in scenario 8, the target schedule (MTS) lay between schedule 3 and 4 but closer to schedule 3.

Furthermore, we examined the design's performance under the setting where the model was misspecified in scenario 9, where schedule 3 was the true MTS, but the data was not simulated from the triangular hazard model. Instead, we assumed the toxicity occurred uniformly over the interval $[10+14(j-1), 10+14j]$ under schedule $s^{(j)}$. In all other scenarios, the actual times to toxicity were simulated assuming the triangular hazard model with the true parameter values shown in Table 3.3.

Regarding the prior selection in the Bayesian estimation procedures, we used the investigator belief as a priori to determine the hyperparameters. In our application, the hazard of toxicity for a single administration vanished after an average of 1.8 unit time and with a range of 0.4 to 10. The upper bound for θ_1 was 0.4. By the methods described in Braun et al. (2005), the hyperparameters for prior distributions $p(\theta_3)$ and $p(\theta_1|\theta_3)$ were derived in the computer programs. The investigators also believed that 12 weeks of KGF would not cause more toxicity, which led us to use the true parameter value 0.03 of θ_2 in scenario 6 of Table 3.1 as the mean of prior distribution $p(\theta_2)$.

3.7.1 Study Result and Conclusion

Tables 3.4, 3.5 and 3.6 display the simulation results for triangular hazard model. Table 3.4 displays the estimated parameter values and corresponding standard errors for all scenarios. Note that β_1 and β_2 were only estimated directly using maximum likelihood method then θ_1 and θ_2 were derived from β_1 and β_2 on the MLE row in Table 3.4. In contrast, θ_1 and θ_2 were only estimated directly using Bayesian approach then β_1 and β_2 were derived from θ_1 and θ_2 on the Bayesian row in Table 3.4. θ_3 was directly estimated using both maximum likelihood and Bayesian methods. The

Table 3.4. **Estimated parameter values of the triangular hazard model. Each entry is the estimated parameter value (standard deviation).**

Scenario	Estimation Method	Estimated Value of θ_1	θ_2	θ_3	β_1	β_2
1	MLE	0.44 (0.249)	0.074 (0.054)	1.76 (0.681)	0.265 (0.249)	0.090 (0.086)
	Bayesian	0.35 (0.074)	0.089 (0.017)	1.81 (0.368)	0.258 (0.032)	0.061 (0.017)
2	MLE	0.48 (0.276)	0.039 (0.027)	1.84 (0.772)	0.142 (0.143)	0.056 (0.042)
	Bayesian	0.26 (0.043)	0.045 (0.011)	1.74 (0.523)	0.179 (0.043)	0.031 (0.009)
3	MLE	0.44 (0.261)	0.029 (0.022)	1.75 (0.778)	0.147 (0.148)	0.047 (0.032)
	Bayesian	0.27 (0.057)	0.032 (0.008)	1.78 (0.522)	0.117 (0.044)	0.021 (0.006)
4	MLE	0.35 (0.268)	0.054 (0.026)	1.68 (0.783)	0.261 (0.249)	0.059 (0.039)
	Bayesian	0.27 (0.073)	0.055 (0.012)	1.77 (0.521)	0.194 (0.066)	0.033 (0.008)
5	MLE	0.36 (0.254)	0.047 (0.024)	1.63 (0.797)	0.234 (0.244)	0.058 (0.034)
	Bayesian	0.27 (0.072)	0.042 (0.010)	1.77 (0.354)	0.159 (0.057)	0.031 (0.007)
6	MLE	0.33 (0.258)	0.037 (0.023)	1.58 (0.792)	0.222 (0.254)	0.042 (0.044)
	Bayesian	0.27 (0.072)	0.031 (0.008)	1.77 (0.351)	0.117 (0.035)	0.021 (0.006)
7	MLE	0.36 (0.262)	0.061 (0.026)	1.71 (0.792)	0.284 (0.263)	0.059 (0.043)
	Bayesian	0.27 (0.064)	0.059 (0.012)	1.78 (0.374)	0.221 (0.063)	0.042 (0.008)
8	MLE	0.51 (0.259)	0.039 (0.024)	1.85 (0.702)	0.153 (0.148)	0.076 (0.049)
	Bayesian	0.26 (0.056)	0.046 (0.011)	1.76 (0.399)	0.205 (0.065)	0.032 (0.007)
9	MLE	0.27 (0.185)	0.048 (0.029)	1.39 (0.721)	0.348 (0.325)	0.085 (0.065)
	Bayesian	0.27 (0.057)	0.032 (0.008)	1.78 (0.284)	0.117 (0.025)	0.021 (0.005)

estimates by Bayesian approach are posterior means. Tables 3.5 and 3.6 summarize the frequency at which each schedule was selected as the MTS. The first row of each scenario in Tables 3.5 and 3.6 contains the percentages of simulations in which each schedule is identified as the MTS while the second row of each scenario contains the mean percentages of subjects assigned to a given schedule among the simulations.

We can see from Table 3.4 that the estimates by the maximum likelihood method have larger biases in most scenarios and also have larger corresponding standard error estimates across all scenarios. We suspect this fact is due to the small sample size in our Phase I trials and informative priors used in the Bayesian approach. However, the parameter estimates in Table 3.4 are still close to the true parameters values, which provide confidence for us to interpret the results in Tables 3.5 and 3.6.

Table 3.5. **Performance of the triangular model with 30 patients by *Maximum Likelihood*. Each entry on the 1st row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on 2nd line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	76.7	14.0	7.3	1.3	0.7	0
	68.3	17.2	7.2	4.2	1.8	1.3
2	22.7	42.6	19.6	8.7	5.1	1.3
	20.6	38.2	15.4	10.6	9.4	5.8
3	10.1	22.1	41.5	13.2	7.6	5.5
	15.7	27.5	33.2	10.8	7.2	5.6
4	2.4	11.1	19.8	34.3	21.2	11.2
	5.7	13.9	18.1	32.9	17.9	11.5
5	0.7	5.3	11.2	23.1	38.7	21.0
	3.6	10.9	13.2	22.8	30.1	19.4
6	0.	4.7	6.3	10.0	25.3	53.7
	4.6	8.1	9.9	11.4	13.4	42.6
7	10.7	35.0	32.2	11.7	7.7	2.7
	14.0	33.3	29.7	10.3	8.3	4.4
8	0.0	13.3	36.0	24.7	16.5	8.5
	4.8	14.4	29.1	20.4	18.8	12.7
9	14.7	25.5	35.3	15.6	6.9	2.0
	13.6	21.1	30.5	17.5	9.0	8.3

By comparing the first row within each scenario in both Tables 3.5 and 3.6, there is no marked difference between the final recommendations by two estimation methods. In addition, there is little difference in the final recommendation whether the assumed model is right or not. However, the Bayesian approach performs consistently better than the maximum likelihood method in terms of the percentage of identifying the correct MTS in the simulation trials. More specifically, in scenarios 1-2, the Bayesian approach has a much higher percentage (over 15% more) of identifying the optimal schedule 1 or 2 as the MTS than the maximum likelihood method. In scenarios 3-4, the Bayesian approach has a moderately higher percentage (between $3\% - 10\%$

Table 3.6. **Performance of the triangular model with 30 patients by *Bayesian Approach*. Each entry on the 1st row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on 2nd line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	94.5	4.5	1.0	0	0	0
	83.5	11.2	4.2	1.1	0	0
2	10.8	79.2	9.3	0.7	0	0
	15.7	67.1	11.0	5.1	1.1	0
3	0.8	29.2	54.1	13.1	2.8	0
	8.7	26.3	43.7	12.9	5.3	3.1
4	0	5.5	34.4	38.2	18.5	3.4
	3.3	6.2	31.3	35.7	16.1	8.4
5	0	0	14.6	27.7	43.1	14.6
	3.3	3.3	12.5	29.8	36.1	18.2
6	0	0	1.5	14.6	24.6	59.3
	3.3	3.3	4.4	14.3	25.4	49.3
7	11.2	38.7	36.6	9.3	4.2	0
	15.3	32.7	28.2	13.6	6.8	3.4
8	2.1	18.1	46.7	25.7	6.2	1.2
	4.0	25.1	36.6	27.2	5.2	1.9
9	5.1	26.7	43.3	21.8	3.1	0
	9.2	29.3	36.5	19.1	5.1	0.8

more) of identifying the optimal schedule as the MTS than the maximum likelihood method. In scenario 7, the pattern of the results is similar as those in scenarios 2-3. The results in scenario 8 follows the same pattern as those in scenarios 3-4. But the results from both schedules 2 and 3 in scenario 7 or schedules 3 and 4 in scenario 8 are very similar by both estimation methods.

To compare the average subject assignment within a trial by both estimation methods, we look at second row within each scenario in both Tables 3.5 and 3.6. We find that the highest average percentage of subjects is assigned to the true MTS by both estimation methods across all scenarios. More specifically, in scenarios 1-2,

the highest average percentage of subjects assigned to the true MTS is much larger (over 15%) than the average percentages of subjects assigned to other schedules. However, in other scenarios, there usually exists at least one schedule whose average percentage of subjects assigned to the specified schedule is very close to the highest average percentage of subjects assigned to the true MTS in a given scenario. This fact is expected since the rates of toxicity between neighboring schedules are closer as the true optimal schedule becomes longer and it is harder for any model based algorithms to distinguish the schedules.

To assess the effect of the model misspecification on identifying the MTS, we compare scenario 3 to scenario 9 within Table 3.5, then within Table 3.6 when the data are not from the assumed model but from the same misspecified model in two simulation studies. The performance of the algorithm in scenario 9 is similar to that in scenario 3 even though scenario 3 has a higher percentage of identifying the schedule 3 as MTS and has more subjects assigned to schedule 3 during a trial on average by both estimation methods. This fact is expected since the subject assignment during a trial is determined by the timing when a subject is having a toxicity but the final schedule recommendation is determined by the overall rate of toxicity in a trial.

Table 3.7. **Average number of observed toxicities (out of 30 subjects) in the simulated trials using the triangular hazard model.**

Estimation Method	Scenario								
	1	2	3	4	5	6	7	8	9
MLE	8.98	7.32	6.23	11.69	10.87	9.44	9.77	8.32	6.24
Bayesian	7.29	6.35	5.97	11.58	10.43	9.96	9.71	8.25	5.39

Table 3.7 displays an interesting side note on the average number of observed toxicities out of the total 30 patients. The numbers were very close for both methods.

There was no clear pattern on which method had more observed toxicities across all scenarios. However, the maximum likelihood method had slightly more toxicities in all but scenario 6. Note that the scenario 8 had more subjects who experienced toxicities than scenario 3 but less than scenario 4. This is due to the fact that scenario 8 had a higher threshold toxicity rate $p_\omega = 0.3$ than scenario 3 ($p_\omega = 0.2$) but lower threshold rate than scenario 4 ($p_\omega = 0.4$).

Table 3.8. **Average number of subjects (out of 30 subjects) had reassignment in the simulated trials using the triangular hazard model.**

Estimation Method	Scenario 1	2	3	4	5	6	7	8	9
MLE	0.78	1.87	4.65	5.23	7.22	8.45	2.86	5.98	5.46
Bayesian	0.69	1.56	4.04	5.13	6.84	8.32	2.28	5.23	4.91

Table 3.8 displays another interesting side note on the average number of subjects received treatment schedule reassignments out of the total 30 patients. The numbers were very close by both estimation methods with Bayesian approach having slightly less numbers across all scenarios. We note that the number of subjects had schedule reassignments increased as the treatment schedule became longer as is expected. We also note that the number of subjects received schedule reassignments was larger in scenario 9 than that in scenario 3. This fact may be due to the reason that more subjects assigned to longer schedules than MTS in scenario 9 than those in scenario 3.

In this chapter, we used both the maximum likelihood and Bayesian approaches to estimate the parameters of a triangular hazard model proposed for optimal schedule finding in multiple treatment schedule setting. We first proposed a new procedure to derive MLEs for the change-point and boundary parameters. The large sample properties of the proposed MLEs were also derived. Then we showed these results

could not be extended to a setting in which multiple administrations are given to each subjects and the number of the administrations varies among the subjects. To address this problem, we used the constrained MLEs of reparameterized triangular hazard function for optimal treatment schedule finding. Via simulation, we demonstrated both maximum likelihood and Bayesian methods performed well under a variety of settings, including the setting where the model is misspecified. However, even in the misspecified prior case, the Bayesian approach had a higher percentage identifying the correct MTS than the maximum likelihood method.

CHAPTER IV

A Non-mixture Cure Model for Optimal Treatment Schedule Finding

4.1 Introduction

Despite the advantage of our proposed mixture cure model over the existing triangular hazard model in term of interpreting the cure fraction directly, the mixture cure model still has its limitations. First, the marginal distribution of the mixture cure model is a mixture distribution that complicates maximum likelihood estimation. We introduced a latent variable cure status in the likelihood function and used the EM algorithm to estimate the parameters of interest. From a computing resource perspective, this estimation procedure is inefficient compared to the procedure of maximizing the log-likelihood function directly. Furthermore, we modeled the event rate as a function of treatment schedule in Chapter II, resulting in a population hazard function that does not have a proportional hazards (PH) structure, which is a frequently used feature for most survival models. Even if in the situation where the PH assumption is not correct, the mixture cure model does not provide a natural structure for testing the departure from PH assumption.

In this chapter, we propose a non-mixture cure model for optimal treatment schedule finding in early-phase clinical trials using adaptive designs, which overcomes the limitations just mentioned. We use both maximum likelihood and Bayesian methods

to estimate the unknown parameters and identifying the optimal treatment schedule. Subject accural, data monitoring and outcome-adaptive decision-making are done sequentially through the simulation studies as in Chapters II and III.

4.2 Model Specification

We first consider the proposed model in a single administration setting. Using the same notation as in Chapter II, at evaluation time t^{cur}, the amount of time that a subject has been observed is denoted by T and the indicator of whether or not a subject is observed with a toxicity prior to time t^{cur} is denoted by C. Let $\theta, \boldsymbol{\phi}$ denote the parameters of interest. Then the survival function for T, also known as the population survival function, is given by

$$S_p(t|\theta, \boldsymbol{\phi}) \;=\; \exp(-\theta F(t|\boldsymbol{\phi})),\ \theta > 0 \tag{4.1}$$

where $F(t|\boldsymbol{\phi})$ is a cumulative distribution function (CDF). Since $S_p(\infty) \;=\; \exp(-\theta) \;>\; 0$, the population survival function $S_p(t|\theta, \boldsymbol{\phi})$ is not a proper survival function. The cure rate is $S_p(\infty) \;=\; \exp(-\theta)$, with a corresponding event rate, i.e. the probability of toxicity is $1 \;-\; \exp(-\theta)$. Furthermore, the population density function is given by

$$f_p(t|\theta, \boldsymbol{\phi}) \;=\; \theta f(t|\boldsymbol{\phi}) \exp(-\theta F(t|\boldsymbol{\phi}))$$

where $f(t|\boldsymbol{\phi})$ is the probability density function (PDF) corresponding to $F(t|\boldsymbol{\phi})$. Therefore, the corresponding population hazard function is given by

$$h_p(t|\theta, \boldsymbol{\phi}) \;=\; \theta f(t|\boldsymbol{\phi}),$$

and the non-mixture cure model yields a multiplicative hazard function in θ and $f(t|\boldsymbol{\phi})$. This population hazard function has the PH structure if covariate effects are modeled through θ. But the population hazard function derived from a mixture

cure model will not have the PH structure even if the covariate effect is modeled through cure fraction. Furthermore, the PH property in the non-mixture cure model is computationally attractive as both maximum likelihood and Bayesian approaches are relatively easy to implement. In addition, even if the PH assumption is not correct, the non-mixture cure model provides a natural structure for testing the departure from PH assumption.

In the multiple schedule setting, we assume that k treatment schedules, $s^{(1)}, ..., s^{(k)}$, are investigated in a trial where $s^{(j)}=(s_1, s_2, ..., s_{m_j})$ and that the jth schedule has a total of m_j administrations. Furthermore, $s^{(j)}$ is nested in $s^{(j+1)}$ for each $j = 1, ..., k-1$. We assume that the form of $f(.)$ does not change with successive administrations. We define the total hazard of toxicity at time t for a subject treated with schedule $s^{(k)}$ to be

$$h_k(t|\boldsymbol{\phi}, \theta, s^{(k)}) = \theta_k \frac{\sum_{l=1}^{m_k} f(t - s_l|\boldsymbol{\phi})}{m_k},$$

where $f(t|\boldsymbol{\phi})$ was the same as that defined in the single administration setting. We also put constraints that the total hazard increases as the number of administrations increases i.e. $\theta_{k+1}/m_{k+1} \geq \theta_k/m_k$. Note that the total hazard is not a sum of hazards from each administration as was done in the mixture cure model case. We will further explain the rationale for this approach after we derive the survival function. Then the cumulative hazard function (CHF) up to t for a subject treated with schedule $s^{(k)}$ is

$$H_k(t|\boldsymbol{\phi}, \theta, s^{(k)}) = \theta_k \frac{\sum_{l=1}^{m_k} F(t - s_l|\boldsymbol{\phi})}{m_k},$$

where $F(.)$ is defined as in the single administration setting. Therefore the survival function up to t for a subject treated with schedule $s^{(k)}$ is given by

$$S_k(t|\boldsymbol{\phi}, \theta, s^{(k)}) = \exp[-\theta_k \frac{\sum_{l=1}^{m_k} F(t - s_l|\boldsymbol{\phi})}{m_k}].$$

We now explain why the CHF $H_k(.)$ is modeled as the average of the CHF from each administration instead of a sum. In order for $S_k(\infty) = \exp(-\theta_k)$ to be the cure rate of treatment schedule k when t approaches ∞, the CHF $H_k(t)$ must approach θ_k. Since $F(t)$ is chosen as a CDF and approaches 1 as t approaches ∞, $\sum_{l=1}^{m_k} F(t - s_l)/m_k$ approaches 1 as t approaches ∞ and the CHF $H_k(t)$ approaches θ_k as t approaches ∞.

In our application, we specify a parametric form for $F(.)$ where

$$F(t|\alpha, \gamma) \;=\; 1 - \exp[-t^{\alpha} \exp(\gamma)] \tag{4.2}$$

follows a Weibull distribution with $\alpha \geq 2$ so that the population hazard function $h_p(.)$ has the required non-monotonic pattern that increases to a peak then decreases. Therefore, the hazard of toxicity attributed to a single administration is modeled by a Weibull pdf. (See Figure 4.1)

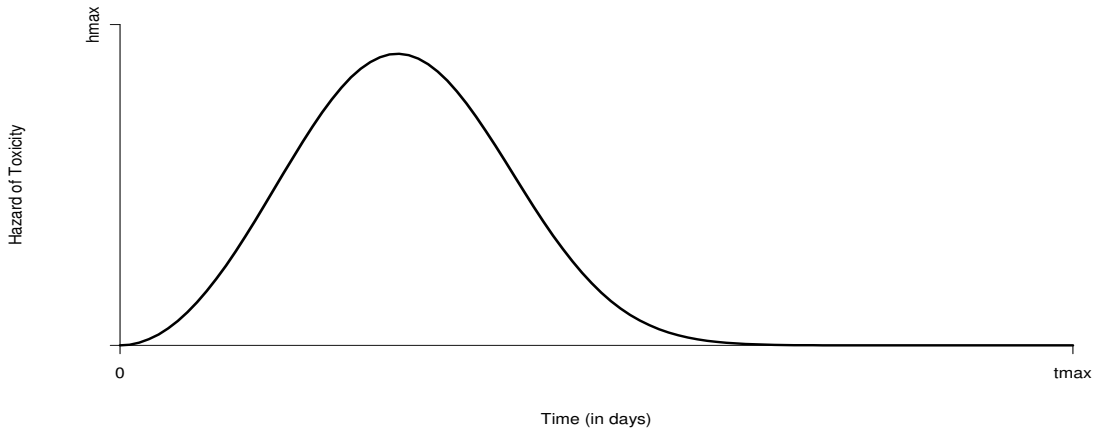
As we discussed in the early part of this section, modeling covariate effects through θ_k would allow the population hazard function to have a PH structure. In our application, treatment schedule is the only covariate to consider. We chose to adopt a linear regression model for log-transformed θ_k i.e. $\log(\theta_k) = \beta_0 + \beta_1 k$ for simplicity. Since the multiple schedules we investigated are nested, $m_k \;=\; m_1 k$. Then $\theta_k/m_k = (\exp(\beta_0)/m_1)(\exp(\beta_1)^k/k)$ is a non-decreasing function of k when $\beta_1 > 0$. Therefore, the constraint $\theta_{k+1}/m_{k+1} \geq \theta_k/m_k$ mentioned earlier is satisfied. Thus, the probability of toxicity at time t is given by

$$G_k(t|\boldsymbol{\phi}, \theta, s) = 1 \;-\; \exp[-\exp(\beta_0 + \beta_1 k)\frac{\sum_{l=1}^{m_k} F(t - s_l|\boldsymbol{\phi})}{m_k}]$$

and the probability of toxicity increases as the treatment schedules become longer.

Using the same notation as in Chapter II, we represent the observed data by $D_{ik} \;=\; (t_{ik}, c_{ik})$ for the observed time and censoring indicator of the ith sub-

Figure 4.1. A Weibull pdf as the hazard function for a single administration of an agent



ject assigned to schedule $s^{(k)}$, respectively, $i = 1, ..., n_k$, where n_k is the number of subjects assigned to schedule $s^{(k)}$. Then the likelihood function of the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)$, $\boldsymbol{\phi} = (\alpha, \gamma)$ can be written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\phi} \mid \boldsymbol{D}) &= \prod_{k=1}^{K} \prod_{i=1}^{n_k} (h_k(t_{ik}|\boldsymbol{\theta}))^{c_{ik}} S_k(t_{ik}|\boldsymbol{\theta}) \\ &= \prod_{k=1}^{K} \prod_{i=1}^{n_k} [\exp(\beta_0 + \beta_1 k) \frac{\sum_{l=1}^{m_k} f(t - s_l|\boldsymbol{\phi})}{m_k}]^{c_{ik}} \\ &\quad \exp[-\exp(\beta_0 + \beta_1 k) \frac{\sum_{l=1}^{m_k} F(t - s_l|\boldsymbol{\phi})}{m_k}] \end{aligned} \tag{4.3}$$

where $\boldsymbol{D} = \{D_{ik}, i = 1, ..., n_k; k = 1, ..., K\}$. Then the log likelihood is given by

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\phi} \mid \boldsymbol{D}) &= \sum_{k=1}^{K}\sum_{i=1}^{n_k}[c_{ik}\log(\sum_{l=1}^{m_k} f(t - s_l|\boldsymbol{\phi})] \\ &+ \sum_{k=1}^{K}\sum_{i=1}^{n_k}[c_{ik}(\beta_0 + \beta_1 k)] - \sum_{k=1}^{K}\sum_{i=1}^{n_k}[c_{ik}log(m_k)] - \\ &\sum_{k=1}^{K}\sum_{i=1}^{n_k}[\exp(\beta_0 + \beta_1 k)\frac{\sum_{l=1}^{m_k} F(t - s_l|\boldsymbol{\phi})}{m_k}]. \end{aligned} \tag{4.4}$$

4.3 Estimation and Schedule Finding by Maximum Likelihood Method

Consider maximization of the log-likelihood function $\ell(\boldsymbol{\beta}, \boldsymbol{\phi} \mid \boldsymbol{D})$ in (4.4) with respect to the distribution parameters $\boldsymbol{\phi}$ and regression coefficients $\boldsymbol{\beta}$. Since the Weibull family of distributions $F(t|\boldsymbol{\phi})$ satisfy the regularity conditions for MLEs to exist, the first derivatives about the parameters for gradient vector and the second derivatives about the parameters for Hessian matrix can be derived from the log-likelihood function directly. Thus Newton-Raphson method can be used to estimate the parameters. For our implementation, we have used SAS built-in non-linear optimization procedures called NLPTR which performs well for small- to medium-sized problems and allows users to specify the linear constraints (include the boundary constraints) on parameters of interest. The estimation of the probability of toxicity and identification of the optimal treatment schedule at each evaluation are defined in Section 2.4.

4.4 Estimation and Schedule Finding by Bayesian Approach

Our schedule-finding algorithm based on a Bayesian approach begins with independent informative priors for $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. The informative priors may be obtained either based on historical data from previous single administration studies or by elicitation from investigators. Since the posterior distributions can not be solved

analytically under the assumed model, we compute the posterior quantities using Markov chain Monte Carlo (MCMC) methods. Specifically, a Metropolis-Hastings algorithm (Robert and Casella, 1999; Gelman et al., 2004) is used. We experiment with different starting values and are convinced that the chains converge and cover the entire posterior distribution using multiple sequences and plots. We eliminate a total of 1000 iterations as burn-in and then generate additional 3000 samples for summarization.

4.4.1 Priors Based on Historical Data

If dose-toxicity data for a single administration are available from previous studies, then these data may be used to obtain the priors on $\boldsymbol{\phi}$ in the multiple schedule trials. Denote the time of ith subject in the historical trial by t_{hi} and the toxicity indicator of this subject by c_{hi} and all historical data by $D_h = \{(t_{hi}, c_{hi}), i = 1, 2, ...n_h\}$, where n_h is the number of subjects in the historical trial. Then the likelihood function of the available historical data is

$$\begin{aligned} L_h(\boldsymbol{\phi}, \theta \mid D_h) &= \prod_{i=1}^{n_h} (f_p(t_{hi}))^{c_{hi}} (S_p(t_{hi}))^{1-c_{hi}} \\ &= \prod_{i=1}^{n} [\theta f(t_{hi})]^{c_{hi}} \exp[-\theta F(t_{hi})], \end{aligned}$$

where θ is the cure fraction parameter attributed to one administration and is not the function of the $\boldsymbol{\beta}$ in multiple administration setting. Assume a vague prior on all parameters of interest before the historical data are observed, then the posterior of $\boldsymbol{\phi}, \theta$ given the historical data is then $f_{po}(\boldsymbol{\phi}, \theta \mid D_h) \propto L_h(\boldsymbol{\phi}, \theta \mid D_h)$ and the prior used at the start of the schedule-finding trial is $f_{po}(\boldsymbol{\phi}, \theta \mid D_h)$. The informative priors based on historical data in cure model setting have been studied extensively by Chen et al. (1999). We refer readers to their work for more details.

An alternative way of using the available historical data to define priors is to estimate the parameters of interest by either Bayesian approach or maximum likelihood method. Refer to subsection 2.5.1 for details. Let $\widehat{\mu}_{\cdot}$ and $\widehat{\sigma}^2_{\cdot}$ denote the respective mean and variance of the prior distribution for a parameter of interest derived from the historical data where $'.'$ represents each of parameters α, γ, β_0, and β_1.

We then select specific functional forms for the prior distributions. Because $\alpha > 2$, we assume $\alpha - 2$ has a gamma distribution with parameters (c_1, d_1) such that α has mean c_1/d_1 and variance c_1/d_1^2. Given the prior mean and variance of α that are derived from historical data, We set $c_1/d_1 + 2 = \widehat{\mu}_\alpha$ $c_1/d_1^2 = \widehat{\sigma}^2_\alpha$, then find $c_1 = (\widehat{\mu}_\alpha - 2)^2/\widehat{\sigma}^2_\alpha$ and $d_1 = (\widehat{\mu}_\alpha - 2)/\widehat{\sigma}^2_\alpha$. Since γ may be any real number, we assume γ follows a normal distribution with hyperparameters $(\mu_\gamma, \sigma_\gamma)$ where we set μ_γ equal to parameter estimate and σ_γ equal to the standard deviation of parameter estimate from historical data.

Furthermore, if dose-toxicity data for a single course consisting of multiple administrations are also available from previous studies, then these data can be used to obtain the priors on $\boldsymbol{\beta}$ in addition to the α, γ in the multiple schedule trials. Following similar arguments as above, we assume β_0 follows a normal distribution with mean equal to $\widehat{\mu}_{\beta_0}$ and variance equal to $\widehat{\sigma}^2_{\beta_0}$ that are the estimates based on the historical data. Since $\beta_1 > 0$, we assume β_1 has a gamma distribution with parameters (c_2, d_2) such that β_1 has mean c_2/d_2 and variance c_2/d_2^2. Given the prior mean and variance of α that are derived from historical data, We set $c_2/d_2 = \widehat{\mu}_{\beta_1}$ $c_2/d_2^2 = \widehat{\sigma}^2_{\beta_1}$, then find $c_2 = (\widehat{\mu}_{\beta_1})^2/\widehat{\sigma}^2_{\beta_1}$ and $d_2 = (\widehat{\mu}_{\beta_1})/\widehat{\sigma}^2_{\beta_1}$.

When the individual subject data from trials of the single administration are available but no data available for a single course, the source for the informative priors on the parameters of interest can be a mixture of historical data and elicitation

from the investigators. For example, in our assumed model, the priors for α, γ may be from historical data while the priors on $\beta_i, i = 0, 1$ are elicited from investigators.

4.4.2 Elicited Priors

When individual subject data from trials of the single administration or a single course are not available, informative priors must be elicited from investigators. This may be done in various ways, with the particular elicitation method tailored to the clinical setting and investigators' level of technical expertise. We employed the following method in our simulation trials.

When no historical data are available, we consider the specific forms of the prior distributions with slightly different hyperparameterization from the case when historical data are available. We assume $\alpha - 2$ has a Gamma distribution with mean μ_α and variance $\mu_\alpha \delta_\alpha$, as we need $\alpha \geq 2$ to create our non-monotonic hazard. We assume the remaining parameters each have a similar prior distribution as that in subsection 4.4.1.

With regard to the cure fraction parameters $\boldsymbol{\beta}$, we ask the investigators to specify an *a priori* value, P_k, for the cumulative probability of toxicity for schedule $k, k = 1, 2, \ldots, K$. Based upon the simple linear regression model $E\{\log[-\log(1 - P_k)]\} = b_0 + b_1 k$, we use ordinary least squares to find estimates of b_0 and b_1; we let μ_{β_0} equal the estimate of b_0 and μ_{β_1} equal the larger of 0.01 and the estimate of b_1.

With regard to the hazard shape parameters $\boldsymbol{\phi}$, we ask the investigators to specify an *a priori* value for the limiting cumulative probability of toxicity for a single administration. We denote this value Q_0 and note that Q_0 must be less than the value of P_1 elicited earlier. We also ask investigators to select two time points t_1 and t_2 and supply *a priori* values Q_1 and Q_2 for the cumulative probabilities of toxicity at t_1 and t_2, respectively, for a single administration. Based upon Equation (4.1),

we first derive the value $\theta^* = -\log(1 - Q_0)$. Plugging θ^* and Equation (4.2) into Equation (4.1), some algebra gives us two equations in terms of two parameters a and g:

$$\begin{aligned} \log\{-\log[1 + \log(1 - Q_1)/\theta^*]\} &= a\log(t_1) + g \\ \log\{-\log[1 + \log(1 - Q_2)/\theta^*]\} &= a\log(t_2) + g \end{aligned}$$

If we let $\hat{a}$ and $\hat{g}$ denote the respective solutions to a and g in the above equations, we set $\mu_\alpha = max\{2.01, \hat{a}\}$ and $\mu_\gamma = \hat{g}$.

4.4.3 Calibrating the Prior Distributions of Parameters of Interest

We emphasize that sufficient attention must be given to selection of the variance hyperparameter values, as they cannot be made arbitrarily large, as is often done with Bayesian analysis of large data sets. In our research setting, very few data are available, especially at the beginning of a trial. If the prior distribution is spread over too broad range, the algorithm will run poorly, severely hindering our algorithm's ability to identify optimal schedules during a trial. Therefore, sensitivity analysis of priors on parameters of interest is essential in order for the data to dominate the prior distribution in adaptive early-phase clinical trials. As a result, those initially estimated hyperparameters still need fine tuning for priors to work in conjunction with the data to allow the schedule-finding algorithm provide a safe and reliable design. However, the calibration methods are different depending on how the informative priors are derived.

When historical data are available, the following calibration methods are used. Recall the prior on α follows a gamma distribution with parameters (c_1, d_1). We set $c_1 = a\widehat{c_1}$ and $d_1 = a\widehat{d_1}$ where $\widehat{c_1}$ and $\widehat{d_1}$ are the initial estimates of (c_1, d_1) derived from the historical data. The tuning constant a scales the values of (c_1, d_1) and modulates

the variability of $f(\alpha)$. In addition, the prior on γ follows a normal distribution $N(\mu_\gamma, \sigma_\gamma)$. We set $\mu_\gamma = \widehat{\mu}_\gamma$, $\sigma_\gamma = b\widehat{\sigma}_\gamma$ where $\widehat{\mu}_\gamma$ and $\widehat{\sigma}_\gamma$ are the initial estimates of $(\mu_\gamma, \sigma_\gamma)$ based on historical data. b is the tuning constant in the sensitivity analysis of priors on γ. Similarly, the prior on β_0 follows a normal distribution $N(\mu_{\beta_0}, \sigma_{\beta_0})$ where we set $\mu_{\beta_0} = \widehat{\beta}_0$, $\sigma_{\beta_0} = d\widehat{\sigma}(\beta_0)$ and d is a tuning constant used to modulate the variances of β_0 in the sensitivity analysis of the prior on β_0. Since the prior on β_1 follows a gamma distribution with parameters (c_2, d_2). We set $c_2 = e\widehat{c_2}$ and $d_2 = e\widehat{d_2}$ where $\widehat{c_2}$ and $\widehat{d_2}$ are the initial estimates of (c_2, d_2) derived from the historical data. The tuning constant e scales the values of (c_2, d_2) and modulates the variability of $f(\beta_1)$. By simulating the toxicity times of a small number of subjects, we can compare the prior means for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ to their respective posterior values and evaluate the effects of a small amount of data on priors.

When no historical data available, we adopt the following calibration methods. Our approach is to set $\delta_\alpha = 1.0$ and $\sigma_* = 0.10\mu_*$, where the asterisk represents each of parameters of interest, γ, β_0, and β_1. Through trial-and-error with several small simulation studies, we modulate the variances until we find suitable values. For example, by simulating the first five subjects to experience toxicity on the shortest schedule, we can compare the posterior means for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ to their respective prior means to determine values of the variances that allow the data to overcome the prior.

Another consideration of the influence of the priors is to examine the resulting value of $F(\omega|\boldsymbol{\phi}, s^{(j)})$ for each schedule $s^{(j)}$ to determine whether the prior may produce pathological behavior by placing too much of the probability mass of $F(\omega|\boldsymbol{\phi}, s^{(j)})$ near 0 or 1 because the consequent distribution of $F(\omega|\boldsymbol{\phi}, s^{(j)})$ determine which schedule to be identified as MTS.

We also used misspecified priors such that the prior means are different from the

true parameters of interest. The simulation results changed slightly, but overall, the final conclusions were relatively unchanged. Thus the proposed model is insensitive to the misspecified priors and different starting values as long as the priors are informative at the beginning of a trial but do not dominate the data at later points in the trial.

4.5 Application to KGF trial

In this section, we investigate the performance of the proposed non-mixture cure model in optimal treatment schedule finding via simulation studies using both Bayesian and maximum likelihood approaches. All results are produced in SAS.

The KGF trial is described in the motivating example of Chapter II. In this section, we implement this trial using our proposed non-mixture cure model under different study set up scenarios. In order to compare the performance of the maximum likelihood and the Bayesian approaches, we use the same true parameter values for both methods. To be consistent with the simulation set up for other proposed models in previous chapters, we rescale the time such that 1 unit time in Braun et al. (2005) as 0.1 unit time in our simulation studies. Then, we can compare the performance of different proposed models by the same estimation method in scenario 9. The true parameter values are shown in Table 4.1, which also contains the actual day 100 ($\omega = 10$ in simulations) probability of toxicity for each schedule under all scenarios.

To be consistent with application set up in the previous chapters, we use very similar study settings in this section. We summarize the main ideas as follows. We consider 6 therapy schedules in our simulation studies, $s^{(1)}, ..., s^{(6)}$, in which $s^{(k)}$ do not have natural units and $s^{(k)} = \{s_{lk}, l = 1, ..., m_k\}$ for $k = 1, ..., 6$. We study the

Table 4.1. **True parameter values of the non-mixture cure model for simulation studies**

Scenario	α	γ	β_0	β_1	True Toxicity Prob of Schedule 1	2	3	4	5	6	Threshold Prob of Toxicity
1	3	0	-1.97	0.47	0.20	0.30	0.43	0.60	0.77	0.90	0.2
2	3	0	-2.27	0.39	0.14	0.20	0.28	0.38	0.51	0.65	0.2
3	3	0	-2.40	0.30	0.11	0.15	0.20	0.26	0.33	0.42	0.2
4	3	0	-2.11	0.36	0.16	0.22	0.30	0.40	0.52	0.65	0.4
5	3	0	-2.47	0.36	0.11	0.16	0.22	0.30	0.40	0.52	0.4
6	3	0	-2.82	0.36	0.08	0.11	0.16	0.22	0.30	0.40	0.4
7	3	0	-2.27	0.49	0.15	0.24	0.36	0.52	0.69	0.85	0.3
8	3	0	-3.26	0.69	0.07	0.14	0.26	0.45	0.69	0.90	0.3
9	na	na	-2.40	0.30	0.11	0.15	0.20	0.26	0.33	0.42	0.2

design with a maximum sample size of 30 patients. In each simulation, the subject interarrival times are assumed to be uniformly distributed from 12 to 16 days.

We examine the design's performance in nine scenarios using the criterion specified in Section 2.4. In the first six scenarios, schedule $s^{(j)}$ is optimal under the jth scenario for $j = 1, ..., 6$. In scenario 7, the true MTS is located midway between schedule 2 and 3, while in scenario 8, the target schedule (MTS) lay between schedule 3 and 4 but much closer to schedule 3. In scenario 9, we examine the design's performance under model misspecification, where schedule 3 is the true MTS, but the data is not simulated from the assumed non-mixture cure model with Weibull distribution. Instead, we model the toxicity to occur uniformly over the interval $[10 + 14(j - 1), 10 + 14j]$ under schedule $s^{(j)}$. Except in scenario 9, the actual times to toxicity are simulated assuming the parameter values shown in Table 4.1 under the non-mixture cure model with Weibull distribution of parameter values $\alpha = 3$ and $\gamma = 0$;

When using a Bayesian approach, All the priors are informative, which mean their

variations are relatively small compared to the conventional priors for the studies with large sample size. With regard to the prior distributions for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$, we use the elicited values from the investigators as in Table 4.1. For example, the investigators supply the values $P_1 = 0.08$, $P_2 = 0.11$, $P_3 = 0.16$, $P_4 = 0.22$, $P_5 = 0.30$, $P_6 = 0.40$, corresponding to the scenario 6 in Table 4.1. Thus, they believe the longest schedule, $\boldsymbol{s}^{(6)}$, is optimal, a belief that leads to a misspecified prior for the first five scenarios. Similarly, we can use any other row of probabilities of toxicity to specify the priors for the $\boldsymbol{\beta}$ as long as the row we choose is the investigators' belief. The investigators also believe that one administration has a limiting cumulative probability $Q_0 = P_1/6$ (one-sixth of the shortest schedule), with corresponding cumulative probabilities of toxicity $Q_1 = Q_0/4$ and $Q_2 = Q_0/2$ at times $t_1 = 6$ days and $t_2 = 9$ days, respectively. From these elicited values, we used the methods described in Section prior elicitation to estimate the mean hyperparameter values $\mu_\alpha = 2.9$, $\mu_\gamma = -0.1$, $\mu_{\beta_0} = -2.9$ and $\mu_{\beta_1} = 0.36$.

4.5.1 Study Result & Conclusion

Tables 4.2, 4.3 and 4.4 display the simulation results for our current model. Table 4.2 displays the estimated parameter values and corresponding standard deviations under all scenarios. Tables 4.3 and 4.4 summarize the respective frequency at which each schedule is chosen as MTS by maximum likelihood and Bayesian approaches. Within each scenario in Table 4.3 and 4.4, each entry on the first row is the percentage of simulations in which a given schedule is identified as the MTS while the entry on the second row is the average percentage of subjects assigned to a given schedule during a trial.

We can see from Table 4.2 that the biases of parameter estimates by maximum likelihood are larger than those by Bayesian approach in most scenarios. Further-

Table 4.2. **Estimated parameter values of the non-mixture cure model. Each entry is the estimated parameter value (standard deviation).**

Scenario	Estimation Method	Estimated Value of α	γ	β_0	β_1
1	MLE	2.97 (1.55)	-0.18 (0.25)	-2.10 (1.59)	0.53 (0.43)
	Bayesian	3.06 (1.13)	-0.10 (0.13)	-2.00 (1.12)	0.54 (0.27)
2	MLE	3.23 (1.71)	-0.19 (0.38)	-2.85 (1.86)	0.52 (0.51)
	Bayesian	3.31 (1.18)	0.06 (0.21)	-2.50 (1.16)	0.44 (0.22)
3	MLE	3.47 (1.73)	0.15 (0.35)	-2.93 (2.15)	0.46 (0.46)
	Bayesian	3.28 (1.22)	-0.07 (0.19)	-2.53 (1.18)	0.37 (0.29)
4	MLE	3.14(1.88)	-0.19 (0.28)	-2.67 (2.05)	0.51 (0.57)
	Bayesian	3.22 (1.26)	0.07 (0.23)	-2.22 (1.19)	0.41 (0.25)
5	MLE	3.26 (1.81)	0.17 (0.37)	-3.00 (2.49)	0.51 (0.60)
	Bayesian	3.11 (1.23)	0.09 (0.23)	-2.57 (1.34)	0.43 (0.23)
6	MLE	3.41 (1.45)	-0.17 (0.38)	-3.22 (3.03)	0.48 (0.59)
	Bayesian	3.27 (1.15)	-0.11 (0.27)	-2.98 (1.27)	0.44 (0.34)
7	MLE	2.81 (2.18)	-0.15 (0.47)	-2.80 (2.10)	0.61 (0.62)
	Bayesian	3.13 (1.77)	0.13 (0.31)	-2.35 (1.28)	0.56 (0.38)
8	MLE	2.92 (2.72)	0.17 (0.49)	-3.65 (2.09)	0.81 (0.89)
	Bayesian	3.15 (1.43)	-0.07 (0.33)	-3.39 (1.60)	0.72 (0.65)

more, the standard errors by maximum likelihood are consistently greater than those by Bayesian approach among all scenarios. Those facts may be due to the small sample size in our simulation studies and informative priors used in our Bayesian procedures. Despite the above facts, the parameter estimates are very close to the true values provide confidence for us to interpret the results in Tables 4.3 and 4.4.

After a direct comparison of the two estimation methods (first line in Table 4.3 vs first line in Table 4.4 within each scenario), we find that Bayesian approach perform consistently better than maximum likelihood method when the priors chosen in the Bayesian procedures are informative at the beginning of a trial but do not dominate the data at later points in a study. However, both methods lead to similar conclusions as far as the final recommendation for the optimal schedule. Specifically, in scenarios 1 & 2, the Bayesian approach has a much higher percentage (over 15% more) of

Table 4.3. **Performance of the non-mixture cure model with 30 patients by *maximum likelihood method*. Each entry on the $1st$ row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on $2nd$ line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	72.2	25.6	1.8	0.4	0	0
	63.3	23.3	7.2	4.1	2.1	0
2	21.8	46.3	19.7	8.7	5.3	1.2
	28.4	36.2	17.2	9.2	6.4	2.6
3	9.7	23.9	42.9	13.4	6.2	4.8
	11.9	23.4	35.8	15.4	8.8	4.6
4	5.2	8.8	22.7	41.1	16.4	5.9
	6.8	14.1	19.1	37.2	14.3	8.4
5	2.6	7.4	10.4	19.9	40.2	19.5
	3.6	8.1	12.2	21.8	33.2	21.2
6	2.9	4.1	6.7	11.4	21.7	53.2
	3.3	5.6	7.4	18.6	25.6	39.5
7	13.7	42.1	39.1	4.7	0.4	0
	17.1	34.4	32.1	9.5	4.3	2.6
8	3.7	15.5	44.3	26.1	9.4	1.0
	3.5	23.6	36.1	25.1	8.7	3.0
9	16.1	22.2	37.4	17.3	5.0	2.0
	16.2	17.9	32.9	16.7	10.2	6.1

identifying the true optimal schedule as the MTS than the maximum likelihood. For example, Bayesian approach has about 91% and 69% of identifying schedule 1 and 2 as the MTS while the maximum likelihood has about 72% and 46% in scenario 1 and 2, respectively. We suspect this fact is due to the following reasons. First, we start a trial at the lowest schedule and only allow incremental escalation during a trial. Second, Bayesian estimation approach produces much smaller variations of parameter estimates than maximum likelihood estimation method. Therefore, it is easier for Bayesian method to distinguish schedule 1 or 2 as MTS than maximum likelihood in scenarios 1 and 2 (when the schedule 1 or 2 is the true optimal schedule).

Table 4.4. **Performance of the non-mixture cure model with 30 patients by *Bayesian approach*. Each entry on the $1st$ row is the percentage of simulations a schedule chosen as MTS, with the average percentage of patients assigned to a given schedule on $2nd$ line within each scenario.**

	Schedule (number of weeks)					
Scenario	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
1	91.3	8.7	0	0	0	0
	77.2	22.6	0.2	0.1	0	0
2	14.9	68.8	15.6	0.7	0.	0
	15.9	51.4	18.2	11.9	3.6	0
3	4.4	17.4	53.1	18.3	6.3	0.5
	10.8	22.3	38.7	15.9	9.2	3.1
4	1.2	6.4	29.4	44.1	15.8	3.1
	5.1	9.6	23.2	40.3	14.8	7.0
5	0	2.8	5.2	29.8	45.1	17.1
	3.4	6.8	11.8	22.1	37.7	18.2
6	0	2.8	4.8	6.9	23.6	61.9
	3.3	5.2	9.3	14.3	21.9	46.0
7	9.6	45.7	38.1	6.6	0	0
	16.6	37.5	29.2	9.6	5.1	2.1
8	2.4	12.6	45.8	28.6	8.4	2.1
	4.6	11.0	39.0	29.6	9.2	6.5
9	4.8	29.2	41.7	20.8	3.5	0
	9.8	21.3	35.7	21.9	8.2	3.1

Furthermore, in scenarios 3-6, the Bayesian approach has a moderately higher percentage (about 5% to 10% more) of identifying the true optimal schedule as the MTS than the maximum likelihood. In scenario 7, both methods produce very similar results because the true MTS is midway between schedules 2 and 3. However, in scenario 8, the Bayesian approach has a higher percentage of identifying schedule 3 or 4 as the MTS than the maximum likelihood, even though the percentage of identifying both schedule 3 and 4 as the MTS are very similar for both methods because the true optimal schedule is somewhere closer to schedule 3 than to schedule 4. These results are expected as our algorithm should work well in optimal sched-

ule finding even when no true treatment schedule exists among the given multiple schedules, but there only existis a schedule that is close to the true MTS or the true MTS is between two existing schedules.

We then compare the second line in Table 4.3 to second line in Table 4.4 within each scenario and find that the highest average percentage of subjects in each scenario in a simulation trial is assigned to the true optimal schedule by both estimation methods. More specifically, under scenario 1, on average, about 63% (the highest percentage) of subjects is assigned to schedule 1 in a simulation trial using the maximum likelihood while about 77% (the highest percentage) of subjects is assigned to schedule 1 using the Bayesian method. Furthermore, under scenarios 5 and 6, the Bayesian approach achieves the optimal schedule faster than the maximum likelihood because the up-down-scheme used in the maximum likelihood at the beginning of a simulation trial slows down the escalation process by grouping when the longer schedule (schedule 5 or 6) is the true optimal schedule. We also note that more subjects are assigned to the schedules shorter than true MTS than those schedules longer than the true MTS. This result indicates that the safety constraints described in Section 2.4 are implemented fairly well in our simulation studies.

To assess the impact of model misspecification on optimal treatment schedule finding, we first compare scenario 3 to scenario 9 in Table 4.3, then in Table 4.4. In general, we find that the ability of our algorithm to correctly identify the MTS is relatively unaffected by model misspecification using both estimation methods, although, fewer subjects are assigned to the true MTS when our assumed model does not reflect actual toxicity times. This result is expected, as schedule assignment during the study will be impacted by when each subject is having toxicity, while the final decision of the study is only impacted by the overall rate of toxicity.

Table 4.5. **Average number of observed toxicities (out of 30 subjects) in the simulated trials using the non-mixture cure model.**

Estimation Method	Scenario 1	2	3	4	5	6	7	8	9
MLE	8.85	8.05	6.58	11.22	10.91	9.06	9.93	9.15	6.41
Bayesian	7.94	7.12	6.31	12.09	11.15	9.92	10.11	8.45	5.98

An interesting note is about the average number of observed toxicities out of the total 30 patients, which is displayed in Table 4.5. Overall, the numbers of observed toxicities are very similar for both methods with the differences between the two methods are less than 1 across all scenarios. We also observe that on one hand, the numbers of observed toxicity are slightly larger in scenarios 1-3 by the maximum likelihood than those numbers by the Bayesian approach. On the other hand, the numbers of observed toxicity are smaller in scenarios 4-6 by the maximum likelihood compared to those numbers by the Bayesian approach. We suspect this fact may be due to the reason that more subjects are assigned to shorter schedules in scenarios 1-3 while more subjects are assigned to longer schedules in scenarios 4-6 using the Bayesian approach compared to using the maximum likelihood method.

Furthermore, we note the following pattern: the decreasing mean number of observed toxicities from scenarios 1-3 and from scenarios 4-6 but the average number of toxicity in scenarios 4-6 are larger than those in scenarios 1-3 using both methods. These facts are due to the simulation set up. We assume that the threshold of probability of toxicity is 0.2 for scenarios 1-3 while the threshold is 0.4 for scenarios 4-6. In addition, we find that the numbers in scenario 3 are very similar to those in scenario 9 by both estimation methods because the probability of toxicity set up in two scenarios are exactly same even though the data are simulated under different models.

Table 4.6. **Average number of subjects (out of 30 subjects) had reassignment in the simulated trials using the *non-mixture cure model*.**

Estimation Method	Scenario 1	2	3	4	5	6	7	8	9
MLE	0.36	1.76	4.36	5.47	6.95	7.24	3.08	4.69	4.57
Bayesian	0.12	1.63	3.37	5.61	7.09	7.58	2.46	4.31	3.54

Another interesting side note is in regard to the average number of subjects out of 30 subjects who received a reassignment to their originally assigned schedule, which is displayed in Table 4.6. The numbers are very similar using both methods with the Bayesian method having slightly larger numbers for scenarios 1-3 and having slightly smaller numbers for scenarios 4-6. Overall, the number of subjects with treatment reassignment increases as the true optimal treatment schedule becomes longer, which is expected. Furthermore, we note the numbers in scenario 3 are very close to those in scenario 9 by both approaches because the probability of toxicity set up in two scenarios are exactly same and both scenarios have schedule 3 as the true MTS although the data are simulated from different models. It is nonetheless interesting to note that the working model affects very weakly on the results.

We have developed a non-mixture cure model with additive hazard to identify an optimal schedule among a fixed number of nested treatment schedules using both maximum likelihood and Bayesian approaches. Subject accrual, data monitoring and outcome adaptive decision-making are done sequentially and continuously throughout Phase I trials. Via simulation, we have demonstrated the excellent operating characteristics of our algorithm when the assumed model is correct or misspecified, as well as when the prior is correctly or incorrectly specified.

CHAPTER V

Summary and Future Research

5.1 Summary

This dissertation extends the optimal treatment schedule finding method proposed in Braun et al. (2005) by using both maximum likelihood and Bayesian methods, by proposing more smooth parametric models. More specifically, we have developed three classes of additive hazard models to identify an optimal schedule among a fixed number of possible nested treatment schedules. The first class of the proposed parametric models is the pseudo non-mixture cure model. The triangular hazard model discussed in Chapter III is an example of this class of models. The second class of the proposed parametric cure models is the mixture cure model where the treatment schedule effect is modeled through the cure fraction. The sectional Weibull hazard model discussed in Chapter II is an application of this class of models. The third class of the proposed parametric models is the non-mixture cure model. The Weibull density function (as a hazard) model we discussed in Chapter IV is an illustration of this class of model. We demonstrated the applications of these parametric models through simulation studies and compared their performance by two different estimation methods.

There are similarities among the three classes of proposed models. We first com-

pare the first class of models to the third class. Returning to Equation (4.1), the survivor function of a single administration, we note that the triangular model of Braun et al. (2005) can be viewed as a pseudo non-mixture cure model. Both θ_1 and θ_3 in the triangular model share the roles of our parameters α and γ and relate most to when toxicity occurs, while θ_2 in the triangular model relates most to the cumulative probability of toxicity like our parameter θ in the non-mixture cure model which we then generalize to accommodate multiple administrations.

We then compare the second class of models to the third class. There is a distinct mathematical connection between a mixture cure model and a non-mixture model. We use the single administration setting to illustrate the mathematical connection. Suppose that p is the event rate and $S(.)$ is the survival function for subjects experiencing toxicity. Then from the population survival function (4.1) in the non-mixture cure model, we obtain the survival function for those with toxicity as

$$S(t) = P(T > t|T < \infty) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}$$

where $p = 1 - \exp(-\theta)$. Then, we can write $S_p(t) = 1 - p + pS(t)$. Thus $S_p(t)$ is a mixture cure model with event rate p and survival function $S(t)$ for those subjects experiencing toxicity. This shows that every non-mixture cure model defined as (4.1) can be written as a mixture cure model. This result also implies that every mixture cure model corresponds to some model of the form (4.1) for some θ and $F(.)$. If the covariate effects are modeled through θ, then the entire population is modeled as a proportional hazards (PH) model in a non-mixture cure model whereas only those with toxicity can be modeled with a PH structure in a mixture cure model.

Even though there is a mathematical relationship between the non-mixture cure model and the mixture cure model, the non-mixture cure model has some distinct advantages over the mixture cure model from application perspective. As we discuss

in the previous paragraph, the non-mixture cure model facilitates the PH structure naturally which is a frequently used property in most clinical trials. For example, the non-mixture cure model can directly incorporate subject-specific covariates into regression model for the cure fraction, while still allowing the pattern of toxicity to remain constant among subjects, i.e. proportional hazards. Even in applications that the PH assumption may be violated, the non-mixture cure model still provide a natural structure for testing the departure from the PH assumption. Furthermore, the non-mixture cure model allows for a broader family of parametric models, such as the toxicity time patterns beyond the up-and-down (non-monotonic) pattern we required because parametric models are more attractive in adaptive early-phase clinical trials. Most importantly, the non-mixture cure model offers certain technical advantages when developing maximum likelihood or Bayesian estimation procedures because of the usual properties of standard lifetime distributions, i.e. continuity, infinite support on positive real line, etc. Therefore, we prefer a non-mixture cure model with Weibull family of distributions in optimal treatment schedule finding.

Via simulation, we demonstrate both maximum likelihood and Bayesian approaches performed well under broad range of scenarios we investigated, including the case when the model was misspecified. But with the maximum likelihood method, the estimated parameters had a larger variability compared to the results by the Bayesian approach due to small sample size of early-phase clinical trials. With the Bayesian approach, we had to use informative priors in our application due to the same reason. As a result, we had to carefully choose a tuning parameter to allow the prior provide enough information at the beginning of an adaptive early-phase trial but not dominate the posterior at later point in the given trial.

However, even in the case when the priors were misspecified, the Bayesian ap-

proach still had a higher percentage of identifying the correct MTS than the maximum likelihood method. Therefore, we would prefer a Bayesian approach over a maximum likelihood method in adaptive early-phase clinical trials because a maximum likelihood method requires a two-stage design. Before any toxicity occurs, a traditional up-down scheme or a Bayesian approach is used to initiate the trial. Once a toxicity occurs and more than 15 subjects are enrolled in our simulated trials (the number of the subjects enrolled in a trial at switching point may be different for different application problems, the 15 is applicable to our application in the KGF trial), then the trial switches to a maximum likelihood method. With a Bayesian approach, no such two-stage design is necessary.

Furthermore, comparing the three classes of proposed models via simulation, Bayesian approaches slightly overestimated the probabilities of toxicity so that more subjects were assigned to the schedules shorter than the true MTS compare to the schedules longer than the true MTS in our simulation studies. Overall, a Bayesian approach performed fairly well for any assumed models in our simulation studies where the toxicity rates of the treatment schedules were moderate. We also conclude that what matters are the toxic probabilities of each schedule and not so much the models generating these probabilities.

5.2 Future Research

Via simulation, we have demonstrated the excellent operating characteristics of our algorithms no matter whether the assumed models are correct or misspecified. However, we have not been able to determine which proposed model is optimal from the perspective of statistical criteria for model selection. Therefore, we plan to further investigate statistical methods for model selection in multiple treatment

schedule setting. Since our simulations have demonstrated the Bayesian approaches constantly performed better than maximum likelihood methods in adaptive early phase clinical trials, we plan to use Bayesian model selection methods. Berger and Pericchi (1997) listed many advantages of Bayesian approaches to model selection; the advantage most relevant to our discussion is that Bayesian model selection does not require nested models (our three proposed models are not nested) nor standard distributions and regular asymptotic results (two of our proposed models have a change point in the hazard function).

Seltman et al. (2001) used Bayes factors and posterior model probabilities for selecting optimal models in the case of survival models with a cure fraction. Schwarz (1978) derived the Bayesian Information Criterion (BIC) as a large sample approximation to twice the logarithm of the Bayes factor. Another model selection procedure is based on a cross-validation predictive check (Stone, 1974; Geisser, 1993) and is called the conditional predictive ordinate (CPO) (Gelfand et al., 1992; Geisser, 1993 and Gelfand, 1995). One attractive feature of the CPO is that it does not require proper priors. As a result, it has been used extensively in the literature for model selection. Chen et al. (2002) used for non-mixture cure model selection. Dey et al. (1995) used for determining the number of components in a mixture distribution. Gelfand (1995) demonstrated good performance of various Bayesian model determination techniques in investigating nonlinear mixed-effects models on a small data set. However, little research exists on the small-sample performance of Bayesian model selection techniques in the area of cure models. Furthermore, Yu (2004) conducted a simulation study to compare the performance of the three Bayesian criteria: BIC, posterior model probabilities, and the CPO statistic, in model selection of cure models. Their research demonstrated that the CPO performed reasonably well in

most settings and much better than both Bayes factors and BIC.

Therefore, we choose CPO as our criterion for model selection. Specifically, we will calculate the CPO for each observation under the three proposed models and derive the B* statistic proposed by Gelfand et al. (1992) related to the CPO of each model. The B* statistic is more useful than the CPO when comparing more than two models simultaneously; the model with the larger B* statistic is judged the better fitting model.

Under our model, one could observe only patients who receive the shortest sequence, $\boldsymbol{s}^{(1)}$ to make predictions about any $\boldsymbol{s}^{(j)}, j \geq 1$. We are exploring the possibility of loosening strong homogeneity assumption on the hazard of toxicity at each administration. For example, we consider develop models that account for inhomogeneity across schedules. Back to the motivating example, suppose one course of treatment is composed of administrations at day 1, day 2 and day 3 then 4 days off, then another 3 days of administrations. If we assume the hazard of toxicity at day 1 follows a Weibull PDF, then we could model the hazard of toxicity at later administrations follow the similar Weibull PDF pattern with a different normalization factor depending on the dose response relationship, which might possibly be determined from the historical data. If no related historical data were available, we could also solicit the hazard of toxicity pattern from the investigators in order to determine our hazard function structure for later administrations.

Braun et al. (2005) have recently updated their algorithm to allow the dose to vary among administrations by modeling each of their triangular parameters to vary as a parametric function of dose (Braun et al. (2007)). We plan to make similar updates to our algorithm by including dose as a covariate in our model for the cure fraction. Such an approach would assume proportional hazards among doses, although we

could also allow the parameters α and γ to vary by dose if proportional hazards was not a reasonable assumption. We can also easily generalize our model to allow for hazards that change with each administration by modeling the hazard parameters as a function of administration number.

We could also extend our models to allow for optimal treatment schedule finding with combinations of two agents where both agents have a multiple treatment schedule. In this scenario, our outcome remains the time to toxicity; however, the non-mixture cure model would incorporate main effects of both agents into the cure fraction, as well as a term for any possible interaction between the agents. The more challenging aspect of this design is how to incorporate both agents into the time to toxicity hazard, as the two agents will likely differ in both the number of administrations, as well as the times of administration. Nonetheless, once we have a reasonable model, the Bayesian estimation procedures developed in this dissertation could be used in this design.

Most importantly, our model can be adopted for any clinical trial in which investigators wish to measure the impact of multiple administrations on a binary outcome. Thus, our algorithm could be used in a Phase II study seeking to determine how many administrations are necessary for a desired rate of efficacy, or in a Phase III study comparing two different schedules or doses of the same agent or two different agents in a large sample of (randomized) subjects. Furthermore, if our methods were applied to a large cohort of subjects like that in a Phase III trial, we could model the single-administration hazard non-parametrically with standard techniques rather than forcing a parametric model on the event times.

Even though our proposed models have been focused on Phase I trials evaluating the safety of multiple treatment schedules, we could also extend our methodology

to evaluate simultaneously the safety and efficacy of multiple administrations like that proposed by Thall and Cook (2004) for single administration setting. In this case, our responses would be bivariate outcomes with a safety outcome as time to events and an efficacy outcome as categorical or continuous. If the efficacy outcome is also time to events and a non-mixture cure model is a reasonable assumption in the application, then our non-mixture cure model could be extended to a bivariate non-mixture cure model with a shared frailty to account for the correlation between the failure times. New Bayesian estimation procedures need to be developed for the bivariate case.

APPENDICES

APPENDIX A

Proofs for Theorems in This Dissertation

A.1 Proofs for Theorems in Chapter III

Proof of Theorem 1: From the log likelihood function (3.4), we derive the first derivative about θ_1 as

$$\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_1}Q_1(\theta_1,\theta_3) - \theta_2\frac{\partial}{\partial\theta_1}Q_2(\theta_1,\theta_3). \tag{A.1}$$

First, if $\theta_1 \in (0, T_1]$, we can simplify $Q_1(\theta_1,\theta_3)$ and $Q_2(\theta_1,\theta_3)$ as

$$\begin{aligned} Q_1(\theta_1,\theta_3) &= \sum_{i=1}^{n}\log\left[\frac{\theta_3 - T_i}{\theta_3-\theta_1}\right] \\ Q_2(\theta_1,\theta_3) &= \frac{1}{2}\sum_{i=1}^{n}\left[\theta_3 - \frac{(\theta_3-T_i)^2}{\theta_3-\theta_1}\right]. \end{aligned}$$

The first order derivatives of $Q_1(\theta_1,\theta_3)$ and $Q_2(\theta_1,\theta_3)$ about θ_1 are

$$\frac{\partial}{\partial\theta_1}Q_1(\theta_1,\theta_3) = \frac{n}{\theta_3-\theta_1}$$

$$\frac{\partial}{\partial\theta_1}Q_2(\theta_1,\theta_3) = -\frac{\sum_{i=1}^{n}(\theta_3-T_i)^2}{2(\theta_3-\theta_1)^2}$$

We can see for $\theta_1 \in (0, T_1]$, $\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta}) > 0$, therefore $\ell_n(\theta_1)$ is strictly increasing over $(0, T_1]$. The maximum value of $\ell_n(\theta_1)$ is attained at T_1.

Second, if $\theta_1 \in [T_n, \theta_3)$, $Q_1(\theta_1, \theta_3)$ and $Q_2(\theta_1, \theta_3)$ can be simplified as

$$\begin{aligned} Q_1(\theta_1, \theta_3) &= \sum_{i=1}^{n} \log\left[\frac{T_i}{\theta_1}\right] \\ Q_2(\theta_1, \theta_3) &= \frac{1}{2}\sum_{i=1}^{n}\left[\frac{T_i^2}{\theta_1}\right]. \end{aligned}$$

The first order derivatives of $Q_1(\theta_1, \theta_3)$ and $Q_2(\theta_1, \theta_3)$ about θ_1 are

$$\frac{\partial}{\partial\theta_1} Q_1(\theta_1, \theta_3) = -\frac{n}{\theta_1}$$

$$\frac{\partial}{\partial\theta_1} Q_2(\theta_1, \theta_3) = -\frac{\sum_{i=1}^{n} T_i^2}{2{\theta_1}^2}$$

In order to have $\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta}) \leq 0$ in the interval $[T_n, \theta_3)$, We need to have $\theta_1 \geq \frac{\theta_2 \sum_{i=1}^{n} T_i^2}{2n}$. Since $T_i \leq T_n$, $\frac{\theta_2 \sum_{i=1}^{n} T_i^2}{2n} \leq \frac{\theta_2 T_n^2}{2}$. When $\theta_1 \in [T_n, \theta_3)$, $\theta_1 > T_n$. Then $\frac{\theta_2 T_n^2}{2} \leq \frac{\theta_2 {\theta_1}^2}{2}$. From the assumption $\theta_2 < \frac{1}{\theta_1}$, $\frac{\theta_2 {\theta_1}^2}{2} \leq \frac{\theta_1}{2} < \theta_1$. Therefore, $\ell_n(\theta_1)$ is decreasing. This will guarantee that the maximum value of $\ell_n(\theta_1)$ is attained at T_n when $\theta_1 \in [T_n, \theta_3)$.

Finally, if θ_1 in $[T_r, T_{r+1}]$ where $1 \leq r \leq (n-1)$, we can simplify $Q_1(\theta_1, \theta_3)$ and $Q_2(\theta_1, \theta_3)$ as

$$Q_1(\theta_1, \theta_3) = \sum_{i=1}^{r} \log\left[\frac{T_i}{\theta_1}\right] + \sum_{i=r+1}^{n} \log\left[\frac{\theta_3 - T_i}{\theta_3 - \theta_1}\right] \tag{A.2}$$

$$Q_2(\theta_1, \theta_3) = \frac{1}{2}\sum_{i=1}^{r}\left[\frac{T_i^2}{\theta_1}\right] + \frac{1}{2}\sum_{i=r+1}^{n}\left[\theta_3 - \frac{(\theta_3 - T_i)^2}{\theta_3 - \theta_1}\right]. \tag{A.3}$$

The first order derivatives of $Q_1(\theta_1, \theta_3)$ and $Q_2(\theta_1, \theta_3)$ about θ_1 are

$$\frac{\partial}{\partial\theta_1} Q_1(\theta_1, \theta_3) = -\frac{r}{\theta_1} + \frac{n-r}{\theta_3 - \theta_1} = \frac{-r\theta_3 + n\theta_1}{\theta_1(\theta_3 - \theta_1)} \tag{A.4}$$

$$\frac{\partial}{\partial\theta_1} Q_2(\theta_1, \theta_3) = -\frac{\sum_{i=1}^{r} T_i^2}{2{\theta_1}^2} - \frac{\sum_{i=r+1}^{n}(\theta_3 - T_i)^2}{2(\theta_3 - \theta_1)^2} \tag{A.5}$$

For $\theta_1 \in [T_r, T_{r+1}]$, if $-r\theta_3 + n\theta_1 \geq 0$, then $\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta}) > 0$, so $\ell_n(\theta_1)$ is strictly increasing over $[T_r, T_{r+1}]$. The maximum value of $\ell_n(\theta_1)$ is attained at T_{r+1}. If $-r\theta_3 + n\theta_1 < 0$,

then $\frac{\partial}{\partial\theta_1}\ell_n(\boldsymbol{\theta})$ may <= 0 then > 0. Furthermore, the second order derivatives of $Q_1(\theta_1,\theta_3)$ and $Q_2(\theta_1,\theta_3)$ about θ_1 are

$$\frac{\partial^2}{\partial\theta_1^2}Q_1(\theta_1,\theta_3) = \frac{r}{\theta_1^2}+\frac{n-r}{(\theta_3-\theta_1)^2} \tag{A.6}$$

$$\frac{\partial^2}{\partial\theta_1^2}Q_2(\theta_1,\theta_3) = \frac{\sum_{i=1}^{r}T_i^2}{{\theta_1}^3}-\frac{\sum_{i=r+1}^{n}(\theta_3-T_i)^2}{(\theta_3-\theta_1)^3} \tag{A.7}$$

Under the assumption $\theta_2 < \min(\frac{1}{\theta_1},\ \frac{1}{\theta_3-\theta_1})$, $\frac{\partial^2}{\partial\theta_1^2}\ell_n(\boldsymbol{\theta}) \geq 0$ for all $\theta_1 \in [T_r, T_{r+1}]$. So $\ell_n(\theta_1)$ is decreasing until reaching minimum, then increasing over $[T_r, T_{r+1}]$. Therefore, the maximum value of $\ell_n(\theta_1)$ is attained at T_r or T_{r+1} when $\theta_1 \in [T_r, T_{r+1}]$. The proof of Theorem 1 is complete.

Proof of Theorem 2: To prove item 1, we first substitute $\widehat{\theta_1} = T_{\widehat{r}(\theta_2,\theta_3)}$ in the likelihood function, then simplify the equation (3.5) and write it in term of $M(\theta)$ as following

$$\widehat{r}(\theta_2,\theta_3) = \arg\underset{r\in\{1,\ldots,n\}}{max}\{M(T_r,\theta_2,\theta_3)\} \tag{A.8}$$

Where

$$M(T_r,\theta_2,\theta_3) = \prod_{i=1}^{r}\frac{T_i}{T_r}\exp(-\frac{\theta_2 T_i^2}{2T_r})\prod_{i=r+1}^{n}\frac{\theta_3-T_i}{\theta_3-T_r}\exp(-\frac{\theta_2(\theta_3-T_i)^2}{2(\theta_3-T_r)}) \tag{A.9}$$

Under assumptions $\theta_2 < \min(\frac{1}{\theta_3-T_r},\ \frac{1}{T_r})$, then

$$\prod_{i=1}^{r}\frac{T_i}{T_r}\prod_{i=r+1}^{n}\frac{\theta_3-T_i}{\theta_3-T_r} > M(T_r,\theta_2,\theta_3) > \prod_{i=1}^{r}\frac{T_i}{T_r}\prod_{i=r+1}^{n}\frac{\theta_3-T_i}{\theta_3-T_r}\exp(-1)$$

As θ_3 approaches $+\infty$, $\frac{\theta_3-T_i}{\theta_3-T_r}$ approaches 1. So, for θ_3 large enough, $\prod_{i=1}^{r}\frac{T_i}{T_r}$ is dominant on both sides of inequality. We know $T_i < T_r$ for $i < r$. Then, the smaller the r is, the larger the $M(T_r,\theta_2,\theta_3)$ is. Thus, r decreases when θ_3 increases.

To prove item 2, let $\theta_3 \to +\infty$, assume θ_1 not approach T_1, then exists some $r \neq 1$ such that θ_1 approach T_r, but $\prod_{i=1}^{r}\frac{T_i}{T_{r+1}} < 1$ while $\prod_{i=2}^{n}\frac{(\theta_3-T_i)}{(\theta_3-T_1)}$ approach 1 as

$\theta_3 \to +\infty$. Therefore, for large enough θ_3, $L_n(T_r, \theta_3) < L_n(T_1, \theta_3)$, which contradicts that $T_r (r \neq 1)$ is the MLE. Thus, $\widehat{r}(\theta_2, \theta_3) = 1$.

To prove item 3, if $\theta_3 = T_n$, assume $\theta_1 \neq T_n$, then exists a term $\theta_3 - T_n = T_n - T_n = 0$ in the likelihood function $L_n(\theta_1, \theta_3)$ such that $L_n(\theta_1, \theta_3) = 0$. So θ_1 has to be T_n. Thus, $\widehat{r}(\theta_2, \theta_3) = n$.

To prove item 4, we apply the result from Theorem 1 that $\widehat{r}(\theta_2, \theta_3)$ is a function of θ_2 and θ_3 that can only take the discrete values in $[1, n]$. From item 1 of this Theorem, $\widehat{r}(\theta_2, \theta_3)$ is a decreasing function of θ_3. Therefore, $\widehat{r}(\theta_2, \theta_3)$ has $(n-1)$ discontinuity points $\{\theta_{3,r}, r = n-1, n-2, ..., 1\}$ when r values step down from n to $n-1$, $n-1$ to $n-2$, ..., from 2 to 1.

To derive $\theta_{3,r}$ as defined in (3.7), let $\widehat{r}(\theta_2, \theta_3) = r$, then $\widehat{\theta_1} = T_r$. So $M(T_r, \theta_2, \theta_3) \geq M(T_{r-1}, \theta_2, \theta_3)$, $M(T_r, \theta_2, \theta_3) \geq M(T_{r+1}, \theta_2, \theta_3)$ Thus, $\theta_{3,r}$ is the solution of equation $M(T_r, \theta_2, \theta_3) = M(T_{r+1}, \theta_2, \theta_3)$. The other endpoints in equation (3.7) can be derived in a similar fashion. The proof of Theorem 2 is complete.

Proof of Lemma 1: Let $S_{\theta_1} = (\theta_1 - \delta, \theta_1 + \delta)$ where δ stands for $\delta(\theta_1, \theta_{1(0)})$ as in condition 1. Let $L_k(\theta_1) = \sum_{i=1}^{k} \log f(t_i, \theta_1)$. Choose any η, define $\Omega = [0, \theta_3] \bigcap [\theta_1 : |\theta_1 - \theta_{1(0)}| \geq \eta]$. Since Ω is compact and $\bigcup \{S_{\theta_1} : \theta_1 \in \Omega\} \supset \Omega$, there exists a finite large number M and a finite set $\{\theta_{1,1}, ..., \theta_{1,M}\} \subset \Omega$ such that $\bigcup_{i=1}^{M} S_i \supset \Omega$.

Under condition 1, by law of large numbers, for any $\varepsilon > 0$ there exist integers $N(\theta_{1,i}, \varepsilon)$ such that for $i \in \{1, ..., M\}$ and $n \geq \underset{i}{max} N(\theta_{1,i}, \varepsilon)$,

$$P_{\theta_{1(0)}}[\bigcup_{k \geq n} \{\sum_{j=1}^{k} \{\log f(t_j, \theta_1) - \log f(t_j, \theta_{1(0)})\} < 0\}] > 1 - \varepsilon/M. \tag{A.10}$$

Then

$$
\begin{aligned}
P_{\theta_{1(0)}}[\bigcup_{k\geq n}\{|\theta_{1,k}-\theta_{1(0)}|\geq\eta\}] & \\
&\leq P_{\theta_{1(0)}}[\bigcup_{k\geq n}\{\sup_{\theta_1\in\Omega}L_k(\theta_1)>L_k(\theta_{1(0)})\}] \\
&\leq \sum_{i=1}^{M}P_{\theta_{1(0)}}[\bigcup_{k\geq n}\{\sup_{\theta_1\in S_i}L_k(\theta_1)>L_k(\theta_{1(0)})\}] \\
&\leq M*\varepsilon/M=\varepsilon.
\end{aligned}
$$

This proves Lemma 1.

Proof of Lemma 2: Let $U_i = X_i - \theta_1, i = 1, ..., n$. Then U_is are i.i.d with p.d.f $f_u(u)$ and $f_u(0) = \alpha > 0$. We then apply the result of Lemma 7 in Yao (1986) and derive the result of Lemma 2.

***Proof of Theorem 4*:** To prove the consistency of the MLEs, we will show CDF $F(t|\boldsymbol{\theta})$ and pdf $f(t|\boldsymbol{\theta})$ satisfy required conditions. First, it can be proved that i) $f(t|\boldsymbol{\theta})$ is a uniformly continuous density which vanishes on $(-\infty, 0)$ and $(\theta_3, +\infty)$ and is positive on $[0, \theta_3]$, ii) $\int_{-\infty}^{\infty}|\log f(t,\boldsymbol{\theta})|f(t)dt \ < \ +\infty$. Second, following the results of Wald (1949), we have $E\log f(T,\boldsymbol{\theta}) < E\log f(T,\boldsymbol{\theta}_0)$ for any $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is the true parameter values. Therefore, the condition 1 holds. Third, let Ω_c denote the constrained parameter space, then Ω_c is a closed subset of the 3-dimensional parameter space. Fourth, through detailed calculation, we can show

$$\mathbf{E}\{(\frac{\partial}{\partial\theta_i})\log f(T|\boldsymbol{\theta})\} \ = \ 0 \ for \ i=1,2,3.$$

From the results of Lemma 1 and Wald (1949), the MLEs $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{1n}, \widehat{\theta}_{2n}, \widehat{\theta}_{3n})$ exist and are consistent.

To derive the limiting distributions of the MLEs, we need to apply the results

from the special cases. When $\theta_3 = \theta_{3(0)}$ is known, utilizing the transformation

$$V = \begin{cases} T & ; \ T \leq 0 \ or \ T > \theta_3 \\ T^2/\theta_1 & ; \ 0 < T \leq \theta_1 \\ \theta_3 - (\theta_3 - T)^2/(\theta_3 - \theta_1) & ; \ \theta_1 < T \leq \theta_3, \end{cases} \tag{A.11}$$

we create a random variable V resembling an exponential distribution as defined in equation (2) of Yao (1986). Then, applying the results from Proposition 6 of Yao (1986) and Lemma 2, we have the following result, $\{n(\widehat{\theta}_{1:3} - \theta_{1(0)}), n^{\frac{1}{2}}(\widehat{\theta}_{2:3} - \theta_{2(0)})\}$ converges in distribution to a random vector (Z_1, Z_2) where Z_1 and Z_2 are independent, Z_1 is defined in Definition 3, Z_2 is a normal random variable with mean 0 and variance m_{22}^{-1}. We already know $\widehat{\theta_1}$ is a consistent estimator of θ_1, applying Lemma 5 of Chernoff and Rubin (1956), we derive the following result

$$\widehat{\theta}_{1(n)} - \widehat{\theta}_{1:3} = o(n^{-1}),$$

$\widehat{\theta_2}$ follows the classical MLE result:

$$\widehat{\theta}_{2(n)} - \widehat{\theta}_{2:3} = o(n^{-\frac{1}{2}}).$$

We can rewrite

$$n(\widehat{\theta}_{1(n)} - \theta_{1(0)}) = n(\widehat{\theta}_{1(n)} - \widehat{\theta}_{1:3}) + n(\widehat{\theta}_{1:3} - \theta_{1(0)}),$$
$$n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)}) = n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \widehat{\theta}_{2:3}) + n^{\frac{1}{2}}(\widehat{\theta}_{2:3} - \theta_{2(0)}).$$

Then, we prove the following result,

$$\widehat{\theta}_{1(n)} - \theta_{1(0)} \quad \leq_p \quad n^{-1}, \tag{A.12}$$

$$\widehat{\theta}_{2(n)} - \theta_{2(0)} \quad \leq_p \quad n^{-\frac{1}{2}}. \tag{A.13}$$

We also show the limiting distribution of $\{n(\widehat{\theta}_{1(n)} - \theta_{1(0)}), n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)})\}$ is the same as $\{n(\widehat{\theta}_{1:3} - \theta_{1(0)}), n^{\frac{1}{2}}(\widehat{\theta}_{2:3} - \theta_{2(0)})\}$. Therefore, $\{n(\widehat{\theta}_{1(n)} - \theta_{1(0)}), n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)})\}$

converges in distribution to a random vector (Z_1, Z_2) where Z_1 and Z_2 are independent, Z_1 is defined in Definition 3, Z_2 is normal random variables with mean 0 and variance m_{22}^{-1}.

When $\theta_1 = \theta_{1(0)}$ is known, we show that the PDF $f(t|\boldsymbol{\theta})$ resembles a truncated reverse Weibull distribution because $f(t|\boldsymbol{\theta}) \neq 0$ only when $0 < t \leq \theta_3$. We say a random variable T has a reverse Weibull distribution if $T = -X + 2\mu$ where X is a random variable following a three-parameter Weibull distribution with CDF $F_x(x) = 1 - \exp[-(\frac{x-\mu}{\alpha})^{\gamma}]$. In our model, the PDF $f(t, \boldsymbol{\theta})$ resembles a reverse Weibull distribution as t approaches the upper boundary θ_3.

Applying the results from Theorem 1 and Theorem 3 of Smith (1985), we have the following results : $\{n^{\frac{1}{2}}(\widehat{\theta}_{2:1} - \theta_{2(0)}), (n \log n)^{\frac{1}{2}}(\widehat{\theta}_{3:1} - \theta_{3(0)})\}$ converges in distribution to a random vector (Z_2, Z_3) where Z_2 and Z_3 are independent, Z_2 and Z_3 are normal random variables with common mean 0 and respective variances m_{22}^{-1} and m_{33}^{-1}. Furthermore,

$$\begin{aligned} \widehat{\theta}_{2(n)} - \widehat{\theta}_{2:1} &\leq_p (n \log n)^{-\frac{1}{2}}, \\ \widehat{\theta}_{3(n)} - \widehat{\theta}_{3:1} &\leq_p n^{-\frac{1}{2}} (\log n)^{-1} \end{aligned}$$

We can rewrite

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)}) &= n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \widehat{\theta}_{2:1}) + n^{\frac{1}{2}}(\widehat{\theta}_{2:1} - \theta_{2(0)}), \\ (n \log n)^{\frac{1}{2}}(\widehat{\theta}_{3(n)} - \theta_{3(0)}) &= (n \log n)^{\frac{1}{2}}(\widehat{\theta}_{3(n)} - \widehat{\theta}_{3:1}) + (n \log n)^{\frac{1}{2}}(\widehat{\theta}_{3:1} - \theta_{3(0)}). \end{aligned}$$

Then, we prove the following result,

$$\begin{aligned} \widehat{\theta}_{2(n)} - \theta_{2(0)} &\leq_p n^{-\frac{1}{2}}, \\ \widehat{\theta}_{3(n)} - \theta_{3(0)} &\leq_p (n \log n)^{-\frac{1}{2}}. \end{aligned}$$

We also show the limiting distribution of $\{n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)}), (n\log n)^{\frac{1}{2}}(\widehat{\theta}_{3(n)} - \theta_{3(0)})\}$ is the same as $\{n^{\frac{1}{2}}(\widehat{\theta}_{2:1} - \theta_{2(0)}), (n\log n)^{\frac{1}{2}}(\widehat{\theta}_{3:1} - \theta_{3(0)})\}$. Therefore,$\{n^{\frac{1}{2}}(\widehat{\theta}_{2(n)} - \theta_{2(0)}), (n\log n)^{\frac{1}{2}}(\widehat{\theta}_{3(n)} - \theta_{3(0)})\}$ converges in distribution to a random vector (Z_2, Z_3) where Z_2 and Z_3 are independent, Z_2 and Z_3 are normal random variables with common mean 0 and respective variances m_{22}^{-1} and m_{33}^{-1}.

It remains to show that $n(\widehat{\theta}_{1(n)} - \theta_{1(0)})$ and $(n\log n)^{\frac{1}{2}}(\widehat{\theta}_{3(n)} - \theta_{3(0)})$ are asymptotically independent. Since results in (3.18) and (3.20), it suffices to show $n(\widehat{\theta}_{1:3} - \theta_{1(0)})$ and $(n\log n)^{\frac{1}{2}}(\widehat{\theta}_{3:1} - \theta_{3(0)})$ are asymptotically independent. Since Z_1 follows a distribution of two independent random walks with each random walk as a sum of iid exponential random variables, Z_1 has a distribution without a Gaussian component. Z_3 is normal, following the result from Remark 1 of Smith (1985), Z_1 and Z_3 are independent. The proof of Theorem 4 is complete.

***Proof of Corollary 2*:** It is similar to the proof of Theorem 2 with the following modifications.

To prove item 2, let $\theta_3 \to +\infty$, assume θ_1 not approaching T_1, then there exists some $r > 1$ such that θ_1 approach T_r. We can deduce contradiction from following two cases. Either $\prod_{i=1}^{r} \frac{{t_i}^{c_i}}{{t_{r+1}}^{c_i}} < 1$ while $\prod_{i=2}^{k} \frac{(\theta_3 - t_i)^{c_i}}{(\theta_3 - t_1)^{c_i}}$ approach 1. Or $(\prod_{i=1}^{r} \exp(-\frac{\theta_2 T_i^2}{2T_r}))^{1-C_i} < \exp(-\frac{\theta_2}{2})^{1-C_i}$. In either case, for large enough θ_3, $L_k(T_r, \theta_3) < L_k(T_1, \theta_3)$, which contradicts that $T_r (r > 1)$ is the MLE of θ_1. Thus, $\widehat{r}(\theta_2,\ \theta_3) = 1$.

To prove item 3, if $\theta_3 = T_k$, assume $\theta_1 \neq T_k$, then there exists a term $\theta_3 - T_k = T_k - T_k = 0$ in either $(K - 1 + S_0)^{1-C_i}$ part of the likelihood function $L_k(\theta_1, \theta_3)$ or $(\theta_3 - T_k)^{C_i}$ part of likelihood function such that $L_k(\boldsymbol{\theta}) = 0$, therefore, $\theta_1 = T_k$. Thus, $\widehat{r}(\theta_2,\ \theta_3) = k$. This completes the proof of Corollary 2.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahn, C. (1998). An evaluation of Phase I cancer clinical trial. *Statistics in Medicine*, 17:1537–1549.

Babb, J., Rogatko, A., and Zacks, S. (1998). Cancer Phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine*, 17:1103–1120.

Berger, J. and Pericchi, L. (1997). Objective bayesian methods for model selection: introduction and comparison. Technical report, Department of statistics, Duke University.

Berkson, J. and Gage, R. (1952). Survival curve for cancer patients following treatment. *J. Amer. Statist. Assoc.*, 47:501–515.

Boag, J. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Royal Statist. Soc.*, 11B:15–44.

Braun, T. M., Levine, J. E., and Ferrara, J. (2003). Determining a maximum cumulative dose: Dose reassignment within the tite-crm. *Controlled Clinical Trials*, 24:669–681.

Braun, T. M., Thall, P. F., Nguyen, H., and de Lima, M. (2007). Simultaneously optimizing dose and schedule of a new cytotoxic agent. *Clinical Trials*, 4:113–124.

Braun, T. M., Thall, P. F., and Yuan, Z. (2005). Determining a maximum tolerated schedule of a cytotoxic agent. *Biometrics*, 61:335–343.

Cam, L. L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Annuals of Mathematical Statistics*, 41:802–28.

Chen, M. H., Ibrahim, J. G., and Lipsitz, S. (2002). Bayesian methods for missing covariates in cure rate models. *Lifetime Data Analysis*, 8:117–146.

Chen, M. H., Ibrahim, J. G., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of American Statistical Association*, 94:909–919.

Chernoff, H. and Rubin, H. (1956). The estimation of the location of a discontinuity in density. *Proceedings of 3rd Berkeley Symposium on Mathetical Statistics and Probability*, 1:19–37.

Cheung, Y. and Chappell, R. (2000). Sequential designs for Phase I clinical trials with late-onset toxicities. *Biometrics*, 56:1177–1182.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society B*, 39:1–38.

Dey, D., Kuo, L., and Sahu, S. (1995). A bayesian predictive approach to determining the number of components in a mixture distribution. *Statistics and Computing*, 5:297–305.

Dumbgen, L. (1994). Confidence sets for a change-point via randomization methods. *Change-Point Problems*, 23:108–129.

Eubank, R. L. and Speckman, P. L. (1994). Nonparametric estimation of functions with jump discountinuities. *Change-Point Problems*, 23:130–144.

Farewell, V. T. (1977a). The combined effect of breast cancer risk factors. *Cancer*, 40:931–936.

Farewell, V. T. (1977b). A model for a binary variable with time censored observations. *Biometrika*, 64:43–46.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *Canad. Jour. Statist.*, 14:257–262.

Garrett-Mayer, E. (2006). The continual reassessment method for dose-finding studies: a tutorial. *Clinical Trials*, 3:57–71.

Gasparini, M. and Eisele, J. (2000). A curve-free method for Phase I clinical trials. *Biometrics*, 56:609–615.

Geisser, S. (1993). *Predictive Inference: An Introduction*. London: Chapman & Hall.

Gelfand, A. (1995). Model determination using sampling-based methods. In Gilds, W., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, pages 145–161. London: Chapman & Hall.

Gelfand, A., Dey, D., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 4*, pages 147–167. Oxford: Oxford University Press.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis, 2nd edition*. Chapman and Hall/CRC.

Ghitany, M., Maller, R., and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Jour. Mult. Analysis*, 49:218–241.

Hall, P. and Wang, J. Z. (1999). Estimating the end-point of a probability distribution using minimum-distance methods. *Bernoulli*, 5:177–189.

Hinkley, D. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57:1–17.

Hinkley, D. (1972). Time-ordered classification. *Biometrika*, 59:509–523.

Jandhyala, V. K. and Fotopoulos, S. (1999). Capturing the distributional behaviour of the maximum likelihood estimator of a change-point. *Biometrika*, 86:129–140.

Korn, E. L., Midthune, D., and Chen, T. (1994). A comparison of two Phase I trial designs. *Statistics in Medicine*, 13:1799–1806.

Kuk, A. and Chen, C. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79:531–541.

Larson, M. B. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risk data. *Applied Statistics*, 34:201–211.

Li, C. and Taylor, J. M. G. (2002). A semi-parametric accelerated failure time cure model. *Statistics in Medicine*, 21:3235–3247.

Li, C., Taylor, J. M. G., and Sy, J. (2001). Identifiability of cure models. *Statistics and Probability Letters*, 54:389–395.

Maller, R. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, 79:731–739.

Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Chichester: John Wiley and Sons.

Nguyen, H. T., Rogers, G. S., and Walker, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika*, 71:299–304.

O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: A practial design for Phase I clinical trials in cancer. *Biometrics*, 46:33–48.

O'Quigley, J. and Shen, L. (1996). Continual reassessment method: A practial design for Phase I clinical trials in cancer. *Biometrics*, 52:673–684.

Peng, Y. and Dear, K. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56:237–243.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

Rosenberger, W. and Haines, L. (2002). Competing designs for phase i clinical trials: a review. *Statistics in Medicine*, 21:2757–2770.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464.

Seltman, H., Greenhouse, J., and Wasserman, L. (2001). Bayesian model selection: analysis of a survival model with a surviving fraction. *Statistics in Medicine*, 20:1681–1691.

Shao, Q. and Zhou, X. (2004). A new parametric model for survival data with long-term survivors. *Statistics in Medicine*, 23:3525–3543.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–90.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of Royal Statistical Society, Series B*, 36:111–147.

Storer, B. E. (1989). Design and analysis of Phase I clinical trials. *Biometrics*, 45:925–937.

Stylianou, M. and Flournoy, N. (2002). Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics*, 58:171–177.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56:227–236.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51:899–907.

Thall, P. F. and Cook, J. D. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, 60:684–693.

Tsodikov, A. (1998). A proportional hazard model taking account of long-term survivors. *Biometrics*, 54:1508–1516.

Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of American Statistical Association*, 98:1063–1078.

Vu, H., Maller, R., and Zhou, X. (1998). Asysmptotic properties of a class of mixture models for failure data: the interior and boundary cases. *Ann. Inst. Statist. Math*, 50:627–653.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annual of Mathematical Statistics*, 20:595–601.

Weiss, L. and Wolfowitz, J. (1973). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Annuals of Statistics*, 1:944–7.

Woodroofe, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Annuals of Mathematical Statistics*, 43:113–22.

Woodroofe, M. (1974). Maximum likelihood estimation of a translation parameter of a truncated distribution ii. *Annuals of Statistics*, 2:474–88.

Wu, C. (1983). On the convergence properties of the EM algorithm. *Annual Statistics*, 11:95–103.

Yakovlev, A. and Tsodikov, A. (1996). *Stochastic models of tumor latency and their biostatistical applications*. River Edge, New Jersey: World Scientific.

Yao, Y. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communication in Statistics - theory and method*, 15:2455–2466.

Yao, Y. C. and Huang, D. (1994). On almost sure behavior of change-point estimators. *Change-Point Problems*, 23:359–372.

Yu, M. (2004). *Joint Longitudinal-Survival-Cure Model with Application to Prostate Cancer Studies*. PhD thesis, Department of Biostatistics, University of Michigan.