Spring 5-17-2010

Statistical Modeling of Data From Smoking Cessation Clinical Trials

Yimei Li
University of Pennsylvania, yimeili@mail.med.upenn.edu

Statistical Modeling of Data From Smoking Cessation Clinical Trials

Abstract
In smoking cessation clinical trials, subjects commonly experience a series of lapse and recovery episodes of varying lengths. Any quit episode may become permanent, in the sense that the subject stops smoking for good, and any lapse may also become permanent, in the sense that the subject abandons the quit attempt entirely. Individual quit patterns may reflect the effects of treatment and measured and unmeasured covariates.

To describe this complex data structure, we propose a multivariate time-to-event model that i) incorporates alternating recurrent events of two types, each with the possibility of "cure", ii) allows for the modifying effects of treatment and covariates, and iii) reflects within-subject correlation via frailties. Specifically, we introduce a novel cure-mixture frailty model in which the cure probability follows a binary regression and the time to event given not cured is determined by a proportional hazard model. We then extend it to data with recurring events of two alternating types, where we assume that each type of event has a gamma frailty, and we link the frailties by means of a Clayton copula. In my first project, I fit this model to data from a smoking cessation drug trial. In my second project, I developed a Bayesian method to predict individual long-term smoking behavior from observed short-term quit/relapse patterns. In my third project, I investigated the theoretical properties of the survival distribution, evidently not previously described, that arises from our cure-mixture frailty model.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Epidemiology & Biostatistics

First Advisor
Daniel F Heitjan

Keywords
frailty model, multiple events, recurrent events, copula, prediction, Bayesian inference

Subject Categories
Biostatistics

STATISTICAL MODELING OF DATA FROM SMOKING CESSATION CLINICAL TRIALS

Yimei Li

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2010

Supervisor of Dissertation

*Signature*_____

Daniel F. Heitjan

Professor of Biostatistics

Graduate Group Chairperson

*Signature*_____

Daniel F. Heitjan, Professor of Biostatistics

Dissertation Committee

Mary E. Putt, Associate Professor of Biostatistics

E. Paul Wileyto, Assistant Professor of Psychiatry

Robert Schnoll, Assistant Professor of Psychiatry

Acknowledgments

The bulk of my thanks goes to my advisor, Professor Daniel F. Heitjan. He provided me many insightful comments and advice, guided me through all the challenges and difficulties, and always encouraged and supported me during the entire process of completing my doctoral work. What's most important is that he taught me how to be an excellent statistician and this experience will definitely benefit me for my lifetime. In addition, I thank the other members of my committee, Professor E. Paul Wileyto, Marry E. Putt and Robert Schnoll, for their help to improve the quality of this work.

I would like to express my gratitude to Prof. Justine Shults for her advising my master project, Prof. Jonas Ellenberg for his career suggestions, Prof. Joshua P. Metlay for the opportunity of working in his research team, and other faculty members at Penn who have offered me tremendous help throughout my Ph.D. years.

Finally, I want to dedicate this dissertation to my parents and my husband. Without their love and support, this thesis would never have come into existence.

ABSTRACT

STATISTICAL MODELING OF DATA FROM SMOKING CESSATION CLINICAL TRIALS

Yimei Li

Daniel F. Heitjan

In smoking cessation clinical trials, subjects commonly experience a series of lapse and recovery episodes of varying lengths. Any quit episode may become permanent, in the sense that the subject stops smoking for good, and any lapse may also become permanent, in the sense that the subject abandons the quit attempt entirely. Individual quit patterns may reflect the effects of treatment and measured and unmeasured covariates.

To describe this complex data structure, we propose a multivariate time-to-event model that i) incorporates alternating recurrent events of two types, each with the possibility of "cure", ii) allows for the modifying effects of treatment and covariates, and iii) reflects within-subject correlation via frailties. Specifically, we introduce a novel cure-mixture frailty model in which the cure probability follows a binary regression and the time to event given not cured is determined by a proportional hazard model. We then extend it to data with recurring events of two alternating types, where we assume that each type of event has a gamma frailty, and we link the frailties by means of a Clayton copula. In my first project, I fit this model to data from a smoking cessation drug trial. In my second project, I developed a Bayesian method to predict individual long-term smoking behavior from observed short-term quit/relapse patterns. In my third project, I investigated the theoretical properties

of the survival distribution, evidently not previously described, that arises from our cure-mixture frailty model.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In a typical smoking cessation trial, each patient designates a target quit date (TQD). Counseling and/or drug treatment begin prior to the TQD, and treatment may run for eight or more weeks post-TQD, with follow-up data gathered at intervals during treatment and at the end of treatment (EOT). The traditional primary outcome variable is a measure of abstinence at EOT, which we code as a binary variable and analyze using well-established methods such as logistic regression. Alternatively, some trials use the length of the period of abstinence prior to the initial failure, which we would then analyze using univariate time-to-event models.

However, these approaches are potentially inefficient in that they ignore the wealth of behavioral information in the raw daily smoking records. Commonly, a quit attempt consists of a series of setbacks and recoveries, potentially culminating in success. This unique feature poses challenging methodological problems on statistical modeling. For instance, the multiple transitions between smoking and abstinence induce two critical endpoints, time to lapse and time to recovery, and one need to model such

pair of recurrent events jointly. Moreover, the possibility of permanence of an episode needs to be incorporated into the analysis, which the traditional time-to-event model cannot handle.

Therefore, the first aim of this dissertation is to develop statistical methods to describe the complex data structure arising from smoking cessation clinical trials. In Chapter 2, we propose a flexible parametric model to describe alternating-states recurrent-event data where there is a possibility of cure with each type of event. We begin by introducing a novel cure-mixture frailty model in which a common frailty influences both the cure probability and the hazard function given not cured. We then extend our model to data with recurring events of two alternating types. We assume that each type of event has a gamma frailty, and we link the frailties by a Clayton copula. We illustrate the model with an analysis of data from two smoking cessation trials comparing bupropion and placebo. We also conduct simulation studies to demonstrate that our method performs well in repeated samples of practical size.

Another interesting characteristic about smoking cessation trials is that although the outcome at the end of this period is an important indicator of treatment success, substantial uncertainty remains on how an individual's smoking behavior will evolve over time. It is of great interest to predict subjects' further outcomes, because this will help researchers to optimize the intervention to meet the smoking cessation needs of different individuals. Therefore as the second aim of this dissertation, we develop a method to predict individual long-term smoking cessation success based on his/her short-term clinical observations. In Chapter 3, we describe such a Bayesian method for prediction, based on our cure-mixed frailty model proposed in Chapter 2. More

specifically, it's a two-stage prediction algorithm that first uses importance sampling to generate subject-specific frailties from their posterior distributions conditional on the observed data, then samples predicted future smoking behavior trajectories from the estimated model parameters and sampled frailties. We apply the method to the same smoking cessation data as in Chapter 2. The results suggest that the predictions from our method are better than those from a variety of empirical prediction methods.

In Chapter 4, we divert from smoking cessation trials to some theoretical derivations about a new survival distribution we discovered during the work in Chapter 2. We denote it as the Complementary Mixture Pareto II distribution (CMPII) and describe its properties in detail. We demonstrate its connections with existing distributions including the Pareto and exponential. We also derive its expectation, variance, characteristic function and moments.

Chapter 2

A Multivariate Cure-Mixture Frailty Model for Smoking Cessation Data

2.1 Introduction

Subjects in smoking cessation trials commonly transit several times between lapse and recovery. An important feature of such data is that any quit episode may become permanent, in the sense that the subject stops smoking for good, and any lapse may also become permanent, in the sense that the subject abandons the quit attempt entirely. The notion that subjects exhibit both temporary and permanent quits is well established in smoking cessation research (USDHHS (1990)) and has found support in recent statistical analyses of smoking cessation data sets (Cook et al. (2002); Mannan and Koval (2003); Banerjee and Carlin (2004); Yu and Peng (2008); Luo et

al. (2008)). Another important aspect of such data is that an individual's quit pattern may reflect underlying factors that are not readily encoded in covariates. This manifests itself as correlation between repeated outcomes, such as series of quit and lapse durations within individuals. Thus a comprehensive model for a smoking cessation data set would also need to include the possibility of correlation between events within individuals. We consider these elements — recurrent events, cure modeling, and modeling of correlation — in turn below.

Recurrent events. Recurrent-event data arise when subjects can have repeated episodes of the event in question; common examples in medicine include coronary infarctions, relapses of cancer, and loss of viral control in HIV infection. Most analyses of multiple failure times have used extended versions of the Cox model, although the same extensions apply to parametric models. Unlike their single-event counterparts, there are many structural choices to make in describing the time origin and at-risk status for an individual episode (Hosmer and Lemeshow (1999); Kelly and Lim (2000)). In smoking cessation studies, the appropriate modeling strategy is one in which the subject returns to the risk set only at the end of the preceding event (Andersen et al. (1993); Therneau and Hamilton (1997)).

Cure models. In survival analysis we typically assume that all subjects are genuinely at risk and will eventually experience an event. In some applications, however, a fraction may either be or become nonsusceptible. An example is time to recurrence of cancer among patients treated with surgery, where we may find that some patients never recur, presumably because the operation removed every vestige of the tumor (Withers et al. (1995)). *Cure models* were devised to describe such data. The perti-

nent literature begins with Farewell (1983), who used logistic regression to model the cured fraction and Weibull regression to model the non-cured event rate. Kuk and Chen (1992) proposed a Cox model for non-cured survival, and Peng et al. (1998) modeled the survival component with a generalized F. Attempts to model such data semiparametrically encounter vexing identifiability problems, because nonparametric survival estimates need not decline to zero (Peng and Carriere (2002)). Sy and Taylor (2000) therefore proposed a proportional hazards model with a zero-tail constraint, and Li and Taylor (2002) described non-cured survival with a semiparametric accelerated failure time model, which handles identifiability more gracefully.

Modeling heterogeneity/correlation. A popular device for extending the classical survival model to account for between-unit heterogeneity (within-unit correlation) is the *frailty model*, first proposed by Vaupel et al. (1979). A typical frailty model assumes that all subjects' survival distributions follow a common form up to a random subject-specific effect, known as a frailty. One then integrates out this latent factor to derive a new marginal distribution for the data. The introduction of frailties can describe lack of fit to classical survival models, between-subject heterogeneity, and within-subject correlation in clustered, multivariate or recurrent outcomes. Practitioners commonly assume that frailties follow a distribution from the power variance function family, such as the gamma, inverse Gaussian, or positive stable law (Hougaard (1986, 1994)).

Several authors have proposed extending the basic frailty model to incorporate the possibility of cure. Aalen (1988) included distributions with a nonsusceptible subgroup and developed these models further using a class of compound Poisson

distributions (Aalen (1992)). Longini and Halloran (1996) proposed frailty mixture models, and Price and Manatunga (2001) elaborated the idea of applying the frailty concept to cure models. Chen et al. (1999, 2002) andYin and Ibrahim (2005) described models for cancer recurrence in which a latent Poisson variable gives the number of surviving malignant clones in a treated subject, with the subject's overall time to recurrence being the minimum of the times to clinical observability generated by the clones; a subject with zero clones is cured.

In this chapter we propose a model to describe data on series of lapse and recovery episodes from subjects participating in a smoking cessation trial. In §2.2, we introduce a frailty model with cure fraction, its extensions, and estimation methods. In §2.3, we apply the proposed model to a smoking cessation data set. In §2.4 we present simulation results, and we offer concluding comments in §2.5.

2.2 Statistical Models

2.2.1 A Cure-Mixture Frailty Model

First consider univariate time-to-event data. We let T_i and $S_i(t_i)$ denote a single event time with its survival function for subject i, and assume that the subject is either cured (with probability π_i) or has a proper survival function $S_i^*(t_i)$ (with probability $1-\pi_i$):

$$S_i(t_i) = \pi_i + (1-\pi_i)S_i^*(t_i). \tag{2.2.1}$$

To account for heterogeneity, we assume that each subject has an unknown, latent frailty, denoted b_i, that affects both the cure fraction and the survival hazard. We further assume that, conditional on b_i, the cure probability follows a binary regression with a complementary log-log (cloglog) link, and time to event given not cured is determined by a proportional hazard model with a constant baseline hazard. Thus,

$$\pi_i(b_i) = \exp[-b_i \exp(-\eta_{\pi_i})], \quad S_i^*(t_i|b_i) = \exp[-t_i b_i \exp(\eta_{\sigma_i})], \tag{2.2.2}$$

where η_{π_i} and η_{σ_i} are linear predictors for the cure fraction and event rate, respectively. We represent the linear predictors as

$$\eta_{\pi_i} = \alpha_\pi + \beta_\pi V_i, \quad \eta_{\sigma_i} = \alpha_\sigma + \beta_\sigma W_i \ , \tag{2.2.3}$$

where the predictor vectors V_i and W_i may overlap. Letting h_i^* be the hazard function associated with S_i^*, we can rewrite Equations (2.2.1) and (2.2.2) as

$$\ln(-\ln \pi_i(b_i)) = \ln b_i + \alpha_\pi + \beta_\pi V_i, \quad \ln h_i^*(t_i|\ b_i) = \ln b_i + \alpha_\sigma + \beta_\sigma W_i.$$

In our parameterization, a higher value of the frailty b_i is associated with both a decreased cure probability and an increased hazard of experiencing the event. This is a natural and plausible way to allow the underlying frailty to affect both cure and survival.

Following common practice (Duchateau and Janssen (2008)), we assume that the frailty follows a gamma distribution with shape and scale both equal to $1/\theta$, ensuring that $\mathrm{E}(b) = 1$ and $\mathrm{Var}(b) = \theta$ ($\theta > 0$). Integrating out b_i, the marginal survival

function is

$$\begin{aligned}\widetilde{S}_i(t_i) &= \int S_i(t_i|\ b_i) f(b_i) db_i = \int S_i(t_i|\ b_i) \frac{b_i^{1/\theta-1}}{e^{-b_i/\theta}\theta^{1/\theta}\Gamma(1/\theta)} db_i \\ &= \left(\frac{1}{1+\theta e^{-\eta_{\pi_i}}}\right)^{1/\theta} + \left(\frac{1}{1+\theta t_i e^{\eta_{\sigma_i}}}\right)^{1/\theta} - \left(\frac{1}{1+\theta e^{-\eta_{\pi_i}}+\theta t_i e^{\eta_{\sigma_i}}}\right)^{1/\theta} \quad (2.2.4)\end{aligned}$$

with corresponding density

$$\widetilde{f}_i(t_i) = e^{\eta_{\sigma_i}}\left[\left(\frac{1}{1+\theta t_i e^{\eta_{\sigma_i}}}\right)^{1/\theta+1} - \left(\frac{1}{1+\theta e^{-\eta_{\pi_i}}+\theta t_i e^{\eta_{\sigma_i}}}\right)^{1/\theta+1}\right]. \quad (2.2.5)$$

Therefore, the log likelihood of the observed data is

$$l = \sum_i [d_i \ln \widetilde{f}_i(t_i) + (1-d_i)\ln \widetilde{S}_i(t_i)], \quad (2.2.6)$$

where for subject i, t_i is the observed time (minimum of event and censoring times) and d_i is the indicator that an event has occurred. We can rewrite (2.2.4) as

$$\widetilde{S}_i(t_i) = \widetilde{\pi}_i + (1-\widetilde{\pi}_i)\widetilde{S_i^*}(t_i),$$

where

$$\widetilde{\pi}_i = \left(\frac{1}{1+\theta e^{-\eta_{\pi_i}}}\right)^{1/\theta}, \quad (2.2.7)$$

and

$$\widetilde{S_i^*}(t_i) = \frac{(1+\theta t_i e^{\eta_{\sigma_i}})^{-1/\theta} - (1+\theta e^{-\eta_{\pi_i}}+\theta t_i e^{\eta_{\sigma_i}})^{-1/\theta}}{1-(1+\theta e^{-\eta_{\pi_i}})^{-1/\theta}}. \quad (2.2.8)$$

Because $\widetilde{S_i^*}(t_i)$ is a proper survival function, the marginal cure fraction is $\widetilde{\pi}_i$ and the marginal non-cured survival function is $\widetilde{S_i^*}(t_i)$, with corresponding marginal non-cured hazard function

$$\widetilde{h_i^*}(t_i) = e^{\eta_{\sigma_i}} \frac{(1+\theta t_i e^{\eta_{\sigma_i}})^{-1/\theta-1} - (1+\theta e^{-\eta_{\pi_i}}+\theta t_i e^{\eta_{\sigma_i}})^{-1/\theta-1}}{(1+\theta t_i e^{\eta_{\sigma_i}})^{-1/\theta} - (1+\theta e^{-\eta_{\pi_i}}+\theta t_i e^{\eta_{\sigma_i}})^{-1/\theta}}. \quad (2.2.9)$$

We note that the cure probability and hazard given not cured are originally defined at the subject level, conditional on the frailty. Thus although we have also derived the marginal cure fraction and non-cure hazard, as a rule we are more interested in the effects of predictors at the individual level, and we will interpret the parameters in this conditional sense.

2.2.2 Modeling Repeated Events

One can extend the model to repeated events by using the underlying frailty to induce dependence among the event times. That is, we assume that for each subject, given the susceptibility b_i the event times are independent with survival function given by (2.2.2) and (2.2.3). Denoting t_{ij} as the length of the j-th episode for subject i, we have

$$S_{ij}(t_{ij}|\ b_i) = \pi_{ij}(b_i) + [1 - \pi_{ij}(b_i)]S^*_{ij}(t_{ij}|\ b_i), \tag{2.2.10}$$

$$\pi_{ij}(b_i) = \exp[-b_i \exp(-\eta_{\pi_{ij}})], \quad S^*_{ij}(t_{ij}|b_i) = \exp[-t_{ij} b_i \exp(\eta_{\sigma_{ij}})], \tag{2.2.11}$$

where the linear predictors can include both subject-level and episode-level predictors. The log likelihood is then

$$l = \sum_i \ \ln\left[\int \exp\left(\sum_j l_{ij}\right) f(b_i)\ db_i\right], \tag{2.2.12}$$

where

$$\begin{aligned} l_{ij} &= d_{ij} \ln f_{ij} + (1 - d_{ij}) \ln S_{ij} \\ &= d_{ij} \ln\left(-\frac{\partial S_{ij}}{\partial t_{ij}}\right) + (1 - d_{ij}) \ln S_{ij}. \end{aligned}$$

One can in principle derive closed-form expressions for the likelihood terms, but the derivation quickly becomes tedious as the number of events increases. Moreover, the analytical formulas involve numerous subtractions among quantities like those in Equations (2.2.4) and (2.2.5), and therefore may lead to loss of precision when evaluated numerically. Therefore in our implementation of the model we have relied on numerical quadrature, which although mathematically only approximate is not subject to cancelation errors and therefore may be more reliable in practice.

2.2.3 Modeling Alternating States

As we have indicated, subjects in smoking cessation trials transit between smoking and abstinence, often several times during the course of treatment. Those currently in state 1 (abstinent) are at risk for events of type 1 (lapse), while those currently in state 2 (smoking) are at risk for events of type 2 (recovery of abstinence). Hougaard (2000) terms this situation "alternating states" (Figure 2.1).

To describe such data, we further extend the model by positing the existence of two correlated frailties $(b_i^{(1)}, b_i^{(2)})$ for each subject, where $b_i^{(k)}$ pertains to events of type k. As in Equations (2.2.10) and (2.2.11), the survival function for episode j of type k for subject i is

$$S_{ij}^{(k)}(t_{ij}^{(k)}|b_i^{(k)}) = \pi_{ij}^{(k)}(b_i^{(k)}) + [1 - \pi_{ij}^{(k)}(b_i^{(k)})]S_{ij}^{*(k)}(t_{ij}^{(k)}|b_i^{(k)}),$$

where

$$\pi_{ij}^{(k)}(b_i^{(k)}) = \exp[-b_i^{(k)} \exp(-\eta_{\pi_{ij}}^{(k)})],$$

$$S_{ij}^{*(k)}(t_{ij}^{(k)}|b_i^{(k)}) = \exp[-t_{ij}^{(k)} b_i^{(k)} \exp(\eta_{\sigma_{ij}}^{(k)})], \quad k = 1, 2, \tag{2.2.13}$$

and $\eta_{ij}^{(k)}$ and $\sigma_{ij}^{(k)}$ refer to the linear predictors for cure fraction and hazard, respectively, of subject i, episode j of type k. To complete the model, we must specify the joint distribution of the frailties. A popular choice in hierarchical models is the log-normal, which can flexibly represent a range of complex dependence structures. We prefer however to use marginal gamma distributions, which are more commonly associated with frailty models. Smith et al. (1982) proposed a five-parameter bivariate gamma distribution, but its intractability complicates implementation. An alternative approach is to apply a copula, which allows for correlation while retaining the specified marginals. We will use the Clayton copula (Clayton (1978)), whose joint cumulative distribution function (CDF) is:

$$F(b_i^{(1)}, b_i^{(2)}) = \begin{cases} \max[(F_1(b_i^{(1)})^{-\alpha} + F_2(b_i^{(2)})^{-\alpha} - 1)^{-1/\alpha}, 0] & \text{if } \alpha \neq 0, \\ F_1(b_i^{(1)})F_2(b_i^{(2)}) & \text{if } \alpha = 0; \end{cases}$$

where $F_k(\cdot)$, $k = 1, 2$, is the CDF of a gamma (shape=$1/\theta_k$, scale=$1/\theta_k$) distribution, and α is an association parameter defined on $(-1, +\infty)$. The copula approaches the minimum, product and maximum, as α approaches -1, 0 and $+\infty$, respectively. The parameter α measures the strength of association and is linked to Kendall's τ by $\tau = \alpha/(\alpha + 2)$. The corresponding density is

$$f(b_i^{(1)}, b_i^{(2)}) = \begin{cases} (1+\alpha)\ \max[(F_1(b_i^{(1)})^{-\alpha} + F_2(b_i^{(2)})^{-\alpha} - 1)^{-1/\alpha-2}, 0] \\ \qquad \times F_1(b_i^{(1)})^{-\alpha-1} F_2(b_i^{(2)})^{-\alpha-1} f_1(b_i^{(1)}) f_2(b_i^{(2)}) & \text{if } \alpha \neq 0, \\ f_1(b_i^{(1)})\ f_2(b_i^{(2)}) & \text{if } \alpha = 0; \end{cases}$$

where $f_k(\cdot)$, $k = 1, 2$, is the gamma density with shape and scale both equal to $1/\theta_k$.

The log likelihood is

$$l = \sum_i \ln\left[\int \exp\left(\sum_j l_{ij}\right) f(b_i^{(1)}, b_i^{(2)}) db_i^{(1)} db_i^{(2)}\right], \tag{2.2.14}$$

where

$$l_{ij} = \text{Indicator(type 1 event)} \cdot l_{ij}^{(1)} + \text{Indicator(type 2 event)} \cdot l_{ij}^{(2)},$$

$$l_{ij}^{(k)} = d_{ij}^{(k)} \ln\left(-\frac{\partial S_{ij}^{(k)}}{\partial t_{ij}^{(k)}}\right) + (1 - d_{ij}^{(k)}) \ln S_{ij}^{(k)}, \quad k = 1, 2.$$

2.2.4 Estimation

As indicated above, we advocate computing the likelihood by numerical integration. A straightforward approach is to use bivariate Gaussian quadrature (Abramowitz and Stegun (1972)), as implemented, for example, in SAS Proc NLMIXED. Unfortunately, we cannot apply Gaussian quadrature directly in Proc NLMIXED unless the random effects are normally distributed. Nelson et al. (2006) described a simple computational method using the probability integral transformation (PIT), and Liu and Huang (2007) applied it to various frailty proportional hazards models. Recently, Liu and Yu (2008) proposed another integration method that reformulates the conditional likelihood on non-normal random effects. One can implement both PIT and likelihood reformulation (LR) in Proc NLMIXED, but we favor LR because it offers reduced computing times and a potentially broader range of application.

Following the notation in the previous section, the likelihood from subject i is

$$L_i = \int \exp\left(\sum_j l_{ij}\right) f(b_i^{(1)}, b_i^{(2)}) db_i^{(1)} db_i^{(2)} \tag{2.2.15}$$

Multiplying and dividing the integrand by the standard bivariate normal density $\phi(\cdot,\cdot)$ gives

$$L_i = \int \exp\left(\sum_j l_{ij}\right) \frac{f(e^{a_i^{(1)}}, e^{a_i^{(2)}}) e^{a_i^{(1)}+a_i^{(2)}}}{\phi(a_i^{(1)}, a_i^{(2)})} \phi(a_i^{(1)}, a_i^{(2)})\ da_i^{(1)} da_i^{(2)}, \tag{2.2.16}$$

where $a_i^{(k)} = \ln b_i^{(k)}$, $k = 1, 2$; the extra $e^{a_i^{(1)}+a_i^{(2)}}$ arises from the Jacobian. We can then apply Gaussian quadrature in SAS Proc NLMIXED to integrate

$$\exp\left(\sum_j l_{ij}\right) \frac{f(e^{a_i^{(1)}}, e^{a_i^{(2)}}) e^{a_i^{(1)}+a_i^{(2)}}}{\phi(a_i^{(1)}, a_i^{(2)})}.$$

Specifically, the new conditional log likelihood for subject i consists of three elements:

$$l_i^A = \sum_j l_{ij},$$

$$l_i^B = \ln f(e^{a_i^{(1)}}, e^{a_i^{(2)}}) + a_i^{(1)} + a_i^{(2)},$$

$$l_i^C = -0.5\ (a_i^{(1)})^2 - 0.5\ (a_i^{(2)})^2 - \ln(2\pi),$$

leading to

$$L_i = \int \exp(l_i^A + l_i^B - l_i^C)\ \phi(a_i^{(1)}, a_i^{(2)})\ da_i^{(1)} da_i^{(2)}.$$

We have obtained maximum likelihood estimates (MLEs) using quasi-Newton optimization in SAS Proc NLMIXED, with covariance estimates from inversion of a finite-differences approximation to the Hessian (Dennis and Schnabel (1983)). As a check on the validity of standard errors (SEs), we have also computed estimated sample variances using a parametric bootstrap (Efron and Tibshirani (1993)). Specifically, denoting ξ as the parameter vector, we estimated the model from the actual data to obtain the MLE $\hat{\xi}$. We then generated a large number Q of data sets simulated

under the assumed model with ξ fixed at $\hat{\xi}$, obtaining from data set q the new estimate $\hat{\xi}^{*(q)}$. The estimated variance-covariance matrix is then the sample variance of the bootstrapped estimates $\hat{\xi}^{*(q)}$ $(q = 1, \ldots, Q)$. In simulations, confidence intervals (CIs) based on bootstrap SEs reliably gave near-nominal coverage probabilities.

2.2.5 A Marginal Approach

An alternative to likelihood-based modeling is a *marginal* approach (Yu and Peng (2008)), in which one estimates the model as though the correlated events are independent, then constructs a covariance matrix that adjusts for potential correlation (Williams (2000)). For our model, we assume the survival function for the length $T_{ij}^{(k)}$ of episode j of type k on subject i to be

$$\begin{aligned}\widetilde{S_{ij}^{(k)}}(t_{ij}^{(k)}) &= \left(\frac{1}{1+\theta^{(k)}e^{-\eta_{\pi_{ij}}^{(k)}}}\right)^{1/\theta^{(k)}} + \left(\frac{1}{1+\theta^{(k)}t_{ij}^{(k)}e^{\eta_{\sigma_{ij}}^{(k)}}}\right)^{1/\theta^{(k)}} \\ &\quad - \left(\frac{1}{1+\theta^{(k)}e^{-\eta_{\pi_{ij}}^{(k)}}+\theta^{(k)}t_{ij}^{(k)}e^{\eta_{\sigma_{ij}}^{(k)}}}\right)^{1/\theta^{(k)}}, \qquad (2.2.17)\end{aligned}$$

for $i = 1, \ldots, m$, $j = 1, \ldots, n_i^{(k)}$, $k = 1, 2$, where m is the total number of subjects and $n_i^{(k)}$ is the number of episodes of type k on subject i. Here $\eta_{\pi_{ij}}^{(k)}$ and $\eta_{\sigma_{ij}}^{(k)}$ refer to linear predictors for subject i, event j of type k, for cure fraction and event rate given not cured, respectively. Note that the marginal distribution of an event time $T_{ij}^{(k)}$ follows Equation (2.2.17) in all of the models in §2.2.1, §2.2.2 and §2.2.3. But in §2.2.2 and §2.2.3, the likelihood is calculated based on the joint distribution of $\{T_{ij}^{(k)}; j = 1, 2, \ldots, n_i^{(k)}\}$, instead of their marginal distributions.

We denote $U(\hat{\xi})$ as the estimated score function and $I(\hat{\xi})$ as the estimated infor-

mation matrix. Then we estimate the variance-covariance matrix of $\hat{\xi}$ by the Williams robust estimate $I^{-1}(\hat{\xi})\widehat{\text{Var}}_W(U(\hat{\xi}))I^{-1}(\hat{\xi})$, where

$$\begin{aligned}\widehat{\text{Var}}_W(U(\xi)) &= \widehat{\text{Var}}(\sum_{i=1}^{m}\sum_{k=1}^{2}\sum_{j=1}^{n_i^k} U_{ij}^{(k)}(\xi)) \\ &= \frac{m}{m-1}\sum_{i=1}^{m}(U_i(\xi)-\bar{U}(\xi))(U_i(\xi)-\bar{U}(\xi))^T, \qquad (2.2.18)\end{aligned}$$

$U_{ij}^{(k)}(\xi)$ is the score for episode j of type k on subject i, $U_i(\xi) = \sum_{k=1}^{2}\sum_{j=1}^{n_i^k} U_{ij}^{(k)}(\xi)$ is the score function for subject i, and $\bar{U}(\xi) = \sum_{i=1}^{m} U_i(\xi)/m$ is the mean of the individual scores.

Note that observing a set of recurrent event times over a fixed length of observation C induces dependent censoring in that the censoring time of T_{ij} is $C - T_{i1} - \cdots - T_{i,j-1}$ (for $j \geq 2$) (Huang and Chen (2003); Chen et al. (2004)). Specifically, the excess of shorter events observed for later episodes (except the first episode) can bias frequentist data summaries such as marginal model fits and simple Kaplan-Meier (KM) estimates. But because the data are coarsened at random, in the sense that censoring times depend only on fully observed data, likelihood-based inferences that ignore the randomness of the coarsening mechanism are nevertheless valid (Heitjan and Rubin (1991); Heitjan (1993, 1994)). We will revisit this issue in the simulations.

2.3 Application

The data are from two clinical trials of bupropion for smoking cessation — one ($n = 555$) conducted at Georgetown University (Lerman et al. (2002)) and the other ($n = 559$) at Brown University (Brown et al. (2007)). Because the designs were

nearly identical, the two studies have been published in combined analyses (Wileyto et al. (2005)). Briefly, eligible participants were randomized to receive 10 weeks of treatment with either sustained-release bupropion (Zyban, 300mg/day) or placebo, plus seven sessions of in-person behavioral group counseling. Subjects began taking assigned medication on the day of the first counseling session, and were instructed to smoke as normal until the target quit date (TQD), scheduled for 14 days after the initiation of drug treatment. Telephone follow-ups were conducted at the end of treatment (EOT) and 6 and 12 months post-TQD. Before each counseling session and during follow-up assessments, participants reported the number of cigarettes smoked each day since the previous assessment. From the daily cigarette consumption records we constructed the outcomes time to lapse (a day that includes smoking) and time to recovery from a lapse (an entire day without smoking).

We restrict our analysis to the treatment phase of the study, an 8-week period between TQD and EOT. Among the 1114 randomized subjects, 757 succeeded in quitting on the TQD and 357 failed. Experience has shown that subjects who are unable to quit on the TQD have sharply different outcome patterns, and therefore we excluded them from this analysis.

The 757 subjects had varying number of episodes, from 1 to 42. Among them, 314 subjects (42%) had only one episode, implying that they never lapsed during the entire 8 weeks. Ninety-five subjects (13%) had only two episodes, with an average censoring time 42, implying that they lapsed shortly after the initial quit and then continued smoking to the end of the study. Most of the rest of the subjects experienced many episodes with short durations, on average 8 episodes of length 7 days.

Figure 2.2 depicts the KM curves in the two groups for the two types of events, disregarding episode number and clustering. This is a rough way to understand the pattern of the data, because as we have indicated the censoring induces a potential bias in event times. Nevertheless we see that all curves exhibit an initial rapid decline followed by a plateau. The recovery events decline steeply to a nearly horizontal plateau, signaling the existence of a cured fraction. For lapse events the decline is more gradual and the plateau less perfectly flat, making the hypothesis of a cure model less certain but still plausible. For lapse events the curve for the intervention group appears to decline more slowly and to a higher plateau compared to the placebo group, suggesting possible drug effects on both cure fraction and hazard of lapse given not cured.

We analyzed the data under three models. First we fit the *Independent Frailty by Episode and Type (IFET)* model, which assumes separate, independent frailties for each event, adjusting for potential correlation using the cluster-correlated robust variance estimate of Williams (2000) (§2.2.5). The second model (*IFT*, for *Independent Frailty by Type*) assumes separate, independent frailties by event type (§2.2.2). That is, we again model the two types of events separately, but let the recurrent episodes of each type on one subject share a common underlying frailty, as shown in Equations (2.2.10) and (2.2.11). The third model (*CF* for *Correlated Frailty*) extends IFT by allowing the possibility of correlation between the two frailties through the gamma Clayton copula (§2.2.3). In all three models we assume that the linear predictor includes only an intercept and a drug treatment effect. Results appear in Table 2.1.

In the IFET model, robust SEs exceed naïve SEs by 9%–48% percent for all parameters pertaining to recovery events, whereas for lapse events, robust SEs associated with the variance and the intercept in the cure component are deflated slightly (15%–19%). The differences do not affect statistical significance, however. The Akaike Information Criterion (AIC) suggests a substantially better fit with IFT. Applying the parametric bootstrap with $Q = 20$ changes the SEs only slightly and has no effect on statistical significance. Examination of the CF results reveals that incorporating the association parameter α has little effect on the parameter estimates or the quality of the fit; the log likelihood increases by only 0.2 and the AIC by 1.6, and the 95% Wald CI for α includes 0 ($[-0.334, 0.417]$). Thus we henceforth restrict attention to the IFT model.

To check the model fit, we calculated the marginal survival function (Equation 2.2.4) based on the MLEs, for the two types of events and two treatment arms separately. In Figure 2.3, we compare these fitted marginal survival curves to KM estimates from the first episode only. We limited our analysis to the first episode because KM estimates from later episodes ($j \geq 2$) are potentially biased due to dependent censoring, as indicated in §2.2.5. The plots suggest that the model fits the data well. Based on Equation (2.2.7), the estimated marginal cure probabilities for lapse are 0.39 for bupropion and 0.27 for placebo, and for recovery are 0.21 for bupropion and 0.23 for placebo. For recovery events, these estimates are very close to the plateau levels in the KM plots in Figure 2.3. For lapse events, these estimates are slightly lower than the graphical estimates, presumably because the curves are not completely flat by the end of the observation period.

Based on the IFT model results, bupropion increases the probability of permanent quit, with cure probability 0.11 for drug and 0.01 for placebo (for a standard individual with frailty 1). For non-cured events bupropion reduces the hazard of lapse, with hazard ratio (HR) 0.56. Bupropion has no significant effect on the probability of abandoning quit attempts (0.15 for drug vs. 0.17 for placebo for a standard individual), nor does it accelerate recovery for those who persist in trying to quit (HR=1.15 for drug vs. placebo). Note that the individual-level cure probabilities are quite different from the marginal cure probabilities, especially for lapse events. The individual and marginal probabilities agree when the frailty variance is zero but can diverge substantially when the frailty variance increases. For our data the estimated frailty variances are 1.63 for lapse and 0.26 for recovery, whence the disparity between marginal and individual quit probabilities.

In the CF model we expected that shorter times to recovery would be correlated with longer times to relapse, reflecting stronger addiction and implying a negative association parameter. The estimated α is slightly positive, although not significant. We postulate that individuals with a heavier addiction may have a stronger impulse to quit, which induces a higher frailty to recovery, while at the same time the heavy addiction makes abstinence harder and thus corresponds to a higher frailty to lapse.

2.4 Simulation

We conducted a series of simulations to evaluate the performance of our method in repeated samples. Each simulation consisted of 100 replications of 800 subjects.

We fixed the length of follow-up at $C = 60$ and assigned treatment randomly with probability 1/2, as in the trial. We chose the parameters to be similar to estimates from the data: For the lapse parameters, we set the intercept at -1.5 and the drug coefficient at 0.7, corresponding to a cure probability of 0.01 for placebo and 0.11 for drug. We set the intercept and drug coefficients in the survival part at -1 and -0.6, giving a baseline hazard of 0.37 and HR of 0.55 for drug vs. placebo. For recovery events, we set the intercept and drug coefficients in the cure component to be 0.5 and -0.1, and those in the survival part to be -1 and 0.1. We fixed the frailty variances at 1 and 0.3 and the association parameter at 1.

For each simulated subject, we first generated two frailties from the Clayton copula using function `mvdc()` from the package `copula` in `R` 2.6.2 (Yan (2007)). This gave an average correlation of 0.35 in the 100 replicates, with range from 0.21 to 0.41. Next, we generated lapse times $T_j^{(1)}$ and recovery times $T_j^{(2)}$. To incorporate the cure fraction, we drew random variates $R_j^{(k)}, k = 1, 2$ from a Uniform (0,1); if $R_j^{(k)} < \pi_j^{(k)}$, $T_j^{(k)}$ was changed to a number beyond the end of study ($T_j^{(k)} = C + 10$), representing a "cure". The recurrent event times were then lined up as: $T_1^{(1)}, T_1^{(2)}, T_2^{(1)}, T_2^{(2)}, \cdots$, and the cumulative time was used to determine censoring. That is, if the sum Ψ_l of times to the l-th event was less than C, and the sum Ψ_{l+1} of times to the $l+1$-st event was greater than C, then the $l+1$-st event was censored at $C - \Psi_l$. We fitted the CF model in Proc NLMIXED.

Simulation results appear in Table 2.2. Bias for the CF model is modest, ranging from -0.010 to 0.014, and all of the estimates except those for α are tightly distributed around their means. Hessian-based 95% CIs achieve roughly nominal coverage, with

values ranging from 89% to 98%. Thus the overall performance of the estimates appears to be satisfactory. Moreover, 94% of the α estimates are significant at level 0.05, suggesting that the CF model has substantial power to detect a moderate underlying correlation.

We also fit the IFT model (fixing $\alpha = 0$) to the simulated data (Table 2.2). Estimated biases under the incorrect IFT model are 3 to 30 times larger than under the correct CF model, and the coverage of 95% CIs is much poorer, ranging as low as 39%. Thus the CF model can provide a substantially better fit than the IFT model when the frailty correlation is moderate.

To better understand the power in testing the significance of α, we conducted two other simulations (results not shown) where the true α was set to be 0.4 or 0. With $\alpha = 0.4$, the numerical correlation is around 0.17, and 29 percent of the α estimates are significant. With α=0, 10% of the estimates are significant, so the type I error is somewhat inflated. We also investigated the likelihood ratio test (LRT), which gives power 95%, 40% and 6% at α=1, 0.4, and 0, respectively, suggesting its superiority over the Wald test.

To investigate potential bias in the IFET (marginal) model (as suggested in §2.2.5), we estimated both the IFT and IFET models when α was set to 0 (Table 2.3). As expected, biases in the parameter estimates under the IFT model are small (absolute value 0 to 0.019), whereas biases under the IFET model range as high as 0.405. Accordingly, 95% CIs have good coverage under the IFT model (91%–97%) and poor coverage under the IFET model (0%–78%).

Lastly, we estimated the parametric bootstrap SE for the CF model in each sce-

nario. In all cases it turns out to be close to the Hessian-based SE ($\pm$ 0.005), with 95% CIs based on both approaches achieving near-nominal coverage probability. So the reliability of the simple Hessian-based SE seems adequate in these settings. For some other parameter values that we investigated, however, the Hessian-based SEs give reduced coverage probabilities; by contrast, the bootstrap method is well calibrated in every case.

2.5 Discussion

We have proposed a novel cure-mixture frailty model that allows for correlated events of alternating types, as one encounters in smoking cessation studies. Our model is unique in positing that a common frailty influences both the cure fraction and the survival rate. The use of a cloglog link for the cure fraction and a constant hazard for the non-cured survival offers analytic convenience, but because we use numerical integration and optimization even with small numbers of events, one could easily extend the model by substituting other, less tractable link functions and survival distributions. Moreover, although the models we estimated include only a randomization indicator in the linear predictor, it is straightforward to incorporate baseline and episode-specific predictors. In particular, this would allow for estimating trends across episodes.

Our method offers an alternative to mover-stayer Markov models with alternating states and a possibility of cure (Cook et al. (2002); Mannan and Koval (2003)) by incorporating time-to-event models for correlated outcomes. Use of the Clayton copula

maintains the simplicity of marginally gamma frailties while avoiding the complexity of the bivariate gamma. Moreover, as with the link and hazard specification, one can readily substitute different and more complex frailty models. A potential disadvantage is that with dimension three or higher the Clayton copula allows only positive association.

As evaluation of the likelihood requires numerical integration in all but the simplest cases, parameter estimation is computationally intensive. We have employed the LR method of Liu and Yu (2008), implemented in SAS Proc NLMIXED with the recommended 30 quadrature points and default quasi-Newton optimization. Simulation results suggest that the procedure works well, and moreover we confirmed results for fitting the IFET model in an independent implementation in `R`.

In simulations with smaller sample sizes (e.g., $m = 200$), estimates sometimes failed to converge. One can avoid this problem by identifying better starting values, reparameterizing to avoid boundary constraints, or employing a more robust optimization technique such as Nelder-Mead. Changing optimization algorithms should be a last resort, however, as the default quasi-Newton method typically attains convergence in several minutes compared to possibly several hours for Nelder-Mead.

With this numerically intensive approach, the largest feasible number of simulation replicates was 100, and thus estimates of frequentist operating characteristics have substantial uncertainty. Nevertheless, we were able to establish that one can fit the model reliably in samples of practical size, that departures from underlying assumptions can cause severe bias, and that SEs derived from a finite-differences estimate of the Hessian can be accurate. Because for some parameter settings the

Hessian-based SEs underestimated sampling variability, we recommend that in key analyses one employ the parametric bootstrap, which gave properly calibrated CIs in every simulation we attempted.

We applied our model to smoking data collected during the relatively brief 8-week treatment period, where time-to-lapse curves had not reached a plateau and consequently the cure fraction might be poorly estimated. Nevertheless we are convinced that there is a cured fraction, as evidence from other trials suggests strongly that this is a real feature of smoking cessation data. It is well known that when cure is a possibility, failure to incorporate it in the model can cause severe bias in survival hazard parameters (Sy and Taylor (2000); Price and Manatunga (2001)). The consequences of incorrectly assuming a cured fraction are perhaps less well known but amenable to investigation in any specific case. In fact we find that the cure model is most helpful as an analytic tool when there is enough censoring to create some ambiguity about it; otherwise, one can easily classify subjects as cured or not and then estimate cure and survival models directly. Nevertheless, we recommend restricting the use of the model to problems where the possibility of cure has a strong basis in the science.

There is a growing literature on the analysis of smoking cessation data. Banerjee and Carlin (2004) extended cure models to allow for spatial correlation using a Bayesian approach, and Yu and Peng (2008) applied a univariate cure model, adjusting for clustering by a one-step jackknife. Both of these models focused on lapse events only. Yu and Peng's model is similar to our IFET model, but substitutes a jackknife for the sandwich variance estimate. A marginal approach can work well with clustered survival data but, as our simulations show, is subject to bias when applied

to recurrent events. Another recent article by Luo et al. (2008) used a discrete-time stochastic model to describe the transitions among smoking, transient quitting, and permanent quitting. A major difference with our work is that they model visit-to-visit transitions of smoking status at lengthy intervals, whereas we model transitions in more nearly continuous time. Moreover, their model fails to account for the probability of abandoning the quit attempt.

A potential further application of our model is to estimate individual frailties as a means of characterizing subjects with respect to amenability to treatment. This is important because nicotine addiction is most helpfully thought of as a chronic disease that may take years to successfully treat. Estimating patients' underlying chances of success may ultimately lead to better prediction of outcomes and personalization of therapies.

Although we developed our model to describe cigarette consumption histories in a randomized smoking cessation trial, it is potentially applicable in observational studies and in research on cessation of other substances of abuse. The modeling of alternating states will also be applicable in the study of relapsing-remitting diseases such as multiple sclerosis and depression, and diseases such as heart failure that involve repeated hospitalizations.

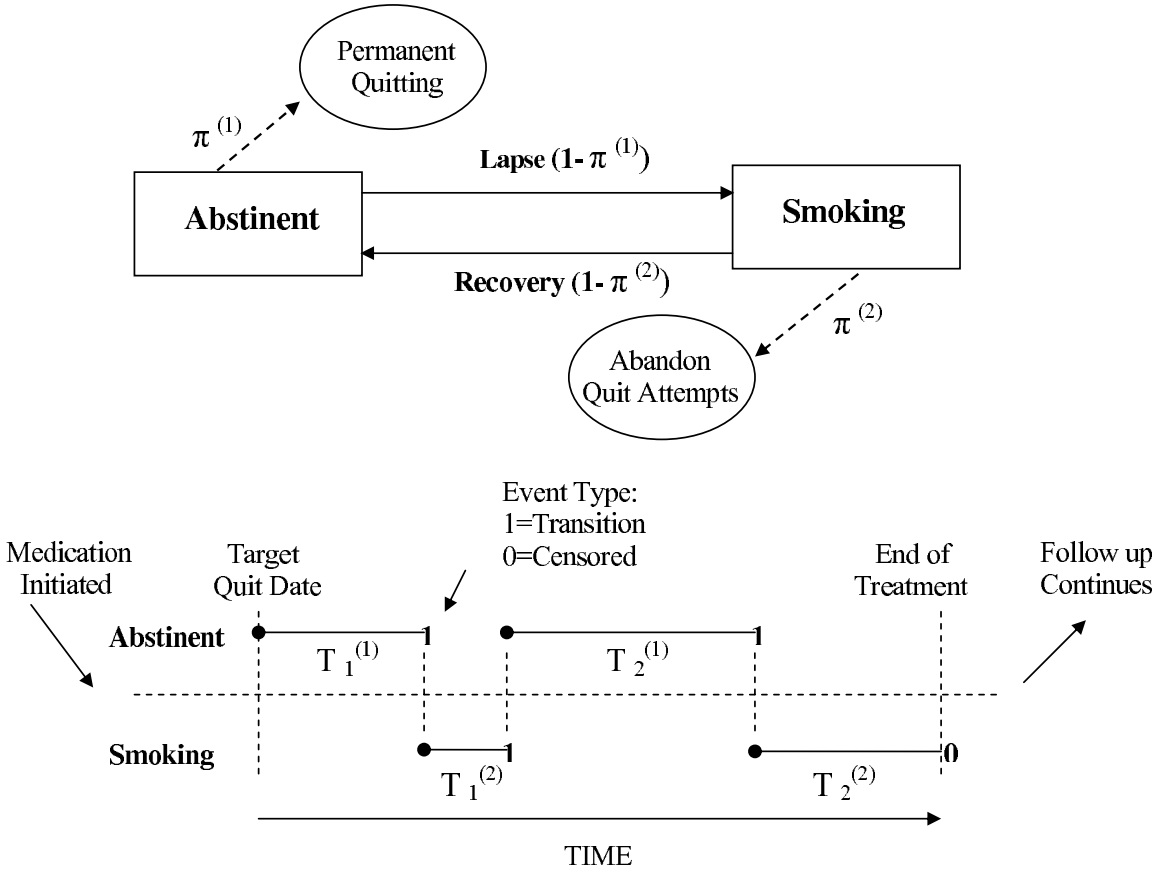

Figure 2.1: Alternating states data.

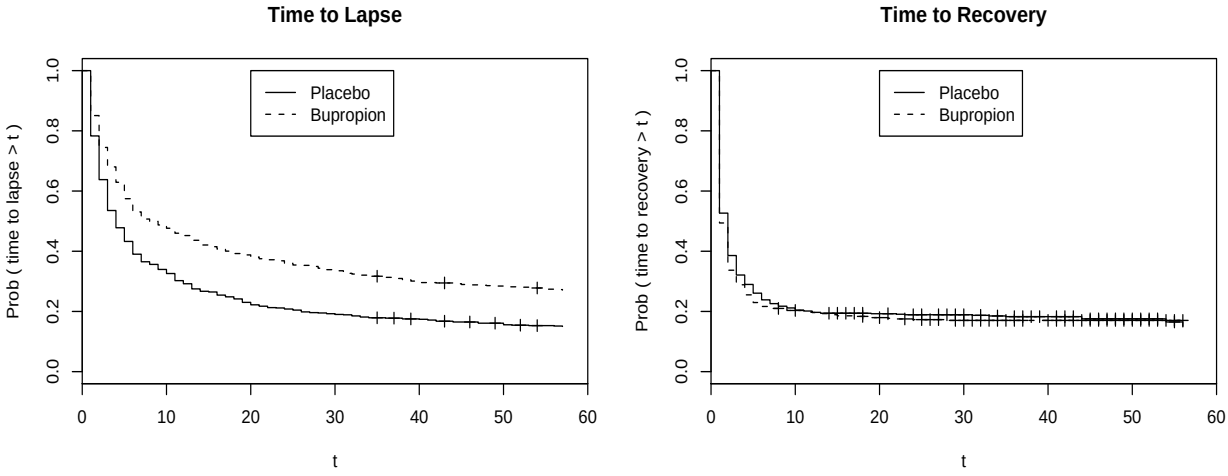

Figure 2.2: Kaplan-Meier estimates for the two event types, by treatment arms.

Table 2.1: Parameter estimates from the bupropion data.

	IFET		IFT		CF	
	Estimate	SE(adjusted)	Estimate	SE	Estimate	SE
Lapse						
Cure fraction						
Intercept	−1.627	0.210(0.171)	−1.479	0.203	−1.475	0.207
Drug	0.669	0.141(0.150)	0.680	0.221	0.681	0.220
Survival model						
Intercept	−1.581	0.059(0.072)	−2.096	0.090	−2.096	0.090
Drug	−0.447	0.082(0.126)	−0.586	0.122	−0.587	0.122
Variance	0.808	0.125(0.106)	1.634	0.144	1.629	0.157
Recovery						
Cure fraction						
Intercept	−0.734	0.062(0.071)	−0.578	0.062	−0.570	0.074
Drug	−0.019	0.081(0.088)	−0.059	0.091	−0.057	0.091
Survival model						
Intercept	−0.815	0.048(0.055)	−0.932	0.060	−0.942	0.075
Drug	0.117	0.069(0.082)	0.141	0.087	0.140	0.087
Variance	0.201	0.029(0.044)	0.255	0.043	0.261	0.044
Association	—	—	—	—	0.041	0.191
Log likelihood	−8024.1		−7716.2		−7716.0	
AIC	16068.2		15452.4		15454.0	

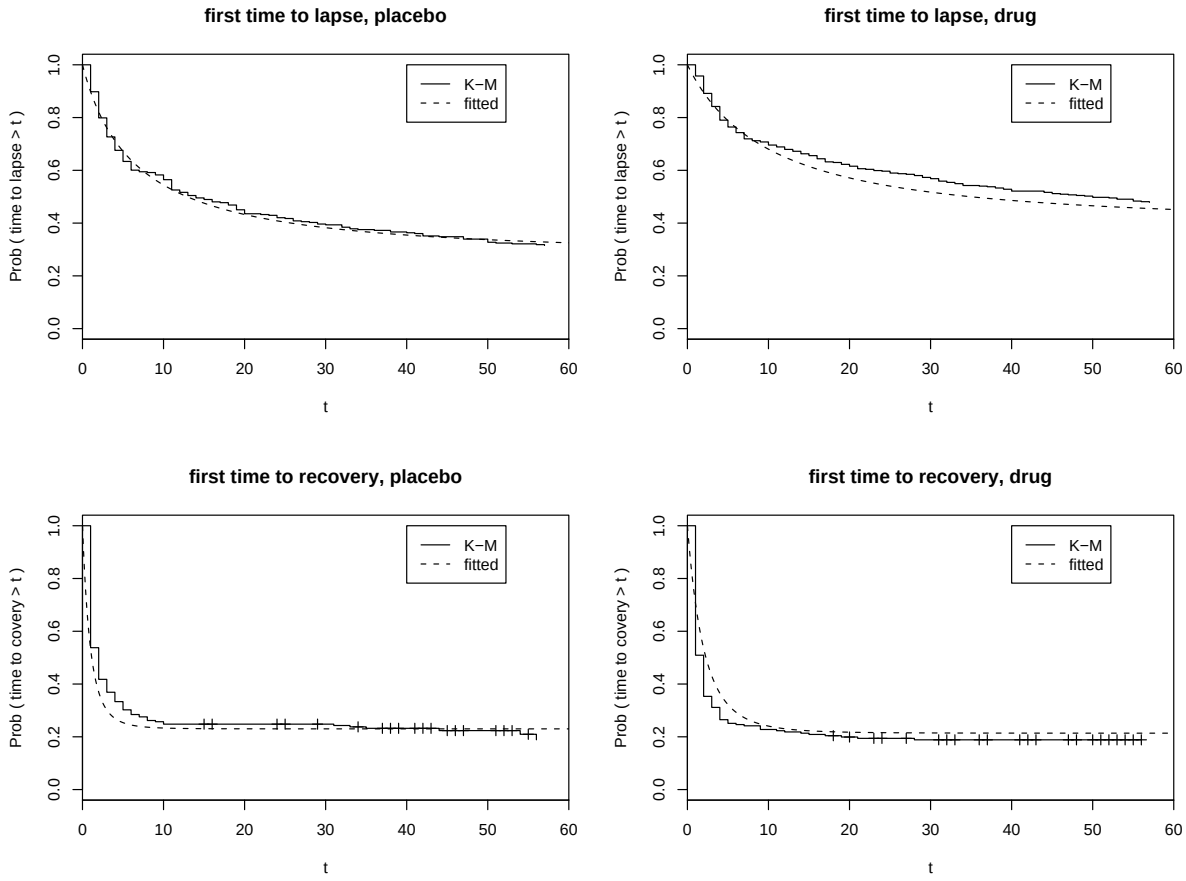

Figure 2.3: Fitted marginal survival curves and Kaplan-Meier estimates for the first episode.

Table 2.2: Results of simulations with $\alpha = 1$.

		CF			IFT		
Parameter	True value	Bias	$\sqrt{MSE}$	95 % CI CP (%)	Bias	$\sqrt{MSE}$	95 % CI CP (%)
Lapse							
cure fraction							
intercept	−1.5	0.010	0.113	93	−0.052	0.127	88
drug	0.7	0.000	0.119	93	0.025	0.123	95
survival model							
intercept	−1.0	−0.010	0.078	89	0.026	0.082	86
drug	−0.6	−0.005	0.095	93	−0.022	0.097	94
variance	1.0	0.001	0.079	96	0.030	0.084	96
Recovery							
cure fraction							
intercept	−0.5	0.000	0.049	97	−0.099	0.108	39
drug	−0.1	−0.003	0.072	96	−0.038	0.083	90
survival model							
intercept	−1.0	0.002	0.056	94	0.099	0.110	41
drug	0.1	−0.006	0.066	98	0.022	0.070	93
variance	0.3	0.004	0.040	95	−0.056	0.063	61
Association	1.0	0.014	0.343	96	—	—	—

Note: "CP" = Coverage probability.

Table 2.3: Results of simulations with $\alpha = 0$.

		IFT			IFET		
Parameter	True value	Bias	$\sqrt{MSE}$	95 % CI CP (%)	Bias	$\sqrt{MSE}$	95 % CI CP (%)
Lapse							
cure fraction							
intercept	−1.5	0.019	0.093	97	0.241	0.271	49
drug	0.7	−0.014	0.116	97	−0.117	0.166	72
survival model							
intercept	−1.0	0.001	0.064	91	0.332	0.340	0
drug	−0.6	0.003	0.094	95	0.112	0.148	54
variance	1.0	−0.016	0.099	91	−0.405	0.413	0
Recovery							
cure fraction							
intercept	−0.5	−0.004	0.054	93	−0.252	0.266	2
drug	−0.1	−0.008	0.076	95	−0.055	0.100	78
survival model							
intercept	−1.0	0.001	0.051	95	0.278	0.285	1
drug	0.1	0.000	0.070	95	−0.059	0.101	70
variance	0.3	−0.004	0.039	93	−0.079	0.104	44
Association	0	—	—	—	—	—	—

Note: "CP" = Coverage probability.

Chapter 3

Prediction of Individual Long-Term Smoking Outcomes

3.1 Introduction

In smoking cessation clinical trials, subjects commonly receive treatment over a defined length of time (often 8–12 weeks), with follow-up assessments taken at 6 and 12 months after baseline. The outcome at the end of treatment is an important but imperfect indicator of efficacy, in that some subjects who quit early are unable to sustain their initial success, whereas some who fail early later succeed. Therefore it is of interest clinically to predict long-term smoking outcome from short-term data. And as we indicated in the end of Chapter 2, estimating patients' underlying chances of success may ultimately lead to personalization of therapies and optimization of treatments to reverse short-term failure and sustain short-term success.

The unique features of smoking cessation data complicate statistical analysis, as

we described in Chapter 2. First, subjects commonly experience a series of lapse and recovery episodes of varying lengths, alternating between smoking and abstinence. Moreover, any quit episode may be temporary, followed by a lapse, but it may also become permanent, in that the subject quits smoking forever. Conversely, any lapse may be temporary, leading to a recovery, but it may also become permanent, in that the subject gives up the quit attempt altogether. One can model such possibilities of permanence by "cure" probabilities in time-to-event models. Another important feature of such data is heterogeneity, in that outcomes depend critically on unobserved factors that are not encoded in baseline covariates. We can model such heterogeneity by random effects, or frailties as they are denoted in survival analysis.

There is a growing statistical literature on the modeling of complex smoking cessation patterns (Cook et al. (2002); Mannan and Koval (2003); Banerjee and Carlin (2004); Yu and Peng (2008); Luo et al. (2008)). In Chapter 2, we proposed a cure-mixture frailty model that allows for recurrent, correlated events of two alternating types, with a possibility of cure in each episode (Li et al. (2010)). Our analysis sought to provide a more accurate and nuanced representation of the activities of the treatments in question by disentangling their effects on the daily hazards and long-term cure probabilities of lapse and recovery. Although in our real-data example longer-term outcomes were collected, we based our analysis on data only from the eight-week treatment phase.

In this chapter we propose using our model to make long-term predictions. We develop a two-stage Bayesian algorithm that first imputes frailties then generates predicted smoking trajectories given the frailties, in both stages conditioning on baseline

predictors, short-term outcomes and parameter estimates. We illustrate the method with an analysis of data from two smoking cessation trials we described before, comparing our predictions to those from a variety of *ad hoc* methods. We moreover perform an array of analyses to evaluate the sensitivity of our predictions to uncertainties in the modeling.

3.2 Review of the Smoking Cessation Trials and the Cure-Mixture Frailty Model

We use the same data as the analysis in Chapter 2, which come from two randomized trials of oral bupropion for smoking cessation. Briefly, subjects who reported smoking at least 10 cigarettes/day were randomized to receive ten weeks of either twice-daily bupropion or placebo. All subjects received seven sessions of in-person behavioral group counseling. The study collected a range of baseline predictors of smoking behavior, including the Fagerström Test for Nicotine Dependence (FTND) score (Heatherton et al. (1991)), the Center for Epidemiologic Studies Depression (CESD) scale (Radloff (1977)), and race and sex (Dale et al. (2001)).

Participants began taking assigned medication on the day of the first counseling session, and all were to quit smoking on the target quit date (TQD), two weeks later. Prior to each counseling session between TQD and the end of treatment (EOT, eight weeks after TQD) participants reported numbers of cigarettes smoked each day since the previous assessment, using a time-line follow-back (TLFB) questionnaire.

From the reported daily cigarette consumption, we constructed for each subject the on-treatment series of times to lapse (a day that includes smoking) and times to recovery (an entire day without smoking). Telephone assessment followed at EOT and 6 months and 12 months post-TQD, where subjects were asked if they had smoked in the last 7 days. Self-reported quits were verified by tests for smoking by-products in the breath or saliva. As we indicated in Chapter 2, our analysis will be based on 757 subjects who succeeded in quitting on TQD.

We now review our model for smoking cessation data with alternating states and cure probabilities (Figure 2.1) (Li et al. (2010)). We denote $T_{ij}^{(k)}$ and $S_{ij}^{(k)}$ as the event time and its survival function for episode j of type k for subject i ($k = 1$ for lapse, $k = 2$ for recovery). To incorporate the unobserved risk factors, we assume that each subject has two latent frailties $(b_i^{(1)}, b_i^{(2)})$ and conditional on them, the survival function $S_{ij}^{(k)}$ is:

$$S_{ij}^{(k)}(t_{ij}^{(k)}|b_i^{(k)}) = \pi_{ij}^{(k)}(b_i^{(k)}) + [1 - \pi_{ij}^{(k)}(b_i^{(k)})]S_{ij}^{*(k)}(t_{ij}^{(k)}|b_i^{(k)}), \tag{3.2.1}$$

where

$$\pi_{ij}^{(k)}(b_i^{(k)}) = \exp[-b_i^{(k)} \exp(-\eta_{\pi_{ij}}^{(k)})], \tag{3.2.2}$$

$$S_{ij}^{*(k)}(t_{ij}^{(k)}|b_i^{(k)}) = \exp[-t_{ij}^{(k)} b_i^{(k)} \exp(\eta_{\sigma_{ij}}^{(k)})], \quad k = 1, 2. \tag{3.2.3}$$

That is, each episode either becomes permanent (with probability $\pi_{ij}^{(k)}$) or ends according to a proper survival function $S_{ij}^{*(k)}(\cdot)$ (with probability $1 - \pi_{ij}^{(k)}$). The linear predictors $\eta_{\pi_{ij}}^{(k)}$ and $\eta_{\sigma_{ij}}^{(k)}$ contain subject-level or episode-level covariates, such as intercept, treatment assignment, and baseline factors.

We link the two frailties by the Clayton copula (Clayton (1978)), so that their joint density is

$$\begin{aligned} f(b_i^{(1)}, b_i^{(2)}) &= (1+\alpha)[F_1(b_i^{(1)})^{-\alpha} + F_2(b_i^{(2)})^{-\alpha} - 1]^{-1/\alpha-2} \\ &\quad \times F_1(b_i^{(1)})^{-\alpha-1} F_2(b_i^{(2)})^{-\alpha-1} f_1(b_i^{(1)}) f_2(b_i^{(2)}), \end{aligned} \tag{3.2.4}$$

where $F_k(\cdot)$, $k = 1, 2$, is the cumulative distribution function (CDF) of a gamma (shape=$1/\theta_k$, scale=$1/\theta_k$) distribution, $f_k(\cdot)$, $k = 1, 2$, is the corresponding gamma density, and α is an association parameter defined on $(-1, +\infty)$.

Letting ξ refer to the parameter vector of covariate effects in the linear predictors, the frailty variances θ_k $(k = 1, 2)$ and the association parameter α, one can obtain the maximum likelihood estimate (MLE) $\hat{\xi}$ using SAS Proc NLMIXED. For details see Chapter 2.

3.3 Prediction Methods

3.3.1 Sampling Individual Frailties

A natural way to construct predictions is to sample in two stages: First, draw the subject-level frailties from their posterior distribution; and second, generate individual future outcomes conditional on the sampled frailties. Denoting the observed data for subject i as $y_i \equiv \{t_{ij}^{(k)} : j = 1, 2, \ldots, n_i; k = 1, 2\}$, and the frailty vector as $b_i \equiv \{b_i^{(1)}, b_i^{(2)}\}$, the posterior distribution of the frailties for subject i is

$$p(b_i|y_i, \xi) = \frac{p(y_i|b_i, \xi) f(b_i|\xi)}{\int p(y_i|b_i, \xi) f(b_i|\xi) db_i}. \tag{3.3.1}$$

Here $f(b_i|\xi)$ is the copula density from (3.2.4), and $p(y_i|b_i,\xi)$ is the conditional density of the vector of event times given b_i for subject i:

$$p(y_i|b_i,\xi) = \prod_{j=1}^{n_i} \left[\left(-\frac{\partial S_{ij}^{(1)}}{\partial t_{ij}^{(1)}}\right)^{d_{ij}^{(1)}} \cdot \left(S_{ij}^{(1)}\right)^{(1-d_{ij}^{(1)})}\right]^{\delta_{ij}} \times \left[\left(-\frac{\partial S_{ij}^{(2)}}{\partial t_{ij}^{(2)}}\right)^{d_{ij}^{(2)}} \cdot \left(S_{ij}^{(2)}\right)^{(1-d_{ij}^{(2)})}\right]^{1-\delta_{ij}}, \qquad (3.3.2)$$

where $d_{ij}^{(k)}, k = 1, 2$, is the indicator that an event has occurred; and δ_{ij} is the indicator that the episode is of the first type (time to lapse). The numerator can be readily calculated, but the integral in the denominator cannot. To avoid this complication we resort to importance sampling (Tanner (1993)).

Importance sampling proceeds as follows: First, one designates an approximate posterior $p^*(b_i)$ from which one can readily sample and whose density is easy to calculate. One then generates a large number (say, M) of sample draws $b_i^{[m]}$, $m = 1, \ldots, M$, from this distribution. At each sampled value, one computes the true value of a function proportional to the posterior density (the numerator) and the value of the approximate posterior density. Their ratio $r^{[m]}$ is then normalized as $w^{[m]}$, which becomes a weight for use in subsequent calculation:

$$r^{[m]} = \frac{p(y_i|b_i^{[m]},\xi) f(b_i^{[m]}|\xi)}{p^*(b_i^{[m]})}, \quad w^{[m]} = \frac{r^{[m]}}{\sum_{m=1}^{M} r^{[m]}}. \qquad (3.3.3)$$

To generate a posterior sample of b_i one can use sampling/importance resampling (SIR) (Rubin (1987)). In SIR, one first draws M values by importance sampling, then selects Q of these, $Q \ll M$, with replacement using probabilities $w^{[m]}$. The Q values so selected $(b_i^{*[q]}, q = 1, \ldots, Q)$ represent a random sample from an approximation to the target posterior (Rubin (1987); Smith et al. (1982)).

One could also calculate some summary measures of b_i from its posterior distribution. For example, to calculate the posterior mean of a function $g(b_i)$, one evaluates the function at all sampled values of b_i, say $g(b_i^{[m]})$, $m = 1, \ldots, M$. The approximate mean is then

$$\mathrm{E}[g(b_i)] \approx \sum_{m=1}^{M} w^{[m]} g(b_i^{[m]}). \tag{3.3.4}$$

For example, the posterior mean of the first frailty $b_i^{(1)}$ can be simply approximated by $\sum_{m=1}^{M} w^{[m]} b_i^{(1)[m]}$.

3.3.2 Predicting Future Outcomes

To predict the outcome at one year post-TQD for subject i, we first obtain posterior samples of b_i using SIR and then generate the transition times after EOT under the assumed model with $\hat{\xi}$ and individual frailties $b_i^{*[q]}$ $(q = 1, \ldots, Q)$. If this subject's last observation is an event time, the generating procedure is straightforward: Set the next event type to be opposite to the last observed event and generate a sequence of events of alternating types: $T_{ij}^{(k)}$. To incorporate the cure probability, we draw a sequence of random variates from a Uniform $(0, 1)$; if the jth variate of type k in the sequence is less than $\pi_{ij}^{(k)}$, we re-set $T_{ij}^{(k)} = \infty$, representing a "cure". The recurrent event times are then lined up as $T_{i1}^{(1)}, T_{i1}^{(2)}, T_{i2}^{(1)}, T_{i2}^{(2)}, \cdots$, and the cumulative time is used to determine the censoring.

If subject i's last observation is censored, we must first simulate whether this episode is permanent or temporary. Denoting the last observed censoring time as t_i,

the conditional cure probability given observed data is

$$\pi_i = \Pr(\text{cured}|T_{ij}^{(k)} > t_i) = \frac{\Pr(\text{cured})}{\Pr(T_{ij}^{(k)} > t_i)} = \frac{\pi_{ij}^{(k)}}{\pi_{ij}^{(k)} + (1 - \pi_{ij}^{(k)})S_{ij}^{*(k)}(t_i)}. \qquad (3.3.5)$$

Again we draw a random variate from Uniform $(0, 1)$. If it is less than π_i, the last observed status is permanent and no further events are generated; otherwise, we generate further episodes by the above procedure, with the starting event type the same as the last observed event type. Because given the frailties the event times are exponential, the conditional distribution of the time from EOT until the first post-EOT event is the same as if the time origin for the event were at EOT.

Because we derive the event times from daily smoking records, the smallest time unit is one day. Thus although in principle the last pre-EOT episode should always be censored, in practice some subjects' last episode is an event. Moreover with our model of continuous times, there is a possibility of multiple events within one day. We therefore reconstruct the simulated data so that the transition times are in units of days. Specifically, we decide the daily smoking status based on $t_{ij}^{(k)}$, applying the rule that any smoking during a day indicates non-abstinence (Figure 3.1). We then obtain the subject's smoking status (1 for smoking, 0 for non-smoking) at one year by evaluating whether the simulated last seven days are all smoke-free. This renders predictions comparable to the actual verified outcomes.

Repeating this process Q times, we obtain Q draws from an approximate predictive posterior. We estimate the individual probabilities of smoking p_i by counting the proportion of draws that find subject i smoking within seven days of the one-year milestone. These predictive probabilities are in fact samples from the predictive pos-

terior distribution and henceforth is referred to as the SPP method. We then compare our predictions to the true, verified values of smoking at one year and estimate an ROC curve (Fawcett (2006)). Specifically, we apply different thresholds to p_i and for each we calculate the sensitivity and specificity. We obtain the ROC curve by connecting the points, using AUC (area under the curve) as an index of predictive accuracy (Hanley and Mcneil (1982)).

We have compared our predictions under the frailty model to predictions from a range of logistic regression models estimated using the year one outcomes (see the detailed description below). As such models are predicting the same outcomes on which they were developed, they are subject to over-fitting. In these cases we applied a bootstrap correction to adjust the AUC (Harrell et al. (1996)).

Analysis was done in `R` 2.6.2.

3.4 Results

3.4.1 Model Selection

We first fit the frailty model to the treatment period data to generate a best-fitting model for the short-term outcomes. Following previous work (Wileyto et al. (2005)), we dichotomized FTND to be positive if the raw score ≥ 6 (representing greater dependence) and CESD to be positive if the raw score ≥ 16 (representing more severe depression). To allow for potential time trends within subjects, we included episode number and its interaction with treatment as predictors. We recoded episode number

for each type of event as 1, 2, 3 and ≥ 4 to attenuate the influence of a handful of subjects who had many episodes. We deleted data from six subjects who had missing predictors, leaving us with 751 observations.

Because estimating the joint model is numerically demanding (i.e. *CF* model in Chapter 2), our initial screening analyses fit separate models for lapse and recovery events (i.e. *IFT* model in Chapter 2). For each type of event, we began with a saturated model including drug, sex, race, FTND, CESD, episode number and the episode-by-treatment interaction. We then reduced the model in a backward fashion, eliminating variables whose p-values were larger than 0.05. This procedure removed race, CESD and the episode-by-treatment interaction for both types of episode, indicating no association of outcome with race and depression, and no differential effect of treatment across episodes. The remaining variables were then included in a joint model for the two types of event.

Table 3.1 shows estimates from the selected model. Bupropion increases the chance of permanent cure and decreases the hazard of lapse, with hazard ratio (HR) $e^{-0.609}$=0.54. Female sex is associated with a smaller chance of permanent cure and a higher hazard of lapse (HR=$e^{0.243}$=1.28). The cure probability does not change significantly over time, but the hazard of lapse increases, with HR $e^{0.146}$=1.16 for episode j vs. episode $j-1$. For recovery events, bupropion has no significant effect on the probability of abandoning quit attempts, nor does it accelerate recovery for those who persist in trying to quit (HR=$e^{0.128}$=1.14). Subjects with elevated FTND are more likely to give up, and have longer time to recovery even if they persist (HR=$e^{-0.198}$=0.82). The probability of abandoning quit attempts increases

over time, and the hazard of recovery decreases (HR=$e^{-0.105}$=0.90 for episode j vs. episode $j-1$). Even after controlling for these baseline factors there is substantial variability in the data (reflected by the frailty variance estimates), but no significant association between the lapse and recovery frailties.

3.4.2 Sampling Individual Frailties

We approximated the posterior distribution of the frailties for each subject, given the observed data and parameter estimates from the final model. We used independent gamma distributions with mean 1 and variance $\hat{\theta}_k$, $k = 1, 2$, as the approximating proposal distribution $p^*(b_i)$ for the importance sampling. To secure better coverage of the tails of the true posterior, we multiplied the approximate variance by 2. We then generated $M = 8,000$ draws from this gamma distribution and calculated the importance ratios. Finally, we sampled from the posterior using SIR with $Q = M/10$ (Smith and Gelfand (1992)).

We conducted some diagnostic evaluations of the quality of the importance sampling approximation: First, we calculated posterior means for various values of M to determine whether they had stabilized (Tanner (1993)). Second, we varied the scale factor for the approximation to see whether increasing it affected the estimates (Tanner (1993)). Third, we examined the histogram of the log importance weights, where a long right tail suggests that the approximation may be deficient (Gelman et al. (1995)). We also applied the method to one simulated data set where we knew the true frailties. We then estimated the frailties by the posterior mean and compared

them to the true values. Points in the scatter plot of estimated by true frailties lay close to the 45° line, suggesting the method is adequate.

We illustrate the result of the sampling for one subject in Figure 3.2. For each frailty, the solid line depicts its log marginal prior density, and the dashed line depicts a smoothed approximation to its log marginal posterior density.

3.4.3 Probabilities of Permanent Recovery and Lapse

We generated transition times after EOT conditional on the sampled frailties. An important feature of this process is the possibility of permanence for an episode. At a given time, a subject can be in one of four states: permanently recovered, temporarily recovered, temporarily lapsed and permanently lapsed. The probabilities of the two permanent states are of particular interest. From the real clinical data, it is impossible to tell whether a status is temporary or permanent, but in the simulated data we can distinguish them. We therefore evaluated each subject's status every two weeks after EOT and averaged them across 800 SIR samples to obtain estimated probabilities of permanent recovery and lapse.

Figure 3.3 shows these probabilities for three selected subjects in each arm. We note that because both permanent states are absorbing, the probabilities increase over time. Moreover subjects in the same arm can have quite different "cure" probabilities. For example, subject 1 in the bupropion arm transited 22 times between smoking and abstinence, with the final event being a lapse. Thus this subject struggled to make a sustained quit attempt, and eventual failure is the most likely outcome. Compare the

graph from subject 8, who never lapsed during the entire eight weeks of treatment and is very likely a permanent success. In the placebo group, subject 46 abstained for 10 days, relapsed for 33, then made a second quit attempt of 14 days that was censored at EOT. This subject therefore had a moderate probability of permanent recovery (0.4). By contrast, the situation for subject 15 is very clear — after a brief one day quit he smoked till EOT, and thus was almost certainly a failure.

Figure 3.4 shows the graphs averaged across subjects within treatment arms. The bupropion arm has a high probability of permanent recovery (0.63) and a relatively low probability of permanent lapse (0.36). These figures are reversed in the placebo arm (0.44 for permanent recovery; 0.55 for permanent lapse), reflecting a strong treatment effect. Note that because the probabilities for both arms sum nearly to 1, the model suggests strongly that the vast majority of subjects will have reached an absorbing state by the end of one year.

3.4.4 Predicting One-Year Outcomes

We predicted the probability of smoking at one year using samples from the predictive posterior distribution as described (the *SPP* method). Using the predictive probabilities and actual outcomes for all the subjects, we constructed the ROC curve and calculated its AUC. For comparison, we considered predictions from three logistic models: The first, *lr.frailty*, includes baseline covariates and individual estimated mean frailties from their posterior distribtuion; the second, *lr.baseline*, includes baseline covariates only; the third, *lr.eot*, includes baseline covariates and the smoking

status at EOT, defined as abstinence in the week leading up to EOT. As we indicated above, because we estimated these models using the actual year one outcome, we calculated a bootstrap-adjusted AUC to penalize for over-fitting.

Results appear in Figure 3.5. Model *lr.frailty* gives the best predictions (AUC = 0.825, adjusted AUC = 0.811), suggesting that the underlying unobserved individual frailties are helpful in predicting the outcome. The ROC of *lr.baseline* is only slightly above the diagonal (AUC = 0.540, adjusted AUC = 0.501), indicating that baseline covariates are useless for predicting outcomes. The methods *lr.eot* and *SPP* are much superior to *lr.baseline*, although slightly inferior to *lr.frailty*. Because the logistic regression models use the actual outcome to estimate parameters, whereas *SPP* uses only data from the treatment phase, *SPP* would have a substantial advantage in real applications. The logistic models would not be beneficial in practice unless one had a prior data set that gave valid parameter estimates, which in many contexts would be unattainable.

To investigate whether we can predict the outcome at an earlier point in the treatment period, we built two other cure-mixture frailty models using only 2-week and 4-week data, then generated the probability of smoking at one year post-TQD using the *SPP* approach. Figure 3.6 shows that the prediction using less information is not as good as that using the full treatment period, but still is far superior to predictions using baseline data only. Evidently the subjects already exhibit distinctive quit patterns by Week 4, so that we gain relatively little by observing them for four weeks more.

As our approach fixed the model parameters at the MLE $\hat{\xi}$, it failed to account

for uncertainty of estimation. To explore sensitivity to this feature of our method, in a separate analysis we sampled values of ξ from a multivariate t distribution with location $\hat{\xi}$, dispersion matrix $\widehat{\mathrm{Var}}(\hat{\xi})$, and degrees of freedom 5. For each draw, we sampled one complete set of trajectories. The predictive probabilities were very similar to those from the simpler version (Figure 3.7), and ROC curves mostly overlapped, with AUC changing by no more than 4%.

We note moreover that our model implicitly made the questionable assumption that the drug effect, observed only over eight weeks of treatment, persists 10 months more into the future. Therefore in another analysis, we coded drug as 0 after EOT for all the subjects, assuming that the drug has no lingering effect. In the resulting predictions, subjects in the placebo group had identical predictive probabilities as before, since they were not affected by assumptions about drug effect, but subjects in the drug group had slightly higher predictive probabilities (on average 0.06 increase). Nevertheless, the ROC curves were almost the same, with no change in AUC.

3.5 Discussion

We have described a prediction method based on a cure-mixture frailty model for smoking cessation trials. Although such models have been used in various contexts, most work has focused on making population inferences and little has been done on predicting individual future outcomes. Lack of software is a potential obstacle. For example, SAS Proc NLMIXED yields frailty estimates based on the posterior mode, but its method is applicable only in the case of normally distributed random

effects, and only the posterior mode is available rather than the complete posterior distribution.

Recently, Luo et al. (2009) proposed a Bayesian model for smoking behavior that predicts subject-specific probabilities using a discrete-time mixed model with a latent cured state. Our model differs from theirs in accommodating a more nearly continuous time scale and assuming a cure model for recovery as well as lapse. Moreover, they did not validate their predictions in a real data application or compare it to more conventional *ad hoc* methods, as we did. Our approach is also far less computationally intensive.

In our application, we chose importance sampling/SIR over other popular sampling methods such as Markov Chain Monte Carlo (MCMC). Our rationale is that importance sampling/SIR is simple to implement and avoids the issues of testing for convergence of the Markov chain. A potential drawback is its reliance on the choice of the approximating distribution — a poor choice can lead to inefficiency or even bias in posterior moments. To address this concern we evaluated several numerical diagnostics and also proved the validity of our approximate distribution analytically (see Appendix 3.6).

Our prediction methods perform reasonably well in terms of AUC, but sensitivity is generally not as good as specificity; that is, the predictions tend to under-estimate the probability of smoking at one year. One explanation is that the self-reported daily cigarette counts have inherent bias. It is not unusual for objective biochemical tests to contradict subject reports of abstinence (Gorber et al. (2009)). Therefore, our method potentially offers more accurate prediction if better data records are available.

3.6 Appendix: The Validity of the Approximate Distribution in Importance Sampling

We denote $w(b)$ as the weight function, which is the ratio of the true posterior distribution $p(b|y,\xi)$ and the approximate distribution $p^*(b)$. We will prove that $\mathrm{E}[w(b)]$ is finite and thus the approximation based on importance sampling is accurate (Geweke (1989)). We prove the result initially for $b^{(1)}$ marginally; the same proof applies to $b^{(2)}$.

The marginal posterior density of $b^{(1)}$ is:

$$p(b^{(1)}|y,\xi) = \frac{p(y|b^{(1)},\xi)f(b^{(1)}|\xi)}{\int p(y|b^{(1)},\xi)f(b^{(1)}|\xi)db_i} = k_1 \cdot p(y|b^{(1)},\xi)f(b^{(1)}|\xi), \tag{3.6.1}$$

where k_1 is a constant. Recall that we propose $f(b^{(1)}|\xi)$ as the approximate distribution, leading to the weight function

$$w(b^{(1)}) = \frac{p(b^{(1)}|y,\xi)}{f(b^{(1)}|\xi)} = k_1 \cdot p(y|b^{(1)},\xi), \tag{3.6.2}$$

and

$$\mathrm{E}[w(b^{(1)})] = \int_0^\infty w(b^{(1)})p(b^{(1)}|y,\xi)db^{(1)} = \int_0^\infty k_1^2 \cdot p^2(y|b^{(1)},\xi) \cdot f(b^{(1)}|\xi)db^{(1)}. \tag{3.6.3}$$

Since we are considering the function of the first frailty $b^{(1)}$ here, all the terms involving the second type of event will contribute only a constant, so that

$$\begin{aligned} p(y|b^{(1)},\xi) &= k_2 \cdot \prod_{j=1}^{n}\left(-\frac{\partial S_j^{(1)}}{\partial t_j^{(1)}}\right)^{d_j^{(1)}} \cdot \left(S_j^{(1)}\right)^{(1-d_j^{(1)})} \\ &= k_2 \cdot \prod_{j=1}^{n-1}\left(-\frac{\partial S_j^{(1)}}{\partial t_j^{(1)}}\right) \cdot S_n^{(1)}, \end{aligned} \tag{3.6.4}$$

where k_2 is a constant and the last equation holds because only the last event is censored. We have

$$\begin{aligned} -\frac{\partial S_j^{(1)}}{\partial t_j^{(1)}} &= (1-\pi_j^{(1)}) \cdot \exp[-b^{(1)} \exp(\eta_{\sigma_j}^{(1)}) t_j^{(1)}] \cdot b^{(1)} \exp(\eta_{\sigma_j}^{(1)}) t_j^{(1)} \\ &< \exp[-b^{(1)} \exp(\eta_{\sigma_j}^{(1)}) t_j^{(1)}] \cdot b^{(1)} \exp(\eta_{\sigma_j}^{(1)}) t_j^{(1)} \\ &= \exp(-b^{(1)} c_j) b^{(1)} c_j; \end{aligned} \tag{3.6.5}$$

and

$$\begin{aligned} S_n^{(1)} &= \pi_n^{(1)} + (1-\pi_n^{(1)}) \exp[-b^{(1)} \exp(\eta_{\sigma_n}^{(1)}) t_n^{(1)}] \\ &< \exp[-b^{(1)} \exp(\eta_{\sigma_n}^{(1)}) t_n^{(1)}] \\ &= \exp(-b^{(1)} c_n), \end{aligned} \tag{3.6.6}$$

where c_j, $j = 1, 2, \cdots, n$, are constants. Plugging (3.6.5) and (3.6.6) into (3.6.4) yields

$$\begin{aligned} p(y|b^{(1)}, \xi) &< k_2 \cdot \prod_{j=1}^{n-1} \exp(-b^{(1)} c_j) b^{(1)} c_j \cdot \exp(-b^{(1)} c_n) \\ &= k_2 \exp(-b^{(1)} \sum_{j=1}^{n} c_j) (b^{(1)})^{n-1} \prod_{j=1}^{n-1} c_j \\ &= k_3 \exp(-b^{(1)} k_4) (b^{(1)})^{k_5}, \end{aligned} \tag{3.6.7}$$

where k_3, k_4 and k_5 are positive constants. Therefore,

$$\begin{aligned} & k_1^2 \cdot p^2(y|b^{(1)}, \xi) \cdot f(b^{(1)}|\xi) \\ < \; & k_1^2 \cdot k_3^2 \exp(-2k_4 b^{(1)}) (b^{(1)})^{2k_5} \cdot \frac{(b^{(1)})^{1/\theta^{(1)}-1} \exp(-b^{(1)}/\theta^{(1)})}{\Gamma(1/\theta^{(1)}) (\theta^{(1)})^{1/\theta^{(1)}}} \\ = \; & \frac{k_1^2 \cdot k_3^2}{\Gamma(1/\theta^{(1)}) (\theta^{(1)})^{1/\theta^{(1)}}} \cdot \exp[-(2k_4 + 1/\theta^{(1)}) b^{(1)}] \cdot (b^{(1)})^{2k_5 + 1/\theta^{(1)} - 1} \\ = \; & k_6 \exp(-k_7 b^{(1)}) (b^{(1)})^{k_8 - 1}, \end{aligned} \tag{3.6.8}$$

where k_6, k_7 and k_8 are positive constants.

Note that $\exp(-k_7 b^{(1)})(b^{(1)})^{k_8-1}$ is the kernel of a gamma density, so the integral of this function over $[0, \infty)$ is finite, and accordingly

$$\begin{aligned} \mathrm{E}[w(b^{(1)})] &= \int_0^\infty k_1^2 \cdot p^2(y|b^{(1)}, \xi) \cdot f(b^{(1)}|\xi)\, db^{(1)} \quad \text{(by (12))} \\ &< \int_0^\infty k_6 \exp(-k_7 b^{(1)})(b^{(1)})^{k_8-1}\, db^{(1)} \quad \text{(by (17))} \\ &< \infty. \end{aligned} \tag{3.6.9}$$

Table 3.1: Parameter estimates in the selected cure-mixture frailty model.

	Estimate	SE	P-value	95% CI
Lapse				
Cure fraction				
Intercept	−1.185	0.322	<0.001	[−1.818, −0.552]
Drug	0.671	0.184	<0.001	[0.310, 1.032]
Sex: female	−0.646	0.187	<0.001	[−1.014, −0.278]
Episode number	0.078	0.091	0.389	[−0.100, 0.257]
Survival model				
Intercept	−2.561	0.130	<0.001	[−2.815, −2.306]
Drug	−0.609	0.120	<0.001	[−0.845, −0.373]
Sex: female	0.243	0.120	0.043	[0.008, 0.478]
Episode number	0.146	0.031	<0.001	[0.084, 0.207]
Variance	1.439	0.222	<0.001	[1.003, 1.876]
Recovery				
Cure fraction				
Intercept	−1.072	0.158	<0.001	[−1.382, −0.763]
Drug	−0.045	0.101	0.656	[−0.243, 0.153]
FTND	0.397	0.105	<0.001	[0.191, 0.602]
Episode number	0.116	0.039	0.003	[0.040, 0.193]
Survival model				
Intercept	−0.648	0.139	<0.001	[−0.920, −0.375]
Drug	0.128	0.099	0.195	[−0.067, 0.323]
FTND	−0.198	0.097	0.041	[−0.389, −0.008]
Episode number	−0.105	0.029	<0.001	[−0.161, −0.049]
Variance	0.407	0.077	<0.001	[0.256, 0.558]
Association	0.245	0.291	0.401	[−0.327, 0.816]

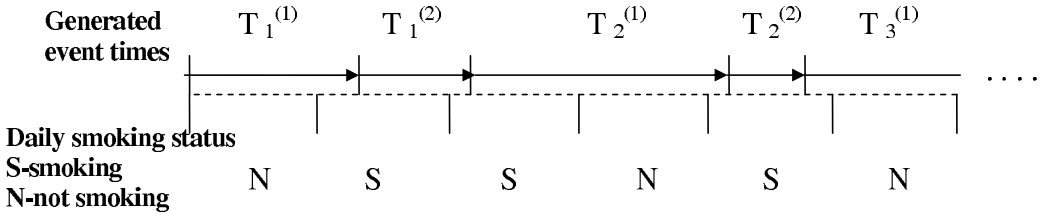

Figure 3.1: Reconstructing continuous time to daily smoking status.

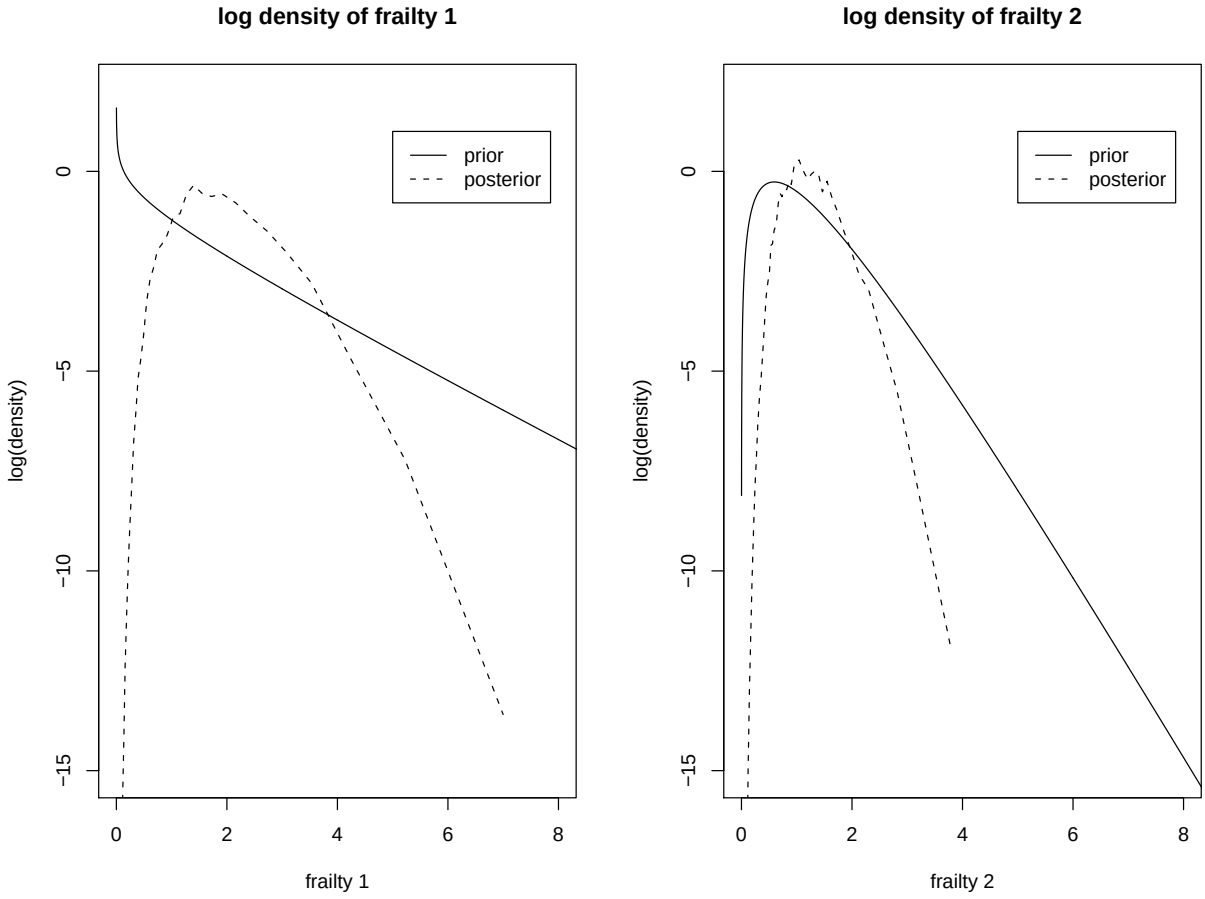

Figure 3.2: Frailty distributions for one subject.

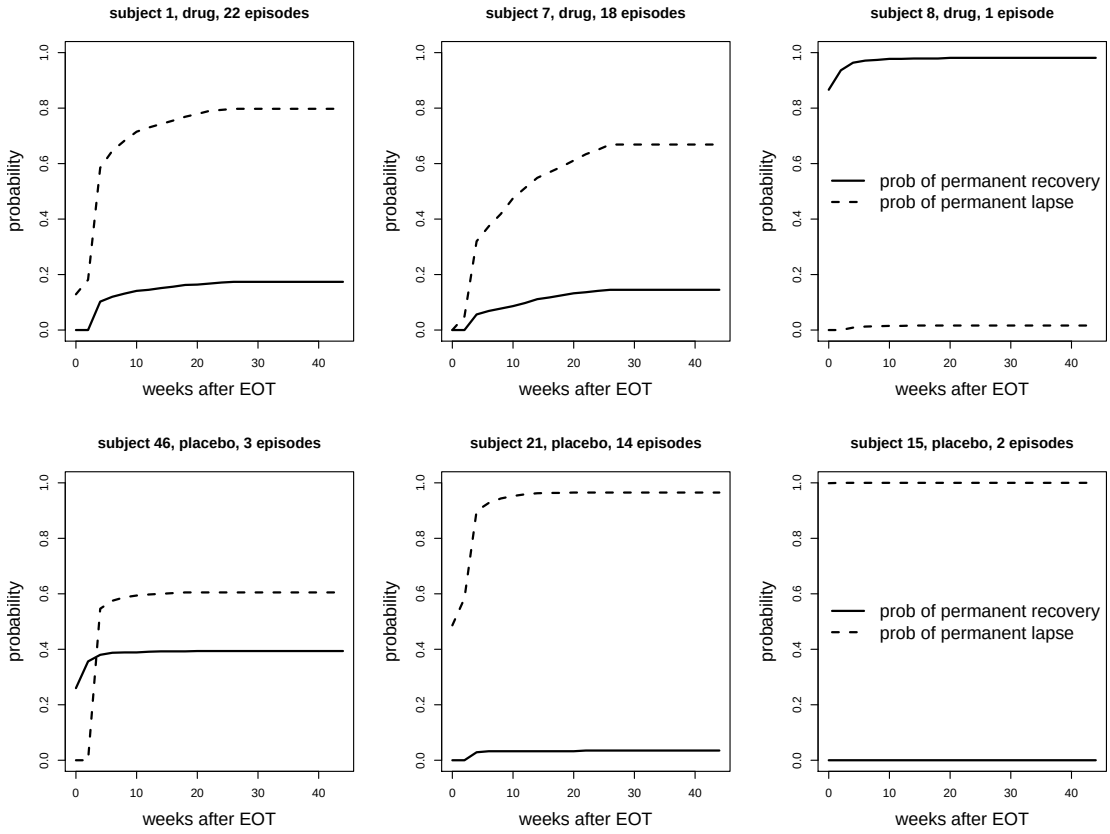

Figure 3.3: Probabilities of permanent recovery and lapse for different subjects. Solid line indicates probability of permanent recovery; dashed line indicates probability of permanent lapse.

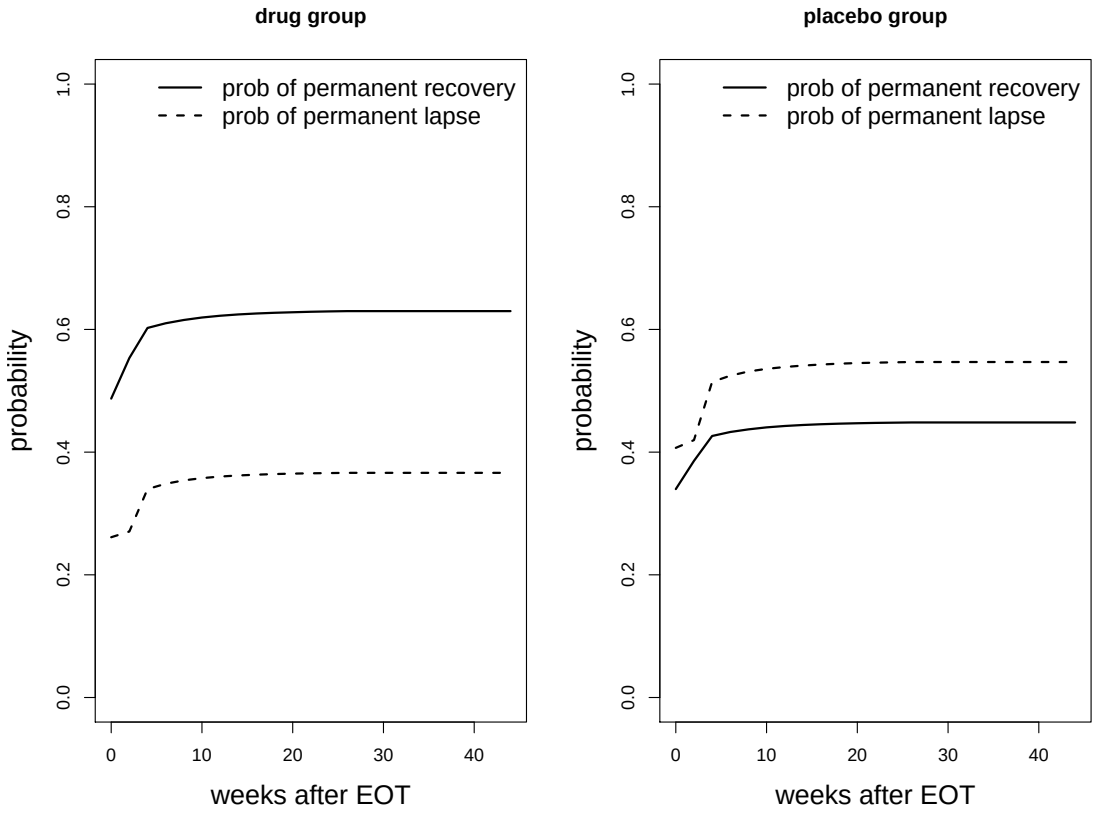

Figure 3.4: Probabilities of permanent recovery and lapse in two treatment arms. Solid line indicates probability of permanent recovery; dashed line indicates probability of permanent lapse.

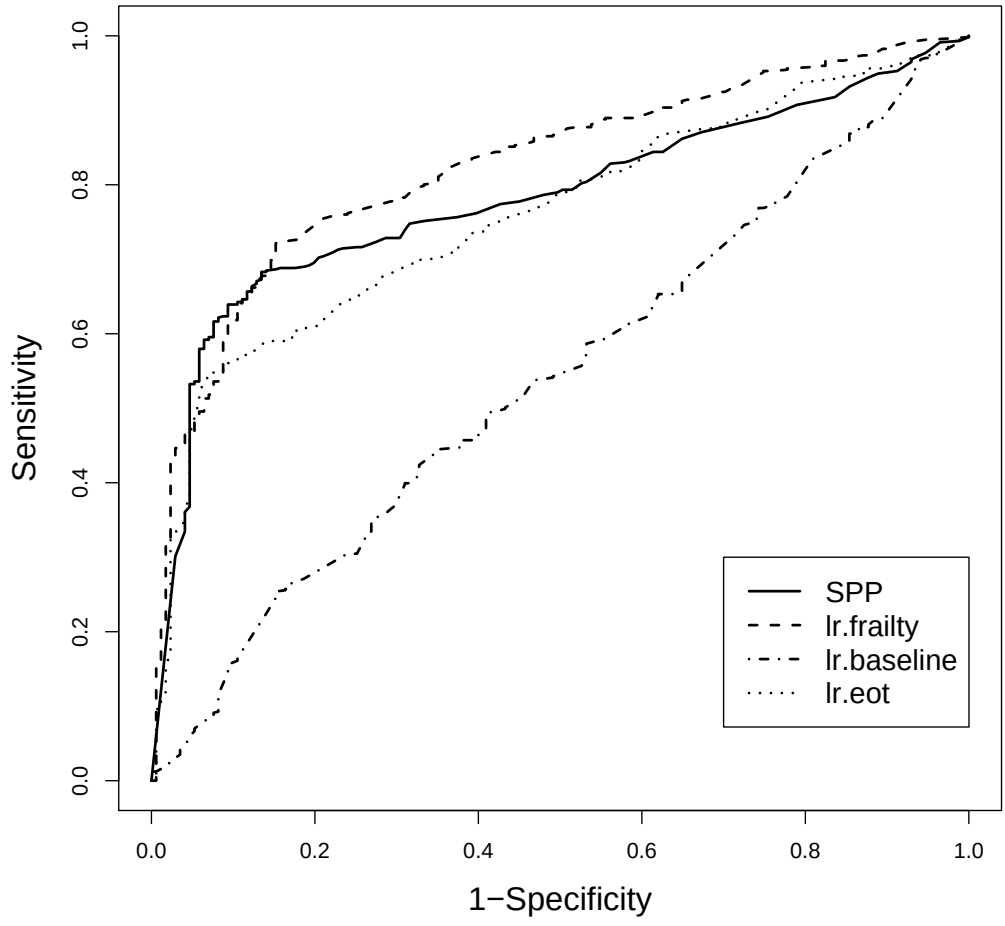

Figure 3.5: ROC curves for predictions using a variety of methods. SPP, AUC=0.782; lr.frailty, AUC=0.825, adjusted AUC=0.811; lr.baseline, AUC=0.540, adjusted AUC=0.501; lr.eot, AUC=0.759, adjusted AUC=0.738.

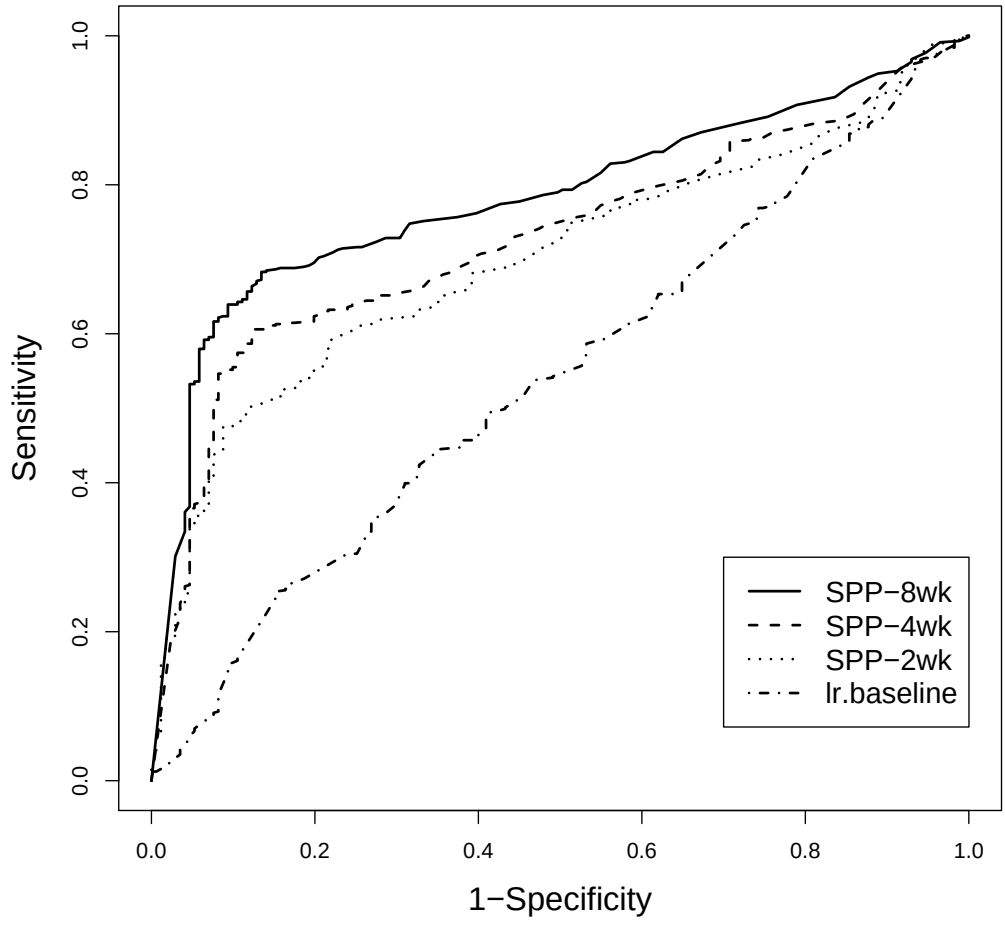

Figure 3.6: ROC curves for predictions using 2, 4, 8 week data. SPP-8wk, AUC=0.782; SPP-4wk, AUC=0.728; SPP-2wk, AUC=0.700; lr.baseline, AUC=0.540, adjusted AUC=0.501.

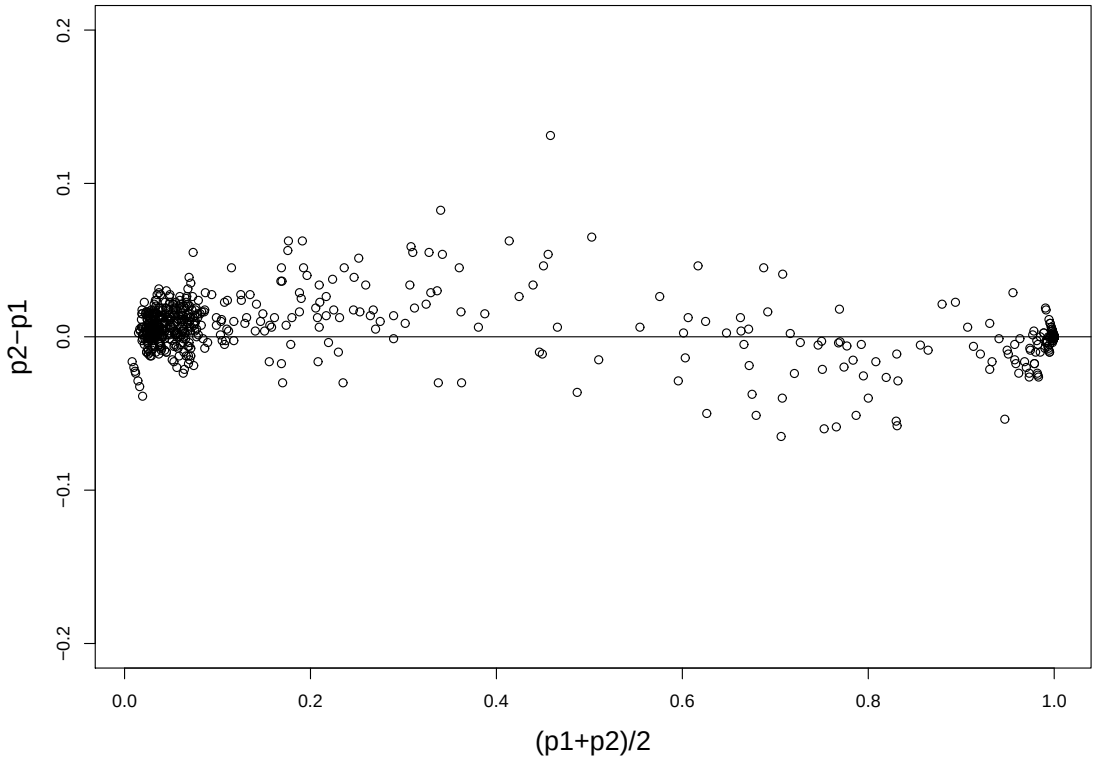

Figure 3.7: Predictive probabilities from two versions of sampling. p1: with model parameters fixed at the MLE; p2: with model parameters sampled from a multivariate t approximate posterior distribution.

Chapter 4

A Complementary Mixture Pareto II Distribution

4.1 Introduction

In a typical survival analysis, we assume that all subjects will eventually experience the event of interest if followed for a sufficiently long time. In some cases, however, it is preferable to entertain models that allow a fraction of subjects to be considered cured, or insusceptible, in the sense that their survival time for the event in question is essentially infinite. These "cure" models are useful in a number of applied contexts, including pediatric oncology, where a fraction of patients can be effectively cured of their disease in the sense that they will never experience a recurrence. Another useful modeling concept in survival analysis is the frailty model, which arises when one assumes that a single, latent random effect (or frailty) simultaneously affects an array of possible survival outcomes on a subject. A setting where both frailties

and cure models are relevant is smoking cessation trials, where subjects may make a series of quit attempts over a treatment period, each episode potentially resulting in a permanent success. Because the durations of an individual's repeated quit attempts are potentially correlated, one might also incorporate frailties in the modeling.

In Chapter 2, we proposed a novel cure-mixture frailty model that we applied to data from two smoking cessation clinical trials (Li et al. (2010)). Our research suggests that the distribution for survival given not cured arising under this model has not previously been described in the statistical literature. Therefore in this chapter we present a more detailed analysis of the distribution, which we denote the Complementary Mixture Pareto II distribution.

Let X denote an event time and $S(x)$ its survival function. We assume that with probability $1-\pi$ the event time is finite, so that $S(x)$ takes the form

$$S(x) = \pi + (1-\pi)S^*(x), \tag{4.1.1}$$

where $S^*(x)$ is a proper survival function. We assume moreover that, conditional on a latent random frailty γ, the cure probability and survival function given not cured take the following forms:

$$\pi(\gamma) = \exp(-\gamma e^{-\eta_\pi}), \quad S^*(x|\gamma) = \exp(-x\gamma e^{\eta_\sigma}), \tag{4.1.2}$$

where η_π is a linear predictor for the cure probability, and η_σ is a linear predictor for event rate given not cured. That is, we assume that conditionally on the frailty γ the probability of cure follows a generalized linear model with a complementary log-log (cloglog) link; that survival given not cured is exponential with a proportional hazard regression; and that the natural log of γ affects both linear predictors with coefficient

1. We moreover assume that the frailty γ represents a random draw from a gamma distribution with shape and scale both equal to $1/\theta$. If the cure probability were fixed at 0, this would be the standard model of a gamma-distributed exponential hazard (Duchateau and Janssen (2008)).

This cure-mixture frailty model is unique in two respects: First, it assumes the cloglog function instead of ubiquitous logistic link to model the cure probability; and second, the underlying frailty γ affects both the cure probability and the hazard given not cured. These assumptions lead to analytical simplification when we integrate out the frailty γ in (4.1.1) and (4.1.2) to obtain the marginal survival function

$$\widetilde{S}(x) = \int S(x)dF(\gamma) = \widetilde{\pi} + (1-\widetilde{\pi})\widetilde{S^*}(x), \tag{4.1.3}$$

where

$$\widetilde{\pi} = \left(\frac{1}{1+\theta e^{-\eta_\pi}}\right)^{1/\theta}, \tag{4.1.4}$$

and

$$\widetilde{S^*}(x) = \frac{(1+\theta x e^{\eta_\sigma})^{-1/\theta} - (1+\theta e^{-\eta_\pi}+\theta x e^{\eta_\sigma})^{-1/\theta}}{1-(1+\theta e^{-\eta_\pi})^{-1/\theta}}. \tag{4.1.5}$$

Note that $\widetilde{S^*}(x)$ is monotone decreasing with $\widetilde{S^*}(0) = 1$ and $\widetilde{S^*}(+\infty) = 0$, so that $\widetilde{S^*}(x)$ is a proper survival function. Therefore the marginal cure probability is $\widetilde{\pi}$ and the marginal survival function given not cured is $\widetilde{S^*}(x)$. The conditional cure probability and non-cured survival in (4.1.2) are defined at the individual level, whereas the marginal cure probability and non-cured survival in (4.1.4) and (4.1.5) represent average population quantities.

Because $\widetilde{S^*}(x)$ is a proper survival function, it is of interest to know whether it belongs to any existing family of survival distributions, and if not, to identify its

unique properties. The remainder of this chapter is devoted to an investigation of the characteristics of a random variable X following the survival function $\widetilde{S^*}(x)$. The discussion will invoke these densities:

Exponential (λ):

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0,\ \lambda \in (0, \infty); \tag{4.1.6}$$

Pareto I (α, σ):

$$f(x) = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{-(\alpha+1)}$$

$$x \geq \sigma,\ \alpha \in (0, \infty),\ \sigma \in (0, \infty); \tag{4.1.7}$$

Pareto II (μ, σ, ξ):

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-(1/\xi+1)}$$

$$x \geq \mu,\ \mu \in (-\infty, \infty),\ \sigma \in (0, \infty),\ \xi \in (0, \infty). \tag{4.1.8}$$

4.2 General Characteristics

For simplicity, we denote $a = e^{\eta_\sigma}$, $b = e^{\eta_\pi}$, $c = 1 - (1 + \theta/b)^{-1/\theta}$; then

$$\widetilde{S^*}(x) = \frac{1}{c}[(1 + \theta a x)^{-1/\theta} - (1 + \theta/b + \theta a x)^{-1/\theta}],$$

$$x \geq 0,\ a \in (0, \infty),\ b \in (0, \infty),\ \theta \in (0, \infty). \tag{4.2.1}$$

Here a is a parameter associated with the conditional hazard (conditional on the frailty) given not cured; a higher value of a implies a higher conditional hazard of experiencing the event. Parameter b is associated with the conditional cure probability, where a larger value implies a larger conditional cure probability. The term θ is the

variance of the frailty, and c is a normalizing constant. The corresponding marginal density is

$$\widetilde{f}^*(x) = \frac{a}{c}[(1+\theta ax)^{-1/\theta-1} - (1+\theta/b+\theta ax)^{-1/\theta-1}], \quad x \in (0,\infty), \tag{4.2.2}$$

and the hazard function is

$$\widetilde{h}^*(x) = a\frac{(1+\theta ax)^{-1/\theta-1} - (1+\theta/b+\theta ax)^{-1/\theta-1}}{(1+\theta ax)^{-1/\theta} - (1+\theta/b+\theta ax)^{-1/\theta}} \quad x \in (0,\infty). \tag{4.2.3}$$

We can rewrite (4.2.2) as

$$\begin{aligned} \widetilde{f}^*(x) &= \frac{1}{c}\cdot a(1+\theta ax)^{-\frac{1}{\theta}-1} - \frac{(1+\theta/b)^{-\frac{1}{\theta}}}{c}\cdot\frac{a}{1+\theta/b}\left(1+\frac{\theta a}{1+\theta/b}x\right)^{-\frac{1}{\theta}-1} \\ &= \alpha f_1(x) - \beta f_2(x), \end{aligned} \tag{4.2.4}$$

where

$$\alpha = \frac{1}{c},\ \beta = \frac{(1+\theta/b)^{-\frac{1}{\theta}}}{c};$$

$$f_1 \sim \text{Pareto II}\left(\mu=0, \sigma=\frac{1}{a}, \xi=\theta\right),$$

$$f_2 \sim \text{Pareto II}\left(\mu=0, \sigma=\frac{1+\theta/b}{a}, \xi=\theta\right).$$

and

$$\alpha > 1,\ \beta > 0,\ \alpha-\beta = 1.$$

So our density can be represented as a weighted sum of two Pareto II densities, with the difference of the weights equal to 1. We therefore denote it as the Complementary Mixture Pareto II distribution (CMPII).

Note that the first derivative of the density is

$$\begin{aligned} \widetilde{f}^{*\prime}(x) &= -\frac{a^2(1+\theta)}{c}\left[(1+\theta ax)^{-(\frac{1}{\theta}+2)} - \left(1+\frac{\theta}{b}+\theta ax\right)^{-(\frac{1}{\theta}+2)}\right] \\ &< 0, \quad \forall a,b,\theta,x, \end{aligned}$$

so that the density function is always decreasing with time x.

Figure 4.1 plots the log density and log hazard function at different values of a and θ when $b = 1$. Figure 4.2 depicts how the density and hazard function vary with b when a and θ are held fixed. Note that at any given time point, the hazard function increases with b and appears to converge to a distribution when $b \to \infty$. In fact,

$$\lim_{b\to\infty} \widetilde{S}^*(x) = \lim_{d\to 0} \frac{(1+\theta ax)^{-1/\theta} - (1+d\theta+\theta ax)^{-1/\theta}}{1-(1+d\theta)^{-1/\theta}} = \lim_{d\to 0} \frac{J_1}{J_2}.$$

Because $\lim_{d\to 0} J_1 = \lim_{d\to 0} J_2 = 0$, we have

$$\lim_{b\to\infty} \widetilde{S}^*(x) = \lim_{d\to 0} \frac{\partial J_1/\partial d}{\partial J_2/\partial d} = \lim_{d\to 0} \frac{(1+d\theta+\theta ax)^{-\frac{1}{\theta}-1}}{(1+d\theta)^{-\frac{1}{\theta}-1}} = (1+\theta ax)^{-\frac{1}{\theta}-1}, \qquad (4.2.5)$$

which is a Pareto II distribution ($\mu = 0$, $\sigma = \frac{1}{a(1+\theta)}$, $\xi = \frac{\theta}{1+\theta}$).

4.3 The Special Case of $b \to 0$

Letting $b \to 0$ in (4.2.1), the marginal non-cured survival function becomes

$$\lim_{b\to 0} \widetilde{S^*}(t) = (1+\theta at)^{-\frac{1}{\theta}}, \qquad (4.3.1)$$

which is a Pareto II distribution ($\mu = 0$, $\sigma = 1/a$, $\xi = \theta$). This is the same as the marginal survival function arising from the standard frailty model (without cure probability) with exponential baseline hazard (Duchateau and Janssen (2008)). This is because when $b \to 0$, the conditional cure probability goes to 0 in our cure-mixture frailty model (4.1.1), so that our model reduces to a standard frailty model with constant hazard conditional on the frailty. In this case, the density function is

$$\widetilde{f}^*(x) = a(1+\theta ax)^{-(\frac{1}{\theta}+1)}, \qquad (4.3.2)$$

with first derivative

$$\widetilde{f}^{*\prime}(x) = -a^2(1+\theta)(1+\theta a x)^{-(\frac{1}{\theta}+2)} < 0, \quad \forall a, \theta, x,$$

so that the density function is always decreasing with time x.

Figures 4.3 and 4.4 depict the log density and hazard functions at different parameter values of a and θ when $b \to 0$. We see that when θ is fixed and a decreases, the density function decreases at a slower rate, and the hazard is smaller at any given time. This is because a smaller value of a implies a smaller individual-level conditional hazard, which manifests itself as a smaller population-level marginal hazard. In particular, when $a \to 0$, the marginal hazard also goes to 0. When $a \to \infty$, the marginal non-cured survival becomes

$$\lim_{b\to 0, a\to\infty} \widetilde{S^*}(x) = (\theta a x)^{-\frac{1}{\theta}}, \tag{4.3.3}$$

which is a Pareto I distribution ($\alpha = 1/\theta$, $\sigma = 1/(\theta a)$).

Similarly, the density and hazard function vary with θ when a is fixed. It is interesting to note that the log density is a linear function of time and the hazard is constant when θ is close to 0. This is because $\theta \to 0$ means that the variance of the frailty also approaches 0, which implies that there is no between-subject heterogeneity, and therefore the population survival function equals the individual survival function, which is exponential with constant hazard. That is,

$$\lim_{b\to 0, \theta\to 0} \widetilde{S^*}(x) = e^{-ax}, \tag{4.3.4}$$

which is the exponential distribution with $\lambda = a$.

We summarize the relationships of CMPII with other known distributions in Figure 4.5.

4.4 Characteristic Function and Moments

We first derive the expectation and variance of the CMPII distribution.

Theorem 1. *The expectation of a variate X following the CMPII distribution is*

$$\mathrm{E}(X) = \frac{1}{a(1-\theta)} \cdot \frac{1-(1+\theta/b)^{-\frac{1}{\theta}+1}}{1-(1+\theta/b)^{-\frac{1}{\theta}}}, \quad \theta < 1.$$

The second moment is

$$\mathrm{E}(X^2) = \frac{2}{a^2(1-\theta)(1-2\theta)} \cdot \frac{1-(1+\theta/b)^{-\frac{1}{\theta}+2}}{1-(1+\theta/b)^{-\frac{1}{\theta}}}, \quad \theta < 1/2,$$

and therefore the variance is

$$\mathrm{Var}(X) = \frac{2}{a^2(1-\theta)(1-2\theta)} \cdot \frac{1-(1+\theta/b)^{-\frac{1}{\theta}+2}}{1-(1+\theta/b)^{-\frac{1}{\theta}}} - \frac{1}{a^2(1-\theta)^2} \cdot \left[\frac{1-(1+\theta/b)^{-\frac{1}{\theta}+1}}{1-(1+\theta/b)^{-\frac{1}{\theta}}}\right]^2, \quad \theta < 1/2.$$

See the Appendix 4.6.1 for a proof. We note that the second factor in the expectation formula is always smaller than 1, so that the expectation satisfies

$$\mathrm{E}(X) < \frac{1}{a(1-\theta)} = \mathrm{E}(X_1 \sim f_1). \tag{4.4.1}$$

We next derive the characteristic function $\mathrm{E}_X(e^{itx})$, where i is the square root of -1.

Theorem 2. *The characteristic function of the CMPII distribution is*

$$\phi(t) = \frac{1}{c\theta}\left(-\frac{it}{\theta a}\right)^{\frac{1}{\theta}} \left[\exp\left(-\frac{it}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it}{\theta a}\right) - \exp\left(-\frac{it(1+\theta/b)}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it(1+\theta/b)}{\theta a}\right)\right].$$

Here $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function, defined as

$$\Gamma(s,z) = \int_z^\infty t^{s-1}e^{-t}dt,$$

where s and z are complex parameters (Temme (1996) and Paris (2002)).

As is well known, the characteristic function can be used to calculate nth moments of a random variable X:

$$\mathrm{E}X^n = i^{-n}\frac{\partial\phi^n}{\partial t^n}\bigg|_{t=0}$$

Applying this formula we obtain the following result:

Theorem 3. *If X follows the CMPII distribution, then its nth moment is:*

$$\mathrm{E}X^n = \frac{n!}{ca^n(1-\theta)(1-2\theta)\cdots(1-n\theta)} - \frac{n!(1+\theta/b)^{n-\frac{1}{\theta}}}{ca^n(1-\theta)(1-2\theta)\cdots(1-n\theta)}, \quad (\theta < \frac{1}{n}).$$

One can verify that the first and second moments calculated in Theorem 1 agree with those dictated by Theorem 3.

4.5 Summary

We have described CMPII, a three-parameter survival distribution that we derived from a cure-mixture frailty model. One can view the CMPII as a variation of the Pareto family whose density arises as a linear combination of two Pareto II densities. The CMPII family includes as special cases several known distributions.

In Appendix 4.6, we provide the proofs of the Theorems in Section 4.4. The easiest way to derive the characteristic function and moments would seem to be deriving based on the formulas for the Pareto II distribution, similar to what we did

to derive expectation and variance (Appendix 4.6.1). However, we were unable to find these formulas for the Pareto II distribution in a literature search, and therefore we elected to derive them for the CMPII directly (see Appendices 4.6.2 and 4.6.3). The derivations are complicated but may serve as examples for similar derivations of other distributions.

Pareto distributions have been known for several decades (Arnold (1983); Arnold and Laguna (1977)) and have been applied in various areas of science, especially in economics to describe income distributions and in medical statistics to describe survival distributions. Naturally this new CMPII distribution can be applied to these areas as well. Our earlier work (Li et al. (2010)) is one possible use; other potential applications deserve further investigation.

4.6 Appendix: Proof the Theorems about CMPII Distribution

4.6.1 Proof of Theorem 1

In this proof, we will use some facts about Pareto II distribution (Johnson et al. (1994)): If a survival time X follows Pareto II (μ, σ, ξ), then

$$\mathrm{E}(X) = \frac{\sigma}{1-\xi}, \quad \xi < 1; \tag{4.6.1}$$

$$\mathrm{E}(X^2) = \frac{2\sigma^2}{(1-\xi)(1-2\xi)}, \quad \xi < 1/2. \tag{4.6.2}$$

If X follows the CMPII distribution, then its expectation is

$$\begin{aligned}
\mathrm{E}(X) &= \alpha \mathrm{E}(X_1 \sim f_1) - \beta \mathrm{E}(X_2 \sim f_2) \quad \text{(by equation (4.2.4))} \\
&= \frac{1}{c} \cdot \frac{1}{a(1-\theta)} - \frac{(1+\theta/b)^{-\frac{1}{\theta}}}{c} \cdot \frac{1+\theta/b}{a(1-\theta)} \quad \text{(by equation (4.6.1))} \\
&= \frac{1}{ca(1-\theta)} \left(1 - (1+\theta/b)^{-\frac{1}{\theta}+1} \right) \\
&= \frac{1}{a(1-\theta)} \cdot \frac{1-(1+\theta/b)^{-\frac{1}{\theta}+1}}{1-(1+\theta/b)^{-\frac{1}{\theta}}}, \quad \theta < 1. \qquad (4.6.3)
\end{aligned}$$

Similarly, the second moment of X is

$$\begin{aligned}
\mathrm{E}(X^2) &= \alpha \mathrm{E}(X_1^2, X_1 \sim f_1) - \beta \mathrm{E}(X_2^2, X_2 \sim f_2) \\
&= \frac{1}{c} \cdot \frac{2}{a^2(1-\theta)(1-2\theta)} - \frac{(1+\theta/b)^{-\frac{1}{\theta}}}{c} \cdot \frac{2(1+\theta/b)^2}{a^2(1-\theta)(1-2\theta)} \\
&= \frac{2}{ca^2(1-\theta)(1-2\theta)} \left(1 - (1+\theta/b)^{-\frac{1}{\theta}+2} \right) \\
&= \frac{2}{a^2(1-\theta)(1-2\theta)} \cdot \frac{1-(1+\theta/b)^{-\frac{1}{\theta}+2}}{1-(1+\theta/b)^{-\frac{1}{\theta}}}, \quad \theta < 1/2. \qquad (4.6.4)
\end{aligned}$$

Accordingly, the variance of X is

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \\
&= \frac{2}{a^2(1-\theta)(1-2\theta)} \cdot \frac{1-(1+\theta/b)^{-\frac{1}{\theta}+2}}{1-(1+\theta/b)^{-\frac{1}{\theta}}} \\
&\quad - \frac{1}{a^2(1-\theta)^2} \cdot \left[\frac{1-(1+\theta/b)^{-\frac{1}{\theta}+1}}{1-(1+\theta/b)^{-\frac{1}{\theta}}} \right]^2, \quad \theta < 1/2. \qquad (4.6.5)
\end{aligned}$$

4.6.2 Proof of Theorem 2

In this proof, we use the fact that

$$\int_0^\infty (x+w)^v e^{-ux} dx = u^{-v-1} e^{wu} \Gamma(v+1, wu); \qquad (4.6.6)$$

where u, v, w are complex numbers with constraints $|\arg w| < \pi$, Re $u > 0$; and $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function (Gradshteyn and Ryzhik (1980)).

Based on equation (4.2.4), we have

$$\phi(t) = \mathrm{E}(e^{itX}) = \alpha \mathrm{E}(e^{itX}, X \sim f_1) - \beta \mathrm{E}(e^{itX}, X \sim f_2). \tag{4.6.7}$$

Now

$$\begin{aligned}
\mathrm{E}(e^{itX}, X \sim f_1) &= a \int_0^\infty e^{itx}(1+\theta a x)^{-\frac{1}{\theta}-1} dx \\
&= a(\theta a)^{-\frac{1}{\theta}-1} \int_0^\infty \left(x + \frac{1}{\theta a}\right)^{-\frac{1}{\theta}-1} e^{itx} dx \\
&= a(\theta a)^{-\frac{1}{\theta}-1}(-it)^{\frac{1}{\theta}} \exp\left(-\frac{it}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it}{\theta a}\right), \quad \text{use (4.6.6)} \\
&= \frac{1}{\theta}\left(-\frac{it}{\theta a}\right)^{\frac{1}{\theta}} \exp\left(-\frac{it}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it}{\theta a}\right), \qquad (4.6.8)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(e^{itX}, X \sim f_2) &= \frac{a}{1+\theta/b} \int_0^\infty e^{itx}\left(1 + \frac{\theta a}{1+\theta/b} x\right)^{-\frac{1}{\theta}-1} dx \\
&= \frac{a}{1+\theta/b}\left(\frac{\theta a}{1+\theta/b}\right)^{-\frac{1}{\theta}-1} \int_0^\infty \left(x + \frac{1+\theta/b}{\theta a}\right)^{-\frac{1}{\theta}-1} e^{itx} dx \\
&= \frac{a}{1+\theta/b}\left(\frac{\theta a}{1+\theta/b}\right)^{-\frac{1}{\theta}-1} (-it)^{\frac{1}{\theta}} \\
&\quad \cdot \exp\left(-\frac{it(1+\theta/b)}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it(1+\theta/b)}{\theta a}\right), \quad \text{use (4.6.6)} \\
&= \frac{1}{\theta}\left(-\frac{it}{\theta a}\right)^{\frac{1}{\theta}} (1+\theta/b)^{\frac{1}{\theta}} \\
&\quad \cdot \exp\left(-\frac{it(1+\theta/b)}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it(1+\theta/b)}{\theta a}\right); \qquad (4.6.9)
\end{aligned}$$

Substituting (4.6.8) and (4.6.9) into (4.6.7) gives

$$\begin{aligned}
\phi(t) &= \frac{1}{c\theta}\left(-\frac{it}{\theta a}\right)^{\frac{1}{\theta}} \left[\exp\left(-\frac{it}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it}{\theta a}\right)\right. \\
&\quad \left. - \exp\left(-\frac{it(1+\theta/b)}{\theta a}\right)\Gamma\left(-\frac{1}{\theta}, -\frac{it(1+\theta/b)}{\theta a}\right)\right]. \qquad (4.6.10)
\end{aligned}$$

4.6.3 Proof of Theorem 3

We first note a series of lemmas.

Lemma 1.

$$\frac{\partial^n(z^{-s}\Gamma(s,z))}{\partial^n z} = (-1)^n z^{-s-n}\Gamma(s+n,z)$$

Proof. See Gradshteyn and Ryzhik (1980). □

Lemma 2.

$$-z^{-s}\Gamma(s,z)|_{z=0} = \frac{1}{s},\ (s<0).$$

Proof.

$$z^{-s}\Gamma(s,z) = z^{-s}\int_z^\infty t^{s-1}e^{-t}dt = \int_1^\infty y^{s-1}e^{zy}dy \quad (\text{let } y=\frac{t}{z}),$$

$$-z^{-s}\Gamma(s,z)|_{z=0} = \int_1^\infty y^{s-1}dy = -\frac{1}{s},\ (s<0).$$

□

Lemma 3. *Define* $g(z,s) = e^z z^{-s}\Gamma(s,z)$. *Then*

$$\frac{\partial g(z,s)}{\partial z} = -g(z,s+1) + g(z,s);$$

$$\left.\frac{\partial g(z,s)}{\partial z}\right|_{z=0} = \frac{-1}{s(s+1)},\ (s<-1).$$

Proof.

$$\begin{aligned}
\frac{\partial g(z,s)}{\partial z} &= \frac{\partial(e^z z^{-s}\Gamma(s,z))}{\partial z} = e^z\frac{\partial(z^{-s}\Gamma(s,z))}{\partial z} + e^z(z^{-s}\Gamma(s,z)) \\
&= e^z(-z^{-s-1}\Gamma(s+1,z)) + e^z(z^{-s}\Gamma(s,z)) \text{ (by Lemma 1)} \\
&= -g(z,s+1) + g(z,s);
\end{aligned}$$

$$\begin{aligned}\left.\frac{\partial g(z,s)}{\partial z}\right|_{z=0} &= -z^{-s-1}\Gamma(s+1,z)|_{z=0} + z^{-s}\Gamma(s,z)|_{z=0} \\ &= \frac{1}{s+1} - \frac{1}{s} \text{ (by Lemma 2)} \\ &= \frac{-1}{s(s+1)},\ (s<-1).\end{aligned}$$

□

Lemma 4.

$$\frac{\partial g^n(z,s)}{\partial z^n} = \sum_{k=0}^{n}(-1)^k\binom{n}{k}g(z,s+k).$$

Proof. Denoting $A_n = \partial g^n(z,s)/\partial z^n$, we prove the lemma by induction. Supposing

it is true for n, then for $n+1$,

$$\begin{aligned}
A_{n+1} &= \frac{\partial A_n}{\partial z} = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{\partial g(z,s+k)}{\partial z} \\
&= \sum_{k=0}^{n}(-1)^k\binom{n}{k}[g(z,s+k)-g(z,s+k+1)] \text{ (by Lemma 3)} \\
&= \sum_{k=0}^{n}(-1)^k\binom{n}{k}g(z,s+k)+\sum_{k=0}^{n}(-1)^{k+1}\binom{n}{k}g(z,s+k+1) \\
&= \sum_{k=0}^{n}(-1)^k\binom{n}{k}g(z,s+k)+\sum_{l=1}^{n+1}(-1)^l\binom{n}{l-1}g(z,s+l) \text{ (let } l=k+1) \\
&= g(z,s)+\sum_{k=1}^{n}(-1)^k\binom{n}{k}g(z,s+k) \\
&\quad +\sum_{l=1}^{n}(-1)^l\binom{n}{l-1}g(z,s+l)+(-1)^{n+1}g(z,s+n+1) \\
&= g(z,s)+\sum_{k=1}^{n}(-1)^k\binom{n}{k}g(z,s+k) \\
&\quad +\sum_{l=1}^{n}(-1)^l\binom{n}{l}\frac{l}{n-l+1}g(z,s+l)+(-1)^{n+1}g(z,s+n+1) \\
&= g(z,s)+\sum_{k=1}^{n}(-1)^k\binom{n}{k}\frac{n+1}{n-k+1}g(z,s+k)+(-1)^{n+1}g(z,s+n+1) \\
&= g(z,s)+\sum_{k=1}^{n}(-1)^k\binom{n+1}{k}g(z,s+k)+(-1)^{n+1}g(z,s+n+1) \\
&= \sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}g(z,s+k),
\end{aligned}$$

so it is true for $n+1$ as well. □

Lemma 5.

$$\left.\frac{\partial g^n(z,s)}{\partial z^n}\right|_{z=0} = \sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\frac{1}{(s+k)(s+k+1)},\ (s+n<0).$$

Proof.

$$\begin{aligned}
\left.\frac{\partial g^n(z,s)}{\partial z^n}\right|_{z=0} &= \left.\frac{\partial}{\partial z}\left(\frac{\partial g^{n-1}(z,s)}{\partial z^{n-1}}\right)\right|_{z=0} \\
&= \left.\frac{\partial}{\partial z}\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}g(z,s+k)\right|_{z=0} \text{ (by Lemma 4)} \\
&= \left.\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\frac{\partial g(z,s+k)}{\partial z}\right|_{z=0} \\
&= \sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\frac{-1}{(s+k)(s+k+1)} \text{ (by Lemma 3).}
\end{aligned}$$

□

Lemma 6.

$$\left.\frac{\partial g^n(z,s)}{\partial z^n}\right|_{z=0} = \frac{-n!}{s(s+1)\cdots(s+n)},\ (s+n<0).$$

Proof. Setting $B_n(s) = \partial g^n(z,s)/\partial z^n|_{z=0}$, we again prove the lemma using induction.

If it is true for n, then for $n+1$,

$$\begin{aligned}
B_{n+1}(s) &= \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{1}{(s+k)(s+k+1)} \text{ (by Lemma 5)} \\
&= \frac{1}{s(s+1)}+\sum_{k=1}^{n-1}(-1)^k\binom{n}{k}\frac{1}{(s+k)(s+k+1)}+(-1)^n\frac{1}{(s+n)(s+n+1)} \\
&= \frac{1}{s(s+1)}+\sum_{k=1}^{n-1}(-1)^k\left[\binom{n-1}{k}+\binom{n-1}{k}\frac{k}{n-k}\right]\frac{1}{(s+k)(s+k+1)} \\
&\quad +(-1)^n\frac{1}{(s+n)(s+n+1)} \\
&= \frac{1}{s(s+1)}+\sum_{k=1}^{n-1}(-1)^k\binom{n-1}{k}\frac{1}{(s+k)(s+k+1)} \\
&\quad +\sum_{k=1}^{n-1}(-1)^k\binom{n-1}{k}\frac{k}{n-k}\frac{1}{(s+k)(s+k+1)}+(-1)^n\frac{1}{(s+n)(s+n+1)} \\
&= B_n(s)+\sum_{k=1}^{n-1}(-1)^k\binom{n-1}{k-1}\frac{1}{(s+k)(s+k+1)}+(-1)^n\frac{1}{(s+n)(s+n+1)} \\
&= B_n(s)-\sum_{l=0}^{n-2}(-1)^l\binom{n-1}{l}\frac{1}{(s+l+1)(s+l+2)}+(-1)^n\frac{1}{(s+n)(s+n+1)} \\
&\quad (\text{let } l=k-1) \\
&= B_n(s)-\sum_{l=0}^{n-1}(-1)^l\binom{n-1}{l}\frac{1}{(s+l+1)(s+l+2)} \\
&\quad +(-1)^{n-1}\frac{1}{(s+n)(s+n+1)}+(-1)^n\frac{1}{(s+n)(s+n+1)} \\
&= B_n(s)-\sum_{l=0}^{n-1}(-1)^l\binom{n-1}{l}\frac{1}{(s+l+1)(s+l+2)} \\
&= B_n(s)-B_n(s+1) \\
&= \frac{-n!}{s(s+1)\cdots(s+n)}-\frac{-n!}{(s+1)(s+2)\cdots(s+n+1)} \text{ (because it is true for } n) \\
&= \frac{-(n+1)!}{s(s+1)\cdots(s+n+1)},
\end{aligned}$$

so it is true for $n+1$ as well. □

Based on these lemmas we can prove Theorem 3 easily. Denote $z_1=-(it)/(\theta a)$,

$z_2 = -[it(1+\theta/b)]/(\theta a)$ and $s = -1/\theta$, then

$$\phi(t) = \frac{\alpha}{\theta} z_1^{-s} e^{z_1} \Gamma(s, z_1) - \frac{\beta}{\theta} z_2^{-s} e^{z_2} \Gamma(s, z_2) = \frac{\alpha}{\theta} g(z_1, s) - \frac{\beta}{\theta} g(z_2, s),$$

$$\begin{aligned}
\left.\frac{\partial \phi^n}{\partial t^n}\right|_{t=0} &= \frac{\alpha}{\theta} \left.\frac{\partial g^n(z_1, s)}{\partial z_1^n}\right|_{z_1=0} \left(\frac{\partial z_1}{\partial t}\right)^n - \frac{\beta}{\theta} \left.\frac{\partial g^n(z_2, s)}{\partial z_2^n}\right|_{z_2=0} \left(\frac{\partial z_2}{\partial t}\right)^n \\
&= \frac{\alpha}{\theta} \frac{-n!}{s(s+1)\cdots(s+n)} \left(\frac{-i}{\theta a}\right)^n \\
&\quad -\frac{\beta}{\theta} \frac{-n!}{s(s+1)\cdots(s+n)} \left(\frac{-i(1+\theta/b)}{\theta a}\right)^n \text{ (by Lemma 6)} \\
&= \alpha \frac{i^n}{a^n} \frac{n!}{(1-\theta)(1-2\theta)\cdots(1-n\theta)} - \beta \frac{i^n}{a^n} (1+\theta/b)^n \frac{n!}{(1-\theta)(1-2\theta)\cdots(1-n\theta)},
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}X^n &= i^{-n} \left.\frac{\partial \phi^n}{\partial t^n}\right|_{t=0} \\
&= \alpha \frac{1}{a^n} \frac{n!}{(1-\theta)(1-2\theta)\cdots(1-n\theta)} - \beta \frac{1}{a^n} (1+\theta/b)^n \frac{n!}{(1-\theta)(1-2\theta)\cdots(1-n\theta)} \\
&= \frac{n!}{ca^n(1-\theta)(1-2\theta)\cdots(1-n\theta)} - \frac{n!(1+\theta/b)^{n-\frac{1}{\theta}}}{ca^n(1-\theta)(1-2\theta)\cdots(1-n\theta)}, \quad (\theta < \frac{1}{n}).
\end{aligned}$$

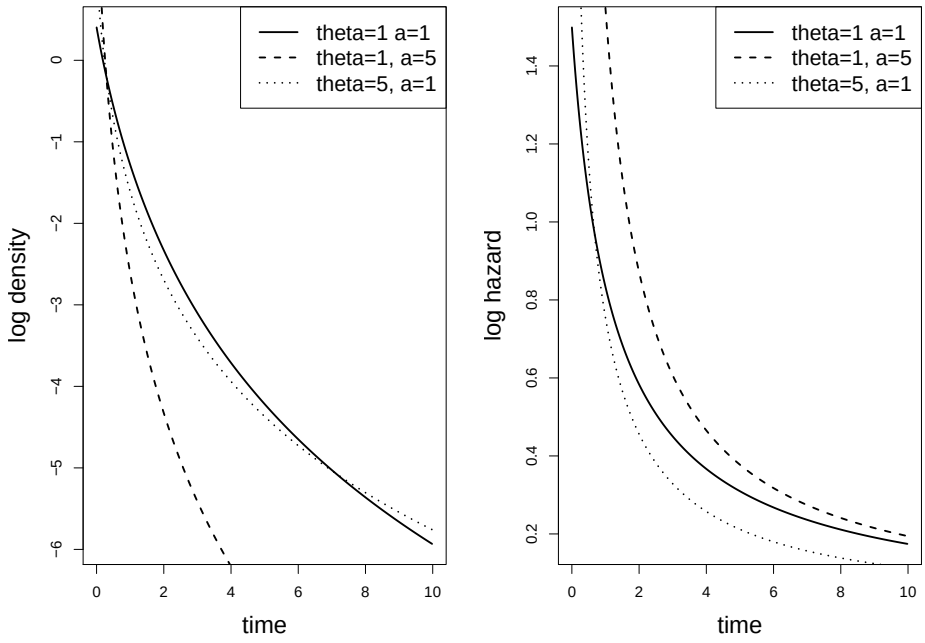

Figure 4.1: Log density and hazard function when $b = 1$.

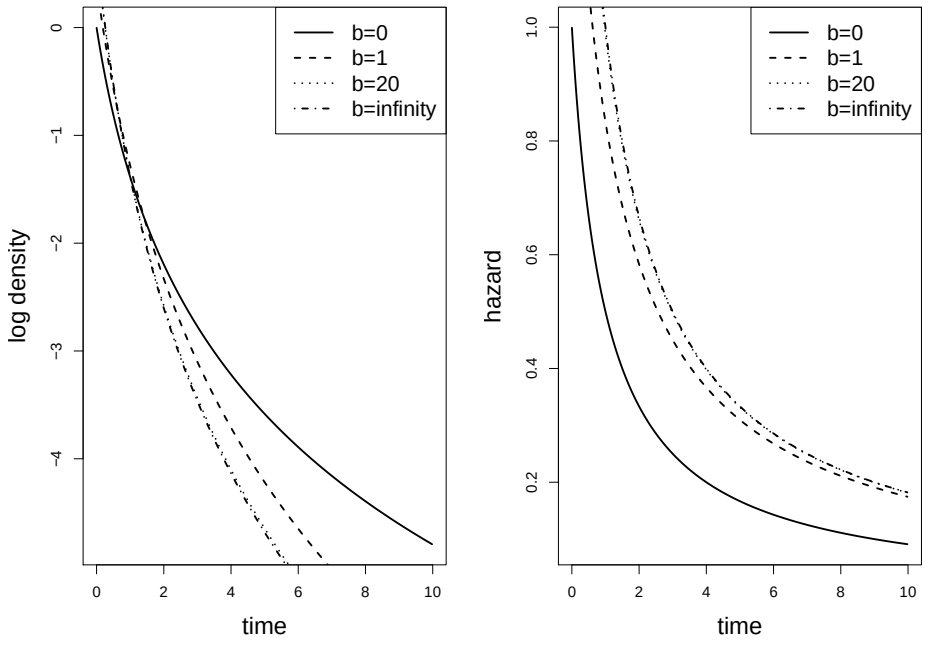

Figure 4.2: Log density and hazard function when $a = 1$ and $\theta = 1$.

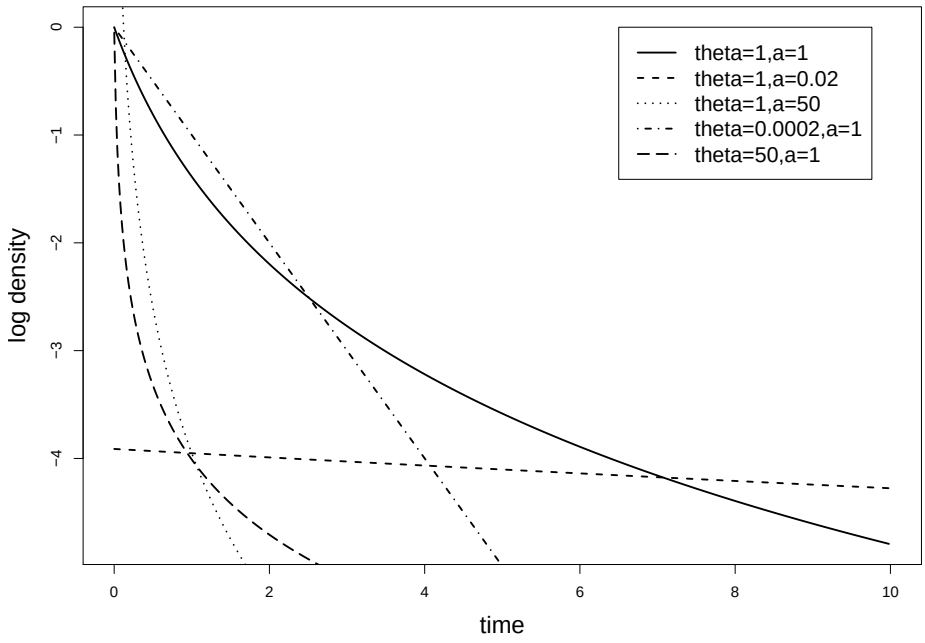

Figure 4.3: Log density function when b $\rightarrow$ 0.

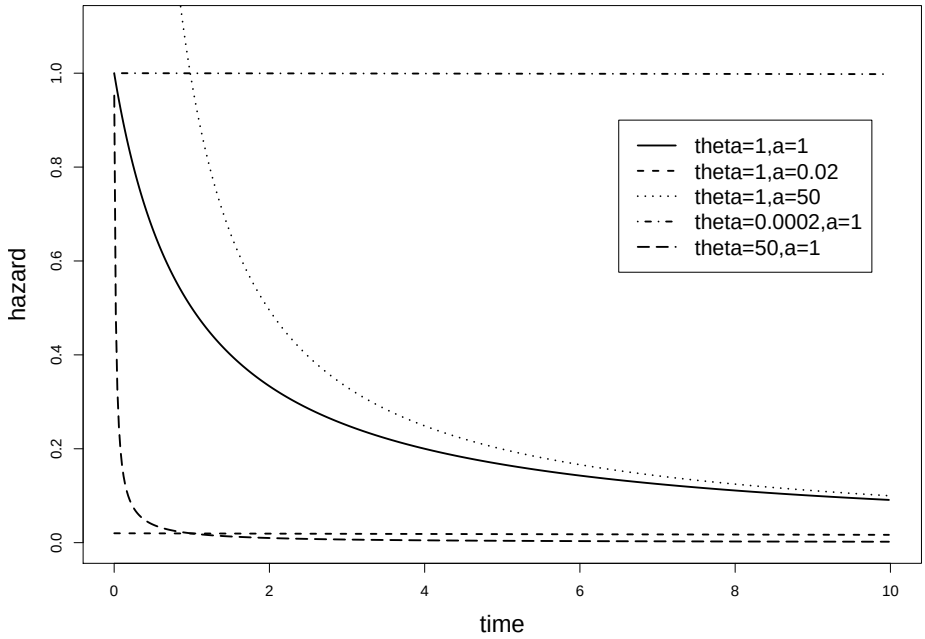

Figure 4.4: Hazard function when b $\rightarrow$ 0.

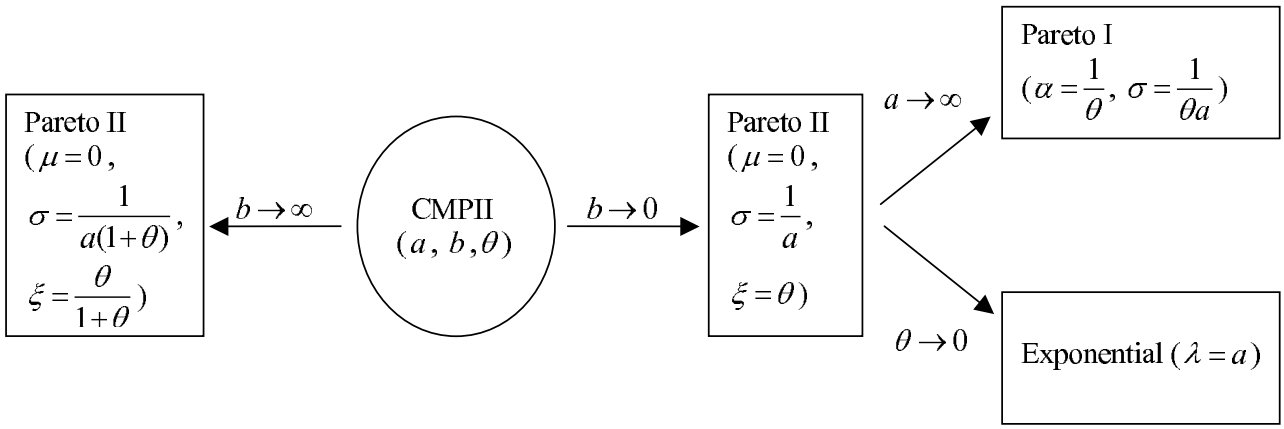

Figure 4.5: Relationship of CMPII distribution and other distributions.

Chapter 5

Conclusion

We have proposed a multivariate cure-mixture frailty model to describe the recurrent time-to-event data with alternating states and a possibility of cure, as the data from smoking cessation clinical trials. And based on it, we further developed a Bayesian method to predict individuals' long-term smoking behaviors from the short-term observed data.

Our model is superior to the traditional approach in several aspects. First, it makes efficient use of the behavioral information in the raw daily records. By analyzing the length of abstinent and smoking events, our model describes the complete process of transition between smoking states, not merely the length of the first attempt or smoking behavior in the neighborhood of EOT. Second, it provide more insights about the drug action than simply answering the question whether the drug is effective for assisting smoking cessation. By incorporating a cure probability, our model allows one to describe drug effects on both the daily hazard and the long-term cure probability; clinically the drug would be of greater interest if it could increase

the underlying cure probability besides temporarily delay the transition into a lapse. And by analyzing two types of events jointly, one could examine whether the drug is effective at inhibiting lapses, promoting recoveries, or both, and thus have a better understanding about the mechanism of the drug action.

Our model describes transition times between smoking and quitting, and an alternative way of representing the data is to model the binary daily smoking status. As subjects have repeated measurements over 56 days, one could construct a mixed-effects logistic regression with a random intercept at subject level. To reflect the potential dependence of the current smoking status on the history, one could introduce a time series correlation structure among the outcomes, or simply include the previous smoking status, or some summaries of the history data as predictors. And similar to what have done in Chapter 3, such binary models could also be utilized for prediction. Ideally by exploring all possible predictors, we wish to find a set of them that can completely explain the probability of smoking without the need of a random effect, so that one can predict the next outcome by evaluating the history data only.

Another variation of the current model would be to describe the daily cigarette consumption directly, rather than the lengths of abstinence and smoking events. This would answer a slightly different but equally important question: whether the drug would help reduce the intensity of the smoking rate. Similar to the idea of cure model with survival analysis, a count regression for such data needs to accommodate a possibility of true abstinence, reflected by excess zeros observed in the daily cigarette counts within a subject. Such methods are known as zero-inflated Poisson or negative binomial regression (Lambert (1992)). Moreover, as the cigarette counts are self-

reported and thus subject to "heaping" (Wang and Heitjan (2008)), the model may also need to incorporate the potential rounding errors.

Nevertheless, the methodologies presented in this dissertation provide a valuable tool for understanding the drug effects and individual behavior patterns in smoking cessation clinical trials. And although tailored for the smoking cessation data, the methods are potentially applicable in other studies that involve recurrent events and possibilities of cure.

Bibliography

Aalen, O.O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*; **7**: 1121–1137.

Aalen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*; **2**: 951–972.

Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions*. Dover: New York.

Andersen, P.K., Borgan, O., Grill, R.D. and Keilding N. (1993). *Statistical Models Based on Counting Process*. Springer- Verlag: New York.

Arnold, B.C. (1983). *Pareto Distributions*. International Cooperative Publishing House, Fairland, MD.

Arnold, B.C. and Laguna, L. (1977). *On Generalized Pareto Distributions with Applications to Income Data*. International Studies in Economics, Monograph #10, Department of Economics, Iowa State University, Ames, IA.

Banerjee, S. and Carlin, B.P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*; **60**: 268–275.

Brown, R.A., Niaura, R., Lloyd-Richardson, E.E., Strong, D.R., Kahler, C.W., Abrantes, A.M., Abrams, D. and Miller, I.W. (2007). Bupropion and cognitive-behavioral treatment for depression in smoking cessation. *Nicotine & Tobacco Research*; **9**: 721–730.

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

Chen, M.H., Ibrahim, J.G. and Sinha, D. (2002). Bayesian inference for multivariate survival data with a cure fracture. *Journal of Multivariate Analysis*; **80**: 101–126.

Chen, Y.Q., Wang, M.C. and Huang, Y. (2004). Semiparametric regression analysis on longitudinal pattern of recurrent gap times. *Biostatistics*; **5**:277–290.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*; **65**: 141–152.

Cook, R.J., Kalbfleisch, J.D. and Yi, G.Y. (2002). A generalized mover-stayer model for panel data. *Biostatistics*; **3**: 407–420.

Dale, L.C., Glover, E.D., Sachs, D.P., Schroeder, D.R., Offord, K.P., Croghan, I.T. and Hurt, R.D. (2001). Bupropion for smoking cessation: Predictors of successful outcome. *Chest*; **119**: 1357–1364.

Dennis, J.E. and Schnabel, R.B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall.

DiDonato, A.R. and Morris, A.H. (1986). Computation of the incomplete gamma function ratios and their inverse. *ACM Transactions on Mathematical Software*; **12**: 377–393.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer: New York.

Efron, B. and Tibshirani, R.J. (1983). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Farewell, V.T. (1983). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*; **38**: 1041–1046.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit Lett*; **27**: 861–874.

Gelman, A., Carlin, J., Stern, H. and Rubin, D.B. (1995). *Bayesian Data Analysis*. Chapman & Hall: London.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*; **57**: 1317–1339.

Gorber, S.C., Schofield-Hurwitz, S., Hardt, J., Levasseur, G. and Tremblay, M. (2009). The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine Tob Res*; **11(1)**: 12–24.

Gradshteyn, I.S. and Ryzhik, I.M. (1980). *Tables of Integrals, Series, and Products*. Acedemic Press, New York.

Hanley, J.A. and Mcneil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*; **143**: 29–36.

Harrell, F.E., Lee, K.L. and Mark, D.B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*; **15**: 361–387.

Heatherton, T.F., Kozlowski, L.T., Frecher, R.C. and Fagerström, K.O. (1991). The Fagerström Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. *Br J Addict*; **86**: 1119–1127.

Heitjan, D.F. and Rubin, D.B. (1991). Ignorability and coarse data. *The Annals of Statistics*; **19**: 2244–2253.

Heitjan, D.F. (1993). Ignorability and coarse data: Some biomedical examples. *Biometrics*; **49**: 1099–1109.

Heitjan, D.F. (1994). Ignorability in general incomplete-data models. *Biometrika*; **81**: 701–708.

Hosmer, D.W. and Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley & Sons: New York.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*; **71**: 75–83.

Hougaard, P. (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics*; **50**: 1178–1188.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag: New York.

Huang, Y. and Chen, Y.Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Analysis*; **9**: 293–303.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Wiley: New York.

Kelly, P.J. and Lim, L.L. (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine*; **19**: 13–33.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*; **34**: 1–14.

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

Lerman, C., Roth, D., Kaufmann, V., Audrain, J., Hawk, L., Liu, A., Niaura, R. and Epstein, L. (2002). Mediating mechanisms for the impact of bupropion in smoking cessation treatment. *Drug and Alcohol Dependence*; **67**: 219–223.

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

Li, Y., Wileyto, E.P. and Heitjan, D.H. (2010). Modeling smoking cessation data with alternating states and a cure fraction using frailty models. *Statistics in Medicine*; **29**: 627–638.

Liu, L. and Huang, X. (2007). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* ; **27**: 2665–2683.

Liu, L. and Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* ; **27**: 3105–3124.

Longini, I.M. and Halloran, M.E. (1996). A frailty mixture model for estimating vaccine efficacy. *Applied Statistics*; **45**: 165–173.

Luo, S., Crainiceanu, C.M., Louis, T.A. and Chatterjee, N. (2008). Analysis of smoking cessation patterns using a stochastic mixed-effects model with a latent cured state. *Journal of the American Statistical Association*; **103**: 1002–1013.

Luo, S., Crainiceanu, C.M., Louis, T.A. and Chatterjee, N. (2009). Bayesian inference for smoking cessation with a latent cure state. *Biometrics*; **65**: 970–978.

Mannan, H.R. and Koval, J.J. (2003). Latent mixed Markov modelling of smoking transition using Monte Carlo bootstrapping. *Statistical Methods in Medical Research*; **12**: 125–246.

Nelson, K.P., Lipsitz, S.R., Fitzmaurice, G.M., Ibrahim, J., Parzen, M. and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* ; **15**: 39-57.

Paris, R.B. (2002). A uniform asymptotic expansion for the incomplete gamma function. *Journal of Computational and Applied Mathematics* **148**, 323–339.

Peng, Y., Dear, K.B.G., Denham, J.W. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*; **17**: 813–830.

Peng, Y. and Carriere, K.C. (2002). An empirical comparison of parametric and semiparametric cure models. *Biometrical Journal*; **44**: 1002–1014.

Price, D.L. and Manatunga, A.K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*; **20**: 1515–1527.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurements*; **1**: 385–401.

Rubin, D.B. (1987). A noniterative sampling/important resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *J Am Stat Assoc*; **82**: 543–546.

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *Am Stat*; **13**: 84–88.

Smith, O.E., Adelfang, S.I. and Tubbs, J.D. (1982). A bivariate Gamma probability distribution with application to gust model. NASA Technical Memorandum 82483, National Aeronautics and Space Administration.

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

Tanner, M.A. (1993). *Tools for Statistical Inference* (2nd edn). Springer: Berlin, New York, NY.

Temme, N.M. (1996). Uniform asymptotics for the incomplete gamma functions starting from negative values of the parameters. *Methods and Applications of Analysis*; **3**, 335–344.

Therneau, T.M. and Hamilton, S.A. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine*; **16**: 2029–2047.

U.S. Department of Health and Human Services (1990). *The Health Benefits of Smoking Cessation*. USDHHS, PHS, CDC, Center for Chronic Disease Prevention and Health Promotion, Ofiice on Smoking and Health; DHHS Publication No. (CDC) 90-8416.

Vaupel, J.W., Manton, K.G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography*; **16**: 439–454.

Wang, H. and Heitjan, D.F. (2008). Modeling heaping in self-reported cigarette counts. *Statistics in Medicine*; **27**: 3789–3804.

Wileyto, E.P., Patterson, F., Niaura, R., Epstein, L.H., Brown, R.A., Audrain-McGovern, J., Hawk, L.W. and Lerman, C. (2005). Recurrent event analysis of laps and covery in a smoking cessation clinical trial using bupropion. *Nicotine & Tobacco Research*; **7(2)**: 257–268.

Williams, R.L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*; **56**: 645–646.

Withers, H.R., Peters, L.J., Taylor, J.M., Owen, J.B., Morrison, W.H., Schultheiss, T.E., Keane, T., O'Sullivan, B., Van Dyk, J. and Gupta, N. (1995). Local control of carcinoma of the tonsil by radiation therapy: An analysis of patterns of fractionation in nine institutions. *International Journal of Radiation Oncology, Biology, Physics*; **33**: 548–562.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*; **21**: Issue 4.

Yin, G. (2005). Bayesian cure rate frailty modes with application to a root canal therapy study. *Biometrics*; **61**: 552–558.

Yin, G. and Ibrahim, J.G. (2005). A class of Bayesian shared gamma frailty models with multivariate failure time data. *Biometrics*; **61**: 208–216.

Yu, B. and Peng, Y. (2008). Mixture cure models for multivariate survival data. *Computational Statistics & Data Analysis*; **52**: 1524–1532.