INFERENCE ON COMPETING RISKS IN BREAST CANCER DATA

by

Sarah R. Haile

BA, The College of Wooster, 2003

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Sarah R. Haile

It was defended on

July 22, 2008

and approved by

Thesis Advisor:
Jong-Hyeon Jeong, PhD
Associate Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Joseph P. Costantino, DrPH
Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Abdus S. Wahed, PhD
Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
(Joyce) Chung-Chou Ho Chang, PhD
Associate Professor
School of Medicine
University of Pittsburgh

Jong-Hyeon Jeong, PhD

INFERENCE ON COMPETING RISKS IN BREAST CANCER DATA

Sarah R. Haile, PhD

University of Pittsburgh, 2008

While nonparametric methods have been well established for inference on competing risks data, parametric methods for such data have not been developed as much. Because the cumulative incidence functions are improper by their nature, flexible distribution families accommodating improperness are needed for modeling competing data more accurately. Additionally, different types of events present in a competing risks setting may be correlated, yet current inference methods do not permit inferring such data taking into account the correlation between failure time distributions. This work first presents two new distributions which are well-suited for modeling competing risks data. In existing inference procedures for competing risks data, it appears that the correlation between failure time distributions of competing events are fixed as a constant. In the second part of this dissertation, a novel approach is proposed which allows researchers to model competing risks data by taking the correlation into account by estimating it. The methods are illustrated by analyzing survival data from a breast cancer trial of the National Surgical Adjuvant Breast and Bowel Project. Simulation studies are also presented for each of the proposed new distributions.

Public Health Significance: Competing risks occur often in many clinical studies, and must be accounted for whenever researchers are interested in only one type of event. For example, researchers may be interested in investigating only local recurrences of breast cancer, but must also take into account all other possible types of events as competing. Parametric methods are not currently as well established as other methods for competing risks data. Development of flexible parametric inference procedures suitable for modeling competing risks data would provide more accurate information, which will serve to improve patient care in clinical settings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Jong-Hyeon Jeong, for his help and support throughout this process. I would also like to thank those I have worked with at the National Surgical Adjuvant Breast and Bowel Project: Joseph Costantino, Greg Yothers, and Stephanie Land. The experience I gained while working at the NSABP has been invaluable in the completion of this work, and will continue to benefit me in the future.

I have greatly enjoyed my time in Pittsburgh, and must thank my friends and colleagues (biostatistics, and otherwise) for the good times we have had, and the support that only other graduate students could provide. My family deserves special thanks for their love and unfailing support. Finally, I must thank my fiancé, Martin Stolle, whose love and support has enabled me to bring this work to fruition, and has made my time in Pittsburgh very special. I will be forever grateful.

1.0 INTRODUCTION

In most settings where survival analysis is used, it is assumed that there is only one event of interest. However, in many situations, investigators may wish to infer survival or hazard rates for events of interest under two or more competing types. Since generally only the event that occurred first is analyzed, the occurrence of either event precludes the observation of other events. For the case where two competing events are present, the cumulative probability of events of primary interest differs from the same quantity in situations without competing events. In particular, the cumulative incidence function have to been used to correctly estimate the cumulative probability.

Parametric methods are more accurately able to predict future behavior of a situation, if the model has been correctly specified, than the corresponding nonparametric procedures. While nonparametric methods have been well-defined for competing risks data (Korn and Dorey, 1992; Pepe and Mori, 1993), parametric methods are not as well established. The nature of the cumulative incidence function being improper indicate that new distributions may need to be developed for more flexible yet precise models. As the hazard function may take on a number of different shapes which describe the underlying processes of failure time distributions, it may be useful to develop various improper distributions which can capture a variety of different hazard shapes.

Additionally, the current methods of inference for competing risks data assume that cause-specific event times are independent. This is a strong assumption and may not be correct in practice. A direct approach was proposed by Jeong and Fine (2006), which assumed a specific dependence structure among cause-specific event times.

To this end, we propose two new distributions which are applicable to competing risks data and can capture a variety of hazard shapes, including unimodal. Additionally, we

formulate a competing risks model for joint inference on two correlated lifetimes.

First, new 3-parameter and 4-parameter Gompertz distributions are proposed, which are suitable to model improper distributions for competing risks data, and, more importantly, are able to capture both monotone and unimodal hazard shapes. Related regression models are developed which can incorporate covariates using a generalized odds rate transformation for the link function. In Chapter 3, we extend the common 2-parameter Gompertz distribution to a three parameter distribution, give some of its basic properties and show that they are regular in that its parameter space does not depend on the support of the distribution. In Chapter 4, the Gompertz distribution is extended to a four parameter one. For both new distributions, simulation results are presented to assess the properties of the maximum likelihood estimates of the parameters. Both models and their covariate extensions of each are applied to a dataset from a breast cancer trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP). Finally, we discuss benefits and limitations of the two models.

Second, a copula-based model for joint modeling on competing risks data is proposed. This method allows for parametric inference on competing risks data without the independence assumption between competing failure time distributions. Furthermore, using this method, one can make inference about the level of dependence between competing events. We apply the approach to data from a breast cancer trial of the NSABP and discuss the approach in comparison to previously-established inference methods.

In Chapter 2, we present background details pertinent to this work. Chapters 3 and 4 are devoted to each of the extended Gompertz distributions. The new approach to modeling competing risks data jointly is presented in Chapter 5. Chapter 6 concludes with a discussion of the proposed methods and possibilities for future work.

2.0 BACKGROUND

In this chapter, we present important background information relating to the rest of this work. Competing risks are defined and the related important quantities are discussed. Maximum likelihood estimation for competing risks data is presented, as well as nonparametric estimation the cumulative incidence function. Parametric methods for survival data are introduced, and the (2-parameter) Gompertz distribution is presented. A review of assumptions made in analysis of competing risks data is made. Finally, previous works relating to inference on competing risks are discussed.

2.1 THE CUMULATIVE INCIDENCE FUNCTION

In some experiments, a subject may fail due to one of K different causes, for some $K \geq 2$. For example, patients in a breast cancer trial may experience a recurrence of the original cancer or die while still in remission. Alternately, investigators may be interested in cause-specific mortality, e.g. death from heart disease, death from cancer, or death by some other cause. Since generally only the event that occurred first is of interest, the occurrence of either event precludes the observation of the other events.

It has been demonstrated (Korn and Dorey, 1992; Pepe and Mori, 1993; Gooley *et al.*, 1999) that the complement of the nonparametric Kaplan-Meier (1958) estimate (1-KM) does not yield an unbiased estimate of the cumulative probability of failure due to the cause-specific event of interest. This occurs because the Kaplan-Meier estimates depend on the hazard of failure from the event of interest alone, and fail to take into account other causes

of failure. The cumulative incidence function (CI), while similar to the 1-KM estimator, is an unbiased estimator of the same quantity where competing risks are present.

2.1.1 Important Quantities

We define two basic quantities to analyze competing risk data, the cause-specific hazard (denoted $h_k(t)$) and the cumulative incidence function ($F_k(t)$). The cause-specific hazard for event of type k is the probability that an event of type k occurs just after time t, given that the subject survived until time t, or had a prior event that was *not* of type k. The cumulative incidence function for event k, $F_k(t) = P\left(T \leq t, K = k\right)$, can be interpreted as the marginal cumulative probability of observing an event of type k by time t.

Let $T = \min(T_1, \ldots, T_K)$ be the time to the first event, and K be the total number of event types. We assume that no subject can experience two different types of events at the same time point. The cause-specific hazard rate for the kth event type can be defined as

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P\left(t < T < t + \Delta t, K = k | T \geq t\right)}{\Delta t}. \tag{2.1}$$

Thus, for small Δt, it can be noted that

$$\frac{P\left(t < T < t + \Delta t, K = k\right)}{\Delta t} \approx p(T \geq t) h_k(t). \tag{2.2}$$

Integrating both sides, the cumulative incidence function for the kth event can be calculated as

$$F_k(t) = \int_0^t S(u) dH_k(u). \tag{2.3}$$

It should be noted that the cumulative incidence function, $F_k(t)$, is improper since $\lim_{t \to \infty} F_k(t) = P\left(K = k\right)$ and $\sum_k F_k(t) = F(t)$. To characterize this behavior for each of the cumulative incidence functions, it is beneficial to look at $\lim_{t \to \infty}(1 - F_k(t))$, representing the proportion of individuals that will never experience an event of type k. This is often referred to as the cure fraction for an event type, and can be found explicitly using a parametric modeling approach.

2.1.2 Maximum Likelihood Estimation

Using the traditional approach to parametric inference, we assume that the competing events are uncorrelated. Therefore the likelihood (for $K = 2$) can be constructed as the product of two cause-specific likelihoods as

$$L_{CS} = \prod_{i=1}^{n} \left[f_1(t;\psi_1)^{\delta_{1i}} S_1(t;\psi_1)^{1-\delta_{1i}} f_2(t;\psi_2)^{\delta_{2i}} S_2(t;\psi_2)^{1-\delta_{2i}} \right], \tag{2.4}$$

for some vector of parameters, ϕ_k, for event type k. It is assumed in this method that all distributions are proper. This approach will be denoted "cause-specific" (CS) in later sections.

An alternative method was presented in Jeong and Fine (2006). This method, denoted the direct method (D), does not assume independence, but does require improper distributions. Here, the likelihood function for parametric inference can be constructed as:

$$L_D = \prod_{i=1}^{n} \left[\left\{ \prod_{k=1}^{K} f_k(t;\psi_k)^{\delta_{ki}} \right\} \left\{ 1 - \sum_{k=1}^{K} F_k(t;\psi_k) \right\}^{(1-\sum_{k=1}^{K}\delta_{ki})} \right], \tag{2.5}$$

for K distinct event types, vector of parameters ψ_k characterizing F_k, vector of covariates covariate $\mathbf{Z}_i = \mathbf{Z}_i$, and $f_k = dF_k/dt$. In the case where $K = 2$, then this reduces to

$$L_D = \prod_{i=1}^{n} f_1(t_i,\psi_1)^{\delta_{1i}} f_2(t_i,\psi_2)^{\delta_{2i}} \left\{1 - F_1(t_i,\psi_1) - F_2(t_i,\psi_2)\right\}^{1-\delta_{1i}-\delta_{2i}}. \tag{2.6}$$

Unlike the likelihood function from the standard methodology (CS), the expression (2.5) cannot be factored into the product of cause-specific functions. Therefore, misspecification of one of the cause-specific hazard functions may affect the estimates of the others (Jeong and Fine, 2006).

Taking the first derivative of $\log L_D$ with respect to $\psi = (\psi_1, \psi_2, \ldots, \psi_K)$ under the given model, the maximum likelihood estimates (MLEs), $\hat{\psi}_k$ may be determined using a numerical algorithm, such as the Newton-Raphson method. The maximum likelihood estimator of the cumulative incidence function is then $F_k(t, \hat{\psi}_k)$, by using the invariance property of MLEs.

We assume certain regularity conditions for the MLEs, including consistency and asymptotic normality. Thus, the observed information matrix can be derived by taking the second

derivatives of the log of the likelihood function with respect to ψ. Applying the delta method, we can find the variance of $F_k(t, \hat{\psi}_k)$ by evaluating the expression

$$\widehat{var}(F_k(t;\hat{\psi}_k)) = \left(\frac{\partial F_k(t;\psi_k)}{\partial \psi_k}\right)\Bigg|_{\psi_k=\hat{\psi}_k} \widehat{var}(\hat{\psi}_k) \left(\frac{\partial F_k(t;\psi_k)}{\partial \psi_k}\right)^T\Bigg|_{\psi_k=\hat{\psi}_k}. \tag{2.7}$$

When $\psi_k = (\rho_k, \kappa_k)$, which characterizes the Gompertz distribution, $\partial F_k(t;\psi_k)/\partial\psi_k$ is equivalent to

$$\left(\frac{\partial F_k(t;\psi_k)}{\partial \rho_k}, \frac{\partial F_k(t;\psi_k)}{\partial \kappa_k}\right).$$

This method is also applicable when there are covariates involved.

The delta method can also be applied to estimate the variance of the estimated cure fraction (see Section 2.1.1), $\widehat{\text{var}}(CF)$. If the estimated value of the cure fraction is equal to $\phi(\psi)$, then

$$\widehat{\text{var}}(CF) = \left(\frac{\partial \phi(\psi)}{\partial \psi}\right)\Bigg|_{\psi_k=\hat{\psi}_k} \text{var}(\hat{\psi}) \left(\frac{\partial \phi(\psi)}{\partial \psi}\right)^T\Bigg|_{\psi_k=\hat{\psi}_k}. \tag{2.8}$$

2.1.3 Nonparametric Estimation of the CI Function

The cumulative incidence function can be estimated nonparametrically as

$$F_k(t) = \sum_{i=1}^{s} \hat{S}(t_i) d\hat{H}_k(t_i), \tag{2.9}$$

where $s = \max_i(t_i < t)$. Let n be the total number of patients on study and n_i the number of patients at risk beyond time t_i. Further, define e_i as the number of patients who fail due to the cause of interest at time t_i, r_i as the number who fail due to all other causes at time t_i, and c_i as the number censored at that time (using the notation of Gooley *et al.*, 1999). Then, $h_1(t_j)$, the hazard rate for events of type 1, can be estimated as $\frac{e_j}{n_{j-1}}$, the proportion of patients at risk at time t_j who fail due to the cause of interest at time t_j.

To estimate $S(t)$, the Kaplan-Meier estimate of the survival function (for any possible event) may be used:

$$KM(t) = \prod_{i=1}^{s} \left(1 - \frac{e_i}{n_{i-1}} - \frac{r_i}{n_{i-1}}\right). \tag{2.10}$$

So, the nonparametric estimator of the cumulative incidence function for the event of interest is given by

$$F_1(t) = \sum_{i=1}^{s} \frac{e_i}{n_{i-1}} \left\{ \prod_{i=1}^{s} \left(1 - \frac{e_i}{n_{i-1}} - \frac{r_i}{n_{i-1}} \right) \right\}. \tag{2.11}$$

Note that the complement of the Kaplan-Meier estimate for the event of interest would be

$$1 - KM_1(t) = 1 - \prod_{i=1}^{s} \left(1 - \frac{e_i}{n_{i-1}} \right),$$

which does not account for the hazard of failing due to any other causes, and therefore overestimates the failure rate of the event of interest in the presence of competing events.

2.2 PARAMETRIC INFERENCE

2.2.1 Distributions

Although nonparametric methods, such as that of Kaplan and Meier (1958), and semi-parametric methods, such as the proportional hazards model presented in Cox (1972), are used extensively in practice, there are many parametric models that are also applicable to survival data. Parametric modeling may result in estimates of the hazard function and other quantities that accurately fit the data, assuming that the model is correctly specified. Nonparametric modeling does not require assumptions to be made about the nature of the data, but, while it may closely fit the data, does not give the additional insight about data that parametric methods allow. There are several distributions that are commonly applied to survival data. Five such distributions are reviewed below.

1. The *exponential* distribution is characterized by the survival function,

$$S(t) = \exp(-\lambda t),$$

for $\lambda > 0$. The density function is $f(t) = \lambda \exp(-\lambda t)$. This distribution has a constant hazard function, $h(t) = \lambda$. The exponential distribution has the *memoryless* property, given by $P(T \geq t + s | T \geq t) = P(T \geq s)$.

2. The *Weibull* distribution, one of the more popular models for survival data, is characterized by the survival function,

$$S(t) = \exp\{-(\rho t)^{\kappa}\}.$$

In this model, ρ, the scale parameter, and κ, the index parameter, must both be greater than 0. The hazard function, $h(t) = \rho^{\kappa}\kappa t^{\kappa-1}$, is relatively flexible and allows for increasing, decreasing, or constant hazard shapes. Note that the exponential distribution is a special case of the Weibull, where $\kappa = 1$.

3. Mudholkar *et al.* (1996) presented a 3-parameter Weibull distribution characterized by the survival function

$$S(t) = \left\{1 - \lambda\left(\frac{t}{\sigma}\right)^{1/\alpha}\right\}^{1/\lambda},$$

for $\alpha, \sigma > 0$ and λ real and finite. The hazard function,

$$h(t) = \frac{(t/\sigma)^{\frac{1}{\alpha}-1}}{\alpha\sigma S(t)}$$

allows for a wide range of hazard shapes, including monotone increasing and decreasing, bathtub and unimodal. The regular Weibull distribution is a special case, where $\lambda \to \infty$.

4. The Gompertz (1825) distribution has a survival function,

$$S(t) = \exp\left[-\frac{\kappa}{\rho}(e^{\rho t} - 1)\right],$$

and a hazard function $h(t) = \kappa e^{\rho t}$. A more detailed discussion of this distribution can be found in Section 2.2.2.

5. A relatively new parametric model is that described in Jeong (2006). Related to the Weibull distribution, this parameterization of Hougaard's family of stable distributions (1986) is characterized by the survival function

$$S(t) = \exp\left[-\frac{\theta^{1-\alpha}\{(\rho t)^{\kappa} + \theta\}^{\alpha} - \theta}{\alpha}\right],$$

where, $(\rho t)^{\kappa}$ is the cumulative hazard function of the Weibull distribution ($\kappa > 0$ and $\rho > 0$), $\theta > 0$ and $-\infty < \alpha < \infty$. This model reduces to the Weibull distribution as

$\alpha \to 0$. Jeong's distribution is regular, in that the domain of t does not depend on α, and has a very flexible hazard function, given by

$$h(t) = \frac{\kappa(\rho t)^{\kappa}\theta^{1-\alpha}}{t\left\{(\rho t)^{\kappa} + \theta\right\}^{1-\alpha}}.$$

For the one-sample case with right censoring, the likelihood function for any parametric model can be constructed as

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}. \tag{2.12}$$

2.2.2 Gompertz Distribution

The Gompertz distribution was first proposed by Benjamin Gompertz in 1825. (Its statistical properties are also explained in some detail in Garg, Rao and Redmond (1970).) The Gompertz (G2) survival function can be written as

$$S_{G2}(t) = \exp\left\{-\frac{\kappa}{\rho}(e^{\rho t} - 1)\right\}, \tag{2.13}$$

with the hazard function given by $h_{G2}(t) = \kappa e^{\rho t}$. While κ must be positive, ρ may take on any finite real value. A plot of a possible hazard shape captured by this distribution is presented in Figure 1.

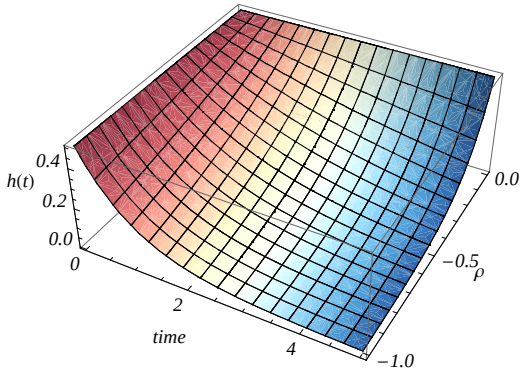


Figure 1: Plot of Gompertz Hazard Function with $\kappa = 0.5$

In the case where ρ is positive, the Gompertz distribution is proper; that is, its survival function ranges between 0 and 1. On the other hand, if $\rho < 0$, then this distribution is improper, with $\min S(t) = \exp(\kappa/\rho)$ as $t \to \infty$. As will be discussed in Section 2.1.1, this property makes the Gompertz distribution well-suited to modeling competing risks data.

2.2.3 Regression Models

Parametric models can be extended to include a number of time-independent covariates. The accelerated failure-time model (AFT) states that survival time for an individual at time t with a vector of covariates $\mathbf{Z}$ is the same as that of an individual with a baseline survival function at time $t \exp(\mathbf{Z}^T\beta)$, for β, a vector of regression coefficients. That is,

$$S(t|\mathbf{Z}) = S_0\left\{t \exp(\mathbf{Z}^T\beta)\right\}, \tag{2.14}$$

implying that $h(t|\mathbf{Z}) = \exp(\mathbf{Z}^T\beta)h_0[t \exp(\mathbf{Z}^T\beta)]$.

Alternatively, the linear model representation states that the relationship between the values of the covariates and survival time is linear on a log scale, that is, $\log(T) = u(T)+\mathbf{Z}^T\beta$. Here $u(T)$ is an invertible and monotonically increasing function, $\mathbf{Z}$ is a $p \times 1$ vector of time-independent covariates, and β is a $p \times 1$ regression parameter vector. This model can be generalized so that $\gamma(F(t|\mathbf{Z})) = u(t)+\mathbf{Z}^T\beta$ (Cheng *et al.*, 1995; Fine, 1999), for $\gamma(t)$, a known increasing function. If γ is the logit function, then a proportional odds model is used, and the regression parameters correspond to a log-odds ratio per unit increase in the covariates. If, however, γ is the c-log-log function $(\ln(-\ln(1-\nu)))$, then proportional hazards are assumed.

Under competing risks, the proportional hazards model may be considered to model the subdistribution hazard function of events of type k as presented in Fine and Gray (1999):

$$h_k^{CI}(t;\mathbf{Z}) = h_{k0}(t) \exp(\mathbf{Z}^T\beta_k), \tag{2.15}$$

where $h_k^{CI}(t;\mathbf{Z})$ is the cause-specific hazard function for the hazard of the kth event type. This cause-specific hazard function can be evaluated based on the improper random variable $T^* = I(K = k) \times T + I(K \neq k) \times \infty$, which is equivalent to T if an event of type k was observed for the subject, and ∞ otherwise (Gray, 1988). At $T^* = \infty$, a mass point exists, equal to $1 - F_k(\infty;\mathbf{Z})$, the "cure fraction" for the kth event.

Jeong and Fine (2007) proposed an extension of their direct parameterization method (Jeong and Fine, 2006) that can handle covariates. They employed the generalized odds

rate regression model of Dabrowska and Doksum (1988) that encompasses both the proportional hazards and proportional odds models as special cases. This transformation takes the following form:

$$g(\nu) = \log\left[\frac{(1-\nu)^{-\phi}-1}{\phi}\right], \tag{2.16}$$

where $F_k(t;\mathbf{Z})$ is the (improper) cumulative distribution function for events of type k, and $H_k(t)$ is the baseline cumulative hazard function.

Fine (1999) considered the transformation model (equation (2.16)) such that, for events of type k,

$$g_k(F_k(t;\mathbf{Z})) = \ln H_k(t) + \mathbf{Z}^T\beta_k. \tag{2.17}$$

If we solve this equation for F_k, we arrive at a cumulative incidence function for events of type k which accounts for the covariates $\mathbf{Z}$:

$$F_k(t;\mathbf{Z}) = 1 - \{1 + \phi_k \exp(\mathbf{Z}^T\beta_k)H_k(t)\}^{-1/\phi_k}. \tag{2.18}$$

When $\phi_k = 1$, the transformation assumes proportional odds. That is,

$$F_k(t;\mathbf{Z}) = \frac{e^{\mathbf{Z}^T\beta_k}H_k(t)}{1+e^{\mathbf{Z}^T\beta_k}H_k(t)}. \tag{2.19}$$

Additionally, proportional hazards are assumed as $\phi_k \to 0$:

$$F_k(t;\mathbf{Z}) = 1 - \exp\{-\exp(\mathbf{Z}^T\beta_k)H_k(t)\}. \tag{2.20}$$

With or without covariates, the maximum likelihood estimates of the parameters (and their standard errors) can be estimated using the techniques given in Section 2.1.2. Suppose that the (2-parameter) Gompertz distribution is assumed for the cumulative incidence function for event types 1 and 2. The likelihood function then involves four parameters, two for each event type, and all four are estimated simultaneously in the direct approach (Jeong and Fine, 2006). The cumulative incidence function can then be estimated by $F_k(t,\hat{\psi}_k)$ for events of type k $(k = 1, 2)$, where $\hat{\psi}_k$ denotes the MLEs for some vector of parameters ψ_k. Because the models for each event type are fitted jointly, model misspecification of one event type may affect the validity of the estimates for the other event type (Jeong and Fine, 2006). Therefore, there is a need to choose a distribution that is as flexible as possible for modeling competing risks data.

2.3 ASSUMPTIONS

Underlying this discussion of analysis of competing risks data is a number of important assumptions. First, we assume that the set of K types of competing events are mutually exclusive and exhaustive. A second assumption is that subjects can experience only one type of event at any particular time point. It is this second point that we would like to discuss in greater detail (Jeong, 2008).

Figure 2 shows the setup of a general competing risks problem. The failure times for the event of interest (T_1) are shown along the horizontal axis, while the failure times for other events (T_2) are shown along the vertical axis. We define T, the observed failure time as $T = \min(T_1, T_2)$. The times where $T_1 = T_2$ are represented by the diagonal line. The area above that line (with dashed lines) represents times when $T_1 < T_2$, or the event of interest is observed first, while the area below the line represents times when $T_1 > T_2$, or some other event is observed first.

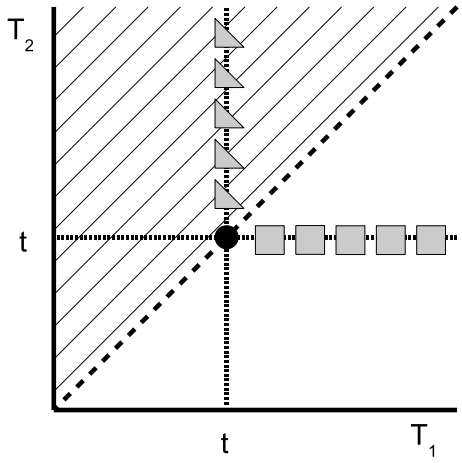


Figure 2: Schema of Competing Events

Suppose we want to determine the probability that $T = t$, that is, $Pr(T = t)$. This is equivalent to the probability that $\min(T_1, T_2) = t$. Therefore, this probability can be denoted

as:

$$\begin{aligned} Pr(T = t) &= Pr(\min(T_1, T_2) = t) \\ &= Pr(T_1 = t, T_2 \geq t) + Pr(T_1 \geq t, T_2 = t) - Pr(T_1 = t, T_2 = t). \quad (2.21) \end{aligned}$$

In the continuous case, we would make no distinction between $Pr(T > t)$ and $Pr(T \geq t)$, as both are considered equivalent to $\int_t^\infty f_T(x)dx$. If the failure times are discrete (which is a reasonable assumption), then the two quantities are only equivalent if

$$Pr(T_1 = T_2 = t) = 0,$$

and some care must be taken in calculating them. Notice that this situation is equivalent to the assumption that only one type of event may be observed on a subject at any particular time point.

Looking at Figure 2 again, suppose we are interested in estimating the probability that $T = t$. From equation (2.21), this is equivalent to $Pr(T_1 = t, T_2 \geq t)$ (shown with triangles and the circle in Figure 2) plus $Pr(T_1 \geq t, T_2 = t)$ (shown with squares and the circle) minus the probability $Pr(T_1 = t, T_2 = t)$ (only the circle). On the other hand, if we use $Pr(T_1 = t, T_2 \geq t) + Pr(T_1 \geq t, T_2 = t)$, then the probability $Pr(T_1 = T_2 = t)$ (in the circle) is not included. Thus, these two quantities are only equivalent if the probability that $T_1 = t$ and $T_2 = t$ is 0.

It is also clear from the form of the likelihood that we assume that a subject cannot experience two different events at the same time. Suppose that under the regular assumptions, we have the following likelihood:

$$L_R = \prod_{i=1}^{n} f_1(t_i)^{\delta_{1i}} f_2(t_i)^{\delta_{2i}} \left\{S(t_i, t_i)\right\}^{1-\delta_{1i}-\delta_{2i}}, \quad (2.22)$$

where $S(t_1, t_2)$ is equivalent to the probability $Pr(T_1 > t, T_2 > t)$, and

$$f_k(t_i) = \frac{-\partial}{\partial t_k} S(t_1, t_2)|_{T_1=T_2=t}.$$

In the case where a subject has an event of type k at time t_i (that is, $\delta_{ki} = 1$), then the contribution to the likelihood for that subject at time i is $f_k(t_i) = Pr(T_k = t_i)$. On the other hand, if a subject is censored at time t_i, then his contribution is $S(t_i, t_i) = Pr(T > t_i)$. Both

Table 1: Comparison of Event Indicators under Regular (δ) and Alternate (Δ) assumptions.

Event Observed	Regular	Alternate
Type 1	δ_1	$\Delta_1 = \delta_1(1-\delta_2)$
Type 2	δ_2	$\Delta_2 = \delta_2(1-\delta_1)$
Types 1 & 2	0	$\Delta_{12} = \delta_1\delta_2$
Neither 1 nor 2	$1-\delta_1-\delta_2$	$1-\Delta_1-\Delta_2-\Delta_{12}$
Assumptions	$\delta_1, \delta_2 \in \{0,1\}$	$\Delta_1, \Delta_2, \Delta_{12} \in \{0,1\}$
	$\delta_1\delta_2 = 0$	$(\delta_1, \delta_2, \delta_1\delta_2 \in \{0,1\})$

of these contributions have good interpretations. Suppose, however, that a subject experiences both types of events at time t_i. Then his contribution will be $f_1(t_i)f_2(t_i)S(t_i,t_i)^{-1}$, which has no logical interpretation.

Now, we suspend the assumption that a subject cannot experience two different types of events at the same time point. For this situation, we employ a different set of event type indicators. Recall that in the standard methodology, δ_k indicates only that an event of type k occurred, and makes no indication about the status of other events. We present an alternate notation to clarify the events which occur at a specific time point. Under this alternate assumption, we have Δ_k, which is defined in Table 2.3. Note that if $\Delta_{12} = 0$, which is the regular assumption, then $\delta_1\delta_2 = 0$ and $\delta_k = \Delta_k$.

Thus, under the alternate assumptions, we have a likelihood of the following form:

$$L_A = \prod_{i=1}^{n} f_1(t_i)^{\Delta_{1i}} f_2(t_i)^{\Delta_{2i}} f_{12}(t_i)^{\Delta_{12}} \left\{S(t_i,t_i)\right\}^{1-\Delta_{1i}-\Delta_{2i}-\Delta_{12i}}, \tag{2.23}$$

where

$$f_{12}(t_i) = \frac{\partial^2}{\partial t_1 \partial t_2} S(t_1,t_2)|_{T_1=T_2=t}.$$

Here, a subject contributes $f_k(t_i)$ to the likelihood if she experiences only an event of type k at time t_i. Similarly, she contributes $f_{12}(t_i) = Pr(T_1 = t_i, T_2 = t_i)$ if she experiences both types of events at time t_i. Finally, if she is censored at time t_i then the contribution

to the likelihood is $S(t_i, t_i)$. Thus, the two likelihoods are equivalent except both events are observed for a subject at a particular time. However, if it is assumed that the probability that both events are observed simultaneously, $Pr(T_1 = t, T_2 = t)$, is 0, then $\delta_1\delta_2$ is always equal to 0, and therefore the two likelihoods are equivalent.

The question then is: does it make a difference which set of assumptions are appropriate? In many situations, a subject never experiences two different types of events simultaneously. Thus, in these cases, the two likelihoods are equivalent and the more restrictive set of assumptions (regular) can be used. There may be cases, however, when the structure of the data is such that one or more subjects experiences more than one event at a time. In such cases, care must be taken to clarify the assumptions and use the correct form of the likelihood.

2.4 LITERATURE REVIEW

The study of competing risks traces back to a paper in 1959 by D. R. Cox. Using data that had previously been presented by Mendenhall & Hader (1958), Cox presented and discussed a number of models that could be used to analyze data where the failures were classified into two types. Additionally, he suggests the exponential and Weibull distributions for parametric models for fitting parametric forms to this kind of data. He also noted that as only $T = \min\{T_1, \ldots, T_K\}$ is observable, the marginal distributions are non-identifiable, as discussed in Tsiatis (1975).

Gail (1975) introduced notation for the competing risks setting and presented assumptions. In addition, Gail noted that the parameterization of $S(t)$ is crucial to the fit and interpretation of the model. Prentice *et al.* (1978) present an overview of competing risk methods up to that point, and present an alternative approach based on cause-specific hazard functions, shown to be a basic estimable quantity in this setting.

Larson and Dinse (1985) introduced a parametric mixture model framework, where the mixing parameters correspond to the marginal probability of failure for each type. Gray (1988) developed a class of nonparametric tests for comparing the cumulative incidence

among different groups. Benichou and Gail (1990) discussed fully parametric analysis of the cumulative incidence function, using exponential or piecewise exponential distributions.

Korn and Dorey (1992) discussed the Kaplan-Meier estimator (Kaplan and Meier, 1958), noting that 1-KM overestimates the true proportion of survival estimates in the presence of competing risks. Later, Pepe and Mori (1993) compared nonparametric methods for competing risks data. Bryant and Dignam (2004) proposed a semi-parametric method for cumulative incidence functions which was more efficient than related non-parametric methods.

Jeong (2006) presented a new parametric family for modeling competing risks data that is based on Hougaard's family of stable distributions, and had the generalized Weibull distribution as a special case. Jeong and Fine (2006) proposed a direct, rather than cause-specific, method for drawing inference in the presence of competing risks. Recently, Jeong and Fine (2007) proposed a method to incorporate covariates in such parametric inference.

3.0 AN EXTENSION OF THE GOMPERTZ DISTRIBUTION

In this chapter, we present the first of two extensions of the Gompertz distribution. Both distributions were developed to allow more flexible modeling of cause-specific cumulative incidence functions. Specifically, the Gompertz distribution (as presented previously in Section 2.2.2) cannot capture unimodal hazards. Each of models presented in this chapter represent a different approach to creating a more flexible Gompertz model. The new distribution developed in this chapter extends the Gompertz distribution from two to three parameters.

The first of extended models has three parameters, and is implied directly from the original Gompertz (G2) distribution. As noted earlier, the G2 model is characterized by the hazard function $h_{G2}(t) = \kappa e^{\rho t}$. The 3-parameter Gompertz distribution (G3) was developed so that $h_{G3}(t)$ is able to capture unimodal hazards, and contains $h_{G2}(t)$ as a special case.

3.1 THREE PARAMETER GOMPERTZ DISTRIBUTION

Suppose that $h_{G3}(t; \kappa, \rho, \eta)$ is of the form $\kappa e^{\rho t} e^{-\eta g(t)}$ for some function $g(t)$. Then the maximum value of $h_{G3}(t)$ over t can be found as the solution to $\rho - \eta g'(t) = 0$. We can then choose $g(t)$ so that $h_{G3}(t)$ is integrable. Here, we let $g(t) = e^{\rho t}$. This choice of g was made primarily so that $h(t)$ is easily integrable.

Thus, $h_{G3}(t; \kappa, \rho, \eta) = \kappa e^{\rho t - \eta e^{\rho t}}$. This hazard function leads to a survival distribution given by

$$S_{G3}(t; \kappa, \rho, \eta) = \exp\left\{-\frac{\kappa}{\rho\eta}\left(e^{-\eta} - e^{-\eta e^{\rho t}}\right)\right\}. \tag{3.1}$$

When modeling cause-specific cumulative incidence functions, F_k can be expressed as $1 - S(t; \rho_k, \kappa_k, \eta_k)$.

3.1.1 Properties

The G3 distribution is characterized by the hazard function $\kappa e^{\rho t - \eta e^{\rho t}}$. This implies that the regular Gompertz model is a special case of the G3 model, where $\eta = 0$. Thus, we can test whether the fit of G3 model is significantly better than that of the G2 by performing a Wald-type test of $H_o : \eta = 0$.

The parameter space for this distribution is as follows: $\kappa > 0$, and ρ and η both real and finite. In this parameter space, the distribution is regular; that is, the domain of t is not dependent on any of the parameters. A proof will be presented in Section 3.1.2.

An example of the G3 hazard function is shown in Figure 3, for parameters $\kappa = 0.5$, $\rho = -0.8$, and η in the interval $[-1, 2]$. Notice that at $\eta = 0$, the hazard is monotone, but that for positive values of η, it has a unimodal shape.

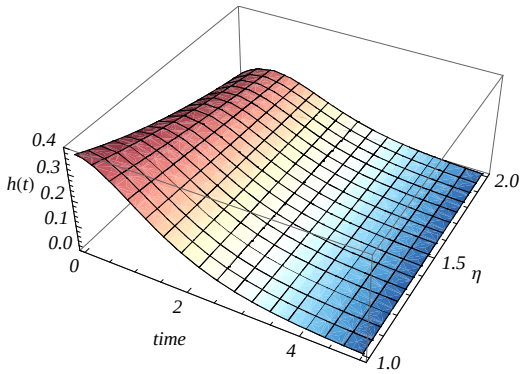


Figure 3: Plot of 3-parameter Gompertz Hazard Function with $\kappa = 1$ and $\rho = -1$

Given the relation noted in the previous section, we can calculate t^*, the value of t at which the hazard function reaches its maximum, as the solution to $\rho - \eta\rho e^{\rho t} = 0$. Thus, the hazard function for the 3-parameter Gompertz distribution occurs at

$$t^* = \frac{1}{\rho} \ln\left(\frac{1}{\eta}\right). \tag{3.2}$$

This distribution, like the G2 model, may be improper. For the G3 model, we can show that it is proper only where both $\rho > 0$ and $\eta < 0$. The cure fraction, defined as $\lim_{t\to\infty} S(t)$, is thus an important quantity here. This quantity can be interpreted as the proportion of subjects who never experience the event under consideration. For the 3-parameter Gompertz model, the cure fraction is equal to

$$\lim_{t\to\infty} S_{G3}(t) = \begin{cases} 0 & \text{if } \rho > 0 \text{ and } \eta < 0; \\ \exp\left(\frac{-\kappa}{\rho\eta} e^{-\eta}\right) & \text{if } \rho > 0 \text{ and } \eta > 0; \\ \exp\left\{\frac{-\kappa}{\rho\eta}(e^{-\eta} - 1)\right\} & \text{if } \rho < 0. \end{cases} \tag{3.3}$$

3.1.2 Regularity

We can show that the G3 model is regular in that $\psi_3 = \{\kappa, \rho, \eta\}$ are not boundary parameters. To prove this, we need only show that $S_{G3}(t; \psi_3)$ is in the interval $[0, 1]$ for all values of the parameter space. Since

$$S_{G3}(t; \kappa, \rho, \eta) = \exp\left\{-\frac{\kappa}{\rho\eta}\left(e^{-\eta} - e^{-\eta e^{\rho t}}\right)\right\},$$

this is equivalent to $\frac{\kappa}{\rho\eta}\left(e^{-\eta} - e^{-\eta e^{\rho t}}\right) > 0$. There are four cases, depending on whether ρ and η are positive or negative; we show only one here. In all cases, assume $\kappa > 0$.

First, note that $\frac{\kappa}{\rho\eta}\left(e^{-\eta} - e^{-\eta e^{\rho t}}\right)$ is equivalent to $\frac{\kappa}{\rho\eta} e^{-\eta}\left(1 - e^{\eta - \eta e^{\rho t}}\right)$. Suppose $\rho < 0$ and $\eta > 0$. Thus, $e^{-\eta}/\eta > 0$, and therefore, $1 - \exp(\eta - \eta e^{\rho t})$ is less than 0. This simplifies to $\eta(1 - e^{\rho t}) > 0$. Since η is assumed in this case to be positive, $1 - e^{-\rho t}$ must also be, which is true wherever $\rho < 0$. Thus, all conditions hold for negative ρ and positive η. The other three cases can be proved in a similar fashion. The 3-parameter Gompertz distribution is therefore regular in its parameter space.

3.2 SIMULATION STUDY

The 3-parameter Gompertz distribution as previously presented in this chapter is a new distribution. It is of great importance to determine if this distribution produces unbiased maximum likelihood estimates for each of the parameters. However, due to the complexity of its distributional form, it is impossible to compute MLEs analytically. Thus, it is necessary to perform a simulation study to determine whether or not the estimates are biased.

3.2.1 Data Generation Algorithm

Data sets of size n were generated using the following method. As survival data generally contain some amount of censored data, we considered three scenarios: 5%, 10% and 20% censoring. Failure times were generated according to a 3-parameter Gompertz distribution, using a modifed form of the probability integral transform method (Casella and Berger, 2002, Chapter 2) which accounted for the improper form of the distribution.

First, $n/2$ observations, denoted U_1, were generated from a uniform distribution on the interval $(0, 1 - \text{CF}_1)$, where CF_1 is equal to the cure fraction for the given set of parameters according to equation (3.3). Another $n/2$ observations, U_2, are generated in a similar fashion from $\text{UNIF}(0, 1 - \text{CF}_2)$. Next, we calculated T_1 and T_2 according the probability integral transform as

$$T_{ki} = \frac{1}{\rho_k} \log \left[-\frac{1}{\eta_k} \log \left\{ e^{-\eta_k} + \frac{\eta_k \rho_k}{\kappa_k} \log(1 - U_{ki}) \right\} \right]$$

for $k = 1, 2$ and $i = 1, \ldots, \dfrac{n}{2}$. For observations in T_1, the event indicators were set as $\delta_1 = 1$ and $\delta_2 = 0$. Conversely, for observations in T_2, we set the event indicators as $\delta_1 = 0$ and $\delta_2 = 1$. Then, to combine the data, let $T = \{T_{11}, \ldots, T_{1\frac{n}{2}}, T_{21}, \ldots, T_{2\frac{n}{2}}\}$ and combine the event indicators in a similar fashion. Finally, to specify the desired censoring proportion, generate n observations, denoted C, from a uniform distribution on the interval $(2, c_{\max})$, where $c_{\max}$ is chosen to guarantee the desired proportion. If $C_i < T_i$, set: $T_i = C_i$, $\delta_{1i} = 0$ and $\delta_{1i} = 0$.

3.2.2 Optimization

For the simulation study, we completed 1000 iterations for datasets with $n = 1000$, and three different censoring levels: 5%, 10% and 20%. The parameters used to generate the data were

$$\psi_1 = \{0.021, -0.107, 0.118\}$$

and

$$\psi_2 = \{0.014, -0.013, -1.118\}.$$

The optimization for each iteration was a two-step process. First, a cause-specific model was used for each type of event. That is, we maximized $\ln L_k = \sum_i \{\delta_{ik} \ln(f(t_i; \psi_k)) + (1 - \delta_{ik}) \ln(S(t_i; \psi_k))\}$ for each $k = 1, 2$, using initial values of $\{1, 1, 1, 1\}$. This is equivalent to using a cause-specific approach of obtaining MLEs, which does not take into account the correlation between the parameters of each type. This approach does, however, provide good initial estimates, denoted $\tilde{\psi}_1$ and $\tilde{\psi}_2$, for the direct approach. Thus, the results of each cause-specific model became the initial values for the joint model. Second, the direct model was used to estimate parameters for both types of events simultaneously, based on the likelihood function (2.6), giving MLEs denoted $\hat{\psi}_1$ and $\hat{\psi}_2$.

The likelihood functions may be optimized using any optimization procedure. While the Newton-Raphson algorithm is commonly used, the method is very sensitive to choice of starting values. Given the complexity of this distribution, the results of the optimization may be more accurate using another method, such as Nelder-Mead (Nelder and Mead, 1965) or conjugate gradient. For the purposes of this simulation, the Nelder-Mead simplex method was used. Data generation and optimization where performed SAS using PROC IML and PROC NLP. Sample code can be found in the Appendix.

3.2.3 Results

The results from the simulation study indicate that there is no significant bias in the maximum likelihood estimates produced by 3-parameter Gompertz distribution. Specific results for each censoring proportion can be found in Table 2. Three values are given for each parameter at each censoring level: bias (computed as $\frac{1}{n}\sum \hat{\beta} - \beta$), the standard error of

Table 2: Results of 3-parameter Gompertz Simulation (1000 simulated data sets with $n = 1000$)

	5% Censoring			10% Censoring			20% Censoring		
	Bias	SE	p	Bias	SE	p	Bias	SE	p
κ_1	51.80	(669.22)	0.47	89.57	(758.63)	0.45	110.78	(851.81)	0.45
ρ_1	0.06	(1.10)	0.48	-0.01	(1.74)	0.50	0.09	(0.20)	0.33
η_1	1.60	(2.51)	0.26	1.24	(2.74)	0.32	0.56	(3.34)	0.43
κ_2	55.18	(495.28)	0.46	5.86	(114.66)	0.48	18.93	(191.46)	0.46
ρ_2	0.01	(0.0001)	0.00	0.00	(0.36)	0.50	0.00	(0.28)	0.50
η_2	0.81	(2.50)	0.37	1.94	(2.31)	0.20	1.96	(3.59)	0.29

the estimates, and p-values (computed from the Wald statistic $Z = \left(\frac{1}{n}\sum\hat{\beta} - \beta\right)/\text{SE}(\hat{\beta})$, where $\beta \in \{\psi_1, \psi_2\}$). As evident in the table, the parameter estimates are approximately unbiased, and the standard errors generally increase as the censoring proportion increases.

3.3 APPLICATION TO DATA

The 3-parameter Gompertz model can be readily applied for parametric estimation of the cumulative incidence function, where the event of interest is local and regional recurrences of breast cancer. One data set comes from a clinical trial (B-14) conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP). This phase III trial investigated the use of tamoxifen in breast cancer patients who had negative axillary lymph nodes and estrogen receptor positive breast cancer. There were 1453 patients randomized to the placebo arm, 1413 of which were eligible and had follow-up. Additionally, 1439 patients were randomized to the tamoxifen arm, but only 1404 were eligible with follow-up. The mean time on study for both treatment groups was approximately 20.4 years. Fisher *et al.* (1989) reported that

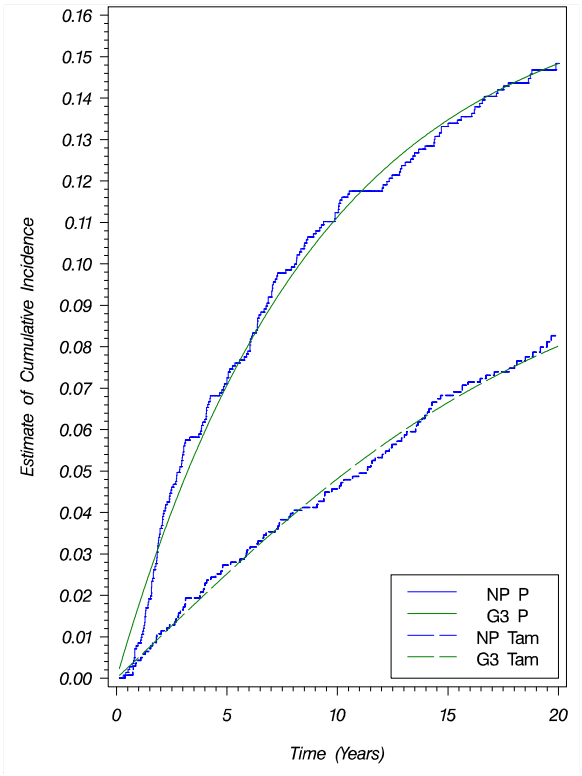
patients treated with tamoxifen had a better outcome than those treated with placebo. This analysis will focus on the effects of treatment and tumor size on the cumulative probabilities of local-regional recurrence and other events.

For this analysis, we use a cohort of a 2817 eligible patients on either the placebo or tamoxifen treatment arms for whom there was follow-up and who had known pathological tumor size. There were 1413 patients from the placebo treatment arm and 1410 from the tamoxifen arm.

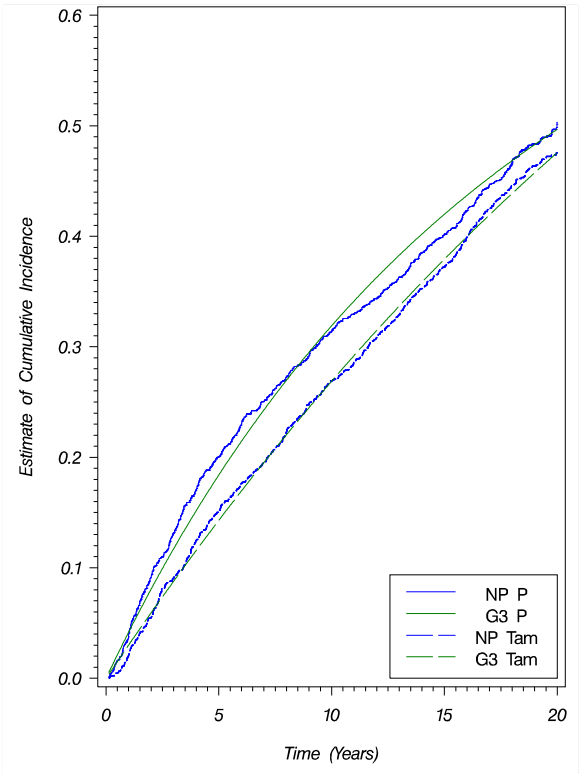
3.3.1 Simple Model

To begin with, we apply the model to the B-14 dataset from the NSABP. At this stage, we will stratify on treatment; however, when we are interested in the effect of covariates, we will compare to an unstratified model. Maximum likelihood estimates of the parameters were obtained using the direct likelihood presented in Section 2.1.2 and the survival function given in Section 3.1. Then parametric estimates of the cumulative incidence functions could be obtained by substituting the MLEs $(\hat{\kappa}_k, \hat{\rho}_k, \hat{\eta}_k)$ into the 3-parameter Gompertz cumulative distribution function, $1 - S_k(t; \hat{\kappa}_k, \hat{\rho}_k, \hat{\eta}_k)$ (from equation (3.1)). Nonparametric estimates of the cumulative incidence were computed using a SAS macro (Medical College of Wisconsin Division of Biostatistics, 2007). Figure 4 shows the estimated cumulative incidence for each event type from the NSABP B-14 dataset. It is clear from the plots that the parametric estimates calculated using the 3-parameter Gompertz distribution are very close to the estimates made without a distributional assumption for the data.

Maximum likelihood estimates of the parameters for the stratified model are given in Table 3. The cause-specific survival functions for both types of events and both treatment arms are improper. As is evident from equation (3.3), the 3-parameter Gompertz distribution is proper only when $\hat{\rho} > 0$ and $\hat{\eta} < 0$. For both types of event on the placebo arm, $\hat{\rho}$ is negative. For the tamoxifen arm, $\hat{\rho}$ is negative for other events, and both $\hat{\rho}$ and $\hat{\eta}$ are positive for local-regional recurrences.



(a) Local-regional Events



(a) Other Events

Figure 4: Plot of nonparametric (NP) and 3-Parameter Gompertz (G3) estimates of the hazard of (a) local-regional and (b) other events by treatment arm (Placebo vs. Tamoxifen) from NSABP B-14

Table 3: NSABP B-14 Parameter Estimates for G3 Model, by Treatment. For Local-Regional recurrences, $k = 1$, while for all other events, $k = 2$.

Var.	Placebo	Tamoxifen
κ_1	0.021 (0.021)	0.011 (0.009)
ρ_1	-0.107 (0.059)	0.064 (0.040)
η_1	0.118 (1.175)	0.807 (0.577)
κ_2	0.014 (0.035)	8.372 (0.009)
ρ_2	-0.012 (0.015)	-0.001 (0.002)
η_2	-1.118 (2.432)	5.619 (0.072)

3.3.2 Models with Covariates

Next, we were interested in the effect of two covariates: treatment and tumor size. Treatment (trt) was coded 0 for patients on the placebo arm, and 1 for patients on tamoxifen. Tumor size (tsize) was measured in centimeters. Both of these models were compared with a model without covariates.

The first model looked only at treatment as a predictor for the two event types. The estimates for this model can be found in Table 4. We can test the proportional odds and proportional hazards assumptions based on the parameter estimates for this model, using a Wald-type statistic of the following form:

$$Z_{\phi_k} = \frac{\hat{\phi}_k - \phi_{k0}}{SE(\hat{\phi}_k)}, \tag{3.4}$$

and $\phi_{k0} = 0$ if testing the proportional hazards assumption, and $\phi_{k0} = 1$ for the proportional odds assumption. Such a test statistic will test the hypothesis $H_0 : \hat{\phi}_k = \phi_{k0}$. As we will be testing both assumptions simultaneously, a Bonferroni correction will be used. For the test of proportional hazards (odds) assumptions, the p-values are 0.0077 (0.0023) for local-regional recurrences and 0.0125 (0.0102) for other events. Thus, we reject both model assumptions for both event types, and we should focus our attention on the model which estimates ϕ.

Next, we fit a model for both treatment and tumor size (measured in centimeters). As was the case with the treatment-only model, we reject both the proportional hazards and proportional odds assumptions. The p-values for the proportional hazards (odds) models are 0.0017 ($< .0001$) and 0.0029 (0.0002) for local-regional recurrences and other events respectively.

It is also of interest to determine whether treatment and tumor size were significant predictors of either of the two event types. For the treatment only model, $\hat{\beta}_{trt} = -1.386$ ($p < .0001$) for recurrences, and -0.513 ($p = .0008$) for other events. Similarly in the model with both treatment and tumor size, $p < .0001$ for treatment and $p = .0091$ for tumor size for local-regional recurrences, while for other events, $p = .0008$ for treatment and $p < .0001$ for tumor size. Thus, both treatment and tumor size have significant effects on recurrence for these patients. In both models, $\hat{\beta}_{trt}$ is negative, indicating a protective effect of tamoxifen on both local-regional recurrences and all other events, though the degree of the effect varies by event type. As $\hat{\beta}_{tsize}$ is positive for both event types, increased tumor size is associated with increased risk for recurrences and all other events.

For both events, the results are relatively consistent across models. Both treatment and tumor size are statistically significant predictors of recurrence and other events for the patients in B-14. Further, treatment is shown to decrease the recurrence rate among patients, while an increase in tumor size is associated with an increase in recurrence rates. These results are important for breast cancer patients, as well as investigators.

Table 4: NSABP B-14 G3 Parameter Estimates for all models: Model without Covariates (NoCovs); Treatment only (Trt); and Treatment and Tumor Size (Trt, Tsize). For Local-Regional recurrences, $k = 1$, while for all other events, $k = 2$. In the models with covariates, a separate intercept has been fit for each event type.

Var.	NoCovs	Trt	Trt, Tsize
$\hat{\kappa}_1$	0.014 (0.013)	0.059 (0.457)	0.063 (2.017)
$\hat{\rho}_1$	-0.085 (0.048)	-0.101 (0.022)	-0.105 (0.020)
$\hat{\eta}_1$	0.173 (1.054)	3.014 (1.015)	3.634 (1.024)
$\hat{\phi}_1$	–	15.811 (5.930)	19.292 (6.139)
$\hat{\beta}_{10}$	–	1.775 (7.854)	1.643 (32.117)
$\hat{\beta}_{1trt}$	–	-1.386 (0.323)	-1.500 (0.324)
$\hat{\beta}_{1tsize}$	–	–	0.276 (0.117)
$\hat{\kappa}_2$	0.016 (0.026)	0.006 (0.028)	0.038 (2.192)
$\hat{\rho}_2$	-0.005 (0.005)	0.006 (0.008)	0.188 (0.050)
$\hat{\eta}_2$	-0.804 (1.582)	-30.743 (51.075)	-0.001 (0.007)
$\hat{\phi}_2$	–	6.371 (2.091)	5.932 (1.337)
$\hat{\beta}_{20}$	–	-28.564 (51.284)	-0.530 (58.322)
$\hat{\beta}_{2trt}$	–	-0.513 (0.164)	-0.479 (0.147)
$\hat{\beta}_{2tsize}$	–	–	0.344 (0.065)

4.0 A SECOND EXTENSION OF THE GOMPERTZ MODEL

4.1 FOUR PARAMETER GOMPERTZ DISTRIBUTION

Jeong (2006) proposed a parameterization of Hougaard's family of distributions (1986), and used the generalized Weibull distribution to model the cumulative incidence function. The generalized Weibull distribution can be used in many cases, as it has a flexible hazard function that can be monotonically decreasing or increasing, unimodal, or U-shape.

We propose an extension of Jeong's model that will account for a "cured population" for each of the possible cause-specific events. Additionally, Jeong's model assumed that the distribution came from a family with expected value 1 (Hougaard, 1986), while this model is more flexible. Finally, this 4-parameter Gompertz model (G4) has a flexible hazard, which can be either monotone or unimodal.

From the distribution presented by Hougaard (1986), a survival function can be developed from equation 4.2:

$$S(t) = \exp\left[-\frac{\delta\theta^{\alpha}}{\alpha}\left\{\left(\frac{s}{\theta}+1\right)^{\alpha}-1\right\}\right]. \tag{4.1}$$

Then we reparameterize δ so that $\delta = \theta^{2-\alpha}$. This substitution allows for a reduction in the dimension of the parameter space as well as permitting the expected value of the underlying function to be θ. Furthermore, noting that s in equation (4.1) represents the cumulative hazard function, here we will adopt that of the Gompertz distribution to allow the proposed distribution to be improper. Hence, we have a 4-parameter survival function related to the Gompertz distribution that can be written as follows:

$$S_{G4}(t;\alpha,\theta,\rho,\kappa) = \exp\left[-\frac{\theta^{2}}{\alpha}\left\{\left(\frac{\kappa}{\theta\rho}(e^{\rho t}-1)+1\right)^{\alpha}-1\right\}\right]. \tag{4.2}$$

The parameters space is as follows: $\theta, \kappa > 0$, and α and ρ must be finite. In this parameter space, the distribution is regular, which will be shown in Section 4.1.2.

4.1.1 Properties

The hazard function for this model can be obtained directly from (4.2), as

$$h_{G4}(t;\alpha,\theta,\kappa,\rho) = \frac{-\frac{d}{dt}S_{G4}(t)}{S_{G4}(t)} = \theta\kappa e^{\rho t}\left\{\frac{\kappa}{\theta\rho}(e^{\rho t}-1)+1\right\}^{\alpha-1}. \tag{4.3}$$

Setting the first derivative of $h_{G4}(t)$ equal to zero, we can find the maximum of the hazard function, occurring at

$$t^* = \frac{1}{\rho}\ln\left\{\frac{\kappa-\theta\rho}{\alpha\kappa}\right\}, \tag{4.4}$$

provided that $\frac{\kappa-\theta\rho}{\alpha\kappa} > 0$. Figure 5 shows the flexibility of the hazard function, and its unimodality.

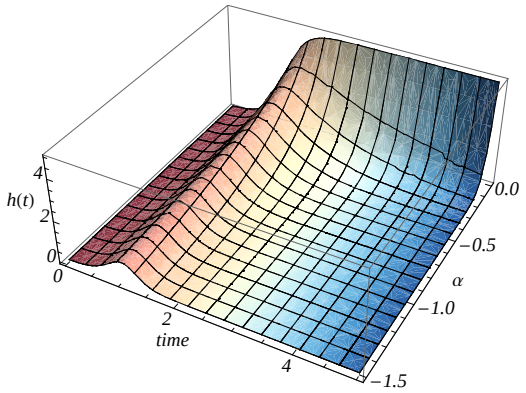


Figure 5: Plot of 4-parameter Gompertz Hazard Function with $\theta = 1$, $\kappa = 0.02$, and $\rho = 5$

The cure fraction is also an important quantity here since this distribution is improper, wherever ρ or α is negative. For the 4-parameter Gompertz model, the cure fraction is equal to

$$\lim_{t\to\infty} S_{G4}(t) = \begin{cases} 0 & \text{if } \rho > 0 \text{ and } \alpha > 0; \\ \exp\left(\frac{\theta^2}{\alpha}\right) & \text{if } \rho > 0 \text{ and } \alpha < 0; \\ \exp\left[\left(\frac{-\theta^2}{\alpha}\right)\left\{(1-\frac{\kappa}{\theta\rho})^{\alpha}-1\right\}\right] & \text{if } \rho < 0. \end{cases} \tag{4.5}$$

Thus, for the case where both ρ and α are greater than zero, the distribution is proper. If either ρ or α is less than zero, the survival function for this distribution is improper. Since the cumulative incidence function is generally improper (see Section 2.1.1), it is not necessary to use proper distributions to model competing risks data. It should be noted that even for the proper case of this distribution, it may be improper for the length of time that is of interest (for example, 25 years), and will therefore be in all cases applicable.

4.1.2 Regularity

Because $S_{G4}(t)$ is a survival function, it must only take on values between 0 and 1. That is,

$$\frac{\theta^2}{\alpha}\left\{\left(\frac{\kappa}{\theta\rho}(e^{\rho t}-1)+1\right)^{\alpha}-1\right\}>0.$$

This must be true for all values in the parameter space.

If $\alpha > 0$, then $\left(\frac{\kappa}{\theta\rho}(e^{\rho t}-1)+1\right)^{\alpha}-1$ must be greater than 0. So, $\left(\frac{\kappa}{\theta\rho}(e^{\rho t}-1)+1\right)^{\alpha} > 1$, implying that $\frac{\kappa}{\theta\rho}(e^{\rho t}-1) > 0$. Since κ and θ are both greater than 0, we need only to look at ρ. Where $\rho > 0$, this implies that $e^{\rho t}-1 > 0$ or $e^{\rho t} > 1$, which is always true. Conversely, if $\rho < 0$, then $e^{\rho t}-1 < 0$ and $e^{\rho t} < 1$, which is also always true. Thus, the conditions hold for all $\alpha > 0$. A similar proof can be constructed for the case where $\alpha < 0$.

4.2 SIMULATION STUDY

The 4-parameter Gompertz distribution as previously presented in this chapter is a new distribution. As with the G3 model, we are interested in the unbiasedness of the maximum likelihood estimates. The G4 distribution is however quite complex and MLEs cannot be computed analytically.

4.2.1 Data Generation and Optimization

Data sets were simulated according to an algorithm like that presented in Section 3.2.1. At the probability integral transform step, the failure times where generated as:

$$T_{ki} = \frac{1}{\rho_k} \log \left[\frac{\theta_k \rho_k}{\kappa_k} \left\{ \left(\left(1 - \frac{\alpha_k}{\theta_k^2} \log(1 - U_{ki}) \right)^{1/\alpha_k} \right) - 1 \right\} + 1 \right]$$

for $k = 1, 2$, and $i = 1 \ldots n$. For the simulation study, we completed 1000 iterations for datasets with $n = 1000$, and three different censoring levels: 5%, 10% and 20%. The parameters used to generate the data were

$$\psi_1 = \{0.2, 0.4, 0.15, -0.1\}$$

and

$$\psi_2 = \{0.1, 0.5, 4.0, -0.08\}.$$

As in the G3 case, he optimization for each iteration was a two-step process. First, a cause-specific model was used for each type of event. That is, we maximized $\ln L_k = \sum_i \{\delta_{ik} \ln(f(t_i; \psi_k)) + (1 - \delta_{ik}) \ln(S(t_i; \psi_k))\}$ for each $k = 1, 2$, using initial values of $\{1, 1, 1, 1\}$. This is equivalent to using a cause-specific approach of obtaining MLEs, but does not take into account the correlation between the parameters of each type. This approach does, however, provide good initial estimates, denoted $\tilde{\psi}_1$ and $\tilde{\psi}_2$, for the direct approach. Thus, the results of each cause-specific model became the initial values for the direct model. Second, the direct model was used to estimate parameters for both types of events simultaneously, using the direct approach given in equation (2.6), which would give the MLEs denoted $\hat{\psi}_1$ and $\hat{\psi}_2$.

The likelihood functions may be optimized using any optimization procedure. While the Newton-Raphson algorithm is commonly used, the method is very sensitive to choice of starting values. Given the complexity of this distribution, the results of the optimization may be more accurate using another method, such as Nelder-Mead (Nelder and Mead, 1965) or conjugate gradient. For the purposes of this simulation, the Nelder-Mead simplex method was used. Data for the simulation was generated using SAS PROC IML, while optimizations were performed using PROC NLP. Sample code can be found in the Appendix.

Table 5: Results of G4 Simulation (1000 simulated data sets with $n = 1000$)

	5% Censoring			10% Censoring			20% Censoring		
	Bias	SE	p	Bias	SE	p	Bias	SE	p
α_1	-0.03	(0.37)	0.23	-0.18	(4.40)	0.24	0.05	(0.83)	0.74
θ_1	0.37	(0.56)	0.63	1.10	(4.95)	0.71	0.54	(1.36)	0.67
κ_1	0.51	(3.15)	0.72	0.53	(2.40)	0.71	0.50	(1.80)	0.69
ρ_1	0.46	(1.04)	0.67	0.56	(1.37)	0.67	0.42	(1.00)	0.67
α_2	-0.01	(0.15)	0.24	0.01	(0.16)	0.73	0.06	(0.15)	0.68
θ_2	0.12	(0.11)	0.57	0.10	(0.12)	0.61	0.09	(0.11)	0.61
κ_2	0.52	(4.95)	0.73	0.91	(7.76)	0.73	1.04	(2.10)	0.66
ρ_2	0.29	(0.17)	0.52	0.36	(0.25)	0.54	0.36	(0.25)	0.54

4.2.2 Results

The results from the simulation study indicate that there is no significant bias in the maximum likelihood estimates produced by the 4-parameter Gompertz distribution. Specific results for each censoring proportion can be found in Table 5. Three values are given for each parameter at each censoring level: bias (computed as $\frac{1}{n}\sum \hat{\beta} - \beta$), the standard error of the estimates, and p-values (computed from the Wald statistic $Z = \text{bias}/SE(\hat{\beta})$, where $\beta \in \{\psi_1, \psi_2\}$). As evident in the table, the parameter estimates are approximately unbiased, and the standard errors generally increase as the censoring proportion increases.

4.3 EXAMPLE

The 4-parameter Gompertz model can be readily applied for parametric estimation of the cumulative incidence function, where the event of interest is local and regional recurrences of breast cancer. One data set comes from a clinical trial (B-14) conducted by the National Sur-

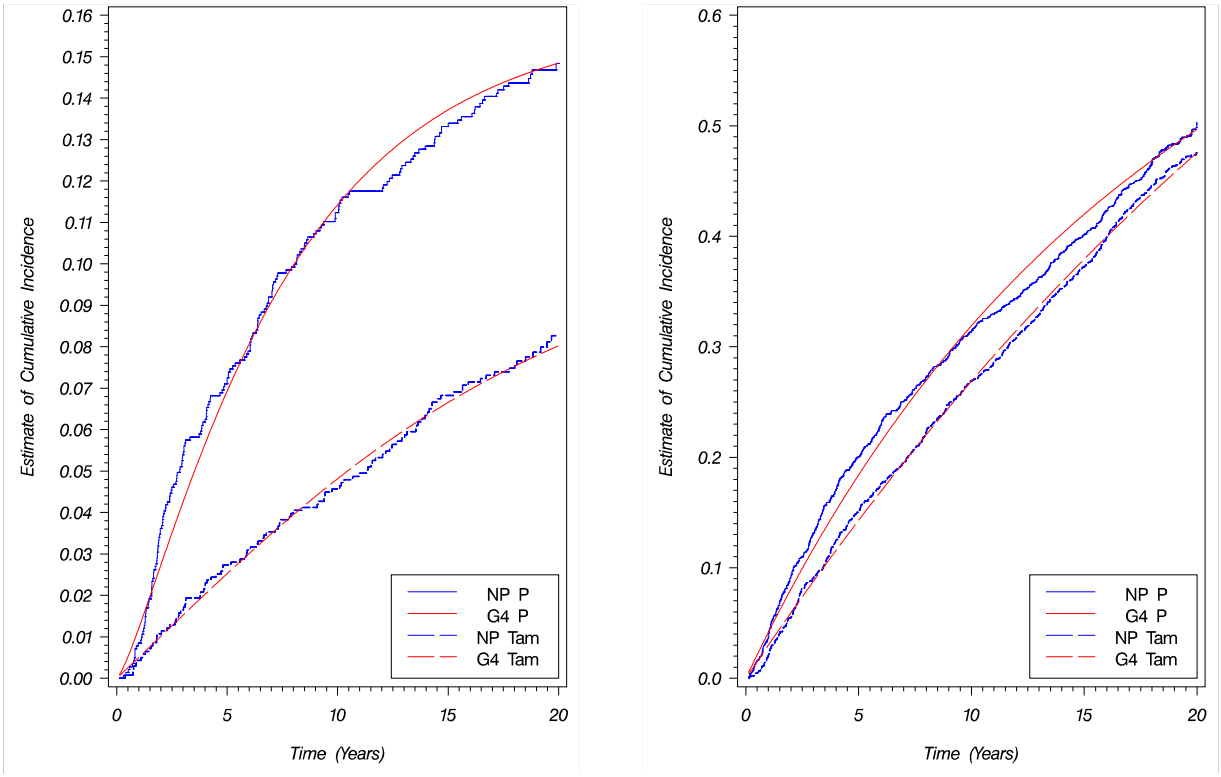
gical Adjuvant Breast and Bowel Project (NSABP). This phase III trial investigated the use of tamoxifen in breast cancer patients who had negative axillary lymph nodes and estrogen receptor positive breast cancer. There were 1453 patients randomized to the placebo arm, 1413 of which were eligible and had follow-up. Additionally, 1439 patients were randomized to the tamoxifen arm, but only 1404 were eligible with follow-up. The mean time on study for both treatment groups was approximately 20.4 years. Fisher *et al.* (1989) reported that patients treated with tamoxifen had a better outcome than those treated with placebo. This analysis will focus on the effects of treatment and tumor size on the cumulative probabilities of local-regional recurrence and other events.

For this analysis, we use a cohort of a 2817 eligible patients on either the placebo or tamoxifen treatment arms for whom there was follow-up and who had known pathological tumor size. There were 1413 patients from the placebo treatment arm and 1410 from the tamoxifen arm.

4.3.1 Simple Model without Covariates

First, we apply the model without covariates to the data. Two models are presented, one which stratifies by treatment, and another which models the entire dataset. Maximum likelihood estimates of the parameters were obtained using the direct likelihood presented in Section 2.1.2 and the survival function given in Section 4.1. Then parametric estimates of the cause-specific cumulative incidence functions could be obtained by substituting the MLEs $(\hat{\alpha}_k, \hat{\theta}_k, \hat{\rho}_k, \hat{\kappa}_k)$ into the 4-parameter Gompertz cumulative distribution function, $1 - S_k^{G4}(t; \hat{\alpha}_k, \hat{\theta}_k, \hat{\rho}_k, \hat{\kappa}_k)$ (from equation (4.2)). Nonparametric estimates of the cumulative incidence were computed using a SAS macro (Medical College of Wisconsin Division of Biostatistics, 2007). Figure 6 shows the estimated cumulative incidence for each event type using data from NSABP B-14. It is clear from the plots that the parametric estimates calculated using the 4-parameter Gompertz distribution are very close to the estimates made without assuming a distributional form for the data.

Maximum likelihood estimates of the parameters for the model stratified by treatment group are presented in Table 6. The cause-specific survival functions for both types of events



(a) Local-Regional Events

(b) Other Events

Figure 6: Plot of nonparametric (NP) and 4-Parameter Gompertz (G4) estimates of the hazard of (a) local-regional and (b) other events by treatment arm (Placebo vs. Tamoxifen) from NSABP B-14

Table 6: NSABP B-14 Parameter Estimates for G4 Model, by Treatment. For Local-Regional recurrences, $k = 1$, while for all other events, $k = 2$.

Var.	Placebo	Tamoxifen
κ_1	0.008 (0.011)	0.002 (0.006)
ρ_1	6.415 (3.261)	10.391 (8.606)
α_1	-0.019 (0.010)	-0.004 (0.004)
θ_1	0.058 (0.016)	0.025 (0.011)
κ_2	0.119 (0.032)	12.271 (29.183)
ρ_2	0.086 (0.084)	-0.020 (0.013)
α_2	0.170 (0.309)	1.177 (0.085)
θ_2	0.444 (0.080)	0.000 (0.001)

are improper, since $\alpha < 0$ for local-regional recurrences, and $\rho < 0$ for all other events.

4.3.2 Models with Covariates

Next, we were interested in the effect of two covariates: treatment and tumor size. Treatment (trt) was coded 0 for patients on the placebo arm, and 1 for patients on tamoxifen. Tumor size (tsize) was measured in centimeters. For comparison, we also present here the parameter estimates for a model without covariates, and without stratification. The first model looked only at treatment as a predictor for the two event types. The second model examined both treatment and tumor size. The estimates for these models can be found in Table 7.

We can test the proportional odds and proportional hazards assumptions based on the parameter estimates for this model, using a Wald-type statistic of the following form:

$$Z_{\phi_k} = \frac{\hat{\phi}_k - \phi_{k0}}{se(\hat{\phi}_k)}, \tag{4.6}$$

and $\phi_{k0} = 0$ if testing the proportional hazards assumption, and $\phi_{k0} = 1$ for the proportional odds assumption. Such a test statistic will test the hypothesis $H_0 : \hat{\phi}_k = \phi_{k0}$, and we use

Table 7: NSABP B-14 G3 Parameter Estimates for all models: Model without Covariates (NoCovs); Treatment only (Trt); and Treatment and Tumor Size (Trt, Tsize). For Local-Regional recurrences, $k = 1$, while for all other events, $k = 2$. In the models with covariates, a separate intercept has been fit for each event type.

Var.	NoCovs	Trt	Trt, Tsize
$\hat{\kappa}_1$	0.005 (0.007)	0.952 (2.249)	0.262 (0.973)
$\hat{\rho}_1$	8.096 (3.774)	-0.026 (0.026)	-0.020 (0.026)
$\hat{\alpha}_1$	-0.011 (0.006)	1.789 (0.224)	1.929 (0.246)
$\hat{\theta}_1$	0.041 (0.010)	0.00005 (0.0003)	0.000001 (0.0001)
$\hat{\phi}_1$	–	25.031 (7.596)	29.185 (8.084)
$\hat{\beta}_{10}$	–	-1.458 (4.968)	-1.100 (0.137)
$\hat{\beta}_{1trt}$	–	-1.836 (0.367)	-1.928 (0.375)
$\hat{\beta}_{1tsize}$	–	–	0.430 (0.155)
$\hat{\kappa}_2$	11.178 (17.216)	0.109 (0.648)	0.080 (0.0001)
$\hat{\rho}_2$	-0.018 (0.009)	3.843 (1.095)	4.147 (1.139)
$\hat{\alpha}_2$	1.054 (0.050)	0.115 (0.060)	0.065 (0.026)
$\hat{\theta}_2$	0.002 (0.003)	0.264 (1.564)	0.165 (0.0001)
$\hat{\phi}_2$	–	14.265 (4.719)	8.839 (1.924)
$\hat{\beta}_{20}$	–	-0.878 (11.881)	-1.387 (0.0001)
$\hat{\beta}_{2trt}$	–	-0.872 (0.243)	-0.683 (0.184)
$\hat{\beta}_{2tsize}$	–	–	0.445 (0.078)

a Bonferroni correction to account for the multiple testing. For the test of the proportional hazards assumptions for the treatment only model, the p-values are 0.0010 and 0.0025 for local-regional recurrences and other events respectively. For the test of the proportional odds assumptions, the p-values are 0.0016 and 0.0049 respectively. Thus, we reject both model assumptions for both event types, and we should focus our attention on the model which estimates ϕ.

Next, we fit a model for both treatment and tumor size (measured in centimeters). The parameters ρ and α are each tested to determine if either of the cause-specific distributions are improper. Based on these tests, it cannot be rejected that either of the distributions are proper. Additionally, ϕ must be tested to determine if we can assume either proportional odds or hazards. The p-values for the proportional hazards (odds) models are 0.0003 (0.0005) and < 0.0001 (< 0.0001) for local-regional recurrences and other events respectively. So, both the proportional odds and hazards assumptions are rejected for this model.

It is also of interest to determine whether treatment and tumor size were significant predictors of either of the two event types. For the treatment only model, $\hat{\beta}_{trt} = -1.84$ ($p < .0001$) for recurrences, and -0.87 ($p = .0001$) for other events. Similarly in the model with both treatment and tumor size, $p < .0001$ for treatment and $p = .0003$ for tumor size for local-regional recurrences, while for other events, $p < .0001$ for both event types. Thus, both treatment and tumor size have significant effects on recurrence for these patients. In both models, $\hat{\beta}_{trt}$ is negative, indicating a protective effect of tamoxifen on both local-regional recurrences and all other events, though the degree of the effect varies by event type. As $\hat{\beta}_{tsize}$ is positive for both event types, increased tumor size is associated with increased risk for recurrences and all other events.

For both events, the results are relatively consistent across models. Both treatment and tumor size are statistically significant predictors of recurrence and other events for the patients in B-14. Further, treatment is shown to decrease the recurrence rate among patients, but an increase in tumor size is associated with an increase in recurrence rates. These results are important for both investigators and breast cancer patients.

4.4 DISCUSSION

In Chapters 3 and 4, we have presented two novel extensions of the Gompertz distribution. The addition of a third parameter, η, or third and fourth parameters, α and θ, allow the models to capture unimodal hazards, while still containing the original model as a special case. This feature is important as it protects against model misspecification when performing parametric analysis of competing risks data. Since the Gompertz distribution is well-suited to fitting cumulative incidence functions, these new distributions would broaden the scope of parametric inference on competing risks data..

For the G3 model, the effect of η on the hazard rate is not straightforward. The relationship between $h_{G2}(t)$ and $h_{G3}(t)$ is such that $h_{G3}(t)$ is not a simple function of $h_{G2}(t)$ and the additional parameter η. Rather, the two hazard functions can be related as

$$h_{G3}(t) = h_{G2}(t) \exp\left\{-\frac{\eta}{\kappa} h_{G2}(t)\right\}, \tag{4.7}$$

which is also a function of κ. Thus, the effect of η on the hazard rate is neither additive nor multiplicative, and the hazard of the G3 model cannot be determined solely as a function of the hazard of the G2 model and the additional parameter. The G3 model, however, remains well-defined because it captures the standard Gompertz model as a special case. However, while G3$(\kappa, \rho, 0)$ is equivalent to G2(κ, ρ), there is no such relation between the G4 and G3 models. It might be beneficial to have a G4 model that is more closely related to its counterparts. This is clearly an area for future research.

Because the models are not directly related, assessing the goodness-of-fit of one model of the other is not simply a matter of testing the significance of a parameter. Some other method must therefore be used to assess the fit of the distributions to the data, such as the Akaike information criteria (Akaike, 1974).

5.0 JOINT MODELING OF COMPETING RISKS

In the previous chapters, we have used the direct approach of Jeong and Fine (2006) (presented in section 2.1.2) to make inference about the parameters for the two distinct event types. This method assumes that the times T_1, the time to an event of type 1, and T_2, the time to an event of type 2, are not independent. However, this approach specifies that the overall survival function be

$$\begin{aligned} S_D(t_1, t_2) &= 1 - F_1(t_1) - F_2(t_2) \\ &\approx S_1(t_1) + S_2(t_2) - 1. \end{aligned} \tag{5.1}$$

With the direct approach, it is assumed that the cumulative incidence functions for two types of events are additive: $F(t) = F_1(t) + F_2(t)$. It is unclear whether this assumption is always applicable. Another approach, called "cause-specific," assumes that the hazards for two competing events are mutually exclusive, and thus that the parameters for each cause-specific event can be estimated separately. The overall survival function assumed in the cause-specific case is

$$S_{CS}(t_1, t_2) = S_1(t_1)S_2(t_2). \tag{5.2}$$

This is equivalent to the cause-specific hazard rates being additive: $h(t) = h_1(t) + h_2(t)$.

Both of these methods make specific assumptions about the form of the joint distribution of T_1 and T_2. (This chapter will assume that we have a situation with only two competing risks.) Here, we present a generalized approach to the modeling of competing risks which encompasses both the direct (D) and cause-specific (CS) methods as special cases. This joint method (J) is simple to employ, and estimates a parameter which allows investigators to test the independence of the two lifetime distributions under competing risks.

5.1 RELATED WORK

Bivariate survival data were first discussed by Clayton (1978), where Clayton developed a nonparametric method for bivariate lifetables, and presented a method of estimating the association between the two correlated lifetime distributions. Oakes (1982) later presented an alternate association parameter which measured concordance between the two lifetime distributions. Oakes (1989) developed theory for bivariate survival modeling, including the use of frailty to construct an Archimedean copula, which could be used to model bivariate survival data.

Zheng and Klein (1995) presented a nonparametric copula-graphic estimator of the marginal distributions for competing risks data. Bunea and Bedford (2002) present an Archimedean copula model used on dependent competing risks models. Escarela and Carrière (2003) fit competing risks models using an assumed copula, and Weibull marginal distributions.

More recently, Jeong and Fine (2006) proposed a direct modeling approach for the cumulative incidence function, contrasting it with the cause-specific approach which assumes independence. Chen *et al.* (2007) proposed three two-sample tests for correlated competing risks data. Nonparametric estimators of the cumulative hazard and incidence functions were presented by Cheng *et al.* (2007) for bivariate competing risks, and also discussed summary statistics for the dependence between failure times. Finally, Cheng and Fine (2008) discuss nonparametric estimation of a cause-specific cross-hazard ratio for bivariate survival data.

To date, copulas have not been applied to bivariate competing risks models in cases where the marginals are modeled with improper distributions. In this chapter, we present such an approach.

5.2 COPULAS

We begin with a discussion of copulas, as the joint method will be developed using an Archimedean copula. However, as the distributions used in competing risks model are im-

proper, and copula models assume proper marginals, it is necessary to confirm that it is possible to use copulas (as defined in the literature) with improper marginal distributions. A literature review revealed no work to this point on copulas for improper marginals. It is therefore necessary to describe copulas in some detail and discuss their use with improper marginals before the joint approach to modeling competing risks is presented.

Copulas were introduced in Sklar (1959) as bivariate functions that combine two univariate distribution functions, the marginal distributions, which are uniform on the unit interval. Though an overview of copulas is presented here, readers may wish to refer to Nelsen (2006); Oakes (1989); Genest and MacKay (1986); Genest and Rivest (1993). A number of definitions are needed to properly define a copula (adapted from Nelsen (2006)).

5.2.1 Basic Properties

Let S_1 and S_1 by subsets of the real line $[-\infty, \infty]$. Further, let $\mathcal{C}$ by a real function with domain $S_1 \times S_2$ and range $[-\infty, \infty]$.

$\mathcal{C}$-Volume: Suppose $B = [u_1, u_2] \times [\nu_1, \nu_2]$ is contained in the domain of $\mathcal{C}$. Then the $\mathcal{C}$-Volume of B is defined as

$$V_{\mathcal{C}}(B) = \mathcal{C}(u_2, \nu_2) + \mathcal{C}(u_1, \nu_2) + \mathcal{C}(u_2, \nu_1) - \mathcal{C}(u_1, \nu_1).$$

2-increasing: Here, the notion of nondecreasing functions is extended to bivariate functions. A function $\mathcal{C}$ is said to be 2-increasing if $V_{\mathcal{C}}(B) \geq 0$ for all rectangles B which are inside the domain of $\mathcal{C}$.

Grounded: Suppose a_1 and a_2 are the least elements of S_1 and S_2 respectively. Then, a function $\mathcal{C}$ is said to be grounded if $\mathcal{C}(a_1, \nu) = 0$ and $\mathcal{C}(u, a_2) = 0$.

It can be shown that if $\mathcal{C}$ is a grounded, 2-increasing function, then the function is nondecreasing in both S_1 and S_2. One example of a function that is both grounded and 2-increasing is $\mathcal{C}(u, \nu) = \min(u, \nu)$, pictured in Figure 7.

Using the previous definitions, both *subcopula* and *copula* can now be defined. A subcopula is a function $\mathcal{C}'$ such that;

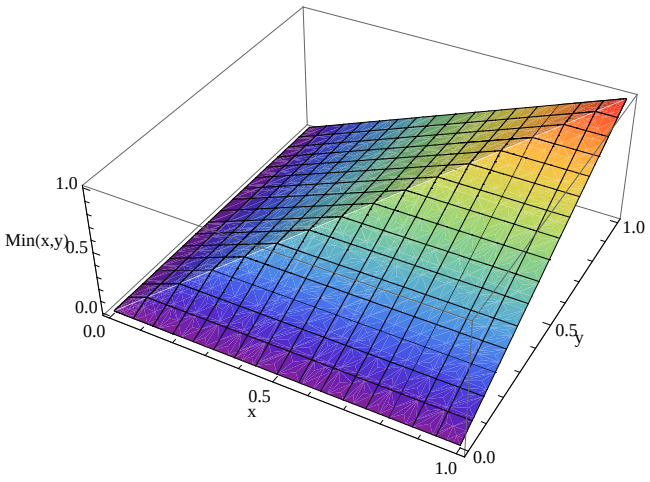


Figure 7: $\mathcal{C} = \min(x, y)$ is a good example of a 2-increasing, grounded function

i) the domain of $\mathcal{C}'$ is $S_1 \times S_2$, where S_1 and S_2 are subsets of the interval $[0, 1]$ and contain both 0 and 1,

ii) $\mathcal{C}'$ is both grounded and 2-increasing, and

iii) $\mathcal{C}'(u, 1) = 1$ and $\mathcal{C}'(1, \nu) = 1$ for all u in S_1 and ν in S_2.

A related function is the copula, $\mathcal{C}$, which is defined as a subcopula, $\mathcal{C}'$, which has a domain $[0, 1] \times [0, 1]$.

Sklar's theorem (Sklar, 1959) is important to the use of copulas in that it shows that a copula can be used to create a joint distribution function from two marginal distributions. Suppose that F_{12} is a joint distribution function that has marginals denoted F_1 and F_2. Then, Sklar's theorem states that there exists a copula F_{12} such that

$$F_{12}(x, y) = \mathcal{C}(F_1(x), F_2(y))$$

for all real x and y. The copula will be unique for continuous F_1 and F_2. If F_1 and F_2 are not continuous, then F_{12} is determined by their ranges. Conversely, if F_{12} is a copula, and F_1 and F_2 are distribution functions, then the function F_{12} defined previously is a joint distribution function for marginals F_1 and F_2.

5.2.2 Archimedean Copulas

One important class of copulas is known as Archimedean copulas. These are widely applicable, as they are easy to construct, and the family encompasses a wide variety of functions. Archimedean copulas take the form

$$\mathcal{C}(u, \nu) = p[q(u) + q(\nu)] \tag{5.3}$$

where u and ν are uniformly distributed on the interval $[0, 1]$, and $p(\cdot)$ is a decreasing non-negative function such that $p(0) = 1$ and $p''(\cdot) \geq 0$, and q is the inverse function of p. That is,

$$p(0) = 1; p(u) \geq 0; p'(u) \leq 0; p''(u) \geq 0. \tag{5.4}$$

The function p is known as the generator for an Archimedean copula. According to Theorem 4.1.4 in Nelsen (2006), such a function can be used to construct an Archimedean copula if and only if p is also convex.

One such generator gives the Clayton copula as presented in Oakes (1989). Oakes discusses the use of copulas in the survival setting, which varies slightly from their use with cumulative distribution functions. In that case, a copula model may be constructed as in equation (5.3). Clayton (1978) proposed that $p(u) = (u + 1)^{-1/(\lambda-1)}$. As λ tends to 1, it is evident that $\mathcal{C}(u, \nu) = u\nu$, representing the case where u and ν are independent. When $\lambda = 0$, then $\mathcal{C}(u, \nu)$ achieves its lower Fréchet bound of $u + \nu - 1$ (Fréchet, 1951; Hoeffding, 1940).

5.2.3 Copulas for Improper Marginals

The discussion of copulas to this point has assumed that the marginal distributions are proper, with a range on the entire interval $[0, 1]$. As we are modeling competing risks data here, the marginal distributions will not be proper. This is, however, not an issue that must be addressed with new methods. As improper distributions are just subsets of proper distributions, we would like the copulas of such distributions to be subsets of the copula of the related proper distributions. The simplest way to accomplish this goal is to use previously-established copula models, such as the Clayton copula.

5.3 JOINT APPROACH

The joint approach (J) to inference on competing risks data involves using an Archimedean copula to specify the overall survival function of the two cause-specific events. Let T_1 be the time until an event of type 1 and, similarly, let T_2 be the time until an event of type 2. Then we can model the joint survival function of T_1 and T_2 by using a Clayton copula in this form:

$$S_J(t_1, t_2) = (S_1(t_1)^{1-\lambda} + S_2(t_2)^{1-\lambda} - 1)^{\frac{1}{1-\lambda}}, \tag{5.5}$$

where $S_k = 1 - F_k(t_k)$, and λ allows us to estimate the dependence between T_1 and T_2. This copula is a bivariate survival function by Sklar's theorem. As λ tends to 1, $S_J(t,t) \to S_{CS}$, and when $\lambda = 0$, the joint survival function is approximately equal to S_D.

A key component to the joint approach is a definition of the likelihood which works in general, and maintains the usual form in both of the special cases. Suppose $S(t_1, t_2)$ is some definition of the joint survival distribution. Then the likelihood can be defined generally as

$$L = \prod_{i=1}^{n} \left[\left\{ \frac{-d}{dt_{1i}} S(t_{1i}, t_{2i})|_{t_{1i}=t_{2i}=t_i} \right\}^{\delta_{1i}} \left\{ \frac{-d}{dt_{2i}} S(t_{1i}, t_{2i})|_{t_{1i}=t_{2i}=t_i} \right\}^{\delta_{2i}} S(t_i, t_i)^{1-\delta_{1i}-\delta_{2i}} \right]. \tag{5.6}$$

We can use this method to construct a likelihood for the the joint approach. If

$$S(t_1, t_2) = S_J(t_1, t_2) = (S_1(t_1)^{1-\lambda} + S_2(t_2)^{1-\lambda} - 1)^{\frac{1}{1-\lambda}},$$

then we calculate the negative first derivatives as $f_1(t)S_1(t)^{-\lambda}S_J(t,t)^{\lambda}$ and $f_2(t)S_2(t)^{-\lambda}S_J(t,t)^{\lambda}$ respectively. The corresponding form of the generalized likelihood given in equation (5.6) is then:

$$\begin{aligned} L_J &= \prod_{i=1}^{n} \Big[\left\{ f_1(t_i) S_1(t_i)^{-\lambda} S_J(t_i, t_i)^{\lambda} \right\}^{\delta_{1i}} \left\{ f_2(t_i) S_2(t_i)^{-\lambda} S_J(t_i, t_i)^{\lambda} \right\}^{\delta_{2i}} \\ &\qquad \times \left\{ (S_1(t_i)^{1-\lambda} + S_2(t_i)^{1-\lambda} - 1)^{\frac{1}{1-\lambda}} \right\}^{1-\delta_{1i}-\delta_{2i}} \Big] \\ &= \prod_{i=1}^{n} \big[f_1(t_i)^{\delta_{1i}} S_1(t_i)^{-\lambda\delta_{1i}} f_2(t_i)^{\delta_{2i}} S_2(t_i)^{-\lambda\delta_{2i}} \\ &\qquad \left(S_1(t_i)^{1-\lambda} + S_2(t_i)^{1-\lambda} - 1 \right)^{\frac{1}{1-\lambda} - \delta_{1i} - \delta_{2i}} \big]. \end{aligned} \tag{5.7}$$

If we were to use the generalized form of the likelihood and assume additive cumulative incidence (that is, use the direct approach), then $S(t_1, t_2) = S_D(t_1, t_2) = 1 - F_1(t_1) - F_2(t_2)$. Then the first derivatives are $f_1(t)$ and $f_2(t)$ respectively. The likelihood for the direct approach then is

$$L_D = \prod_{i=1}^{n} \left[\{f_1(t_i)\}^{\delta_{1i}} \{f_2(t_i)\}^{\delta_{2i}} \{1 - F_1(t_i) - F_2(t_i)\}^{1-\delta_{1i}-\delta_{2i}} \right], \tag{5.8}$$

which is exactly what was presented in Section 2.1.2. It should be noted that f_1 and f_2, are improper pseudo-survival distributions, and the direct approach is not appropriate when the marginal survival distributions are proper.

The same procedure can also be used to derive the likelihood used with the cause-specific approach. In this case, $S(t_1, t_2) = S_{CS}(t_1, t_2) = S_1(t_1)S_2(t_2)$ and the corresponding first derivatives are $f_1(t)S_2(t)$ and $S_1(t)f_2(t)$ respectively. The likelihood is then constructed as:

$$\begin{aligned} L_{CS} &= \prod_{i=1}^{n} \left[\{f_1(t_i)S_2(t_i)\}^{\delta_{1i}} \{S_1(t_i)f_2(t_i)\}^{\delta_{2i}} \{S_1(t_i)S_2(t_i)\}^{1-\delta_{1i}-\delta_{2i}} \right] \\ &= \prod_{i=1}^{n} \left[f_1(t_i)^{\delta_{1i}} f_2(t_i)^{\delta_{2i}} S_1(t_i)^{1-\delta_{1i}} S_2(t_i)^{1-\delta_{2i}} \right], \end{aligned} \tag{5.9}$$

which also corresponds with the likelihood presented in Section 2.1.2. In this situation, both f_k and S_k are proper distributions, but the product $f_k(t)S_{3-k}(t)$ is improper.

Thus, it is clear that both the direct model and the cause-specific model are special cases of the joint model. We can use this fact to test whether either the direct or cause-specific approach is correct by using a Wald-type statistic of the form

$$Z = \frac{\hat{\lambda} - \lambda_0}{\mathrm{SE}(\hat{\lambda})}, \tag{5.10}$$

where λ_0 is either 0 for the direct approach or 1 for the cause-specific approach, which can then be compared to the standard normal distribution to obtain a p-value.

Once the maximum likelihood estimation has been completed, there is still the question of how to correctly estimate the cumulative incidence function for each of the event types. Suppose some distribution has been assumed for the marginals, and that the MLEs are $\hat{\psi}_1$

and $\hat{\psi}_2$ for events of type 1 and 2, respectively. It was stated earlier that the negative first derivative of S_J with respect to t_k is equal to

$$S_J(t;\hat{\psi}_1,\hat{\psi}_2)^{\lambda}S_k(t;\hat{\psi}_k)^{-\lambda}f_k(t;\hat{\psi}_k).$$

Thus, we can calculate the cumulative incidence function for events of type k as

$$F_k(t) = \int_0^t S_J(u;\hat{\psi}_1,\hat{\psi}_2)^{\lambda}S_k(u;\hat{\psi}_k)^{-\lambda}f_k(u;\hat{\psi}_k)du. \tag{5.11}$$

While this cannot always be computed analytically, numerical integration techniques such as Simpson's rule or a similar method can be used. For a pair of consecutive failure times observed in study, t_i and t_{i+1}, we can approximate the integral on that interval as:

$$\int_{t_i}^{t_{i+1}} g(u)du \approx \frac{t_{i+1}-t_i}{6}\left[g(t_i) + 4g\left(\frac{t_i+t_{i+1}}{2}\right) + g(t_{i+1})\right], \tag{5.12}$$

where $g(t) = S_J(t;\hat{\psi}_1,\hat{\psi}_2)^{\lambda}S_k(t;\hat{\psi}_k)^{-\lambda}f_k(t;\hat{\psi}_k)$. Consecutive sums of such integrals form an approximation for F_k.

One clear advantage that joint modeling has over the other two approaches is that λ can be used to estimate a global measure of dependence between the two failure times (Oakes, 1989; Genest and Rivest, 1993). One often-used measure of dependence is Kendall's τ (1938), a measure of concordance. Genest and Rivest (1993) show that, with a copula of the form presented previously, τ can be calculated as

$$\tau = 4\int_0^1 \frac{q'(v)}{q(v)}dv + 1. \tag{5.13}$$

For the Clayton model, Kendall's τ is then equivalent to $\dfrac{\lambda-1}{\lambda+1}$, as shown in Figure 8. Under the assumptions of the direct approach, where $\lambda \to 1$, then $\tau = 0$. Additionally, under the assumptions of the cause-specific or independence approach, where $\lambda = 0$, then $\tau = -1$. This implies that for all values of $\lambda \geq 1$, there is a positive correlation between the two failure time distributions. On the contrary, for values of λ that are less than one, then the correlation is negative.

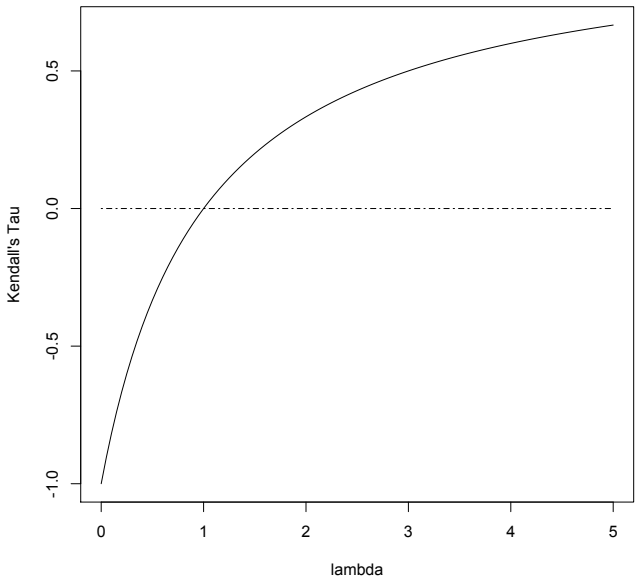


Figure 8: Plot of Kendall's τ as a function of λ

5.4 EXAMPLE WITH NSABP B-14

The joint modeling approach can be easily applied to parametric estimation of the cumulative incidence function. One data set comes from a clinical trial (B-14) conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP). This phase III trial investigated the use of tamoxifen in breast cancer patients who had negative axillary lymph nodes and estrogen receptor positive breast cancer. Here, the event of interest (type 1) was local-regional recurrences, while all other events were considered to be of type 2. There were 1453 patients randomized to the placebo arm, 1413 of which were eligible and had follow-up. Additionally, 1439 patients were randomized to the tamoxifen arm, but only 1404 were eligible with follow-up. The mean time on study for both treatment groups was approximately 20.4 years. Fisher *et al.* (1989) reported that patients treated with tamoxifen had a better outcome than those treated with placebo. For this analysis, we use a cohort of a 2817 eligible

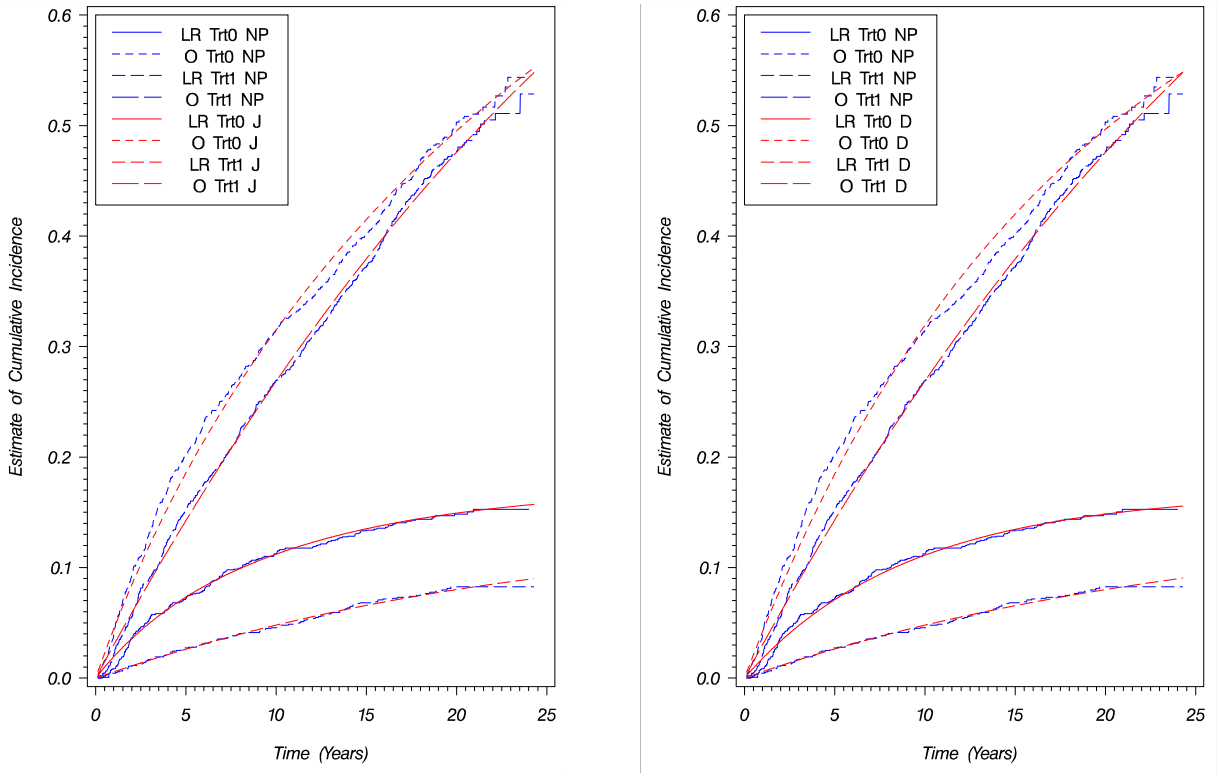
patients on either the placebo or tamoxifen treatment arms for whom there was follow-up and who had known pathological tumor size. There were 1413 patients from the placebo treatment arm and 1410 from the tamoxifen arm. All analysis was stratified on treatment group.

In this analysis, the marginal distributions are modeled using the 2-parameter Gompertz distribution for simplicity. The joint approach can of course be used with any possible marginal distributions, including the 3 and 4-parameter Gompertz models. The maximum likelihood estimates for both the joint and direct approaches are presented in Table 8. Similarly, the plot of the estimated cause-specific cumulative incidence functions are shown in Figure 9 for both approaches.

Table 8: Maximum Likelihood Estimates for Joint Modeling of NSABP B-14 by Treatment (using 2-parameter Gompertz marginal distributions)

	Joint Model		Direct Model	
	Placebo	Treatment	Placebo	Treatment
	MLE (SE)	MLE (SE)	MLE (SE)	MLE (SE)
$\hat{\kappa}_1$	0.021 (0.003)	0.006 (0.001)	0.019 (0.002)	0.006 (0.001)
$\hat{\rho}_1$	0.008 (0.023)	0.033 (0.094)	-0.102 (0.014)	-0.036 (0.017)
$\hat{\kappa}_2$	0.047 (0.003)	0.030 (0.002)	0.043 (0.003)	0.030 (0.002)
$\hat{\rho}_2$	0.002 (0.007)	0.019 (0.011)	-0.024 (0.007)	0.006 (0.007)
$\hat{\lambda}$	3.323 (1.073)	1.997 (3.101)		

It is important for the comparison of the three approaches to consider whether $\hat{\lambda}$ is significantly different from 0 (for the direct approach) or from 1 (for the cause-specific approach). To test this, a Wald-type test as presented earlier may be used. For the placebo group, $\hat{\lambda} = 3.323$ with a standard error of 1.073, while for the tamoxifen group, $\hat{\lambda} = 1.997$ with a standard error of 3.101. The Bonferroni adjusted p-values (presented in Table 9) for the tests of λ compared with the direct (cause-specific) approach are 0.0007 (0.0094) for the placebo group and 0.1299 (0.1869) for the tamoxifen group. Thus, for the placebo group, the joint model is significantly different from both the direct and cause-specific models, and



(a) Joint Modeling

(b) Direct Modeling

Figure 9: Comparison of (a) Joint and (b) Direct modeling approaches for NSABP B-14 data by treatment using the G2 distribution to model the marginals

neither additive cause-specific hazards nor additive cause-specific cumulative incidence can be assumed. For the tamoxifen group, however, the joint model is significantly different from neither the direct nor the cause-specific models, and it is unclear which assumption is more appropriate. In such a case, any of the approaches may be appropriate. Given that the joint approach is significantly different from the other two approaches for one of the treatment arms, it is more appropriate to use the joint model here.

Table 9: Wald statistics and Bonferroni-Adjusted p-values for Tests of λ

Treatment	Joint vs. Direct ($\lambda_0 = 0$)	Joint vs. Cause-Specific ($\lambda_0 = 1$)
Placebo	3.0102 (0.0007)	2.078 (0.0094)
Tamoxifen	0.644 (0.1299)	0.322 (0.1869)

5.5 DISCUSSION

The joint method presented here generalizes previously-established inference methods of inference for competing risks data. This novel approach, new for inference on cumulative incidence functions, encompasses both of the previous inference methods. Further, it allows investigators to test the appropriateness of the two possible assumptions which may be used in analysis of such data: additive cause-specific hazards, and additive cause-specific cumulative incidences. The method is robust, and may be used with any improper marginal distributions. The joint approach, however, requires the use of non-standard probability distribution function for each of the types of events. While with either the direct or the cause-specific methods, the cause-specific densities are simply calculated as the negative first derivatives of the marginal survival distributions, the joint method requires that the cause-specific densities be computed as the negative first derivatives of the joint survival distribution. Therefore, for the joint approach, the cause-specific densities are functions of all the parameters for all types of events. This fact means that the densities explicitly account for all event types simultaneously, yet makes the densities more difficult to interpret than in

either of the other approaches. Despite this disadvantage, the method for computing such densities is simple, and may be computed for any possible marginal distribution. Further, since Kendall's τ can be computed based on the results of parametric analysis using the joint method, we now have a global dependence statistic which relates tho two types of events.

6.0 DISCUSSION AND FUTURE WORK

Competing risks arise naturally in many situations where multiple failure types are present, yet investigators are interested primarily in only a subset of the event types. Parametric methods for such data are important as they may provide more precise estimates of important quantities and allow for prediction. In cases of model misspecification, however, parametric methods may estimate important quantities incorrectly, thereby introducing invisible bias in inference about the data. Flexible models are therefore needed which can accurately capture the features of the data. In the case of competing risks data, we are interested primarily in two properties of a model: ability to capture various hazard shapes, and the ability to capture improper distributions.

The first aim of this work has been to develop models which are flexible and appropriate for the analysis of competing risks data. To this end, two novel distributions have been presented, which are extensions of the Gompertz distribution to 3 and 4 parameters. Each model has the ability to capture improperness and a variety of hazard shapes, including monotone and unimodal. These novel extensions of the Gompertz distribution were developed using different approaches, and are therefore unrelated, except by their relationship to the traditional Gompertz distribution. The 3-parameter Gompertz distribution was developed explicitly so that the hazard can capture unimodal shapes and contains the regular Gompertz distribution as a special case. The other extension, the 4-parameter Gompertz, was developed on the basis of Hougaard's family of stable distributions. Both models capture a broad range of hazard shapes and both are improper in a subset of their parameter space. Because the two distributions are not closely related, one potential path for future research would be to develop a different 4-parameter distribution which contains the 3-parameter model as a special case.

While the extensions of the Gompertz distribution help to prevent model misspecification, they do not affect the inference methods used in competing risks settings. The second aim of this work has been to develop a generalized inference method which accounts for the correlation between the cause-specific marginal distributions. A copula approach was used, and the marginal distributions were modeled with a (2-parameter) Gompertz distribution, which can capture the improper nature of the cumulative incidence function. This joint modeling approach includes two important special cases which have been established previously: the direct approach and the cause-specific (independence) approach. While it is possible to test whether the proposed approach is significantly better than either of the other two approaches based on a Wald-type test on the dependence parameter, there are no current methods available to assess the goodness-of-fit of such models. This is a possible area for future research.

Nonparametric methods of estimating the joint survival function may also be of interest. Currently, the copula-graphic method (Zheng and Klein, 1995) makes it possible to nonparametrically estimate marginal distributions in cases where a joint survival function is known. A nonparametric method to estimate the joint survival function when estimates of the marginals are known would be beneficial, especially if it could be used to estimate a measure of dependence between the two failure time distributions, such as Kendall's τ, in a manner similar to that of the parametric joint approach. Investigators are also commonly interested in the effect of covariates, such as treatment group, on survival times. For this reason, the development of a covariate extension to the joint modeling approach may be of interest.

APPENDIX

SAS CODE

A.1 OPENING CODE STATEMENTS

The statements presented in this section are necessary to run all of the optimization and simulation code. The code for the incidence macro can be found in the last section of the appendix.

```
libname reg 'C:\Regression';
%include 'C:\datasteps.sas';
%inc "C:\incidence.sas"; *Reads in the CI macro from Dr. John Klein;

%let bounds4=theta&k.>0, kappa&k.>0;
%let b4=theta1>0, kappa1>0, theta2>0, kappa2>0;
%let parms4=%str(alpha&k.=1, theta&k.=1, kappa&k.=1, rho&k.=1);
%let parms14=%str(alpha1 theta1 kappa1 rho1);
%let parms24=%str(alpha2 theta2 kappa2 rho2);
%let LL4=%str(
kappa=kappa&k; rho=rho&k; theta=theta&k; alpha=alpha&k;
R = (kappa/(rho*theta))*(exp(rho*t)-1) + 1;
   Q = -(1-R**alpha)*(theta**2)/alpha;
       LL = log(kappa*theta)*d&k + rho*d&k*t + (alpha-1)*d&k*log(R) - Q);
%let LLd4=%str(R1 = (kappa1/(rho1*theta1))*(exp(rho1*t)-1) + 1;
   R2 = (kappa2/(rho2*theta2))*(exp(rho2*t)-1) + 1;
   Q1 = -(1-R1**alpha1)*(theta1**2)/alpha1;
   Q2 = -(1-R2**alpha2)*(theta2**2)/alpha2;
       f1 = d1*(-Q1 + log(theta1*kappa1) + rho1*t + (alpha1-1)*log(R1));
   f2 = d2*(-Q2 + log(theta2*kappa2) + rho2*t + (alpha2-1)*log(R2));
       surv=exp(-Q1)+exp(-Q2)-1;
if 1-d1-d2=0 then surv=1;
if 1-d1-d2^=0 and surv<0 then surv=.001;
```

```
   LL = f1 + f2 + (1-d1-d2)*log(surv));
%let LLj4=%str(R1 = (kappa1/(rho1*theta1))*(exp(rho1*t)-1) + 1;
   R2 = (kappa2/(rho2*theta2))*(exp(rho2*t)-1) + 1;
   Q1 = -(1-R1**alpha1)*(theta1**2)/alpha1;
   Q2 = -(1-R2**alpha2)*(theta2**2)/alpha2;
S1=exp(-Q1);
S2=exp(-Q2);
logh1 = log(theta1*kappa1) + rho1*t + (alpha1-1)*log(R1);
logh2 = log(theta2*kappa2) + rho2*t + (alpha2-1)*log(R2);
S12=(S1**(1-lambda)+S2**(1-lambda)-1)**(1/(1-lambda));
LL = d1*(logh1 - (1-lambda)*Q1 + lambda*log(S12)) +
d2*(logh2 - (1-lambda)*Q2  + lambda*log(S12)) + (1-d1-d2)*log(S12)
);
%let haz4=%str(R1 = (kappa1/(rho1*theta1))*(exp(rho1*t)-1) + 1;
   R2 = (kappa2/(rho2*theta2))*(exp(rho2*t)-1) + 1;
   h1=theta1*kappa1*exp(rho1*t)*(R1**(alpha1-1));
   h2=theta2*kappa2*exp(rho2*t)*(R2**(alpha2-1));
   CH1= (theta1*theta1/alpha1)*(R1**alpha1-1);
   CH2= (theta2*theta2/alpha2)*(R2**alpha2-1);
);

%let bounds3=kappa&k.>0;
%let b3=kappa1>0,kappa2>0;
%let parms3=%str(kappa&k.=1, rho&k.=-1, eta&k.=1);
%let parms13=%str(kappa1 rho1 eta1);
%let parms23=%str(kappa2 rho2 eta2);
%let LL3=%str(
kappa=kappa&k; rho=rho&k; eta=eta&k;
Q=(kappa/(eta*rho))*(exp(-eta)-exp(-eta*exp(rho*t)));
LL = d&k*(log(kappa) + rho*t - eta*exp(rho*t)) - Q);
%let LLd3=%str(Q1=(kappa1/(eta1*rho1))*(exp(-eta1)-exp(-eta1*exp(rho1*t)));
Q2=(kappa2/(eta2*rho2))*(exp(-eta2)-exp(-eta2*exp(rho2*t)));
f1 = d1*(log(kappa1) + rho1*t - eta1*exp(rho1*t) - Q1);
   f2 = d2*(log(kappa2) + rho2*t - eta2*exp(rho2*t) - Q2);
       surv=exp(-Q1)+exp(-Q2)-1;
if 1-d1-d2=0 then surv=1;
if 1-d1-d2^=0 and surv<0 then surv=.001;
   LL = f1 + f2 + (1-d1-d2)*log(surv)
);
%let LLj3=%str(Q1=(kappa1/(eta1*rho1))*(exp(-eta1)-exp(-eta1*exp(rho1*t)));
Q2=(kappa2/(eta2*rho2))*(exp(-eta2)-exp(-eta2*exp(rho2*t)));
S1=exp(-Q1);
S2=exp(-Q2);
logh1 = log(kappa1) + rho1*t - eta1*exp(rho1*t);
logh2 = log(kappa2) + rho2*t - eta2*exp(rho2*t);
S12=(S1**(1-lambda)+S2**(1-lambda)-1)**(1/(1-lambda));
LL = d1*(logh1 - (1-lambda)*Q1 + lambda*log(S12)) +
```

```
d2*(logh2 - (1-lambda)*Q2  + lambda*log(S12)) + (1-d1-d2)*log(S12)
);
%let haz3=%str(h1=kappa1*exp(rho1*t-eta1*exp(rho1*t));
h2=kappa2*exp(rho2*t-eta2*exp(rho2*t));
ch1=(kappa1/(eta1*rho1))*(exp(-eta1)-exp(-eta1*exp(rho1*t)));
ch2=(kappa2/(eta2*rho2))*(exp(-eta2)-exp(-eta2*exp(rho2*t)));
);

%let bounds2=kappa&k.>0;
%let b2=kappa1>0,kappa2>0;
%let parms2=%str(kappa&k.=1, rho&k.=-1);
%let parms12=%str(kappa1 rho1);
%let parms22=%str(kappa2 rho2);
%let LL2=%str(kappa=kappa&k; rho=rho&k;
LL = d&k*(log(kappa) + rho*t) - (kappa/rho)*(exp(rho*t)-1));
%let LLd2=%str(Q1=(kappa1/rho1)*(exp(rho1*t)-1);
Q2=(kappa2/rho2)*(exp(rho2*t)-1);
  f1 = d1*(log(kappa1) + rho1*t - Q1);
   f2 = d2*(log(kappa2) + rho2*t - Q2);
   surv=exp(-Q1)+exp(-Q2)-1;
if 1-d1-d2=0 then surv=1;
if 1-d1-d2^=0 and surv<0 then surv=.001;
  LL = f1 + f2 + (1-d1-d2)*log(surv));
%let LLj2=%str(Q1=(kappa1/rho1)*(exp(rho1*t)-1);
Q2=(kappa2/rho2)*(exp(rho2*t)-1);
S1=exp(-Q1);
S2=exp(-Q2);
logh1 = log(kappa1) + rho1*t;
logh2 = log(kappa2) + rho2*t;
S12=(S1**(1-lambda)+S2**(1-lambda)-1)**(1/(1-lambda));
LL = d1*(logh1 - (1-lambda)*Q1 + lambda*log(S12)) +
d2*(logh2 - (1-lambda)*Q2  + lambda*log(S12)) + (1-d1-d2)*log(S12)
);
%let haz2=%str(h1=kappa1*exp(rho1*t);
h2=kappa2*exp(rho2*t);
ch1=(kappa1/rho1)*(exp(rho1*t)-1);
ch2=(kappa2/rho2)*(exp(rho2*t)-1)
);
```

A.2 OPTIMIZATION CODE (DIRECT METHOD)

```
/* Produces a data set of initial values for the Gompertz
parameters, one event at a time. Using data variable from
```

```
%GompRegNLP.

k = event type
lp = line search precision, (same as LSPRECISION option
in PROC NLP) */
%macro Init(k,p, lp=.0001, method=newrap);
proc nlp data=&data tech=&method
outest=opar&k
        cov=2 vardef=n noprint
LSPRECISION=&lp MAXFUNC=5000 maxiter=2500;
       max LL;
   bounds &&bounds&p.;
       parms &&parms&p.;
   &&LL&p;
run;
Data ginit&k;
set opar&k;
where _Type_ in ('PARMS','TERMINAT');
keep _name_ &&parms&k.&p _RHS_ ;

run;
*proc print data=ginit&k noobs;
*run;
Data ginit&k;
set ginit&k;
_type_='parms';
drop _name_ _RHS_;
run;
%mend Init;

/* Produces a dataset containing initial values for
covariate parameters using PHREG. Variable data comes
from %GompRegNLP.

k = event number (1,2, ...) */
%macro Phreg(k);
ods output ParameterEstimates=beta&k;
ods listing close;
PROC PHREG data=&data;
model t*d&k(0)=&vars / ties=efron;
run;
ods listing;
ods output close;

Data beta&k;
set beta&k;
beta=trim(right(put(&k,3.)))||trim(left(put(_n_,3.)));
```

```
keep Variable Estimate beta;
run;
%mend Phreg;


/* Performs Gompertz regression with out covariates, for two event types,
assuming event indicators are assigned as d1 and d2 respectively.

Data = dataset name
p = number of parameters in Gompertz model (2, 3, or 4)
vars = covariates (leave blank if regression without covariates is requested)
phi = value of phi desired to use a specific value, or . to estimate
(ignored if no covariates)
method = optimization method (same as TECH option in PROC NLP)
    outvar = dataset containing variance estimates
ls = line search precision (LSPRECISION in PROC NLP)
maxf = maximum number of function calls permitted
maxit = maximum number of iterations permitted

A dataset containing the estimated parameters will be saved
as &data._&vars._&p.&phi  */

%macro GompRegNLP(data, p, vars=v1, phi=., method=NRRIDG, outvar=varest,
ls=.001, maxf=1000, maxit=500);
title 'Estimating Initial Values';
%Init(1,&p);
%Init(2,&p);

%if &vars= %then %do;
Data parinit(type=est);
merge ginit1 ginit2;
if kappa1^=.;
by _type_;
run;

title "&p Par. Gompertz Regression";
PROC NLP data=&data tech=&method inest=parinit outest=parms
        cov=2 vardef=n pcov LSPRECISION=&ls MAXFUNC=&maxf maxiter=&maxit;
       max LL;
       parms &&parms1&p. &&parms2&p.;
   bounds &&b&p.;
   &&LLd&p.;
    run;

%let dataname=&data._&p.est;
%end;
```

```
%else %do;
%phreg(1);
%phreg(2);
Data beta;
set beta1 beta2;
run;
proc transpose data=beta out=beta prefix=beta; id beta; run;
proc transpose data=beta out=names; run;
PROC SQL; *Getting Names of Betas;
select count(variable)
into :numvars
from beta1;
select _name_
into :betas separated by " "
from names;
select variable
into :varc separated by ""
from beta1; *variable names in one string without space;
select _name_
into :beta1-:beta%left(%eval(2*&numvars))
from names;
select variable
into :z1-:z%left(&numvars)
from beta1;
quit;

data beta;
  set beta;
keep &betas;
run;
%let zb_1 = 0;
%let zb_2 = 0;
%if &numvars=1 %then %do;
%let zb_1 = &beta1*&z1;
%let zb_2 = &beta2*&z1;
%end;
%if &numvars>1 %then %do i = 1 %to &numvars;
%let zb_1 = &zb_1 + &beta%left(&i)*&z%left(&i);
%let zb_2 = &zb_2 + &beta%left(%eval(&numvars+&i))*&z%left(&i);
%end;
%put &numvars;
%put &zb_1;
%put &zb_2;
Data ginit(type=est);
merge ginit1 ginit2 beta;
run;
%if &phi=. %then %do;
```

```
title "&p Par Gompertz Regression estimating phi";
Data bounds(type=est);
keep _type_ &&parms1&p. &&parms2&p.
phi1 phi2 beta10 beta20 &betas;
data parinit(type=est);
set ginit bounds;
if _type_='parms' then do;
phi1=-.5; phi2=-.5;
beta10=-.5; beta20=-.5;
end;
run;
PROC NLP data=&data tech=&method inest=parinit outest=parms
        cov=2 vardef=n pshort pcov
lsprecision=&ls MAXFUNC=&maxf maxiter=&maxit;
       max LL;
       parms &&parms1&p. &&parms2&p.
phi1 phi2 beta10 beta20 &betas;
bounds &&b&p.;
zb1=beta10+&zb_1;
zb2=beta20+&zb_2;
&&haz&p.;
M1=1+phi1*exp(zb1)*ch1;
M2=1+phi2*exp(zb2)*ch2;
f1=d1*(-(1+1/phi1)*log(M1)+zb1+log(h1));
f2=d2*(-(1+1/phi2)*log(M2)+zb2+log(h2));
surv=M1**(-1/phi1)+M2**(-1/phi2)-1;
if 1-d1-d2=0 then surv=1;
if 1-d1-d2^=0 and surv<0 then surv=.001;
cdf=(1-d1-d2)*log(surv);
    LL = f1 + f2 + cdf;
    run;
%end;

%else %if &phi=0 %then %do;
title "&p Par Gompertz Regression with phi= 0";
Data bounds(type=est);
keep _type_ &&parms1&p. &&parms2&p.
beta10 beta20 &betas;
data parinit(type=est);
set ginit bounds;
if _type_='parms' then do;
beta10=-.5; beta20=-.5;
end;
run;
*phi=0;
PROC NLP data=&data tech=&method inest=parinit outest=parms
        cov=2 vardef=n pshort pcov lsprecision=&ls MAXFUNC=&maxf maxiter=&maxit;
```

```
    max LL;
    parms &parms1&p. &&parms2&p.
beta10 beta20 &betas;
bounds &&b&p.;
&&haz&p.;
zb1=beta10+&zb_1;
zb2=beta20+&zb_2;
    f1=d1*(-exp(zb1)*ch1+zb1+log(h1));
f2=d2*(-exp(zb2)*ch2+zb2+log(h2));
surv=exp(-exp(zb1)*ch1)+exp(-exp(zb2)*ch2)-1;
if 1-d1-d2=0 then surv=1;
if 1-d1-d2^=0 and surv<0 then surv=.001;
cdf=(1-d1-d2)*log(surv);
   LL = f1 + f2 + cdf;
    run;
Data parms;
set parms;
phi1=&phi; phi2=&phi;
run;
%end;

%else %do;

title "&p Par Gompertz Regression with phi=" &phi;
Data bounds(type=est);
keep _type_ &&parms1&p. &&parms2&p.
beta10 beta20 &betas;
data parinit(type=est);
set ginit bounds;
if _type_='parms' then do;
beta10=-.5; beta20=-.5;
end;
run;*phi=anything else;
PROC NLP data=&data tech=&method inest=parinit outest=parms
        cov=2 vardef=n pshort pcov
lsprecision=&ls MAXFUNC=&maxf maxiter=&maxit;
       max LL;
       parms &&parms1&p. &&parms2&p.
beta10 beta20 &betas;
zb1=beta10+&zb_1;
zb2=beta20+&zb_2;
&&haz&p.;
M1=1+&phi*exp(zb1)*ch1;
M2=1+&phi*exp(zb2)*ch2;
f1=d1*(-(1+1/&phi)*log(M1)+zb1+log(h1));
f2=d2*(-(1+1/&phi)*log(M2)+zb2+log(h2));
surv=M1**(-1/&phi)+M2**(-1/&phi)-1;
```

```
if 1-d1-d2=0 then surv=1;
if 1-d1-d2^=0 and surv<0 then surv=.001;
cdf=(1-d1-d2)*log(surv);
    LL = f1 + f2 + cdf;
    run;
Data parms;
set parms;
phi1=&phi; phi2=&phi;
run;
%end;

%if &phi^=. %then %let dataname=&data._&varc._&p.&phi;
%if &phi=. %then %let dataname=&data._&varc._&p.est;
%end;

data &dataname;
set parms;
where _Type_='PARMS' or _Type_='STDERR';
drop _tech_ _name_ _rhs_ _iter_;
run;

data &outvar;
set parms;
where _Type_='COV2: H' and _rhs_^=.;
drop _tech_ _type_ _iter_ _rhs_;
run;
%mend GompRegNLP;
```

A.3 OPTIMIZATION CODE (JOINT METHOD)

```
/* Produces a data set of initial values for the Gompertz
parameters, one event at a time. Using data variable from
%GompJointNLP.

k = event type
lp = line search precision, (same as LSPRECISION option
in PROC NLP)
*/
option mprint;
%macro Init(k,p, lp=.0001, method=newrap); *by trt version;
%put k= &k;
%put p= &p;
proc sort data=&data; by trt; run;
proc nlp data=&data tech=&method
```

```
outest=opar&k
        cov=2 vardef=n pshort pcov
LSPRECISION=&lp MAXFUNC=5000 maxiter=2500;
       max LL;
   by trt;
   bounds &&bounds&p.;
       parms &&parms&k.&p.;
   &&LL&p;
run;
Data ginit&k;
set opar&k;
where _Type_ in ('PARMS');
keep TRT &&parms&k.&p _RHS_ ;

run;
*proc print data=ginit&k noobs;
*run;
Data ginit&k;
set ginit&k;
_type_='parms';
drop _name_ _RHS_;
run;
%mend Init;

/* Performs joint modeling of the Gompertz distribution, for two event types,
assuming event indicators are assigned as d1 and d2 respectively.

Data = dataset name
p = number of parameters in Gompertz model
method = optimization method (same as TECH option in PROC NLP)
outparms = name of dataset containing parameter estimates
ls = line search precision (LSPRECISION in PROC NLP)
maxf = maximum number of function calls permitted
maxit = maximum number of iterations permitted*/
%macro GompJointNLP(data, p, method=NEWRAP, outparms=parmest,
ls=.1, maxf=10000, maxit=10000);
title 'Estimating Initial Values';
%Init(1,&p);
%Init(2,&p);

Data parinit(type=est);
merge ginit1 ginit2;
if kappa1^=.;
lambda=2;
run;

title "&p Par. Joint Gompertz Regression";
```

```
PROC NLP data=&data tech=&method inest=parinit outest=parms
        cov=2 vardef=n pcov LSPRECISION=&ls MAXFUNC=&maxf maxiter=&maxit;
       max LL;
   by trt;
       parms &&parms1&p. &&parms2&p. lambda;
   bounds &&b&p., lambda>0;
   &&LLj&p.;
    run;

data &outparms;
set parms;
where _Type_='PARMS' or _type_='STDERR';
drop _tech_ _name_ _rhs_ _iter_;
run;


data tmp;
set &outparms;
where _type_='PARMS';
drop _type_;
run;
%mend GompJointNLP;
```

A.4 SIMULATION CODE

```
/* Produces a data set of initial values for the Gompertz
parameters, one event at a time. Using data variable from
%gompertznlp.

k = event type
lp = line search precision, (same as LSPRECISION option
in PROC NLP)
*/
%macro Init(k,p, lp=.0001, method=newrap); *by trt version;
%put k= &k;
%put p= &p;
*proc sort data=&data; *by trt; *run;
proc nlp data=&data tech=&method
outest=opar&k
        cov=2 vardef=n noprint
LSPRECISION=&lp MAXFUNC=5000 maxiter=2500;
       max LL;
   bounds &&bounds&p.;
       parms &&parms&k.&p.;
```

```
   &&LL&p;
run;
Data ginit&k;
set opar&k;
where _Type_ in ('PARMS');
keep &&parms&k.&p _RHS_ ;

run;
Data ginit&k;
set ginit&k;
_type_='parms';
drop _name_ _RHS_;
run;
%mend Init;

%let path=C:\;
%macro GompNLPsim(data, method=NMSIMP, ls=.1, maxf=2000, maxit=1000);
title 'Estimating Initial Values';
%Init(1,&p);
%Init(2,&p);

Data parinit(type=est);
merge ginit1 ginit2;
if kappa1^=.;
run;

title "&p Par. Gompertz Regression";
PROC NLP data=&data tech=&method inest=parinit outest=parms
        cov=2 vardef=n noprint LSPRECISION=&ls MAXFUNC=&maxf maxiter=&maxit;
       max LL;
       parms &&parms1&p. &&parms2&p.;
   bounds &&b&p.;
   &&LLD&p.;
    run;

data parmest;
set parms;
where _Type_='PARMS';
n_sim=&i;
drop _tech_ _type_ _name_ _rhs_ _iter_;
run;

data conv;
set parms;
where _Type_='TERMINAT';
n_sim=&i;
CONV=_name_;
```

```
keep n_sim CONV;
run;

proc sort data=parmest; by n_sim; run;
proc sort data=conv; by n_sim; run;

data sim;
merge parmest conv;
by n_sim;
run;

%if %EVAL(&i=1) %then %do;
data outparms;
set sim;
run;
%end;
%else %if %EVAL(&i>1) %then %do;
Proc APPEND data=sim base=outparms; run;
%end;
%mend GompNLPsim;


%MACRO SIMLOOP(p, n_sim, cens, n=1000, method=NMSIMP);
%do i = 1 %to &n_sim;
%put "***ITERATION INFORMATION: "&p &i &cens;
%if (&i=%eval(&n_sim/2) or &i=%eval(&n_sim/4) or &i=%eval(3*&n_sim/4))
and &n_sim>10 %then %do;
DM 'CLEAR LOG';
%end;

%if &p=2 %then %do;
proc iml;
n=&n; *n=1000;
i=&i; *i=1;
seed_d1=15213;
seed_d2=44691;
seed_c1=40489;
seed_c2=8001;
k1=0.0187725206; k2=0.0433120015;
r1=-0.10155527; r2=-0.024246275;
cens05=275;
cens10=175;
cens20=80;
if &cens=.05 then cmax=cens05;
if &cens=.10 then cmax=cens10;
if &cens=.20 then cmax=cens20;
LimF1=(1-exp(k1/r1))*(r1<0) + 1*(r1>0);
```

```
LimF2=(1-exp(k2/r2))*(r2<0) + 1*(r2>0);
u1=uniform(repeat(seed_d1*i,n/2,1))*LimF1;
t1=log(1-(r1/k1)*log(1-U1))/r1;
u2=uniform(repeat(seed_d2*i,n/2,1))*LimF2;
t2=log(1-(r2/k2)*log(1-U2))/r2;
c1=uniform(repeat(seed_c2*i,n/2,1))*cmax;
c2=uniform(repeat(seed_c2*i,n/2,1))*cmax;
d1=(t1<c1);
d2=2*(t2<c2);
cens_p=1-(d1//(d2/2))[:];
t=(t1||c1)[,><]//(t2||c2)[,><];
d1=d1//j(n/2,1,0);
d2=j(n/2,1,0)//(d2/2);
data=t||d1||d2;
create thisdata var{t d1 d2};
append from data;
free data;
quit;
%end;

%if &p=3 %then %do;
proc iml;
n=&n; *n=1000;
i=&i; *i=1;
seed_d1=15213;
seed_d2=44691;
seed_c1=40489;
seed_c2=8001;
k1=0.0208374818; k2=0.0143013467;
r1=-0.107226626; r2=-0.01249532;
e1=0.1179076961; e2=-1.117743967;
cens05=390;
cens10=185;
cens20=82;
if &cens=.05 then cmax=cens05;
if &cens=.10 then cmax=cens10;
if &cens=.20 then cmax=cens20;
LimF1=(1-exp(-(k1/(e1*r1))*(exp(-e1)-1)))*(r1<0) +
(r1>0)*((e1>0)*(1-exp(-(k1*exp(-e1))/(r1*e1)))+(e1<0)*(1));
LimF2=(1-exp(-(k2/(e2*r2))*(exp(-e2)-1)))*(r2<0) +
(r2>0)*((e2>0)*(1-exp(-(k2*exp(-e2))/(r2*e2)))+(e2<0)*(1));
u1=uniform(repeat(seed_d1*i,n/2,1))*LimF1;
t1=log(-log(exp(-e1)+(r1*e1/k1)*log(1-U1))/e1)/r1;
u2=uniform(repeat(seed_d2*i,n/2,1))*LimF2;
t2=log(-log(exp(-e2)+(r2*e2/k2)*log(1-U2))/e2)/r2;
c1=uniform(repeat(seed_c2*i,n/2,1))*cmax;
c2=uniform(repeat(seed_c2*i,n/2,1))*cmax;
```

```
d1=(t1<c1);
d2=2*(t2<c2);
cens_p=1-(d1//(d2/2))[:];
t=(t1||c1)[,><]//(t2||c2)[,><];
d1=d1//j(n/2,1,0);
d2=j(n/2,1,0)//(d2/2);

data=t||d1||d2;
create thisdata var{t d1 d2};
append from data;
free data;
quit;
%end;


%if &p=4 %then %do;
proc iml;
n=&n; *n=1000;
i=&i; *i=1;
seed_d1=15213;
seed_d2=44691;
seed_c1=40489;
seed_c2=8001;
a1 = 0.2;
a2 = 0.1;
th1 = 0.4 ;
th2 = 0.5;
k1 = 0.15;
k2 = 4;
r1 = -0.1;
r2 = -0.08;
cens05=69;
cens10=36;
cens20=16;
if &cens=.05 then cmax=cens05;
if &cens=.10 then cmax=cens10;
if &cens=.20 then cmax=cens20;
LimF1=(1-exp((-th1*th1/a1)*((1-(k1/(r1*th1)))**a1 - 1)))*(r1<0)
+ ((1-exp(th1*th1/a1))*(a1<0) + (1)*(a1>0))*(r1>0);
LimF2=(1-exp((-th2*th2/a2)*((1-(k2/(r2*th2)))**a2 - 1)))*(r2<0)
+ ((1-exp(th2*th2/a2))*(a2<0) + (1)*(a2>0))*(r2>0);
u1=uniform(repeat(seed_d1*i,n/2,1))*LimF1;
t1=log((th1*r1/k1)*((1-(a1/(th1#th1))*log(1-u1))##(1/a1)-1)+1)/r1;
u2=uniform(repeat(seed_d2*i,n/2,1))*LimF2;
t2=log((th2*r2/k2)*((1-(a2/(th2#th2))*log(1-u2))##(1/a2)-1)+1)/r2;
c1=uniform(repeat(seed_c2*i,n/2,1))*cmax;
c2=uniform(repeat(seed_c2*i,n/2,1))*cmax;
```

```
d1=(t1<c1);
d2=2*(t2<c2);
cens_p=1-(d1//(d2/2))[:];
t=(t1||c1)[,><]//(t2||c2)[,><];
d1=d1//j(n/2,1,0);
d2=j(n/2,1,0)//(d2/2);

data=t||d1||d2;
create thisdata var{t d1 d2};
append from data;
free data;
quit;
%end;

%GompNLPsim(thisdata, method=&method, ls=.001, maxf=5000, maxit=3000);
%end;

DM 'CLEAR LOG';

%let censi=%sysevalf(100*&cens); *as an integer;
Data outparms&method.&n_sim._&censi.;
set outparms;
run;

Data converged;
set outparms;
where CONV ^='PROBLEMS';
run;

PROC MEANS data=converged;
var &&parms1&p. &&parms2&p.;
*var kappa1 rho1 kappa2 rho2;
output out=means;
run;

proc transpose data=means out=means; id _stat_; run;

Data means;
set means;
where _name_ ^in ("_FREQ_" "_TYPE_");
drop min max;
run;

ods rtf file="&path.\SimResults&method.&n._&p.sim._&censi..rtf" style=minimal;
proc iml;
use means;
read all var {mean} into mean;
```

```
    read all var {std} into std;
read all var {n} into nlist;
p=&p;
n=min(nlist);
print n p;
if p=2 then do;
k1=0.0187725206; k2=0.0433120015;
r1=-0.10155527; r2=-0.024246275;
base=k1//r1//k2//r2;
end;
if p=3 then do;
k1=0.0208374818; k2=0.0143013467;
r1=-0.107226626; r2=-0.01249532;
e1=0.1179076961; e2=-1.117743967;
base=k1//r1//e1//k2//r2//e2;
end;
if p=4 then do;
a1 = 0.2;
a2 = 0.1;
th1 = 0.4 ;
th2 = 0.5;
k1 = 0.15;
k2 = 4;
r1 = -0.1;
r2 = -0.08;
base=k1//r1//a1//th1//k2//r2//e2//a2//th2;
end;
*base = {0.2, 0.4, 0.15, -0.1, 0.1, 0.5, 4, -0.08};
bias=mean-base;
Z=bias/std;
pz2=probnorm(z)/2;
Pval=(Z<=0)#(probnorm(z)) + (Z>0)#(1-probnorm(z));
print "OPTIMIZATION METHOD: &method";
print "CENSORING: &cens";
print "NUMBER OF SIMULATIONS: &n_sim";
print "NUMBER SIMULATIONS CONVERGED: " n;
print base mean bias std z pz2 pval;
quit;
ods rtf close;
%MEND SIMLOOP;
```

A.5 MACRO FOR NONPARAMETRIC ESTIMATES OF THE CUMULATIVE INCIDENCE FUNCTION

This macro can be found at http://www.biostat.mcw.edu/software/SoftMenu.html.

```
%macro incid(data,group,relp,trm,time,out=);
/* arguments are: dataset name,
                  group variable(1,2,...),
                  relapse indicator(1-yes, 0-no)
                  trm indicator(1-yes, 0-no),
                  dataset name for relapse incidence
                  dataset name for trm incidence  */
data lc_one; set &data;
a=&time;
b=&relp;
c=&trm;
d=&group;
keep a b c d;

 proc sort data=lc_one; by descending a;
 /*proc sort data=lc_one; by d;*/
 proc iml;
  use lc_one;
read all into x;
  n=nrow(x);
  ngrp=max(x[ ,4]);
  gnum=J(1,ngrp,0);
  do k=1 to n;
  gnum[1,x[k,4]] =  gnum[1,x[k,4]]+1;
  end;
t=J(1,n,0);
t[1,1]=x[1,1];
ntime=1;
tnow=x[1,1];
do j=2 to n;
if x[j,1] <tnow then do;
ntime=ntime+1;
t[1,ntime]=x[j,1];
tnow=x[j,1];
end;
end;

relap=J(ntime,ngrp,0);
trm=J(ntime,ngrp,0);
atrisk=J(ntime,ngrp,0);

do k=1 to ntime;
```

```
do j=1 to n;
ax=x[j,1];
at=t[1,k];
ag=x[j,4];
ad=x[j,3];
ar=x[j,2];
if x[j,1] =t[1,k] then do;
relap[k,ag]=relap[k,ag]+ar;
trm[k,ag]=trm[k,ag]+ad;
end;
if x[j,1] >= t[1,k] then atrisk[k,ag]=atrisk[k,ag]+1;
end;
end;

lfs=J(ntime,ngrp,-1);
  ci_rel=J(ntime,ngrp,-1);
  ci_trm=J(ntime,ngrp,-1);
   vci_rel=J(ntime,ngrp,-1);
  vci_trm=J(ntime,ngrp,-1);
index=J(ntime,1,0);
  tt=t(t[1,1:ntime]);

do ig=1 to ngrp;
p=1;
cr=0;
cd=0;
do j=1 to ntime ;
index[j,1]=j;
k=ntime-j+1;
  if atrisk[k,ig] >0 then do;
cr=cr+relap[k,ig]*p/atrisk[k,ig];
cd=cd+trm[k,ig]*p/atrisk[k,ig];
p=p*(1-(trm[k,ig]+relap[k,ig])/atrisk[k,ig]);
lfs[k,ig]=p;
ci_rel[k,ig]=cr;
ci_trm[k,ig]=cd;
end;
else do;
lfs[k,ig]=.;
ci_rel[k,ig]=.;
ci_trm[k,ig]=.;
end;
end;
do j=1 to ntime;
know=ntime-j+1;
vr=0;  vd=0;
if ci_rel[know,ig] = . then do;
```

```
vci_rel[know,ig]=.;
vci_trm[know,ig]=.;
end;
else do;
do k=1 to ntime;
jnow=ntime-k+1;
if tt[jnow,1] <= tt[know,1] then do;
wr=(trm[jnow,ig]+relap[jnow,ig])/atrisk[jnow,ig]**2;
wr=wr*lfs[jnow,ig]*(ci_rel[know,ig]-ci_rel[jnow,ig])**2;
q=(relap[jnow,ig]/atrisk[jnow,ig]**2)*lfs[jnow,ig]**2;
q=q*(1-2*(ci_rel[know,ig]-ci_rel[jnow,ig]));
vr=vr+wr+q;
wd=(trm[jnow,ig]+relap[jnow,ig])/atrisk[jnow,ig]**2;
wd=wd*lfs[jnow,ig]*(ci_trm[know,ig]-ci_trm[jnow,ig])**2;
q=trm[jnow,ig]/atrisk[jnow,ig]**2*lfs[jnow,ig]**2;
q=q*(1-2*(ci_trm[know,ig]-ci_trm[jnow,ig]));
vd=vd+wd+q;
end;
end;
end;
 vci_rel[know,ig]=sqrt(vr);
 vci_trm[know,ig]=sqrt(vd);
end;


end;
nn=ngrp*ntime;
yout=j(nn,6,0);
k=0;
do is=1 to ngrp;
do it=1 to ntime;
k=k+1;
yout[k,1]=tt[it,1];
yout[k,2]=is;
yout[k,3]=ci_rel[it,is];
yout[k,4]=vci_rel[it,is];
yout[k,5]=ci_trm[it,is];
yout[k,6]=vci_trm[it,is];
end;  end;

create dout from yout;
append from yout;
close dout;
quit;
data io;  set dout;
time=col1;
group=col2;
```

```
CI1=col3;
SE_CI1=col4;
CI2=col5;
SE_CI2=col6;
if CI1 =. then se_ci1=.;
if CI2=. then se_ci2=.;
drop col1-col6;
proc sort data=io;  by time;
proc sort data=io;  by group;
*proc print data=io;
data &out;  set io;
run;
%mend;
```

BIBLIOGRAPHY

Akaike, H. (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, **19**, 716–723.

Benichou, J. and Gail, M. H. (1990) Estimates of absolute cause-specific risk in cohort studies. *Journal of the Royal Statistical Society, Series A*, **46**, 813–826.

Bryant, J. and Dignam, J. A. (2004) Semiparametric models for cumulative incidence functions. *Biometrics*, **60**, 182–190.

Bunea, C. and Bedford, T. (2002) The effect of model uncertainty on maintenance optimization. *Reliability, IEEE Transactions on*, **51**, 486–493.

Casella, G. and Berger, R. L. (2002) *Statistical Inference*. Duxbury, 2nd edn.

Chen, B. E., Kramer, J. L., Greene, M. H. and Rosenberg, P. S. (2007) Competing risks analysis of correlated failure time data. *Biometrics*.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995) Analysis of transformation models with censored data. *Biometrika*, **82**, 835–845.

Cheng, Y. and Fine, J. (2008) Nonparametric estimation of cause-specific cross hazard ratio with bivariate competing risks data. *Biometrika*, **95**, 233.

Cheng, Y., Fine, J. and Kosorok, M. (2007) Nonparametric association analysis of bivariate competing risks data. *Journal of the American Statistical Association*, **102**, 1407 – 1415.

Clayton, D. (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.

Cox, D. R. (1959) The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society, Series B*, **21**, 411–421.

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

Dabrowska, D. M. and Doksum, K. A. (1988) Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, **83**, 744–749.

Escarela, G. and Carrière, J. (2003) Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, **12**, 333 – 349.

Fine, J. P. (1999) Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **61**, 817–830.

Fine, J. P. and Gray, R. J. (1999) A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.

Fisher, B., Costantino, J., Redmond, C., Poisson, R., Bowman, D., Couture, J., Dimitrov, N., Wolmark, N., Wickerham, D., Fisher, E., Margolese, R., Robidoux, A., Shibata, H., Terz, J., Paterson, A., Feldman, M., Farrar, W., Evans, J., Lickley, H. and Ketner, M. (1989) A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *New England Journal of Medicine*, **320**, 479–484.

Fréchet, M. (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. de Lyon III ser. fasc. 14A*, 53–77.

Gail, M. (1975) A review and critique of some models used in competing risk analysis. *Biometrics*, **31**, 209–222.

Garg, M. L., Rao, B. R. and Redmond, C. K. (1970) Maximum-likelihood estimation of the parameters of the gompertz survival function. *Applied Statistics*, **19**, 152–159.

Genest, C. and MacKay, J. (1986) The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, **40**, 280–283.

Genest, C. and Rivest, L.-P. (1993) Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034–44.

Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513–580.

Gooley, T. A., Leisenring, W., Crowley, J. and Storer, B. E. (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, **18**, 695–706.

Gray, R. J. (1988) A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, **16**, 1141–1154.

Hoeffding, W. (1940) Maßtabinvariante Korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, **5**, 179–233.

[Reprinted as Scale-Invariant correlation theory. In: Fisher NI, Sen PK (eds) The Collected Works of Wassily Hoeffding. Springer, New York pp 57–107].

Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.

Jeong, J.-H. (2006) A new parametric family for modelling cumulative incidence functions: application to breast cancer data. *Journal of the Royal Statistical Society, Series A*, **169**, 289–303.

Jeong, J.-H. (2008) Parametric inference on competing risks. In *ENAR Meeting*.

Jeong, J.-H. and Fine, J. P. (2006) Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **55**, 187–200.

Jeong, J.-H. and Fine, J. P. (2007) Parametric regression on cumulative incidence function. *Biostatistics*, **8**, 184–196.

Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **35**, 457–481.

Kendall, M. G. (1938) A new measure of rank correlation. *Biometrika*, **30**, 81–03.

Korn, E. L. and Dorey, F. J. (1992) Applications of crude incidence curves. *Statistics in Medicine*, **11**, 813–829.

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

Medical College of Wisconsin Division of Biostatistics (2007) *SAS Macro For Cumulative Incidence Functions*. URL http://www.biostat.mcw.edu/software/incidence.txt.

Mudholkar, G. S., Srivastava, D. K. and Kollia, G. D. (1996) A generalization of the weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, **91**, 1575–1583.

Nelder, J. and Mead, R. (1965) A simplex method for function minimization. *Computer Journal*, **7**, 308–313.

Nelsen, R. B. (2006) *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2nd edn. ISBN: 0-387-28659-4.

Oakes, D. (1982) A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B*, **44**, 414–422.

Oakes, D. (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, **84**, 487–493.

Pepe, M. S. and Mori, M. (1993) Kaplan-meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, **12**, 737–751.

Prentice, R. L., Kalbfleisch, J. D., A. V. Peterson, J., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

SAS Institute Inc. (2004) *SAS/STAT 9.1 Users Guide*. SAS Institute Inc., Gary, NC. URL www.sas.com/products/stat/index.html.

Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

Tsiatis, A. (1975) A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, **72**, 20–22.

Zheng, M. and Klein, J. P. (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, **82**, 127–138.