The International Journal of Biostatistics

Volume 7, Issue 1

2011

Article 26

A First Passage Time Model for Long-Term Survivors with Competing Risks

Ruimin Xu, *University of Guelph*

Paul D. McNicholas, *University of Guelph*

Anthony F. Desmond, *University of Guelph*

Gerarda A. Darlington, *University of Guelph*

Recommended Citation:

Xu, Ruimin; McNicholas, Paul D.; Desmond, Anthony F.; and Darlington, Gerarda A. (2011) "A First Passage Time Model for Long-Term Survivors with Competing Risks," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 26.

DOI: 10.2202/1557-4679.1224

A First Passage Time Model for Long-Term Survivors with Competing Risks

Ruimin Xu, Paul D. McNicholas, Anthony F. Desmond, and Gerarda A. Darlington

Abstract

We investigate a competing risks model, using the specification of the Gompertz distribution for failure times from competing causes and the inverse Gaussian distribution for failure times from the cause of interest. The expectation-maximization algorithm is used for parameter estimation and the model is applied to real data on breast cancer and melanoma. In these applications, our models compare favourably with existing techniques. The proposed method provides a useful technique that may be more broadly applicable than existing alternatives.

KEYWORDS: competing risks, EM algorithm, inverse Gaussian distribution, long-term survivors, Louis' method

Author Notes: This work was supported by respective NSERC Discovery Grants held by McNicholas, Desmond, and Darlington, and by the University Research Chair in Computational Statistics held by McNicholas.

1 Introduction & Background

In medical studies, a major question of interest is the proportion of patients who are cured of a given disease and never experience an event of interest, such as disease reoccurrence. In addition, the survival function of those patients not cured is also of interest. The cure rate model originally proposed by Boag (1949) is often used to model data sets for which cure is a possibility. Farewell (1982, 1986) also investigated this model, and obtained results quite different than from standard techniques, such as the Cox proportional hazards model (Cox, 1972). In many medical data sets, however, a sizeable proportion of individuals under study may die (fail) from a cause other than the cause of interest. In such studies, subjects under investigation are exposed to more than one possible event. A two-component competing risks (CR) model, with one component representing the lifetimes for the competing causes and the other component representing the lifetimes for the cause of interest, is appropriate for modeling such data.

Let T denote the time to failure either from the cause of interest or from a competing cause. The survival function for the CR model is then given by

$$S(t) = p_1 S_1(t) + (1 - p_1) S_2(t),$$

where $S_1(t)$ denotes the conditional survival function for failure due to a competing cause, $S_2(t)$ denotes the conditional survival function for failure due to the cause of interest, p_1 is the probability of failure from a competing cause, and $1 - p_1$ is the probability of failure from the cause of interest in the presence of other risks.

The idea of studying survival time data based on the notion of an underlying stochastic process leading to ‘failure’ has been studied recently by Aalen and co-workers (see Aalen and Gjessing, 2001, for example), and earlier by Whitmore in a series of papers (see Lee and Whitmore, 2006, and the references therein for an excellent review). A stimulating discussion of this ‘process’ point of view is given in the recent text by Aalen, Borgan, and Gjessing (2008). A prior discussion of the use of stochastic process approaches is given in the text by Andersen, Borgan, Gill, and Keiding (1993). While this latter text is canonical in developing the counting process paradigm, the emphasis has been to a large extent on the classical Cox proportional hazards model. Lee and Whitmore (2006) consider a family of stochastic models for the underlying process leading to the failure of interest. Here, we study only one of these, namely a Wiener process $X(t)$, such that when $X(t)$ reaches a barrier at c , say, the failure of interest occurs. Thus, $T_2 = \inf\{u \geq 0 : X(u) = c\}$ is the time to failure due to the cause of interest; this is well known to have an inverse Gaussian (IG) distribution, if the drift is positive (i.e., towards the barrier). Here we assume $X(0) = 0$ and $c > 0$.

Thus far, only a single failure event of interest has been considered. In the analysis of event time data, two or more competing risks are often operating and the analysis must allow for this. Kalbfleisch and Prentice (2002) give a good discussion of various approaches to the CR problem, where time and cause of death are recorded for each individual. If one or more competing risks exist, considering a multidimensional Wiener process to model the various risks simultaneously is possible. Such an approach is discussed, for example, by Lee and Whitmore (2006). In the two illustrative examples discussed in Section 3, death due to breast cancer and recurrence of melanoma are the respective events of interest, and the competing risk is ‘cure’ or death from other causes. The two classic data sets we use have been the subject of many publications, but for the most part the competing risks aspect has been either ignored or cursorily treated.

An alternative approach to competing risks (see, for example, Kalbfleisch and Prentice, 2002) postulates the existence of latent failure times (T_1, T_2) , with a corresponding joint survivor function $S(t_1, t_2)$. Only $T = \min\{T_1, T_2\}$, along with the cause of death, is observed. With this approach, independence of the competing risks is frequently assumed. Even if a dependence model for the competing risks is used, there is a well known identifiability issue. Cox (1959) and Tsiatis (1975) show that for every distribution having dependent T_i , there exists a distribution with independent T_i that yields the same distribution of observable failure times T . Kalbfleisch and Prentice (2002) argue against the latent failure time models commonly used in the competing risks literature and recommend against them, except in exceptional circumstances. In contrast, the mixture model we use here does not need to make essentially untestable assumptions about the nature of the dependence between the competing risks, unlike the latent-failure approach. Moreover, component survival functions and associated mixing proportions can be estimated directly from the data, which is again in contrast to the latent failure modeling approach. Another useful explanation is in terms of a gamble at time $t = 0$ between dying from the cause of scientific interest with probability $1 - p_1$ versus dying from other causes with probability p_1 ; this is essentially the interpretation by Larson and Dinse (1985). Historically, this is reminiscent of arguments of Daniel Bernoulli. Our use of mixture models for competing risks is not new. Other examples of earlier work include Larson and Dinse (1985), Gordon (1990), and Ng and McLachlan (1998); see also McLachlan and Peel (2000) for other advantages of the mixture model approach.

2 Methodology

2.1 Model Assumption

We use the Gompertz distribution to model deaths from other causes. The Gompertz distribution reflects the intuition that ‘cured’ individuals should follow the hazard behaviour of the general population who have not experienced the event of interest. For example, Andersen et al. (1993) compare the failure behaviour for the CR group in the melanoma data with that for the general Danish population, which is commonly taken to be Gompertz beyond a certain age. The authors test this assumption using log-rank statistics and find no significant difference (Andersen et al., 1993, page 338).

The probability density function for the Gompertz distribution is given by

$$f_1(t) = \kappa e^{\alpha t} \exp\{-(\kappa/\alpha)(e^{\alpha t} - 1)\},$$

for $t \geq 0$, where $\kappa, \alpha > 0$. The corresponding survival function is then given by

$$S_1(t) = \exp\{-(\kappa/\alpha)(e^{\alpha t} - 1)\}.$$

With $\alpha = 0$, the Gompertz distribution reduces to an exponential distribution; with $\alpha < 0$, the Gompertz distribution is improper.

The distribution of lifetimes for the cause of interest is assumed to follow an IG distribution with probability density function

$$f_2(t | \delta, \nu) = \frac{1}{\sqrt{2\nu\pi t^3}} \exp\left\{-\frac{(1 - \delta t)^2}{2\nu t}\right\},$$

for $t \geq 0$ where $\delta, \nu > 0$. The corresponding survival function has the form

$$S_2(t) = 1 - \Phi\left[\frac{\delta t - 1}{\sqrt{\nu t}}\right] - e^{2\delta/\nu} \Phi\left[-\left(\frac{\delta t + 1}{\sqrt{\nu t}}\right)\right],$$

where $\Phi(z)$ is the standard normal cumulative distribution function.

The IG distribution has many interesting properties. Unlike the Gompertz distribution, whose hazard rate either increases, decreases, or remains constant, the IG distribution has a non-monotonic hazard rate; it first increases, then decreases, approaching a non-zero constant as the lifetime goes to infinity. This property is very appealing for data for which the occurrence of early events dominates the lifetimes, as is shown in Section 3.

2.2 Parameter Estimation

Parameter estimates in the CR model are obtained using the maximum likelihood method. Let the observed data be of the form (t_i, D_i) for the i th individual, $i = 1, 2, \dots, n$, where $D_i=2$ implies that the i th individual was observed to fail from the cause of interest at time t_i during the follow-up period, $D_i=1$ implies that the i th individual was observed to fail from a competing cause at time t_i and $D_i=0$ implies that the i th individual was right censored at time t_i . The log-likelihood function is given by

$$\begin{aligned} \log \mathcal{L}(p_1, \alpha, \kappa, \delta, \nu | \mathbf{t}) = & \sum_{i:D_i=1} \log p_1 + \sum_{i:D_i=1} \log f_1(t_i | \alpha, \kappa) + \sum_{i:D_i=2} \log(1 - p_1) \\ & + \sum_{i:D_i=2} \log f_2(t_i | \delta, \nu) + \sum_{i:D_i=0} \log [p_1 S_1(t_i | \alpha, \kappa) + (1 - p_1) S_2(t_i | \delta, \nu)]. \end{aligned}$$

The expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) is used to find maximum likelihood estimates of the model parameters. Two latent variables are introduced for right-censored times. The first latent variable is an indicator variable z_i depending on whether the i th individual eventually fails from a competing cause or from the cause of interest. The second latent variable represents the failure times t_{2i} for the i th individual, conditional on that individual eventually failing from the cause of interest.

In the E-step, the conditional expected value of the complete-data log-likelihood is computed based on the observed data and the current parameter estimates. Given the parameters $\Psi^{(k)} = (p_1^{(k)}, \alpha^{(k)}, \kappa^{(k)}, \delta^{(k)}, \nu^{(k)})$ at the k th iteration, the expected complete-data log-likelihood is given by

$$\begin{aligned} Q(\Psi | \Psi^{(k)}) = & \sum_{i:D_i=1} \log p_1^{(k)} + \sum_{i:D_i=1} \log f_1(t_i | \alpha^{(k)}, \kappa^{(k)}) + \sum_{i:D_i=2} \log(1 - p_1^{(k)}) \\ & + \sum_{i:D_i=2} \log f_2(t_i | \delta^{(k)}, \nu^{(k)}) + \sum_{i:D_i=0} (1 - \mathbb{E}^{(k)}[z_i]) \log p_1^{(k)} \\ & + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i] \log(1 - p_1^{(k)}) + \sum_{i:D_i=0} (1 - \mathbb{E}^{(k)}[z_i]) \log S_1(t_i | \alpha^{(k)}, \kappa^{(k)}) \\ & + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i] \mathbb{E}^{(k)}[\log f_2(t_{2i} | \delta^{(k)}, \nu^{(k)})], \end{aligned}$$

where $\mathbb{E}^{(k)}[z_i]$ denotes the conditional probability that individual i dies from the cause of interest given right-censored time t_i . We have

$$\mathbb{E}^{(k)}[z_i] = \mathbb{E}^{(k)}[z_i | t_i] = \frac{(1 - p_1^{(k)}) S_2(t_i | \delta^{(k)}, \nu^{(k)})}{p_1^{(k)} S_1(t_i | \alpha^{(k)}, \kappa^{(k)}) + (1 - p_1^{(k)}) S_2(t_i | \delta^{(k)}, \nu^{(k)})}.$$

The necessary expected values are given by Whitmore (1983) based on the right-censored time t_i . In the M-step, the parameter estimates are obtained by maximizing the expected value of complete-data log-likelihood. Details are given in Appendix A.

The estimated standard errors for these parameter estimates are obtained using the method of Louis (1982). That is, the observed information matrix is the difference between the conditional expectation of the complete-data information matrix given the observed data and the covariance of the complete-data score statistics with respect to the missing data. The estimated standard errors for the parameter estimates are then the square roots of the diagonals of the inverse of the observed information matrix. Detailed calculations are given in Appendix B.

3 Data Analysis

3.1 Breast Cancer Data

3.1.1 The Data

Boag (1949) presents data for 121 patients undergoing treatment for breast cancer. The patients were originally divided into four categories:

1. Patients who died of causes other than breast cancer during follow up ($D = 1$);
2. Patients who died of breast cancer during follow up ($D = 2$);
3. Patients who were alive and free of breast cancer at the time of follow up ($D = 0$);
4. Patients who were alive at the time of follow up, but were suffering from persistence of breast cancer ($D = 0$).

Ng and McLachlan (1998) classified the data into three groups by combining the last two categories into one: patients who were still alive at the end of the follow up period, which corresponds to the censored observations. The data are presented in Table 3, Appendix C. The numbers in each category were 18 ($D = 1$), 78 ($D = 2$), and 25 ($D = 0$), respectively, and times to event are recorded in months.

Amongst these data, one very early (0.3 month) death was observed, both for death from the competing causes and for death from breast cancer. For a medical trial concerning breast cancer, such an early death (within 10 days) would be very rare. Because the IG distribution is very sensitive to the earliest death time, as investigated by Balka (2005), the earliest death time from breast cancer is excluded in the following analysis.

3.1.2 The Analysis

The CR model is applied to the breast cancer data, with the specification of the Gompertz distribution for the death times from competing causes and the IG distribution for the death times from breast cancer, respectively. The parameter estimates are given in Table 1; for comparison, we include results from using a Newton-Raphson algorithm and, as expected, the estimates are almost identical. The estimate of the probability of dying from competing causes, $\hat{p}_1$, is 0.322; this is very close to that (0.329) obtained by Ng and McLachlan (1998), who specified the Gompertz distribution for deaths from competing causes and the exponential distribution for deaths from breast cancer, respectively.

Table 1: Parameter estimates, with standard errors in parentheses, for the breast cancer data with the earliest death time excluded. To facilitate comparison, we include results from the Newton-Raphson approach in addition to our EM algorithm results.

	$\hat{\alpha}$	$\hat{\kappa}$	$\hat{\delta}$	$\hat{\nu}$	$\hat{p}_1$
EM Algorithm	-0.00079 (0.0029)	0.00486 (0.0011)	0.0188 (0.0039)	0.0285 (0.0046)	0.3025 (0.046)
Newton-Raphson Algo.	-0.00079 (0.0030)	0.00486 (0.0012)	0.0188 (0.0039)	0.0285 (0.0046)	0.3024 (0.046)

Figure 1 plots the conditional survival functions for death times from competing causes, using the CR model proposed in this paper and that proposed by Ng and McLachlan (1998); the conditional survival functions for death times from breast cancer, the cause of interest, using these same two models, are also plotted. Both plots show that the CR model with the Gompertz distribution and the IG distribution proposed in this paper result in similar survival curves for deaths from competing causes and deaths from the cause of interest, respectively, compared to that proposed by Ng and McLachlan (1998). The survival curve for death times from breast cancer using the IG distribution may better represent the data than using the exponential distribution; this may be due to the early occurrence of deaths from breast cancer that dominate among all the death times.

As Kaplan-Meier plots cannot be directly produced in the presence of competing risks, cumulative incidence for both competing events and the event of interest are plotted against the corresponding failure probabilities by using the proposed distributions for CR model (Figure 2). Clearly, both models fit the data very well.

Figure 1: The conditional survival functions for deaths from competing causes (left) and from breast cancer (right), using the CR model with the specification of the Gompertz and the IG distribution, and the specification of the Gompertz and exponential distribution.

Figure 2: Cumulative incidence versus failure probability using the Gompertz distribution (left) and the IG distribution (right) for the competing events of breast cancer data.

3.1.3 A Note on Data Treatment

In terms of data treatment, the merging of groups 3 and 4 might raise concern. Group 4 (infected but alive) contains only one observation, which is the largest censored time at 228 months. The parameter estimates with this one observation dropped were computed and were effectively identical to the estimates in Table 1.

3.2 Melanoma Data

The melanoma data from Andersen et al. (1993) are available as part of the boot package (Canty and Ripley, 2008, Davison and Hinkley, 1997) for the R software (R Development Core Team, 2010). The data consist of seven measurements made on 205 patients with malignant melanoma removed by surgery between 1962 and 1977. Event times are recorded in days for 71 observed death times; 57 were related to malignant melanoma and 14 were not. One measurement available for every patient is the thickness of the tumour. Thick tumours are considered to increase the risk of death from melanoma; thus, the analysis presented here will focus on tumours that are at least 1.94 mm thick (the median tumour thickness). With this condition, the data set is reduced to 106 patients, with 46 melanoma deaths and 7 deaths from other causes observed. The parameter estimates for these data are presented in Table 2. Here, the probability of dying from competing causes is estimated at 0.107. Figure 3 plots the conditional survival function, for tumour thickness ≥ 1.94 mm, for death times from competing causes using the Gompertz distribution, and from melanoma using the IG distribution.

Table 2: Parameter estimates, with standard errors in parentheses, for patients with tumour thickness ≥ 1.94 mm.

$\hat{\alpha}$	$\hat{\kappa}$	$\hat{\delta}$	$\hat{\nu}$	$\hat{p}_1$
-0.00051	0.00067	5.27×10^{-10}	0.00077	0.107
(0.0004)	(0.00037)	(8.28×10^{-5})	(0.000113)	(0.039)

Cumulative incidence for both competing events and the event of interest were plotted against the corresponding failure probabilities by using the proposed distributions for CR model (Figure 4). Clearly, both models again fit the data very well.

4 Conclusions

In this paper, a CR model is introduced where the IG distribution is assumed for the cause of interest and the Gompertz distribution is assumed for the competing causes. Parameter estimates are obtained using the EM algorithm and standard errors are computed using Louis' method. To demonstrate the use of the proposed model, this approach is applied to breast cancer and malignant melanoma data. Similar estimates are observed when this model is compared to a CR model with the Gompertz distribution for the failure times from the competing risks and the

Figure 3: Plot of the conditional survival function for deaths from competing causes (left) and from melanoma (right) for tumour thickness ≥ 1.94 mm.

Figure 4: Cumulative incidence versus failure probability using the Gompertz distribution (left) and the IG distribution (right) for the competing events of melanoma data.

exponential distribution for the failure times from the cause of interest. Thus, the proposed method provides a technique that may be useful, not only in situations where currently available models are valid, but also as an important alternative that may be more broadly applicable given the properties of the IG distribution.

In the melanoma example, the proposed model was used with a specific data subset based on tumour thickness. In future work, the proposed model may be generalized, where regression modelling can be considered for parameters of the IG and Gompertz distributions. This would allow for a comparison of data subgroups based on categorical covariates or consideration of the form of the relationship, when continuous covariates are available.

Balka, Desmond, and McNicholas (2011) consider cure models in the presence of covariates without regard to the competing risk aspect. Extensions of this

approach, to allow for competing risks, are currently under investigation. In addition, the more complex suite of first hitting time models treated in Balka, Desmond, and McNicholas (2009) are currently being developed to account for competing risks.

A Details on the EM Algorithm

Solving $\partial Q(\Psi | \Psi^{(k)}) / \partial \Psi = 0$ for Ψ gives

$$\begin{aligned} p_1^{(k+1)} &= \frac{\sum_{i:D_i=1} + \sum_{i:D_i=0} (1 - \mathbb{E}^{(k)}[z_i])}{n}, \\ \delta^{(k+1)} &= \frac{\sum_{i:D_i=2} + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i]}{\sum_{i:D_i=2} t_i + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i] \mathbb{E}^{(k)}[t_{2i}]}, \\ \mathbf{v}^{(k+1)} &= \frac{\sum_{i:D_i=2} (1/t_i) + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i] \mathbb{E}^{(k)}[1/t_{2i}] - \delta^{(k+1)} (\sum_{i:D_i=2} + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i])}{\sum_{i:D_i=2} + \sum_{i:D_i=0} \mathbb{E}^{(k)}[z_i]}, \end{aligned}$$

where $D_i = 0$ is the set of right-censored times, $D_i = 1$ is the set of observed failure times from the competing risks, and $D_i = 2$ is the set of observed failure times from the cause of interest. For the right-censored times t_i with $D_i = 0$,

$$\mathbb{E}[z_i] = \frac{(1 - p_1)S_2(t_i | \delta, \mathbf{v})}{p_1 S_1(t_i | \alpha, \kappa) + (1 - p_1)S_2(t_i | \delta, \mathbf{v})}.$$

The formulae for $\mathbb{E}[t_{2i}]$ and $\mathbb{E}[1/t_{2i}]$ for right-censored times are given by Whitmore (1983).

In this M-step, however, there is no closed-form solution for α and κ . A Newton-Raphson method is thus used. Let $\alpha^{(k,m)}$ and $\kappa^{(k,m)}$ be the m th sub-step to get the parameter estimates $\alpha^{(k+1)}$ and $\kappa^{(k+1)}$ within this M-step. We have

$$\begin{aligned} \alpha^{(k,m+1)} &= \alpha^{(k,m)} + \frac{T_{A1} + \Delta_1 / \alpha^{(k,m)} - \Delta_1 B'(\alpha^{(k,m)}) / B(\alpha^{(k,m)})}{\Delta_1 / (\alpha^{(k,m)})^2 + (\Delta_1 B''(\alpha^{(k,m)}) B(\alpha^{(k,m)}) - \Delta_1 (B'(\alpha^{(k,m)}))^2) / B^2(\alpha^{(k,m)})}, \\ \kappa^{(k,m+1)} &= (\alpha^{(k,m+1)} \Delta_1) / B(\alpha^{(k,m+1)}), \end{aligned}$$

where $T_{A1} = \sum_{i:D_i=1} t_i$, $\Delta_1 = \sum_{i:D_i=1}$, and

$$B(\alpha^{(k,m)}) = \sum_{i:D_i=1} (e^{\alpha^{(k,m)} t_i} - 1) + \sum_{i:D_i=0} (1 - \mathbb{E}[z_i]) (e^{\alpha^{(k,m)} t_i} - 1).$$

B Louis' Method Applied to the CR Model

The information functions, based on the complete-data log-likelihood, with respect to the parameters $\Psi = (p_1, \alpha, \kappa, \delta, \nu)$ for the CR model are:

$$\begin{aligned}
 -\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial p_1^2} &= \sum_{i:D_i=1} \frac{1}{p_1^2} + \sum_{i:D_i=2} \frac{1}{(1-p_1)^2} + \sum_{i:D_i=0} \frac{1-z_i}{p_1^2} + \sum_{i:D_i=0} \frac{z_i}{(1-p_1)^2} \\
 &= \frac{\sum_{i=1}^n z_i}{(1-p_1)^2} + \frac{n - \sum_{i=1}^n z_i}{p_1^2}, \\
 -\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \alpha^2} &= \frac{2\kappa}{\alpha^3} \left[\sum_{i:D_i=1} (e^{\alpha t_i} - 1) + \sum_{i:D_i=0} (1-z_i)(e^{\alpha t_i} - 1) \right] \\
 &\quad - \frac{2\kappa}{\alpha^2} \left[\sum_{i:D_i=1} t_i e^{\alpha t_i} + \sum_{i:D_i=0} (1-z_i) t_i e^{\alpha t_i} \right] + \frac{\kappa}{\alpha} \left[\sum_{i:D_i=1} t_i^2 e^{\alpha t_i} + \sum_{i:D_i=0} (1-z_i) t_i^2 e^{\alpha t_i} \right], \\
 -\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \kappa^2} &= \sum_{i:D_i=1} \frac{1}{\kappa^2}, \\
 -\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \delta^2} &= \sum_{i:D_i=2} t_i + \sum_{i:D_i=0} z_i t_{2i}, \\
 -\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \nu^2} &= - \sum_{i:D_i=2} \left[\frac{1}{2\nu^2} - \frac{(1-\delta t_i)^2}{\nu^3 t_i} \right] - \sum_{i:D_i=0} z_i \left[\frac{1}{2\nu^2} - \frac{1/t_{2i} - 2\delta + \delta^2 t_{2i}}{\nu^3} \right].
 \end{aligned}$$

Using Louis' method, the observed information functions for $\Psi = (p_1, \alpha, \kappa, \delta, \nu)$ are

$$\begin{aligned}
 -\frac{\partial^2 \log L(\Psi | \mathbf{t})}{\partial p_1^2} &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[-\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial p_1^2} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\frac{\partial \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial p_1} \right] \\
 &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[\frac{\sum_{i=1}^n z_i}{(1-p_1)^2} + \frac{n - \sum_{i=1}^n z_i}{p_1^2} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[-\frac{\sum_{i=1}^n z_i}{1-p_1} + \frac{n - \sum_{i=1}^n z_i}{p_1} \right] \\
 &= \frac{n}{(1-p_1)p_1} - \frac{1}{(1-p_1)^2 p_1^2} \sum_{i:D_i=0} (1 - \mathbb{E}[z_i]) \mathbb{E}[z_i], \\
 -\frac{\partial^2 \log L(\Psi | \mathbf{t})}{\partial \alpha^2} &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[-\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \alpha^2} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\frac{\partial \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \alpha} \right] \\
 &= \frac{2\kappa}{\alpha^3} B(\alpha) - \frac{2\kappa}{\alpha^2} B'(\alpha) + \frac{\kappa}{\alpha} B''(\alpha) - \sum_{i:D_i=0} \left[\frac{\kappa}{\alpha^2} (e^{\alpha t_i} - 1) - \frac{\kappa}{\alpha} t_i e^{\alpha t_i} \right]^2 (1 - \mathbb{E}[z_i]) \mathbb{E}[z_i],
 \end{aligned}$$

where $B(\alpha) = \sum_{i:D_i=1} (e^{\alpha t_i} - 1) + \sum_{i:D_i=0} (1 - \mathbb{E}[z_i]) (e^{\alpha t_i} - 1)$, $B'(\alpha)$ and $B''(\alpha)$ are the first and second derivatives of $B(\alpha)$.

$$\begin{aligned}
 -\frac{\partial^2 \log L(\Psi | \mathbf{t})}{\partial \kappa^2} &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[-\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \kappa^2} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\frac{\partial \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \kappa} \right] \\
 &= \sum_{i:D_i=1} \frac{1}{\kappa^2} - \sum_{i:D_i=0} \left(\frac{e^{\alpha t_i} - 1}{\alpha} \right)^2 (1 - \mathbb{E}[z_i]) \mathbb{E}[z_i], \\
 -\frac{\partial^2 \log L(\Psi | \mathbf{t})}{\partial \delta^2} &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[-\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \delta^2} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\frac{\partial \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial \delta} \right] \\
 &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[\sum_{i:D_i=2} t_i + \sum_{i:D_i=0} z_i t_{2i} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\frac{1}{v} \sum_{i:D_i=2} (1 - \delta t_i) + \frac{1}{v} \sum_{i:D_i=0} z_i (1 - \delta t_{2i}) \right] \\
 &= \frac{1}{v} \left[\sum_{i:D_i=2} t_i + \sum_{i:D_i=0} \mathbb{E}[z_i] \mathbb{E}[t_{2i}] \right] + \frac{1}{v^2} \sum_{i:D_i=0} (\mathbb{E}[z_i])^2 (1 - \delta \mathbb{E}[t_{2i}])^2 \\
 &\quad - \frac{1}{v^2} \sum_{i:D_i=0} (\mathbb{E}[z_i]) (1 - 2\delta \mathbb{E}[t_{2i}] + \delta^2 \mathbb{E}[t_{2i}^2]), \\
 -\frac{\partial^2 \log L(\Psi | \mathbf{t})}{\partial v^2} &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[-\frac{\partial^2 \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial v^2} \right] - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\frac{\partial \log L_c(\Psi | \mathbf{z}, \mathbf{t}_2)}{\partial v} \right] \\
 &= \mathbb{E}_{\mathbf{z}, \mathbf{t}_2} \left[-\sum_{i:D_i=2} \left(\frac{1}{2v^2} - \frac{(1 - \delta t_i)^2}{v^3 t_i} \right) - \sum_{i:D_i=0} z_i \left(\frac{1}{2v^2} - \frac{1/t_{2i} - 2\delta + \delta^2 t_{2i}}{v^3} \right) \right] \\
 &\quad - \mathbb{V}_{\mathbf{ar}_{\mathbf{z}, \mathbf{t}_2}} \left[\sum_{i:D_i=2} \left(-\frac{1}{2v} + \frac{(1 - \delta t_i)^2}{2v^2 t_i} \right) + \sum_{i:D_i=0} z_i \left(-\frac{1}{2v} + \frac{(1 - \delta t_{2i})^2}{2v^2 t_{2i}} \right) \right] \\
 &= -\frac{(v + 4\delta)(1 - p_1)n}{2v^3} + \frac{1}{v^3} \sum_{i:D_i=2} \left(\frac{1}{t_i} + \delta^2 t_i \right) + \frac{1}{v^3} \sum_{i:D_i=0} \mathbb{E}[z_i] (\mathbb{E}[1/t_{2i}] + \delta^2 \mathbb{E}[t_{2i}]) \\
 &\quad - \sum_{i:D_i=0} \mathbb{E}[z_i] \left\{ \frac{1}{4v^4} \mathbb{E}[1/t_{2i}^2] - \left(\frac{\delta}{v^4} + \frac{1}{2v^3} \right) \mathbb{E}[1/t_{2i}] \right. \\
 &\quad \left. + \left(\frac{1}{4v^2} + \frac{6\delta^2}{4v^4} + \frac{\delta}{v^3} \right) - \left(\frac{\delta^3}{v^4} + \frac{\delta^2}{2v^3} \right) \mathbb{E}[t_{2i}] + \frac{\delta^4}{4v^4} \mathbb{E}[t_{2i}^2] \right\} \\
 &\quad + \sum_{i:D_i=0} (\mathbb{E}[z_i])^2 \left(-\frac{v + 2\delta}{2v^2} + \frac{1}{2v^2} \mathbb{E}[1/t_{2i}] + \frac{\delta^2}{2v^2} \mathbb{E}[t_{2i}] \right)^2.
 \end{aligned}$$

The standard errors are then obtained by taking the square roots of the inverse of the observed information with respect to the corresponding parameters.

C Breast Cancer Data

Table 3: The breast cancer data of Boag (1949), with time in months.

Survival Times								Censored Times ($D=0$)	
Died of the Cancer ($D=2$)						Died of Other Causes ($D=1$)			
0.3	12.2	17.5	28.2	41	78	0.3	110	111	136
5.0	12.3	17.9	29.1	42	80	4.0	111	112	141
5.6	13.5	19.8	30.0	44	84	7.4	112	113	143
6.2	14.4	20.4	31	46	87	15.5	132	114	167
6.3	14.4	20.9	31	48	89	23.4	162	114	177
6.6	14.8	21.0	32	48	90	46		117	179
6.8	15.7	21.0	35	51	97	46		121	189
7.5	16.2	21.1	35	51	98	51		123	201
8.4	16.3	23.0	38	52	100	65		129	203
8.4	16.5	23.6	39	54	114	68		131	203
10.3	16.8	24.0	40	56	126	83		133	213
11.0	17.2	24.0	40	60	131	88		134	228
11.8	17.3	27.9	41	78	174	96		134	

References

- Aalen, O. O., Ø. Borgan, and H. K. Gjessing (2008): *Survival and Event History Analysis: A Process Point of View*, Statistics for Biology and Health, New York: Springer.
- Aalen, O. O. and H. K. Gjessing (2001): “Understanding the shape of the hazard rate: a process point of view,” *Statistical Science*, 16, 1–22.
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993): *Statistical Models Based on Counting Processes*, New York: Springer-Verlag New York, Inc.
- Balka, J. (2005): *Wiener Process Cure Rate Models*, Ph.D. thesis, University of Guelph, Guelph, Ontario, Canada.
- Balka, J., A. F. Desmond, and P. D. McNicholas (2009): “Review and implementation of cure models based on first hitting times for Wiener processes,” *Lifetime Data Analysis*, 15, 147–176.
- Balka, J., A. F. Desmond, and P. D. McNicholas (2011): “Bayesian and likelihood inference for cure rates based on defective inverse Gaussian regression models,” *Journal of Applied Statistics*, 38, 127–144.

- Boag, J. W. (1949): "Maximum likelihood estimates of the proportion of patients cured by cancer therapy," *Journal of the Royal Statistical Society. Series B*, 11, 15–53.
- Canty, A. and B. Ripley (2008): *boot: Bootstrap R (S-Plus) Functions.*, r package version 1.2-34.
- 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.
- Davison, A. C. and D. V. Hinkley (1997): *Bootstrap Methods and Their Applications*, Cambridge: Cambridge University Press.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, 39, 1–38.
- Farewell, V. T. (1982): "The use of mixture models for the analysis of survival data with long-term survivors," *Biometrics*, 38, 1041–1046.
- Farewell, V. T. (1986): "Mixture models in survival analysis: Are they worth the risk?" *The Canadian Journal of Statistics*, 14, 257–262.
- Gordon, N. H. (1990): "Application of the theory of finite mixtures for the estimation of 'cure' rates of treated cancer patients," *Statistics in Medicine*, 9, 397–407.
- Kalbfleisch, J. and R. Prentice (2002): *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons, 2nd edition.
- Larson, M. G. and G. E. Dinse (1985): "A mixture model for the regression analysis of competing risks data," *Journal of the Royal Statistical Society. Series C*, 34, 201–211.
- Lee, M.-L. T. and G. A. Whitmore (2006): "Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary," *Statistical Science*, 21, 501–513.
- Louis, T. A. (1982): "Finding observed information using the EM algorithm," *Journal of the Royal Statistical Society. Series B*, 44, 98–130.
- McLachlan, G. J. and D. Peel (2000): *Finite Mixture Models*, New York: John Wiley & Sons.
- Ng, S. K. and G. J. McLachlan (1998): "On modifications to the long-term survival mixture model in the presence of competing risks," *Journal of Statistical Computation and Simulation*, 61, 77–96.
- R Development Core Team (2010): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

- Tsiatis, A. (1975): “A nonidentifiability aspect of the problem of competing risks,” *Proceedings of the National Academy of Sciences*, 72, 20–22.
- Whitmore, G. A. (1983): “A regression method for censored inverse-Gaussian data,” *The Canadian Journal of Statistics*, 11, 305–315.