isid/ms/2003/21
July 29, 2003
http://www.isid.ac.in/~statmath/eprints

Parametric models for subsurvival functions

Isha Dewan
S. B. Kulathinal

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Parametric models for subsurvival functions

Isha Dewan, Indian Statistical Institute, India
AND
S.B.Kulathinal, National Public Health Institute, Finland
Date : June 04, 2003

Abstract
The competing risks data consist of time to failure and cause of failure. Nonparametric techniques for estimating and testing cause-specific hazards or subsurvival functions have been developed. However, very little work has been done in specifying parameteric models for the cause-specific hazards or for the subsurvival functions. A difficulty in specifying a parametric subsurvival distribution is its very nature - it is not a proper survival function. This difficulty can be overcome by specifying a parametric model for the cause-specific hazards which could assume any suitable form of the hazard of the standard survival distributions. This leads to parametric forms for the subsurvival functions or subdensity functions. The parameters can be estimated using standard techniques. Three sets of published data are reanalysed here and parametric models are fitted.

1 Introduction

The competing risks data consist of the failure time (T) and the cause of failure (δ). The competing risks data are generally modelled in two ways: (i) using latent lifetimes and (ii) using the joint distribution of (T, δ). In (i), the latent lifetimes are generally assumed to be independent, which is an untestable assumption. However, statistical analysis for independent competing risks data under various parametric models has been considered (David and Moeschberger, 1978). The techniques used for analysing right censored data can be implemented if the approach (i) is used and standard parametric, semiparametric or nonparametric models can be used without any complications. Even when latent failures are not independent multivariate parametric forms can be assumed and statistical analysis can be done (Moeschberger, 1974). In (ii) subsurvival functions or cause specific hazard rates are used to model the competing risks data (Kalbfleisch and Prentice, 1980 and Deshpande, 1990). Nonparametric techniques for estimating and testing cause-specific hazards or subsurvival functions have been developed. There has been some discussion about the use of Cox's proportional hazards model, which is a semiparametric model, for the cause-specific hazards (Crowder, 2001). We refer to Crowder (2001) and Kalbfleisch and Prentice (2002) for the detailed discussion on various issues related to competing risks. However, there has been very little done in specifying parametric models for the cause-specific hazards or for the subsurvival functions. A difficulty in specifying a parametric subsurvival distribution is because it is an improper survival function. One way to overcome this difficulty is to specify a parametric model for the cause-specific hazards which could in practice take any form of the hazard of the standard survival distributions. Unfortunately, this approach has not been widely used in the competing risks literature and hence there are no well documented and standard parametric distributions for subsurvival functions. Standard methods can be employed for fitting parametric subsurvival functions. Analytical methods like maximum likelihood estimation can be used to estimate the unknown parameters. A plot of time versus cause-specific hazard can be used to have initial guess of a parametric model for the cause-specific hazards. More sophisticated goodness of fit tests can be used to test whether the model fit the data well. In section 2, we suggest some parametric models for subsurvival/subdensity functions. In section 3, we find maximum likelihood estimators of the parameters of the distributions. In section 4, we reanalyse the data given in Hoel (1972) and Nair (1993) and fit parameteric subdistributions. We conclude with a discussion on the use of parametric models.

2 Parametric subsurvival functions

In this section, we propose some relationships between the cause-specific hazards and time and derive corresponding parametric models for subsurvival/subdensity functions. We assume that k competing risks are acting simultaneously and T denotes the failure time and δ, which can take one of the values in $\{1, 2, ..., k\}$, denotes the cause of failure. The joint distribution of (T, δ) are defined here by the subsurvival functions

$$S_i(t) = pr(T > t, \delta = i), \quad i = 1, 2, \ldots, k$$

and equivalently by the incidence functions or subdistribution functions

$$F_i(t) = pr(T \leq t, \delta = i) = p_i - S_i(t) \quad i = 1, 2, \ldots, k,$$

where $p_i = pr(\delta = i)$. The survival function of T is given by

$$S(t) = pr(T > t) = \sum_{i=1}^{k} S_i(t)$$

and the distribution function is given by

$$F(t) = pr(T \leq t) = \sum_{i=1}^{k} F_i(t).$$

Let $f_i(t), i = 1, 2, ..., k$ denote the subdensity functions and $f(t) = \sum_{i=1}^{k} f_i(t)$ denote the density function of T whenever they exist. The cause-specific hazard rate for cause i is defined as

$$h_i(t) = f_i(t)/S(t-), \quad i = 1, \ldots, k, \tag{1}$$

and the cumulative cause specific hazard function is

$$H_i(t) = \int_0^t h_i(u)du$$

The crude hazard rate for cause i is defined as

$$r_i(t) = f_i(t)/S_i(t-) \ \ i = 1, \ldots, k, .$$

The hazard rate of T is

$$h(t) = f(t)/S(t-) = \sum_{i=1}^{k} h_i(t).$$

It is easy to see that the survival function of T can be expressed as

$$S(t) = \exp(-\int_0^t h(u)du) = \prod_{i=1}^{k} \exp(-\int_0^t h_i(u)du), \tag{2}$$

and

$$f_i(t) = h_i(t)S(t), \ \ i = 1, \ldots, k. \tag{3}$$

The subsurvival function can then be obtained as

$$S_i(t) = \int_t^\infty f_i(u)du.$$

In most cases it is not possible to write the subsurvival function in a closed form, but we do have nice expressions for subdensity functions. The conditional probability function

$$\Phi_i(t) = pr(\delta = i \mid T \geq t), \ \ = 1, 2, \ldots, k$$

has been studied by Dewan , Deshpande and Kulathinal (2004) in case of two competing risks. The shape of this function helps in studying the dependence between T and δ. Deshpande (1990) proposed two semi-parametric models for subsurvival functions when $k = 2$. The first model corresponds to the independence of T and δ and under this model

$$F_1(t) = pF(t) \ \text{ and } \ F_2(t) = (1-p)F(t).$$

In this case, the conditional probability function $\Phi_1(t) = p$. The second model is defined as

$$F_1(t) = \frac{(F(t))^\theta}{2} \ \text{ and } \ F_2(t) = F(t) - \frac{(F(t))^\theta}{2}, 1 \leq \theta \leq 2. \tag{4}$$

For this model $F_1(\infty) = F_2(\infty) = 1/2$. Both subdistribution functions are increasing to $1/2$ as $t \to \infty$ and $F_1(t) > F_2(t)$ for $0 < t < \infty$ when $1 < \theta \leq 2$. When $\theta = 1$, T and δ are independent and $F_1(t) = F_2(t)$. A more general model was proposed recently by Dewan *et al.* (2004) where $F_1(t) = pF^\theta(t)$, $F_2(t) = F(t) - pF^\theta(t)$, where $1 \leq \theta \leq 2$, and $0 \leq p \leq 0.5$. Here, $p = pr(\delta = 1)$ and

$$\Phi_1(t) = \frac{p(1 - F^\theta(t))}{1 - F(t)}$$

which is an increasing function of t. We will refer to this model as Weibull subdistribution family when $F(t)$ corresponds to the Weibull distribution. In this section, we model the cause-specific hazards using the standard and commonly used parametric distributions in the survival analysis and derive corresponding subdensity functions or subsurvival functions wherever possible. Commonly used parametric distributions are constant hazard, linear hazard and power hazard. We introduce a general model and all the three models mentioned can be derived as special cases of it.

2.1 General power model

When the cause-specific hazard rate due to cause i is of the form

$$h_i(t) = \lambda_i + \gamma_i \alpha_i t^{\alpha_i - 1} \tag{5}$$

where $\lambda_i,\ \gamma_i,\ \alpha_i \geq 0$, it is said to belong to the family of general power model. When all the k cause-specific hazard rates belong to the same family then

$$\begin{aligned}
h(t) &= \sum_{i=1}^{k} (\lambda_i + \gamma_i \alpha_i t^{\alpha_i - 1}) \\
S(t) &= \exp(-\sum_{i=1}^{k} \lambda_i t - \sum_{i=1}^{k} \gamma_i t^{\alpha_i}) \\
f_i(t) &= (\lambda_i + \gamma_i \alpha_i t^{\alpha_i - 1}) \exp(-\sum_{i=1}^{k} \lambda_i t - \sum_{i=1}^{k} \gamma_i t^{\alpha_i}).
\end{aligned}$$

The subsurvival function can be obtained by integrating the subdensity function given above but it is not in a nice closed form. We consider special cases of the general power model below.

2.1.1 Constant hazard rate: $\gamma_i = 0$

When $\gamma_i = 0$ in (5),

$$h_i(t) = \lambda_i. \tag{6}$$

Under this model, $h(t) = \sum_{i=1}^{k} \lambda_i = \lambda$ and $S(t) = \exp(-\lambda t)$. Also

$$\begin{aligned}
S_i(t) &= \lambda_i / \lambda \exp(-\lambda t) \\
f_i(t) &= \lambda_i \exp(-\lambda t).
\end{aligned}$$

Note that $p_i = \lambda_i/\lambda$. When $k = 2$, $\Phi_1(t) = \lambda_1/(\lambda_1 + \lambda_2)$. This is the simplest special case referred to as a constant hazard rate. In this case, the time to failure T and the cause of failure δ are independent and the study of competing risks problem is simplified to a great extent. For example, when $i = 2$ and T and δ are independent, testing for equality of subsurvival functions $S_1(t) = S_2(t)$ reduces to testing $pr(\delta = 1) = pr(\delta = 2)$. In reliability studies, constant hazard rate characterises the exponential distribution. It reflects the lack of memory property or the no-ageing aspect of the distribution. Further, if a random variable T has any arbitrary continuous distribution then the cumulative hazard function $H(t)$ has exponential distribution with parameter 1. A more general model is a piecewise constant hazard model and it can be specified as

$$h_i(t) = \lambda_{il},\ \ t'_{l-1} < t \leq t'_l, \tag{7}$$

where $l = 0, 1, ..., m$ and $t'_0 = 0$. The change points $(t'_1 < ... < t'_m)$ and m can be assume to be known or unknown.

2.1.2 Linear hazard rate: $\alpha_i = 2$

A more general model than the constant hazard rate is given by linear hazard rate obtained by setting $\alpha_i = 2$ in (5) and is given by

$$h_i(t) = \lambda_i + \gamma_i' t, \tag{8}$$

where $\gamma_i' = 2\gamma_i$. Then $h(t) = \sum_{i=1}^k (\lambda_i + \gamma_i' t) = \lambda + \gamma t$ and $S(t) = \exp(-\lambda t - \gamma t^2/2)$, where $\lambda = \sum_{i=1}^k \lambda_i$ and $\gamma = \sum_{i=1}^k \gamma_i$. This gives

$$f_i(t) = (\lambda_i + \gamma_i' t)\exp(-\lambda t - \gamma t^2/2), i = 1, 2, ..., k.$$

Then

$$S_i(t) = \sqrt{2\pi\gamma}\exp\frac{\lambda^2}{2\gamma}\int_{t+a/\gamma}^{\infty}[\lambda_i - \frac{\gamma_i\lambda}{\gamma} + \gamma_i z]\phi^*(z)dz$$

where $\phi^*(z)$ is the density function of a normal variable with mean zero and variance λ. In particular for $k = 2$ $\frac{f_1(t)}{f_2(t)} = \frac{\gamma_1+\lambda_1}{\gamma_2+\lambda_2}$. Then it is easy to see that $\frac{f_1(t)}{f_2(t)}$ is increasing in t iff $\lambda_1/\lambda_2 < \gamma_1/\gamma_2$, that is, the ratio of slopes is greater than the ratio of the intercepts. This implies that $S_1(t)/S_2(t)$ is increasing in t.

2.1.3 Weibull hazard rate: $\lambda_i = 0$

When $\lambda_i = 0$ in (5), the cause-specific hazards have the form

$$h_i(t) = \gamma_i \alpha_i t^{\alpha_i - 1}. \tag{9}$$

Note that this is the Weibull hazard rate with scale paramter γ_i and shape parameter α_i. Under this model,

$$\begin{aligned} h(t) &= \sum_{i=1}^k [\gamma_i \alpha_i t^{\alpha_i - 1}], \\ S(t) &= \exp(-\sum_{i=1}^k \gamma_i t^{\alpha_i}), \\ f_i(t) &= (\gamma_i \alpha_i t^{\alpha_i - 1})\exp(-\sum_{i=1}^k \gamma_i t^{\alpha_i}). \end{aligned}$$

It is not possible to write the subsurvival function in a closed form. When $k = 2$, we have $\frac{f_1(t)}{f_2(t)} = \frac{\alpha_1\gamma_1}{\alpha_2\gamma_2} t^{\alpha_1 - \alpha_2}$. This is constant in t iff $\alpha_1 = \alpha_2$ and increasing in t iff $\alpha_1 > \alpha_2$. This implies T and δ are independent for $\alpha_1 = \alpha_2$ and the conditional probability function $\Phi_1(t)$ is an increasing function of t for $\alpha_1 > \alpha_2$. Hence the value of the shape parameter determines the nature of dependence between T and δ. The Weibull hazard rate is increasing for $\alpha_i > 1$ and constant for $\alpha_i = 1$. The Weibull distribution arises theoretically as the limiting distribution of the minimum of a large number of independent nonnegative random variables. It is very often used in reliability because of the simplicity of the density, survivor and hazard functions.

2.2 Pareto distribution

Suppose cause-specific hazard is of the form

$$
\begin{aligned}
h_i(t) &= \frac{\theta_i}{\alpha_i + t}, \quad i = 1, \ldots, k, \\
h(t) &= \sum_{i=1}^{k} \frac{\theta_i}{\alpha_i + t}, \\
S(t) &= \exp - \sum_{i=1}^{k} [\theta_i \log \frac{\alpha_i + t}{\alpha_i}], \\
f_i(t) &= \frac{\theta_i}{\alpha_i + t} \exp[- \sum_{i=1}^{k} \theta_i \log \frac{\alpha_i + t}{\alpha_i}], \quad i = 1, \ldots, k.
\end{aligned} \tag{10}
$$

Again , for $k = 2$,we have $\frac{f_1(t)}{f_2(t)} = \frac{\theta_1(\alpha_2+t)}{\theta_2(\alpha_1+t)}$. This is constant in t iff $\alpha_1 = \alpha_2$ and increasing in t iff $\alpha_1 > \alpha_2$. This implies T and δ are independent for $\alpha_1 = \alpha_2$ and the conditional probability function $\Phi_1(t)$ is an increasing function of t for $\alpha_1 > \alpha_2$. Note that the hazard function in this case is decreasing in t and tends to zero as $t \to \infty$. But the object will eventually fail as the survival at ∞ has probability zero.

3 Fitting parametric subsurvival functions

Let (T_j, δ_j) $j = 1, 2, ..., n$ be the competing risks data obtained from n independent and identical units. A naive estimator of the subsurvival function is the empirical estimator given by

$$S_{in}(t) = \frac{1}{n} \sum_{j=1}^{n} I(T_j > t, \delta_j = i), \quad i = 1, \ldots, k, \tag{11}$$

and Nelson-Aalen estimator (Crowder, 2001) of cumulative cause-specific hazard is given by

$$\hat{H}_i(t) = \sum_{\{j | T_{(j)} \leq t\}} d_{ij}/n_j^*, \quad i = 1, \ldots, k \tag{12}$$

where $T_{(j)}$ is the jth ordered failure time, n_j^* is the number of individuals at risk prior to $T_{(j)}$ and d_{ij} is the number of failures from ith cause at $T_{(j)}$. To fit a parametric model, we need to estimate the unknown parameters of the model and evaluate the model by substituting these parameters. We use maximum likelihood method to estimate the parameters for the models introduced in the earlier section. Using (1), (2) and (3), the likelihood function is given by

$$
\begin{aligned}
L &= \prod_{j=1}^{n} \prod_{i=1}^{k} [f_i(t_j)]^{I(\delta_j = i)} \\
&= \prod_{i=1}^{k} \prod_{j=1}^{n} [h_i(t_j)]^{I(\delta_j = i)} \prod_{j=1}^{n} S(t_j) \\
&= \prod_{i=1}^{k} \prod_{j=1}^{n} [h_i(t_j)]^{I(\delta_j = i)} exp(- \int_0^{t_j} h_i(u) du).
\end{aligned} \tag{13}
$$

In practical situations, all the cause specific hazard rates need not belong to the same family of the distributions. An appropriate distribution can be assumed for each cause-specific hazard and can easily be fitted since the likelihood factorises as a product of cause-specific hazard for each risk. For the purpose of illustration, we assume that all the cause-specific hazards belong to the same family of distributions.

3.1 General Power Model

When all the cause-specific hazards are assumed to belong to the family of general power model, the likelihood function is given by

$$L = \prod_{i=1}^{k}\prod_{j=1}^{n}[\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1}]^{I(\delta_j=i)}\exp(-\lambda_i t_j - \gamma_i t_j^{\alpha_i})$$

and the logarithm of the likelihood function is

$$LogL = \sum_{i=1}^{k}\sum_{j=1}^{n}[I(\delta_j=i)\log(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1}) - \lambda_i t_j - \gamma_i t_j^{\alpha_i}].$$

The likelihood equations are given by

$$\begin{aligned}
\sum_{j=1}^{n}[\frac{I(\delta_j=i)}{\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1}} - t_j] &= 0 \\
\sum_{j=1}^{n}[\frac{I(\delta_j=i)\alpha_i t_j^{\alpha_i-1}}{\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1}} - t_j^{\alpha_i}] &= 0 \\
\sum_{j=1}^{n}[\frac{I(\delta_j=i)\gamma_i(t_j^{\alpha_i-1} + \alpha_i t_j^{\alpha_i-1}\log t_j)}{\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1}} - \gamma_i t_j^{\alpha_i}\log t_j] &= 0.
\end{aligned}$$

The second derivatives are given by

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial\lambda_i^2} &= -\sum_{j=1}^{n}\frac{I(\delta_j=i)}{(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1})^2} \\
\frac{\partial^2 \log L}{\partial\lambda_i\partial\gamma_i} &= -\sum_{j=1}^{n}\frac{I(\delta_j=i)\alpha_i t_j^{\alpha_i-1}}{(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1})^2} \\
\frac{\partial^2 \log L}{\partial\lambda_i\partial\alpha_i} &= -\sum_{j=1}^{n}\frac{I(\delta_j=i)\gamma_i[t_j^{\alpha_i-1} + \alpha_i t_j^{\alpha_i-1}\log t_j]}{(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1})^2} \\
\frac{\partial^2 \log L}{\partial\gamma_i^2} &= -\sum_{j=1}^{n}\frac{I(\delta_j=i)(\alpha_i t_j^{\alpha_i-1})^2}{(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1})^2} \\
\frac{\partial^2 \log L}{\partial\alpha_i^2} &= -\sum_{j=1}^{n}I(\delta_j=i)[\frac{\gamma_i^2(t_j^{\alpha_i-1} + \alpha_i t_j^{\alpha_i-1}\log t_j)^2}{(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1})^2} - \frac{\gamma_i(2t_j^{\alpha_i-1}\log t_j + \alpha_i t_j^{\alpha_i-1}(\log t_j)^2)}{(\lambda_i + \gamma_i\alpha_i t_j^{\alpha_i-1})}] \\
&- \sum_{j=1}^{n}\gamma_i t_j^{\alpha_i}(\log t_j)^2
\end{aligned}$$

$$\frac{\partial^2 \log L}{\partial\alpha_i\partial\gamma_i} = -\sum_{j=1}^{n}[I(\delta_j=i)[-\frac{(t_j^{\alpha_i-1}+\alpha_i t_j^{\alpha_i-1}\log t_j)}{(\lambda_i+\gamma_i\alpha_i t_j^{\alpha_i-1})}+\frac{\gamma_i(t_j^{\alpha_i-1}+\alpha_i t_j^{\alpha_i-1}\log(t_j))\alpha_i t_j^{\alpha_i-1}}{(\lambda_i+\gamma_i\alpha_i t_j^{\alpha_i-1})^2}]$$
$$- \sum_{j=1}^{n} t_j^{\alpha_i}\log(t_j).$$

3.2 Constant hazard rate

The likelihood is given by

$$\begin{aligned} L &= \prod_{i=1}^{k}\prod_{j=1}^{n}\lambda_i^{I(\delta_j=i)} exp(-\lambda_i t_j) \\ &= \prod_{i=1}^{k}\lambda_i^{n_i} exp(-\lambda T), \end{aligned}$$

where $n_i = \sum_{j=1}^{n} I(\delta_j = i)$, $\lambda = \sum_{i=1}^{k}\lambda_i$ and $T = \sum_{j=1}^{n} t_j$. Hence, $\hat{\lambda}_i = n_i/T$. The information matrix is $n/\lambda \ \ diag(1/\lambda_1, 1/\lambda_2, ..., 1/\lambda_k)$. When $(t'_1, ..., t'_m)$ are known in case of the piecewise constant hazard, the maximum likelihood estimates are

$$\hat{\lambda_{il}} = \frac{\sum_{j=1}^{n} I(\delta_j=i)I(t_j\in(t'_{l-1},t'_l])}{\sum_{j=1}^{n}(t_j-t'_{l-1})I(t_j\in(t'_{l-1},t'_l])+(t'_l-t'_{l-1})\sum_{j=1}^{n}I(t_j>t'_l)} \tag{14}$$

3.3 Linear hazard rate

The likelihood is given by

$$L = \prod_{i=1}^{k}\prod_{j=1}^{n}[\lambda_i+\gamma'_i t_j]^{I(\delta_j=i)}\exp(-\lambda_i t_j - \gamma'_i t_j^2/2) \tag{15}$$

The likelihood equations are given by

$$\begin{aligned} \sum_{j=1}^{n}[\frac{I(\delta_j=i)}{\lambda_i+\gamma'_i t_j} - t_j] &= 0 \\ \sum_{j=1}^{n}[\frac{t_j I(\delta_j=i)}{\lambda_i+\gamma'_i t_j} - \frac{t_j^2}{2}] &= 0 \end{aligned}$$

We also have

$$\begin{aligned} \frac{\partial^2 LogL}{\partial\lambda_i^2} &= -\sum_{j=1}^{n}\frac{I(\delta_j=i)}{(\lambda_i+\gamma'_i t_j)^2} \\ \frac{\partial^2 LogL}{\partial\lambda_i^2} &= -\sum_{j=1}^{n}\frac{t_j^2 I(\delta_j=i)}{(\lambda_i+\gamma'_i t_j)^2} \\ \frac{\partial^2 LogL}{\partial\lambda_i\partial\lambda_i} &= -\sum_{j=1}^{n}\frac{t_j I(\delta_j=i)}{(\lambda_i+\gamma'_i t_j)^2} \end{aligned}$$

3.4 Weibull hazard rate

The likelihood is given by

$$L = \prod_{i=1}^{k} \prod_{j=1}^{n} [\gamma_i \alpha_i t_j^{\alpha_i - 1}]^{I(\delta_j = i)} \exp(-\gamma_i t_j^{\alpha_i}) \tag{16}$$

The likelihood equations are given by

$$\begin{aligned} \sum_{j=1}^{n} [I(\delta_j = i)[\frac{1}{\alpha_i} + \log t_j] - \gamma_i t_j^{\alpha_i} \log t_j] &= 0 \\ \sum_{j=1}^{n} [\frac{I(\delta_j = i)}{\gamma_i} - t_j^{\alpha_i}] &= 0 \end{aligned}$$

We also have

$$\begin{aligned} \frac{\partial^2 LogL}{\partial \gamma_i^2} &= -\sum_{j=1}^{n} \frac{I(\delta_j = i)}{\gamma_i^2} \\ \frac{\partial^2 LogL}{\partial \alpha_i^2} &= \sum_{j=1}^{n} [-\frac{I(\delta_j = i)}{\alpha_i^2} - \gamma_i t_j^{\alpha_i} (log t_j)^2] \\ \frac{\partial^2 LogL}{\partial \alpha_i \partial \gamma_i} &= -\sum_{j=1}^{n} t_j^{\alpha_i} log t_j \end{aligned}$$

3.5 Pareto distribution

The likelihood is given by

$$L = \prod_{i=1}^{k} \prod_{j=1}^{n} (\frac{\theta_i}{\alpha_i + t_j})^{I(\delta_j = i)} \exp -\theta_i [\log \frac{\alpha_i + t_j}{\alpha_i}] \tag{17}$$

The likelihood equations are

$$\begin{aligned} \sum_{j=1}^{n} [\frac{-I(\delta_j = i)}{\alpha_i + t_j} + \frac{\theta_i t_j}{\alpha_i(\alpha_i + t_j)}] &= 0 \\ \sum_{j=1}^{n} [I(\delta_j = i) \frac{1}{\theta_i} - \log \frac{\alpha_i + t_j}{\alpha_i}] &= 0. \end{aligned}$$

We also have

$$\begin{aligned} \frac{\partial^2 LogL}{\partial \alpha_i^2} &= \sum_{j=1}^{n} [I(\delta_j = i) \frac{1}{(\alpha_i + t_j)^2} + \theta_i [\frac{1}{(\alpha_i + t_j}^2 - \frac{1}{\alpha_i^2}]] \\ \frac{\partial^2 LogL}{\partial \theta_i^2} &= -\sum_{j=1}^{n} \frac{I(\delta_j = i)}{\theta_i^2} \\ \frac{\partial^2 LogL}{\partial \alpha_i \partial \theta_i} &= \sum_{j=1}^{n} \frac{t_j}{\alpha_i(\alpha_i + t_j)} \end{aligned}$$

3.6 Weibull subdistribution family

Consider a situation where units are exposed to two risks that is $k = 2$. Let the overall distribution function be $F(t) = 1 - \exp(-\lambda t^{\alpha})$, the Weibull distribution function in the expression (4). The likelihood equation for θ is

$$n_1/\theta + \sum_{j=1}^{n} I(\delta_j = 1)\log(z_j) - \sum_{j=1}^{n} I(\delta_j = 2) z_j^{\theta-1}(1 + \theta\log(z_j))/(2 - \theta z_j^{\theta-1}) = 0, \qquad (18)$$

where $z_j = z_j(\lambda, \alpha, t_j) = 1 - \exp(-\lambda t_j^{\alpha})$. The second derivative w.r.t θ is

$$\begin{aligned} \frac{\partial^2 LogL}{\partial\theta^2} &= -n_1/\theta^2 - \sum_{j=1}^{n} I(\delta_j = 2) z_j^{\theta-1}/(2 - \theta z_j^{\theta-1}) \\ & \quad [\log(z_j)(1 + \theta\log(z_j)) + \log(z_j) + (1 + \theta\log(z_j))^2 z_j^{\theta-1}/(2 - \theta z_j^{\theta-1})]. \end{aligned}$$

The M.L.E.'s of λ and α , the parameters of Weibull distribution can be found in the usual way.

4 Illustrations

4.1 Mortality data

We consider the mortality data given in Hoel (1972). The data were obtained from a laboratory experiment on two groups of RFM strain male mice which had received a radiation dose of 300 r at an age of 5-6 weeks. The first group of mice lived in a conventional laboratory environment, while the second group was in a germ-free environment. The causes of death were grouped into three classes - thymic lymphoma (risk=1), reticulum cell sarcoma (risk=2) and all other causes (risk=3). Following table summarises the size of the group and number of deaths due to various risks

Table 1: Group sizes and number of deaths due to various risks for Hoel's data (1972)

Environment	Size	Thymic lymphoma (risk=1)	Reticulum cell sarcoma (risk=2)	Other (risk=3)
Conventional	99	22	38	39
Germ-free	82	29	15	38

4.1.1 Conventional laboratory environment

Two risks

We first look at the autopsy data for 99 RFM mice in the conventional environment. We combine the deaths due to cancer, that is deaths due to risks 1 and 2 into one group and refer to it as risk 1 and treat the data as two risks problem. We will refer to risk 3 as risk 2 here. Figure 1a shows the Nelson-Aalen estimates of the cumulative cause-specific hazard rates for cancer ($H_1(t)$), other causes ($H_2(t)$) and overall hazard ($H(t)$). It also shows $\hat{p}_1 H(t)$ and $\hat{p}_2 H(t)$ where $\hat{p}_i$ is an empirical probability of dying due to risk i. It is clear that $H_i(t)$ is quite close to $\hat{p}_i H(t)$, $i = 1, 2$. Hence, T and δ may be treated independent and a parametric distribution can be fitted for the overall survival function. Note that similar conclusion was drawn in Dewan *et al.* (2004). A Weibull distribution was fitted to the overall survival function and it is shown in Figure 1b. The maximum likelihood estimate of the scale parameter is 1.32×10^{-7} and of the shape parameter is 2.54. A parametric distribution for the subsurvival functions is then

$$S_i(t) = p_i \exp(-\lambda t^{\alpha}),$$

where λ and α are specified using the maximum likelihood estimates. Application of Kolmogorov-Smirnov test showed nonsignificant difference between the empirical and Weibull distribution for the overall death. Also, Kolmogorv-Smirnov type tests were applied to test the fit of $F_i(t) = p_i F(t)$, where $F(t)$ was the Weibull distribution fitted for the overall death. Once again, the difference was not significant supporting the earlier observations that the death time and the cause of death are independent. **Three risks**

The same data were analysed using three risks as defined above. As it is clear from Figure 2a, an estimate of cause-specific hazard for risk 1 is a piecewise constant hazard and this model was fitted for risk 1. The fitted model for risk 1 is

$$\begin{aligned} h_1(t) &= 5.83 \times 10^{-5}, \quad t \leq 179 \\ &= 0.0016, \quad 179 < t \leq 282 \\ &= 0.0002, \quad 282 < t \leq 343 \\ &= 0.0009, \quad 343 < t \leq 441 \\ &= 0, \quad t > 441. \end{aligned}$$

The Weibull distribution fits well for reticulum cell sarcoma (risk=2) as can be seen from Figure 2b. It also fits well for other causes (risk=3), see Figure 2c. The cumulative cause-specific hazards for risk 2 and risk 3 are, respectively

$$\begin{aligned} H_2(t) &= 1.41 \times 10^{-23} t^{8.07} \\ H_3(t) &= 2.70 \times 10^{-6} t^{1.92}. \end{aligned}$$

The overall hazard is then the sum of a piece-wise constant hazard and two Weibull hazards. Figures 2d show the Nelson-Aalen estimate and the fitted parametric distribution for the overall hazard. The curves seem to be quite close to each other in all the cases. Application of goodness of fit tests showed no significant difference between the estimates of the cause-specific hazards and fitted models.

4.1.2 Germ-free environment

Next we look at the autopsy data for 82 RFM mice in the germ-free environment. Again we combine the deaths due to cancer into one group and consider it as a two risk (cancer and other) problem. Figure 3a-3c shows the Nelson-Aalen estimates of cumulative hazard functions. We fitted Weibull distribution to each cause-specific hazard. The maximum likelihood estimates of the scale and shape parameters of Weibull distribution are $(6.03 \times 10^{-6}, 1.77)$ and $(5.41 \times 10^{-15}, 4.87)$ for risks 1 and 2, respectively. The fitted distributions are shown in Figure 3a-3b alongwith their Nelson-Aalen estimates and the curves are close to each other for both the risks. The sum of these two Weibull distributions is a good fit for the overall hazard also (see Figure 3c). Goodness-of-fit tests showed no significant difference.

4.2 Failure data on switches

Consider the data on the times to failure, in millions of operations, and modes of failure of 37 switches, obtained from a reliability study conducted at AT&T, given in Nair (1993). There are two possible modes of failure, denoted by A (risk=1) and B (risk=2), for these switches. The number of failures due to risk 1 is 17 and that due to risk 2 are 20. Dewan *et al.* (2004) concluded that T and δ are not independent in this case. Assuming that $p_1 = p_2 = 1/2$, Weibull subdistribution function given in (4) was fitted. The maximum likelihood estimates of the scale and shape parameters of Weibull distribution for the overall hazard are $(0.028, 4.47)$ and θ is is taken as 1.265. Figure 4a show $-\log(H(t)$ against $\log(t)$ and the fitted Weibull model. Figure 4b shows the empirical estimates and the fitted subdistributions for the two risks. As it can be seen from the figures and also confirmed by the goodness-of-fit tests, it may be concluded that the two subdistribution functions belong to the family of weibull subdistribution functions.

5 Discussion

We have demonstrated the use of standard parametric survival models in case of competing risks by modeling cause-specific hazards or subsurvival functions. Any standard parametric survival distributions like Gompertz-Makeham distribution, generalised Weibull distribution *etc.* can be employed easily for the cause-specific hazards which can then be used to specify the subdensity or subsurvival functions. From various examples, we see that parametric distributions fit well to subsurvival functions as well as overall survival functions. Once a parameteric model is justified, optimal estimation and testing procedures can easily be worked out using standard techniques (Casella and Berger, 1990)

References

Casella , G. and Berger, R.L. (1990). *Statistical Inference.*

Duxbury, California. Crowder, M. J. (2001). *Classical Competing Risks.* Chapman and Hall/CRC, London.

David, H.A. and Moeschberger, M.L. (1978). *The Theory of Competing Risks.* Griffin, London.

Deshpande, J. V. (1990). A test for bivariate symmetry of dependent competing risks. *Biometrical Journal* **32**, 736-746.

Dewan I., Deshpande J.V. and Kulathinal S.B. (2004). On testing dependence between time to failure and cause of failure via conditional probabilities. *Scandinavian J. Statistics* (to appear).

Hoel, D. G. (1972). A representation of mortality data by competing risks. *Biometrics* **28**, 475-488.

Kalbfleisch, J.D. and Prentice, R.L. (1980). *The statistical analysis of failure time data.* John Wiley, New York.

Kalbfleisch, J.D. and Prentice, R.L. (2002). *The statistical analysis of failure time data.* Second Edition, John Wiley, New Jersey.

Moeschberger, M.L. (1974). Life tests under dependent competing causes of failure. *Technometrics* **16**, 39-47.

Nair, V. N. (1993). Bounds for reliability estimation under dependent censoring. *International Stat. Review* **61**, 169-182.

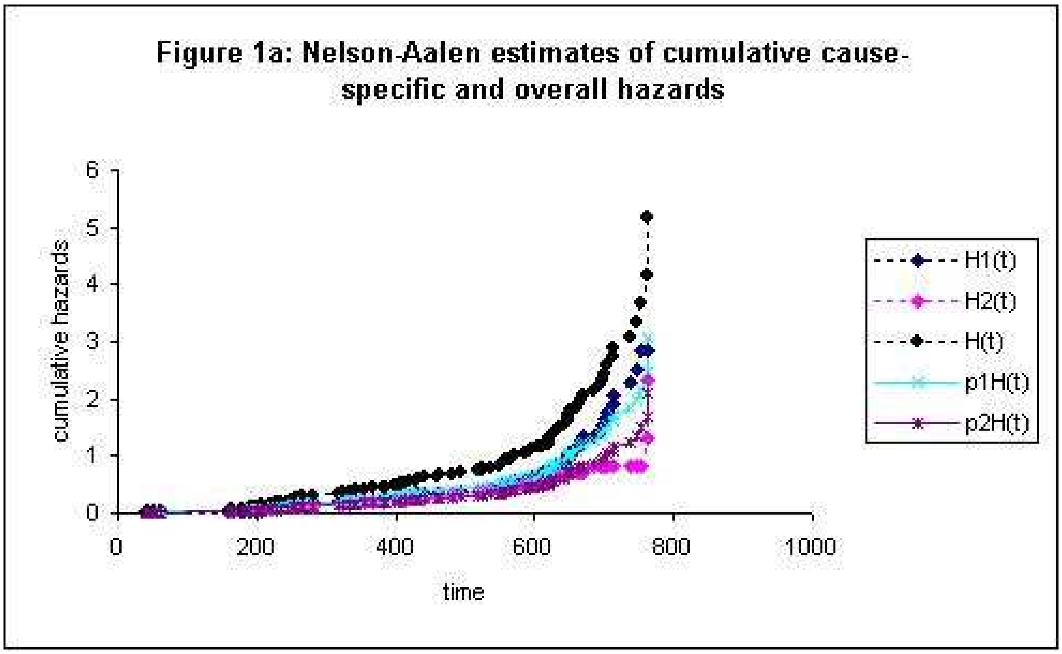


Figure 1a: Nelson-Aalen estimates of cumulative cause-specific and overall hazards
cumulative hazards
0
1
2
3
4
5
6
0
200
400
600
800
1000
time
H1(t)
H2(t)
H(t)
p1H(t)
p2H(t)

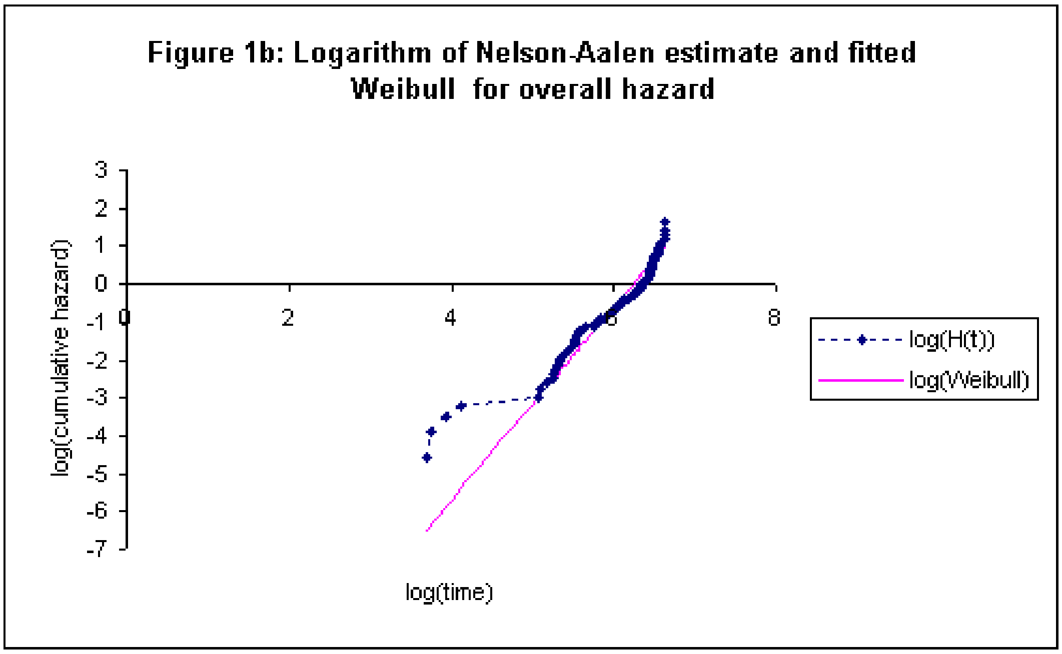


Figure 1b: Logarithm of Nelson-Aalen estimate and fitted Weibull for overall hazard
log(cumulative hazard)
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
0
2
4
6
8
log(time)
log(H(t))
log(Weibull)

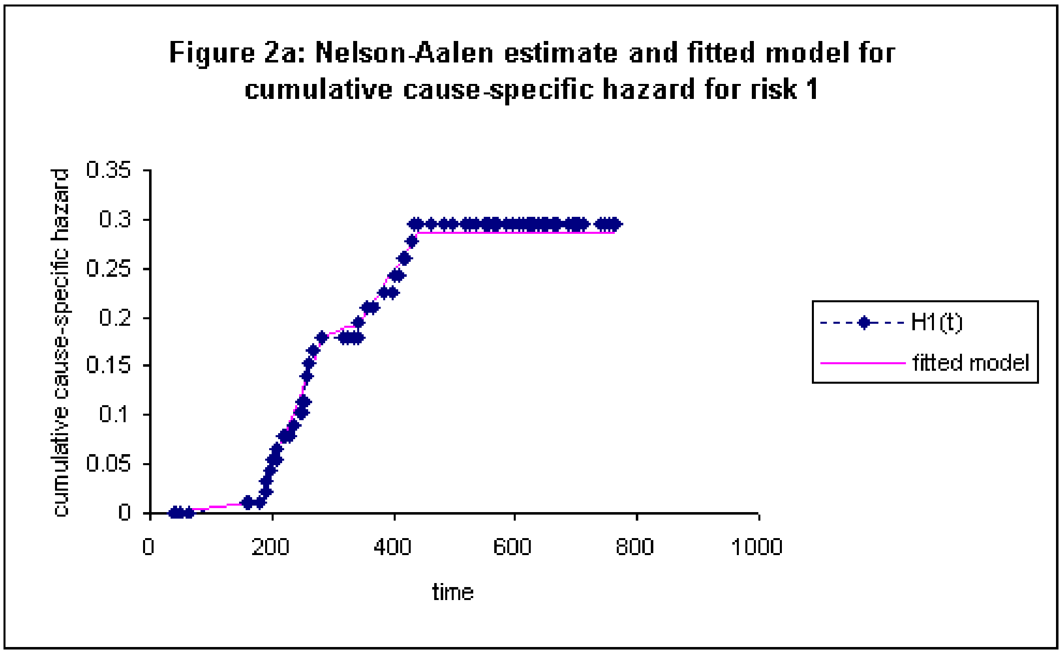


Figure 2a: Nelson-Aalen estimate and fitted model for cumulative cause-specific hazard for risk 1
cumulative cause-specific hazard
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
200
400
600
800
1000
time
H1(t)
fitted model

Figure 2b: Logarithm of Nelson-Aalen estimate and Weibull cumulative cause-specific hazard of risk 2

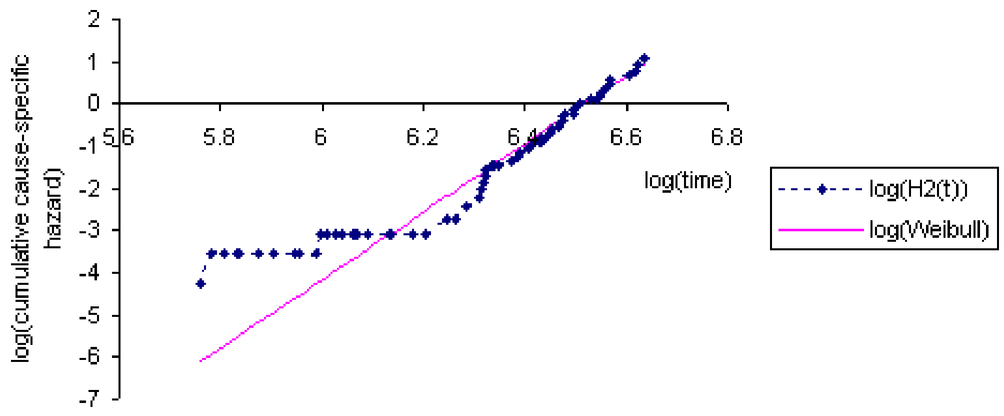
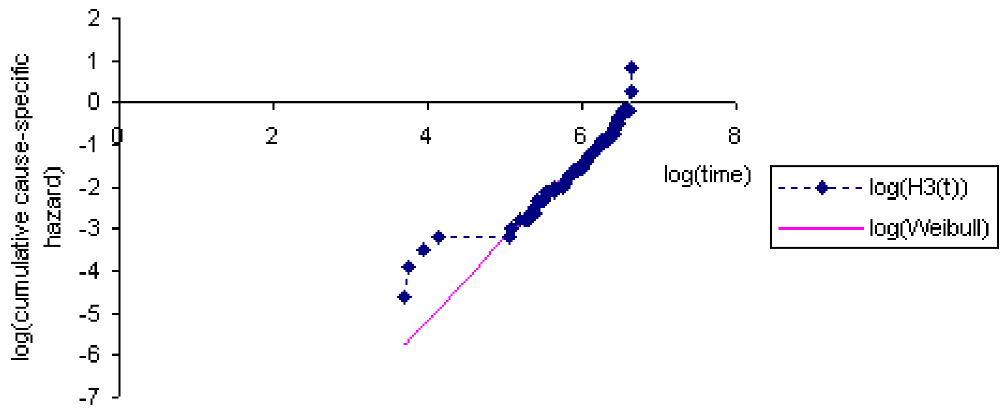


Figure 2c: Logarithm of Nelson-Aalen estimate and Weibull cumulative cause-specific hazard of risk 3



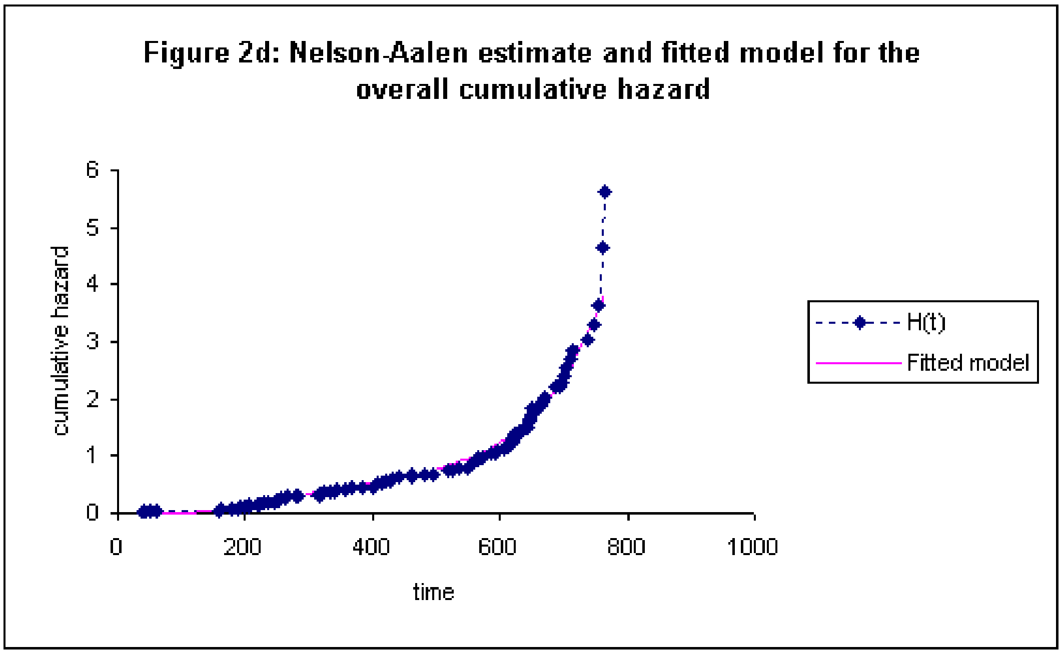


Figure 2d: Nelson-Aalen estimate and fitted model for the overall cumulative hazard
cumulative hazard
time
0
1
2
3
4
5
6
0
200
400
600
800
1000
H(t)
Fitted model

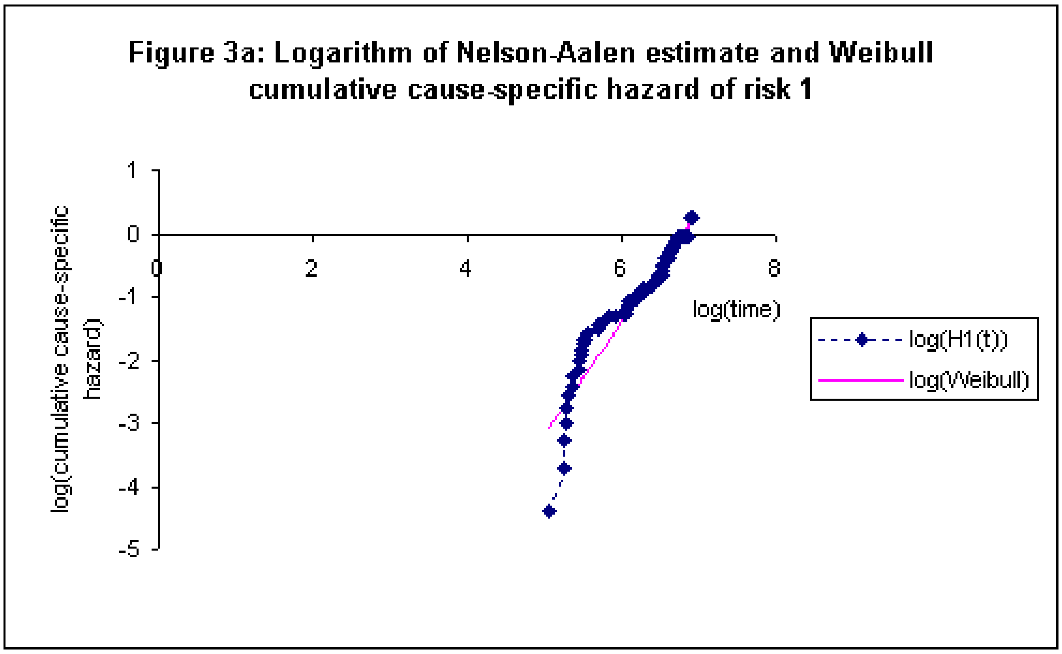


Figure 3a: Logarithm of Nelson-Aalen estimate and Weibull cumulative cause-specific hazard of risk 1
log(cumulative cause-specific hazard)
log(time)
1
0
-1
-2
-3
-4
-5
0
2
4
6
8
log(H1(t))
log(Weibull)

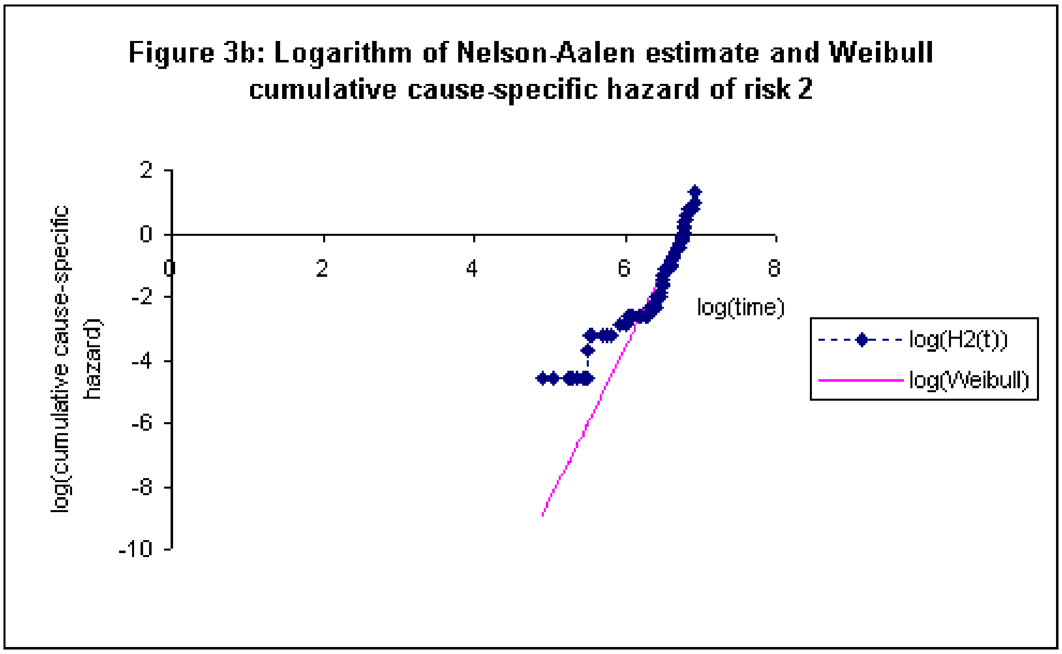


Figure 3b: Logarithm of Nelson-Aalen estimate and Weibull cumulative cause-specific hazard of risk 2
log(cumulative cause-specific hazard)
log(time)
log(H2(t))
log(Weibull)
2
0
-2
-4
-6
-8
-10
0
2
4
6
8

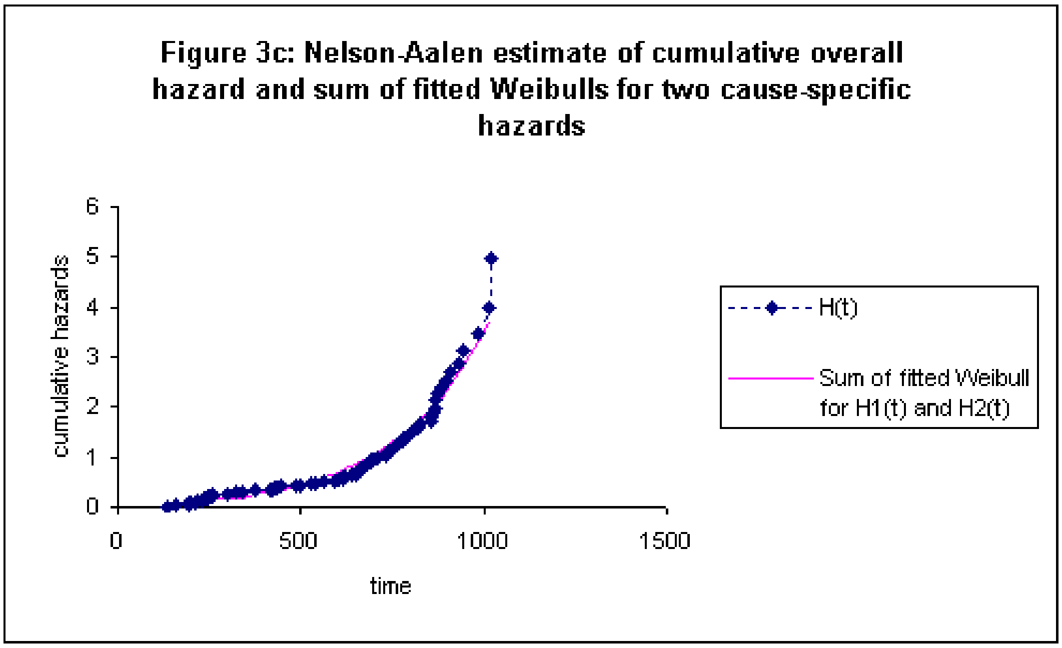


Figure 3c: Nelson-Aalen estimate of cumulative overall hazard and sum of fitted Weibulls for two cause-specific hazards
cumulative hazards
time
H(t)
Sum of fitted Weibull for H1(t) and H2(t)
6
5
4
3
2
1
0
0
500
1000
1500

Figure 4a: Logarithm of Nelson-Aalen estimate and Weibull cumulative overall hazard

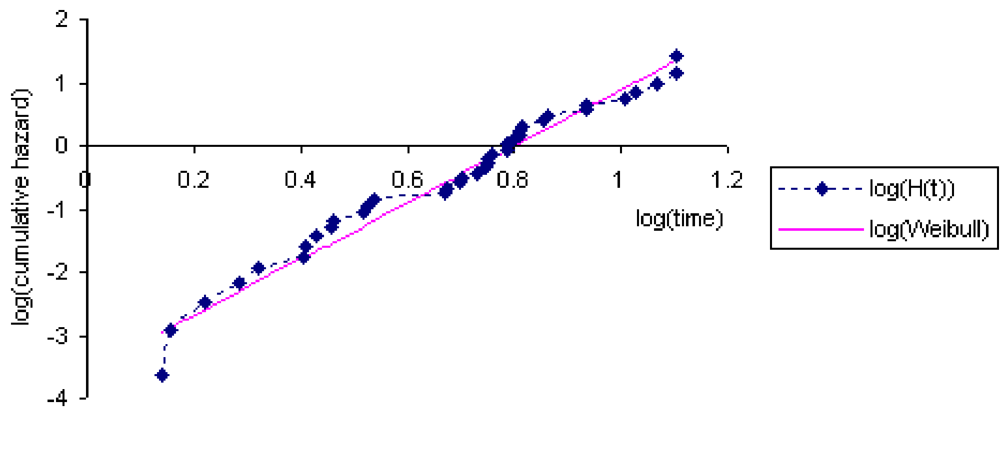


Figure 4b: Empirical subsurvival functions and Weibull subsurvival functions