Current Status Data with two Competing risks and Missing failure types

Technical Report No. ASU/2018/2
Dated 02 August, 2018

Tamalika Koley

and

Anup Dewanji

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Current Status Data with two Competing risks and Missing failure types

Tamalika Koley[1] and Anup Dewanji[2]
Applied Statistics Unit, Indian Statistical Institute
203, B.T. Road, Kolkata-700108
email[1] : tamalikakoley@gmail.com
email[2] : dewanjia@isical.ac.in

Abstract

Missing cause of failure is a common problem in competing risks data. Here we consider a general missing pattern in which one observes a set of possible causes containing the true cause. In this work, we focus on the parametric analysis of current status data with two competing risks and the above-mentioned missing pattern. We make some simpler assumptions on the conditional probability of observing a set of possible causes of failure given the true cause and carry out maximum likelihood estimation of the model parameters. Asymptotic properties of the maximum likelihood estimators are also discussed. Simulation studies are performed to study the finite sample properties of the estimators. Finally, the method is illustrated through some real data sets.

Keywords Monitoring time, Sub-distribution function, Masking probabilities, Identifiability, Maximum likelihood estimation.

1 Introduction

Current status data is a type of survival data in which individual under study is observed at a particular time point (monitoring time point) to note the status (the event of interest has occurred or not by that time, termed as 'failure' or 'survival', respectively) of the individual at the monitoring time point. Failure on an individual may be one of several distinct types; such data are commonly referred to as competing risks data (See Kalbfleisch and Prentice (2002)). Various studies on current status data with competing risks are already present in the literature, most of which focused on nonparametric estimation of the corresponding sub-distribution functions. See, for example, Hudgens et al. (2001), Jewell et al. (2003), Jewell and Kalbfleisch (2004), Maathuis (2006) and many others. In competing risks data, lack of information on the cause or type of failure is a common phenomenon. Typically, when the cause of failure is not exactly observed, one observes a set of possible causes containing the true cause. For example, in usual carcinogenicity studies (See Dewanji et al. (1993)), there are three different types of end-points representing three competing risks. These are (i) death with cancer present and the cause is also cancer, (ii) death with cancer present and the cause is not cancer and finally, (iii) death without cancer. In the third case, clearly, the death is due to some cause other than the cancer. But, between the first two cases, pathologists are often confused in determining whether the death is due to cancer or some other causes when cancer is present, although it is certain that the cancer is present at death. Such general missing pattern was first studied by Flehinger et al. (1998) under the assumption that the cause-specific hazard functions of the competing risks are proportional. They extended the work of Flehtnjer et al. (1998) in which a subsample of those with missing failure types is subjected to a validation procedure to determine the true cause. Dewanji and Sengupta (2003) considered nonparametric estimation of the different cause-specific hazards, based on the above mentioned missing pattern in failure types and without any assumption of proportional hazards due to different types.

In the context of current status data with competing risks, consideration of missing failure types has not received any attention, as much as we know. In this work, for simplicity, we consider parametric analysis of such missing data with two competing risks. It is to be noted that initial consideration of survival data with competing risks and missing failure types dealt with two competing risks (See, for example, Goetghebeur and Ryan (1990); Lo (1991); Mukerjee and Wang (1993)) in which case the failure type is either exactly observed or completely missing. In Section 2, the data is described along with the relevant notation and the corresponding likelihood is constructed. Section 3 discusses the model identifiability under different assumptions. Section 4 derives the asymptotic results. Section 5 presents a detailed simulation study to investigate the finite sample properties of the estimators. Section 6 illustrates the proposed estimation procedure through analysis of some real data sets. Finally, Section 7 ends with some concluding remarks.

2 The Data and the Likelihood

Let T be the random variable representing the failure incidence time subject to $m(=2)$ competing risks. Let $J \in \{1, 2\}$ be the random variable denoting the true cause of failure and $G \in \{\{1\}, \{2\}, \{1, 2\}\}$ denote the observed set of possible causes containing the true cause. Write $\mathcal{G}$ as the support of G, which is $\{\{1\}, \{2\}, \{1, 2\}\}$ in this case. Misspecification may occur if an individual experiences the failure but the observed set of possible causes does not contain the corresponding true cause. We do not consider such type of misspecification in this work. Note that when G is the singleton set J, there is no missingness in the failure type, while $G = \{1, 2\}$ represents no information on failure type (that is, failure type is completely missing) for $m = 2$.

The sub-distribution function for failure due to cause j is given by

$$F_j(t; \underset{\sim}{\theta}) = P\left[T \leq t, J = j\right] = \int_0^t f_j(u; \underset{\sim}{\theta}) du,$$

for $j = 1, 2$, where $f_j(\cdot; \underset{\sim}{\theta})$ is the corresponding sub-density functions with $\underset{\sim}{\theta}$ being the vector of parameters associated with the failure time distribution. The overall distribution function of T is the sum of these two sub-distribution functions given by $F(t; \underset{\sim}{\theta}) = \sum_{j=1}^{2} F_j(t; \underset{\sim}{\theta})$ with the survivor function $S(t; \underset{\sim}{\theta}) = P\left[T > t\right] = 1 - F(t; \underset{\sim}{\theta})$. Let X denote the monitoring time assumed to be independent of the random vector (T, J) not involving any common parameter. Also, if $T \leq X$, we observe G instead of J and if $T > X$, we write G as the empty set ϕ. Therefore, the support $\mathcal{G}$ of G now includes also the empty set ϕ. Hence, for the n individuals under study, we observe (x_i, g_i), the realization of the random vector (X, G) for the i^{th} individual, for $i = 1, \cdots, n$. Let us define the indicator variable $\delta_{gi} = I\{g = g_i\}$, for $g \in \mathcal{G}$, and write $\delta_i = \sum_{g \in \mathcal{G}} \delta_{gi}$, for $i = 1, \cdots, n$, which is equal to 1, if the i^{th} individual is observed to have failed on or before the monitoring time x_i and 0, otherwise. For $j = 1, 2$ and $g \in \mathcal{G} - \phi$, let us define the conditional probability of observing g as the set of possible causes of failure, given that the true cause is j and all other related information, as

$$p_{gj}(x, t) = \lim_{\delta t \to 0} P\left[G = g \mid t < T \leq t + \delta t \leq X = x, J = j\right],$$

if $g \ni j$ and 0, otherwise, since there is no misspecification in the data, as described before. Thus, for a fixed j, $\sum_{g \ni j} p_{gj}(x, t) = 1$, with $t \leq x$. These are also known as masking probabilities (Basu, 2009) while the cause of failure is said to be masked. Note that, when $p_{gj}(x, t) = p$, independent of g, j, x and t, we have missing completely at random (MCAR) and when $p_{g,j}(x, t)$

may depend on g and x but not on j and t, then we have missing at random (MAR). The missingness is said to be ignorable in the above two situations. In other situations, the missingness is non-ignorable. We are interested in such non-ignorable missing patterns, when $p_{g,j}(x,t)$ depends on either j or t or both (See Little and Rubin (1987)). Note that

$$\begin{aligned} P\left[T \le x, G = g\right] &= \sum_{j \in g} \int_0^x \lim_{\delta t \to 0} \frac{P\left[G = g, t < T \le t + \delta t \le X = x, J = j\right]}{\delta t} dt \\ &= \sum_{j \in g} \left[\int_0^x \lim_{\delta t \to 0} \frac{P\left[t < T \le t + \delta t \le X = x, J = j\right]}{\delta t} \times \right. \\ &\qquad \left. P\left[G = g \mid t < T \le t + \delta t \le X = x, J = j\right] dt \right] \\ &\propto \sum_{j \in g} \int_0^x f_j(t; \underset{\sim}{\theta}) p_{gj}(x,t) dt \end{aligned} \tag{2.1}$$

Throughout this paper, we consider the simpler situation when the masking probabilities $p_{gj}(x,t)$ depend on g and j but not on the failure and monitoring time points t and x, respectively; that is, $p_{gj}(x,t) = p_{gj}$, for all $j = 1, 2$ and $g \ni j$. Let us write

$$p_j = p_{\{j\}j} = P\left[G = \{j\} \mid T = t \le X = x, J = j\right], \text{ for } j = 1, 2.$$

Then, clearly, $P\left[G = \{1,2\} \mid T = t \le X = x, J = j\right] = 1 - p_j$, for $j = 1, 2$.

Theorem 2.1. The density of the observed random vector (X, G) with respect to the dominating probability measure $H \times \mu$, where H, μ are the distribution function of X and counting measure, respectively, is given by,

$$f^*(x,g) = \begin{cases} \left\{ \int_0^x f_1(u; \underset{\sim}{\theta}) p_1 du \right\} h(x), & \text{if } g = \{1\}, T \le X = x \\ \left\{ \int_0^x f_2(u; \underset{\sim}{\theta}) p_2 du \right\} h(x), & \text{if } g = \{2\}, T \le X = x \\ \left\{ \int_0^x \left[f_1(u; \underset{\sim}{\theta})(1-p_1) + f_2(u; \underset{\sim}{\theta})(1-p_2) \right] du \right\} h(x), & \text{if } g = \{1,2\}, T \le X = x \\ S(x; \underset{\sim}{\theta}) h(x), & \text{if } g = \phi, T > X = x. \end{cases}$$

Here, $h(\cdot)$ is the probability density corresponding to the distribution function $H(\cdot)$.

Proof: For $g = \{1\}, T \le X = x$,

$$\begin{aligned} f^*(x, g = \{1\}) &= P\left[T \le x, G = \{1\}\right] h(x) \\ &= P\left[T \le x, G = \{1\}, J = 1\right] h(x) \\ &= \left[\int_0^x P\left[G = \{1\} \mid T = t \le x, J = 1\right] f_1(t; \underset{\sim}{\theta}) dt \right] h(x) \\ &= \left[\int_0^x p_1 f_1(t; \underset{\sim}{\theta}) dt \right] h(x). \end{aligned}$$

Similarly, for $g = \{2\}, T \le X = x$, $f^*(x, g = \{2\}) = \left[\int_0^x p_2 f_2(t; \underset{\sim}{\theta}) dt \right] h(x)$.
Now for $g = \{1,2\}, T \le X = x$,

$$\begin{aligned} f^*(x, g = \{1,2\}) &= P\left[T \le x, G = \{1,2\}\right] h(x) \\ &= P\left[T \le x, G = \{1,2\}, J = 1\right] h(x) + P\left[T \le x, G = \{1,2\}, J = 2\right] h(x) \\ &= \left[\int_0^x (1-p_1) f_1(t; \underset{\sim}{\theta}) dt + \int_0^x (1-p_2) f_2(t; \underset{\sim}{\theta}) dt \right] h(x) \end{aligned}$$

Finally, for $T > X = x$,

$$f^*(x, g = \phi) = P\left[T > x\right] h(x) = S(x; \underset{\sim}{\theta}) h(x).$$

Now, we verify that $\int_0^{\infty} \sum_g f^*(x,g)dx = 1$. We can write

$$\begin{aligned}
\int_0^{\infty} \sum_g f^*(x,g)dx &= \int_0^{\infty} \left[\int_0^x p_1 f_1(t; \underset{\sim}{\theta})dt\right] h(x) + \int_0^{\infty} \left[\int_0^x p_2 f_2(t; \underset{\sim}{\theta})dt\right] h(x) \\
&+ \left[\int_0^x (1-p_1) f_1(t; \underset{\sim}{\theta})dt + \int_0^x (1-p_2) f_2(t; \underset{\sim}{\theta})dt\right] h(x) + S(x; \underset{\sim}{\theta})h(x). \\
&= \int_0^{\infty} \left[F(x; \underset{\sim}{\theta}) + S(x; \underset{\sim}{\theta})\right] h(x)dx \\
&= \int_0^{\infty} h(x)dx = 1.
\end{aligned}$$

□

Note that the likelihood function of the observed data $(\underset{\sim}{x}, \underset{\sim}{g})$, with $\underset{\sim}{x} = (x_1, \cdots, x_n)^T$ and $\underset{\sim}{g} = (g_1, \cdots, g_n)^T$, is given by the product of these densities as

$$L(\underset{\sim}{\theta}, p_1, p_2 \mid \underset{\sim}{x}, \underset{\sim}{g}) \propto \prod_{i=1}^{n} f^*(x_i, g_i)$$

$$= \prod_{i=1}^{n} \left[\left\{ \int_0^{x_i} p_1 f_1(t; \underset{\sim}{\theta})dt \right\}^{\delta_{\{1\}i}} \left\{ \int_0^{x_i} p_2 f_2(t; \underset{\sim}{\theta})dt \right\}^{\delta_{\{2\}i}} \right.$$

$$\left. \left\{ \int_0^{x_i} (1-p_1) f_1(t; \underset{\sim}{\theta})dt + \int_0^{x_i} (1-p_2) f_2(t; \underset{\sim}{\theta})dt \right\}^{\delta_{\{1,2\}i}} S(x_i; \underset{\sim}{\theta})^{1-\delta_i} \right]. \quad (2.2)$$

It is clear from the likelihood function (2.2) that there is no closed form expressions for the maximum likelihood estimators (MLEs) of the model parameters. So estimation is carried out using some numerical maximization method subject to the parameters being identifiable. We have used **optim** function in R software to obtain the maximum likelihood estimates of the parameters. The variance-covariance matrix of the estimators is calculated from the inverse of the hessian matrix computed at the MLE.

3 Identifiability Issues

We assume that the sub-density functions $f_j(\cdot; \underset{\sim}{\theta})$, for $j = 1, 2$, satisfy the standard identifiability condition given by

$$f_j(x; \underset{\sim}{\theta}^{(1)}) = f_j(x; \underset{\sim}{\theta}^{(2)}), \text{ for all } x, j = 1, 2 \iff \underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}. \tag{3.1}$$

In the following subsections, we make different assumptions on the two probability terms p_1 and p_2 and study the corresponding identifiability issues.

3.1 p_1 and p_2 both known

Here, the parameter of interest is only the vector $\underset{\sim}{\theta}$. In order to investigate identifiability of the parameter vector $\underset{\sim}{\theta}$ based on the data type mentioned in the previous section, we set the corresponding probability distributions as equal for two values of $\underset{\sim}{\theta}$ and then try to prove that these two values of $\underset{\sim}{\theta}$ are equal. Suppose $\underset{\sim}{\theta}^{(1)}$ and $\underset{\sim}{\theta}^{(2)}$ are two values of the parameter vector $\underset{\sim}{\theta}$ and consider the identity

$$f^*(x, g \mid \underset{\sim}{\theta}^{(1)}) = f^*(x, g \mid \underset{\sim}{\theta}^{(2)}), \tag{3.2}$$

for all $x \in \text{Dom}X$, the domain of X, and $g \in \mathcal{G}$. We need to prove that $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}$. Putting $g = \{1\}$ in both sides of (3.2), we get $F_1(x; \underset{\sim}{\theta}^{(1)}) = F_1(x; \underset{\sim}{\theta}^{(2)})$, for all $x \in \text{Dom}X$. Similarly, putting $g = \{2\}$ in both sides of (3.2), we get $F_2(x; \underset{\sim}{\theta}^{(1)}) = F_2(x; \underset{\sim}{\theta}^{(2)})$, for all $x \in \text{Dom}X$. Hence, from the identifiability condition 3.1, we have $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}$. Thus, the model is identifiable.

3.2 p_1 and p_2 both unknown

In order to investigate the identifiability issue, let us consider two choices of the parameter vector $(\underset{\sim}{\theta}^{(1)}, p_1^{(1)}, p_2^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, p_1^{(2)}, p_2^{(2)})$ along with the identity

$$f^*(x, g \mid \underset{\sim}{\theta}^{(1)}, p_1^{(1)}, p_2^{(1)}) = f^*(x, g \mid \underset{\sim}{\theta}^{(2)}, p_1^{(2)}, p_2^{(2)}), \qquad (3.3)$$

for all $x \in \text{Dom}X$ and $g \in \mathcal{G}$. Clearly, the vector of parameters $(\underset{\sim}{\theta}, p_1, p_2)$ is non-identifiable if we can find two different sets $(\underset{\sim}{\theta}^{(1)}, p_1^{(1)}, p_2^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, p_1^{(2)}, p_2^{(2)})$ such that

$p_1^{(1)} F_1(x; \underset{\sim}{\theta}^{(1)}) = p_1^{(2)} F_1(x; \underset{\sim}{\theta}^{(2)})$

$p_2^{(1)} F_2(x; \underset{\sim}{\theta}^{(1)}) = p_2^{(2)} F_2(x; \underset{\sim}{\theta}^{(2)})$

and $F_1(x; \underset{\sim}{\theta}^{(1)}) + F_2(x; \underset{\sim}{\theta}^{(1)}) = F_1(x; \underset{\sim}{\theta}^{(2)}) + F_2(x; \underset{\sim}{\theta}^{(2)})$,

for all x. These three equations give

$$\frac{p_1^{(1)} F_1(x; \underset{\sim}{\theta}^{(1)})}{p_2^{(1)} F_2(x; \underset{\sim}{\theta}^{(1)})} = \frac{p_1^{(2)} F_1(x; \underset{\sim}{\theta}^{(2)})}{p_2^{(2)} F_2(x; \underset{\sim}{\theta}^{(2)})} = k(x), \text{ say}, \qquad (3.4)$$

possibly a function of $p_1^{(1)}, p_1^{(2)}, p_2^{(1)}, p_2^{(2)}, \underset{\sim}{\theta}^{(1)}$ and $\underset{\sim}{\theta}^{(2)}$ as well, which implies

$$F_1(x; \underset{\sim}{\theta}^{(1)}) = k(x) \frac{p_2^{(1)}}{p_1^{(1)}} F_2(x; \underset{\sim}{\theta}^{(1)}),$$

$$F_1(x; \underset{\sim}{\theta}^{(2)}) = k(x) \frac{p_2^{(2)}}{p_1^{(2)}} F_2(x; \underset{\sim}{\theta}^{(2)})$$

$$\text{and } (1 + k(x) \frac{p_2^{(1)}}{p_1^{(1)}}) F_2(x; \underset{\sim}{\theta}^{(1)}) = (1 + k(x) \frac{p_2^{(2)}}{p_1^{(2)}}) F_2(x; \underset{\sim}{\theta}^{(2)}), \qquad (3.5)$$

for all x. Using (3.5) and $p_2^{(1)} F_2(x; \underset{\sim}{\theta}^{(1)}) = p_2^{(2)} F_2(x; \underset{\sim}{\theta}^{(2)})$, it is clear that

$\frac{1+k(x)\frac{p_2^{(2)}}{p_1^{(2)}}}{1+k(x)\frac{p_2^{(1)}}{p_1^{(1)}}} = \frac{p_2^{(2)}}{p_2^{(1)}}$, since $F_2(x; \underset{\sim}{\theta}^{(2)}) \neq 0$, for some $x \in \text{Dom}X$.

$\iff k(x) = \frac{p_1^{(2)} p_1^{(1)} (p_2^{(2)} - p_2^{(1)})}{p_2^{(2)} p_2^{(1)} (p_1^{(1)} - p_1^{(2)})}$, for all $x \in$ Dom X.

From this expression of $k(x)$, it is clear that, if the model is identifiable, $k(x)$ will be of the form $0/0$, which contradicts (3.4) since $k(x) > 0$, for some $x \in \text{Dom}X$. Hence, the model is not identifiable. For the identifiability of the model parameter $\underset{\sim}{\theta}$, we need to find some conditions on the probabilities p_1, p_2.

Theorem 3.1. The necessary and sufficient condition for the model to be identifiable is $p_2 = cp_1$, where $c > 0$ is some known constant.

Proof: Suppose that the model parameter $\underset{\sim}{\theta}$ is identifiable. From (3.5) and noting that $k(x) = k$, independent of x, we have, for all x,

$(1 + k\frac{p_2^{(1)}}{p_1^{(1)}}) = (1 + k\frac{p_2^{(2)}}{p_1^{(2)}})$, since $F_2(x; \underset{\sim}{\theta}^{(1)}) = F_2(x; \underset{\sim}{\theta}^{(2)})$, from identifiability of $\underset{\sim}{\theta}$. This implies, $\frac{p_2^{(1)}}{p_1^{(1)}} = \frac{p_2^{(2)}}{p_1^{(2)}} \iff p_2 = cp_1$, where $c > 0$ is some known constant.

Conversely, let the probabilities p_1 and p_2 satisfy $p_2 = cp_1$ for some known constant c. To check the identifiability of the parameters, let $(\underset{\sim}{\theta}^{(1)}, p_1^{(1)})$ and $(\underset{\sim}{\theta}^{(2)}, p_1^{(2)})$ be two choices of the parameter vector $(\underset{\sim}{\theta}, p_1)$ and consider the identity (3.3). First putting $g = \{1, 2\}$ in both sides of (3.3), we get

$$(1-p_1^{(1)})F_1(x;\underset{\sim}{\theta}^{(1)})+(1-cp_1^{(1)})F_2(x;\underset{\sim}{\theta}^{(1)}) = (1-p_1^{(2)})F_1(x;\underset{\sim}{\theta}^{(2)})+(1-cp_1^{(2)})F_2(x;\underset{\sim}{\theta}^{(2)})$$

$$\text{or, } F(x;\underset{\sim}{\theta}^{(1)}) - cp_1^{(1)}F(x;\underset{\sim}{\theta}^{(1)}) = F(x;\underset{\sim}{\theta}^{(2)}) - cp_1^{(2)}F(x;\underset{\sim}{\theta}^{(2)}), \tag{3.6}$$

for all $x \in \text{Dom}X$. Next we put $g = \phi$ in (3.3) and obtain $S(x;\underset{\sim}{\theta}^{(1)}) = S(x;\underset{\sim}{\theta}^{(2)})$, for all $x \in \text{Dom}X$. Using this in (3.6), we get $p_1^{(1)} = p_1^{(2)}$. Finally, putting $g = \{1\}$ in both sides of (3.3) and considering $p_1^{(1)} = p_1^{(2)}$, clearly,

$F_1(x;\underset{\sim}{\theta}^{(1)}) = F_1(x;\underset{\sim}{\theta}^{(2)})$, for all $x \in \text{Dom}X$. Similarly, it can be shown that $F_2(x;\underset{\sim}{\theta}^{(1)}) = F_2(x;\underset{\sim}{\theta}^{(2)})$, for all $x \in \text{Dom}X$. Thus the model is identifiable.

□

Note that the assumption $p_2 = cp_1$, for some known constant $c > 0$, induces a restriction on p_1 as given by $0 \leq p_1 \leq \min(1, 1/c)$.

3.3 Introducing Heterogeneity and Uncertainty

It may be of interest to study if p_1 and p_2 can be allowed to be different for different individuals thereby modeling the missing probabilities to be heterogeneous. This is usually done by considering the n values of the pair (p_1, p_2) for the n individuals to be independent realizations from a common distribution, say, $\nu(p_1, p_2)$. Let us denote the marginal means of p_1 and p_2 by μ_1 and μ_2, respectively. Clearly, the density of the observed random vector (X, G) is given by

$$f^{**}(x, g) = \int_0^1 \int_0^1 f^*(x, g) \mathrm{d}\nu(p_1, p_2), \tag{3.7}$$

where $f^*(x, g)$ is as in Theorem 2.1. The density function $f^{**}(x, g)$ clearly reduces to

$$f^{**}(x,g) = \begin{cases} \mu_1 F_1(x;\underset{\sim}{\theta}) h(x), & \text{if } g = \{1\}, T \leq X = x \\ \mu_2 F_2(x;\underset{\sim}{\theta}) h(x), & \text{if } g = \{2\}, T \leq X = x \\ \left\{ (1-\mu_1) F_1(x;\underset{\sim}{\theta}) + (1-\mu_2) F_2(x;\underset{\sim}{\theta}) \right\} h(x), & \text{if } g = \{1,2\}, T \leq X = x \\ S(x;\underset{\sim}{\theta}), & \text{if } g = \phi, T > X = x. \end{cases}$$

Note that this $f^{**}(x, g)$ has the same algebraic form as that of $f^*(x, g)$, as in Theorem 2.1, with p_1 and p_2 replaced by μ_1 and μ_2, respectively. Therefore, this general modeling of heterogeneity is indistinguishable with the homogeneous model, except the interpretation. In view of Theorem 3.1, this heterogeneous model with parameter vector $(\mu_1, \mu_2, \underset{\sim}{\theta})$ is identifiable if and only if

$\mu_2 = c\mu_1$, for some known constant $c > 0$. As before, the vector of independent parameters is now $(\mu_1, \underset{\sim}{\theta})$ with the restriction $0 \leq \mu_1 \leq \min(1, 1/c)$. The corresponding likelihood function can be obtained by the product of these density functions over all the observed (x_i, g_i), for $i = 1, \cdots, n$.

As a simpler alternative, one can have the joint distribution $\nu(p_1, p_2)$ to be completely known. In that case, the two means μ_1 and μ_2 also become known and the model becomes indistinguishable with that of known p_1 and p_2, as in Section 3.1, except the interpretation. For example, $\nu(p_1, p_2)$ may be modeled as two independent Beta distributions with known parameter values, or two independent Uniform distributions with known limits.

On the other hand, one may be interested in modeling uncertainty in the missing probabilities (that is, p_1 and p_2), however, assumed to be common for all the individuals. This is usually modeled in similar manner by assuming a distribution $\nu(p_1, p_2)$, say, for the common probability pair (p_1, p_2). The resulting likelihood function is given by

$$\int_0^1 \int_0^1 L(\underset{\sim}{\theta}, p_1, p_2 \mid \underset{\sim}{x}, \underset{\sim}{g}) \mathrm{d}\nu(p_1, p_2), \tag{3.8}$$

where $L(\underset{\sim}{\theta}, p_1, p_2 \mid \underset{\sim}{x}, \underset{\sim}{g})$ is the likelihood of the form (2.2). It is to be noted that, the pair (p_1, p_2) being random and common to all the n individuals, there is induced dependence between the individuals and, hence, the likelihood cannot be written as a product of n density terms. Instead, the likelihood becomes a complicated function involving the moments of (p_1, p_2) of different orders. This makes the investigation into the identifiability issue rather difficult, in particular, in the presence of unknown parameters associated with $\nu(p_1, p_2)$. For simplicity, we assume that the joint distribution $\nu(p_1, p_2)$ is completely known, as before, and then the identifiability of $\underset{\sim}{\theta}$ follows easily. In particular, let $\underset{\sim}{\theta}^{(1)}$ and $\underset{\sim}{\theta}^{(2)}$ be two values of the parameter vector $\underset{\sim}{\theta}$ and consider the identity

$$\int_0^1 \int_0^1 L(\underset{\sim}{\theta}^{(1)}, p_1, p_2 \mid \underset{\sim}{x}, \underset{\sim}{g}) \mathrm{d}\nu(p_1, p_2) = \int_0^1 \int_0^1 L(\underset{\sim}{\theta}^{(2)}, p_1, p_2 \mid \underset{\sim}{x}, \underset{\sim}{g}) \mathrm{d}\nu(p_1, p_2), \tag{3.9}$$

for all $(\underset{\sim}{x}, \underset{\sim}{g})$. Consider $g_i = \delta_{\{1\}i}$ for all i so that $F_1(x; \underset{\sim}{\theta}^{(1)}) = F_1(x; \underset{\sim}{\theta}^{(2)})$ for all x. Similarly, the case $g_i = \delta_{\{2\}i}$ for all i gives $F_2(x; \underset{\sim}{\theta}^{(1)}) = F_2(x; \underset{\sim}{\theta}^{(2)})$ for all x. This implies $F(x; \underset{\sim}{\theta}^{(1)}) = F(x; \underset{\sim}{\theta}^{(2)})$ for all x and hence $\underset{\sim}{\theta}^{(1)} = \underset{\sim}{\theta}^{(2)}$.

4 Asymptotic Results

Let us write $\underset{\sim}{\psi} = (\underset{\sim}{\theta}, p_1)$ with $0 < p_1 < 1$, to represent the vector of the model parameters, although in some cases p_1 may not be relevant (See Sections 3.1 and 3.3). As mentioned before, the likelihood function of $L(\underset{\sim}{\psi} \mid \underset{\sim}{x}, \underset{\sim}{g})$ can be written as the product of n densities $f^*(x, g)$, or $f^{**}(x, g)$, corresponding to the n iid random vectors (x, g), except for the study of uncertainty in Section 3.3. Let us assume that the failure time distribution satisfies the regularity conditions (See Lehmann and Casella (1998, p449)) and all the third order derivatives $\frac{\partial^3}{\partial\theta_k \partial\theta_l \partial\theta_m} f_j(x; \underset{\sim}{\theta})$, for $j = 1, 2$, exist and are continuous function of $\underset{\sim}{\theta}$. To investigate the consistency and asymptotic normality of the MLE of the parameter vector $\underset{\sim}{\psi}$, let us first note the following conditions.

R1 The parameter vector $\underset{\sim}{\psi}$ is identifiable with respect to the density function $f^*(\cdot, \cdot; \underset{\sim}{\psi})$, or $f^{**}(\cdot, \cdot; \underset{\sim}{\theta})$. This follows from the results in Section 3.

R2 The parameter space of the vector $\underset{\sim}{\psi} = (\underset{\sim}{\theta}, p_1)$ is open since the same for the vector $\underset{\sim}{\theta}$ is open.

R3 The observations (x_i, g_i), for $i = 1, \cdots, n$, are iid random vectors with common density $f^*(\cdot, \cdot; \underset{\sim}{\psi})$, or $f^{**}(\cdot, \cdot; \underset{\sim}{\theta})$.

R4 The sample space of the random vector (X, G), that is the support of the density $f^*(\cdot,\cdot;\underset{\sim}{\psi})$, or $f^{**}(\cdot,\cdot;\underset{\sim}{\theta})$ is independent of the parameter vector $\underset{\sim}{\psi}$.

R5 Note that the failure time distribution satisfies the regularity conditions so that the partial derivatives $\frac{\partial^3}{\partial\theta_k\partial\theta_l} f_j(x;\underset{\sim}{\theta})$, $\frac{\partial^3}{\partial\theta_k\partial\theta_l\partial\theta_m} f_j(x;\underset{\sim}{\theta})$, for $j = 1, 2$, exist and are continuous. So, the corresponding derivatives of $\int_0^x f_j(x;\underset{\sim}{\theta}) = F_j(x;\underset{\sim}{\theta})$ are obtained by differentiating under the integral sign so that the third order derivatives of $f^*(\cdot,\cdot;\underset{\sim}{\psi})$, or $f^{**}(\cdot,\cdot;\underset{\sim}{\theta})$ also exist and are continuous.

R6 Note that the sub-distribution functions $F_j(\cdot)$ are probability distributions and are bounded above by 1. So, the third order derivatives if exist are also bounded above by a function of (x, g) in the neighbourhood of the true value $\underset{\sim}{\psi}_0$ of the parameter vector $\underset{\sim}{\psi}$.

R7 The Fisher information matrix is assumed to be positive definite.

Theorem 4.1. The MLE of the parameter vector $\underset{\sim}{\psi}$, given by $\underset{\sim}{\hat{\psi}}$, with $\underset{\sim}{\psi}_0$ being the true value, satisfies

i. $\underset{\sim}{\hat{\psi}} \xrightarrow{P} \underset{\sim}{\psi}_0$,

ii. $\sqrt{n}(\underset{\sim}{\hat{\psi}} - \underset{\sim}{\psi}_0)$ is asymptotically a mean zero normal vector with variance-covariance matrix given by the inverse of the hessian matrix computed at the MLE.

Proof: The proof directly follows from Theorem 7.5.2 of Lehmann and Casella (1998, p463-465) using **R1-R7.**

5 Simulation Studies

In order to investigate the finite sample properties of the MLEs of the parameters of interest, we carry out several simulation studies. First we simulate n observations from the assumed failure time distribution $F(\cdot)$ and then the corresponding monitoring times are simulated from a known distribution $H(\cdot)$. If for a particular observation, the failure time (t) is less than the monitoring time (x), then the corresponding set (g) of possible causes of failure is generated by using the probability distribution $P\,[G = g \mid T \leq x]$, for $g \in \{\{1\}, \{2\}, \{1, 2\}\}$. This probability is given by $\frac{\sum_{j \in g} p_{gj} F_j(x)}{F(x)}$, where $F_j(\cdot)$ is the sub-distribution function of the failure time due to cause j, for $j = 1, 2$, and $F(\cdot) = F_1(\cdot) + F_2(\cdot)$. Thus we get the simulated data set (x_i, g_i), for $i = 1, \cdots, n$. We carry out 10000 simulations to get 10000 such data sets.

For each simulated data set, we obtain the maximum likelihood estimates of the model parameters along with their standard errors. Average of the standard errors over the 10000 simulations (denoted by ASE) are computed. Sample standard errors (denoted by SSE) are computed from the sample variance of the 10000 estimates of the parameters from as many simulations. Cover percentage of the asymptotic 95% confidence interval of a parameter based on normal approximation is estimated by the proportion of times these intervals contain the true value in the 10000 simulations (denoted by CP).

The simulation is carried out for three different choices of sample size $n = 50$, 150 and 250. First, the failure time is taken to be exponential with rate parameter $\lambda = 1$ and the two types of failures occurring with rate ratio $6 : 4$. The monitoring time is also taken to be exponential with rate parameter 0.8. We consider the cases described in Sections 3.1-3.2 one by one: first (i) $(p_1, p_2) = (0.9, 0.8)$ and (0.6,0.7) which are assumed known in the analysis, and then (ii) $p_1 = 0.9$, $p_2 = cp_1$ with $c = 0.9$ and 1, and only c is assumed known in the analysis. Note that the case $c = 1$ implies MCAR. The corresponding results are presented in Tables 1-3, respectively. For case (i), with the decrease in $p_j (j = 1, 2)$ values, the standard errors of the estimates increase (See Table 1), as expected, since higher p_j values means less number of observations with missing causes. As expected, bias and standard error (See ASE and SSE) of the estimates seem to decrease and the values of the ASE and SSE become more similar with the increase in sample size . Also, the estimated coverage probabilities become closer to 0.95 as the sample size increases providing evidence in favour of asymptotic normality of the MLEs of the parameters, as proved in Section 4. The results are seen to be sensitive to misspecification of the value of c used in the analysis (See Tables 2-3). There seems to be a pattern in the nature of bias due to misspecification of c. If the value of c used for analysis is larger (smaller) than the true one, there is a tendency to under-(over-)estimate p_1, possibly because of the opposing push imposed on the value of p_2. This in turn assigns more (less) number of missing causes to cause 1 and hence λ_1 is over-(under-)estimated, while under-(over-)estimating λ_2. Similar pattern, although mildly, is also observed in the corresponding standard errors. We have carried out similar simulation studies for different choices of the model parameters and the results are found to be qualitatively similar, hence not reported.

Next we simulate data from Weibull distribution with two choices for the shape parameter as $\theta = 0.8$ (decreasing failure rate) and $\theta = 1.2$ (increasing failure rate) and scale parameter $(\sum_{j=1}^{2} \lambda_j^{\theta})^{-1}$, where $\lambda_1 = 0.6, \lambda_2 = 0.4$. We consider only the case of Section 3.2 with $p_1 = 0.9$ and $p_2 = cp_1$, where $c = 0.9$, and the analysis assumes c to be known. After generating one simulated dataset, we first fit Weibull distribution and then the exponential distribution to the simulated data to obtain the estimates of the corresponding parameters. Using these estimates we compute the MLEs of the sub-distribution functions at $10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}$ percentiles (denoted by τ_i) of the true Weibull distribution with the above-mentioned shape and scale parameters. The results on bias, the corresponding ASE and SSE, and the estimated coverage probabilities (CPs) are presented in Tables 4 and 5 for the Weibull fit for $\theta = 0.8$ and 1.2, respectively. The results for the exponential fit for $\theta = 0.8$ and 1.2 are presented in Tables 6 and 7, respectively. When the Weibull distribution is fitted to the data, bias of the estimates seems to decrease with increase in sample size, as expected. However, while fitting exponential distribution, the results indicate some bias in the estimates, as is evident with increasing sample size. The results on ASE, SSE and CP, for the Weibull fit, are qualitatively similar, as expected. The same holds for the exponential fit as well except the CP values, as expected. Interestingly, the standard errors seem to increase at the higher percentiles, possibly indicating the difficulty in estimating tail probabilities.

Table 1: Simulation results on $\left\{ \underset{\sim}{\hat{\psi}} = (\hat{\lambda}_1, \hat{\lambda}_2) \right\}$ for Exponential($\lambda = 1$) failure time distribution with $\lambda_1 = 0.6$, $\lambda_2 = 0.4$, and the value of (p_1, p_2) is assumed known in the analysis.

n	para-meters	$p_1 = 0.9, p_2 = 0.8$				$p_1 = 0.6, p_2 = 0.7$			
		$\underset{\sim}{\hat{\psi}}$	ASE	SSE	CP	$\underset{\sim}{\hat{\psi}}$	ASE	SSE	CP
50	λ_1	0.622	0.177	0.178	0.943	0.623	0.186	0.185	0.941
	λ_2	0.417	0.142	0.144	0.936	0.418	0.153	0.151	0.937
150	λ_1	0.606	0.097	0.098	0.947	0.607	0.102	0.102	0.948
	λ_2	0.406	0.078	0.078	0.949	0.404	0.084	0.084	0.945
250	λ_1	0.604	0.075	0.075	0.949	0.606	0.079	0.080	0.952
	λ_2	0.401	0.060	0.060	0.950	0.403	0.065	0.066	0.946

Table 2: Simulation results on $\left\{\underset{\sim}{\hat{\psi}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{p}_1)\right\}$ for exponential(1) failure time distribution with $\lambda_1 = 0.6$, $\lambda_2 = 0.4$ and $p_1 = 0.9$, $p_2 = c\,p_1$ with $c = 0.9$.

Value of c used in the analysis	n	parameters	MLE $\underset{\sim}{\hat{\psi}}$	ASE	SSE	CP
0.8	50	λ_1	0.591	0.171	0.164	0.931
		λ_2	0.438	0.147	0.141	0.958
		p_1	0.934	0.084	0.064	0.995
0.9	50	λ_1	0.616	0.175	0.169	0.950
		λ_2	0.413	0.142	0.136	0.939
		p_1	0.897	0.076	0.068	0.941
1	50	λ_1	0.640	0.178	0.173	0.959
		λ_2	0.389	0.221	0.130	0.916
		p_1	0.862	0.073	0.067	0.953
0.8	150	λ_1	0.582	0.096	0.097	0.927
		λ_2	0.433	0.082	0.082	0.945
		p_1	0.943	0.043	0.041	0.777
0.9	150	λ_1	0.610	0.098	0.099	0.950
		λ_2	0.405	0.079	0.079	0.948
		p_1	0.899	0.039	0.040	0.929
1	150	λ_1	0.635	0.099	0.101	0.955
		λ_2	0.380	0.076	0.076	0.918
		p_1	0.864	0.037	0.038	0.882
0.8	250	λ_1	0.579	0.074	0.071	0.929
		λ_2	0.431	0.063	0.064	0.943
		p_1	0.944	0.033	0.032	0.702
0.9	250	λ_1	0.607	0.075	0.073	0.953
		λ_2	0.402	0.061	0.061	0.945
		p_1	0.900	0.031	0.030	0.934
1	250	λ_1	0.632	0.076	0.074	0.944
		λ_2	0.377	0.058	0.059	0.910
		p_1	0.864	0.029	0.029	0.811

Table 3: Simulation results on $\left\{ \underset{\sim}{\hat{\psi}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{p}_1) \right\}$ for exponential(1) failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4$ and $p_1 = 0.9$, $p_2 = c\,p_1$ with $c = 1$.

Value of c used in the analysis	n	parameters	MLE $\underset{\sim}{\hat{\psi}}$	ASE	SSE	CP
0.9	50	λ_1	0.591	0.169	0.153	0.943
		λ_2	0.439	0.145	0.148	0.940
		p_1	0.922	0.084	0.055	0.981
1	50	λ_1	0.614	0.172	0.156	0.966
		λ_2	0.416	0.140	0.142	0.930
		p_1	0.886	0.082	0.058	0.990
1.1	50	λ_1	0.635	0.175	0.159	0.971
		λ_2	0.395	0.135	0.137	0.912
		p_1	0.856	0.080	0.062	0.963
0.9	150	λ_1	0.579	0.094	0.090	0.938
		λ_2	0.433	0.081	0.082	0.944
		p_1	0.933	0.036	0.031	0.820
1	150	λ_1	0.605	0.096	0.092	0.956
		λ_2	0.408	0.078	0.079	0.951
		p_1	0.893	0.034	0.029	0.984
1.1	150	λ_1	0.627	0.098	0.094	0.961
		λ_2	0.385	0.075	0.076	0.926
		p_1	0.860	0.033	0.029	0.857
0.9	250	λ_1	0.578	0.072	0.071	0.924
		λ_2	0.424	0.061	0.061	0.955
		p_1	0.936	0.027	0.025	0.684
1	250	λ_1	0.604	0.074	0.072	0.954
		λ_2	0.398	0.059	0.058	0.945
		p_1	0.897	0.026	0.023	0.981
1.1	250	λ_1	0.626	0.075	0.073	0.963
		λ_2	0.376	0.057	0.056	0.902
		p_1	0.864	0.025	0.022	0.779

Table 4: Simulation results on $\{\hat{F}_j(\tau_i),\ j = 1, 2,\ i = 1, \cdots, 5\}$ for Weibull failure time distribution (shape parameter $\theta = 0.8$) with $\lambda_1 = 0.6,\ \lambda_2 = 0.4$ and $p_1 = 0.9$, $p_2 = cp_1$ with known $c = 0.9$ ($c = 0.9$ is assumed known in the analysis).

		Cause 1				Cause 2			
n	τ_i	Bias×10^2	ASE	SSE	CP	Bias×10^2	ASE	SSE	CP
	0.051	0.369	0.019	0.043	0.524	0.208	0.018	0.031	0.643
	0.178	0.247	0.038	0.062	0.726	0.275	0.040	0.046	0.847
50	0.534	0.370	0.071	0.071	0.973	0.398	0.079	0.060	0.985
	1.270	0.453	0.101	0.086	0.967	0.102	0.122	0.080	0.983
	2.395	0.384	0.116	0.094	0.970	0.187	0.144	0.092	0.980
	0.051	0.080	0.009	0.024	0.533	0.040	0.104	0.018	0.655
	0.178	0.233	0.021	0.035	0.755	0.171	0.021	0.028	0.851
150	0.534	0.210	0.039	0.040	0.934	0.194	0.040	0.035	0.979
	1.270	0.319	0.053	0.049	0.960	0.078	0.058	0.046	0.983
	2.395	0.042	0.060	0.055	0.961	0.132	0.067	0.052	0.978
	0.051	0.057	0.007	0.010	0.553	0.019	0.010	0.014	0.681
	0.178	0.091	0.016	0.021	0.756	0.158	0.016	0.020	0.880
250	0.534	0.038	0.030	0.031	0.945	0.114	0.030	0.028	0.975
	1.270	0.042	0.041	0.039	0.953	0.014	0.043	0.038	0.977
	2.395	0.018	0.050	0.049	0.956	0.119	0.050	0.043	0.976

Table 5: Simulation results on $\{\hat{F}_j(\tau_i),\ j = 1, 2,\ i = 1, \cdots, 5\}$ for Weibull failure time distribution (shape parameter $\theta = 1.2$) with $\lambda_1 = 0.6,\ \lambda_2 = 0.4$ and $p_1 = 0.9$, $p_2 = cp_1$ with known $c = 0.9$ ($c = 0.9$ is assumed known in the analysis).

		Cause 1				Cause 2			
n	τ_i	Bias$\times 10^2$	ASE	SSE	CP	Bias$\times 10^2$	ASE	SSE	CP
	0.171	0.046	0.021	0.042	0.593	0.031	0.019	0.026	0.718
	0.396	0.591	0.044	0.063	0.762	0.356	0.040	0.043	0.850
50	0.824	0.264	0.077	0.080	0.928	0.086	0.072	0.064	0.948
	1.468	0.637	0.096	0.098	0.948	0.510	0.097	0.087	0.949
	2.240	0.146	0.101	0.104	0.953	0.153	0.105	0.097	0.949
	0.171	0.017	0.011	0.024	0.618	0.016	0.010	0.015	0.753
	0.396	0.166	0.025	0.037	0.800	0.083	0.023	0.025	0.901
150	0.824	0.182	0.044	0.046	0.928	0.063	0.041	0.036	0.972
	1.468	0.301	0.054	0.057	0.945	0.330	0.055	0.050	0.972
	2.240	0.079	0.058	0.061	0.944	0.134	0.061	0.058	0.957
	0.171	0.003	0.009	0.019	0.645	0.011	0.008	0.012	0.757
	0.396	0.139	0.019	0.028	0.807	0.042	0.018	0.019	0.914
250	0.824	0.016	0.034	0.035	0.932	0.032	0.032	0.027	0.948
	1.468	0.075	0.042	0.043	0.948	0.073	0.043	0.037	0.949
	2.240	0.035	0.045	0.048	0.953	0.059	0.047	0.043	0.949

Table 6: Simulation results on $\{\hat{F}_j(\tau_i),\ j=1,2,\ i=1,\cdots,5\}$ for Weibull failure time distribution (shape parameter $\theta=0.8$) with $\lambda_1=0.6,\ \lambda_2=0.4$ and $p_1=0.9$, $p_2=cp_1$ with known $c=0.9$, but exponential fit ($c=0.9$ is assumed known in the analysis).

		Cause 1				Cause 2			
n	τ_i	Bias$\times 10^2$	ASE	SSE	CP	Bias$\times 10^2$	ASE	SSE	CP
50	0.051	2.408	0.009	0.009	0.282	1.732	0.008	0.008	0.359
	0.178	3.479	0.029	0.029	0.678	2.490	0.025	0.026	0.720
	0.534	2.025	0.061	0.062	0.900	1.418	0.054	0.057	0.897
	1.270	0.975	0.085	0.085	0.947	0.744	0.080	0.082	0.938
	2.395	1.750	0.093	0.093	0.937	1.282	0.091	0.093	0.935
150	0.051	2.481	0.005	0.005	0.020	1.788	0.004	0.005	0.068
	0.178	3.675	0.016	0.017	0.377	2.638	0.014	0.015	0.489
	0.534	2.269	0.035	0.037	0.866	1.595	0.031	0.033	0.887
	1.270	0.999	0.049	0.051	0.930	0.791	0.047	0.049	0.936
	2.395	1.949	0.054	0.056	0.916	1.487	0.053	0.056	0.930
250	0.051	2.519	0.004	0.004	0.002	1.818	0.003	0.004	0.006
	0.178	3.784	0.012	0.013	0.189	2.727	0.011	0.011	0.294
	0.534	2.457	0.027	0.029	0.798	1.756	0.024	0.026	0.840
	1.270	0.865	0.038	0.040	0.933	0.656	0.036	0.038	0.934
	2.395	1.933	0.042	0.043	0.914	1.429	0.041	0.043	0.929

Table 7: Simulation results on $\{\hat{F}_j(\tau_i),\ j=1,2,\ i=1,\cdots,5\}$ for Weibull failure time distribution (shape parameter $\theta = 1.2$) with $\lambda_1 = 0.6,\ \lambda_2 = 0.4$ and $p_1 = 0.9$, $p_2 = cp_1$ with known $c = 0.9$, but exponential fit ($c = 0.9$ is assumed known in the analysis).

		Cause 1				Cause 2			
n	τ_i	Bias$\times 10^2$	ASE	SSE	CP	Bias$\times 10^2$	ASE	SSE	CP
	0.171	3.148	0.029	0.052	0.684	2.327	0.027	0.037	0.833
	0.396	3.649	0.052	0.067	0.842	3.047	0.049	0.050	0.930
50	0.824	2.919	0.079	0.077	0.941	3.252	0.078	0.066	0.968
	1.468	0.677	0.093	0.092	0.958	2.452	0.096	0.085	0.956
	2.240	1.567	0.098	0.098	0.957	1.353	0.103	0.093	0.960
	0.171	3.073	0.016	0.029	0.492	2.382	0.015	0.021	0.629
	0.396	3.796	0.030	0.038	0.723	3.334	0.028	0.030	0.813
150	0.824	2.629	0.045	0.045	0.913	3.325	0.044	0.039	0.913
	1.468	0.006	0.053	0.056	0.932	2.341	0.055	0.050	0.937
	2.240	1.925	0.056	0.061	0.910	1.532	0.059	0.057	0.942
	0.171	3.248	0.012	0.023	0.319	2.551	0.011	0.016	0.411
	0.396	4.017	0.023	0.030	0.587	3.610	0.022	0.023	0.639
250	0.824	2.615	0.035	0.036	0.884	3.569	0.034	0.030	0.857
	1.468	0.323	0.041	0.043	0.937	2.498	0.042	0.038	0.924
	2.240	2.283	0.043	0.046	0.907	1.726	0.045	0.043	0.940

In another simulation study, we introduce uncertainity in p_1 and p_2 (See end of Section 3.3) by assuming a completely known $\nu(p_1, p_2)$. In particular, we consider simulation of data from exponential failure time model with $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$, as before, and two independent Beta distributions for p_1 and p_2 having means 0.9 and 0.8, respectively. The resulting estimates, averaged over 500 simulations, are similar to those presented in Table 1 and, hence, not reported.

6 Data Analysis

This section illustrates the proposed methodology using two real data sets. We first consider the much analyzed data set on onset of menopause with two types (See Krailo and Pike (1983)) and induce missing type at random to study the effect of different extent of missing data. Next we consider a data set on hearing loss due to two major types, collected from Ali Yavar Jung National Institute of Speech and Hearing Disabilities, Eastern Regional Center (See Banik et al. (2018)), while some observations are with missing information on the type of hearing loss.

6.1 Menopause Data

The dataset on onset of menopause due to two types, natural and operative, from the National Center for Health Statistics' Examination survey, was analyzed originally by MacMahon and Worcester (1966). Since this data does not have missing failure types, we modify this data by inducing some observations with missing type using different choices of the probabilities p_1 and p_2 as discussed in Section 2. At any monitoring time point x_i, there are d_{ji} number of individuals experiencing menopause of type $j = 1, 2$ (that is, natural or operative, respectively). We randomly impute the number of individuals failed with observed set of failure types g, for $g = \{1\}, \{2\}, \{1, 2\}$, out of these d_{ji} failures, by using a trinomial distribution, with cell probabilities $(p_{gj}, g = \{1\}, \{2\}, \{1, 2\})$. In particular, $p_{\{j\}j} = p_j$ and $p_{\{1,2\}j} = 1 - p_j$, for $j = 1, 2$. Once a data set with some missing failure types is generated, we analyze the same first by discarding the missing types, and then by retaining them (and using the methods of the present paper) to estimate the model parameters. This procedure is repeated 50 times and average of the parameter estimates along with the corresponding average standard errors are noted. For the sake of simplicity, we assume exponential menopause time distribution with time-independent cause-specific hazards λ_1 and λ_2 for natural and operative menopause, respectively. The results are presented in Table 8. In all the cases, comparing with the results of the original data analysis before imputation, we see that, when we ignore the missing data, the estimates of λ_1 and λ_2 are biased, as expected. On the other hand, when the missing failure types are included in the analysis, the estimates are similar to those obtained from the original data analysis, specially when the values of p_1 and p_2 used for the analysis matches with those used for imputation. Although the estimates seem to be mildly sensitive to such mismatch, these are evidently better than the ones obtained by ignoring the missing types.

6.2 Hearing Loss Data

Hearing loss is more common than most people realize that can develop at any age and can be caused by many different factors. It can be categorized into three general different types depending on which part of the ear is affected: (i) Sensorineural hearing loss (SNHL) is the most common type and occurs when there is a problem with the sensory and/or neural structures in the inner ear, (ii) Conductive hearing loss occurs when there are obstructions in the outer or middle ear. As a result of this, sound does not properly conduct through the outer ear to the middle ear, and, finally, (iii) Mixed hearing loss is a combination of sensorineural and conductive hearing loss. In addition to some irreversible hearing loss caused by a problem with the inner ear, there is also an issue with the outer or middle ear, which makes the hearing loss worse (See Deutsch and Richards (1979, p37)). The mixed hearing loss can be interpreted as following: one of the first two types (that is, SNHL and Conductive) appears first and the other one is acquired later, but before the monitoring time, so that the mixed type is observed at the monitoring time point, even though there is a possibility, albeit small, of co-existence of both the types occurring simultaneously. For the sake of illustration, we rule out the later possibility and, in our study, the event of interest is defined as the hearing loss due to the type occurring first. This is a competing risk setup with two competing risks being SNHL and Conductive. Thus, the mixed hearing loss can be interpreted as the missing type where we are not able to identify the type occurring first.

The data consists of 795 respondents, collected from the Department of Speech and Hearing, Ali Yavar Jung National Institute of Speech and Hearing Disabilities (Divyangjan), Eastern Regional Centre, who appeared for hearing loss diagnosis in the months of January and February in 2016. The monitoring time is taken to be the age at which (s)he appeared for diagnosis. We analyze this data by first fitting exponential distribution and then Weibull distribution for the event or failure time, as described above (See Section 5 for the description of the model parameters). Next we perform a goodness of fit test based on a modified χ^2 statistic which is defined as $\chi^2_M = \sum_{i=1}^{n} \sum_g \frac{(\delta_{gi} - p^*_{gi})^2}{p^*_{gi}(1-p^*_{gi})}$, with $p^*_{gi} = p_j F_j(x_i)$, if $g = \{j\}$, for $j = 1, 2$, $p^*_{gi} = (1 - p_1)F_1(x_i) + (1 - p_2)F_2(x_i)$, if $g = \{1, 2\}$ and $p^*_{gi} = S(x_i)$, if $g = \phi$, for $i = 1, \cdots, n$. The corresponding p-values are calculated using Monte Carlo simulation method, as described in the following. First we estimate the model parameters under the assumed parametric model and then calculate the modified χ^2 statistic based on the observed data. Then we simulate a large number, say $K = 500$, of datasets of the same size n from the assumed model with the estimated model parameters and compute the modified χ^2 statistic for each simulated dataset. This step gives an estimate of the null distribution of the modified χ^2 statistic. The p-value is estimated as the proportion of times (out of K) the value of the modified χ^2 statistic exceeds its observed value (See Hope, 1968). Finally, we have calculated the AIC (Akaike information criterion) and AICc (Akaike information criterion corrected) values (See Burnham and Anderson (2002)) for the two different models as given by

$$AIC = 2p - 2ln(\hat{L}), AICc = AIC + \frac{2p^2 + 2p}{n - p - 1},$$

where p is the number of estimated parameters in the model, n is the sample size and $\hat{L}$ is the maximum value of the likelihood function for the model. The results are presented in Table 9, for the exponential and Weibull fits. The p-values indicate that Weibull distribution fits the data which is also supported by the AIC and AICc values.

Table 8: Analysis of the Menopause Data with imputed missing types under exponential model*

p_1 and p_2 used for imputation	Type of Analysis	MLE $\times 10^2$ with standard errors$\times 10^2$		
		λ_1	λ_2	p_1^+
$p_1 = p_2 = 0.9$	Ignoring missing data	0.409 (0.022)	0.467 (0.024)	-
	$p_1 = p_2 = 0.9$ known	0.441 (0.023)	0.500 (0.025)	-
	$p_1 = 0.9, p_2 = 0.8$ known	0.428 (0.023)	0.523 (0.026)	-
	$p_2 = c\,p_1, c = 0.9$, known	0.419 (0.022)	0.533 (0.026)	0.959 (0.011)
	$p_2 = c\,p_1, c = 1$, known	0.445 (0.023)	0.506 (0.025)	0.902 (0.011)
$p_1 = p_2 = 0.8$	Ignoring missing data	0.370 (0.021)	0.427 (0.023)	-
	$p_1 = p_2 = 0.8$ known	0.443 (0.024)	0.505 (0.026)	-
	$p_1 = 0.9, p_2 = 0.8$ known	0.406 (0.023)	0.544 (0.027)	-
	$p_2 = c\,p_1, c = 0.9$, known	0.420 (0.023)	0.532 (0.027)	0.849 (0.016)
	$p_2 = c\,p_1, c = 1$, known	0.443 (0.024)	0.508 (0.026)	0.801 (0.015)
$p_1 = p_2 = 0.7$	Ignoring missing data	0.332 (0.019)	0.385 (0.021)	-
	$p_1 = p_2 = 0.7$ known	0.444 (0.025)	0.506 (0.027)	-
	$p_1 = 0.9, p_2 = 0.8$ known	0.385 (0.023)	0.566 (0.028)	-
	$p_2 = c\,p_1, c = 0.9$, known	0.420 (0.024)	0.528 (0.027)	0.745 (0.018)
	$p_2 = c\,p_1, c = 1$, known	0.443 (0.025)	0.506 (0.027)	0.703 (0.017)
$p_1 = 0.9, p_2 = 0.8$	Ignoring missing data	0.411 (0.022)	0.423 (0.022)	-
	$p_1 = 0.9, p_2 = 0.8$ known	0.443 (0.024)	0.506 (0.025)	-
	$p_1 = p_2 = 0.9$ known	0.490 (0.028)	0.481 (0.025)	-
	$p_2 = c\,p_1, c = 0.9$, known	0.446 (0.024)	0.506 (0.026)	0.894 (0.014)
	$p_2 = c\,p_1, c = 1$, known	0.470 (0.025)	0.482 (0.025)	0.846 (0.013)

*Analysis of original data gives the MLE$\times 10^2$, with standard error$\times 10^2$ in parantheses, of λ_1 and λ_2 as 0.444 (0.023) and 0.508 (0.025), respectively.

$^+$ MLE and standard error of p_1 are presented as they are.

Table 9: Analysis of the hearing loss data

Model	Type of Analysis	MLE×10^2 of parameters with standard error×10^2 in parantheses							
		λ_1	λ_2	θ^+	p_1^+	χ^2_M	p-value	AIC	AICc
Exponential	$p_1 = 0.9, p_2 = 0.8$,	8.78	7.24	-	-	5335.117	0	1429.465	1429.481
	known beforehand	(0.491)	(0.125)	-	-				
	$p_2 = cp_1$,	8.711	7.94	-	0.901	5317.1	0	1431.224	1431.255
	$c = 0.8$, known	(0.490)	(0.138)	-	(0.014)				
	$p_2 = cp_1$,	8.860	6.480	-	0.886	5321.706	0	1431.226	1431.256
	$c = 1$, known	(0.492)	(0.113)	-	(0.138)				
Weibull	$p_1 = 0.9, p_2 = 0.8$,	9.576	0.261	0.694	-	2289.601	0.81	1394.045	1394.045
	known beforehand	(0.739)	(0.062)	(0.043)	-				
	$p_2 = cp_1$,	9.467	0.320	0.697	0.907	2278.369	0.88	1396.029	1396.08
	$c = 0.8$, known	(0.734)	(0.087)	(0.046)	(0.014)				
	$p_2 = cp_1$,	9.713	0.246	0.701	0.886	2263.93	0.87	1395.928	1395.979
	$c = 1$, known	(0.742)	(0.056)	(0.042)	(0.014)				

$^+$ MLE and standard error of θ and p_1 are presented as they are.

7 Concluding Remarks

In this work, we consider the simple situation when the masking probabilities $p_{gj}(x,t)$, for $j = 1, 2$ and $g \ni j$, depend on j and g but are independent of both the failure and monitoring time points leading to two unknowns p_1 and p_2 in addition to the model parameters. Different cases related to the assumption on p_j's are discussed and analyzed in this work. It turns out that all the parameters are estimable if and only if p_1 and p_2 satisfy the relation

$p_2 = cp_1$, for some known constant c. We have also tried to analyze the data by introducing heterogeneity and uncertainity to the missing probabilities. Note that, for the case $p_2 = cp_1$ with $c = 1$ (that is, $p_1 = p_2$), we have missing at random since the masking probabilities depend on g, but not on j (See Section 2). The analysis by ignoring the missing causes, however, still gives biased estimates as the right censored observations with $\delta = 0$ become over-represented.

Our simulation studies indicate that the results are sensitive to misspecification of the value of c, although our analysis with the menopause data (See Section 6.1) seems to indicate mild sensitivity. The question regarding the choice of c, therefore, remains to be an important one. One can possibly try different choices and select one based on some criteria like AIC or AICc (See Section 6.2).

In this work, we have considered only two competing risks. A more general case considers $m > 2$ competing risks. The support $\mathcal{G}$ for the observed set of possible causes of failure, consists of all possible subsets of $\{1, \cdots, m\}$ and is of dimension $2^m - 1$. For a particular j, there are 2^{m-1} observed sets of possible failure causes containing the true cause j. Assuming the masking probabilities p_{gj}'s to be independent of both the monitoring and failure time points, a particular choice can be taken as

$$p_{gj} = \begin{cases} \alpha, & \text{if } g = \{j\} \\ \frac{1-\alpha}{2^{m-1}-1}, & \text{if } g \neq \{j\}, \end{cases}$$

so that $\sum_{g \ni j} p_{gj} = 1$, for $j = 1, \cdots, m$. Different values of α represent different degrees of missingness.

ACKNOWLEDGEMENT

We like to thank the Ali Yavar Jung National Institute of Speech and Hearing Disabilities (Divyangjan), Eastern Regional Centre, for providing us with the hearing loss data for our analysis.

References

Banik, P., Jaiswal, J., Koley, T., and Chatterjee., I. (2018). *Explanatory analysis of tertiary audiological data.* To be submitted.

Basu, S. (2009). Inference about the masking probabilities in the competing risks model. *Communications in Statistics - Theory and Methods*, 38:2677–2690.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer-Verlag.

Deutsch, L. J. and Richards, A. M. (1979). *Elementary Hearing Science.* University Park Press.

Dewanji, A., Krewski, D., and Goddard, M. J. (1993). A weibull model for the estimation of tumorigenic potency. *Biometrics*, 49(2):367–377.

Dewanji, A. and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics*, 59:1063–1070.

Flehtnjer, B. J., Reiser, B., and Yashchin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika*, 85(1):151–164.

Goetghebeur, E. and Ryan, L. (1990). A modified log rank test for competing risks with missing failure type. *Biometrika*, 77(1):207–211.

Hope, A. C. A. (1968). A simplified monte carlo significance test procedure. *Journal of Royal Statistical Society, Series B*, 30.

Hudgens, M., Satten, G., and Longini I.M., J. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*, 57:74–80.

Jewell, N. and Kalbfleisch, J. (2004). Maximum likelihood estimation of ordered multinomial parameters. *Biostatistics*, 5:291–306.

Jewell, N., Van Der Laan, M., and Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika*, 90:183–197.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data, 2nd edition.* New York:Wiley.

Krailo, M. and Pike, M. C. (1983). Estimation of the distribution of age at natural menopause from prevalence data. *American Journal of Epidemiology*, 117:356–361.

Lehmann, E. and Casella, G. (1998). *Theory of Point Estimation, Second edition.* Springer.

Little, A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Lo, S. (1991). Estimating a survival function with incomplete cause-of-death data. 39(2):217–235.

Maathuis, M. H. (2006). *Nonparametric estimation for current status data with competing risks.* Diss. University of Washington.

MacMahon, B. and Worcester, J. (1966). Age at menopause. united states–1960-1962. *National Center for Health Statistics; Vital and Health Statistics, Series 11: Data from the National Health Survey, no. 19,*, Washington, DC: DHEW Publication no. (HSM):66–1000.

Mukerjee, H. and Wang, J.-L. (1993). Nonparametric maximum likelihood estimation of an increasing hazard rate for uncertain cause-of-death data. *Scandinavian Journal of Statistics*, 20(1):17–33.