Use of Additional Information for Current Status Data with Two Competing Risks and Missing Failure Types

Technical Report No. ASU/2020/3
Dated 05 March, 2020

Tamalika Koley
and
Anup Dewanji

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Use of Additional Information for Current Status Data with Two Competing Risks and Missing Failure Types

Tamalika Koley
email: tamalikakoley@gmail.com

Anup Dewanji
email: dewanjia@isical.ac.in

Abstract

In practice, the cause of failure for some subjects may be missing or uncertain in competing risks data. This type of uncertainty on the failure type arises mainly in tumorigenicity data where it becomes difficult to determine the true cause of death in presence of tumor. Both parametric and non-parametric analyses of such uncertain failure type in current status data with two competing risks require some model assumptions to deal with the identifiability issues of the model parameters. In this work, we attempt to alleviate this identifiability problem using some additional data and without any such model assumption. In particular, we first consider additional information from a validation sample which ascertain the uncertain failure type. Next, we consider additional information in the form of some prior knowledge on the missing probabilities. We consider parametric maximum likelihood estimation of the model parameters and non-parametric maximum likelihood estimation of the sub-distribution function. We investigate the associated large sample properties theoretically and the finite sample properties through simulation. We also consider analysis of some real data set for the purpose of illustration.

Keywords— Monitoring time, Masking probability, Validation sample, Identifiability, Maximum likelihood estimation, Bootstrap.

1 Introduction

In an extreme form of interval censoring, known as current status data, only the information on the survival status of the individuals is available at a single observation time point (known as monitoring time point). The status here refers to the occurrence/non-occurrence of the event of interest usually termed as failure and survival, respectively. The individuals experiencing the event are exposed to the risk of more than one causes called the competing risks (See Kalbfleisch and Prentice (2002)). For example, a breast cancer patient may die due to causes unrelated to the disease, termed as competing risk events. Both parametric and nonparametric maximum likelihood estimation of sub-distribution functions based on current status data with competing risks have been extensively covered in the literature. See, for example, Hudgens et al. (2001), Jewell et al. (2003), Jewell and Kalbfleisch (2004), Maathuis (2006) and many others.

In most applications, the cause of failure is assumed to be known. However, in practice, the cause of failure for some subjects may be missing or uncertain. This type of uncertainty in the information on failure type arises, for example, in tumorigenicity data where it becomes difficult to determine the true cause of death in presence of tumor. Previously, many authors have worked with right censored survival data with two failure types and the information on failure types is either exactly available or completely missing (See, for example, Miyakawa (1984); Racine-Poon and Hoel (1984); Lo (1991); Mukerjee and Wang (1993)). Goetghebeur and Ryan (1990, 1995), Dewanji (1992) and Lu and Tsiatis (2001)have considered the regression problem in this context. However, in all the above studies, missingness meant complete missing of the information on failure causes. But sometimes it is quite possible that the experimentalists can determine a subset of all causes containing the true cause of failure. Flehinger et al. (1998) have studied such general pattern of missing failure types with the assumption of proportional hazards due to different types. On the otherhand, Dewanji and Sengupta (2003) considered the same estimation problem but without any proportional hazards assumption. They have developed an EM algorithm for non parametric maximum likelihood estimation of survival functions. They have also proposed Nelson-Aalen type estimator based on the information on the probability that one of them is responsible.

Not much work related to the analysis of missing failure types in the context of current status data with competing risks is present in the literature. Koley and Dewanji (2018a) and Koley and Dewanji (2019) considered this problem in presence of two competing risks and time independent missing probabilities and carried out parametric and non-parametric methods of estimation, respectively. However, there were some identifiability issues of model parameters and the likelihood function in the parametric and non-parametric analyses, respectively. Some additional modeling assumption on the probabilities of missing failure type has been required to achieve identifiability. On the other hand, additional information may sometime be available which can be judiciously used to alleviate the identifiability problem without requiring to make additional modeling assumptions. For example, information on the missing cause may be available on some individuals in a validation sample through some stringent diagnostic procedure. Alternatively, one may have some direct or indirect information on the missing probabilities. In this work, we consider these two cases of additional information separately with two competing risks. We first develop maximum likelihood methods to estimate the model parameters in a parametric set-up. Then we consider similar situations in a non-parametric set-up and carry out non-parametric maximum likelihood estimation of the sub-distribution functions. Throughout this work, the missing probabilities are assumed to be time independent. We also assume that there is no misspecification in the data related to the failure types. In Section 2, the data is described along with the corresponding likelihood function. Section 3 considers parametric estimation in presence of the two different cases of additional information as discussed before, while Section 4 discusses non-parametric maximum likelihood estimation of the two sub-distribution functions. Section 5 presents some simulation studies to investigate the finite sample properties of the proposed estimators. We illustrate the estimation procedure of the second case of additional information only with a hearing loss data in Section 6, since there is no validation sample in this data. Finally, Section 7 ends with some concluding remarks.

2 The Data and the Likelihood

Let T be the random variable representing the failure time of event subject to $m = 2$ competing risks and $J \in \{1, 2\}$ denote the true cause of failure. Also, let X denote the monitoring time which is assumed to be independent of the random vector (T, J). In current status data with competing risks, an individual is observed at the monitoring time and the observation consists of the monitoring time $X = x$, $\delta = 1$ or 0 indicating whether the event has occurred or not along with $J = j$, the cause of failure whenever $\delta = 1$. As discussed in Section 1, there may be some individuals with missing information on the cause of failure. This problem of missing information is defined by the random variable G which represents the observed set of possible causes of failure. Therefore, instead of J, we observe G and $G \in \mathcal{G} = \{\{1\}, \{2\}, \{1, 2\}\}$, the collection of observed sets of possible causes. When G is a singleton set, $G = \{J\}$ (that is, there is no missingness), whereas $G = \{1, 2\}$ represents complete missing. As in Koley and Dewanji (2018a) and Koley and Dewanji (2019), we make the assumption of no misspecification in the observed set of possible causes in the sense that G always contains the true cause. The sub-distribution function for failure due to cause j is given by $F_j(t) = P[T \leq X = t, J = j] = \int_0^t f_j(u)\mathrm{d}u$, where $f_j(\cdot)$ is the corresponding sub-density function. This gives the overall distribution function as $F(t) = \sum_{j=1}^{2} F_j(t)$ and $S(t) = 1 - F(t)$ as the survivor function.

For a particular individual if $T \leq X$, we observe the set G and, if $T > X$, we denote G to be the empty set, ϕ. So the support of G is given by $\{\phi, \{1\}, \{2\}, \{1,2\}\}$. Thus our data is of the form (x_i, g_i), which is a realization of the random vector (X, G) and the suffix i represents the observation from the ith individual, for $i = 1, \cdots, n$. Let us define the indicator variable $\delta_{gi} = \mathbb{I}\left[T \leq X = x_i, G = g\right]$, for $i = 1, \cdots, n$, and $g \in \mathcal{G} \setminus \phi$. Also, define the conditional probability of observing g as the set of possible causes of failure, given that the true cause $J = j$ and all other related information, as

$$p_{gj}(x,t) = P\left[G = g \mid T = t \leq X = x, \delta = 1, J = j\right],$$

if $g \ni j$ and 0, otherwise. This gives $\sum_{g \ni j} p_{gj}(x,t) = 1$, for fixed j with $t \leq x$. These probabilities are termed as masking probabilities since the true cause is masked in the observed set of failure. Throughout this work we have assumed these masking probabilities to be time independent (both monitoring and failure time) but dependent on the true cause j, for $j = 1, 2$. This type of missing pattern is non-ignorable. Let us write

$$p_j = p_{\{j\}j}, \text{ for } j = 1, 2.$$

Thus, for a fixed j, $p_{\{1,2\}j} = 1 - p_j$. Noting that the failure time distribution with two competing risks is of primary interest, the likelihood contributions for the data are of the form $P\left[T \leq x, G = g\right] = \sum_{j \in g} p_{\{g\}j} F_j(x)$ and $P\left[T > X = x\right] = S(x)$, for $x \in$ DomX and $g \in \mathcal{G}$. Hence, the likelihood of the data is

$$L(p_1, p_2) \propto \prod_{i=1}^{n} \Big[\{p_1 F_1(x_i)\}^{\delta_{\{1\}i}} \{p_2 F_2(x_i)\}^{\delta_{\{2\}i}} \times \{(1-p_1)F_1(x_i) + (1-p_2)F_2(x_i)\}^{\delta_{\{1,2\}i}} S(x_i)^{1-\delta_i} \Big]. \tag{2.1}$$

3 Parametric Estimation

In parametric estimation, we make some assumption on failure time distribution and then the model parameters are estimated. Suppose $\underset{\sim}{\theta}$ is the parameter vector of interest corresponding to the failure time distribution and the parameters associated with the monitoring time distribution are not of interest. The likelihood function (2.1) is then given by

$$L(\underset{\sim}{\theta}, p_1, p_2) \propto \prod_{i=1}^{n} \Bigg[\Big\{ p_1 F_1(x_i; \underset{\sim}{\theta}) \Big\}^{\delta_{\{1\}i}} \Big\{ p_2 F_2(x_i; \underset{\sim}{\theta}) \Big\}^{\delta_{\{2\}i}} \times \Big\{ (1-p_1) F_1(x_i; \underset{\sim}{\theta}) + (1-p_2) F_2(x_i; \underset{\sim}{\theta}) \Big\}^{\delta_{\{1,2\}i}} S(x_i; \underset{\sim}{\theta})^{1-\delta_i} \Bigg]. \tag{3.1}$$

The following two subsections discuss two cases of additional information separately and develop maximum likelihood methods to estimate the model parameters.

3.1 Estimation via Validation Sample Approach

Let us write the set of observations with missing cause as $V = \big\{ i \colon \delta_{\{1,2\}i} = 1 \big\}$. We draw a random validation sample S_V of size n_V from V with $n_V \leq |V|$, where $|V|$ denotes the cardinality of the set V. After validation, we can observe if the true cause of the failure for an individual is either 1 or 2. The likelihood L_V of this validation subsample thus consists of the following two probability terms:

$$\begin{aligned} P\left[J=1 \mid G=\{1,2\}, T \leq x\right] &= \frac{P\left[J=1, G=\{1,2\}, T \leq x\right]}{P\left[G=\{1,2\}, T \leq x\right]} \\ &= \frac{(1-p_1) F_1(x; \underset{\sim}{\theta})}{(1-p_1) F_1(x; \underset{\sim}{\theta}) + (1-p_2) F_2(x; \underset{\sim}{\theta})} \end{aligned}$$

and

$$\begin{aligned} P\left[J=2 \mid G=\{1,2\}, T \leq x\right] &= \frac{P\left[J=2, G=\{1,2\}, T \leq x\right]}{P\left[G=\{1,2\}, T \leq x\right]} \\ &= \frac{(1-p_2) F_2(x; \underset{\sim}{\theta})}{(1-p_1) F_1(x; \underset{\sim}{\theta}) + (1-p_2) F_2(x; \underset{\sim}{\theta})}. \end{aligned}$$

Define the indicator variable $\delta_i^* = \mathbb{I}\left[J=1, i \in S_V, T \leq X = x_i, G = g_i = \{1,2\}\right]$, indicating if the ith individual is in the validation sample with the true failure type being 1 or not, for $i = 1, \cdots, n$. Thus, the validation sample likelihood is given by

$$L_V \propto \prod_{i \in S_V} \left[\left\{ \frac{(1-p_1)F_1(x_i; \underset{\sim}{\theta})}{(1-p_1)F_1(x_i; \underset{\sim}{\theta}) + (1-p_2)F_2(x_i; \underset{\sim}{\theta})} \right\}^{\delta_i^*} \right.$$

$$\left. \times \left\{ \frac{(1-p_2)F_2(x_i; \underset{\sim}{\theta})}{(1-p_1)F_1(x_i; \underset{\sim}{\theta}) + (1-p_2)F_2(x_i; \underset{\sim}{\theta})} \right\}^{(1-\delta_i^*)} \right].$$

Therefore, the total likelihood can be written as $L = L(\underset{\sim}{\theta}, p_1, p_2) \times L_V$, which is

$$\prod_{i=1}^{n} \left[\left\{ p_1 F_1(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{1\}i}} \left\{ p_2 F_2(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{2\}i}} \left\{ (1-p_1)F_1(x_i; \underset{\sim}{\theta}) + (1-p_2)F_2(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{1,2\}i}} \right.$$

$$\left. \times S(x_i; \underset{\sim}{\theta})^{1-\delta_i} \right] \times \prod_{v \in S_V} \left[\left\{ \frac{(1-p_1)F_1(x_v; \underset{\sim}{\theta})}{(1-p_1)F_1(x_v; \underset{\sim}{\theta}) + (1-p_2)F_2(x_v; \underset{\sim}{\theta})} \right\}^{\delta_v^*} \right.$$

$$\left. \times \left\{ \frac{(1-p_2)F_2(x_v; \underset{\sim}{\theta})}{(1-p_1)F_1(x_v; \underset{\sim}{\theta}) + (1-p_2)F_2(x_v; \underset{\sim}{\theta})} \right\}^{(1-\delta_v^*)} \right].$$

Finally, define the indicator variable $\delta_i^{**} = \mathbb{I}\left[T \leq X = x_i, g_i = \{1, 2\}, i \in S_V\right]$, indicating if the i^{th} individual in the sample is in the validation sample or not, for $i = 1, \cdots, n$. Then, the total likelihood simplifies to

$$L \propto \prod_{i=1}^{n} \left[\left\{ p_1 F_1(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{1\}i}} \left\{ p_2 F_2(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{2\}i}} \left\{ (1-p_1)F_1(x_i; \underset{\sim}{\theta}) + (1-p_2)F_2(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{1,2\}i}(1-\delta_i^{**})} \right.$$

$$\left. \times \left\{ (1-p_1)F_1(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{1,2\}i}\delta_i^{**}\delta_i^*} \left\{ (1-p_2)F_2(x_i; \underset{\sim}{\theta}) \right\}^{\delta_{\{1,2\}i}\delta_i^{**}(1-\delta_i^*)} S(x_i; \underset{\sim}{\theta})^{1-\delta_i} \right]. \tag{3.2}$$

Numerical maximization of this log-likelihood function gives the MLEs $(\hat{\underset{\sim}{\theta}}, \hat{p}_1, \hat{p}_2)$ of the model parameters $(\underset{\sim}{\theta}, p_1, p_2)$. For this purpose, we use the R function *optim()*. We now attempt to establish the asymptotic properties of the MLEs by writing the likelihood function L as a product of n densities. Let us write $\underset{\sim}{\eta} = (\underset{\sim}{\theta}, p_1, p_2)$ to represent the vector of model parameters with $0 < p_1, p_2 < 1$. Define the variable $\epsilon \in \{1, \cdots, 6\}$ in the following way:

$$\epsilon = \begin{cases} 1, & \text{if } T \leq X = x, g = \{1\}. \\ 2, & \text{if } T \leq X = x, g = \{2\}. \\ 3, & \text{if } T \leq X = x, g = \{1,2\}\text{, the sample is not included in the validation sample.} \\ 4, & \text{if } T \leq X = x, g = \{1,2\}\text{, the sample is included in the validation sample.} \\ & \qquad\qquad\qquad\qquad\qquad \text{with true cause 1.} \\ 5, & \text{if } T \leq X = x, g = \{1,2\}\text{, the sample is included in the validation sample} \\ & \qquad\qquad\qquad\qquad\qquad \text{with true cause 2.} \\ 6, & \text{if } T > X = x. \end{cases}$$

Thus, the combined data of the original sample and the validation sample can be represented in terms of the observations on (x, ϵ) from n individuals. Also, define π to be the conditional probability of inclusion in the validation sample. In other words, given an individual observed to have failed with $g = \{1, 2\}$ as the observed set of possible causes, the individual will be sampled for validation with probability π, depending on $|V|$ and n_V. This probability is assumed to be same for all individuals of this type, which is a valid assumption for simple random sampling, for example. The density function of the random vector $(X, \mathcal{E})$, corresponding to the observation (x, ϵ), with respect to the dominating measure $H \times \mu$, where H is the distribution function of X and μ, the counting measure, is given by

$$f^*(x, \epsilon; \underset{\sim}{\eta}) = \begin{cases} p_1 F_1(x; \underset{\sim}{\theta}) h(x), & \text{if } \epsilon = 1 \\ p_2 F_2(x; \underset{\sim}{\theta}) h(x), & \text{if } \epsilon = 2 \\ \left\{ (1 - p_1) F_1(x; \underset{\sim}{\theta}) + (1 - p_2) F_2(x; \underset{\sim}{\theta}) \right\} h(x)(1 - \pi), & \text{if } \epsilon = 3 \\ (1 - p_1) F_1(x; \underset{\sim}{\theta}) h(x) \pi, & \text{if } \epsilon = 4 \\ (1 - p_2) F_2(x; \underset{\sim}{\theta}) h(x) \pi, & \text{if } \epsilon = 5 \\ S(x; \underset{\sim}{\theta}) h(x), & \text{if } \epsilon = 6. \end{cases}$$

Here, $h(\cdot)$ is the density function of the monitoring time random variable X. Also, since $\underset{\sim}{\eta}$ is the parameter vector of interest, we keep the dependence of f^* on only $\underset{\sim}{\eta}$ for notational simplicity. It is easy to check that this density integrates to 1 through the following claim.

Claim 3.1.

$$\int_0^\infty \sum_{\epsilon=1}^{6} f^*(x, \epsilon; \underset{\sim}{\eta}) dx = 1$$

Proof.

$$\begin{aligned}\int_0^\infty \sum_{\epsilon=1}^{6} f^*(x,\epsilon;\underset{\sim}{\eta})dx &= \int_0^\infty p_1F_1(x;\underset{\sim}{\theta})h(x)dx + \int_0^\infty p_2F_2(x;\underset{\sim}{\theta})h(x)dx \\ &+ \int_0^\infty \left\{(1-p_1)F_1(x;\underset{\sim}{\theta}) + (1-p_2)F_2(x;\underset{\sim}{\theta})\right\} h(x)(1-\pi)dx \\ &+ \int_0^\infty (1-p_1)F_1(x;\underset{\sim}{\theta})h(x)\pi dx + \int_0^\infty (1-p_2)F_2(x;\underset{\sim}{\theta})h(x)\pi dx \\ &+ \int_0^\infty S(x;\underset{\sim}{\theta})h(x)dx = \int_0^\infty h(x)dx = 1.\end{aligned}$$

□

Clearly, the likelihood function L of (3.2) is proportional to the product of these density functions; that is,

$$L \propto \prod_{i=1}^{n} f^*(x_i,\epsilon_i;\underset{\sim}{\eta}),$$

where ϵ_i is the observed value of ϵ for the ith individual. The terms involving π and the set of parameters associated with the monitoring time factor out. In this work, we are interested in the estimation of $\underset{\sim}{\eta}$ consisting of the parameters associated with the failure time distribution (that is, $\underset{\sim}{\theta}$) and the probabilities p_1, p_2. Let us assume that the failure time distribution satisfies the standard regularity conditions (See Lehmann and Casella, 1998, p449). In addition, consider the following conditions.

R1 The set of model parameters $\underset{\sim}{\eta} = (\underset{\sim}{\theta}, p_1, p_2)$ is identifiable with respect to the density function $f^*(\cdot,\cdot;\underset{\sim}{\eta})$. This is clear from the form of $f^*(\cdot,\cdot;\underset{\sim}{\eta})$.

R2 The set of model parameters $\underset{\sim}{\eta} = (\underset{\sim}{\theta}, p_1, p_2)$ is open since $\underset{\sim}{\theta}$ is open and $0 < p_1, p_2 < 1$.

R3 The sample space of the random vector $(X, \mathcal{E})$ is independent of the model parameters.

R4 The failure time distribution satisfies the standard regularity conditions so that the partial derivatives $\frac{\partial^2}{\partial\theta_k\partial\theta_l} f_j(x;\underset{\sim}{\theta})$, $\frac{\partial^3}{\partial\theta_k\partial\theta_l\partial\theta_m} f_j(x;\underset{\sim}{\theta})$, 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 $f^*(\cdot,\cdot;\underset{\sim}{\eta})$ also exist and are continuous.

R5 The Fisher information matrix is assumed to be positive definite.

Theorem 3.1. *The MLE of the parameter vector $\underset{\sim}{\eta}$, denoted by $\underset{\sim}{\hat{\eta}}$, with $\underset{\sim}{\eta_0}$ being the true value, satisfies*

1. $\underset{\sim}{\hat{\eta}} \xrightarrow{P} \underset{\sim}{\eta_0}$,

2. $\sqrt{n}(\underset{\sim}{\hat{\eta}} - \underset{\sim}{\eta_0})$ *is asymptotically a mean zero normal vector with variance-covariance matrix estimated by the inverse of the hessian matrix computed at the MLE.*

Proof. With the assumption of failure time distribution satisfying the regularity conditions and the conditions **R1-R5**, proof of the theorem directly follows from Theorem 7.5.2 of Lehmann and Casella (1998, p463-465). □

3.2 Indirect Information on the Masking Probabilities

In this subsection, we consider the case when some specific type of indirect information is available on the masking probabilities $1 - p_1$ and $1 - p_2$. Suppose we have additional knowledge on the conditional probabilities, given a failure and $G = \{1, 2\}$, of the failure type, denoted by $q_j = P\left[J = j \mid \delta = 1, G = \{1, 2\}\right]$, for $j = 1, 2$, with $q_1 + q_2 = 1$, where δ represents the indicator random variable indicating failure or not. Using Bayes' rule, we then get

$$\begin{aligned} 1 - p_j &= P\left[G = \{1, 2\} \mid \delta = 1, J = j\right] \\ &= \frac{q_j P\left[\delta = 1, G = \{1, 2\}\right]}{q_j P\left[\delta = 1, G = \{1, 2\}\right] + P\left[\delta = 1, G = \{j\}\right]}, \text{ for } \quad j = 1, 2, \end{aligned} \tag{3.3}$$

as $P[J = j | \delta = 1, G = \{j\}] = 1$ under the assumption of no misspecification, as discussed in Section 2. In the following, we exploit a simple multinomial structure for the frequency of observations with four possible values of G and then the invariance property to obtain an estimate of p_j, denoted by $\hat{p}_j$, for $j = 1, 2$. Use of the delta method thereafter allows the study of asymptotic results for these estimates $\hat{p}_j$, for $j = 1, 2$.

Let us write the random indicator variable $\delta_g = \mathbb{I}[T \leq X, G = g]$, indicating if the possible set of causes for failure is g or not, for $g \in \mathcal{G}$. Recall the definition of δ_{gi}, for $g \in \mathcal{G}$, in Section 2, which turns out to be the observed value of the random variable δ_g for the ith individual. Then, clearly, the random vector $\underset{\sim}{Y} = (\delta_{\{1\}}, \delta_{\{2\}}, \delta_{\{1,2\}}, \delta_\phi)'$ is a multinomial distribution with sample size 1 and the four cell probabilities $(r_{\{1\}}, r_{\{2\}}, r_{\{1,2\}}, r_\phi)'$, where

$$\begin{aligned} r_g &= P\left[T \leq X, G = g\right] \\ &= \int_0^\infty \Big\{ \sum_{j \in g} p_{gj} F_j(x; \underset{\sim}{\theta}) \Big\} h(x) \mathrm{d}x, \end{aligned}$$

for $g \in \mathcal{G} \setminus \phi$ and $r_\phi = P\left[T > X\right] = \displaystyle\int_0^\infty S(x; \underset{\sim}{\theta}) h(x) \mathrm{d}x$. Note that the data consists of n iid realizations $\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_n$ from the random vector $\underset{\sim}{Y}$ with mean vector

$$\underset{\sim}{r} = \begin{pmatrix} r_{\{1\}} \\ r_{\{2\}} \\ r_{\{1,2\}} \\ r_{\phi} \end{pmatrix} \quad \text{and the variance-covariance matrix,}$$

$$\Sigma^* = \begin{bmatrix} r_{\{1\}}(1-r_{\{1\}}) & r_{\{1\}}r_{\{2\}} & r_{\{1\}}r_{\{1,2\}} & r_{\{1\}}r_{\phi} \\ r_{\{2\}}r_{\{1\}} & r_{\{2\}}(1-r_{\{2\}}) & r_{\{2\}}r_{\{1,2\}} & r_{\{2\}}r_{\phi} \\ r_{\{1,2\}}r_{\{1\}} & r_{\{1,2\}}r_{\{2\}} & r_{\{1,2\}}(1-r_{\{1,2\}}) & r_{\{1,2\}}r_{\phi} \\ r_{\phi}r_{\{1\}} & r_{\phi}r_{\{2\}} & r_{\phi}r_{\{1,2\}} & r_{\phi}(1-r_{\phi}) \end{bmatrix}.$$

Define

$$\overline{\underset{\sim}{Y}} = \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{Y_i} = \frac{1}{n}\begin{pmatrix} f_{\{1\}} \\ f_{\{2\}} \\ f_{\{1,2\}} \\ f_{\phi} \end{pmatrix} = \underset{\sim}{f}, \quad \text{say,}$$

where $\underset{\sim}{Y_i}$ denotes the random variable representing the $\underset{\sim}{Y}$ vector for the i^{th} individual and $f_g = \sum_{i=1}^{n} \mathbb{I}\left[g_i = g\right]$ represents the total number of individuals with g as the observed set of possible causes in the sample of size n, for $g \in \mathcal{G}$. Following the multivariate generalization of the central limit theorem (Trotter, 1959), we have $\sqrt{n}(\underset{\sim}{f} - \underset{\sim}{r})$ converges in distribution to a multivariate normal random vector with mean vector $\underset{\sim}{0}$ and the variance-covariance matrix Σ^*.

Note that the MLE of the cell probability r_g is given by $\hat{r}_g = \dfrac{f_g}{n}$, for $g \in \mathcal{G}$, using the property of multinomial distribution. Then the MLE of $\underset{\sim}{r}$, denoted by $\underset{\sim}{\hat{r}}$ is given by $\underset{\sim}{f}$. Now, for a given j, using (3.3), the masking probability p_j is a function of the vector $\underset{\sim}{r}$, that is, $p_j = b_j(\underset{\sim}{r})$, say, since the event $\{\delta = 1\}$ is same as the event $\{T \leq X\}$. Using the Invariance property of MLE, we have $\hat{p}_j = b_j(\underset{\sim}{\hat{r}}) = b_j(\underset{\sim}{f})$, for $j = 1, 2$. Clearly, the function $b_j(\underset{\sim}{r})$ admits continuous partial derivatives with respect to $\underset{\sim}{r}$. Then, following the multivariate delta method (Cox (2005)) and defining $\underset{\sim}{b}(\cdot) = (b_1(\cdot), b_2(\cdot))'$, we have

$$\sqrt{n}(\underset{\sim}{b}(\underset{\sim}{f}) - \underset{\sim}{b}(\underset{\sim}{r})) \xrightarrow{d} N\left(\underset{\sim}{0}, J_b(\underset{\sim}{r})\Sigma^* J_b(\underset{\sim}{r})^T\right),$$

where $J_b(\underset{\sim}{r})$ is the 2×4 matrix of the partial derivatives of $\underset{\sim}{b}(\underset{\sim}{r})$ with respect to $\underset{\sim}{r}$. Note that this asymptotic variance can be consistently estimated by evaluating it at the MLE $\underset{\sim}{f}$ of $\underset{\sim}{r}$.

Claim 3.2. *The masking probability p_j can be consistently estimated by $\hat{p}_j$, for $j = 1, 2$.*

Proof. For a given j, suppose p_j^0 denote the true value of the masking probability p_j. Note that

$$\hat{p}_j = b_j(\underset{\sim}{\hat{r}}) = \frac{f_{\{j\}}}{f_{\{j\}} + q_j f_{\{1,2\}}}$$

$$\xrightarrow{P} \frac{P\left[G = \{j\}, \delta = 1\right]}{P\left[G = \{j\}, \delta = 1\right] + q_j P\left[G = \{1, 2\}, \delta = 1\right]}$$

$$= \frac{P\left[G = \{j\}, \delta = 1\right]}{P\left[G = \{j\}, \delta = 1\right] + P\left[J = 1 \mid G = \{1, 2\}, \delta = 1\right] P\left[G = \{1, 2\}, \delta = 1\right]}$$

$$= P\left[G = \{j\} \mid J = 1, \delta = 1\right] = p_j^0.$$

Note that $P\left[\delta = 1, G = \{j\}\right] = P\left[\delta = 1, G = \{j\}, J = j\right]$, since there is no misspecification of failure types, as discussed before. This holds for both $j = 1$ and 2. □

Hence, the p_j's can be estimated consistently from (3.3) using information on the q_j's. We suggest treating these estimated p_j's as known in order to estimate $\underset{\sim}{\theta}$ by maximizing the likelihood function (3.1). Note that, since the unknown p_j's are estimated first and then treated as known, the resulting estimate $\underset{\sim}{\widetilde{\theta}}$ of $\underset{\sim}{\theta}$ is not the usual MLE. Therefore, the observed information matrix may not be proper for the purpose of obtaining the corresponding standard errors. For this purpose, we suggest using the bootstrap procedure as described below.

From the observed sample $\{(x_i, g_i),\ i = 1, \cdots, n\}$, a bootstrap sample of size n is drawn by simple random sampling with replacement. Denote this bootstrap sample by $\{(x_i^l, g_i^l),\ i = 1, \cdots, n\}$. Obtain a bootstrap estimate $\underset{\sim}{\widetilde{\theta}}^l$ of $\underset{\sim}{\theta}$ by the above method but based on the bootstrap sample. This is repeated a large number B, say 1000, of times to obtain B number of bootstrap estimates of $\underset{\sim}{\theta}$. The standard deviation of these B bootstrap estimates gives the standard error of the estimate $\underset{\sim}{\widetilde{\theta}}$. In general, a histogram of these bootstrap estimates gives an estimate of the sampling distribution of $\underset{\sim}{\widetilde{\theta}}$. Note that this bootstrap procedure also gives the standard errors of the estimates of p_1 and p_2.

Suppose the dimension of the vector $\underset{\sim}{\theta}$ is $s \times 1$. Let $\nabla\widetilde{\theta}_l(\cdot, \cdot)$ represents the vector of partial derivatives of $\widetilde{\theta}_l(p_1, p_2)$ with repect to p_1 and p_2 and is assumed to be stochastically bounded in a neighbourhood of the true value (p_1^0, p_2^0), that is, $O_p(1)$, for $l = 1, \cdots, s$. This assumption is numerically validated in a limited way, as demonstrated in the following table.

Table 1: Numerical calculation of the vector of partial derivatives $\nabla\widetilde{\theta}_l(\cdot,\cdot)$ at different values of (p_1, p_2) with increasing sample size, n.

n	(p_1, p_2)	$\frac{\partial\widetilde{\lambda}_1(p_1,p_2)}{\partial p_1}$	$\frac{\partial\widetilde{\lambda}_1(p_1,p_2)}{\partial p_2}$	$\frac{\partial\widetilde{\lambda}_2(p_1,p_2)}{\partial p_1}$	$\frac{\partial\widetilde{\lambda}_2(p_1,p_2)}{\partial p_2}$
50	(0.761,0.635)	0.418	0.273	0.418	0.273
150		0.395	0.219	0.395	0.219
250		0.402	0.225	0.402	0.225
50	(0.763,0.637)	0.419	0.274	0.419	0.274
150		0.400	0.220	0.400	0.220
250		0.406	0.226	0.406	0.226
50	(0.761,0.637)	0.418	0.274	0.418	0.278
150		0.396	0.220	0.396	0.220
250		0.403	0.226	0.403	0.226
50	(0.763,0.635)	0.419	0.273	0.419	0.273
150		0.399	0.218	0.399	0.218
250		0.405	0.224	0.405	0.224

Theorem 3.2. *The approximate MLE* $\widetilde{\underset{\sim}{\theta}} = \widetilde{\underset{\sim}{\theta}}(\hat{p}_1, \hat{p}_2)$*, say, is a consistent estimator of* $\underset{\sim}{\theta}$ *under the condition that the partial derivatives* $\widetilde{\underset{\sim}{\theta}}_l(\cdot,\cdot)$ *is* $O_p(1)$*, for* $l = 1, \cdots, s$.

Proof. Suppose $\underset{\sim}{\theta}_0 = (\theta_l^0; l = 1, \cdots, s)$ is the true values of the parameter vector $\underset{\sim}{\theta}$. Then, for a given l, using Taylor's Theorem (Malik and Arora, 1992, p544), we have

$$\widetilde{\theta}_l(\hat{p}_1, \hat{p}_2) - \theta_l^0 = \widetilde{\theta}_l(p_1^0, p_2^0) - \theta_l^0 + \nabla\widetilde{\theta}_l(\xi_1, \xi_2) \cdot (\hat{p}_1 - p_1^0, \hat{p}_2 - p_2^0)', \tag{3.4}$$

for some (ξ_1, ξ_2) which is a convex combination of the two points $(\hat{p}_1, \hat{p}_2)$ and (p_1^0, p_2^0), and $\nabla\widetilde{\theta}_l(\xi_1, \xi_2)$ is the gradient vector of $\widetilde{\theta}_l(p_1, p_2)$ computed at (ξ_1, ξ_2). When p_1 and p_2 are known, $\widetilde{\theta}_l(p_1, p_2)$ is a consistent estimator of θ_l, for $l = 1, \cdots, s$. From Claim 3.2, $\hat{p}_j$ consistently estimates the true value p_j^0, for $j = 1, 2$. Since the elements of the vector $\nabla\widetilde{\theta}_l(\xi_1, \xi_2)$ are $O_p(1)$, we have

$$\nabla\widetilde{\theta}_l(\xi_1, \xi_2) \cdot (\hat{p}_1 - p_1^0, \hat{p}_2 - p_2^0)' \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

Then, using Slutsky's Theorem (See Davidson, 1994, p314) and (3.4), we have,

$$\widetilde{\theta}_l(\hat{p}_1, \hat{p}_2) \xrightarrow{P} \theta_l^0,$$

for all $l = 1, \cdots, s$, which implies $\widetilde{\underset{\sim}{\theta}}(\hat{p}_1, \hat{p}_2) \xrightarrow{P} \underset{\sim}{\theta}_0$. □

Here also we use the notation $\underset{\sim}{\eta}$ for the set of all the parameters p_1, p_2 and $\underset{\sim}{\theta}$, in line of the same used in Section 3.1, for ease of notation in presenting the simulation results in

Section 5.1.

4 Non-parametric Estimation

In this section, we develop methods to estimate the sub-distribution functions non-parametrically in presence of the two kinds of additional information discussed in Section 3 for parametric estimation.

4.1 Estimation via Validation Sample Approach

Suppose we have K fixed monitoring time pointes denoted by $\tau_1 < \cdots < \tau_K$ with n_i individuals being monitored at τ_i, for $i = 1, \cdots, K$. At any monitoring time τ_i, define $\delta^l_{gi} = 1$, if the lth individual observed at τ_i is found to have failed with g as the observed set of possible causes, for $l = 1, \cdots, n_i$ and $g = \{1\}, \{2\}, \{1,2\}$. Let us also write the set of individuals observed at τ_i with missing failure causes (that is, $g = \{1, 2\}$) as $V_i = \{l \colon \delta^l_{\{1,2\}i} = 1\}$, for $i = 1, \cdots, K$. Let us draw a validation sample S_{V_i} of size m_i from V_i with $m_i \leq \mid V_i \mid$. For the individuals in this validation sample, the true cause of failure can be determined possibly with some more stringent diagnostic procedure. Note that the conditional probability of the true cause being $J = j$, given the original observation $\{G = \{1, 2\}, T \leq X = \tau_i\}$, is

$$P\left[J = j \mid G = \{1,2\}, T \leq X = \tau_i\right] = \frac{(1-p_j)F_j(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)},$$

for $j = 1, 2$, and for $i = 1, \cdots, K$. Thus, the likelihood of observations from the K validation samples is

$$L^*_V \propto \prod_{i=1}^{K}\Bigg[\prod_{v\in S_{V_i}}\left\{\frac{(1-p_1)F_1(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)}\right\}^{\delta^*_{vi}} \times\left\{\frac{(1-p_2)F_2(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)}\right\}^{(1-\delta^*_{vi})}\Bigg]$$

$$= \prod_{i=1}^{K}\Bigg[\left\{\frac{(1-p_1)F_1(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)}\right\}^{d^*_i} \times\left\{\frac{(1-p_2)F_2(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)}\right\}^{(m_i-d^*_i)}\Bigg], \tag{4.1}$$

where $\delta^*_{vi} = \mathbb{I}\left[J = 1, T \le X = \tau_i, G = g_v = \{1,2\}\right]$, for $v \in S_{V_i}$, and $d^*_i = \sum_{v \in S_{V_i}} \delta^*_{vi}$, giving the number of type 1 failures in the ith validation sample, for $i = 1, \cdots, K$. Then the total likelihood is given by

$$\begin{aligned}
L &= L_O \times L^*_V \\
&\propto \prod_{i=1}^{K} \Bigg[\{p_1 F_1(\tau_i)\}^{d_{\{1\}i}} \{p_2 F_2(\tau_i)\}^{d_{\{2\}i}} \{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)\}^{d_{\{1,2\}i}} S(\tau_i)^{n_i - d_i} \\
&\qquad\qquad \times \left\{ \frac{(1-p_1)F_1(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)} \right\}^{d^*_i} \\
&\qquad\qquad\quad \times \left\{ \frac{(1-p_2)F_2(\tau_i)}{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)} \right\}^{(m_i - d^*_i)} \Bigg],
\end{aligned}$$

where L_O is the likelihood of the original data, as given by (2.1), d_{gi}'s, for $g \in \mathcal{G} \setminus \phi$, are the number of individuals failed with observed set of possible causes g out of the n_i individuals monitored at τ_i and $d_i = \sum_{g \in \mathcal{G} \setminus \phi} d_{gi}$. Writing this total likelihood in terms of the vector $\underset{\sim}{\gamma}$ consisting of parameters $p_1,\ p_2$ and the $F_j(\tau_i)$'s, we have

$$\begin{aligned}
L(\underset{\sim}{\gamma}) \propto \prod_{i=1}^{K} \Bigg[&\{p_1 F_1(\tau_i)\}^{d_{\{1\}i}} \{p_2 F_2(\tau_i)\}^{d_{\{2\}i}} \{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)\}^{d_{\{1,2\}i} - m_i} \\
&\times \{(1-p_1)F_1(\tau_i)\}^{d^*_i} \{(1-p_2)F_2(\tau_i)\}^{m_i - d^*_i} S(\tau_i)^{n_i - d_i} \Bigg]. \qquad (4.2)
\end{aligned}$$

To check the identifiability of the likelihood, let $\underset{\sim}{\gamma}^{(1)}$ and $\underset{\sim}{\gamma}^{(2)}$ be two values of the parameter vector $\underset{\sim}{\gamma}$ and consider the equality

$$\log L(\underset{\sim}{\gamma}^{(1)}) = \log L(\underset{\sim}{\gamma}^{(2)}). \qquad (4.3)$$

For a fixed i, equating coefficients of $d_{\{1\}i}$ and d^*_i from both sides of (4.3) and then adding them up, we get, $F_1^{(1)}(\tau_i) = F_1^{(2)}(\tau_i)$. Proceeding in the similar manner after equating coefficients of $d_{\{2\}i}, m_i - d^*_i$ from both sides of (4.3), we have $F_2^{(1)}(\tau_i) = F_2^{(2)}(\tau_i)$. This holds for all $i = 1, \cdots, K$ and hence, $\underset{\sim}{F}^{(1)} = \underset{\sim}{F}^{(2)}$, where $\underset{\sim}{F}$ is the vector with entries $F_j(\tau_i)$, for $j = 1, 2$ and $i = 1, \cdots, K$. Using $F_1^{(1)}(\tau_i) = F_1^{(2)}(\tau_i)$ and equating the coefficient of $d_{\{1\}i}$ from both sides of (4.3), for a given i, we obtain $p_1^{(1)} = p_1^{(2)}$. Similarly, it can be shown that $p_2^{(1)} = p_2^{(2)}$ and thus the model is identifiable. Hence the total likelihood (4.2) can be maximized with respect to $p_1,\ p_2$ and the $F_j(\tau_i)$'s to obtain their MLEs.

We reparametrize the model in terms of 'interval hazards', λ_{ji}'s, to transform the constraint optimization problem into an unconstrained one (similar to Koley and Dewanji (2018b)). In particular, $\lambda_{ji} = [F_j(\tau_i) - F_j(\tau_{i-1})]/[1 - \sum_{j=1}^{2} F_j(\tau_{i-1})]$ is the cause-specific interval hazard of failure between τ_{i-1} and τ_i due to cause j, given survival beyond τ_{i-1}, for $j = 1, 2$ and $i = 1, \cdots, K$ with $\tau_0 = 0$. The sub-distribution functions $F_1(\tau_i)$ and $F_2(\tau_i)$ can be written in terms of the interval hazards. From the likelihood (4.2), written in terms of λ_{ji}'s, p_1 and p_2, it is clear that there is no closed form for the estimators. Hence, numerical maximization method is used for estimation of parameters. The *optim* function in R software is used to maximize the likelihood.

In order to study the asymptotic properties, we write the full data, including those observed in the validation sample, in the following manner. Let us define the variable $\epsilon_i \in \{1, \cdots, 6\}$, for an observation at τ_i, in the following way (See Section 3.1):

$$\epsilon_i = \begin{cases} 1, & \text{if } T \le \tau_i, g = \{1\}, \\ 2, & \text{if } T \le \tau_i, g = \{2\}, \\ 3, & \text{if } T \le \tau_i, g = \{1,2\}, \text{ the individual is not included in the validation sample}, \\ 4, & \text{if } T \le \tau_i, g = \{1,2\}\text{, the individual is included in the validation sample} \\ & \qquad \text{with true cause 1}, \\ 5, & \text{if } T \le \tau_i, g = \{1,2\}\text{, the individual is included in the validation sample} \\ & \qquad \text{with true cause 2}, \\ 6, & \text{if } T > \tau_i, \end{cases}$$

for $i = 1, \cdots, K$. Let π_i be the probability of inclusion in the validation sample at τ_i, for $i = 1, \cdots, K$. We assume this probability to be same for all the individuals observed at τ_i. For a given i, there are n_i independent realizations from the identical distribution of ϵ_i as given by

$$f_i^*(\epsilon_i; \underset{\sim}{\gamma}) = \begin{cases} p_1 F_1(\tau_i), & \text{if } \epsilon_i = 1 \\ p_2 F_2(\tau_i), & \text{if } \epsilon_i = 2 \\ \{(1-p_1)F_1(\tau_i) + (1-p_2)F_2(\tau_i)\}\,(1-\pi_i), & \text{if } \epsilon_i = 3 \\ (1-p_1)F_1(\tau_i)\pi_i, & \text{if } \epsilon_i = 4 \\ (1-p_2)F_2(\tau_i)\pi_i, & \text{if } \epsilon_i = 5 \\ S(\tau_i), & \text{if } \epsilon_i = 6, \end{cases}$$

with respect to the dominating measure μ, the counting measure, for $i = 1, \cdots, K$. The likelihood (4.2) can then be re-written as the product of these densities, as given by

$$L(\underset{\sim}{\gamma}) = \prod_{i=1}^{K} \prod_{l=1}^{n_i} f_i^*(\epsilon_i^l; \underset{\sim}{\gamma}), \tag{4.4}$$

where ϵ_i^l is the lth realization of the random variable ϵ_i, for $l = 1, \cdots, n_i$ and $i = 1, \cdots, K$.

Note that the sub-distribution functions $F_1(\tau_i)$ and $F_2(\tau_i)$ in $f_i^*(\epsilon_i, \underset{\sim}{\gamma})$ can be written in terms of the interval hazards, λ_{ji}'s. The parameter vector $\underset{\sim}{\gamma}$ now consists of p_1, p_2 and the vector $\underset{\sim}{\lambda}$ of the interval hazards λ_{ji}'s. The representation of likelihood function in (4.4) as the product of the density functions is useful in establishing the asymptotic properties of the MLE $\hat{\underset{\sim}{\gamma}} = (\hat{p}_1, \hat{p}_2, \hat{\underset{\sim}{\lambda}})$. Let us consider the following properties:

R1 Suppose $\underset{\sim}{\gamma}^{(1)} = (p_1^{(1)}, p_2^{(1)}, \underset{\sim}{\lambda}^{(1)})$ and $\underset{\sim}{\gamma}^{(2)} = (p_1^{(2)}, p_2^{(2)}, \underset{\sim}{\lambda}^{(2)})$ with $\underset{\sim}{\gamma}^{(1)} \neq \underset{\sim}{\gamma}^{(2)}$ be two values of the parameter vector $\underset{\sim}{\gamma}$. Then the corresponding density functions $f_i^*(\epsilon_i \mid \underset{\sim}{\gamma}^{(1)})$ and $f_i^*(\epsilon_i \mid \underset{\sim}{\gamma}^{(2)})$ are not equal for some ϵ_i and i. The likelihood function being the product of these densities, we have $L(\underset{\sim}{\gamma}^{(1)}) \neq L(\underset{\sim}{\gamma}^{(2)})$.

R2 For each $i = 1, \cdots, K$, $E_{i,\underset{\sim}{\gamma}}\left[\frac{\partial}{\partial(\underset{\sim}{\gamma})} \log f_i^*(\epsilon_i \mid \underset{\sim}{\gamma})\right] = 0$ and the matrix obtained by taking double derivatives with respect to the parameter vector $\underset{\sim}{\gamma}$, $E_{i,\underset{\sim}{\gamma}}\left[-\frac{\partial^2}{\partial\underset{\sim}{\gamma}\partial\underset{\sim}{\gamma}'} \log f_i^*(\epsilon_i \mid \underset{\sim}{\gamma})\right] = E_{i,\underset{\sim}{\gamma}}\left[(\frac{\partial}{\partial\underset{\sim}{\gamma}} \log f_i^*(\epsilon_i \mid \underset{\sim}{\gamma}))^2\right] = \mathcal{I}_i(\underset{\sim}{\gamma})$, say, is assumed to be positive definite.

R3 For a given i, the density function $f_i^*(\epsilon_i \mid \underset{\sim}{\eta})$ is linear in p_1 and p_2 and quadratic in each component of $\underset{\sim}{\lambda}$. Hence, it is continuous in each component of the parameter vector $\underset{\sim}{\gamma}$ admitting all third order partial derivatives, which are bounded by functions with finite expectations.

Theorem 4.1. *Under the assumption* $\frac{n_i}{n} \to w_i$, *where* w_i*'s are positive constants for all* $i = 1, \cdots, K$ *with* $\sum_{i=1}^K w_i = 1$, *we have, under* ***R1-R3***,

1. $\hat{\underset{\sim}{\gamma}} \xrightarrow{P} \underset{\sim}{\gamma}_0$,
2. $\sqrt{n}(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_0)$ *is asymptotically a normal random vector with mean vector* $\underset{\sim}{0}$ *and variance-covariance matrix* $\left[\sum_{i=1}^k w_i \mathcal{I}_i(\underset{\sim}{\gamma}_0)\right]^{-1}$.

Proof. The proof of the Theorem is similar to Theorem 3.2 of Koley and Dewanji (2018b) and, hence, is skipped here. □

Since $\hat{\underset{\sim}{\gamma}}$ is a consistent estimator of $\underset{\sim}{\gamma}$, from the first part of Theorem 4.1, using invariance property and the Weak Law of Large Numbers, $-\frac{1}{n}\frac{\partial^2}{\partial\underset{\sim}{\gamma}\partial\underset{\sim}{\gamma}'} \log L(\underset{\sim}{\gamma}) = -\sum_{i=1}^K \frac{n_i}{n}\frac{1}{n_i}\sum_{l=1}^{n_i}\frac{\partial^2}{\partial\underset{\sim}{\gamma}\partial\underset{\sim}{\gamma}'} \log f_i^*(\epsilon_i^l; \underset{\sim}{\gamma})$, evaluated at $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}$, can be taken as a consistent estimator of $\sum_{i=1}^K w_i \mathcal{I}_i(\underset{\sim}{\gamma}_0)$.

4.2 Indirect Information on masking probabilities

Suppose we have some indirect information on the probabilities p_1 and p_2 of the following nature. Define the conditional probability, as in Section 3.2,

$$q_j = P\left[J = j \mid G = \{1,2\}, \delta = 1\right], \quad \text{for} \quad j = 1, 2,$$

where δ represents the indicator random variable indicating failure or not (See Section 2). It is clear that $q_1 + q_2 = 1$, since it is assumed that there is no misspecification in the data. Suppose we have additional knowledge on this conditional probability q_1 or q_2. Using Bayes' rule, as in Section 3.2, we have the masking probability

$$\begin{aligned}
1 - p_j &= P\left[G = \{1,2\} \mid J = j, \delta = 1\right] \\
&= \frac{P\left[G = \{1,2\}, \delta = 1\right] P\left[J = j \mid G = \{1,2\}, \delta = 1\right]}{\sum_{g' \ni j} P\left[G = g', \delta = 1\right] P\left[J = j \mid G = g', \delta = 1\right]} \\
&= \frac{P\left[G = \{1,2\}, \delta = 1\right] q_j}{P\left[G = \{j\}, \delta = 1\right] + P\left[G = \{1,2\}, \delta = 1\right] q_j}, \quad \text{for } j = 1, 2. \qquad (4.5)
\end{aligned}$$

Similar to the Claim 3.2 in Section 3.2, we have the following claim. Although the two claims are identical, the proof is little different due to the difference in the data configuration.

Claim 4.1. *The masking probability $1 - p_j$ can be consistently estimated by*

$$\frac{q_j f_{\{1,2\}}}{f_{\{j\}} + q_j f_{\{1,2\}}}, \quad \textit{for } j = 1, 2,$$

where $f_g = \sum_{i=1}^K d_{gi}$ is the number of failures ($\delta = 1$) with g as the observed set of possible causes in the sample, for $g \in \mathcal{G} \setminus \phi$.

Proof. Let us write $r_{gi} = P\left[G = g \mid X = \tau_i\right]$ so that its MLE is $\hat{r}_{gi} = \dfrac{d_{gi}}{n_i}$, for $g \in \mathcal{G} \setminus \phi$ and $i = 1, \cdots, K$, following the same multinomial argument as that in the proof of Claim 3.2 but for the observations at τ_i only. Note that $r_g = P\left[G = g\right] = \sum_{i=1}^K r_{gi} P\left[X = \tau_i\right] = \sum_{i=1}^K r_{gi} \dfrac{n_i}{n}$ so that its MLE is $\hat{r}_g = \sum_{i=1}^K \hat{r}_{gi} \dfrac{n_i}{n} = \sum_{i=1}^K \dfrac{d_{gi}}{n} = \dfrac{f_g}{n}$, for $g \in \mathcal{G} \setminus \phi$. Since $1 - p_j$, as given in (4.5), is a continuous function of $\underset{\sim}{r} = (r_g, g \in \mathcal{G} \setminus \phi)$, namely,

$$1 - p_j = b_j(\underset{\sim}{r}) = \frac{r_{\{1,2\}} q_j}{r_{\{j\}} + r_{\{1,2\}} q_j},$$

we have the MLE of $1 - p_j$ as $b_j(\underset{\sim}{\hat{r}})$ by using the invariance property, for $j = 1, 2$. Also, note that, by the multinomial generalization of the central limit theorem (Trotter, 1959),

$\{\sqrt{n}_i(\hat{r}_{gi} - r_{gi}), g \in \mathcal{G} \setminus \phi\} = \sqrt{n}_i(\hat{\underset{\sim}{r}}_i - \underset{\sim}{r}_i)$, say, converges in distribution to a normal random vector with mean vector $\underset{\sim}{0}$ and variance-covariance matrix $\Sigma_i = \text{Diag}\,(r_{gi}, g \in \mathcal{G} \setminus \phi) - \left(\left(r_{gi} r_{g'i}\right)\right)_{g,g' \in \mathcal{G}\setminus\phi}$. Therefore,

$$\sqrt{n}(\hat{\underset{\sim}{r}} - \underset{\sim}{r}) = \sqrt{n} \sum_{i=1}^{K} \frac{n_i}{n} (\hat{\underset{\sim}{r}}_i - \underset{\sim}{r}_i) = \sum_{i=1}^{K} \sqrt{\frac{n_i}{n}} \times \sqrt{n}_i(\hat{\underset{\sim}{r}}_i - \underset{\sim}{r}_i),$$

converges in distribution to a normal random vector with mean vector $\underset{\sim}{0}$ and variance-covariance matrix $\sum_{i=1}^{K} w_i \Sigma_i = \Sigma$, say, where $\frac{n_i}{n} \to w_i$ as $n \to \infty$. Hence, using the multivariate delta method (Cox, 2005), we have

$$\sqrt{n}(b_j(\hat{\underset{\sim}{r}}) - b_j(\underset{\sim}{r})) \xrightarrow{d} N\left(0, b_j'(\underset{\sim}{r})^T \Sigma b_j'(\underset{\sim}{r})\right),$$

where $b_j'(\underset{\sim}{r}) = \frac{\partial}{\partial \underset{\sim}{r}} b_j'(\underset{\sim}{r})$ is the 3×1 vector of the first order partial derivatives of $b_j(\underset{\sim}{r})$ with respect to $\underset{\sim}{r}$. Hence, the MLE $b_j(\hat{\underset{\sim}{r}})$ is also consistent. □

The estimated p_j's can be treated as known quantities in further estimation of the $F_j(\tau_i)$'s, for $j = 1, 2$ and $i = 1, \cdots, K$, using the method involving reparametrization of the $F_j(\tau_i)$'s in terms of the interval hazards λ_{ji}'s. Clearly, the resulting estimate, denoted by $\tilde{\underset{\sim}{\lambda}}$, of $\underset{\sim}{\lambda}$, the set of λ_{ji}'s, is not the usual MLE, as discussed in Section 3.2, and so the standard error of the estimates cannot be obtained directly from the observed information matrix. Instead, the bootstrap method is used for obtaining the standard errors. This involves drawing a simple random sample of the same size n_i with replacement from the set of observations (that is, the g_{il}'s) at each τ_i, called a bootstrap sample, for $i = 1, \cdots, K$. Thus, there are K bootstrap samples at the K monitoring times. For our purpose, we call this a set of K bootstrap samples. Then, the model parameters p_1, p_2 and λ_{ji}'s are estimated from this set of K bootstrap samples following the same procedure as described above and using the information on q_1 and q_2. This procedure is repeated B times to obtain B such sets of K bootstrap samples along with the corresponding sets of parameter estimates. The sample standard deviation of these estimates over the B sets of K bootstrap samples gives an estimate of the corresponding standard error.

Theorem 4.2. *If the elements of the gradient vector $\nabla\widetilde{\lambda}_{jl}(\cdot,\cdot)$ with respect to p_1 and p_2 are $O_p(1)$ in a neighbourhood of the true value (p_1^0, p_2^0), for all $j = 1, 2$ $l = 1, \cdots, K$, then the estimate of $\underset{\sim}{\lambda}$, obtained by the method described above and denoted by $\widetilde{\underset{\sim}{\lambda}}(\hat{p}_1, \hat{p}_2)$, is a consistent estimator of $\underset{\sim}{\lambda}$.*

Proof. The proof follows the similar arguments as those used for Theorem 3.2. Suppose

(p_1^0, p_2^0) and $\underset{\sim}{\lambda}_0 = (\lambda_{ji}^0; j = 1,2\ i = 1, \cdots, K)$ are the true values of the parameter vectors (p_1, p_2) and $\underset{\sim}{\lambda}$, respectively. Then, for given j and i, using Taylor's Theorem (Malik and Arora, 1992), we have

$$\widetilde{\lambda}_{ji}(\hat{p}_1, \hat{p}_2) - \lambda_{ji}^0 = \widetilde{\lambda}_{ji}(p_1^0, p_2^0) - \lambda_{ji}^0 + \triangledown\widetilde{\lambda}_{ji}(\xi_1, \xi_2) \cdot (\hat{p}_1 - p_1^0, \hat{p}_2 - p_2^0)', \tag{4.6}$$

where (ξ_1, ξ_2) is a convex combination of the two points $(\hat{p}_1, \hat{p}_2)$ and (p_1^0, p_2^0) and $\triangledown\widetilde{\lambda}_{ji}(\xi_1, \xi_2)$ is the gradient vector, or the vector of partial derivatives of the function $\widetilde{\lambda}_{ji}(p_1, p_2)$ with respect to p_1 and p_2, computed at (ξ_1, ξ_2). When p_1 and p_2 are known, $\widetilde{\lambda}_{ji}$ is a consistent estimator of λ_{ji}, for $j = 1, 2,\ i = 1, \cdots, K$. This follows from Theorem 3.2 of Koley and Dewanji (2019) with known p_1 and p_2. From Claim 4.1, $\hat{p}_j$ consistently estimates the true value p_j^0, for $j = 1, 2$. Since the elements of the vector $\triangledown\widetilde{\lambda}_{ji}(\xi_1, \xi_2)$ are $O_p(1)$, so

$$\triangledown\widetilde{\lambda}_{ji}(\xi_1, \xi_2) \cdot (\hat{p}_1 - p_1^0, \hat{p}_2 - p_2^0)' \xrightarrow{P} 0.$$

Then, using Slutsky's Theorem (See (Davidson, 1994, p314)) and (4.6), we have

$$\widetilde{\lambda}_{ji}(\hat{p}_1, \hat{p}_2) \xrightarrow{P} \lambda_{ji}^0,$$

for all $j = 1, 2\ i = 1, \cdots, K$, which implies $\widetilde{\lambda}(\hat{p}_1, \hat{p}_2) \xrightarrow{P} \underset{\sim}{\lambda}_0$. □

5 Simulation Studies

We conduct several simulation studies to investigate the performance of the estimators described in Sections 3 and 4.

5.1 Parametric Estimation

First we specify the distributional assumption for the failure time random variable T. Next we simulate n observations from the assumed lifetime distribution $F(\cdot; \underset{\sim}{\theta})$ with the corresponding monitoring times being simulated from a known distribution $H(\cdot)$ of X. Once an observation is found to be a failure (that is, its failure time is less than the monitoring time $X = x$), then the set of possible causes of failure g is simulated with probability $P\left[G = g \mid T \le x\right]$, for $g = \{1\}, \{2\}, \{1, 2\}$. This probability is given by $\dfrac{\sum_{j \in g} p_{gj} F_j(x; \underset{\sim}{\theta})}{F(x; \underset{\sim}{\theta})}$, where $F_j(\cdot; \underset{\sim}{\theta})$ is the sub-distribution function due to cause j, for $j = 1, 2$. Thus, we simulate data of the form (x_i, g_i), for $i = 1, \cdots, n$. For each simulation study, we carry out this simulation 10000 times to get 10000 such simulated data sets. For each such simulated data set, we estimate the related model parameters by using the methods of Sections 3.1 and 3.2, requiring observation from a validation sample and information on the q_j's, respectively, along with the corresponding standard errors. We also obtain the asymptotic 95% confidence interval for each parameter based on normal approximation to check if this contains the true value of the parameter. Then, we consider the average of the 10000 estimates for each parameter and also the average of the corresponding standard errors, denoted by ASE. Bias for each parameter estimator is estimated by subtracting the true value of the parameter from the corresponding average of 10000 estimates and absolute value is reported. The sample standard error, denoted by SSE, for each parameter estimator is calculated as the standard deviation of the corresponding 10000 estimates. Cover percentage of an estimator is estimated by the proportion of times the corresponding asymptotic 95% confidence intervals contain the true value. This is denoted by CP. Three different choices of sample size, $n = 50, 150$ and 250, are considered for each simulation study to understand the behavior of these estimators with moderate to large sample size.

In validation sample approach, after obtaining the set V in each simulated data set, the validation sample subset S_V is generated by drawing randomly (without replacement) a fraction f of individuals from V. Different choices of p_1, p_2 and f are taken as ($p_1 = p_2 = 0.8$), ($p_1 = 0.9, p_2 = 0.8$) and $f = 0.2$, 0.3 and 0.5 to study their impact on estimates. First, the failure time distribution is taken to be exponential with rate parameter 1 and two types of failures occurring with rate ratio 6: 4. Next, failure time is generated from Weibull distribution with the same rate parameter 1 and two different choices of the shape parameter, denoted by b in the tables, as 0.8 (decreasing failure rate) and 1.2 (increasing failure rate). In Tables 2-4, we present the results associated with different sets of model parameters corresponding to the method of Section 3.1 using validation sample corresponding to the three failure time distributions, respectively. As expected, absolute value of the bias of the estimates, and also the corresponding ASE and SSE, decrease with sample size. Also the values of ASE and SSE become similar to each other with increasing sample size. For a fixed n, there is a gain in precision (given by smaller ASE and SSE) due to increasing fraction f, as expected. Also, the CP values are closer to 0.95 with increasing sample size, providing evidence in favour of asymptotic normality.

For the approach described in Section 3.2 using indirect information on the masking probabilities, we consider different choices of q_1 and q_2 for a simulation study to investigate the performance of the corresponding estimators. The failure time distribution is again taken to be exponential with rate parameter 1 and the two types of failures with rate ratio 6: 4. Next, failure time is generated from Weibull distribution with the same rate parameter 1 and two different choices of the shape parameter, denoted by b in the tables, as 0.8 and 1.2. Note that the true values of q_1 and q_2 can be obtained, at the monitoring time $X = x$, as

$$
\begin{aligned}
q_j(x) &= P\left[J = j \mid T \leq X = x, G = \{1,2\}\right] \\
&= \frac{(1-p_j)F_j(x)}{(1-p_1)F_1(x) + (1-p_2)F_2(x)},
\end{aligned}
$$

for $j = 1, 2$. It can be checked that these quantities are independent of x for both exponential and Weibull failure time distribution with common shape parameter. With $p_1 = 0.9$ and $p_2 = 0.8$, the true value of (q_1, q_2) thus equals (0.429, 0.571) for exponential failure time distribution, (0.409,0.591) for Weibull distribution with shape parameter 0.8 and, finally, (0.449,0.551) for Weibull distribution with shape parameter 1.2. Along with the true choice, we also consider two wrong choices of (q_1, q_2), taken as $(0.3, 0.7)$ and $(0.6, 0.4)$, for analysis and the results are presented in Tables 5-7, for the three assumptions on failure time distribution, respectively. The standard errors of the estimates are computed using the bootstrap procedure as discussed in Section 3.2. As expected, absolute value of the bias, SSE and ASE decrease with increase in sample size for correct choices of q_1 and q_2. The CP values are also closer to 0.95 with increasing sample size, providing evidence in favour of asymptotic normality. However, there seems to be bias in the estimates for wrong choices of q_1 and q_2.

5.2 Non-parametric Estimation

The failure time distribution is taken to be exponential with rate 1 and the two types of failures occurring with rate ratio $6:4$. The five fixed monitoring time points, $\tau_1, \cdots, \tau_5$, are chosen, as before, to be the $10^{th}, 25^{th}, 50^{th}, 75^{th}$ and 90^{th} quantiles of the true exponential failure time distribution with rate 1. Of the n observations under study, first n_1 observations are monitored at τ_1, next n_2 observations at τ_2 and so on. For a given i, suppose an individual monitored at τ_i is observed to have failed; then the corresponding observed set of possible causes is generated with probability

$$
P\left[G = g \mid T \leq X = \tau_i\right] = \frac{\sum_{j \in g} p_{gj} F_j(\tau_i)}{F(\tau_i)},
$$

for $g \in \mathcal{G} \setminus \phi$, $i = 1, \cdots, K$, resulting in a simulated data set $\{(x_l, g_l); l = 1, \cdots, n\}$. Clearly, each x_l is one of $\tau_1, \cdots, \tau_K$ and g_l is taken to be the empty set ϕ for a censored individual. This is carried out 10000 times to obtain 10000 such simulated data sets. For each simulated data set, we obtain the estimates of the sub-distribution functions at the five monitoring time points using the methods discussed in Sections 4.1 and 4.2, along with the corresponding standard errors. Then, we compute the bias by subtracting the true value of the sub-distribution function from the corresponding average of these estimates over 10000 simulation data sets and record the absolute value of this bias. We also compute the average of the standard errors over these 10000 simulated data sets and denote it by ASE. The sample standard errors, denoted by SSEs, are also computed as the square root of the sample variances of these 10000 estimates. Finally, approximate 95% confidence intervals are constructed by using normal approximation of the estimators and then the cover percentage is estimated by the proportion of times these intervals contain the true values. We denote it by CP. We take the n_i's to be equal. Three different choices of n_i's as 50, 150 and 250 are considered in each simulation study to understand the behaviour of the estimators with increasing sample size.

For validation sample approach, a fraction f of the individuals from V_i are drawn by simple random sampling without replacement to generate a validation sample S_{V_i} (See Section 4.1), for $i = 1, \cdots, K$. Different values of $f = 0.3$ and 0.5 are considered to study their impact on the estimates (See Table 8 and 9, respectively). For the analysis with indirect information on masking probabilities, we consider the following steps. Taking true $p_1 = 0.9$ and $p_2 = 0.8$, the values of q_1 and q_2 are computed as 0.429 and 0.571, respectively. Two wrong choices of q_1 and q_2 are taken as (0.3,0.7) and (0.6,0.4) and then the analyses are carried out with the correct and the two wrong choices of (q_1, q_2) (See Tables 10-12, respectively).

In the analysis with validation sample (Tables 8 and 9 corresponding to $f = 0.3$ and 0.5, respectively), we notice that absolute bias, ASE and SSE decrease with increasing sample size, as expected. Also, the values of ASE and SSE become more similar with increasing sample, while the CP values become closer to 0.95 giving evidence in favour of asymptotic normality. Again, with increase in f value, the number of observations included in each validation sample increases and, hence, the standard errors of the estimates decrease or remain same. In the second case with indirect information on masking probabilities, the absolute bias, ASE and SSE decrease with increase in sample size for correct specification of q_1 and q_2, as expected. However, there seems to be bias in the estimates for wrong choices of q_1 and q_2, as is evident with larger sample size in Tables 11 and 12. Note that both ASE and SSE values become similar to each other, and the estimated CPs become closer to 0.95, as the sample size increases providing some evidence in favour of asymptotic normality (except for the cases with wrong choice of q_1 and q_2).

Table 2: Simulation results on $\hat{\eta}$ from the validation sample approach for exponential failure time distribution with $\underset{\sim}{\lambda_1} = 0.6, \lambda_2 = 0.4$.

(n, f)	Para-meters	$p_1 = p_2 = 0.8$				$p_1 = 0.9, p_2 = 0.8$			
		|Bias|	ASE	SSE	CP	|Bias|	ASE	SSE	CP
$(50, 0.2)$	λ_1	0.053	0.205	0.204	0.960	0.043	0.199	0.202	0.965
	λ_2	0.035	0.173	0.172	0.958	0.028	0.166	0.155	0.959
	p_1	0.006	0.192	0.141	0.983	0.005	0.195	0.111	0.992
	p_2	0.033	0.316	0.183	0.978	0.019	0.280	0.171	0.988
$(150, 0.2)$	λ_1	0.027	0.109	0.108	0.954	0.022	0.107	0.106	0.959
	λ_2	0.017	0.092	0.093	0.952	0.014	0.090	0.090	0.953
	p_1	0.005	0.088	0.081	0.974	0.004	0.092	0.070	0.987
	p_2	0.013	0.139	0.115	0.973	0.008	0.123	0.104	0.984
$(250, 0.2)$	λ_1	0.025	0.083	0.083	0.950	0.021	0.081	0.082	0.943
	λ_2	0.017	0.069	0.069	0.951	0.012	0.067	0.067	0.950
	p_1	0.004	0.064	0.063	0.946	0.003	0.062	0.055	0.980
	p_2	0.008	0.093	0.088	0.966	0.004	0.085	0.084	0.956
$(50, 0.3)$	λ_1	0.061	0.194	0.198	0.932	0.055	0.192	0.200	0.933
	λ_2	0.036	0.159	0.162	0.936	0.038	0.156	0.158	0.940
	p_1	0.011	0.153	0.121	0.983	0.002	0.163	0.100	0.994
	p_2	0.010	0.249	0.164	0.922	0.005	0.224	0.152	0.977
$(150, 0.3)$	λ_1	0.041	0.106	0.104	0.963	0.029	0.103	0.106	0.953
	λ_2	0.026	0.087	0.088	0.945	0.020	0.084	0.086	0.941
	p_1	0.003	0.073	0.074	0.930	0.001	0.071	0.062	0.974
	p_2	0.008	0.106	0.100	0.975	0.004	0.096	0.093	0.930
$(250, 0.3)$	λ_1	0.033	0.081	0.082	0.952	0.023	0.079	0.080	0.948
	λ_2	0.020	0.066	0.067	0.950	0.016	0.064	0.065	0.948
	p_1	0.002	0.056	0.057	0.937	0.001	0.049	0.047	0.954
	p_2	0.006	0.077	0.078	0.965	0.004	0.070	0.070	0.963
$(50, 0.5)$	λ_1	0.083	0.195	0.214	0.912	0.083	0.194	0.209	0.955
	λ_2	0.055	0.158	0.170	0.929	0.044	0.153	0.156	0.960
	p_1	0.005	0.126	0.112	0.914	0.003	0.135	0.089	0.995
	p_2	0.005	0.199	0.146	0.910	0.002	0.184	0.133	0.992
	λ_1	0.064	0.106	0.113	0.927	0.040	0.102	0.105	0.945
	λ_2	0.038	0.085	0.086	0.949	0.027	0.082	0.085	0.941

Continued on next page

Table 2 – continued from previous page

(n, f)	Parameters	$p_1 = p_2 = 0.8$				$p_1 = 0.9, p_2 = 0.8$			
		|Bias|	ASE	SSE	CP	|Bias|	ASE	SSE	CP
$(150, 0.5)$	p_1	0.002	0.064	0.067	0.975	0.002	0.055	0.052	0.916
	p_2	0.003	0.085	0.088	0.922	0.001	0.081	0.082	0.912
$(250, 0.5)$	λ_1	0.052	0.081	0.085	0.950	0.037	0.078	0.080	0.950
	λ_2	0.035	0.065	0.066	0.950	0.025	0.063	0.066	0.947
	p_1	0.001	0.050	0.051	0.940	0.001	0.040	0.041	0.958
	p_2	0.001	0.064	0.067	0.977	0.001	0.061	0.062	0.940

Table 3: Simulation results on $\underset{\sim}{\hat{\eta}}$ from the validation sample approach for Weibull failure time distribution (shape parameter $b = 0.8$) with $\lambda_1 = 0.6, \lambda_2 = 0.4$.

(n, f)	Parameters	$p_1 = p_2 = 0.8$				$p_1 = 0.9, p_2 = 0.8$			
		|Bias|	ASE	SSE	CP	|Bias|	ASE	SSE	CP
$(50, 0.2)$	λ_1	0.050	0.261	0.245	0.942	0.046	0.262	0.251	0.932
	λ_2	0.043	0.222	0.209	0.940	0.042	0.221	0.214	0.922
	b	0.091	0.283	0.305	0.971	0.074	0.279	0.303	0.963
	p_1	0.013	0.190	0.142	0.979	0.005	0.193	0.110	0.990
	p_2	0.017	0.279	0.173	0.981	0.018	0.245	0.159	0.991
$(150, 0.2)$	λ_1	0.016	0.140	0.138	0.948	0.015	0.139	0.138	0.944
	λ_2	0.014	0.121	0.116	0.943	0.013	0.120	0.113	0.958
	b	0.026	0.144	0.146	0.956	0.023	0.144	0.141	0.961
	p_1	0.006	0.085	0.083	0.963	0.003	0.091	0.069	0.986
	p_2	0.017	0.279	0.173	0.981	0.011	0.112	0.099	0.990
$(250, 0.2)$	λ_1	0.001	0.106	0.104	0.950	0.010	0.105	0.105	0.945
	λ_2	0.008	0.092	0.093	0.946	0.017	0.110	0.108	0.956
	b	0.012	0.109	0.109	0.952	0.017	0.110	0.108	0.956
	p_1	0.001	0.063	0.062	0.917	0.003	0.061	0.054	0.980
	p_2	0.001	0.084	0.082	0.950	0.004	0.077	0.076	0.956
$(50, 0.3)$	λ_1	0.037	0.249	0.244	0.932	0.055	0.246	0.255	0.938
	λ_2	0.041	0.210	0.205	0.924	0.039	0.205	0.210	0.909
	b	0.071	0.278	0.288	0.979	0.102	0.291	0.325	0.966
	p_1	0.006	0.149	0.122	0.980	0.004	0.157	0.094	0.993
	p_2	0.011	0.215	0.151	0.930	0.005	0.202	0.145	0.987
	λ_1	0.015	0.134	0.135	0.933	0.018	0.133	0.125	0.940

Continued on next page

(n, f)	Para-meters	$p_1 = p_2 = 0.8$ |Bias|	ASE	SSE	CP	$p_1 = 0.9, p_2 = 0.8$ |Bias|	ASE	SSE	CP
	λ_2	0.016	0.115	0.116	0.934	0.010	0.114	0.113	0.940
$(150, 0.3)$	b	0.029	0.144	0.146	0.954	0.026	0.144	0.153	0.958
	p_1	0.006	0.071	0.070	0.919	0.004	0.068	0.061	0.983
	p_2	0.005	0.095	0.091	0.976	0.005	0.088	0.087	0.945
	λ_1	0.009	0.103	0.104	0.948	0.012	0.102	0.105	0.955
	λ_2	0.009	0.089	0.088	0.938	0.009	0.088	0.089	0.943
$(250, 0.3)$	b	0.019	0.110	0.111	0.951	0.015	0.110	0.111	0.944
	p_1	0.001	0.055	0.056	0.925	0.002	0.048	0.047	0.940
	p_2	0.001	0.070	0.072	0.958	0.002	0.065	0.064	0.952
	λ_1	0.025	0.235	0.239	0.938	0.037	0.238	0.242	0.934
	λ_2	0.044	0.200	0.201	0.924	0.042	0.201	0.200	0.924
$(50, 0.5)$	b	0.090	0.286	0.314	0.969	0.084	0.284	0.314	0.963
	p_1	0.003	0.125	0.109	0.983	0.003	0.134	0.086	0.995
	p_2	0.004	0.171	0.130	0.981	0.007	0.161	0.129	0.983
	λ_1	0.017	0.131	0.130	0.939	0.008	0.129	0.133	0.941
	λ_2	0.012	0.112	0.111	0.937	0.013	0.111	0.110	0.948
$(150, 0.5)$	b	0.025	0.144	0.145	0.953	0.021	0.143	0.149	0.945
	p_1	0.001	0.063	0.064	0.931	0.002	0.054	0.051	0.929
	p_2	0.003	0.078	0.079	0.922	0.002	0.074	0.077	0.926
	λ_1	0.016	0.100	0.100	0.947	0.004	0.099	0.099	0.953
	λ_2	0.006	0.086	0.087	0.950	0.009	0.086	0.085	0.951
$(250, 0.5)$	b	0.019	0.109	0.109	0.953	0.013	0.109	0.109	0.953
	p_1	0.001	0.049	0.049	0.938	0.001	0.039	0.039	0.953
	p_2	0.002	0.060	0.061	0.925	0.001	0.057	0.058	0.939

Table 4: Simulation results on $\hat{\underset{\sim}{\eta}}$ from the validation sample approach for Weibull failure time distribution (shape parameter $b = 1.2$) with $\lambda_1 = 0.6, \lambda_2 = 0.4$.

(n, f)	Para-meters	$p_1 = p_2 = 0.8$ |Bias|	ASE	SSE	CP	$p_1 = 0.9, p_2 = 0.8$ |Bias|	ASE	SSE	CP
	λ_1	0.037	0.177	0.175	0.930	0.026	0.175	0.174	0.961
	λ_2	0.023	0.167	0.164	0.935	0.031	0.164	0.163	0.920
$(50, 0.2)$	b	0.155	0.410	0.439	0.974	0.147	0.422	0.502	0.962

Continued on next page

Table 4 – continued from previous page

(n, f)	Para-meters	$p_1 = p_2 = 0.8$ |Bias|	ASE	SSE	CP	$p_1 = 0.9, p_2 = 0.8$ |Bias|	ASE	SSE	CP
	p_1	0.005	0.192	0.142	0.979	0.004	0.196	0.113	0.989
	p_2	0.045	0.360	0.192	0.980	0.019	0.298	0.174	0.990
	λ_1	0.011	0.097	0.098	0.944	0.010	0.096	0.095	0.940
	λ_2	0.011	0.093	0.092	0.942	0.007	0.092	0.091	0.943
$(150, 0.2)$	b	0.051	0.200	0.214	0.944	0.040	0.198	0.196	0.957
	p_1	0.005	0.088	0.084	0.976	0.002	0.091	0.070	0.985
	p_2	0.011	0.155	0.124	0.975	0.012	0.138	0.110	0.988
	λ_1	0.006	0.074	0.074	0.954	0.003	0.073	0.072	0.952
	λ_2	0.007	0.071	0.072	0.943	0.007	0.070	0.070	0.954
$(250, 0.2)$	b	0.028	0.150	0.151	0.951	0.027	0.150	0.152	0.952
	p_1	0.001	0.065	0.063	0.943	0.002	0.064	0.057	0.980
	p_2	0.008	0.104	0.094	0.967	0.004	0.092	0.087	0.968
	λ_1	0.033	0.165	0.174	0.919	0.027	0.167	0.170	0.931
	λ_2	0.036	0.154	0.162	0.911	0.034	0.156	0.158	0.928
$(50, 0.3)$	b	0.194	0.430	0.482	0.973	0.147	0.408	0.450	0.967
	p_1	0.005	0.156	0.126	0.979	0.005	0.162	0.097	0.993
	p_2	0.014	0.281	0.172	0.975	0.004	0.238	0.160	0.984
	λ_1	0.010	0.093	0.096	0.943	0.007	0.093	0.093	0.945
	λ_2	0.015	0.088	0.090	0.944	0.014	0.088	0.090	0.931
$(150, 0.3)$	b	0.052	0.200	0.212	0.965	0.042	0.199	0.215	0.939
	p_1	0.003	0.074	0.075	0.932	0.002	0.073	0.062	0.988
	p_2	0.005	0.117	0.106	0.971	0.002	0.105	0.095	0.971
	λ_1	0.008	0.072	0.075	0.947	0.004	0.071	0.071	0.946
	λ_2	0.008	0.068	0.067	0.953	0.006	0.069	0.071	0.934
$(250, 0.3)$	b	0.033	0.151	0.152	0.949	0.027	0.150	0.154	0.954
	p_1	0.001	0.057	0.057	0.935	0.001	0.049	0.046	0.960
	p_2	0.004	0.083	0.083	0.932	0.001	0.075	0.075	0.940
	λ_1	0.030	0.163	0.164	0.946	0.030	0.163	0.162	0.941
	λ_2	0.033	0.152	0.148	0.941	0.034	0.151	0.157	0.918
$(50, 0.5)$	b	0.151	0.403	0.418	0.970	0.156	0.414	0.456	0.970
	p_1	0.002	0.130	0.111	0.981	0.004	0.141	0.087	0.991
	p_2	0.016	0.230	0.151	0.980	0.004	0.211	0.148	0.988

Continued on next page

Table 4 – continued from previous page

(n, f)	Para-meters	$p_1 = p_2 = 0.8$				$p_1 = 0.9, p_2 = 0.8$			
		|Bias|	ASE	SSE	CP	|Bias|	ASE	SSE	CP
$(150, 0.5)$	λ_1	0.008	0.091	0.090	0.949	0.007	0.091	0.090	0.944
	λ_2	0.009	0.085	0.085	0.942	0.009	0.085	0.085	0.954
	b	0.043	0.198	0.208	0.963	0.034	0.198	0.203	0.959
	p_1	0.001	0.065	0.067	0.928	0.001	0.056	0.050	0.976
	p_2	0.001	0.091	0.091	0.923	0.004	0.086	0.083	0.931
$(n = 250, f = 0.5)$	λ_1	0.005	0.070	0.069	0.950	0.003	0.070	0.070	0.949
	λ_2	0.009	0.066	0.066	0.942	0.003	0.065	0.065	0.949
	b	0.030	0.150	0.148	0.952	0.017	0.149	0.149	0.956
	p_1	0.001	0.051	0.050	0.942	0.001	0.040	0.042	0.925
	p_2	0.001	0.068	0.068	0.931	0.001	0.065	0.065	0.934

Table 5: Simulation results on $\underset{\sim}{\hat{\eta}}$ using information on q_1 and q_2 for exponential(1) failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4,\ p_1 = 0.9, p_2 = 0.8$ having the true (q_1, q_2) in the left panel.

n	Para-	$q_1 = 0.429, q_2 = 0.571$				$q_1 = 0.3, q_2 = 0.7$				$q_1 = 0.6, q_2 = 0.4$			
	meters	|Bias|	ASE	SSE	CP	|Bias|	ASE	SSE	CP	|Bias|	ASE	SSE	CP
50	λ_1	0.015	0.173	0.161	0.942	0.002	0.171	0.160	0.950	0.046	0.179	0.174	0.964
	λ_2	0.016	0.134	0.125	0.942	0.031	0.139	0.130	0.962	0.011	0.129	0.126	0.917
	p_1	0.004	0.054	0.055	0.900	0.025	0.041	0.042	0.775	0.037	0.068	0.067	0.951
	p_2	0.005	0.102	0.105	0.900	0.037	0.112	0.115	0.926	0.042	0.088	0.087	0.806
150	λ_1	0.008	0.093	0.092	0.954	0.014	0.075	0.072	0.960	0.032	0.096	0.093	0.953
	λ_2	0.005	0.073	0.076	0.951	0.024	0.075	0.072	0.960	0.015	0.071	0.072	0.912
	p_1	0.002	0.030	0.030	0.946	0.026	0.023	0.022	0.698	0.037	0.039	0.039	0.890
	p_2	0.004	0.057	0.058	0.945	0.037	0.064	0.064	0.926	0.049	0.046	0.045	0.729
250	λ_1	0.002	0.071	0.073	0.952	0.014	0.071	0.069	0.943	0.027	0.073	0.073	0.941
	λ_2	0.002	0.055	0.056	0.951	0.022	0.058	0.057	0.957	0.020	0.054	0.053	0.907
	p_1	0.001	0.023	0.023	0.947	0.028	0.017	0.017	0.591	0.035	0.030	0.030	0.826
	p_2	0.001	0.044	0.043	0.945	0.033	0.049	0.049	0.920	0.051	0.035	0.035	0.644

Table 6: Simulation results on $\underset{\sim}{\hat{\eta}}$ using information on q_1 and q_2 for Weibull failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4,\ b = 0.8,\ p_1 = 0.9, p_2 = 0.8$ having the true (q_1, q_2) in the left panel.

n	Para-	$q_1 = 0.409, q_2 = 0.591$				$q_1 = 0.3, q_2 = 0.7$				$q_1 = 0.6, q_2 = 0.4$			
	meters	10 × |Bias|	ASE	SSE	CP	10 × |Bias|	ASE	SSE	CP	10 × |Bias|	ASE	SSE	CP
	λ_1	0.463	0.235	0.232	0.922	0.180	0.240	0.223	0.943	0.722	1.192	0.241	0.953
	λ_2	0.535	2.980	0.198	0.933	0.596	0.217	0.195	0.951	0.086	0.198	0.195	0.928
50	b	0.939	0.521	0.306	0.977	0.810	0.550	0.302	0.972	0.724	0.527	0.291	0.978
	p_1	0.035	0.075	0.053	0.981	0.223	0.066	0.038	0.946	0.438	0.085	0.072	0.938
	p_2	0.025	0.106	0.096	0.969	0.298	0.110	0.105	0.957	0.500	0.099	0.078	0.911
	λ_1	0.153	0.121	0.130	0.928	0.092	0.118	0.127	0.931	0.473	0.125	0.128	0.942
	λ_2	0.085	0.108	0.109	0.938	0.269	0.113	0.112	0.943	0.168	0.101	0.103	0.923
150	b	0.186	0.152	0.144	0.956	0.199	0.152	0.149	0.967	0.325	0.152	0.143	0.965
	p_1	0.012	0.041	0.030	0.968	0.240	0.037	0.023	0.865	0.405	0.048	0.038	0.898
	p_2	0.020	0.059	0.056	0.942	0.289	0.061	0.061	0.926	0.539	0.055	0.042	0.822
	λ_1	0.049	0.092	0.099	0.930	0.106	0.089	0.095	0.923	0.390	0.096	0.099	0.934
	λ_2	0.048	0.083	0.084	0.956	0.264	0.086	0.083	0.944	0.230	0.077	0.081	0.914
250	b	0.035	0.111	0.109	0.949	0.161	0.112	0.112	0.955	0.142	0.113	0.110	0.959
	p_1	0.006	0.032	0.025	0.966	0.237	0.027	0.018	0.790	0.427	0.036	0.030	0.799
	p_2	0.018	0.046	0.043	0.946	0.310	0.047	0.048	0.901	0.535	0.041	0.031	0.733

Table 7: Simulation results on $\hat{\underset{\sim}{\eta}}$ using information on q_1 and q_2 for Weibull failure time distribution with $\lambda_1 = 0.6, \lambda_2 = 0.4$, $b = 1.2$, $p_1 = 0.9, p_2 = 0.8$ having the true (q_1, q_2) in the left panel.

n	Para-	$q_1 = 0.449, q_2 = 0.551$				$q_1 = 0.3, q_2 = 0.7$				$q_1 = 0.6, q_2 = 0.4$			
	meters	$10 \times$ |Bias|	ASE	SSE	CP	$10 \times$ |Bias|	ASE	SSE	CP	$10 \times$ |Bias|	ASE	SSE	CP
	λ_1	0.432	0.162	0.172	0.939	0.097	0.160	0.164	0.945	0.445	0.164	0.164	0.949
	λ_2	0.423	0.162	0.157	0.931	0.443	0.155	0.153	0.948	0.163	0.158	0.151	0.941
50	b	1.786	1.319	0.509	0.995	1.279	1.067	0.430	0.990	1.506	1.290	0.453	0.986
	p_1	0.022	0.084	0.055	0.983	0.280	0.075	0.041	0.931	0.328	0.084	0.067	0.970
	p_2	0.094	0.122	0.114	0.974	0.483	0.123	0.129	0.917	0.383	0.122	0.093	0.950
	λ_1	0.112	0.086	0.090	0.940	0.072	0.086	0.088	0.942	0.218	0.087	0.091	0.931
	λ_2	0.108	0.081	0.084	0.932	0.304	0.084	0.083	0.930	0.071	0.079	0.082	0.929
150	b	0.439	0.215	0.212	0.960	0.335	0.214	0.199	0.968	0.435	0.213	0.211	0.966
	p_1	0.022	0.047	0.030	0.974	0.295	0.041	0.023	0.837	0.310	0.052	0.040	0.932
	p_2	0.019	0.067	0.061	0.961	0.406	0.068	0.068	0.910	0.453	0.064	0.048	0.900
	λ_1	0.052	0.065	0.069	0.941	0.079	0.069	0.068	0.930	0.235	0.067	0.065	0.946
	λ_2	0.092	0.062	0.064	0.933	0.240	0.060	0.063	0.935	0.119	0.060	0.063	0.932
250	b	0.337	0.157	0.156	0.958	0.272	0.156	0.157	0.962	0.232	0.156	0.153	0.959
	p_1	0.004	0.036	0.024	0.968	0.314	0.039	0.017	0.747	0.298	0.040	0.029	0.913
	p_2	0.008	0.051	0.049	0.960	0.393	0.052	0.052	0.883	0.451	0.049	0.036	0.855

Table 8: Simulation results on $\left\{\hat{F}_j(\tau_i), j = 1, 2\right\}$ and $(\hat{p}_1, \hat{p}_2)$ for exponential(1) failure time distribution with 6: 4 rate ratio for the two failure types and $(p_1, p_2) = (0.9, 0.8)$ in presence of validation sample (f=0.3). The five τ_i's are 0.105, 0.288, 0.693, 1.386 and 2.303.

	Cause 1				Cause 2			
n_i	$10 \times$ |Bias|	ASE	SSE	CP	$10 \times$ |Bias|	ASE	SSE	CP
50	0.032	0.058	0.068	0.880	0.038	0.069	0.067	0.997
	0.061	0.056	0.079	0.931	0.040	0.049	0.075	0.929
	0.028	0.067	0.080	0.957	0.015	0.060	0.078	0.937
	0.014	0.073	0.074	0.963	0.007	0.068	0.076	0.969
	0.015	0.075	0.074	0.963	0.132	0.073	0.079	0.967
	p_1 : |Bias| = 0.0009 , ASE=0.046 , SSE=0.045 , CP=0.918							
	p_2 : |Bias| = 0.0017 , ASE=0.068 , SSE=0.068 , CP=0.931							
150	0.003	0.020	0.020	0.923	0.008	0.016	0.016	0.919
	0.008	0.030	0.030	0.936	0.008	0.025	0.025	0.929
	0.005	0.038	0.038	0.945	0.012	0.034	0.035	0.957
	0.013	0.042	0.041	0.951	0.006	0.039	0.038	0.947
	0.014	0.042	0.041	0.957	0.020	0.041	0.040	0.960
	p_1 : |Bias| = 0.0002 , ASE=0.028 , SSE= 0.028, CP=0.926							
	p_2 : |Bias| = 0.0009 , ASE=0.042 , SSE= 0.042, CP=0.937							
250	0.001	0.015	0.015	0.941	0.003	0.013	0.013	0.924
	0.004	0.023	0.023	0.938	0.006	0.020	0.020	0.938
	0.002	0.030	0.030	0.953	0.009	0.026	0.026	0.949
	0.008	0.033	0.033	0.949	0.037	0.030	0.030	0.952
	0.007	0.033	0.032	0.955	0.009	0.032	0.031	0.958
	p_1 : |Bias| = 0.0001 , ASE=0.022 , SSE= 0.022 , CP=0.936							
	p_2 : |Bias| = 0.0006 , ASE=0.032 , SSE=0.032 , CP=0.952							

Table 9: Simulation results on $\left\{\hat{F}_j(\tau_i), j = 1, 2\right\}$ and $(\hat{p}_1, \hat{p}_2)$ for exponential(1) failure time distribution with 6 : 4 rate ratio for the two failure types and $(p_1, p_2) = (0.9, 0.8)$ in presence of validation sample (f=0.5). The five τ_i's are 0.105, 0.288, 0.693, 1.386 and 2.303.

	Cause 1				Cause 2			
n_i	$10 \times$ |Bias|	ASE	SSE	CP	$10 \times$ |Bias|	ASE	SSE	CP
50	0.049	0.057	0.068	0.870	0.029	0.069	0.066	0.988
	0.037	0.054	0.073	0.940	0.035	0.049	0.070	0.930
	0.029	0.067	0.078	0.936	0.013	0.059	0.073	0.932
	0.012	0.072	0.073	0.965	0.012	0.067	0.072	0.968
	0.030	0.074	0.072	0.972	0.089	0.072	0.075	0.972
	p_1 : |Bias| = 0.0007 , ASE = 0.041 , SSE= 0.042 , CP = 0.916							
	p_2 : |Bias| = 0.0015 , ASE = 0.063, SSE = 0.065 , CP = 0.934							
150	0.004	0.020	0.019	0.926	0.003	0.016	0.017	0.916
	0.005	0.030	0.029	0.941	0.006	0.025	0.025	0.932
	0.009	0.038	0.039	0.942	0.016	0.033	0.033	0.944
	0.001	0.042	0.040	0.961	0.005	0.038	0.036	0.961
	0.016	0.042	0.040	0.966	0.016	0.041	0.038	0.966
	p_1 : |Bias| = 0.0003, ASE = 0.024, SSE = 0.023 , CP = 0.921							
	p_2 : |Bias| = 0.0005 , ASE = 0.037, SSE = 0.038 , CP = 0.948							
250	2.64×10^{-4}	0.015	0.015	0.933	0.001	0.013	0.013	0.918
	0.004	0.023	0.022	0.945	0.003	0.019	0.019	0.946
	0.005	0.029	0.029	0.952	0.004	0.026	0.026	0.952
	7.54×10^{-4}	0.032	0.032	0.950	5.56×10^{-5}	0.029	0.029	0.949
	0.013	0.033	0.033	0.944	0.004	0.031	0.031	0.948
	p_1 : |Bias| = 0.002 , ASE = 0.019, SSE = 0.019 , CP = 0.946							
	p_2 : |Bias| = 0.0002 , ASE = 0.029 , SSE = 0.029 , CP = 0.951							

Table 10: Simulation results on $\left\{\hat{F}_j(\tau_i), j = 1, 2\right\}$ and $(\hat{p}_1, \hat{p}_2)$ for exponential(1) failure time distribution with 6: 4 rate ratio for the two failure types and true $(q_1, q_2) = (0.429, 0.571)$. The five τ_i's are 0.105, 0.288, 0.693, 1.386 and 2.303.

	Cause 1				Cause 2			
n_i	10 × |Bias|	ASE	SSE	CP	10 × |Bias|	ASE	SSE	CP
50	0.006	0.018	0.033	0.891	0.022	0.017	0.027	0.831
	0.017	0.032	0.051	0.843	0.008	0.031	0.040	0.844
	0.018	0.044	0.062	0.812	0.053	0.042	0.050	0.859
	0.015	0.046	0.062	0.836	0.036	0.046	0.055	0.863
	0.028	0.051	0.058	0.895	0.051	0.052	0.058	0.906
	p_1 : |Bias × 10| = 0.006, ASE = 0.024, SSE = 0.025, CP = 0.937							
	p_2 : |Bias × 10| = 0.014, ASE = 0.046, SSE = 0.047, CP = 0.936							
150	0.002	0.013	0.020	0.901	0.006	0.013	0.017	0.924
	0.007	0.021	0.029	0.922	0.005	0.020	0.025	0.959
	0.005	0.028	0.037	0.848	0.017	0.028	0.032	0.889
	0.006	0.031	0.038	0.875	0.025	0.030	0.034	0.900
	0.024	0.034	0.038	0.900	0.023	0.034	0.038	0.907
	p_1 : |Bias × 10| = 0.006, ASE = 0.014, SSE = 0.013, CP = 0.960							
	p_2 : |Bias × 10| = 0.005, ASE = 0.026, SSE = 0.026, CP = 0.953							
250	0.001	0.010	0.016	0.961	0.005	0.010	0.013	0.954
	0.003	0.016	0.022	0.926	0.003	0.016	0.019	0.954
	0.003	0.023	0.030	0.957	0.011	0.022	0.026	0.920
	0.005	0.025	0.030	0.920	0.004	0.025	0.028	0.935
	0.023	0.027	0.031	0.934	0.021	0.027	0.030	0.938
	p_1 : |Bias × 10| = 0.006, ASE = 0.011, SSE = 0.011, CP = 0.944							
	p_2 : |Bias × 10| = 0.001, ASE = 0.020, SSE = 0.020, CP = 0.948							

Table 11: Simulation results on $\left\{\hat{F}_j(\tau_i), j = 1, 2\right\}$ and $(\hat{p}_1, \hat{p}_2)$ for exponential(1) failure time distribution with 6: 4 rate ratio for the two failure types and wrong $(q_1, q_2) = (0.3, 0.7)$. The five τ_i's are 0.105, 0.288, 0.693, 1.386 and 2.303.

	Cause 1				Cause 2			
n_i	10 × |Bias|	ASE	SSE	CP	10 × |Bias|	ASE	SSE	CP
50	0.012	0.018	0.032	0.666	0.001	0.017	0.030	0.710
	0.053	0.031	0.047	0.779	0.031	0.031	0.041	0.824
	0.085	0.044	0.062	0.814	0.059	0.043	0.056	0.842
	0.149	0.046	0.062	0.828	0.101	0.047	0.056	0.878
	0.128	0.052	0.058	0.900	0.166	0.052	0.057	0.902
	p_1 : |Bias| = 0.029 , ASE = 0.018, SSE = 0.017, CP = 0.587							
	p_2 : |Bias| = 0.033 , ASE = 0.051, SSE = 0.052, CP = 0.916							
150	0.011	0.013	0.021	0.743	0.019	0.013	0.017	0.844
	0.024	0.020	0.029	0.819	0.044	0.020	0.026	0.856
	0.098	0.028	0.038	0.826	0.074	0.028	0.033	0.886
	0.014	0.031	0.037	0.866	0.132	0.031	0.035	0.875
	0.017	0.033	0.039	0.867	0.171	0.033	0.036	0.891
	p_1 : |Bias| = 0.028 , ASE = 0.010, SSE = 0.010, CP = 0.252							
	p_2 : |Bias| = 0.035 , ASE = 0.030, SSE = 0.030, CP = 0.820							
250	0.020	0.010	0.015	0.792	0.021	0.010	0.013	0.857
	0.047	0.016	0.023	0.824	0.049	0.016	0.019	0.879
	0.084	0.022	0.029	0.853	0.085	0.023	0.025	0.879
	0.137	0.025	0.030	0.860	0.122	0.025	0.029	0.860
	0.175	0.027	0.032	0.851	0.173	0.027	0.030	0.878
	p_1 : |Bias| = 0.028 , ASE = 0.008, SSE = 0.008, CP = 0.094							
	p_2 : |Bias| = 0.034 , ASE = 0.023, SSE = 0.023, CP = 0.701							

Table 12: Simulation results on $\left\{\hat{F}_j(\tau_i), j = 1, 2\right\}$ and $(\hat{p}_1, \hat{p}_2)$ for exponential(1) failure time distribution with 6: 4 rate ratio for the two failure types and wrong $(q_1, q_2) = (0.6, 0.4)$. The five τ_i's are 0.105, 0.288, 0.693, 1.386 and 2.303.

	Cause 1				Cause 2			
n_i	10 × |Bias|	ASE	SSE	CP	10 × |Bias|	ASE	SSE	CP
50	0.032	0.018	0.033	0.686	0.034	0.017	0.030	0.757
	0.050	0.032	0.049	0.751	0.041	0.030	0.047	0.810
	0.155	0.043	0.065	0.784	0.159	0.041	0.063	0.850
	0.232	0.045	0.061	0.791	0.255	0.045	0.060	0.838
	0.177	0.051	0.062	0.862	0.160	0.051	0.060	0.876
	p_1 : |Bias| = 0.035 , ASE = 0.030, SSE = 0.031, CP = 0.845							
	p_2 : |Bias| = 0.048 , ASE = 0.037, SSE = 0.037, CP = 0.674							
150	0.035	0.013	0.021	0.760	0.028	0.012	0.016	0.827
	0.058	0.020	0.031	0.783	0.050	0.020	0.025	0.848
	0.123	0.028	0.040	0.802	0.121	0.027	0.031	0.870
	0.183	0.030	0.039	0.807	0.176	0.030	0.035	0.863
	0.195	0.033	0.038	0.867	0.190	0.033	0.035	0.875
	p_1 : |Bias| = 0.034 , ASE = 0.018, SSE = 0.018, CP = 0.527							
	p_2 : |Bias| = 0.051 , ASE = 0.021, SSE = 0.021, CP = 0.328							
250	0.029	0.010	0.016	0.755	0.028	0.010	0.012	0.845
	0.062	0.016	0.023	0.804	0.056	0.016	0.020	0.851
	0.126	0.022	0.030	0.830	0.129	0.022	0.026	0.848
	0.193	0.025	0.031	0.802	0.196	0.024	0.028	0.825
	0.211	0.027	0.031	0.821	0.220	0.027	0.029	0.825
	p_1 : |Bias| = 0.035, ASE = 0.014, SSE = 0.014, CP = 0.281							
	p_2 : |Bias| = 0.050, ASE = 0.016, SSE = 0.016, CP = 0.182							

6 Data Analysis

Let us consider the hearing loss data, described in Koley and Dewanji (2018b). Since we do not have any validation sample in this example, we consider only the method of Section 3.2 for illustration using additional information on q_1 and q_2. We consider different choices of q_1 and q_2 as $(q_1 = 0.02, q_2 = 0.98)$, $(q_1 = 0.1, q_2 = 0.9)$, $(q_1 = 0.3, q_2 = 0.7)$, $(q_1 = 0.5, q_2 = 0.5)$, $(q_1 = 0.7, q_2 = 0.3)$, $(q_1 = 0.9, q_2 = 0.1)$ and $(q_1 = 0.98, q_2 = 0.02)$. To compute the standard error of all the estimates, we carried out the bootstrap procedure with B =1000 replications. First we fit exponential and then Weibull distribution to the data. We perform a goodness of fit test based on a modified χ^2 statistics, χ^2_M (say), as described in Koley and Dewanji (2018a) with the corresponding p-values obtained similarly via simulation (See Hope (1968)). Also, AIC values (See Burnham and Anderson (2003)) of both the models are computed to compare them for fitting the given data. The results are presented in Table 13.

In both the models, we observe that, as q_j values increases, the corresponding estimate of λ_j increases. This is possibly because, with the increase in q_j, more of the missing causes are assigned to cause j, for $j = 1, 2$. Also, as expected from (3.3), the p_j estimates decrease with increase in the corresponding q_j values. Interestingly, the standard errors are seen to increase with q_1 for all the parameter estimates, possibly because the true value of q_1 may be small. The p-values seem to indicate that the Weibull model fits the data much better at least in comparison with the exponential model. This result is also supported by the AIC values corresponding to both the models. Also, by looking at the p-values and also the AIC values, the choice of $q_1 = 0.02,\ q_2 = 0.98$ seems to give the best fit out of all the choices. This explains why the standard errors increase with increasing q_1.

For the wide range of choices for (q_1, q_2) values, the estimates of λ_1 are much larger than those of λ_2 for both exponential and Weibull models. Therefore, the incidence of SNHL seems to be more frequent than conductive hearing loss. Also, from the estimate of the shape parameter in the Weibull model ($\hat{b} = 0.688$), the incidence of both the types seems to decrease with age.

For non-parametric estimation using the method of Section 4.2, we split the range of monitoring times into five sub-intervals as $(0, 10], (10, 30], (30, 50], (50, 70]$ and (70,90] for ease of illustration. We consider different choices of (q_1, q_2) as $(0.02, 0.98)$, $(0.1, 0.9)$, (0.3,0.7), (0.5,0.5), (0.7,0.3), (0.9,0.1) and (0.98,0.02) and carry out the analysis with the results being presented in Table 14.

We note that a decrease in the q_j value means less missing data for cause j, as a result of which estimate of p_j increases, as seen in Table 14, for $j = 1, 2$. Also, as argued in the parametric analysis, the estimates of $F_j(\tau_i)$'s are found to increase with q_j value, for $j = 1, 2$.

To determine the choice of (q_1, q_2) that best fits the data, we compute the corresponding AIC values. From Table 14, we see that the AIC value is the least for $(q_1, q_2) = (0.02, 0.98)$, thus fitting the data best out of all other choices, as in the parametric analysis.

Table 13: Parametric analysis of the Hearing Loss Data using the method of Section 3.2 with additional information on q_1 and q_2. The standard errors are given in parentheses.

Model	Choice of (q_1, q_2)	MLE$\times 10^2$ with standard error $\times 10^2$					χ^2_M	p-value	AIC
		λ_1	λ_2	b^+	p_1^+	p_2^+			
Exponential	(0.02,0.98)	7.869	1.635	-	0.997	0.349	5316.452	0	1417.274
		(0.485)	(0.183)	-	(0.0004)	(0.0520)			
	(0.1,0.9)	7.956	1.548	-	0.986	0.368	5316.048	0	1417.274
		(0.513)	(0.181)	-	(0.002)	(0.053)			
	(0.3,0.7)	8.172	1.334	-	0.960	0.429	5311.797	0	1417.274
		(0.496)	(0.155)	-	(0.006)	(0.056)			
	(0.5,0.5)	8.389	1.112	-	0.935	0.512	5312.119	0	1417.274
		(0.482)	(0.127)	-	(0.009)	(0.054)			
	(0.7,0.3)	8.625	0.902	-	0.912	0.636	5317.199	0	1417.274
		(0.513)	(0.118)	-	(0.011)	(0.055)			
	(0.9,0.1)	8.827	0.683	-	0.889	0.840	5324.913	0	1417.276
		(0.541)	(0.110)	-	(0.014)	(0.031)			
	(0.98,0.02)	8.912	0.597	-	0.881	0.963	5322.511	0	1417.277
		(0.520)	(0.105)	-	(0.015)	(0.008)			
Weibull	(0.02,0.98)	8.209	0.837	0.688	0.997	0.349	2251.914	0.82	1377.863
		(3.23×10^{-3})	(6.34×10^{-4})	(2.15×10^{-4})	(3.74×10^{-4})	(4.71×10^{-2})			
	(0.1,0.9)	8.341	0.780	0.690	0.986	0.368	2256.536	0.756	1377.865
		(1.65×10^{-2})	(9.13×10^{-3})	(4.571×10^{-4})	(1.98×10^{-3})	(5.05×10^{-2})			
	(0.3,0.7)	8.665	0.630	0.690	0.960	0.428	2254.683	0.788	1377.867
		(0.046)	(0.023)	(0.0001)	(0.005)	(0.054)			
	(0.5,0.5)	8.999	0.490	0.691	0.935	0.512	2254.376	0.776	1377.871
		(0.055)	(0.027)	(0.0002)	(0.009)	(0.052)			
	(0.7,0.3)	9.770	0.254	0.618	0.912	0.636	2134.997	0.736	1380.016
		(0.062)	(0.027)	(0.0002)	(0.011)	(0.053)			
	(0.9,0.1)	10.152	0.173	0.617	0.889	0.840	2106.644	0.158	1380.229
		(0.113)	(0.027)	(0.021)	(0.013)	(0.043)			
	(0.98,0.02)	10.343	0.141	0.615	0.881	0.963	2098.757	0.054	1380.438
		(2.56×10^{-7})	(2.60×10^{-3})	(1.85×10^{-8})	(1.65×10^{-2})	(8.89×10^{-3})			

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

Table 14: Non-parametric analysis of Hearing loss data under fixed monitoring time using the method of Section 4.2 with additional information on q_1 and q_2. The standard errors are given in parentheses.

Type of Analysis		MLE with standard error							AIC
		Monitoring time (τ)							
		10	30	50	70	90	p_1	p_2	
$q_1 = 0.02, q_2 = 0.98$	F_1	0.436	0.678	0.678	0.678	0.740	0.997	0.349	1140.817
		(0.023)	(0.024)	(0.024)	(0.024)	(0.025)	(0.004×10^{-1})	(0.049)	
	F_2	0.033	0.077	0.250	0.260	0.260			
		(0.009)	(0.024)	(0.027)	(0.025)	(0.025)			
$q_1 = 0.1, q_2 = 0.9$	F_1	0.439	0.686	0.686	0.686	0.748	0.986	0.368	1142.64
		(0.025)	(0.024)	(0.024)	(0.024)	(0.024)	(0.002)	(0.052)	
	F_2	0.030	0.072	0.243	0.252	0.252			
		(0.009)	(0.023)	(0.026)	(0.024)	(0.024)			
$q_1 = 0.3, q_2 = 0.7$	F_1	0.445	0.704	0.704	0.704	0.768	0.960	0.428	1147.896
		(0.024)	(0.022)	(0.022)	(0.022)	(0.023)	(0.005)	(0.055)	
	F_2	0.025	0.061	0.225	0.232	0.232			
		(0.008)	(0.021)	(0.024)	(0.023)	(0.023)			
$q_1 = 0.5, q_2 = 0.5$	F_1	0.448	0.724	0.724	0.724	0.792	0.935	0.512	1152.666
		(0.025)	(0.022)	(0.022)	(0.022)	(0.023)	(0.008)	(0.050)	
	F_2	0.022	0.051	0.205	0.209	0.209			
		(0.008)	(0.019)	(0.025)	(0.023)	(0.023)			
$q_1 = 0.7, q_2 = 0.3$	F_1	0.450	0.750	0.750	0.750	0.822	0.912	0.636	1153.928
		(0.024)	(0.023)	(0.023)	(0.023)	(0.024)	(0.011)	(0.050)	
	F_2	0.020	0.043	0.178	0.178	0.178			
		(0.007)	(0.018)	(0.022)	(0.021)	(0.021)			
$q_1 = 0.9, q_2 = 0.1$	F_1	0.451	0.790	0.790	0.790	0.870	0.889	0.834	1154.468
		(0.024)	(0.022)	(0.022)	(0.024)	(0.019)	(0.014)	(0.029)	
	F_2	0.018	0.033	0.130	0.130	0.130			
		(0.007)	(0.013)	(0.019)	(0.019)	(0.019)			
$q_1 = 0.98, q_2 = 0.02$	F_1	0.452	0.809	0.810	0.812	0.897	0.880	0.963	1454.364
		(0.023)	(0.024)	(0.024)	(0.026)	(0.018)	(0.014)	(0.008)	
	F_2	0.017	0.029	0.102	0.102	0.102			
		(0.006)	(0.012)	(0.018)	(0.018)	(0.018)			

7 Concluding Remarks

In this work, we have focused on both parametric and non-parametric estimation methods that use the additional information available either through validation sampling or through some prior/subject-matter indirect information on the masking probabilities. Also, the masking probabilities are assumed to be time independent, which may be generalized to include time dependence requiring more modeling and data. We have worked with two competing risks; one may extend this to more than two competing risks. The monitoring time random variable is assumed to be independent of the random vector (T, G); it would be interesting to study the dependent case in the context of current status data with competing risks and missing failure types. Finally, in the validation sample approach, there is possibility of recalling the time of failure T, possibly with some imperfection, which will certainly benefit the precision of the estimates. The corresponding estimation procedure will involve further modeling. Relaxation of the assumption that there is no misspecification of possible causes may also be of interest, but will require a different kind of modeling and analysis. One can also consider presence of other types of additional information different from the two discussed in this work. It may be of interest to incorporate covariates and perform regression analysis of the data.

In non-parametric estimation we consider fixed and finite number of monitoring times. The proposed estimation technique along with the asymptotic results can be extended to the case when the n_i's are random with n_i/n converging to w_i in probability as $n \to \infty$, for $i = 1, \ldots, K$. Re-parametrization is considered to evade the monotonic constraint property of the sub-distribution functions, converting the optimization problem into a simpler one. The method utilizing indirect information on masking probabilities can be extended to the case of random monitoring time, since after estimating the masking probabilities, the problem becomes similar to the one discussed in Koley and Dewanji (2019). In case of validation sample approach, generalization to random monitoring time is a challenge that we intend to take up as a future work.

Acknowledgement

We would 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

Burnham, K. P. and Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Cox, C. (2005). Delta method. *Encyclopedia of biostatistics*, 2.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika*, 79(4):855–857.

Dewanji, A. and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics*, 59(4):1063–1070.

Flehinger, 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.

Goetghebeur, E. and Ryan, L. (1995). Analysis of competingrisks survival data when some failure types are missing. *Biometrika*, 82:821–833.

Hope, A. C. (1968). A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 582–598.

Hudgens, M. G., Satten, G. A., and Longini Jr, I. M. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*, 57(1):74–80.

Jewell, N. P. and Kalbfleisch, J. D. (2004). Maximum likelihood estimation of ordered multinomial parameters. *Biostatistics*, 5(2):291–306.

Jewell, N. P., Van Der Laan, M., and Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika*, 90(1):183–197.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data, second edition*. John Wiley & Sons.

Koley, T. and Dewanji, A. (2018a). Current status data with two competing risks and missing failure types. `https://www.isical.ac.in/~asu/TR/TechRepASU201802.pdf`.

Koley, T. and Dewanji, A. (2018b). Revisiting non-parametric maximum likelihood estimation of current status data with competing risks. *Sankhya B*.

Koley, T. and Dewanji, A. (2019). Current status data with two competing risks and missing failure types: A non-parametric approach. https://www.isical.ac.in/~asu/TR/TechRepASU201901.pdf.

Lehmann, E. C. and Casella, G. (1998). G.(1998). theory of point estimation.

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

Lu, K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, 57(4):1191–1197.

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

Malik, S. C. and Arora, S. (1992). *Mathematical analysis*. New Age International.

Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *IEEE TRANSACTIONS ON RELIABILITY*, 33(4):293–296.

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*, pages 17–33.

Racine-Poon, A. H. and Hoel, D. G. (1984). Nonparametric estimation of the survival function when cause of death is uncertain. *Biometrics*, 40(4):1151–1158.

Trotter, H. (1959). An elementary proof of the central limit theorem. *Archiv der Mathematik*, 10(1):226–234.