Ann Inst Stat Math (2012) 64:69–85
DOI 10.1007/s10463-010-0303-8

Random partition masking model for censored and masked competing risks data

Qiqing Yu · G. Y. C. Wong · Hao Qin · Jiaping Wang

Received: 13 May 2008 / Revised: 21 October 2009 / Published online: 22 May 2010

Abstract We consider the parametric estimation with right-censored competing risks data and with masked failure cause. We propose a new model, called the random partition masking (RPM) model. The existing model based on the so called symmetry assumption, but the RPM model does not need the symmetry assumption. We propose a wide class of parametric distribution families of the failure time and cause, which does not need the assumption of independence between the components of the system. We also study the asymptotic properties of the maximum likelihood estimator under the new model, and apply our procedure to a medical and an industrial data sets.

Keywords Right-censorship · Competing risks model · MLE · Consistency

1 Introduction

We consider the estimation problem based on right-censored (RC) competing risks data with masked failure cause, called RMCR data hereafter. The background about RMCR data is introduced in Sect. 1.1, the existing models and assumptions in the literature are given in Sect. 1.2, and finally our objects of the paper are stated in Sect. 1.3.

Q. Yu (✉) · H. Qin · J. Wang
Department of Mathematical Sciences, SUNY,
Binghamton, NY 13902, USA
e-mail: qyu@math.binghamton.edu

G. Y. C. Wong
Strang Cancer Prevention Center, 428 E 72nd Street,
New York, NY 10021, USA

1.1 Background on RMCR data

The study of RMCR data dates back to Friedman and Gertsbakh (1980). Real examples of RMCR data in the reliability context and medical research can be found in Dinse (1982), Reiser et al. (1995), and Flehinger et al. (2001). Sen et al. (2001) and Flehinger et al. (2001). An extensive review of the literature is given by Mukhopadhyay (2006).

An RMCR observation consists of the failure time and the associated failure cause of a J-component series system, a system that stops functioning as soon as one of its constituent J components fails. We assume that the system under study is non-repairable and the system lifetime is subject to censoring. Then the observation on the failure time T and failure cause C of such systems can be described as follows. Let the random variable X_j denote the lifetime of the jth component, $j = 1, \ldots, J$. By the definition of a series system, it follows that $T = \min\{X_1, \ldots, X_J\}$. It is assumed that the probability of a system failure due to simultaneous failures of two or more distinct components is 0 (which is true when X_j's are continuous and independent), and thus there exists a unique positive integer $C \in \{1, \ldots, J\}$ associated with each system failure time T, say $X_C = T$.

In engineering applications, T could be the failure time of a system and C the part of the system that causes the failure, and in such cases, $X_1, \ldots, X_J$ can be independent. In medical research, T could be the survival time until death of a patient and C the cause of death, and in such cases, $X_1, \ldots, X_J$ are likely to be dependent. It is quite often in medical researches and industry applications that the failure time is right censored by a censoring variable R. Let V be the minimum of T and R, and δ ($\stackrel{def}{=} \mathbf{1}_{(T \le R)}$) the indicator function of the event $\{T \le R\}$.

Let $\mathcal{C}_r$ be the range of C, that is, $\mathcal{C}_r = \{1, \ldots, J\}$, and let $\mathcal{J}$ be the collection of all the subsets of $\mathcal{C}_r$, including $\mathcal{C}_r$ but not the empty set $\emptyset$. In examining a failed system, one may first check parts one by one in detecting the failure cause and may stop at some point due to cost saving if it makes no sense economically to continue. In this case, one may end up with only knowing which ones are not the cause of failure. Thus one can say that at the failure time T the failure cause C is masked by $\mathcal{M}$, a subset of $\mathcal{C}_r$. If $T \le R$, $\mathcal{M}$ is the observation on C. Notice that as in most papers on RMCR data, we assume that there is no hope of acquiring any information about its future cause of failure for an RC observation. Moreover, we do not consider the case that there are stage-2 data available. That is, we assume that there is no re-examination on a sub-sample of the original n observations for additional data on the true failure causes, which is the case in Dinse (1986).

1.2 The current models

Let $F_{T,C}(t, c)$ be the cumulative distribution function (cdf) of (T, C), i.e. $F_{T,C}(t, c) = P\{T \le t, C \le c\}$, and denote by $f_{T,C}$ its density function (df) and by $S_T(t) = P(T > t)$ the survival distribution function of T. Denote in an obvious way the "df"s $f_{\mathcal{M}|T,C}(A|t, c) = P\{\mathcal{M} = A|T = t, C = c\}$, $f_{\mathcal{M}|C}(A|c) = P\{\mathcal{M} = A|C = c\}$, and the "cdf" $F_{T,C,\mathcal{M},R}$, though $\mathcal{M}$ is a random set, not a random variable.

As pointed out by Craiu and Reiser (2006, p.22[21]), almost all of the research concerned with masking makes the following two conditions, S1 and S2:

S1 (S1a) $\forall\, A$ in $\mathcal{J}$, $\nu_c(A)$ is constant in c pertaining to A, where $\nu_c(A) \stackrel{def}{=} f_{\mathcal{M}|C}(A|c)$.
(S1b) $\forall\, A$ in $\mathcal{J}$ and $\forall\, c, t > 0$, $\nu_{t,c}(A) = \nu_c(A)$, where $\nu_{t,c}(A) \stackrel{def}{=} f_{\mathcal{M}|T,C}(A|t, c)$.
S2 $(\mathcal{M}, T)$ and R are independent $((\mathcal{M}, T) \perp R)$ (see Mukhopadhyay 2006, p. 806[14]).

S1 is called the symmetry assumption (see Flehinger et al. 1996). Moreover, it is also assumed (see Mukhopadhyay 2006, p. 810[1−8]) the condition S3 below.

S3 $(T_1, C_1, \mathcal{M}_1), \ldots, (T_{n_1}, C_{n_1}, \mathcal{M}_{n_1})$, $T_{n_1+1}, \ldots, T_n$ are independent. $T_1, \ldots, T_n$ are i.i.d. copies of T, $\delta_1 = \cdots = \delta_{n_1} = 1$ and $\delta_{n_1+1} = \cdots = \delta_n = 0$.

By S3, the likelihood of the observed RMCR data is given by (1) below.

$$\left(\prod_{i=1}^{n_1}\left\{\int_{t=V_i\le u,\ c\in\mathcal{M}_i} \mathrm{d}F_{T,C,\mathcal{M},R}(t,c,\mathcal{M}_i,u)\right\}\right)\prod_{i>n_1}^{n} P\{T > V_i = R\}, \tag{1}$$

$$= \left(\prod_{i=1}^{n_1}\left\{\sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C,\mathcal{M}}(t,c,\mathcal{M}_i)S_R(V_i-)\right\}\right)\prod_{i>n_1}^{n} S_T(V_i)f_R(V_i) \quad \text{(by S2)}$$

$$\propto \left(\prod_{i=1}^{n_1}\left\{\sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C}(t,c)\nu_{t,c}(\mathcal{M}_i)\right\}\right)\prod_{i>n_1}^{n} S_T(V_i) \quad (\stackrel{def}{=}\Lambda)$$

$$= \left(\prod_{i=1}^{n_1}\left\{\sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C}(t,c)\nu_c(\mathcal{M}_i)\right\}\right)\prod_{i>n_1}^{n} S_T(V_i) \quad (\stackrel{def}{=}\Lambda_c) \quad \text{(by S1b)}$$

$$\propto \left(\prod_{i=1}^{n_1}\left\{\sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C}(t,c)\right\}\right)\prod_{i>n_1}^{n} S_T(V_i) \quad (\stackrel{def}{=}\mathcal{L}_c) \quad \text{(by S1a)}$$

(see, e.g., Flehinger et al. 2001, p. 502–504 and Sen et al. 2001, p.525[4] for more details on the derivation above).

The $\mathcal{L}_c$ is a desirable likelihood function, and together with S1, S2, S3 and (1), it forms a model for RMCR data, called the Conditional Masking Probability (CMP) model as it is based on $f_{\mathcal{M}|T,C}$, even though it vanishes in $\mathcal{L}_c$. In the paper we will refer to it as CMP Model 1. It has been argued that the symmetry assumption is misleading (see Lin and Guess 1994; Guttman et al. 1995). Thus some people try to make MLE or Bayesian inferences based on the likelihood Λ_c in (1) assuming only S2, S3 and S1b without S1a (see, e.g., Flehinger et al. 2001, p.502–504; Mukhopadhyay and Basu 2007, p. 333; Kuo and Yang 2000; Craiu and Duchesne 2004; Lawless 2003, and Craiu and Reiser 2006). Likelihoods (1) and Λ_c, together with the assumptions S1b, S2 and S3 actually form the second model for RMCR data. Hereafter, we will refer to it as CMP Model 2.

1.3 The objects of the paper

There are two objects about RMCR data in this paper: (1) to propose a realistic model; (2) to propose more general and convenient parametric families.

In the literature, there are two CMP models (see Sect. 1.2). CMP model 2 is not that useful unless there are stage-2 data or prior information (for justification, see Example 11), especially in the case of a non-parametric approach. However, CMP Model 1 is more useful. In the literature in order to justify CMP Model 1 or the likelihood $\mathcal{L}_c$, people make use of the symmetry assumption, which "is done purely for mathematical convenience without practical justification" (see Mukhopadhyay and Basu 2007, p.331[15]). In this paper, we propose a new realistic RMCR model, called the random partition masking (RPM) model, which does not need the symmetry assumption nor S2 to justify $\mathcal{L}_c$. Moreover, we show that under certain additional assumptions, the RPM model satisfies the symmetry assumption. Thus our new model gives a practical justification to CMP Model 1 for the first time.

In the literature, many distribution families of (T, C) are based on the assumption S4 below (see e.g., Flehinger et al. 2001; Nagai 2004).

S4 $X_1, \ldots, X_J$ are independent and X_j's are continuous.

In medical research S4 often fails. Thus this approach does not always work. Since $F_{T,C}$ is of main interest, it is better to consider the family of $F_{T,C}$ directly, though as noticed by Kalbfleisch and Prentice (2002) a parametric model for the dependency is hard to specify. Craiu and Duchesne (2004), Lawless (2003), and Craiu and Reiser (2006) propose a special parametric model in such case, where the cause specific hazard functions $\lambda_c(t)(\stackrel{def}{=} f_{T,C}(t, c)/S_T(t-))$ are piecewise constant. We propose a more general and more convenient way to specify parametric models and apply them to analyze some real data.

The paper is organized as follows. In Sect. 2, we propose the RPM model. In Sect. 3, we propose a general form of parametric families for $F_{T,C}$ and study the MLE of the parameters involved in the parametric form of $F_{T,C}$. In Sect. 4, we establish the asymptotic properties of the MLE. In Sect. 5, we present some simulation results. Data analysis of real data sets using the new parametric form is presented in Sect. 6. In Sect. 7 we compare the RPM model to the CMP models.

2 A new model for RMCR data

In order to formulate the RPM model, we first define some notations. Notice that each value W of $\mathcal{M}$ is associated with at least one partition $\{W, W^c\}$, where $W^c = \mathcal{C}_r \setminus W$. Let $\mathcal{P}$ be the collection of all partitions of $\mathcal{C}_r$ that satisfy the following conditions: $P_h \in \mathcal{P}$ implies that $P_h = \{P_{h1}, \ldots, P_{hk_h}\}$, $P_{hi} \in \mathcal{J}$, $\cup_{i=1}^{kh} P_{hi} = \mathcal{C}_r$ and $P_{hi} \cap P_{hj} = \emptyset$ $\forall i \neq j$. By definition, for each given partition P_h and given C, there exists an i such that $C \in P_{hi}$. For instance, $P_1 = \{\{1\}, \{2\}, \ldots, \{J\}\}$, $P_2 = \{\{1\}, \{2\}, \{3, 4, \ldots, J\}\}$ and $P_3 = \{\mathcal{C}_r\}$ are three such partitions.

The P_2 can be interpreted as follows: In the process of determining the cause of failure in a J-component series system, exactly two steps will be taken. Steps 1 and 2

can determine whether the failure is due to causes 1 and 2, respectively. If the failure is not due to these two causes, no further investigation will be taken for cost saving. However, it is only one of the six examination schemes corresponding to P_2 and each has two steps. The first step can be either of the three inspections:

(1) whether the cause is due to part 1;
(2) whether the cause is due to part 2;
(3) whether the cause is not due to parts 1 and 2.

The second step can be either of the 2 remaining inspections. Thus P_2 corresponds to total of 6 examination schemes. All the 6 of them result in $\mathcal{M} = \{1\}$ if $C = 1$, $\{2\}$ if $C = 2$, $\{3, 4, \ldots, J\}$, otherwise.

In other words, an inspection scheme for the system corresponds to a partition. After an inspection scheme is chosen, that is, after a partition $(P_{h1}, \ldots, P_{hk_h})$ is chosen, $\mathcal{M}$ can be uniquely determined. Notice that an inspection or partition P_h may not simply be an examination procedure, but include information obtained from the description of the symptoms of the failed system from the user or the symptoms of the patient collected in a check list filled by the user or the patient. Moreover, for a particular observation, if the failure cause is detected at the first step, there is no need to continue the inspection scheme.

It is obvious that $||\mathcal{P}||$, the number of all distinct partitions denoted by $n_{\mathcal{P}}$, is finite. Thus one can order these partitions as $P_1, P_2, \ldots, P_{n_{\mathcal{P}}}$. It is easy to define a random variable, say Δ, taking values in $\{1, \ldots, n_{\mathcal{P}}\}$ with the df f_Δ. The value of $f_\Delta(h)$ can be viewed as the proportion in the population that inspection scheme h has been taken. Then $\mathcal{M} = P_{hj}$ if $\Delta = h$ and $C \in P_{hj}$. We shall make use of the assumption A1 below,

A1 $(T, C) \perp (R, \Delta)$.

While (V, δ) is the observation on T, the observation on C is not defined in the literature if $T > R$, as $\mathcal{M}$ is missing if $T > R$. We define the observable random variable on C by

$$\mathcal{M}^o = \begin{cases} \mathcal{C}_r & \text{if } T > R, \\ P_{hj} & \text{if } T \le R, \Delta = h \text{ and } C \in P_{hj}. \end{cases} \tag{2}$$

$$\text{Let } (V_i, \delta_i, \mathcal{M}_i), i = 1, \ldots, n, \text{ be i.i.d. copies of } (V, \delta, \mathcal{M}^o). \tag{3}$$

Statements (2) and (3) together with A1 are the new model that we are proposing for analyzing RMCR data. We call it the random partition masking model, as it is based on the random partition of the set $\mathcal{C}_r$. Under the RPM model, the full likelihood function is

$$\mathcal{L}_{\text{full}} = \prod_{i=1}^{n} \left(\left\{ \sum_{h=1}^{n_{\mathcal{P}}} f_\Delta(h) \mathbf{1}_{(\mathcal{M}_i \in \mathrm{P}_h)} S_{R|\Delta}(V_i - |h) \sum_{t=V_i, c \in \mathcal{M}_i} f_{T,C}(t, c) \right\}^{\delta_i} \times [f_R(V_i) S_T(V_i)]^{1-\delta_i} \right)$$

$$= \mathcal{L}(F_{T,C}) \prod_{i=1}^{n} \left(\left(\sum_{h=1}^{n_{\mathcal{P}}} f_{\Delta}(h) \mathbf{1}_{(\mathcal{M}_i \in \mathrm{P}_h)} S_{R|\Delta}(V_i - |h) \right)^{\delta_i} [f_R(V_i)]^{1-\delta_i} \right),$$

where

$$\mathcal{L}(F_{T,C}) = \prod_{i=1}^{n} \left(\left\{ \sum_{t=V_i, c \in \mathcal{M}_i} f_{T,C}(t,c) \right\}^{\delta_i} \left[\int_{t>V_i, c \in \mathcal{C}_r} \mathrm{d}F_{T,C}(t,c) \right]^{1-\delta_i} \right), \quad (4)$$

$S(t-) = \lim_{x \uparrow t} S(x)$, $F_{T,C}$ is a bivariate cdf with its df $f_{T,C}$, f_R is the df of R, and $S_{R|\Delta}$ is the conditional survival function of R given Δ. Since the full likelihood can be written as two factors and only the factor $\mathcal{L}$ involves $F_{T,C}$, in order to find the MLE of $F_{T,C}$, or to estimate the parameters in $F_{T,C}$, it suffices to define the likelihood function by $\mathcal{L}(F_{T,C})$.

If one further assumes S4 then the likelihood function becomes

$$\mathcal{L}(F_{T,C}) = \prod_{i=1}^{n} \left(\sum_{j \in \mathcal{M}_i} \left\{ f_{X_j}(V_i) \prod_{k:\, k \neq j} S_{X_k}(V_i) \right\} \right)^{\delta_i} \left(\prod_{j=1}^{J} S_{X_j}(V_i) \right)^{1-\delta_i}.$$

The following example illustrates the relation between $\mathcal{J}$, $\mathcal{P}$, $\mathcal{M}$, and f_{Δ}.

Example 1 Suppose that $J = 3$; $X_i \sim f(t;\theta_i) = \theta_i e^{-\theta_i t}$, $t > 0$, $i = 1, 2, 3$; $T = \min\{X_1, X_2, X_3\}$, and $R \sim U(a,b)$, the uniform distribution on the interval (a,b). Suppose that A1 and S4 hold. The $\mathcal{J}$ consists of 7 elements: $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$. The collection $\mathcal{P}$ consists of 5 partitions: $P_1 = \{\{1\},\{2\},\{3\}\}$, $P_2 = \{\{1\},\{2,3\}\}$, $P_3 = \{\{2\},\{1,3\}\}$, $P_4 = \{\{3\},\{1,2\}\}$, $P_5 = \{\{1,2,3\}\}$. One may assume $\Delta = Z + 1$, where $Z \sim bin(4, p)$. Notice that f_{Δ} specifies a distribution on $\mathcal{P}$, not on $\mathcal{J}$. Also $\Delta \neq \mathcal{M}$, as

$$\begin{aligned} f_{\mathcal{M}}(A) &= \sum_{h} f_{\mathcal{M}|\Delta}(A|h) f_{\Delta}(h) = \sum_{h:A \in P_h} P\{C \in A\} f_{\Delta}(h) \\ &= P\{C \in A\} \sum_{h:\ A \in P_h} f_{\Delta}(h). \end{aligned}$$

3 The parametric estimation

3.1 Forms of parametric families

Unlike the form of parametric families for the RMCR data considered in the literature, we propose a general form for parametric families. In particular, a parametric distribution family for (T, C) can be written as

$$f_{T,C}(t,c) = p_c f(t|c;\gamma) \quad \text{where } \gamma = (p_1, \ldots, p_{J-1}, \gamma_J, \ldots, \gamma_k) \quad (5)$$

is a $k \times 1$ vector with $k \geq J$ or a function of ψ, say $\gamma = g(\psi)$, where ψ is a $k_1 \times 1$ parameter vector with $k_1 \leq k$ (see the end of Examples 2, 3 and 4), $\mathbf{p} = (p_1, \ldots, p_J)$ is a probability vector, and $f(\cdot|c, \gamma)$ is a df depending on (c, γ). The reason is as follows.

1. An arbitrary df $f_{T,C}$ satisfies $f_{T,C}(t, c) = f_C(c) f_{T|C}(t|c)$. Thus $f(\cdot|c; \gamma)$ corresponds to $f_{T|C}$ and p_c corresponds to $f_C(c)$. Hence, form (5) always holds.
2. If $T \perp C$, then

$$f_{T,C}(t, c) = p_c f_T(t), \quad \text{where } p_c = f_C(c), \tag{6}$$

which is naturally of form (5). For instance, under S4, it is often that the proportional hazards (PH) models have been assumed for $X_1, \ldots, X_J$ (see Flehinger et al. 2001). This class of distributions satisfies (6). In particular, if $S_{X_j}(t) = (S_o(t))^{\theta_j}$, where S_o is a survival function which is either a given form (e.g., in Example 8) or a parametric form (e.g., in Example 9), then

$$\begin{aligned} f_{T,C}(t, c) &= h_{X_c}(t) \prod_{j=1}^{J} S_{X_j}(t) \quad \text{(as hazard function } h = f/S) \\ &= \theta_c h_o(t)(S_o(t))^{\theta_1 + \cdots + \theta_J} \quad \text{(due to the PH model)} \\ &= p_c \beta h_o(t)(S_o(t))^{\beta} \end{aligned} \tag{7}$$

where $\beta = \sum_{j=1}^{J} \theta_j$, $p_c = \theta_c/\beta = f_C(c)$ and $f_{T|C}(t|c) = f_T(t) = \beta h_o(t)(S_o(t))^{\beta}$. Example 1 is one such special case with $S_o(t) = e^{-t}$, $t > 0$.
3. In the literature, $f_{T,C}$ is specified through the distributions of X_j's under S4, which can also be expressed in form (5) (see Example 2). The class of distribution families to be considered is larger than those induced by the distributions on X_j's under assumption S4. In fact, in medical research, S4 often does not hold and then it is difficult to specify $f_{T,C}$ through the distributions of X_j's.
4. It is quite convenient and easy to specify a parametric family for $f_{T,C}$ in the form of (5) (see Examples 3 and 4).

Example 2 Suppose $J = 2$, $X_1 \sim Exp(\theta)$ and $X_2 \sim U(a, b)(0 \leq a < b)$, and $X_1 \perp X_2$. Then T and C are dependent. In fact, $f_{T,C}(t, 1) = \theta e^{-\theta t}\{\mathbf{1}_{(t \in (0,a))} + \mathbf{1}_{(t \in [a,b])} \frac{b-t}{b-a}\}$ and $f_{T,C}(t, 2) = \frac{\mathbf{1}_{(t \in (a,b))}}{b-a} e^{-\theta t}$. Moreover, $f_{T|C}(t|1) = f_{T,C}(t, c)/p_1$ and $f_{T|C}(t|2) = f_{T,C}(t, c)/p_2$ with $p_2 = \frac{e^{-a\theta} - e^{-b\theta}}{\theta(b-a)}$ and $p_1 = 1 - p_2$. Notice that $f_C(2) = p_2 > 0$ and $f_T(t) > 0$ for $t \in (0, b)$. However, $f_{T,C}(a/2, 2) = 0 < f_T(a/2) f_C(2)$. Thus $C \not\perp T$ even though condition S4 holds. In this example, $\gamma = (p_1, a, b, \theta) = g(\psi)$, where $\psi = (a, b, \theta)$ (see (5)).

Example 3 Suppose that $J = 3$, $f_{T|C}(t|i) = \beta_i e^{-\beta_i t}$, $t > 0$. Thus T and C are dependent. In this example, $\psi = \gamma = (p_1, p_2, \beta_1, \beta_2, \beta_3)$ (see (5)).

Example 4 Suppose that $J = 3$, $T|(C = c) \sim U(a_c, b_c)$, $c = 1, 2, 3$. The parameter γ consists of a_c's, b_c's and p_c's. In this example, $\psi = \gamma$ (see (5)).

In Examples 2 and 4 we try to model the situation that certain parts of a system will not fail in the early stage or some diseases will not happen in the childhood.

3.2 The MLE

Under the RPM model, the parametric MLE with RMCR data maximizes the likelihood function in (4):

$$\mathcal{L}(\gamma) = \prod_{i=1}^{n} \left\{ \left(\sum_{t=V_i, c\in\mathcal{M}_i} f_{T,C}(t,c;\gamma) \right)^{\delta_i} \left(\int_{t>V_i, c\in\mathcal{C}_r} \mathrm{d}F_{T,C}(t,c;\gamma) \right)^{1-\delta_i} \right\}.$$

Verify that as far as the likelihood is concerned, $\mathcal{L}$ is the same as $\mathcal{L}_c$ in (1). We shall make use of the following identifiability assumptions in addition to A1.

A2 The matrix $(\phi(W_1), \ldots, \phi(W_m))$ is of rank J, where $W_1, \ldots, W_m$ are all the distinct values of $\mathcal{M}^o$ with $P(\mathcal{M}^o = W_j) > 0$, $j = 1, \ldots, m$, $\phi(A) = (\mathbf{1}_{(1\in A)}, \ldots, \mathbf{1}_{(J\in A)})'$ with $A \in \mathcal{J}$ and B' is the transpose of the matrix B.

A3 The degree of freedom of the parameters related to f_C should be at least one.

A4 The family $\{f(t,c;\theta) : \theta \in \Theta\}$ to which $f_{T,C}$ belongs satisfies that $\theta = \theta_*$ iff $f(t,c;\theta) = f(t,c;\theta_*)$ a.e. on $[0,b]\times\mathcal{C}_r$, for some $b \in [0,\tau]$, where $\tau = \sup\{t : F_R(t) < 1\}$ and $P(T \in [0,b]) > 0$.

Hereafter, we shall present Examples 5–10, which are helpful in understanding these identifiability assumptions. The detailed proofs of the arguments in some of these examples are given in Yu et al. (2009). First we provide an example of a consistent MLE.

Example 5 Assume the conditions in Example 1. Thus $f_{T,C} = f_T f_C$. Suppose that $P(\delta = 1) > 0$, θ_1, θ_2 and θ_3 are unknown parameters and $\mathcal{M}_i$'s are always of the forms $\{1\}$, $\{2\}$, $\{3\}$, $\{2,3\}$ if $\delta_i = 1$. Let $\eta_{ij} = \mathbf{1}_{(\mathcal{M}_i=\{j\})}$, $j = 1, 2, 3$, and let $\overline{Y} = (\overline{\eta}_1, \overline{\eta}_2, \overline{\eta}_3, \overline{\delta}, \overline{V})$ be the sample means of η_{ij}'s, δ_i's and V_i's. A medical research data set given by Dinse (1986) has such form (see also Sect. 6); and it is the case when conditional on $\delta = 1$, Δ only takes values 1 and 2, corresponding to the partitions P_1 and P_2 defined in Example 1. Then it can be shown that the MLE of p_j's and β are $\hat{p}_1 = \frac{\overline{\eta}_1}{\overline{\delta}}$, $\hat{p}_2 = \frac{\overline{\eta}_2}{\overline{\eta}_2+\overline{\eta}_3}\frac{\overline{\delta}-\overline{\eta}_1}{\overline{\delta}}$, $\hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2$, $\hat{\beta} = \overline{\delta}/\overline{V}$. It follows from the invariance of the MLE that $\hat{\theta}_j = \hat{p}_j\hat{\beta}$. Since the MLE has an explicit form and is a smooth function of the sample mean $\overline{Y}$, its strong consistency and asymptotic efficiency can be easily established under the RPM model.

However, there are cases that there exists no consistent MLE such as in Example 6.

Example 6 Assume the assumptions in Example 1 except that $R \sim bin(1, 1/2)$.

(1) Suppose that either $\mathcal{M}_i = \{1\}$ or $\mathcal{M}_i = \{2,3\}$ if $\delta_i = 1$ (*i.e.*, $f_\Delta(2) = 1$), and θ_i's are all unknown parameters. Then an MLE of (p_1, p_2, β) is $\hat{p}_1 = \overline{\eta}_1/\overline{\delta}$, $\hat{\beta} = \overline{\delta}/\overline{V}$ and $\hat{p}_2 = p$ for an arbitrary $p \in [0, 1-\hat{p}_1]$. Consequently, $\hat{p}_1$ and $\hat{\beta}$ are consistent, but not $\hat{p}_2$.

(2) Suppose that $f_{\Delta|R}(1|0) = 1$ and $f_{\Delta|R}(5|1) = 1$. Hence $\mathcal{M}_i = \mathcal{C}_r \ \forall i$, and θ_i's are all unknown parameters. Then an MLE of (p_1, p_2, β) is $(\hat{p}_1, \hat{p}_2, \hat{\beta}) = (p, q, \overline{\delta}/\overline{V})$ for each (p, q) with $p, q \geq 0$ and $p + q \leq 1$. Thus $\hat{p}_c$'s are inconsistent.

In both cases of Example 6, A2 fails and the MLE $\hat{p}_2$ is inconsistent. Notice that in case (2), even though $(\phi(\{1\}), \phi(\{2\}), \phi(\{3\}))$ is of rank 3 $(= J)$ where $P(\mathcal{M} = \{j\}) > 0$ for $j = 1, \ldots, 3$, A2 still fails, as $\mathcal{M}^o$ only takes one value $\mathcal{C}_r$ and thus $(\phi(W_1))$ has rank $1 < J$.

If one defines P_1 to be the partition with components $\{j\}$, $j = 1, \ldots, J$, then a sufficient condition of A2 is $f_{\Delta|R}(1|t) > 0$ if $F_R(t) > 0$ with $t < \tau$. In fact, $\{j\}$, $j = 1, \ldots, J$, are J distinct values of $\mathcal{M}^o$, and the $J \times J$ matrix $A = (\phi(\{1\}), \ldots, \phi(\{J\}))$ is the $J \times J$ identity matrix. Notice that in Example 5, $f_{\Delta|R}(1|r) \in (0, 1)$, where $r \in [a, b]$. Thus A2 holds. Moreover if $f_\Delta(j) = 1$ and $j \neq 1$, then $\mathcal{M}^o$ does not have J or more distinct values and thus the first matrix in A2 is of rank $< J$. Consequently, A2 does not hold and it leads to inconsistency as in Example 6.

Assumption A3 eliminates the case that a bivariate distribution in the RMCR estimation problem can be converted to a univariate distribution such as in Example 7 below.

Example 7 Let the notations and assumptions be the same as in Example 1, except that $\theta_i = i\theta$, $i = 1, 2, 3$, with unknown parameter θ. Even under the assumptions in cases (1) and (2) of Example 6 (that is, even when A2 fails), θ can be identified through S_T. Verify that $f_C(i) = i/6$, $i = 1, 2, 3$. Thus the degree of freedom of the parameters related to f_C is zero in this case. The MLE of θ is $\overline{\delta}/(6\overline{V})$. It is essential a univariate distribution problem.

Now we consider how to find the MLE in some special cases based on a random sample of size n from the RPM model. A closed-form solution can be found in Example 5. However, in general there is no closed form solution for the MLE such as in Example 3.

Example 3 (continued). Assume that A1 and A2 hold. There is no closed form solution to the MLE. The MLE of γ is a zero point of $\frac{\partial \ln \mathcal{L}(\gamma)}{\partial \gamma} = 0$, where $\frac{\partial \ln \mathcal{L}(\gamma)}{\partial \gamma}$ is derived in a technical report (see Yu et al. 2009). One can use the Newton–Raphson method to find the MLE.

If there is no closed form solution for the MLE, one can also use the Self-consistent (SC) algorithm to find the MLE (see Example 8 below).

Example 8 Suppose that A1, A2 and S4 hold, and X_j has the survival function $(S_o(t))^{\theta_j}$, where S_o is a known survival function with $\theta_j > 0$. Thus $f_{T,C} = f_T f_C$. The MLE of β maximizes $\ln \prod_{i=1}^n (\beta h_o(V_i))^{\delta_i} \prod_{i=1}^n (S_o(V_i))^\beta$ or $\sum_i \delta_i \ln(\beta h_o(V_i)) + \beta \sum_i \ln S_o(V_i)$. Thus the MLE $\hat{\beta} = \overline{\delta}/\overline{\ln S_o(V)}$, where $\overline{\ln S_o(V)} = \sum_i \ln S_o(V_i)/n$. The MLE $\hat{p}_c$ maximizes $\prod_i (\sum_{c \in \mathcal{M}_i} p_c)$ and can be obtained by the SC algorithm (see Turnbull 1976):

1. Assign mass $p_c^{(1)} = 1/J$ to each point c in $\mathcal{C}_r$.

2. For $h \geq 1$, update $p_c^{(h+1)}$ by $p_c^{(h+1)} = \sum_{W \in \mathcal{J}} \frac{\sum_{i=1}^n \mathbf{1}_{(\mathcal{M}_i = W)}}{n} \frac{p_c^{(h)} \mathbf{1}_{(c \in W)}}{\sum_{k=1}^J p_k^{(h)} \mathbf{1}_{(k \in W)}}$, $c = 1, \ldots, J$; stop at convergence.

Sometimes, one may combine both the SC and the Newton–Raphson algorithms as follows.

Example 9 Continuing from the previous example with the same notations. Suppose that $J = 3$, $F_{T|C}$ is $U(a, b)$ and (3), A1 and A2 hold. Verify that the MLE of a is $\hat{a} = \min\{V_i : \delta_i = 1, i \in \{1, \ldots, n\}\}$; the MLE of p_j's are the same as in Example 8. Finally, to find the MLE of b it suffices to maximize $\prod_{i=1}^n [(1 - \frac{V_i - a}{b-a})^{\mathbf{1}_{(a<V_i<b)}}]^{1-\delta_i} (\frac{1}{b-a})^{\sum_{i=1}^n \delta_i}$, $b \geq max_i\{V_i : \delta_i = 1\}$; the MLE $\hat{b}$ of b should satisfy the equation $D(b) = 0$, where $D(b) = \sum_i \frac{1-\delta_i}{b-V_i} - \frac{n}{b-\hat{a}}$. A zero point can be obtained by the Newton–Raphson method.

Notice that A4 holds for Examples 1–5, 7, 8, and 9, but not in case (2) of the following example.

Example 10 Suppose that $J = 2$, $R \equiv 0$ and T has range $\{0, 1\}$. Let $p = f_C(1)$ and $p_{kc} = f_{T|C}(k|c)$. Consider two cases: (1) $f_{T,C}(0, 1) + f_{T,C}(0, 2) + f_{T,C}(1, 1) = 1$, (2) no further assumption is imposed. Then A3 holds in case (1) but not in case (2).

Proof Case (1). Since $p_{01} + p_{11} = 1$ and $p_{02} = 1$, $f_{T,C}(0, 1) = pp_{01}$ and $f_{T,C}(0, 2) = 1 - p$, then $pp_{01} = p^* p_{01}^*$ and $1 - p = 1 - p^*$ iff $p = p^*$ and $p_{01} = p_{01}^*$. Thus A4 holds.

Case (2). Since $p_{0c} + p_{1c} = 1$ for $c = 1, 2$, $f_{T,C}(0, 1) = pp_{01}$ and $f_{T,C}(0, 2) = (1 - p)p_{02}$, then $pp_{01} = p^* p_{01}^*$ and $(1 - p)p_{02} = (1 - p^*)p_{02}^*$ do not imply that $p = p^*$ and $p_{02} = p_{02}^*$. For instance, take $p^* = p_{01}^* = p_{02}^* = 1/2$, the solutions to $pp_{01} = 1/4$ and $(1 - p)p_{02} = 1/4$ are not unique. Thus A4 fails and the parameters p and $f_{T|C}(k|c)$ are not identifiable. □

4 Asymptotic properties of the MLE

The theorems about the consistency and asymptotic normality of the MLE are given in this section. Since this is a finite dimensional parametric estimation problem, under certain regularity conditions, their proofs are quite standard. For a better presentation, we shall relegate their proofs to Yu et al. (2009).

Theorem 1 *Suppose that (1) A1, A2 and A4 hold; (2) T is continuous and the df of (T, C) belongs to a parametric distribution family $\{f(t, c; \theta) : \theta \in \Theta\}$, where Θ is a compact subset of $\mathcal{R}^k$, k is a positive integer; (3) $\ln \sum_{c \in B} f(t, c; \theta)$ and $\ln S_T(; \theta)$ are continuous in θ uniformly in t in the sense that given ψ, $\forall \epsilon > 0$, $\exists \eta > 0$ which is independent of t such that $|\ln \sum_{c \in B} f(t, c; \theta) - \ln \sum_{c \in B} f(t, c; \psi)| < \epsilon$ and $|\ln S_T(; \theta) - \ln S_T(; \psi)| < \epsilon$ whenever $|\psi - \theta| < \eta$, where $B \in \mathcal{J}$; (4) $|E(H(\theta_o))| < \infty$, where θ_o denotes the true value of θ and $H(\theta) = \frac{1}{n} \ln \mathcal{L}(\theta)$. Then the MLE of θ is consistent.*

Let $I(\theta) = E(\frac{\partial^2 H(\theta)}{\partial\theta\partial\theta'})$. Notice that

$$\frac{\partial H(\theta)}{\partial\theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i \frac{\sum_{c\in\mathcal{M}_i}\frac{\partial}{\partial\theta}f(V_i,c;\theta)}{\sum_{c\in\mathcal{M}_i}f(V_i,c;\theta)} + (1-\delta_i)\frac{\frac{\partial}{\partial\theta}S_T(V_i;\theta)}{S_T(V_i;\theta)}\right\} = \frac{1}{n}\sum_i h_i(\theta),$$

where $h_i(\theta) = h(V_i, \mathcal{M}_i, \delta_i, \theta)$ and $h(t, B, \delta, \theta) = \left\{\delta\frac{\sum_{c\in B}\frac{\partial}{\partial\theta}f(t,c;\theta)}{\sum_{c\in B}f(t,c;\theta)} + (1-\delta)\frac{\frac{\partial}{\partial\theta}S_T(t;\theta)}{S_T(t;\theta)}\right\}$.

Theorem 2 *Suppose that the assumptions in Theorem 1 hold, $|I(\theta_o)| \neq 0$, and $h(t, B, \delta, \theta)$ is continuous in θ uniformly over t, in the sense that for each $B \in \mathcal{J}$ and δ, given ψ, $\forall\epsilon > 0$, $\exists\eta > 0$ which is independent of t such that $|h(t, B, \delta, \theta) - h(t, B, \delta, \psi)| < \epsilon$ whenever $|\theta - \psi| < \eta$. Then the MLE of θ is asymptotically efficient.*

5 Simulation results

We have carried out simulation results to see whether the MLE and the estimate of its standard deviation provided by Theorem 2 are close to the true values for moderate sample sizes. The RPM model can easily be implemented in a simulation study. For instance, in Example 1, one can proceed as follows.

1. Generate random vector (T, C) and random variable R independently, say (t, c, u).
2. Set $V = \min\{t, u\}$ and $\delta = \mathbf{1}_{(t\le u)}$.
3. If $\delta = 0$, set $\mathcal{M}^o = \{1, 2, 3\}$.
4. If $\delta = 1$, generate a random number Δ from $Z + 1$ where $Z \sim bin(4, p)$, with $p = \frac{1}{2} + \frac{1}{4}F_R(u)$. Choose partition P_h if $\Delta = h$. Let P_{hj}'s be the sets in the partition P_h, then among them $\exists!$ P_{hj} such that $c \in P_{hj}$. Define $\mathcal{M}^o = P_{hj}$. By now, an observation $(V, \delta, \mathcal{M}^o)$ is obtained, no matter whether $\delta = 0$ or 1.
5. Repeat steps 1, 2, 3 and 4 n times, and obtain $(V_1, \delta_1, \mathcal{M}_1), \ldots, (V_n, \delta_n, \mathcal{M}_n)$.

This scheme contains partially masked $\mathcal{M}_i$ such as $\{1, 2\}$, as well as unmasked $\mathcal{M}_i$'s, that is, $\{j\}$'s. One can replace $bin(4, p)$ by other distributions on $\{0, 1, 2, 3, 4\}$, but should make sure that A2 is satisfied. The aforementioned scheme assumes that R and Δ are dependent.

In Table 1 we present simulation results under the assumptions in Example 5. In particular, assume $T \sim Exp(1)$, $J = 3$, $(C - 1) \sim bin(2, 0.5)$, $R \sim U(1, 2)$ (or $U(1, 3)$ in the second block of Table 1) (so that the censoring rate c changes), Δ and R are independent, and $Z \sim bin(1, 0.7)$ (or $bin(1, 0.5)$ in the second block) with P_h defined in Example 1 (so that the masking rate m changes). For each case, we repeated 10,000 times. The simulation results are given in Table 1. The row with $\hat{\gamma}_n(SE)$ gives the sample means of the MLE's $\hat{\gamma}_n = (\hat{p}_1, \hat{p}_2, \hat{\beta})$ with a sample size of n and their standard errors (SE's) in brackets. The rows with $\hat{\sigma}_{\hat{\gamma}}$ and (SE) give the empirical estimates of the Cramer–Rao lower bound (CRLB) (see Ferguson 1996) and their SE's respectively. It is seen that the MLE almost obtains the CRLB for the standard deviations of the estimators even with a sample size of 50.

Table 1 MLE under exponential distribution with $T \perp C$

True value	$\beta = 1$	$p_1 = 0.25$	$p_2 = 0.5$
$(m, c) = (0.637, 0.233)$			
$\hat{\gamma}_{50}$ (SE)	1.014 (0.165)	0.249 (0.070)	0.501 (0.135)
(SE) $\hat{\sigma}_{\hat{\gamma}}$	(0.025) 0.162	(0.007) 0.069	(0.024) 0.124
$\hat{\gamma}_{100}$ (SE)	1.006 (0.114)	0.250 (0.049)	0.500 (0.093)
(SE) $\hat{\sigma}_{\hat{\gamma}}$	(0.012) 0.114	(0.004)0.049	(0.011) 0.090
$(m, c) = (0.476, 0.159)$			
$\hat{\gamma}_{50}$ (SE)	1.016 (0.158)	0.250 (0.067)	0.499 (0.102)
(SE) $\hat{\sigma}_{\hat{\gamma}}$	(0.025) 0.155	(0.007) 0.066	(0.010) 0.098
$\hat{\gamma}_{100}$ (SE)	1.007 (0.109)	0.250 (0.047)	0.499 (0.071)
(SE) $\hat{\sigma}_{\hat{\gamma}}$	(0.013) 0.109	(0.003) 0.047	(0.005) 0.070

Table 2 MLE under exponential distribution with $T \not\perp C$

γ	$p_1 = 0.25$	$p_2 = 0.5$	$\beta_1 = 1$	$\beta_2 = 2$	$\beta_3 = 3$
$\hat{\gamma}_{58}$	0.244 (0.117)	0.513 (0.482)	1.220 (0.590)	2.142 (0.639)	3.292 (1.436)
$\hat{\sigma}_{\hat{\gamma}_{58}}$	(0.038) 0.063	(0.115) 0.092	(0.368) 0.538	(0.261) 0.643	(1.873) 1.787
$\hat{\gamma}_{400}$	0.248 (0.025)	0.501 (0.034)	1.027 (0.155)	2.014 (0.220)	3.031 (0.465)
$\hat{\sigma}_{\hat{\gamma}_{400}}$	(0.001) 0.025	(0.001) 0.034	(0.035) 0.152	(0.025) 0.222	(0.079) 0.474

In Table 2, we present simulation results under the assumptions in Example 3. In particular, $(C - 1) \sim bin(2, 0.5)$; $T|C \sim Exp(C)$ with mean C; $Z \sim bin(1, 0.5)$; $R \sim U(1, 2)$. The censoring rate $c = 0.091$ and masking rate $m = 0.450$. We only present the sample means of the MLE's and their SE for sample sizes 58 (same as the data in Dinse 1986) and 400. $\hat{\gamma}_n$ and $\sigma_{\hat{\gamma}_n}$ represent the averages of the MLE and the estimate of its standard deviation (SD) of sample size n in 10000 replications, and SE represents the corresponding standard error.

The results in these tables suggest that the approximation of the MLE to the true parameters are quite satisfactory for moderate sample sizes $n = 58$ and the estimates of the SD's are good for sample size $n = 400$.

6 Data analysis

We consider two sets of data here: one is a medical research data set and the other is an industrial data set. We consider the parametric set-ups in Example 3. Denote H_o: $\beta_1 = \beta_2 = \beta_3 (= \beta)$, and $H_1 : H_o$ does not hold.

6.1 A medical data set

Dinse (1986) provides a data set in the medical research and studies the non-parametric estimation of two age-specific descriptors of disease development and the subsequent

Table 3 Time to death (in days) and status at death, with respect to NRVD for 58 female RFM mice

Type	$\mathcal{M}_i$	Time (T)						
Absent	{1}	231	444	468	473	527	550	593
	{1}	600	610	650	655	660	715	720
	{1}	752	785	832	838	859	891	896
	{1}	904	931	952	998			
Incidental	{2}	559	595	598	603	765	783	794
	{2}	811	856	870	883	897	975	978
	{2}	991	1005	1023	1026	1053		
Fatal	{3}	500	591	713	751	778	784	786
	{3}	796						
Unknown	{2, 3}	593	735	816	848	850	1048	

Table 4 MLE based on Dinse's medical research data

	p_1	p_2	$\hat{\beta}$ or β_1	β_2	β_3
H_o:	0.431 (0.065)	0.400 (0.068)	0.151 (0.001)		
H_1:	0.431 (0.065)	0.400 (0.068)	0.154 (0.627)	0.149 (1.082)	0.152 (2.119)

effects of a disease on longevity: prevalence and mortality. The data in his Table 1 are not formulated as the masked competing risks (MCR) data form, but can be converted to MCR data as in Table 3.

The data can be interpreted as three causes of death: (1) the disease is not present at death and the death is due to other causes; (2) the death is not due to the disease though the disease is present; (3) the death is due to the disease. There are 4 types of $\mathcal{M}_i$'s and they can be classified as in Table 3.

We obtain the MLE through the Newton–Raphson algorithm based on the data in Dinse (1986). Under H_o it can be shown that $T \perp C$ and thus it becomes the case in Example 5. Then the MLE has closed form solution (see Example 5). The standard deviation of $\hat{\mathbf{p}}$ can be obtained by the inverse of the Fisher information matrix. For the data given in Table 3 with T_i replaced by $\ln T_i$, the MLE under H_o or H_1 is presented in Table 4.

The results in Table 4 have a strong indication that $T \perp C$ for this data set: (1) the MLEs $\hat{p}_c$ do not change under these two models; (2) the MLEs $\hat{\beta}_c$'s are almost identical under these two models; (3) a χ^2 test of degree of freedom 1 for testing H_o results in a test statistic value of 10^{-4} (for details see Yu et al. 2009).

6.2 An industrial data set

Consider the MCR data set about 682 PS/2 computer systems in Reiser et al. (1995) with 8 failed $(V_i, \mathcal{M}_i)$: (1, {1}), (1, {1}), (1, {1, 3}), (1, {1, 2, 3}), (16, {3}), (17,

{2, 3}), (21, {2}), (222, {2}), 348 right-censored at 67, 246 right-censored at 200, 26 right-censored at 800, and 54 right-censored at 4000.

For this data set with T_i replaced by $2+\ln T_i$ (as $T_i = 1$ for some i), the MLE under the assumptions in Example 3 is

$\hat{p}_1$	$\hat{p}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.987 (0.009)	0.007 (0.008)	0.00001 (0.000007)	0.004 (0.007)	0.092 (0.084)

It can be shown that under H_o the MLE $(\hat{p}_1, \hat{\beta}) = ((4-\sqrt{2})/7, \overline{\delta}/\overline{V}) = (0.369, 0.000026)$ and $\hat{p}_2 = \hat{p}_1$, as the likelihood is $\mathcal{L} = p_1^4(1-2p_1)(1-p_1)^2\beta^{n\overline{\delta}}e^{-n\overline{V}}$. Unlike the data in § 6.1, the current data set suggests that H_o is unlikely to be true, as the MLE's of $\hat{p}_1$ under H_o and H_1 satisfy $|0.987 - 0.369| > 3 \times 0.009$.

Remark 1 In data analysis, an interesting issue is to construct a goodness-of-fit test procedure to check whether the parametric model assumption is valid for the data. A possible procedure is the Kolmogorov test statistic $\sum_{t,c} |F_{T,C}(t,c;\hat{\theta}) - \hat{F}_{T,C}(t,c)|$ where $\hat{F}_{T,C}$ is the generalized MLE (GMLE) of $F_{T,C}$. However, the GMLE is not uniquely defined and the one defined in the literature is erratic (Dinse 1982, p.425$_{11}$). Moreover we shall show in a forthcoming paper that the GMLE defined in the literature is inconsistent if T is continuous.

7 Comparison between the RPM model and CMP models

In this paper, we propose the RPM model. We shall first examine CMP Model 1 under the RPM model in the following three lemmas. For a better presentation, the proofs of the lemmas in this section are given in Yu et al. (2009).

Lemma 1 $\forall A \in \mathcal{J}\, v_c(A) = \begin{cases} 0 & \text{if } c \notin A \\ \sum_{h:\ A \in P_h} f_\Delta(h) & \text{if } c \in A. \end{cases}$ *Thus S1a holds.*

Lemma 2 $\forall A \in \mathcal{J}\, v_{t,c}(A) = \begin{cases} \sum_{h:\ A \in P_h} f_\Delta(h) & \text{if } c \in A \text{ and } f_{T,C}(t,c) > 0 \\ 0 & \text{if } c \notin A \text{ or } f_{T,C}(t,c) = 0. \end{cases}$

Lemma 3 *Assumption S1 is valid iff the following assumption holds*

A5 *for each* t, $f_T(t) > 0$ *implies* $f_{T,C}(t,c) > 0$ *for each* $c \in \mathcal{C}_r$.

Remark A5 is valid in Examples 1, 3, 5–8, but not in Examples 2, 4 and 9.

It can help us appreciate the RPM model by comparing masking and right censoring, which have a strong analog. The random set $\mathcal{M}$ in masking corresponds to a random interval $\mathcal{I}$ in right censoring, where $\mathcal{I} \overset{def}{=} \begin{cases} (V,\infty) \text{ if } T>V \\ [V,V] \text{ if } T \le V, \end{cases}$ which is also a random set. C and T are the variables of interests in masking and right censoring, respectively, but $C \in \mathcal{M}$ and $T \in \mathcal{I}$. In the RPM model, a new random variable Δ chooses a partition $P_h (= (P_{h1}, \ldots, P_{hk_h}))$ of $\mathcal{C}_r$, then $\mathcal{M} = P_{hj}$ if $C \in P_{hj}$; whereas in the right censorship model, a random variable V induces a partition $\{(-\infty, V), [V, V], (V, \infty)\}$

of $(-\infty, \infty)$, then $\mathcal{I}$ is determined by the interval that T falls in. The likelihood Λ in (1) for RMCR data also corresponds to a likelihood $\mathcal{L}_{rc}$ for the RC data, where

$$\mathcal{L}_{rc} = \prod_{i=1}^{n} (f_T(V_i) P\{\mathcal{I} = [V_i, V_i] | T = V_i\})^{\delta_i} \left(\int_{V_i}^{\infty} f_T(t) P\{\mathcal{I} = (V_i, \infty) | T = t\} \mathrm{d}t \right)^{1-\delta_i}$$
$$\propto \prod_{i=1}^{n} (f_T(V_i))^{\delta_i} (S_T(V_i))^{1-\delta_i} \quad \text{if } P\{\mathcal{I} = (V_i, \infty) | T = t\} \text{ is free of } t.$$

The latter corresponds to the conditions that $P(\mathcal{M} = A | C = c)$ is free of c in masking and $P(\mathcal{M} = A | T = t, C = c)$ is free of t in the CMP models. Such a right censorship model is probably not an attractive alternative to the standard right censorship model.

The difference between the CMP models and the RPM model can be listed as follows.

1. The CMP Model 1 does not allow the case that $f_T(t) S_R(t-) > 0$ but $f_{T,C}(t, c) = 0$ for some $c \in \mathcal{C}_r$ (see Lemma 3 and Examples 2, 4 and 9). However, the RPM model has no such restriction.
2. Under the RPM model S2 implies $\Delta \perp R$, as S2 says $\mathcal{M} \perp R$. The CMP Model relies on S2. In fact, the first equality in the likelihood in (1) does not hold if S2 fails, as $f_{T,C,\mathcal{M},R} \neq f_{T,C,\mathcal{M}} f_R$ without S2. However, the RPM model is valid without S2, as long as A1 holds (see step 4 of the simulation procedure for generating $\mathcal{M}$ in Sect. 5).
3. It is much simpler to implement a simulation study with the RPM model than with CMP Model 1. Under CMP Model 1, in order to generate $\mathcal{M}$, one has to solve for $f_{\mathcal{M}|C}$'s subject to constraints $\sum_{A: A \in \mathcal{J}} f_{\mathcal{M}|C}(A|c) = 1$ and S1a. This can be done but is not as simple as the procedure for generating $\mathcal{M}$ under the RPM model (see step 4 in Sect. 5).
4. The CMP models are based on S3 too. It can be easily shown that assumption S3 is a false statement, unless that it is further assumed that n_1 is not random. The disproof of S3 as stated is very similar to the proof that the order statistics of i.i.d. random variables are dependent.
5. In this paper, to avoid the ugly assumption S3, we define $\mathcal{M}^o$ and impose (3). Mukhopadhyay (2006, p.806[13]) maybe the first person who points out that $\mathcal{M} \neq \mathcal{C}_r$ if $T > R$. $\mathcal{M}$ is what we know about C at the failure time T and $\mathcal{M}^o$ is what we know about C at the observable time V. Under the RPM model, if $T \leq R$, $C \in A \in P_h$ and $\Delta = h$ then $\mathcal{M}^o = \mathcal{M} = A$.
6. The CMP Model 1 does not have the restriction A1, which is critical in the RPM model. Thus neither of these two models is a special case of the other. They both are useful alternative to each other.
7. The CMP Model 2 has more parameters than the RPM model, and thus it is more flexible. However, unless additional constraints are imposed, the parameter under CMP Model 2, can be non-identifiable under assumptions A1, A2 and S4 (see Example 11).

Example 11 Suppose that $J \geq 2$; there is no censoring; X_j has the survival function $S(t) = e^{-\lambda_j(t)^{\beta_j}}$ where $t > 0$; X_j's are independent; $\mathcal{M}_i \in \{\{1\}, \{2\}, \mathcal{C}_r\}$ with $f_{\mathcal{M}}(\mathcal{C}_r) > 0$ and $f_{\mathcal{M}}(\{j\}) > 0$; and $\beta_i = \beta_j$ for some $i \neq j$. Then the parameter under Λ_c is not identifiable. For instance, let $J = 2$, write $\beta = \beta_1 = \beta_2$, and $\psi = \sum_k \lambda_k$, then $f_C(j) = \lambda_j/\psi$. Let $\theta = (p_1, p_2, \mu_{11}, \mu_{22}, \mu_{13}, \mu_{23})$ be the parameter corresponding to the true value $(f_C(1), f_C(2), \nu_1(\{1\}), \nu_2(\{1\}), \nu_{13}(\mathcal{C}_r), \nu_{23}(\mathcal{C}_r))$, denoted by θ_o. Without S1a, $\Lambda_c(\beta, \psi, \theta_o) = \Lambda_c(\beta, \psi, \theta)$ for each $\theta = \left(p, 1-p, \frac{u_1}{p}, \frac{u_2}{1-p}, 1-\frac{u_1}{p}, 1-\frac{u_2}{1-p}\right)$ with $p \in [u_1, u_1+u_3]$, where $u_1 = f_{\mathcal{M}}(\{1\}), u_2 = f_{\mathcal{M}}(\{2\})$, and $u_3 = f_{\mathcal{M}}(\mathcal{C}_r) > 0$. Thus the parameter θ is not identifiable. This example basically shows that even under the assumption of the popular Weibull distribution family, which is quite common in real data, the parameters are not identifiable if $\beta_i = \beta_j$ for two coordinates of $(\beta_1, \ldots, \beta_J)$ (which is the case in the data analysis in Sect. 6.1). Thus CMP Model 2 is not that useful, unless stage-2 data or prior information are available.

Acknowledgments Prof. Yu is partially supported by NSF Grant DMS-0803456. The authors thank the editor, the associate editor and referees for helpful comments.

References

Craiu, R. V., Duchesne, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure. *Biometrika, 91*, 543–558.

Craiu, R. V., Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Analysis, 12*, 21–33.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics, 38*, 417–431.

Dinse, G. E. (1986). Nonparametric prevalence and mortality estimators of animal experiments with incomplete cause-of-death data. *JASA, 81*, 328–336.

Ferguson, T. S. (1996). *A course in large sample theory*. New York: Chapman & Hall.

Flehinger, B. J., Reiser, B., Yashchin, E. (1996). Inference about defects in the presence of masking. *Technometrics, 38*, 247–255.

Flehinger, B. J., Reiser, B., Yashchin, E. (2001). Statistical analysis for masked data. In N. Balakrishnan, C. R. Rao (Eds.), *Handbook of statistics* (Vol. 20, pp. 499–522). Amsterdam: Elsevier.

Friedman, L., Gertsbakh, I. B. (1980). Maximum likelihood estimation in a minimum type model with exponential and Weibull failure modes. *JASA, 75*, 460–465.

Guttman, I., Lin, D. K. J., Reiser, B., Usher, J. S. (1995). Dependent masking and system life data analysis: Bayesian inference for two-component systems. *Lifetime Data Analysis, 1*, 87–100.

Kalbfleisch, J. D., Prentice, R. I. (2002). *The statistical analysis of failure time data* (2nd ed.). Hoboken: Wiley.

Kuo, L., Yang, T. Y. (2000). Bayesian reliability modeling for masked system lifetime data. *Statistics & Probability Letters, 47*, 229–241.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). New York: Wiley.

Lin, D. K. J., Guess, F. M. (1994). System life data analysis with dependent partial knowledge on the exact cause of system failure. *Microelectronics and Reliability, 34*, 535–544.

Mukhopadhyay, C. (2006). Maximum likelihood analysis of masked series system lifetime data. *Journal of Statistical Planning and Inference, 136*, 803–838.

Mukhopadhyay, C., Basu, S. (2007). Bayesian analysis of masked series system lifetime data. *Communications in Statistics-Theory and Methods, 36*, 329–348.

Nagai, Y. (2004). Maximum likelihood analysis of masked data in competing risk models with an environmental stress. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E87-A*(12), 3389–3396.

Reiser, B., Guttman, I., Lin, D. K. J., Usher, J. S., Guess, F. M. (1995). Bayesian inference for masked system lifetime data. *Journal of the Royal Statistical Society, Ser C Applied Statistics, 44*, 79–90.

Sen, A., Basu, S., Benerjee, M. (2001). Analysis of masked failure time data under competing risks. In N. Balakrishnan, C. R. Rao (Eds.), *Handbook of statistics* (Vol. 20, pp. 523–540). Amsterdam: Elsevier.

Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *Journal of Royal Statistics Society, Ser. B, 38*, 290–295.

Yu, Q. Q., Wong, G. Y. C., Qin, H., Wang, J. (2009). *Technical report for "Random partition masking model for censored and masked competing risks data."* http://math.binghamton.edu/qyu.