About Conditional Masking Probability Models

by Qiqing Yu[1], Hao Qin[1] and Jiaping Wang[2]

Department of Mathematical Sciences, SUNY, Binghamton, NY 13902, USA
Correspondent email address: qyu@math.binghamton.edu

First version 12/2/09
Current version 3/16/10
AMS 1991 subject classification: Primary 62 J 05; Secondary 62 G05.
Short title: CMP model for MCR data.
Key Words: Right-censorship, competing risks models.

Abstract. We study the existing models for right-censored competing risks data and with masked failure cause. By introducing a new random variable hidden behind the current models, we give a practical interpretation of the symmetry assumption made by almost all researchers in this field. We further point out that the drawback of the symmetry assumption is that it has a strong restriction on the underlying distribution function to be studied. Moreover, we correct an assumption in the current models.

1. *Department of Mathematical Sciences, SUNY, Binghamton, NY* 13902, *USA*;

2. *Department of Biostatistics, UNC at ChapelHill, NC* 27514, *USA*

1. Introduction. We study the existing models for right-censored (RC) competing risks data with masked failure cause, called RMCR data hereafter. Friedman and Gertsbakh (1980) are the pioneers on the study of RMCR. Since then the study of competing risks data has become an important field in survival analysis and reliability theory. Mukhopadhyay (2006) gives an extensive review of the literature.

An RMCR observation consists of the observation on the failure time T and the observation on the associated failure cause C of a J-component series system. A J-component series system is a system that stops functioning as soon as one of its constituent J components fails. Let X_j be the lifetime of the j^{th} component, $j = 1, ..., J$. We assume that $P(X_i = X_j) = 0$ if $i \neq j$. Then (T, C) satisfies $T = \min\{X_1, ..., X_J\} = X_C$. In engineering and medical applications, X_1, ..., X_J may or may not be independent.

In reality, the failure time may be right censored by a censoring variable R, and the observation on T is (V, δ), where V is the minimum of T and R, and δ ($\stackrel{def}{=} \mathbf{1}_{(T\leq R)}$) is the indicator function of the event $\{T \leq R\}$. Moreover, one may not be able to observe C exactly. For instance, in examining a failed system, one may narrow down the failure cause to a subset $\mathcal{M}$ of $\mathcal{C}_r$ $(= \{1, ..., J\})$ and stop for various practical reasons. It is then said that at the failure time T the failure cause C is masked by $\mathcal{M}$. If $T \leq R$, $\mathcal{M}$ is the observation on C. Otherwise, we assume that $\mathcal{M}$ is missing and we know nothing about C.

Hereafter, denote $\mathcal{J}$ the collection of all the subsets of $\mathcal{C}_r$, including $\mathcal{C}_r$ but not the empty set $\emptyset$. Denote $S_T(t) = P(T > t)$, $F_{T,C}(t, c) = P\{T \leq t, C \leq c\}$ with its density function (df) $f_{T,C}$, $f_{\mathcal{M}|T,C}(A|t, c) = P\{\mathcal{M} = A|T = t, C = c\}$ and $f_{\mathcal{M}|C}(A|c) = P\{\mathcal{M} = A|C = c\}$. Moreover, denote the "cdf" $F_{T,C,\mathcal{M},R}$ in an obvious way, though $\mathcal{M}$ is a random set, not a random variable. The so-called symmetry assumption made by almost all of the research concerned with masking is S1 as follows.

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)).

In the literature, the n observations on T are specified by (V_1, δ_1), ..., (V_n, δ_n) and the n observations on C are specified by $\mathcal{M}_i$ if $\delta_i = 1$ and are missing otherwise. In order to specified an independent assumption on observations, it is assumed in the literature (see Mukhopadhyay, 2006) that

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

Based on S1, S2 and S3, the likelihood of $f_{T,C}$ becomes

$$\begin{aligned}
\mathcal{L} =& \Big(\prod_{i=1}^{n_1}\{\int_{t=V_i\leq u,\ c\in\mathcal{M}_i} dF_{T,C,\mathcal{M},R}(t, c, \mathcal{M}_i, u)\}\Big) \prod_{i>n_1}^{n} P\{T > V_i = R\} \quad \text{(by (S3))} \quad (1.1)\\
=& \Big(\prod_{i=1}^{n_1}\{\sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C,\mathcal{M}}(t, c, \mathcal{M}_i)S_R(V_i-)\}\Big) \prod_{i>n_1}^{n} S_T(V_i)f_R(V_i) \quad \text{(by S2)}\\
\propto& \Big(\prod_{i=1}^{n_1}\{\sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C}(t, c)\nu_{t,c}(\mathcal{M}_i)\}\Big) \prod_{i>n_1}^{n} S_T(V_i) \quad (\stackrel{def}{=} \Lambda)
\end{aligned}$$

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

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

(see, *e.g.*, Flehinger *et al*, 2001 and Sen *et al*, 2001 for more details on the derivation above).

Yu *et al.* (2010) call the model that assumes S1, S2 and S3 the Conditional Masking Probability (CMP) model 1. The model leads to the likelihood $\mathcal{L}_c$. Moreover, they name the model that assumes S1b, S2 and S3 the CMP model 2. The latter model leads to the likelihood Λ_c in (1.1).

The CMP Model 2 has more parameters than the CMP model 1, and thus it is more flexible. Guttman *et al.* (1995) and Kuo and Yang (2000) make MLE and Bayesian inferences based on CMP model 2. However, unless additional constraints are imposed, the parameter under CMP Model 2, can be non-identifiable (see Example 11 in Yu *et al.*, 2010). Thus CMP model 2 is useful under the assumption that there are stage-2 data or prior information.

In the literature in order to justify the likelihood Λ_c or $\mathcal{L}_c$, 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]). Yu *et al.* (2010) study its practical justification under the new random partition masking (RPM) model they proposed. Since the CMP models are not special cases of the RPM model, we shall reinvestigate the practical justification of the symmetry assumption without the restriction of the RPM model in the short note.

2. The Main Results. The original formulation of the CMP models is based on S3 as well as S1 and S2 (see (1.1)). Notice that S3 is related to $P(T\le R)$. If $P(T\le R)=1$, there is no censoring, thus it is not of interests, as long as the RMCR data are concerned. If $P(T\le R)=0$ then it implies that there exists a point t such that $P(R\le t<T)=1$. It follows that the right-censored sample does not provide any information for T and C and should be deleted. Thus the CMP models are only relevant if $P(T\le R)\in(0,1)$.

Proposition 1. *S3 is false if* $P(T\le R)\in(0,1)$.

In view of Proposition 1, S3 is a false statement. However, its role can be replaced by another assumption to be specified next. An interesting feature of the RMCR data is that

1. T is not always observable and is censored by R, which is not always observable too.
2. C is not always observable and is masked by $\mathcal{M}$, which may be missing too.

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 set on C by $\mathcal{M}^o$ where $\mathcal{M}^o=\begin{cases}\mathcal{M} & \text{if } T\le R\\ \mathcal{C}_r & \text{if } T>R.\end{cases}$ Mukhopadhyay (2006) maybe the first person who points out that $\mathcal{M}\ne\mathcal{C}_r$ if $T>R$. If one defines $\mathcal{M}=\mathcal{C}_r$ when $T>R$, then S2 becomes false. But he does not define $\mathcal{M}^o$. $\mathcal{M}$ 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. S3 should be replaced by the assumption as follows

$$(V_i,\delta_i,\mathcal{M}_i)\text{'s are i.i.d. copies of } (V,\delta,\mathcal{M}^o). \qquad (2.1)$$

The likelihood of the observed RMCR data can be written as

$$\Lambda = \prod_{i=1}^{n} \big(\{ \sum_{t=V_i,\ c\in\mathcal{M}_i} f_{T,C}(t,c)\nu_{t,c}(\mathcal{M}_i)\}\big)^{\delta_i} (S_T(V_i))^{1-\delta_i},$$

which is the same as the Λ in (1.1).

To understand the CMP models, we first define some notations. Notice that $\mathcal{M}$ is associated with at least one random partition of $\mathcal{C}_r$, namely, $P_r = \begin{cases} \{\mathcal{M},\mathcal{M}^c\} & \text{if } \mathcal{M} \neq \mathcal{C}_r, \\ \{\mathcal{M}\} & \text{otherwise} \end{cases}$, where $\mathcal{M}^c = \mathcal{C}_r \setminus \mathcal{M}$. 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 $P_h = \{P_{h1}, ..., P_{hk_h}\}$, $P_{hi} \in \mathcal{J}$, $\cup_{i=1}^{k_h} P_{hi} = \mathcal{C}_r$ and $P_{hi} \cap P_{hj} = \emptyset\ \forall\ i \neq j$. Notice that $\{\mathcal{C}_r, \emptyset\} \notin \mathcal{P}$. By definition, for each given partition P_h and given $C = c$, there exists an i such that $C = c \in P_{hi}$. For instance, $P_1 = \{\{1\},\{2\}, ..., \{J\}\}$, $P_2 = \{\{1\},\{2\},\{3,4,...,J\}\}$ 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 $\mathcal{M} = \begin{cases} \{1\} & \text{if } C = 1 \\ \{2\} & \text{if } C = 2 \\ \{3,4,...,J\} & \text{otherwise.} \end{cases}$

In other words, an inspection scheme for the system corresponds to a partition. After an inspection scheme is chosen, that is, after a partition $(P_{h1}, ..., P_{hk_h})$ 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.

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, ..., $P_{n_{\mathcal{P}}}$. Let Δ be a random variable taking values in $\{1, ..., n_{\mathcal{P}}\}$ with df f_Δ. The value of $f_\Delta(h)$ can be viewed as the proportion in the population that inspection schemes corresponding to partition P_h have been taken. Then

$$(V, \delta, \mathcal{M}^o) = \begin{cases} (v, 0, \mathcal{C}_r) & \text{if } T > R = v, \\ (v, 1, P_{hj}) & \text{if } T = v \leq R,\ \Delta = h \text{ and } C \in P_{hj}. \end{cases} \tag{2.2}$$

CMP model 1 should be formulated by assumptions (2.1) and (2.2) together with S1 and S2, rather than S1, S2 and S3. CMP model 2 should be formulated by assumptions (2.1) and (2.2) together with S1b and S2, rather than S1b, S2 and S3.

To give a practical justification of the symmetry assumption, it suffices to study the relation between S1, S2 and Δ, we make use of the following assumptions:

A1 $(T, C) \perp (R, \Delta)$.
A2 for each t, $f_T(t) > 0$ implies $f_{T,C}(t, c) > 0$ for each $c \in \mathcal{C}_r$.
A3 For each $A \in \mathcal{J}$, $\sum_{h:\ A \in P_h} f_{\Delta|C}(h|c)$ is constant in c if $c \in A$.
A4 For each $A \in \mathcal{J}$, $\sum_{h:\ A \in P_h} f_{\Delta|T,C}(h|t, c) = \sum_{h:\ A \in P_h} f_{\Delta|C}(h|c)$ if $c \in A$.
A5 $(\mathbf{1}_{(C \in A, A \in P_\Delta)}, T) \perp R\ \forall\ A \in \mathcal{J}$, where $P_\Delta = P_h$ if $\Delta = h$.

Lemma 1. $\nu_c(A) = \begin{cases} \sum_{h:\ A \in P_h} f_{\Delta|C}(h|c) & \textit{if } f_C(c) > 0 \\ \textit{undefined} & \textit{if } f_C(c) = 0 \end{cases} \forall\ A \in \mathcal{J}$ *and* $c \in A$.

Lemma 2. $\nu_{t,c}(A) = \begin{cases} \sum_{h:\ A \in P_h} f_{\Delta|T,C}(h|t, c) & \textit{if } c \in A \textit{ and } f_{C|T}(c|t) > 0 \\ \textit{undefined} & \textit{if } f_{C|T}(c|t) = 0\ \&\ f_T(t) > 0, \end{cases} \forall\ A \in \mathcal{J}.$

Lemma 3. *Assumption S1b implies A2.*

Proof. By Lemma 2, S1b implies A4. In A3 and A4, $f_{\Delta|C}(h|c)$ and $f_{\Delta|T,c}(h|t, c)$ are needed. If $f_T(t) > 0$ but $f_{C|T}(c|t) = 0$ for some t, which is possible, then $f_{\Delta|T,C}(h|t, c) = \frac{f_{\Delta,T,C}(h,t,c)}{f_{T,C}(t,c)} = \frac{0}{0}$. One might consider to make a new convention to $\frac{0}{0}$. Since $f_{\Delta|T,C}(h|t, c) = \frac{0}{0}$ for each h and they should satisfy the constraint $\sum_h f_{\Delta|T,C}(h|t, c) = 1$ and thus $\sum_h \frac{0}{0} = 1$. As a consequence, $f_{\Delta|T,C}(h|t, c) = \frac{0}{0} = 1/||\mathcal{P}||$. However,

$$\frac{1}{||\mathcal{P}||} \neq 1 = \sum_h f_{\Delta|T,C}(h|t, c) = \frac{\sum_h f_{\Delta,T,C}(h, t, c)}{f_{T,C}(t, c)} = \frac{f_{T,C}(t, c)}{f_{T,C}(t, c)} = \frac{0}{0}.$$

That is, $\frac{0}{0} = 1 \neq \frac{1}{||\mathcal{P}||}$. As a conclusion, S1b implies A2. □

Theorem 1. Under formulation (2.2),

1. S1a <=> A3;
2. S1b <=> A2 and A4;
3. S2 <=> A5.

Proof. The theorem follows from Lemmas 1, 2, and 3. □

By Theorem 1, CMP model 1 is equivalent to a model that assumes (2.1), (2.2), A2, A3, A4 and A5, and CMP model 2 is equivalent to a model that assumes (2.1), (2.2), A2, A4 and A5.

The disadvantage of the CMP models is that it requires A2 and thus it is more restrictive on $F_{T,C}$. Yu *et al.* (2010) propose a different model that the RMCR data are generated by (2.1) assuming (2.2) and A1. It leads to the same likelihood function $\mathcal{L}_c$ in (1.1), but not Λ_c. They call the model the random partition masking (RPM) model. The RPM model does not require A2 and allows $F_{T,C}$ to be arbitrary.

Proposition 2. (Yu *et al.* (2010)). *When $J = 2$, CMP model 1 implies A1, A2 and $\Delta \perp R$. Thus it is a special case of the RPM model if $J = 2$.*

Proposition 3. *Suppose $J = 3$. CMP model 1 is not a special case of the RPM model and the RPM model is not a special case of CMP model 1.*

Concluding Remark. Our study indicates that both the old CMP model 1 and the RPM model can be put in the same frame work, making use of the random vector (Δ, R). In view of Proposition 3, in general, the RPM model and CMP model 1 are two different models. Our study in this paper shows that they both have their own merits. The RPM but not the CMP model allows $F_{T,C}$ arbitrary. On the other hands, the CMP models weaken the assumption $(T, C) \perp (\Delta, R)$. This is clear now by Proposition 3.

We are not proposing a new model for the RMCR data. Instead, we are trying to understand what the old CMP model 1 means in terms of their symmetry assumption. Thus the estimation and the inferences of the parameters in the model based on the new interpretation will remain the same as the old one.

The study is useful in studying the properties of the estimators. For instance, in the current literature, it is not known that the CMP models have the restriction A2 on $F_{T,C}$. Moreover, after one has a better understanding on the models, one may design a more realistic model for the RMCR data, taking advantages of both the CMP models and the RPM model. The authors are working on the new model.

3. Proofs.

Proof of Proposition 1. S3 intends to say that the observations are independent and the first n_1 have exact T_i's and the remaining have right-censored T_i's. Thus it follows that δ_1, ..., δ_n are independent and the first n_1 δ_i's are 1 and the rest of them are 0. S3 can be interpreted in two ways. (1) n_1 is fixed and thus there are two independent samples; (2) n_1 is random, and $n_1 = n$ with positive probability.

In case (1) $P\{R \geq T\} > 0$ means that with positive probability there are observations with $\delta_i = 1$ among those $n - n_1$ observations. It contradicts S3 as S3 implies that $\delta_{n_1+1} = \cdots = \delta_n = 0$ w.p.1. Thus S3 is false.

In case (2), the n_1^{th} and $(n_1+1)^{th}$ observations cannot be independent, as

$$\begin{aligned} P(\delta_1 = 1)P(\delta_2 = 0) =& P(\delta_1 = 1, \delta_2 = 0) && \text{(by S3)} \\ =& P(\delta_1 = 1, \delta_2 = 0, ..., \delta_n = 0) && \text{(by the definition of } n_1) \\ =& P(\delta_1 = 1)P(\delta_2 = 0) \cdots P(\delta_n = 0) && \text{(by S3)} \end{aligned}$$

It follows that $P(\delta_i = 0)$ is either 0 or 1, violating the assumption that $P(\delta_i = 0) \in (0, 1)$. □

Proof of Lemma 1. For each $A \in \mathcal{J}$ and $c \in A$, if $P(C = c) = 0$ then $\nu_c(A) = P\{\mathcal{M} = A|C = c\}$ is undefined; otherwise

$$\begin{aligned} \nu_c(A) =& P\{\mathcal{M} = A|C = c\} \\ =& \sum_h P\{\mathcal{M} = A, \Delta = h|C = c\} \\ =& \sum_h P\{C \in A, A \in P_h, \Delta = h|C = c\} \text{ (by the definitions of } \Delta \text{ and } \mathcal{M}) \\ =& \sum_{h:\ A \in P_h} P\{C \in A, \Delta = h|C = c\} \qquad \text{(as } A \text{ and } P_h \text{ are not random)} \\ =& \sum_{h:\ A \in P_h} P\{C = c, \Delta = h|C = c\} \\ =& \sum_{h:\ A \in P_h} f_{\Delta|C}(h|c). \ \square \end{aligned}$$

Proof of Lemma 2. For each $A \in \mathcal{J}$ and $c \in A$, if $f_{T,C}(t, c) = 0$ and $f_T(t) > 0$, $\nu_{t,c}(A) = P\{\mathcal{M} = A|C = c, T = t\}$ is undefined.

For each $A \in \mathcal{J}$ and $c \in A$, if $f_{T,C}(t,c) > 0$ then

$$\begin{aligned}
\nu_{t,c}(A) =& P\{\mathcal{M} = A | C = c, T = t\} \\
=& \sum_h P\{\mathcal{M} = A, \Delta = h | C = c, T = t\} \\
=& \sum_h P\{C \in A, A \in P_h, \Delta = h | C = c, T = t\} \text{ (by the definitions of } \mathcal{M} \text{ and } \Delta) \\
=& \sum_{h:\ A \in P_h} P\{C = c, \Delta = h | C = c, T = t\} \qquad \text{(as } A \text{ and } P_h \text{ are not random)} \\
=& \sum_{h:\ A \in P_h} P\{\Delta = h | C = c, T = t\} \\
=& \sum_{h:\ A \in P_h} f_{\Delta|T,C}(h|t,c). \ \square
\end{aligned}$$

Proof of Proposition 2. When $J = 2$, $||\mathcal{P}|| = 2$. Let $P_1 = \{\{1\},\{2\}\}$ and $P_0 = \{\mathcal{C}_r\}$. S1 says $f_{\mathcal{M}|T,C}(\mathcal{C}_r|t,c) = f_{\mathcal{M}}(\mathcal{C}_r)\ \forall\ (t,c)$. Thus $(T,C) \perp \mathbf{1}_{(\mathcal{M}=\mathcal{C}_r)}\ (= \mathbf{1}_{(\Delta=0)})$. It implies that $(T,C) \perp \mathbf{1}_{(\Delta=1)}$, and thus $(T,C) \perp \Delta$. Moreover, S2 yields $(T, \mathbf{1}_{(C=c,\Delta=1)}) \perp R$ for $c = 1,2$. Thus $(T, C, \mathbf{1}_{(\Delta=1)}) \perp R$, which yields $(T, C, \mathbf{1}_{(\Delta=0)}) \perp R$. Thus $(T,C,\Delta) \perp R$. As a summary, if $J = 2$, CMP model 1 requires A1, $R \perp \Delta$ and A2 (see Lemma 3). $\square$

Proof of Proposition 3. Since the RPM model does not require A2, but CMP models do, the RPM model is not a special case of the CMP models.

We shall now find a case that it belongs to the CMP model, but not the RPM model. In particular, we shall present a case that S1 and S2 hold but not A1.

Assume A2 so that S1 is possible to hold. Assume $(C,\Delta) \perp R$ and $T \perp \Delta$. Then S2 holds and $\nu_{t,c}(A) = \sum_{A \in P_h} f_{\Delta|T,C}(h|t,c) = \sum_{A \in P_h} f_{\Delta|C}(h|c) = \nu_c(A)\ \forall\ A$ with $c \in A$.

If $J = 3$, $\mathcal{J}$ contains 7 elements: $\mathcal{C}_r$ and elements of forms $\{2,3\}$, $\{1,3\}$, $\{1,2\}$, $\{1\}$, $\{2\}$ and $\{3\}$, denoted by A_0, ..., A_6. Verify that $||\mathcal{P}|| = 5$ and $\mathcal{P}$ contains $\mathcal{C}_r$ (P_0), $\{\{1\},\{2\},\{3\}\}$ (P_4), $P_c = \{\{c\},\{i,j\}\}\ \forall$ distinct (c,i,j) with $i < j$ and $c,i,j \in \mathcal{C}_r$. Verify that

$$\begin{cases} \nu_{t,c}(A_i) = a_i, & c \in A_i \text{ and } a_i \text{ is a constant, } i = 0,...,6 \text{ (by S1)}, \\ \sum_{A_i} \nu_{t,c}(A_i) = 1, & c \in \mathcal{C}_r \text{ (as } \nu_{t,c} \text{ is a conditional density function)}. \end{cases} \tag{3.1}$$

Since $\Delta \perp T$, (3.1) yields

$$\begin{aligned}
& f_{\Delta|C}(0|1) = f_{\Delta|C}(0|2) = f_{\Delta|C}(0|3) = a_0 \\
& f_{\Delta|C}(1|2)\ \big\{ = \nu_{t,2}(\{2,3\}) = \nu_{t,3}(\{2,3\}) \big\}\ = f_{\Delta|C}(1|3) = a_1, \\
& f_{\Delta|C}(2|1)\ \big\{ = \nu_{t,1}(\{1,3\}) = \nu_{t,3}(\{1,3\}) \big\}\ = f_{\Delta|C}(2|3) = a_2, \\
& f_{\Delta|C}(3|1)\ \big\{ = \nu_{t,1}(\{1,2\}) = \nu_{t,2}(\{1,2\}) \big\}\ = f_{\Delta|C}(3|2) = a_3, \\
& f_{\Delta|C}(4|1) + f_{\Delta|C}(1|1) = \nu_{t,1}(\{1\}) = a_4, \\
& f_{\Delta|C}(4|2) + f_{\Delta|C}(2|2) = \nu_{t,2}(\{2\}) = a_5, \\
& f_{\Delta|C}(4|3) + f_{\Delta|C}(3|3) = \nu_{t,3}(\{3\}) = a_6.
\end{aligned} \tag{3.2}$$

Verify that if $a_0 = 0.1$, $a_1 = a_2 = a_3 = 0.2$ and $a_4 = a_5 = a_6 = 0.5$ then (3.1) holds. If one further sets $\begin{array}{ll} f_{\Delta|C}(4|1) = 0.1, & f_{\Delta|C}(1|1) = 0.4, \\ f_{\Delta|C}(4|2) = 0.2, & f_{\Delta|C}(2|2) = 0.3, \\ f_{\Delta|C}(4|3) = 0.3, & f_{\Delta|C}(3|3) = 0.2. \end{array}$ then (3.2) still holds and thus S1 holds. However, it is obvious that $\Delta \not\perp C$, and thus A1 fails. This implies that CMP model 1 is not a special case of the RPM model if $J = 3$. □

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.

Flehinger, B.J., Reiser, B., and Yashchin, E. (2001). Statistical analysis for masked data. In: Balakrishnan N, Rao CR (eds) Handbook of Statistics Vol. 20 499-522. Elsevier Science. Amsterdam.

Friedman, L. and 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., Dependent masking and system life data analysis: Bayesian inference for two-component systems. *Lifetime Data Analysis* 1 87-100.

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

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

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

Sen, A. Basu, S. and Benerjee, M. (2001). Analysis of masked failure time data under competing risks. In: Balakrishnan N, Rao CR (eds) Handbook of Statistics 20 523-540. Elsevier Science. Amsterdam.

Yu, Q.Q., Wong, G.Y.C., Qin, H. and Wang, J. (2010). Random partition masking model for censored and masked competing risks data. *Annals of Institure of Mathematical Statistics* (Accepted). ftp://ftp.math.binghamton.edu/pub/qyu/para.mle.pdf