

A Bayesian approach to competing risks analysis with masked cause of death

Ananda Sen,^a Mousumi Banerjee,^{b*†} Yun Li^b and Anne-Michelle Noone^c

Cause-specific analyses under a competing risks framework have received considerable attention in the statistical literature. Such analyses are useful for comparing mortality patterns across racial and/or age groups. Earlier work in the statistical literature focused on the situation when the cause of death is known. A challenging twist to the problem arises when the cause of death is not known exactly, but can be narrowed down to a set of potential causes that do not necessarily act independently. This phenomenon, referred to as *masking*, is often the result of incomplete or partial information on death certificates and/or lack of routine autopsy on every patient. In this article we propose a semiparametric Bayesian approach for analyzing competing risks survival data with masked cause of death. The models proposed do not assume independence among the causes, and are valid for an arbitrary number of causes. Further, the Bayesian approach is flexible in allowing a general pattern of missingness for the cause of death. We illustrate our methodology using breast cancer data from the Detroit Surveillance, Epidemiology, and End Results registry. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: competing risks; masked cause of death; Markov chain Monte Carlo; semiparametric Bayesian analysis

1. Introduction

Cause-specific analysis under a competing risks framework has received considerable attention in the statistical literature [1–3]. David and Moeschberger [4] in their seminal book introduced and integrated the theory behind the general competing risks model. Since then numerous investigations dealing with the issues and nuances of a competing risks model have been carried out [5–9]. Much of the earlier work in this area have focused on the situation when the exact cause of death is known. A challenging twist to this problem arises when the cause of death is not known exactly but can be narrowed down to a set of potential causes. This phenomenon, referred to as *masking*, is often the result of incomplete or partial information on death certificates and/or lack of routine autopsy on every patient. Masked data frequently arise in population-based registries, such as the *Surveillance, Epidemiology, and End Results* (SEER) cancer registry that captures incidence and mortality data on cancers in the United States. Approximately 16 per cent of the deceased breast cancer cases in the SEER registry do not have exact information on the underlying cause of death due to missing or partial information on the state death certificate. Thus their cause of death is not known exactly. However, for many of these women information on related causes of death is captured in six additional fields in SEER. The related causes of death provide a set of ‘potential’ causes that a woman could have died from. This presents an example of masked survival data. The survival study of heart transplant patients [10] and the longitudinal study on breast cancer patients [11] are two other illuminating examples of masked survival data available in the literature. Masked survival data are also quite common in industrial applications [12–14], where most of the methodological work has focused on fully parametric models without covariate adjustment. Developments under a semiparametric framework that incorporates covariate information have been detailed in [15–19].

In this article, we develop a semiparametric Bayesian approach in the framework of cause-specific hazard function, where the partial information about the cause of death is incorporated by means of latent variables. We propose to use a simulation-based approach that exploits recent advances in Markov Chain Monte Carlo (MCMC) techniques, to implement the Bayesian methodology.

^aDepartment of Statistics and Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, MI, U.S.A.

^bDepartment of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

^cDepartment of Biostatistics, Bioinformatics, and Biomathematics and Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, U.S.A.

*Correspondence to: Mousumi Banerjee, Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

†E-mail: mousumib@umich.edu

Contract/grant sponsor: Blue Cross Blue Shield Foundation of Michigan

The prior specifications follow standard practice in semiparametric Bayesian analysis. Section 2 describes the model formulation. In Section 3 we present details of the Bayesian analysis, specifically prior models for the various parameters, and details of the MCMC technique for posterior analyses. To illustrate our methodology, in Section 4 we present analyses of a cohort of breast cancer cases from the Detroit SEER registry. A hybrid of Bayesian and frequentist methodology is used in the implementation steps in order to obtain the estimates of model parameters. Finally, Section 5 contains some discussion and an appraisal of the approach adopted in the current article in contrast to other approaches used in this context.

2. Model formulation

Earlier work in competing risks analysis utilized a *series system* (observing the minimum of several lifetimes) formulation in terms of latent failure times. Tsiatis [20] warned about the identifiability issue that underlies such a formulation. Since then, substantial efforts have been made to formulate a model that has direct links to the observables.

In order to understand the basics of a competing risks formulation under masked cause of death, we first need to lay down some notational preliminaries. Suppose we have n subjects under study each of whom have J competing risks of death acting on them. In what follows, let T, C denote the failure time and cause, respectively. One key ingredient in the competing risks analysis is the cause-specific hazard defined as:

$$\lambda_j(t) = \lim_{h \rightarrow 0} \frac{\Pr(t \leq T \leq t+h, C=j | T \geq t)}{h}, \quad j = 1, \dots, J.$$

The corresponding cumulative incidence function is

$$F_j(t) = \Pr(T \leq t, C=j) = \int_0^t \lambda_j(u) R(u) du, \tag{1}$$

where $R(u) = \Pr(T \geq u)$ is the overall survival function. The associated overall hazard rate $\lambda(t) = -d \log R(t)$ is related to the cause-specific hazard rates by the equation $\lambda(t) = \sum_{j=1}^J \lambda_j(t)$. Note that under the (non-testable) assumption of independence of the risks, the series system formulation mentioned above is identifiable and in that case the cause-specific hazard λ_j matches the marginal hazard function associated with the j th latent lifetime random variable. In recent research on survival analysis with masked cause of death, (1) has sometimes been used as a descriptor of the cause-specific nature of the framework [14].

Cause-specific hazard-based analysis and cumulative incidence function-based analysis are two alternative, albeit somewhat different avenues of handling competing risks data. Fine and Gray [21] describe how the cause-specific regression formulation can be turned into a regression model for cumulative incidence function via a complimentary log–log transformation. Many authors, however, noted that despite this connection the co-variate effects based on one formulation can qualitatively be quite distinct from that based on the other. Our strategy in this article is to focus on a single framework, namely the cause-specific formulation. One can certainly develop a parallel Bayesian methodology under a regression setting for the cumulative incidence function. We do not pursue this in the current article.

In the case of masked cause of death, the cause can be known up to a *minimum random subset* $S \subset \{1, 2, \dots, J\}$ of the J risks. Note that there are a total of $2^J - 1$ possible such S 's. If the exact cause of death is known to be j , then $S = \{j\}$ is a singleton. On the other extreme, if no knowledge is available on the cause of death, $S = \{1, \dots, J\}$, resulting in *full masking* of the cause. For each individual we observe the vector (T, S, X, Δ) , where X is the observed collection of covariate values for the individual, and Δ is his/her censoring status ($= 1$ if died, $= 0$ if censored). Note that $\Delta = 0$ would imply $S = \{1, \dots, J\}$, by definition.

Throughout this article we assume that the censoring distribution (a) is independent of, and (b) does not share any parameter with the overall survival distribution. In our context of studying death registry, this seems like a reasonable assumption. The joint likelihood function of the observed data $(T_i, S_i, X_i, \Delta_i)_{i=1}^n$ can be written as:

$$L = \prod_{i=1}^n \left\{ \sum_{j \in S_i} P(S_i | T_i, X_i, C_i = j) \lambda_j(T_i, X_i) \right\}^{\Delta_i} \exp \left\{ - \sum_{j=1}^J \int_0^{T_i} \lambda_j(t, X_i) dt \right\}. \tag{2}$$

Given the knowledge of time and true cause of death, the *masking probability* $q_j(s_j; X_i, T_i) \equiv P(S_i = s_j | T_i, X_i, C_i = j)$ signifies the chance of the cause being masked by s_j . This quantity is difficult to handle as no natural estimate of it can be found based upon the sampled data. Further, treating them as completely unspecified nuisance parameters gives rise to an over-parameterized problem. A compromise adopted in the literature is the so-called 'symmetry' assumption that entails an equal chance of observing the same masked subset of risk components irrespective of the true cause (within the masking subset) of death, i.e.

$$q_j(s_j; X_i, T_i) = q_{j'}(s_j; X_i, T_i), \quad j, j' \in s_j. \tag{3}$$

In addition to the symmetry assumption, Guess *et al.* [22] and Lin *et al.* [23] further assumed q to be independent of T, X and cited industrial examples where it was reasonable to make such assumptions. Dewanji and Sengupta [24] derived Nelson–Aalen-type estimators of the cause-specific cumulative hazard functions for discrete survival data without assuming (3), but their likelihood-based inference made explicit use of the symmetry assumption. Flehinger *et al.* [25, 26] avoided making the assumption (3) under the assumption of availability of data from a second-stage autopsy on a subset of deceased individuals in order to

pinpoint their true cause of death. In this article, we shall provide a Bayesian solution to the problem that does not need to rely on any assumption such as (3) or any second-stage data. Further, we incorporate heterogeneity in the masking probabilities by allowing them to depend on subject-level covariates.

Note that we shall work under the cause-specific formulation of (2), which precludes any *a priori* assumption of independence among risks. In this sense, our solution can be thought of as an extension of Kuo and Yang [27]. Henceforth, for simplicity of exposition, we shall suppress the argument for the q -probability and simply use appropriate indices to indicate its functional dependence on relevant quantities.

For the cause-specific hazards, we choose the popular proportional hazards (PH) formulation, namely,

$$\lambda_j(u, X) = \lambda_{0j}(u) \exp(\beta_j' X). \quad (4)$$

Note here that λ_{0j} is the baseline hazard specific to cause j , and is allowed to vary across different causes. Similar comment applies for the regression coefficient β_j . Also note that we use the entire collection X of covariates in formulating (4) merely for the simplicity of exposition, and the set of predictors for different causes need by no means be identical.

3. Bayesian approach

In order to carry out the Bayesian analysis, we need to prescribe the prior models for various parameters.

3.1. Masking probabilities

Turning our attention to the masking probabilities first, we note that for a given (latent) cause of death $C=j$ there are in all 2^{J-1} potential subsets S that would contain j . We shall explore two models for q 's that we describe below.

qDir: Here we model the vector of the masking probabilities as a Dirichlet random variable. More specifically,

$$q_j(s_i; X_i, T_i) \sim \text{Dir}_{2^{J-1}}(\theta_j), \quad j = 1, \dots, J.$$

This model posits that the masking probabilities are independent of the death times as well as the covariate level. It is, of course, possible to extend this model to the subject level, where the masking probabilities sharing the same latent cause j of death for different subjects are viewed as independent random realizations from the above Dirichlet distribution. This is akin to a random subject effect specification, and accounts for more model variation than a specification free of subject effect.

More specifically, given the (latent) true cause of death j , let $s_j = \{s_{j1}, \dots, s_{jk}\}$ denote the collection of potential subsets of causes which contain j . Note that k is at most 2^{J-1} . In this article, we entertain two sub-models in the class *qDir*. The first assumes

$$(q_j(s_{j1}), \dots, q_j(s_{jk})) \sim \text{Dir}_k(\theta_j), \quad (5)$$

while the second formulation entails

$$(q_{ij}(s_{j1}), \dots, q_{ij}(s_{jk})) \sim \text{Dir}_k(\theta_j). \quad (6)$$

In (5), the masking probabilities are identical for different subjects sharing the same underlying cause of death, drawn from the Dirichlet distribution all at once. In contrast, under (6), the masking probabilities for different subjects sharing the same latent cause of death are obtained using independent draws from the Dirichlet and hence are likely distinct.

qReg: As explained earlier, there is no natural empirical estimate for the masking probabilities. Thus, in order to borrow strength indirectly from the auxiliary information, we assume a logistic regression model for the masking probabilities with random regression coefficients. For the two-cause case, Lu and Tsiatis [19] and Tsiatis *et al.* [28] considered a similar regression model, albeit under a frequentist framework. Further, they assumed a version of the symmetry assumption (3) which was required for the large-sample inference to work.

3.2. Baseline hazard

In the PH framework, we can model the baseline hazard and the regression parameters separately. A common prior model for the baseline hazard is the Gamma process popularized by Kalbfleisch [29]. Despite its limitations, it is used on a regular basis, primarily due to its attractive statistical properties. We shall pursue two variants of the baseline hazard formulation in our case.

PB: The proportional baseline (PB) postulates the relation

$$\Lambda_{0j}(u) = \Lambda_0(u) p_j, \quad j = 1, \dots, J,$$

where Λ_{0j} is the cumulative baseline hazard for risk j , and Λ_0 is a nonnegative, nondecreasing function common to all risks. One important consequence of the PB model is the independence of the time and cause of death [30]. In this case, we assume the prior specification

$$\begin{aligned} \Lambda_0(u) &\sim \text{GP}(c\omega(u), c), \\ p_j &\sim \text{Gamma}(\alpha_j, \delta_j), \quad j = 1, \dots, J, \end{aligned}$$

where all variables are assumed to be independent. In the Gamma process specification, the mean process $\omega(u)$ is viewed as an initial estimate of Λ_0 , whereas c is a measure of degree of belief in the prior specification. Small values of c would reflect a higher level of uncertainty about the prior specification, and vice versa.

NPB: Non-proportional baseline model (NPB) does not assume any relation between the cause-specific baselines. In this case, we model each baseline as an independent Gamma process.

3.3. Regression coefficients

The regression coefficients are typically modeled as independent Normal or t variables, unless there is *a priori* reason to envision them as skewed. In the latter case skew-normal or skew- t are reasonable prior choices for the β parameters. Also note that our methodology allows complete flexibility to choose different distributions for the β parameters associated with the same covariate, but different risks.

All group of parameters described above are further assumed to be stochastically independent. Elicitation of the prior parameters has remained an inherent part of the prior selection process. Since a strong prior information can drive the direction of analysis in a significant manner, it is imperative that reasonably accurate information on the prior parameters be obtained wherever possible. In our context, one naive but often useful prior elicitation method may consist of extracting summary information (e.g. mean, variance, percentile) on cause-specific survival from historical data or previous case studies and then translating the information into parameter values for the prior distribution. In situations where such information is unavailable, one acceptable strategy is to make the prior information sufficiently 'diffused' (by assuming a large variance), so that the effect of the prior on the end-analysis is not over-bearing.

3.4. Posterior analysis

The posterior estimators of the various parameters are not in tractable form. We adopt the simulation-based MCMC technique that relies on generating random draws from the relevant full conditional distributions in an iterative manner. The Bayesian approach is greatly facilitated by data augmentation technique whereby exact cause of death C_i is imputed for the subject i for whom the cause is missing. The imputation probabilities are easily calculated from (2) as

$$\Pr(C_i = j | X_i, S_i = s_i, T_i) = \frac{q_j(s_i; X_i, T_i) \lambda_j(T_i, X_i)}{\sum_{j' \in s_i} q_{j'}(s_i; X_i, T_i) \lambda_{j'}(T_i, X_i)}, \quad j \in s_i. \quad (7)$$

Note that (7) has appeared in the literature under the name *diagnostic probability* [25, 26]. The full conditional distributions are obtained from the likelihood augmented by the latent cause generated from (7).

3.5. Full conditionals

q_j: In the case of Dirichlet prior for the masking probabilities, the full conditional distributions are also Dirichlet, whether the modeling is at the subject-level or not. No such conjugacy obtains in the case of *qreg* model. It is important to note, however, that either for the *qdir* or the *qreg* model, the full conditional distribution of the masking probabilities are stochastically independent of those of the baseline intensity and the regression parameters β .

Baseline: The full conditional distributions of the baseline cumulative intensities are in tractable form for the Gamma process prior. They can be derived using straightforward adaptations of the representation in Kalbfleisch [29] to the competing risks case. For example, in the *PB* model, the full conditional distribution for Λ_0 is given for the *no tie* case by

$$\Lambda_0(t_i) \propto \sum_{j=1}^J \sum_{l=1}^i [X_{lj} + \Delta_{lj}(C_i = j)] U_{lj},$$

where

$$\begin{aligned} X_{lj} &\sim \text{Gamma}(c\omega(t_j) - c\omega(t_{j-1}), c + A_{lj}) \\ g(u_{lj}) &\propto (u_{lj})^{-1} [\exp(-(c + A_{l+1,j})u_{lj}) - \exp(-(c + A_{l,j})u_{lj})] \\ A_{lj} &= \sum_{m=l}^n \exp(\beta_j x_m). \end{aligned}$$

The representation in the case with ties is more tedious, and will not be presented here. Note that the constants of proportionality p_j for the *PB* model have gamma full conditionals.

β : The full conditionals for β are not in standard form. They are log-concave, however, if the corresponding priors are so too. Log-concavity, while not essential to run the MCMC, certainly enhances its speed by enabling one to employ specialized algorithms such as the ones developed by Gilks and Wild [31].

4. Analysis of breast cancer data

4.1. Description of the data

We analyzed data from a cohort study of women newly diagnosed with breast cancer between January 1990 and December 1998 in the Detroit SEER catchment area. During this time frame a total of 10620 women were diagnosed with breast cancer in the Detroit SEER area. Survival for these women was measured from the time at diagnosis to time of last follow-up or death, with a median follow-up of 13.2 years. For illustrative purposes, we analyze data from a sample of 500 women diagnosed with local or regional stage breast cancer from this population.

In order to illustrate our methodology, we over-sampled the ‘masked’ cases from the original population so that our final sample contains a sizable representation of the women with masked cause of death. Approximately 46 per cent of the women in our analytic cohort died during the course of follow-up, either due to *breast cancer*, *other cancer*, or *non-cancer related* causes. The exact cause of death was known for 56 per cent of these patients, with breast cancer and non-cancer causes largely contributing to these deaths. The underlying cause of death was missing for the remaining group. Of these, however, partial information was available on 35 per cent of the patients based on related cause of death fields reported in the state death certificate. This resulted in masked survival data. Specifically, in this data set, 30 per cent of the women with missing cause of death had (breast cancer, other cancer) as their masked (set) cause of death (for these patients the related cause of death fields listed other cancers or ‘multiple cancers’), 5 per cent had (breast cancer, non-cancer) as their masked (set) cause of death (for these patients the related cause of death fields listed ‘acute myocardial infarction’, ‘congestive heart failure’, ‘atherosclerosis’ etc.). In the original population, only about 0.01 per cent of women had (other cancer, non-cancer) as their masked (set) cause of death. Therefore, we dropped this small group of subjects from the sampling frame during sample selection, and further imposed model restrictions that excludes such a masking group (specifics explained later). For 65 per cent of the women with missing cause of death, the cause was completely masked (unknown, i.e. no additional information could be retrieved from the related cause of death fields).

Note that in this example we are dealing with three competing causes of death. All patients in the study were originally diagnosed with breast cancer. Thus, breast cancer is the primary cause of interest (our ‘main’ cause). Death from other cancers could either be related to the original breast cancer, or could be from independent diagnosis of some other cancer. Either way, it is scientifically natural to group the other cancers together. Subsequently, all other (non-cancer) causes of death were lumped in a single group. Breaking the non-cancer causes up in subgroups would have led to severe sparseness and have burdened the model with further parameterization for no added scientific value.

4.2. Analytical framework

For the survival regression model, we considered as covariates race (African American vs Caucasian), stage (local or regional), and age (in years) at diagnosis (continuous), all of which are well-established predictors of overall survival [32]. Table I shows the distribution of the cause of death categories across the levels of the covariates. The distribution in the sample (Table I) is fairly consistent with that in the original population.

For our analysis, we considered three models that are constructed using a combination of specifications described in Section 3. The details are provided below.

Model 1: This model uses a PB formulation which induces a cause-specific hazard structure given by

$$\begin{aligned}\lambda_j(u) &= \lambda_0(u)p_j \exp(\beta_{0j} + \beta_{1j}\text{Age} + \beta_{2j}\text{Race} + \beta_{3j}\text{Stage}), \quad j=1, \\ &= \lambda_0(u)p_j \exp(\beta_{0j} + \beta_{1j}\text{Age} + \beta_{2j}\text{Race}), \quad j=2,3,\end{aligned}$$

where the indices j refer to the three causes of death, namely, breast cancer, other cancer and non-cancer related. We further assume that the masking probabilities are specified by the Dirichlet distribution described as $qDir$ specification of (5) in Section 3. Note, however, that our prior specification excludes the masking subset (other cancer, non-cancer), so that $\mathbf{qBCa} \sim \text{Dir}_4(\theta_1, \theta_2, \theta_3, \theta_4)$, $\mathbf{qOCa} \sim \text{Dir}_3(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3)$, and $\mathbf{qNca} \sim \text{Dir}_3(\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3)$, where $\theta_i, \tilde{\theta}_i, \bar{\theta}_i$ are Dirichlet parameters chosen by the analyst *a priori*.

Model 2: This model uses a nonrestrictive general baseline specification (NPB) for the cause-specific hazard functions. However, the model still uses a Dirichlet specification for the masking probabilities, albeit at the subject level as in (6).

Model 3: In addition to the NPB hazard, this model also incorporates a regression structure for the masking probabilities. The masking probabilities are modeled under a multinomial logistic framework with random regression coefficients. Denoting by $q_{ij} = (q_{ij1}, q_{ij2}, \dots, q_{ijk})$ the k -dimensional vector of masking probabilities for the i th subject and j th cause, the multinomial logistic model imposes a log-linear structure on the set of logits given by:

$$\log\left(\frac{q_{ijl}}{q_{ijk}}\right) = \gamma'_{jl}Z_i, \quad l=1, \dots, k-1, \quad j=1, 2, 3. \quad (8)$$

Note that in (8), k refers to the number of potential masking subsets containing j , and hence, is at most 2^{j-1} . In our example, k equals 4 for BCa, and equals 3 for both OCa or NCa. Further, Z_i is a vector describing individual i that may or may not have an overlap with the vector of covariate levels X_i used to describe the cause-specific survival model. Also note that γ_{jl} 's are allowed to differ between causes, ensuring further flexibility. Notwithstanding the capability and flexibility of the model above,

Table I. Distribution of cause of death across levels of covariates*.

	Cause						
	BCa	OCa	NCa	Masked			Censored
				BCa/OCa	BCa/NCa	BCa/OCa/NCa	
Race (per cent AA)	29.1	6.7	7.3	16.1	20.0	14.5	14.8
Stage (per cent regional)	61.8	26.7	21.8	51.6	60.0	43.5	25.6
Mean age (Years)	59.3	67.3	75.6	64.8	61.6	65.7	57.9
(SD)	(13.5)	(11.8)	(9.2)	(13.4)	(9.4)	(13.6)	(12.6)

*BCa: breast cancer; OCa: other cancer; NCa: non-cancer.

it is difficult to identify useful predictors to be used in (8). This is due to the attempt to model a latent quantity on which no information is available. In this article, we pursue this modeling primarily for illustrative purposes, and for comparing the results with the other models. We report our findings based on a model that uses stage and dichotomized age (<65 years, ≥65 years) as covariates Z_j in (8).

4.3. Prior parameterization

Dirichlet parameters: All Dirichlet parameters are taken to be equal to 1.

Gamma process parameters: The mean process $\omega(u)$ is taken to be proportional to u , i.e. $\omega(u)=ru$, we assume $c=0.01$, $r=0.5$. In this formulation, r is the proportionality constant attached to the mean function of the cumulative baseline hazard. An alternative interpretation of r is that it is the mean of the baseline hazard rate which is constant by our choice of the functional form. The quantity c on the other hand is the weight associated with guessing the baseline cumulative intensity to be equal to its mean. Since $c=E[\Lambda]/\text{Var}[\Lambda]$, the reciprocal of c can also be viewed as an over-dispersion factor. We have run our model multiple times with small and large values of the over-dispersion factor with virtually identical estimates of the parameters. Of course, if either r or c is too small, thereby making the prior information too negligible, convergence issues were encountered. For our illustration, we further chose p_j to be degenerate at 1, although a random p_j specification such as the one mentioned in Section 3 can easily be incorporated with a couple of additional lines of code.

Regression parameters: All regression parameters (β) for the survival model are assumed to be $N(0,10^6)$, the large standard deviation (10^3) ensuring the prior specification to be diffused. The regression parameters (γ) for Model 3 are assumed to be $N(0,1)$, a specification that provided faster mixing.

4.4. Implementation

The analysis was carried out in WinBUGS version 1.4.1 [33] (sample code attached in the Appendix). In order to avoid computational difficulties and facilitate faster convergence, we adopted a hybrid multiple imputation method. The steps followed for the posterior calculations are detailed below.

Step 1: For each subject with masked cause of death, an exact (latent) cause of death was generated with equal probability of selection for each eligible cause.

Step 2: For each model, MCMC was carried out (till convergence) to obtain posterior estimates of the model parameters using the pseudo-complete data with the ‘imputed’ exact causes from step 1. For Models 1 and 2, the chains were run with a burn-in of 2000 and thinning interval of 2. For Model 3, a burn-in of 4000 was used to provide a greater stabilization time. Posterior characteristics were based on 2000 iterations. The summary statistics of the parameters were saved from the converged model.

Step 3: The cause of death for each subject with masked cause of death was redrawn using the imputation probabilities from (7) and the estimates in Step 2.

Step 4: Steps 2 and 3 were repeated five times independently to generate five sets of estimates.

Step 5: The estimates were combined using Rubin’s formula [34] for multiple imputation. A normal distribution-based calculation using the multiple imputation standard error is used to approximate the 95 per cent credible intervals.

The MCMC simulations were monitored using trace plots and other relevant diagnostics to ensure convergence. The cumulative hazard plots and the barplots described later were generated in the general purpose statistical package *R* [35].

4.5. Results

Table II demonstrates the posterior summary statistics for the regression parameters. It is evident, based on the credible intervals, that *stage* is a significant factor for breast cancer mortality. For all three models considered here, there is a significantly higher (between 3.3 and 4.5 times greater on average) hazard of death for patients with regional stage cancer, compared with local stage cancer. In our data, the African American women tend to have a higher hazard (between 1.3 and 1.5 times larger) of death due to breast cancer compared with Caucasians, although the 95 per cent credible intervals for all models include 1, thereby

Table II. Posterior mean, median and credible intervals for hazard ratios.									
	Breast cancer								
	Age			Race (Ref: Caucasian)			Stage (Ref: local)		
	Mean	Median	95 per cent CI	Mean	Median	95 per cent CI	Mean	Median	95 per cent CI
Model 1	1.001	1.001	(0.982, 1.019)	1.528	1.539	(0.913, 2.558)	4.542	4.528	(2.672, 7.720)
Model 2	1.011	1.011	(0.993, 1.029)	1.354	1.361	(0.785, 2.333)	3.585	3.581	(2.291, 5.613)
Model 3	1.010	1.010	(0.994, 1.027)	1.490	1.499	(0.901, 2.465)	3.311	3.304	(2.103, 5.214)
Other Cancer									
Model 1	1.048	1.048	(1.019, 1.078)	0.775	0.795	(0.249, 2.415)			
Model 2	1.035	1.035	(1.013, 1.057)	1.306	1.319	(0.627, 2.718)			
Model 3	1.033	1.033	(1.011, 1.055)	1.112	1.127	(0.536, 2.308)			
Non Cancer									
Model 1	1.123	1.122	(1.091, 1.155)	0.960	0.982	(0.394, 2.341)			
Model 2	1.107	1.107	(1.079, 1.136)	0.712	0.730	(0.270, 1.882)			
Model 3	1.112	1.112	(1.084, 1.140)	0.725	0.752	(0.297, 1.770)			

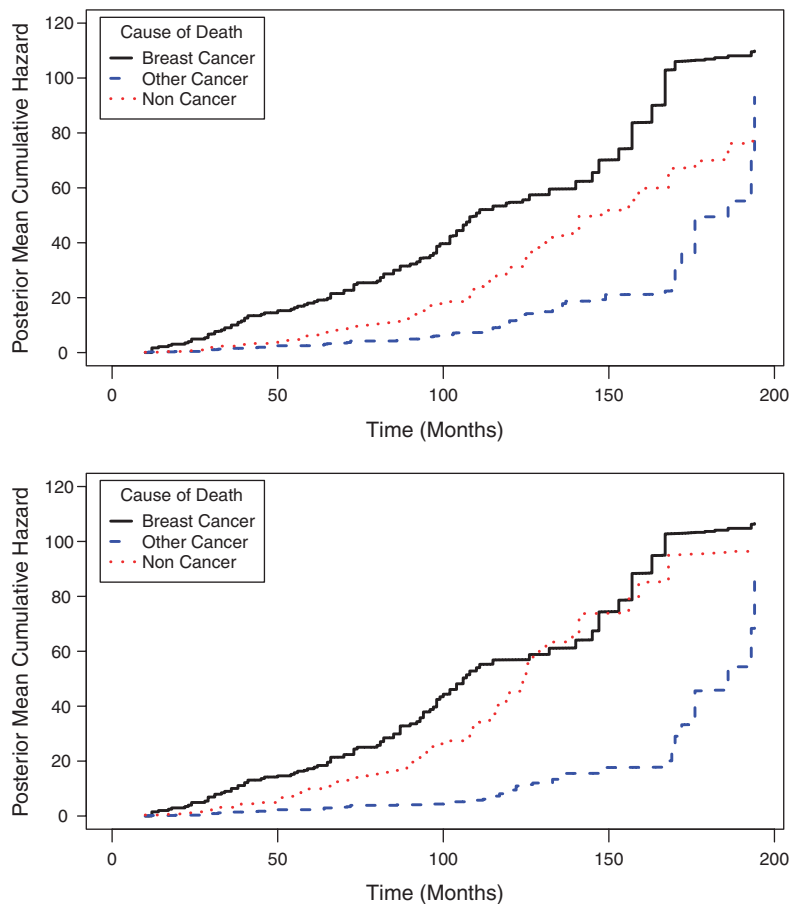


Figure 1. Baseline Cumulative Hazard Plots for Model 2 (top) and Model 3 (bottom).

rendering the difference insignificant at 5 per cent level. While older women tended to show a slightly higher risk compared with their younger counterparts, age did not turn out to be a significant predictor of breast cancer mortality. The posterior distributions of all the regression coefficients turned out to be fairly symmetric.

For both other cancer and non-cancer related deaths, the picture is quite clear. Irrespective of the model used, for either cause, age turns out to be the only significant predictor of death, with 3–5 per cent higher risk for other cancer, and 11–12 per cent higher risk for non-cancer related deaths for every one year increase in age.

Figure 1 shows the cumulative baseline hazard plots for the three causes under Model 2 and Model 3. Both plots, especially the plot corresponding to Model 3, show indications confirming an NPB model. A formal model selection is carried out using

Table III. Log-likelihood and DIC.			
	Posterior mean log-likelihood	Log-likelihood at posterior means	DIC
Model 1	-1670.21	-1610.75	3459.36
Model 2	-1599.28	-1500.64	3395.86
Model 3	-1577.57	-1514.94	3280.39

the *deviance information criterion* (DIC) popularized by Spiegelhalter *et al.* [36]. The DIC values shown in Table III favor Model 3 slightly over the other two models.

It is of interest to investigate, for the subset of individuals with masked cause of death, the posterior likelihood of each (latent) cause and how these change across the characteristic profile of the subject. These are precisely the ‘imputation’ probabilities in (7). Figures 2 and 3 demonstrate these probabilities (expressed as percentages) for the subset of *completely* masked cases stratified by race, stage, and age group. Each subplot is based on the mean of the posterior probabilities for the subset of individuals falling in that classification category. The plots show the marked influence of the model used to compute these probabilities. The discrepancy in diagnostic probabilities across models is more prominent for the subgroups which are sparsely populated. For subgroups having a sizeable number of individuals, the diagnostic probabilities maintain the relative ranking across all three models. It is also interesting to compare the probability distributions across profiles. For example, for a Caucasian woman less than 65 years of age with regional stage cancer, there is a 58 per cent chance of breast cancer being the primary cause of death (using Model 3). In comparison, for an individual with the same age and race, but local stage cancer, the likelihood of breast cancer being the true (latent) cause of death is 28 per cent. For Caucasians older than 65 years, on the other hand, other cancer and non-cancer related deaths are more likely to be the true cause of death. For African American women older than 65 years, the inference may not be too reliable due to the small number of individuals in these categories.

One interesting point to note here is that the posterior probabilities in Figures 2 and 3 are quite distinct from the observed proportions for each cause in the same subgroups for the subjects with known cause of death. For example, for the subgroup of Caucasians younger than 65 years diagnosed with local stage cancer (15 subjects), the proportions of deaths due to BCa, OCa, and NCa, are 33.7, 25, and 41.7 per cent, respectively. Similarly, there is a 50-50 split between BCa and NCa as the observed causes of death in the subgroup of Caucasian subjects who are 65 years or older and are diagnosed with regional stage cancer (6 subjects). These are drastically different from the estimated posterior probabilities in the corresponding subgroups. This may be construed as an indirect evidence that for our data set the imputation probabilities of (7) are strongly dependent on S_j , the masking subset of causes.

5. Concluding remarks

In this article, we have proposed a semi-parametric Bayesian approach to analyzing survival data with masked cause of death. ‘Masking’ is commonly encountered when analyzing data from large registries and population-based studies. The methods we used to carry out the inference are quite similar to those in the missing data literature. Specifically, we used *multiple imputation* to impute the propensity of a cause to be the primary reason for death when the actual cause of death is masked. The distribution of these (latent) true causes is then used to create a ‘complete’ data set for carrying out the competing risks regression. This process is carried out multiple times, and the estimates obtained from each pseudo-complete data set are combined using the usual multiple-imputation formulas. Unlike in the usual missing data framework, however, the data-augmentation is carried out in a Bayesian way.

To the extent we trust our model specification, our approach offers some flexibility in modeling. In the literature, competing risks analysis with missing cause of death is often handled using either a *symmetry* structure on the q -probabilities, or the availability of a second-stage data that are typically used to carry out likelihood-based inference. Based on a sample run of our data using Model 1, for the subset of individuals with *completely masked* cause of death, we estimated the full conditional distributions of the q -probabilities. There was substantial non-overlap between the probabilities. Specifically, using 1, 2, 3 to denote breast cancer, other cancer, and non-cancer causes, respectively, the distribution of $\Pr(s=\{1,2,3\} | C=2, X)$ is completely separated from the distributions of $\Pr(s=\{1,2,3\} | C=1, X)$ and $\Pr(s=\{1,2,3\} | C=3, X)$, in violation of the symmetry assumption. In the approach used in this article, we do not need to assume any *symmetry* structure on the q -probabilities.

The Bayesian paradigm provides a unified framework for carrying out estimation and predictive inference. Further, it lends itself readily to straightforward model selection procedures. The models we have used do not pre-suppose any independence among the causes of death, and are valid for an arbitrary number of causes, thereby entertaining the possibility of partial masking. By taking recourse to a simulation-based approach, we have avoided some complexities of analytical inference. The possibility of entertaining a flexible modeling environment and the ease of implementation make this, in our view, an appealing approach to analyze competing risks survival data with masked cause of death.

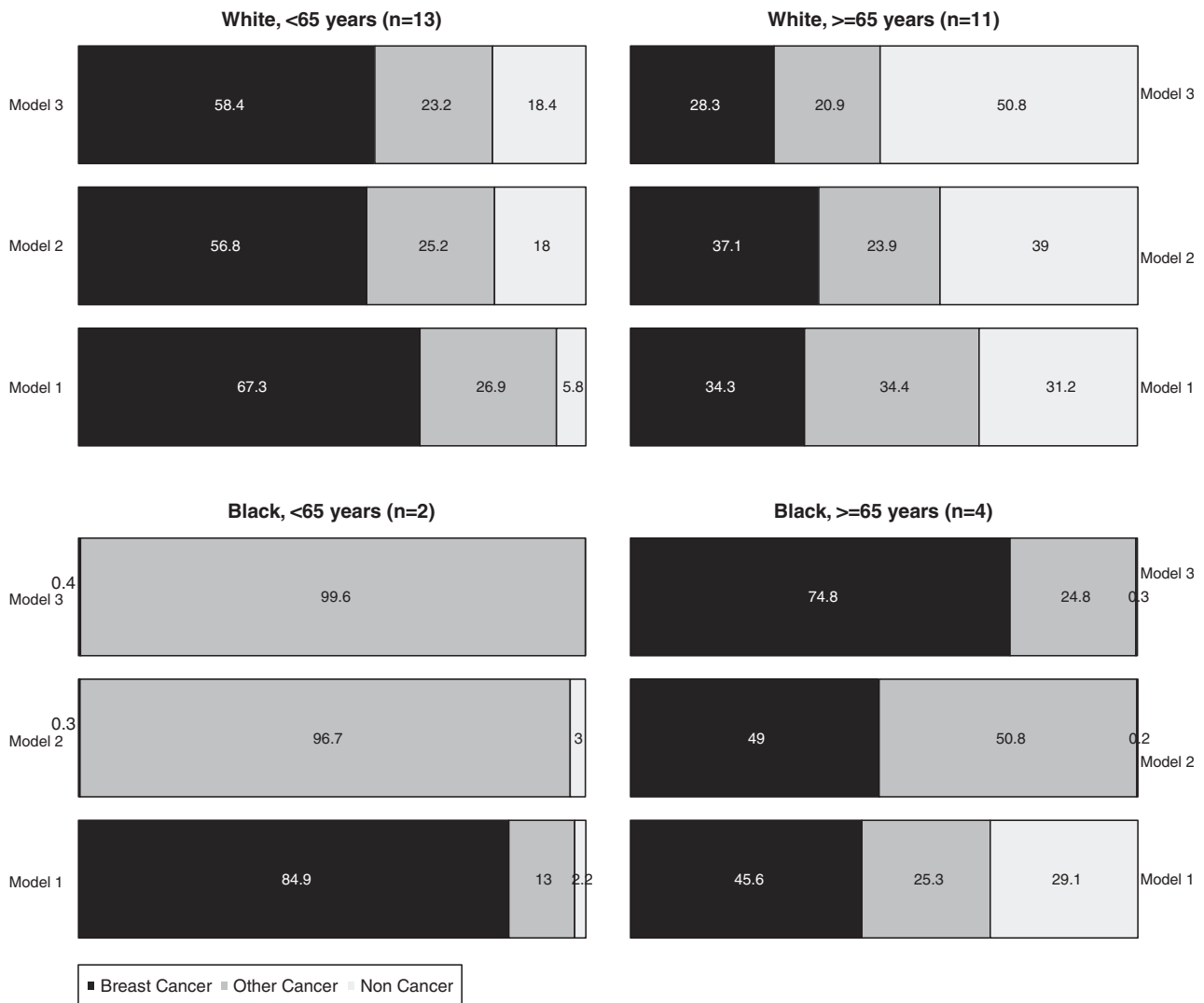


Figure 2. Posterior probabilities (expressed as percentage) of the latent causes for the completely masked cause of death with regional stage cancer.

It is perhaps fair to put the modeling framework in proper perspective. While the Bayesian approach avoids the ‘symmetry’ assumption imposed on the masking probabilities, it puts a modeling structure (Dirichlet or logistic) on them. The symmetry assumption implies missing at random mechanism that is un-testable based on observed data alone. While the Bayesian framework does not need to rely on this mechanism, the parametric formulation of prior is equally un-testable and needs to be taken in good faith.

One drawback of the approach presented here is the computation time. The semi-parametric Bayesian model involves generation of random points from a stochastic process. The number of generations grows proportionately to the number of observations. Consequently, the computing time becomes burdensome for a moderately large number of observations. For our data with 500 subjects, generation of a single MCMC chain consisting of 6000 realizations took between 7 and 8 h. In order to complete the five chains for the multiple imputation, the computing time required was a little less than two days on a dedicated machine. This feature, inherent in any simulation-based approach, counterbalances to some extent the flexibility and general applicability of the methodology.

The model in (2) in its fully non-specified form is over-parameterized. Over-parameterization is not a formal technical problem in Bayesian context since in this framework the parameters are assumed to be random draws from some distribution. Over-parameterization can certainly lead to convergence issues in many situations, however. Part of the slow convergence rates we encounter in our simulation is likely attributable to this reason.

Despite some of the shortcomings, the modeling approach seems to offer great promise in providing a flexible framework for analyzing competing risks survival data with missing cause of death. Among individuals with a cancer diagnosis, one would ideally like to estimate cause-specific mortality metrics in the presence of other causes of death. Such measures better reflect the influence of cancer treatment interventions than all-cause survival, which can vary widely by race or other covariates of interest. By employing a cause-specific framework in the presence of competing risks, one is able to assess the association of a covariate

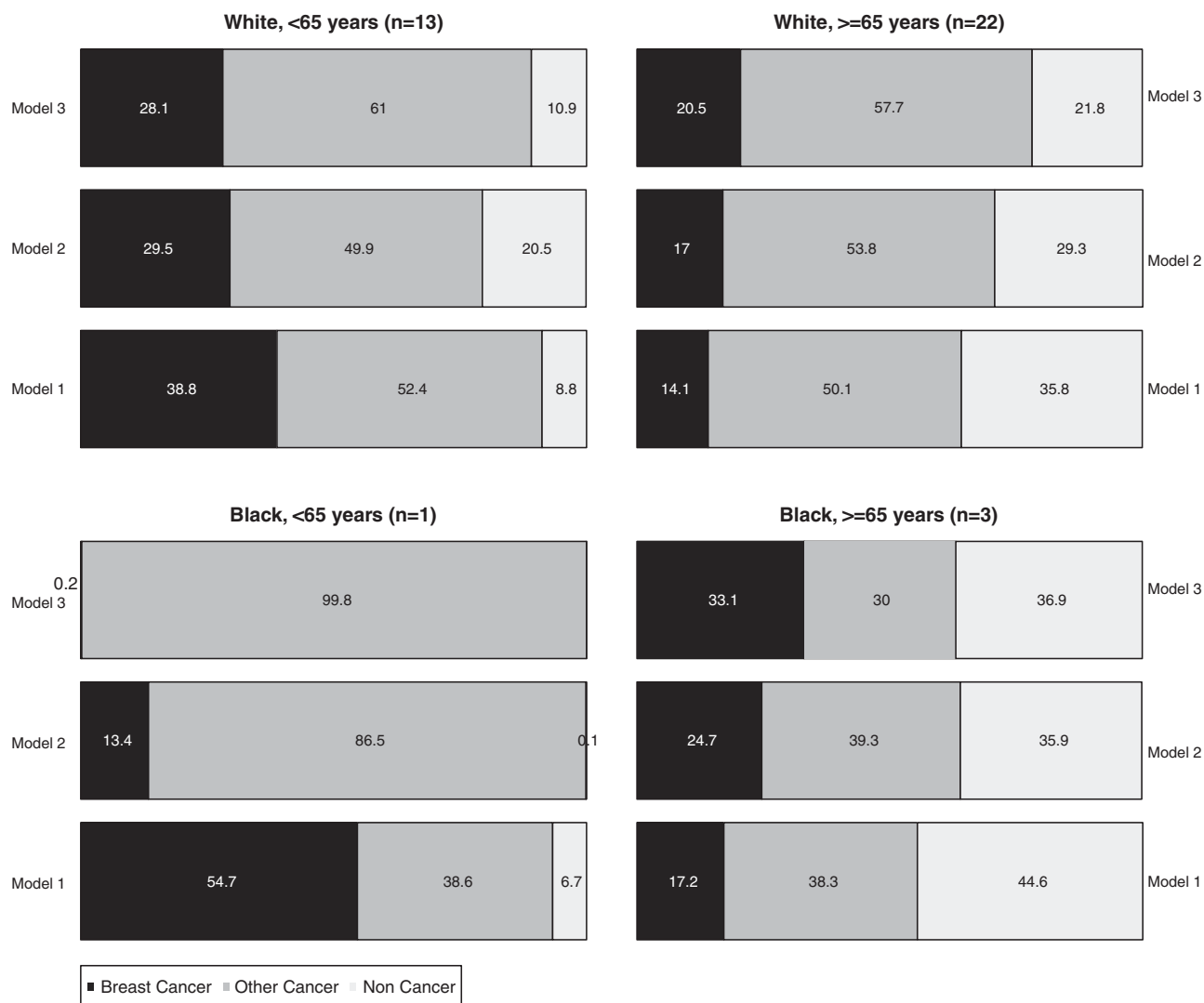


Figure 3. Posterior probabilities (expressed as percentage) of the latent causes for the completely masked cause of death with local stage cancer.

with mortality from the primary cause of interest more reliably, unaffected by the association of this covariate with mortality from other causes. The calculation of cause-specific mortality endpoints, such as time from diagnosis until death due to cancer and the proportion of deaths attributable to the cancer, requires knowledge of the cause of death, which may only be partially available in population data that rely on state death certificates. Other examples of missing cause of failure under competing risks appear in areas such as reliability engineering. In this article, we have used our methodology to analyze data from the SEER cancer registry, where the partial information about the cause of death is incorporated by means of latent variables.

The data analysis presented here serves only as an illustration of our methodology. In particular, we over-sampled the number of patients with masked cause of death for illustrative purposes. This resulted in a non-random sampling from the original SEER database. Therefore, based on our analysis, we cannot readily generalize our findings to the population of breast cancer patients. However, in order to assess the effect of the non-random sampling on the inference, we ran a parallel competing risks analysis in SAS 9.2 [37] following Lunn and McNeil [38] on all subjects in the entire database who are either censored or have a known cause of death. Essentially the model is a stratified PH model with age, race, stage, and cause indicator as covariates. In order to gauge the differential effects of age, race, stage across different causes, we also included interaction terms of these variables with cause. Note that the stratified PH model differs from the Bayesian model proposed in this article in that the latter assumes the same regression structure on all causes. Despite the above difference, this parallel analysis provides a comparison between the results obtained using the entire data set and the sample used in the article. The point estimates of the hazard ratio for age are 1.008, 1.052, and 1.12, respectively, for BCa, OCa, and NCa, which are remarkably close to the corresponding posterior estimates based on our Bayesian model. For BCa, the hazard ratio for race based on the stratified PH model is 1.6, which is quite close to the estimate obtained from our Bayesian analysis. Frequentist estimate of the hazard ratio for stage with respect to breast cancer mortality is 4.3, which is somewhat different in magnitude from the Bayesian estimate. Clearly, the confidence intervals obtained from the frequentist analysis of the entire data set are substantially narrower than the corresponding credible intervals from the

Bayesian analysis of the smaller sample. This resulted in a conclusion of significant racial difference for breast cancer mortality based on the frequentist analysis of the larger data set which we failed to achieve in the smaller data set.

Appendix A

The following is the WinBUGS code for the competing risks model using Model 2. This assumes availability of a *complete* data set with all known causes. The following notation is used in the code: N : number of subjects, T : number of distinct death times, M : number of potential causes. There are two cause of death variables associated with each subject. One, labeled as $cause[i]$ for the i th subject, takes values 0, 1, 2, 3, denoting censoring or death due to BCa, OCa, NCa, respectively. These are a mix of the observed and the estimated causes (for the ones whose exact cause of death is unknown). The observed subset of causes is denoted by the variable $S[i]$. The vector $obs.t$ contains observed times for all subjects, whereas Z 's represent covariate information.

```
model {

for (i in 1:N)
{
  q[i,1,2] <- 0
  q[i,1,3] <- 0
  q[i,2,1] <- 0
  q[i,2,3] <- 0
  q[i,3,1] <- 0
  q[i,3,2] <- 0
  q[i,4,1] <- 0
  q[i,4,2] <- 0
  q[i,4,3] <- 0
  q[i,5,3] <- 0
  q[i,6,2] <- 0
  q[i,7,1] <- 0
  q[i,7,2] <- 0
  q[i,7,3] <- 0
}

for (i in 1:N){

  fail[i,1] <- equals(cause[i], 1)
  fail[i,2] <- equals(cause[i], 2)
  fail[i,3] <- equals(cause[i], 3)

  indi[i,1,1]<-equals(s[i],1)
  indi[i,1,2]<-equals(s[i],5)*equals(cause[i],1)
  indi[i,1,3]<-equals(s[i],6)*equals(cause[i],1)
  indi[i,1,4]<-equals(s[i],8)*equals(cause[i],1)
  indi[i,2,1]<-equals(s[i],2)
  indi[i,2,2]<-equals(s[i],5)*equals(cause[i],2)
  indi[i,2,3]<-equals(s[i],7)*equals(cause[i],2)
  indi[i,2,4]<-equals(s[i],8)*equals(cause[i],2)
  indi[i,3,1]<-equals(s[i],3)
  indi[i,3,2]<-equals(s[i],6)*equals(cause[i],3)
  indi[i,3,3]<-equals(s[i],7)*equals(cause[i],3)
  indi[i,3,4]<-equals(s[i],8)*equals(cause[i],3)
  indi[i,4,1]<-equals(s[i],4)
  indi[i,4,2]<-0
  indi[i,4,3]<-0
  indi[i,4,4]<-0
}

for (i in 1:N){ for (j in 1:T){

Y[i,j] <- step(obs.t[i] - t[j] + eps)

for (k in 1:M){

dN[i,j,k] <- Y[i,j]*step(t[j+1] - obs.t[i] - eps)*fail[i,k] }

} }

}
```

```

for (i in 1:N)
{
  sumbase[i,1] <-inprod(Y[i,],dL01[])
  sumbase[i,2] <-inprod(Y[i,],dL02[])
  sumbase[i,3] <-inprod(Y[i,],dL03[])
}

for (j in 1:T) {
  for (i in 1:N) {

dN[i, j, 1] ~ dpois(Idt[i, j, 1]) Idt[i, j, 1] <-
Y[i,j]*exp(beta0[1] + beta1[1]*z1[i] + beta2[1]*z2[i] +
beta3*z3[i])*dL01[j]
}

for (i in 1:N) {

dN[i, j, 2] ~ dpois(Idt[i, j, 2])

Idt[i, j, 2] <- Y[i,j]*exp(beta0[2] + beta1[2]*z1[i] +
beta2[2]*z2[i])*dL02[j]
}
for (i in 1:N) {

dN[i, j, 3] ~ dpois(Idt[i, j, 3])

Idt[i, j, 3] <- Y[i,j]*exp(beta0[3] + beta1[3]*z1[i] +
beta2[3]*z2[i])*dL03[j]
}}

for (j in 1:T) {
  dL01[j] ~ dgamma(mu[j,1], c)
  dL02[j] ~ dgamma(mu[j,2], c)
  dL03[j] ~ dgamma(mu[j,3], c)
  for (k in 1:M) {
    mu[j,k] <- dL0.star[j,k]*c
    dL0.star[j,k] <- r[k]*(t[j+1] - t[j])
  }
}

for (i in 1:N)
{
  indi[i, cause[i] , 1:4] ~ dmulti(v[i, cause[i] , 1:4], 1)
}

for (i in 1:N)
{
  v[i, 1, 1:4] ~ ddirch(alpha1[])
  v[i, 2, 1:3] ~ ddirch(alpha2[])
  v[i, 3, 1:3] ~ ddirch(alpha2[])
}

for(j in 1:4)
{
  alpha1[j] <- 1/2
}

for(j in 1:3)
{
  alpha2[j] <- 1/2
}

for (k in 1:M)
{ beta0[k] ~ dnorm(0.0,0.000001)
  beta1[k] ~ dnorm(0.0,0.000001)
  beta2[k] ~ dnorm(0.0,0.000001)
}

```

```

beta3 ~ dnorm(0.0,0.000001)

for (i in 1:N) {
  q[i, 1, 1] <- v[i, 1,1]
  q[i, 5, 1] <- v[i, 1,2]
  q[i, 6, 1] <- v[i, 1,3]
  q[i, 8, 1] <- v[i, 1,4]
  q[i, 2,2] <- v[i, 2,1]
  q[i, 5,2] <- v[i, 2,2]
  q[i, 8,2] <- v[i, 2,3]
  q[i, 3,3] <- v[i, 3,1]
  q[i, 6,3] <- v[i, 3, 2]
  q[i, 8,3] <- v[i, 3, 3]
}

for (i in 1:N)
{
  u[i,1] <- exp(beta0[1] + beta1[1]*z1[i] + beta2[1]*z2[i] + beta3*z3[i])
    *sumbase[i,1]
  w[i,1] <- q[i,s[i],1]*exp(beta0[1] + beta1[1]*z1[i] + beta2[1]*z2[i] +
    beta3*z3[i])*dL01[obstrank[i]]
  u[i,2] <- exp(beta0[2] + beta1[2]*z1[i] + beta2[2]*z2[i])
    *sumbase[i,2]
  w[i,2] <- q[i,s[i],2]*exp(beta0[2] + beta1[2]*z1[i] + beta2[2]*z2[i] )
    *dL02[obstrank[i]]
  u[i,3] <- exp(beta0[3] + beta1[3]*z1[i] + beta2[3]*z2[i] )
    *sumbase[i,3]
  w[i,3] <- q[i,s[i],3]*exp(beta0[3] + beta1[3]*z1[i] + beta2[3]*z2[i] )
    *dL03[obstrank[i]]

  lik1[i] <- sum(w[i,])
  lik2[i] <- -sum(u[i,])

  #If the subject is censored (s[i] = 4, then lik1=0 and only
  # lik2 contributes (1st piece)
  #If the subject is not censored, then the log-likelihood contribution
  # is log(lik1)+lik2 (2nd piece)

  loglik[i] <- lik2[i]*equals(s[i], 4) + (log(lik1[i]+eps) +
    lik2[i])*(1-equals(s[i], 4))
}

finloglik <- sum(loglik[1:N])
#for those in subset (12)
#p11 = probability of being cause = 1
for (i in (N1+1):N2){
  p11[i] <- w[i,1] / (w[i,1] + w[i,2])}

#for those in subset (13)
#p21 = probability of being cause = 1
for (i in (N2+1):N3)
{
  p21[i] <- w[i,1] / (w[i,1] + w[i,3])
}

#for those in subset (123)
#p41[,1] = probability of being cause = 1
#p42[,2] = probability of being cause = 2
#p43[,3] = probability of being cause = 3
for (i in (N4+1):N)
{
  p41[i] <- w[i,1]/(w[i,1] + w[i,2] + w[i,3])
  p42[i] <- w[i,2]/(w[i,1] + w[i,2] + w[i,3])
  p43[i] <- 1 - p41[i] - p42[i]
}
}

```

Acknowledgements

We thank the Editor for timely processing of the manuscript. We also gratefully acknowledge the thorough reading of an earlier version by the two anonymous referees whose several valuable and insightful suggestions have helped improve the presentation considerably. The research of Mousumi Banerjee and Anne-Michelle Noone was partially supported by a grant from the Blue Cross Blue Shield Foundation of Michigan.

References

1. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure time data in the presence of competing risks. *Biometrics* 1978; **34**:541–554.
2. Lawless JF. *Statistical Models and Methods for Lifetime Data*. Wiley: New York, 1982.
3. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data* (2nd edn). Wiley: New York, 2002.
4. David HA, Moeschberger ML. *The Theory of Competing Risks* Griffin's Statistical Monographs and Courses, vol. 39. Charles W. Griffin: London, 1978.
5. Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association* 1993; **88**:400–409.
6. Pepe MS, Mori M. Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* 1993; **12**:737–751.
7. Moeschberger ML, Klein JP. Statistical methods for dependent competing risks. In *Lifetime Data: Models in Reliability and Survival Analysis*, Jewell NP, Kimber AC, Lee MT, Whitmore GA (eds). Kluwer Academic Publishers: Boston, 1996; 233–242.
8. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine* 1999; **18**:695–706.
9. Crowder MJ. *Classical Competing Risks*. Chapman & Hall, CRC: London, 2001.
10. Greenhouse JB, Wolfe RA. A competing risks derivation of a mixture model for the analysis of survival data. *Communications in Statistics: Theory and Methods* 1984; **13**:3133–3154.
11. Cummings FJ, Gray R, Davis TE, Tormey DC, Harris JE, Falkson GG, Arsenequ J. Tamoxifen versus placebo: double blind adjuvant trial in elderly woman with stage II breast cancer. *National Cancer Institute Monographs* 1986; **1**:119–123.
12. Guttman I, Lin DK, Reiser B, Usher JS. Dependent masking and system life data analysis: Bayesian inference for two-component systems. *Lifetime Data Analysis* 1995; **1**:87–100.
13. Basu S, Sen A, Banerjee M. Bayesian analysis of competing risks with partially masked cause-of-failure. *Applied Statistics* 2003; **52**:77–93.
14. Craiu RV, Reiser B. Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Analysis* 2006; **12**: 21–33.
15. Lagakos SW, Louis TA. Use of tumor lethality to interpret tumorigenicity experiments lacking cause of death data. *Applied Statistics* 1988; **37**:169–179.
16. Goetghebeur E, Ryan L. Analysis of competing risks survival data when some failure types are missing. *Biometrika* 1995; **82**:821–833.
17. Andersen J, Goetghebeur E, Ryan L. Analysis of survival data under competing risks with missing cause of death information: application and implications for study design. In *Lifetime Data: Models in Reliability and Survival Analysis*, Jewell NP, Kimber AC, Lee MT, Whitmore GA (eds). Kluwer Academic Publishers: Boston, 1996; 13–19.
18. Andersen J, Goetghebeur E, Ryan L. Missing cause of death information in the analysis of survival data. *Statistics in Medicine* 1996; **15**:2191–2201.
19. Lu K, Tsiatis AA. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* 2001; **57**:1191–1197.
20. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, U.S.A.* 1975; **72**: 20–22.
21. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.
22. Guess FM, Usher JS, Hodgson TJ. Estimating system and component reliabilities under partial information on cause of failure. *Journal of Statistical Planning and Inference* 1991; **29**:75–85.
23. Lin D, Usher J, Guess F. Exact maximum likelihood estimation using masked system data. *IEEE Transactions on Reliability* 1993; **42**: 631–635.
24. Dewanji A, Sengupta D. Estimation of competing risks with general missing pattern in failure type. *Biometrics* 2003; **59**:1063–1070.
25. Flehinger BJ, Reiser B, Yashchin E. Survival with competing risks and masked causes of failure. *Biometrika* 1998; **85**:151–164.
26. Flehinger BJ, Reiser B, Yashchin E. Parametric modeling for survival with competing risks and masked failure causes. *Lifetime Data Analysis* 2002; **8**:177–203.
27. Kuo L, Yang T. Bayesian reliability modeling for masked system lifetime data. *Statistics and Probability Letters* 2000; **47**:229–241.
28. Tsiatis AA, Davidian M, McNeney B. Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. *Biometrika* 2002; **89**:238–244.
29. Kalbfleisch JD. Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B* 1978; **40**:214–221.
30. Kochar SC, Proschan F. Independence of time and cause of failure in the multiple dependent competing risks model. *Statistica Sinica* 1991; **1**:295–299.
31. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 1992; **41**:337–348.
32. Simon MS, Banerjee M, Crossley HM, Vigneau FD, Noone AM, Schwartz K. Racial differences in breast cancer survival in the Detroit metropolitan area. *Breast Cancer Research and Treatment* 2006; **97**:149–155.
33. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.

34. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
35. R Development Core Team. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN: 3-900051-07-0. Available from: <http://www.R-project.org>.
36. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 2002; **64**:583–639.
37. SAS Institute. SAS version 9.2, Cary, NC, 2008.
38. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; **51**:524–532.