Sankhyā : The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series A, Pt. 3, pp. 31-47

LATENT WAITING TIME MODELS FOR BIVARIATE EVENT TIMES WITH CENSORING

By SUJIT K. GHOSH
Northy Carolina State University, Raleigh
and
ALAN E. GELFAND
University of Connecticut, Storrs

SUMMARY. Multivariate event time data arise frequently in both medical and industrial settings. In such data sets: event times may be associated with quite different occurrences, event times can not be considered as independent - the distribution of time to occurrence of one event may change after the occurrence of another, events can occur simultaneously, available covariate information may provide useful explanation. Censoring in some of the observations, both partial and complete, occurs. Focusing on the bivariate case, we formulate models rich enough to accommodate these features. In the spirit of fatal shock models our classes are built using latent waiting times which are assumed to follow general proportional hazards or accelerated life models. We adopt a Bayesian perspective for inference using simulation based fitting which routinely handles censoring. Since a wide range of model specifications can be introduced, we propose a generic model selection criterion for choosing among bivariate event time models. We conclude with the analysis of a sample of such event times for patients with a clinical diagnosis of AIDS.

1. Introduction

Multivariate event time data arise frequently in both medical and industrial settings. For example, medical researchers might examine time to first hospitalization following organ transplantation and time to organ failure Production analysts might be interested in time-to-failure measurements for machinery components. The motivating data for the work in this article is a sample of 169 patients with Center for Disease Control(CDC) defined AIDS at time of entry into the study. For each patient, a pair of event times (measured in days) are observed: time to first hospitalization after clinical diagnosis of AIDS and time to treatment with Zidovudine (AZT) after diagnosis. Data are collected over a $2\frac{1}{2}$ year period with all patients alive but with censoring of times at the end

AMS (1991) *subject classification.* 62C10
Key words and phrases. Accelerated life models, fatal shock models, model choice, proportional hazards models.

of this period. From a health care policy perspective, we ask if patient care in terms of time to treatment, is affected by covariate information, in particular by patient age, sex and race? Other covariate information such as disease severity would be of interest but was unavailable.

Several noteworthy points are

- Event times are for quite different occurrences. We are not considering times to recurrence of an event such as a tumor nor are we considering times to occurrence of the same sort of event, e.g., times to metastasis at distinct anatomic sites or times to diabetic retinopathy in each eye of a patient.

- Event times can not be considered as independent. The introduction of say AZT treatment presumably would be expected to lengthen survival time, hence time to hospitalization if it has not yet occurred. The model for time to occurrence of one event may change after occurrence of another.

- Events can occur simultaneously i.e., event times for an individual may match apart from censoring. In the AIDS data set, for the 169 patients who had not died at the time the experiment was stopped, 13 had uncensored ties in event times. These ties do not arise from grouping of the data. Rather, assuming continuous event times, an appropriate model must specify a singular joint density, i.e., place positive mass on sets of Lebesgue measure 0 in the sample space of joint event times.

In the literature, with a full probabilistic specification, dependence among event times for an individual has been introduced in essentially two ways. One, which seems most appropriate for the case when times are for the same type of event, assumes, the event times are assumed conditionally independent given an individual random effect or frailty, becoming marginally dependent upon integrating over this effect. See for example Clayton and Cuzick (1985) Hougaard (1986) or Oakes (1989). For continuous event times this approach yields ties with probability zero. The second way, which we adopt, is to introduce a latent set of independent waiting time variables and define the observable event times as functions of these variables. Observed event times sharing common latent waiting times will be dependent. If suitable functions are used, ties in observed event times occur with positive probability.

The marginal model approach (see, e.g., Diggle, Liang and Zeger 1994, Chap. 8) only requires a partial model specification, i.e., a mean and covariance but is not considered here because it can not accommodate a singular joint distribution for event times.

The latent variable approach in the bivariate case has a substantial literature by now. Distribution-theoretic research activity in this regard, motivated by the basic problem of extending univariate exponential models to bivariate dependent versions appears in Freund (1961), Marshall and Olkin (1967), Block and Basu

(1974), Proschan and Sullo (1975) and references therein. Limited attention is given to inference under these models. Also conspicuous is the complete absence of any data analysis. Leurgans, Tsai and Crowley (1982) extended the Freund model to include censoring. Klein, Keiding and Kamby (1989) noted that the Marshall and Olkin model allows semiparametric extensions using proportional hazards regression models (Cox, 1972). Klein et al. assumed a proportional baseline hazards structure in order to develop the customary partial likelihood which they used to fit the model. In addition, they confined themselves to censoring through competing risks.

Our contribution is to propose a very general bivariate class of models built through proportional hazards or accelerated failure time structure which includes all of the work of the previous paragraph. We adopt a Bayesian approach to fit these models. We assume rather vague priors so that our inference will resemble a full likelihood analysis except that our interval estimates will be more reliable by avoiding possibly inappropriate asymptotics. Censoring, which can be partial or complete is routinely accommodated. We use simulation methods to fit models.

We implement such fitting for a range of models suitable for the aforementioned AIDS data set and thus require a criterion for selecting among models. We develop a utility-based model choice criterion applicable to arbitrary modeling of multivariate event time data with censoring. Assuming that utility for a model is captured through predictive performance, among a set of models, we select the one which maximizes an expected predictive utility. The criterion yields a goodness of fit piece as well as a piece which penalizes for model complexity.

In section 2 we develop broad classes of bivariate event time models which can accommodate our introductory concerns. Section 3 details the data augmentation scheme to fit these models. Section 4 focuses upon model choice among models within these classes, in the presence of censoring. In section 5 we examine the motivating AIDS data set. Our analysis does not constitute a rigorous medical study of these data but rather is intended to illustrate the proposed approach.

2. The Models

We formulate models which capture the interaction between two life history events using so called fatal shocks as in Marshall and Olkin (1967). Following Leurgans, Tsai and Crowley (1980), in the most general setting we allow the possibility of separate clocks to guide the two waiting times. Hence our models need not be Markovian. Let i denote the individuals, $i = 1, 2, \cdots, n$. For the ith individual we consider events 1 and 2 with event times (T_{1i}, T_{2i}). The associated state vector at time t is $(Z_{1i}(t), Z_{2i}(t))$ where $Z_{ji}(t) = 0$ if $t < T_{ji}, = 1$ if $t \geq T_{ji}$. We assume individual level time dependent intensities. In particular, $\lambda_1^{(i)}(t)$ is the intensity associated with the outcome that event 1 occurs but event 2

does not. Similarly $\lambda_2^{(i)}(t)$ is the intensity associated with the outcome that event 2 occurs but event 1 does not. $\lambda_{12}^{(i)}(t)$ is the intensity associated with the simultaneous occurrence of events 1 and 2.

2.1 *Generalized Marshall-Olkin Models.* Suppose for each individual we introduce three independent latent waiting time variables U_{1i}, U_{2i} and U_{12i} having intensities $\lambda_1^{(i)}, \lambda_2^{(i)}$ and $\lambda_{12}^{(i)}$ respectively and define $T_{1i} = \min(U_{1i}, U_{12i}), T_{2i} = \min(U_{2i}, U_{12i})$. That is, as in Marshall and Olkin (1967), we envision the ith individual as a two component system for whom three underlying independent Poisson processes describe the occurrence of *shocks* to each or both of the components. The U_i's are viewed as waiting times to the first shock under the respective latent processes. T_{1i} and T_{2i} are dependent, can match and can be interpreted in terms of competing risks, e.g., T_{1i} records the time to event 1 whether it occurred with event 2 or not. For $\lambda_1^{(i)}$ we write

$$\lambda_1^{(i)}(t) = h_1(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_1) \qquad \ldots (2.1)$$

where h_1 is a baseline hazard, g is a covariate link function, $\boldsymbol{x}_i$ is a vector of covariates for individual i and $\boldsymbol{\beta}_1$ is a coefficient vector (no intercept). Similarly,

$$\lambda_2^{(i)}(t) = h_2(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_2) \qquad \ldots (2.2)$$

and

$$\lambda_{12}^{(i)}(t) = h_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_{12}). \qquad \ldots (2.3)$$

The hazards/intensities in (2.1)-(2.3) along with the independence of the U's determine that the marginal intensities of T_{1i} and T_{2i} are $h_1(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_1) + h_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_{12})$ and $h_2(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_2) + h_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_{12})$ respectively. In this model there is only one clock for T_{1i} and T_{2i}. More precisely, though T_{1i} and T_{2i} are dependent, the occurrence of, e.g., T_{1i} does not start a second clock with regard to the waiting time for T_{2i}. The special case where $g \equiv 1$ (no covariates) and the h's are constants yields the original Marshall-Olkin model. Model III of Klein et al. (1989), a proportional baseline hazards model, arises with $h_l(t) = \lambda_l h_{12}(t)$, l=1,2. We call a model defined through (2.1)-(2.3) as a generalized Marshall-Olkin (GMO) model.

Since the waiting times are for possibly very different events the use of distinct coefficient vectors allows the possibility that the waiting times may be affected differently by the covariates. Interpretation of the $\boldsymbol{\beta}$'s requires care. For instance, if the lth component of $\boldsymbol{x}_i$ is age and $\beta_{1l} > \beta_{12l}$ then the risk of event 1 occurring without event 2 is more increased by age than the risk of event 1 occurring with event 2. The competing risk idea may facilitate. If, e.g., both β_{1l} and $\beta_{12l} > 0$ the risk associated with event 1 increases in age. As the ensuing development shows we could as easily use accelerated life forms for these $\lambda^{(i)}$'s, e.g., $\lambda_1^{(i)}(t) = h_1(tg(\boldsymbol{x}_i^T\boldsymbol{\beta}_1))g(\boldsymbol{x}_i^T\boldsymbol{\beta}_1)$.

2.2 *Generalized Proschan-Sullo Models.* Suppose for each individual we introduce five independent waiting time variables, $U_{1i}, U_{2i}, U_{12i}, U'_{1i}, U'_{2i}$. Let U_{1i}, U_{2i} and U_{12i} have the intensities as in (2.1)-(2.3) but in addition assume for U'_{1i} and U'_{2i} proportional hazards models,

$$h'_1(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}'_1) \text{ and } h'_2(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}'_2) \qquad \ldots (2.4)$$

respectively. For individual i, if event 1 occurs first and occurs at the time t_{1i}, then $t_{1i} = T_{1i} = U_{1i} > \min(U_{2i}, U_{12i}) = T_{2i}$ but the distribution of time to event 2 is revised to that of $T'_{2i} = \min(U_{12i}, U'_{2i} + t_{1i})$. That is, after the occurrence of event 1 the intensity for the waiting time until the second event is changed; the incremental waiting time, $T'_{2i} - t_{1i} = \min(U_{12i} - t_{1i}, U'_{2i})$. It is as if a second event clock (for the U'_{2i} process) was started at time t_{1i}. Similarly, if for individual i event 2 occurs first at time t_{2i}, then $T_{2i} = U_{2i}$ but now $T'_{1i} = \min(U_{12i}, U'_{1i} + t_{1i})$. As a result, when $T_{2i} > t_{1i}$ the intensity for the time to the second event is revised to $h'_2(t - t_{1i})g(\boldsymbol{x}_i^T\boldsymbol{\beta}'_2) + h_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_{12}), t > t_{1i}$. Similarly when $T_{1i} > t_{2i}$, the intensity for the time to first event is revised to $h'_1(t - t_{2i})g(\boldsymbol{x}_i^T\boldsymbol{\beta}'_1) + h_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}_{12}), t > t_{2i}$.

The special case where $g \equiv 1$ and the h's are constants, yields the Proschan and Sullo (1975) extension to the Marshall-Olkin model. The case where $\lambda_{12} = 0$ yields Freund's (1960) model. More generally, if $h_{12}(t) = 0$ with $g \equiv 1$ we obtain the models mentioned in Leurgans et al. (1982, p.240) including Lagakos, Sommer and Zelen (1978) and Voelkel (1980). The case where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}'_1, \boldsymbol{\beta}_2 = \boldsymbol{\beta}'_2$ may be of interest. It asserts that only the baseline hazard for an event is affected by the occurrence of the other. We call a model defined through (2.1)-(2.4) a generalized Proschan-Sullo (GPS) model. As with a GMO model, interpretation of the coefficients requires care. For instance, continuing the earlier illustration, if $\beta_{2l} < \beta'_{2l}$ the implication is that, after the first event occurs, risk for the second event increases more with age than before.

2.3 *A further extension* Next, for each individual introduce two additional independent latent waiting time variables $U^{(1)}_{12i}$ and $U^{(2)}_{12i}$ having proportional hazards models

$$h^{(1)}_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}^{(1)}_{12}) \text{ and } h^{(2)}_{12}(t)g(\boldsymbol{x}_i^T\boldsymbol{\beta}^{(2)}_{12}) \qquad \ldots (2.5)$$

respectively. Now, for individual i if event 1 occurs first at time t_{1i} then $T_{1i} = U_{1i}$ but now we modify T_{2i} to $T'_{2i} = t_{1i} + \min(U^{(2)}_{12i}, U'_{2i})$, starting a second clock for both the $U^{(2)}_{12i}$ and $U^{(i)}_{12i}$ processes. Similarly, for individual i if event 2 occurs first at time t_{2i} then $T_{2i} = U_{2i}$ and we modify T_{1i} to $T'_{1i} = t_{2i} + \min(U^{(1)}_{12i}, U'_{1i})$. When $T_{2i} > t_{1i}$ the intensity for the time to the second event becomes $h'_2(t - t_{1i})g(\boldsymbol{x}_i^T\boldsymbol{\beta}'_2) + h^{(2)}_{12}(t - t_{1i})g(\boldsymbol{x}_i^T\boldsymbol{\beta}^{(2)}_{12}), t > t_{1i}$. The corresponding expression for the first event becomes $h'_1(t - t_{2i})g(\boldsymbol{x}_i^T\boldsymbol{\beta}'_1) + h^{(1)}_{12}(t - t_{2i})g(\boldsymbol{x}_i^T\boldsymbol{\beta}^{(2)}_{12}), t > t_{2i}$.

Model IV of Klein et al. (1989) considers a special case of (2.5), but we suspect that the most general version is too much to tackle in practice. Data sets will rarely be rich enough to justify *seven* latent proportional hazards models and interpretation of the coefficients will be challenging!

3. The Data Augmentation Scheme

Defining $\Lambda_1^{(i)}(t) = \int_0^t \lambda_1^{(i)}(u)du$ with similar definitions for $\Lambda_2^{(i)}$ and $\Lambda_{12}^{(i)}$, a GMO model determines a joint distribution for (T_{1i}, T_{2i}) with the following properties:

(i) The marginal distribution of T_{1i} has survival function $\exp\{-\Lambda_1^{(i)}(t)-\Lambda_{12}^{(i)}(t)\}$. T_{2i} has survival function $\exp\{-\Lambda_2^{(i)}(t) - \Lambda_{12}^{(i)}(t)\}$.

(ii) $P(T_{1i} < T_{2i}) = E\{\lambda_1^{(i)}(V_i)\ (\lambda_1^{(i)}(V_i) + \lambda_2^{(i)}(V_i) + \lambda_{12}^{(i)}(V_i))^{-1}\}$,
$P(T_{1i} > T_{2i}) = E\{\lambda_2^{(i)}(V_i)\ (\lambda_1^{(i)}(V_i) + \lambda_2^{(i)}(V_i) + \lambda_{12}^{(i)}(V_i))^{-1}\}$,
$P(T_{1i} = T_{2i}) = E\{\lambda_{12}^{(i)}(V_i)\ (\lambda_1^{(i)}(V_i) + \lambda_2^{(i)}(V_i) + \lambda_{12}^{(i)}(V_i))^{-1}\}$.
where $V_i = \min(U_{1i}, U_{2i}, U_{12i})$. Hence the joint distribution of T_{1i} and T_{2i} is singular, placing positive mass on the 45^o line within $\mathbb{R}^+ \times \mathbb{R}^+$.

(iii) The density for (T_{1i}, T_{2i}) on $0 < t_{1i} < t_{2i} < \infty$ is

$$f(t_{1i}, t_{2i}) = \lambda_1^{(i)}(t_{1i}) \left(\lambda_2^{(i)}(t_{2i}) + \lambda_{12}^{(i)}(t_{2i})\right) \exp\left\{-\Lambda_1^{(i)}(t_{1i}) - \Lambda_2^{(i)}(t_{2i}) - \Lambda_{12}^{(i)}(t_{2i})\right\}.$$

On $0 < t_{2i} < t_{1i} < \infty$ it is

$$f(t_{1i}, t_{2i}) = \lambda_2^{(i)}(t_{2i}) \left(\lambda_1^{(i)}(t_{1i}) + \lambda_{12}^{(i)}(t_{1i})\right) \exp\left\{-\Lambda_1^{(i)}(t_{1i}) - \Lambda_2^{(i)}(t_{2i}) - \Lambda_{12}^{(i)}(t_{1i})\right\},$$

and on $0 < t_{2i} = t_{1i} = t_i < \infty$ it is

$$f(t_i) = \lambda_{12}^{(i)}(t_i) \exp\left\{-\Lambda_1^{(i)}(t_i) - \Lambda_2^{(i)}(t_i) - \Lambda_{12}^{(i)}(t_i)\right\}.$$

Given observed data $(t_{1i}, t_{2i}), i = 1, \cdots, n$ with no censoring the likelihood would be a product of appropriate terms from (iii). Suppose t_{si} is a fixed censoring time for the ith individual, e.g., time to stopping the study minus time to entrance into the study for the ith individual. Then, censoring can occur in the following ways: $t_{si} < \min(T_{1i}, T_{2i}), T_{1i} < t_{si} < T_{2i}$, or $T_{2i} < t_{si} < T_{1i}$. The contribution to the likelihood under such censoring can be immediately calculated using (iii).

To clarify, the likelihood arising from $\boldsymbol{T} = (\boldsymbol{T}_1, \cdots, \boldsymbol{T}_n)$ with $\boldsymbol{T}_i = (T_{1i}, T_{2i})$ can be viewed as a missing data likelihood; the latent $\boldsymbol{U} = (\boldsymbol{U}_1, \cdots, \boldsymbol{U}_n)$ with

$\boldsymbol{U}_i = (U_{1i}, U_{2i}, U_{12i})$ are the complete data. The conditional independence of the U's given the h's and $\boldsymbol{\beta}$'s yields a routine complete data likelihood,

$$\prod_{i=1}^{n} \lambda_1^{(i)}(u_{1i})\lambda_2^{(i)}(u_{2i})\lambda_{12}^{(i)}(u_{12i}) \exp\left\{-\Lambda_1^{(i)}(u_{1i}) - \Lambda_2^{(i)}(u_{2i}) - \Lambda_{12}^{(i)}(u_{12i})\right\} \quad \ldots (3.1)$$

In other words, given $\boldsymbol{U}$ we obtain three separate proportional hazards models.

With no censoring, $\boldsymbol{U}_i$ is constrained to the set $\mathcal{R}_i$ which arises as follows. If $t_{1i} = T_{1i} < T_{2i} = t_{2i}$, $U_{1i} = t_{1i}$ and $\min(U_{2i}, U_{12i}) = t_{2i}$; if $t_{2i} = T_{2i} < T_{1i} = t_{1i}, U_{2i} = t_{2i}$ and $\min(U_{1i}, U_{12i}) = t_{1i}$; if $T_{1i} = T_{2i} = t_i, U_{12i} = t_i$ and $U_{1i} > t_i, U_{2i} > t_i$. Suppose censoring is present. Then $\mathcal{R}_i$ is modified as follows. If $t_{si} < \min(T_{1i}, T_{2i})$ then $U_{1i} > t_{si}, U_{2i} > t_{si}$ and $U_{12i} > t_{si}$; if $T_{1i} = t_{1i} < t_{si}$ then $U_{2i} = t_{2i}$ and $U_{1i} > t_{si}, U_{12i} > t_{si}$. Hence, given T_i and perhaps t_{si}, simple restrictions are introduced on $\boldsymbol{U}_i$.

With the distributional assumptions for U_{1i}, U_{2i} and U_{12i} given by (2.1) - (2.3) we can integrate out the $\boldsymbol{U}_i$ explicitly. Suppose the unknown baseline hazard $h_1(t)$ belongs to the parametric family $h_1(t;\boldsymbol{\theta}_1)$ and similarly $h_2(t) = h_2(t;\boldsymbol{\theta}_2)$, $h_{12}(t) = h_{12}(t;\boldsymbol{\theta}_{12})$. In section 5 we use constant (exponential) and Weibull hazards. With $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{12})$, denote the resulting incomplete data likelihood at $\boldsymbol{T} = \boldsymbol{t}$ by $L(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{t})$. With a prior on $(\boldsymbol{\beta}, \boldsymbol{\theta})$, $f(\boldsymbol{\beta}, \boldsymbol{\theta})$ the Bayesian model becomes

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}; \boldsymbol{t}) f(\boldsymbol{\beta}, \boldsymbol{\theta}). \quad \ldots (3.2)$$

The above marginalization over $\boldsymbol{U}$ verifies that a legitimate Bayesian model is associated with (3.2). That is, given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, $\boldsymbol{T}_i$ has a singular distribution; the form in (3.2) is *not* the joint density of $\boldsymbol{T}, \boldsymbol{\beta}$ and $\boldsymbol{\theta}$. However, (3.1) provides a joint density for $\boldsymbol{U}$. Hence (3.1), restricted such that $\boldsymbol{U}_i \in \mathcal{R}_i$, multiplied by $f(\boldsymbol{\beta}, \boldsymbol{\theta})$ is proportional to the joint density of $\boldsymbol{U}, \boldsymbol{\beta}$ and $\boldsymbol{\theta}$ given $\boldsymbol{T}$. Integrating over $\boldsymbol{U}$ yields (3.2), so that (3.2) is proportional to the posterior of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ given $\boldsymbol{T}$.

We assume $f(\boldsymbol{\beta}, \boldsymbol{\theta}) = f(\boldsymbol{\beta}) \cdot f(\boldsymbol{\theta})$ with $f(\boldsymbol{\beta})$ flat. We further assume $f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}_1) \cdot f(\boldsymbol{\theta}_2) \cdot f(\boldsymbol{\theta}_{12})$. Under Weibull hazard for $h_1(t) = \alpha_1\gamma_1 t^{\gamma_1 - 1}$, $\boldsymbol{\theta}_1 = (\alpha_1, \gamma_1)$ so we take $f(\boldsymbol{\theta}_1) = f(\alpha_1) \cdot f(\gamma_1)$ with $f(\alpha_1) = Ga(c_1, d_1)$ and $f(\gamma_1) = Ga(c_2, d_2)$. We take $f(\alpha_1)$ be rather vague but take $f(\gamma_1)$ to have mean 1 (roughly expressing prior indifference as to whether the hazard is increasing or decreasing) with large variance (say 10). We adopt similar expressions for h_2 and h_{12}. The Gibbs sampler (Gelfand and Smith, 1990) provides a convenient tool for fitting (3.2). The full conditional density for $\boldsymbol{\beta}$ is log concave since $L(\boldsymbol{\beta}, \boldsymbol{\beta}; \boldsymbol{t})$ is, following Dellaportas and Smith (1993). Efficient componentwise sampling of the $\boldsymbol{\beta}$'s may be implemented using the adaptive rejection sampling algorithm of Gilks and

Wild (1992). The full conditional distributions for α_1, α_2 and α_{12} are mixtures of Gamma distributions. For instance for α_1, the full conditional is proportional to

$$\alpha_1^{l_1+c_1-1} \exp\{-\alpha_1 \left(\sum_{i=1}^{n} t_{1i}^{\gamma_1} g(\boldsymbol{x}_i^T \boldsymbol{\beta}_1) + d_1 \right)\}.$$
$$\left(\prod_{i \in \mathcal{S}} (\alpha_1 \gamma_1 t_{1i}^{\gamma_1 - 1} g(\boldsymbol{x}_i^T \boldsymbol{\beta}_1) + \alpha_{12} \gamma_{12} t_{1i}^{\gamma_{12}-1} g(\boldsymbol{x}_i^T \boldsymbol{\beta}_{12}), \right),$$

where l_1 is the number of observed t_{1i} such that $t_{1i} < t_{2i}$ and $\mathcal{S}$ is the set of all observed t_{1i} such that $t_{1i} > t_{2i}$. Lastly, since the full conditionals for the γ's are nonstandard densities, we introduce a Metropolis step into the Gibbs loop for each γ. In particular, for γ_1 the full conditional is proportional to

$$\gamma_1^{l_1+c_2-1} \exp \left\{ \gamma_1 (\sum_i \log t_{1i} - d_2) - \sum_i e_i \exp(\gamma_1 \log t_{1i}) \right\}$$
$$\left(\prod_{i \in \mathcal{S}} (\alpha_1 \gamma_1 t_{1i}^{\gamma_1 - 1} g(\boldsymbol{x}_i^T \boldsymbol{\beta}_1) + \alpha_{12} \gamma_{12} t_{1i}^{\gamma_{12}-1} g(\boldsymbol{x}_i^T \boldsymbol{\beta}_{12}) \right),$$

where $e_i = \alpha_1 g(\boldsymbol{x}_i^T \boldsymbol{\beta}) > 0$. If $t_{1i} \leq 1$ the expression in brackets is ≤ 0. When $t_{1i} > 1$ we can show that

$$\gamma_1 \sum_{i: t_{1i} > 1} (\log t_{1i} - e_i \exp(\gamma_1 \log t_{1i})) < -\gamma_1 \sum_{i: t_{1i} > 1} d_i \log t_{1i} + \text{ const.}$$

so that a convenient Gamma proposal density can be used.

A GPS model can be handled similarly. Again, the likelihood for the incomplete data $\boldsymbol{T}$ with censoring can be developed from the full data likelihood for the latent $\boldsymbol{U} = (\boldsymbol{U}_1, \cdots, \boldsymbol{U}_n)$ where now $\boldsymbol{U}_i = (U_{1i}, U_{2i}, U_{12i}, U'_{1i}, U'_{2i})$, which has a product form density analogous to (3.1). Five Weibull hazards and five sets of regression coefficients must be sampled.

4. **Model choice**

The GMO and GPS classes provide a wide range of models for bivariate event time data. For example, a GMO model can be specified with, e.g., constant or Weibull baseline hazards and with all or perhaps subsets of the covariates. Similar choices arise for a GPS model. Also, there need not be any nesting between a particular GMO model and a particular GPS model. A GMO model with Weibull hazards and four covariates has a total of 18 parameters ($3 \times 2 = 6$ baseline parameters + $3 \times 4 = 12$ coefficients); a GPS model with constant

hazards and two covariates has a total of 15 parameters ($5\times1=5$ baseline hazard parameters + $5\times2=10$ coefficients).

To choose among such models, we propose a criterion which is appropriate to any Bayesian model specification for bivariate event time data, possibly with partial or complete censoring of some of the $\boldsymbol{T}_i$'s. The criterion is developed using the following steps (with details below) and extends ideas in Gelfand and Ghosh (1998). We propose a suitable utility which rewards predictive performance. Then, for a given model we obtain the maximum expected utility. Finally, we select over the proposed choices, the one yielding the largest maximized expected utility. The resulting criterion rewards both goodness-of-fit and parsimony, penalizes inadequacy but also overfitting. It can be routinely computed from the output of the simulation-based fitting of the model. Customary screening criteria such as the marginal density of the data evaluated at the observations or the BIC (see, e.g., Kass and Raftery, 1995) can be criticized as in Gelfand and Ghosh (1998).

The criterion is developed from the following loss (negative utility) function. Suppose first that the data $\boldsymbol{t}_{obs}$ is collected without censoring and our unknown is an uncensored replication, $\boldsymbol{t}_{i,rep}$ of $\boldsymbol{t}_{i,obs}$, i.e., under the given model, $\boldsymbol{t}_{i,rep}$ is an unobservable realization from the same distribution as $\boldsymbol{t}_{i,obs}$. Let

$$\begin{aligned}\mathcal{L}(\boldsymbol{t}_{i,rep},\boldsymbol{a}_i;\boldsymbol{t}_{i,obs}) &= (\log\boldsymbol{t}_{i,rep}-\log\boldsymbol{a}_i)^T A(\log\boldsymbol{t}_{i,rep}-\log\boldsymbol{a}_i)\\ &+ k(\log\boldsymbol{t}_{i,obs}-\log\boldsymbol{a}_i)^T A(\log\boldsymbol{t}_{i,obs}-\log\boldsymbol{a}_i)\end{aligned} \quad \ldots(4.1)$$

where $\boldsymbol{a}_i$ is an action in $I\!R^+\times I\!R^+$ and A is positive definite matrix. The general quadratic loss in (4.1) provides a second order approximation to an arbitrary convex loss. The first term on the right hand side in (4.1) discourages $\boldsymbol{a}_i$ from being too far from $\boldsymbol{t}_{i,rep}$. The second term discourages $\boldsymbol{a}_i$ from being too far from $\boldsymbol{t}_{i,obs}$ which seems sensible given $\boldsymbol{t}_{i,obs}$ and $\boldsymbol{t}_{i,rep}$ have the same distribution. A common A in (4.1) is natural since analogous comparisons are made across the two terms. The value of k reflects how much weight we place on the components of $\mathcal{L}$. Losses of this form have been called balanced loss functions by Zellner (1994). Since the times are positive, conversion to the log scale is appropriate under quadratic loss. In fact, to simplify notation in what follows below, let $\boldsymbol{v}_{i,rep}=\log\boldsymbol{t}_{i,rep}$, $\boldsymbol{v}_{i,obs}=\log\boldsymbol{t}_{i,obs}$ and $\boldsymbol{b}_i=\log\boldsymbol{a}_i$.

Let $\tilde{\boldsymbol{\mu}}_i=E(\boldsymbol{v}_{i,rep}\mid\boldsymbol{t}_{obs})$, and $\tilde{\Sigma}_i=E\left((\boldsymbol{v}_{i,rep}-\tilde{\boldsymbol{\mu}}_i)(\boldsymbol{v}_{i,rep}-\tilde{\boldsymbol{\mu}}_i)^T\mid\boldsymbol{t}_{obs}\right)$. It is straightforward to show that for action $\boldsymbol{a}_i$ the expected posterior loss is

$$E\left(\mathcal{L}(\boldsymbol{t}_{i,rep},\boldsymbol{a}_i;\boldsymbol{t}_{i,obs})\mid\boldsymbol{t}_{obs}\right)=trA\tilde{\Sigma}_i+(\tilde{\boldsymbol{\mu}}_i-\boldsymbol{b}_i)^TA(\tilde{\boldsymbol{\mu}}_i-\boldsymbol{b}_i)+k(\boldsymbol{v}_{i,obs}-\boldsymbol{b}_i)^TA(\boldsymbol{v}_{i,obs}-\boldsymbol{b}_i) \quad \ldots(4.2)$$

When $A=I$, the $\boldsymbol{b}_i$ which minimizes (4.2) is $(k+1)^{-1}(\tilde{\boldsymbol{\mu}}_i+k\boldsymbol{v}_{i,obs})$ and at this $\boldsymbol{b}_i$ (4.2) becomes $tr\tilde{\Sigma}_i+(k+1)^{-1}k(\tilde{\boldsymbol{\mu}}_i-\boldsymbol{v}_{i,obs})^T(\tilde{\boldsymbol{\mu}}_i-\boldsymbol{v}_{i,obs})$. Since the $\boldsymbol{T}_i$ are conditionally independent given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ we follow standard practice and

cumulate the losses over i to obtain

$$\sum_{i=1}^{n} tr\tilde{\Sigma}_i + (k+1)^{-1}k\sum_{i=1}^{n}(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{i,obs})^T(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{i,obs}). \quad \ldots (4.3)$$

In (4.3) the first term reflects predictive variability. It will be large for a poor model, will diminish as we move to better models, but then would be expected to increase as we begin to overfit. The second term is a goodness-of-fit measure comparing what we observed with what the model predicts. We select the model producing the smallest value of (4.3).

In the case that observation i is censored, the *actual* $\boldsymbol{t}_{i,obs}$ is restricted to a set say $C_{i,obs}$. For example, if $t_{si} \leq \min(t_{1i}, t_{2i}), C_{i,obs} = \{(t_{1i}, t_{2i}) : t_{1i} \geq t_{si}, t_{2i} \geq t_{si}\}$. If $t_{1i} < t_{si} \leq t_{2i}, C_{i,obs} = \{(t_{1i}, t_{2i}) : t_{1i} = t_{1i,obs}, t_{2i} \geq t_{si}\}$. If $t_{2i} < t_{si} \leq t_{1i}, C_{i,obs} = \{(t_{1i}, t_{2i}) : t_{1i} \geq t_{si}, t_{2i} = t_{2i,obs}\}$. For completeness we can define $C_{i,obs} = \{(t_{1i,obs}, t_{2i,obs})\}$ when there is no censoring. We thus extend the loss function to

$$\begin{aligned}\mathcal{L}(\boldsymbol{t}_{i,rep}, \boldsymbol{a}_i; C_{i,obs}) &= \inf_{\boldsymbol{t}\in C_{i,obs}} \mathcal{L}(\boldsymbol{t}_{i,rep}, \boldsymbol{a}_i; \boldsymbol{t}) \\ &= (\boldsymbol{v}_{i,rep} - \boldsymbol{b}_i)^T A(\boldsymbol{v}_{i,rep} - \boldsymbol{b}_i) + k\inf_{\boldsymbol{t}\in C_{i,obs}} (\boldsymbol{v} - \boldsymbol{b}_i)^T A(\boldsymbol{v} - \boldsymbol{b}_i).\end{aligned}$$

$$\ldots (4.4)$$

where $\boldsymbol{v} = \log \boldsymbol{t}$. Analogous to (4.2), the posterior expected loss at $\boldsymbol{a}_i$ is

$$trA\tilde{\Sigma}_i + (\tilde{\boldsymbol{\mu}}_i - \boldsymbol{b}_i)^T A(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{b}_i) + k \inf_{\boldsymbol{t}\in C_{i,obs}} (\boldsymbol{v} - \boldsymbol{b}_i)^T A(\boldsymbol{v} - \boldsymbol{b}_i). \quad \ldots (4.5)$$

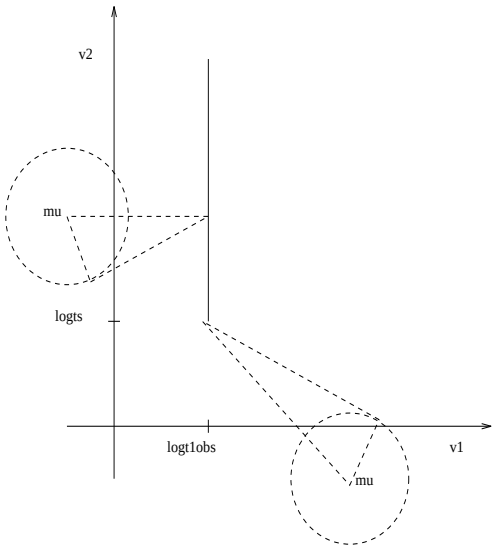

Figure 1: Geometry for minimization of (4.5) under censoring where $t_{Ii,obs} < t_{si} < t_{2i}$

When $A = I$ it is easy to establish that the minimizing $\boldsymbol{b}_i$ for (4.5) occurs along the line connecting $\tilde{\boldsymbol{\mu}}_i$ to $\boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i}$, the point in $C_{i,obs}$ closest to $\tilde{\boldsymbol{\mu}}_i$. Figure 1 provides geometric elaboration for one choice of $C_{i,obs}$ using two illustrative $\tilde{\boldsymbol{\mu}}_i$'s. All $\boldsymbol{b}_i$'s on a given circle contribute the same amount to the middle term in (4.4). Thus (4.5) is minimized along the line of shortest distance from $\tilde{\boldsymbol{\mu}}_i$ to $C_{i,obs}$.

If $\tilde{\boldsymbol{\mu}}_i \in C_{i,obs}$ then $\boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i} = \tilde{\boldsymbol{\mu}}_i$ and (4.5) becomes $tr\tilde{\Sigma}_i$.

Determining the overall minimum of (4.5) requires finding the minimizing radius which is $(k+1)^{-1}k\sqrt{(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})^T(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})}$. Insertion into (4.4) yields $tr\tilde{\Sigma}_i + (k+1)^{-1}k(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})^T(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})$. As in (4.3) we sum over i to obtain,

$$\sum_{i=1}^{n} tr\tilde{\Sigma}_i + (k+1)^{-1}k\sum_{i=1}^{n}(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})^T(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i}) \qquad \ldots (4.6)$$

which extends (4.3) to provide a general model screening criterion in the presence of censoring. In (4.6) we denote $\sum_{i=1}^{n} tr\tilde{\Sigma}_i$ by P (for penalty), $\sum_{i=1}^{n}(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})^T(\tilde{\boldsymbol{\mu}}_i - \boldsymbol{v}_{\bar{\boldsymbol{\mu}}_i})$ by G (for goodness-of-fit) and the total by D_k.

To compute (4.6) only requires the values of $\tilde{\boldsymbol{\mu}}_i$ and $tr\tilde{\Sigma}_i$ which in turn requires expectations with respect to the predictive distributions of the $\boldsymbol{t}_{i,rep}$ given $\boldsymbol{t}_{obs}$. Such expectations are computed by Monte Carlo integration using samples from these distributions. Following the discussion below (3.2), $\boldsymbol{t}_{i,rep}$ given $\boldsymbol{t}_{obs}$ has a singular distribution but it can be readily sampled. In particular if $\boldsymbol{\theta}^*, \boldsymbol{\beta}^*$ is a draw from the posterior $f(\boldsymbol{\theta}, \boldsymbol{\beta} \mid \boldsymbol{t}_{obs})$ and $\boldsymbol{u}_i^*$ is drawn from $f(\boldsymbol{u}_i \mid \boldsymbol{\theta}^*, \boldsymbol{\beta}^*)$ then $\boldsymbol{t}_i^*$, as a function of $\boldsymbol{u}_i^*$, is a draw from the desired distribution.

It is obvious that the argument leading to (4.6) is valid in more than two dimensions so (4.6) provides a general model choice criterion for multivariate data models with possible censoring.

5. A bivariate event time example

We return to the motivating example introduced in section 1. A sample of 169 patients entered the registry after January 1, 1989 and were still alive at the end of follow up on December 31, 1991. All of these patients had received a Center for Disease Control (CDC) defined AIDS diagnosis at the time of entry ($t = 0$). For these individuals the two event times (measured in days) that we study are time to first hospitalization after clinical diagnosis of AIDS and time to treatment with Zidovudine (AZT). A tie occurs if AZT treatment is initialized at the time of hospitalization. Interest focuses on the association between the times as well as the effect of patient covariates on the times.

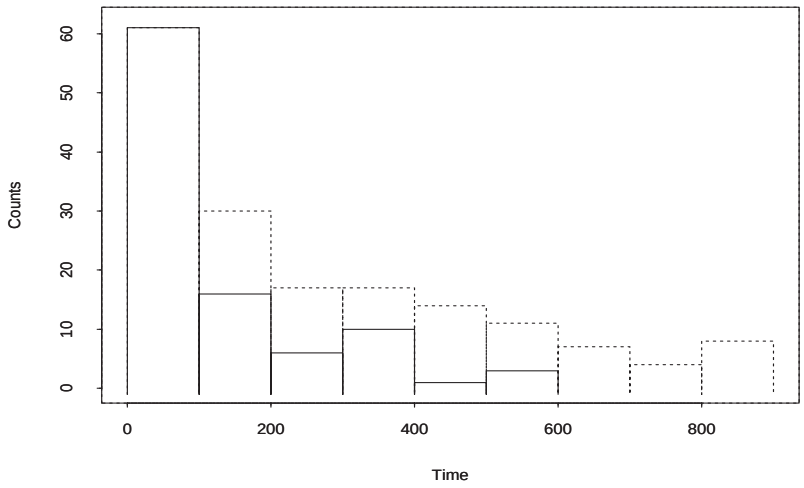

Figure 2: Histogram of censored and uncensored times to hospitalization.

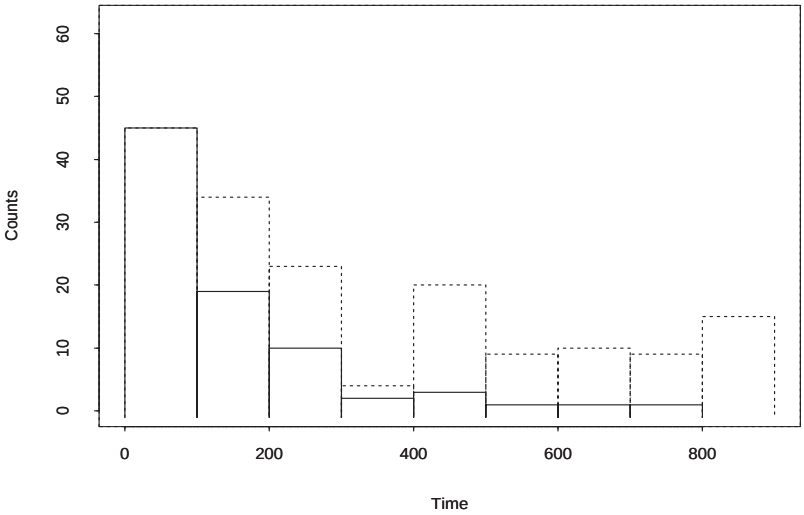

Figure 3: Histogram of censored and uncensored times to AZT treatment.

With an exploratory perspective, Figure 2 presents a histogram of the censored and uncensored hospitalization times. Rectangles below solid lines record frequencies of uncensored times. Rectangles above solid lines but below dashed lines record frequencies of censored times. Figure 3 presents a similar histogram for the AZT times. Alternatively, we can also look at the Kaplan-Meier estimate of the event times. Censoring is heavy; 72 out of the 169 hospitalization times were censored, 87 of the 169 AZT times were censored, 44 times were jointly censored.

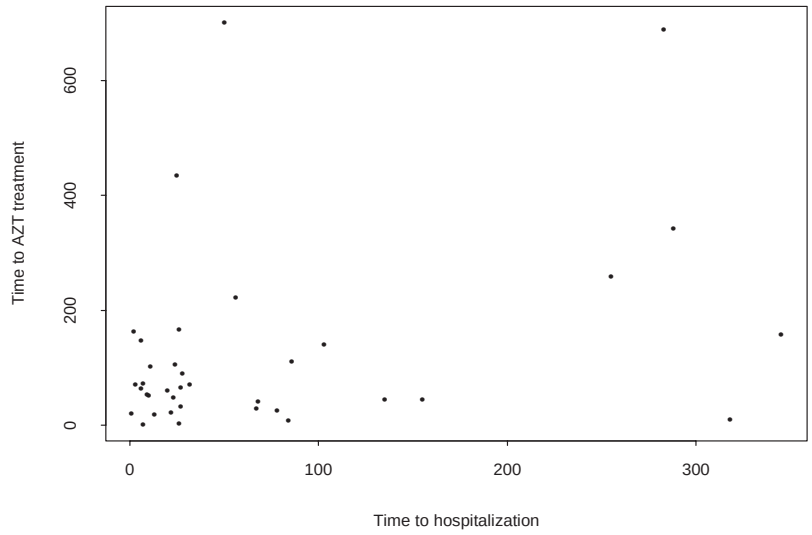

Figure 4: Scatterplot of uncensored times

Figure 4 provides a scatter plot of the uncensored pairs (54 points) and suggests weak dependence between the times. The Spearman correlation coefficient is 0.204. In these uncensored pairs there are 13 ties. Three covariates were used: sex (two levels), race (three levels - white, black, others) and age (continuous).

We first address the matter of model choice. For illustration six models were considered, (i) an independence model assuming independent Exponential and Weibull proportional hazards for each event time, 10 and 12 parameter models (ii) a GMO with constant (exponential) baseline hazards, a 15 parameter model, (iii) a GMO with Weibull baseline hazards, an 18 parameter model, (iv) a GPS with constant baseline hazards, a 25 parameter model and (v) a GPS with Weibull baseline hazards, a 30 parameter model and (vi) a GPS with Weibull baseline hazards but assuming $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1'$ and $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2'$, a 22 parameter model. Each model was fitted using the Gibbs sampler as described in section 3. The Raftery and Lewis (1995) criterion was used to determine the number of iterations.

One may be concerned that, for instance, a 30 parameter model as in (v) is excessive for the data set. However, at most 20 parameters are devoted to covariate coefficients (relative to the independence model which commits to 8) and we do have essentially 2x169 = 338 observations. Moreover, the Gibbs samplers for all six models were very well behaved using a variety of diagnostics; the models seem well identified. Table 1 demonstrates the use of the model choice criterion from section 4, presenting G, P and D_k for $k = 1, 3, 9$ and ∞.

Table 1. Model choice for the example in section 4.

Model	G	P	D_1	D_3	D_9	D_∞
(i) Independence Exponential	1156.72	660.48	1238.84	1528.02	1701.53	1817.20
(i′) Independence Weibull	1103.86	429.04	980.97	1256.93	1422.51	1532.90
(ii) GMO Exponential	876.37	593.29	1031.47	1250.57	1382.02	1469.66
(iii) GMO Weibull	823.21	256.64	668.24	874.05	997.53	1079.85
(iv) GPS Exponential	817.62	544.38	953.19	1157.59	1280.24	1362.00
(v) GPS Weibull	744.51	244.7	617.00	803.13	914.81	989.26
(vi) GPS (Weibull) ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1', \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2'$)	787.17	251.80	645.38	842.18	960.25	1038.97

Recalling that small values of D are preferred, the independence model is the poorest. Models (iii) and (v) are substantially better than models (ii) and (iv); model (i′) is substantially better than model (i). The more flexible Weibull hazards are needed. Model (v) emerges as best though the parsimonious model (vi) is a close competitor. Overfitting relative to models (iii) and (vi) is not suggested since P is smallest for (v).

Table 2 summarizes the posterior distributions for the 30 parameters in model (v), providing for each, the posterior mean and a 95% equal tailed interval estimate. Event 1 (hence subscript 1) refers to time to hospitalization, event 2 to time to AZT treatment. All the Weibull index parameters are about the same. More importantly, all are greater than one implying all hazards are increasing, reasonable in our context. The scale parameters are all quite small because they are inversely related to the means of the t^γ which are of the order of 10^4 or 10^5.

Table 2. Summary of marginal posteriors for model (v).

Scale	λ_1	λ_2	λ_{12}	λ_1'	λ_2'
Mean	12.0e-05*	8.23e-05	11.0e-05	41.0e-05	22.0e-05
Lower 95%	3.60e-05	3.0e-05	3.5e-05	17.0e-05	6.60e-05
Upper 95%	25.0e-05	15.0e-05	26.0e-05	84.0e-05	49.0e-05
Index	γ_1	γ_2	γ_{12}	γ_1'	γ_2'
Mean	1.701	1.685	1.685	1.688	1.691
Lower 95%	1.500	1.473	1.479	1.489	1.476
Upper 95%	1.928	1.906	1.893	1.929	1.940
Age	β_1	β_2	β_{12}	β_1'	β_2'
Mean	0.001	-0.001	-0.021	-0.006	-0.016
Lower 95%	-0.004	-0.008	-0.041	-0.024	-0.033
Upper 95%	0.005	0.007	-0.006	0.011	-0.001
Sex (M)	β_1	β_2	β_{12}	β_1'	β_2'
Mean	-0.165	-0.062	0.178	0.284	0.119
Lower 95%	-0.249	-0.211	-0.174	0.060	-0.127
Upper 95%	-0.069	0.078	0.514	0.537	0.353
Race (W)	β_1	β_2	β_{12}	β_1'	β_2'
Mean	-0.500	-0.146	-1.033	-1.106	-0.717
Lower 95%	-0.764	-0.438	-1.725	-1.606	-1.209
Upper 95%	-0.287	0.229	-0.498	-0.592	-0.258
Race (B)	β_1	β_2	β_{12}	β_1'	β_2'
Mean	-0.243	-0.046	-1.071	-0.983	-0.654
Lower 95%	-0.392	-0.259	-1.665	-1.476	-1.035
Upper 95%	-0.122	0.202	-0.673	-0.509	-0.309

* 12.0e-05 $= 12.0 \times 10^{-5}$

Turning to the regression coefficients we find the following. For age, since β_1 and β_2 are insignificant, age doesn't affect time to hospitalization when it occurs before time to AZT treatment and vice versa. With β_{12} significantly negative, when both events occur together, risk decreases with age. Since β_2' is significantly negative, risk for AZT treatment after hospitalization decreases with age. With regard to sex, since β_1 is significantly negative, males are at less risk for hospitalization than females when this occurs prior to AZT treatment. Interestingly, since β_1' is significantly positive, the conclusion is reversed for hospitalization when it occurs after AZT treatment. As for race, with the "other" group as baseline, notice that for both blacks and whites $\beta_1, \beta_{12}, \beta_1'$ and β_2' are all significantly negative. It follows that both blacks and whites are at less risk for hospitalization than those in the "other" group and, to a lesser extent, for AZT treatment. Also, since β_1' is significantly smaller than β_1 and β_2' is significantly smaller than β_2 the gap widens in waiting for the second event.

Acknowledgments. The authors thank Cindy Christiansen, Thomas J. McLaughlin and Stephen B. Soumerai of the Department of Ambulatory Care and Prevention, Harvard Medical School; and Harvard Pilgrim Health Care for providing the AIDS data set and for numerous important conversations. The authors also thank Carl Morris of the Department of Statistics at Harvard University for valuable discussions in focusing this effort. The work of both authors was sup-

ported in part by NSF Grant DMS 9301316. The work of the second author was also supported by AHCPR Grant HS 07118 under Principal Investigator Carl Morris. Lastly, the authors acknowledge the support of AHCPR Grant 06893 in enabling the development of the data set.

References

Block, H.W. and Basu, A.P. (1974). A continuous bivariate exponential extension, *Journal American Statistical Association*, **69**, 1031-1037.

Clayton, D.G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model, *Journal of Royal Statistical Society, series A*, **148**, 82-117.

Cox, D.R. (1972). "Regression models and life tables" (with discussion), *Journal of Royal Statistical Society, series B*, **34**, 187-220.

Dellaportas, P. and Smith, A.F.M. (1993). "Bayesian inference for generalized linear and proportional hazard models via Gibbs sampler", *Applied Statistics*, **42**, 443-460.

Diggle, P.J., Liang, K-Y. and Zeger, S.L. (1994). "Analysis of Longitudinal data", Oxford, Oxford University Press.

Freund, J.E. (1961). "A bivariate extension of the exponential distribution", *Journal of American Statistical Association*, **56**, 971-977.

Gelfand, A.E. and Ghosh, S.K. (1998). "Model choice: A minimum posterior predictive loss approach," *Biometrika*, **85** (to appear).

Gelfand, A.E. and Smith, A.F.M. (1990). "Sampling based approaches to calculating marginal densities", *Journal of American Statistical Association*, **85**, 398-409.

Gilks, W.R. and Wild, P. (1992). "Adaptive rejection sampling for Gibbs sampling", *Applied Statistics*, **41**, 337-348.

Hougaard, P. (1986). "A class of multivariate failure time distributions", *Biometrika*, **73**, 671-678.

Kass, R.E. and Raftery, A.E. (1995). "Bayes factors", *Journal of American Statistical Association*, **90**, 773-795.

Klein, J.P., Keiding, N. and Kamby, C. (1989). "Semiparametric Marshall-Olkin models applied to the occurrence of Metastasis at multiple sites after breast cancer", *Biometrics*, **45**, 1073-1086.

Lagakos, S.W., Sommer, C.J. and Zelen, M. (1978). "Semi-Markov models for partially censored data", *Biometrika*, **65**, 311-317.

Leurgans, S., Tsai, W.Y. and Crowley, J. (1982). "Freund's bivariate exponential distribution and censoring", In *Survival Analysis*, J. Crowley and R.A. Johnson (eds.) 230-242, Hayward, California, Institute of Mathematical Statistics.

Marshall, A.W. and Olkin, I. (1967). "A multivariate exponential distribution", *Journal of American Statistical Association*, **62**, 30-40.

Oakes, D. (1989). "Bivariate survival models induced by frailties", *Journal of American Statistical Association*, **84**, 487-493.

Proschan, F. and Sullo, P. (1974). "Estimating the parameters of a bivariate exponential distribution in several sampling situations", In *Reliability and Biometry*, Proschan, F. and Serfling, R.J. (eds.) SIAM, 423-440.

Raftery, A.E. and Lewis, S.M. (1995). "The number of iterations, convergence diagnostics and generic Metropolis algorithms", In *Practical Markov Chain Monte Carlo* (W.R. Gilks, D.J. Spiegelhalter and S. Richardson, eds.), London, Chapman and Hall, 115-130.

Voelkel, J. (1980). "Multivariate counting processes and the probability of being in response function", Ph.D dissertation, Statistics Department, University of Wisconsin-Madison.

Zellner, A. (1994). "Bayesian and non-Bayesian estimation using balanced loss functions", In *Statistical Decision Theory and Related topics V*, (eds. S.S. Gupta and J.O. Berger), New York, Springer Verlag, 377-390.

Sujit K. Ghosh
Department of Statistics
NC State University
Raleigh, NC 27695-8203
U.S.A.
e-mail : sghosh2@eos.ncsu.edu

Alan E. Gelfand
Department of Statistics
University of Connecticut
Storrs, CT 06269-3120
U.S.A.
e-mail :gelfand@stat.uconn.edu