Sensitivity analysis for informative censoring in parametric survival models

FOTIOS SIANNIS*
MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 2SR, UK
fotios.siannis@mrc-bsu.cam.ac.uk

JOHN COPAS
Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

GUOBING LU
MRC HSRC, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK

SUMMARY
Most statistical methods for censored survival data assume there is no dependence between the lifetime and censoring mechanisms, an assumption which is often doubtful in practice. In this paper we study a parametric model which allows for dependence in terms of a parameter $\delta$ and a bias function $B(t, \theta)$. We propose a sensitivity analysis on the estimate of the parameter of interest for small values of $\delta$. This parameter measures the dependence between the lifetime and the censoring mechanisms. Its size can be interpreted in terms of a correlation coefficient between the two mechanisms. A medical example suggests that even a small degree of dependence between the failure and censoring processes can have a noticeable effect on the analysis.

Keywords: Sensitivity analysis; informative censoring; proportional hazard models.

1. INTRODUCTION

Methods for analyzing censored survival data are highly developed and widely used. Although many different models and approaches have been studied in the literature, they almost invariably assume that the censoring is non-informative or ignorable. This is most clearly seen in the way censored observations enter the likelihood function. If an observation is censored at time $c$, the contribution to the likelihood is just the probability that lifetime $T$ exceeds $c$. The fact that the censoring has occurred when it did has not altered the distribution of $T$, hence the censoring mechanism is irrelevant for inference about the distribution of $T$.

In many applications, however, the assumption of ignorable censoring is at best an approximation and at worst seriously misleading. In a medical trial for example, a patient may be withdrawn from treatment because their condition is deteriorating or they are showing side effects which need alternative treatment. In this case withdrawal at time $c$ may indicate death is likely sooner than might have been expected.

*To whom correspondence should be addressed.
otherwise. Conversely, a patient may withdraw because they are feeling better and in no need of continuing treatment. In this case, the event that censoring has occurred may increase the expected remaining lifetime. Lagakos (1979) gives a number of such examples where the assumption of non-informative censoring is questionable. It is clear that if such dependence is ignored, the resulting inference will be biased.

In this paper we study the bias induced by informative censoring by embedding censored survival data in a competing risks framework. For each individual we assume there is a potential random censoring time \( C \) and a potential random lifetime \( T \). The censoring is non-informative if \( C \) and \( T \) are independent (conditional on values of covariates) and the parameters that govern their distributions are distinct. We observe the time \( Y = \min(T, C) \), and the censoring indicator \( I = 1 \) if \( T \leq C \) and \( I = 0 \) if \( T > C \). Unfortunately, the joint distribution of \( Y \) and \( I \) is not sufficient to identify the joint distribution of \( T \) and \( C \), so that unless we make additional assumptions it is impossible to estimate the level of dependence. Tsiatis (1972) demonstrated that if we have a model with dependent risks, then a proxy model always exists in which \( T \) and \( C \) are independent and which gives the same joint distribution of the observables \( Y \) and \( I \).

Heckman and Honore (1989) showed that if a non-parametric proportional hazard model for the joint distribution of \( T \) and \( C \) with covariates is assumed, then the estimation of the dependence is technically possible. However, their method is too dependent on unverifiable model assumptions to be useful in practice. Fisher and Kanarek (1974) proposed a rather different model in which the remaining lifetime \((T - C)\) after censoring at time \( C \) is assumed to be equivalent to surviving an extra time \( \alpha(T - C) \) had the censoring not occurred, for some constant factor \( \alpha \). Moeschberger and Klein (1995) review models for dependent competing risks, while Hogan and Laird (1997) and Little (1995) review methods that simultaneously model the outcome of interest and the missing process with both longitudinal and failure-time outcomes. More recently, Scharfstein et al. (1999) and Scharfstein and Robins (2002) propose a semi-parametric approach for longitudinal and time to event data, introducing a sensitivity analysis based on unmeasured factors. Of these papers, that of Scharfstein and Robins (2002) is most closely related to our work, although the two approaches have rather different objectives (see the final section for further comments).

Given that we are unable to estimate the level of dependence between the lifetime and censoring mechanisms, and hence in practice are unable to fit a definitive model, the next best thing is to develop a sensitivity analysis. Our aim is to link a sensitivity analysis to already established parametric methods of analyzing survival data. The association between the failure time and the censoring processes is introduced by modelling the conditional distribution of \( C \) given the (possibly unobserved) value of \( T \). This is achieved by allowing the parameter of the marginal distribution of \( C \) to depend on \( T \) through a bias function \( B(t, \theta) \) and a dependence parameter \( \delta \). A relatively simple sensitivity analysis on the parameter of interest is developed, based on linear approximations for small values of \( \delta \). Essentially, the maximum likelihood estimate for a given (small) value of \( \delta \) is approximately equal to the estimate that would be obtained assuming ignorable censoring (with \( \delta = 0 \)), plus \( \delta \) times a sensitivity index which depends on the observed pattern of censored observations. With these approximations we avoid having to fit the joint distribution of \( T \) and \( C \) explicitly; all parameter estimates required in calculating the sensitivity index are available from the standard analysis.

We illustrate our methods by re-analyzing the survival data of multiple myeloma patients discussed by Krall et al. (1975). In this study, patients were followed up for many months, and there were a substantial number of censored observations. Exact independence between \( T \) and \( C \) (i.e. \( \delta = 0 \)) seems a very strong assumption, although the fact that the original authors did not report any concerns about informative censoring suggests that \( \delta \) is unlikely to be large. The sensitivity indices calculated for these data suggest that estimates of median survival can be quite sensitive to small departures from the ignorability assumption made in the original analysis, especially when predicting survival for high risk patients. This example is probably typical of many studies where censoring occurs for reasons which are
Section 3 develops the resulting sensitivity analysis, discusses the choice of bias function, and shows in section 5. The paper concludes with a discussion in Section 6.

proportional hazards models in Section 4, and illustrated with the analysis of the multiple myeloma data.

not controlled in advance, e.g. withdrawals which are self-selected rather than structural, such as reaching the end of the study. In such cases we suggest that the tacit assumption of ignorability should always be viewed with caution, and that a sensitivity analysis on these lines can be a useful part of routine survival data analysis.

The paper is organized as follows. Section 2 introduces our model for the joint distribution of $T$ and $C$. Section 3 develops the resulting sensitivity analysis, discusses the choice of bias function, and shows how $\delta$ can be interpreted as a correlation coefficient. The sensitivity analysis is extended to parametric proportional hazards models in Section 4, and illustrated with the analysis of the multiple myeloma data in section 5. The paper concludes with a discussion in Section 6.

2. THE MODEL

To establish the notation, let $S_T(t, \theta)$ be the survival function of failure time $T$, where $\theta$ is the unknown parameter of interest. Equivalently, the distribution of $T$ can be specified by its probability density function, its hazard function or its cumulative hazard function: respectively

$$f_T(t, \theta) = -\frac{d}{dt}S_T(t, \theta), \quad h_T(t, \theta) = -\frac{d}{dt}\log S_T(t, \theta), \quad H_T(t, \theta) = -\log S_T(t, \theta).$$

The density function $f_T(t, \theta)$ gives the usual score and information functions, $s_T(t, \theta) = \frac{\partial}{\partial \theta} \log f_T(t, \theta)$ and $i_\theta = \text{Var}_T\{s_T(T, \theta)\}$.

Similarly, the censoring time $C$ is also assumed to follow a parametric model with survival function $S_C(c, \gamma)$, with corresponding functions $f_C(c, \gamma)$, $h_C(c, \gamma)$, $H_C(c, \gamma)$, $s_C(c, \gamma)$ and $i_\gamma$. We assume that $\theta$ and $\gamma$ are distinct parameters so that any information about one tells us nothing about the other. The parameter $\gamma$ is a nuisance parameter so far as inference about $\theta$ is concerned.

This paper is based on the assumption that the conditional distribution of $C$ given $T$ has exactly the same parametric form as its marginal distribution $f_C(c, \gamma)$, but with the parameter allowed to depend on $T$. Explicitly, the conditional density is

$$P(C = c|T = t) = f_C(c, \gamma + \delta i_\gamma^{-\frac{1}{2}}B(t, \theta)).$$

(2.1)

Note that if $\delta = 0$ then $T$ and $C$ are independent and the censoring is ignorable. The parameter $\delta$ can be thought of as measuring the size of the dependence between $T$ and $C$ and the bias function $B(t, \theta)$ as measuring the pattern of this dependence. Our aim is to develop a local sensitivity analysis on the lines of Copas and Eguchi (2001), assuming that $\delta$ is small. We have in mind applications where we suspect there may be modest dependence in the censoring. We take the standard ignorable analysis as our starting point, and see how sensitively our conclusions change as we start to move $\delta$ away from zero.

For simplicity of notation we will assume that $\theta$ and $\gamma$ are scalar parameters, and that $B(t, \theta)$ is a scalar function. More generally $\theta$ and $\gamma$ may be vectors, in which case $B(t, \theta)$ can also be a vector with the same number of components as $\gamma$, so that $\delta$ remains a scalar. All of the formulae below apply to the vector case also, with the obvious changes to notation. By concentrating on the marginal distribution of $T$ we are also ignoring the presence of covariates, but the generalization to a proportional hazards model is straightforward, as shown in Section 4.

The form of the second argument of the function in (2.1) is invariant under an arbitrary linear rescaling of $B(t, \theta)$, so to fix the notation we need to specify location and scale constraints on this function. The joint density function of $T$ and $C$ is

$$f_{T,C}(t, c) = f_T(t, \theta)f_C(c, \gamma + \delta i_\gamma^{-\frac{1}{2}}B(t, \theta))$$

$$\simeq f_T(t, \theta)f_C(c, \gamma)[1 + \delta i_\gamma^{-\frac{1}{2}}s_C(c, \gamma)B(t, \theta)].$$

(2.2)
Integrating over $T$ gives the marginal density of $C$ to be

$$
\int_0^\infty f_T(t, \theta) f_C(c, \gamma) \left[ 1 + \delta i_\gamma^{-1} s_C(c, \gamma) B(t, \theta) \right] dt = f_C(c, \gamma) + \delta i_\gamma^{-1} s_C(c, \gamma) \int_0^\infty B(t, \theta) f_T(t, \theta) dt. \tag{2.3}
$$

A natural constraint on $B(t, \theta)$ is therefore

$$
E_T[B(T, \theta)] = \int_0^\infty B(t, \theta) f_T(t, \theta) dt = 0, \tag{2.4}
$$

so that the marginal density of $C$ in (2.3) equals $f_C(c, \gamma)$ up to first order terms in $\delta$. We will also require that $B(t, \theta)$ has finite variance, so without any loss of generality we can assume that

$$
\text{Var}_T[B(T, \theta)] = E_T[B^2(T, \theta)] = 1. \tag{2.5}
$$

3. Sensitivity analysis

3.1 Local sensitivity analysis

Suppose that we have a random sample of $n$ observations. For the $i$th observation, let $t_i$ be the value of $\min(T_i, C_i)$ and $I_i$ be the value of the censoring indicator defined to be 1 if $T_i \leq C_i$ and 0 if $T_i > C_i$, $i = 1, 2, \ldots, n$. Then the log-likelihood function is

$$
L(\theta, \gamma) = \sum_{i=1}^n I_i \log P(T = t_i \cap T < C) + (1 - I_i) \log P(C = t_i \cap C < T). \tag{3.1}
$$

Evaluating each of these probabilities from (2.2), and expanding up to first order terms in $\delta$, we find after a little algebra that

$$
L_\delta(\theta, \gamma) \simeq L_0(\theta, \gamma) + \delta i_\gamma^{-1} \sum_{i=1}^n \left\{ (1 - I_i) \mu(t_i, \theta) s_C(t_i, \gamma) - I_i B(t_i, \theta) \frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \right\} \tag{3.2}
$$

where

$$
\mu(t, \theta) = \int_t^\infty \frac{B(u, \theta) f_T(u, \theta) du}{S_T(t, \theta)}.
$$

The leading term in (3.2) is

$$
L_0(\theta, \gamma) = \sum_{i=1}^n \left\{ I_i \log h_T(t_i, \theta) + (1 - I_i) \log h_C(t_i, \gamma) - H_T(t_i, \theta) - H_C(t_i, \gamma) \right\}, \tag{3.3}
$$

the likelihood function for all the parameters which we would have if $T$ and $C$ were assumed to be independent. This is seen to be the sum of two parts, the usual log-likelihood from the observed values of $T$ assuming that $C$ is ignorable, and the usual log-likelihood from the observed values of $C$ assuming that $T$ is ignorable. Moreover, in (3.2) the term which is multiplied by $\delta$ depends on both parameters, so in practice we need to substitute estimates for these. As the term is already a multiple of $\delta$ it makes no difference, to this order of approximation, whether the estimates $(\hat{\delta}_\theta, \hat{\gamma})$ or $(\hat{\theta}_0, \hat{\gamma}_0)$ are used. For
Sensitivity analysis for informative censoring in parametric survival models

Assuming that $\delta$ is fixed, let $\hat{\theta}_0$ be the value of $\theta$ which maximizes (3.2) and $\hat{\theta}_0$ be the usual maximum likelihood estimate from (3.3). Then, differentiating (3.2) with respect to $\theta$ and rearranging terms gives

$$\hat{\theta}_0 - \hat{\theta}_0 \simeq \delta \frac{i(\theta)}{\sum_{i=1}^n \left(1 - I_i \frac{\partial \mu(t_i, \theta)}{\partial \theta} s_C(t_i, \gamma) - I_i \frac{\partial B(t_i, \theta)}{\partial \theta} \frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \right)},$$  

(3.4)

where $i(\theta) = -\frac{\partial^2 L_0(\theta, \gamma)}{\partial \theta^2}$ is the observed information for $\theta$ based on the ignorable part of the likelihood. Analogous approximations can also be developed for the estimates $\hat{\gamma}_0$ and $\hat{\gamma}_0$ of the parameter $\gamma$.

### 3.2 Choice of $B(t, \theta)$

In principle, the bias function $B(t, \theta)$ can be chosen to reflect the pattern of dependence of most concern in the context of the particular application. For example, if $\gamma$ is a rate parameter of the distribution of $C$, then we could choose $B(t, \theta)$ to be an increasing, or a decreasing, function of $t$ to reflect the possibility that censoring is more likely for unusually large, or unusually small, values of $t$. For a general method, however, we present two arguments which suggest a particular functional form for $B(t, \theta)$.

The first is based on sample heterogeneity, that non-ignorability is the result of correlation between patient-specific random effects in the distributions of $T$ and $C$. Suppose that for a given subject, $T$ and $C$ are independent with probability density functions taking the respective forms

$$g_T(t, \theta + \epsilon_T t^{\frac{1}{2}}); \quad g_C(c, \gamma + \epsilon_C c^{\frac{1}{2}}),$$  

(3.5)

where $\epsilon_T$ and $\epsilon_C$ are (assumed small) random effects with means zero, variances $\sigma_T^2$ and $\sigma_C^2$, and covariance $\sigma_{TC}$. We assume that all three of these second moments are small, and of the same order of magnitude. Then the marginal distribution of $T$ is, for any fixed $t$,

$$f_T(t, \theta) = E \left\{ g_T(t, \theta + \epsilon_T t^{\frac{1}{2}}) \right\}$$

$$\simeq g_T(t, \theta) + \frac{\sigma_T^2}{2t} \frac{\partial^2 g_T(t, \theta)}{\partial \theta^2},$$

with a similar approximation for $f_C(c, \gamma)$. The joint distribution is

$$f_{T, C}(t, c) = E \left\{ g_T(t, \theta + \epsilon_T t^{\frac{1}{2}}) g_C(c, \gamma + \epsilon_C c^{\frac{1}{2}}) \right\}$$

$$\simeq f_T(t, \theta) f_C(c, \gamma) + \sigma_{TC}(\theta t^{\frac{1}{2}}) \frac{1}{2} \frac{\partial g_T(t, \theta)}{\partial \theta} \frac{\partial g_C(c, \gamma)}{\partial \gamma}$$

$$\simeq f_T(t, \theta) f_C(c, \gamma) \left[ 1 + \sigma_{TC}(\theta t^{\frac{1}{2}}) \frac{1}{2} s_T(t, \theta) s_C(c, \gamma) \right].$$

With the appropriate definition of $\delta$, this is just the same as (2.2) with $B(t, \theta)$ equal to the standardized score function

$$B(t, \theta) = t^{\frac{1}{2}} s_T(t, \theta).$$  

(3.6)
The second argument is based on the idea that, since the form of dependence is completely unknown, our assumptions about it should be as weak as possible so far as the information about $\theta$ is concerned. Following Bickel et al. (1992), Pagan and Ullah (1999) or van der Vaart (2000), we can find the efficient information bound for estimating $\theta$ in the light of uncertainty about the choice of $B = B(t, \theta)$. Extending the small-$\delta$ approximations leading to (3.4), the bound is approximately

$$\inf_B \left\{ \int \left( \frac{\partial B(T, \theta)}{\partial \theta} \right)^2 \right\},$$

(3.7)

where the minimization is taken over all functions $B$ satisfying the standardization constraints (2.4) and (2.5). But differentiating (2.4) shows that $E(\partial B/\partial \theta)$ is just minus the covariance of $B$ with $s_T(T, \theta)$, whose square is maximized when one is a linear function of the other. The bound is therefore attained at (3.6).

Notice that with this choice of $B$, the first order term in (2.2) gives a nice mathematical symmetry between the competing risks, in the sense that the conditional distribution of $C$ given $T$ has the same form as the conditional distribution of $T$ given $C$.

### 3.3 Sensitivity analysis based on the proportional hazard structure

The choice of $B(t, \theta)$ in (3.6) further simplifies in the special case when the $T$ and $C$ distributions have the parametric proportional hazards structure

$$h_T(t, \theta) = e^{\theta h^*_T(t)}; \quad h_C(c, \gamma) = e^{\gamma h^*_C(c)},$$

(3.8)

where $h^*_T(t)$ and $h^*_C(c)$ are known baseline hazard functions (e.g. Weibull hazards). This means that the role of the parameters $\theta$ and $\gamma$ is to increase or decrease the overall intensities of the $T$ and $C$ processes, not to alter how the hazards vary over time. The way the parameters $\theta$ and $\gamma$ enter into (3.8) reflects the usual log–linear parametrization of proportional hazards modelling. If (3.8) holds, it is easy to see that

$$s_T(c, \gamma) = 1 - H_T(t, \theta), \quad s_C(c, \gamma) = 1 - H_C(c, \gamma) \quad \text{and} \quad t_0 = t_T = 1,$$

(3.9)

which, together with (3.6), gives

$$B(t, \theta) = 1 - H_T(t, \theta).$$

(3.10)

The joint distribution of $T$ and $C$ now has the symmetric form

$$f_{T,C}(t, c) \simeq f_T(t, \theta) f_C(c, \gamma) \left[ 1 + \delta [1 - H_C(c, \gamma)][1 - H_T(t, \theta)] \right].$$

(3.11)

Also, $\mu(t, \theta) = -H_T(t, \theta)$, so the log-likelihood simplifies to

$$L_0(\theta, \gamma) \simeq L_0(\theta, \gamma) + \delta t_T^{-1} \sum_{i=1}^n \left\{ H_T(t_i, \theta)H_C(t_i, \gamma) - I_i H_C(t_i, \gamma) - (1 - I_i) H_T(t_i, \theta) \right\},$$

(3.12)

giving

$$\hat{\theta}_0 - \hat{\theta}_0 \simeq \delta t(\theta)^{-1} t_T^{-1} \sum_{i=1}^n \left\{ H_T(t_i, \theta)H_C(t_i, \gamma) - (1 - I_i) H_T(t_i, \theta) \right\}. $$

(3.13)

We shall see in Section 4 that these expressions generalize in a straightforward way to proportional hazards models in which the parameters $\theta$ and $\gamma$ become linear functions of covariates.
3.4 Sensitivity indices and confidence intervals

The general form of the bias approximations (3.4) and (3.13) is that, up to linear terms in \( \delta \), the maximum likelihood estimate \( \hat{\theta}_3 \) is approximately equal to \( \hat{\theta}_0 \) plus \( \delta \) times a factor which can be estimated from the standard ignorable analysis. We write this as

\[
\hat{\theta}_3 - \hat{\theta}_0 = \delta U + O(\delta^2),
\]

where \( U \) is the first order correction factor, which we will call the sensitivity index. In many applications, \( \theta \) is simply a convenient way of parametrizing interpretable quantities such as the median survival or the proportion surviving beyond some given time. If we are interested in some function \( J(\theta) \), say, then the corresponding first order sensitivity analysis for \( J \) is

\[
J(\hat{\theta}_3) \approx J(\hat{\theta}_0) + \delta J'(\hat{\theta}_0) U.
\]

For a confidence interval, the approximations leading to (3.7) give

\[
\{\text{Var}(\hat{\theta}_3)\}^{1/2} = \{r(\theta)\}^{-1/2} + O(\delta^2).
\]

Hence, retaining only linear terms in \( \delta \), the asymptotic confidence interval for \( \theta \) is approximately

\[
\hat{\theta}_0 - \delta U \pm z_{\alpha} \{r(\theta)\}^{-1/2},
\]

where \( z_{\alpha} \) is the appropriate standard normal percentage point. This is just the standard confidence interval for the ignorable model shifted by the bias correction in (3.14). Of course this is only valid if \( \delta \) is small—in general, non-ignorability will affect the variance as well as the bias in \( \hat{\theta} \).

For a local sensitivity analysis in practice, we envisage fixing a maximum but small value for \( \delta \) to give \((-\delta, \delta)\) as a plausible range of dependence parameters that we wish to consider. This leads to \( \hat{\theta}_0 \pm \delta U \) as the plausible range of values of the estimate of \( \theta \), and

\[
\hat{\theta}_0 \pm \left\{ \delta |U| + z_{\alpha} \{r(\theta)\}^{-1/2} \right\}
\]

as the corresponding conservative confidence interval for \( \theta \) itself.

3.5 Interpretation of \( \delta \)

For the sensitivity analysis to be useful, we need to have a way of judging whether the dependence implied by any particular value of \( \delta \) is large or small. An obvious measure of association is the correlation coefficient between \( T \) and \( C \), but this may not be appropriate if the dependence is non-linear and the distributions highly skewed. We might want to consider instead the correlation between some function of \( T \), say \( A(T, \theta) \), and some function of \( C \), say \( D(C, \gamma) \). For completeness we allow these transformations to depend also on the parameters so that, for example, we can accommodate the choices

\[
A = H_T(T, \theta); \quad D = H_C(C, \gamma),
\]

these being the monotonic transformations which transform the marginal distributions of \( T \) and \( C \) to the canonical form of two unit exponential distributions.

We show in the Appendix that, if \( T \) and \( C \) are observations from (2.2), then up to linear terms in \( \delta \),

\[
|\text{Corr}(A(T, \theta), D(C, \gamma))| \leq |\delta|.
\]
This establishes that $\delta$ can be interpreted as an upper bound to all correlations between the process generating $T$ and the process generating $C$. Given smoothness restrictions needed to justify our approximations, this bound holds for all transformations $A$ and $D$ as well as for all choices of bias function $B$ satisfying the standardization constraints (2.4) and (2.5).

Inequality (3.17) is attained for the particular transformations $A = B(T, \theta)$ and $D = s_C(c, \gamma)$. Under proportional hazards (3.8), and with the bias function (3.10), both of these functions are linear in $H_T(T, \theta)$ and $H_C(C, \gamma)$ respectively, so $A$ and $D$ are just the canonical transformations (3.16).

We again emphasize that these results are approximations for small values of $\delta$.

4. PROPORTIONAL HAZARDS MODELLING WITH COVARIATES

All of the above theory generalizes to models in which we have a vector of covariates $x$ measured with each observation. The familiar proportional hazards model is obtained by replacing $\theta$ and $\gamma$ in (3.8) by linear combinations of $x$, giving the hazard functions

$$h_T(t, \theta, x) = e^{\theta^{T}x}h_T^*(t); \quad h_C(c, \gamma, x) = e^{\gamma^{T}x}h_C^*(c).$$

(4.1)

For a fully parametric analysis, we assume that the form of the baseline hazard functions are known (e.g. two Weibull models). This assumes that both $T$ and $C$ follow proportional hazards in $x$, although this is clearly not necessary for the general formulation. All of the distributions and functions discussed in earlier sections are now conditional on $x$. For simplicity we will assume that the bias function depends on $x$ only through the value of the linear predictor $\theta^{T}x$, and that the dependence parameter $\delta$ does not depend on $x$, but again neither of these is a necessary restriction.

Extending the previous notation in the obvious way, and assuming (3.10), the joint density of $T$ and $C$ becomes

$$f_{T,C}(t, c, x) \simeq f_T(t, \theta, x)f_C(c, \gamma, x)[1 + \delta[1 - H_C(c, \theta, x)][1 - H_T(t, \gamma, x)]],$$

and the log-likelihood follows immediately from (3.12) as

$$L_4(\theta, \gamma) \simeq L_0(\theta, \gamma) + \delta \sum_{i=1}^{n} \left[H_T(t_i, \theta, x_i)H_C(t_i, \gamma, x_i) - I_iH_C(t_i, \gamma, x_i) - (1 - I_i)H_T(t_i, \theta, x_i)\right] .$$

The corresponding approximation to the bias in the estimating vector $\hat{\theta}$ is

$$\hat{\delta}_4 - \hat{\delta}_0 \simeq \delta(\hat{\theta})^{-1}\sum_{i=1}^{n} \left[x_i[H_T(t_i, \theta, x_i)H_C(t_i, \gamma, x_i) - (1 - I_i)H_T(t_i, \theta, x_i)]\right].$$

(4.2)

5. EXAMPLE

As an example, we re-analyze the data discussed in Krall et al. (1975). This study reported the survival times of 65 multiple myeloma patients who were diagnosed and treated with alkylating agents at the West Virginia University Medical Center. There were 16 concomitant variables for every patient, but we follow these authors by selecting just four of them as useful indicators for prognosis. See the cited paper for details of the background to this study and the selected covariates.

First we use these data to illustrate Section 3. Ignoring the covariates, marginal Kaplan–Meier plots for the lifetimes and censoring times suggest that simple exponential survival models give a reasonable fit to the marginal distributions of $T$ and $C$. This suggests the marginal models

$$f_T(t, \theta) = e^{\theta}e^{-\theta^{T}t}; \quad f_C(c, \gamma) = e^{\gamma}e^{-\gamma^{T}c}.$$  

(5.1)
Sensitivity analysis for informative censoring in parametric survival models

Of the 65 patients in the study, 48 (74%) were observed to die while 17 (26%) were censored. Under the assumption of ignorable censoring, the standard maximum likelihood analysis gives estimated median survival of 22.6 months for $T$ and 93.6 months for $C$.

For this simple model, the transformations in (3.16) are both linear, so from Section 3.5 the dependence parameter $\delta$ is equal to the ordinary correlation between $T$ and $C$ (for small $\delta$). Further, expression (3.13) simplifies to

$$
\hat{\theta}_0 \doteq \hat{\theta}_0 + \delta \sum_{i=1}^{65} \left\{ \hat{\gamma}^2 t_i^2 - (1 - I_i) t_i \right\} \sum_{i=1}^{65} t_i.
$$

For example, if we take ($-0.3, +0.3$) as our range for $\delta$, (5.2) leads to a range for the estimated median lifetime of $22.6 \pm 1.8$ months, suggesting that the marginal analysis is fairly robust to small departures from ignorability. In assessing whether $\delta = 0.3$ can be considered `small’, it may be helpful to note that, even if were able to observe the actual values of both $T$ and $C$ for the patients in this study, it is quite likely that we would not be able to detect a correlation of this magnitude.

For the full analysis, we now allow $\theta$ and $\gamma$ to be linear functions of the covariates as in Section 4. The model is (4.1) with constant baseline hazard functions for both processes. If we include an intercept term in the covariate vector $x$, there is no loss of generality in assuming that $\hat{h}_T^*(t) = \hat{h}_C^*(t) \equiv 1$ for all $t$, so that $w(x) = \theta^T x$ and $z(x) = \gamma^T x$ are the log-hazard rates for the $T$ and $C$ processes respectively. Estimating $\theta$ and $\gamma$ by the standard ignorable analysis for these data gives Figure 1 as the scatter plot of the fitted log-hazard rates for the 65 patients in this study. Note that these are positively correlated under the independence model: patients with a high risk of death (high $w(x)$) tend to be those with a high value for the intensity of censoring (high $z(x)$).
Fig. 2. Graph of $S(x)$ against $z(x)$ for $\delta = 0.3$.

Fig. 3. Min/max sensitivity analysis on the survival with respect to $z(x)$ ($|\delta| = 0.3$).
The model predicts that a patient with covariates \( x \) will have an exponential survival time with rate \( w(x) \). Our interest here is to see how estimates of this failure rate might be affected by unexplained dependence in the censoring. Let \( \hat{w}_3(x) \) be the value of \( w(x) \) when \( \theta \) is estimated by \( \hat{\theta}_0 \). Then (4.2) leads to the approximation

\[
\hat{w}_3(x) - \hat{w}_0(x) \simeq \delta U(x) = \delta \frac{\sum_{i=1}^n \{e^{z(x)} t_i^2 - (1 - I_i) t_i\}}{\sum_{i=1}^n t_i},
\]

where \( \delta \) is now the partial correlation between \( T \) and \( C \) given \( x \). This is a surprisingly simple formula, essentially the same as (5.2). The sensitivity index \( U(x) \) depends on \( x \) only through \( z(x) \). The greater the hazard of being censored (the greater is \( z(x) \)), the more sensitive is the estimate of \( w(x) \) to dependence between \( T \) and \( C \).

Figure 2 is a graph of the absolute value of the sensitivity index \( U(x) \) in (5.3) plotted against \( z(x) \). Note that for some values of \( z(x) \), those near \(-5\), the sensitivity is zero, meaning that the corresponding estimates of \( w(x) \) are essentially unaffected by small amounts of non-ignorability. Patients with high values of \( z(x) \) are those whose estimated survival is most sensitive to ignorability assumptions. This is easy to see in Figure 3, where the sensitivity analysis on the survival curves of the patients with the highest and the smallest values of \( z(x) \) are presented. The ranges shown are for \( \delta \in (-0.3, +0.3) \) as before. It is clear from this figure that for high risk patients, when \( z(x) \) is large, our inferences could be seriously misleading had we wrongly assumed ignorable censoring.

To give a simpler interpretation of these calculations, first note that the median survival, \( M(x) \) say, is inversely proportional to the failure rate \( w(x) \), so (5.3) can be rewritten in the obvious notation as

\[
\frac{\hat{M}_3(x) - \hat{M}_0(x)}{\hat{M}_0(x)} \simeq \delta U(x).
\]

This formula gives the proportional change in the median survival for a non-zero value of \( \delta \) compared to its value when \( \delta = 0 \). For \( \delta = 0.3 \), for example, the proportional change in the median for the patients in this study varies from zero (when \( z(x) = -5.2 \)) to 23% (when \( z(x) = -2.9 \)).

6. DISCUSSION

It is customary, when reporting censored survival studies, to give much attention to the survival times and explanatory factors, but little or no attention to the reasons for observations being censored. When ignorability can safely be assumed this is of no consequence, since the likelihood (3.3) factors into terms involving \( \theta \) and terms involving \( \gamma \), so that inference about \( \theta \) can be made without the need to model the distribution of \( C \). If the assumption of ignorability is questionable, and little or nothing is known about the censoring process, we suggest that a local sensitivity analysis on the lines discussed here can be a useful addition to standard parametric survival analysis. By using the approximations developed here, we avoid having to fit a dependent competing risks model, all that we need for estimating the sensitivity index are the two complementary standard censored survival analyses of \( T \) censored by \( C \) and of \( C \) censored by \( T \).

We have made two important assumptions, both of which can be tested with reference to the myeloma data of Section 5. Firstly, we have argued that little information is available from the data about the degree of dependence between \( T \) and \( C \), so we have suggested a sensitivity approach considering different but fixed values of \( \delta \), instead of a full analysis estimating \( \delta \) along with the other parameters in the usual way. If we fit the joint model (2.2) with (5.1) to the myeloma data (ignoring covariates), we find that the profile log-likelihood for \( \delta \) is extremely flat, confirming that the data give almost no information about...
Fig. 4. (a) Solid line: exact $\hat{\theta}_{\delta}$; semi-dotted line: approximate $\hat{\theta}_{\delta}$; horizontal dotted line: $\hat{\theta}_{\delta} = 0$; vertical dotted lines: cut-off points for $|\delta| = 0.3$. (b) The two lines in the middle represent the exact and approximate $\hat{\theta}_{\delta}$ exactly as in the top graph. The dashed lines represent the 95% confidence interval for values of $\delta \in [-0.3, 0.3]$ for the exact (short dashed line) and the approximate (long dashed line) model.

Numerically, we find $\hat{\delta} = -0.02$ but with a very large standard error. This is similar to the analogous discussion of ignorability in missing data problems in Copas and Li (1997), where confidence intervals for the parameter corresponding to $\delta$ are extremely wide, and estimates are found to be very sensitive to minor changes in the choice of model. See Chapter 7 of Crowder (2001) for a comprehensive review of identifiability issues within the competing risks framework.

A second assumption we have made is that linear approximations for small $\delta$ are sufficiently accurate for the values we wish to consider, up to $\delta = \pm 0.3$ in the example. Figure 4 tests this for the myeloma data by comparing our approximations with the exact conditional maximum likelihood estimates of $\theta$ and $\gamma$ given $\delta$ found from the full likelihood (3.1) with (5.1). Figure 4(a) shows the exact (solid line) and the approximate (stippled line) estimates of $\theta$ for $|\delta| \leq 1$. The horizontal dotted line represents the standard estimate $\hat{\theta}_0$ and the two vertical dotted lines mark the limits for $\delta$ we have used in Section 5. The agreement is reasonably close within this range, but very poor for larger values of $\delta$. In Figure 4(b) we compare the confidence interval for $\theta$ derived numerically from the full log-likelihood function using the standard asymptotic 95% likelihood threshold (solid and semi-dashed lines), with the approximate 95% confidence interval (3.15) (stippled and full-dashed lines). Again, the agreement is quite close within the range $\pm 0.3$. 

As Figure 4(a) shows, our approximations are only valid for small $\delta$. In missing data problems, Copas and Eguchi (2001) argue that if strong non-ignorability is suspected, so that large values of $\delta$ need to be entertained, then the resulting sensitivity bounds are likely to be too wide to be useful in practice. Peterson (1976) developed bounds on the joint and marginal survival functions, which were the maximum possible ones for any kind of dependence between the failure and the censored times. These bounds can be too wide to provide any useful information. If $T$ and $C$ are suspected of being highly correlated, then the only sensible approach is to study the censoring process in its own right and model the data properly using an explicit dependent risks model.

The sensitivity analysis depends on the choice of the bias function $B(t, \theta)$. For routine use, not requiring any special knowledge of the censoring process, we have suggested that $B$ can be taken as the standardized score function of the distribution of $T$. However, this should not over-ride any expert judgement of the practical context. In the analogous problem of missing data, Copas and Eguchi (2001) develop a rather more general model for the mis-specification of ignorability, and show that the analogous standardized score is the ‘worst case’ bias function so far as inference about $\theta$ is concerned.

Our formulation is based on the parametric form of the conditional distribution of $C$ given $T$ in (2.1). Essentially, mis-specification is modelled as a perturbation on a parameter of the distribution, rather than on the values of $T$ or $C$ directly. This is also reflected in (3.5), where model parameters are perturbed by patient-specific random effects. Scharfstein and Robins (2002) use a different formulation in which dependence between $T$ and $C$ is modelled as a bias function multiplying a non-parametric conditional hazard of the $C$ process. In our notation, and simplifying their formulation by excluding covariates, their model is essentially

$$
h_C(c|t, t > c) = h_0(c) \exp\{q(c, t)\}
$$

(6.4)

where $h_C(c|t, t > c)$ is the conditional hazard function of $C$ given $T = t > c$, and $h_0(c)$ is an unknown baseline function. This compares with our model (2.2), where the corresponding conditional hazard for $C$ is

$$
h_C(c, \gamma + \delta \frac{1}{2} B(t, \theta)) \approx h_C(t, \gamma) \exp\left\{\delta \frac{1}{2} B(t, \theta) \frac{\partial}{\partial \gamma} \log h_C(c, \gamma)\right\}.
$$

(6.5)

This is of the same general form, but with a parametric function for the baseline. Specifying $q(t, T)$ in (6.4) implies a choice for $\delta B(t, \theta)$ in (6.5). As well as the different formulation, however, there is also a difference in the aims of the analysis. Scharfstein and Robins (2002) aim to estimate a definitive non-ignorable model; our interest is in using the standard analysis to assess robustness to small departures from ignorability.

ACKNOWLEDGEMENTS

The authors thank the referees for their helpful comments. The first and third authors acknowledge the support of the EPSRC, with a research studentship and research fellowship respectively.

APPENDIX

In the notation of Section 3.5, the first order approximation to the correlation between transformed variables $A$ and $D$ can be shown to be

$$
\text{Cor}(A, D) = \frac{\delta t}{\sigma_A \sigma_D} \int_0^\infty D(c, \gamma) x_C(c, \gamma) f_C(c, \gamma) dc \int_0^\infty A(t, \theta) B(t, \theta) f_T(t, \theta) dt,
$$

(A.1)
where \( \sigma_A \) and \( \sigma_D \) are the corresponding standard deviations of \( A \) and \( D \). We look at the two integrals in (A.1) separately. First,

\[
\frac{1}{\sigma_D} \left| \int_0^\infty D(c, \gamma) s_C(c, \gamma) f_C(c, \gamma) \, dc \right| = W_1(c) \left[ \int_0^\infty [s_C(c, \gamma)]^2 f_C(c, \gamma) \, dc \right]^{1/2} \leq \left[ \int_0^\infty [s_C(c, \gamma)]^2 f_C(c, \gamma) \, dc \right]^{1/2} \tag{A.2}
\]

where

\[
W_1(c) = \frac{\int_0^\infty D(c, \gamma) s_C(c, \gamma) f_C(c, \gamma) \, dc}{\sigma_D \left[ \int_0^\infty [s_C(c, \gamma)]^2 f_C(c, \gamma) \, dc \right]^{1/2}}
\]

is the absolute value of the correlation between \( D(c, \gamma) \) and \( s_C(c, \gamma) \), and so is less than one. Similarly,

\[
\frac{1}{\sigma_A} \left| \int_0^\infty A(t, \theta) B(t, \theta) f_T(t, \theta) \, dt \right| = W_2(t) \left[ \int_0^\infty B^2(t, \theta) f_T(t, \theta) \, dt \right]^{1/2} \leq 1 \tag{A.3}
\]

where

\[
W_2(t) = \frac{\int_0^\infty A(t, \theta) B(t, \theta) f_T(t, \theta) \, dt}{\sigma_A \left[ \int_0^\infty B^2(t, \theta) f_T(t, \theta) \, dt \right]^{1/2}}
\]

is the absolute value of the correlation between \( A(t, \theta) \) and \( B(t, \theta) \). The inequality (A.3) follows from the fact that \( W_2(t) \) is less than one and from the standardization constraint (2.5). Therefore, substituting (A.2) and (A.3) into (A.1) gives

\[
|\text{Corr}(A, D)| \leq |\delta| \left[ \int_0^\infty [s_C(c, \gamma)]^2 f_C(c, \gamma) \, dc \right]^{1/2} = |\delta|. \tag{A.4}
\]

In particular, \(|\text{Corr}(T, C)| \leq |\delta|\).

REFERENCES


*Received October 30, 2002; first revision September 10, 2003; second revision February 17, 2004; accepted for publication May 6, 2004*