Informative Censoring in Piecewise Exponential Survival Models

N. D. Staplin[1], A. C. Kimber[1], D. Collett[2], and P.J. Roderick[3]

[1]School of Mathematics and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, U.K
[2]Statistics and Clinical Audit, NHS Blood and Transplant, Bristol, U.K
[3]School of Medicine, University of Southampton, Southampton, U.K

Abstract

There are often reasons to suppose that there is dependence between the time to event and time to censoring, or informative censoring, for survival data, particularly when considering medical data. This is because the decision to treat or not is often made according to prognosis, usually with the most ill patients being prioritised. Due to identifiability issues, sensitivity analyses are often used to assess whether non-informative censoring can lead to misleading results. In this paper, a sensitivity analysis method for piecewise exponential survival models is presented. This method assesses the sensitivity of the results of standard survival models to small amounts of dependence between the time to failure and time to censoring variables. It uses the same assumption about the dependence between the time to failure and time to censoring as previous sensitivity analyses for both standard parametric survival models and the Cox model. However, the method presented in this paper allows the use of more flexible models for the marginal distributions whilst remaining computationally simple. A simulation study is used to assess the accuracy of the sensitivity analysis method and identify the situations in which it is suitable to use this method. The study found that the sensitivity analysis performs well in many situations, but not when the data has a high proportion of censoring.

1 Introduction

A feature of survival data is that the end point of interest may not be observed for some individuals, leading to censored observations. In general, standard

methods assume that the time to failure and time to censoring variables are independent and therefore any censoring is non-informative. However there are many cases where this assumption may not be valid. In this paper, the survival of patients on the waiting list for a liver transplant will be considered, where those who are removed for transplantation are censored. Generally, these are the most ill patients on the list so it is unlikely that time to death and time to transplantation are independent. If there is dependence between the time to failure and time to censoring variables, then the censoring is said to be informative (Collett 2003). In particular informative right censoring will be considered.

If there is informative censoring in a data set, but it is assumed that it is non-informative, then the results of the analysis may be biased. The direction of this bias depends on the sign of the correlation between the time to failure and time to censoring variables. With non-informative censoring, those who are censored are representative of the individuals at risk at the time of censoring. If there is a positive (negative) association between the two variables then those who are censored would have a smaller (larger) expected survival time than those who remain at risk. Standard methods would then over(under)estimate the survivor function. The magnitude of the bias will tend to increase as the number of censored observations increases.

It is not possible to use the observed data to determine whether a dataset has informative censoring or the extent of the dependence between the time to failure and time to censoring variables; see Tsiatis (1975). Therefore, an alternative is to use a sensitivity analysis. This allows any parameters that control the dependence between the time to failure and time to censoring variables to be varied over a range of sensible values so the effect that this has on the inferences can be seen. From here it can be assessed whether informative censoring is likely to be an issue in the analysis.

In recent years, various sensitivity analyses have been suggested to assess the effect of informative censoring on the inferences obtained when analysing survival data, including Zhang and Heitjan (2006), Ruan and Gray (2008) and Huang and Zhang (2008). Siannis et al. (2005) introduce a sensitivity analysis for informative censoring in parametric survival models. This approximates the change in the parameter estimates obtained if a small amount of dependence between the time to failure and time to censoring variables is assumed instead of independence. This particular sensitivity analysis appeals because it produces an approximation that is straightforward to apply with results that are easy to interpret. The same assumption about the joint distribution of the time to failure and time to censoring variables is used in Siannis (2011) to develop a sensitivity analysis for the Cox proportional hazards model. This approach is more flexible and can be applied to a wider range of data sets, but is more computationally intensive.

In this paper, we use piecewise exponential models which lead to sensitivity analyses that are more flexible than those for the standard parametric models and computationally simpler than those for the Cox model. We use the same assumption about the dependence between the time to failure and time to cen-

soring as Siannis et al. (2005), but our method may be preferable as it combines the strengths of each of the methods in these papers.

The following section outlines the model used and assumptions made. The equations needed to conduct the sensitivity analysis are derived in Section 3. The method is then applied to data on the time from registration to death on the waiting list for patients registered for a liver transplant in Section 4. This data set is a good exemplar data set as it is found that standard parametric models do not provide a good fit for time to censoring. The details of a simulation study carried out to assess the accuracy of the sensitivity analysis method are given in Section 5. This allows both the strengths and weaknesses of the methods to be identified. A discussion of the method is given in Section 6.

2 Notation and model

The form of the joint distribution of T, the failure time and C, the censoring time, is needed so that the dependence between the two variables can be assessed. However, only $Y = \min(T, C)$ and an indicator function $I(T \leq C)$ are observed. This means that some additional assumptions will be required to identify the joint distribution.

A piecewise parametric model with piecewise constant hazard functions will be used for the marginal distributions of both T and C. This is known as the piecewise exponential model and was introduced in Breslow (1974). This means that the hazard is constant over a given interval, but may vary between intervals.

As intervals have been introduced into the model, a piecewise approach is required to generate the log-likelihood. The piecewise approach requires a time variable corresponding to each interval for each individual which can be obtained using the observation time, y_i, for each individual. Therefore the exposure time for individual i in intervals j is defined as

$$\begin{aligned} y_{ij} &= a_j - a_{j-1} \qquad & j &= 1, \ldots, N_i - 1 \\ y_{ij} &= y_i - a_{N_i - 1} \qquad & j &= N_i, \end{aligned}$$

where a_j is the endpoint of the jth interval. Here N_i denotes the number of the interval in which individual i experiences either failure or the censoring of interest at time y_i. Once having experienced one of these events, individual i has no further exposure in later intevals.

In this framework, there are three possible times that may be observed for each individual in any of the intervals. These are T, the failure time, C_I, the censoring that occurs within an interval, and C_E, the censoring at the end of an interval. The censoring at the end of each interval, which has been introduced by the use of a piecewise model, is assumed to be independent of any censoring that takes place in the intervals.

Two indicator variables are needed, first to distinguish between a failure time and a censored time and then to distinguish between the two different types of

censoring. These indicator variables are

$$I_{ij} = \begin{cases} 1, & \text{if } i\text{th individual fails in } j\text{th interval} \\ 0, & \text{if } i\text{th individual does not fail in } j\text{th interval} \end{cases}$$

and, when $I_{ij} = 0$,

$$Z_{ij} = \begin{cases} 1, & \text{if } i\text{th individual is censored before the end of } j\text{th interval} \\ 0, & \text{if } i\text{th individual is censored at the end of } j\text{th interval.} \end{cases}$$

As we are using a piecewise exponential model, we can take advantage of the lack of memory between the intervals. If we condition on $T > a_{j-1}$, then the survivor function $S_T(t|T > a_{j-1}) = S_T(t - a_{j-1})$ for the jth interval depends only on the parameter associated with that interval, θ_j. The mean of the distribution in the jth interval is given by θ_j^{-1} Let $t_j = t - a_{j-1}$ be the amount of time passed in the jth interval, then the survivor function can be denoted by $S_T(t_j, \theta_j)$. For the ith individual, the survivor function for the jth interval would be $S_T(y_{ij}, \theta_j)$. The density, hazard and integrated hazard functions for T in the jth interval,

$$f_T(t_j, \theta_j) = -\frac{d}{dt} S_T(t_j, \theta_j), \quad h_T(t_j, \theta_j) = -\frac{d}{dt} \log S_T(t_j, \theta_j),$$
$$\text{and} \quad H_T(t_j, \theta_j) = -\log S_T(t_j, \theta_j),$$

also only depend on θ_j. The score and information functions for the density function $f_T(t_j, \theta_j)$ are defined by

$$s_T(t, \theta_j) = \frac{\partial}{\partial \theta_j} \log f_T(t_j, \theta_j) \quad \text{and} \quad i_{\theta_j} = \text{Var}_T\{s_T(T, \theta_j)\}.$$

Similarly, if we condition on $C > a_{j-1}$ then the survivor function $S_C(c|C > a_{j-1})$ for C_I in the jth interval only depends on the nuisance parameter, γ_j. For ease of notation, without ambiguity subscript C will be used for functions relating to C_I. Let $c_j = c - a_{j-1}$ and the survivor function for C_I in the jth interval can be denoted $S_C(c_j, \gamma_j)$. There are the corresponding functions $f_C(c_j, \gamma_j)$, $h_C(c_j, \gamma_j)$, $H_C(c_j, \gamma_j)$, $s_C(c, \gamma_j)$ and i_{γ_j} for C_I.

It is now necessary to make an assumption concerning the conditional distribution of C_I given T, so that we can identify the joint distribution of T and C_I. As in Siannis et al. (2005), Siannis (2004) and Siannis (2011), we assume that the conditional distribution of C_I given T has the same parametric distribution as the marginal distribution of C_I. However, the parameter of the conditional density is allowed to depend on T. Therefore, the conditional density in the jth interval can be written explicitly as

$$f_{C|T}(c_j|t_j, \gamma_j, \delta, \theta_j) = f_C(c_j, \gamma_j + \delta i_{\gamma_j}^{-1/2} B(t_j, \theta_j)),$$

where i_{γ_j} is the information function for C_I. The dependence between T and C_I is defined by δ and $B_T(t, \theta)$. These can be thought of, respectively, as a

correlation coefficient, that quantifies the amount of dependence between the two processes, and a bias function which gives a form to this dependence.

As we will let the parameters vary between the intervals, we will have the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ with θ_j and γ_j being the scalar parameters in the jth of the m intervals in our model.

3 Sensitivity analysis

In this section, a sensitivity analysis is developed that can be applied to piecewise parametric models. The aim is to approximate the difference between the parameter estimates from the dependent model and those from the independent model. As piecewise parametric survival models will be assumed for the marginal distributions of the failure and censoring variables, there will be different parameter values in each of the intervals used. Therefore, let $\hat{\theta}_{\delta j}$ denote the parameter estimate in the jth interval for the model that assumes informative censoring and similarly $\hat{\theta}_{0j}$ is the parameter estimate for the jth interval from the model that assumes non-informative censoring.

The log-likelihood function $\ell_\delta(\boldsymbol{\theta}, \boldsymbol{\gamma})$ when T and C_I are dependent is

$$\ell_\delta(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\Big\{I_{ij}\log K_1(y_{ij}) + Z_{ij}(1-I_{ij})\log K_2(y_{ij}) + (1-I_{ij})(1-Z_{ij})\log K_3(y_{ij})\Big\}, \tag{1}$$

where

$$\begin{aligned} K_1(y_{ij}) &= \int_{y_{ij}}^{\infty} f_{T,C}(y_{ij}, u)du, \\ K_2(y_{ij}) &= \int_{y_{ij}}^{\infty} f_{T,C}(u, y_{ij})du \\ \text{and} \quad K_3(y_{ij}) &= \int_{y_{ij}}^{\infty}\int_{y_{ij}}^{\infty} f_{T,C}(t,c)dt\,dc. \end{aligned} \tag{2}$$

These can be thought of as the likelihood contributions for each of the three types of observations that may occur in each interval. To avoid having integrals in the above contributions that cannot be evaluated analytically, the joint density for T and C_I in the jth interval is written

$$\begin{aligned} f_{T,C}(t_j, c_j) &= f_T(t_j, \theta_j) f_C(c_j, \gamma_j + \delta i_{\gamma_j}^{-1/2} B(t_j, \theta_j)) \\ &\simeq f_T(t_j, \theta_j) f_C(c_j, \gamma_j)[1 + \delta i_{\gamma_j}^{-1/2} s_C(c_j, \gamma_j) B(t_j, \theta_j)]. \end{aligned}$$

So, when the approximations of the contributions in (2) are substituted into

(1), the log-likelihood becomes:

$$\ell_\delta(\boldsymbol{\theta},\boldsymbol{\gamma}) \simeq \ell_0(\boldsymbol{\theta},\boldsymbol{\gamma}) - \delta \sum_{i=1}^{n}\sum_{j=1}^{m} i_{\gamma_j}^{-1/2}\Big\{ I_{ij} B_T(y_{ij},\theta_j)\frac{\partial}{\partial\gamma_j} H_C(y_{ij},\gamma_j) + (1-I_{ij})(1-Z_{ij})\frac{\partial}{\partial\gamma_j} H_C(y_{ij},\gamma_j)\mu(y_{ij},\theta_j) - Z_{ij}(1-I_{ij}) s_C(y_{ij},\gamma_j)\mu(y_{ij},\theta_j)\Big\}, \tag{3}$$

where

$$\mu(y_{ij},\theta_j) = \frac{\int_{y_{ij}}^{\infty} f_T(u,\theta_j)B_T(u,\theta_j)du}{S_T(y_{ij},\theta_j)}$$

and

$$\ell_0(\boldsymbol{\theta},\boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\Big\{ Z_{ij}(1-I_{ij})\log h_C(y_{ij},\gamma_j) - H_C(y_{ij},\gamma_j) + I_{ij}\log h_T(y_{ij},\theta_j) - H_T(y_{ij},\theta_j)\Big\}.$$

Note that $\ell_0(\boldsymbol{\theta},\boldsymbol{\gamma})$ is the log-likelihood in the non-informative censoring model.

A proportional hazards structure is used to simplify some of the terms in (3), so that the hazard functions of T and C_I become

$$h_T(t_j,\theta_j) = e^{\theta_j} h_T^*(t_j) \quad \text{and} \quad h_C(c_j,\gamma_j) = e^{\gamma_j} h_C^*(c_j),$$

where $h_T^*(t_j)$ and $h_C^*(c_j)$ are baseline hazard functions. Consequently, the score and information functions become

$$s_T(t_j,\theta_j) = 1 - H_T(t_j,\theta_j), \quad s_C(c_j,\gamma_j) = 1 - H_C(c_j,\gamma_j)$$
$$\text{and} \quad i_{\theta_j} = i_{\gamma_j} = 1.$$

If the bias function is assumed to be the standardized score function, $B(t_j,\theta_j) = i_{\theta_j}^{-1/2} s_T(t_j,\theta_j)$, as in Siannis et al.(2005), then

$$B(t_j,\theta_j) = 1 - H_T(t_j,\theta_j) \qquad \text{and} \qquad \mu(t_j,\theta_j) = H_T(t_j,\theta_j).$$

An expression for the difference in the parameter estimates for the jth of the m parameters of interest can be obtained using Taylor expansions of the score functions

$$r_0(\hat{\theta}_{0j}) = \frac{\partial}{\partial\theta_j}\ell_0(\boldsymbol{\theta},\boldsymbol{\gamma})\Big|_{\hat{\theta}_{0j}} \quad \text{and} \quad r_\delta(\hat{\theta}_{\delta j}) = \frac{\partial}{\partial\theta_j}\ell_\delta(\boldsymbol{\theta},\boldsymbol{\gamma})\Big|_{\hat{\theta}_{\delta j}}. \tag{4}$$

These are the score functions for the jth interval under the assumption of non-informative and informative censoring respectively. The score functions given in (4) are expanded about θ_j and set equal to zero at $\hat{\theta}_j$ to give

$$r_0(\hat{\theta}_{0j}) \simeq r_0(\theta_j) - (\hat{\theta}_{0j} - \theta_j) i_j(\boldsymbol{\theta}) = 0$$
$$r_\delta(\hat{\theta}_{\delta j}) \simeq r_\delta(\theta_j) - (\hat{\theta}_{\delta j} - \theta_j) i_j(\boldsymbol{\theta}) = 0 \tag{5}$$

where

$$i_j(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \theta_j^2} \ell_0(\boldsymbol{\theta}, \boldsymbol{\gamma}).$$

Rearranging the two equations in (5) gives

$$(\hat{\theta}_{\delta j} - \hat{\theta}_{0j}) i_j(\theta) \simeq r_\delta(\theta_j) - r_0(\theta_j). \tag{6}$$

If the values from the proportional hazards structure are substituted into (3) and the resulting form of the log-likelihood is then used in (6), the approximation to the difference in the parameter estimates for the jth of the m parameters becomes:

$$\hat{\theta}_{\delta j} - \hat{\theta}_{0j} \simeq \delta i_j(\boldsymbol{\theta})^{-1} \sum_{i \in R_j} H_T(y_{ij}, \theta_j) \Big[H_C(y_{ij}, \gamma_j) - (1 - I_{ij}) Z_{ij} \Big],$$

where R_j is the risk set in the jth interval.

Covariates also need to be included in the model, so that the hazard function of the ith individual in the jth interval is defined to be

$$h_{ij} = \exp(\alpha_j + \sum_{k=1}^{p} \beta_k x_{ik}).$$

See Friedman (1982) for a full description of such models. This model is equivalent to splitting $\boldsymbol{\theta}_j' \mathbf{x}$ into an intercept for each interval, α_j, and a component for the p covariates included in the model, given by $\sum_{k=1}^{p} \beta_k x_{ik}$, which remains constant over the intervals. Here, $\boldsymbol{\theta}_j$ is the vector of parameters for the jth interval, the first element corresponds to the intercept term in the jth interval and the other elements to the covariates that remain constant for all j. Similarly, we can define $\boldsymbol{\gamma}_j$ for the model for time to censoring including covariates.

In order to incorporate covariates in the sensitivity analysis for piecewise parametric models, the scalar parameters θ_j and γ_j in the jth interval are replaced by $w_j(\mathbf{x}) = \boldsymbol{\theta}_j' \mathbf{x}$ and $z_j(\mathbf{x}) = \boldsymbol{\gamma}_j' \mathbf{x}$ respectively. This approach is similar to that used in Siannis (2004) who perform a sensitivity analysis on the function $w(\mathbf{x}) = \boldsymbol{\theta}' \mathbf{x}$ rather than $\boldsymbol{\theta}$ when incorporating covariates. This means that the equation that can be used to carry out the sensitivity analysis is

$$\hat{w}_{\delta j}(\mathbf{x}) - \hat{w}_{0j}(\mathbf{x}) \simeq \delta \frac{\sum_{i \in R_j} H_T(y_{ij}, w_j(\mathbf{x})) [H_C(y_{ij}, z_j(\mathbf{x})) - (1 - I_{ij}) Z_{ij}]}{\sum_{i \in R_j} H_T(y_{ij}, w_j(\mathbf{x}))}, \tag{7}$$

where the change in the parameter estimates is estimated for a covariate vector of interest, $\mathbf{x}$, when a suitable value for the corresponding nuisance function, $z_j(\mathbf{x})$, has been substituted into the equation.

When $w_j(\mathbf{x}) = \boldsymbol{\theta}_j' \mathbf{x}$ and $z_j(\mathbf{x}) = \boldsymbol{\gamma}_j' \mathbf{x}$ the hazard and integrated hazard functions for T and C with piecewise exponential marginal distributions can be expressed as:

$$h_T(t_j, \boldsymbol{\theta}_j, \mathbf{x}) = e^{w_j(\mathbf{x})} \qquad h_C(c_j, \boldsymbol{\gamma}_j, \mathbf{x}) = e^{z_j(\mathbf{x})}$$
$$H_T(t_j, \boldsymbol{\theta}_j, \mathbf{x}) = e^{w_j(\mathbf{x})} t_j \qquad H_C(c_j, \boldsymbol{\gamma}_j, \mathbf{x}) = e^{z_j(\mathbf{x})} c_j \tag{8}$$

The approximation for $\hat{w}_{\delta j}(\mathbf{x}) - \hat{w}_{0j}(\mathbf{x})$ when conducting a sensitivity analysis on $w_j(\mathbf{x})$ is obtained by substituting the functions from (8) into (7). This then gives:

$$\hat{w}_{\delta j}(\mathbf{x}) - \hat{w}_{0j}(\mathbf{x}) \simeq \delta \frac{\sum_{i \in R_j} \{e^{\hat{z}_{0j}(\mathbf{x})} y_{ij}^2 - y_{ij}(1 - I_{ij}) Z_{ij}\}}{\sum_{i \in R_j} y_{ij}}. \tag{9}$$

The value of $z_{0j}(\mathbf{x})$ used in (9) is $\hat{z}_{0j}(\mathbf{x})$, the estimated linear predictor for time to censoring under the assumption of non-informative censoring. The sensitivity analysis equation in (9) may be applied to each interval individually, which allows the particular intervals with the largest estimated changes in parameter estimates to be identified. It should be noted that the sensitivity analysis equation in (9) only applies to the given covariate vector, $\mathbf{x}$.

4 Liver registration data

Liver transplantation is the only treatment option for patients whose livers are failing. In the UK, those judged suitable for a transplant are registered with NHS Blood and Transplant, but a national shortage of organ donors means that some patients die waiting for their transplant. Analyses of factors affecting time to death informs policy for the selection of patients for transplantation.

Data provided by NHS Blood and Transplant contain information on 4594 adult patients who were first registered for an elective liver transplant between 1 January 2000 and 31 December 2008 in the UK. Patients are followed until they are removed from the waiting list, whether that be due to death, receiving a liver transplant or other reasons. Patients that were still on the list at the time of compiling the data set were censored, but we can treat this censoring as independent, along with any end of interval censoring. Removal from the list for transplantation is a form of potentially informative censoring. This is because, generally, the patients that have the worst prognosis and are therefore closer to death are the ones that receive transplants. Also, some patients were removed from the list for reasons other than death or transplantation. A patient removed due to deteriorating condition was assumed to have died on the day of removal. Otherwise the patient was non-informatively censored at the time of removal.

Of the variables whose values are known at registration, patient age, ethnicity, primary liver disease(PLD) category, and UKELD (UK model for End-stage Liver Disease) score at registration were found to significantly affect the hazard of death following registration. The UKELD score is a measure of disease

Table 1: Patient ethnicity levels

Level	Patient ethnicity
1	White (n=4077)
2	Asian or Asian-British (n=332)
3	Black or Black-British (n=103)
4	Chinese/Oriental (n=34)
7	Other (n=48)

Table 2: Primary liver disease groupings that are used in the Liver Registration data set. The numbers of patients included in each group are also given.

Level	Primary liver disease group
1	Primary biliary cirrhosis (PBC) (n=580)
2	Primary sclerosing cholangitis (PSC) (n=434)
3	Alcoholic liver disease (ALD) (n=1142)
4	Auto-immune + cryptogenic disease (AID) (n=523)
5	Hepatitis C cirrhosis (HCV) (n=686)
6	Hepatitis B cirrhosis (HBV) (n=162)
7	Cancers (n=208)
8	Metabolic liver disease (n=196)
9	Other liver diseases (n=489)
10	Acute hepatic failure (n=130)

severity based on several clinical measurements, the details of which are given in Barber et al. (2011). The groupings used for patient ethnicity and PLD are given in Tables 1 and 2 respectively. These variables were identified using parametric survival models that assume non-informative censoring.

Of all the patients in the dataset, 2650 had full information for the variables that were found to be significant. This is because for some of the earlier years in the time period considered, less patient information had to be supplied to the UK Transplant Registry (UKTR). The patients with missing data are not considerably different from the rest when considering the covariates for which information was available, so they will be excluded from the analysis. There were 423 events in this reduced data set, 304 were observed deaths and 119 were removals due to deteriorating condition. There were 1899 potentially informatively censored observations where patients had been removed from the list for transplantation. The remaining 328 observations were patients that were censored for non-informative reasons.

4.1 Application of sensitivity analysis

It is assumed that the lifetime and censoring variables each have piecewise exponential marginal distributions. The starting values for the cut points are the quantiles of the distribution of the time to death variable. The models with the cut points that give the largest value of the likelihood were found for 3 and 4

intervals. The log-cumulative hazard plots were then examined to check if the assumed model is appropriate. The 3-interval model was found to be appropriate and the 4-interval model did not seem to give any improvement. Therefore in the interest of parsimony, we used the 3-interval model with cut points at 40 and 165 days.

To determine whether these models gave a significantly better fit than the corresponding standard Weibull models, the differences in $-2\log\hat{L}$ for the models were found. If the true hazard is Weibull, then the difference in $-2\log\hat{L}$ for the Weibull model and the piecewise exponential model should be approximately χ^2_{m-2}, where m is the number of intervals in the piecewise exponential distribution. The piecewise exponential model was significantly better than the Weibull for time to censoring ($p < 0.0001$) but not for time to death (p=0.85). As the same form of model must be used for both time to death and time to censoring when applying the sensitivity analysis, then the use of piecewise exponential models for the marginal distributions of the failure and censoring variables is justified for the Liver Registration data.

Expression (9) can be used to apply the sensitivity analysis to this data set. However, (9) gives the change in parameter estimates for a specified $\mathbf{x}$. But $\mathbf{x}$ varies across patients, so the difference between $\hat{w}_{\delta j}(\mathbf{x})$ and $\hat{w}_{0j}(\mathbf{x})$ for all the individuals in the jth interval needs to be assessed. The simplest way to do this is to plot the estimated value of $\hat{w}_{\delta j}(\mathbf{x})-\hat{w}_{0j}(\mathbf{x})$ against the entire range of values that $\hat{z}_j(\mathbf{x})$, the estimated linear predictor for $\mathbf{x}$ for time to censoring in the jth interval, takes across all the individuals in the jth interval. The distributions of $\hat{z}_j(\mathbf{x})$ for the Liver Registration data are shown by the boxplots in Figure 1. We see that the median value of $\hat{z}_j(\mathbf{x})$ decreases across the intervals, which shows that the hazard of censoring is generally smaller in the later intervals. We also see that the majority of patients have values of $\hat{z}_j(\mathbf{x})$ that fall in the middle of the observed range for each interval, with only a small number at either of the extremes.

The results of the sensitivity analysis are shown in Figure 2. This figure shows the plots with two different values of δ: 0.3, the largest value of δ used for the sensitivity analysis in Siannis et al. (2005), and 0.2, a more conservative value. It can be seen from Figure 2 that the second and third intervals have larger estimated values of $\hat{w}_{\delta j}(\mathbf{x}) - \hat{w}_{0j}(\mathbf{x})$ than the first interval. The largest values of $\hat{w}_{\delta j}(\mathbf{x}) - \hat{w}_{0j}(\mathbf{x})$ are observed for the patients with the largest values of $\hat{z}_j(\mathbf{x})$ or the highest hazards of censoring. However if we consider the distributions of $\hat{z}_j(\mathbf{x})$ shown in Figure 1, then we can see that only a small number of individuals will have these large changes in $\hat{w}_{\delta j}(\mathbf{x}) - \hat{w}_{0j}(\mathbf{x})$. This means the effect of informative censoring is small for the majority of patients in the Liver Registration data. However, as some individuals have a large estimated change in the linear predictors, then any inferences may be misleading if non-informative censoring was assumed, and there is even a moderate amount of dependence between the time to death and time to censoring variables.

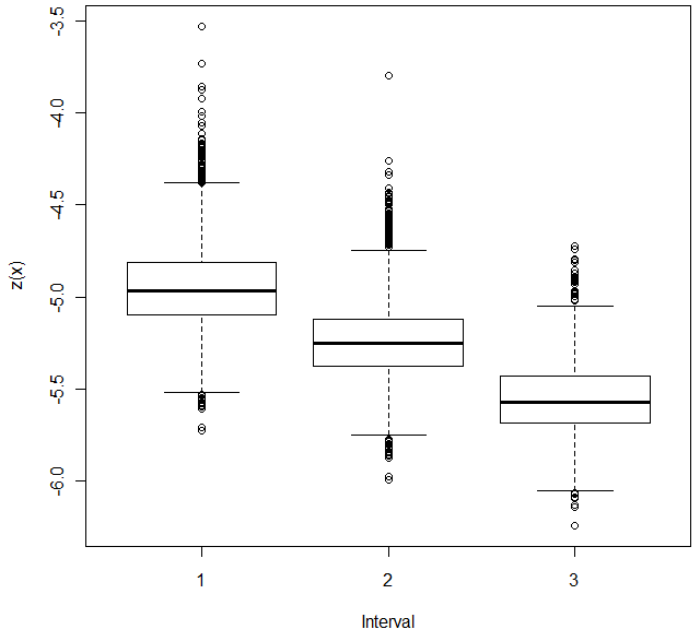

Figure 1: Boxplots showing the distribution of $\hat{z}_j(\mathbf{x})$ in each of the intervals for the Liver Registration data.

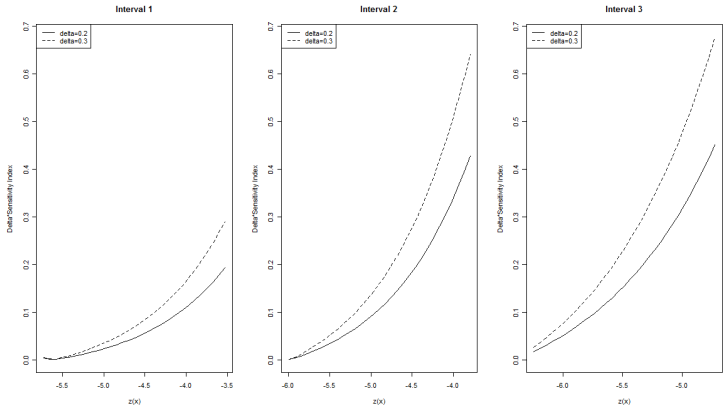

Figure 2: Plot of sensitivity analysis expression in (9) for observed values of $\hat{z}_j(\mathbf{x})$ for the Liver Registration data in each of the three intervals with $\delta = 0.2, 0.3$.

4.2 Comparing results to those obtained by fitting dependent model

To assess the validity of the method in Section 4.1 for this dataset, the model that accounts for dependence is fitted to the dataset as well. The simplification used to obtain (3) was necessary to get a closed form of the equations for the sensitivity analysis, but is not necessary when fitting the dependent model. So, now the likelihood for the dependent model will be derived. The joint density of T and C_I in the jth interval can be written as

$$f_{T,C}(t_j, c_j) = f_T(t_j, \theta_j) f_{C|T}(c_j | t_j, \gamma_j, \delta, \theta_j) \tag{10}$$

As in Section 3 it is assumed that

$$f_{C|T}(c_j | t_j, \gamma_j, \delta, \boldsymbol{\theta}_j) = f_C(c_j, \gamma_j + \delta i_{\gamma_j}^{-1/2} B(t_j, \boldsymbol{\theta}_j, \mathbf{x})), \tag{11}$$

with $i_{\gamma_j} = 1$ and $B(t_j, \boldsymbol{\theta}_j, \mathbf{x}) = 1 - e^{\boldsymbol{\theta}_j' \mathbf{x}} t$ under the proportional hazards structure.

In addition, it is assumed that both T and C_I have piecewise exponential marginal models, so in the jth interval

$$f_T(t_j, \boldsymbol{\theta}_j, \mathbf{x}) = e^{\boldsymbol{\theta}_j' \mathbf{x}} \exp(-e^{\boldsymbol{\theta}_j' \mathbf{x}} t_j) \quad \text{and} \quad f_C(c_j, \eta) = e^{\eta} \exp(-e^{\eta} c_j), \tag{12}$$

where the linear combination $\boldsymbol{\gamma}_j^T \mathbf{x}$ is replaced by a scalar parameter η, as it is just a nuisance parameter.

If (10) and (11) are combined, and then the exponential forms in (12) are substituted into the resulting equation, the joint distribution can be written as

$$f_{T,C}(t, c) = e^{\boldsymbol{\theta}_j' \mathbf{x}} e^{-\exp\{\boldsymbol{\theta}_j' \mathbf{x}\} t_j} e^{\eta + \delta(1 - \exp\{\boldsymbol{\theta}_j' \mathbf{x}\} t_j)} e^{-\exp(\eta + \delta(1 - \exp\{\boldsymbol{\theta}_j' \mathbf{x}\} t_j)) c_j}. \tag{13}$$

The parameter estimates for the full model will be obtained by finding the maximum likelihood estimates of the likelihood in (1) but with contributions

$$K_1(y_{ij}) = e^{\boldsymbol{\theta}_j' \mathbf{x}_{ij}} e^{-\exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} y_{ij}} e^{-\exp\left(\eta + \delta\left(1 - \exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} y_{ij}\right)\right) y_{ij}}$$

$$K_2(y_{ij}) = \int_{y_{ij}}^{\infty} e^{\boldsymbol{\theta}_j' \mathbf{x}_{ij}} e^{-\exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} u} e^{\eta + \delta\left(1 - \exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} u\right)} \times e^{-\exp\left(\eta + \delta\left(1 - \exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} u\right)\right) y_{ij}} du$$

and

$$K_3(y_{ij}) = \int_{y_{ij}}^{\infty} e^{\boldsymbol{\theta}_j' \mathbf{x}_{ij}} e^{-\exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} u} e^{-\exp\left(\eta + \delta\left(1 - \exp\{\boldsymbol{\theta}_j' \mathbf{x}_{ij}\} u\right)\right) y_{ij}} du.$$

These were obtained by substituting the form of the joint distribution given in (13) into (2).

The model fitted gives $\delta = 0.1863$ with 95% confidence interval (-0.0394, 0.4120), this shows that even after making identifying assumptions there is

little information about the value of δ as the 95% confidence interval for the estimate is wide. For the Liver Registration data set, it is not even possible to infer that δ is significantly different from zero. This highlights the need for a sensitivity analysis that allows values of δ to be varied in a sensible range and indicates whether inferences based on the parameter of interest are likely to be affected by the assumption of differing amounts of dependence between the time to failure and time to censoring variables.

The sensitivity analysis described in Section 4.1 is applied with $\delta = 0.1863$ so that estimated change in $\hat{w}_\delta(\mathbf{x})$ and $\hat{w}_0(\mathbf{x})$ given by the sensitivity analysis can be directly compared to the observed change in the estimated linear predictors. This will us allow us to assess how accurate the sensitivity analysis is at estimating the change in $\hat{w}_\delta(\mathbf{x})$ and $\hat{w}_0(\mathbf{x})$. When the sensitivity analysis was applied to the Liver Registration data set using $\delta = 0.1863$, the largest estimated change in the estimated linear predictors was 0.42 for the final interval. When calculating the difference between $\hat{w}_\delta(\mathbf{x})$ and $\hat{w}_0(\mathbf{x})$ for each of the individuals in the data set, the largest difference observed over all intervals was 0.25.

This result shows that for the Liver Registration data, the sensitivity analysis overestimates the change in the estimated linear predictors. However, only a small number of the patients in the data will have a discrepancy that is this large. We already know that the sensitivity analysis gives the largest changes in $\hat{w}_\delta(\mathbf{x})$ and $\hat{w}_0(\mathbf{x})$ for the patients with the largest values of $\hat{z}_j(\mathbf{x})$. From Figure 1, we know that only a small number of patients have values of $\hat{z}_j(\mathbf{x})$ that are that large. So, for the majority of individuals in the Liver Registration data the discrepancy between the results of the sensitivity analysis and the change in $\hat{w}_\delta(\mathbf{x})$ and $\hat{w}_0(\mathbf{x})$ using the results of the dependent model is small.

To see if the sensitivity analysis always overestimates the change in $\hat{w}_\delta(\mathbf{x})$ and $\hat{w}_0(\mathbf{x})$, a simulation study is carried out. The details of this simulation study are given in the next section.

5 Simulations

The aim of the simulation study is to see how much the sensitivity analysis either over or under estimates the actual change in the estimated linear predictors at a given level of δ. A combination of different values of the parameters θ, γ and δ will be used. This will give data sets that have differing amounts of censoring, from small to large. The data are simulated from a 2-interval piecewise exponential distribution with no other covariates. An arbitrary cut-point is chosen for each situation to give approximately equal numbers of events in the two intervals. For each different combination of θ, γ and δ, 500 replicates are simulated. In all the simulations, it is assumed that $n = 2000$. Then, the dependent model is fitted and the sensitivity analysis applied to each simulated data set so that the results can be compared.

When fitting the dependent model, the value of δ will be fixed. This is because there is very little information about δ in the data, even after identifying

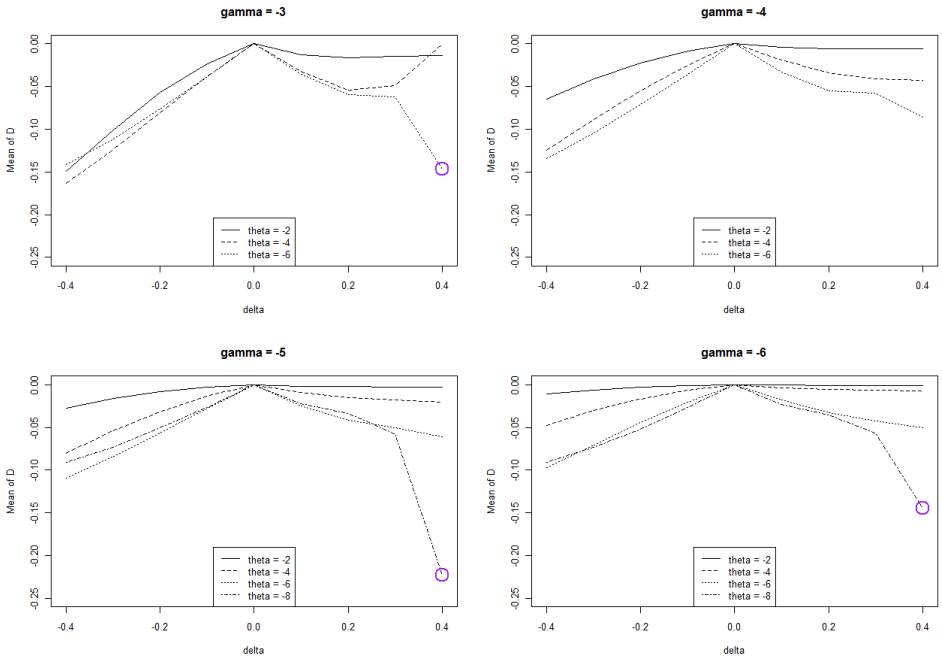

Figure 3: Effect of different combinations of the parameters δ, θ and γ on mean of $\mathbf{D}$ observed. The plots show the effect of θ on the mean observed as δ increases, for each different level of γ

assumptions have been made, so there is much variation in the values of δ obtained. This would make it difficult to make meaningful comparisons between different parameter combinations. The amount of dependence assumed in the sensitivity analysis was the fixed value of δ used when fitting the dependent model.

For each replication the parameter estimates from the dependent model, $\hat{\boldsymbol{\theta}}_{\delta}^{(d)}$, were found along with the parameter estimates approximated by the sensitivity analysis, $\hat{\boldsymbol{\theta}}_{\delta}^{(s)}$. The value $\mathbf{D} = (\hat{\boldsymbol{\theta}}_{\delta}^{(d)} - \hat{\boldsymbol{\theta}}_0) - (\hat{\boldsymbol{\theta}}_{\delta}^{(s)} - \hat{\boldsymbol{\theta}}_0)$ is of interest. The element of $\mathbf{D}$ with the largest magnitude is found as this corresponds to the largest discrepancy between the results of the dependent model and the results of the sensitivity analysis. If this term is negative, then the sensitivity analysis overestimates the change in the parameter estimates. The mean values of the largest element of $\mathbf{D}$ were calculated along with a 95% confidence interval for the mean. These results are summarised graphically in Figure 3. The plots in Figure 3 show the effect of θ on the mean observed as δ increases, at each different level of γ.

The majority of the means observed in Figure 3 are negative, which means that generally the sensitivity analysis overestimates the change in the parameter

estimates. From the plots in Figure 3, it can be seen that generally we observe the larger means when δ is greater than 0.3, and γ and θ are similar in size or $\gamma > \theta$. The greater the difference between γ and θ, the bigger the mean difference we observe. As the size of γ relative to θ increases, the hazard rate of censoring is also increasing relative to the hazard rate for failure. So the simulated data sets would generally contain an increasing proportion of censored observations. Therefore we observe the largest changes in the mean of $\mathbf{D}$ when there is a relatively large proportion of censored observations in the data set. Also as the magnitude of δ increases, the size of the mean also increases, especially in the situations with relatively large amounts of censoring.

To establish the effects of the individual parameters on the simulation results, an analysis of variance model that included all the main effects and interactions between θ, γ and δ was fitted. The ANOVA shows that there is a significant 3 factor interaction between δ, θ and γ ($p < 0.0001$). As θ increases the mean observed generally decreases. There is a greater rate of decrease for larger values of γ when δ is negative. This is because as θ increases there is a greater decrease in the proportion of censored observations in the simulated data sets with larger values of γ. Conversely, for positive δ, the decrease in the proportion of censored observations as θ increases is greatest for the simulated data sets with smaller values of γ, so these values have the greatest rate of decrease in the mean. The rate of decrease in the mean is also affected by the magnitude of δ. For larger magnitudes of δ, the rate of decrease in the mean as θ increases is larger. This makes intuitive sense as the effect of censored observations on the results of the sensitivity analysis increases as the magnitude of δ increases.

There are a handful of situations that are found to have means that are significantly different from most of the other means. These are when $\theta = -6$ and $\gamma = -3$, $\theta = -8$ and $\gamma = -5$ and $\theta = -8$ and $\gamma = -6$, all for $\delta = 0.4$. These situations are highlighted in Figure 3 by the circles on some of the plots. It is easy to see that the means in these cases are larger than the other means.

Further investigations revealed that in these cases, some of the data sets had large outlying values that caused a large increase in the value of the sensitivity index, U. This meant that the sensitivity analysis performed particularly badly for these data sets, resulting in an increased mean for $\mathbf{D}$. This tells us that the accuracy of the sensitivity analysis is affected by the size of the observations included in each interval. This was observed in section 4.1 as the widest intervals had the largest estimated changes in the parameter estimates. So the accuracy of the sensitivity analysis for piecewise exponential models can be improved by dividing the time into a larger number of small intervals.

The results of the simulation study carried out in this section suggest that the sensitivity analysis is not a good approximation of the change in parameter estimates when there is heavy censoring and δ becomes large. This could help to explain why the sensitivity analysis overestimated the actual change in parameter estimates for the Liver Registration dataset. In this dataset there is heavy censoring, with 71.7% of patients having a potentially informatively censored time and a further 12.4% having a non-informatively censored time.

However, even though some situations have been identified where the sen-

sitivity analysis does not give a good approximation to the dependent model, the simulation study in this section shows that there are many situations when the sensitivity analysis does provide a reasonable approximation to the dependent model. This means that while the sensitivity analysis in Section 3 was not as accurate as we would have hoped for the Liver Registration data, it is still suitable for application in other situations.

6 Discussion

Cox proportional hazards models are often used when modelling times to events in medical data. However, if there are potentially informatively censored observations in the data, the sensitivity analyses used to assess the effect of these observations on the results of the standard Cox model can be more computationally intensive than those that use parametric models. The method outlined in this paper uses a model that can be viewed as a "simple computational approximation to the Cox nonparametric model" (Aitken et al. 1983). In fact, as stated in Aitken et al. (1983), the piecewise exponential model with intervals defined by the observed events is equivalent to the Cox model. Therefore, the piecewise exponential model is as flexible as the Cox model but has the added advantage that the form of the baseline hazard is known which makes sensitivity analyses easier to apply.

The sensitivity analysis presented in this paper only incorporates covariates using a linear predictor. This means the effect of informative censoring on the individual parameter estimates cannot be estimated using the method derived in this paper. However, one objective of the analysis of medical data is to be able to inform patients about their expected survival time. As it is the linear predictor that drives the survival process, then from a patient perspective our analysis considers the right quantity.

There is no objective method that identifies when the changes in the estimated linear predictors given by the sensitivity analysis have a significant effect on inferences about the effect of covariates on the hazard of death. Therefore it may be more useful to look at the effect on another value that uses the linear predictor for time to death rather than the linear predictor itself. For example, in the liver transplantation setting, the effect on the median survival time or the estimated survival function might be considered, as these could potentially be used to inform patients of their likely survival time on the waiting list without a transplant based on their disease severity etc.

Finally, if the sensitivity analysis does suggest that the assumption of informative censoring has an effect on inferences about the effect of covariates on the hazard of the event of interest then inverse probability of censoring weighted (IPCW) methods can be easily applied in practice. IPCW estimators (Robins and Rotnitzky 1992, Robins 1993, Robins and Finkelstein 2000) do not require the amount of dependence between time to event and time to censoring to be specified, which means fewer untestable assumptions are required to implement these methods.

Acknowledgements

The authors would like to thank NHS Blood and Transplant who funded the research studentship of Natalie Staplin.

References

[1] M. Aitken, N. Laird, and B. Francis. A reanalysis of the Stanford heart transplant data. *Journal of the American Statistical Association*, 78:264–274, 1983.

[2] K. Barber, S. Madden, J. Allen, D. Collett, J. Neuberger, and A. E. Gimson. Elective liver transplant list mortality; Development of a United Kingdom end-stage liver disease (UKELD) score. *Transplantation*, 92:469–476, 2011.

[3] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974.

[4] D. Collett. *Modelling Survival Data in Medical Research*. Chapman & Hall/CRC, second edition, 2003.

[5] M. Friedman. Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, 10:101–113, 1982.

[6] F. Hu and J. V. Zidek. The weighted likelihood. *The Canadian Journal of Statistics*, 30:347–371, 2002.

[7] X. Huang and N. Zhang. Regression survival analysis with an assumed copula for dependent censoring: A sensitivity analysis approach. *Biometrics*, 64:1090–1099, 2008.

[8] J. M. Robins. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pages 24–33. Alexandria, VA: American Statistical Association, 1993a.

[9] J. M. Robins and D. H. Finkelstein. Correcting for non-compliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56:779–788, 2000.

[10] J. M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In N. Jewell and K. Dietz, editors, *AIDS Epidemiology: Methodological Issues*, pages 297–331. Boston: Birkhauser, 1992.

[11] P. K. Ruan and R. J. Gray. Sensitivity analysis of progression-free survival with dependent withdrawal. *Statistics in Medicine*, 27:1180–1198, 2008.

[12] D. O. Scharfstein and J. M. Robins. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*, 89:617–634, 2002.

[13] F. Siannis. Applications of a parametric model for informative censoring. *Biometrics*, 60:704–714, 2004.

[14] F. Siannis. Sensitivity analysis for multiple right censoring processes: Investigating mortality in psoriatic arthritis. *Statistics in Medicine*, 30:356–367, 2011.

[15] F. Siannis, J. Copas, and G. Lu. Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6:77–91, 2005.

[16] A. A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72:20–22, 1975.

[17] J. Zhang and D. F. Heitjan. A simple local sensitivity analysis tool for nonignorable coarsening: Application to dependent censoring. *Biometrics*, 62:1260–1268, 2006.