Brazilian Journal of Probability and Statistics (2001), 15, pp. 107–119.

STRATIFICATION POINTS FOR THE PROPORTIONAL HAZARDS MODEL

Enrico A. Colosimo
Departmento de Estatística – ICEx, Universidade Federal de Minas Gerais, C. P. 702, 30161-970 – Belo Horizonte, MG, Brazil. Email: enricoc@est.ufmg.br

Summary
An important generalization of the proportional hazards model is the stratified model arising when the data for some specified reason are divided into strata. The stratified model is used in situations where an important covariate or covariates do not satisfy the proportional hazards assumption. This paper presents a simulation study showing the impact on the maximum partial likelihood estimators when the strata are misspecified or an unstratified analysis is performed in a nonproportional hazards situation. The results of this simulation show that in general there is a large bias in the estimate of the regression parameter for the covariate associated with the nonproportionality when stratification points are misspecified in the stratified analysis. Also the estimate of this parameter in the unstratified analysis has a large bias when the stratified analysis is the correct procedure. These results suggest the necessity of guidelines for the division of the data into strata when there is an indication in favor of the stratified analysis. A criterion is proposed as guideline for such division.

Key words: Cox model, model misspecification, stratification criterion, stratified model.

1 Introduction

In the past years, special attention has been given to the proportional hazards model proposed by Cox (1972). This model provides a flexible method for exploring the association of covariates with failure rates and for studying the effect of a covariate of interest, such as treatment, while adjusting for other covariates. It also allows for time-dependent covariates.

The most popular form of the proportional hazards model, for covariates not dependent on time, uses the exponential form for the relative hazard, so that the hazard function is given by

$$\lambda(t) = \lambda_0(t) exp(\beta^t Z) \tag{1.1}$$

where $\lambda_0(t)$, the baseline hazard function, is an unknown non-negative function of time, Z is the covariate vector and β is a p-vector of parameters to be estimated.

Estimation of β is based on the partial (Cox, 1975) or marginal (Kalbfleisch and Prentice, 1973) loglikelihood which in the absence of ties is written for the model (1.1) as

$$\sum_{i=1}^{n} \left(\delta_i \ (\beta^t Z_i) - log(\sum_{j \in R_i} exp(\beta^t Z_j)) \right) \tag{1.2}$$

where $R_i = \{k | t_k \geq t_i\}$ is the risk set at time t_i and δ_i is the failure indicator.

A basic assumption of the model (1.1) is the proportional hazards property. That is, the ratio $\lambda^i(t)/\lambda^j(t)$ does not depend on t for the i^{th} and j^{th} $(i \neq j)$ individuals at risk at time t. In other words, the baseline hazard function $\lambda_0(t)$ and the parametric portion $exp(\beta^t Z)$ of the model act in a multiplicative fashion on the hazard. Model (1.1) also assumes an exponential form for the relative hazard and that no important covariates are excluded from this parametric portion of the model.

Efforts have been made to develop useful goodness-of-fit techniques. These goodness-of-fit techniques include graphical techniques (Kay, 1977; Schoenfeld, 1982; Arjas, 1988; Barlow and Prentice, 1988; Grambsch and Therneau, 1994; Hess, 1995) and goodness-of-fit tests (Schoenfeld, 1980; Andersen, 1982; Wei, 1984; Moreau, O'Quigley and Mesbah, 1985; O'Quigley and Pessione, 1989; Horowitz and Neumann, 1992; Klein and Moescheberger, 1997). Almost all of these techniques are designed to check the proportional hazards assumption.

When the proportional hazards model is inappropriate for a particular survival data set, a reasonable alternative is to use the stratified proportional hazards model. The main use of the stratified model occurs when the proportional hazards assumption is violated overall but holds within strata. Also in other situations of modeling assumption violations a stratified model might be useful. For instance, when the parametric portion of the model assumes different forms in different time periods.

A stratified analysis consists of splitting the data into k strata according to an indication of modeling assumption violation. The proportional hazards model (1.1) is then expressed for the i^{th} stratum as

$$\lambda_i(t) = \lambda_{0_i}(t) exp(\beta^t Z)$$

for $i = 1, \ldots, k$. The baseline hazards functions $\lambda_{0_1}, \ldots, \lambda_{0_k}$, are allowed to be arbitrary and are completely unrelated.

Stratification does not provide any difficulty in the estimation of β. A partial loglikelihood (1.2) is constructed for each stratum and the estimation of β is based on the sum of these partial loglikelihoods (Kalbfleisch and Prentice, 1980). A simple way to split the survival time data set into strata is to consider the different levels of a categorical covariate. Each level defines a stratum. In this way this covariate is naturally excluded from the parametric form of the model. However, continuous covariates provide an extra difficulty in the stratified analysis.

Stratifying a model according to a continuous covariate implies possible inclusion of a parameter term for this covariate. Otherwise information would be lost since the observations within each stratum have different values for this covariate. This is not an issue when the stratification is done according to a categorical covariate. Also, and more important, there are no guidelines in the literature for the division into strata.

In this paper the effect of stratum misspecification on the estimates of the regressor parameters is studied, and a criteria is proposed as guideline for the division of the survival data set into strata. In Section 2, the impact on the maximum partial likelihood estimators is studied if strata are misspecified or an unstratified analysis is done when a stratified analysis is the correct procedure. The criterion presented in Section 3 is based on the expression of the likelihood ratio test when the proportional hazards model is approximated by a fully parametric model. The criterion performance is evaluated in Section 4 and applied to the well known Stanford heart transplant data in Section 5. Some remarks are made in Section 6 about using the proposed criterion in situations different from those examined in the previous sections.

2 Bias study

Consider a survival time study including two covariates, say Z_1 and Z_2. Suppose that Z_2 is a continuous covariate with support on the interval $[0,1]$ such that

$$\lambda(t) = \begin{cases} exp(\beta_1 Z_1 + \beta_2 Z_2) & \text{if } 0 \leq Z_2 \leq c \\ t^{\rho-1} exp(\beta_1 Z_1 + \beta_2 Z_2) & \text{if } c < Z_2 \leq 1 \end{cases} \tag{2.1}$$

for $\rho > 0$ and $0 < c < 1$. That is, the survival time T has an exponential distribution with parameter $exp(\beta_1 Z_1 + \beta_2 Z_2)$ if $0 \leq Z_2 \leq c$ and a Weibull distribution with parameters ρ and $[exp(\beta_1 Z_1 + \beta_2 Z_2)/\rho]^{1/\rho}$ if $c < Z_2 \leq 1$.

A question of interest is related to the properties of the maximum partial likelihood estimators when the fitted model is misspecified. In particular, the effect of fitting an unstratified model or a stratified model with misspecified strata when the model (2.1) is the correct one.

Struthers and Kalbfleisch (1986) showed for a general situation of model misspecification that $\widehat{\beta}$ for the proportional hazards model, converges in probability to a well-defined constant vector β^*. The expression for β^* is difficult to derive and takes a very complicated form in the situation of interest in this study.

This study is developed via a Monte Carlo simulation. Sample of sizes 40 and 80 and two values of ρ are considered. Using $\rho = 2$ the Weibull portion of the model (2.1) has a distribution with increasing hazard function and for $\rho = 0.5$ the hazard is decreasing. For each generated sample, stratified proportional hazards models are fitted using different stratification points. The stratification points are points in the support of Z_2 and each one divides the survival data set into two strata. Different correlation structures are used for the covariates Z_1 and Z_2

Independent random samples of T under the model (2.1) and U, a censoring time with a uniform distribution on $(0, \theta)$, are generated and the lifetime $min(T, U)$ and failure indicator $\delta = I_{\{T<U\}}$ are recorded. A censoring fraction of 25% is achieved in this study by controlling the value of the parameter θ. The covariate Z_2 has a uniform distribution on the interval $(0, 1)$.

The way the data are generated, the stratification point $Z_2 = c$ is the correct value to split the data into two strata. Table 1 illustrates some consequences of the wrong choice of a stratification point when $c = 0.5$, Z_1 has a Bernoulli distribution with parameter 0.5, and Z_1 and Z_2 are independent. In this study, the bias in $\widehat{\beta}_1$ and $\widehat{\beta}_2$ is estimated over 5000 repetitions of this process and the true values of the parameters are $\beta_1 = \beta_2 = 1$. The simulation standard error for the bias estimate is below 0.02 in all cases.

Table 1 shows the bias of the parameter estimates using different stratification points. There is a very large bias in the estimates of β_2 when the wrong stratification points are used in the stratified analyses. Also, in the unstratified analysis $\widehat{\beta}_2$ has a large bias. In general, the bias increases for stratification points further from the correct one.

The coefficient β_2 is underestimated for $\rho = 2$ and overestimated for $\rho = 0.5$ when the stratified analysis is misspecified and when the unstratified analysis is used. On the other hand, estimates of β_1 are reasonably close to the true value for all analyses. The same conclusions were obtained by Monte Carlo simulations using as the correct stratification point values different from 0.5.

Similar conclusions for the estimates of β_2 are obtained when Z_1 is a continuous covariate generated from a uniform $(0, 1)$ and Z_1 and Z_2 are correlated. These results are not presented in the paper but it is available upon request. The bias for the estimate of β_1 it is very small when compared to the bias associated to $\widehat{\beta}_2$. There is an indication that the $\widehat{\beta}_1$ is more biased for the sample of size 40 than for the sample of size 80.

Table 1
Bias of the estimated coefficients. Z_1 *has a Bernoulli distribution which is independent of* Z_2*.*

Sample Size	Stratif. Point	$\rho = 2$		$\rho = 0.5$	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
n=40	unstrat.	−0.025	−1.586	0.103	2.150
	0.1	−0.028	−1.563	0.094	2.145
	0.2	−0.039	−1.485	0.088	1.914
	0.3	−0.032	−1.095	0.092	1.512
	0.4	0.020	−0.922	0.134	0.513
	0.5	0.085	0.074	0.078	0.081
	0.6	−0.026	−1.479	0.113	1.788
	0.7	−0.035	−2.168	0.172	3.401
	0.8	0.009	−2.237	0.131	3.468
	0.9	−0.064	−2.012	0.121	3.030
n=80	unstrat.	0.062	−1.774	−0.067	2.463
	0.1	0.046	−1.960	−0.055	2.534
	0.2	0.062	−2.004	−0.092	2.472
	0.3	0.051	−1.885	−0.101	2.307
	0.4	0.047	−0.914	−0.116	1.505
	0.5	0.032	0.034	0.030	0.030
	0.6	0.040	−1.570	−0.091	2.072
	0.7	0.040	−1.981	−0.057	3.136
	0.8	0.096	−2.097	−0.108	3.265
	0.9	0.083	−1.956	−0.078	2.887

3 A criterion for model stratification

Consider the following situation of nonproportional hazards

$$\lambda(t) = \begin{cases} \lambda_{0_1}(t)\ exp(\beta_1 Z_1 + \beta_2 Z_2) & \text{if } Z_2 \leq c \\ \lambda_{0_2}(t)\ exp(\beta_1 Z_1 + \beta_2 Z_2) & \text{if } Z_2 > c \end{cases} \tag{3.1}$$

where Z_2 is the continuous covariate of interest, c is a known constant that belongs to the support of Z_2 and $\lambda_{0_1}(t) \neq \lambda_{0_2}(t)$ for some $t \in [0, \infty)$.

The interest here is to derive a criterion in order to find the correct stratification point c. The criterion proposed approximates the baseline hazard function in model (1.1) by a step function. This is the approach considered by Andersen (1982) and Barbosa, Colosimo and Louzada-Neto (1996) in different contexts. Even though this parametric approach does not have the flexibility of the proportional hazards model, some applied studies described in the literature have obtained similar values for the estimates of the parameter of interest β when both models are used. See Selmer (1990) for a study comparing these two models. On the other hand, the parameters in the parametric model are estimated by simple maximum likelihood techniques.

Consider a survival time data set with $p \geq 1$ covariates including a continuous covariate Z. Let k be the number of strata arising when the observations are grouped according to Z, $(-\infty, c_1), [c_1, c_2), \ldots, [c_{k-1}, \infty)$ and $\mathcal{C} = \{c_1, \ldots, c_{k-1}\}$, the finite set of stratification points. Consider also m prespecified time intervals, $[0, b_1), \ldots, [b_{m-1}, \infty)$. The baseline hazard function $\lambda_0(t)$ is approximated in each stratum by $\lambda_{0_i}(t)$, $i = 1, 2, \ldots, k$, that is constant in each of the intervals $[b_{j-1}, b_j)$, $j = 1, \ldots, m$, where $b_0 = 0$ and $b_m = \infty$. Therefore mk additional parameters $\lambda = (\lambda_{11}, \ldots, \lambda_{mk})^t$ are introduced in the model by taking $\lambda_{0_i}(t) = \lambda_{ij}$ when $t \in [b_{j-1}, b_j)$.

Let t_{ijl} be the l^{th} observed survival time in the j^{th} time interval and in the i^{th} stratum defined by Z. Correspondingly consider Z_{ijl}, the p-vector of covariates and δ_{ijl}, the failure indicator, $i = 1, \ldots, k, j = 1, \ldots, m$, and $l = 1, \ldots, n_{ij}$.

It is appropriate at this point to make a reparametrization in order to express the likelihood for β and λ in a more convenient form. Consider for any value c^0 on the support of Z_2

$$\begin{aligned} \lambda_{ij} &= \lambda_{1j} + \psi_{ij}, \quad i = 1, \ldots, c_0; \quad j = 1, \ldots, m; \\ \lambda_{ij} &= \lambda_{kj} + \eta_{ij}, \quad i = c_0 + 1, \ldots, k; \quad j = 1, \ldots, m; \end{aligned}$$

where $\psi_{1j} = 0$ and $\eta_{kj} = 0$ for $j = 1, \ldots, m$ and $c_0 = max\{c^{'} : c^{'} \in \mathcal{C}$ and $c^{'} \leq c^0\}$ is the considered stratification point.

The expression of the loglikelihood for β, λ, ψ and η has the following form

$$\begin{aligned} l(\beta, \lambda, \psi, \eta) &= \sum_{i=1}^{c_0} \sum_{j=1}^{m} d_{ij} \; log(\lambda_{1j} + \psi_{ij}) + \sum_{i=c_0+1}^{k} \sum_{j=1}^{m} d_{ij} \; log(\lambda_{kj} + \eta_{ij}) \\ &+ \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{l=1}^{n_{ij}} \delta_{ijl}(\beta^t Z_{ijl}) - \sum_{i=1}^{c_0} \sum_{j=1}^{m} (\lambda_{1j} + \psi_{ij}) \sum_{o=1}^{m} \sum_{l=1}^{n_{io}} exp(\beta^t Z_{iol}) S_{iolj} \\ &- \sum_{i=c_0+1}^{k} \sum_{j=1}^{m} (\lambda_{kj} + \eta_{ij}) \sum_{o=1}^{m} \sum_{l=1}^{n_{io}} exp(\beta^t Z_{iol}) S_{iolj} \end{aligned}$$

where d_{ij} is the number of failures in the i^{th} stratum and time interval $[b_{j-1}, b_j)$ and S_{iolj} denotes the time that observation (i, o, l) spent in the time interval $[b_{j-1}, b_j)$.

The observed information matrix $\mathcal{I}(\beta, \lambda, \psi, \eta)$ can be viewed as a partitioned matrix

$$\mathcal{I}(\beta, \lambda, \psi, \eta) = \begin{bmatrix} \mathcal{I}_{\psi,\eta} & \mathcal{I}_{\beta,\lambda,\psi,\eta} \\ \mathcal{I}_{\beta,\lambda,\psi,\eta} & \mathcal{I}_{\beta,\lambda} \end{bmatrix}.$$

It is interesting to observe that the largest component of this matrix, $I_{\psi,\eta}$, is a diagonal matrix. This property of $\mathcal{I}(\beta, \lambda, \psi, \eta)$ simplifies future computations.

The interest here is to derive a criterion to search for the correct stratification point when (3.1) is the correct model. Even though there is no interest in a hypothesis test, this criterion is easier to understand using the well known terminology of tests.

Consider the following hypothesis

$$\begin{aligned} H_0(c^0) \quad : \quad & \psi_{ij} = 0, \quad i = 2, \ldots, c_0 \quad ; j = 1, \ldots, m \quad \text{and} \\ & \eta_{ij} = 0, \quad i = c_0 + 1, \ldots, k-1 \quad ; j = 1, \ldots, m. \end{aligned} \tag{3.2}$$

The continuous covariate Z_2 initially defines $k > 2$ strata in the considered partition. Now for a given c^0 in the support of Z_2, the hypothesis $H_0(c^0)$ says that just 2 strata arise when the observations are grouped according to Z_2.

The expression of the likelihood ratio test for a given c^0 in the support of Z_2

$$Q(c^0) = -2[l(\widehat{\lambda}_0, \widehat{\beta}_0, c_0) - l(\widehat{\lambda}, \widehat{\beta}, \widehat{\psi}, \widehat{\eta})] \tag{3.3}$$

where $\widehat{\lambda}_0$ and $\widehat{\beta}_0$ are the estimates of λ and β under H_0, is a natural statistic to be used as a criterion in searching for the best stratification point.

Under H_0 and for the correct stratification c, $Q(c)$ converges to a chi-square distribution with $m(k-2)$ degrees of freedom, under some regularity conditions (see Cox and Hinkley (1974), Chapter 9). Even though there is no interest here in the expression of Q as a test statistic, the result above gives information to establish a decision criterion. That is, to pick c^0 in the support of Z_2 such that $Q(c^0)$ is minimum. However, since $l(\widehat{\lambda}, \widehat{\beta}, \widehat{\psi}, \widehat{\eta})$, the loglikelihood evaluated at the unrestricted estimates, does not depend on c^0, an equivalent criterion that is simpler in its form is

pick $\boldsymbol{c^0}$ in the support of $\boldsymbol{Z_2}$ such that $\boldsymbol{l(\widehat{\lambda}_0, \widehat{\beta}_0, c_0)}$ is maximum.

The number of strata k does not need to be specified in order to use the criterion, so that $l(\widehat{\lambda}_0, \widehat{\beta}_0, c_0)$ can be exchanged by $l(\widehat{\lambda}_0, \widehat{\beta}_0, c^0)$ in the criterion above. This is a nice feature of this criterion, since being more precise in the decision taken means a large number of values in the support of Z_2 to be considered by the criterion. In the extreme, by considering c^0 equal to each distinct value of the support of Z_2, a maximum likelihood estimate of the stratification point for a two strata model is obtained. For each fixed c^0, the number of parameters to be estimated is always the same and equal to $2m + p$.

4 Criterion performance

The proposed criterion is applied in some situations under the model (2.1). The results are based on a Monte Carlo simulation using 1000 repetitions. Ten strata are considered according to the division of the data into the set of stratification points $\mathcal{C} = \{0.1, \ldots, 0.9\}$ defined by the covariate Z_2. The time axis is divided into two intervals for the application of the criterion. In all situations considered below the median of the failure times T for a typical sample is chosen as the time split. All repetitions that include an empty stratum are removed from the simulation.

Table 2 shows the percentage of correct decisions based on the proposed criterion for the situation presented in Table 1 and for Z_1 generated from a uniform $(0, 1)$. Z_1 and Z_2 are generated with correlations 0, 0.5 and 0.8. The results in Table 2 show that the performance of the proposed criterion is reasonably good for the designs used in the simulation. The percentage of correct decisions is over 50% for the sample of size 80 and over 70% for the sample of size 120 for all designs and values of ρ. The results are better for $\rho = 0.5$ than $\rho = 2$ and for the designs where Z_1 and Z_2 are independent than correlated. There is a small reduction of the percentage of the correct decision as the correlation of Z_1 and Z_2 increases in both sample sizes and values of ρ.

Table 2
Percentage of correct decision for the proposed criterion.

Design	Sample Size	Percentage of Correct Decision $\rho = 2$	$\rho = 0.5$
Z_1 binary	80	81.6	88.0
Z_1,Z_2 indep.	120	92.2	97.0
Z_1 contin.	80	69.6	89.7
Z_1,Z_2 indep.	120	78.9	92.3
Z_1 contin.	80	56.4	68.3
corr $(Z_1$,$Z_2) = 0.5$	120	76.0	84.8
Z_1 contin.	80	50.8	71.3
corr $(Z_1$,$Z_2) = 0.8$	120	82.8	91.2

The overall bias due to the use of the criterion or any criterion in general, may be measured by

$$B = \sum_{i \in \mathcal{C}} P(c_i \text{ is chosen})\, B_i \tag{4.1}$$

where B_i is the bias associated with using stratification point c_i. The values of B_i are obtained from the simulations in Section 2 and the values of $P(c_i$ is chosen) from the simulations in this section.

The overall bias B using the proposed criterion and the bias using the unstratified analysis for the estimate $\widehat{\beta}_2$ are shown in Table 3 for samples of size 80. The values in this table show a great improvement in using the proposed criterion and fitting the stratified proportional hazards model since they are much smaller than the bias associated with the unstratified model.

Table 3
Bias in $\widehat{\beta}_2$ using the unstratified analysis and the proposed criterion.

Design	Bias in $\widehat{\beta}_2$ using :			
	Unstratified Analysis		Proposed Criterion	
	$\rho = 2$	$\rho = 0.5$	$\rho = 2$	$\rho = 0.5$
Z_1 binary Z_1,Z_2 indep.	-1.774	2.463	-0.627	0.436
Z_1 contin. Z_1,Z_2 indep.	-1.669	2.241	-0.416	0.286
Z_1 contin. corr (Z_1,Z_2) = 0.5	-1.633	2.296	-0.583	0.845
Z_1 contin. corr (Z_1,Z_2) = 0.8	-1.712	3.133	-0.699	0.698

5 Real data set

The Stanford Heart Transplant data set (Crowley and Hu, 1977) has been used as an illustration for several techniques in survival analysis. One question of interest is related to the fit of the covariate waiting time to transplant in the proportional hazards model when age, mismatch score and previous surgery are also used as covariates.

Arjas (1988) studying this data set considered two strata based on an arbitrary split of the covariate waiting time to transplant (up to 20 days and larger than 20 days). Based on the plot $log\widehat{\Lambda}(t)$ vs t and also on the plot proposed by his paper, he concluded that the proportional hazards assumption did not hold for these strata.

As an illustration of the criterion presented in the previous sections, it is applied to the Stanford heart transplant data set. Six strata defined by the covariate waiting time to transplant are initially considered as $\mathcal{C} = \{10, 20, 30, 40, 60\}$. For the application of the parametric criterion, two time intervals are defined by the median of the failure times (63 days).

Table 4 shows the values of the conditional expected number of failures given the history until up time t (Schoenfeld, 1980) and the respective observed number of failures for each of the six strata. As one can notice the largest difference between expected and observed number of failures in this table occurs for the stratum defined by $[10, 20)$ days. For the other strata the differences are relatively small. Table 5 shows the loglikelihood values under the specified stratification points for the application of the proposed criterion. The criterion picks 20 days as the best stratification point among those considered in the partition.

Table 4
Expected and observed number of failures for six strata defined by the covariate waiting time to transplant.

Stratum	Expec. Value	Observ. Value
$[0, 10)$	11.40	12
$[10, 20)$	5.39	9
$[20, 30)$	7.34	5
$[30, 40)$	7.42	6
$[40, 60)$	6.26	5
≥ 60	3.19	5

Table 5
Loglikelihood values under the specified stratification points for the application of the proposed criterion.

Stratif. Points	$l(\widehat{\beta}_0, \widehat{\lambda}_0, c_0)$
10	−280.86
20	−278.59
30	−280.90
40	−280.75
60	−279.61

The proposed criterion is flexible in the sense that any other stratification point can be included in the decision rule without being explicitly in the partition. Other stratification points have been considered, such as 5, 15, 25 and 35 days, but the stratification point 20 days is still the best selection according to the criterion.

6 Discussion

A criterion was proposed as guideline for the division of the data into strata in a situation of nonproportional hazards. The main features of the proposed criterion is : (i) it approximates the baseline hazard function by a step function; (ii) it uses an expression derived from the likelihood ratio test as a criterion to find the best stratification point; and (iii) it does not require specification of the total number of strata in the general partition for its derivation.

In a situation such as was examined in Section 3, where there are two covariates and one stratification point, it was showed that the criterion has a nice performance. The criterion can be easily extended for other situations of interest. It can be used in the way it was proposed for situations with more than two covariates provided just one covariate defines the stratification.

The situation of most interest is the one with more than one stratification point. For instance, with two stratification points (three strata), say c_1 and c_2, and two covariates, the stratified model is specified by

$$\lambda(t) = \begin{cases} \lambda_{0_1}(t)exp(\beta_1 z_1 + \beta_2 z_2) & \text{if } z_2 \leq c_1 \\ \lambda_{0_2}(t)exp(\beta_1 z_1 + \beta_2 z_2) & \text{if } c_1 < z_2 \leq c_2 \\ \lambda_{0_3}(t)exp(\beta_1 z_1 + \beta_2 z_2) & \text{if } z_2 > c_2 \end{cases}$$

for $c_1 < c_2$ both on the support of z_2. Considering k possible stratification points of interest in the set $\mathcal{C}$, there will be $\binom{k}{2}$ possible combinations of stratification points to be included in the decision rule of the parametric criterion and the criterion expression is still the same.

An extension of the model (3.1), where the effect of one covariate varies freely between strata, such as

$$\lambda(t) = \begin{cases} \lambda_{0_1}(t)exp(\beta_1 z_1 + \beta_2 z_2) & \text{if } z_2 \leq c \\ \lambda_{0_2}(t)exp(\beta_3 z_1 + \beta_2 z_2) & \text{if } z_2 > c \end{cases}$$

where $\beta_1 \neq \beta_3$. This information can be included in the derivation of the criterion. The likelihood can be modified to incorporate this extra information and the remainder of the derivations follow in the same way as in Section 3.

A last question that concerns the application of the criterion must be discussed. What happens if the proportional hazards model is correct and a stratified model chosen by the criterion is fitted? If the unstratified model is correct it implies that any stratified model is also correct. However, it is expected that some loss of efficiency in the estimation of β will occur when stratification is used unnecessarily.

For instance, under the conditions of Table 1 and considering $\lambda(t) = t^{\rho-1}exp(\beta_1 Z_1 + \beta_2 Z_2)$ as the correct model, there is no additional bias in the estimates of the parameters when a stratified model is fitted. It means that the overall bias measure B defined in Section 5 is approximately zero for the unstratified as well as for any stratified model. However, the simulation standard deviations that estimate the standard errors of the estimated coefficients are smaller under the unstratified model than any stratified model. In a simulation with 5000 repetitions, sample of size 80 and two values for ρ (2 and 0.5) the simulation standard deviations for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in the unstratified model are 0.287 and 0.523 respectively and for a stratified model are around 0.291 and 0.884 respectively for $\rho = 0.5$. For $\rho = 2.0$, the simulation standard deviations for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are 0.302 and 0.539 respectively for the unstratified model and around 0.309 and 0.849 respectively for a stratified model. These values are an indication that this loss of efficiency is not very mild. Therefore a careful analysis must be done using goodness-of-fit techniques to avoid using the proposed criterion and consequently a stratified analysis unnecessarily.

Acknoledgements

The author wish to thank Prof. Barry Storer for his valuable comments, and constructive suggestions. The author is partially supported by the Brazilian agencies Conselho Nacional de Pesquisa, CNPq.

References

Andersen, P. K. (1982). Testing goodness of fit for Cox's regression and life model. *Biometrics*, **38**, 67–77.

Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *J. Am. Statist. Assoc.*, **83**, 204–212.

Barbosa, E. P., Colosimo, E.A. and Louzada-Neto, F. (1996). Analysis of accelerated life tests with a piecewise exponential distribution via generalized linear modelling. *IEEE Transactions on Reliability*, **45**, 619–623.

Barlow, W. E. and Prentice, R. A. (1988). Residuals for relative risk regression. *Biometrika*, **75**, 65–74.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269–76.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *J. Am. Statist. Assoc.*, **72**, 27–36.

Grambsch, P. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**, 515–526.

Hess, K. R. (1995). Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Statistics in Medicine*, **14**, 1707–1723.

Horowitz, J. L. and Neumann, G. R. (1992). A generalized moments specification test of the proportional hazards model. *J. Am. Statist. Assoc.*, **87**, 234–240.

Kalbfleisch, J. D. and Prentice, R. A. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, **60**, 267–278.

Kalbfleisch, J. D. and Prentice, R. A. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Kay, R. (1977). Proportional hazards regression models and the analysis of censored data. *Appl. Statist.*, **26**, 227–37.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Moreau, T., O'Quigley, J. and Mesbah, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. *Appl. Statist.*, **34**, 212–218.

O'Quigley, J. and Pessione, F. (1989). Score tests for homogeneity of regression effect in the proportional hazards model. *Biometrics*, **45**, 135–144.

Schoenfeld, D. A. (1980). Chi-squared goodness of fit tests for the proportional hazards regression model. *Biometrika*, **67**, 145–153.

Schoenfeld, D. A. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, **69**, 239–241.

Selmer, R. (1990). A comparison of Poisson regression models fitted to multiway summary table and Cox's survival model using data from a blood pressure screening in the city of Bergen, Norway. *Statistics in Medicine*, **9**, 1157–65.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazards models. *Biometrika*, **73**, 363–367.

Wei, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Am. Statist. Assoc.*, **79**, 649–652.