Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1132

Published online 26 August 2010 in Wiley Online Library

Change Detection in the Cox Proportional Hazards Models from Different Reliability Data

Zhiguo Li,[a]*† Shiyu Zhou,[b] Crispian Sievenpiper[c] and Suresh Choubey[c]

The Proportional Hazards (PH) model is an important type of failure time regression model which relates the occurrence probability of critical failures to influential factors. However, little research work has been done on detecting changes in the PH models fitted based on different sets of reliability data. This paper develops the methods for change detection in the Cox PH models, also known as Semiparametric PH model, for reliability prediction and/or assessment of the time-to-failure data collected from different subjects. The effectiveness of the developed methods is illustrated through numerical studies and real-world data analysis. The developed technique possesses wide applicability to the systems and processes where the Cox PH model fits the reliability data well. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: change detection; the Cox PH model; reliability analysis

1. Introduction

Reliability modeling, which is the process to build a descriptive or predictive model for the reliability of a component, subsystem, or a system, has always been a significant part of reliability assessment and prediction. Among the failure-time regression models that relate the failure times of a component or a system to influential factors (covariates), such as customer usage patterns and environmental stresses, Proportional Hazards (PH) model[1] is an important survival model widely used in reliability engineering[2–5]. According to the form of the baseline hazard function, the PH model can be classified as the Parametric PH model or the Semiparametric (Cox) PH model. The former assumes a parametric distribution, such as the Weibull distribution, for the baseline hazard function[6–8], whereas the latter has an unspecified baseline hazard function[9–11]. Both types of PH models are widely used in modeling different types of reliability data. Based on PH models, some researchers[12–14] have developed optimal Condition-Based Maintenance (CBM) policies, where the maintenance decision is determined based on the information or signals collected through condition monitoring of the equipment or estimates of the machine reliability.

The PH model lays a mathematical foundation for predicting the failure occurrence, testing the prediction accuracy, and developing optimal maintenance policy. However, many interesting technical challenges exist in applying this model to the reliability data, which could be collected from different machines/locations and normally evolve over time due to the underlying degradation mechanism. Thus, detection of changes in the PH models from different reliability data sets is critical for the practical application of this model.

One of the challenges is that the PH model is often estimated based on historical data. To utilize such a model for failure prediction and other applications, however, we should make sure that this model is up-to-date and reflects the current system behavior. After all, system characteristics, such as the failure occurrence rate and the influence of covariates, are subject to changes. As a motivating example, a reliability engineer who designs maintenance and service policies based on a PH model for failure events needs to make sure that the adopted PH model, which is often fitted using the historical reliability data, should fit well with the currently observed reliability data. Only if the adopted model differs significantly from the present data should both the PH model and the associated service policies be updated.

Another example that motivates the detection of changes in PH model is the remote service of machine in the field. A customer service engineer may want to analyze the time-to-failure data collected from different machines in the field. Only if some machines

[a] *Xerox Research Center Webster, Webster, NY 14580, U.S.A.*
[b] *Department of Industrial and Systems Engineering, University of Wisconsin, Madison, WI 53706, U.S.A.*
[c] *GE HealthCare, Global Service Technology, Waukesha, WI 53186, U.S.A.*
* *Correspondence to: Zhiguo Li, Xerox Research Center Webster, Webster, NY 14580, U.S.A.*
† *E-mail: zhiguo.li.wisc@gmail.com*

Copyright © 2010 John Wiley & Sons, Ltd.

show significant differences in the failure behavior from others after the effects of the usage patterns are adjusted, a customer visit will be scheduled to find out the causes of the change, which normally is costly. In addition, the machines with similar behaviors can be grouped together and the same service policies can be implemented.

The problem of testing if the present data fits well with a PH model fitted from the historical data can be formulated as an off-line change detection problem. The general change detection problem has been an active research topic in many applications, such as quality control, segmentation of signals, and CBM of industrial equipment[15]. In change detection, different distributions or mathematical/statistical models are fitted to the data collected from a static or dynamic system and then tools or algorithms are proposed to detect the changes in the parameters of the distribution or model. For instance, in the field of statistics and quality control, some methods are developed to handle change-point problems or detect homogeneous regions of a random field[16]. These include Shewhart Control Charts, the CUSUM method for sequential change-point detection[17–19]. In more complex applications, some change detection algorithms are developed based on different mathematical/statistical models, such as *regression models*, *ARMA (ARIMA) models* in time series analysis, *state–space models* for dynamic systems, and *controlled semi-Markov processes*[20]. In Computer Science, some authors developed a framework for quantifying the differences between two data sets for change detection purposes under several *data mining models* and algorithms, such as dt-models, list-models, and cluster-models[21, 22]. However, to the best of our knowledge, there is little work on change detection in the time-to-failure data under Semiparametric (Cox) PH models. *In this paper, we shall develop the methods for off-line change detection in Cox PH models fitted from different sets of reliability data.* Similar to Reference[22], the term *deviation* will be used for the difference between two Cox PH models in this paper.

The remainder of the paper is organized as follows. In Section 2, the problem formulation and the statistical procedure to measure the differences between the Cox PH models are presented. The details of each step of the proposed procedure can be found in Section 3, where we also illustrate the effectiveness of the developed procedure through numerical case studies and real-world data analysis. Finally, we conclude the paper in Section 4.

2. Measuring deviation in the Cox PH models for different sets of reliability data

2.1. Brief review of the Cox PH model

Denote T as the time-to-failure. Let $h[t|\mathbf{Z}(t)]$ be the hazard (rate) function at time t with covariate vector $\mathbf{Z}(t)$, the basic Cox PH model[23] is as follows:

$$h[t|\mathbf{Z}(t)] = h_0(t)\exp[\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}(t)] = h_0(t)\exp\left[\sum_{k=1}^{p}\beta_k Z_k(t)\right], \tag{1}$$

where $h_0(t)$ is the *baseline* hazard (rate) function, and the PH model means that the hazard rate of a subject is proportional to its baseline hazard rate $h_0(t)$, which is the basic assumption of the Cox PH model. In the model, $\boldsymbol{\beta}$ is the coefficient vector and $\mathbf{Z}(t) = [Z_1(t), Z_2(t), \ldots, Z_p(t)]^{\mathrm{T}}$ is the covariate vector. $Z_i(t)$, $i = 1, 2, \ldots p$, is a *time-dependent* covariate if its value varies with time in a part life. If the value of $Z_i(t)$ does not change over time, we denote it as Z_i.

The coefficient vector $\boldsymbol{\beta}$ is estimated by maximizing the partial likelihood. Based on the partial maximum likelihood estimator $\hat{\boldsymbol{\beta}}$, the Breslow estimator of the *baseline cumulative hazard function*, $\hat{H}_0(t, \hat{\boldsymbol{\beta}})$ can be estimated directly from the data and thus the estimated *baseline survival function* $\hat{S}_0(t, \hat{\boldsymbol{\beta}}) = \exp[-\hat{H}_0(t, \hat{\boldsymbol{\beta}})]$ can be obtained. From $\hat{H}_0(t, \hat{\boldsymbol{\beta}})$, we can obtain a non-smooth or smooth estimate of the baseline hazard $h_0(t)$ through several methods. These methods often give rise to different estimate results of $h_0(t)$. Obviously, a Cox PH model is determined by the baseline function $h_0(t)$ and the covariate coefficients $\boldsymbol{\beta}$. By statistically comparing the fitted baseline function and the covariate coefficients across different Cox PH models, the model deviations can be determined. In this paper, a systematic methodology is developed to compare the covariate coefficients (assuming the covariates are the same) for different Cox PH models. Since the survival function (or equivalently the cumulative hazard function) is of main interest to the reliability engineers and also easy to be interpreted, and the estimate of the baseline survival (cumulative hazard) function is more straightforward compared with that of the baseline hazard rate function, we also develop a method for the comparison of $S_0(t)$ or equivalently, $H_0(t)$ in this study.

In biomedical applications, the comparison of the survival functions or cumulative hazard functions of two treatments is a very common problem solved through statistical tests based on comparing the survival curves or the cumulative hazard functions over the entire observational period. In the case that no other covariates have impact on the survival, the log-rank test[24] is widely used for the abovementioned aim, whereas the testing of treatment effects with other covariates is performed in the context of regression models, typically the Cox PH model, which can adjust the effects of the other covariates. As an example, in Reference[25] is proposed the technique for building a simultaneous confidence band, i.e. a collection of confidence intervals for all times in a specified time interval $[t_1, t_2]$ with the simultaneous coverage probability $1-\alpha$, where α is the confidence level, under the stratified Cox PH model[23]. The selection of time points t_1 and t_2 will affect the obtained simultaneous confidence band. Although the simultaneous confidence band is used to assess the overall probability that the two baseline survival functions differ for a given time interval $[t_1, t_2]$, it cannot be used to show at what fixed times these two curves are statistically different since it only gives an overall probability.

In this paper, we propose a procedure for constructing the pointwise confidence region for the difference between the two cumulative hazard curves to answer the question 'at what times are these two baseline survival functions different?', which is an

Copyright © 2010 John Wiley & Sons, Ltd.

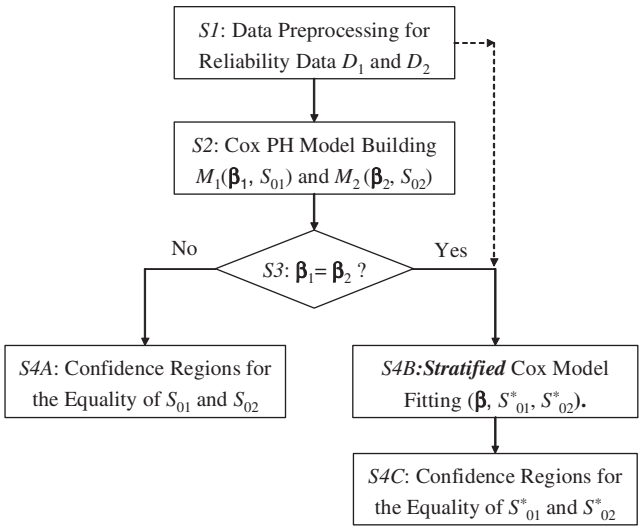

Figure 1. The procedure for measuring deviations between the Cox models

important question in industrial applications such as reliability analysis. For example, regardless of whether the overall survival functions are statistically different over the entire observational time period, reliability engineers may be interested in finding out whether one group of machines has higher early survival or higher long-term survival compared to the other group. The answer will help to implement different service and maintenance policies on different machines, and this information can also be fed back to the design engineers. Compared to the simultaneous confidence band, the pointwise confidence region is built by grouping together all fixed time points at which the two baseline functions differ to answer the abovementioned question. The details are discussed in Sections 2.4 and 2.5.

2.2. *Steps for measuring deviations between Cox PH models*

A diagram of the complete procedure for measuring the deviation in the Cox model is illustrated in Figure 1.

The detailed steps are as follows:

Step 1: Preprocess two time-to-failure data sets D_1 and D_2;

Step 2: Construct two Cox models M_1 and M_2 for D_1 and D_2, respectively. For simplicity, we use $M_1(\boldsymbol{\beta}_1, S_{01}(t))$ and $M_2(\boldsymbol{\beta}_2, S_{02}(t))$ to denote the models, since the Cox model consists of two parts: the exponential part $\exp[\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}(t))]$ and the baseline survival function $S_0(t)$ adjusted for the effects of covariates. Proceed to step 3.

In some cases, the data analyst may assume that the regression coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the same based on the engineering knowledge of the system/process. That is, there is no justification for the changes in the effects of the covariates. In this case, a *stratified Cox model* will be fitted to the *combined data* with D_1 and D_2 as the strata. In this case, go to step 4B following the dashed line in Figure 1.

Step 3: Check whether the coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the same or not in the first place. As stated in Section 2.1, the coefficient vector $\boldsymbol{\beta}$ can be estimated by maximizing the partial likelihood, *which does not depend on the baseline part*. Thus, the effects of covariates can be compared without knowing $S_0(t)$. If $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, go to step 4A; otherwise go to step 4B.

Step 4A: (When $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$): Compare the baseline survival curves $S_{01}(t_0)$ and $S_{02}(t_0)$ and construct a $(1-\alpha)\times 100\%$ pointwise confidence region for the equality of two baseline survival curves, i.e. $S_{01}(t_0) = S_{02}(t_0)$, at each fixed time point t_0. Thus the time intervals when the two curves are different can be identified. Note that this step can be skipped since it can be concluded that the two Cox PH models are different simply based on the fact of $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$.

Step 4B: (When $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$): Fit a *stratified* Cox model[23] to the *combined data* consisting of D_1 and D_2 as

$$h_j^*[t|\mathbf{Z}(t)] = h_{0j}^*(t)\exp[\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}(t)], \quad j = 1, 2 \tag{2}$$

where the regression coefficients are assumed to be the same in the two data sets (*strata*), whereas the baseline parts are not related to each other and thus can be different. Some details of the stratified Cox model are discussed in Section 2.5.

Step 4C: Construct a $(1-\alpha)\times 100\%$ pointwise confidence region for the equality of two baseline survival curves $S_{01}^*(t_0) = S_{02}^*(t_0)$ at each fixed time point t_0. *Remark*: if the overall probability of the two survival functions over the entire observational period is of interest, we need to build a simultaneous confidence band for the difference of the two baseline survival curves. Refer to Reference[25] for details.

Note that, in Steps 4A and 4C, the range of the fixed time t_0 is $[0, \tau]$, where τ is defined as the smaller value of the largest failure times in D_1 and D_2, respectively, because the nonparametric estimates of survival function $S(t)$ and cumulative hazard function $H(t)$ are not defined after τ[23] and thus we cannot compare the two baseline functions when $t > \tau$.

Copyright © 2010 John Wiley & Sons, Ltd.

2.3. *Measuring the deviation in the effects of covariates*

In this subsection, our approach for measuring the differences between the effects of covariates is presented. To measure the differences in the coefficient vector $\boldsymbol{\beta}$, a global hypothesis can be formulated as $H_0:\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$ vs $H_1:\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2$. Here the term 'global' is used to differentiate this test with the local test when only a subset of $\boldsymbol{\beta}$ is of interest. Now we have the following lemma and the proof can be found in Appendix A.

Lemma 1

Let M_1, M_2 be two Cox PH models induced by the two reliability data sets D_1, D_2 and $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are the coefficient vectors of M_1, M_2, respectively. Under the null hypothesis $H_0:\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$, we have the statistic

$$X^2=(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2)^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}_1+\hat{\boldsymbol{\Sigma}}_2)^{-1}(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2), \tag{3}$$

which has a chi-squared distribution with p degrees of freedom and p is the length of $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$. Here $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are the partial maximum likelihood estimators of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively; $\hat{\boldsymbol{\Sigma}}_1$ and $\hat{\boldsymbol{\Sigma}}_2$ are the estimators of variance-covariance matrices for $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$, respectively. The proof of this lemma can be found in Appendix A.

The null hypothesis will be rejected when the value of statistic X^2 exceeds the α% upper percentile of the chi-squared distribution, where α is a given significance level. Thus statistic X^2 proposed in Lemma 1 can be used to detect the differences in $\boldsymbol{\beta}$.

One may also be interested in the property of this statistic, such as its power to reject a *false* null hypothesis. Given a significance level α, the power of hypothesis testing can be estimated as

$$\text{Estimated power}=Pr(X^2(\hat{\tau})>\chi^2_{p,1-\alpha}), \tag{4}$$

where $X^2(\hat{\tau})$ is a *non-central* χ^2 random variable with p degrees of freedom and the estimated non-centrality parameter $\hat{\tau}=(\mathbf{b}_1-\mathbf{b}_2)^{\mathrm{T}}(\mathbf{S}_1+\mathbf{S}_2)^{-1}(\mathbf{b}_1-\mathbf{b}_2)$, where $\mathbf{b}_j$ and $\mathbf{S}_j$, $j=1,2$, are the estimates of the coefficient vector and variance–covariance matrix from the data.

If a subset of the coefficient vector $\boldsymbol{\beta}_a\subset\boldsymbol{\beta}$ is of interest, that is, we try to test if $H_0:\boldsymbol{\beta}_{1a}=\boldsymbol{\beta}_{2a}$ vs $H_1:\boldsymbol{\beta}_{1a}\neq\boldsymbol{\beta}_{2a}$ (*local test*), the statistic in Equation (3) can be modified as follows:

$$X^2_a=(\hat{\boldsymbol{\beta}}_{1a}-\hat{\boldsymbol{\beta}}_{2a})^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}_{1,aa}+\hat{\boldsymbol{\Sigma}}_{2,aa})^{-1}(\hat{\boldsymbol{\beta}}_{1a}-\hat{\boldsymbol{\beta}}_{2a}), \tag{5}$$

where $\hat{\boldsymbol{\beta}}_{1a}$ and $\hat{\boldsymbol{\beta}}_{2a}$ are the same subsets of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$, $\hat{\boldsymbol{\Sigma}}_{1,aa}$ and $\hat{\boldsymbol{\Sigma}}_{2,aa}$ are their $q\times q$ submatrices of $\hat{\boldsymbol{\Sigma}}_1$ and $\hat{\boldsymbol{\Sigma}}_2$, where q is the length of $\hat{\boldsymbol{\beta}}_{1a}$ or $\hat{\boldsymbol{\beta}}_{2a}$. The statistic in Equation (5) follows a large sample chi-squared distribution with q degrees of freedom under H_0. Similar to Equation (4), the power can be estimated. In the following subsection, we will develop the procedure for measuring the differences between the baseline parts of the Cox model.

2.4. *Measuring the differences in the baseline survival curves when* $\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2$

Note that this step can be skipped if we are only interested in concluding that the two Cox PH models are different based on the fact of $\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2$. The following procedure is designed to further understand if and at what times the baseline survival functions of the two models differ from each other.

When there are no covariates in the data set, the rank tests such as the log-rank test, Mann–Whitney–Wilcoxon, and Kruskal–Wallis test can be used to compare two survival curves from D_1 and D_2 over the entire time period[23]. In case the covariates exist, we need to use the regression technique, such as the Cox model, to adjust the effects of the covariates on the occurrence of failure events so that we can understand the changes in the 'real' performance of the machines.

In this subsection, we will compare the baseline survival functions of two models $M_1(\boldsymbol{\beta}_1,S_{01}(t))$ and $M_2(\boldsymbol{\beta}_2,S_{02}(t))$ induced by D_1 and D_2. For this, we will consider the null hypothesis $H_0:S_{01}(t_0)=S_{02}(t_0)$ against the alternative $H_1:S_{01}(t_0)\neq S_{02}(t_0)$ for each fixed time point t_0 in the range $[0,\tau]$, where τ is defined as the smaller value of the largest failure times in D_1 and D_2. The confidence region is built based on this testing by unifying all fixed times at which this test does NOT reject the hypothesis of no difference between two baseline survival curves. Note that this null hypothesis is not a test for equality of the two survival curves over the entire observational time period.

This set of hypotheses is equivalent to $H_0:H_{01}(t_0)=H_{02}(t_0)$ against the alternative $H_1:H_{01}(t_0)\neq H_{02}(t_0)$ considering the fact that $S_{0j}(t)=\exp[-H_{0j}(t)]$, $j=1,2$. However, only the test on H_{01} and H_{02} will be studied in this paper because the estimated baseline cumulative hazard function tends to converge faster than the estimated baseline survival function[26].

Let $N_j(t)$, $j=1,2$ denote the counting process which counts the number of failure events in the jth data set at or prior to time t. And $Y_j(t)$, $j=1,2$ is the number of failure times that are NOT less than time t in the jth data set.

The Breslow's estimator[24] for the baseline cumulative hazard rate given the Cox models is

$$\hat{H}_{0j}(t)=\int_0^t\frac{dN_j(u)}{S_j^{(0)}(\hat{\boldsymbol{\beta}}_j,u)},\quad j=1,2, \tag{6}$$

Copyright © 2010 John Wiley & Sons, Ltd.

where $S_j^{(0)}(\hat{\boldsymbol{\beta}}_j, u) = \sum_{i=1}^{n_j} Y_{ij}(u)\exp(\hat{\boldsymbol{\beta}}_j \mathbf{Z}_{ij}(u))$, n_j is the number of failures in the jth data set, and $Y_{ij}(u)$ is the indicator for the ith failure time in the data set j being not less than time u. Thus we have the statistic for $H_0 : H_{01}(t_0) = H_{02}(t_0)$ against $H_1 : H_{01}(t_0) \neq H_{02}(t_0)$ as

$$\Delta\hat{H}_0(t) = \hat{H}_{01}(t) - \hat{H}_{02}(t). \tag{7}$$

The test of $H_0 : H_{01}(t_0) = H_{02}(t_0)$ cannot be rejected at an α level if we have $|\Delta\hat{H}_0(t)/\sqrt{\text{var}[\Delta\hat{H}_0(t_0)]}| \leqslant z_{\alpha/2}$, where $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution. The variance of this statistic can be derived according to Corollary VII.2.4 in [24] as follows:

$$\text{var}[\Delta\hat{H}_0(t_0)] = \sum_{j=1}^{2}\left\{\int_0^{t_0} \frac{dN_j(u)}{[S_j^{(0)}(\hat{\boldsymbol{\beta}}_j, u)]^2} + \mathbf{W}_j^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_j, t_0)\hat{\Sigma}_j \mathbf{W}_j(\hat{\boldsymbol{\beta}}_j, t_0)\right\}, \tag{8}$$

where $\mathbf{W}_j^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_j, t_0) = \int_0^{t_0} \tilde{\mathbf{Z}}_j(\hat{\boldsymbol{\beta}}_j, u)\, d\hat{H}_{0j}(u)$, $j = 1, 2$,

$$\tilde{\mathbf{Z}}_j(\hat{\boldsymbol{\beta}}_j, u) = \frac{\mathbf{S}_j^{(1)}(\hat{\boldsymbol{\beta}}_j, u)}{S_j^{(0)}(\hat{\boldsymbol{\beta}}_j, u)},$$

and

$$\mathbf{S}_j^{(1)}(\hat{\boldsymbol{\beta}}_j, u) = \sum_{i=1}^{n_j} Y_{ij}(u)\mathbf{Z}_{ij}\exp(\hat{\boldsymbol{\beta}}_j \mathbf{Z}_{ij}(u)).$$

Finally, a $(1-\alpha)\times 100\%$ pointwise confidence region for the times when $S_{01}(t) = S_{02}(t)$ can be derived based on the statistic as $\{t_0 : -z_{\alpha/2} \leqslant \Delta\hat{H}_0(t_0)/\sqrt{\text{var}[\Delta\hat{H}_0(t_0)]} \leqslant z_{\alpha/2}\}$. Namely, the set of times when the two baseline curves are the same contains all the time points in the abovementioned confidence region.

2.5. *Measuring the differences between the baseline survival curves when* $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$

Now let us consider the case that $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the same. The conclusion $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ could come from the hypothesis testing stated in Section 2.3 or the assumption of no change in the effects of covariates based on the engineering knowledge of the system.

If $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, then a stratified Cox model can be fitted to the combined data set consisting of D_1 and D_2, $D = \{D_1, D_2\}$, as $h_j^*[t|\mathbf{Z}(t)] = h_{0j}^*(t)\exp[\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}(t)]$, $j = 1, 2$. In biomedical analysis, the stratified Cox model is the most widely used conditional model for modeling clustered survival data from multiple study centers, because this model possesses 'ease of computation and the applicability across a wide variety of settings'[27]. In our study, if we view different machines or time periods as study centers in clinical trials, then the stratified Cox model is an appealing tool to model the time-to-failure data from different machines or time periods.

As stated in Section 2.1, for the stratified Cox model, the regression coefficients are assumed to be the same in the two data sets whereas the baseline parts can be different. Following the similar logic as in Section 2.4, we can construct a $(1-\alpha)\times 100\%$ confidence region for the equality of two baseline survival curves $S_{01}^*(t_0) = S_{02}^*(t_0)$ at each fixed time point t_0.

The method is similar to that in Section 2.4. We try to test $H_0 : H_{01}^*(t_0) = H_{02}^*(t_0)$ against $H_1 : H_{01}^*(t_0) \neq H_{02}^*(t_0)$ with the statistic in Equation (7). However, the differences are: (1) all $\hat{\boldsymbol{\beta}}_j$s in the formulas (6)–(8) should be replaced by $\hat{\boldsymbol{\beta}}$ since now we have the same coefficient vector for two data sets and (2) the variance of the statistic for $H_0 : H_{01}^*(t_0) = H_{02}^*(t_0)$ against $H_1 : H_{01}^*(t_0) \neq H_{02}^*(t_0)$ is different from Equation (8) since here estimating both baseline cumulative hazards depends on $\hat{\boldsymbol{\beta}}$ and thus the estimates are not independent. To make it clear, we listed all formulae in this case as follows[24].

The statistic for $H_0 : H_{01}^*(t_0) = H_{02}^*(t_0)$ against $H_1 : H_{01}^*(t_0) \neq H_{02}^*(t_0)$ is

$$\Delta\hat{H}_0^*(t_0) = \hat{H}_{01}^*(t_0) - \hat{H}_{02}^*(t_0), \tag{9}$$

which is based on the Breslow's estimator for the baseline cumulative hazard rate

$$\hat{H}_{0j}^*(t) = \int_0^t \frac{dN_j(u)}{S_j^{(0)}(\hat{\boldsymbol{\beta}}, u)}, \quad j = 1, 2, \tag{10}$$

where $S_j^{(0)}(\hat{\boldsymbol{\beta}}, u) = \sum_{i=1}^{n_j} Y_{ij}(u)\exp(\hat{\boldsymbol{\beta}}\mathbf{Z}_{ij}(u))$. The variance of this statistic is,

$$\text{var}[\Delta\hat{H}_0(t_0)] = \sum_{j=1}^{2}\int_0^{t_0} \frac{dN_j(u)}{[S_j^{(0)}(\hat{\boldsymbol{\beta}}, u)]^2} + \mathbf{W}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}, t_0)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}\mathbf{W}(\hat{\boldsymbol{\beta}}, t_0), \tag{11}$$

Copyright © 2010 John Wiley & Sons, Ltd.

where $\mathbf{W}^{\mathrm{T}}(\hat{\boldsymbol{\beta}}, t_0) = \int_0^{t_0} \tilde{\mathbf{Z}}_1(\hat{\boldsymbol{\beta}}, u)\, \mathrm{d}\hat{H}^*_{01}(u) - \int_0^{t_0} \tilde{\mathbf{Z}}_2(\hat{\boldsymbol{\beta}}, u)\, \mathrm{d}\hat{H}^*_{02}(u)$,

$$\tilde{\mathbf{Z}}_j(\hat{\boldsymbol{\beta}}, u) = \frac{\mathbf{S}_j^{(1)}(\hat{\boldsymbol{\beta}}, u)}{S_j^{(0)}(\hat{\boldsymbol{\beta}}, u)},$$

and

$$\mathbf{S}_j^{(1)}(\hat{\boldsymbol{\beta}}, u) = \sum_{i=1}^{n_j} Y_{ij}(u)\mathbf{Z}_{ij} \exp(\hat{\boldsymbol{\beta}}\mathbf{Z}_{ij}(u)).$$

Similarly, a $(1-\alpha)\times 100\%$ pointwise confidence region for the times when $S_{01}(t) = S_{02}(t)$ can be obtained as $\{t_0 : -z_{\alpha/2} \leqslant \Delta\hat{H}^*_0(t_0) / \sqrt{\mathrm{var}[\Delta\hat{H}^*_0(t_0)]} \leqslant z_{\alpha/2}\}$.

3. Case studies

To show the effectiveness of the proposed methods, we carried out the following numerical case studies (Sections 3.1 and 3.2) and real-world data analysis (Section 3.3). In the studies, without loss of generality, we shall use the event sequence data as an example and follow the procedure presented in Reference[11] to build the Cox PH model for the extracted time-to-failure data. The time-dependent covariates incorporated in the Cox PH model for the event sequence are different event types, which are represented by step functions—the covariate value is zero before the occurrence time and one after that. Readers can refer to Reference[11] for details about the Cox PH model fitting from event sequence data.

In numerical case studies, to generate the simulated data, the Weibull distribution[28], widely used in reliability analysis, is used as the baseline part of the Cox model. When only the scale parameter of the Weibull distribution is changed, we can have a hazard function proportional to the old hazard function—the PH change. If the shape parameter is changed, the resulted hazard will not be proportional to the original one—the non-PH change. Both cases can be addressed with the proposed technique.

In the numerical studies, two event types (covariates) *A* and *B* and their interactions are assumed to be significant in the Cox PH model. For a single hypothetical failure event sequence, we generated $N = 1000$ time intervals. Event type *A* or *B* may occur at most once in each time interval. Some details of the simulation will now be discussed. The distributions used to generate the simulated data and the corresponding parameters are summarized in Table I.

1. For each time interval, the occurrences of events *A* and *B* are assumed to be independent of one another. We also assumed that *A* and *B* occur within 50% and 40% of all the time intervals respectively, that is the occurrence of events *A* and *B* in each time interval independently follow a Bernoulli distribution with $p = 0.5$ and 0.4, respectively.
2. For those time intervals during which events *A* and/or *B* occur, we generated the occurrence times of the corresponding event according to a specified distribution. Without loss of generality, for *A*, the assumed distribution of the occurrence time was log normal with $\mu_A = 2$, $\sigma_A = 1$; for *B*, it was an exponential distribution with $\lambda_B = 5$. In total, $N = 1000$ sets of time-dependent covariates (events *A* and *B*) are generated.
3. The shape parameter and scale parameter of the baseline Weibull distribution are denoted as *a* and *b*, respectively, and its hazard function is $h_0(t) = a^{-b}bt^{b-1}$. Three sets of Weibull parameters are used in the studies and they are listed in the last three rows of Table I.
4. Assuming that the coefficient vector $\boldsymbol{\beta}$ in the Cox PH model is known, we can follow the method in Reference[29] to generate the failure times (the length of time intervals). In the study, we assume that the time-dependent covariates of events *A* and *B* as well as their interactions are significant.
5. The censoring percentage is around 5%, which means about 5% of the generated failure times are censored. The censoring distribution is the uniform distribution $U[0, h]$, where *h* is the upper limit and its value should be set to obtain a 5% censoring percentage.

Table I. The distributions and corresponding parameters used in the studies

Term	Description	Distribution	Parameter	
Event *A*	Occurs or not	Bernoulli	$p_A = 0.5$	
	Occurrence time	Log-normal	$\mu_A = 2$	$\sigma_A = 1$
Event *B*	Occurs or not	Bernoulli	$p_B = 0.4$	
	Occurrence time	Exponential	$\lambda_B = 5$	
Baseline	Set I	Weibull	$a = 100$	$b = 3$
	Set II	Weibull	$a = 80$	$b = 2$
	Set III	Weibull	$a = 100$	$b = 2$

Copyright © 2010 John Wiley & Sons, Ltd.

Table II. The estimated power values in (a) Cases I and II when $\boldsymbol{\beta}_2 = c \cdot \boldsymbol{\beta}_1$ and (b) in Case I when $\boldsymbol{\beta}_{2,i} = c \cdot \boldsymbol{\beta}_{1,i}$, $i = A$, B or $A \times B$

(a) Case	c	Estimated power	Estimated local power		
			A	B	$A \times B$
I	0.75	0.312	**0.207**	**0.283**	**0.236**
	0.50	0.905	**0.419**	**0.569**	**0.395**
	0.25	0.999	**0.710**	**0.872**	**0.523**
II	0.75	0.330	0.213	0.279	0.237
	0.50	0.895	0.444	0.580	0.377
	0.25	0.999	0.688	0.827	0.549
(b) Covariate					
A	0.75	0.089	**0.215**	0.120	0.174
	0.5	0.282	**0.432**	0.152	0.170
	0.25	0.560	**0.684**	0.107	0.171
B	0.75	0.131	0.121	**0.270**	0.167
	0.5	0.443	0.126	**0.576**	0.199
	0.25	0.813	0.128	**0.828**	0.174
$A \times B$	0.75	0.089	0.131	0.154	**0.229**
	0.5	0.197	0.14	0.151	**0.371**
	0.25	0.361	0.137	0.143	**0.532**

In the following studies, two cases will be investigated to show the effectiveness of the proposed technique for measuring the deviations in the Cox model: (i) two different Cox models M_1 and M_2 for D_1 and D_2 (Section 3.1) and (ii) the stratified model M^* for the data set $D = \{D_1, D_2\}$ (Section 3.2).

Note that it is typical that the Cox PH model is fitted to the data set containing 50–100 subjects in survival analysis[23] and the tests for the model parameter $\boldsymbol{\beta}$, such as the Wald test, score test, and likelihood ratio test, are performed based on the asymptotic properties of estimator $\hat{\boldsymbol{\beta}}$. In industrial applications, it is very common that the real-world data set contains more data points. Therefore, the results presented in this paper can be applied to the data set with sample size 50–100 or more in practice.

3.1. *Different Cox models for two data sets*

In this subsection, two different Cox models M_1 and M_2 are fitted to the simulated data sets D_1 and D_2. When we generate the data, we assume $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$. Two cases are considered for the baseline parts: (I) the two data sets share the same baseline parameters and (II) the two data sets have different baseline parts. Following the procedure illustrated in Figure 1, we will test the hypothesis $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ vs $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ and then find the $(1-\alpha) \times 100\%$ pointwise confidence region for the times when $S_{01}(t) = S_{02}(t)$. The significance level for both tests is $\alpha = 0.05$.

We assume the coefficient vector $\boldsymbol{\beta}_1 = [0.75, 1.0, 0.8]^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = c \cdot \boldsymbol{\beta}_1$ in both cases, where $c = 0.75$, 0.5, and 0.25, respectively. The Weibull parameters in Case I for both data sets are $a = 100$ and $b = 3$. In Case II, the Weibull parameters D_1 are $a = 100$ and $b = 3$, but $a = 80$ and $b = 2$ are assumed for D_2.

First, we generate 1000 pairs of D_1 and D_2 and calculate the statistic values in Equation (3). The estimated type-I error using the simulated data sets are 0.054 and 0.053, respectively. Both results are close to the assumed confidence level $\alpha = 0.05$. The estimated global power values for the hypothesis testing of $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ vs $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ in both cases are summarized in Table II(a). The power results are calculated by counting the number of pairs for which the change in $\boldsymbol{\beta}$ is correctly detected by the proposed procedure. The estimated local powers for $H_0 : \boldsymbol{\beta}_{1,i} = \boldsymbol{\beta}_{2,i}$ vs $H_1 : \boldsymbol{\beta}_{1,i} \neq \boldsymbol{\beta}_{2,i}$, $i = A$, B or $A \times B$, are also listed there. Based on the results, we can see that the power values are pretty similar in these two cases, that is whether the baseline parts are the same or not does not impact the power values for testing the equality of coefficient vectors. As the whole coefficient vector is changed, the global power value is greater than the corresponding local power values.

We also calculated the global and local powers in Case I when only the effect of one covariate changes its value: $\boldsymbol{\beta}_{2,i} = c \cdot \boldsymbol{\beta}_{1,i}$, $i = A$, B or $A \times B$. The results can be found in Table II(b). The results tell us that the local test corresponding to the changed coefficient provides a greater power value than the global test, and the local powers are very close to those bold values in Table II(a).

Next we find the $(1-\alpha) \times 100\%$ pointwise confidence region for the times when $S_{01}(t) = S_{02}(t)$ in two cases. An example for a pair of D_1 and D_2 can be found in Figure 2, which is the plot of standardized statistic $\Delta\hat{H}_0(t_0) / \sqrt{\mathrm{var}[\Delta\hat{H}_0(t_0)]}$ against t_0 when the two data sets share the same baseline part. Intuitively, there should be no points outside of the 'control limits' $z_{\alpha/2}$ and $z_{-\alpha/2}$ since the two data sets are assumed to have the same baseline survival curve. Both Figure 2(a) with the sample size $N = 200$ and Figure 2(b) with $N = 500$ show that there are no instances when $S_{01}(t) \neq S_{02}(t)$ since the whole curve of the standardized difference in the cumulative hazards falls within the limits.

Copyright © 2010 John Wiley & Sons, Ltd.

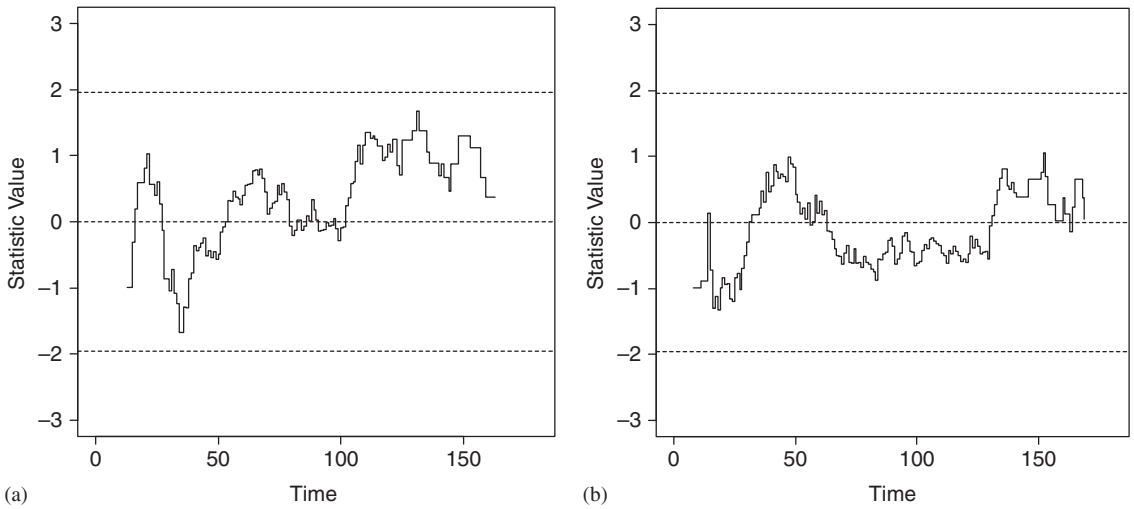

Figure 2. Standardized difference in the baseline cumulative hazards in Case I: (a) $N=200$ and (b) $N=500$

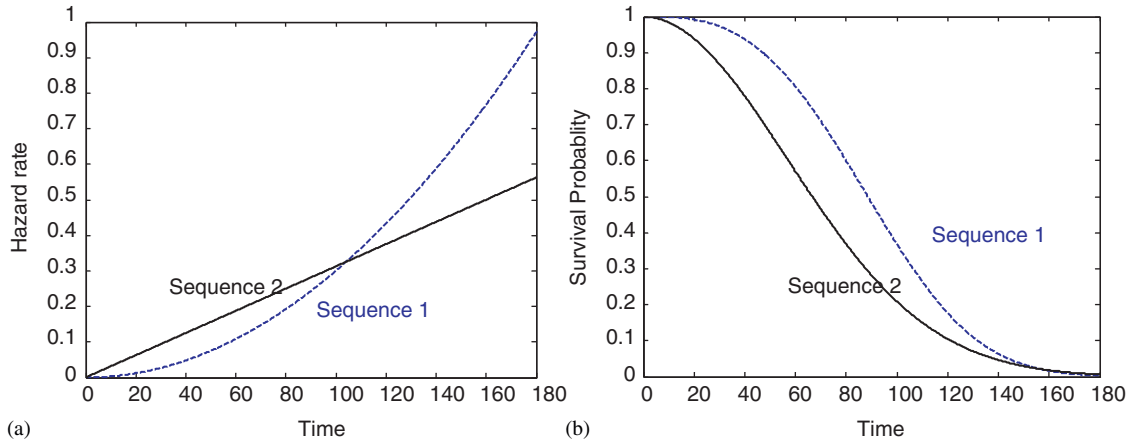

Figure 3. The baseline hazard function and survival function for Case II: (a) hazard rate function and (b) survival function

For the case that the baseline parts are different, the baseline hazard function and survival function of two Weibull distributions are plotted in Figure 3.

An example is plotted in Figure 4 for the standardized difference against t_0. From this figure we can obtain a 95% pointwise confidence region for the times when $S_{01}(t) \neq S_{02}(t)$ by ensuring that all those points fall outside of the 'control limits'. When the sample size N is 200, the region is given by $C_{200}=\{t_0|t_0 \in [16.86, 116.62]\}$. The interpretation of the result is that, at each fixed time point in this range, the two baseline survival functions are statistically different, and we can conclude that D_1 has higher early survival than D_2. A larger sample size $N=500$ can provide a more accurate estimate of the baseline cumulative function with smaller variance and the region in this case is $C_{500}=\{t_0|t_0 \in [3.93, 139.71]\}$, which is consistent with Figure 3(b).

3.2. The stratified Cox model for two data sets

In this subsection, the data sets are generated under the same coefficient vector $\boldsymbol{\beta}=[0.75, 1.0, 0.8]^{\mathrm{T}}$ but with different baseline parts. The data set D_1 has Weibull parameters $a=100$, $b=3$. Two cases are considered for D_2 here: (III) $a=80$, $b=2$ and (IV) $a=100$, $b=2$. Note that the Weibull parameters in Case III are the same as Case II in Section 3.1.

Let us begin with Case III with a pair of D_1 and D_2. When $N=200$, the value of statistic X^2 is 1.77, which means that we cannot reject the null hypothesis $H_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Now we fit a stratified Cox model to the data and then a 95% confidence region for the times when $S_{01}(t) \neq S_{02}(t)$ is given by $C_{200}=\{t_0|t_0 \in [7.57, 118.80]\}$. For the same data, $N=500$ gives the value of statistic X^2 as 4.71, the 95% confidence region is $C_{500}=\{t_0|t_0 \in [6.69, 125.10]\}$. The plot of the standardized difference in the baseline cumulative function $\Delta\hat{H}_0^*(t_0)/\sqrt{\mathrm{var}[\Delta\hat{H}_0^*(t_0)]}$ is illustrated in Figure 5.

For Case IV, the baseline hazard function and survival function for the two data sets are found in Figure 6. Notice that there is an intersection of the two baseline survival function in Figure 6(b). When $N=200$, the value of statistic X^2 is 2.37, and the 95% confidence region for the times when $S_{01}(t) \neq S_{02}(t)$ is $C_{200}=\{t_0|t_0 \in [11.17, 68.64] \cup [118.41, 131.81] \cup [133.12, 138.78]\}$.

Copyright © 2010 John Wiley & Sons, Ltd.

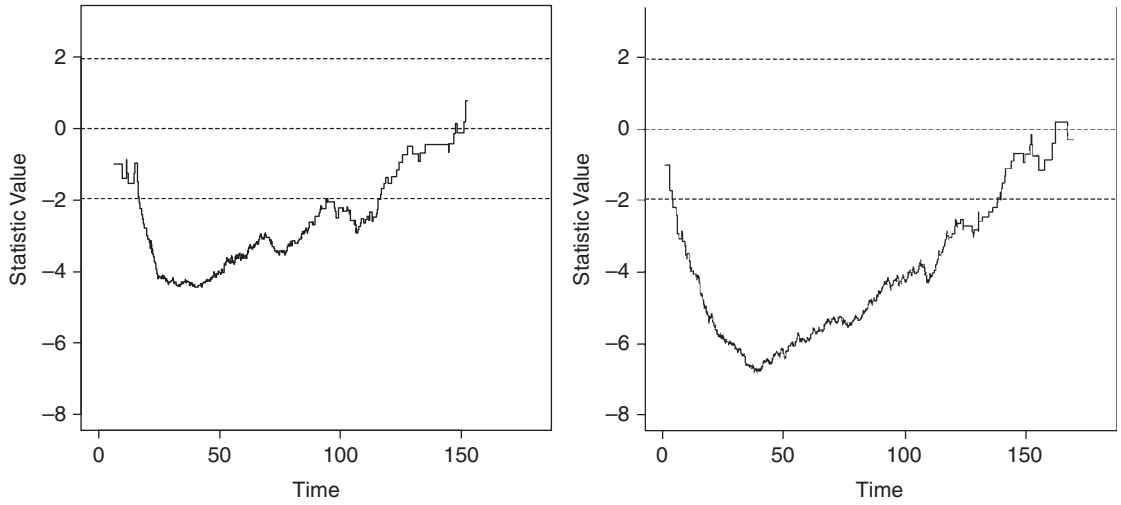

Figure 4. Standardized difference in the baseline cumulative hazards in Case II: (a) $N=200$ and (b) $N=500$

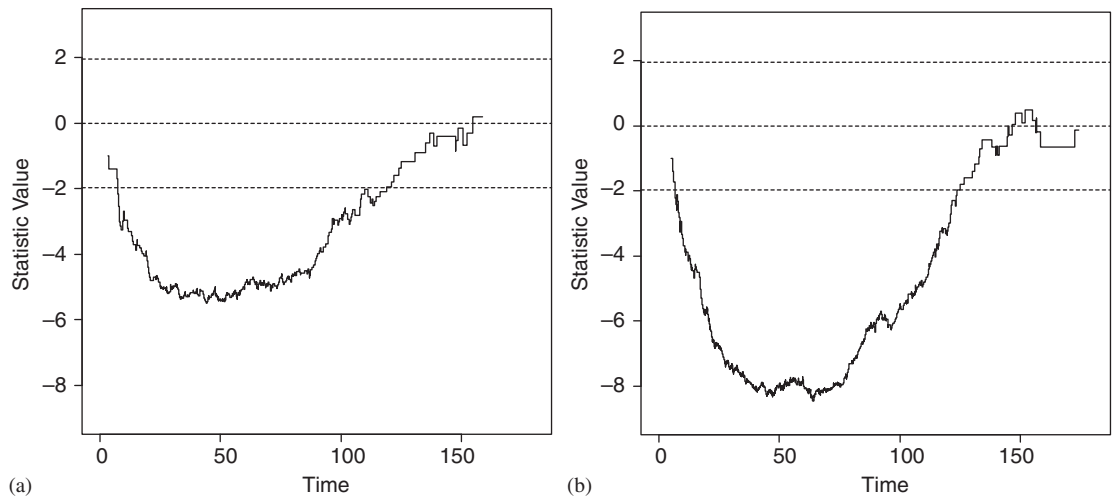

Figure 5. Standardized difference in the baseline cumulative hazards in Case III: (a) $N=200$ (b) $N=500$

With N=500, we have the value of statistic X^2 as 4.61, the confidence region is $C_{500}=\{t_0|t_0\in[8.43,91.28]\cup[118.55,163.98]\}$. The plot of the statistic $\Delta\hat{H}_0^*(t_0)/\sqrt{\text{var}[\Delta\hat{H}_0^*(t_0)]}$ is found in Figure 7. Again, the results are consistent with Figure 6(b).

3.3. *Real event log analysis*

In this subsection, we use real log files from a computer tomography (CT) machine to show the procedure for measuring the deviations in the Cox PH model. CT log files record the occurrences of failure events and covariates and thus time-to-failure data can be extracted from event logs. Log file D_1 has 106 failure events and 13 censored time intervals, whereas there are 68 failure events and 50 censored time intervals in D_2. Here censoring results from the preventive maintenance activities that would cause replacement of the component of interests before its failure. Preventive maintenance activities are either scheduled on a regular time basis or evoked by failures of other components/subsystems in the machine. Variable selection procedure was implemented[11] and finally two Cox models were fitted to the data. The results are listed in Table III.

Notice that three covariates ($Z_A(t)$, $Z_B(t)$, and $Z_C(t)$) and all their interactions ($Z_A(t)\times Z_B(t)$, $Z_A(t)\times Z_C(t)$, $Z_B(t)\times Z_C(t)$, and $Z_A(t)\times Z_B(t)\times Z_C(t)$) are incorporated in the two models. Using AIC in a stepwise algorithm ('*stepAIC*' in *R*), we found that all the terms should be kept in the final models M_1 and M_2. The results for testing the PH assumption are listed in Table IV. All p-values are greater than the 5% significance level, and thus we could conclude that the PH assumption holds for all covariates in two models.

The statistic value for the test $H_0:\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$ vs $H_1:\boldsymbol{\beta}_1\neq\boldsymbol{\beta}_2$ is 21.43, which tells us that there is a significant change in the effects of the covariates. Indeed, we can see that the coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are different from Table III.

Copyright © 2010 John Wiley & Sons, Ltd.

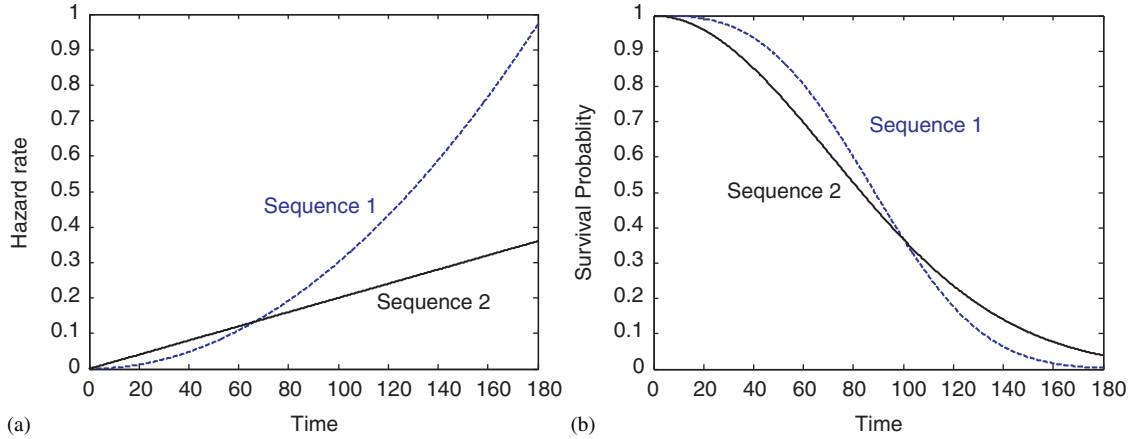

Figure 6. The baseline hazard function and survival function for Case IV: (a) Hazard rate function (b) Survival function

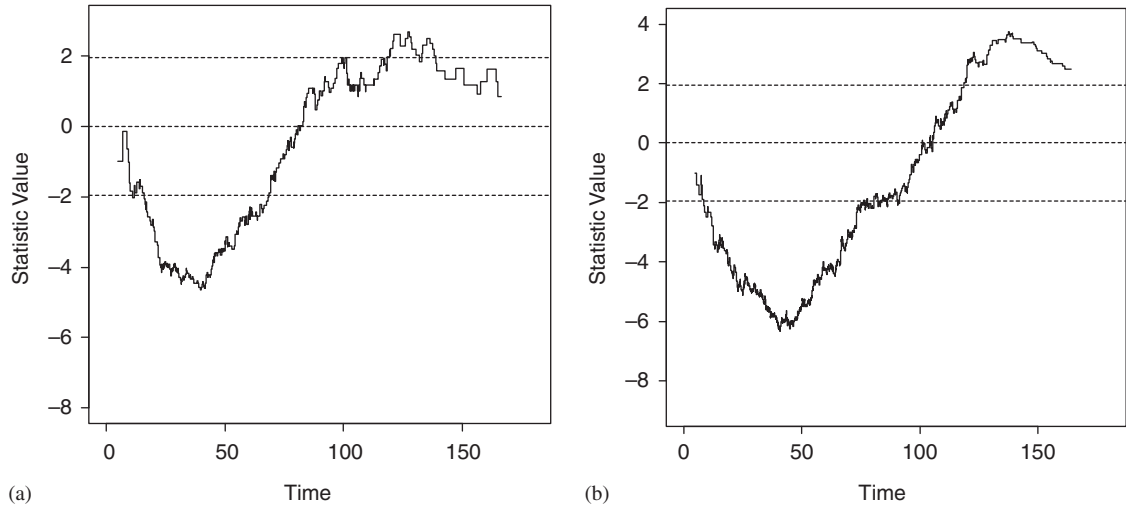

Figure 7. Standardized difference in the baseline cumulative hazards in Case IV: (a) $N=200$ (b) $N=500$

Table III. The Cox PH models for two event logs

	Cox model 1			Cox model 2		
Term	**b**	se	*p*-value	**b**	se	*p*-value
Z_A	0.507	0.343	0.140	0.524	0.470	0.260
Z_B	0.719	0.364	0.0480	−0.061	0.472	0.900
Z_C	0.272	0.406	0.500	0.176	0.513	0.730
Z_{AB}	−0.348	0.544	0.520	0.221	0.706	0.750
Z_{AC}	2.426	0.651	0.002	−0.196	0.837	0.820
Z_{BC}	0.738	0.639	0.250	0.658	0.794	0.410
Z_{ABC}	−3.158	0.969	0.001	−2.396	1.164	0.039

Figures 8(a) and (b) illustrate that there is some difference between the two baseline survival functions $S_{01}(t)$ and $S_{02}(t)$. However, the plot of standardized difference in the baseline cumulative functions in Figure 8(c) shows that there are no instances when $S_{01}(t) \neq S_{02}(t)$ since the whole curve falls within the limits. Although the curve in Figure 8(c) moves to the upper control limit very quickly and closely, there is statistically no significant change in the baseline survival functions. In summary, the results reveal that the usage patterns represented by the model coefficients are different but the inherent machine performance remains the same.

Copyright © 2010 John Wiley & Sons, Ltd.

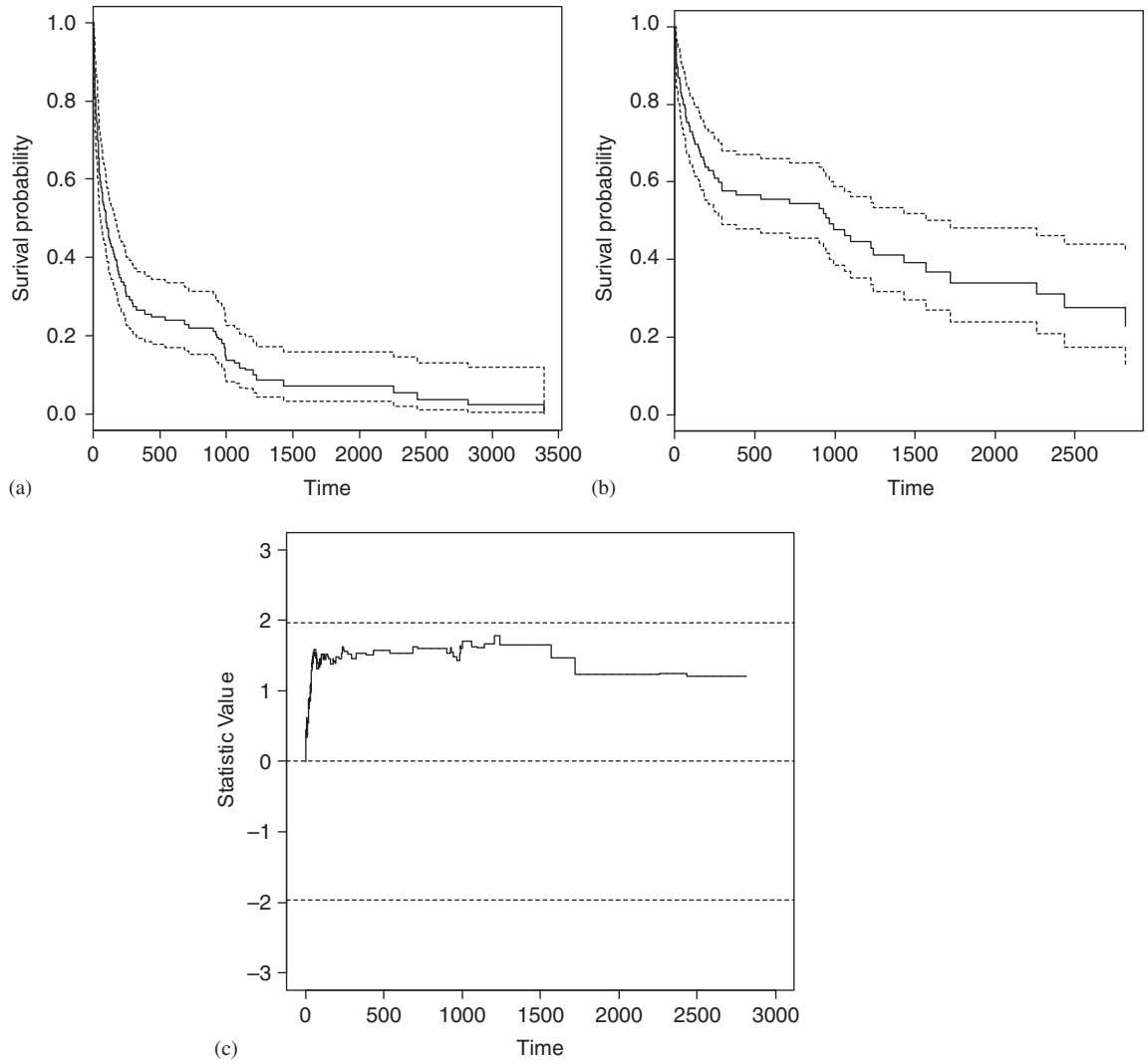

Figure 8. Comparison of the baseline survival functions for event logs: (a) D_1; (b) D_2; and (c) standardized difference

Table IV. Test results for the proportional hazards assumption

	Model 1		Model 2	
Term	Chi-square	*p*-value	Chi-square	*p*-value
Z_A	0.526	0.468	0.025	0.873
Z_B	0.602	0.438	0.960	0.327
Z_C	1.159	0.282	0.881	0.348
Z_{AB}	0.749	0.387	0.356	0.551
Z_{AC}	0.468	0.494	3.167	0.075
Z_{BC}	0.010	0.919	0.225	0.635
Z_{ABC}	1.912	0.167	0.125	0.724
GLOBAL	13.728	0.056	8.206	0.315

4. Concluding remarks

In this paper, we develop the methods to quantify the differences between the Cox PH models, which can relate the occurrence of failure events to the covariates and is widely used in reliability modeling. In this study, a systematic methodology is developed to compare the covariate coefficients and the baseline survival (cumulative hazard) functions for different Cox PH models. The effectiveness of the developed technique is illustrated through numerical case studies and real-world data analysis with log files

Copyright © 2010 John Wiley & Sons, Ltd.

collected from CT scanners. The developed procedure can be used to detect the changes and thus update the reliability model in practice where the Cox PH model fits the reliability data.

A very interesting open issue is, when $\beta_1 \neq \beta_2$, to construct the simultaneous confidence band for the difference between the baseline survival (cumulative hazard) functions of the two Cox models if we are interested in testing if the two baseline survival (cumulative hazard) functions are equal over the entire time period. Lin *et al.*[30] proposed a method to construct a simultaneous confidence band for survival curves under the PH model through a simulation procedure. The results along this direction based on this result will be reported in a future work.

Acknowledgements

The financial support for this study was provided by GE Healthcare and the National Science Foundation grant #0757683.

References

1. Meeker WQ, Escobar LA. *Statistical Methods for Reliability Data*. Wiley: New York, 1998.
2. Bendell A, Walley M, Wightman DW, Wood LM. Proportional hazards modelling in reliability analysis—An application to brake discs on high speed trains. *Quality and Reliability Engineering International* 1986; **2**:45–52.
3. Bendell A, Wightman DW, Walker EV. Applying proportional hazards modeling in reliability. *Reliability Engineering and System Safety* 1991; **34**:35–53.
4. Kumar D, Klefsjö B. Proportional hazards model: An application to power supply cables of electric mine loaders. *International Journal of Reliability, Quality and Safety Engineering* 1994; **1**(3):337–352.
5. Leitao ALF, Newton DW. Proportional hazards modelling of aircraft cargo door complaints. *Quality and Reliability Engineering International* 1989; **5**:229–238.
6. Jardine AKS, Anderson PM, Mann DS. Application of the Weibull proportional hazard model to aircraft and marine engine failure data. *Quality and Reliability Engineering International* 1987; **3**:77–82.
7. Jardine AKS, Ralston P, Reid N, Stafford J. Proportional hazards analysis of diesel engine failure data. *Quality and Reliability Engineering International* 1989; **5**:207–216.
8. Love CE, Guo R. Application of Weibull proportional hazards modeling to bad-as-old failure data. *Quality and Reliability Engineering International* 1991; **7**:149–157.
9. Bendell A. Proportional hazards modeling in reliability assessment. *Reliability Engineering* 1985; **11**:175–183.
10. Dale CJ. Application of the proportional hazards model in the reliability field. *Reliability Engineering* 1985; **10**:1–14.
11. Li Z, Zhou S, Choubey S, Sievenpiper C. Failure event prediction using cox proportional hazard model driven by frequent failure signatures. *IIE Transactions on Quality and Reliability Engineering* 2007; **39**:303–315.
12. Banjevic D, Jardine AKS, Makis V, Ennis M. A control-limit policy and software for condition-based maintenance optimization. *INFOR* 2001; **39**:32–50.
13. Makis V, Jardine AKS. Optimal replacement in the proportional hazards model. *INFOR* 1991; **30**:172–183.
14. Makis V, Jardine AKS. Computation of optimal policies in replacement models. *Journal of Mathematics Applied in Business and Industry* 1992; **3**:169–175.
15. Basseville M, Nikiforov IV. *Detection of Abrupt Changes*: *Theory and Application*. Prentice-Hall: Upper Saddle River, NJ, 1993.
16. Brodsky B, Darkovsky B. *Nonparametric Methods in Change Point Problems*. Kluwer Academic: Netherlands, 1993.
17. Page E. Continuous inspection schemes. *Biometrika* 1954; **41**:100–115.
18. Page E. A test for a change in a parameter occurring at an unknown point. *Biometrika* 1955; **42**:523–527.
19. Page E. On problem in which a change in a parameter occurs at an unknown point. *Biometrika* 1957; **44**:248–252.
20. Benveniste A, Metivier M, Prioutet P. *Adaptive Algorithms and Stochastic Approximations*. Series on Applications of Mathematics, Balakrishnan AV, Karatzas I, Yor M (eds.). Springer: New York, 1990.
21. Ganti V, Gehrke J, Ramakrishnan R. Mining data streams under block evolution. *ACM SIGKDD Explorations* 2002; **3**(2):1–10.
22. Ganti V, Gehrke J, Ramakrishnan R, Loh W-Y. A framework for measuring differences in data characteristics. *Journal of Computer and System Sciences* 2002; **64**:542–578.
23. Klein JP, Moeschberger ML. *Survival Analysis*: *Techniques for Censored and Truncated Data*. Springer: New York, 2003.
24. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: New York, 1993.
25. Zhang M-J, Klein JP. Confidence bands for the difference of two survival curves under proportional hazards model. *Lifetime Data Analysis* 2001; **7**:243–254.
26. Klein JP, Zhang M-J. Confidence regions for the equality of two survival curves. *Technical Report*, Division of Biostatistics, Medical College of Wisconsin, 1996.
27. Glidden DV, Vittinghoff E. Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine* 2004; **23**:369–388.
28. Wasserman GS. *Reliability Verification, Testing and Analysis in Engineering Design*. Marcel Dekker: New York, 2003.
29. Leemis LM, Shih L-H, Reynertson K. Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Statistics and Probability Letters* 1990; **10**(4):335–339.
30. Lin DY, Fleming TR, Wei LJ. Confidence bands for survival curves under the proportional hazards model. *Biometrika* 1994; **81**:73–81.

Appendix A

A.1. Proof of Lemma 1

As M_1, M_2 are two Cox PH models induced by the two reliability data sets D_1 and D_2, we have (Theorem VII.2.2 in Reference[24]),

$$(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) \xrightarrow{D} N(0, \boldsymbol{\Sigma}_j), \quad j = 1, 2, \tag{A1}$$

Copyright © 2010 John Wiley & Sons, Ltd.

where $\boldsymbol{\beta}_j$ is the true parameter vector for the model j. Under null hypothesis $H_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, since the random variables $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are independent of each other, this yields

$$(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) \xrightarrow{D} N(0, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2). \tag{A2}$$

As the covariance matrix $(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$ can be estimated consistently by $(\hat{\boldsymbol{\Sigma}}_1 + \hat{\boldsymbol{\Sigma}}_2)$, this leads to the test statistic given by Equation (3)

$$X^2 = (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)^{\mathrm{T}} (\hat{\boldsymbol{\Sigma}}_1 + \hat{\boldsymbol{\Sigma}}_2)^{-1} (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2), \tag{A3}$$

which has a chi-squared distribution with p degrees of freedom and p is the length of $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$ (refer to the large-sample tests based on likelihood theory in Reference[23]).

Under alternative H_1: $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, the statistic (A3) follows a non-central chi-squared distribution with the estimated non-central parameter $\hat{\tau}$, thus the power of the hypothesis testing given the significance level α can be estimated as

$$1 - \hat{\beta} = Pr(X^2(\hat{\tau}) > \chi^2_{p,1-\alpha}), \tag{A4}$$

where $X^2(\hat{\tau})$ is a *non-central* χ^2 random variable with p degrees of freedom and the estimated non-centrality parameter is $\hat{\tau} = (\mathbf{b}_1 - \mathbf{b}_2)^{\mathrm{T}} (\mathbf{S}_1 + \mathbf{S}_2)^{-1} (\mathbf{b}_1 - \mathbf{b}_2)$, where $\mathbf{b}_j$ and $\mathbf{S}_j$, $j = 1,2$, are the estimates of the coefficient vector and covariance matrix from the data. □

Authors' biographies

Zhiguo Li is a Research Staff Member with the Xerox Research Center Webster in NY. He obtained his BE and ME in Automotive Engineering at Tsinghua University in 1999 and 2002, respectively, and obtained his MS in Statistics and PhD in Industrial Engineering at the University of Wisconsin-Madison in 2007. His research interests focus on post-sales service, machine diagnostics and prognostics, reliability modeling and prediction of complex engineering systems. Dr Li is a member of IIE, INFORMS and Sigma Xi.

Shiyu Zhou is an associate professor in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. He got his BS and MS in Mechanical Engineering from the University of Science and Technology of China in 1993 and 1996, respectively, and got his Masters in Industrial Engineering and PhD in Mechanical Engineering from the University of Michigan in 2000. Dr Zhou's research interests are the in-process quality and productivity improvement methodologies by integrating statistics, system and control theory, and engineering knowledge. His research is sponsored by the National Science Foundation, Department of Energy, Department of Commerce, and industries. He is a recipient of the CAREER Award from the National Science Foundation and the Best Application Paper award from IIE Transactions in 2006. Dr Zhou is a member of IIE, INFORMS, ASME, and SME.

Crispian Sievenpiper holds Masters degrees in Computer Science and Statistics and a doctorate in Systems and Industrial Engineering. His research interests include enhanced medical scanner service delivery, global development, and deployment of new services, and sustainable improvement of machine utilization. He currently works as a Research Architect for GE Healthcare in Global Service.

Suresh Choubey is currently a Chief Scientist and Manager of Informatics at GE Healthcare CSE—Applied Science Lab. He is MBB certified. He received MS and PhD degrees in Computer Science from the Center for Advanced Computer Studies, University of Louisiana at Lafayette. He has authored/co-authored several journal (Pattern Recognition, JASIS, IIE) and conference papers. His areas of research interests are Machine Early Health, Patient Early Health, Asset Intelligence, Home Healthcare Informatics and Service Dispatch Prediction and holds multiple patents.

Copyright © 2010 John Wiley & Sons, Ltd.