Testing for Covariate Effect in the Cox Proportional Hazards Regression Model

Karthik Devarajan
Division of Population Science
Fox Chase Cancer Center
Philadelphia, PA 19111
karthik.devarajan@fccc.edu

Nader Ebrahimi
Division of Statistics
Northern Illinois University
DeKalb, IL 60115
nader@math.niu.edu

Abstract

This paper presents methods for testing covariate effect in the Cox Proportional Hazards (PH) Model based on Kullback-Leibler (KL) divergence and Renyi's information measure. Renyi's measure is referred to as the information divergence of order γ ($\gamma \neq 1$) between two distributions. In the limiting case $\gamma \rightarrow 1$, Renyi's measure becomes Kullback-Leibler (KL) divergence. In our case, the distributions correspond to the baseline and one possibly due to a covariate effect. Our proposed statistics are simple transformations of the parameter vector in the Cox PH model, and are compared with the Wald, likelihood ratio (LR) and Score tests that are widely used in practice. Finally, the methods are illustrated using two real-life data sets.

1 Introduction

The Cox proportional hazards (PH) model (Cox, 1972) offers a method for exploring the association of covariates with the survival time outcome variable often seen in medical and engineering studies. The observed data consist of independent observations on the triple $(X, \delta, \mathbf{z})$, where X is the minimum of a survival and censoring time pair (T, C) , $\delta = I(T \leq C)$ is the indicator of the event that a death has been observed and $\mathbf{z} = (z_1, z_2, \dots, z_p)'$ is a p -vector of covariates. The non-negative random variables T and C denote the survival and censoring times respectively, and are assumed to be independent conditional on the covariates.

The Cox PH model is perhaps most widely used in censored survival data analysis. For a given vector of observed covariates $\mathbf{z}$, the hazard function of T is modeled as

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}) \quad (1.1)$$

where $\lambda_0(t)$ is an unspecified function of time and $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$ is a row vector of unknown regression parameters. Alternate forms of the Cox PH model include

$$S(t|\mathbf{z}) = (S_0(t))^{\exp(\boldsymbol{\beta}'\mathbf{z})} \quad (1.2)$$

and

$$\Lambda(t|\mathbf{z}) = \Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{z}) \quad (1.3)$$

where $S_0(t)$ and $\Lambda_0(t)$ are the baseline survival and cumulative hazard functions, respectively. In (1.2) and (1.3), the survival and cumulative hazard functions of T are modeled respectively.

Checking for covariate effect in (1.1) is a matter of great interest. Several procedures have been described that include the Wald, LR and Score tests among others (see Klein and Moeschberger (2003)). In this paper, we develop tests for covariate effect in the Cox PH model (1.1) that are easy to compute as well as perform better than these statistical tests. Our tests are based on KL divergence as well as Renyi's divergence. In Section 2, we derive expressions for our test statistics based on these measures. In Section 3, we discuss the implementation of the proposed tests and give two illustrative examples. Section 4 compares the performance of the proposed tests with existing tests in terms of power and consistency via extensive simulation studies. Section 5 provides some concluding remarks.

2 Testing for Covariate Effect in the Cox PH Model

In this case, our null hypothesis of interest is non-parametric and is given by

$$H_0 : \Lambda(t|\mathbf{z}) = \Lambda_0(t). \quad (2.1)$$

The alternative is given by

$$H_1 : \Lambda(t|\mathbf{z}) = \exp(\boldsymbol{\beta}'\mathbf{z})\{\Lambda_0(t)\}, \quad (2.2)$$

where the terms are as defined earlier. We observe that the null hypothesis corresponds to $\boldsymbol{\beta} = \mathbf{0}$ indicating no covariate effect. We can now re-write our hypotheses as

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ against } H_1 : \boldsymbol{\beta} \neq \mathbf{0}. \quad (2.3)$$

We use KL divergence (see Kullback (1978), Ebrahimi et al. (1994) and Ebrahimi and Soofi (2004) for more details) to discriminate between the null and alternative hypotheses.

In general, the KL divergence between two distribution functions F and F_0 is

$$I(F : F_0) = \int_0^\infty f(x) \left\{ \log \frac{f(x)}{f_0(x)} \right\} dx, \quad (2.4)$$

where f and f_0 are the density functions corresponding to F and F_0 respectively. $I(F : F_0)$ is also known as the directed divergence and measures the discrepancy between F and F_0 in the direction of F . It is well known that $I(F : F_0) \geq 0$ and the equality holds if and only if $F = F_0$. Similarly, one can also define the directed divergence

$$I(F_0 : F) = \int_0^\infty f_0(x) \left\{ \log \frac{f_0(x)}{f(x)} \right\} dx \quad (2.5)$$

which measures the discrepancy between F and F_0 in the direction of F_0 . Combining (2.4) and (2.5), we can derive the divergence measure given by

$$J(F : F_0) = I(F : F_0) + I(F_0 : F) \quad (2.6)$$

measuring the difficulty of discriminating between F_0 and F .

Using (1.2), we can write the density functions under H_1 and H_0 as

$$f(t|\mathbf{z}) = \lambda_F(t|\mathbf{z})S(t|\mathbf{z}) \quad (2.7)$$

and

$$f_0(t) = \lambda_0(t)S_0(t), \quad (2.8)$$

respectively, where λ_F is the hazard function under H_1 . Using the above equations and (2.4) and (2.5), the directed divergence $I_1(F : F_0|\mathbf{z})$ given covariate vector $\mathbf{z}$, is given by

$$\begin{aligned} I_1(F : F_0|\mathbf{z}) &= \int_0^\infty f(t|\mathbf{z}) \log \left\{ \frac{f(t|\mathbf{z})}{f_0(t)} \right\} dt \\ &= \int_0^\infty f(t|\mathbf{z}) \log \left\{ \frac{\lambda_F(t|\mathbf{z})S(t|\mathbf{z})}{\lambda_0(t)S_0(t)} \right\} dt \\ &= \exp(-\boldsymbol{\beta}'\mathbf{z}) + \boldsymbol{\beta}'\mathbf{z} - 1, \end{aligned} \quad (2.9)$$

and the directed divergence $I_2(F_0 : F|\mathbf{z})$ given covariate vector $\mathbf{z}$, is given by

$$\begin{aligned} I_2(F_0 : F|\mathbf{z}) &= \int_0^\infty f_0(t) \log \left\{ \frac{f_0(t)}{f(t|\mathbf{z})} \right\} dt \\ &= \int_0^\infty f_0(t) \log \left\{ \frac{\lambda_0(t)S_0(t)}{\lambda_F(t|\mathbf{z})S(t|\mathbf{z})} \right\} dt \\ &= \exp(\boldsymbol{\beta}'\mathbf{z}) - \boldsymbol{\beta}'\mathbf{z} - 1. \end{aligned} \quad (2.10)$$

The divergence measure $J(F : F_0|\mathbf{z})$ given the covariate vector $\mathbf{z}$ is given by

$$\begin{aligned} J(F : F_0|\mathbf{z}) &= I_1(F : F_0|\mathbf{z}) + I_2(F_0 : F|\mathbf{z}) \\ &= \exp(\boldsymbol{\beta}'\mathbf{z}) + \exp(-\boldsymbol{\beta}'\mathbf{z}) - 2. \end{aligned} \quad (2.11)$$

Another information measure that one may use to discriminate between H_0 and H_1 is Renyi's measure. This measure generalizes KL divergence and is referred to as the information divergence of order γ between two distributions F and F_0 and corresponding densities f and f_0 . Conditional on the covariate vector $\mathbf{z}$, this measure is defined as

$$\begin{aligned} R_\gamma(F : F_0|\mathbf{z}) &= \frac{1}{\gamma - 1} \log \int \left(\frac{f(x|\mathbf{z})}{f_0(x|\mathbf{z})} \right)^{\gamma-1} f(x|\mathbf{z}) dx \\ &= \frac{1}{1 - \gamma} \log [(1 - \gamma) \exp(\boldsymbol{\beta}'\mathbf{z}) + \gamma] - \boldsymbol{\beta}'\mathbf{z}, \end{aligned} \quad (2.12)$$

where γ is the order of information and $\gamma \neq 1$. In the limit $\gamma \rightarrow 1$, (2.9) and (2.12) coincide. Information divergence of order γ , $R_\gamma(F : F_0|\mathbf{z})$, is symmetric only for the case $\gamma = \frac{1}{2}$, i.e., $R_{\frac{1}{2}}(F : F_0|\mathbf{z}) = R_{\frac{1}{2}}(F_0 : F|\mathbf{z})$. This is better known as the Bhattacharya distance. For our application, we consider several values of γ in the interval $(0, 1)$. For more details about Renyi's measure, the interested reader is referred to Asadi et al. (2006).

It should be emphasized that an important feature of our proposed measures I_1, I_2, J and R_γ is that all these information measures are invariant under any non-singular transformation on the failure time data.

Let us denote the observed data by $(X_j, \delta_j, \mathbf{z}_j), j = 1, \dots, n$, with $X_j = \min(T_j, C_j)$, the minimum of a failure and censoring time pair, $\delta_j = I(T_j \leq C_j)$ is the indicator of the event that a death has been observed and $\mathbf{z}_j = (z_{1j}, \dots, z_{pj})$ is a p -vector of covariates associated with an individual j . Since evaluations of (2.9)-(2.12) require complete knowledge of $\boldsymbol{\beta}$, then $I_1(F : F_0|\mathbf{z}), I_2(F_0 : F|\mathbf{z}), J(F : F_0|\mathbf{z})$ and $R_\gamma(F : F_0|\mathbf{z})$ are not operational. We operationalize these measures by replacing $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$, the maximum partial likelihood (MPL) estimate based on $(X_j, \delta_j, \mathbf{z}_j), j = 1, \dots, n$.

Using (2.9) - (2.12), our discrimination information statistics based on KL divergence, J and Renyi's measure are

$$\hat{I}_1(F : F_0|\mathbf{z}) = \sum_{i=1}^n \left[\exp(-\hat{\boldsymbol{\beta}}' \mathbf{z}_i) + \hat{\boldsymbol{\beta}}' \mathbf{z}_i - 1 \right], \quad (2.13)$$

$$\hat{I}_2(F_0 : F|\mathbf{z}) = \sum_{i=1}^n [\exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_i) - \hat{\boldsymbol{\beta}}' \mathbf{z}_i - 1], \quad (2.14)$$

$$\hat{J}(F : F_0|\mathbf{z}) = \sum_{i=1}^n \left[\exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_i) + \exp(-\hat{\boldsymbol{\beta}}' \mathbf{z}_i) - 2 \right], \quad (2.15)$$

and

$$\hat{R}_\gamma(F : F_0|\mathbf{z}) = \frac{1}{\gamma - 1} \sum_{i=1}^n \log \left[(1 - \gamma) \exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_i) + \gamma \right] - \hat{\boldsymbol{\beta}}' \mathbf{z}_i, \quad (2.16)$$

respectively. It should be noted that one can use (2.13)-(2.16) to perform the global hypothesis of $H_0 : \beta = \beta_0$. For global testing, one can easily verify that $\hat{\beta}$ must be replaced by $\hat{\beta} - \beta_0$ in equations (2.13)-(2.16). Often, one is interested in testing a hypothesis about a subset of the β 's, i.e., local tests. The hypothesis is then $H_0 : \beta_1 = \beta_{10}$, where $\beta = (\beta'_1, \beta'_2)'$. Here β_1 is a $q \times 1$ vector of the β 's of interest and β_2 is the vector of the remaining $p - q$ β 's. In this situation I_1, I_2, J and R_γ are obtained by simply replacing β with $\beta - \beta_0$ in equations (2.10)-(2.12), where $\beta_0 = (\beta'_{10}, \beta'_2)'$. Consequently, $\hat{I}_1, \hat{I}_2, \hat{J}$ and $\hat{R}_\gamma$ can be obtained by estimating $\beta - \beta_0$ with $\hat{\beta} - \hat{\beta}_0$, where $\hat{\beta}$ and $\hat{\beta}_0$ are the MPL estimators of β under H_1 and H_0 respectively.

The following theorem gives asymptotic distributions of our proposed test statistics.

Theorem 1 *$\hat{I}_1, \hat{I}_2, \hat{J}$ and $\hat{R}_\gamma$ are maximum likelihood estimators and are asymptotically normal with means I_1, I_2, J and R_γ respectively.*

Proof: As seen from (2.13)-(2.16), $\hat{I}_1, \hat{I}_2, \hat{J}$ and $\hat{R}_\gamma$ are simple transformations of the MPL estimator $\hat{\beta}$. Using the fact that $\hat{\beta}$ is asymptotically multivariate normal with mean β and the invariance property of maximum likelihood estimators, we see that the proposed statistics are in fact maximum likelihood estimators and are asymptotically normal with the means stated in the theorem. See Klein and Moeschberger (2003) for more details and for the variance-covariance matrix of $\hat{\beta}$. This concludes the proof.

Using Theorem 1, we compute the exact variances of the proposed statistics $\hat{I}_1, \hat{I}_2$ and $\hat{J}$ under the null hypothesis. For simplicity, we derive the variance estimators for the case of a single covariate z . The extension to multiple covariates is straightforward. Given that $\hat{\beta}$ is asymptotically normal with mean $\beta = 0$ and variance σ^2 under the null, and using the moment generating function of the normal random variable (i.e., if $\hat{\beta} \sim N(0, \sigma^2)$, then for z_i fixed, $i = 1, \dots, n$, $E[\exp(\hat{\beta}z_i)] = \exp(\sigma^2 z_i^2 / 2)$ and $E[\hat{\beta} \exp(\hat{\beta}z_i)] = z_i \sigma^2 \exp(\sigma^2 z_i^2 / 2)$, we

have,

$$\begin{aligned}
Var(\hat{I}_1) &= V \left(\sum_{i=1}^n [\exp(-\hat{\beta}' z_i) + \hat{\beta}' z_i - 1] \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n Cov \left([\exp(-\hat{\beta}' z_i) + \hat{\beta}' z_i - 1], [\exp(-\hat{\beta}' z_j) + \hat{\beta}' z_j - 1] \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \exp[(z_i + z_j)^2 \sigma^2 / 2] - \exp[(z_i^2 + z_j^2) \sigma^2 / 2] + \sigma^2 z_i z_j [1 - \exp(z_i^2 \sigma^2 / 2) - \exp(z_j^2 \sigma^2 / 2)].
\end{aligned}$$

Similarly,

$$\begin{aligned}
Var(\hat{I}_2) &= V \left(\sum_{i=1}^n [\exp(\hat{\beta}' z_i) - \hat{\beta}' z_i - 1] \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n Cov \left([\exp(\hat{\beta}' z_i) - \hat{\beta}' z_i - 1], [\exp(\hat{\beta}' z_j) - \hat{\beta}' z_j - 1] \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \exp[(z_i + z_j)^2 \sigma^2 / 2] - \exp[(z_i^2 + z_j^2) \sigma^2 / 2] + \sigma^2 z_i z_j [1 - \exp(z_i^2 \sigma^2 / 2) - \exp(z_j^2 \sigma^2 / 2)]
\end{aligned}$$

and

$$\begin{aligned}
Var(\hat{J}) &= V \left(\sum_{i=1}^n [\exp(\hat{\beta}' z_i) + \exp(-\hat{\beta}' z_i) - 2] \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n Cov \left([\exp(\hat{\beta}' z_i) + \exp(-\hat{\beta}' z_i) - 2], [\exp(\hat{\beta}' z_j) + \exp(-\hat{\beta}' z_j) - 2] \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n 2 \exp[(z_i + z_j)^2 \sigma^2 / 2] - 4 \exp[(z_i^2 + z_j^2) \sigma^2 / 2] + 2 \exp[(z_i - z_j)^2 \sigma^2 / 2].
\end{aligned}$$

For our measure based on Renyi's divergence $\hat{R}_\gamma$, the variance estimator is obtained using the delta method, simply by expanding $R_\gamma(\hat{\beta})$ around β .

$$\begin{aligned}
Var(\hat{R}_\gamma) &= V \left(c \sum_{i=1}^n \log [(1 - \gamma) \exp(\hat{\beta}' z_i) + \gamma] - \hat{\beta}' z_i \right) \\
&= c^2 \sum_{i=1}^n \sum_{j=1}^n Cov \left(\log [(1 - \gamma) \exp(\hat{\beta}' z_i) + \gamma] - \hat{\beta}' z_i, \log [(1 - \gamma) \exp(\hat{\beta}' z_j) + \gamma] - \hat{\beta}' z_j \right)
\end{aligned}$$

where $c = \frac{1}{\gamma - 1}$. Using the first three terms in the expansion and the fact that $E(\hat{\beta}^{2k+1}) = 0$

for k odd and $E(\hat{\beta}^{2k}) = \frac{(2k)!}{2^k k!} \sigma^{2k}$ for k even, we obtain a reasonable approximation to the variance estimator given by

$$Var(\hat{R}_\gamma) \approx \left(\frac{1}{\gamma - 1} \right)^2 \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \gamma^2 z_i^2 z_j^2 \sigma^4 + \frac{5}{12} (2\gamma^2 - \gamma)^2 z_i^3 z_j^3 \sigma^6.$$

3 Implementation of the Proposed Tests and Examples

3.1 Implementation

In this section, we describe the implementation of our proposed test statistics for testing (2.1) against (2.2). One can also follow similar steps for global testing as well as local testing. We apply the following steps to implement $\hat{I}_1(F : F_0|\mathbf{z})$, $\hat{I}_2(F_0 : F|\mathbf{z})$, $\hat{J}(F : F_0|\mathbf{z})$, and $\hat{R}_\gamma(F : F_0|\mathbf{z})$:

Step 1: Use partial likelihood procedure to estimate β (see Klein & Moeschberger (2003)). Denote the estimate by $\hat{\beta}$.

Step 2: Replace β by $\hat{\beta}$ in equations (2.10)-(2.12) to obtain (2.13) - (2.16).

Step 3: From Theorem 1, it is clear that $\hat{I}_1$, $\hat{I}_2$, $\hat{J}$ and $\hat{R}_\gamma$ in (2.13) - (2.16) have asymptotically normal distributions under H_0 . As shown in the previous section, the means and variances of $\hat{I}_1$, $\hat{I}_2$, $\hat{J}$ and $\hat{R}_\gamma$ can easily be obtained using Theorem 1. In fact, under H_0 , the means of all the proposed test statistics are zero.

Step 4: (a) Using $\hat{I}_i$, reject H_0 if $\frac{(\hat{I}_i)^2}{\hat{Var}(\hat{I}_i)} > \chi_p^2$, $i = 1, 2$, (b) Using $\hat{J}$, reject H_0 if $\frac{(\hat{J})^2}{\hat{Var}(\hat{J})} > \chi_p^2$, (c) Using $\hat{R}_\gamma$, reject H_0 if $\frac{(\hat{R}_\gamma)^2}{\hat{Var}(\hat{R}_\gamma)} > \chi_p^2$ where p is the number of parameters in the model. Equivalently, we can also use the normalized z -scores and the standard normal distribution to test our hypothesis for each proposed test.

Remark: From Theorem 1, it is clear that the tests specified above using $\hat{I}_1$, $\hat{I}_2$, $\hat{J}$ and $\hat{R}_\gamma$ are all consistent tests.

We implemented our methods using the R statistical system (R Development Core Team (2005)). The MPLE $\hat{\beta}$ was computed using the function *coxph* and the variance-covariance matrix of $\hat{\beta}$ under H_0 ($Var(\hat{\beta}_0)$) was computed using the function *vcov*. All computations for the simulations, including those pertaining to the variances of the proposed statistics, were programmed in R. we utilized the parallel implementation in the *snow* package in R using the High Performance Computing cluster at Fox Chase Cancer Center.

3.2 Examples

In this section we illustrate our approach using two real-life examples. For Renyi's measure, we only compute $\hat{R}_\gamma$ for $\gamma = .5$ which gives us the symmetric case simply for the purpose of illustration.

Example 1: The data consist of times to remission (in weeks) of 42 leukemia patients from Kalbfleisch & Prentice (2002). This data set is available in the R statistical system (R Development Core Team (2005)). The patients are randomized into two groups; a control group and the treatment group for which the drug 6-MP was administered. Each group consists of 21 patients. The covariate is group membership specified by the indicator variable z where $z = 0$ denotes the control group and $z = 1$ denotes the treatment group.

We are interested in testing for covariate effect i.e., if there is any significant treatment effect ($z = 1$) relative to the control group ($z = 0$). Here $p = 1$ and the null hypothesis is

$$H_0 : \beta = 0$$

corresponding to no covariate effect against the two-sided alternative

$$H_1 : \beta \neq 0.$$

The Cox PH model was fit to this data. The estimate of the regression coefficient $\hat{\beta}$ is 1.57 with a standard error of 0.412 and a significant p -value of 1.4e-04. The estimate of the relative risk is $e^{\hat{\beta}} = 4.81$ i.e., the risk is 4.81 times higher for a patient in the control group relative to that of a patient in the treatment group. Using the MPL estimate of β , $\hat{\beta} = 1.57$, our estimate of divergence given treatment group membership i.e., $z = 1$ is given by

$$\hat{J}(F : F_0|z = 1) = \exp(\hat{\beta}) + \exp(-\hat{\beta}) - 2 = 3.02.$$

Also, $\hat{J}(F : F_0|z = 0) = 0$. Consequently from (2.15), $\hat{J}(F : F_0|z) = 63.51$ and $\frac{\hat{J}^2}{\hat{Var}(\hat{J})} = 128.07$ which gives a p -value $< 1e-09$ using an asymptotic χ^2 distribution with 1 degree of freedom under the null hypothesis. Hence, we reject the null hypothesis of no covariate effect

at a significance level 0.05. Similarly from (2.13), $\hat{I}_1(F : F_0|z) = 47.14$ and $\frac{\hat{I}_1^2}{\hat{Var}(\hat{I}_1)} = 322.55$ which gives a p -value $< 1e-09$, and from (2.14), $\hat{I}_2(F_0 : F|z) = 16.34$ and $\frac{\hat{I}_2^2}{\hat{Var}(\hat{I}_2)} = 38.92$ which gives a p -value $< 1e-09$. Finally, using (2.16), Renyi's divergence of order γ for $\gamma = 0.5$ is $\hat{R}_{0.5}(F : F_0|z) = 11.83$ and $\frac{\hat{R}_{0.5}^2}{\hat{Var}(\hat{R}_{0.5})} = 106.57$ which also gives a p -value $< 10^{-9}$. In this example, all our proposed tests confirm that the treatment has a positive effect. In the case of Wald, LR and Score tests, the computed statistics (and p -values) were 14.5 (1.38e-04), 16.4 (5.26e-05) and 17.3 (3.28e-05), respectively. Hence, these standard tests confirm the results from our proposed statistics.

Example 2: We now consider the leukemia data described in Feigl & Zelen (1965). This data set consists of uncensored survival times of 33 patients who died from acute myelogenous leukemia, and is also available in the R statistical system (R Development Core Team (2005)). The patients were also factored into 2 groups according to the presence or absence of a morphologic characteristic of white blood cells (AG). The white blood cell count at the time of diagnosis was also measured for each patient (WBC).

We are interested in testing the effects of AG and WBC on the survival time of patients. Here, AG is a binary variable (present/absent) and WBC is a continuous covariate. Also, $p = 2$ and the null hypothesis is

$$H_0 : \beta = \mathbf{0}$$

corresponding to no effects due to AG and WBC against the two-sided alternative

$$H_1 : \beta \neq \mathbf{0}$$

that at least one covariate has an effect on survival time. We fit the Cox PH model to this data with AG (z_1), and log(WBC) (z_2), as our covariates, and performed a global test of the null hypothesis. AG is coded as 1 for present and 0 for absent. Based on this, the MPL estimates of the regression coefficients are $\hat{\beta}_1 = -1.069$ and $\hat{\beta}_2 = 0.368$ with corresponding p -values 0.013 and 0.017, respectively. That is, AG has a negative effect and WBC has a positive

effect on survival time, and both effects are statistically significant. Now, our estimate of the divergence measure given by (2.15) is $\hat{J}(F : F_0|\mathbf{z}) = 901190.8$. Consequently, $\frac{\hat{J}^2}{\widehat{Var}(\hat{J})} = 8.76$ which gives a p -value of about .0031 using an asymptotic χ^2 distribution with 2 degrees of freedom under the null hypothesis. Hence, we reject the null hypothesis of no covariate effect at a significance level 0.05. Similarly from (2.13), $\hat{I}_1(F : F_0|\mathbf{z}) = 716.38$ and $\frac{\hat{I}_1^2}{\widehat{Var}(\hat{I}_1)} = 6.91$ which gives a p -value of 0.0086. Also, $\hat{I}_2(F_0 : F|\mathbf{z}) = 900474.4$ and $\frac{\hat{I}_2^2}{\widehat{Var}(\hat{I}_2)} = 9.8$ which gives a p -value of about .0017. Finally, using (2.16), Renyi's divergence of order γ for $\gamma = 0.5$ is $\hat{R}_{0.5}(F : F_0|\mathbf{z}) = 177.78$ and $\frac{\hat{R}_{0.5}^2}{\widehat{Var}(\hat{R}_{0.5})} = 10.1$ which gives a p -value of .0015. It is obvious that all proposed tests confirm an effect due to the covariates. In the case of Wald, LR and Score tests, the computed statistics (and corresponding p -values) were, respectively, 15.1 (0.000537), 15.6 (0.000401) and 16.5 (0.000263). Hence, these standard tests confirm the results from our proposed statistics. It is worth noting here that if we were interested in performing a test of the local hypothesis $H_0 : \beta_i = 0$ vs. $H_1 : \beta_i \neq 0, i = 1$ or 2 , then one could use the approach described in Section 2 on page 6.

4 Comparison of the Proposed Test with Existing Methods

We compared the proposed test statistics $\hat{I}_1(F : F_0|\mathbf{z}), \hat{I}_2(F_0 : F|\mathbf{z}), \hat{J}(F : F_0|\mathbf{z})$ and $\hat{R}_\gamma(F : F_0|\mathbf{z})$ for testing covariate effect in the Cox PH model with existing methods such as the Wald, Likelihood Ratio and Score tests. For Renyi's divergence, we considered $\gamma = 0.1, 0.25, 0.5$ and 0.75 in our simulations. We used empirical estimates of power and consistency based on 5000 simulations for various sample sizes and censoring patterns. We considered the two sample problem (binary covariate case) as well the continuous covariate case.

For the two sample problem, simulations were performed using randomly generated data from exponential distributions with means 1 and 0.5. Sample sizes ranged from 5 to 20 per

group (total sample size ranged from 10 to 40). Independent censoring based on a uniform distribution over a fixed interval $(0, B)$ was applied to result in 25% censored observations. For the continuous covariate case, we randomly generated data from exponential distributions with means 0.5, 1 or 2 using the method outlined in Bender et al. (2005). For a sample of size n , we created a covariate vector by randomly permuting the vector $(1/n, 2/n, \dots, n/n)$ once or generating a single sample from a standard normal or a uniform $(0, 1)$ distribution and using this covariate vector across all simulations. Sample sizes ranged from 10 to 50 and simulations were performed in the manner outlined for the two sample problem.

For each sample size, we computed the power and consistency at various levels of the Type I error α (i.e., $\alpha = 0.01, 0.05, 0.1, 0.15, 0.2, 0.25$) based on 5000 simulations for each test. The simulation results for the two sample problem are tabulated in Tables 4.1 through 4.4. For the continuous covariate case, results are tabulated in Tables 4.5 through 4.8. In each table, the first row corresponding to each test statistic indicates the power and the second row (numbers enclosed within paranthesis) indicates the consistency for each α . These results demonstrate that the proposed tests outperform all the leading tests in terms of power. In particular, we observe from the simulation results that in most cases, $\hat{I}_1$ has better performance relative to all other proposed tests in terms of both power and consistency. This is true across sample sizes and censoring proportions. Also, as $\gamma \rightarrow 1$, there is an improvement in the consistency of $\hat{R}_\gamma$. There is also an overall improvement in its power and in some cases, higher values of γ result in more power than $\hat{I}_1$.

Overall, we find that for some loss in consistency, there is a substantial increase in power for almost all of the proposed tests relative to the three standard tests. The only exception to this is $\hat{I}_2$. It is interesting to note that the realized α for the proposed tests stabilize at around a Type I error cut-off of 0.15. This is not observed in the case of the three standard tests. It is well known that the score test is the most powerful test, asymptotically, under the proportional hazards assumption. However, for small to moderate sample sizes, we find

that the proposed tests demonstrate better performance. In particular, we recommend using our tests for small sample sizes. For larger sample sizes, we found that the performance of the proposed and standard tests are very similar (data not shown). In that case, use of either test (proposed or standard) would be appropriate. For the continuous covariate case, the overall results were similar across these three different approaches used for creating the covariate vector. This demonstrates that the proposed tests are insensitive to the choice of the covariate structure. Furthermore, for larger censoring proportions (such as 50% censored observations), we found the performance of the proposed and standard tests to be similar overall (data not shown). It should be mentioned that the critical value for the Wald statistic was based on the normalized z -score and those for all other test statistics in this study were based on a Chi-square distribution with 1 degree of freedom at a significance level of 0.05.

Table 4.1: Power Comparison and Consistency of Proposed Tests: Binary Covariate (25% censoring)

N (per group)	5	5	5	5	5	5
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.1488 (0.0368)	0.1856 (0.0544)	0.2098 (0.0638)	0.2374 (0.0752)	0.2504 (0.0988)	0.2728 (0.1122)
$\hat{I}_2$	0.0478 (0.0476)	0.0478 (0.0668)	0.0524 (0.0804)	0.0636 (0.091)	0.1022 (0.1104)	0.1088 (0.1152)
$\hat{J}$	0.1128 (0.0492)	0.1594 (0.0848)	0.1936 (0.1148)	0.219 (0.137)	0.245 (0.1536)	0.258 (0.1698)
Wald	0 (0)	0.0424 (0.0214)	0.1278 (0.0806)	0.1976 (0.132)	0.2616 (0.1804)	0.3262 (0.2404)
Score	0.0378 (0.01)	0.1212 (0.057)	0.1948 (0.119)	0.2686 (0.1692)	0.3278 (0.2136)	0.3736 (0.2718)
LR	0.0418 (0.0106)	0.1146 (0.053)	0.199 (0.1176)	0.2598 (0.1638)	0.312 (0.2034)	0.3724 (0.2704)
$\hat{R}_{0.1}$	0.054 (0.047)	0.11 (0.074)	0.127 (0.094)	0.168 (0.127)	0.187 (0.151)	0.206 (0.171)
$\hat{R}_{0.25}$	0.096 (0.053)	0.136 (0.08)	0.176 (0.116)	0.197 (0.14)	0.223 (0.166)	0.259 (0.184)
$\hat{R}_{0.5}$	0.127 (0.055)	0.18 (0.093)	0.222 (0.123)	0.255 (0.149)	0.284 (0.179)	0.318 (0.201)
$\hat{R}_{0.75}$	0.156 (0.066)	0.207 (0.089)	0.247 (0.119)	0.282 (0.145)	0.32 (0.163)	0.332 (0.189)

Table 4.2: Power Comparison and Consistency of Proposed Tests: Binary Covariate (25% censoring)

N (per group)	10	10	10	10	10	10
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.2094 (0.0498)	0.2726 (0.076)	0.3082 (0.098)	0.338 (0.1274)	0.3648 (0.1484)	0.398 (0.1606)
$\hat{I}_2$	0.033 (0.0502)	0.0878 (0.0764)	0.141 (0.1024)	0.183 (0.1282)	0.2234 (0.1568)	0.2634 (0.1708)
$\hat{J}$	0.1728 (0.059)	0.2434 (0.1012)	0.2848 (0.127)	0.3144 (0.1534)	0.3458 (0.185)	0.3732 (0.2134)
Wald	0.0532 (0.0126)	0.1652 (0.048)	0.2804 (0.1134)	0.3522 (0.165)	0.416 (0.222)	0.4642 (0.2704)
Score	0.0646 (0.0238)	0.1932 (0.0684)	0.2874 (0.1226)	0.3544 (0.1806)	0.4242 (0.2304)	0.4696 (0.2842)
LR	0.0586 (0.022)	0.1814 (0.0664)	0.2808 (0.1204)	0.35 (0.1784)	0.4164 (0.2282)	0.4656 (0.2804)
$\hat{R}_{0.1}$	0.0768 (0.059)	0.1348 (0.092)	0.1928 (0.112)	0.2368 (0.1362)	0.2654 (0.1644)	0.3044 (0.1858)
$\hat{R}_{0.25}$	0.1066 (0.0594)	0.1736 (0.095)	0.2288 (0.1178)	0.2654 (0.1442)	0.3022 (0.1712)	0.3412 (0.1904)
$\hat{R}_{0.5}$	0.1484 (0.0592)	0.222 (0.0978)	0.279 (0.13)	0.3168 (0.1488)	0.354 (0.1702)	0.388 (0.1942)
$\hat{R}_{0.75}$	0.1898 (0.0496)	0.253 (0.0936)	0.3044 (0.1236)	0.3478 (0.148)	0.382 (0.164)	0.4098 (0.198)

Table 4.3: Power Comparison and Consistency of Proposed Tests: Binary Covariate (25% censoring)

N (per group)	15	15	15	15	15	15
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.2336 (0.038)	0.3136 (0.0624)	0.3566 (0.0878)	0.3876 (0.111)	0.4178 (0.1254)	0.4448 (0.1554)
$\hat{I}_2$	0.073 (0.0318)	0.1486 (0.054)	0.1956 (0.0826)	0.2472 (0.1082)	0.29 (0.1328)	0.3312 (0.1514)
$\hat{J}$	0.1904 (0.0308)	0.275 (0.0788)	0.3212 (0.107)	0.364 (0.1348)	0.3916 (0.1534)	0.4232 (0.1742)
Wald	0.0698 (0.0046)	0.1986 (0.0298)	0.3154 (0.0834)	0.3878 (0.138)	0.4502 (0.1794)	0.4912 (0.229)
Score	0.0892 (0.007)	0.2112 (0.0362)	0.3186 (0.091)	0.3892 (0.1424)	0.4516 (0.1888)	0.4912 (0.2364)
LR	0.0772 (0.007)	0.2004 (0.0316)	0.3096 (0.0858)	0.383 (0.1412)	0.4482 (0.1896)	0.4898 (0.233)
$\hat{R}_{0.1}$	0.118 (0.052)	0.207 (0.088)	0.253 (0.108)	0.299 (0.131)	0.354 (0.153)	0.392 (0.171)
$\hat{R}_{0.25}$	0.146 (0.052)	0.234 (0.084)	0.284 (0.11)	0.336 (0.134)	0.382 (0.153)	0.421 (0.186)
$\hat{R}_{0.5}$	0.199 (0.051)	0.264 (0.081)	0.335 (0.118)	0.383 (0.136)	0.418 (0.163)	0.446 (0.192)
$\hat{R}_{0.75}$	0.232 (0.05)	0.304 (0.082)	0.369 (0.111)	0.41 (0.141)	0.441 (0.167)	0.472 (0.189)

Table 4.4: Power Comparison and Consistency of Proposed Tests: Binary Covariate (25% censoring)

N (per group)	20	20	20	20	20	20
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.3176 (0.0556)	0.4006 (0.0914)	0.4446 (0.1184)	0.481 (0.1328)	0.5138 (0.1548)	0.5354 (0.1698)
$\hat{I}_2$	0.158 (0.0536)	0.2424 (0.081)	0.301 (0.1088)	0.3726 (0.1474)	0.4076 (0.1602)	0.438 (0.1898)
$\hat{J}$	0.2716 (0.065)	0.3724 (0.1004)	0.4168 (0.1246)	0.4472 (0.1548)	0.4814 (0.1818)	0.516 (0.1942)
Wald	0.1226 (0.0092)	0.2792 (0.0666)	0.3978 (0.1084)	0.4786 (0.151)	0.5396 (0.2)	0.5756 (0.2538)
Score	0.137 (0.0164)	0.2914 (0.0726)	0.3976 (0.1112)	0.4786 (0.163)	0.5378 (0.208)	0.5756 (0.2564)
LR	0.1302 (0.0164)	0.2762 (0.0706)	0.3932 (0.1112)	0.472 (0.163)	0.5336 (0.2078)	0.5702 (0.2562)
$\hat{R}_{0.1}$	0.1928 (0.0478)	0.2988 (0.0814)	0.3784 (0.1078)	0.4296 (0.1406)	0.4652 (0.1726)	0.4978 (0.2044)
$\hat{R}_{0.25}$	0.2316 (0.0482)	0.3374 (0.084)	0.4094 (0.109)	0.4522 (0.147)	0.4856 (0.1772)	0.511 (0.2066)
$\hat{R}_{0.5}$	0.2754 (0.0488)	0.3888 (0.0808)	0.4414 (0.1172)	0.4768 (0.144)	0.5098 (0.1806)	0.5388 (0.206)
$\hat{R}_{0.75}$	0.3148 (0.0502)	0.4226 (0.0876)	0.467 (0.119)	0.4968 (0.149)	0.5234 (0.1726)	0.547 (0.1994)

Table 4.5: Power Comparison and Consistency of Proposed Tests: Continuous Covariate (25% censoring)

N	10	10	10	10	10	10
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.1238 (0.0234)	0.1468 (0.037)	0.1704 (0.0468)	0.1908 (0.0562)	0.2148 (0.064)	0.2254 (0.067)
$\hat{I}_2$	0.006 (0.0384)	0.0114 (0.048)	0.015 (0.0642)	0.0168 (0.0786)	0.017 (0.0828)	0.0236 (0.0872)
$\hat{J}$	0.0988 (0.0514)	0.1318 (0.0692)	0.1558 (0.083)	0.1736 (0.103)	0.194 (0.1238)	0.2138 (0.1392)
Wald	0 (0)	0.0688 (0.0338)	0.1798 (0.0984)	0.2528 (0.1422)	0.3204 (0.202)	0.381 (0.2496)
Score	0.0178 (0.0136)	0.1112 (0.0582)	0.1964 (0.1166)	0.2678 (0.1718)	0.3304 (0.216)	0.3818 (0.2724)
LR	0.0276 (0.0174)	0.117 (0.0582)	0.202 (0.1244)	0.273 (0.1756)	0.3306 (0.2184)	0.3854 (0.2748)
$\hat{R}_{0.1}$	0.0256 (0.0652)	0.0524 (0.1004)	0.0804 (0.1088)	0.1188 (0.1316)	0.1488 (0.154)	0.1744 (0.17)
$\hat{R}_{0.25}$	0.0512 (0.0608)	0.0992 (0.0944)	0.1408 (0.1296)	0.1744 (0.1444)	0.2076 (0.1716)	0.2344 (0.1968)
$\hat{R}_{0.5}$	0.1148 (0.0608)	0.1724 (0.1028)	0.216 (0.1448)	0.2548 (0.178)	0.2868 (0.2016)	0.3172 (0.2304)
$\hat{R}_{0.75}$	0.1516 (0.054)	0.2084 (0.0996)	0.256 (0.1264)	0.2828 (0.1548)	0.3056 (0.1884)	0.3328 (0.2084)

Table 4.6: Power Comparison and Consistency of Proposed Tests: Continuous Covariate (25% censoring)

N	20	20	20	20	20	20
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.2244 (0.0274)	0.288 (0.049)	0.3266 (0.0608)	0.358 (0.0784)	0.3784 (0.0932)	0.4112 (0.1092)
$\hat{I}_2$	0.0056 (0.0486)	0.0406 (0.0698)	0.08 (0.0902)	0.1142 (0.1036)	0.1368 (0.1298)	0.1748 (0.1376)
$\hat{J}$	0.1738 (0.058)	0.2468 (0.0896)	0.292 (0.1116)	0.3262 (0.1334)	0.3594 (0.1514)	0.382 (0.1748)
Wald	0.06 (0.01)	0.1824 (0.0596)	0.3032 (0.109)	0.3818 (0.1536)	0.4546 (0.2022)	0.5162 (0.2644)
Score	0.078 (0.0138)	0.1964 (0.0662)	0.3034 (0.1166)	0.3836 (0.1676)	0.4546 (0.215)	0.5162 (0.2708)
LR	0.0766 (0.0138)	0.1962 (0.0626)	0.3016 (0.1148)	0.3818 (0.168)	0.4508 (0.2166)	0.5138 (0.2706)
$\hat{R}_{0.1}$	0.0796 (0.0588)	0.1376 (0.0908)	0.1916 (0.1208)	0.2352 (0.146)	0.2752 (0.1668)	0.314 (0.1864)
$\hat{R}_{0.25}$	0.1176 (0.0568)	0.1912 (0.1)	0.2492 (0.1308)	0.2896 (0.1588)	0.3336 (0.18)	0.3668 (0.2068)
$\hat{R}_{0.5}$	0.1756 (0.0628)	0.2652 (0.104)	0.3204 (0.1432)	0.3604 (0.1708)	0.3936 (0.2024)	0.4368 (0.22)
$\hat{R}_{0.75}$	0.2304 (0.0544)	0.306 (0.1044)	0.3596 (0.138)	0.3956 (0.1668)	0.4372 (0.1896)	0.4632 (0.2092)

Table 4.7: Power Comparison and Consistency of Proposed Tests: Continuous Covariate (25% censoring)

N	30	30	30	30	30	30
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.2738 (0.0288)	0.3542 (0.0598)	0.4004 (0.0868)	0.4426 (0.107)	0.4662 (0.1258)	0.4956 (0.135)
$\hat{I}_2$	0.0604 (0.0414)	0.1084 (0.0562)	0.1666 (0.0768)	0.2152 (0.0906)	0.2642 (0.1036)	0.3026 (0.1192)
$\hat{J}$	0.2262 (0.0448)	0.2948 (0.0726)	0.3656 (0.1008)	0.4006 (0.1336)	0.4368 (0.1464)	0.4666 (0.1684)
Wald	0.0838 (0.0078)	0.2286 (0.0456)	0.3598 (0.085)	0.4432 (0.1406)	0.5076 (0.1848)	0.5546 (0.2418)
Score	0.0902 (0.0102)	0.2342 (0.049)	0.3594 (0.0932)	0.4432 (0.1522)	0.5074 (0.196)	0.5506 (0.2512)
LR	0.0896 (0.0104)	0.232 (0.0484)	0.3552 (0.0928)	0.441 (0.1504)	0.5058 (0.1906)	0.5508 (0.2528)
$\hat{R}_{0.1}$	0.1416 (0.054)	0.2372 (0.094)	0.286 (0.122)	0.3376 (0.1452)	0.3824 (0.1804)	0.4292 (0.2096)
$\hat{R}_{0.25}$	0.1848 (0.054)	0.276 (0.0984)	0.3304 (0.1276)	0.38 (0.164)	0.4324 (0.1948)	0.4656 (0.214)
$\hat{R}_{0.5}$	0.252 (0.0492)	0.33 (0.0992)	0.392 (0.146)	0.442 (0.1732)	0.48 (0.1988)	0.5252 (0.2212)
$\hat{R}_{0.75}$	0.2924 (0.0484)	0.374 (0.1012)	0.4368 (0.1416)	0.4732 (0.17)	0.5172 (0.194)	0.5412 (0.2136)

Table 4.8: Power Comparison and Consistency of Proposed Tests: Continuous Covariate (25% censoring)

N	50	50	50	50	50	50
Type I Error α	0.01	0.05	0.1	0.15	0.2	0.25
$\hat{I}_1$	0.475 (0.0368)	0.5756 (0.0558)	0.6196 (0.0946)	0.6582 (0.1192)	0.683 (0.14)	0.7038 (0.1582)
$\hat{I}_2$	0.203 (0.0462)	0.336 (0.0766)	0.4314 (0.095)	0.4988 (0.118)	0.5586 (0.1278)	0.5904 (0.157)
$\hat{J}$	0.4132 (0.048)	0.529 (0.0834)	0.5836 (0.104)	0.6224 (0.1358)	0.664 (0.1624)	0.6834 (0.1798)
Wald	0.1722 (0.0036)	0.4146 (0.0476)	0.5718 (0.0934)	0.6436 (0.1456)	0.7038 (0.19)	0.746 (0.2446)
Score	0.1804 (0.0056)	0.421 (0.0502)	0.5662 (0.094)	0.6432 (0.1538)	0.7018 (0.1976)	0.744 (0.2486)
LR	0.1722 (0.0038)	0.4184 (0.0498)	0.568 (0.0936)	0.641 (0.1502)	0.7 (0.1958)	0.744 (0.2484)
$\hat{R}_{0.1}$	0.298 (0.0488)	0.4172 (0.0792)	0.4888 (0.1096)	0.5392 (0.1424)	0.5796 (0.1584)	0.6112 (0.1812)
$\hat{R}_{0.25}$	0.3496 (0.046)	0.458 (0.0804)	0.5216 (0.1144)	0.5668 (0.1432)	0.6032 (0.1692)	0.6376 (0.1836)
$\hat{R}_{0.5}$	0.4036 (0.0464)	0.5076 (0.0944)	0.5644 (0.1188)	0.6004 (0.1416)	0.6352 (0.1704)	0.6592 (0.1956)
$\hat{R}_{0.75}$	0.4536 (0.0504)	0.5444 (0.0932)	0.5908 (0.1216)	0.6312 (0.1396)	0.654 (0.1636)	0.6824 (0.192)

5 Concluding Remarks

In this paper, we proposed several tests for covariate effect in the Cox PH model based on KL divergence and Renyi's divergence. These tests are useful for discriminating between the baseline distribution and one possibly due to the covariates in the Cox PH model. An important advantage of these discrimination information functions is that all of them are invariant under any non-singular transformation of the failure time data.

Our proposed tests are moderately easy to compute. An attractive feature of the proposed methods is that the test statistics are simple functions of the MPLE $\hat{\beta}$ and independent of the baseline distribution. Using the results derived in this paper, one can easily implement our tests in a readily available, open source programming environment such as the R statistical system. All these tests, particularly $\hat{I}_1$ and $\hat{R}_\gamma$ as $\gamma \rightarrow 1$, perform very well in terms of power compared with other leading tests, and in terms of both power and consistency relative to the other proposed tests. For small to moderate sample sizes, the increase in power we observe is substantial relative to the loss in consistency for all these tests compared with the three standard tests. Based on our results, the proposed tests appear to be an attractive alternative to the standard tests for small sample sizes.

While we have specifically focused on testing for any covariate effect in the widely used Cox PH model using a class of discrimination information functions, one could be interested in applying similar methods to other survival models such as the accelerated failure time (AFT) model (Etezadi-Amoli & Ciampi, 1985) or the additive hazards (AH) model (Tibshirani & Ciampi, 1983). Unlike the Cox PH model, the AFT model allows for crossing hazard curves. For a test of the null hypothesis of no covariate effect against the alternative specified by the AFT model, the discrimination information functions (analogous to the quantities specified in (2.9)-(2.12)) would be dependent on the baseline hazard. This is also the case with the AH model. In that situation, one would have to apply spline approximations for parameter estimation as described in Devarajan (1999), and then follow the methods de-

scribed in this paper for testing the covariate effect in these models. Devarajan & Ebrahimi (2001) utilized such an approach for testing the goodness of fit of the Cox PH model against a generalization of the Cox PH model that allows for crossing hazards. A detailed discussion of this method is beyond the scope of this article and interested reader is referred to the paper referenced above.

Acknowledgement The authors would like to thank the referees for their helpful comments. The work of the first author was supported in part by NIH grant P30 CA 06927 and an appropriation from the Commonwealth of Pennsylvania. The authors would like to acknowledge the services provided by the High Performance Computing Facility at Fox Chase Cancer Center.

References

- Asadi, M., Ebrahimi, N. and Soofi, E. (2006). Dynamic generalized information measures. *Statistics and Probability letters*, 71, 85-98.
- Bender, R., Augustin, T. and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models, *Statistics in Medicine*, 24, 1713-1723.
- Cortinas Abrahantes, J. and Burzykowski, T. (2005). A version of EM algorithm for proportional hazards model with random effects. *Biometrical Journal*, 47, 847-862.
- Devarajan, K. (1999). Inference for a non-proportional hazards regression model and applications, Ph.D. Dissertation, Division of Statistics, Northern Illinois University, DeKalb, Illinois.
- Devarajan, K., Ebrahimi, N. (2001). Goodness of fit testing for the Cox proportional hazards model using Kullback-Leibler discrimination function. *Goodness-of-Fit Tests and Model Validity, Statistics for Industry and Technology*, 235-251, Birkhauser Publishers.

- Ebrahimi, N., Habibullah, M. and Soofi, E. (1994). Testing exponentiality based on Kullback-Leibler information. *JRRS-B*, 54, 739-748.
- Ebrahimi, N. and Soofi, E. (2004). Information functions for reliability. In *Mathematics Reliability*, edited by Soyer, R., Mazzuchi, T.A. and Singpurwalla, N., Kluwer.
- Etezadi-Amoli, J. and A. Ciampi (1985). A General Model for Testing the Proportional Hazards and the Accelerated Failure Time Hypothesis in the Analysis of Censored Survival Data with Covariates, *Communications in Statistics-A*, 14, 651-667.
- Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21, 826-838.
- Kalbfleisch, J.D. and R.L. Prentice (2002). *The Statistical Analysis of Failure Time Data*, second edition. John Wiley, New York.
- Klein, John, and Moeschberger, M.L. (2003). *Survival analysis, techniques for censored and truncated data*, second edition. Springer, New York.
- Kullback, S. (1978). *Information Theory and Statistics*, Dover Publications, 1978.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raja Rao, B. and Talwalker, S. (1991). Random censorship, competing risks and a simple proportional hazards model, *Biometrical Journal*, 33, 461-483.
- Tibshirani, R.J. and Ciampi, A. (1983). A family of proportional- and additive-hazards models for survival data, *Biometrics*, 39, 141-147.