An updated algorithm for moderate censoring in time-to-event data using rank-based regression

Milind A. Phadnis PhD[1*]

[1]*Department of Biostatistics and Data Science, University of Kansas Medical Center, Kansas City, KS, USA*

ABSTRACT

Aim: To propose an updated algorithm with an extra step added to the Newton-type algorithm used in robust rank based non-parametric regression for minimizing the dispersion function associated with Wilcoxon scores in order to account for the effect of covariates.

Methodology: The proposed accelerated failure time approach is aimed at incorporating right random censoring in survival data sets for low to moderate levels of censoring. The existing Newton algorithm is modified to account for the effect of one or more covariates. This is done by first applying Mantel scores to residuals obtained from a regression model, and second by minimizing the dispersion function of these scored residuals. Diagnostic check of the model fit is performed by observing the distribution of the residuals and suitable Bent scores are considered in the case of skewed residuals. To demonstrate the efficacy of this method, a simulation study is conducted to compare the power of this method under three different scenarios: non-proportional hazard, proportional and constant hazard, and proportional but non-constant hazard.

Results: In most situations, this method yielded reasonable estimates of power for detecting an association of the covariate with the response as compared to popular parametric and semi-parametric approaches. The estimates of the regression coefficient obtained from this method were evaluated and were found to have low bias, low mean square error, and adequate coverage. In a real-life example pertaining to pancreatic cancer study, the proposed method performed admirably well and provided a more realistic interpretation about the effect of covariates (age and Karnofsky score) compared to a standard parametric (lognormal) model.

Conclusion: In situations where there is no clear best parametric fit for time-to-event data with moderate level of censoring, the proposed method provides a robust alternative to obtain regression coefficients (both adjusted and unadjusted) with a performance comparable to that of a proportional hazards model.

Keywords: bent scores, mantel scoring, newton algorithm, rank regression, Wilcoxon scores

* E-mail address: mphadnis@kumc.edu.

1. INTRODUCTION

For interval scaled, non-censored data, Conover and Iman [1] have investigated the properties of regression analysis of the ranks of interval data as an alternative to ordinary least squares analyses. These contributions of Conover and Iman provided an alternative non-parametric rank-transform approach that allowed for the modeling of the impact of multiple continuous and categorical predictors on continuous outcomes. Howard and Koch [2] extended this approach to the univariate analysis of exponentially distributed right censored (survival) data by considering simple regression analysis of log rank scores, showing the performance of the approach to be similar to proportional hazards modeling. Their simulation studies show that in the case where there are no ties in the survival times, this approach was only marginally less powerful than tests from proportional hazards models, but clearly less powerful than a likelihood ratio test for a fully parametric model when the appropriate underlying survival function is employed. When there were tied survival times, this approach proved marginally more powerful than tests from Cox's semi-parametric proportional hazards procedure. While their approach is generally reliable for the testing of associations with survival outcomes, it has the substantial shortcoming of not providing a clinically interpretable parameter quantifying the magnitude of the association between predictors and outcomes, such as the hazard ratio provides for proportional hazards analysis. This shortcoming arises due to the fact that when the response variable is replaced by its logrank score, it is not possible to estimate the true value of the regression coefficient in the original metric. Hence commonly used measures of assessing performance of the method such as bias, mean square error, and coverage cannot be deployed. Also, Howard and Koch [2] did not evaluate the performance of logrank scores when survival data comes from different distributions such as the loglogistic or the lognormal distribution and is hence not generalizable.

Many authors such as Hougaard [3] have commented on the restrictions owing to lack of suitable estimation routines in the non-parametric case for an accelerated failure time model. Several semiparametric estimators accommodating censoring in survival data were proposed such as the modified least squares estimator by Buckley and James [4] and rank-based estimators based on the weighted log-rank statistics by Prentice [5]. The theoretical properties of these estimators were rigorously studied by Tsiatis [6], Ritov [7], Lai and Ying [8] and [9], and Ying [10] among others. Jin, Lin, Wei, and Ying [11] has discussed the reasons why despite theoretical developments, semiparametric approaches are rarely used in real life applications owing to the lack of efficient and reliable computational methods. They discuss how the inference procedure developed by Wei, Ying, and Lin [12] based on the minimum dispersion statistic is difficult and cannot be solved by conventional optimization algorithms. To overcome the limitations of the computational method developed by Lin and Greyer [13] in failing to always find a true minimum for the dispersion statistic, Jin et al., [11] have developed a linear programming method to minimize a convex objective function for the rank estimator based on Gehan [14] type weight function without having to indulge in nonparametric density estimation.

Advances in robust rank-based procedures have spawned a detailed methodology for analyzing linear and nonlinear models in a regression setting. This methodology applies the appropriate scoring function (such as the Wilcoxon scoring function) on the residuals arising out of a log-linear model rather than the response variable thereby allowing the estimation of the regression coefficient. This methodology has also been extended to diverse areas such as time-series analyses, random effects models, and censor-free survival data; however, reliable and easy-to-use developments to extend the approaches to the analysis of right-censored (survival) data have not been investigated using this approach. In the context of the survival data analyses, by estimating the regression coefficient, this method therefore, has the potential to allow the practitioner to derive meaningful measures of the magnitude of the association such as the increase in median survival time (of treatment over placebo).

By replacing the Euclidean (L_2) norm by a rank-based norm, and by minimizing the dispersion function associated with this norm, it is possible to get robust non-parametric estimates of the regression parameters (Hettsmanperger and McKean [15]). Various diagnostic procedures that examine the quality of fit of these models and inference procedures to compute confidence intervals for parameters and their contrasts have also been developed (Hettsmanperger & McKean [15]). With non-censored data, these procedures outperform the traditional least squares methods when there are many outliers and influential points in the data set. The performance of these rank-based approaches is optimized when the underlying error density is known as it is possible to compute the optimal scoring function (McKean and Sievers [16]). These methods can therefore be extended to survival data and optimal scoring functions for many popular distributions used in analyses of time to failure data including exponential, Weibull, loglogistic and lognormal have been calculated. In order to counter the influence of outliers from affecting the model fit, various weighted versions of the rank-based model fit have been proposed (McKean, Terpstra, and Kloke [17]).

Herein, we show how a fully non-parametric approach can be employed to estimate regression coefficients, and assess the impact of the approach across varying censoring rates from relatively low censoring rate as would be observed in an oncology study to a higher censoring rate as observed in cardiovascular outcome studies. Our analyses are focused on right censored survival data expressed as a log-linear model and the performance is assessed via a simulation study.

In Section 2.1, we discuss in brief the general theory associated with the rank based procedures. Hettsmanperger and McKean [15] outline the Newton Raphson algorithm used to obtain the optimal regression parameter estimates. The R code for implementing this algorithm is due to Terpstra and McKean [18]. In Section 2.2, we discuss our motivation for extending these methods to account for right random censoring in survival data. In Section 2.3, for the case where Wilcoxon scores are used as the scoring function (optimal for the logistic error density), we propose the addition of an extra step to this algorithm that incorporates the right random censoring mechanism inherent in survival data so as to reassign the Wilcoxon scores without violating the assumptions required by theory. This approach makes use of the fact that responses that have been censored carry partial information to the effect that an event has not occurred till the time of censoring but is likely to occur at some time in the future. In Section 2.4 we discuss the Bent score function as a diagnostic checking aid (and as an alternative) to the Wilcoxon fit of residuals in the case where the residuals are positively skewed. In Section 3.1 and 3.2, we simulated data from different scenarios reflecting different levels of censoring and different error densities. In Section 3.3, we present results obtained from applying the proposed method to a real-life data from a cohort of patients suffering from pancreatic cancer. The results obtained from our method are compared with those obtained from the traditional approaches that are otherwise used to analyze this data. Concluding remarks are presented in Section 4.

2. MATERIAL AND METHODS

2.1 Rank-based Methods for Linear Models

In this section we give a brief discussion of the theory associated with developing linear models in the context of nonparametric regression that can be used to draw inference.

Let $\mathbf{Y}$ denote a n x 1 vector of responses that follows the linear model:

$$\mathbf{Y} = \mathbf{1}\alpha + \mathbf{X}\boldsymbol{\beta} + \varepsilon \qquad (1)$$

where $\mathbf{1}$ denotes a n x 1 vector of ones, α is an unknown scalar intercept, $\mathbf{X}$ is a n x p matrix of predictors (continuous or categorical), $\boldsymbol{\beta}$ is a p x 1 vector of unknown constant regression coefficients, and ε is the n x 1 vector of random errors. Let Ω be the column space of full rank design matrix $\mathbf{X}$ so that the dimension of Ω is p. The rank-based estimate of $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}}_{\varphi} = \text{Argmin} \; \| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \| = \text{Argmin}\{ \; D_{\varphi}(\boldsymbol{\beta})\} \quad (2)$$

Here, Argmin is the value of $\boldsymbol{\beta}$ that minimizes $D_{\varphi}(\boldsymbol{\beta}) = \| \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \|$ and $\| \cdot \|$ is the pseudo-norm used in the rank-based procedures that has replaced the Euclidean norm of the traditional least squares methods and is given by:

$$\| \mathbf{v} \|_{\varphi} = \sum_{i=1}^{n} a\{R(v_i)\} v_i , \qquad \underset{\sim}{v} \in \Re^{n} \quad (3)$$

where the scores are generated as $a(i) = \varphi\{ i/(n+1)\}$ for a non-decreasing square-integrable function $\varphi(u)$ defined on the interval (0,1) and standardized such that $\int \varphi(u)du = 0$, $\int \varphi^2(u)du = 1$, and $R(v_i)$ is the rank of v_i among $v_1, v_2, v_3, \ldots, v_n$. Using this norm, various scoring functions can be generated such as the sign-pseudo norm of the form $\varphi(u) = \text{sgn}(u - 1/2)$ and the Wilcoxon pseudo-norm of the form $\varphi(u) = \sqrt{12}(u - 1/2)$. Thus in terms of these pseudo-norms, $D_{\varphi}(\boldsymbol{\beta})$ is a convex function of $\boldsymbol{\beta}$ and $D_{\varphi}(\hat{\boldsymbol{\beta}}_{\varphi})$ is the minimized distance between $\mathbf{Y}$ and Ω. As the scores are standardized (they sum to zero) and the ranks are invariant to a constant shift, the intercept cannot be estimated using the norm and is usually estimated as the median of the residuals $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}}_{\varphi}$ in the following way:

$$\hat{\alpha}_S = \text{median} \; (Y_i - x_i^T \hat{\beta}_{\varphi}) \quad (4)$$

Hettsmanperger and McKean (1998) have shown that under some regularity conditions

$$\begin{pmatrix} \hat{\alpha}_s \\ \hat{\boldsymbol{\beta}}_{\varphi} \end{pmatrix} \sim N_{p+1}\left\{ \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} n^{-1}\tau_s^2 & \mathbf{0}^T \\ \mathbf{0}^T & \tau^2_{\varphi}(\mathbf{X}^T\mathbf{X})^{-1} \end{pmatrix} \right\} \quad (5)$$

where $\tau_s = \dfrac{1}{2f(0)}$, $\tau_{\varphi} = \dfrac{1}{\int \varphi(u)\varphi_f(u)du}$, $\varphi_f(u) = \dfrac{-f'\{F^{-1}(u)\}}{f\{F^{-1}(u)\}}$ and $f(\varepsilon)$ is the pdf of ε.

They then applied this result to develop asymptotic test of hypothesis and other inferential procedures. A formal Newton-type algorithm to compute the estimates of the regression parameters by minimizing the dispersion function given in equation (3) has been proposed by Kapenga, McKean, and Vidmar [19] who have programmed the algorithm in the Fortran routine rglm (see Appendix A).

2.2 Scoring Scheme in the Proposed Algorithm

In this section we discuss modifications to this algorithm to accommodate survival data with right random censored observations. It is very important to note that the algorithm in Appendix A applies the Wilcoxon scores on the residuals and not directly on the observations which constitute the survival data. The proposed approach extends results (discussed below) obtained by Mantel [20] that were originally applied directly to survival data, by applying the scoring function to the residuals while retaining the assumptions required by the algorithm discussed in Appendix A.

From equation (A.1) in Appendix A, it can be seen that the scoring function $\mathbf{a}\{R(\hat{\mathbf{e}})\}$ is a vector whose i^{th} component is $a\{R(\hat{e}_i)\}$. Using the formula for $\varphi_f(u)$ defined in the preceding section, Hetsmanperger and McKean [15] showed that for errors which follow a logistic distribution, $a\{R(e)\} = \varphi\{R(e)/(n+1)\} = \varphi(u) = \sqrt{12}(u - 1/2)$ is the optimal scoring function and is called the Wilcoxon scoring function. Let $X_{(1)}, X_{(2)}, X_{(3)},\ldots,X_{(n)}$ be the ordered statistics from a uniform distribution. If all the observations *j* = 1,2,3,…n are uncensored, it follows that $E\{X_{(j)}\} = j/(n+1)$ (see for instance (Casella and Berger, 2002). Furthermore, it can be shown that $(1/n)\sum_{j=1}^{n} E(X_{(j)}) = 1/2$ and $\mathrm{Var}\{E(X_{(j)})\} \cong 1/12\}$. Thus, the Wilcoxon scoring function $\varphi(u) = \sqrt{12}\,(u - 1/2)$ applied over the ranked residuals represents the standardized expected values of the ordered statistics from a Uniform (0, 1) distribution. This scoring function satisfies the assumptions discussed in section 2.1 above. However, it should be noted that no adjustment is made to account for censored observations in the sense that the scoring function does not distinguish between an event and a censored observation.

Mantel [20] has obtained the expected values of the Uniform (0, 1) order statistics in the presence of arbitrary right censoring for survival data. Our proposed modification to the algorithm applies Mantel's method to reflect change in scores for the ranked residuals that are associated with censored observations. As an illustration, consider the following hypothetical survival data sorted in ascending order where 'E' indicates an uncensored (event) observation and 'C' indicates a right censored observation: $T_{(1)} = 1(E),\ T_{(2)} = 2(E),\ T_{(3)} = 4(C),\ T_{(4)} = 6(C),\ T_{(5)} = 7(C),\ T_{(6)} = 8(E),\ T_{(7)} = 10(E),\ T_{(8)} = 12(E),\ T_{(9)} = 15(E),\ T_{(10)} = 18(E)$. In this dataset of 10 observations sorted in ascending order, the first 2 observations are uncensored followed by 3 censored observations and then followed by 5 uncensored observations. For this particular ordering of events and censored observations, applying Mantel's method we get:

$$\text{for } j = 1, 2;\quad E(X_{(j)}) = \frac{j}{n+1} = \frac{R(X_{(j)})}{n+1}$$

$$\text{for } j = 3, 4, 5;\quad E(X_{(j)}) = \frac{n+1+R(X_{(i)})}{2(n+1)}$$

$$\text{for } j = 6, 7, 8, 9, 10;\quad E(X_{(j)}) = \frac{(j-i-k)\dfrac{n+1-R(X_{(i)})}{n+1-i-k} + R(X_{(i)})}{n+1} \qquad (6)$$

where *i* = 2 (first two uncensored observations), *k* = 3 (next three censored observations), $n - i - k = 5$.

Since the first 2 observations are uncensored, they are assigned the scores of 1/11 and 2/11 respectively. The next three events are censored observations and are each assigned a score of 6.5/11 which is the average over the interval 2 through 11 divided by *n*+1. The remaining 5 observations which are uncensored are spread over $n+1-i-k$ = 6 intervals so that the average width into which they would divide the remaining space is $\{n+1-R(X_{(i)})\}/(n+1-i-k) = 1.5$. Thus,

for $j = 6$, $E(X_{(j)}) = \{1\cdot(1.5)+2\}/11 = 3.5/11$; for $j = 7$, $E(X_{(j)}) = \{2\cdot(1.5)+2\}/11 = 5/11$;

for $j = 8$, $E(X_{(j)}) = \{3\cdot(1.5)+2\}/11 = 6.5/11$; for $j = 9$, $E(X_{(j)}) = \{3\cdot(1.5)+2\}/11 = 8/11$;

for $j = 10$, $E(X_{(j)}) = \{4\cdot(1.5)+2\}/11 = 9.5/11$;

The censoring mechanism dictates the allocation of scores to the observations depending on whether they are uncensored or censored values and depending on their order of their occurrence in the data set. It is important to note that with the allocation of these scores, $(1/n)\sum_{j=1}^{n} E(X_{(j)}) = 1/2$ still holds. Further adjustments can be made for tied events. Thus for a consecutive sequence of m ties *j*, *j*+1, *j*+2,…, *j*+(*m*-1), the expected values for each $X_{(j)}$ is averaged across the *m* ties. For tied censored observations, however, no adjustment is necessary reflecting the fact that the empirical distribution does not have any probability between successive uncensored observations and has all its remaining mass at or beyond the later uncensored observation (Mantel 1981). Thus consecutive tied censored observations share the same score (6.5 for the three tied censored observations *j* = 3, 4, 5).

Here, it should be noted that equation (3) calls for $a(i) = \varphi\{i/(n+1)\}$ to be a non-decreasing set of scores, not all equal (Jaeckel [22]). However, the Mantel scoring scheme has assigned scores of 1/11, 2/11, 6.5/11, 6.5/11, 6.5/11, 3.5/11, 5/11, 6.5/11, 8/11 and 9.5/11 respectively to the observations $\mathrm{T}_{(1)} = 1(\mathrm{E})$, $\mathrm{T}_{(2)} = 2(\mathrm{E})$, $\mathrm{T}_{(3)} = 4(\mathrm{C})$, $\mathrm{T}_{(4)} = 6(\mathrm{C})$, $\mathrm{T}_{(5)} = 7(\mathrm{C})$, $\mathrm{T}_{(6)} = 8(\mathrm{E})$, $\mathrm{T}_{(7)} = 10(\mathrm{E})$, $\mathrm{T}_{(8)} = 12(\mathrm{E})$, $\mathrm{T}_{(9)} = 15(\mathrm{E})$, $\mathrm{T}_{(10)} = 18(\mathrm{E})$ that would make the convexity property of $D_{\varphi}(\boldsymbol{\beta})$ not always hold in general (Jaeckel [22]). To overcome this problem, the censored observations $\mathrm{T}_{(3)} = 4(\mathrm{C})$, $\mathrm{T}_{(4)} = 6(\mathrm{C})$, $\mathrm{T}_{(5)} = 7(\mathrm{C})$, which resulted in a score of 6.5/11 need to be assigned new pseudo values. This is based on the assumption that a censored observation is a partially observed value and its true unobserved value is likely more than its observed (censored) value. Thus we need to find two consecutive event observations with respective scores s_1 and s_2 such that the conditions 6.5/11 $\geq s_1$ and 6.5/11 $< s_2$ are met. In this data set, we find that $\mathrm{T}_{(8)} = 12(\mathrm{E})$ and $\mathrm{T}_{(9)} = 15(\mathrm{E})$, two such event observations with respective scores s_1= 6.5/11 and s_2= 8/11. Therefore, the pseudo values for the three censored observations are generated as the average of 12 and 15 leading to a pseudo-value of 13.5. That is, we have now generated the scores as 1/11, 2/11, 3.5/11, 5/11, 6.5/11, 6.5/11, 6.5/11, 6.5/11, 8/11 and 9.5/11 respectively for the observations $\mathrm{T}_{(1)} = 1(\mathrm{E})$, $\mathrm{T}_{(2)} = 2(\mathrm{E})$, $\mathrm{T}_{(3)} = 8(\mathrm{E})$, $\mathrm{T}_{(4)} = 10(\mathrm{E})$, $\mathrm{T}_{(5)} = 12(\mathrm{E})$, $\mathrm{T}_{(6)} = 13.5$(pseudo-value), $\mathrm{T}_{(7)} = 13.5$(pseudo-value), $\mathrm{T}_{(8)} = 13.5$ (pseudo-value), $\mathrm{T}_{(9)} = 15(E)$, $\mathrm{T}_{(10)} = 18(E)$. This results in a value of [1(1) + 2(2) + 8(3.5) + 10(5) + 12(6.5) + 13.5(6.5) + 13.5(6.5) + 13.5(6.5) + 15(8) + 18(9.5)]/11 = 65.023 for $D_{\varphi}(\boldsymbol{\beta})$ and ensures its convexity owing to the observations and their corresponding scores ordered in the same direction in the sum of equation (3).

Every data set will thus have a unique scoring scheme based on the order in which events and censorings occur in the dataset. After the initial Mantel scoring, pseudo values will have to be generated for all the censored observations with their magnitude depending on first finding $s_1$and s_2, and then averaging out the magnitude of the observations corresponding to $s_1$and s_2. In cases where the largest observation in a dataset is an event and the Mantel score for any censored observation exceeds this largest event observation, the pseudo value for this censored observation will be the same as this largest event observation. When the largest observation in a dataset is a censoring, its Mantel score will always be more than that of the largest event observation and so there is no cause for concern.

2.3 Steps of the Proposed Modified Algorithm

In this section we enumerate the steps in our updated algorithm.

Step (i) Obtain an initial estimate of the regression coefficients, $\hat{\boldsymbol{\beta}}^{(0)}$ (say, the least squares estimate) and calculate the initial residuals as $\hat{\mathbf{e}}^{(0)} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{0)}$. Rank these residuals in ascending order. Using the censoring mechanism inherent in the data set, reassign the ranks using the scores described in equation (6). By design, the average of these new ranks is 1/2. Calculate the standard deviation of these new ranks and denoted it by ς . Apply the scoring function $a_{adj}(j) = \varphi_{adj}\{\mathrm{E}(e_{(j)})\} = \{\mathrm{E}(e_{(j)}) - 0.5\}/\varsigma$. Let $\hat{\tau}_{\varphi-adj}^{(0)}$ denote the initial estimate of $\tau_{\varphi-adj}$ based on these residuals. This is obtained by solving:

$$\hat{\tau}_{\varphi-adj}^{(0)} = \frac{2t_{n,\delta}\sqrt{n}}{\{\varphi_{\max}(u)-\phi_{\min}(u)\}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\varphi^{*}\left(\frac{j}{n}\right)-\varphi^{*}\left(\frac{j-1}{n}\right)\right\} I\left(|\hat{e}_{(i)}^{(0)}-\hat{e}_{(j)}^{(0)}| \le t_{n,\delta}\right)} \qquad (7)$$

where $\varphi^{*}(u) = \varphi(u)/\{\varphi_{\max}(u)-\phi_{\min}(u)\}$

and $t_{n,\delta}$ is the δ^{th} quantile of $\hat{G}(t_{n,\delta}) = (1/n)\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\varphi^{*}\left(\frac{j}{n}\right)-\varphi^{*}\left(\frac{j-1}{n}\right)\right\} I\left(|\hat{e}_{(i)}^{(0)}-\hat{e}_{(j)}^{(0)}| \le t_{n,\delta}\right)$

where δ is the bandwidth used to obtain stable estimates of $\tau_{\varphi-adj}$. For moderate sample sizes, where the ratio of n to the number of parameters p exceeds 5, δ = 0.8 yields stable estimates. For more details about the theory associated with equation (7), refer to the text by Hettsmanperger and McKean [15]

Calculate the dispersion function $D_{adj}^{(0)}$ using equation (3) evaluated at $\hat{\mathbf{e}}^{(0)}$. Note that the assumptions of $\int \varphi_{adj}(u)du = 0$ and $\int \varphi_{adj}^{2}(u)du = 1$ are true (see Appendix B for proof) and $a_{adj}(j) \equiv \varphi_{adj}\{\mathrm{E}(X_{(j)})\}$ is a non-decreasing function.

Step (ii) Using the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$ onto the column space of $\mathbf{X}$, obtain the residuals at the 1st iteration of the algorithm using the relation:

$$\hat{\mathbf{e}}_{adj}^{(1)} = \hat{\mathbf{e}}^{(0)} - \hat{\tau}_{\varphi-adj}\,\mathbf{H}\mathbf{a}_{adj}\{R(\hat{\mathbf{e}}^{(0)})\} \qquad (8)$$

where $\mathbf{a}_{adj}\{R(\hat{\mathbf{e}})\}$ denotes the vector whose ith component is $a_{adj}\{R(\hat{e}_{i}^{(0)})\}$.

Step (iii) and (iv) are the same as in the existing algorithm displayed in Appendix A except that we use the notation $D_{adj}^{(k)}$ and $\hat{\tau}_{\varphi-adj}$ in place of $D^{(k)}$ and $\hat{\tau}_{\varphi}$. We retain the notation $\hat{\boldsymbol{\beta}}_{\varphi}$ and $\hat{\alpha}_{s}$ for the estimates of the regression coefficients.

2.4 Bent Scores

McKean, Vidmar, and Sievers [21] have demonstrated that a gain in power in rank based analysis based on Bent scores can be obtained by choosing the specific scoring function appropriate for data. In particular, they have used the B75 scoring for residuals that are positively skewed in a random drug screening experiment (upper quartile of the residuals are assigned a constant score while the remainder of the residuals are a linear function of their ranks). These scores are estimated diagnostically after the initial Wilcoxon fit to the data produced highly skewed residuals. By diagnostically it is meant that the histogram of the residuals obtained from the Wilcoxon fit is used to estimate a reasonable Bent score. The real purpose behind this procedure of retrospectively using the residuals to estimate the scoring function is to investigate what types of scores are appropriate for the data at hand and must be used with caution in the case of small sample experiments (McKean et al., [21]).

In this work, we also investigate the impact of moderate censoring (up to 50-60%) on these scores for the censored observations as compared to the uncensored observations. If more observations are censored, the residuals generated by a Wilcoxon fit are likely to be positively skewed. By using a Bent score function (such as the B75 score function), we are down-weighing the upper quartile tail of the residuals. The Bent scores are composed of two linear pieces; a linearly increasing piece followed by a flat piece as follows:

$$\varphi_{bent}(u) = \begin{cases} \frac{2}{d(2-d)}u - 1 & if\ \ 0 < u < d \\ \frac{d}{2-d} & if\ \ d \le u < 1 \end{cases} \qquad (9)$$

Here *d* denotes the proportion of the flat piece. For more information on how to generate scores, refer to Policello and Hettsmanperger [22]. The actual scores are standardized as in $\int \varphi_{bent}(u)du = 0$ and $\int \varphi_{bent}{}^{2}(u)du = 1$. In our simulation study we have considered d = 0.25 (B75 scores) as an adjustment to the Wilcoxon fit reflecting the extent to which skewness occurs in the distribution of the residuals.

3. RESULTS AND DISCUSSION

3.1 Simulating the Data

Simulation is conducted for the following three scenarios:

(i) The survival times come from a loglogistic distribution with non-proportional and non-constant hazards for the covariate of interest.

(ii) The survival times come from an exponential distribution with constant and proportional hazards for the covariate of interest.

(iii) The baseline error density is loglogistic but the hazards are proportional for the covariate of interest (discussed in brief only).

The first scenario results in an accelerated failure time (AFT) model where we consider a covariate potentially influencing the survival time. In the log-linear scale, therefore, the error density follows a logistic distribution for which we use a Wilcoxon scoring function that is optimal for this distribution (in the uncensored case). Additionally, we make use of a Bent scoring function in the case of positively skewed residuals (when applicable) resulting from the initial Wilcoxon fit. In the scenario where the error distribution arises from an exponential distribution, both the parametric AFT as well as the Cox proportional hazards (PH) model are applicable. In the uncensored case, the Wilcoxon scores have an asymptotic relative efficiency of 75% when applied to exponentially distributed data (Hettsmanperger and McKean, [15]). However, performance with censoring has not been evaluated and we assess performance in the case of 30% censoring. In the third scenario, we have the situation that the Cox PH model yielding proportional hazards for the covariate is most appropriate, though the baseline hazards are generated from the loglogistic distribution. Thus for this case an AFT model may not be the appropriate choice and incorrectly applying it will reduce the power. Still, we briefly assess the performance of using Wilcoxon scoring function when there is 50% censoring in the data just to get an idea of how much power is lost when a mis-specified method is used.

For the first scenario mentioned in Section 1, we simulated data by generating 1 000 independent data sets of sample size N=100 observations from a loglogistic distribution in the following way. First, the number of simulations M was calculated using the formula given in Burton, Altman, Royston, and Holder [23] which is:

$$M = \left(\frac{Z_{1-\alpha/2}\sigma}{\omega} \right)^2 \tag{10}$$

where ω was kept at 5 per cent level of accuracy of the true regression coefficient b. The value of σ (standard deviation of the regression coefficient *b*) was obtained from 50 pilot runs of the simulation. For various values of coefficients b ranging from -2 to 2, M varied from 700-900. So we set M=1 000 as the number of simulations. Performance evaluation measures such as bias of the estimate of the regression coefficient, mean square error of the estimate of the regression coefficient and coverage percentage of the estimate are evaluated by varying the strength of association of the covariate with the survival times, namely *b* = (-1, -0.75, 0.75). The detailed steps used in simulating the data are provided in the Supplementary material.

We have used R for writing the code. After verifying that our code, for uncensored data, yielded results same as obtained by using the R package Rfit written by Kloke and McKean

[24], we modify it to incorporate censoring using our proposed algorithm in order to conduct the simulations.

3.2 Simulation Results

Table 1 displays the type I errors for these simulations. These results show that for our proposed method, the type I errors are inflated when there is more than 50% censoring in the data in the case of a loglogistic (LLG) error distribution though applying Bent scores alleviates them to a considerable extent (around 60%). Also, Wilcoxon scores yield inflated type I error rates when the underlying distribution is exponential (EXPL) for more than 30% censoring.

Table 1. Percentage Type I error rates for N=100, number of replications=10,000

	Bent75 Scores	Wilcoxon Scores		Cox PH		Parametric AFT model		Logrank on response	
Censor %	LLG errors	LLG errors	EXPL errors	LLG errors	EXPL errors	LLG errors	EXPL errors	LLG errors	EXPL errors
0	-	4.45	4.27	5.38	4.93	5.08	4.89	5.16	5.00
30	1.41	4.64[+]	5.68[+]	5.25	5.27	5.44	4.15	5.04	5.05
50	2.82[+]	6.57[+]	12.60[$]	5.03	4.79	5.64	3.34	4.95	5.07
60	4.20[+]	11.01[$]	18.88[$]	5.00	4.79	5.89	3.37	4.85	5.02

[+] Power simulations are conducted for these scenarios and then compared to the standard approaches
[$] Situations with highly inflated alpha are not considered in the simulations

Only those cases in which the empirical type I error rates are close to the nominal alpha of 5% are considered for generating graphs for comparing the power of the proposed method with the traditional approaches. Power graphs for the first scenario (loglogistic distribution with non-proportional hazard) are displayed in Figure 1(a) through Figure 1(c) for three different levels of censoring (30%, 50%, and 60%). The power graph for the second scenario (exponential distribution with proportional and constant hazard with 30% censoring) is displayed in Figure 1(d). Analogously, Table 2 displays the numerical values for the power calculations shown in Figure 1 (a) through (c). Table 3 displays the simulations representing the second (Figure 1 (d)) and third scenarios (discussed briefly). In these tables, the abbreviations used are: BS = Bent scores, WS = Wilcoxon scores, AF = parametric AFT model, PH = Cox proportional hazards model, LR = logrank scores.

Table 2. Power for N=100; # of replications=1000; distribution=loglogistic (Figure 1(a) – (c))

Reg Coef	Power												
	30 % censoring				50 % censoring					60 % censoring			
	WS	AF	PH	LR	BS75	WS	AF	PH	LR	BS75	AF	PH	LR
0.00	4.5	5.4	5.3	5.0	2.8	5.0	5.7	5.0	4.9	4.2	5.9	5.0	4.9
0.20	16.8	23.8	16.1	17.1	8.6	16.5	20.4	15.9	14.4	9.8	17.3	12.8	13.5
0.40	55.6	64.3	52.3	53.4	34.6	51.4	55.0	43.4	43.5	31.2	49.5	37.9	36.3
0.60	87.8	91.7	84.1	83.8	66.4	84.0	85.5	75.6	73.6	65.6	78.8	68.2	66.5
0.80	98.8	99.0	97.2	96.9	88.0	96.4	97.0	92.0	91.5	85.6	95.1	87.2	85.7
1.00	100.0	100.0	99.5	99.6	98.8	99.6	100.0	98.3	98.4	96.9	98.8	97.1	95.7

From Figure 1 and the tables, for the first scenario which represents non-proportional hazards, Wilcoxon scores provide power somewhat less than what is obtained from a parametric fit of an AFT (using the loglogistic distribution) model for 30% and 50% censoring in data. However, they do provide power slightly more than the (incorrectly applied) PH and LR methods. In case of 50% censoring, the B75 scores yield considerably less power than the Wilcoxon scores. For 60% censoring, the Wilcoxon scores cannot be used as the type I error is inflated and using the conservative B75 scores maybe the only alternative. As expected, an incorrectly specified Cox PH model performs less powerfully than our proposed method (in the case of 30-50% censoring) as does the GLM using logrank scores on the response whereas the parametric AFT model performs best.

For the second scenario which represents constant and proportional hazards arising out of an exponential distribution, the Wilcoxon scores perform relatively well compared to the parametric model, the Cox PH model, and the GLM using logrank scores (as demonstrated by Howard and Koch [2]) on the response for 30% censoring in data. Again this is expected because an exponential distribution is a special case for which both PH and parametric AFT models are appropriate (with the regression coefficients related to each other).

Table 3. Power for N=100; # of replications=1000; Second (Fig 1(d)) and third simulation scenario

Reg Coeff	Power							
	Scenario 2: Exponential Distribution [30 % censoring]					Scenario 3: [50 % censoring]		
	BS75	WS	AF	PH	LR	WS	PH	LR
0.00	2.9	5.8	4.2	5.2	5.1	5.0	5.0	5.0
0.25	6.3	9.8	8.4	9.9	8.6	7.4	7.1	8.0
0.50	20.4	21.8	20.8	22.3	22.8	18.2	21.4	20.8
0.75	34.4	40.8	43.0	45.3	44.1	30.9	38.5	39.1
1.00	53.6	63.4	68.9	69.9	66.6	46.6	60.2	58.5
1.25	68.0	81.6	85.5	85.7	84.5	62.6	75.4	77.4
1.50	84.0	90.6	95.8	95.8	93.9	77.4	87.0	90.3
1.75	96.3	97.1	98.8	98.2	98.5	85.6	94.0	96.7
2.00	99.1	99.3	99.9	99.7	99.4	92.4	98.8	98.6
2.25	99.9	100.0	100.0	100.0	99.9	96.7	99.7	99.7

For the third scenario which represents proportional hazards for the covariate but has non-constant baseline hazards (generated from a baseline loglogistic error density with 50% censoring), the Cox PH and the GLM on logrank scores have expectedly much higher power than the (mis-specified) log-linear model Wilcoxon scores. The parametric AFT model is not used here as in this case it is well known that in this scenario it will not perform well. To further assess the performance of the proposed method, performance evaluation measures such as bias of the estimate of the regression coefficient, mean square error of the estimate of the regression coefficient and coverage percentage of the estimate were used. In all scenarios, we obtained low bias, low mean square error, and adequate coverage (at least 87% in all cases). Table 4 displays the results of these performance evaluation measures for the errors arising out of the loglogistic distribution (representing the first scenario) for three different values of the shape parameter, namely, *s* = {0.25, 0.5, 1}. For *s* = 0.25 and 0.5, the hazard function first increases and then decreases whereas for *s* = 1, the hazard is decreasing. Such hazards are often encountered in clinical trials related to cancer research where the loglogistic and lognormal distribution are used extensively to account for non-monotone hazard functions. In such trials, it is important to summarize the improvement in median survival time following a treatment intervention as opposed to merely specifying a hazard ratio from using a Cox PH model (Royston, [25]).

Table 4 Performance evaluation of the proposed method (N=100, replications=1000)

Scenario	$(\hat{b})$	SE $(\hat{b})$	Bias $(\hat{b})$	% Bias $(\hat{b})$	MSE	%Coverage	% power
50% censored True *b* = -0.75 *s* = 0.5	-0.7350	0.0344	0.0150	1.9960	0.0014	91.4	58.4
50% censored True *b* = 0.75 *s* = 0.25	0.7464	0.0225	-0.0036	0.4827	0.0005	95.5	95.1
50% censored True *b* = 0 *s* = 0.25	-0.0036	0.0225	-0.0036	*	0.0005	95.5	4.5
50% censored True *b* = -1 *s* = 1	-1.0097	0.0638	-0.0097	0.9658	0.0041	88.1	43.1

* indicates % bias cannot be calculated as the true value of *b* = 0 yields a divide by 0 error. AFT

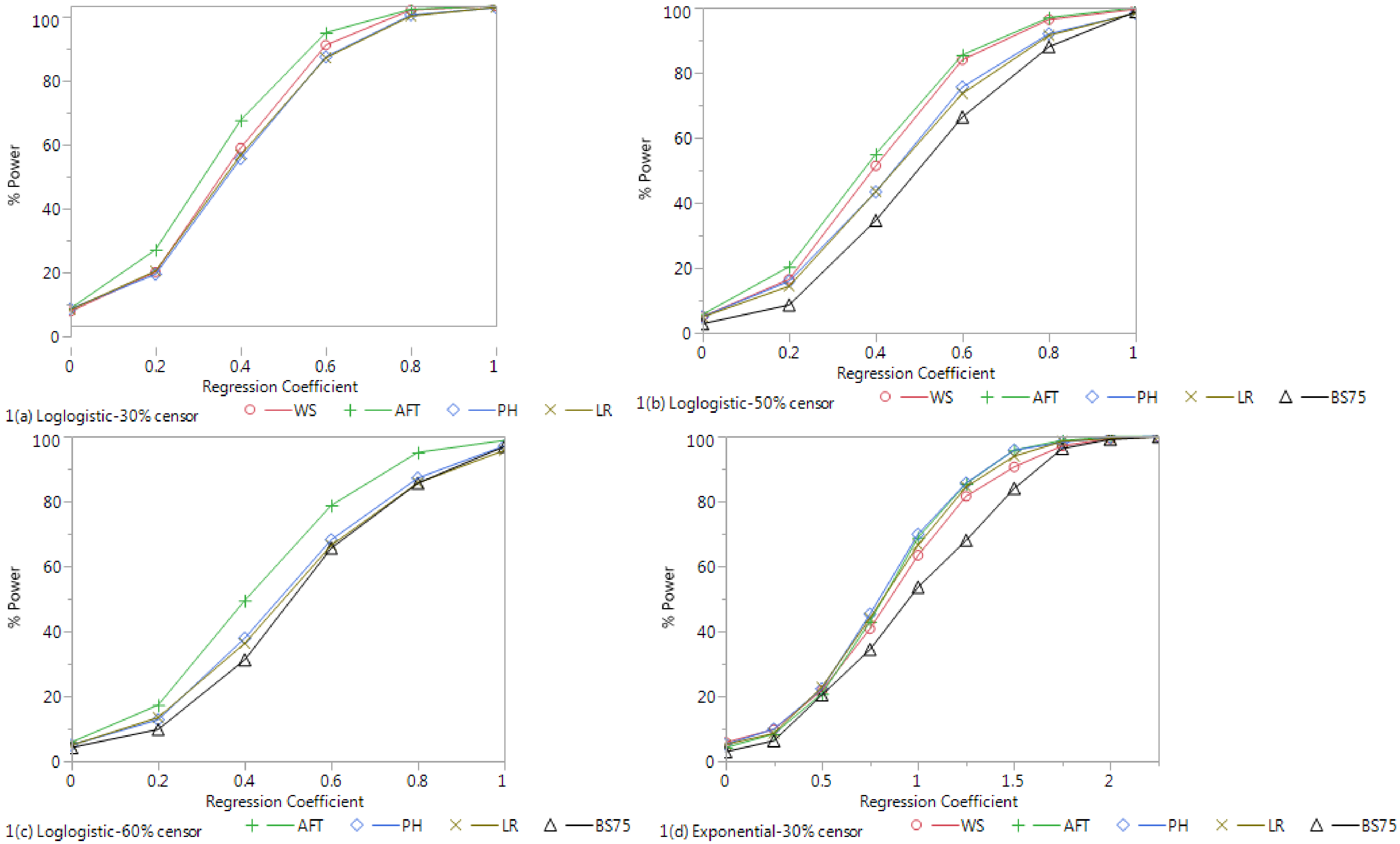

Figure 1 Power graphs for the first (Loglogistic distribution; 30% - 60% censor) and second (Exponential distribution; 30% censor) scenario

3.2 Pancreatic Cancer Study Example

We will demonstrate our method on a data set consisting of 106 patients who were prospectively identified with suspected pancreatic cancer over a 34-month period at the Division of Gastroenterology and Hepatology at the University of Birmingham at Alabama for stent placement [26]. The type of stent placed (plastic or metal) depended on certain evaluation criteria such as presence or absence of liver metastases, whether or not surgery was planned, and the Karnofsky score (K-score) for the patient. The K-score allows patients to be classified in terms of their functional impairment thereby allowing doctors to assess the prognosis in each patient. It is measured on a continuous scale of 0 to 100 in increments of 10 with 100 indicating that the patient shows no evidence of diseases and 0 indicating that the patient faces certain death. Scores between 0-40 represent various gradations of disability and scores between 50-70 represent gradations of self-care ability with assistance. Scores ranging between 80-100 represent gradations of ability to conduct normal activity. Generally, patients with a K-score of more than 70 underwent metal stent placement while those with a score of 70 or lower underwent plastic stent placement, though there were some exceptions. The response measured is the time to death in months. Though other demographic variables and comorbidities are recorded as covariates, prior studies in this field suggest that once the prognosis is made, these are not important predictors of time to death. Thus, we shall initially consider only the K-score as a single continuous predictor of time to death, and later adjust for age as a covariate. This data set contains 68 events (64.2% deaths) while 38 observations (35.8%) were censored due to loss to follow-up. It is expected that all censored observations will die at some stage of pancreotibiliary malignancy, however, due to loss to follow-up there is no option but to treat these observations as censored, thereby carrying incomplete information about these patients.

To analyze these data, various parametric AFT models were fit using the exponential, Weibull, loglogistic, lognormal, and generalized gamma distributions. Table 5 displays the results of these parametric fits with the parameter estimate $\hat{b}$ representing increase in logarithm of time to death per unit increase in the K-score. It can be seen from the log-likelihood and AIC values in this table, that the exponential distribution offers the most parsimonious fit to this dataset. As the K-score has gradations in increments of 10, we also evaluated the increase in time to death per 10-unit increase in the K-score. For the exponential distribution this value was 1.669 (95% CI: 1.438-1.937). We also fit a Cox PH model to this data and this resulted in a hazard ratio (HR) of -0.047 9 (standard error = 0.008 2) per unit increase in the K-score. This corresponds to a HR for time to death of 0.618 (95% CI: 0.527-0.728) per 10-unit increase in the K-score indicating that patients with a higher K-score live longer than those with a lower score. All model fitting assumptions were assessed as per the methods available in standard statistical texts.

Table 5. Parametric fit for the Pancreatic Cancer data (N=106) with K- score as a continuous predictor

Distribution	$(\hat{b})$	SE $(\hat{b})$	Scale/ Shape	*P* value	LL	AIC	$e^{10\hat{b}}$ [95% CI]
Loglogistic	0.0606	0.0109	scale=0.743	<0.001	-134.853	275.707	1.833 [1.480-2.270]
Lognormal	0.0601	0.0104	scale=1.283	<0.001	-133.956	273.913	1.824 [1.488-2.236]
Exponential	0.0512	0.0076	scale=1	<0.001	-134.554	273.109	1.669 [1.438-1.937]
Weibull	0.0511	0.0078	scale=1.005 shape=1	<0.001	-134.553	275.105	1.667 [1.431-1.942]
Generalized Gamma	0.0566	0.010 2	scale=1.187 shape=0.383	<0.001	-133.569	275.138	1.761 [1.442-2.151]

Finally, we fit our proposed method that uses full non-parametric regression using Wilcoxon scoring on the residuals, to this data set (also shown in Table 6). We obtained $\hat{b}$ = 0.045 4 (S.E $(\hat{b})$ = 0.007 67, *P* value < 0.000 1) as the parameter estimate for every one unit increase in the K-score on the logarithmic scale. This corresponds to exp(10 $\hat{b}$) = 1.555 times increase in the time to death per 10-unit increase in K-score (95% CI: 1.314-1.839) again indicating significantly higher longevity for patients with high K-scores as compared to patients with low K-scores.

The Wilcoxon fit of the residuals revealed five outliers with high negative values for the residuals. However, these correspond to five patients who were lost to follow-up immediately after the day of prognosis and hence their survival time was entered in the database as 0.033 months (1 day). All five patients had high Karnofsky scores (four had a score of 90 while one had a score of 80) and these observations correspond to patients about whom the least information was available. The gastroenterologists wanted to ensure that these observations do not influence the interpretation in any way and hence they were removed from the data set. The resulting Wilcoxon fit yielded an estimate of $\hat{b}$ = 0.046 6 (close to the earlier estimate of 0.045 4) with a standard error of 0.008 39 (*P* value < 0.000 1) thereby demonstrating the robustness of the Wilcoxon fit.

As part of a follow-up analysis, the gastroenterologists also wanted to assess the effect of K-score on mortality after adjusting for age. Table 6 shows the results of these analyses in comparison to the best fit parametric (lognormal) AFT model. The lognormal AFT model (second column) suggests that after adjusting for age, every ten unit increase in K-score increases the time to death by a factor of 1.795 whereas the corresponding value for this factor using the proposed model with Wilcoxon scores, is 1.361. However, the lognormal fit also shows age as statistically significant (*P* value=0.041 9) implying that after adjusting for the K-score, every 10-year increase in age decreases the time to death by a factor of 0.773(95% CI: 0.603-0.991), a result that is found to be somewhat surprising by the gastroenterologists. On the other hand, our proposed method with Wilcoxon scoring (third column) does not show age to be statistically significant (*P* value=0.119 1) after adjusting for K-score. The ten-year estimate is found to be 0.8564 (95% CI: 0.705-1.041). The fourth column in Table 6 shows how the results would change if the Normal scores $\varphi(u) = \phi^{-1}(u)$ were used instead of the Wilcoxon scores. If the lognormal distribution were the best fit for the data, then an AFT model would have normally distributed errors, and we could expect comparable results by adopting the Normal scores. On doing so, we find that the parameter estimates for age and K-score are now qualitatively similar to the lognormal model.

Table 6 Parametric and non-parametric fit with two covariates (N = 101)

Covariate specifics		Lognormal AFT	Proposed method (Wilcoxon scores)	Proposed method (Normal scores)
Intercept	b_0	-0.7241	-0.7259	1.5771
	SE(b_0)	1.1609	0.9539	0.8249
	P value	0.5328	0.4466	0.0559
Age	b_1	-0.0258	-0.0155	-0.033 6
	SE(b_1)	0.0127	0.0099	0.0098
	P value	0.0419	0.1191	0.00 6
K-score	b_2	0.0585	0.0459	0.0308
	SE(b_2)	0.0110	0.0085	0.0073
	P value	< 0.001	< 0.001	< 0.001

4. CONCLUSION

Rank based non-parametric methods provide a robust alternative to parametric procedures in terms of their sensitivity to outliers and positive breakdown values for the estimates. In the uncensored case, it is known that the asymptotic efficiency of these methods depends on the optimality of the scoring function used to minimize the dispersion function of the residuals. The Wilcoxon scoring function is optimal for errors from a logistic distribution and reasonably efficient for errors from a normal distribution in a regression setting and hence can be extended to loglogistic and lognormal survival data. The proposed non-parametric method of modifying the Newton-type algorithm used to estimate the regression coefficients appears to work well for moderate random right censoring (up to 50%) in survival data both in the case of proportional and non-proportional hazards. The quality of the model can be assessed by performing a diagnostic check of the distribution of the residuals arising out of the Wilcoxon fit. For severely skewed residuals, the Bent scoring function can be used as an adjustment for higher levels of censoring in the data. In the simulations conducted by us, the B75 scores provided less power than the other methods. In practice, however, one may have to study the distribution of the residuals in greater detail and incorporate other types of Bent scores for modeling particular types of data sets. This procedure is akin to checking the model fits from a Cox PH model or from a parametric fit of the model and should be viewed as a diagnostic checking tool.

In the limited scenarios that we have tested, this method has yielded estimates of the regression coefficients that have low bias, low mean square error, and adequate coverage. In cases where the proportional hazards assumption is not met and there is no clear winner among the popularly used parametric distribution, our proposed method may provide a reasonable alternative non-parametric solution that yields robust estimates of the regression coefficients. Both continuous and categorical predictors may be used allowing the practitioner to draw inferences about the significance of one covariate after adjusting for other covariates in a non-parametric way (though in our simulations we have incorporated only continuous predictors), something which cannot be done in a simple stratified analysis of the standard Kaplan Meier method. It remains to be assessed how this method will perform in the presence of interactions among covariates. This method has also been applied to a real-life data set from a Pancreatic cancer study and it proved to be a robust fit to the outliers present in that data set. Future work aims to compare the performance of this method with the other theoretical nonparametric and semiparametric methods mentioned in Section 1.

ACKNOWLEDGEMENTS

We are grateful to Dr. C. Mel Wilcox for the use of his pancreatic cancer study data.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

CONSENT (WHERE EVER APPLICABLE)

The real-life example discussed is from a previously published abstract and does not require consent from any patients.

REFERENCES

1. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. Am Stat. 1981;35:124-134.

2. Howard G, Koch GG. The GLM logrank test: general linear modeling of logrank scores as a method of analysis for survival data. Commun Stat Simul Comput. 1990;19:903-917.

3. Hougaard P. Analysis of multivariate survival data, New York, NY: Springer; 2001.

4. Buckley J, James I. Linear regression with censored data. Biometrika 1979;66:429-436.

5. Prentice RL. Linear rank tests with censored data. Biometrika 1978;65:167-179.

6. Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. Ann Stat. 1990;18:354-372.

7. Ritov Y. Estimation in a linear regression model with censored data. Ann Stat. 1990;18:303-328.

8. Lai TL, Ying Z. Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. Ann Stat. 1991;19:1370-1402.

9. Lai TL, Ying Z. Rank regression methods for left-truncated and right-censored data. Ann Stat. 1991;19:531-556.

10. Ying Z. A large sample study of rank estimation for censored regression data. Ann Stat. 1993;21:76-99.

11. Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. Biometrika 2003;90(2):341-353.

12. Wei LJ, Ying Z, Lin DY. Linear regression analysis of censored survival data based on rank tests, Biometrika 1990;77:845-851.

13. Lin DY, Greyer CJ. Computational methods for semiparametric linear regression with censored data. J Comput Graph Stat. 1992;1:77-90.

14. Gehan EA. A generalized Wilcoxon test for comparing arbitrarily single censored samples. Biometrika 1965;52:203-223.

15. Hettmansperger TP, McKean JW. Kendalls library of statistics 5: Robust nonparametric statistical methods, New York, NY: Arnold; 1998.

16. McKean JW, Sievers GL. Rank scores suitable for analyses of linear models under asymmetric error distributions. Technometrics 1989;31:207-218.

17. McKean JW, Terpstra JT, Kloke JD. Computational rank-based statistics. Comput Stat. 2009;1:132-140.

18. Terpstra JT, McKean JW. Rank-based analysis of linear models in R. J Stat Softw. 2005;14:1-26.

19. Kapenga JA, McKean JW, Vidmar TJ. Short course on robust statistical procedures for the analysis of linear and nonlinear models. RGLM: Users Manual, American Statistics Association, New Orleans; 1988.

20. Mantel N. Calculation of scores for a Wilcoxon generalization applicable to data subject to arbitrary right censorship. Am Stat. 1981;35:244-247.

21. McKean JW, Vidmar TJ, Sievers GL. A robust two-stage multiple comparison procedure with application to a random drug screen. Biometrics 1989;45:1281-1297.

22. Policello GE, Hettsmanperger TP. Adaptive robust procedures for the one-sample location problem. J Am Stat Assoc. 1976;71:624-633.

23. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Stat Med. 2006;25:4279-4292.

24. Kloke JD, McKean JW. Rfit: Rank based estimation for linear models. R J. 2012;4:57-64.

25. Royston P. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. Stat Neerl. 2001;55:89-104.

26. Wilcox CM, Phadnis MA, Seay T, Varadarajulu S. MO1428 Prospective Evaluation of the Karnofsky Score As a Guide to Placing a Plastic Versus Metal Stent for Pancreaticobiliary Malignancy (4th ed., vol. 73, pp. AB342). Elsevier; 2011.

APPENDIX

A. Newton algorithm by Kapenga et al., [19]

i. Obtain an initial estimate of the regression coefficients, $\hat{\boldsymbol{\beta}}^{(0)}$ (say, least squares estimate) and calculate the initial residuals as $\hat{\mathbf{e}}^{(0)} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(0)}$. Let $\hat{\tau}_{\varphi}^{(0)}$ denote the initial estimate of τ_{φ} based on these residuals. Calculate the dispersion function $D^{(0)}$ evaluated at $\hat{\mathbf{e}}^{(0)}$.

ii. Using the projection matrix $\mathbf{H} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ onto the column space of $\mathbf{X}$, obtain the residuals at the 1st iteration of the algorithm using the relation:

$$\hat{\mathbf{e}}^{(1)} = \hat{\mathbf{e}}^{(0)} - \hat{\tau}_{\varphi}\mathbf{Ha}\{R(\hat{\mathbf{e}}^{(0)})\}$$

where $\mathbf{a}\{R(\hat{\mathbf{e}}^{(0)})\}$ denotes the vector whose i^{th} component is $a\{R(\hat{\mathbf{e}}_i^{(0)})\}$.

iii. Calculate the dispersion function $D^{(1)}$ evaluated at $\hat{\mathbf{e}}^{(1)}$. If $D^{(1)} < D^{(0)}$, this step is considered successful. If not, a linear search can be made along the direction to find a value that minimizes D. In general, the dispersion function at the k^{th} step is denoted by $D^{(k)}$ and a rule to halt the algorithm is established by specifying a tolerance ξ_D such that

$$\frac{D^{(k)} - D^{(k-1)}}{D^{(k-1)}} < \xi_D$$

iv. If $D^{(k)}$ obtains the minimum value for the dispersion function, then find $\hat{\mathbf{Y}}^{(k)} = \mathbf{Y} - \hat{\mathbf{e}}^{(k)}$. Then the optimal estimate of the regression coefficients can be obtained using the relation

$$\hat{\boldsymbol{\beta}}_{\varphi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{Y}}^{(k)}$$

v. Obtain the final estimate of $\hat{\tau}_{\varphi}$ and use it to calculate the standard error of $\hat{\boldsymbol{\beta}}_{\varphi}$ using (5). Obtain $\hat{\alpha}_s$ by finding the median of $\hat{\mathbf{e}}^{(k)}$.

B. Meeting assumptions of Section 2.1

With reference to the proposed method meeting the assumptions in Section 2.1,

$$\int \varphi(u)\,du = \sum_{j=1}^{n} \frac{\mathrm{E}(X_{(j)}) - 0.5}{\varsigma}$$

$$= \frac{1}{\varsigma}\left\{\sum_{j=1}^{n} \mathrm{E}(X_{(j)}) - 0.5n\right\}$$

$$= \frac{1}{\varsigma}\left(\frac{n}{2} - \frac{n}{2}\right)$$

$$= 0$$

Similarly,

$$\int \varphi^2(u)\,du = \sum_{j=1}^{n}\left\{\frac{E(X_{(j)}) - 0.5}{\varsigma}\right\}^2$$

$$= \frac{1}{\varsigma^2}\sum_{j=1}^{n}\left\{\left(\mathrm{E}(X_{(j)}) - \frac{1}{n}\sum_{j=1}^{n} E(X_{(j)})\right)\right\}^2$$

$$= \frac{\mathrm{Var}\left\{\mathrm{E}(X_{(j)})\right\}}{\varsigma^2}$$

$$\mathbf{=1}$$