Parametric Analysis of Renal Failure Data using the Exponentiated Odd Weibull Distribution

Nonhle Channon Mdziniso and Kahadawala Cooray*

Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859, USA

Abstract: In this article, we analyze renal failure data from patients with mesangioproliferative glomerulonephritis (MPGN) which was modeled by [1] non-parametrically using the Kaplan-Meier curve. In their work, they showed that the clinical variables, large increase serum creatinine (LISC) and systolic blood pressure >160 mmHg (SBP>160), and morphological variables, benign nephrosclerosis present (BNP) and interstitial score group 5-6 (IS5-6) were part of the variables which indicated progression to end-stage renal failure (ESRF). Though survival curves associated with these variables may be difficult to model by existing parametric distributions in literature. Therefore, we introduce a four-parameter Odd Weibull extension, the exponentiated Odd Weibull (EOW) distribution which is very versatile in modeling lifetime data that its hazard function exhibits ten different hazard shapes as well as various density shapes. Basic properties of the EOW distribution are presented. In the presence of random censoring, a small simulation study is conducted to assess the coverage probabilities of the estimated parameters of the EOW distribution using the maximum likelihood method. Our results show that the EOW distribution is very convenient and reliable to analyze the MPGN data since it provides an excellent fit for the variables LISC, SBP>160, BNP, and IS5-6. Furthermore, advantages of using the EOW distribution over the Kaplan-Meier curve are discussed. Comparisons of the EOW distribution with other Weibull-related distributions are also presented.

Keywords: Coverage probability, hazard function, maximum likelihood, random censoring, survival function.

1. INTRODUCTION

The Weibull distribution is one of the most popular parametric distributions used in modeling survival data due to its relatively simple hazard, survival, and probability density functions [2]. However, it has some limitations on its hazard function which only accommodates increasing, decreasing, and constant hazard shapes. As a result, extensions of the Weibull distribution have been developed in order to improve flexibility in the hazard function [3, 4].

Specifically, in this paper we first consider a new Weibull extension which has the generalized gamma (GG) [5], exponentiated Weibull (EW) [6], and Odd Weibull (OW) [7] distributions as sub models. The GG, EW, and OW distributions have five common hazard shapes: constant, increasing, decreasing, bathtub, and arc-shape. In addition, all three distributions are Weibull extensions with an extra shape parameter. The GG and EW distributions are well-known distributions in literature. On the other hand, the OW distribution, which was recently developed by [7] considering the distributions of the odds of the Weibull and inverse Weibull distributions, has not yet been studied extensively. Some highlights of the OW distribution in contrast to either GG or EW distribution are as follows. The OW hazard function exhibits 8 different shapes: constant, decreasing, increasing, arc-shape, bathtub, S-shape, inverse-S shape [8], and unimodal. Observe that here we define the arc-shape of a unimodal hazard curve such that there is no inflection point before the mode. Hence arc-shape hazard curves do not have a left tail but unimodal hazard curves do. The OW distribution provides both unimodal and bimodal densities. In addition, in the presence of censoring and truncation, the OW parameters can be estimated in two different ways due to the inverse property of the OW random variable [9]. Observe that the k^{th} integer moment for both GG and EW distributions exist [10, 11]. However, when λ and β of the OW distribution are positive, the k^{th} moments exist, though when both λ and β are negative such that $\lambda\beta < k$, the k^{th} integer moment does not exist [12], meaning that the OW distribution can have heavy right-tail in this case. For example, the extreme value property of the OW distribution is discussed in [8]. Even though the OW has many desirable properties, there is still a need for a single distribution having more advanced hazard shapes for modeling lifetime data. Thus, we develop and study a new four-parameter generalized distribution, the exponentiated Odd Weibull (EOW) distribution which is an extension of the EW and OW distributions. In particular, the hazard function of the EOW distribution produces ten different hazard shapes including more advanced shapes, the M- and W-shapes. In addition, a more generalized version of the EOW distribution is presented for interested readers.

The rest of the article is focused on analyzing renal failure data given in [1] in order to illustrate the flexibility and applicability of the EOW distribution. The renal failure data in this study is from patients with mesangioproliferative glomerulonephritis (MPGN). This MPGN data consist of a record of different clinical and histopathological variables recorded from the time of biopsy on progression to end-stage renal failure (ESRF) and death. Patients with MPGN got their

*Address correspondence to this author at the Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859, USA; Tel: +19897743543; Fax: +19897742414; E-mail: coora1k@cmich.edu

E-ISSN: 1929-6029/18

diagnosis in Norway from April 1988 to December 1990 after a renal biopsy. Vikse *et al.* [1] used the traditional non-parametric, Kaplan-Meier curve to model the data. Motivated by the shapes of their Kaplan-Meier curves, we parametrically fit the MPGN data successfully using the EOW distribution. For comparison purposes, we include the four-parameter exponentiated generalized gamma distribution (EGG) [13] which is an extension of the GG distribution with an extra parameter in order to have the EW as a sub model. However, the EGG distribution does not have any additional density or hazard shapes other than that of either GG or EW distribution.

In addition, the EOW is compared to an alternative exponentiated Odd Weibull (AEOW) distribution introduced in this paper and the Beta-Weibull (BW) distribution [14].

2. NEW STATISTICAL DISTRIBUTIONS

2.1. Generalization of the Three Weibull Extensions

In this section, we present the following six-parameter generalized distribution function which is an extension of the GG, EW, and OW distributions.

$$F(x;\ \lambda,\beta,\theta,\delta,\gamma,k) = \left(\Gamma\left[\log\left\{1+\delta(e^t-1)^\beta\right\},k\right]\right)^\gamma \quad (1)$$

where $t=(x/\theta)^\lambda$, $x>0$, $\theta>0$, $\gamma>0$, $\delta>0$, $k>0$, $\lambda\beta>0$, $\Gamma(v,k)$ is the incomplete gamma function defined by $\Gamma(v,k)=(1/\Gamma(k))\int_0^v x^{k-1}e^{-x}dx$; $k>0$, and $\Gamma(k)$ is the complete gamma function defined by $\Gamma(k)=\int_0^\infty v^{k-1}e^{-v}dv$.

Sub models of (1) which have been studied in literature include the exponentiated generalized gamma distribution [13] (when δ = β = 1), Marshall-Olkin extended Weibull distribution [15] (when k = β = γ = 1), GG distribution (when δ = β = γ = 1), EW distribution (when δ = β = k = 1), OW distribution (when γ = δ = k = 1), and their sub models. When δ = k = 1, we obtain the EOW distribution. Another four-parameter distribution, the gamma Odd Weibull (GOW) distribution with γ = δ = 1 can also be obtained from (1). However, our analysis which is not included in the present paper, showed that the GOW and EOW distributions have similar properties, and hence a similar performance. Therefore the rest of this paper is focused on studying the EOW distribution and its applications in modeling renal failure data.

2.2. EOW Distribution

2.2.1. Density and Hazard Shapes

The cdf, pdf, hazard function, and quantile function of the EOW distribution are respectively given by the following.

$$F_{EOW}(x;\lambda,\beta,\gamma,\theta) = \left\{1-\frac{1}{1+(e^t-1)^\beta}\right\}^\gamma \quad (2)$$

$$f_{EOW}(x;\lambda,\beta,\gamma,\theta) = \frac{\lambda\beta\gamma t e^t}{x}\frac{(e^t-1)^{\beta\gamma-1}}{\{1+(e^t-1)^\beta\}^{\gamma+1}} \quad (3)$$

$$h(x;\lambda,\beta,\gamma,\theta) = \frac{f_{EOW}(x;\lambda,\beta,\gamma,\theta)}{1-F_{EOW}(x;\lambda,\beta,\gamma,\theta)}$$

$$Q(u;\lambda,\beta,\gamma,\theta) = \theta\left[\ln\left\{\left(\frac{u^{1/\gamma}}{1-u^{1/\gamma}}\right)^{1/\beta}+1\right\}\right]^{1/\lambda} \quad (4)$$

where $t=(x/\theta)^\lambda$, $x>0$, $0\le u\le 1$, $\theta>0$, $\gamma>0$, and $\lambda\beta>0$. Observe that the OW is a special case of the EOW distribution when $\gamma=1$. The EW is a special case of the EOW distribution when $\beta=1$. Other special cases from the OW distribution can be found in [7]. Note that when $\beta=-1$, the parameters of the EOW cannot be identified.

The cdf of another distribution having the EW and OW distributions as sub models is given by

$$F(x;\ \lambda_1,\beta_1,\gamma_1,\theta_1) = 1-\frac{1}{1+\left[\frac{1}{\left\{1-\left(1-e^{-t_1}\right)^{\gamma_1}\right\}}-1\right]^{\beta_1}} \quad (5)$$

where $t_1=(x/\theta_1)^{\lambda_1}$, $x>0$, $\theta_1>0$, $\gamma_1>0$, and $\lambda_1\beta_1>0$. We refer to this distribution as the alternative exponentiated Odd Weibull (AEOW) distribution and consider it for comparison purposes in Section 3.

Typical density and hazard shapes for the given parameters of the EOW distribution are shown in Figures **1** and **2**, respectively. Parameters for the EOW density graphs are, respectively, from left to right on first row: ($\lambda=.5,\beta=.5,\gamma=1,\theta=10$), ($\lambda=1.7,\beta=3.5,\gamma=.1,\theta=15$), ($\lambda=9,\beta=.01,\gamma=1,\theta=.5$), ($\lambda=3.33,\beta=.5,\gamma=1,\theta=.3$), and ($\lambda=2.2,\beta=.2,\gamma=2,\theta=3.3$). From left to right on second row: ($\lambda=2.2,\beta=.3,\gamma=2,\theta=3.3$), ($\lambda=-1.8,\beta=-3,\gamma=1.2,\theta=8$), ($\lambda=-2,\beta=-.7,\gamma=.6,\theta=5$), ($\lambda=6,\beta=.5,\gamma=2,\theta=18$), and ($\lambda=8,\beta=.25,\gamma=1,\theta=.6$). Hazard shapes of the EOW distribution include, from left to right on first row: constant ($\lambda=1,\beta=1,\gamma=1,\theta=10$), decreasing ($\lambda=.5,\beta=.5,\gamma=1,\theta=10$), increasing ($\lambda=5,\beta=.5,\gamma=2,\theta=22$), arc-shape ($\lambda=.2,\beta=5,\gamma=1,\theta=10$), bathtub ($\lambda=6,\beta=.1,\gamma=.05,\theta=20$); from left to right on second row: S-shape ($\lambda=.8,\beta=5,\gamma=.1,\theta=15$), inverse-S shape ($\lambda=2,\beta=.3,\gamma=2,\theta=5.9$), M-shape ($\lambda=-1.5,\beta=-3,\gamma=.05,\theta=8$), W-shape ($\lambda=1.2,\beta=3.5,\gamma=.1,\theta=15$), and unimodal ($\lambda=-5,\beta=-.6,\gamma=2.5,\theta=3.5$). Clearly, the EOW has more advanced hazard shapes than the EW and OW distributions. Hence adding a shape parameter on the OW distribution to obtain EOW not only improved the flexibility of the hazard function, but it also created two new shapes: M- and W- shapes. More properties of the EOW distribution are discussed in the subsections that follow.

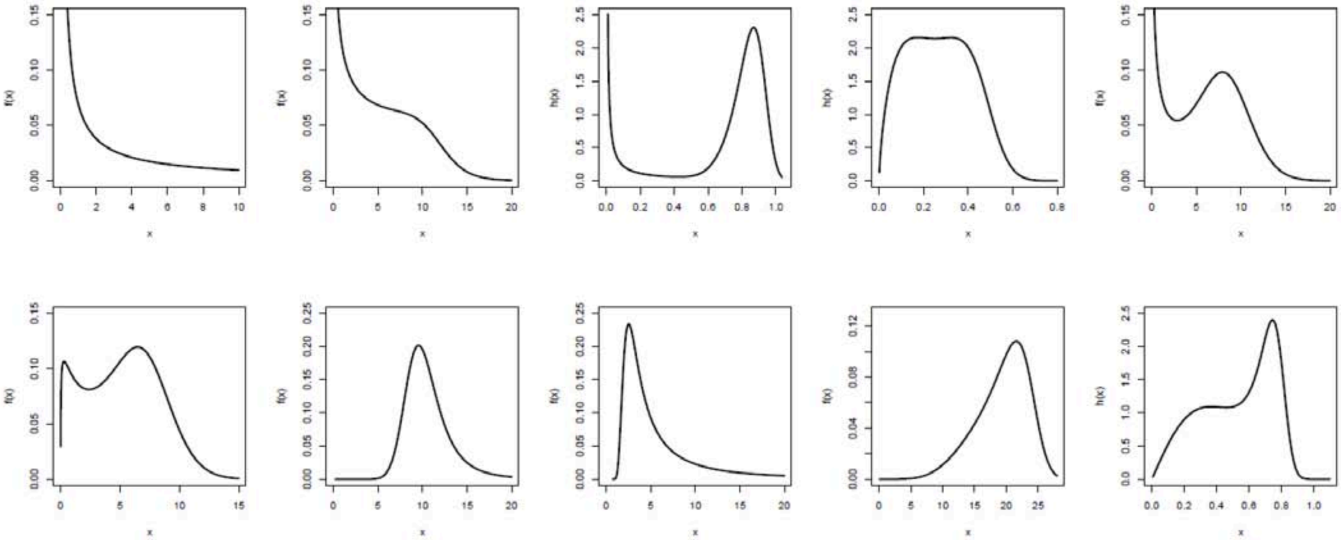

Figure 1: Typical density shapes of the EOW distribution.

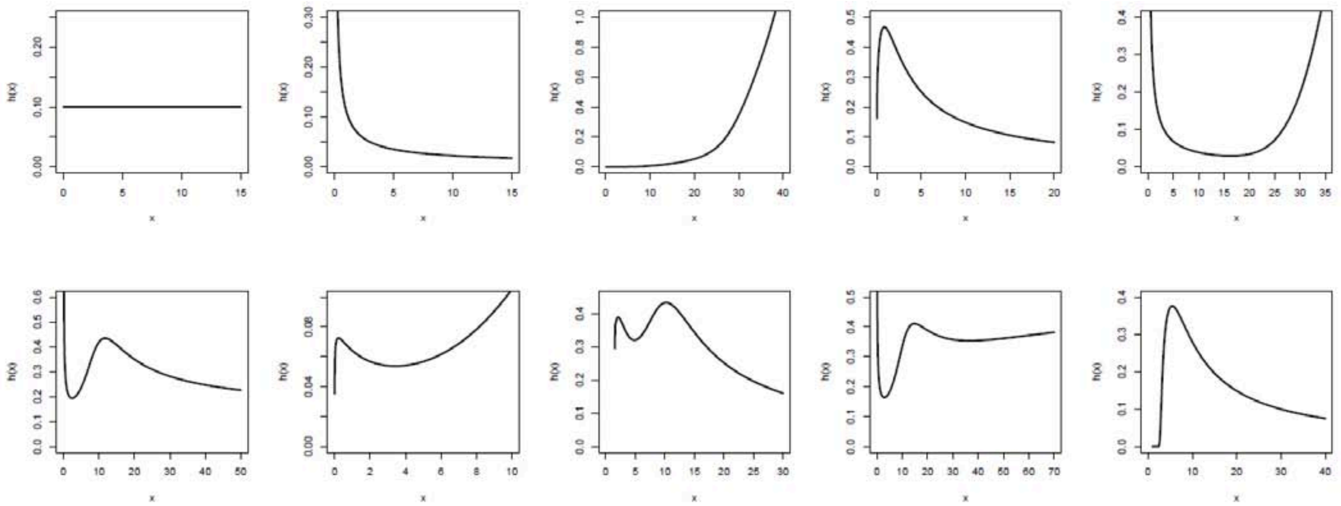

Figure 2: Typical hazard shapes of the EOW distribution.

2.2.2. Moments

The k^{th} positive raw moments of the EOW distribution are given by the following.

$$E[X^k] = \lambda\beta\gamma \int_0^\infty x^{k-1} t e^t \frac{(e^t-1)^{\beta\gamma-1}}{\{1+(e^t-1)^\beta\}^{\gamma+1}}\, dx, \quad (6)$$

where $t = (x/\theta)^\lambda$, $x > 0$, $\theta > 0$, $\gamma > 0$, and $\lambda\beta > 0$. The formula in (6) does not have a closed form solution and have to be evaluated numerically. However, the following theorem provides the condition for the existence of the k^{th} positive raw moments of the EOW distribution.

Theorem 2.1: For $k \in \mathbb{R}^+$, the k^{th} raw moments of the EOW distribution exist finitely for all positive λ and β. When λ and β are negative, the k^{th} raw moments exist finitely when $k < \lambda\beta$.

Proof. Proof can be obtained from the authors upon request.

Furthermore, bounds for the k^{th} positive raw moments are given by the following theorem.

Theorem 2.2: For $k \in \mathbb{R}^+$, the k^{th} raw moments of the EOW distribution have the following finite bounds when $\lambda > 0$ and $\beta > 0$.

$E[(X/\theta)^k] \le \frac{\gamma \cdot 2^{(\beta-1)(\gamma+1)}\Gamma\left(\frac{k}{\lambda}+1\right)}{\beta^{\frac{k}{\lambda}}}$, if $\beta \ge 1 \wedge \beta\gamma \ge 1$;

$E[(X/\theta)^k] \le \frac{\gamma \cdot \Gamma\left(\frac{k}{\lambda}+1\right)}{\beta^{\frac{k}{\lambda}}}$, if $\beta \le 1 \wedge \beta\gamma \ge 1$;

$E[(X/\theta)^k] \ge \frac{\gamma \cdot \Gamma\left(\frac{k}{\lambda}+1\right)}{\beta^{\frac{k}{\lambda}}}$, if $\beta \ge 1 \wedge \beta\gamma \le 1$;

$E[(X/\theta)^k] \ge \frac{\gamma \cdot \Gamma\left(\frac{k}{\lambda}+1\right)}{2^{\gamma+1}\beta^{\frac{k}{\lambda}}}$, if $\beta \le 1 \wedge \beta\gamma \le 1$;

$E[(X/\theta)^k] \le \frac{\beta\gamma \cdot \Gamma\left(\frac{k}{\lambda}+1\right)}{(1+\beta-\beta\gamma)^{\frac{k}{\lambda}+1}}$, if $\beta \le 1 \wedge \frac{1}{\beta} \le \gamma \le 1 + \frac{1}{\beta}$;

$E[(X/\theta)^k] \leq \frac{\beta\gamma\cdot\Gamma\left(\frac{k}{\lambda}+1\right)}{(1+\beta-\beta\gamma)^{\frac{k}{\lambda}+1}}$, if $\beta \geq 1 \wedge \frac{2}{1+\beta} \leq \gamma \leq 1+\frac{1}{\beta}$;

where $x > 0$, $\theta > 0$, $\gamma > 0$, $\lambda\beta > 0$, and $\Gamma(k)$ is the complete gamma function defined by $\Gamma(k) = \int_0^\infty v^{k-1}e^{-v}dv\,; k > 0$.

Proof. Proof can be obtained from the authors upon request.

2.2.3. Parameter Estimation

This subsection presents results on ml estimation of the EOW distribution parameters λ, β, γ, and θ. The derivation is as follows.

Let $t_{1i} = (x_i/\theta)^\lambda$, $t_{2i} = \ln t_{1i}^{1/\lambda}$, $t_{3i} = e^{t_{1i}}$, $t_{4i} = t_{3i} - 1$, $t_{5i} = t_{4i}^{\beta}$, $t_{6i} = t_{4i}^{-\beta}$, $t_{7i} = t_{5i} + 1$, $t_{8i} = \ln t_{4i}$, $t_{9i} = 1 + t_{6i}$, and $t_{10i} = t_{9i}^{\gamma} - 1$. Set $\delta_i = \begin{cases} 0, \text{if censoring;} \\ 1, \text{if death.}\end{cases}$

The EOW loglikelihood function is given by

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \left\{ \ln(\lambda\beta\gamma) + \ln\left(\frac{t_{1i}}{x_i}\right) + t_{1i} + (\beta\gamma - 1)t_{8i} - (\gamma+1)\ln t_{7i} \right\} + \sum_{i=1}^{n}(1-\delta_i)\ln\left\{1 - \left(\frac{t_{5i}}{t_{7i}}\right)^{\gamma}\right\} \quad (7)$$

where $\boldsymbol{\theta} = (\lambda,\beta,\gamma,\theta)'$. Finding the derivatives with respect to $\lambda,\beta,\gamma,\text{and}\,\theta$ of the loglikelihood function in (7) gives the following.

$$\frac{\partial l\boldsymbol{\theta}}{\partial\lambda} = \sum_{i=1}^{n} \delta_i \left\{ \frac{1}{\lambda} + t_{2i}(1+t_{1i}) + (\beta\gamma-1)\frac{t_{1i}t_{2i}}{1-e^{-t_{1i}}} - \beta(\gamma+1)\frac{t_{1i}t_{2i}t_{3i}t_{5i}}{t_{4i}t_{7i}} \right\} - \beta\gamma\sum_{i=1}^{n}(1-\delta_i)\frac{t_{1i}t_{2i}t_{3i}t_{6i}t_{9i}^{\gamma-1}}{t_{4i}t_{10i}} \quad (8)$$

$$\frac{\partial l\boldsymbol{\theta}}{\partial\beta} = \sum_{i=1}^{n} \delta_i \left\{ \frac{1}{\beta} + \gamma t_{8i} - (\gamma+1)\frac{t_{5i}t_{8i}}{t_{7i}} \right\} - \gamma\sum_{i=1}^{n}(1-\delta_i)\frac{t_{6i}t_{8i}t_{9i}^{\gamma-1}}{t_{10i}} \quad (9)$$

$$\frac{\partial l\boldsymbol{\theta}}{\partial\theta} = -\frac{\lambda}{\theta}\sum_{i=1}^{n} \delta_i \left\{ 1 + t_{1i} + (\beta\gamma-1)\frac{t_{1i}t_{3i}}{t_{4i}} - \beta(\gamma+1)\frac{t_{1i}t_{3i}t_{5i}}{t_{4i}t_{7i}} \right\} + \frac{\lambda\beta\gamma}{\theta}\sum_{i=1}^{n}(1-\delta_i)\frac{t_{1i}t_{3i}t_{6i}t_{9i}^{\gamma-1}}{t_{4i}t_{10i}} \quad (10)$$

$$\frac{\partial l\boldsymbol{\theta}}{\partial\gamma} = \sum_{i=1}^{n} \delta_i \left\{ \frac{1}{\gamma} + \beta t_{8i} - \ln t_{7i} \right\} + \sum_{i=1}^{n}(1-\delta_i)\frac{\ln t_{9i}}{t_{10i}} \quad (11)$$

We obtain the ml estimators for $\lambda,\beta,\gamma,\text{and}\,\theta$ by solving the system of equations $\left[\frac{\partial l\boldsymbol{\theta}}{\partial\lambda}, \frac{\partial l\boldsymbol{\theta}}{\partial\beta}, \frac{\partial l\boldsymbol{\theta}}{\partial\theta}, \frac{\partial l\boldsymbol{\theta}}{\partial\gamma}\right]' = \mathbf{0}$. An analytic solution to this system does not exist. However, the following theorem provides the existence of the ml estimators.

Theorem 2.3: The ml estimators for EOW parameters $\lambda,\beta,\gamma,\text{and}\,\theta$ exist.

Proof. Proof can be obtained from the authors upon request.

2.2.4. Coverage Probabilities of the EOW Distribution

In our data analysis, parameters of the EOW distribution are estimated under the maximum likelihood method, which is a large sample procedure for estimating the model parameters. However, survival data more often comes in smaller sample sizes with random censoring. Therefore, it is necessary to assess the coverage probabilities of maximum likelihood (ml) estimators for smaller samples under the random censoring. Hence, a simulation study is conducted to compute the approximate coverage probabilities for estimated parameters of the EOW distribution with 0%, 10%, 20% censoring (cens.), and keeping the intended confidence levels at 90% and 95%. The results, which are summarized in Table **1**, are based on 1000 simulated random samples from the EOW distribution.

The random samples are generated by plugging the known values of parameters $\lambda,\beta,\gamma,\text{and}\,\theta$ (say $\lambda = 2, \beta = 0.2, \gamma = 2, \theta = 1$) to the EOW quantile function given in (4). In addition, n (say $n = 100$) number of ordered uniform random sample from the uniform distribution, u~ U(0,1) is required to substitute as u in the quantile function. In that way, one random sample with size n from the EOW distribution with parameters $\lambda,\beta,\gamma,\text{and}\,\theta$ can be generated. In this simulation study, one thousand such samples are generated to get a single cell value in Table **1**. The approximate $100(1-\alpha)\%$ confidence intervals for parameters $\lambda,\beta,\gamma,\text{and}\,\theta$ are calculated by using $\left(\hat{\lambda} - Z_{\frac{\alpha}{2}}SE_{\hat{\lambda}}, \hat{\lambda} + Z_{\frac{\alpha}{2}}SE_{\hat{\lambda}}\right)$, $\left(\hat{\beta} - Z_{\frac{\alpha}{2}}SE_{\hat{\beta}}, \hat{\beta} + Z_{\frac{\alpha}{2}}SE_{\hat{\beta}}\right)$, $\left(\hat{\gamma} - Z_{\frac{\alpha}{2}}SE_{\hat{\gamma}}, \hat{\gamma} + Z_{\frac{\alpha}{2}}SE_{\hat{\gamma}}\right)$, and $\left(\hat{\theta} - Z_{\frac{\alpha}{2}}SE_{\hat{\theta}}, \hat{\theta} + Z_{\frac{\alpha}{2}}SE_{\hat{\theta}}\right)$, respectively, where $\hat{\lambda}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\theta}$ are the ml estimators for $\lambda,\beta,\gamma,\text{and}\,\theta$, respectively; $SE_{\hat{\lambda}}$, $SE_{\hat{\beta}}$, $SE_{\hat{\gamma}}$, and $SE_{\hat{\theta}}$ are, respectively, the asymptotic standard errors of $\hat{\lambda}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\theta}$ which are taken from the observed information matrix.

From Table **1**, one can clearly see for uncensored data (0% cens.), the approximate coverage probabilities for the parameters get closer to the intended coverage probabilities as the sample size

Table 1: The Approximate Coverage Probabilities of EOW Distribution Based on 1000 Simulations

Parameters		90% intended						95% intended					
$\theta = 1$		0% cens.		10% cens.		20% cens.		0% cens.		10% cens.		20% cens.	
λ, β, γ	n →	100	400	100	400	100	400	100	400	100	400	100	400
2.0, 0.2, 2.0	λ	.92	.91	.45	.31	.19	.05	.96	.95	.52	.38	.23	.07
	β	.92	.91	.50	.44	.39	.16	.96	.96	.57	.50	.52	.21
	γ	.90	.90	.54	.57	.17	.08	.94	.95	.59	.62	.20	.10
	θ	.90	.90	.52	.48	.25	.17	.95	.95	.57	.55	.29	.20
4.0, 0.2, 2.0	λ	.93	.90	.51	.33	.25	.13	.96	.95	.57	.41	.32	.15
	β	.93	.90	.50	.43	.31	.14	.96	.95	.56	.48	.40	.17
	γ	.89	.91	.56	.57	.13	.05	.93	.96	.60	.62	.14	.06
	θ	.89	.89	.57	.49	.23	.16	.94	.95	.62	.54	.27	.18
2.5, 0.5, 0.5	λ	.89	.91	.78	.78	.23	.01	.93	.96	.79	.85	.24	.02
	β	.88	.90	.70	.89	.14	.00	.91	.94	.72	.95	.17	.00
	γ	.92	.92	.58	.63	.07	.00	.95	.96	.61	.72	.09	.00
	θ	.91	.91	.60	.71	.03	.00	.94	.95	.64	.79	.06	.00

increases. When censoring percentage increases and the sample sizes becomes larger, the approximate coverage probabilities for the parameters become smaller than intended coverage probabilities. However, these low probabilities are acceptable since this is an expected result in the presence of censoring. Observe that in Table **1**, the shape parameter values $\lambda, \beta,$ and γ are chosen to represent different hazard and density shapes. Therefore, we only consider the scale parameter θ=1 for all the cases considered here.

3. RENAL FAILURE DATA EXAMPLE

In this section, the parametric distribution, EOW is applied to model MPGN data which was modeled non-parametrically using the Kaplan-Meier curve by [1]. For comparison purposes, the EGG is considered since it is the four-parameter distribution having both GG and EW as sub models. In addition, the EOW is compared to the AEOW in (5) and BW distribution. The MPGN data consists of 273 patients. All patients were followed for a median duration of 34.8 months (0.8 - 68 months). One person died at the beginning of the study, thus our analysis is based on 272 patients.

Vikse *et al.* [1] showed that large increase serum creatinine (LISC) and systolic blood pressure >160 mmHg (SBP>160) were part of the clinical variables which indicated progression to end-stage renal failure (ESRF). Morphological variables indicating progression to ESRF included benign nephrosclerosis Present (BNP) and interstitial score group 5-6 (IS5-6). Moreover, survival curves associated with LISC, SBP>160, BNP, and IS5-6 may be difficult to model by existing parametric distributions in literature. Specifically, these survival curves exhibit an upper S-shape with random censoring. Therefore we analyze the variables; LISC, SBP>160, BNP, and IS5-6 using the more advanced parametric model presented in this paper. Observe that data from other categories of the variables; benign nephrosclerosis absent (BNA), serum creatinine normal (NSC) and moderate increase (MISC), interstitial score 0-1 (IS0-1) and 2-4 (IS2-4), systolic blood pressure <140 mmHg (SBP<140) and 140-160 mmHg (SBP140-160) can be fitted using more simpler and well-known distributions such as lognormal, gamma, and Weibull. Thus, they were only included in this analysis for the sake of completeness of the study. Our study is only focused on analyzing the variable categories; LISC, SBP>160, BNP, and IS5-6. The maximum likelihood estimation method was used to estimate parameters. Our analysis is based on the Akaike information criteria (AIC), Bayesian information criteria (BIC), Kolmogorov Smirnov (KS) test statistic for randomly right censored data defined in [16], and Cramér-von Mises (CM) type distances defined by the sum of the squared KS distances for each complete data.

RESULTS AND DISCUSSIONS

Table **2** shows results of the analysis for the variables; serum creatinine, systolic blood pressure, benign nephrosclerosis, and interstitial score. Results for the variable categories LISC, SBP>160, BNP, and IS5-6 have been highlighted. Note that standard errors of the estimates are given in parenthesis, and the best values from each test have been highlighted. It is clear from the results that the EOW is a better distribution for fitting the variables BNP, LISC, and SBP>160 based on all the four tests since it has the lowest value for each test statistic. In particular, results from the CM

Table 2: Estimated Values of Fitted Distributions for MPGN Data

Distribution		Likelihood estimates				
		Benign nephrosclerosis		**Serum creatinine**		
		Present	Absent	Normal	Mod. increase	**Large increase**
EGG						
	$\hat{\lambda}$	0.1309(0.0911)	0.1155(0.0740)	0.4868(1.1105)	0.6709(1.4880)	0.3801(0.1876
	$\hat{\gamma}$	2.9942(8.0721)	3.0484(6.1908)	1.1435(7.8064)	0.0461(0.1273)	0.1307(0.1236)
	$\hat{k}$	1.9308(3.3715)	2.1812(2.5605)	5.7254(14.416)	19.992(16.557)	9.9563(0.6999)
	$\hat{\theta}$	0.7263(4.8141)	1.1829(6.6357)	10.106(5.5539)	45.000(287.78)	2.5347(3.5036)
	NLL	60.7	72.6	17.5	38.9	57.1
	AIC	129.4	153.2	42.9	85.8	122.1
	BIC	138.8	166.3	55.7	93.8	128.5
	KS	0.2135	0.0232	0.0174	0.1630	0.1745
	CM	0.0476	0.0010	0.0004	0.0300	0.0385
EOW						
	$\hat{\lambda}$	6.1361(0.1691)	0.8577(0.7118)	-7.0812 (2.7745)	1.5335(0.9451)	6.2009(0.1710)
	$\hat{\beta}$	0.0309(0.0113)	0.1209(0.0974)	-0.0805 (0.0545)	0.0726(0.0619)	0.0549(0.0191)
	$\hat{\gamma}$	3.2666(0.5211)	5.8614 (1.5807)	34.118 (35.800)	4.8498(1.2500)	1.6609(0.4061)
	$\hat{\theta}$	30.960(3.2885)	8.4629(14.330)	1.2265(0.8981)	9.4839(8.9007)	33.469(4.1756)
	NLL	**57.7**	71.7	17.3	37.3	**55.3**
	AIC	**123.4**	151.4	42.6	82.6	**118.6**
	BIC	**132.8**	164.4	55.4	90.6	**125.0**
	KS	**0.1368**	0.0149	0.0169	0.1087	**0.1011**
	CM	**0.0209**	0.0006	0.0004	0.0147	**0.0222**
AEOW						
	$\widehat{\lambda_1}$	0.0896(0.2852)	0.0356(0.2717)	0.1191(0.0043)	0.0426(0.2066)	4.5205(0.6304)
	$\widehat{\beta_1}$	1.8640(4.7760)	7.1325(62.416)	0.1986(0.0346)	6.8760(38.537)	0.3684(0.3980)
	$\widehat{\gamma_1}$	4.6899(23.464)	2.4516(7.7504)	52.032(0.1932)	2.5028(5.7230)	0.3121(0.3364)
	$\widehat{\theta_1}$	0.7198(32.486)	0.4813(19.401)	32.722(13.989)	0.2160(5.6569)	90.483(75.528)
	NLL	60.7	72.4	18.7	39.0	56.8
	AIC	129.4	152.8	45.3	85.9	121.6
	BIC	138.7	165.9	58.1	94.0	128.0
	KS	0.2128	0.0219	0.0257	0.1680	0.1286
	CM	0.0475	0.0008	0.0007	0.0320	0.0223
BW						
	$\hat{c}$	0.0658(0.0201)	1.2572(0.6812)	1.4189(0.0002)	2.8226(0.8922)	0.0981(0.0265)
	$\hat{g}$	2.3651(22.431)	5.5216(7.7363)	3.892 (0.3576)	14.220(5.9869)	9.7466(63.521)
	$\hat{a}$	18.503(6.7395)	0.2497(0.2090)	1.5225(0.9546)	0.1162(0.0632)	11.702(4.2577)
	$\hat{b}$	4.9374(4.0999)	0.0037(0.0043)	0.0005(0.0004)	0.0053(0.0024)	4.8356(2.3485)
	NLL	60.8	71.6	17.7	37.4	57.1
	AIC	129.5	151.2	43.4	82.7	122.1
	BIC	130.0	164.3	56.2	90.7	128.5
	KS	0.2151	0.0145	0.0206	0.1046	0.1900
	CM	0.0483	0.0006	0.0005	0.0143	0.0423

(Table 2). Continued.

Distribution		Likelihood estimates					
		Interstitial score			Systolic BP		
		$0-1$	$2-4$	$\mathbf{5-6}$	< 140 mmHg	$140-160$ mmHg	$\mathbf{> 160}$ **mmHg**
EGG							
	$\hat{\lambda}$	0.2573(0.1529)	0.1788(0.3355)	1.2399(2.8242)	0.3126(0.5865)	0.2311(0.1552)	0.2603(0.1508)
	$\hat{\gamma}$	19.806(43.523)	0.7526(3.7933)	0.0182(0.0414)	0.6627(2.3258)	0.6656(0.9638)	0.1257(0.1238)
	$\hat{k}$	1.3469(0.2354)	3.8774(6.9774)	21.903(2.8583)	6.2684(7.7882)	5.1428(4.4912)	11.394(1.399)
	$\hat{\theta}$	2.1791(3.6923)	3.0398(6.0499)	54.990(223.03)	5.0632(49.193)	4.6068(16.722)	1.5117(4.7642)
	NLL	17.4	40.6	56.5	46.6	35.2	43.9
	AIC	42.7	89.1	121.0	101.2	78.4	95.8
	BIC	55.7	96.6	127.2	113.2	87.6	102.4
	KS	0.0167	0.1800	0.1833	0.0272	0.0485	0.3644
	CM	0.0004	0.0348	0.0428	0.0010	0.0029	0.1438
EOW							
	$\hat{\lambda}$	8.2871(0.1918)	2.6359(3.3918)	2.4084(2.9182)	0.8030(1.0534)	0.0667(0.1620)	6.3350(0.1540)
	$\hat{\beta}$	0.1047(0.0786)	0.0703(0.0850)	0.1106(0.1027)	0.1746(0.2143)	1.9796 (4.6743)	0.0303(0.0113)
	$\hat{\gamma}$	3.9240(1.5236)	3.1498(0.8389)	2.2718(0.9309)	7.6257(4.8790)	13.684(30.053)	2.3895(0.5046)
	$\hat{\theta}$	92.601(0.6098)	25.403(17.442)	21.647(19.351)	7.1970(25.392)	3.8172(36.397)	29.556(3.2993)
	NLL	17.4	40.0	**55.3**	46.3	35.1	**39.8**
	AIC	42.8	87.9	**118.6**	100.5	78.2	**87.6**
	BIC	55.8	95.4	**124.9**	112.5	87.4	**94.3**
	KS	0.0171	0.1504	0.1350	0.0211	0.0519	**0.2103**
	CM	0.0004	0.0255	0.0292	0.0008	0.0032	**0.0556**
AEOW							
	$\widehat{\lambda_1}$	0.1526(0.5291)	0.1453(0.2284)	0.0975(0.2629)	0.0571(0.3707)	0.1690(0.06498)	0.1317(0.4665)
	$\widehat{\beta_1}$	4.2135(24.647)	0.2282(0.5500)	2.3822(7.5314)	9.2659(66.503)	0.1911(0.0455)	0.8000(1.1349)
	$\widehat{\gamma_1}$	7.7955(39.912)	21.216(131.14)	3.2247(6.7826)	2.2807(5.2226)	34.010(4.8712)	5.6444(32.687)
	$\widehat{\theta_1}$	0.7329(7.7608)	9.8174(142.71)	0.7853(11.848)	3.5632(76.160)	17.714(1.6897)	2.9887(112.03)
	NLL	17.5	40.6	56.9	46.5	35.2	43.8
	AIC	43.0	89.2	121.8	101.0	78.3	95.6
	BIC	55.9	96.7	128.0	113.0	87.4	102.2
	KS	0.0170	0.1861	0.2087	0.0260	0.0515	0.3690
	CM	0.0004	0.0372	0.0529	0.0010	0.0032	0.1461
BW							
	$\hat{c}$	0.0897(0.0726)	0.3940(1.0821)	5.3667(0.2052)	0.6765(0.0001)	0.5302(0.0919)	0.2557(0.1501)
	$\hat{g}$	0.0047(0.0356)	0.0090(0.0226)	32.202(5.8396)	0.2362(0.0275)	0.0723(0.0647)	0.0001(0.0004)
	$\hat{a}$	189.17(5.7357)	1.8161(63.697)	0.0758(0.0265)	1.4960(0.7989)	5.7361(7.6981)	32.174(5.6359)
	$\hat{b}$	12.468(2.0485)	0.0066(0.0658)	0.0273(0.0172)	0.0015(0.0007)	0.0030(0.0014)	0.0130(0.0169)
	NLL	17.4	40.4	55.5	47.0	35.1	42.9
	AIC	42.7	88.8	119.0	102.0	78.1	93.8
	BIC	55.7	96.3	125.2	114.0	87.2	100.5
	KS	0.0166	0.1792	**0.1265**	0.0324	0.0518	0.3691
	CM	0.0004	0.0346	**0.0277**	0.0014	0.0033	0.1486

Table 3: Survival Rates in Percentages for MPGN Data Based on EOW Distribution

	% Survival beyond given # of months									
Variable	**10 mo.**	**20 mo.**	**30 mo.**	**40 mo.**	**50 mo.**	**60 mo.**	**65 mo.**	**70 mo.**	**80 mo.**	**90 mo.**
BNP	92.81	90.93	89.48	86.88	76.47	39.62	15.70	3.19	0.01	0.00
LISC	78.29	72.80	68.97	63.99	49.90	18.24	5.48	0.80	0.00	0.00
IS5-6	83.63	78.63	73.52	66.41	56.03	42.48	35.10	27.84	15.34	7.04
SBP>160	85.30	82.55	80.43	75.91	57.40	14.59	2.71	0.19	0.00	0.00

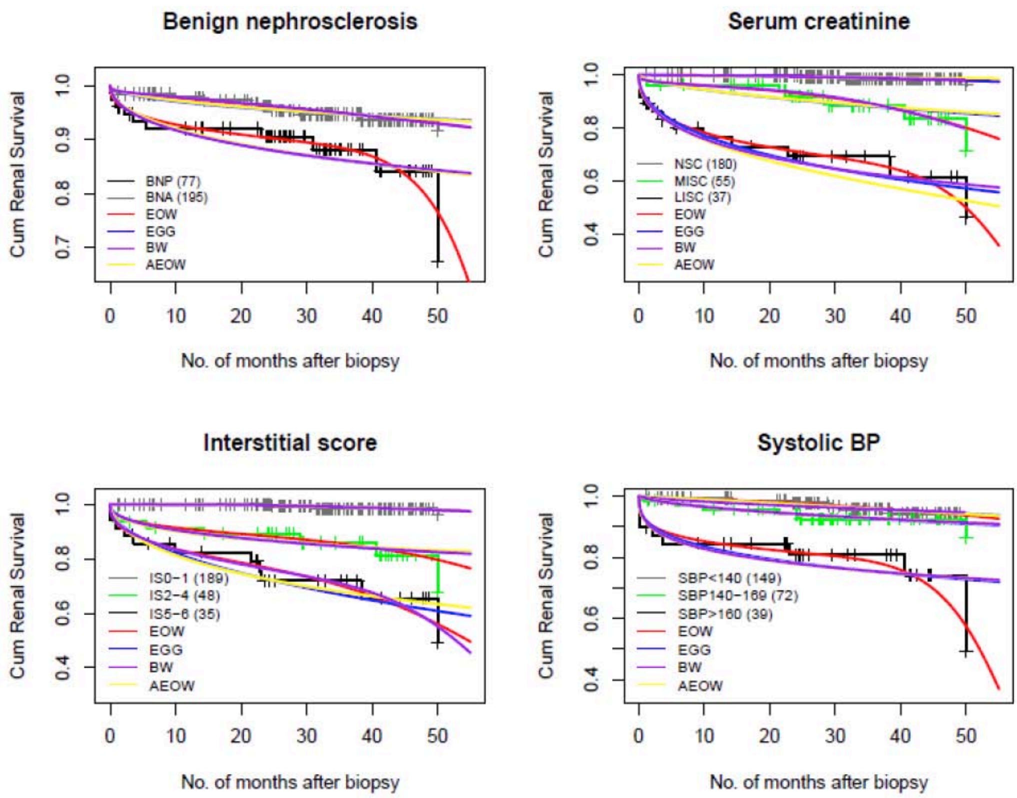

Figure 3: Fitted survival curves of EGG (blue), BW (purple), AEOW (yellow), and EOW (red) distributions along with Kaplan-Meier curves for the MPGN data.

test show that the EOW gives the best overall fit for these variables. EOW also provides the best fit for IS5-6 based on the AIC and BIC values, while the BW distribution performs better based on the KS and CM values. Observe that some of the parameter estimates for the other variables have higher standard errors for the distributions given in Table **2**, for example, estimates of the parameter θ for IS5-6. The reason here is that we are overfitting these variables using four parameters.

Figure **3** shows the fitted survival curves of the distributions; EGG, BW, AEOW, and EOW, along with Kaplan-Meier curves illustrating the effects of benign nephrosclerosis, serum creatinine, interstitial score, and systolic blood pressure on cumulative (Cum) renal survival for patients with MPGN. Note that the number of patients in the categories for each variable are given in parenthesis. Results from the fitted survival curves for the variables show that the EOW distribution outperforms the fit of the distributions considered in all the variables, in particular; LISC, SBP>160, BNP, and IS5-6, which agrees with the results from Table **2**. BW also provides a good fit for IS5-6, which agrees with the KS and CM values from Table **2** for this category. Thus we conclude that the EOW survival function provides a good parametric estimate for the Kaplan-Meier curve used by [1] to model the MPGN data.

Therefore, we can predict renal survival for the subgroups of patients studied using the EOW survival function. In particular, for patients with BNP, the cumulative probability of progression to ESRF in three years is 0.12 based on the EOW survival function, as compared to 0.13 obtained by [1] using the Kaplan-Meier curve. For patients with LISC, the cumulative probability of progression to ESRF in three years is 0.34 based on the EOW survival function, as compared to 0.32 obtained by [1] using the Kaplan-Meier curve. For patients with IS5-6, the cumulative probability of progression to ESRF in three years is 0.30 based on the EOW survival function, as compared to 0.28 obtained by [1] using the Kaplan-Meier curve. In addition, for patients with SBP>160, the cumulative probability of progression to ESRF in three years is 0.22 based on the EOW survival function, as compared to 0.21 obtained by [1] using the Kaplan-Meier curve. As mentioned in [17], the non-continuous nature of the Kaplan-Meier curve emphasizes that they are not smooth functions, but rather step-wise estimates; thus, calculating a point survival can be difficult. However, a point survival can easily be calculated directly using the EOW survival function, which is an advantage of using the EOW survival function over the Kaplan-Meier curve. Table **3** shows survival rates in percentages for MPGN data based on the EOW distribution. Each cell entry gives the percentage of patients who survive beyond the given number of months (mo.) for the variable categories LISC, SBP>160, BNP, and IS5-6. Our results show that patients with IS5-6 have a higher survival rate with 7.04% of patients surviving beyond 90 months after biopsy. Patients with BNP, LISC, and SBP>160 have no chance of surviving beyond 80 months after biopsy. Clearly, one can see effects of the upper-S shaped survival function of these variables which results in a sudden drop in percentage survival for patients with BNP, LISC, and SBP>160 after 50 months. For patients with IS5-6, the highest drop in percentage survival also occurs after 50 months, though it is lower compared to other variables. These results show that the EOW distribution is easy to interpret compared to the Kaplan-Meier curve since it is parametric.

Furthermore, as part of our analysis, we examined the fit of other well-known Weibull-related distributions for these data sets. Our results showed that the distributions give a poor fit for LISC, SBP>160, BNP, IS5-6, and some of the other categories of the variables. For example, the standard 2-parameter Weibull fit gives a poor fit with large standard errors for its parameter estimates. Thus we did not include this work in Table **2**.

4. CONCLUSIONS

The MPGN data [1] was analyzed using the EOW distribution introduced in this study. Results showed that the EOW distribution provides an excellent fit for the variables LISC, SBP>160, BNP, and IS5-6, which may be difficult to model by existing parametric distributions in literature. In particular, based on the distributions that we compared in this paper, the EOW survival function provides the best parametric estimate for the Kaplan-Meier curve used by [1] to model this data. Thus we were able to predict renal survival for the subgroups of patients studied using the EOW survival function. Our analysis showed that patients with IS5-6 have a higher chance of surviving beyond 90 months after biopsy. On the other hand, patients with BNP, LISC, and SBP>160 have no chance of surviving beyond 80 months after biopsy. These results show that the EOW distribution is easy to use and interpret compared to the Kaplan-Meier curve since it is parametric. Therefore, the EOW distribution is very useful for modeling renal failure data under parametric modeling framework.

ACKNOWLEDGEMENTS

The authors wish to specially thank Professor Vikse for providing the MPGN data used in [1].

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

REFERENCES

[1] Vikse BE, Bostad L, Aasarod K, Lysebo DE, and Iversen BM. Prognostic factors in mesangioproliferative glomerulonephritis. Nephrol Dial Transplant 2002; 17: 1603-1613. https://doi.org/10.1093/ndt/17.9.1603

[2] Klein JP and Moeschberger ML. Survival analysis techniques for censored and truncated data. New York: Springer-Verlag; 2003.

[3] Rinne H. The Weibull distribution. Boca Raton; Chapman and Hall, CRC; 2009.

[4] Murthy DNP, Xie M, and Jiang R. Weibull Models. Hoboken; Wiley Interscience: 2004.

[5] Stacy EW. A generalization of the gamma distribution. Ann Math Statist 1962; 33: 1187-1192. https://doi.org/10.1214/aoms/1177704481

[6] Mudholkar GS, Srivastava DK, and Freimer M. The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. Technometrics 1995; 37: 436-445. https://doi.org/10.1080/00401706.1995.10484376

[7] Cooray K. Generalization of the Weibull distribution: the odd Weibull family. Stat Modelling 2006; 6: 265-277. https://doi.org/10.1191/1471082X06st116oa

[8] Jiang H, Xie M, and Tang LC. On the Odd Weibull Distribution. Proc Inst Mech Eng O J Risk Reliab 2008; 222: 583-594. https://doi.org/10.1243/1748006XJRR168

[9] Cooray K. Analyzing grouped, censored, and truncated data using the Odd Weibull family. Commun Stat Theory Methods 2012; 41: 2661-2680. https://doi.org/10.1080/03610926.2011.556294

[10] Johnson NL, Kotz S and Balakrishnan N. Continuous univariate distributions. New York: Wiley; 1994.

[11] Choudhury A. A simple derivation of moments of the exponentiated Weibull distribution. Metrika 2005; 62: 17-22. https://doi.org/10.1007/s001840400351

[12] Cooray K. A study of moments and likelihood estimators of the odd Weibull distribution. Stat Methodol 2015; 26: 72-83. https://doi.org/10.1016/j.stamet.2015.03.003

[13] Cox C and Matheson M. A comparison of the generalized gamma and exponentiated Weibull distributions. Stat Med 2014; 33: 3772-3780. https://doi.org/10.1002/sim.6159

[14] Lee C, Famoye F, and Olumolade O. Beta-Weibull distribution: some properties and applications to censored

data. J Mod Appl Stat Methods 2007; 6: 173-186.
https://doi.org/10.22237/jmasm/1177992960

[15] Marshall AW and Olkin I. A new method of adding a parameter to a family of distributions with applications to the exponential and Weibull families. Biometrika 1997; 84: 641-652.
https://doi.org/10.1093/biomet/84.3.641

[16] Guilbaud O. Exact Kolmogorov-type tests for left-truncated and (or) right-censored data. J Am Stat Assoc 1988; 83: 213-221.
https://doi.org/10.1080/01621459.1988.10478589

[17] Rich JT, *et al.* A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg 2010; 143: 331-336.
https://doi.org/10.1016/j.otohns.2010.05.007

Received on 15-04-2018 Accepted on 02-05-2018 Published on 25-06-2018

https://doi.org/10.6000/1929-6029.2018.07.03.5