A New Generalized-X Family of Distributions: Applications, Characterization and a Mixture of Random Effect Models

Rasool Roozegar¹*, Getachew Tekle², G. G. Hamedani³

*Corresponding author

1 Department of Statistics, Yazd University, Yazd, Iran, rroozegar@yazd.ac.ir
2 Department of Statistics, Yazd University, Yazd, Iran, getch.55tekle@gmail.com
3 Department of Mathematics, Statistics and Computer Science, Marquette University USA, gholamhoss.hamedani@marquette.edu

Abstract

The researchers in applied statistics are recently highly motivated to introduce new generalizations of distributions due to the limitations of the classical univariate distributions. In this study, we propose a new family called new generalized-X family of distributions. A special sub-model called new generalized-Weibull distribution is studied in detail. Some basic statistical properties are discussed in depth. The performance of the new proposed model is assessed graphically and numerically. It is compared with the five well-known competing models. The proposed model is the best in its performance based on the model adequacy and discrimination techniques. The analysis is done for the real data and the maximum likelihood estimation technique is used for the estimation of the model parameters. Furthermore, a simulation study is conducted to evaluate the performance of the maximum likelihood estimators. Additionally, we discuss a mixture of random effect models which are capable of dealing with the overdispersion and correlation in the data. The models are compared for their best fit of the data with these important features. The graphical and model comparison methods implied a good improvement in the combined model.

Key Words: NG-X Family; NG-Weibull distribution; Weibull distribution; Simulation; Combined model; Random effect models; Overdispersion; Correlation.

Mathematical Subject Classification: 60E05, 62E15.

1. Background

Applied statisticians are highly attracted to the probability distribution theory. The development and introduction of new families of distributions have been an additional interest for the expertise in applied statistics. The adaptation, extension, modification and application of additional parameters to the existing models are some of the techniques commonly used.

Theoretically and practically, some distributions have natural flexibility features. This is the main attracting reason for the area to become more interesting.

Recently, Chipepa et al. (2020) proposed a new generalized family of distributions called the odd generalized half logistic Weibull-G family of distributions. They also considered some special baseline distributions to discuss some structural properties of the new proposed family of distributions using the failure time’s data. They have shown this
model performs best among its competitor models.

Ahmad and Hussain (2017) have proposed a new extended flexible Weibull distribution (NEx-FW) to model the lifetime data with bathtub shaped failure rates and offers greater flexibility. They have also mentioned some class of Weibull model generalizations proposed by many scholars in the literature. Such as Beta-Weibull (BW) distribution of Famoye et al. (2005), Kumaraswamy Weibull (KW) distribution proposed by Cordeiro et al. (2010), generalized modified Weibull (GMW) distribution proposed by Carrasco et al. (2008), exponentiated modified Weibull extension (EMWEx) distribution introduced by Sarhan and Apaloo (2013) and flexible Weibull (FWEx) distribution of Bebbington et al. (2007).

In this article we propose a new generalized-$X$ (NG-$X$) family of distributions, that flexibly works for all baseline distributions in reliability, survival analysis, health, biomedical engineering and lifetime study. We illustrated the practicability of the NG-$X$ family by taking a Weibull distribution to be the baseline distribution and call it NG-Weibull. Most of the statistical and practical properties of NG-Weibull are addressed and the numerical study are conducted using the real data and simulation.

The rest part of the article is summarized as follows: Section 2 introduces the new class of distributions with related computations, Section 3 presents a detail computation and graphical illustration of the special member of the proposed family of distributions. Section 4 discusses some basic statistical properties of the family of distributions, Section 5 presents characterization results, the estimation technique and simulation study are discussed in Section 6, application of the distribution to the real data by numerical computation of some important quantities and the comparison of the models are presented in Section 7. Random effect models for overdispersion and correlation are discussed in the Section 8 and lastly a concluding remark with a precise recommendation is given in Section 9.

### 2. New Generalized-$X$ family

In this section, the probability density, cumulative distribution, survival, hazard, cumulative hazard, and reverse hazard functions for the NG-$X$ are computed. The proposed new family has a cumulative distribution function (CDF) and a probability density function (PDF), respectively, as follows:

$$ G(x; \alpha, \delta, \omega) = 1 - \left( \frac{\bar{F}(x; \omega)}{1 - \delta \bar{F}(x; \omega)} \right)^\alpha, \quad \alpha > 0, \delta \in (0, 1), x, \omega \in \mathbb{R}, $$

and

$$ g(x; \alpha, \delta, \omega) = \frac{\alpha(1 - \delta)f(x; \omega)[1 - F(x; \omega)]^{\alpha-1}}{[1 - \delta F(x; \omega)]^\alpha}, \quad \alpha > 0, \delta \in (0, 1), x, \omega \in \mathbb{R}, $$

where $\omega$ is a vector of parameters of a certain baseline distribution, $F(x; \omega)$, and $\bar{F}(x; \omega) = 1 - F(x; \omega)$.

The survival function $S(x)$, hazard or failure rate function $h(x)$, cumulative failure rate function $H(x)$, and reverse failure rate functions $r(x)$ are computed as

$$ S(x; \alpha, \delta, \omega) = \left( 1 - \left[ \frac{(1 - \delta)F(x; \omega)}{1 - \delta F(x; \omega)} \right] \right)^\alpha, $$

$$ h(x; \alpha, \delta, \omega) = \frac{\alpha(1 - \delta)f(x; \omega)}{(1 - F(x; \omega))(1 - \delta F(x; \omega))}, $$

$$ H(x; \alpha, \delta, \omega) = -\log \left\{ 1 - \left[ \frac{(1 - \delta)F(x; \omega)}{1 - \delta F(x; \omega)} \right] \right\}^\alpha, $$

and

$$ r(x; \alpha, \delta, \omega) = \frac{\alpha(1 - \delta)f(x; \omega)[1 - F(x; \omega)]^{\alpha+1}}{(1 - \delta F(x; \omega))(1 - \delta F(x; \omega))^\alpha - [1 - F(x; \omega)]^\alpha}, $$

respectively.
The range of the parameter values are $\alpha > 0$, $\delta \in (0, 1)$, and $x, \omega \in \mathbb{R}$, where $\omega$ is a vector of parameters. In the next section, we discuss a special member of the NG-X family called, New Generalized Weibull (NG-Weibull) distribution.

3. New Generalized-Weibull distribution

Let $F(x; \omega)$ be a CDF of the Weibull random variable (for two parameters case $\theta, \gamma > 0$), $F(x; \omega) = 1 - e^{-\gamma x^\theta}$, $x > 0$ and $f(x; \omega) = \theta \gamma x^{\theta-1} e^{-\gamma x^\theta}$, where $\omega = (\theta, \gamma)$ is a vector of parameters. The CDF of the NG-Weibull distribution is given by:

$$G(x; \alpha, \delta, \omega) = 1 - \left[ \left(1 - \delta \right) \left(1 - e^{-\gamma x^\theta}\right) \right]^{\alpha}, \quad x \geq 0. \quad (3)$$

For $\alpha, \gamma, \theta > 0$, $\delta \in (0, 1)$ and $x > 0$, the PDF, survival function, hazard rate function, cumulative hazard rate function, and reverse failure rate function for the NG-Weibull distribution are given by

$$g(x; \alpha, \delta, \omega) = \frac{\alpha(1 - \delta) \theta \gamma x^{\theta-1} e^{-\alpha \gamma x^\theta}}{\left[1 - \delta(1 - e^{-\gamma x^\theta})\right]^{\alpha+1}}, \quad (4)$$

$$S(x; \alpha, \delta, \omega) = \left[ \frac{e^{-\gamma x^\theta}}{1 - \delta(1 - e^{-\gamma x^\theta})}\right]^\alpha,$$

$$h(x; \alpha, \delta, \omega) = \frac{\alpha(1 - \delta) \theta \gamma x^{\theta-1}}{(1 - \delta(1 - e^{-\gamma x^\theta}))},$$

$$H(x; \alpha, \delta, \omega) = -\log \left[ \frac{e^{-\gamma x^\theta}}{1 - \delta(1 - e^{-\gamma x^\theta})}\right]^\alpha,$$

and

$$r(x; \alpha, \delta, \omega) = \frac{\alpha(1 - \delta) \theta \gamma x^{\theta-1} e^{-\alpha \gamma x^\theta}}{(1 - \delta(1 - e^{-\gamma x^\theta})) \left[1 - \delta(1 - e^{-\gamma x^\theta})\right]^{\alpha} - e^{-\alpha \gamma x^\theta}},$$

respectively.

Graphical expression of PDF and hazard function of the NG-Weibull is displayed below (see Figures 1). In the Figure 1 above, the graphs in part (a) are the plots of the PDF and the graphs in part (b) are the plots of the hazard rate function of NG-Weibull, respectively, to show different patterns and how flexible the distribution is. It’s observed that the PDF (which is given by Eq.(4)) is right and left skewed, symmetric, and decreasing. Moreover, the hazard function has decreasing, increasing, and parabola up patterns.

4. Basic statistical properties

In this section, some of the important statistical properties of the new proposed family of distributions are discussed.

4.1. Quantile function

Quantile function can be used for many purposes in theory and numerical applications in statistics. For example, it can be used to draw simulations. We can obtain the quantile function for the NG-X family of distributions by applying the inversion technique. Thus,

$$Q_{x_n} (u) = G(x; \alpha, \delta, \omega)^{-1},$$

$$1 - \left(1 - \left[ \frac{(1 - \delta)F(x; \omega)}{1 - \delta F(x; \omega)} \right]^{\alpha} \right) = u, \quad 0 < u < 1.$$

By using a simple algebraic application and solving the non-linear equation, this can be computed as follows:
(a) Plot of PDF

\[
Q_{x_n}(u) = F^{-1}\left[\frac{1 - (1 - u)^{\frac{1}{\alpha} - 1}}{-\delta(1 - u)^{\frac{1}{\alpha} - 1}}\right],
\]

where \(u \sim \text{uniform}(0, 1)\).

The median (Med) can be obtained by substituting \(u = 1/2\) in Eq. (5) as:

\[
Med = F^{-1}\left[\frac{1 - (1 - \frac{1}{2})^{\frac{1}{\alpha} - 1}}{-\delta(1 - \frac{1}{2})^{\frac{1}{\alpha} - 1}}\right],
\]

which can be simplified to
(b) Plot of hazard function

\begin{align*}
\alpha &= 1.3, \quad \delta = 0.4, \quad \gamma = 1, \quad \theta = 0.7 \\
\alpha &= 3, \quad \delta = 0.3, \quad \gamma = 0.5, \quad \theta = 1.1 \\
\alpha &= 3.5, \quad \delta = 0.6, \quad \gamma = 1, \quad \theta = 0.9 \\
\alpha &= 1, \quad \delta = 0.8, \quad \gamma = 0.7, \quad \theta = 3.6
\end{align*}

Figure 1: Plot of PDF and hazard function for the NG-Weibull distribution.

\[ Med = F^{-1} \left[ \frac{(1 - \left(\frac{1}{2}\right)^{\frac{1}{\alpha}})}{-\delta((\frac{1}{2})^{\frac{1}{\alpha}} + 1)} \right]. \]

Similarly, one can obtain the lower and upper quartiles by substituting \( u = 1/4 \) and \( u = 3/4 \) in Eq.(5), respectively.

4.2. Skewness and Kurtosis

Based on the formulas given by Galton (1883) and Moors (1988), we can obtain Skewness(Sk) and Kurtosis(K) from Eq. (5) as

\[ Sk = \frac{(Q_3 - 2Q_2 + Q_1)}{(Q_3 - Q_1)} = \frac{(q(0.75) - 2q(0.5) + q(0.25))}{(q(0.75) - q(0.25))}, \]

and

\[ K = \frac{Q_{2\times} - Q_{\frac{3}{2}} + Q_{\frac{1}{2}} - Q_{2}}{Q_{2\times} - Q_{\frac{3}{2}}} = \frac{q(0.875) - q(0.625) + q(0.375) - q(0.125)}{q(0.75) - q(0.25)}, \]

respectively.

4.3. Order statistics

It is well-common that order statistics are widely used in the statistical applications, such as reliability and lifetime testing. Suppose that \( X_1, X_2, ..., X_n \) are a random sample of size \( n \) drawn independently from the NG-X family of distributions with parameters \( \alpha, \delta \) and \( \omega \). Let \( X_{1:n}, X_{2:n}, ..., X_{n:n} \) be the corresponding order statistics. Then, the
density of $X_{i:n}$ for ($i = 1, 2, ..., n$) is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x; \omega) [F(x; \omega)]^{i+j-1},$$

where

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \frac{\alpha(1-\delta)f(x; \omega)[1-F(x; \omega)]^{\alpha-1}}{[1-\delta(1-F(x; \omega))]^{\alpha+1}} \times \left[ 1 - \left[ \frac{1-F(x; \omega)}{1-\delta F(x; \omega)} \right]^{\alpha+i-j} \right].$$

The order statistics of the special member of the family NG-Weibull is derived in the same way as follows:

$$f_{i:nNG-Weibull}(x; \alpha, \delta, \omega) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \frac{\alpha(1-\delta)\theta \gamma x^{\theta-1} e^{-(i+j)\alpha \gamma x^\theta}}{[1-\delta(1-e^{-(i+j)\alpha \gamma x^\theta})]^{\alpha+1}} \times \left[ 1 - \left[ \frac{e^{-(i+j)\gamma x^\theta}}{1-\delta(1-e^{-(i+j)\gamma x^\theta})} \right]^{\alpha+i-j} \right].$$

This can be simplified in Eq.(6) as:

$$f_{i:nNG-Weibull}(x; \alpha, \delta, \omega) = \sum_{j=0}^{n-i} \eta_j f_{NG-Weibull}(x; \alpha, \delta, \omega),$$

(6)

where

$$\eta_j = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j}.$$

4.4. Moments and moment generating functions

The $r^{th}$ central moment of the new proposed family can be obtained as follows:

$$\mu_r' = \int_{-\infty}^{\infty} x^r g(x; \alpha, \delta, \omega) dx$$

$$= \int_{-\infty}^{\infty} x^r \frac{\alpha(1-\delta)[F(x; \omega)]^{\alpha r}}{[1-\delta F(x; \omega)]^{\alpha r+1}} dx$$

$$= \sum_{r=0}^{\infty} \frac{[\alpha(1-\delta)]^{r+1}}{r!} \eta_{r,i+1}.$$

The moment generating function, $M_x(t)$, for the family is given in Eq.(7).

$$M_x(t) = \sum_{r=0}^{\infty} t^r \frac{[\alpha(1-\delta)]^{r+1}}{r!} \eta_{r,i+1},$$

(7)

where

$$\eta_{r,i+1} = \int_{-\infty}^{\infty} x^r f(x; \omega) \left[ \frac{[1-F(x; \omega)]^{\alpha r}}{[1-\delta F(x; \omega)]^{\alpha r+1}} \right] dx.$$
5. Characterization results

This section is devoted to the characterizations of the NG-X distribution in the following directions: (i) based on a simple relationship between two truncated moments; (ii) in terms of the hazard function and (iii) based on the conditional expectation of a function of the random variable. It should be mentioned that for the characterization (i) the CDF is not required to have a closed form.

We present our characterizations (i) – (iii) in three subsections.

5.1 Characterizations based on two truncated moments

In this subsection we present the characterizations of NG-X distribution in terms of a simple relationship between two truncated moments. The first characterization result employs a theorem due to Glänzel [14], see Theorem 5.1.1 below. Note that the result holds also when the interval \( H \) is not closed. Moreover, as mentioned above, it could be also applied when the CDF \( G \) does not have a closed form. As shown in Glänzel [15], this characterization is stable in the sense of weak convergence.

**Theorem 5.1.1.** Let \((\Omega, \mathcal{F}, \mathbf{P})\) be a given probability space and let \( H = [d, e] \) be an interval for some \( d < e \) \((d = -\infty, e = \infty \) might as well be allowed). Let \( X : \Omega \rightarrow H \) be a continuous random variable with the distribution function \( G \) and let \( q_1 \) and \( q_2 \) be two real functions defined on \( H \) such that

\[
E[q_2(X) | X \geq x] = E[q_1(X) | X \geq x] \xi(x), \quad x \in H,
\]

is defined with some real function \( \xi \). Assume that \( q_1, q_2 \in C^1(H) \), \( \xi \in C^2(H) \) and \( G \) is twice continuously differentiable and strictly monotone function on the set \( H \). Finally, assume that the equation \( \xi q_1 = q_2 \) has no real solution in the interior of \( H \). Then \( G \) is uniquely determined by the functions \( q_1, q_2 \) and \( \xi \), particularly

\[
G(x) = \int_d^x C \left[ \frac{\xi'(u)}{\xi(u) q_1(u) - q_2(u)} \right] \exp(-s(u)) \, du,
\]

where the function \( s \) is a solution of the differential equation \( s' = \frac{s'\xi}{\xi q_1 q_2} \) and \( C \) is the normalization constant, such that \( \int_H dG = 1 \).

**Proposition 5.1.1.** Let \( X : \Omega \rightarrow \mathbb{R} \) be a continuous random variable and let \( q_1(x) = [1 - \delta F(x; \omega)]^{\alpha-1} \) and \( q_2(x) = q_1(x) [1 - F(x; \omega)] \) for \( x \in \mathbb{R} \). The random variable \( X \) has PDF (1) if and only if the function \( \xi \) defined in Theorem 5.1.1 has the form

\[
\xi(x) = \frac{\alpha}{\alpha + 1} [1 - F(x; \omega)], \quad x \in \mathbb{R}.
\]

Proof. Let \( X \) be a random variable with pdf (1), then

\[
(1 - G(x)) E[q_1(X) | X \geq x] = (1 - \delta) [1 - F(x; \omega)]^\alpha, \quad x \in \mathbb{R},
\]

and

\[
(1 - G(x)) E[q_2(X) | X \geq x] = \frac{\alpha(1 - \delta)}{\alpha + 1} [1 - F(x; \omega)]^{\alpha+1}, \quad x \in \mathbb{R},
\]

and finally

\[
\xi(x) q_1(x) - q_2(x) = -\frac{1}{\alpha + 1} q_1(u) [1 - F(x; \omega)] < 0 \quad \text{for} \quad x \in \mathbb{R}.
\]

Conversely, if \( \xi \) is given as above, then
\[ s'(x) = \frac{\xi'(x) q_1(x)}{\xi(x) q_1(x) - q_2(x)} = \frac{\alpha f(x;\omega)}{1 - F(x;\omega)}, \]

and hence

\[ s(y) = -\log [1 - F(x;\omega)]^{\alpha}, \quad x \in \mathbb{R}. \]

Now, in view of Theorem 5.1.1, \( X \) has density \( (1) \).

**Corollary 5.1.1.** Let \( X : \Omega \rightarrow \mathbb{R} \) be a continuous random variable and let \( q_1(x) \) be as in Proposition 5.1.1. The PDF of \( X \) is \( (1) \) if and only if there exist functions \( q_2 \) and \( \xi \) defined in Theorem 5.1.1 satisfying the differential equation

\[ \xi'(x) q_1(x) = \alpha f(x;\omega), \quad x \in \mathbb{R}. \]

**Corollary 5.1.2.** The general solution of the differential equation in Corollary 5.1.1 is

\[ \xi(x) = \left[1 - F(x;\omega)\right]^{-\alpha} \left[-\int \alpha f(x;\omega) [1 - F(x;\omega)]^{\alpha-1} (q_1(x))^{-1} q_2(x) + D\right], \]

where \( D \) is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 5.1.1 with \( D = 0 \). However, it should be also noted that there are other triplets \((q_1, q_2, \xi)\) satisfying the conditions of Theorem 5.1.1.

### 5.2 Characterization based on hazard function

It is well known that the hazard function, \( h_G \), of a twice differentiable distribution function, \( G \) with density \( g \), satisfies the first order differential equation

\[ \frac{g'(x)}{g(x)} = \frac{h'_G(x)}{h_G(x)} - h_g(x). \]

For many univariate continuous distributions, this is the only characterization available in terms of the hazard function. The following characterization establishes a characterization of NG-X distribution in terms of the hazard function, which is not of the above trivial form.

**Proposition 5.2.1.** Let \( X : \Omega \rightarrow \mathbb{R} \) be a continuous random variable. The pdf of \( X \) is \( (1) \) if and only if its hazard function \( h_G(x) \) satisfies the differential equation

\[ h'_G(x) - \frac{f'(x;\omega)}{f(x;\omega)} h_G(x) = \frac{\alpha (1 - \delta)^2 f(x;\omega)^2}{\left[F(x;\omega) [1 - \delta F(x;\omega)]\right]^2}, \quad x \in \mathbb{R}. \]

Proof. Is straightforward and hence omitted.

### 5.3 Characterizations based on conditional expectation

The following proposition has already appeared in Hamedani [16], so we will just state it here which can be used to characterize the NG-X distribution.

**Proposition 5.3.1.** Let \( X : \Omega \rightarrow (a, b) \) be a continuous random variable with CDF \( G \). Let \( \psi(x) \) be a differentiable function on \((a, b)\) with \( \lim_{x \to -\infty} \psi(x) = 1 \). Then for \( \delta \neq 1 \),
The computation of maximum likelihood estimators (MLEs) for the model parameters of NG-X is dealt in this section. Let $x_1, x_2, ..., x_n$ be observed values of a sample randomly selected from NG-X with parameters $\alpha$, $\delta$, and $\omega$. Given the PDF of the new proposed family of distributions in Eq. (2) and the total likelihood function (in Eq. (8) below)

$$L(x; \alpha, \delta, \omega) = \prod_{i=1}^{n} \frac{\alpha(1-\delta) f(x_i; \omega) [1-F(x_i; \omega)]^{n-1}}{[1-\delta F(x_i; \omega)]^{n+1}},$$

the log-likelihood function of the respective sample $\log L(x; \alpha, \delta, \omega)$ is given below in Eq. (9) and the model parameters can be estimated by taking the first partial derivative of the $\log L(x; \alpha, \delta, \omega)$ with respect to each model parameter and equating to zero and solving them simultaneously.

$$\log L(x; \alpha, \delta, \omega) = n \log \alpha + n \log (1-\delta) + \sum_{i=1}^{n} \log (f(x_i; \omega))$$

$$+ (\alpha - 1) \sum_{i=1}^{n} \log (F(x_i; \omega)) - (\alpha + 1) \sum_{i=1}^{n} \log (1-\delta F(x_i; \omega)).$$

The partial derivatives of $\log L(x; \alpha, \delta, \omega)$ with respect to each parameter are given below.

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log (1-F(x_i; \omega)) - \sum_{i=1}^{n} \log (1-\delta F(x_i; \omega)),$$

$$\frac{\partial \log L}{\partial \delta} = -\frac{n}{1-\delta} - (\alpha + 1) \sum_{i=1}^{n} \left[ \frac{\partial \log (1-\delta F(x_i; \omega))}{\partial \delta} \right],$$

and

$$\frac{\partial \log L}{\partial \omega} = \sum_{i=1}^{n} \frac{\partial \log (f(x_i; \omega))}{\partial \omega} + (\alpha + 1) \sum_{i=1}^{n} \frac{\partial \log (1-F(x_i; \omega))}{\partial \omega}$$

$$- (\alpha + 1) \sum_{i=1}^{n} \frac{\partial \log (1-\delta F(x_i; \omega))}{\partial \omega}.$$ 

Now, the MLEs of the parameters $\alpha, \delta, \omega$ can be obtained by solving the non-linear equation

$$U_n = \left( \frac{\partial \log L(x; \alpha, \delta, \omega)}{\partial \alpha}, \frac{\partial \log L(x; \alpha, \delta, \omega)}{\partial \delta}, \frac{\partial \log L(x; \alpha, \delta, \omega)}{\partial \omega} \right)^T = 0,$$

using a numerical method.
6.2. Simulation study

In this section, we evaluate the performance of the MLEs for fixed sample size \( n \). A numerical evaluation is carried out to study the performance of the MLEs for the NG-Weibull model (which is a special case of the proposed family). The estimators, biases and the practical mean square errors (MSEs) were performed using the \( R \) software program. The empirical phases are itemized as below:

i. A sequence of random sample \( X_1, X_2, \ldots, X_n \) of sizes; \( n=25, 100, 375, 525, 750, 950 \) and \( 1000 \) are drawn and these random samples are considered for the computation of the above mentioned quantities. The samples are generated from the NG-Weibull distribution by using inversion method.

ii. Three scenarios are considered for the four parameters of the proposed model to evaluate the MLEs for each parameter and sample size iteratively. These are categorized by cases, such as case 1: \( \alpha = 0.7, \delta = 0.4, \theta = 0.5, \gamma = 0.3 \), case 2: \( \alpha = 0.5, \delta = 0.5, \theta = 0.6, \gamma = 1 \), and case 3: \( \alpha = 0.2, \delta = 0.5, \theta = 0.6, \gamma = 0.3 \).

iii. There samples were repeated 1000 times to compute the bias and the MSEs for each parameter.

iv. The computational formulas for bias and MSEs are given as follows based on the formulae for the mean and variance of the parameters:

\[
\hat{\eta} = \frac{1}{1000} \sum_{i=1}^{1000} \eta_i,
\]

\[
\text{Bias}(\hat{\eta}) = \hat{\eta}_i - \eta,
\]

\[
\text{Var}(\hat{\eta}) = \frac{1}{1000} \sum_{i=1}^{1000} (\eta_i - \eta)^2,
\]

and

\[
\text{MSE}(\hat{\eta}) = \text{var}(\hat{\eta}) + (\text{Bias}(\hat{\eta}))^2,
\]

where \( \hat{\eta} = (\hat{\alpha}, \hat{\delta}, \hat{\theta}, \hat{\gamma}) \).

The numerical results are displayed in Table 1 and the plots for parameter estimates, MSEs, absolute biases and biases are presented consecutively in the graphical part (Figures 5-7 in the appendix part). From this result, it can be observed that the parameter values estimated are quite steady and are close to the true parameter values as the sample size increases. And there is a vivid sight that as the sample size increases, the error is minimized as expected. Table 1 displays the results of the simulation study for different simulation cases.

Graphical illustration of the simulation study for different random values of the model parameters is given below. The illustration displays the graphs for the three simulation settings consecutively. The Figure 5 (in the appendix part) plots the estimated parameters, MSEs, absolute biases, and biases for \( \alpha = 0.7, \delta = 0.4, \theta = 0.5, \gamma = 0.3 \). The parameter \( \alpha \) behaves the same way in the above graphs (Figure 5) for the first setting. The Figure 6 (in the appendix part) plots the estimated parameters, MSEs, absolute biases, and biases for \( \alpha = 0.5, \delta = 0.5, \theta = 0.6, \gamma = 1 \) for the second setting. And the Figure 7 (in the appendix part) plots the estimated parameters, MSEs, absolute biases, and biases for \( \alpha = 0.2, \delta = 0.5, \theta = 0.6, \gamma = 0.3 \) for the last setting.

In Figure 6, the parameters \( \alpha \) and \( \gamma \) have similar effect on the flexibility of the graphs for the second case. In Figure 7, these two parameters attain the closeness of their counterpart estimated values with the same rate. The rest of the parameter values (\( \delta \) and \( \theta \)) are close to the true parameter values as the sample size increases for all of the above graphs.

7. An application to failure times data

In this section, the NG-Weibull model is applied to the failure times data and it is compared with five of its competitor models. The data used for the analysis in this section come from the failure times in hours for 101 subjects cited by Chipepa et al. (2020) and the detail description of the data can be obtained from [ (Barlow et al., 1984), (Andrews
Table 1: Summary of simulation study for different simulation settings.

<table>
<thead>
<tr>
<th>n</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>MSEs</td>
<td>Bis.</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>1.37010</td>
<td>4.37813</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>0.61238</td>
<td>0.17316</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}$</td>
<td>0.58640</td>
<td>0.03147</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}$</td>
<td>1.22280</td>
<td>2.63451</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>0.86482</td>
<td>2.00358</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>0.56658</td>
<td>0.11100</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}$</td>
<td>0.58319</td>
<td>0.01303</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}$</td>
<td>0.95740</td>
<td>4.95354</td>
</tr>
<tr>
<td>375</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>0.54035</td>
<td>0.54050</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>0.49905</td>
<td>0.04748</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}$</td>
<td>0.56640</td>
<td>0.03147</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}$</td>
<td>1.22280</td>
<td>2.63451</td>
</tr>
<tr>
<td>750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>0.42371</td>
<td>0.08349</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>0.49221</td>
<td>0.03690</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}$</td>
<td>0.60192</td>
<td>0.00308</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}$</td>
<td>0.95740</td>
<td>4.95354</td>
</tr>
<tr>
<td>950</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>0.41231</td>
<td>0.05518</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>0.48368</td>
<td>0.01854</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}$</td>
<td>0.60549</td>
<td>0.00143</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}$</td>
<td>0.56599</td>
<td>0.09924</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>0.40443</td>
<td>0.03998</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>0.47198</td>
<td>0.02007</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}$</td>
<td>0.60844</td>
<td>0.00148</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}$</td>
<td>0.55861</td>
<td>0.09622</td>
</tr>
</tbody>
</table>

Table 2: Descriptive summary result for the failure times in hours data.

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.240</td>
<td>0.800</td>
<td>1.025</td>
<td>1.450</td>
<td>7.890</td>
</tr>
</tbody>
</table>

and Herzberg, 2012]. Some of the features of the practical data are highlighted in the introduction part of our study based on Moss (2004). The data is given in appendix, data for section 6.2 part with the data name: dataFT. Descriptive summary measures of the data are provided in Table 2 and Figure 5 displays the box plot of the data.

We illustrate the fitting capability of the NG-Weibull to the data by comparing it to three-parameter new extended flexible Weibull (NEx-FW), four-parameter flexible Weibull (FW), five-parameter Kumaraswamy Weibull Poisson (KWP), four-parameter Kumaraswamy Weibull (Ku-W), and a three-parameter Zubair-Weibull (Z-Weibull). The CDFs of the competing models are:

1. New extended flexible Weibull(NEx-FW)

$$G(x; \gamma, \beta, \theta) = 1 - e^{-e^{x^{\gamma}} + x^{\beta}}$$,

$$x, \gamma, \beta, \theta > 0.$$
2. Flexible Weibull (FW)
\[ G(x; \alpha, \gamma, \beta, \theta) = 1 - e^{-e^{\beta x^{\gamma} + \theta x^{\alpha}}}, \quad x, \alpha, \gamma, \beta, \theta > 0. \]

3. Kumaraswamy Weibull Poisson (KWP)
\[ G(x; a, b, c, \beta, \lambda) = \frac{1 - e^{-\lambda[1-(1-e^{-(\beta x^{\gamma} + c a)^b}]}}{1 - e^{-\lambda}}, \quad x, a, b, c, \beta, \lambda > 0. \]

4. Kumaraswamy Weibull (Ku-W)
\[ G(x; a, b, \gamma, \theta) = 1 - (1 - e^{-\gamma x^{\theta}})^b, \quad x, a, b, \gamma, \theta > 0. \]

5. Z-Weibull
\[ F(x; \alpha, \gamma, \theta) = \frac{e^{\alpha(1-e^{-\gamma x^\theta})^2} - 1}{e^\alpha - 1}, \quad \alpha, \gamma, \theta > 0, x \geq 0. \]

For model comparison, we apply the following model adequacy measures or measures for goodness-of-fit test. These are Akaike information criterion (AIC), consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), Cramer-von Mises (CM) statistic and Anderson-Darling (AD) test statistic. Accordingly, the model with the least value for these measures is discriminated as the best fit for the data. Hence, our proposed model NG-Weibull out performs the five competitor models based on this comparison for discrimination, as revealed by results presented on Table 4. Tables 3 and 4 give the summary of the results for MLEs and standard errors (SE.s) and model adequacy measures, respectively, for the fitted models.

From these tables, one can easily observe that there is a clear evidence for the improvement in the NG-Weibull model and its performance indicates it as the selected among the competing models.

The following section deals with the random effects models for the data that are collected for several times repeatedly and which exhibit the two commonly known features. These features in the literature are termed as overdispersion and correlation.

8. Random effects models for overdispersion and correlation

Some of the classical models and general classes of distributions are not convenient to express as an exponential families and hence, cannot hold covariates, and cannot handle overdispersion and hierarchical features or correlations in the data. Hence, other mixture of models should be employed to hold covariates and to handle both overdispersion and correlation simultaneously (Kassahun et al., 2012).

The primary goal of this section is to select the appropriate model which accommodates this issue for the data that are
A New Generalized-X Family of Distributions: Applications, Characterization and a Mixture of Random Effect Models

1. Weibull model:

\[ f(y|\eta, \phi) = e^{\{\phi^{-1}[y\eta - \psi(\eta)] + c(y, \phi)\}}, \quad y \in \mathbb{R}^+, \]  

(10)

for a definite fixed unknown parameters \( \eta \) and \( \phi \) (often called natural or canonical and dispersion parameter, respectively) and for the known functions \( \psi(\cdot) \) and \( c(\cdot, \cdot) \).

From the computation of Molenberghs and Verbeke (2005), it is easy to consider the first two moments by following the function \( \psi(\cdot) \) (based on Eq. (10)) as: \( E(Y) = \mu = \psi'(\eta) \) and \( \text{var}(Y) = \sigma^2 = \phi \psi''(\eta) \). From this, the mean and variance can be related as: \( \sigma^2 = \phi \psi'' \left[ \psi^{-1}(\mu) \right] = \phi \nu(\mu) \), where \( \nu(\cdot) \) is a variance function defining the mean-variance relationship.

Molenberghs et al. (2010) gave a basic example of exponential families for normal, binary, count, and time-to-event cases. The normal one is seen as case of exponential family for its particular feature that it needs an overdispersion parameter which should exceed 1. Thus, it lacks a mean-variance relationship while others held exist.

It is conventional to decompose \( \varphi = \lambda e^\mu \) in the Weibull and Exponential models which allow \( \mu \) being a function of covariates. However, it can be noted that \( \mu \), in this case, is different from the mean since it is a component of mean function. On the other hand, Weibull model does not belong to the exponential family where \( Y \) is replaced by \( Y^\rho \).

Hence, based on Molenberghs et al. (2010) and Oliveira (2014), the density functions for some models with conjugate random, hierarchical, and marginal effects for time-to-event and repeatedly measured data can be given as:

1. Weibull model:

\[ f(y|\varphi, \rho) = \varphi \rho e^{-\varphi y^\rho}, \quad \varphi, \rho, y > 0. \]

It is clear that the Exponential model follows the Weibull model for \( \rho = 1 \).

2. Exponential-gamma

Table 3: MLE estimates of the parameters and the corresponding standard errors (SE. in the parentheses) for the fitted models.

<table>
<thead>
<tr>
<th>Dist.</th>
<th>( \hat{a} ) (SE.)</th>
<th>( b ) (SE.)</th>
<th>( \hat{c} ) (SE.)</th>
<th>( \hat{\alpha} ) (SE.)</th>
<th>( \beta ) (SE.)</th>
<th>( \hat{\gamma} ) (SE.)</th>
<th>( \delta ) (SE.)</th>
<th>( \theta ) (SE.)</th>
<th>( \lambda ) (SE.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNW</td>
<td>1.162 (0.672)</td>
<td>1.892 (0.980)</td>
<td>0.772 (0.257)</td>
<td>0.677 (0.187)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEFW</td>
<td>0.457 (0.044)</td>
<td>0.648 (0.062)</td>
<td>0.011 (0.005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FW</td>
<td>0.630 (0.178)</td>
<td>0.297 (0.626)</td>
<td>0.626 (0.046)</td>
<td>0.355 (0.066)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-W</td>
<td>1.539 (0.911)</td>
<td>2.177 (0.329)</td>
<td>0.355 (0.066)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KW</td>
<td>0.734 (0.172)</td>
<td>0.270 (0.035)</td>
<td>3.313 (0.149)</td>
<td>0.187 (0.005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KWP</td>
<td>0.211 (0.023)</td>
<td>1.300 (1.002)</td>
<td>4.509 (0.035)</td>
<td>0.107 (0.033)</td>
<td>6.652 (3.686)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Model adequacy measures for the fitted models.

<table>
<thead>
<tr>
<th>Dist.</th>
<th>AIC</th>
<th>CAIC</th>
<th>BIC</th>
<th>HQIC</th>
<th>(-2\log L)</th>
<th>CM</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NG-W</td>
<td>211.62</td>
<td>212.03</td>
<td>222.08</td>
<td>215.85</td>
<td>101.81</td>
<td>0.07</td>
<td>0.75</td>
</tr>
<tr>
<td>KW</td>
<td>213.25</td>
<td>213.67</td>
<td>223.71</td>
<td>217.49</td>
<td>102.63</td>
<td>0.14</td>
<td>0.87</td>
</tr>
<tr>
<td>Z-W</td>
<td>213.45</td>
<td>213.70</td>
<td>222.20</td>
<td>216.63</td>
<td>103.73</td>
<td>0.26</td>
<td>1.40</td>
</tr>
<tr>
<td>KWP</td>
<td>215.94</td>
<td>216.58</td>
<td>229.02</td>
<td>221.24</td>
<td>106.97</td>
<td>0.21</td>
<td>1.16</td>
</tr>
<tr>
<td>FW</td>
<td>421.29</td>
<td>421.71</td>
<td>431.75</td>
<td>425.53</td>
<td>206.65</td>
<td>0.36</td>
<td>2.71</td>
</tr>
<tr>
<td>NEFW</td>
<td>514.25</td>
<td>514.50</td>
<td>522.10</td>
<td>517.43</td>
<td>254.13</td>
<td>0.86</td>
<td>5.54</td>
</tr>
</tbody>
</table>

repeatedly measured over time, usually called longitudinally measured data. Since the data are measured repeatedly for several times for the same patient, it characterize some basic quantities like an outcome of interest measured. Thus, repeatedly measured over time, usually called longitudinally measured data. Since the data are measured repeatedly for several times for the same patient, it characterize some basic quantities like an outcome of interest measured. Thus, repeatedly measured over time, usually called longitudinally measured data. Since the data are measured repeatedly for several times for the same patient, it characterize some basic quantities like an outcome of interest measured.
i. Hierarchical model (EGh):
\[ f(y|\varphi, \theta) = \varphi \theta e^{-\varphi \theta y}, \quad \varphi, \theta, y > 0. \]

ii. Random effect model (EGr):
\[ f(\theta|\alpha, \beta) = \frac{\theta^{(\alpha-1)}e^{-\theta/\beta}}{\beta^{\alpha} \Gamma(\alpha)}, \quad \alpha, \beta, \theta, y > 0. \]

iii. Marginal model (EGm):
\[ f(y|\varphi, \alpha, \beta) = \frac{\varphi \alpha \beta}{1 + \varphi \beta y}^{\alpha+1}, \quad \alpha, \beta, \varphi, y > 0. \]

3. Weibull-gamma
i. Hierarchical model (WGh):
\[ f(y|\varphi, \theta, \rho) = \varphi \theta \rho y^{\rho - 1} e^{-\varphi \theta y}, \quad \varphi, \theta, \rho, y > 0. \]

ii. Random effect model (WGr):
\[ f(\theta|\alpha, \beta) = \frac{\theta^{(\alpha-1)}e^{-\theta/\beta}}{\beta^{\alpha} \Gamma(\alpha)}, \quad \alpha, \beta, \theta, y > 0. \]

iii. Marginal model (WGm):
\[ f(y|\varphi, \alpha, \beta, \rho) = \frac{\varphi \rho y^{\rho - 1} \alpha \beta}{1 + \varphi \beta y^{\rho}}^{\alpha+1}, \quad \alpha, \beta, \varphi, \rho, y > 0. \]

Using the decomposition of \( \varphi = \lambda e^\mu \) this can be written as
\[ f(y|\alpha, \beta, \lambda, \rho, \mu) = \frac{\lambda \rho y^{\rho - 1} \alpha \beta}{1 + \lambda \beta y^{\rho}}^{\alpha+1}, \quad \alpha, \beta, \lambda, \rho, \mu, y > 0. \]

4. Weibull-gamma-normal model (WGN or Combined):
\[ f(y|\lambda, \theta, \rho, \mu) = \lambda \theta \rho y^{\rho - 1} e^{-\lambda y e^\mu}, \quad \lambda, \theta, \mu, \rho, y > 0. \] (11)

This closed form of the density function, Eq. (11), is reached by the help of the above decomposition of \( \varphi = \lambda e^\mu \) and by considering the two random effects, gamma random effect (usually represented by \( \theta \)) and the normal random effect (\( \mu \), in which both fixed and the random effect terms are collected).

And it can also be further extended to the computational closed form by taking the expression, usually based on the linear mixed effects model (LMM), \( x_{ij} \beta + z_{ij} b_i \), where \( x_{ij} \)-vector of covariates for the fixed term with \( j^{th} \) measurement for \( i^{th} \) subject or patient with diabetes for our study, \( \beta \)-vector of unknown parameters for fixed term, \( z_{ij} \)-vector of covariates for random term with \( j^{th} \) measurement for the \( i^{th} \) patient with diabetes, \( b_i \)-vector of parameters for random term, where \( i = 1, 2, ..., n \) and \( j = 1, 2, ..., n_i \). Here importantly, the random effect term part, \( b_i \sim N(0, D) \), where 0- a vector of zero means and \( D \)- a variance-covariance matrix.

8.1. The likelihood function for the combined model and parameter estimation

Combining ideas from both gamma and normal random effects for time-to-event data gives the combined Weibull based model and it further gives the closed form of vector of means and variance-covariance matrices which is later given by Molenberghs et al. (2010), though before it’s reported as it has no closed form by Molenberghs et al. (2007).

For repeatedly measured events, the likelihood function for the Weibull model based combined model with gamma
and normal random effects can be given as:

\[
  f \left( y_i | \theta_i, b_i, \Omega \right) = \prod_{j=1}^{n_i} \lambda \omega_j y_i^{\omega_j - 1} e^{-y_i^{\omega_j} \xi + z_i b_i} e^{-\lambda y_i^{\omega_j} \theta_j^{\omega_j} \xi + z_i b_i}.
\]  

(12)

This result can be obtained by taking Eq. (11) and considering the decomposition \( \varphi = \lambda e^\mu \) with \( \mu = x_i \beta + z_i b_i \) and by replacing \( \beta \) by \( \xi \), where \( \Omega = (\lambda, \rho, \theta, \xi)^T \).

The marginal density of \( \theta_{ij} \) gamma random effects based on its important properties like its random effects are independent and which make easy to extend to the multivariate gamma distributions, can be written as:

\[
  f \left( \theta_i | \alpha_i, \beta_i \right) = \prod_{j=1}^{n_i} \frac{1}{\beta_j^{\alpha_j}} \Gamma(\alpha_j) \theta_j^{\alpha_j - 1} e^{-\theta_j/\beta_j}.
\]

And the marginal density for normal random effect is also given by integrating the combined model over the gamma random effect as:

\[
  f \left( b_i | D \right) = \frac{1}{(2\pi)^{q/2} |D|^{1/2}} e^{-(1/2) b_i^T D^{-1} b_i}.
\]

The model parameters can be estimated by following the analytical computations given by Molenberghs et al. (2010). This can be shown as below by integrating over the two random effects, on which the computation is based on the likelihood function of patient \( i \).

\[
  f_i \left( y_i | \varphi, D, \vartheta_i, \Sigma_i \right) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij} | \varphi, b_i, \theta_i) f(b_i | D) f(\theta_i | \vartheta_i, \Sigma_i) db_i d\theta_i.
\]

In this equation, the new additional parameter vector \( \varphi \) groups all parameters in this conditional model for the vector of individual patients \( Y_i \). Hence, the likelihood function is derived as:

\[
  L(\varphi, D, \vartheta, \Sigma) = \prod_{i=1}^{N} f_i(y_i | \varphi, D, \vartheta_i, \Sigma_i),
\]

(13)

\[
  = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij} | \varphi, b_i, \theta_i) f(b_i | D) f(\theta_i | \vartheta_i, \Sigma_i) db_i d\theta_i.
\]

(14)

From Eqs.(13-14), we see that there are three functions to be integrated and there are two random effects \( \theta_i \) and \( b_i \) on which the integrals (say \( N \)) are carried out. This makes the equation too complex to get analytical solution for the parameters.

As the solution, though some of whose series expansion methods exhibit bias, some scholars suggested penalized quasi-likelihood (PQL), marginal quasi-likelihood (MQL), and Laplace approximation which are based on Taylor-series expansion and numerical-integration based methods. These methods have been implemented, for example, in statistical analysis system (SAS) procedures like GLIMMIX and NLMIXED [(Molenberghs and Verbeke, 2005), (Breslow and Lin, 1995)]. GLIMMIX and NLMIXED are commonly known SAS procedures for the estimation of model parameters of generalized linear mixed models and non-linear mixed models, respectively.

8.2. A practical illustration of the models to diabetes mellitus patients data

The following section displays the practical application of the random effects models to health data, diabetes mellitus (DM) patients data and the comparison among these models to discriminate the best fitting model to the data. The data was originally collected from Nigist Ellen Mohamad memorial and referral comprehensive specialized Hospital at Hossana town, South Ethiopia. A random sample of patients is selected from their medical record number (MRN) in the diabetes follow up ward.

The study period is from September 30, 2015 to August 30, 2020. There are 200 diabetes patients who monthly followed for their blood pressure (blood sugar level follow up) and for the corresponding treatments in the ward within the study period. There are thirteen patients who followed only for 3 times and four patients (the 2nd, 8th, 19th, and
200th followed for eight times in the study period. A total of 889 measurements are recorded from the total of 200 sample patients.

Table 5: A summary of descriptive measures of diabetes patients data.

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00</td>
<td>17.00</td>
<td>32.00</td>
<td>31.81</td>
<td>47.00</td>
<td>60.00</td>
<td>292.11</td>
</tr>
</tbody>
</table>

Figure 3: A graph for boxplot of diabetes patients data.

Figure 6 shows a clear display of the five important magnitudes in the data. It displays the quantities the minimum, maximum, median, lower and the upper quartiles of the response variable in the data. The two extreme observations have a moderate influence on the computation of other quantities. However, systematically, their contribution to the analysis is non-ignorable and the imposing action is not taken (authors’ suggestion). From Table 5, which displays summary of basic descriptive measures, we see that the observed variance (292.11 months) is much greater than the observed mean (31.80 months) and it is clear to see that there is extra variability in the data which violates the mean-variance relationship.

And it’s so clear that each patient is subjected to a measurement taken repeatedly over several months which shows there is a dependence among observations taken from each patient (formally, there is correlation). Hence, the need for the mixture models to capture these features simultaneously is evidently supported.

Here below, the result for the six models (Exponential (Ex), Exponential-normal (ExN), Weibull (W), Weibull-normal (WN), Weibull-gamma (WG) and Weibull-gamma-normal (WGN)) is summarized with graphical and other common methods for models comparison.

The above Figure 7 shows a comparison of Exponential and Weibull mixture models for the cases of hierarchical, random effect, and marginal, respectively, and the combined case too, specifically with Exponential and Weibull. The WGN model has the higher upper curve over the histogram and elicits the out-fit of the Weibull-gamma-normal model to handle both over dispersion and correlation simultaneously. Furthermore, Table 6 gives the formal comparison among these models. The result was obtained by estimating the model parameters using the ML by which case the numerical integration of random effects was done, where it presents. The table summarizes the parameter estimates and SE. together with other results.

The Table 6 displays the summary of important results for the six models. When compared to Ex and a mixture of it, W and a mixture of itself are better performing, except WG case. Again except WG, W and its mixture models have similarities in performances, including a combined model.

It is a common sense that WN can handle the correlation among individual patients. However, the combined model has shown an improvement. As a matter of choice, though its BIC is higher than WN, a combined model best fits the data, mainly compared to Ex and its mixture models.

Molenberghs et al. (2007) and Kassahun et al. (2012) have first analysed and applied the combined model for binary data by using beta-binomial as a model from which a random effect for overdispersion and a logistic-normal for normal
PDFs Over Histogram for Models with Random Effects

Figure 4: Graphical comparison of Exponential and Weibull mixture models.

Table 6: Comparison of the models: Ex, ExN, W, WN, WG and WGN.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Ex</th>
<th>ExN</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\xi_0$</td>
<td>125 (0.159)</td>
<td>-3.341(0.319)</td>
<td>120(0.187)</td>
</tr>
<tr>
<td>Sex effect</td>
<td>$\xi_1$</td>
<td>0.032(0.069)</td>
<td>0.012(0.031)</td>
<td>0.044(0.068)</td>
</tr>
<tr>
<td>History effect</td>
<td>$\xi_2$</td>
<td>-0.021(0.129)</td>
<td>-0.073(0.154)</td>
<td>-0.046(0.360)</td>
</tr>
<tr>
<td>Status effect</td>
<td>$\xi_3$</td>
<td>-0.105(0.272)</td>
<td>-0.208(0.150)</td>
<td>-0.207(0.273)</td>
</tr>
<tr>
<td>Weibull shape</td>
<td>$\rho$</td>
<td>1.863(0.268)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD random effect</td>
<td>$\sqrt{d}$</td>
<td>7.9 $\times 10^{-11}$ (0.039)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2log-likelihood</td>
<td></td>
<td>7929.1</td>
<td>7929.1</td>
<td>7550.3</td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>7941.1</td>
<td>7941.1</td>
<td>7564.3</td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td>7969.8</td>
<td>7960.9</td>
<td>7597.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>WN</th>
<th>WG</th>
<th>WGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\xi_0$</td>
<td>-6.428(0.375)</td>
<td>-17.641(1.957)</td>
<td>736(1.674)</td>
</tr>
<tr>
<td>Sex effect</td>
<td>$\xi_1$</td>
<td>0.053(0.072)</td>
<td>0.474(1.626)</td>
<td>0.041(0.081)</td>
</tr>
<tr>
<td>History effect</td>
<td>$\xi_2$</td>
<td>-0.035(0.129)</td>
<td>0.166(3.286)</td>
<td>-0.046(0.280)</td>
</tr>
<tr>
<td>Status effect</td>
<td>$\xi_3$</td>
<td>-0.187(0.189)</td>
<td>-107.270(1.957)</td>
<td>-0.207(2.026)</td>
</tr>
<tr>
<td>Weibull shape</td>
<td>$\rho$</td>
<td>1.769(0.052)</td>
<td>173.010(1.205)</td>
<td>1.861(0.267)</td>
</tr>
<tr>
<td>Gamma parameter</td>
<td>$\alpha$</td>
<td>0.002(7.5$\times 10^{-05}$)</td>
<td>0.002(7.5$\times 10^{-05}$)</td>
<td>225(1.2$\times 10^{-05}$)</td>
</tr>
<tr>
<td>SD random effect</td>
<td>$\sqrt{d}$</td>
<td>8.7$\times 10^{-21}$ (0.040)</td>
<td>4.8$\times 10^{-05}$ (0.002)</td>
<td></td>
</tr>
<tr>
<td>-2log-likelihood</td>
<td></td>
<td>7550.3</td>
<td>9178.2</td>
<td>7550.3</td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>7566.5</td>
<td>9192.2</td>
<td>7564.3</td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td>7587.4</td>
<td>9225.8</td>
<td>7592.7</td>
</tr>
</tbody>
</table>

The scholars Molenberghs et al. (2010) and Oliveira (2014) have followed similar way to indicate the essence of combining random effects to deal with those special features in the data. The current result is supported by the literature and its validation exists.
9. Concluding remarks

A new family of distribution is introduced with the purpose of contributing to the application to survival analysis, lifetimes, health, and biomedical engineering. The proposed family is discussed in depth for a particular sub-family of distributions called NG-Weibull which is formed by taking the ordinary Weibull distribution as the baseline distribution in the NG-X family of distributions. Some basic statistical properties are discussed in detail.

The performance of this special member of the family is assessed by taking five common distributions as competitors and the model discrimination techniques are applied. The evaluation analysis shows the proposed distribution outperforms the rest five counter parts. The simulation study has shown that the parameter values estimated are quite steady and are close to the true parameter values as the sample size increases. It is shown that it has a greater application and preference than counter set models for survival analysis, lifetimes, biomedical engineering.

From the random effects models, we see that except WG, the models behaved similarly. However, there is still a clear discrimination that WN goes further, nearly followed by the combined model. Hence, based on the real data analysis, which is supported by the literature, the combined (WGN) random effect model is observed to be an improvement in its performance to best fit the data.

Multivariate extension will be the future direction for scholars in the same area. Conducting simulation study and testing significance of the regression model parameters for repeatedly measured over-dispersed time-to-event data to deal with overdispersion and correlation will be a task for the researchers in the similar area.

Data availability statement

The first data is available in the appendix and the second data can be obtained from the corresponding author upon request.

Conflicts of interest

The authors declare that there is no any kind of conflict of interest concerning the publication of this article.

Funding statement

This study is supported by the Yazd University, Iran.

References


### Appendix

Data for Section 6.2
dataFT=c(4.69, 0.01, 1.51, 0.02, 7.89, 0.03, 1.11, 0.04, 0.05, 0.06, 0.07, 0.07, 0.08, 0.09, 0.09, 0.10, 0.10, 0.11, 0.11, 0.12, 0.12, 0.13, 0.18, 0.19, 0.20, 0.23, 0.24, 0.24, 0.29, 0.34, 0.35, 0.36, 0.38, 0.40, 0.42, 0.43, 0.52, 0.54, 0.56, 0.60, 0.60, 0.63, 0.65, 0.67, 0.68, 0.72, 0.72, 0.72, 0.73, 0.79, 0.79, 0.80, 0.80, 0.83, 0.85, 0.90, 0.92, 0.95, 0.99, 1.00, 1.01, 1.02, 1.03, 1.05, 1.10, 1.10, 0.03, 1.15, 1.18, 1.20, 1.29, 1.31, 1.33, 1.34, 1.40, 1.43, 1.45, 1.50, 0.02, 1.52, 1.53, 1.54, 1.54, 1.55, 1.58, 1.60, 1.63, 1.64, 1.80, 1.80, 1.81, 2.02, 2.05, 2.14, 2.17, 2.33, 3.03, 3.03, 3.34, 3.4, 4.20, 0.01, 0.02).
Figure 5: Plot for estimated parameters, MSEs, biases and absolute biases for case 1: $\alpha = 0.7, \delta = 0.4, \theta = 0.5, \gamma = 0.3$. 

A New Generalized-$X$ Family of Distributions: Applications, Characterization and a Mixture of Random Effect Models

502
Figure 6: Plot for estimated parameters, MSEs, biases and absolute biases for case 2: $\alpha = 0.5$, $\delta = 0.5$, $\theta = 0.6$, $\gamma = 1$. 

A New Generalized-X Family of Distributions: Applications, Characterization and a Mixture of Random Effect Models
Figure 7: Plot for estimated parameters, MSEs, biases and absolute biases for case 3: $\alpha = 0.2, \delta = 0.5, \theta = 0.6, \gamma = 0.3$. 

A New Generalized-X Family of Distributions: Applications, Characterization and a Mixture of Random Effect Models

504