

Research Article

A New Type 1 Alpha Power Family of Distributions and Modeling Data with Correlation, Overdispersion, and Zero-Inflation in the Health Data Sets

Getachew Tekle D, Rasool Roozegar D, and Zubair Ahmad D

Department of Statistics, Yazd University, P.O. Box: 89175-741, Yazd, Iran

Correspondence should be addressed to Rasool Roozegar; rasool62@gmail.com

Received 18 May 2023; Revised 4 July 2023; Accepted 13 August 2023; Published 31 August 2023

Academic Editor: Alessandro De Gregorio

Copyright © 2023 Getachew Tekle et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the recent era, the introduction of a new family of distributions has gotten great attention due to the curbs of the classical univariate distributions. This study introduces a novel family of distributions called a new type 1 alpha power family of distributions. Based on the novel family, a special model called a new type 1 alpha power Weibull model is studied in depth. The new model has very interesting patterns and it is very flexible. Thus, it can model the real data with the failure rate patterns of increasing, decreasing, parabola-down, and bathtub. Its applicability is studied by applying it to the health sector data, and time-to-recovery of breast cancer patients, and its performance is compared to seven well-known models. Based on the model comparison, it is the best model to fit the health-related data with no exceptional features. Furthermore, the popular models for the data with exceptional features such as correlation, overdispersion, and zero-inflation in aggregate are explored with applications to epileptic seizer data. Sometimes, these features are beyond the probability distribution models. Hence, this study has implemented eight possible models separately to these data and they are compared based on the standard techniques. Accordingly, the zero-inflated Poisson-normal-gamma model which includes the random effects in the linear predictor to handle the three features simultaneously has shown its supremacy over the others and is the best model to fit the health-related data with these features.

1. Introduction

Statistical modeling and predicting real-life events are vital issues in the training and implementation of the health care and health sector in general [1]. Although the classical and modified statistical models have been applied to the data in health applications, they do not provide the best fit when the data show nonmonotonic failure rates. This clearly demands the generalized or extended versions of these classical models. Thus, it motivated many involved researchers to propose new flexible extensions of distributions by adding one or more additional parameters to the baseline distribution.

Correspondingly, our study acquaints with a more flexible family of distributions called a new alpha power type 1 family of distributions by introducing a new parameter α to the exponential type of family of distributions. It is more suitable for skewed data with non-monotonic failure rates and it shows novelties in the area of distribution theory.

Several new developments in the distribution theory are proposed in the literature. Roozegar and Nadarajah [2] intended to study the quadratic hazard rate power series distribution. The same authors Roozegar and Nadarajah [3] were also devoted to propose the power series skew normal class of distributions, by taking power series as a key issue. Very recently, Chesneau et al. [4] introduced a new extended family of distributions called an alternative to the Marshall–Olkin family of distributions. They used five different estimation methods to reveal the alternating capacity of the new family to the existing one and they used the subcases for the regression purpose. The other scholars, Elbatal et al. [5], also suggested a new class for the generalized distributions, which they named it the alpha power Weibull G (APW-G) family.

Shehata et al. [6] introduced a flexible family of distributions for the asymmetric left-skewed bimodal real-life data with special attention to the flexibility patterns of the probability density and hazard functions and they called it a novel two-parameter G family of distributions. They used a copula method to characterize the new family for the special model, the new extension of the exponential distribution.

Some researchers used the techniques such as transformation, extension, and compounding to introduce a new family of distributions. Chesneau et al. [7] proposed a new family of probability distributions, based on a cosine-sine transformation by compounding a baseline distribution with the cosine and sine functions. The other authors, Ahmad et al. [8], applied the same approach by adding a parameter to introduce a new class of probability distributions. They named the newly suggested model as the extended alpha power transformed family of distributions and the extended alpha power transformed Weibull distribution is studied as a special model. A few other scholars who followed this approach, among many, are Ahmad et al. [8]to propose a new exponentiated TX class of distributions. Hussein et al. [9] to introduce a new flexible modified alpha power (MAP) family of distributions by adding two parameters to the baseline model, and El-Sherpieny et al. [10] to suggest a new generating family of distributions.

In line with the introduction of the new family of probability distributions, the exploration of the models for some special features in the data is reasonably needed. Thus, repeatedly measured count outcomes are characterized by three special features. These are the dependence of the individual subject-specific due to the clustering effect or correlation, extravariability due to the counts (overdispersion), and a special case of overdispersion which is said to be zero-inflation [11–15].

One of the motivating gaps this study raises is that Mekonnen et al. [16] used epilepsy data by allowing both correlation and zero-inflation in the data and they analyzed the data by using the Poisson (P), negative-binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negativebinomial (ZINB) models. Their study has two defects: the first one is it could not raise and discuss the issue of overdispersion while the data are overdispersed and the second one is that the mentioned count models alone cannot fully handle the dependence in the data. It is not a straightforward and easy task to handle the three features in the same model. Due to this, many authors in the literature dealt with one or two of these features [12, 16]. Unlike other studies, this study is motivated to model those three features simultaneously.

Moreover, the repeatedly measured count data such as epileptic seizure are senseless to be analyzed by using the distribution models unless they are easy to be expressed in the exponential form. Hence, the need for an update on this gap is raised in this study reasonably and convincingly. The rest part of the paper is organized as follows: The first part of this study introduces a new family of distributions called a new type 1 alpha power family of distributions to the health data. This part is presented in Sections (1–6) in detail. The second part discusses the models for correlation, overdispersion, and zero-inflation in the data (see Section 7). Thus, Section 2 introduces a new type 1 alpha power family of distributions, Section 3 discusses special cases, and Section 4 discusses the basic statistical properties of the new proposed family. Section 5 presents an estimation of the model parameters of the new subfamily of distributions. Section 6 illustrates the application of the new family to the new data set. Section 7 discusses the models for correlation, overdispersion, and zero-inflation, and lastly, Section 8 summarizes the points with concluding remarks.

2. A New Type 1 Alpha Power Family of Distributions

Alzaatreh et al. [17] introduced a popular T-X approach which updates the distributional flexibility of the existing models and it has become a well-known approach in the literature. They proposed the probability density function (PDF) of it as follows:

$$F(z;\boldsymbol{\omega}) = \int_{0}^{W[G(z;\boldsymbol{\omega})]} r(t) \mathrm{d}t, \qquad (1)$$

where the function $W[G(z; \boldsymbol{\omega})]$ fulfills some specified conditions; for detail, see Alzaatreh et al. [17]. In equation (1), $G(z; \boldsymbol{\omega})$ is a baseline cumulative distribution function (CDF) with parameter vector $\boldsymbol{\omega}$ and r(t) is the PDF of the parent model with parameter vector $\boldsymbol{\omega}$. Corresponding to $(z; \boldsymbol{\omega})$, the PDF $f(z; \boldsymbol{\omega})$ is given by the following equation:

$$f(z; \boldsymbol{\omega}) = \left\{ \frac{\mathrm{d}}{\mathrm{d}z} W[G(z; \boldsymbol{\omega})] \right\} r\{W[G(z; \boldsymbol{\omega})]\}, \quad z, \boldsymbol{\omega} \in \mathbb{R}.$$
(2)

Recently, Ahmad et al. [18] applied the T-X approach and proposed an interesting member, which they call the weighted *T*-X Weibull distribution family (WT-XW). They introduced this family by using $W[G(z; \boldsymbol{\omega})] = -\log\{(1 - G(z; \boldsymbol{\omega}))/e^{G(z; \boldsymbol{\omega})}\}$ in equation (1) with $r(t) = \phi \varphi z^{\phi-1} e^{-\varphi z^{\phi}}$, where r(t) is the PDF of the two-parameter Weibull model with the parameters (ϕ and φ). The CDF of the *WT-XW* family is given by the following equation:

$$F(z;\boldsymbol{\omega}) = 1 - \left(\frac{1 - G(z;\boldsymbol{\omega})}{e^{G(z;\boldsymbol{\omega})}}\right), \quad z, \boldsymbol{\omega} \in \mathbb{R},$$
(3)

and the corresponding PDF is given by the following equation:

$$f(z;\boldsymbol{\omega}) = \frac{g(z;\boldsymbol{\omega})}{e^{G(z;\boldsymbol{\omega})}} \{2 - G(z;\boldsymbol{\omega})\}, \quad z,\boldsymbol{\omega} \in \mathbb{R}.$$
(4)

The PDF and CDF of exponential type are given by the following equations:

$$f(z;\boldsymbol{\omega}) = m(z;\boldsymbol{\omega})e^{-M(z;\boldsymbol{\omega})}, \quad \boldsymbol{\omega}, z \in \mathbb{R},$$
(5)

and

$$F(z;\boldsymbol{\omega}) = 1 - e^{-M(z;\boldsymbol{\omega})}, \quad \boldsymbol{\omega}, z \in \mathbb{R},$$
(6)

respectively.

Note that, $M(z; \boldsymbol{\omega})$ must satisfy the following conditions.

- M(z; ω) is a non-negative, differentiable, and increasing function of z.
- (2) $\lim_{z \to -\infty} M(z; \boldsymbol{\omega}) \longrightarrow 0$ and $\lim_{z \to +\infty} M(z; \boldsymbol{\omega}) \longrightarrow +\infty$.

The classical exponential types, Rayleigh, Weibull, and other extended lifetime distributions belong to the class defined in equation (6); see Liao et al. [19]. For further detailed information about similar statistical distributions, see Mehboob Zaidi et al. [20], Afify et al. [21], Reyad et al. [22], and Al-Babtain et al. [23].

Herein after, we introduced an additional parameter α in equation (6), which replaces the exponent term *e* to propose a very flexible family whose CDF and PDF are given by the following equations:

$$F(z; \boldsymbol{\omega}) = 1 - \alpha^{-M(z; \boldsymbol{\omega})}, \quad \alpha > 1, \boldsymbol{\omega}, z \in \mathbb{R},$$
(7)

and

$$f(z; \boldsymbol{\omega}) = m(z; \boldsymbol{\omega}) \alpha^{-M(z; \boldsymbol{\omega})}, \quad \alpha > 1, \boldsymbol{\omega}, z \in \mathbb{R},$$
(8)

respectively, where $\alpha \neq e$ throughout the paper.

If $\alpha = e$, the CDF in equation (7) becomes similar to equation (6). The function $M(z; \omega)$ fulfills the conditions which are given in 1 and 2. It is straightforward that $0 \le F(z; \omega) \le 1$. To make it more clear, let

$$\lim_{z \to -\infty} F(z; \boldsymbol{\omega}) = \lim_{z \to -\infty} \left(1 - \alpha^{-M(z; \boldsymbol{\omega})} \right) = 0, \qquad (9)$$

and

$$\lim_{z \to +\infty} F(z; \boldsymbol{\omega}) = \lim_{z \to +\infty} \left(1 - \alpha^{-M(z; \boldsymbol{\omega})} \right) = 1.$$
(10)

Therefore, based on the results in equations (9) and (10), it is observed that the function $F(z; \boldsymbol{\omega})$ written in equation (7) is a proper CDF. And the expression given in equation (7) is very helpful and is useful to generate new statistical models belonging to the T-X family of distributions.

Next, we proposed a new family called a new type 1 alpha power (NT1AP) family by using $W[G(z; \boldsymbol{\omega})] = -\log\{1 - G(z; \boldsymbol{\omega})/e^{G(z; \boldsymbol{\omega})}\}$ in equation (7). The CDF and PDF of the NT1AP family are given by the following equations:

$$F(z;\boldsymbol{\omega}) = 1 - \alpha^{\log\{(1 - G(z;\boldsymbol{\omega}))/e^{G(z;\boldsymbol{\omega})}\}}, \quad \alpha > 1, \boldsymbol{\omega}, z \in \mathbb{R}, \quad (11)$$

and

$$f(z;\boldsymbol{\omega}) = \frac{\log(\alpha)g(z;\boldsymbol{\omega})(2-G(z;\boldsymbol{\omega}))}{1-G(z;\boldsymbol{\omega})}\alpha^{\log\{(1-G(z;\boldsymbol{\omega}))/e^{G(z;\boldsymbol{\omega})}\}}, \quad \alpha > 1, \boldsymbol{\omega}, z \in \mathbb{R},$$
(12)

respectively.

In addition, the corresponding survival function $S(z; \omega)$ and the hazard function $h(z; \omega)$ are given by the following equation:

$$S(z; \boldsymbol{\omega}) = \alpha^{\log\{(1 - G(z; \boldsymbol{\omega}))/e^{G(z; \boldsymbol{\omega})}\}}, \quad \alpha > 1, \boldsymbol{\omega}, z \in \mathbb{R},$$
$$h(z; \boldsymbol{\omega}) = \frac{\log(\alpha)g(z; \boldsymbol{\omega})(2 - G(z; \boldsymbol{\omega}))}{1 - G(z; \boldsymbol{\omega})}, \quad \alpha > 1, \boldsymbol{\omega}, z \in \mathbb{R},$$
(13)

respectively.

3. A Special Subfamily

In this section, we discuss a special member of the NT1AP family called a new type 1 alpha power Weibull (NT1AP-W)

model. It is introduced by inserting the CDF and PDF of the two-parameter Weibull into equations (11) and (12), necessarily. Thus, let a random variable Z have the Weibull distribution, then its CDF and PDF, respectively, are given by the following equation:

$$G(z;\lambda,\beta) = 1 - e^{-\beta z^{\lambda}}, \quad \lambda,\beta > 0, z \ge 0,$$

$$g(z;\lambda,\beta) = \lambda\beta z^{\lambda-1} e^{-\beta z^{\lambda}}, \quad \lambda,\beta > 0, z > 0,$$
(14)

and the resulting CDF and PDF of the NT1AP-W are given by the following equations:

$$F(z;\boldsymbol{\omega}) = 1 - \alpha^{\left(e^{-\beta z^{\lambda}} - \beta z^{\lambda} - 1\right)}, \alpha > 1, \lambda, \beta > 0, z \ge 0,$$
(15)

and

$$f(z;\boldsymbol{\omega}) = \log(\alpha)\lambda\beta z^{\lambda-1} \left(1 + e^{-\beta z^{\lambda}}\right) \alpha^{\left(e^{-\beta z^{\lambda}} - \beta z^{\lambda} - 1\right)}, \quad \alpha > 1, \lambda, \beta > 0, z > 0,$$
(16)

respectively, where $\boldsymbol{\omega} = (\lambda, \beta)$.

Based on equations (15) and (16), the corresponding $S(z; \boldsymbol{\omega})$ and $h(z; \boldsymbol{\omega})$ are given by the following equations:

$$S(z; \boldsymbol{\omega}) = \alpha^{\left(e^{-\beta z^{\lambda}} - \beta z^{\lambda} - 1\right)}, \quad \alpha > 1, \lambda, \beta > 0, z > 0 \text{ and}$$
$$h(z; \boldsymbol{\omega}) = \log(\alpha)\lambda\beta z^{\lambda-1} \left(1 + e^{-\beta z^{\lambda}}\right), \quad \alpha > 1, \lambda, \beta > 0, z > 0,$$
(17)

respectively.

Following this, the graphical expression of the NT1AP-W model for different parameter values is displayed as follows.

Figure 1(a) displays the plot of $f(z; \alpha, \omega)$ for the NT1AP-W for different scenarios: (*i*) $\alpha = 2.2, \beta = 4.1, \lambda = 0.7$ (blueline), (*ii*) $\alpha = 5.1, \beta = 0.7, \lambda = 1.2$ (purple-line), (*iii*) $\alpha = 1.3, \beta = 7.1, \lambda = 3.5$ (green-line), (*iv*) $\alpha = 1.6, \beta = 5, \lambda = 1.4$ (redline), and (ν) $\alpha = 1.01, \beta = 11, \lambda = 0.6$ (black-line).

Figure 1(b) displays the plot of $h(z; \alpha, \omega)$ for the NT1AP-W for different scenarios: (*i*) $\alpha = 6, \beta = 0.6, \lambda = 53$ (purpleline), (*ii*) $\alpha = 8, \beta = 0.7, \lambda = 0.8$ (cyan-line), (*iii*) $\alpha = 8, \beta = 0.3, \lambda = 1.1$ (green-line), (*iv*) $\alpha = 4.5, \beta = 0.3, \lambda = 0.5$ (redline), and (*v*) $\alpha = 1.5, \beta = 2.1, \lambda = 7.2$ (black-line).

 $f(z; \alpha, \omega)$ of the NT1AP-W is elicited in Figure 1(a) and it has attractive flexible patterns such as (*i*) decreasing, (*ii*)

parabola-down, (*iii*) left-skewed, (*iv*) right-skewed, and (*v*) decreasing-increasing-decreasing-constant (polynomial type). Figure 1(b) illustrates $h(z; \alpha, \omega)$ for different parameter values to show the different patterns and how flexible the distribution is. It has the patterns such as (*i*) increasing, (*ii*) bathtub, (*iii*) parabola-down, and (*iv*) decreasing.

4. Some Basic Statistical Properties of the New Type 1 Alpha Power Family of Distributions

In this section, we discuss the basic statistical properties of the NT1AP family of distributions.

4.1. Order Statistics. Order statistics are widely used in applied statistics such as reliability and lifetime and records. Suppose that Z_1, Z_2, \ldots, Z_n is a random sample of size *n* following the NT1AP family of distributions with parameters (α, ω) and $Z_{1:n}, Z_{2:n}, \ldots, Z_{n:n}$ are its corresponding order statistics. Then, the density function of $Z_{i:n}$ for $(i = 1, 2, \ldots, n)$ is given by the following equation:

$$f_{i:n}(z) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f_{NT1AP}(z; \boldsymbol{\omega}) [F_{NT1AP}(z; \boldsymbol{\omega})]^{i+j-1}.$$
(18)

By substituting the CDF and PDF of the NT1AP (see equations (11) and (12)) into $f_{i:n}(z)$, we obtain the following equation:

$$f_{i:n}(z) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^{j} {\binom{n-i}{j}} [\log(\alpha)]^{i+1} [g(z;\boldsymbol{\omega})]^{i-1} [2 - G(z;\boldsymbol{\omega})]^{i-1} \times [\overline{G}(z;\boldsymbol{\omega})]^{-i+1} \alpha^{\log\left[\overline{G}(z;\boldsymbol{\omega}/e^{\overline{G}(z;\boldsymbol{\omega})}\right]^{i-1}} \left[1 - \alpha^{\log\left[\overline{G}(z;\boldsymbol{\omega}/e^{\overline{G}(z;\boldsymbol{\omega})}\right]}\right]^{i+j-1}, \quad \alpha > 1, z, \boldsymbol{\omega} \in \mathbb{R},$$

$$(19)$$

where $\overline{G}(z; \boldsymbol{\omega}) = 1 - G(z; \boldsymbol{\omega})$.

The simplified form of the order statistics density function is given by the following equation:

$$f_{i:n}(z) = \sum_{j=0}^{n-i} \Phi_j f_{NT1AP}^*(z; \boldsymbol{\omega}),$$
 (20)

where

$$\Phi_j = \frac{n!}{(i-1)! (n-i)!} (-1)^j \binom{n-i}{j},$$
 (21)

and $f^*_{NT1AP}(z; \boldsymbol{\omega})$ is the product of the PDF and the CDF for the *i*th order statistics.

The PDF of the order statistics for the NT1AP can be obtained from equation (20) and its r^{th} central moments and moments generating function are given in the next subsection.





FIGURE 1: Plots of $f(z; \alpha, \omega)$ and $h(z; \alpha, \omega)$ for the NT1AP-W distribution for different scenarios. (a) Visual illustration of $f(z; \alpha, \omega)$ for the NT1AP-W. (b) Visual illustration of $h(z; \alpha, \omega)$ for the NT1AP-W.

4.2. Moments and Moment-Generating Functions. Based on the PDF of the NT1AP family of distributions given above, its r^{th} central moments, $E(z^r) = \mu'_r$, is obtained as follows:

$$\mu_{r}' = \int_{0}^{+\infty} z^{r} \frac{\log(\alpha)g(z;\boldsymbol{\omega})(2-G(z;\boldsymbol{\omega}))}{(1-G(z;\boldsymbol{\omega}))} \alpha^{\log(\bar{G}(z;\boldsymbol{\omega})/e^{G(z;\boldsymbol{\omega})})} dz, \quad \alpha > 1, z, \boldsymbol{\omega} \in \mathbb{R}$$

$$= \int_{0}^{+\infty} [\log(\alpha)]^{i+1} z^{ri-1} [g(z;\boldsymbol{\omega})]^{i-1} [2-G(z;\boldsymbol{\omega})]^{i-1} [\bar{G}(z;\boldsymbol{\omega})]^{-i+1} \alpha^{\log[\bar{G}(z;\boldsymbol{\omega})/e^{G(z;\boldsymbol{\omega})}]^{i-1}} dz \qquad (22)$$

$$= \sum_{i=0}^{+\infty} \left[\frac{\log(\alpha)}{i!} \right]^{i+1} \Psi_{r,i+1},$$

where

$$\Psi_{r,i+1} = \int_{0}^{+\infty} z^{ri-1} [g(z;\boldsymbol{\omega})]^{i-1} [2 - G(z;\boldsymbol{\omega})]^{i-1} [\overline{G}(z;\boldsymbol{\omega})]^{-i+1} \alpha^{\log \left[\overline{G}(z;\boldsymbol{\omega})/e^{G(z;\boldsymbol{\omega})}\right]^{i-1}} dz.$$
(23)

The moment generating function of the NT1AP can be obtained by using the last result of μ'_r in $M_z(t)$ as follows:

$$M_{z}(t) = \sum_{b=0}^{+\infty} \frac{t^{b}}{b!} \mu'_{r}.$$
 (24)

4.3. Mean Deviation and Bonferroni and Lorenz Curves. Let $Z \sim \text{NT1AP}(\alpha, \omega)$; the mean deviation about the mean and the median are defined by the following equations:

$$\delta_{1}(Z) = \int_{0}^{+\infty} |z - \mu| f_{NT1AP}(z) dz \text{ and}$$

$$\delta_{2}(Z) = \int_{0}^{+\infty} |z - M| f_{NT1AP}(z) dz,$$
(25)

respectively, where $z \in \mathbb{R}$, $\mu = E(Z)$, and M = Median(Z) denotes the median.

These can further be expressed as follows:

$$\delta_{1}(Z) = 2\mu F_{NT1AP}(\mu) - 2 \int_{0}^{\mu} z f_{NT1AP}(z) dz \text{ and}$$

$$\delta_{2}(Z) = \mu - 2 \int_{0}^{M} z f_{NT1AP}(z) dz,$$
(26)

respectively, where $m(x) = \int_0^x z f_{NT1AP}(z) dz$ is the first incomplete moment. These measures have been applied to a wide variety of fields, such as reliability, demography, insurance, and medicine [24].

Moreover, let $Z \sim \text{NT1AP} (\alpha, \omega)$; the Bonferroni and Lorenz curves are defined by the following equations:

$$B(p) = \frac{1}{p\mu} \int_{0}^{q} z f_{NT1AP}(z) dz \text{ and}$$

$$L(p) = \frac{1}{\mu} \int_{0}^{q} z f_{NT1AP}(z) dz,$$
(27)

respectively, where $\mu = E(Z)$ and $q = F_{NT1AP}^{-1}(p)$.

The next section deals with the maximum likelihood estimation for the model parameters of the NT1AP-W model.

5. Estimation of the Model Parameters of the NT1AP-W

The method of the maximum likelihood estimation for the model parameters for the NT1AP-W is discussed in this section.

5.1. Maximum Likelihood Estimation. This subsection deals with the computation of the maximum likelihood estimators (MLEs) for the model parameters of the NT1AP-W. Let z_1, z_2, \ldots, z_n be *n* observations of a random sample drawn from the NT1AP-W with parameters α , λ , and β . By using the PDF of the NT1AP-W (see equation (16)), the likelihood function is given by the following equation:

Journal of Probability and Statistics

$$L(z_1, z_2, \dots, z_n | \boldsymbol{\Theta}) = (\lambda \beta \log(\alpha))^n \prod_{i=1}^n z_i^{\lambda-1} \left(1^n + e^{-\beta \sum_{i=1}^n z_i^{\lambda}} \right) \alpha^{\sum_{i=1}^n \left(e^{-\beta z_i^{\lambda}} - \beta z_i^{\lambda-1} \right)},$$
(28)

and its corresponding log-likelihood function $(\log L(z; \Theta))$ is given as follows:

$$\log L(z; \boldsymbol{\Theta}) = n \log(\log(\alpha)) + n \log(\lambda) + n \log(\beta) + (\lambda - 1) \sum_{i=1}^{n} \log(z_i) + n \log(1) - \beta \sum_{i=1}^{n} z_i^{\lambda} + \log(\alpha) \sum_{i=1}^{n} \left(e^{-\beta z_i^{\lambda}} - \beta z_i^{\lambda} - 1 \right),$$
(29)

where $\boldsymbol{\Theta} = (\alpha, \omega)^T$. The model parameters are estimated by taking the first partial derivatives of the log $L(z; \boldsymbol{\Theta})$ with respect to each model parameter and equating them to zero.

Thus, having the $\log L(z; \Theta)$, the partial derivatives of it with respect to each parameter are given by the following equation:

$$\frac{\partial \log L(z; \mathbf{\Theta})}{\partial \alpha} = S_1 = \frac{n}{\log(\alpha)\alpha} - \frac{1}{\alpha} \sum_{i=1}^n \left(e^{-\beta z_i^{\lambda}} - \beta z_i^{\lambda} - 1 \right),$$

$$\frac{\partial \log L(z; \mathbf{\Theta})}{\partial \lambda} = S_2 = \frac{n}{\lambda} + \sum_{i=1}^n \log(z_i) - \beta \sum_{i=1}^n z_i^{\lambda} \log(z_i) - \log(\alpha) \sum_{i=1}^n \left(\beta z_i^{\lambda} \log(z_i) e^{-\beta z_i^{\lambda}} + \beta z_i^{\lambda} \log(z_i) \right),$$

$$\frac{\partial \log L(z; \mathbf{\Theta})}{\partial \beta} = S_3 = \frac{n}{\beta} - \sum_{i=1}^n z_i^{\lambda} - \log(\alpha) \sum_{i=1}^n \left[z_i^{\lambda} e^{-\beta z_i^{\lambda}} + z_i^{\lambda} \right],$$
(30)

where S_i is the *i*th score function for i = 1, 2, 3.

Subsequently, the MLEs of the parameters can be obtained by solving the following nonlinear equation:

$$\mathbf{U}_{\mathbf{n}}(\boldsymbol{\Theta}) = \left(\frac{\partial \log L(z; \boldsymbol{\Theta})}{\partial \alpha}, \frac{\partial \log L(z; \boldsymbol{\Theta})}{\partial \lambda}, \frac{\partial \log L(z; \boldsymbol{\Theta})}{\partial \beta}\right)^{T} = 0,$$
(31)

using numerical methods such as Newton-Raphson or Broyden's methods.

Furthermore, the Fisher information matrix is given by the following equation:

$$I(\mathbf{\Theta}) = \left[I_{\mathbf{\Theta}_i,\mathbf{\Theta}_j}\right] 3 \times 3,\tag{32}$$

where *i*, *j* = 1, 2, 3, and

$$I_{\Theta_i,\Theta_j} = E\left(-\frac{\partial^2 \log L(z;\Theta)}{\partial \Theta_i \partial \Theta_j}\right).$$
 (33)

The total *I* is given by $I_n(\Theta) = nI(\Theta)$, which can be approximated by the following equation:

$$I_n(\widehat{\mathbf{\Theta}}) \approx \left[-\frac{\partial^2 \log L(z; \mathbf{\Theta})}{\partial \mathbf{\Theta}_i \partial \mathbf{\Theta}_j} \mid \mathbf{\Theta} = \widehat{\mathbf{\Theta}} \right] 3 \times 3.$$
(34)

More analytically,

$$I(\boldsymbol{\Theta}) = E_{\boldsymbol{\Theta}} \left(SS^{T} \right) = E_{\boldsymbol{\Theta}} \left[\begin{bmatrix} s_{1} \\ s_{2} \\ s_{3} \end{bmatrix} \begin{bmatrix} s_{1} & s_{2} & s_{3} \end{bmatrix} \right] = E_{\boldsymbol{\Theta}} \left[\begin{bmatrix} s_{1}^{2} & s_{12} & s_{13} \\ s_{21} & s_{2}^{2} & s_{23} \\ s_{31} & s_{32} & s_{3}^{2} \end{bmatrix} \right],$$
(35)

where

$$s_{1}^{2} = \frac{\partial}{\partial \alpha} \left[\frac{\partial \log L(z; \Theta)}{\partial \alpha} \right] = -\frac{n}{\alpha^{2} (\log(\alpha))^{2}} + \sum_{i=1}^{n} \left(e^{-\beta z_{i}^{\lambda}} - \beta z_{i}^{\lambda} - 1 \right),$$

$$s_{12} = \frac{\partial s_{1}}{\partial \lambda} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \alpha \partial \lambda} = -\frac{1}{\alpha} \sum_{i=1}^{n} \left(\beta z_{i}^{\lambda} \log(z_{i}) e^{-\beta z_{i}^{\lambda}} + \beta z_{i}^{\lambda} \log(z_{i}) \right),$$

$$s_{13} = \frac{\partial s_{1}}{\partial \beta} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \alpha \partial \beta} = -\frac{1}{\alpha} \sum_{i=1}^{n} \left(z_{i}^{\lambda} e^{-\beta z_{i}^{\lambda}} + z_{i}^{\lambda} \right),$$

$$s_{21} = \frac{\partial s_{2}}{\partial \alpha} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \lambda \partial \alpha} = -\frac{1}{\alpha} \sum_{i=1}^{n} \left(\beta z_{i}^{\lambda} \log(z_{i}) e^{-\beta z_{i}^{\lambda}} + \beta z_{i}^{\lambda} \log(z_{i}) \right),$$

$$s_{22}^{2} = \frac{\partial s_{2}}{\partial \alpha} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \lambda \partial \lambda} = -\frac{n}{\lambda^{2}} - \beta (1 + \log(\alpha)) \sum_{i=1}^{n} \left[(\log(z_{i}))^{2} z_{i}^{\lambda} \left(e^{-\beta z_{i}^{\lambda}} (1 - \beta z_{i}^{\lambda}) + 1 \right) \right],$$

$$s_{31} = \frac{\partial s_{3}}{\partial \alpha} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \beta \partial \alpha} = -\frac{1}{\alpha} \sum_{i=1}^{n} (z_{i}^{\lambda} (e^{-\beta z_{i}^{\lambda}} + 1)),$$

$$s_{32} = \frac{\partial s_{3}}{\partial \lambda} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \beta \partial \lambda} = -(1 + \log(\alpha)) \sum_{i=1}^{n} (\log(z_{i}) z_{i}^{\lambda}) \sum_{i=1}^{n} (e^{-\beta z_{i}^{\lambda}} (1 - \beta^{2} z_{i}^{\lambda} \log(z_{i})) + 1),$$

$$s_{32} = \frac{\partial s_{3}}{\partial \lambda} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \beta \partial \lambda} = -(1 + \log(\alpha)) \sum_{i=1}^{n} (\log(z_{i}) z_{i}^{\lambda}) \sum_{i=1}^{n} (e^{-\beta z_{i}^{\lambda}} (1 - \beta z_{i}^{\lambda}) + 1),$$

$$s_{32}^{2} = \frac{\partial s_{3}}{\partial \beta} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \beta \partial \lambda} = -(1 + \log(\alpha)) \sum_{i=1}^{n} (\log(z_{i}) z_{i}^{\lambda}) \sum_{i=1}^{n} (e^{-\beta z_{i}^{\lambda}} (1 - \beta z_{i}^{\lambda}) + 1),$$

$$s_{32}^{2} = \frac{\partial s_{3}}{\partial \beta} = \frac{\partial^{2} \log L(z; \Theta)}{\partial \beta \partial \lambda} = -(1 + \log(\alpha)) \sum_{i=1}^{n} (\log(z_{i}) z_{i}^{\lambda}) \sum_{i=1}^{n} (e^{-\beta z_{i}^{\lambda}} (1 - \beta z_{i}^{\lambda}) + 1),$$

For an interval estimation of the model parameters, we need the 3 × 3 observed information matrix $I(\Theta) = \{-I_{rw}\}$, where $r, w = \alpha, \lambda, \beta$. Under the regularity conditions, the multivariate normal $\sqrt{n} (\widehat{\Theta} - \Theta) \sim N_3(\mathbf{0}, I(\widehat{\Theta})^{-1})$ distribution is used to construct approximate confidence intervals for the model parameters. Here, $I(\widehat{\Theta})$ is the total observed information matrix evaluated at Θ . Then, the 100(1- α^*)% confidence intervals for α, λ , and β are given by $\widehat{\alpha} \pm z_{\alpha^*/2} \times \sqrt{\operatorname{var}(\widehat{\alpha})}$, $\widehat{\lambda} \pm z_{\alpha^*/2} \times \sqrt{\operatorname{var}(\widehat{\lambda})}$, and $\widehat{\beta} \pm z_{\alpha^*/2} \times \sqrt{\operatorname{var}(\widehat{\beta})}$, respectively, where the var(.)' s denote the diagonal elements of the $I(\widehat{\Theta})^{-1}$ corresponding to the model parameters, and the $z_{\alpha^*/2}$ is the quantile $(1-\alpha^*/2)$ of the standard normal distribution.

6. An Application to Breast Cancer Data

Breast cancer is one of the most severe diseases in the world and has become the public's everyday agenda in both developed and developing countries. The new data on the timeto-recovery of 686 breast cancer patients were taken from a patient's medical record card that was enrolled from October 2012 to April 2017 in Nigist Elleni Mohamad memorial referral comprehensive hospital (NEMMRCH), Hossana, south Ethiopia; see Figure 2. We illustrated the fitting capacity of the NT1AP-W model to the data by comparing it to the threeparameters exponential flexible Weibull extension (EFWE) of El-Desouky et al. [25], the three-parameters Poisson inverse Weibull (PIW) of Joshi and Kumar [26], the five-parameters exponentiated Weibull-Weibull (EWW) of Hassan and Elgarhy [27], the three-parameters Alpha Power Transformed Weibull (APTW) of Elbatal et al. [28], the five-parameters Kumaraswamy Weibull Poisson (KWP) of Marinho et al. [29], the four-parameters Kumaraswamy Weibull (KW) of Cordeiro et al. [30], and the threeparameters New Weighted Weibull (NWW) of Elsherpieny et al. [31].

The competing models with their corresponding CDFs are given by the following equation.

$$F(z; \alpha, \beta, \lambda) = 1 - e^{-\lambda e^{\left(e^{\alpha z - \langle \beta \rangle z}\right)}}, \quad z \ge 0, \alpha, \beta, \lambda > 0.$$
(37)

(2) EWW:

$$F(z; a, \alpha, \beta, \lambda, \gamma) = \left[1 - e^{-\alpha \left(e^{\lambda z^{\gamma} - 1}\right)^{\beta}}\right]^{a}, \quad z \ge 0, a, \alpha, \beta, \lambda, \gamma > 0.$$
(38)

(3) APTW:

(4) KWP:

$$F(z; \alpha, \theta, \gamma) = \frac{\alpha^{(1-e^{-\gamma z^{\alpha}})} - 1}{\alpha - 1}, \quad z \ge 0, \alpha \ne 1, \theta, \gamma > 0.$$
(39)

$$F(z; a, b, c, \beta, \lambda) = \frac{1 - e^{-\lambda \left[1 - \left(1 - \left(1 - e^{-(\beta z)^{c}}\right)^{a}\right)^{b}\right]}}{1 - e^{-\lambda}}, \quad z \ge 0, a, b, c, \beta, \lambda > 0.$$
(40)

(5) KW:

$$F(z; a, b, \gamma, \theta) = 1 - \left(1 - \left(1 - e^{-\gamma z^{\theta}}\right)^{a}\right)^{b}, \quad z \ge 0, a, b, \gamma, \theta > 0.$$
(41)

(6) NWW:

 $F(z; \alpha, \theta, \gamma) = 1 - e^{-\alpha z^{\theta} + \alpha (\gamma z)^{\theta}}, \quad \alpha, \gamma, \theta > 0, z \ge 0.$ (42)

(7) PIW:

$$F(z; \alpha, \beta, \lambda) = \frac{1}{\left(1 - e^{-\lambda}\right)} \left[1 - e^{-\lambda e^{-\left(\alpha/z\right)^{\beta}}}\right], \quad z \ge 0, \alpha, \beta, \lambda > 0.$$
(43)

The information criteria (IC) such as (i) AIC [32], (ii) CAIC [33], (iii) BIC [34], and (iv) HQIC [35] are used to discriminate the best model. In addition to these criteria, the log-likelihood (-2logL) of the fitted models is also calculated. In all these, the model with the least IC value is taken to be the best model to fit the data.

The MLE of the parameters with their corresponding standard errors and the model adequacy measures for the fitted models are given in Tables 1 and 2, respectively.

The MLEs and standard errors of the NT1AP-W model along with the seven competing models (EFWE, EWW, PIW, KWP, KW, NWW, and APTW) are displayed in Table 1. Table 2 gives the model comparison result (model adequacy measures) for all models considered in this section. The new proposed model NT1AP-W, based on the five criteria, is shown to be the best-performing model among the seven competing models. This shows the new proposed model outperforms the set of similar competing models.

The mathematical expressions of the $F(z; \hat{\alpha}, \hat{\omega})$ and $S(z; \hat{\alpha}, \hat{\omega})$ of the NT1AP-W are given by the following equation:

$$F(z; \hat{\alpha}, \hat{\omega}) = 1 - 1.002^{\left(e^{-1.018z^{3.191}} - 1.018z^{3.191} - 1\right)}, z \ge 0 \text{ and}$$

$$S(z; \hat{\alpha}, \hat{\omega}) = 1.002^{\left(e^{-1.018z^{3.191}} - 1.018z^{3.191} - 1\right)}, z > 0,$$
(44)

respectively.

The quantile-quantile (Q-Q) plots of the estimated CDF and $S(z; \hat{\alpha}, \hat{\omega})$ including the total time on test (TTT) plot, see e.g., Aarset [36], for the estimated CDF, are given in Figure 3.

From Figure 3, it is observed that there is no potential influential observation and the data is linear and normal enough. In addition, the TTT plot shows the model has an increasing shape in this data set.

Furthermore, the semivariogram for the exponential and Gaussian serial correlations for the estimated CDF and $S(z; \hat{\alpha}, \hat{\omega})$ for the NT1AP-W distribution using an epileptic data set (another public health data set) is plotted as follows.

Figure 4 depicts a decreasing (in CDF) and a constant (in $(z; \hat{\alpha}, \hat{\omega})$) correlation patterns for the clustered data set. An epileptic seizure data set, which will be discussed below, owns some special features such as correlation, overdispersion, and zero-inflation. Hence, it needs further investigation using models discussed in Section 7.

TABLE 1: MLEs of the parameters and the corresponding standard errors (SE. in the parentheses) for the fitted models.

Dist.	â	\widehat{b}	ĉ	â	\widehat{eta}	$\widehat{\gamma}$	$\widehat{ heta}$	$\widehat{\lambda}$
	(SE.)	(SE.)	(SE.)	(SE.)	(SE.)	(SE.)	(SE.)	(SE.)
NT1AP-W				1.002 (0.001)	1.018 (0.013)			3.191 (0.212)
EFWE				0.258 (0.004)	0.269 (0.056)			0.028 (0.004)
EWW	1.201 (0.224)	0.384 (0.028)		8.749 (1.127)		1.523 (0.066)		0.002 (0.002)
PIW				0.658 (0.013)	1.841 (0.050)			1.837 (0.147)
KWP	59.176 (8.219)	5.452 (0.633)	1.334 (0.048)		0.064 (0.002)			52.925 (12.167)
KW	73.846 (6.725)	24.751 (1.404)				0.007 (0.009)	1.576 (0.008)	
NWW				62.866 (5.195)		0.005 (0.001)	1.731 (0.009)	
APTW				63.012 (8.126)		0.003 (0.001)	1.711 (0.013)	

TABLE 2: Model adequacy measures for the fitted models.

Dist.	AIC	CAIC	BIC	HQIC	$-2\log L$
NT1AP-W	745.11	745.08	731.52	739.85	375.56
EFWE	1306.48	1306.54	1324.60	1313.49	686.99
EWW	1365.99	1366.10	1388.64	1374.75	715.74
PIW	1515.96	1515.99	1529.55	1521.22	792.73
KWP	1986.40	1986.53	2007.14	1994.56	988.20
KW	2115.38	2115.47	2131.96	2121.91	1053.69
NWW	2740.02	2740.07	2752.46	2744.91	1367.01
APTW	3548.01	3548.07	3560.45	3552.91	1771.01

7. Combined Models for Data with Overdispersion, Correlation, and Zero-Inflation

Many authors in the literature could not model the three special features such as correlation, overdispersion, and zero-inflation in the repeatedly measured count data simultaneously in the same model, for instance, see Mekonnen et al. [16]. Another possible reason for this section is that data such as epileptic seizures are meaningless to be analyzed by using the distribution models unless they are easily able to be expressed in the exponential form (see Section 7.1) to fit the regression models. Even so, some extra parameters to handle some special characteristics in the data should be able to be imposed on the distribution model. Thus, it is a good opportunity for this study to implement the appropriate models for the data with special characteristics, as discussed in the next subsections.

7.1. Combined Models for Correlation and Overdispersion. To handle those features mentioned in the above subsection simultaneously in the same model, we need to follow some procedures as follows. Thus, let Y_{ij} be the outcome with j^{th} observation for i^{th} subject, where i = 1, 2, ..., N and $j = 1, 2, ..., n_i$ and which be with the group of n_i measurements into a vector of Y_i . Conditionally upon q-dimensional random effects $\mathbf{b_i} \sim N(\mathbf{0}, \mathbf{D})$, as an assumption, the outcomes Y_{ij} are independent with the density given by the following equation:

$$f_{ij}(\boldsymbol{y}_{ij} \mid \mathbf{b}_{\mathbf{i}}, \boldsymbol{\zeta}, \boldsymbol{\phi}) = e^{\phi^{-1} \left[\boldsymbol{y}_{ij} \lambda_{ij} - \boldsymbol{\psi} \left(\lambda_{ij} \right) \right] + c \left(\boldsymbol{y}_{ij}, \boldsymbol{\phi} \right)}, \tag{45}$$

where

$$\eta \left[\psi^{T} (\lambda_{ij}) \right] = \eta (\mu_{ij}) = \eta \left[E (Y_{ij} \mid \mathbf{b}_{i}, \boldsymbol{\zeta}) \right] = x_{ij}^{T} \boldsymbol{\zeta} + z_{ij}^{T} \mathbf{b}_{i},$$
(46)

where η (.) is a known link function, x_{ij} and z_{ij} are a p and qdimensional vectors of covariates for fixed and random effects, respectively, ζ is a p-dimensional vector of unknown regression coefficients for fixed effect, and ϕ is a scale parameter for overdispersion. Let further $f(\mathbf{b}_i | D)$ be the probability density function of the $N(\mathbf{0}, \mathbf{D})$ distributed random effects \mathbf{b}_i . Then, the Poisson-normal (PN) distribution model is given by the following equation:

$$Y_{ij} \sim Poi(\lambda_{ij}),$$
 (47)

where

$$\lambda_{ii} = e^{x_{ij}^T \boldsymbol{\zeta} + z_{ij}^T \mathbf{b}_i}.$$
(48)

It is clear that equation (47) treats the feature dependence in the data (correlation) by imposing the normal random effect in the Poisson model. According to Molenberghs et al. [11] and Molenberghs et al. [13], adding another random effect-gamma gives the following equation:

$$Y_{ij} \sim Poi(\theta_{ij}\kappa_{ij}),$$
 (49)

where κ_{ij} is given in equation (46) and $\theta_{ij} \sim \text{Gamma}(\alpha_j, \beta_j)$ in which α_j and β_j are shape and scale parameters, respectively and $\beta_j = 1/\alpha_j$. According to Molenberghs et al. [13]; the combined model which incorporates the normal and gamma random effects, **b**_i and θ_i , respectively, is expressed in the form:

$$f_i(y_{ij} | \mathbf{b}_i, \phi) = \exp\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi)\}.$$
(50)



FIGURE 3: Plot of the Q-Q and TTT for the estimated CDF and $S(z; \hat{\alpha}, \hat{\omega})$ for the NT1AP-W distribution.

The expression for its expectation is given by $E(Y_{ij} | \mathbf{b_i}) = \mu_{ij} = \theta_{ij}\kappa_{ij}$, where $\theta_{ij} \sim w_{ij}(\nu_{ij}, \sigma_{ij}^2)$ and ν_{ij} and σ_{ij}^2 are the mean and the variance of the θ_{ij} , respectively. The

likelihood contribution of the combined model for subject *i* is given by the following equation:

$$f_i(y_{ij} | \nu, \mathbf{D}, \nu_i, \Delta_i) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij} | \nu, \mathbf{b_i}, \theta_i) f_{ij}(\mathbf{b_i} | \mathbf{D}) f_{ij}(\theta_i | \nu_i, \Delta_i) d\mathbf{b_i} d\theta_i,$$
(51)

and its total likelihood is further given by the following equation:



FIGURE 4: Semivariogram for the exponential and Gaussian serial correlations for the estimated CDF and $S(z; \hat{\alpha}, \hat{\omega})$ for the NT1AP-W distribution.

$$L(\nu, \mathbf{D}, \nu, \Delta) = \prod_{i=1}^{N} f_i \Big(y_{ij} \mid \nu, \mathbf{D}, \nu_i, \Delta_i \Big)$$

=
$$\prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij} \Big(y_{ij} \mid \nu, \mathbf{b_i}, \theta_i \Big) f_{ij} \Big(\mathbf{b_i} \mid \mathbf{D} \Big) f_{ij} \Big(\theta_i \mid \nu_i, \Delta_i \Big) d\mathbf{b_i} d\theta_i.$$
 (52)

Thus, equation (49) shows the addition of gamma random effect in the PN-mixture model to handle the overdispersion feature and to form the combined model called Poisson-normal-gamma (PNG), which further handles both correlation and overdispersion simultaneously.

7.2. Combined Models for Zero-Inflation. The next concern is zero-inflation or excess zeroes which may be beyond the P model. It is assumed that there are two processes of zeroes in zero-inflated count models. The first one is where zeroes may arise from point mass (process 1) and as the second one, they may come from the conventional count component (process 2). Now assume that for measurement *i*, process 1 takes the

probability of π_i and $1 - \pi_i$ for process 2, due to Kassahun et al. [37]. We note that process 1 generates only zeroes, whereas process 2 with a designation $f_i(y_{ij} | \mathbf{b}_{1i}, \boldsymbol{\zeta}, \theta_{ij})$ generates counts from a P model, an NB model, a PN model, a generalized linear mixed model (GLMM), or a PNG combined model [37, 38]. The general form of the zeroinflated PN model in mixture is given by the following equation:

$$Y_{ij} \sim \begin{cases} 0, & \text{with } \pi_i, \\ f_i(y_{ij} \mid \mathbf{b_{li}}, \boldsymbol{\zeta}, \theta_{ij}), & \text{with } 1 - \pi_i. \end{cases}$$
(53)

Thus,

$$P(Y_{ij} = y_{ij} | \mathbf{b}_{\mathbf{1i}}, \boldsymbol{\zeta}, \theta_{ij}, \pi_{ij}) = \begin{cases} \pi_{ij} + (1 - \pi_{ij}) f_i(0 | \mathbf{b}_{\mathbf{1i}}, \boldsymbol{\zeta}, \theta_{ij}), & \text{for } y_{ij} = 0, \\ (1 - \pi_{ij}) f_i(y_{ij} | \mathbf{b}_{\mathbf{1i}}, \boldsymbol{\zeta}, \theta_{ij}), & \text{for } y_{ij} > 0. \end{cases}$$
(54)

The Bernoulli model is employed to represent the presence of additional zeros or a zero-inflated element $\pi_{ij} = \pi (x_{2ij}^T \gamma + z_{2ij}^T \mathbf{b}_{2i})$. This is done for the simplest scenario that includes just an intercept, along with the incorporated predictors $x_{2ij}^T \operatorname{and} z_{2ij}^T$ for both fixed and random effects. In addition, the model involves a vector of coefficient γ dedicated to estimate excess zero occurrences, as well as the inclusion of random effects \mathbf{b}_{1i} . For this case, the common link functions for the count and or binary outcomes such as the logit or probit can be used. Note that x_{1ij}^T , z_{1ij}^T , and \mathbf{b}_{1i} in equation (46) are now replaced by x_{1ij}^T, z_{1ij}^{T} , and \mathbf{b}_{1i} , respectively, for the nonzero count part. The predictors in the count and excess-zero component can either be overlapping, a subset of the predictors can be used for excess-zeroes, or different predictors for the two parts that can be used.

By taking the assumption that the random effects are normally distributed and possibly correlated with the correlation parameter ρ , the variance-covariance matrix **D** is given by the following equation:

$$\mathbf{D} = \begin{pmatrix} d_1 & \rho \sqrt{d_1} \sqrt{d_2} \\ \rho \sqrt{d_1} \sqrt{d_2} & d_2 \end{pmatrix}.$$
 (55)

The conditional mean and variance of the zero-inflated PNG (ZIPNG) are given as follows:

$$E(Y_{ij} \mid \mathbf{b_{li}}, \boldsymbol{\zeta}, \theta_{ij}) = \theta_{ij} \kappa_{ij} (1 - \pi_{ij}), \text{ and}$$

$$\operatorname{Var}(Y_{ij} \mid \mathbf{b_{li}}, \boldsymbol{\zeta}, \theta_{ij}) = \theta_{ij} \kappa_{ij} (1 - \pi_{ij}) \Big[1 + \theta_{ij} \kappa_{ij} \Big(\pi_{ij} + \frac{1}{\alpha} \Big) \Big],$$
(56)

respectively.

Generally, the theoretical discussion of this section helps us to understand how the count outcomes (P model) with some special features, mentioned above, are dealt with by integrating the combined models over the respective random effects to describe the three features simultaneously.

7.3. Estimation of the Model Parameters for the Combined Models. The full likelihood for PNG is given in Section 7.1. Corresponding to this contribution, the estimation is performed by integrating over the respective random effects, accumulating the marginal likelihood, and maximizing it analytically, which can also be seen from Molenberghs et al. [11], Molenberghs et al. [13], and Kassahun et al. [37]. Based on Section 7.1, the partially marginalized PNG model is expressed as follows:

$$f(y_{ij} | \mathbf{b}_{\mathbf{ii}}, \boldsymbol{\zeta}) = \int f(y_{ij} | \mathbf{b}_{\mathbf{ii}}, \boldsymbol{\zeta}, \theta_{ij}) f(\theta_{ij} | \alpha_j, \beta_j) d\theta_{ij}$$
$$= \binom{\alpha_j + y_{ij} - 1}{\alpha_j - 1} \binom{\beta_j}{1 + \kappa_{ij}\beta_j}^{y_{ij}} \binom{1}{1 + \kappa_{ij}\beta_j}^{\alpha_j} \kappa_{ij}^{y_{ij}}.$$
(57)

A similar expression for ZIPNG is given by the following equation:

$$f(y_{ij} \mid \mathbf{b_{1i}}, \mathbf{b_{2i}}, \boldsymbol{\zeta}, \boldsymbol{\gamma}) = \mathbf{I}(y_{ij} = 0)\pi_{ij} + (1 - \pi_{ij})\binom{\alpha_j + y_{ij} - 1}{\alpha_j - 1} \binom{\beta_j}{1 + \kappa_{ij}\beta_j}^{y_{ij}} \binom{1}{1 + \kappa_{ij}\beta_j}^{\alpha_j} \kappa_{ij}^{y_{ij}}.$$
(58)

The numerical estimation is performed by using a flexible normal random effects tool, the SAS procedure NLMIXED.

7.4. An Application to Epilepsy Data. Epilepsy is a noncommunicable neurological and human brain disorder. It is treated medically in a psychiatric clinic in a hospital. The data on this issue is collected from Felege Hiwot Referral Hospital (FHRH) which is located in Bahir Dar city, the capital city of Amhara regional state, Ethiopia; see Figure 5. It is 565 km far from the capital city of Ethiopia, Addis Ababa. In this clinical trial study, fifty-three subjects were followed for seven months (not equally for all) and the number of epileptic seizures was collected on a week basis. Furthermore, data on socio-economic, behavioral, and demographic information are collected to determine associated risk factors. Mekonnen et al. [16] used similar data for their analysis with all factors. However, in our study, we only focused on time-related factors for the modeling easiness.

7.4.1. Description of the Data. The numerical and graphical visualizations of the three features are presented here.

The numerical presentation in Table 3 shows 276 (30.2%) of the measurements among the total number of measurements were zero or the number of times that the patients have not shown seizures in their follow-up. This indicates that there are excess zeros in the data. And it is seen that the observed standard deviation (st. deviation, 5.5) is greater than the observed mean (3.1), which further shows the presence of overdispersion in the data. The graphical

	SN	trt	id	sex	age	race	height	weight	bserate	date0	studyweek	nseizw
1	1	0	1204	1	30	2	71	168	4.3 26/	01/1900	1	1
2	2	0	1204	1	30	2	71	168	4.3 26/	01/1900	2	0
3	3	0	1204	1	30	2	71	168	4.3 26/	01/1900	3	3
4	4	0	1204	1	30	2	71	168	4.3 26/	01/1900	4	1
5	5	0	1204	1	30	2	71	168	4.3 26/	01/1900	5	0
6	6	0	1204	1	30	2	71	168	4.3 26/	01/1900	6	1
7	7	0	1204	1	30	2	71	168	4.3 26/	01/1900	7	1
8	8	0	1204	1	30	2	71	168	4.3 26/	01/1900	8	1
9	9	0	1204	1	30	2	71	168	4.3 26/	01/1900	9	0
10	10	0	1204	1	30	2	71	168	4.3 26/	01/1900	10	1
•	•	•		•	•	•	•	•				
•	•	•	•	•	•	•	•	•			•	•
•	•	•	•	•	•	•	•	•				•
915	915	16	601714	1	58	2	74	212	42.3 02/	01/1900	16	1
						-						

FIGURE 5: Epileptic seizure patients' data.

 TABLE
 3:
 Numerical
 visualization
 of
 zero-inflation
 and
 overdispersion.

Zero-ir	iflation	Overdispersion				
#Seizures	Frequency	Percent	Mean	St. deviation		
0	276	30.2				
1	190	13.4				
2	128	9.0				
3	85	6.0				
4	49	3.5				
5	41	2.9				
•	•	•				
•	•	•				
•	•	•				
61	1	0.1				
Total	915	100.0	3.1	5.5		

illustration via histogram in Figure 6 also shows that most of the observations are right skewed to zero and it indicates that most of the observations are oriented to the left of the average of the data.

In Figure 6, the first picture supports Table 3 (illustration of zero-inflation) and the second one shows the individual profile plot or presence of correlation in the data.

7.4.2. Test Result for Overdispersion and Zero-Inflation. The presence of overdispersion was tested by using the function

dispersiontest (Epilepsy1, trafo = 1),
$$(59)$$

of the R package AER and it is distributed asymptotically normal with mean zero and variance 1. The input "Epilepsy1" of the function represents the response variable and

$$trafo = 1, \tag{60}$$

represents the linear transformation. The test statistic with the p value and the alpha

$$(z = 5.725, p - \text{value} = 5.17e - 09, \text{ and alpha} = 1.434),$$

(61)

confirms that the data are overdispersed due to the known feature in the data. This feature as shown next is due to the excess structural zeros in the data. Thus, since the p value is much smaller than the z value and the alpha is greater than 1, overdispersion in the data is confirmed.

The test for excess zeros with the test statistic distributed as Chi-square with 1 df

(Chi – square =
$$1649.742$$
, df = 1, and p – value = $2.22e - 16$),
(62)

shows that there is zero-inflation in the data. The presence of correlation is first detected from the longitudinal nature of the data in which the individuals are exposed for the repetitive measurements and second it is easily visualized from the individual profile plot in Figure 6, the second picture.

For a further analytical description of the data, let Y_{ij} stand for the number of epileptic seizures for the *i*th patient for the *j*th week of the follow-up period and let t_{ij} be the corresponding time-period for the occurrence of the Y_{ij} . For the characterization of the data based on the zero-inflation process, let us assume that the PN generates the counts with the mean λ_{ij} which is given by equation (48) as follows:

$$\ln \lambda_{ij} = (\zeta_0 + b_{1i}) + \zeta_1 t_{ij}, \tag{63}$$

when the counting process is generated from the PNG with the mean $\lambda_{ij} = \theta_{ij}\kappa_{ij}$, it takes the following form:

$$\ln \kappa_{ij} = (\zeta_0 + b_{1i}) + \zeta_1 t_{ij}.$$
 (64)

The probability of zero-inflation is given by logit (π) = $\gamma_0 + b_{2i} + \gamma_1 t_{ij}$, see Subsection 7.2 for the analytical expression.



FIGURE 6: Plot of the histogram and individual profile plot of the epileptic seizure data set.

Γ <i>Φ</i> +	Demonstern	Р	ZIP	PN
Effect	Parameter	Estimate (SE.)	Estimate (SE.)	Estimate (SE.)
Intercept	ζ_0	1.840 (0.066)*	2.179 (0.069)*	1.238 (0.282)*
Slope	ζ_1	$-0.094 \ (0.008)^*$	$-0.089 \ (0.009)^*$	$-0.072 \ (0.009)^*$
Inflation intercept	γ_0		$-1.468 (0.163)^*$	
Inflation slope	γ_1		0.059 (0.015)*	
St. dev. RE	σ			1.004 (0.100)*
Ratio slopes	ζ_1/ζ_0	7.046 (2.362)*	-14.362 (10.644)	4.996 (1.654)*
$-2\log L$		6710.6	5815.1	4058.4
AIC		6718.6	5827.1	4068.4
BIC		6737.9	5856.0	4078.8
Г <i>Ф</i> +	Demonstern	ZIPN	PG	ZIPG
Enect	Parameter	Estimate (SE.)	Estimate (SE.)	Estimate (SE.)
Intercept	ζ_0	1.447 (0.271)	1.859 (0.177)*	1.801 (0.179)*
Slope	ζ_1	$-0.076 \ (0.056)^*$	$-0.096 (0.017)^*$	$-0.086 \ (0.018)^*$
NB parameter	α_1		0.648 (0.040)*	1.479 (0.106)*
St. dev. nonzero RE	$\sqrt{d_1}$	0.944 (0.436)*		
Correlation RE	ρ	-0.055 (0.217)		
St. dev. zero part RE	$\sqrt{d_2}$	2.012 (1.321)*		
Inflation intercept	Yo	$-3.922 (0.663)^*$		-7.691 (2.138)*
Inflation slope	γ_1	0.091 (0.035)*		0.317 (0.113)*
Ratio slopes	ζ_1/ζ_0	16.795 (17.459)	7.663 (6.334)	11.501 (16.602)
Inverse NB parameter	$\beta = 1/\alpha$		1.543 (0.096)*	$0.676 \ (0.048)^*$
$-2\log L$		3742.2	4072.8	4070.8
AIC		3760.2	4082.8	4084.8
BIC		3778.9	4106.9	4118.5
Effect	Daramatar	PNG	ZIPNG	
	Parameter	Estimate (SE.)	Estimate (SE.)	

TABLE 4: Comparison of the P, ZIP, PN, ZIPN, PG, ZIP	G, PNG, and ZIPNG models.
--	---------------------------

Effect	Demonstern	Р	ZIP	PN
Ellect	Parameter	Estimate (SE.)	Estimate (SE.)	Estimate (SE.)
Intercept	ζ_0	1.096 (0.295)*	1.294 (0.287)*	
Slope	ζ_1	$-0.050 (0.013)^*$	$-0.058 (0.031)^*$	
NB parameter	α_1	2.648 (0.282)*	$0.208 \ (0.028)^*$	
St. dev. nonzero RE	$\sqrt{d_1}$		0.967 (0.441)*	
Correlation RE	ρ		0.088 (0.149)	
St. dev. zero part RE	$\sqrt{d_2}$		2.364 (1.658)*	
Inflation intercept	γ_0		-4.744(0.691)	
Inflation slope	γ_1		0.072 (0.034)*	
St. dev. RE	σ	0.991 (0.102)*		
Ratio slopes	ζ_1/ζ_0	2.016 (0.822)*	3.433 (1.650)*	
Inverse NB parameter	$\beta = 1/\alpha$	0.378 (0.377)*	4.820 (0.641)*	
-2logL		3550.2	3487.3	
AIC		3562.2	3507.3	
BIC		3574.6	3528.1	

TABLE 4: Continued.

The table summarizes the parameter estimates and standard errors (SE.) together with model comparison criteria results. The significant parameters are identified by * and the slope is for covariate time.

7.5. Comparison of Models for Overdispersion, Correlation, and Zero-Inflation. In this subsection, the known features (separately and jointly) in the repeatedly measured count data, correlation, overdispersion, and zero-inflation are discussed by using the eight models. More importantly, the ZIP, ZIPN, ZIPG, and ZIPNG models are compared by using the epileptic seizure data, and also their immediate counterparts P, PN, PG, and PNG models are used for the respective comparison. In the zero-inflation version as compared to the nonzero version models, there is a substantial improvement based on the -2logL, AIC, and BIC comparison techniques.

The ZIPG model has shown an improvement relative to the ZIP. The ZIPN is an improvement when it is compared to the ZIP and ZIPG. And the ZIPNG has further shown the best improvement relative to the ZIP, ZIPG, and ZIPN, which shows that it is the best-fitting model for the data. In the absence or when the extrazeroes in the data are ignored, the improvement in the model goes from the P to the PNG, and the PNG is the best-fitting model for the nonzeroinflated data. Generally, the trend of improvement in the models goes from the P to the ZIPNG.

The standard deviation of the random effect (St. dev. RE, represented by σ in Table 4) is significant in the models PN and PNG. This shows the normal random effect is important to add in the P model to handle the correlation in the data and it is more pronounced in the PN and PNG than in other models. The NB parameter α_1 and the inverse NB parameter (β) are significant in the models PG, ZIPG, PNG, and ZIPNG. This implies that it is customary to include the gamma random effect in the P model to deal with the overdispersion in the data and these models are capable of handling overdispersion in the data. The zero-inflation parameter (γ_i) , where i = 0, 1, is significant in the models ZIP, ZIPN, and ZIPNG, which indicates the excess zeroes in the data should not be ignored in the analysis and these models are appropriate for its analysis.

The feature correlation in the repeatedly measured data is common and needs to be handled correctly during the data analysis. As it is displayed in Table 4, it is treated by adding the normal random effect (\mathbf{b}_i) in the P model. The overdispersion is another unavoidable feature and it is dealt with by including the gamma-distributed random effect in the log-linear predictor forming the combined model (PNG) by aggregating the other induced features. Furthermore, the zero-inflation is modeled simultaneously by considering the two processes in the PNG model.

8. Discussion and Conclusion

In this study, the NT1AP family of distributions was introduced. Based on the new family, the new NT1AP-W model was discussed in detail. The need for introducing the new model is due to the reason that the classical and modified statistical models, which have been applied to health applications, do not provide the best fit when the data show nonmonotonic failure rates. The mathematical logic behind the new method for deriving new distributions is to introduce an extraparameter α which gives extraflexibility to the new family and to introduce the new family which is capable of handling different patterns in the data. A new data set (breast cancer) is considered and the proposed model is compared to the recent models.

The NT1AP-W, EFWE, EWW, PIW, KWP, KW, NWW, and APTW models were applied to the abovementioned public health data. The NT1AP-W model has shown its supremacy based on the five adequacy measures. The newly proposed family has several advantages such as (i) the addition of an extra parameter gives great flexibility, (ii) the added extra parameter makes the approach simple to modify the existing distributions, (iii) it is useful to introduce new distributions in the domain of T-X family, and (iv) it is useful to extend the existing distributions with a closed CDF. Based on the findings, the NT1AP-W model is an appropriate model for dealing with the data in health science and other related sciences. Mekonnen et al. [16] used epilepsy data by allowing both correlation and zero-inflation in the data and they analyzed the data by using the P, NB, ZIP, and ZINB models. The second part of this study is motivated by two gaps: the first one is these authors could not raise and discuss the issue of overdispersion while the data are overdispersed and the second one is the mentioned count models which alone cannot fully handle the dependence in the data. Hence, the need for an update on this gap is raised reasonably. It is also misused to analyze epileptic seizure data by using linear mixed models due to the special features present in the data.

Except Molenberghs et al. [11], Kassahun et al. [37], and Molenberghs et al. [13], none of the scholars tried to discuss the issue of the three special features in the data in the aggregate. For instance, Mekonnen et al. [16] discussed only the two features (correlation with less focus and zero-inflation). And the other scholars Hinde and Demétrio [39], Workie and Lakew [40], Dare et al. [41], and Adesina et al. [42] discussed only the count models for the nonclustered data with a lack of attention to the aggregate features. Contrary to these studies, this study has implemented and explored the appropriate models for the data with the three features simultaneously.

The slope ratio ζ_1/ζ_0 is significant in the three nonzero version models P, PN, and PNG, while it is insignificant in the zero version models except in the ZIPNG model. Other studies could show this relationship except the result of Kassahun et al. [37]. Like in the result of Kassahun et al. [37]; the correlation of the random effects (correlation RE, ρ) is unrealistically significant in none of the models. This result contradicts the longitudinal nature of the data that the response variable (number of epileptic seizures) is recorded repeatedly over time and this needs further investigation.

Based on the analysis results, it is noted that the special features such as correlation, overdispersion, and zeroinflation cannot be ignored from the repeatedly measured count data and need to be modeled simultaneously by using the appropriate combined models.

Appendix

Sample Data

The sample of two real datasets (breast cancer patients' data and epileptic seizure patients' data) is displayed under these subsections, respectively.

Data Availability

Due to third-party usage, some ethical issues, and subjects being human beings, the data can obtained from the corresponding author upon request.

Ethical Approval

Epileptic seizure data are obtained from Mekonnen et al. [16] formally, where the ethical issues are kept standard.

Breast cancer data are collected by the researchers by keeping all ethical standard issues including obtaining ethical permission letters and consent of the subjects.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Getachew Tekle, Rasool Roozegar, and Zubair Ahmed conceptualized the study. Getachew Tekle and Zubair Ahmed gathered the data. Getachew Tekle, Rasool Roozegar, and Zubair Ahmed designed the methodology. Getachew Tekle and Rasool Roozegar brought the resources. Rasool Roozegar and Zubair Ahmed supervised the study. Getachew Tekle and Rasool Roozegar analyzed the data. Getachew Tekle and Rasool Roozegar wrote the original draft.

Acknowledgments

This study was supported by the Yazd University, Iran.

References

- A. S. Malehi, F. Pourmotahari, and K. A. Angali, "Statistical models for the analysis of skewed healthcare cost data: a simulation study," *Health economics review*, vol. 5, no. 1, pp. 11–16, 2015.
- [2] R. Roozegar and S. Nadarajah, "The quadratic hazard rate power series distribution," *Journal of Testing and Evaluation*, vol. 45, no. 3, pp. 20150506–20151072, 2016.
- [3] R. Roozegar and S. Nadarajah, "The power series skew normal class of distributions," *Communications in Statistics-Theory* and Methods, vol. 46, no. 22, pp. 11404–11423, 2017.
- [4] C. Chesneau, K. Karakaya, H. S. Bakouch, and C. Kuş, "An alternative to the marshall-olkin family of distributions: bootstrap, regression and applications," *Communications on Applied Mathematics and Computation*, vol. 4, no. 4, pp. 1229–1257, 2022.
- [5] I. Elbatal, M. Elgarhy, and B. M. Golam Kibria, "Alpha power transformed weibull-G family of distributions: theory and applications," *Journal of Statistical Theory and Applications*, vol. 20, no. 2, pp. 340–354, 2021.
- [6] W. A. Shehata, H. Yousof, and M. Aboraya, "A novel generator of continuous probability distributions for the asymmetric left-skewed bimodal real-life data with properties and copulas," *Pakistan Journal of Statistics and Operation Research*, vol. 17, pp. 943–961, 2021.
- [7] C. Chesneau, H. S. Bakouch, and T. Hussain, "A new class of probability distributions via cosine and sine functions with applications," *Communications in Statistics-Simulation and Computation*, vol. 48, no. 8, pp. 2287–2300, 2019.
- [8] Z. Ahmad, G. G. Hamedani, and N. S. Butt, "Recent developments in distribution theory: a brief survey and some new generalized classes of distributions," *Pakistan Journal of Statistics and Operation Research*, vol. 15, no. 1, pp. 87–110, 2019.

- [9] M. Hussein, H. Elsayed, and G. M. Cordeiro, "A new family of continuous distributions: properties and estimation," *Symmetry*, vol. 14, no. 2, p. 276, 2022.
- [10] E. S. A. El-Sherpieny, S. Assar, and T. Helal, "A new generating family of distributions: properties and applications to the Weibull exponential model," *Journal of Modern Applied Statistical Methods*, vol. 19, no. 1, p. 20, 2021.
- [11] G. Molenberghs, G. Verbeke, and C. G. Demétrio, "An extended random-effects approach to modeling repeated, overdispersed count data," *Lifetime Data Analysis*, vol. 13, no. 4, pp. 513–531, 2007.
- [12] W. Kassahun, T. Neyens, C. Faes, G. Molenberghs, and G. Verbeke, "A zero-inflated overdispersed hierarchical Poisson model," *Statistical Modelling*, vol. 14, no. 5, pp. 439–456, 2014.
- [13] G. Molenberghs, G. Verbeke, C. G. Demétrio, and A. M. Vieira, "A family of generalized linear models for repeated measures with normal and conjugate random effects," *Statistical Science*, vol. 25, no. 3, pp. 325–347, 2010.
- [14] R. Roozegar, G. Tekle, and G. Hamedani, "A new generalized-X family of distributions: applications, characterization and a mixture of random effect models," *Pakistan Journal of Statistics and Operation Research*, vol. 18, no. 2, pp. 483–504, 2022.
- [15] G. Tekle, R. Roozegar, and H. Bidram, "A new alpha power type-1 family of distributions and modeling the overdispersed count outcome," *Journal of Statistical Modeling: Theory and Applications*, vol. 3, no. 1, pp. 83–101, 2023.
- [16] F. H. Mekonnen, W. D. Lakew, Z. D. Tesfaye, and P. K. Swain, "Statistical models for longitudinal zero-inflated count data: application to seizure attacks," *African Health Sciences*, vol. 19, no. 3, pp. 2555–2564, 2019.
- [17] A. Alzaatreh, C. Lee, and F. Famoye, "A new method for generating families of continuous distributions," *Metron*, vol. 71, no. 1, pp. 63–79, 2013.
- [18] Z. Ahmad, E. Mahmoudi, S. Dey, and S. K. Khosa, "Modeling vehicle insurance loss data using a new member of TX family of distributions," *Journal of Statistical Theory and Applications*, vol. 19, no. 2, pp. 133–147, 2020.
- [19] Q. Liao, Z. Ahmad, E. Mahmoudi, and G. G. Hamedani, "A new flexible bathtub-shaped modification of the Weibull model: properties and applications," *Mathematical Problems in Engineering*, vol. 2020, Article ID 3206257, 11 pages, 2020.
- [20] S. Mehboob Zaidi, M. M Al Sobhi, M. El-Morshedy, and A. Z Afify, "A new generalized family of distributions: properties and applications," *AIMS Mathematics*, vol. 6, no. 1, pp. 456–476, 2021.
- [21] A. Z. Afify, G. M. Cordeiro, N. A. Ibrahim, F. Jamal, M. Elgarhy, and M. A. Nasir, "The Marshall-Olkin odd Burr III-G family: theory, estimation, and engineering applications," *IEEE Access*, vol. 9, pp. 4376–4387, 2021.
- [22] H. Reyad, M. Ç. Korkmaz, A. Z. Afify, G. G. Hamedani, and S. Othman, "The Fréchet Topp Leone-G family of distributions: properties, characterizations, and applications," *Annals* of *Data Science*, vol. 8, no. 2, pp. 345–366, 2019.
- [23] A. A. Al-Babtain, R. A. Sherwani, A. Z. Afify et al., "The extended Burr-R class: properties, applications, and modified test for censored data," *AIMS Mathematics*, vol. 6, no. 3, pp. 2912–2931, 2021.
- [24] B. Oluyede, S. Pu, B. Makubate, and Y. Qiu, "The gamma-Weibull-G Family of distributions with applications," *Austrian Journal of Statistics*, vol. 47, no. 1, pp. 45–76, 2018.

- [25] B. S. El-Desouky, A. Mustafa, and S. Al-Garash, "The exponential flexible Weibull extension distribution," 2016, https:// arxiv.org/abs/1605.08152.
- [26] R. K. Joshi and V. Kumar, "Poisson Inverse Weibull distribution with theory and applications," *International Journal of Statistics and Systems*, vol. 16, no. 1, pp. 1–16, 2021.
- [27] A. S. Hassan and M. Elgarhy, "A new family of exponentiated Weibull-generated distributions," *International Journal of Machine Intelligence and Applications*, vol. 4, no. 1–D, pp. 135–148, 2016.
- [28] I. Elbatal, Z. Ahmad, M. Elgarhy, and A. M. Almarashi, "A new alpha power transformed family of distributions: properties and applications to the Weibull model," *The Journal of Nonlinear Science and Applications*, vol. 12, no. 1, pp. 1–20, 2018.
- [29] P. R. D. Marinho, R. B. Silva, M. Bourguignon, G. M. Cordeiro, and S. Nadarajah, "Adequacy Model: An R package for probability distributions and general purpose optimization," *PLoS ONE*, vol. 14, no. 8, Article ID e0221487, 2019.
- [30] G. M. Cordeiro, E. M. Ortega, and S. Nadarajah, "The Kumaraswamy Weibull distribution with application to failure data," *Journal of the Franklin Institute*, vol. 347, no. 8, pp. 1399–1429, 2010.
- [31] E. S. A. Elsherpieny, Y. Y. Abdelall, and A. A. Mohamed, "On the exponentiated new weighted Weibull distribution," *International Journal of Engineering and Applied Sciences*, vol. 4, no. 10, 2017.
- [32] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [33] H. Bozdogan, "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.
- [34] G. Schwarz, "Estimating the dimension of a model," Annals of Statistics, vol. 6, no. 2, pp. 461–464, 1978.
- [35] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society: Series B*, vol. 41, no. 2, pp. 190–195, 1979.
- [36] M. V. Aarset, "How to identify a bathtub hazard rate," *IEEE Transactions on Reliability*, vol. 36, no. 1, pp. 106–108, 1987.
- [37] W. Kassahun, T. Neyens, G. Molenberghs, C. Faes, and G. Verbeke, "Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros," *Statistics in Medicine*, vol. 33, no. 25, pp. 4402–4419, 2014.
- [38] T. W. Rakhmawati, G. Molenberghs, G. Verbeke, and C. Faes, "Local influence diagnostics for hierarchical count data models with overdispersion and excess zeros," *Biometrical Journal*, vol. 58, no. 6, pp. 1390–1408, 2016.
- [39] J. Hinde and C. G. Demétrio, "Overdispersion: models and estimation," *Computational Statistics & Data Analysis*, vol. 27, no. 2, pp. 151–170, 1998.
- [40] M. S. Workie and A. M. Lakew, "Bayesian count regression analysis for determinants of antenatal care service visits among pregnant women in Amhara regional state, Ethiopia," *Journal of Big Data*, vol. 5, no. 1, pp. 7–23, 2018.
- [41] R. J. Dare, O. S. Adesina, P. E. Oguntunde, and O. O. Agboola, "Adaptive regression model for highly skewed count data," *International Journal of Mechanical Engineering & Technol*ogy, vol. 10, no. 1, pp. 1964–2197, 2019.
- [42] O. S. Adesina, D. A. Agunbiade, and P. E. Oguntunde, "Flexible Bayesian Dirichlet mixtures of generalized linear mixed models for count data," *Scientific African*, vol. 13, no. 1, Article ID e00963, 2021.