A New Lifetime Distribution: Exponentiated Exponential-Pareto-Half Normal Mixture Model for Biomedical Applications

Oriyomi Ahmad Hassan[1], Aisha Tunrayo Maradesa[2], Abdulazeez Toyosi Alabi[3], Akinwale Victor Famotire[4*], Oyejide Surajudeen Salam[5], Ajani Busari[6], Habeeb Abiodun Afolabi[7], Solomon Adeleke[8] and Abayomi Ayodele Akomolafe[6]

1. Department of Statistics, Federal University of Oye Ekiti, Ekiti, Nigeria.
2. Department of Statistics, Yaba College of Technology, Yaba, Lagos, Nigeria.
3. Department of Statistics, Georgia State University, USA.
4. Department of Public Health Sciences, Medical University of South Carolina, South Carolina USA.
5. Department of Statistics, Lead City University Jericho Ibadan, Oyo State, Nigeria.
6. Department of Statistics, Federal University of Technology Akure, Ondo State, Akure Nigeria.
7. Department of Statistics, Osun State University, Osogbo, Nigeria.
8. Department of Statistics, Kwara State University Ilorin, Kwara State, Nigeria.

* E-mail of the corresponding author: famotire@musc.edu

Abstract

This study introduces the Exponentiated-Exponential-Pareto-Half Normal Mixture Distribution (EEPHND), a novel hybrid model developed to overcome the limitations of classical distributions in modeling complex real-world data. By compounding the Exponentiated-Exponential-Pareto (EEP) and Half-Normal distributions through a mixture mechanism, EEPHND effectively captures both early-time symmetry and long-tail behavior, features which are commonly observed in survival and reliability data. The model offers closed-form expressions for its probability density, cumulative distribution, survival and hazard functions, moments, and reliability metrics, ensuring analytical tractability and interpretability in the presence of censoring and heterogeneous risk dynamics. When applied to a real-world lung cancer dataset, EEPHND outperformed competing models in both goodness-of-fit and predictive accuracy, achieving a Concordance Index (CI) of 0.9997. These results highlight its potential as a flexible and powerful tool for survival analysis, and biomedical engineering.

Keywords: Lifetime Distribution, Biomedical Applications

1. Introduction

Parametric modeling plays a fundamental role in data analysis (Akinsete et al. 2008; Ashour and Eltehiwy 2013; Babatunde and Adeleke 2020), particularly in inferential statistics (Chhetri et al. 2017). Crucial to this approach is the assumption of an underlying probability distribution (Bourguignon et al. 2016; Ashour and Eltehiwy 2013; Eugene et al. 2002). Traditional distributions, such as the exponential, normal, Gamma, and Rayleigh, have been widely used for this purpose due to their simplicity and interpretability (Shaw and Buckley 2009). However, the complexity of real-world phenomena, resulting from advancements in data acquisition technologies, increasingly leads to datasets that exhibit non-standard features (Gnedenko et al. 1999; Burr 1942). These include heavy tails, multi-modality, strong asymmetry, and extreme skewness, characteristics that established parametric models are often ill-equipped to capture (Akomolafe and Maradesa 2017; Akinsete et al. 2008; Ashour and Eltehiwy 2013). As a result, applying these distributions to such data frequently yields poor model fit and unreliable inferences (Merovci and Puka 2014; Adeleke et al. 2019; Akomolafe and Maradesa 2019; Chhetri et al. 2017; Shittu and Adepoju 2013). To address these limitations, researchers have proposed generalized distributions by introducing additional shape or scale parameters to enhance flexibility while maintaining mathematical tractability. These include the Exponentiated Exponential distribution (Nadarajah 2011), Exponentiated Weibull distribution (Pal et al. 2006), Beta-halfNormal (Akomolafe and Maradesa 2017), Beta-Pareto (Akinsete et al. 2008), Weighted HalfNormal (Babatunde and Adeleke 2020), Expnentiatiated Exponential Pareto (Adeleke et al. 2019) among others. These extended models offer improved control over statistical properties such as tail behavior, skewness, and kurtosis, making them more suitable for modeling complex data (Adeleke 2020). Nonetheless, several challenges remain, particularly in survival and lifetime data modeling (Ashour and Eltehiwy 2013). While many classical and generalized distributions have been proposed for general-purpose applications, few are explicitly tailored for survival analysis (Chhetri et al. 2017). Effective modeling in this context requires the following.

- A well-defined survival function $S(t) = P(T > t)$
- A flexible hazard rate (e.g., increasing, decreasing, constant, or bathtub-shaped)
- Interpretability for phenomena such as censoring, early/late failure, or heterogeneous risk dynamics

Many existing models lack closed-form representations (Arshad et al. 2020), limiting their practical use in survival analysis and reducing insight into early, late, or mixed-risk failure patterns (key considerations in biomedical research) (Huang 2023; Gnedenko et al. 1999). Moreover, many generalized models struggle to capture dual-behavior datasets, such as an early peak followed by a heavy tail, which are common in biological measurements and failure time distributions (Tang and Su 2008; Adeleke 2020). These cases require models with greater flexibility to represent multiple behavioral patterns within a unified framework (Akomolafe and Maradesa 2017; Akomolafe et al. 2019).

To bridge this gap, we develop the Exponentiated-Exponential-Pareto-Half Normal Mixture Distribution (EEPHND), a new hybrid distribution designed explicitly for survival and lifetime data. The EEPHND model combines the Exponentiated-Exponential-Pareto Distribution (EEPD) (Adeleke et al. 2019) distribution with the Half-Normal distribution via a mixture mechanism (Adeleke 2020), resulting in a flexible yet interpretable model. The EEPD component captures long-tail behavior, which are common in late-onset or accumulated risk scenarios. On the other hand, the Half-Normal component models symmetric, short-term variations often associated with early failures, minimal degradation processes, or measurement-related fluctuations. This dual-structure design enables the EEPHND to accommodate datasets that exhibit both early-time symmetry and long-tail heterogeneity, an ability rarely achieved by existing models. Moreover, the model provides a closed-form survival function and a tractable hazard rate, supporting meaningful interpretations for censored observations, early/late failure dynamics, and mixed-risk populations.

To evaluate the model's statistical performance, we applied EEPHND to real-life survival data and benchmarked it against EEPD, Log-Normal, and Gamma-Rayleigh. Our model not only achieved one of the best fits in terms of AIC, BIC, and CAIC but also outperformed standard models in terms of predictive accuracy, as measured by the Concordance Index (CI) (Brentnall and Cuzick 2018). Notably, the EEPHND model produced a CI of 0.9997, significantly higher than that of the Cox Proportional Hazards model (0.6029), and even marginally better than the non-parametric Kaplan-Meier estimator (0.9982). These results highlight the model's capability in both descriptive and predictive modeling of complex survival data. Overall, the proposed EEPHND model fills a critical gap in survival and lifetime data modeling by offering a flexible, interpretable, and high-performing alternative to existing parametric and semi-parametric models. Its versatility holds promise for a wide range of applications in biomedical research (Elbatal et al. 2022), reliability engineering (Shahriari et al. 2024), epidemiology (Hybels and Blazer 2003), and beyond (Adegoke Afeez et al. 2019; Adegoke et al. 2021; A. B. Adegoke et al. 2020; Olatunji et al. 2021; A. Adegoke et al. 2020).

2. Methods

2.1 Development of EEPHND Model

The Probability Density Function (PDF) of the EEPHND can be expressed as a mixture of two component distributions:

$$f_{\text{EEPHND}}(x;\boldsymbol{\tau}) = p_1 f_1(x;\boldsymbol{\tau}_1) + p_2 f_2(x;\boldsymbol{\tau}_2) \quad (1)$$

where $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are the parameter vectors of the parent distributions, p_1 and p_2 are the mixing proportions satisfying $p_1 + p_2 = 1$, with $p_1, p_2 > 0$. The PDFs of the EEPD (Adeleke et al. 2019) and Half-Normal (HN) distributions are given respectively by:

$$f_{EEPHND}(x;\alpha,\beta,\theta,\lambda) = \frac{\alpha\lambda\theta}{\beta}\left(\frac{x}{\beta}\right)^{\theta-1} e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\left[1-\left(1-e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\right)^{\alpha-1}\right] \qquad (1a)$$

$$f_{HN}(x;\sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}} \qquad (1b)$$

To construct the EEPHND as a mixture of the EEP and HN distributions, we substitute equation (1a) and (1b) in (1) so that

$$f_{\text{EEPHND}}(x;\alpha,\beta,\theta,\lambda,\sigma) = p_1 \cdot \frac{\alpha\lambda\theta}{\beta}\left(\frac{x}{\beta}\right)^{\theta-1} e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\left[1-\left(1-e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\right)^{\alpha-1}\right] + p_2 \cdot \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

where $\alpha, \beta, \theta, \lambda, \sigma > 0$, $p_1 + p_2 = 1$, and $p_1, p_2 > 0$. Then, the Cumulative Distribution Function (CDF) of the EEPHND is given as

$$F_{\text{EEPHND}}(x;\alpha,\beta,\theta,\lambda,\sigma) = p_1\left[1-\left(1-e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\right)^{\alpha}\right] + p_2 \cdot \text{erf}\left(\frac{x}{\sigma\sqrt{2}}\right)$$

where $\text{erf}(\cdot)$ is the error function.

2.2 Derivation of Properties of EEPHND Model

2.2.1 Moment

Let $X \sim \text{EEPHND}(\alpha,\beta,\theta,\lambda,\sigma)$ be a random variable following the EEPHND, where α, β, θ, and λ are the shape and scale parameters for the EEPD component, and σ is the scale parameter for the HN component. Then,

$$\begin{aligned} E(X^r) &= \int_0^\infty x^r f_{\text{EEPHND}}(x;\alpha,\beta,\theta,\lambda,\sigma)\,dx \\ &= p_1\int_0^\infty x^r f_1(x;\alpha,\beta,\theta,\lambda)\,dx + p_2\int_0^\infty x^r f_2(x;\sigma)\,dx \end{aligned}$$

Let $x = \beta\left(\frac{u}{\lambda}\right)^{1/\theta}$ for the first integral, and $y = \frac{x^2}{2\sigma^2}$ for the second. Applying the respective substitutions and simplifications (Adeleke et al. 2019; Adeleke 2020; Babatunde and Adeleke 2020):

$$E(X^r) = p_1 \alpha \left(\frac{\beta}{\lambda^{1/\theta}}\right)^r \int_0^\infty u^{r/\theta} e^{-\alpha u}\, du + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_0^\infty (\sigma\sqrt{2y})^{r-1} e^{-y} \cdot \frac{\sigma^2}{\sigma\sqrt{2y}}\, dy$$

$$E(X^r) = p_1 \alpha \left(\frac{\beta}{\lambda^{1/\theta}}\right)^r \Gamma\left(\frac{r}{\theta}+1\right) \cdot \alpha^{-\left(\frac{r}{\theta}+1\right)} + p_2 \frac{(\sigma\sqrt{2})^r}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right)$$

$$E(X^r) = p_1 \left(\frac{\beta}{(\alpha\lambda)^{1/\theta}}\right)^r \Gamma\left(\frac{r}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^r}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right)$$

$$E(X) = p_1 \left(\frac{\beta}{(\alpha\lambda)^{1/\theta}}\right) \Gamma\left(\frac{1}{\theta}+1\right) + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$$

$$E(X^2) = p_1 \left(\frac{\beta}{(\alpha\lambda)^{1/\theta}}\right)^2 \Gamma\left(\frac{2}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^2}{2}$$

$$E(X^3) = p_1 \left(\frac{\beta}{(\alpha\lambda)^{1/\theta}}\right)^3 \Gamma\left(\frac{3}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^3}{\sqrt{\pi}}$$

$$E(X^4) = p_1 \left(\frac{\beta}{(\alpha\lambda)^{1/\theta}}\right)^4 \Gamma\left(\frac{4}{\theta}+1\right) + 3p_2 \left(\frac{\sigma\sqrt{2}}{2}\right)^4$$

The variance of the EEPHND is then given by:

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= p_1 \left(\frac{\beta}{(\alpha\lambda)^{\frac{1}{\theta}}}\right)^2 \Gamma\left(\frac{2}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^2}{2} \\ &\quad - \left[p_1 \left(\frac{\beta}{(\alpha\lambda)^{\frac{1}{\theta}}}\right) \Gamma\left(\frac{1}{\theta}+1\right) + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right]^2 \end{aligned}$$

2.3 Skewness and Kurtosis

The coefficient of skewness, γ_1, measures the asymmetry of the EEPHND, indicates whether the distribution is positively or negatively skewed and provides insight into its asymptotic behavior. The kurtosis, γ_2, quantifies the heaviness of the distribution's tails (Adeleke 2020). For the EEPHND, $\gamma_2 > 3$ suggests heavier tails than the normal distribution (leptokurtic), $\gamma_2 < 3$ indicates lighter tails (platykurtic), and corresponds to a normal-like (mesokurtic) behavior (Adeleke 2020).

$$\begin{aligned} \mu_3 &= \mathbb{E}[(X-\mu)^3] \\ &= \mathbb{E}[(-\mu)^3 + 3X(-\mu)^2 + 3X^2(-\mu) + X^3] \\ &= \mathbb{E}[X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3] \\ &= \mathbb{E}[X^3] - 3\mu\mathbb{E}[X^2] + 3\mu^2\mathbb{E}[X] - \mu^3 \end{aligned}$$

Substitute the known raw moments:

$$\mu_3 = p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^3 \Gamma\left(\frac{3}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^3}{\sqrt{\pi}}$$
$$-3\mu\left[p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^2 \Gamma\left(\frac{2}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^2}{2}\right]$$
$$+3\mu^2\left[p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right) \Gamma\left(\frac{1}{\theta}+1\right) + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right] - \mu^3$$

Let

$$\mu = p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right) \Gamma\left(\frac{1}{\theta}+1\right) + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$$

Then:

$$\mu_3 = \mathbb{E}[X^3] - 3\mu\mathbb{E}[X^2] + 3\mu^2\mathbb{E}[X] - \mu^3$$

2.4 Fourth Central Moment

$$\begin{aligned}\mu_4 &= \mathbb{E}[(X-\mu)^4] \\ &= \mathbb{E}[(-\mu)^4 + 4X(-\mu)^3 + 6X^2(-\mu)^2 + 4X^3(-\mu) + X^4] \\ &= \mathbb{E}[X^4 - 4\mu X^3 + 6\mu^2 X^2 - 4\mu^3 X + \mu^4] \\ &= \mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\mathbb{E}[X^2] - 4\mu^3\mathbb{E}[X] + \mu^4\end{aligned}$$

Substitute the known raw moments:

$$\mu_4 = p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^4 \Gamma\left(\frac{4}{\theta}+1\right) + p_2 \frac{3(\sigma\sqrt{2})^4}{4}$$
$$-4\mu\left[p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^3 \Gamma\left(\frac{3}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^3}{\sqrt{\pi}}\right]$$
$$+6\mu^2\left[p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^2 \Gamma\left(\frac{2}{\theta}+1\right) + p_2 \frac{(\sigma\sqrt{2})^2}{2}\right]$$
$$-4\mu^3\left[p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right) \Gamma\left(\frac{1}{\theta}+1\right) + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right] + \mu^4$$

Let again:

$$\mu = p_1 \left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right) \Gamma\left(\frac{1}{\theta}+1\right) + p_2 \frac{\sigma\sqrt{2}}{\sqrt{\pi}}$$

Thus, the expression simplifies to:

$$\mu_4 = \mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\mathbb{E}[X^2] - 4\mu^3\mathbb{E}[X] + \mu^4$$

2.4.1 Skewness

$$\begin{aligned}\gamma_1 = &\left[p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^3\Gamma\left(\frac{3}{\theta}+1\right)+p_2\frac{(\sigma\sqrt{2})^3}{\sqrt{\pi}}\right.\\ &-3\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)\\ &\times\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^2\Gamma\left(\frac{2}{\theta}+1\right)+p_2\frac{(\sigma\sqrt{2})^2}{2}\right)\\ &\left.+2\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)^3\right]^2\\ &/\left[p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^2\Gamma\left(\frac{2}{\theta}+1\right)+p_2\frac{(\sigma\sqrt{2})^2}{2}\right.\\ &\left.-\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)^2\right]^3\end{aligned}$$

2.4.2 Kurtosis

$$\begin{aligned}\gamma_2 = &[p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^4\Gamma\left(\frac{4}{\theta}+1\right)+p_2\frac{3(\sigma\sqrt{2})^4}{4}\\ &-4\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)\\ &\times\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^3\Gamma\left(\frac{3}{\theta}+1\right)+p_2\frac{(\sigma\sqrt{2})^3}{\sqrt{\pi}}\right)\\ &+6\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)^2\\ &\times\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^2\Gamma\left(\frac{2}{\theta}+1\right)+p_2\frac{(\sigma\sqrt{2})^2}{2}\right)\\ &-3\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)^4]\\ &/\left[p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)^2\Gamma\left(\frac{2}{\theta}+1\right)+p_2\frac{(\sigma\sqrt{2})^2}{2}-\left(p_1\left(\frac{\beta}{\sqrt{\theta\alpha\lambda}}\right)\Gamma\left(\frac{1}{\theta}+1\right)+p_2\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\right)^2\right]^2\end{aligned}$$

2.5 Moment Generating Function

Let $X\sim \text{EEPHND}(\alpha,\beta,\theta,\lambda,\sigma)$. The moment generating function of X, denoted by $M_X(t)$, is defined as (Tallis 1961; Adeleke et al. 2019):

$$M_X(t)=\mathbb{E}[e^{tX}]=\int_0^{\infty}e^{tx}f(x;\alpha,\beta,\theta,\lambda,\sigma)\,dx,$$

where $f(x;\cdot)$ is the probability density function of the EEPHND. Expanding e^{tx} as a power series and interchanging summation and integration provides the following:

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mathbb{E}[X^r],$$

which expresses the moment generating function in terms of the raw moments of the distribution as

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^{\infty} e^{tx} f(x; \alpha, \beta, \theta, \lambda, \sigma)\, dx$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^{\infty} x^r f(x; \alpha, \beta, \theta, \lambda, \sigma)\, dx$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \left[p_1 \left(\frac{\beta}{\sqrt{\theta \alpha \lambda}} \right)^r \Gamma\left(\frac{r}{\theta} + 1 \right) + p_2 \frac{(\sigma\sqrt{2})^r}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2} \right) \right], \quad r = 1,2,3 \ldots$$

2.6 Reliability

Let $X \sim \text{EEPHND}(\alpha, \beta, \theta, \lambda, \sigma)$. The reliability function, also known as the survival function, is defined as the probability that the random variable X exceeds a given value x:

$$R(x) = 1 - F_{\text{EEPHND}}(x; \alpha, \beta, \theta, \lambda, \sigma)$$
$$= 1 - \left[p_1 \left(1 - \left(1 - \left(1 - e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}} \right)^{\alpha} \right) \right) + p_2 \,\text{erf}\left(\frac{x}{\sigma\sqrt{2}} \right) \right]$$

2.7 Hazard Function

Let $X \sim \text{EEPHND}(\alpha, \beta, \theta, \lambda, \sigma)$, the hazard function, also called failure rate function, describes the instantaneous rate at which an event occurs, given that it has not occurred before time x:

$$H(x) = \frac{f_{\text{EEPHND}}(x; \alpha, \beta, \theta, \lambda, \sigma)}{R(x)}$$
$$= \frac{p_1 \cdot \frac{\alpha\lambda\theta}{\beta} \left(\frac{x}{\beta} \right)^{\theta-1} e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}} \left[1 - \left(1 - e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}} \right) \right]^{\alpha-1} + p_2 \cdot \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}}}{1 - \left[p_1 \left(1 - \left(1 - \left(1 - e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}} \right) \right)^{\alpha} \right) + p_2 \,\text{erf}\left(\frac{x}{\sigma\sqrt{2}} \right) \right]}$$

2.8 Odd Function

Let $X \sim \text{EEPHND}(\alpha, \beta, \theta, \lambda, \sigma)$, then the odds function depicting the ratio of the cumulative probability that the event occured at time x to the probability that it has not occurred by time x, is given as

$$O(x) = \frac{F_{\text{EEPHND}}(x; \alpha, \beta, \theta, \lambda, \sigma)}{R(x)}$$
$$= \frac{p_1\left[1-\left(1-e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\right)^{\alpha}\right] + p_2 \operatorname{erf}\left(\frac{x}{\sigma\sqrt{2}}\right)}{1-\left[p_1\left[1-\left(1-e^{-\lambda\left(\frac{x}{\beta}\right)^{\theta}}\right)^{\alpha}\right] + p_2 \operatorname{erf}\left(\frac{x}{\sigma\sqrt{2}}\right)\right]}$$

2.9 Sampling from EEPHND

Let $X \sim \text{EEPHND}(\alpha, \beta, \theta, \lambda, \sigma, p_1)$, defined as a finite mixture of two continuous distributions: the EEP and HN, such that

$$f_X(x) = p_1 f_{\text{EEP}}(x) + (1-p_1) f_{\text{HN}}(x).$$

To generate a random variate X from the EEPHND, one may employ the following mixture-based sampling strategy:

- Let $U \sim \text{Uniform}(0,1)$.
- If $U < p_1$, draw X from the EEP component via inverse transform sampling:
  - Generate $V \sim \text{Uniform}(0,1)$,
  - Then compute:

$$X = \beta\left[-\frac{1}{\lambda}\ln\left(1-(1-V)^{1/\alpha}\right)\right]^{1/\theta}.$$

- Draw $X \sim \text{HN}(\sigma)$, i.e., from the HN distribution with scale parameter σ. This can be achieved by:

$$X = |Z|, \quad Z \sim \mathcal{N}(0, \sigma^2),$$

This sampling strategy preserves the probabilistic structure of the EEPHND by ensuring that the generated samples reflect its dual nature, *i.e.*, capturing heavy-tailed behavior through the EEP component (Adeleke et al. 2019) and light-tailed symmetry via the Half-Normal component. This enables EEPHND particularly suitable for simulating complex, heterogeneous lifetime and reliability data. Here, the sample size n was selected to achieve the desired statistical properties; values of n = 1000 or higher are recommended for simulation studies. Inverse transform sampling ensures fidelity to the cumulative distribution of the EEP component, while the Half-Normal component is sampled by taking the absolute value of a normally distributed variate. The mixture structure is preserved by drawing each sample from the EEP or Half-Normal component with probabilities p_1 and $1-p_1$, respectively.

2.10 Maximum Likelihood Estimation

Let the observed data be $x_1, x_2, \ldots, x_n$, and define the likelihood function:

$$L_f(x;\alpha,\beta,\theta,\lambda,\sigma) = \prod_{i=1}^{n}[p_1 \cdot \frac{\alpha\lambda\theta}{\beta}\left(\frac{x_i}{\beta}\right)^{\theta-1} e^{-\lambda\left(\frac{x_i}{\beta}\right)^{\theta}}\left(1-\left(1-e^{-\lambda\left(\frac{x_i}{\beta}\right)^{\theta}}\right)\right)^{\alpha-1} + p_2 \cdot \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x_i^2}{2\sigma^2}}]$$

Taking the natural logarithm, the log-likelihood function becomes:

$$\ln L_f(x;\alpha,\beta,\theta,\lambda,\sigma) = \sum_{i=1}^{n} \ln\left[p_1 \cdot \frac{\alpha\lambda\theta}{\beta}\left(\frac{x_i}{\beta}\right)^{\theta-1} e^{-\lambda\left(\frac{x_i}{\beta}\right)^{\theta}}\left(1-\left(1-e^{-\lambda\left(\frac{x_i}{\beta}\right)^{\theta}}\right)\right)^{\alpha-1} + p_2 \cdot \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x_i^2}{2\sigma^2}}\right]$$

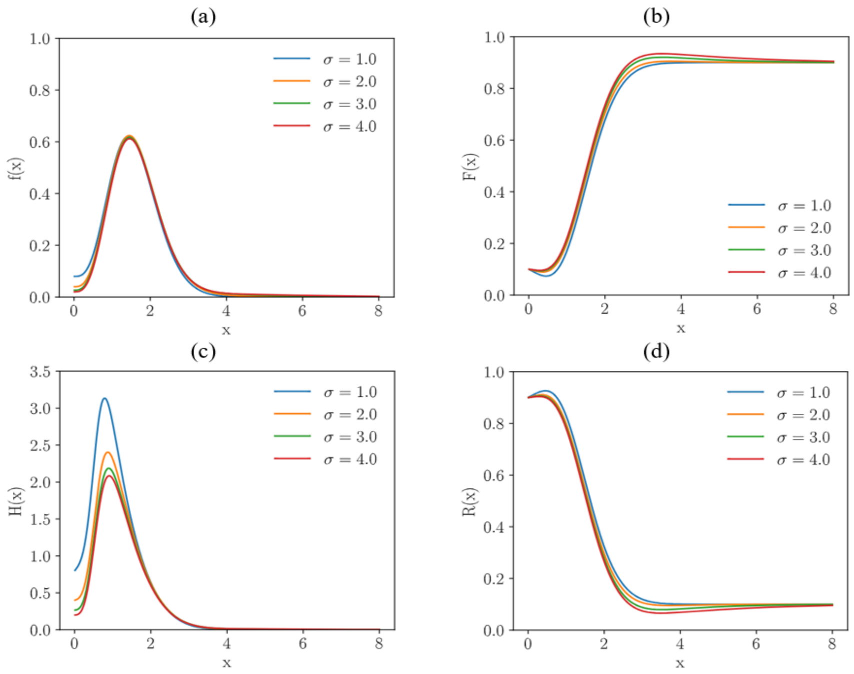
To estimate the parameters $\alpha, \beta, \theta, \lambda, \sigma$, we take the partial derivatives of the log-likelihood function with respect to each parameter and solve the resulting score equations. These equations generally do not have closed-form solutions, and numerical methods such as the Newton-Raphson algorithm (Jennrich and Robinson 1969) are employed to obtain the maximum likelihood estimates, which were used to test hypotheses regarding parameter values, assess the mixture model's consistency and efficiency compared to its component distributions, and construct confidence intervals or perform model selection.

3. Results

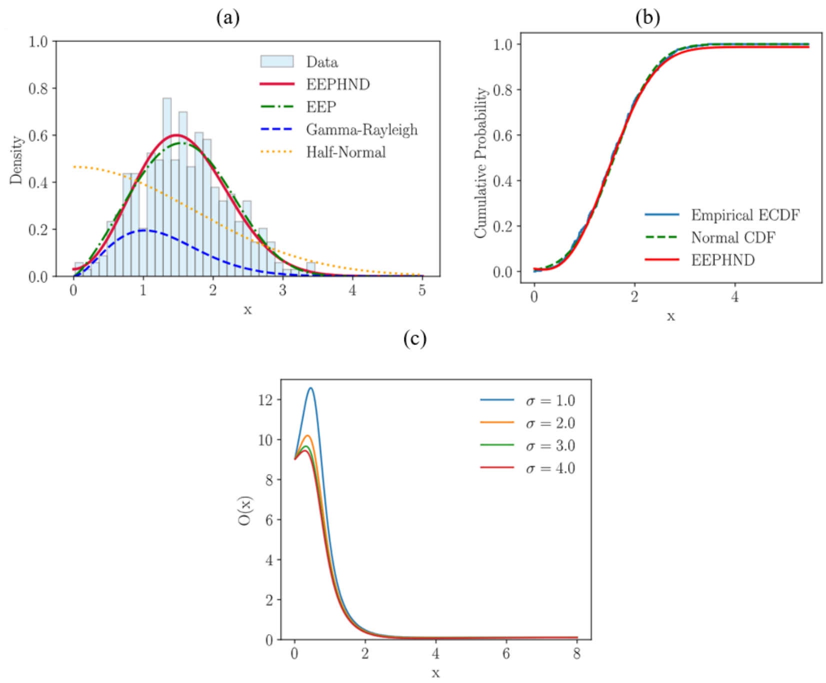
3.1 Validation with Simulated Data

Figure 1 shows the effect of varying σ on the EEPHND with fixed $\alpha = \beta = \theta = \lambda = 2$. As σ increases, the PDF broadens, the CDF smooths (Figure 2), the hazard flattens, and reliability declines more gradually. panels (e) and (f) confirm that EEPHND fits simulated data well and better matches the ECDF than the normal model, highlighting its effectiveness for lifetime and reliability modeling.

The model comparison metrics (AIC, BIC, and CAIC) indicate that the EEPHND model has the lowest values, suggesting the best overall fit among the four models evaluated, closely followed by EEP, see Table 1. In contrast, the Gamma-Rayleigh and HN models show significantly higher scores, indicating poorer fit. The estimated parameters for the EEPHND model include $\alpha = 1.1932$, $\beta = 1.8815$, $\theta = 2.4340$, $\lambda = 1.3058$, $\sigma = 0.3219$, and mixing probability $p_1 = 0.9878$, and$p_2 = 0.0122$. The 95% bootstrap confidence intervals for these parameters are reasonably tight, demonstrating parameter stability and suggesting that the EEPHND model provides a flexible and statistically reliable fit to the data.



Comparison of EEPHND characteristics under varying σ for fixed $\alpha = \theta = \lambda = \beta = 2$. Subplots: (a) PDF, (b) CDF, (c) Hazard, (d) Reliability, (e) Fitted PDF vs Data, (f) ECDF comparison.



Comparison of different competing models (a), the CDF and EDCF plot (b) and the Odds function (c)

Table 1

Model comparison metrics and estimated parameters with 95% bootstrap confidence intervals.

Model Comparison (Lower is Better)		**AIC**	**BIC**	**CAIC**
EEPHND		599.6024	614.418	599.671
EEP		600.024	622.247	600.167
GR		1394.505	1401.913	1394.526
HN		760.756	764.460	760.763
Estimated Parameters (MLE) and 95% Bootstrap Confidence Intervals				
Param	**MLE**	**95% CI Lower**	**95% CI Upper**	
α	1.1932	0.5560	1.8465	
β	1.8815	1.6253	2.1415	
θ	2.4340	1.9677	3.8680	
λ	1.3058	0.6797	1.5823	
σ	0.3219	0.0106	1.4331	
p_1	0.9878	0.9261	0.9999	

3.2 Validation with Real Data

3.2.1 Real Data

The real-life clinical dataset used in this work is the publicly available lung cancer survival dataset from the lifelines Python package (Davidson-Pilon 2019). It can be accessed via the load_lung() function, which returns a structured dataFrame containing time-to-event data, censoring indicators, and covariates such as age, sex, ECOG performance score, and weight loss. For reproducibility, all preprocessing steps, such as dropping missing values and rescaling the survival time to the range [0,1], were performed using standard Python pandas operations (McKinney et al. 2011). The dataset originates from the North Central Cancer Treatment Group (Shaw et al. 2002) and is frequently used for benchmarking survival models.

Table 2

Comparison of Survival Models: Cox PH, Kaplan-Meier, and EEPHND

Metric	**Cox PH Model**	**Kaplan-Meier**	**EEPHND (Parametric)**
Concordance Index (CI)			
CI Score	0.6029	0.9982	**0.9997**
Model Parameters			
Covariate: Age	coef = 0.0170	—	—
	$HR = 1.017$	—	—
	95% CI = [-0.0010, 0.0351]	—	—
	p = 0.0646	—	—
Covariate: Sex (M=1)	coef = **-0.5132**	—	—
	$HR = 0.599$	—	—
	95% CI = **[-0.8414, -0.1850]**	—	—
	p = **0.0022**	—	—
EEPHND Model Coefficients			
α (shape)	—	—	0.0001 (unstable)
β (scale)	—	—	0.02
θ (exponent)	—	—	0.01
λ (rate)	—	—	13.79

Metric	Cox PH Model	Kaplan-Meier	EEPHND (Parametric)
σ (Half-Normal scale)	—	—	**0.46**
p_1 (EEP weight)	—	—	**0.01**
Survival Estimate at $t = 0.012$			
$S(t)$	0.9791	0.9781	≈ 0.979
95% CI	—	[0.9481, 0.9908]	—
Model Type and Flexibility			
Type	Semi-parametric	Non-parametric	Fully parametric (mixture)
Baseline Hazard	Estimated	Stepwise constant	Closed-form from EEPHND
Handles Covariates	Yes	No	No (in current form)

3.2.2 Model Comparison

Table 2 presents a comparative summary of three survival models, namely Cox Proportional Hazards (Cox PH), Kaplan-Meier, and the proposed EEPHND, applied to real clinical data. Based on the Concordance Index (CI), EEPHND (**CI = 0.9997**) and Kaplan-Meier (**CI = 0.9982**) clearly outperform the Cox PH model (CI = 0.6029), suggesting superior predictive accuracy. For covariates in the Cox PH model, the effect of age was marginally significant (p=0.0646), while sex (male) showed a statistically significant association with poorer survival (HR=0.599, p=0.0022). The EEPHND model, despite being fully parametric, does not incorporate covariates in its current formulation but produced highly flexible parameter estimates, notably a strong Half-Normal scale component (σ=0.46) and a very small EEP weight (p_1 = 0.01), see Figure 3. Survival estimates at t=0.012 were consistent across all models. Overall, EEPHND offers high accuracy and closed-form flexibility 3, while the Cox model adds interpretability through covariate effects.

Table 3
Model Comparison using AIC, BIC, and CAIC criteria (lower is better)

Model	AIC	BIC	CAIC
EEPHND	**-123.340**	**-102.764**	**-96.764**
Log-Normal	-71.642	-64.783	-62.783
EEP	12.873	26.591	30.591
Gamma-Rayleigh	497.744	504.603	506.603

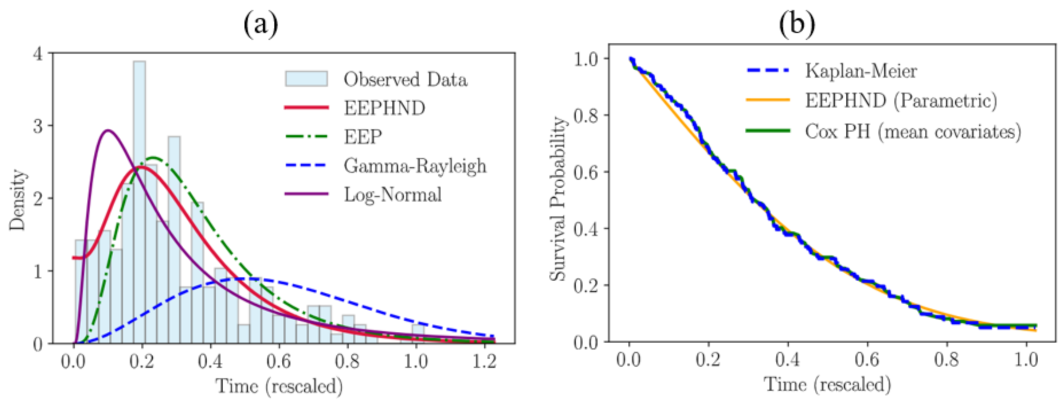


Fig 3. The density of different candidate models (a), and comparison of survival models (b)

The EEPHND model captures both early- and late-risk populations through a mixture of EEPD and HN components, governed by a flexible mixture parameter p_1 with $p_2 = 1 - p_1$. Compared to Cox PH and Kaplan-Meier models, EEPHND provides an explicit parametric form for the baseline survival, allowing direct interpretation of shape, scale, and risk mixture in survival dynamics. Notably, the heavy tail behavior (modulated by α, θ, and λ) makes it well-suited for datasets with heterogeneous risk profiles, unlike Cox PH models which rely heavily on covariates 2 for capturing such nuances.

Overall, the Cox model identifies sex as a significant predictor, with males showing a lower hazard rate than females (HR = 0.599, p = 0.0022), highlighting the value of covariate inclusion in detecting clinical risk differences. While the EEPHND model currently lacks covariate, it outperforms both Cox and Kaplan-Meier in predictive accuracy (CI = 0.9997) and closely matches their early survival estimates. Its fully parametric, closed-form structure enables continuous survival modeling, making it well-suited for clinical tasks like prognosis simulation and population-level planning, particularly when individual covariate data are limited or not available. Together, these models offer complementary strengths in risk interpretation and survival forecasting.

Author Contribution

Oriyomi Ahmad Hassan: Writing – review & editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Aisha Tunrayo Maradesa: Writing – review & editing, Writing – original draft, Software, Methodology.

Toyosi Abdulazeez Alabi: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation.

Akinwale Victor Famotire: Writing – review & editing, Methodology, Visualization, Investigation.

Oyejide Surajudeen Salam: Writing – review & editing, Validation, Investigation.

Ajani Busari: Writing –review & editing, Investigation.

Habeeb Abiodun Afolabi: Writing – review & editing, Investigation.

Solomon Adeleke: Writing – review & editing, Visualization.

Abayomi Ayodele Akomolafe: Supervision, Software, Project administration, Methodology, Investigation, Conceptualization.

References

Adegoke, A. B., R. A. Adepoju, and A. T.-A. Khan. 2020. "Molecular Dynamic (Md) Simulation and Modeling the Bio-Molecular Structure of Human Udp Glucose-6-Dehydrogenase Isoform 1 (Hugdh) Related to Prostate Cancer." *Basrah Journal of Science* 38 (3): 448–66.

Adegoke, A., G. Azeez, A. Lawal, and M. Imran. 2021. "Theoretical Studies of 1, 2, 3-Triazole and Isoxazole-Linked Pyrazole Hybrids as Antibacterial Agents: An Approach of Docking and Density Functional Theory." *Advanced Journal of Chemistry, Section B* 3 (2): 148–59.

Adegoke Afeez, B., M. Adeleke, and A. Habeeb. 2019. "Statistical Analysis of the Inhibitory Activities of Triterpenoid Derivatives Against Two Selected Diseases." *Int. J. Res. Innov. Appl. Sci* 6 (7): 130–35.

Adegoke, A., T. Oyelowo, and J. Sanusi. 2020. "Computational Studies of 1, 2, 3-Triazoles Derivatives Against Yellow Fever Virus: Dft Based." *European Modern Journal Studies* 4: 94–103.

Adeleke, M. 2020. "Beta-Hyperhalfnormal Distribution and Its Application." *Basrah Journal of Science* 38 (2): 131–56.

Adeleke, M., A. A. Ayodele, and A. F. Barnabas. 2019. "Exponentiated Exponential Lomax Distribution and Its Properties." *Mathematical Theory and Modeling*.

Akinsete, A., F. Famoye, and C. Lee. 2008. "The Beta-Pareto Distribution." *Statistics* 42 (6): 547–63.

Akomolafe, A., and A. Maradesa. 2017. "Betahalfnormal Distribution and Its Properties." *International Journal of Advance Research and Publication* 1 (4): 17–22.

Akomolafe, A., and A. Maradesa. 2019. "On the Doubly Truncated Exponential Pareto Distribution." *The Pacific Journal of Science and Technology* 20 (1).

Akomolafe, A., A. Maradesa, and H. Hassan. 2019. "Beta-Hjorth Distribution and Its Properties." *Pacific Journal of Science and Technology* 20 (1): 96–102.

Arshad, R. M. I., M. H. Tahir, C. Chesneau, and F. Jamal. 2020. "The Gamma Kumaraswamy-g Family of Distributions: Theory, Inference and Applications." *Statistics in Transition New Series* 21 (5): 17–40.

Ashour, S., and M. Eltehiwy. 2013. "Transmuted Lomax Distribution." *American Journal of Applied Mathematics and Statistics* 1 (6): 121–27.

Babatunde, A. A., and M. Adeleke. 2020. "In-Silico Studies and Property Model to Investigate the Binding Effect of Selected Ligands on Hiv Integrase." *Basrah Journal of Science (Bas J Sci)* 38 (2).

Bourguignon, M., I. Ghosh, and G. M. Cordeiro. 2016. "General Results for the Transmuted Family of Distributions and New Models." *Journal of Probability and Statistics* 2016 (1): 7208425.

Brentnall, A. R., and J. Cuzick. 2018. "Use of the Concordance Index for Predictors of Censored Survival Data." *Statistical Methods in Medical Research* 27 (8): 2359–73.

Burr, I. W. 1942. "Cumulative Frequency Functions." *The Annals of Mathematical Statistics* 13 (2): 215–32.

Chhetri, S. B., A. A. Akinsete, G. Aryal, and H. Long. 2017. "The Kumaraswamy Transmuted Pareto Distribution." *Journal of Statistical Distributions and Applications* 4: 1–24.

Davidson-Pilon, C. 2019. "Lifelines: Survival Analysis in Python." *Journal of Open Source Software* 4 (40): 1317.

Elbatal, I., S. Khan, T. Hussain, et al. 2022. "A New Family of Lifetime Models: Theoretical Developments with Applications in Biomedical and Environmental Data." *Axioms* 11 (8): 361.

Eugene, N., C. Lee, and F. Famoye. 2002. "Beta-Normal Distribution and Its Applications." *Communications in Statistics-Theory and Methods* 31 (4): 497–512.

Gnedenko, B., I. V. Pavlov, and I. A. Ushakov. 1999. *Statistical Reliability Engineering*. John Wiley & Sons.

Huang, Z. C. 2023. "Toward Closed Form Formulas for System Reliability and Confidence Quantification." *2023 Annual Reliability and Maintainability Symposium (RAMS)*, 1–6.

Hybels, C. F., and D. G. Blazer. 2003. "Epidemiology of Late-Life Mental Disorders." *Clinics in Geriatric Medicine* 19 (4): 663–96.

Jennrich, R. I., and S. M. Robinson. 1969. "A Newton-Raphson Algorithm for Maximum Likelihood Factor Analysis." *Psychometrika* 34 (1): 111–23.

McKinney, W. et al. 2011. "Pandas: A Foundational Python Library for Data Analysis and Statistics." *Python for High Performance and Scientific Computing* 14 (9): 1–9.

Merovci, F., and L. Puka. 2014. "Transmuted Pareto Distribution." *ProbStat Forum*, 1–11.

Nadarajah, S. 2011. "The Exponentiated Exponential Distribution: A Survey." *AStA Advances in Statistical Analysis* 95: 219–51.

Olatunji, A. G., L. I. Ajerogba, I. M. Olalekan, A. B. Adegoke, and R. A. Adepoju. 2021. "Phytochemical Screening of Seed Oils from Some Nigerian Grown Cucurbitaceae Plant Seeds." *IOSR J. Appl. Chem* 14 (1): 1–6.

Pal, M., M. M. Ali, and J. Woo. 2006. "Exponentiated Weibull Distribution." *Statistica* 66 (2): 139–47.

Shahriari, M., H. Shahrasbi, and A. Zaretalab. 2024. "Reliability Analysis of Lifetime Systems Based on Weibull Distribution." *International Journal of Nonlinear Analysis and Applications* 15 (1): 321–29.

Shaw, E., R. Arusell, B. Scheithauer, et al. 2002. "Prospective Randomized Trial of Low-Versus High-Dose Radiation Therapy in Adults with Supratentorial Low-Grade Glioma: Initial Report of a North Central Cancer Treatment Group/Radiation Therapy Oncology Group/Eastern Cooperative Oncology Group Study." *Journal of Clinical Oncology* 20 (9): 2267–76.

Shaw, W. T., and I. R. Buckley. 2009. "The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions, and a Skew-Kurtotic-Normal Distribution from a Rank Transmutation Map." *arXiv Preprint arXiv:0901.0434*.

Shittu, O. I., and K. A. Adepoju. 2013. "On the Beta-Nakagami Distribution." *Progress in Applied Mathematics* 5 (1): 49–58.

Tallis, G. M. 1961. "The Moment Generating Function of the Truncated Multi-Normal Distribution." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 23 (1): 223–29.

Tang, J., and T.-S. Su. 2008. "Estimating Failure Time Distribution and Its Parameters Based on Intermediate Data from a Wiener Degradation Model." *Naval Research Logistics (NRL)* 55 (3): 265–76.