Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics

MARSHALL-OLKIN BURR TYPE X MIXTURE CURE FRACTION MODEL FOR SURVIVAL ANALYSIS

Ibrahim Lauwali[1, 3], Umar Yusuf Madaki[2] and Aminu Suleiman Mohammed[3]

[1]Department of Statistics, School of Science and Technology, Federal Polytechnic Kaura Namoda, P.M.B 1012, Zamfara. ibrahimkaura1982@gmail.com
[2]Department of Mathematics and Statistics, Faculty of Science, Yobe State University, P.M.B 1144, Damaturu, NIGERIA
[3]Department of Statistics, Faculty of Physical Science, Ahmadu Bello University, Zaria.

ABSTRACT

In survival analysis with censored data, models accounting for individuals who never experience the event are termed long-term survival or cure models. The two main approaches are the mixture (standard cure) and non-mixture models. This study used the Marshall–Olkin Burr Type X (MOBX) distribution as the baseline for only mixture cure model (M) approach. We developed the overall survival functions for MOBXM, derived key statistical properties, and estimated parameters using maximum likelihood estimation. A simulation study was carried out to assess the consistency of the developed model estimates. The model is applied to business customer churn data and compared with existing cure models. Evaluation is based on AIC, BIC, and log-likelihood values. Results have shown that the proposed model fit better and capture survival behavior more effectively.

Keywords: Marshall-Olkin Distribution, Burr Type X Distribution, Mixture Cure Fraction Model, Time-to-event data.

1. INTRODUCTION

In contemporary survival analysis, parametric models have gained prominence due to their ability to handle complex data structures and parameters with greater precision Brostrom (2021). Modern statistical advancements have enabled more sophisticated techniques for estimating and interpreting survival data (Burke *et al*, 2019). Survival analysis relies heavily on two core models: the proportional hazards (PH) and accelerated failure time (AFT) models. These assume all subjects will eventually experience the event given sufficient follow-up time. However, real-world data often contains cure individuals who never experience the event, rendering standard models insufficient. Such cases require specialized cure rate models for accurate time-to-event analysis Lambert (2007). Researchers employ two primary models to analyze time-to-event data with a cure fraction. The first is Mixture Cure Model introduced by Boag (1949) and later by Berkson & Gage (1952), this approach divides the population into cured (long-term survivors) and uncured subgroups. And the second is Non-Mixture Cure Model Proposed by kuk *et al*. (1992) and later refined by Ibrahim *et al*. (2001) and Tsodikov (2002), this framework estimates cure rates without explicit subgroup classification.

In this study, we introduced and explored a novel four-parameter survival model named the Marshall-Olkin Burr Type X Cure Fraction Model (MOBXM). The model is constructed by extending the Marshall-Olkin Burr Type X (MOBX) distribution originally developed by Jamal *et al*. (2017), which serves as the baseline for modeling the survival times of uncured individuals. To accommodate the presence of a cured subpopulation, the MOBX distribution is integrated into the standard mixture cure fraction framework. This incorporation results in the formulation of the Marshall-Olkin Burr Type X

Mixture Cure Model (MOBXM). The proposed model is intended to enhance flexibility in analyzing survival data that includes long-term survivors

1.1 Marshall-Olkin Burr Type X (MOBX) Distribution

The Marshall-Olkin (MO) family of distributions is effective in modeling sudden failures due to external shocks but lacks the flexibility to represent diverse hazard rate shapes (Marshall & Olkin 1997). Similarly, the Burr Type X (BX) is limited in term of flexibility to model bathtub-shaped failure rates, restricting its applicability to monotonic hazard rates (Raqab & Kundu, 2005). To overcome these limitations, The MOBX distribution enhances the standard BX distribution by incorporating mechanisms for sudden failures while maintaining its flexible hazard rate shapes, making it more suitable for systems experiencing both gradual deterioration and external shocks (Jamal *et al*., 2017). The CDF, PDF and survival function of MOBX by Jamal *et al*, (2017) are defined (for $x > 0$) by;

$$G_{MOBX}(x;\lambda,\theta,\alpha) = \frac{\left(1-e^{-(\lambda x)^2}\right)^{\theta}}{\alpha+(1-\alpha)\left(1-e^{-(\lambda x)^2}\right)^{\theta}} \tag{1}$$

$$g_{MOBX}(x;\lambda,\theta,\alpha) = \frac{2\alpha\theta\lambda^2 x e^{-(\lambda x)^2}\left(1-e^{-(\lambda x)^2}\right)^{\theta-1}}{\left[\alpha+(1-\alpha)\left(1-e^{-(\lambda x)^2}\right)^{\theta}\right]^2} \tag{2}$$

$$S_{MOBX}(x;\lambda,\theta,\alpha) = \frac{\alpha\left[1-\left(1-e^{-(\lambda x)^2}\right)^{\theta}\right]}{\alpha+(1-\alpha)\left(1-e^{-(\lambda x)^2}\right)^{\theta}} \tag{3}$$

where; $\theta,\alpha > 0$: shape parameters and $\lambda > 0$: scale parameter.

2. MIXTURE CURE FRACTION MODELS

The cure fraction model (Achcar *et al*., 2012), also termed a long-term survival model, extends traditional survival analysis by accounting for individuals who may never experience the event of interest. Two common formulations are the parametric mixture cure model and the non-mixture cure model, both incorporating a parameter for the cure fraction (Berkson & Gage, 1952; Boag, 1949). Advances in medicine often lead to a proportion of the population being "cured," necessitating statistical models that include a cure fraction to evaluate treatment efficacy (Achcar *et al*., 2012; Martinez *et al*., 2013). The mixture cure model assumes the population consists of cured individual (who will never experience the event) and **s**usceptible individuals (who may experience the event). The overall survival function *S(t)* is given by:

$$S(t) = p + (1-p)\cdot S_o(t) \tag{4}$$

ICSSDA 2025

where; $S(t)$: The overall survival function of the population, p: the proportion of the population that is cured ($0 < p < 1$), $1 - p$: The proportion of the population that is not cured follows survival distribution, $S_o(t)$: The survival function of the uncured population.

3. RELATED WORKS

Boag (1949) pioneered the cure fraction concept by estimating the proportion of cancer patients cured via maximum likelihood, laying the foundation for survival models with long-term survivors. Berkson and Gage (1952) expanded this by modeling survival curves with a cured subgroup, emphasizing its utility in medical statistics. Farewell (1982) formalized the mixture cure model, distinguishing between cured and uncured populations using parametric survival distributions. Kuk and Chen (1992) enhanced mixture models by integrating logistic regression for cure probability and Cox regression for survival time, improving covariate analysis. Peng and Dear (2000) proposed a non-mixture cure model as an alternative, linking cure status to bounded cumulative hazard functions, offering flexibility in modeling. Several studies have extended cure rate modeling using flexible baseline distributions. Mazucheli *et al*. (2015) introduced Burr XII-based mixture and non-mixture cure models, showing improved fit for cancer survival data. Khan *et al*. (2013) developed Exponentiated exponential cure models incorporating covariates, estimated via the EM algorithm. Bustos and Vilar (2017) applied a Bayesian approach with the generalized Gompertz distribution, using MCMC to estimate cure fractions from real data.

Usman *et al*. (2021) developed the Nadarajah–Haghighi model for survival data incorporating long-term survivors, providing a flexible approach to handling right-censored datasets and illustrating its applicability in medical survival analysis. Building on this foundation, Aliyu and Usman (2024) extended the framework by formulating a non-mixture cure fraction model based on the Nadarajah–Haghighi distribution, integrating covariates to enhance the model's interpretability and predictive capability for censored medical data. In a related advancement, Usman *et al*. (2022) introduced a mixture cure model using the Weibull–Exponentiated Exponential distribution to analyze right-censored survival data, offering greater flexibility in estimating both cured and uncured population proportions. Expanding on similar methodological ideas, Yakubu *et al*. (2024) applied the Weibull–Exponentiated Exponential cure fraction model to cancer survival data, demonstrating its superior fit and interpretive strength in handling long-term survivors. Further extending the class of cure models, Terna and Adamu (2025) proposed a Lomax–Exponential–based cure fraction model, showcasing its robustness and applicability across diverse real-world datasets. Collectively, these works advance the development of flexible parametric cure models for right-censored survival data, emphasizing the role of generalized baseline distributions in improving model performance and interpretability

4. METHODOLOGY

Let T denote a random variable representing the time until the event of interest occurs, and let t > 0 be a specific observed value of T. The model described in Equation (4) aligns with the classical *mixture cure* model originally introduced by Boag (1949) and later formalized by Berkson and Gage (1952). In this framework, the survival function expresses the probability that the time-to-event exceeds a given time t, while p denotes the proportion of individuals who are cured or considered long-term survivors. Accordingly, the cumulative distribution function (CDF) of T is defined as follows:

$$G(t) = P(T \le t) = 1 - S_o(t) = F_o(t) \quad (5)$$

Therefore, $\lim_{t\to\infty} F_0(t) = 1$, implies that, $\lim_{t\to\infty} F_0(t) = 1 - p$. The probability density function of T is:

$$g(t)=\frac{dF(t)}{dt}=(1-p)\,f_o(t) \tag{6}$$

where $f_o(t)$ is the baseline probability density function for the susceptible individuals, Let us assume censored data. Considering a random sample (t_i,δ_i) of size n, $i=1,2,\cdots,n$, the contribution of the ith subject for the likelihood function is given by:

$$L=\prod_{i=1}^{n}\left[\left(g(t_i)\right)^{\delta_i}\cdot\left(S(t_i)\right)^{1-\delta_i}\right]=\prod_{i=1}^{n}\left[\left((1-p)\,f_o(t_i)\right)^{\delta_i}\cdot\left(p+(1-p)S_o(t_i)\right)^{1-\delta_i}\right] \tag{7}$$

where; δ_i is the censoring indicator variable, denoting δ_i = 1 for an observed lifetime and δ_i = 0 for a censored lifetime.

4.1 Marshall-Olkin Burr Type X with Mixture Cure Model (MOBXM)

Incorporating the MOBX distribution to mixture cure fraction model involves modeling the survival times of susceptible individuals using the MOBX distribution. While accounting for a fraction of individuals who are "cured" and will not experience the event of interest. Substituting (3) in to (4), the model is given as;

$$S_{MOBXM}(t;\lambda,\theta,\alpha,p)=p+(1-p)\cdot\frac{\alpha\left[1-\left(1-e^{-(\lambda t)^2}\right)^{\theta}\right]}{\alpha+(1-\alpha)\left(1-e^{-(\lambda t)^2}\right)^{\theta}} \tag{8}$$

Where; $\theta,\alpha>0$: shape parameters and $\lambda>0$: scale parameter, P: the proportion of the population that is cured ($0 < p < 1$), $1-p$: The proportion of the population that is not cured follows MOBX survival distribution,

4.2 The CDF and PDF of MOBXM

The CDF and PDF of MOBXM are respectively given as;

$$G_{MOBXMCM}(x;\lambda,\theta,\alpha,p)=(1-p)\left(1-\frac{\alpha\left[1-\left(1-e^{-(\lambda x)^2}\right)^{\theta}\right]}{\alpha+(1-\alpha)\left(1-e^{-(\lambda x)^2}\right)^{\theta}}\right) \tag{9}$$

$$f_{MOBXM}(t;\lambda,\theta,\alpha,p)=(1-p)\left(\frac{2\alpha\theta\lambda^2 te^{-(\lambda t)^2}\left(1-e^{-(\lambda t)^2}\right)^{\theta-1}}{\left[\alpha+(1-\alpha)\left(1-e^{-(\lambda t)^2}\right)^{\theta}\right]^2}\right) \tag{10}$$

4.3 Parameter Estimation of MOBXM using MLE

To estimate the parameters: λ, θ, α, and p of the MOBXM by using the Maximum Likelihood Estimation (MLE) method. The goal is to maximize the log- likelihood function**,** given a sample of observed data (Fisher, 1992). Let $X_1, X_2, \ldots, X_n$ be a random sample from the MOBXM distribution, with an observed dataset consisting of both uncensored and censored survival times. Define: $\delta_i = 1$ if the event (failure) is observed for individual i, and $\delta_i = 0$ if the event is censored for individual, i.

ICSSDA 2025

The likelihood function is the product of the contributions from uncensored and censored observations. Contribution of uncensored observations, $\delta_i = 1$. The likelihood contribution is the pdf of the MOBXM model. While contribution of censored observations, $\delta_i = 0$. The likelihood contribution is the survival function of the MOBXM model (Klein & Moeschberger, 2003). The likelihood and the log-likelihood are respectively given as;

$$L_{MOBXMCM}(\lambda,\theta,\alpha,p)\prod_{i=1}^{n}\left(\left[g_{MOBXMCM}(x_i)\right]^{\delta_i}\cdot\left[S_{MOBXMCM}(x_i)\right]^{1-\delta_i}\right) \tag{11}$$

$$\ell_{MOBXMCM}(\lambda,\theta,\alpha,p)=\sum_{i=1}^{n}\left(\left[\delta_i \log g_{MOBXCM}(x_i)+(1-\delta_i)\log S_{MOBXCM}(x_i)\right]\right) \tag{12}$$

5. SIMULATION STUDY

In this section, we assess the performance of the MOBXM model through simulation Study with various sample sizes (n = 20, 50, 100, 200, 500) and the model parameter values. Different sample sizes are chosen to investigate the performance of the models.

Table 1 Estimates, Bias and RMSE of MOBXM for the model Parameter Values

n	Properties	λ = 2.0	θ = 2.0	α = 1.2	p = 0.3
20	Est	2.3617	14.8135	17.8368	0.2830
	Bias	0.3617	12.8135	16.6368	-0.0170
	RMSE	1.6979	41.4901	60.3282	0.1226
50	Est	2.1023	4.7910	5.6336	0.2900
	Bias	0.1023	2.7910	4.4336	-0.0100
	RMSE	1.2031	15.5588	38.8559	0.0974
100	Est	1.9465	2.0358	1.2260	0.2914
	Bias	-0.0535	0.0358	0.0260	-0.0086
	RMSE	0.3571	0.4169	0.7049	0.0545
200	Est	2.0006	2.0220	1.2239	0.2983
	Bias	0.0006	0.0220	0.0239	-0.0017
	RMSE	0.2382	0.3019	0.4423	0.0367
500	Est	1.9993	2.0001	1.1936	0.2995
	Bias	-0.0007	0.0001	-0.0064	-0.0005
	RMSE	0.1687	0.1854	0.2935	0.0241

From Table 1, the results of the simulation study for the Marshall-Olkin Burr Type X Mixture Cure Model demonstrates that the model's parameter estimates become increasingly accurate and consistent as the sample size grows. At small sample sizes n = 20, the estimates for θ and α, are substantially biased and highly variable, with large RMSE values (12.81 and 16.64 for bias, and 41.49 and 60.33 for RMSE, respectively), indicating poor estimation precision. As the sample size increases to n = 50 and n = 100, the biases and RMSEs for all parameters noticeably decrease, especially for λ and p, suggesting improvement in estimation stability. By n = 200, the estimates for all parameters are very close to their true values, with minimal bias and low RMSEs, and at n = 500, the estimates are nearly unbiased (bias < 0.001 in most cases) and the RMSEs are the lowest across all sample sizes. These results confirm the

consistency of the MOBXM parameter estimators, as both bias and RMSE approach zero with increasing sample size, validating the reliability of the model in large-sample settings.

6. ANALYSIS AND RESULTS

To demonstrate the practical application of MOBXM, we analyzed the Business Customer Churn Dataset by Helsen & Schmittlein (1993), which records the duration (in days) until churn for a cohort of subscription service users. Each observation includes the time until churn and a status indicator, where 1 denotes churn and 0 indicates continued subscription. For this study, we focused on modeling the churn behavior to evaluate long-term subscriber retention and identify the presence of a cured (loyal) customer segment.

Table 2. MLEs, Log-likelihood, AIC and BIC of all the competing Mixture Cure models for **Business Customer Churn Dataset**

Model	Param	Est	Log-Lik	AIC	BIC
MOBXM	λ	0.0011	-950.88	1909.76	1920.4
	θ	0.8114			
	α	0.0320			
	P	0.2242			
GEM	λ	0.0090	-952.25	1910.51	1922.95
	α	1.4606			
	P	0.2530			
BXIIM	λ	83.633	-959.71	1925.42	1935.31
	θ	1.1708			
	P	0.0004			
BXM	k	0.0042	-959.75	1925.50	1935.39
	c	0.4773			
	P	0.2571			

Table 2 presents the maximum likelihood estimates (MLEs), log-likelihood, AIC, and BIC values for four competing mixture cure models fitted to the Business Customer Churn dataset. Among the models, the Marshall–Olkin Burr X Mixture (MOBXM) achieved the highest log-likelihood (–950.88) with the lowest AIC (1909.76) and BIC (1920.40), indicating the best overall fit to the data. The Generalized Exponential Mixture (GEM) followed closely, though with slightly higher AIC and BIC values, suggesting a reasonable but less efficient fit compared to MOBXM. In contrast, the Burr XII Mixture (BXIIM) and Burr X Mixture (BXM) recorded lower log-likelihood values and higher information criteria, making them less competitive. Overall, the MOBXM model provides the best balance between fit and parsimony, highlighting its suitability for modeling customer churn survival with the presence of long-term survivors (cured fraction).

ICSSDA 2025

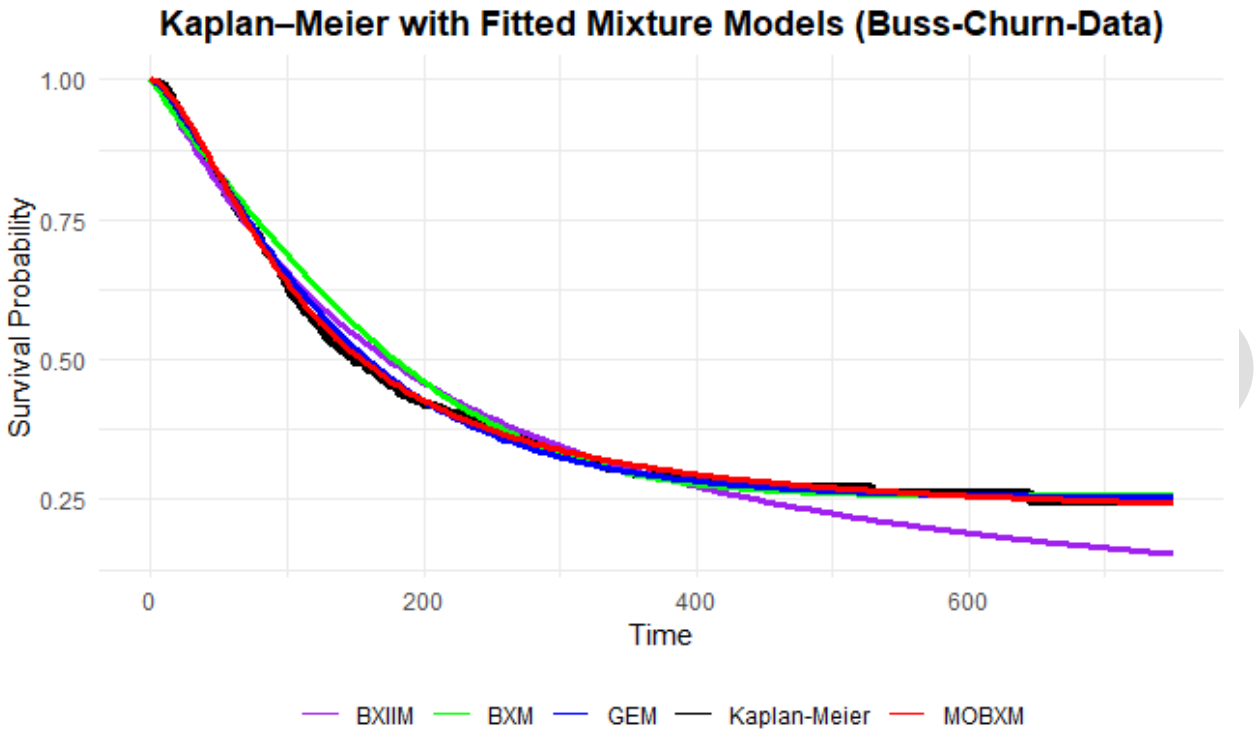

Figure 1: plot of Kaplan-Meier survival curve overlaid of the mixture cure models with the fits of the MLEs for the Business customer churn data.

The figure above showed the Kaplan–Meier survival curve (black) for business customer churn data, compared with four fitted mixture cure models. The Marshall–Olkin Burr X Mixture (MOBXM, red) aligned most closely with the Kaplan–Meier curve across time, confirming its good fit as also shown in the table. The Generalized Exponential Mixture (GEM, blue) and Burr X Mixture (BXM, green) also followed the data reasonably well but deviate slightly. The Burr XII Mixture (BXIIM, purple) diverged more in the long run, underestimating survival probabilities. Overall, the MOBXM model provided the best visual and statistical fit to the customer churn data.

7. CONCLUSION

In survival analysis with long-term survivors, the proposed MOBXM model consistently outperformed existing mixture models when applied to real business churn data. It provided a realistic cure fraction and effectively captured both early customer churn and long-term subscriber retention, making it statistically robust and practically meaningful. Competing models such as BXM and BXIIM produced weaker fits and unrealistic cure estimates, limiting their reliability. The Kaplan–Meier overlay further confirmed the strong alignment of MOBXM with observed survival patterns, validating its credibility. Thus, MOBXM emerges as the most suitable model for analyzing business time-to-event data with a cured (loyal) customer segment. We recommend adopting the MOBXM model for churn analysis, as it provided the most reliable fit. Future work should validate this model on other datasets to confirm its generalizability.

REFERENCE

Achcar, J. A., Barros, M., & Mazucheli, J. (2012). Cure rate models using different distributions for the latency time: A Bayesian approach. *Journal of Applied Statistics, 39*(1), 1–16.

Aliyu, Y. ., & Usman, U. (2024). Non-mixture Cure Fraction Model in the Presence of Right Censored Survival Data and Covariates Based on Nadarajah-Haghighi Distribution with Applications to Medical Data. *Thailand Statistician*, *23*(1), 41–63.

Berkson, J., & Gage, R. P. (1952). Survival Curve for Cancer Patients Following Treatment. Journal of the American Statistical Association, 47(259), 501–515.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 11(1), 15-53

Broström, G. (2021). Parametric models (Chap. 8). In *Event History Analysis with R* (2nd ed., pp. …). Chapman & Hall/CRC.

Burke, K., Jones, M. C. & Noufaily, A. (2019). A flexible parametric modelling framework for survival analysis. *arXiv preprint arXiv:1901.03214*.

Farewell, V. T. (1982). The Use of Mixture Models for the Analysis of Survival Data with Long Term Survivors. *Biometrics, 38*(4), 1041–1046.

Fisher, R. A. (1992). "On the Mathematical Foundations of Theoretical Statistics." Philosophical Transactions of the Royal Society of London. Series A, 222, 309–368.

Helsen, K., & Schmittlein, D. C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science, 12*(4), 395-414

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). Bayesian semi-parametric models for survival data with a cure fraction. Biometrics, 57(2), 383-388.

Jamal, F., Tahir, M. H., Alizadeh, M., & Nasir, M. A. (2017). On Marshall-Olkin Burr X family of distribution *Tbilisi Mathematical Journal, 10*(4), 175–199

Khan, S., Pal, S., & Siddiqua, J. A. (2013). The exponentiated exponential mixture and non-mixture cure rate model in the presence of covariates. *Computational Statistics & Data Analysis, 60*, 156–169.

Klein, J. P., & Moeschberger, M. L. (2003). Survival Analysis. Techniques for Censored and Truncated Data (2nd Ed.).

Kuk, A. Y., & Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3), 531-541.

Lambert, P. C. (2007). Modeling of the cure fraction in survival studies. *International Journal of Biostatistics, 3*(1), 1–23.

ICSSDA 2025

Marshall, A. W., & Olkin, I. (1997). A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families. *Biometrika*, *84*(3), 641 652.

Martinez, E. Z., Achcar, J. A., Jácome, A. A., & Santos, J. S. (2013). Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Computer Methods and Programs in Biomedicine*, *112*(1), 58–67.

Mazucheli, E. A., Coelho-Barros, E. A, & Achcar, J. A. (2015). *Mixture and non-mixture cure rate model considering the Burr XII distribution.* In *Stochastic Models, Statistics and Their Applications* (pp. 217–224).

Raqab, M. Z., & Kundu, D. (2005). Burr Type X distribution: Revisited. *Journal of Probability and* Statistical Science, 3(1), 127–142.

Terna, G. I., & Adamu, A. U. (2025). *Cure fraction models based on a Lomax-exponential distribution with applications*. Royal Statistical Society Nigeria Local Group At: Nigeria.

Tsodikov, A. (1998). A Proportional Hazards Model Taking Account of Long-Term Survivors. Biometrics, 54, 1508-1516.

Tsodikov, A. D., Ibrahim, J. G., & Yakovlev, A. Y. (2003). Estimating Cure Rates from Survival Data: An Alternative to Two-Component Mixture Models. *Journal of the American Statistical Association*, *98*(464), 1063–1078.

Usman, U., Suleiman, S., Aliyu, Y., & Magaji, B. (2022). Mixture cure model for right censored survival data with Weibull exponentiated exponential distribution. *Pakistan Journal of Statistics*, 38(4), 431–450.

Usman, U., Suleiman, S., Magaji Arkilla, B., & Aliyu, Y. (2021). Nadarajah-Haghighi Model for Survival Data with Long Term Survivors in the Presence of Right Censored Data. *Pakistan Journal of Statistics and Operation Research*, *17*(3), 695-709.

Yakubu, A., Muhammad, N. I., Falgore, J. Y., & Rabiu, A. (2024). The Weibull-Exponentiated Exponential Cure Fraction Model for Right Censored Survival Data with Applications to Cancer Data. *African Multidisciplinary Journal of Sciences and Artificial Intelligence, 1*(2), 711-736.

ICSSDA 2025