Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Ode *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.004

JOURNAL OF STATISTICAL SCIENCES & DATA ANALYTICS
Department of Statistics, Ahmadu Bello University, Zaria, Nigeria | www.jossda.abu.edu.ng www.jossda.com.ng

A Proposed Odd Rayleigh-Exponential Distribution with Shared Gamma Frailty Survival Model

O. Ode[1], M. Musa Tasi'u[2], A. Usman[2] and I. A. Sadiq[2]
[1]*Department of Biostatistics, Federal University of Health Sciences, Otukpo, Nigeria*
[2]*Department of Statistics, Ahmadu Bello University, Zaria, Nigeria*

Corresponding Author:
O. Ode, Department of Biostatistics, Federal University of Health Sciences, Otukpo, Nigeria.
oga.ode@fuhso.edu.ng | +234 703 528 4876 | ORCID: 0009-0001-3160-526X

Abstract

Clustered survival data frequently exhibit unobserved heterogeneity that violates the independence assumption underlying conventional proportional hazards models, resulting in inefficient estimation and potentially misleading inference. This study develops a shared Gamma frailty model based on the Odd Rayleigh–Exponential Distribution (OR-ED) to accommodate latent cluster-level dependence in time-to-event data. The proposed framework extends the OR-ED to a proportional hazards frailty setting by deriving closed-form expressions for the marginal joint survival function and marginal likelihood through the Laplace transform of the Gamma frailty distribution. The corresponding score equations were derived, and model parameters were estimated using the maximum likelihood estimation method. The proposed model is assessed through Monte Carlo simulation experiments under varying cluster sizes and censoring levels and is compared with the corresponding OR-ED model that ignores frailty. The simulation results demonstrate that incorporating shared frailty substantially improves parameter estimation, produces hazard functions that closely follow the underlying data-generating mechanism, and provides more reliable inference than the no-frailty counterpart when unobserved heterogeneity is present. Graphical diagnostics further support the adequacy of the Gamma frailty assumption in representing latent cluster effects. Overall, the proposed OR-ED shared Gamma frailty model provides a tractable and flexible framework for analysing clustered survival data and offers an effective alternative for applications involving correlated failure times and unobserved heterogeneity.

Keywords: Clustered dataset; correlated event times; Laplace transform; proportional hazards; shared frailty; survival model; unobserved heterogeneity.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

1. Introduction

Classical survival models generally assume that event times are independent. In practice, however, individuals within the same family, household, community, or other naturally occurring cluster often share genetic, environmental, or behavioural characteristics that induce dependence among their survival times. Such intra-cluster correlation violates the independence assumption of standard survival models, making them unsuitable for clustered time-to-event data. As noted by Mbona *et al*. (2023), these data are more appropriately analysed using multivariate survival models, particularly shared frailty models.

Frailty models extend conventional survival regression by incorporating an unobserved multiplicative random effect into the hazard function to account for unmeasured heterogeneity among individuals or clusters. This latent frailty is typically assumed to have a unit mean and finite variance, with larger frailty values corresponding to higher failure risk. By capturing unobserved risk factors, frailty models provide a flexible framework for analysing correlated survival data.

Time-to-event analysis examines the duration from a defined starting point to the occurrence of an event of interest (Mbona *et al*., 2023). A widely used approach is the proportional hazards (PH) model of Cox (1972), which assumes constant hazard ratios over time. In the frailty approach, multivariate survival data consist of correlated observations within the same individual or cluster (Arnheim, 2014), analogous to the clustered structure encountered in linear mixed-effects models.

Frailty modelling is motivated by the practical difficulty of measuring all factors influencing disease progression or failure risk (Pickles & Crouchley, 1995; Gutierrez, 2002; Gorfine & Hsu, 2011; Hazarika & Mahanta, 2016; Mahanta & Hazarika, 2018). It extends conventional survival models by introducing a multiplicative random effect into the baseline hazard to account for unobserved heterogeneity, and is therefore often regarded as a random-effects survival model. Among its forms, the PH frailty model is the most widely studied and applied because it provides a theoretically sound method for analysing dependent survival data when observed covariates alone are insufficient.

Frailty models have become fundamental in multivariate survival analysis (Pickles & Crouch-ley, 1995; Gutierrez, 2002; Wienke *et al*., 2005; Govindarajulu *et al*., 2011; Gorfine & Hsu, 2011; Hazarika & Mahanta, 2016; Mahanta & Hazarika, 2018; Balan & Putter, 2020). The Gamma frailty model is particularly popular due to its compatibility with the PH approach and the availability of closed-form survival and hazard functions through the Laplace transform, which links conditional and marginal survival quantities (Balan & Putter, 2020). Shared frailty induces positive dependence among clustered observations, while ignoring it may lead to biased estimation and misleading inference (Abdulkarimova, 2013).

Survival analysis models time-to-event outcomes using survival or hazard functions as the primary response (Govindarajulu *et al*., 2011). Although the Cox PH model is widely used because of its flexibility, it assumes that all relevant heterogeneity is explained by observed covariates. Frailty models relax this assumption by incorporating latent random effects, thereby extending conventional survival models to clustered and multivariate settings. As noted by Gutierrez (2002), frailty models are analogous to random-effects regression models but specifically designed for time-to-event data.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

Several methodological developments have strengthened frailty modelling. Gorfine and Hsu (2011) proposed frailty-based competing risk models for clustered survival data with parametric and semi-parametric MLE, demonstrating good performance through simulation. Ripatti and Palmgren (2000) developed a multiplicative frailty model with multivariate lognormal frailties using Laplace approximations and penalised partial likelihood, whereas Stefanescu and Turnbull (2006) highlighted the computational challenges of MLE for multivariate lognormal frailty models, particularly with varying cluster sizes.

Recent studies have broadened the range of frailty distributions. Yslas (2025) proposed phase-type frailty models for univariate and multivariate survival data, preserving analytical tractability through closed-form expressions and EM-based estimation. Likewise, Liyanage and Karunarathna (2022) developed a joint frailty model for two correlated survival outcomes, reporting improved bias, coverage probability, and mean squared error over univariate PH models, with Gamma frailty performing best under exponential survival and Normal frailty under Weibull survival.

Gallardo *et al*. (2025) further introduced a multivariate shared frailty model based on the truncated normal distribution with parametric and non-parametric baseline hazards. Using closed-form Laplace transforms and EM estimation, the model demonstrated consistent parameter estimation and superior performance on medical datasets.

Motivated by these developments, this study proposes a shared Gamma frailty model based on the OR-ED of Ode *et al*. (2025). By combining the flexibility of the OR-ED with the analytical advantages of Gamma frailty, the proposed model extends existing frailty methodologies beyond traditional baseline distributions and contributes to the growing literature on multivariate survival modelling. It also provides a practical alternative to the estimation challenges reported by Stefanescu and Turnbull (2006) for multivariate lognormal frailty models.

2. Materials and Methods

2.1. Odd Rayleigh-Exponential Distribution

The CDF and PDF of the OR-ED are respectively expressed as:

$$F(x;\theta,\lambda) = 1 - \exp\left(-\frac{(e^{\lambda x}-1)^2}{2\theta^2}\right) \tag{1}$$

$$f(x;\theta,\lambda) = \frac{\lambda}{\theta^2}(e^{\lambda x}-1)(e^{\lambda x})\exp\left(-\frac{(e^{\lambda x}-1)^2}{2\theta^2}\right) \tag{2}$$

The hazard, cumulative hazard, and survival functions are given, respectively, as follows:

$$h(x;\theta,\lambda) = \frac{\lambda}{\theta^2}(e^{\lambda x}-1)(e^{\lambda x}) \tag{3}$$

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

$$H(x;\theta,\lambda) = \frac{(e^{\lambda x}-1)^2}{2\theta^2} \tag{4}$$

$$S(x;\theta,\lambda) = \exp\left[-\frac{(e^{\lambda x}-1)^2}{2\theta^2}\right] \tag{5}$$

where $\lambda > 0$, $\theta > 0$ are scale parameters, and $0 < x < \infty$.

2.2. Proportional Hazards (PH) Frailty Model

The OR-ED falls within the PH class and admits closed-form expressions for its survival and hazard functions, a property that facilitates its extension to multivariate survival settings. Following Gutierrez (2002), a distribution belongs to the PH family if its hazard function can be expressed as:

$$h_0(x \mid Z) = Zh_0(x) \tag{6}$$

with the corresponding cumulative hazard

$$H_0(x) = \int_0^x h_0(s)ds \tag{7}$$

and conditional survival for an individual, given frailty, Z

$$S(x \mid Z) = \exp(-ZH_0(x)) \tag{8}$$

Here, $h_0(x)$ denotes the baseline hazard, $H_0(x)$ is the baseline cumulative hazard, x represents time-to-event, and Z is a multiplicative frailty term (typically Gamma-distributed) that captures unobserved heterogeneity at the individual or cluster level by scaling the hazard upward or downward relative to observed covariates.

The hazard function is fundamental to survival analysis, describing the instantaneous risk of failure. In practice, unobserved heterogeneity often causes individuals with similar covariates to experience different risk levels, motivating the inclusion of frailty (random effects) in survival models.

The conditional survival function corresponds to an individual with a frailty value of $Z = 1$. For a heterogeneous population, inference is based on the marginal survival function, obtained by integrating the conditional survival over the frailty distribution:

$$\bar{S}(x) = E[\exp(-ZH(x))] \tag{9}$$

This marginal survival has a population-averaged interpretation and may be viewed as a weighted combination of individual survival curves. The associated marginal hazard function is derived as:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

$$h(x) = \frac{d}{dx}[-\log S(x)] \tag{10}$$

2.3. The Laplace Transform of Frailty Distributions

The Laplace transform serves as an important tool for connecting conditional hazard functions to their corresponding marginal hazards and survival functions. For a non-negative random variable *Z*, the Laplace transform is defined as:

$$L_Z(s) = \int_0^{\infty} e^{-sZ} f(z)dz \tag{11}$$

where *f(z)* denotes the PDF of *Z*, and *s* is a non-negative real number.

2.4. OR-ED Shared Frailty Model

In a shared frailty approach, individuals belonging to the same cluster are assumed to be influenced by a common latent frailty component. Let cluster *k* contain *n* individuals with survival times ($x_{k1}, \ldots, x_{kn}$). For the *i*th individual in cluster *k*, let $\mathbf{x}_{ki}$ denote the vector of observed covariates, and β the corresponding regression coefficients. Introducing the linear predictor $\mathbf{x}^T\beta$ and the associated covariate effect $\exp(\mathbf{x}^T\beta)$, along with a cluster-specific frailty term Z_k, the conditional hazard function given $Z_k = z$ is expressed as

$$h_{ki}(x \mid Z_k = z) = z h_{0i}(x) exp(\mathbf{x}_{ki}^{\mathbf{T}}\beta) \tag{12}$$

where $h_{0i}(x)$ represents the OR-ED baseline hazard with parameters (λ_i, θ_i), and $z > 0$ is a shared multiplicative effect capturing unobserved heterogeneity within cluster *k*.

The corresponding cumulative conditional hazard function is given as

$$H_{ki}(x \mid Z_k = z) = z H_{0i}(x) exp(\mathbf{x}_{ki}^{\mathbf{T}}\beta) \tag{13}$$

where $H_{0i}(x)$ denotes the OR-ED baseline hazard.

Equations (12 & 13) conform to the multiplicative structure of the PH model. Conditional on the shared frailty Z_k, individual survival times within a cluster are assumed independent. Unlike univariate survival data, clustered survival data consist of independent clusters containing correlated observations that share a common frailty. In Equation (12), the regression component captures observed heterogeneity, while the frailty term Z_k accounts for unobserved risk factors affecting survival.

2.4.1. Conditional and marginal joint survival

Under the assumption that survival times within a cluster are conditionally independent given the shared frailty Z_k, the conditional joint survival function for cluster *k* is obtained as the product of the individual conditional survival functions:

$$S_k(x_{k1,\ldots,}x_{kn} \mid Z_k = z) = \prod_{i=1}^{n} \exp(-H_{ki}(x_{ki} \mid z)) \tag{14}$$

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

$$= \exp(-z\sum_{i=1}^{n} H_{0i}(x_{ki}) exp(\mathbf{x}_{ki}{}^{\mathbf{T}}\beta)) \quad (15)$$

Let the cumulative hazard for cluster *k* be defined by:

$$W_k = \sum_{i=1}^{n} H_{0i}(x_{ki}) exp(\mathbf{x}_{ki}{}^{\mathbf{T}}\beta) \quad (16)$$

Under this definition, the joint conditional survival in equation (15) can be rewritten as:

$$S_k(x_k \mid z) = \exp(-zW_k) \quad (17)$$

The marginal joint survival is derived by integrating the conditional survival over the frailty distribution, as shown:

$$S_k(x_k) = \int_0^\infty S_k(x_k \mid z) g(z) dz = \int_0^\infty \exp(-zW_k) g(z) dz \quad (18)$$

Assume that the frailty term follows a Gamma distribution with shape and rate parameters both equal to α, ensuring a unit mean $E(Z_k) = 1$, and $Var\ (Z_k) = \mathbf{1}/\boldsymbol{\alpha}$. The PDF of the Gamma frailty is given as:

$$g(z) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\alpha z}, \ \ z > 0 \quad (19)$$

Substituting Equation (19) into Equation (18) yields:

$$S_k(x_k) = \int_0^\infty \exp(-zW_k) \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\alpha z} dz \quad (20)$$

which simplifies to:

$$S_k(x_k) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \int_0^\infty z^{\alpha-1} e^{-z(\alpha+W_k)} dz \quad (21)$$

Evaluation of this integral lies on the standard Gamma integral identity:

$$\int_0^\infty z^{\nu-1} e^{-\mu z} dz = \Gamma(\nu) \mu^{-\nu} \ for \Re(\nu), \Re(\mu) > 0 \quad (22)$$

By setting $\nu = \alpha$, and $\mu = \alpha + W_k$, the expression reduces to:

$$S_k(x_k) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \Gamma(\alpha)(\alpha + W_k)^{-\alpha} \quad (23)$$

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

and equivalently:

$$S_k(x_k) = \left(\frac{\alpha}{\alpha + W_k} \right)^{\alpha} \tag{24}$$

Consequently, the marginal joint survival function under the Gamma frailty model is obtained in closed form as:

$$S_k(x_{k1}, ..., x_{kn}) = \left(\frac{\alpha}{\alpha + \sum_{i=1}^{n} H_{0i}(x_{ki}) exp(\mathbf{x}_{ki}{}^{\mathbf{T}} \beta)} \right)^{\alpha} \tag{25}$$

This result follows directly from integrating out the frailty term using the Gamma integral identity, yielding a tractable expression for clustered survival data.

2.4.2. Conditional and marginal cluster likelihood with right-censoring

Let δ_{ki} denote the event indicator for the *i*th individual in cluster *k*, where δ_{ki} = 1 if the event of interest is observed and δ_{ki} = 0 if the observation is right-censored. Conditional on a given frailty value $Z = z$, the likelihood contribution of individual *i* in cluster *k* can be written as

$$\ell_{ki}(x_{ki} \mid z) = (h_{ki}(x_{ki} \mid z))^{\delta_{ki}} \exp(-H_{ki}(x_{ki} \mid z)) \tag{26}$$

Aggregating across all individuals within cluster k, the conditional likelihood of the cluster, given $Z = z$, is obtained as:

$$\prod_{i=1}^{n} (z h_{0i}(x_{ki}) e^{(\mathbf{x}_{ki}{}^{\mathbf{T}}\boldsymbol{\beta})})^{\delta_{ki}} \exp(-z \sum_{i=1}^{n} H_{0i}(x_{ki}) e^{\mathbf{x}_{ki}{}^{\mathbf{T}}\boldsymbol{\beta}}) \tag{27}$$

To eliminate the frailty term *z*, the conditional likelihood is integrated with respect to the Gamma density defined in Equation (19). Let $d_k = \sum_{i=1}^{n} \delta_{ki}$ represent the total number of observed events in cluster *k*, and recall the cluster-specific cumulative hazard defined in Equation (16). The resulting marginal likelihood for cluster *k* is therefore expressed as:

$$L_k(\Theta) = \int_0^{\infty} (\prod_{i=1}^{n} (h_{0i}(x_{ki}) e^{(\mathbf{x}_{ki}{}^{\mathbf{T}}\boldsymbol{\beta})})^{\delta_{ki}}) z^{d_k} e^{-zW_k} g(z) dz \tag{28}$$

$$= (\prod_{i=1}^{n} (h_{0i}(x_{ki}) e^{\delta_{ki}(\mathbf{x}_{ki}{}^{\mathbf{T}}\boldsymbol{\beta})}) \frac{\alpha^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} z^{\alpha - 1 + d_k} e^{-z(\alpha + W_k)} dz \tag{29}$$

Evaluation of this expression relies on the standard Gamma integral identity

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

$$\int_0^\infty z^{\nu-1}e^{-\mu z}dz = \frac{\Gamma(\nu)}{\mu^\nu}, \ \ \mu > 0 \tag{30}$$

Applying this identity with parameters $\nu = \alpha + d_k$ and $\mu = \alpha + W_k$ yields:

$$\int_0^\infty z^{\alpha-1+d_k}e^{-z(\alpha+W_k)}dz = \Gamma(\alpha+d_k)(\alpha+W_k)^{-(\alpha+d_k)} \tag{31}$$

Hence, the marginal likelihood contribution of cluster *k* takes the closed-form expression:

$$L_k(\Theta) = (\prod_{i=1}^{n}(h_{0i}(x_{ki})e^{\delta_{ki}(\mathbf{x}_{ki}{}^{\mathrm{T}}\boldsymbol{\beta})})\frac{\alpha^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+d_k)}{(\alpha+W_k)^{(\alpha+d_k)}} \tag{32}$$

The likelihood for the entire sample is obtained by multiplying the marginal likelihoods for all clusters. Consequently, the associated log-likelihood function, used for estimating the marginal parameters via MLE, is:

$$\ell(\Theta) = \sum_{k=1}^{K}\log L_k(\Theta) \tag{33}$$

$$= \sum_{k=1}^{K}\{\sum_{i=1}^{n}\delta_{ki}[\log h_0(x_{ki};\lambda,\theta) + \mathbf{x}_{ki}{}^{\mathrm{T}}\beta] \ + \alpha\log\alpha + \log\Gamma(\alpha+d_k) - \log\Gamma(\alpha) - (\alpha+d_k)\log(\alpha+W_k)\}$$

(34)

2.5. MLE of the OR-ED Shared Gamma Frailty Model

To obtain the score equation for the regression parameter vector β, define the following quantities:

δ_{ki} = event indicator, x_{ki} = observed time, $d_k = \sum_i \delta_{ki}$, $\boldsymbol{\delta}_k = (\delta_{ki},...,\delta_{kn})^T$, $\mathbf{X}_k$ = design matrix, $\mathbf{H}_{0,k} = (H_0(x_{k1},...,H_0(x_{kn}))^T$, and $e_k = \exp(X_k\beta)$

Under these definitions, the score contribution corresponding to cluster *k* can be expressed as:

$$\nabla_{\boldsymbol{\beta}}\ell_k = \frac{\partial\ell(\Theta)}{\partial\beta} = X_k{}^T\boldsymbol{\delta}_k - \frac{\alpha+d_k}{\alpha+W_k}X^T{}_k(H_{0,k}\circ \mathrm{e}_k) \tag{35}$$

Aggregating the cluster-specific contributions over all clusters leads to the estimating equation:

$$\nabla_{\boldsymbol{\beta}}\ell = \frac{\partial\ell(\Theta)}{\partial\beta} = \sum_{k=1}^{K}\left[X_k{}^T\boldsymbol{\delta}_k - \frac{\alpha+d_k}{\alpha+W_k}X^T{}_k(H_{0,k}\circ \mathrm{e}_k)\right] \tag{36}$$

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

In these expressions, the symbol "$\circ$" denotes the Hadamard (elementwise) product, rather than the standard matrix multiplication. The expression $\nabla_{\boldsymbol{\beta}}\ell(\Theta)$ represents the vector of partial derivatives of the log-likelihood with respect to the regression parameter vector β, $X^T{}_k\boldsymbol{\delta}_k$ is the covariate vector corresponding to observed events, and $(-\frac{\alpha+d_k}{\alpha+W_k}X^T{}_k(H_{0,k}\circ \mathrm{e}_k))$ captures the frailty-adjusted expected covariate contribution over the risk set. In clustered survival modelling, individual likelihood contributions are typically represented as vectors to maintain consistency across cluster-level calculations.

The log-likelihood function differentiated with respect to the frailty parameter α is given as:

$$\frac{\partial\ell(\Theta)}{\partial\alpha}=\sum_{k=1}^{K}\left[\log\alpha+1+\psi(\alpha+d_k)-\psi(\alpha)-\log(\alpha+W_k)-\frac{\alpha+d_k}{\alpha+W_k}\right] \tag{37}$$

where $\psi(.)$ denotes the digamma function, arising from differentiating the logarithm of the Gamma function.

Estimation of the baseline parameters λ and θ, both scalars, proceeds similarly.

$$\frac{\partial\ell(\Theta)}{\partial\lambda}=\sum_{k=1}^{K}\left[\delta_k{}^T\frac{\partial\log\mathbf{h}_{0,k}}{\partial\lambda}-\frac{\alpha+d_k}{\alpha+W_k}\left(\frac{\partial\mathbf{H}_{0,k}}{\partial\lambda}\right)^T e_k\right] \tag{38}$$

$$\frac{\partial\ell(\Theta)}{\partial\theta}=\sum_{k=1}^{K}\left[\delta_k{}^T\frac{\partial\log\mathbf{h}_{0,k}}{\partial\theta}-\frac{\alpha+d_k}{\alpha+W_k}\left(\frac{\partial\mathbf{H}_{0,k}}{\partial\theta}\right)^T e_k\right] \tag{39}$$

Closed-form solutions for these estimating equations are generally intractable, and all likelihood equations are inherently non-linear. Consequently, parameter estimation is implemented using numerical optimisation techniques to maximise the log-likelihood function.

3. Results and Discussions

3.1. Validity Plots of Frailty vs. No-Frailty

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

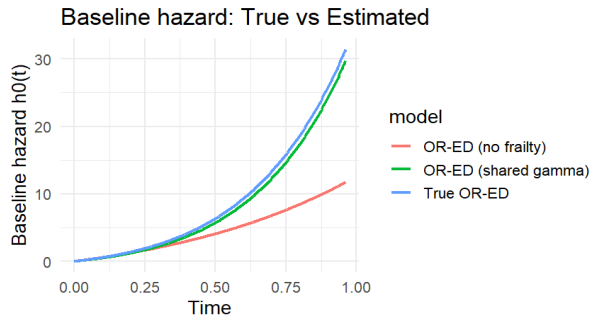

Figure 1. Baseline hazard

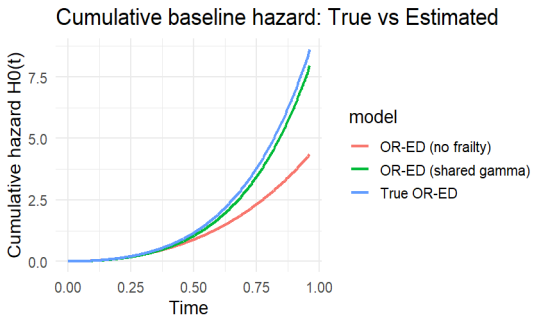

Figure 2. Cumulative baseline hazard

Figures 1 and 2 compare the true OR-ED baseline and cumulative hazard functions with estimates from the no-frailty and shared Gamma frailty models. The shared frailty model closely reproduces the true curves, whereas the no-frailty model consistently underestimates both functions by ignoring between-cluster heterogeneity. This demonstrates the ability of the proposed frailty model to capture latent cluster-level variability and reduce estimation bias.

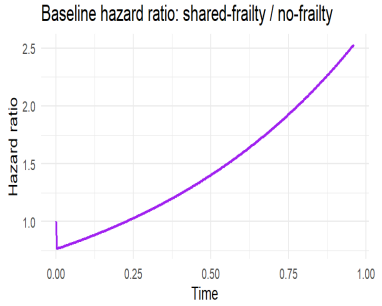

Figure 3. Baseline hazard ratio

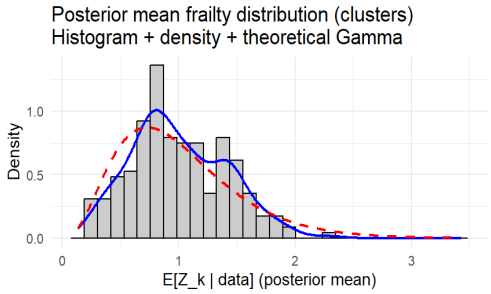

Figure 4. Posterior mean of frailty distribution

Figure 3 shows that the hazard ratio between the shared frailty and no-frailty models increases over time, indicating that ignoring frailty leads to progressively greater underestimation of risk. Figure 4, on the other hand, demonstrates that the estimated cluster-specific frailties closely follow the assumed Gamma distribution, providing graphical evidence that the shared Gamma frailty model adequately captures latent heterogeneity across clusters.

3.2. Simulation Results on the OR-ED Shared Gamma Frailty Model

Table 1 reports simulation results based on 120 clusters with four individuals per cluster (480 observations) under a censoring rate of 15%, meaning that 15% of individuals did not experience the event or were lost to follow-up. The true parameter values used in the simulation, including a frailty variance of 0.50, are shown in the second column, while standard errors (SE), *z*-statistics, *p*-values, and 95% confidence intervals (CIs) are presented for the shared Gamma frailty model. Blank entries indicate parameters not defined for a given model. When frailty is ignored, the OR-ED yields remarkably biased parameter estimates, reflecting that failure to account for cluster-level heterogeneity distorts baseline parameters and compromises inference. In contrast, the shared Gamma frailty model closely recovers the

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.004

true values, with small SEs and highly significant *p*-values (< 0.001). Although the frailty variance (0.582) is slightly overestimated, this does not affect the results and still confirms the presence of unobserved heterogeneity.

Table 1. Frailty-based simulation results on the OR-ED for 120 clusters

Parameter	True Value	OR-ED (No Frailty) Estimate	OR-ED (Shared Frailty) Estimate (SE)	Z	*p*-value	95% CI
λ	0.90	0.2256	1.0059 (0.1133)	8.879	< 0.001	0.7838, 1.2279
θ	1.10	0.2050	1.2977 (0.1965)	6.605	< 0.001	0.9126, 1.6829
α	2.00	-	1.7179 (0.1973)	8.707	< 0.001	1.3312, 2.1046
β_1	0.60	0.4230	0.5581 (0.0673)	8.298	< 0.001	0.4269, 0.6892
β_2	-0.40	-0.3672	-0.3889 (0.1157)	-3.363	< 0.001	-0.6166, -0.1613
β_3	0.80	0.6265	0.9157 (0.2173)	4.214	< 0.001	0.4898, 1.3416
$1/\alpha$	0.50	-	0.5820	-	-	0.5942, 1.2758

Similarly, Table 2 presents a second Monte Carlo experiment with 200 clusters of eight individuals each and a higher censoring rate of 20%. The no-frailty model again exhibits substantial bias, particularly for the baseline parameters, whereas the shared frailty model achieves excellent parameter recovery. Ignoring frailty forces the model to compensate by shrinking the hazard scale, leading to unreliable estimates. Regression coefficients are generally better recovered under the frailty specification, while the frailty parameter itself shows good recovery, with narrow CIs and strong statistical significance ($p < 0.001$).

Table 2. Frailty-based simulation output on the OR-ED for 200 clusters

Parameter	True Value	OR-ED (No Frailty) Estimate	OR-ED (Shared Frailty) Estimate (SE)	Z	*p*-value	95% CI
λ	1.3	0.5685	1.3594 (0.1149)	11.83	< 0.001	1.134, 1.585
θ	0.6	0.2467	0.6752 (0.0843)	8.01	< 0.001	0.510, 0.840
α	3.2	-	3.7145 (0.5737)	6.48	< 0.001	2.590, 4.839
β_1	0.4	0.4138	0.5025 (0.0307)	16.38	< 0.001	0.442, 0.563
β_2	-0.7	-0.5482	-0.6908 (0.0606)	-11.39	< 0.001	-0.810, -0.572
β_3	1.4	1.3647	1.6109 (0.1084)	14.86	< 0.001	1.398, 1.824
$1/\alpha$	0.3125	-	0.2692	-	-	0.206, 0.386

From these results, it is evident that neglecting frailty in clustered datasets leads to underestimated baseline hazard parameters and false precision. Incorporating a shared Gamma frailty component restores unbiasedness, yields interpretable covariate effects, and accurately captures between-cluster heterogeneity, thereby ensuring valid statistical inference.

4. Conclusion

This study proposed and developed a shared Gamma frailty survival model based on the OR-ED, extending its applicability to clustered time-to-event data. By incorporating the OR-ED in the PH frailty approach, closed-form expressions for the marginal survival and likelihood functions were obtained, enabling likelihood-based inference. Simulation results

succinctly showed that ignoring frailty in the presence of clustering leads to considerable bias in baseline hazard parameters, distorted hazard shapes, and unreliable covariate effects. In contrast, the shared Gamma frailty model consistently recovered the true parameter values, with small SEs and significant *p*-values. Graphical diagnostics further confirmed that the frailty model accurately captures latent cluster-level heterogeneity and corrects the downward bias observed in no-frailty models. Generally, the proposed OR-ED shared frailty model provides a tractable and statistically robust tool for analysing clustered survival data, offering valid inference where conventional models fail.

Funding: This work was supported by the Tertiary Education Trust Fund (TETFund), Nigeria.

References

Abdulkarimova, U. (2013). Frailty Models for Modelling Heterogeneity. *McMaster University, Ontario*, 1 – 76.

Arnheim, J. (2014). Introduction to Frailty Models. *McDougall Scientific Ltd., Ontario*, 1 – 9. www.mcdougallscientific.com

Balan, T. A., & Putter, H. (2020). A Tutorial on Frailty Models. *Statistical Methods in Medical Research*, 29(11), 3424 – 3454. http://dx.doi.org/10.1177/0962280220921889

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187 – 220. http://links.jstor.org/sici?sici=0035-9246%281972%2934%3A2%3C187%3ARMAL%3E2.0.CO%3B2-6

Gallardo, D. I., Gomez, Y. M., Santibanez, J. L., Venegas, O., & Bourguignon, M. (2025). The Multivariate Shared Truncated Normal Frailty Model with Application to Medical Data. *Scientific Reports*, 15(30099). https://doi.org/10.1038/s41598-025-15903-y

Gorfine, M., & Hsu, L. (2011). Frailty-Based Competing Risks Model for Multivariate Survival Data. *Biometrics*, 67(2), 415 – 426. https://doi.org/10.1111/j.1541-0420.2010.01470.x

Govindarajulu, U. S., Lin, H., Lunetta, K. L., & Agostino, R. B. D. (2011). Frailty Models: Applications to Biomedical and Genetic Studies. *Tutorial in Biostatistics*, 30, 2754 – 2764. https://doi.org/10.1002/sim.4277

Gutierrez, R. G. (2002). Parametric Frailty and Shared Frailty Survival Models. *The Stata Journal*, 2(1), 22 – 44.

Hazarika, J., & Mahanta, K. K. (2016). Studies on Survival Analysis using Frailty Models under Bayesian Mechanism and its Further Scope. *IOSR Journal of Mathematics*, 12(4), 15 – 21. https://doi.org/10.9790/5728-1204051521

Liyanage, J. C., & Karunarathna, G. H. S. (2022). Shared Frailty Model for Joint Survival Data: A Simulation Study. *Sri-Lankan Journal of Applied Statistics*, 23(2). http://doi.org/10.4038/sljastats.v23i2.8071

Mahanta, K. K., & Hazarika, J. (2018). A Bayesian Analysis of Univariate Proportional Hazard, Accelerated Failure Time and Proportional Odds Model Under Frailty Approach. *International Journal of Creative Research Thoughts*, 6(1), 755 – 765. ISSN: 2320 - 2882

Mbona, S. V., Mwambi, H., & Ramroop, S. (2023). The Importance of the Frailty Effect in Survival Models: For Multidrug-Resistant Tuberculosis Data. *The Open Public Health Journal*, 20, 1 – 8. http://dx.doi.org/10.2174/18749445-v16-230912-2023-76

McGilchrist, C. A. (1993). REML Estimation for Survival Models with Frailty. *Biometrics*, 49(1), 221 – 225. http://www.jstor.org/stable/2532615?origin=JSTOR-pdf

Ode, O., Tasi'u, M., Usman, A., & Sadiq, I. A. (2025). A Novel Odd Rayleigh-Exponential Distribution (OR-ED) and its Application to Lifetime Datasets. *J. Stat. Sci. Comput. Intell*., 1(3), 262 – 282. https://doi.org/10.64497/jssci.91

Pickles, A., & Crouchley, R. (1995). A Comparison of Frailty Models for Multivariate Survival Data. *Statistics in Medicine*, 14, 1447 – 1461.

Ripatti, S., & Palmgren, J. (2000). Estimation of Multivariate Frailty Models Using Penalised Partial Likelihood. *Biometrics*, 56(4), 1016 – 1022. https://www.jstor.org/stable/2677032

Stefanescu, C., & Turnbull, B. W. (2006). Multivariate Frailty Models for Exchangeable Survival Data with Covariates. *Technometrics*, 48(3), 411 – 417, https://doi.org/10.1198/004017006000000048

Wienke, A., Arbeev, K. G., Locatelli, I., & Yashin, A. I. (2005). A Comparison of Different Bivariate Correlated Frailty Models and Estimation Strategies. *Mathematical Biosciences*, 198, 1 – 13. https://doi.org/10.1016/j.mbs.2004.11.010

Yslas, J. (2025). Phase-type Frailty Models: A Flexible Approach to Modelling Unobserved Heterogeneity in Survival Analysis. *Scandinavian Actuarial Journal*. https://doi.org/10.1080/03461238.2025.2520586