

Comparison of Semi-Parametric Shared Frailty Models for Bees' Survival

Patience Isiaho^{1,2}, Daisy Salifu^{2*}, Samuel Mwalili¹, Henri E. Z. Tonnang²

¹Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

²Data Management, Modelling, and Geo-Information Unit, International Centre of Insect Physiology and Ecology (ICIPE), Nairobi, Kenya

Email: *dsalifu@icipe.org

How to cite this paper: Isiaho, P., Salifu, D., Mwalili, S. and Tonnang, H.E.Z. (2024) Comparison of Semi-Parametric Shared Frailty Models for Bees' Survival. *Journal of Data Analysis and Information Processing*, 12, 267-288.
<https://doi.org/10.4236/jdaip.2024.122015>

Received: April 15, 2024

Accepted: May 26, 2024

Published: May 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Survival analysis is a fundamental tool in medical science for time-to-event data. However, its application to colony organisms like bees poses challenges due to their social nature. Traditional survival models may not accurately capture the interdependence among individuals within a colony. Frailty models, accounting for shared risks within groups, offer a promising alternative. This study evaluates the performance of semi-parametric shared frailty models (gamma, inverse normal, and positive stable-in comparison to the traditional Cox model using bees' survival data). We examined the effect of misspecification of the frailty distribution on regression and heterogeneity parameters using simulation and concluded that the heterogeneity parameter was more sensitive to misspecification of the frailty distribution and choice of initial parameters (cluster size and true heterogeneity parameter) compared to the regression parameter. From the data, parameter estimates for covariates were close for the four models but slightly higher for the Cox model. The shared gamma frailty model provided a better fit to the data in comparison with the other models. Therefore, when focusing on regression parameters, the gamma frailty model is recommended. This research underscores the importance of tailored survival methodologies for accurately analyzing time-to-event data in social organisms.

Keywords

Correlated Failure Times, Frailty, Survival Analysis, Unobserved Heterogeneity

1. Introduction

Honey bees, (tribe Apini), refer to a group of insects in the family Apidae (order

Hymenoptera) that in a broad sense includes all bees that make honey. The Western Honey Bee (*Apis mellifera*), which is domesticated for honey production and crop pollination, is the most well-known type of honey bee. Only members of the genus *Apis* are considered true honey bees [1]. However, other species of bees, such as the stingless bees belonging to the genus *Meliponula* and the Indian stingless *Tetragonula iridipennis* have also been kept by humans to produce honey.

Colony collapse disorder (CCD), originally identified in the United States in 2006, resulted in considerable colony losses and posed severe challenges for agricultural pollination, a crucial service provided by the North American beekeeping industry. CCD can be caused by a many factors like disease, parasites (*Varroa destructor* and *Tropilaelaps clareae*) and possibly extensive use of broad-spectrum chemical pesticides. The effects of chemical pesticides on beneficial and nontarget insects are well documented by [2] [3] [4] and they are one of the main causes of the unprecedented decline of bee pollinators globally [5]. The entomopathogenic fungi *Beauveria bassiana* (Balsamo) Vuillemin and *Metarhizium anisopliae* (Metschnikoff) Sorokin are created and utilized as biopesticides globally. Based on their durability in the field and compatibility with the environment, these entomopathogens offer safer alternatives to chemical pest management [6]. However, it is essential to evaluate the safety of entomopathogens on the survival of bees and thus the context of the data being considered in this study.

Bee survival time has been modelled using Kaplan-Meier estimator [7], Generalized Linear Model (GLM) [8] and Cox proportional hazard model [9] with the assumption of independent and identically distributed data. These assumptions may not be realistic because bees are social insects and thus they live together in large, well-structured family groups such that the survival duration of a bee may depend on the survival duration of its nest-mates. This brings about heterogeneity or variability in the population which is regarded as one of the main factors influencing variation in biological applications. Assessing this heterogeneity may be challenging since it can be unobservable. This unobserved heterogeneity might stem from factors like genetic variations, differences in exposure to environmental stressors, or other unknown or unmeasured factors. [10] noted that the power to detect treatment differences is reduced by heterogeneity. Therefore, random effects, called frailties in the context of hazard models should be included in the model to account for the unobserved heterogeneity in survival modeling of organisms that live in colonies.

Clayton [11] initially introduced a model to account for the dependence between survival times within a cluster, although he did not employ the term “frailty.” Vaupel *et al.* [12] later coined the term “frailty” to describe this concept, specifically in the context of a random effects model for survival data aimed at tackling unobserved heterogeneity within a population. The classical widely used frailty model is based on a proportional hazards model conditional on the frailty. This loosely translates that an individual’s hazard depends additionally on an

unobservable variable Z that takes a constant value over time [13]. Let Z be a non-negative random variable, which is assumed to act multiplicatively on a baseline hazard function $\lambda_0(t)$, $t > 0$. In the univariate case, the hazard of an individual with frailty Z is specified as:

$$\lambda(t|Z) = Z\lambda_0(t)$$

By incorporating the frailty term into the model, we can accurately evaluate the effects of the covariates without underestimating or overestimating the parameters [14].

Frailty models account for the correlation structure in the data and the frailty terms are introduced to model the dependence between observations. Many frailty distributions have been studied by various authors, including the log-normal and the Power Variance Function (PVF) family, which comprises Gamma, Inverse Gaussian, Positive Stable and Compound Poisson distributions [15] [16]. The PVF family was considered in this study because of the tractable Laplace transform which simplifies the model manipulation and enables closed-form expressions for marginal functions [17]. The association between failure times was assessed using Kendall's tau for clustered data, a non-parametric measure advantageous for its mathematical simplicity in the context of the PVF family's Laplace transform properties. Owing to the latent nature of the frailty, it can be challenging to find the proper distribution of the frailty term for a given data set, and its misspecification can result in biased estimators, reduced efficiency, and misleading conclusions [18].

Frailty models, more specifically shared frailty models, have been commonly used in different studies to accommodate familial or, more generally, clustered survival data. Some of the studies include research by [19] [20] [21]. The objective of this study was therefore to assess the performance of different semi-parametric shared frailty models in the analysis of the bees' survival data. Specifically, the gamma, inverse Gaussian, and positive stable models are explored as frailty distributions. We also examined the effect of frailty distribution misspecification on regression and heterogeneity parameters estimate through a simulation study.

2. The Semi-Parametric Shared Frailty Model

Maximization of the partial likelihood is used to fit semiparametric hazard models without frailty terms [22]. The contribution of the unobserved frailty components, however, has to be taken into consideration for semiparametric frailty models.

Suppose there are S groups with n_i individuals within the i^{th} group ($i = 1, \dots, S$) in a survival study. Let T_{ij} denote the survival time and δ_{ij} denote the censoring indicator for the j^{th} individual ($j = 1, \dots, n_i$) in the i^{th} group. δ_{ij} takes a value of 0 if T_{ij} has not been observed and 1 when it is observed. Let X_{ij} a $p \times 1$, denote observable covariate vector for the j^{th} individual in the i^{th} group. The hazard rate function for the j^{th} individual in the i^{th} group at time t in the semi-parametric shared frailty model:

$$\lambda(t | X_{ij}) = \lambda_0(t) Z_i \exp(\beta^T X_{ij})$$

where Z_i is the “frailty” or unobserved random effect of the i^{th} group, β is a vector of unknown regression coefficients and $\lambda_0(t)$ is an unspecified baseline hazard function. Z_1, \dots, Z_S are assumed to be independent and have a common probability density function (pdf) $f(z)$, such that $E(Z_i) = 1$ for identifiability. As addressed by [23], the above model characterizes two sources of variation: the group variation described by Z_i , and the individual variation characterized by $\lambda_0(t) \exp(\beta^T X_{ij})$.

Individuals in the i^{th} group are assumed to be conditionally independent, given both the observed covariates and the group-specific frailty, ensuring that the analysis accounts for both measurable risk factors and unobserved group-specific effects on survival times. The joint survival distribution of failure times is therefore given by: $T_{i1}, T_{i2}, \dots, T_{in_i}$ conditional on X_{i1}, \dots, X_{in_i} is:

$$\begin{aligned} S(t_{i1}, \dots, t_{in_i}) &= \int_0^\infty P(T_{i1} > t_{i1}, \dots, T_{in_i} > t_{in_i} | z) f(z) dz \\ &= \int_0^\infty \exp\left\{-z \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta^T X_{ij})\right\} f(z) dz \\ &= \mathcal{L}_f\left(\sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta^T X_{ij})\right) \end{aligned} \tag{1}$$

where $\mathcal{L}_f = \int_0^\infty e^{-zs} f(z) dz$ is the Laplace transform of the pdf of the frailty $f(z)$ and $\Lambda_0(t) = \int_0^t \lambda_0(u) du$.

2.1. Frailty Distributions: The Power Variance Function (PVF) Family

A more general family of infinitely divisible PVF distributions is the power variance function (PVF) family, with the Laplace transform described by [23] [34]:

$$L(c; \alpha, y, m) = \exp\left[-\alpha \operatorname{sign}(m) \left\{1 - \left(\frac{\gamma}{\gamma + c}\right)^m\right\}\right] \tag{2}$$

where $\operatorname{sign}(m)$ is the sign of m , and $m > -1$ and $m \neq 0$. Particular cases include: The gamma frailty ($m \rightarrow 0$ with $m < 0$); inverse Gaussian distribution ($m = -\frac{1}{2}$); compound poisson DISTRIBUTION ($m > 0$) and the positive stable distribution ($\gamma \rightarrow 0$).

The log-normal distribution has often been used for frailty models, although it is not part of the PVF family.

1) **Gamma Frailty** As gamma distributions have a simple Laplace transform and thus simple expressions of likelihood functions, they are frequently used to model frailty. The pdf of Gamma distribution is:

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

where $\Gamma(\alpha)$ is the gamma function, θ is a rate parameter in the inverse scale,

and α is a shape parameter. For the identifiability problem, to restrict $\alpha = \theta$, we take $\mathbf{E}(Z) = 1$. This leads to:

$$f(z) = \frac{\theta^\theta z^{\theta-1} e^{-\theta z}}{\Gamma(\theta)}, \quad \theta > 0, z > 0$$

Then, the corresponding Laplace transform is:

$$\mathcal{L}(u) = (1 + \theta u)^{-1/\theta}$$

2) Positive stable frailty

The positive stable model is an appealing alternative when the data is not well-fitted by the gamma frailty specification [23]. In contrast to the gamma model, which exhibits time-invariant predictive hazard ratios [25], this model's predictive hazard ratio drops to 1 with time [26]. The pdf of positive stable distribution is given by:

$$f(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{k!} (-z^{-\alpha}\theta/\alpha)^k \sin(\alpha k\pi)$$

where $0 < \alpha < 1$, $z > 0$, and $\theta > 0$. α and θ are unknown parameters.

The mean of this positive stable distribution is infinite, therefore making its variance undetermined. To address the identifiability problem on the parameters we restrict $\alpha = \theta$. Then, the pdf becomes:

$$f(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta+1)}{k!} (-z)^{(k\theta)} \sin(\theta k\pi),$$

where $0 < \theta < 1$, $z > 0$. The corresponding Laplace transform is:

$$\mathcal{L}(s) = \exp(-s^\theta)$$

3) Power Variance Function frailty

The PVF distribution was proposed by [27] and [24]. Gamma, Inverse Gaussian, and Positive Stable distributions are its special cases, making it a three-parameter family. We denote the PVF distribution by PVF $(\alpha; \delta; \theta)$. According to [24] the pdf is given as:

$$f(z) = -\exp(-\theta z + \delta\theta^\alpha/\alpha) \frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha+1)}{k!} (-z^{-\alpha}\delta/\alpha)^k \sin(\alpha k\pi)$$

when $0 < \alpha < 1$, $\delta > 0$, and $\theta \geq 0$.

We set $\delta = \theta^{1-\alpha}$ corresponding to $\mathbf{E}(Z) = \delta\theta^{\alpha-1} = 1$ for identifiability. Therefore, the Laplace transform is given as:

$$\mathcal{L}(s) = \exp\left[\theta \left\{1 - (1 + s/\theta)^\alpha\right\}/\alpha\right]$$

This reduces to the Laplace transform of the gamma frailty If $\alpha \rightarrow 0$ and when $\alpha = 1/2$, it reduces to the Laplace transform of inverse Gaussian frailty. This suggests that the PVF contains many other fascinating frailty models as special cases. Use of the PVF frailty model in actual practice is therefore desirable.

2.2. Estimation of Parameters

The development of R packages for the PVF distribution family has been somewhat limited. Notably, the “frailtySurv [28] [29]” and “parfm [30]” packages include most of these distributions, with the former utilizing a pseudo full likelihood approach and the latter offering fully parametric models. To enhance this suite, the “frailtyEM” package has been introduced, employing the Expectation-Maximization (EM) algorithm to fit semiparametric shared frailty models, thus providing comprehensive support for various data scenarios within the PVF distribution family.

The EM algorithm was suggested by [31] and is frequently applied when there is unobserved data. The application of the EM algorithm to analyze the survival of bees using frailty models was used because of its ability to handle unobserved heterogeneity, manage incomplete data, offer flexibility in model specification, provide computational efficiency, improve estimation accuracy, and exhibit robustness to model assumptions.

The EM algorithm iterates between two steps: The Expectation step and the Maximization step. The expectation step determines the expected values of the unobserved frailties conditional on the observed data and the current parameter estimates. The expected values from the E-step are taken into account as the actual information in the maximization step, and new estimates of the parameters of interest are then derived by maximizing the likelihood given the expected values.

The EM algorithm for Semiparametric frailty models is described by [32]. We first consider the complete data log-likelihood with the frailties assumed to be observed randomly variables. The full data log-likelihood follows from the joint density of y and z , with y containing the observed times t_{ij} (the minimum between the failure and right-censoring time) and the censoring indicators δ_{ij} . With $\lambda_0(\cdot)$ the unspecified baseline hazard function and observing the z_i 's we have:

$$l_{full}(\theta, \beta, \Lambda_0) = l_{full,1}(\lambda_0(\cdot), \beta) + l_{full,2}(\theta) \tag{3}$$

with:

$$l_{full,1}(\lambda_0(\cdot), \beta) = \sum_{i=1}^s \sum_{j=1}^{n_i} \left[\delta_{ij} \log(\lambda_0(t_{ij}) z_i \exp(x_{ij}^T \beta)) - \Lambda_0(t_{ij}) z_i \exp(x_{ij}^T \beta) \right] \tag{4}$$

as the log-likelihood of y conditional on the frailties, which is used to estimate β and λ_0 . And;

$$l_{full,2}(\theta) = \sum_{i=1}^s \log g(z_i) \tag{5}$$

is used to estimate θ . The full data log-likelihood is then solved using the EM algorithm.

Log-likelihood, theoretically, can be used to compare various frailty distributions like the power variance function (PVF) family. Frailties are latent, making it difficult to predict them because the data is usually insufficient to allow for the

construction of a well-informed judgment. Within the PVF, the model with the highest log-likelihood is selected.

3. Application to Bees Data

3.1. Available Data

This study utilized secondary data from a study that was conducted to assess the effect of entomopathogenic fungi (biopesticides) on the African stingless bee (*Meliponula ferruginea* Cockrell) and the Western honey bee (*Apis mellifera* L.) [33]. The study included six entomopathogenic fungi namely five isolates of *Metarhizium anisopliae* (ICIPE 7, ICIPE 20, ICIPE 62, ICIPE 69, and ICIPE 78), and one isolate of *Beauveria bassiana* (ICIPE 284) plus control consisting of unexposed bees. The study was conducted to evaluate the safety of entomopathogenic fungi which are globally being used as biopesticides for pest control as an alternative to harmful chemicals. The rationale for the study is that conidia of entomopathogenic fungi applied on flowering plants could be unsafe for forager bees hence the need to assess the nontarget effect of biopesticides on honey bees and stingless bees. In the study, young bees of the same age were exposed to 1×10^8 conidia/ml using a micro-spray tower before placing them in cages to monitor survival under laboratory conditions. The bees were randomly allocated to 111 cages (clusters) with 25 or 30 bees per cage depending on the availability of the bees. The bees were monitored every 24 hours over 10 days. The 10-day limit was set because literature indicates that the confinement of small groups of bees secluded from their queen under artificial conditions renders them stressful. Besides, the biopesticides are only active from 3 days up to 14 days. The outcome of interest was the time to death, which was recorded for each bee in the cages (clusters). The survival data were collected by monitoring the bees every 24 hours over the 10-day period. Bees were considered dead if they exhibited no movement when prodded with a fine brush. If a bee was still alive at the end of the 10-day period, it was censored from the analysis. Out of a total of 3080 observations, 2443 (79.8%) were censored. More details on study methodology can be found in [33].

3.2. Exploratory Data Analysis

Figure 1 shows the Kaplan-Meier estimates of the overall population (black) compared to each cluster (gray). It was evident that each cluster differed from each other and the overall survival curve. However, the cluster-specific survival curves are not controlled for potential variations in the covariate distribution across clusters, and hence the need to account for the unobserved variation.

3.3. Parameter Estimates Comparison

The parameter estimates for the Cox model and the Gamma, Positive stable, and Inverse Gaussian frailty models are given in **Table 1**. The Hazard Ratio (HR) for all the covariates was close for the four models but slightly higher for the Cox

Table 1. Hazard ratio (exp(coef)) and standard error estimates (SE) from the Cox model, the gamma, positive stable, and inverse Gaussian frailty models.

Factor	Cox Model		Gamma		Positive stable		Inverse Gaussian	
	exp (coef)	SE	exp(coef)	SE	exp(coef)	SE	exp(coef)	SE
Bees' species vs Apis								
Meliponula	0.55*	0.08	0.54*	0.10	0.52*	0.09	0.55*	0.10
Treatment vs control								
ICIPE 20	2.64*	0.17	2.58*	0.19	2.53*	0.19	2.57*	0.19
ICIPE 284	1.43	0.19	1.42	0.21	1.44	0.20	1.42	0.21
ICIPE 62	1.82*	0.18	1.82*	0.20	1.81*	0.19	1.82*	0.20
ICIPE 69	2.45*	0.17	2.46*	0.19	2.47*	0.18	2.47*	0.19
ICIPE 7	2.66*	0.17	2.62*	0.19	2.62*	0.18	2.62*	0.19
ICIPE 78	1.83*	0.18	1.83*	0.20	1.71*	0.20	1.83*	0.20

*Significant estimate ($p < 0.001$).

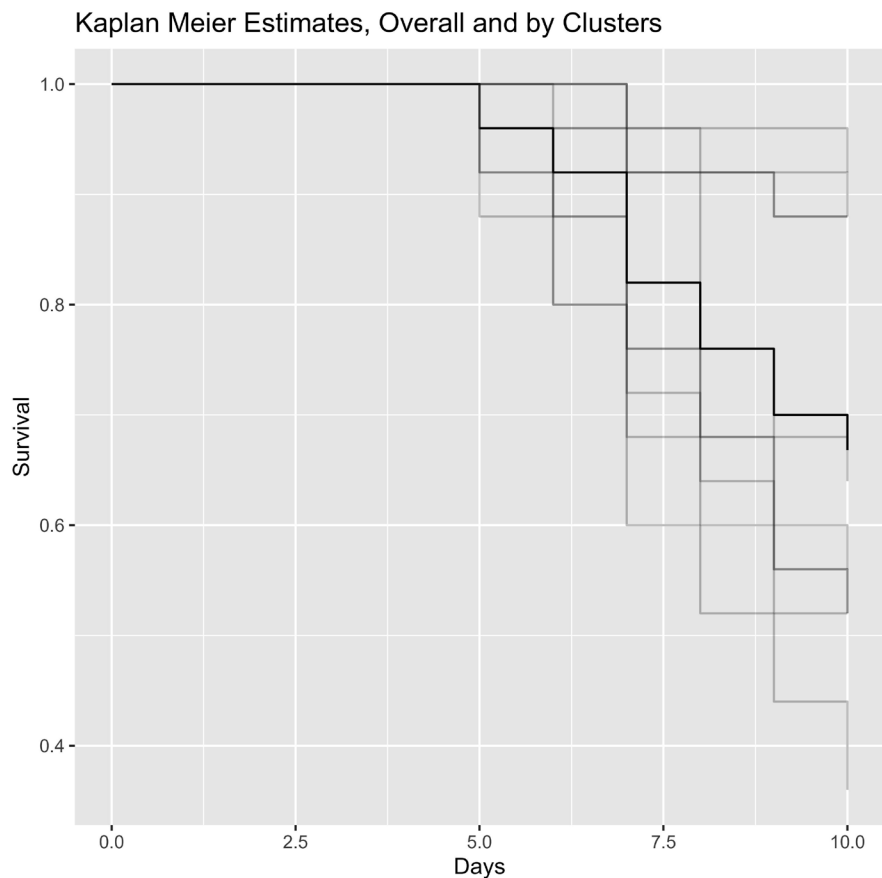


Figure 1. Kaplan Meier estimates corresponding to each cluster (grey) and the overall estimate (black).

model. The similarities between the Cox model and shared frailty models could be explained by the fact that the frailty variance for the frailty models was quite

low, that is, 0.067 and 0.065 for gamma and inverse Gaussian frailty models respectively as shown in **Table 2**. The variance for the positive stable distribution is undefined. However, it is easy to obtain Kendall’s τ , and in this case, it is lower than in the gamma frailty model. The positive stable frailty predicts marginal model with proportional hazards where the marginal hazard ratios are an attenuated version of the conditional hazard ratios shown in the output. Although the frailty variance was small, the differences in mortality rates between clusters were statistically significant (Commenges-Andersen test for heterogeneity, $p < 0.05$ and likelihood ratio test, $p < 0.05$) for all frailty models.

Sensitivity Analysis

We conducted a sensitivity analysis to assess the robustness of the parameter estimates. It’s worth noting that the frailtyEM package does not have a direct function for sensitivity analysis. Therefore, we manually conducted the sensitivity analysis, which included two components: bias analysis and variance analysis.

1) Bias Analysis

For the bias analysis, we perturbed the parameter estimates by adding random noise to the original estimates. We then re-estimated the models multiple times and calculated the average bias across simulations. The perturbation was based on the standard errors of the parameter estimates. The average bias across simulations is presented in **Table 3**.

2) Variance Analysis

Table 2. Frailty summary of gamma, positive stable and inverse gaussian frailty models; theta is the heterogeneity parameter, LRT is the Likelihood ratio test, and ca test is the Commenges-Andersen test.

	Gamma		Positive Stable		Inverse Gaussian	
	Estimate	95% C.I	Estimate	95% C.I	Estimate	95% C.I
Var(Z)	0.067	[0.006, 0.158]	—	—	0.065	[0.006, 0.160]
Kendall’s τ	0.032	[0.03, 0.073]	0.021	[0.062, 1.000]	0.030	[0.003, 0.065]
theta	14.963	[6.330, 159.564]	45.783	[15.246, 3835.841]	15.317	[6.235, 169.560]
LRT	0.0132		0.0226		0.0137	
ca test	0.026		0.026		0.026	

Table 3. Average bias in parameter estimates.

Factor	Cox Model	Gamma	Positive Stable	Inverse Gaussian
Meliponula	-0.001	0.002	0.001	0.000
ICIPE 20	-0.003	0.001	0.001	-0.001
ICIPE 284	-0.001	0.000	0.000	0.001
ICIPE 62	0.000	-0.001	-0.001	0.000
ICIPE 69	-0.001	-0.001	0.001	0.000
ICIPE 7	-0.001	-0.001	0.000	0.000
ICIPE 78	0.001	-0.002	0.002	0.000

For the variance analysis, we resampled the parameter estimates and calculated the variability across simulations. The average variance across simulations is presented in **Table 4**.

The sensitivity analysis indicates that the parameter estimates obtained from the Cox and frailty models are robust to perturbations in the input data. The average bias and variance across simulations are small, indicating that the estimates are unbiased and consistent.

3.4. Comparing the Performance of the Fitted Semi-Parametric Frailty Models

3.4.1. Visualization of the Marginal and Conditional Cumulative Hazard from Different Frailty Distributions

Predicted cumulative hazard curves of two selected individuals, one from the *Apis* species and another from the *Meliponula* species, subjected to ICIPE 20 treatment are displayed in **Figure 2**. The marginal cumulative hazard of the gamma and inverse gaussian frailty models appeared as a “dragged down” version of their conditional cumulative hazards. As described in [34], this is a well-known effect observed in frailty models. If the frailty variance is larger, this selection effect is stronger. Particularly, depending on the variance of the frailty, the marginal hazard may seem to increase, peak, and then decrease beyond a certain time, even if the conditional hazard is increasing.

This is however different for the positive stable frailty model. Unlike other distributions, which imply non-proportional risks at the marginal level, the positive stable distribution implies proportional hazards both unconditionally and conditionally on frailty. Therefore, this is the only distribution in which the frailty effect does not obscure the potential violation of the proportional hazards. On average, individuals with higher hazards die earlier than those with lower hazards. This is particularly true for all frailty distributions.

3.4.2. Visualization of the Conditional and Marginal Hazard Ratios from Different Distributions

Figure 3 shows the predicted hazard ratio curves of two selected individuals, one from the *Apis* Species and another from the *Meliponula* species, both subjected to ICIPE 20 treatment.

Table 4. Average variance in parameter estimates.

Factor	Cox Model	Gamma	Positive Stable	Inverse Gaussian
Meliponula	0.0001	0.0003	0.0002	0.0002
ICIPE 20	0.0002	0.0001	0.0002	0.0002
ICIPE 284	0.0001	0.0001	0.0001	0.0002
ICIPE 62	0.0001	0.0002	0.0002	0.0002
ICIPE 69	0.0001	0.0002	0.0001	0.0002
ICIPE 7	0.0001	0.0002	0.0001	0.0001
ICIPE 78	0.0002	0.0002	0.0002	0.0002

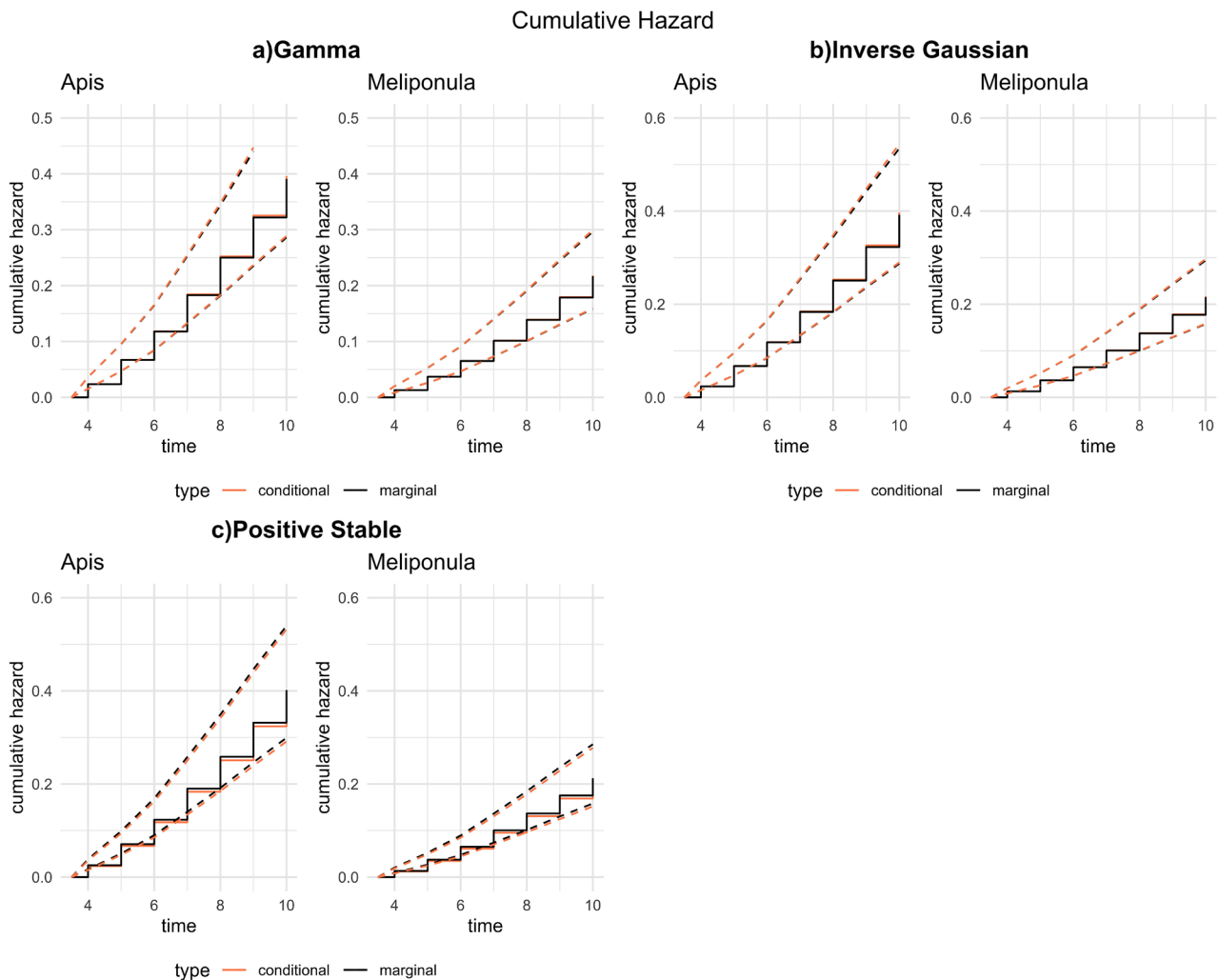


Figure 2. Predicted conditional and marginal cumulative hazard for (a) Gamma, (b) Positive stable, and (c) Inverse gaussian frailty models for Apis species and Meliponula species treated with ICIPE 20 entomopathogenic fungi.

The gamma and inverse gaussian frailty models shrink the hazard ratio towards 1. However it can be seen that this effect is slightly more pronounced for the gamma frailty model. The positive stable frailty model exhibits constant “average” shrinkage. This is because unlike the PVF distributions, it predicts both the marginal and conditional models with proportional hazards. The marginal hazard ratios are an attenuated version of the conditional hazard ratios shown in the output. This type of behavior from the positive stable is often seen as a strength of the model [23].

3.4.3. Profile Log Likelihood

The results in **Table 5** show the heterogeneity parameter θ was significant in all three frailty models. Despite the estimates of θ being somewhat small, it suggested the failure times within the cages were significantly correlated, that is, $\theta > 0$. The gamma frailty model had the highest log-likelihood and hence was the best-fitting model among the three.

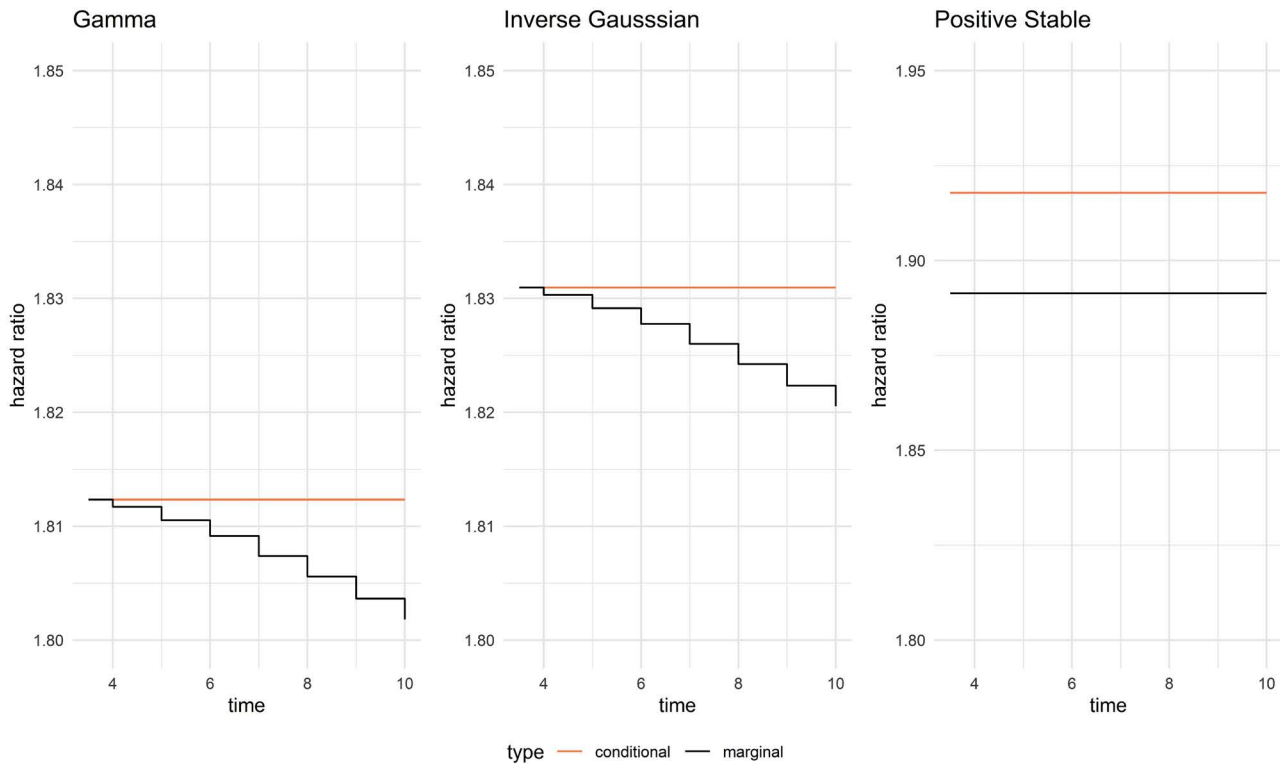


Figure 3. Predicted conditional and marginal hazard ratio for (a) Gamma, (b) Positive stable, and (c) Inverse Gaussian frailty models for *Apis* species and *Meliponula* species treated with ICIPE 20 entomopathogenic fungi.

Table 5. Profile log-likelihood for Gamma, Positive stable and Inverse Gaussian Frailty models.

	Association parameter	P-value	No-frailty Log-likelihood	Log-likelihood
Gamma	0.067	0.0132	-5007.947	-5005.483
Positive Stable	0.021	0.0226	-5007.947	-5005.942
Inverse Gaussian	0.065	0.0137	-5007.947	-5005.511

4. Simulation Study

A simulation study was performed to investigate the effect of frailty distribution misspecification on the regression coefficient and heterogeneity parameters. The bias and mean squared error (MSE) of the regression parameter (treatment log hazard, $\hat{\beta}$) and the heterogeneity parameter (θ) estimates around the true initial values were estimated.

We varied the total sample size ($N = 100, 400$), the number of individuals in each cluster ($n_i = 10, 20$) and heterogeneity parameter ($\theta = 0.2, 2$) to examine the effect of changing the variance of the frailty distributions. Frailties were generated from four distributions: gamma, log-normal, inverse Gaussian and positive stable densities. The log-normal distribution which according to [15] is expressed as:

$$f(z) = \frac{1}{z\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log z - \mu)^2}{2\sigma^2}\right)$$

To ensure a log-normal distribution with mean 1 and variance θ , we set $\mu = -\log(\theta+1)/2$ and $\sigma^2 = \theta+1$. The mean and variance are expressed as:

$$E(Z) = \exp\left(\frac{1}{2}(\sigma^2 - \mu)\right) = 1$$

$$\text{Var}(Z) = \frac{\exp(2\sigma^2) - \exp(\sigma^2)}{\sigma^2} = \theta$$

To allow for comparability, the mean and variance were fixed to 1 and θ respectively for the four distributions.

In all settings, the treatment effect parameter was set to $-\log(\beta) = 0.5$, so that the treatment group's risk rate is twice as high as the risk rate for individuals in the control group. Half of the individuals in each cluster were assigned to treatment and the censoring rate was 40%. For each parameter setting 1000 datasets were generated where N was the fixed sample size. The simulated data were fitted with frailty models. The spread and the bias around were determined using the mean squared error (MSE). The bias and mean squared error are respectively defined as:

$$\text{Bias}(\beta) = \bar{\beta} - \beta_1$$

and

$$\text{MSE}(\beta) = \text{Bias}^2(\beta) + \text{Var}(\hat{\beta})$$

where β_1 is the true value of log hazard treatment effect and $\bar{\beta} = \sum_i \hat{\beta}^i / 1000$ is the mean of $\hat{\beta}^i$'s estimated in the i^{th} simulation.

Similarly, the bias and mean squared error (MSE) for the heterogeneity parameter were obtained as above replacing accordingly.

Statistical analysis was conducted using R version 4.1.3. The semi-parametric Cox marginal and frailty models were fitted using *emfrail* procedure in frailtyEM package [35]. Statistical tests were performed at a 5% level of significance.

4.1. Simulation Results for Comparing the Performance of Different Frailty Models

The results of the simulations are summarized in **Table 6**. The results show that the models performed well in estimating the regression parameter. This suggests that the choice of frailty distribution does not have an effect on the regression parameter.

For the heterogeneity parameter θ , it was observed that the bias range for the gamma frailty model was between 0.4349 and 1.3651 and between 0.0809 and 6.2323 for the inverse gaussian frailty model. For a 10-cluster scenario, the bias decreased with increasing size of true θ . Similarly, this trend was observed with 5 clusters scenario. Furthermore, for a particular true θ , the bias increased when the number of clusters decreased from 10 to 5. The mean squared error (MSE) was increased with an increase in the magnitude of true θ and a decrease in the number of clusters for the gamma frailty model. For the inverse gaussian frailty

Table 6. Simulation results for estimated regression coefficient (β) and heterogeneity parameter (θ) from correctly specified gamma, inverse gaussian and positive stable frailty models. i = number of clusters and n_i = cluster size.

	$i = 20, n_i = 5$		$i = 10, n_i = 10$		$i = 20, n_i = 20$		$i = 10, n_i = 40$	
	$\theta = 0.2$	$\theta = 2$	$\theta = 0.2$	$\theta = 2$	$\theta = 0.2$	$\theta = 2$	$\theta = 0.2$	$\theta = 2$
<u>Gamma</u>								
β								
Bias	0.0033	0.0033	0.0404	0.0404	0.0129	0.0129	0.0111	0.0111
MSE	0.0902	0.0902	0.0848	0.0848	0.0197	0.0197	0.0205	0.0205
θ								
Bias	0.1769	0.1275	0.1779	0.1415	0.1855	1.7174	0.1885	0.8544
MSE	0.0342	1.6929	0.0351	2.2926	0.0348	8.4969	2.3108	0.0359
<u>Inverse Gaussian</u>								
β								
Bias	0.0033	0.0033	0.0404	0.0404	0.0129	0.0129	0.0111	0.0111
MSE	0.0902	0.0902	0.0848	0.0848	0.0197	0.0197	0.0205	0.0205
θ								
Bias	0.5301	1.2699	0.4349	1.3651	0.6250	1.1749	0.5082	1.2918
MSE	0.6603	1.9919	0.3618	2.0361	0.5285	1.5184	0.3200	1.7306
<u>Positive Stable</u>								
β								
Bias	0.0124	0.0124	0.0359	0.0359	0.0055	0.0055	0.0076	0.0076
MSE	0.0742	0.0742	0.0863	0.0863	0.0256	0.0256	0.0369	0.0369

model, the mean squared error (MSE) was decreased with an increase in the magnitude of true θ and a decrease in the number of clusters.

The simulation results demonstrated that increasing the cluster size and decreasing the number of clusters did not affect the MSEs of the estimates of the treatment effect and the frailty variance. The results also showed that when the frailty distribution is inverse gaussian as opposed to the gamma frailty distribution, the biases in the estimations of the frailty variance are slightly larger.

By increasing the number of clusters from 5 and 10 to 20 and 40, while keeping the number of individuals in each cluster at 10 and 20, we were able to investigate the effectiveness of clusters and their influence on the estimations. The simulation's results demonstrated that MSEs for both estimates slightly decreased when compared to the original simulation's outcomes, as expected.

4.2. Effect of Frailty Misspecification on Regression Coefficient and Heterogeneity Parameter

Gamma models are selected primarily on theoretical and computational tractability rather than biological validity. Because cluster effects can have different distributions, it is crucial to assess how well a gamma frailty model performs.

A gamma frailty model was fitted to clustered data generated by log-normal, inverse gaussian, and positive stable distributions to evaluate the influence and sensitivity to misspecification of the frailty distribution on regression coefficient and heterogeneity parameter. It was important to examine the performance of a gamma frailty model when the cluster effects have other distributions. **Table 7** shows the results of misspecification. Generally, the mean squared error (MSE) increased as the variance of the frailty increased. As the sample size increased, the MSE decreased in all settings (that is, when $N = 100$ and $N = 400$).

Table 7. Simulation results for estimated regression coefficient (β) and heterogeneity parameter (θ) from misspecified gamma frailty model. i = number of clusters and n_i = cluster size.

	$i = 20, n_i = 5$		$i = 10, n_i = 10$		$i = 20, n_i = 20$		$i = 10, n_i = 40$	
	$\theta = 0.2$	$\theta = 2$	$\theta = 0.2$	$\theta = 2$	$\theta = 0.2$	$\theta = 2$	$\theta = 0.2$	$\theta = 2$
Gamma								
β								
Bias	0.0033	0.0033	0.0404	0.0404	0.0129	0.0129	0.0111	0.0111
MSE	0.0902	0.0902	0.0848	0.0848	0.0197	0.0197	0.0205	0.0205
θ								
Bias	0.1769	0.1275	0.1779	0.1415	0.1855	1.7174	0.1885	0.8544
MSE	0.0342	1.6929	0.0351	2.2926	0.0348	8.4969	2.3108	0.0359
Log-normal								
β								
Bias	0.0010	0.0133	0.0011	0.0367	0.0031	0.0032	0.0053	0.0075
MSE	0.0724	0.0743	0.0827	0.0869	0.0184	0.0229	0.0185	0.0191
θ								
Bias	0.5301	1.2699	0.4349	1.3651	0.6250	1.1749	0.5082	1.2918
MSE	0.6603	1.9919	0.3618	2.0361	0.5285	1.5184	0.3200	1.7306
Inverse-Gaussian								
β								
Bias	0.0106	0.0084	0.0205	0.0173	0.0165	0.0159	0.0207	0.0132
MSE	0.0862	0.0870	0.1046	0.0977	0.0218	0.0214	0.0234	0.0218
θ								
Bias	1.1179	1.7214	1.030	1.7889	1.7383	1.7197	1.235	1.7964
MSE	2.4764	3.0827	1.8606	3.2559	4.3447	2.9825	1.8648	3.2369
Positive Stable								
β								
Bias	0.0113	0.0113	0.0227	0.0227	0.0070	0.0075	0.0335	0.0335
MSE	0.0758	0.0758	0.1236	0.1236	0.0192	0.0191	0.0226	0.0226
θ								
Bias	0.5327	2.233	0.5966	2.3966	1.7174	7.6200	3.9390	5.7390
MSE	3.0017	8.1595	3.2610	8.6488	8.4969	79.1822	23.1695	40.5900

Except for the positive stable frailty model, the MSE of θ is at the same level as that of other misspecified frailties, where the true frailty model is a gamma distribution. However, the distinction between the MSEs provided by the true gamma and other misspecified frailty models gets more pronounced as the frailty variance θ increases.

4.2.1. Regression Coefficient

The bias from this misspecified model and the correctly specified model did not vary substantially suggesting the robustness of the gamma frailty with respect to log-normal, inverse gaussian and positive stable distributions (Table 5).

Within a certain cluster scenario, *i.e.* either 5, 10, 20, or 40, the bias significantly increased with an increase in the size of the true θ . On the other hand for a particular θ , the bias increased by a small margin when the clusters increased from 5 to 10 and 20 to 40. This implied that the regression coefficient was not greatly affected by cluster size. We also note that under the misspecified models, the MSE tended to be slightly smaller than the correctly specified model.

4.2.2. Heterogeneity Parameter

Sensitivity to misspecification of the frailty distribution was evaluated about the estimated heterogeneity parameter. Moderate to high bias was observed for each of the assumed true θ . The bias was much higher for the misspecified models compared to the correctly specified frailty model. An increase in bias was observed when the true θ increased across the clusters. These findings demonstrate that the misspecified gamma frailty model did not accurately predict the underlying true heterogeneity parameter.

5. Discussion

The data utilised in this study had high percentage of right censored subjects because of some justifiable practical constraints namely the subjects (bees) were observed for a limit of 10 days because studies have shown that confinement of a small group of honey bees secluded from their queen under controlled conditions renders them stressful and consequently more vulnerable to *Beauveria bassiana* and *Metarhizium anisopliae* ([36] as cited [33]). Further, virulent fungi isolates belonging to *Beauveria bassiana* and *Metarhizium anisopliae* species have shown their pathogenicity in less than 10 days (median lethal time < 10 days) after inoculation or exposure to study organisms under laboratory conditions [37] [38]. Therefore, we ordinarily incorporated the right censoring in the analysis as all surviving subjects practically reached the end of the study.

From the statistical analysis results, parameter estimates (hazard ratio) for all the covariates were close for the three frailty models (gamma, inverse gaussian, and positive stable frailty models) but slightly higher for the marginal Cox PH model. This small difference was attributed to the fact that the cluster random effect though significant was very small (that is 0.067 for gamma, 0.065 for inverse gaussian and 0.021 Kendall's τ for positive stable frailty model). For the

positive stable distribution, the variance is not defined but Kendall's τ is easily obtained [35]. The frailty variance was notably small. However, both the one-sided likelihood ratio test and the Commenges-Andersen test for heterogeneity indicated a significant difference in mortality rates among clusters, even after adjusting for covariates. This finding was further supported by the confidence interval for the frailty variance, which excluded zero. This subtle effect is ecologically important, highlighting how minor stressors can impact bee health and ecosystem stability. Our findings underscore the complexity of bee-environment interactions and the need for ongoing research into bee conservation.

The positive stable model offers the advantage of being able to fit both proportional hazards models and Weibull models. Because of its ability to fit proportional hazards models implies that if the conditional model exhibits proportional hazards so does the marginal distribution. This aspect becomes particularly advantageous when viewing the model as a random effects model. From a theoretical standpoint, stable distributions are highly desirable as they consistently adhere to proportional hazards for the covariates. However, it is unfortunate that the fit of these distributions is unsatisfactory in several applications [15]. The stable model tends to be the most reasonable choice for regression models. On the other hand, the gamma and more general PVF models may yield peculiar outcomes when evaluating the effect of covariates. This implies that the influence of non-proportional hazards can sometimes overshadow the degree of dependence, making it crucial to interpret the results with caution.

Furthermore, alternative frailty models can be compared based on their log-likelihoods to pick a well-fitting frailty distribution from a larger class of frailty distributions. However, since the frailties are latent, choosing which frailty distribution is most appropriate requires an assessment of the frailty distribution on the data at hand. The gamma frailty model, which can be estimated by the base R survival package, is the most straightforward and well-understood frailty model [35]. From our data, the gamma frailty model had the highest log-likelihood, hence it was the best-fitting model. Other studies where the gamma frailty model performed well include a study done by Gachau [39] and Adham and AlAhmadi [40].

A simulated study further investigated the effect of misspecification of the frailty distribution on the prediction of the heterogeneity parameters and regression coefficients. We considered the true frailty distribution from gamma distribution, inverse Gaussian distribution and positive stable distribution. We focused only on the gamma distribution as the assumed frailty distribution partially because it had the highest log-likelihood and also because it is the distribution that is most frequently used in practice and its closed form of the marginal likelihood can be obtained under the Cox proportional hazard shared frailty models. Our simulation results showed that the bias of the treatment log hazard was comparatively small and similar to that of the correctly specified model. This showed that, despite the fact that different frailty distributions can result in

distinctly different correlation structures, misspecification of the frailty distribution did not significantly alter the regression coefficient estimate. [41] and [42] found by simulation that regression coefficient estimates were hardly influenced by frailty misspecification. They assumed inverse Gaussian their true frailty distribution. This shows that even when the random effects interpretation is unappealing, the gamma frailty approach is a reasonable choice in terms of performance.

In comparison to the correctly specified model, the bias was more pronounced for large heterogeneity parameter θ values. This was slightly affected by the number of clusters taken into account. The statistical simulation studies showed that when the true frailty is inverse Gaussian or log-normal frailty, the gamma frailty working model provides a satisfactory prediction of θ in terms of the MSE. When the true frailty distribution is the positive stable, it can be challenging to use a gamma frailty working model to predict specific frailties utilizing shared frailty models. These findings are in line with the research by [39].

The gamma distribution is the most widely used frailty distribution for Cox proportional hazard shared frailty models, and most statistical software utilizes it as its default frailty distribution for fitting the models. In the Cox proportional hazard shared frailty models, using a gamma frailty distribution can result in a reliable frailty prediction. However, if a working gamma frailty distribution is utilized when the underlying real frailty distribution is an extreme distribution, such as the positive stable frailty, the prediction on frailty terms should be biased.

6. Conclusions

The parameter estimates from the frailty models gamma, inverse Gaussian, and positive stable considered in this study were almost indistinguishable from those estimated from a Cox PH model. There was sufficient evidence to suggest heterogeneity between the clusters in frailty models and conclude that events were dependent within clusters and independent across clusters. Therefore, the frailty models were the most appropriate for statistical inference for these data.

The estimate of the regression coefficient (treatment effect) was only slightly biased when a gamma frailty model is assumed, while the underlying true frailty distribution is log-normal, inverse gaussian, and positive stable.

From the simulation study, we concluded that the heterogeneity parameter was more sensitive to frailty distribution misspecification compared to the regression parameter estimate. In this regard, when the regression parameters are of primary interest, the gamma frailty model can be an effective option in real data analysis. However, the frailty variable is not observable, so the distributional assumption cannot be evaluated to see if it holds true.

Availability of Data and R Code

The data file and R code used for this study can be found at Bee survival.

Acknowledgements

The authors gratefully acknowledge the financial support for this research by the following organizations and agencies: The Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Australian Centre for International Agricultural Research (ACIAR); the Norwegian Agency for Development Cooperation (Norad); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. The first author received a scholarship from the project, Combatting Arthropod Pests for better Health, Food and Climate Resilience (CAP-Africa, Project number: RAF-3058 KEN-18/0005) funded by Norwegian Agency for Development Cooperation (Norad). The views expressed herein do not necessarily reflect the official opinion of the donors.

Disclosure Statement

The authors declare no conflict of interest.

Notes on Contributor(s)

The authors confirm their contribution to the paper as follows: Conceptualization: Daisy Salifu; Methodology: Patience Isiaho, Daisy Salifu, Samuel Mwalili; Software: Patience Isiaho; Validation: Patience Isiaho, Daisy Salifu, Samuel Mwalili; Data Curation: Patience Isiaho; Formal Analysis: Patience Isiaho; Writing - Original Draft: Patience Isiaho; Writing - Review & Editing: Patience Isiaho, Daisy Salifu, Samuel Mwalili, Henri E.Z. Tonnang; Supervision: Daisy Salifu, Samuel Mwalili, Henri E.Z. Tonnang; Funding acquisition: Daisy Salifu, Henri E.Z. Tonnang. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Buchmann, A., Moore, K. and Fisher, D. (2010) EXPERIENCING FILM TOURISM: Authenticity Fellowship. *Annals of Tourism Research*, **37**, 229-248. <https://doi.org/10.1016/j.annals.2009.09.005>
- [2] Brittain, C.A., Vighi, M., Bommarco, R., Settele, J. and Potts, S.G. (2010) Impacts of a Pesticide on Pollinator Species Richness at Different Spatial Scales. *Basic and Applied Ecology*, **11**, 106-115. <https://doi.org/10.1016/j.baae.2009.11.007>
- [3] Wiest, L., Buleté, A., Giroud, B., Fratta, C., Amic, S., Lambert, O., Pouliquen, H. and Arnaudguilhem, C. (2011) Multi-Residue Analysis of 80 Environmental Contaminants in Honeys, Honeybees and Pollens by One Extraction Procedure Followed by Liquid and Gas Chromatography Coupled with Mass Spectrometric Detection. *Journal of Chromatography A*, **1218**, 5743-5756. <https://doi.org/10.1016/j.chroma.2011.06.079>

- [4] Ndakidemi, B., Mtei, K. and Ndakidemi, P.A. (2016) Impacts of Synthetic and Botanical Pesticides on Beneficial Insects. *Agricultural Sciences*, **7**, 364-372. <https://doi.org/10.4236/as.2016.76038>
- [5] Kumar, A., Joseph, S., Tsechansky, L., Privat, K., Schreiter, I.J., Schüth, C. and Graber, E.R. (2018) Biochar Aging in Contaminated Soil Promotes Zn Immobilization Due to Changes in Biochar Surface Structural and Chemical Properties. *Science of the Total Environment*, **626**, 953-961. <https://doi.org/10.1016/j.scitotenv.2018.01.157>
- [6] Shah, P.A. and Pell, J.K. (2003) Entomopathogenic Fungi as Biological Control Agents. *Applied Microbiology and Biotechnology*, **61**, 413-423. <https://doi.org/10.1007/s00253-003-1240-8>
- [7] Boff, S., Scheiner, R., Raizer, J. and Lupi, D. (2021) Survival Rate and Changes in Foraging Performances of Solitary Bees Exposed to a Novel Insecticide. *Ecotoxicology and Environmental Safety*, **211**, Article 111869. <https://doi.org/10.1016/j.ecoenv.2020.111869>
- [8] Parish, J.B., Scott, E.S., Correll, R. and Hogendoorn, K. (2019) Survival and Probability of Transmission of Plant Pathogenic Fungi through the Digestive Tract of Honey Bee Workers. *Apidologie*, **50**, 871-880. <https://doi.org/10.1007/s13592-019-00697-6>
- [9] Van Dooremalen, C., Gerritsen, L., Cornelissen, B., van der Steen, J.J.M., van Langevelde, F. and Blacquière, T. (2012) Winter Survival of Individual Honey Bees and Honey Bee Colonies Depends on Level of Varroa destructor Infestation. *PLOS ONE*, **7**, e36285. <https://doi.org/10.1371/journal.pone.0036285>
- [10] Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. and Sylvester, R. (2002) The Shared Frailty Model and the Power for Heterogeneity Tests in Multi-center Trials. *Computational Statistics Data Analysis*, **40**, 603-620. [https://doi.org/10.1016/S0167-9473\(02\)00057-9](https://doi.org/10.1016/S0167-9473(02)00057-9)
- [11] Clayton, D.G. (1978) A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, **65**, 141-151. <https://doi.org/10.1093/biomet/65.1.141>
- [12] Vaupel, J.W., Manton, K.G. and Stallard, E. (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, **16**, 439-454. <https://doi.org/10.2307/2061224>
- [13] Gorfine, M. and Zucker, D.M. (2023) Shared Frailty Methods for Complex Survival Data: A Review of Recent Advances. *Annual Review of Statistics and Its Application*, **10**, 51-73. <https://doi.org/10.1146/annurev-statistics-032921-021310>
- [14] Hsu, L., Gorfine, M. and Malone, K. (2007) On Robustness of Marginal Regression Coefficient Estimates and Hazard Functions in Multivariate Survival Analysis of Family Data When the Frailty Distribution Is Mis-Specified. *Statistics in Medicine*, **26**, 4657-4678. <https://doi.org/10.1002/sim.2870>
- [15] Duchateau, L. and Janssen, P. (2008) The Frailty Model. Springer Science & Business Media, New York.
- [16] Wienke, A. (2010) Frailty Models in Survival Analysis. Chapman and Hall/CRC, New York. <https://www.taylorfrancis.com/books/9781420073911> <https://doi.org/10.1201/9781420073911>
- [17] Balan, T.A. and Putter, H. (2020) A Tutorial on Frailty Models. *Statistical Methods in Medical Research*, **29**, 3424-3454. <http://journals.sagepub.com/doi/10.1177/0962280220921889>
- [18] Li, Y., Dick, A.W., Glance, L.G., Cai, X. and Mukamel, D.B. (2007) Misspecification

- Issues in Risk Adjustment and Construction of Outcome-Based Quality Indicators. *Health Services and Outcomes Research Methodology*, **7**, 39-56.
<https://doi.org/10.1007/s10742-006-0014-z>
- [19] Goethals, K., Ampe, B., Berkvens, D., *et al.* (2009) Modeling Interval-Censored, Clustered Cow Udder Quarter Infection Times through the Shared Gamma Frailty Model. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 1-14.
<https://doi.org/10.1198/jabes.2009.0001>
- [20] Yavuz, A. and Lambert, P. (2016) Semi-Parametric Frailty Model for Clustered Interval-Censored Data. *Statistical Modelling*, **16**, 360-391.
<https://doi.org/10.1177/1471082X16655631>
- [21] Zike, D.T., Fenta, H., Workie, D. and Swain, P. (2018) Determinants of Under-Five Mortality in Ethiopia: An Application of Cox Proportional Hazard and Frailty Models. *Biostatistics*, **10**, 123-136. <https://doi.org/10.5336/biostatic.2018-60550>
- [22] Cox, D.R. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**, 187-202.
<https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- [23] Hougaard, P. (2000) Analysis of Multivariate Survival Data. Springer New York, New York. <https://doi.org/10.1007/978-1-4612-1304-8>
- [24] Hougaard, P. (1986) Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika*, **73**, 387-396.
<https://doi.org/10.1093/biomet/73.2.387>
- [25] Fine, J.P., Glidden, D.V. and Lee, K.E. (2003) A Simple Estimator for a Shared Frailty Regression Model. *Journal of the Royal Statistical Society: Series B (Methodological)*, **65**, 317-329. <https://doi.org/10.1111/1467-9868.00388>
- [26] Oakes, D. (1989) Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*, **84**, 487-493.
<https://doi.org/10.1080/01621459.1989.10478795>
- [27] Tweedie, M.C.K. (1984) An Index Which Distinguishes between Some Important Exponential Families. In: Ghosh, J.K. and Roy, J., Eds., *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, Indian Statistical Institute, Calcutta, 579-604.
- [28] Gorfine, M., Zucker, D.M. and Hsu, L. (2006) Prospective Survival Analysis with a General Semiparametric Shared Frailty Model: A Pseudo Full Likelihood Approach. *Biometrika*, **93**, 735-741. <https://doi.org/10.1093/biomet/93.3.735>
- [29] Monaco, J.V., Gorfine, M. and Hsu, L. (2017) FrailtySurv: General Semi-Parametric Shared Frailty Model. R Package Version, 1.
- [30] Munda, M., Rotolo, F. and Legrand, C. (2012) Parfm: Parametric Frailty Models in R. *Journal of Statistical Software*, **51**, 1-20. <https://doi.org/10.18637/jss.v051.i11>
- [31] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 1-22.
<https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- [32] Klein, J.P. (1992) Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm. *Biometrics*, **48**, 795-806.
<https://doi.org/10.2307/2532345>
- [33] Omuse, E.R., Niassy, S., Wagacha, J.M., Ong'Amo, G.O., Lattorff, H.M.G., Kiatoko, N., Mohamed, S.A., Subramanian, S., Akutse, K.S. and Dubois, T. (2022) Susceptibility of the Western Honey Bee *Apis mellifera* and the African Stingless Bee *Meli-*

- ponula ferruginea* (Hymenoptera: Apidae) to the Entomopathogenic Fungi *Metarhizium anisopliae* and *Beauveria bassiana*. *Journal of Economic Entomology*, **115**, 46-55. <https://doi.org/10.1093/jee/toab211>
- [34] Aalen, O.O., Borgan, Ø. and Gjessing, H.K. (2008) Survival and Event History Analysis. Springer, New York. <https://doi.org/10.1007/978-0-387-68560-1>
- [35] Balan, T.A. and Putter, H. (2019) FrailtyEM: An R Package for Estimating Semiparametric Shared Frailty Models. *Journal of Statistical Software*, **90**, 1-29. <https://doi.org/10.18637/jss.v090.i07>
- [36] Alves, S.B., Marchini, L.C., Pereira, R.M. and Baumgratz, L.L. (1996) Effects of Some Insect Pathogens on the Africanized Honey Bee, *Apis Mellifera* L. (Hym., Apidae). *Journal of Applied Entomology*, **120**, 559-564. <https://doi.org/10.1111/j.1439-0418.1996.tb01652.x>
- [37] Seiedy, M., Saboori, A., Allahyari, H., Talaie-Hassanloui, R. and Tork, M. (2010) A Laboratory Investigation on the Virulence of Two Isolates of the Entomopathogenic Fungus *Beauveria bassiana* against the Twospotted Mite, *Tetranychus urticae* (Acari: Tetranychidae). *International Journal of Acarology*, **8**, 527-532. <https://doi.org/10.1080/01647954.2010.519718>
- [38] Oreste, M., Bubici, G., Polisenio, M., Triggiani, O. and Tarasco, E. (2012) Pathogenicity of *Beauveria bassiana* (Bals.-Criv.) Vuill. and *Metarhizium anisopliae* (metschn.) sorokin against *Galleria mellonella* L. and *Tenebrio molitor* L. in Laboratory Assays. *Journal of Zoology*, **95**, 43-48.
- [39] Susan Gachau, W. (2014) Frailty Models with Applications in Medical Research: Observed and Simulated Data. Ph.D. Thesis, University of Nairobi, Nairobi. <http://erepository.uonbi.ac.ke/handle/11295/73552>
- [40] Adham, S.A. and AlAhmadi, A.A. (2016) Gamma and Inverse Gaussian Frailty Models: A Comparative Study. *International Journal of Mathematics and Statistics Invention*, **4**, e4105.
- [41] Glidden, D.V. and Vittinghoff, E. (2004) Modelling Clustered Survival Data from Multicentre Clinical Trials. *Statistics in Medicine*, **23**, 369-388. <https://doi.org/10.1002/sim.1599>
- [42] Su, Y.-R. and Wang, J.-L. (2016) Semiparametric Efficient Estimation for Shared-Frailty Models with Doubly-Censored Clustered Data. *Annals of Statistics*, **44**, 1298-1331. <https://doi.org/10.1214/15-AOS1406>