Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 18(1) 2025: 73-82
http://ejournal.undip.ac.id/index.php/media_statistika

ANALYSIS MULTILEVEL SURVIVAL DATA USING COVARIATE-ADJUSTED FRAILTY PROPORTIONAL HAZARDS MODEL

Krismona Sandelvia, Adhitya Ronnie Effendie
Department of Mathematics, Gadjah Mada University, Yogyakarta, Indonesia

e-mail: adhityaronnie@ugm.ac.id

DOI: 10.14710/medstat.18.1.73-82

Article Info:
Received: 1 June 2025
Accepted: 15 October 2025
Available Online: 16 October 2025

Keywords:
Multilevel Survival; Frailty; Covariate-Adjusted Proportional Hazards; Bayesian MCMC.

Abstract: Multilevel survival data is time-to-event data with a hierarchical or nested structure. This study aims to model the data using the Covariate-Adjusted Frailty Proportional Hazards method, which is an extension of the Cox proportional hazards model with the addition of random effects (frailty). Parameter estimation is performed using a Bayesian approach via Markov Chain Monte Carlo (MCMC). This method is applied to analyze repeated observations of Chronic Granulomatous Disease (CGD) infections, with frailty represented by the hospital and the patient. The results of the data analysis indicate that both hospital and patient frailty significantly influence the time to infection, with patient frailty having a greater effect. Additionally, the treatment variable rINF-g significantly in reducing the risk of serious infection for CGD patients by 64.44%.

1. INTRODUCTION

Survival analysis models the time-to-event outcome, widely applied across various fields (Kleinbaum & Klein, 2005). Data often features a multilevel structure (e.g., patients nested within hospitals), which induces dependence. Ignoring this hierarchy violates the independence assumption and results in biased estimates (Hox, 2002).

To address this, the frailty model introduces a random effect that accounts for unobserved heterogeneity among individuals or groups (Vaupel et al., 1979). The Cox proportional hazards model with frailty is widely used to analyze multilevel survival data, allowing for variation in risk between groups. In medical research, the frailty model is essential for differentiating between the observed effect of covariates and the unobserved heterogeneity (such as genetic factors or lifestyle) that can bias parameter estimates if ignored (Kiprotich et al., 2025; Yslas, 2025). By incorporating the frailty term, the model accurately captures the variation in risk that is not explained by the observed covariates, thus leading to more reliable inferences in survival studies (Kiprotich et al., 2025). Subsequent developments led to the Covariate-Adjusted Frailty model, where the frailty variance flexibly depends on cluster-level covariates, estimated using a Bayesian approach (Zhang et al., 2020).

Previous studies have demonstrated the effectiveness of these models. Noh et al. (2006) applied hierarchical generalized linear models to analyze kidney disease data, meanwhile, Liu et al. (2011) used a shared frailty model on organ transplant data. Zhou et al. (2015) developed a covariate-adjusted frailty proportional hazards model for breast cancer survival data in Iowa, United States, and highlighted the importance of covariate effects at the regional level.

This study aims to apply the Covariate-Adjusted Frailty Proportional Hazards model with a Bayesian approach to multilevel survival data of patients with Chronic Granulomatous Disease (CGD). This method was chosen because of its ability to handle hierarchical data structures, accommodate heterogeneity at the patient and hospital levels, and flexibly model the influence of covariates at the cluster level. It is expected that this analysis will provide a deeper understanding of the factors influencing the time to infection in CGD patients and demonstrate the advantages of the multilevel survival approach further. This study contributes by incorporating cluster-level covariates (hospital categories) into the frailty variance, extending standard applications to recurrent event data in CGD, unlike previous single-event focuses (Zhou et al., 2015).

2. LITERATURE REVIEW

2.1. Covariate-Adjusted Frailty Proportional Hazards Model

The Covariate-Adjusted Frailty Proportional Hazards model extends the Cox Proportional Hazards model by incorporating random effects (frailty) to account for unobserved heterogeneity between individuals or groups (Vaupel et al., 1979). In this model, frailty variance can be influenced by covariates at the cluster level, providing greater flexibility in multilevel survival data analysis (Zhou et al., 2015). Recent advancements include the multivariate shared truncated normal frailty model with application to medical data (Diego et al., 2025) and Bayesian multivariate survival tree approach based on three frailty models (Porndumnernsawat et al., 2025). Additionally, neural network-based frailty models have been proposed for complex correlated outcomes (Lee et al., 2025).

Multilevel survival data is time-to-event data with a nested structure, where units at the lower level (e.g., patients) are grouped at a higher level (e.g., hospitals) (Hox, 2002). For example, in health studies, patients are grouped based on the hospital where they were treated. This type of data structure requires a special model so that the analysis accommodates the dependence between observations within a group.

2.2. Specifications of the Covariate-Adjusted Frailty Proportional Hazards Model in Multilevel Survival Data

The Covariate-Adjusted Frailty PH model was used to analyze multilevel survival data with three levels, repeated observation time (level 1) nested within patients (level 2), nested within hospitals (level 3). This model makes it possible to identify unobserved heterogeneity at the patient and hospital levels, and allows for hospital-level frailty variance to be influenced by group covariates.

Let $i = 1, \dots, n_j$ denote the number of patients admitted to hospital $j = 1, \dots, J$ and $k = 1, \dots, K_{ij}$ denote the random number of observations of several failure times for patient i in hospital j, the hazard function of the frailty proportional hazard model is as follows:

$$h(t; i, j, k) = h_0(t) \exp(\boldsymbol{x}_{ujk}^T \boldsymbol{\beta} + u_{ij} + v_j) \quad (1)$$

with: h_0: baseline hazard function; $\boldsymbol{x}_{ijk}$: covariate vector for patient i in hospital j at observation k corresponding to parameter $\boldsymbol{\beta}$; $u_{ij} \sim N(0, \sigma_u^2)$: patient frailty; $v_j \sim N(0, \sigma_v^2)$: hospital frailty.

The baseline hazard $h_0(t)$ follows the Weibull distribution. This choice is preferred for its flexibility, allowing the hazard rate to increase ($\rho > 1$), decrease ($\rho < 1$), or remain constant ($\rho = 1$) over time (Kleinbaum & Klein, 2005; Carroll, 2003).

The Weibull hazard baseline function is expressed as:

$$h_0(t) = \lambda \rho t^{\rho-1} \tag{2}$$

with λ: scale parameter and ρ: shape parameter.

The survival function for the Weibull distribution is:

$$S(t) = \exp(-\lambda t^{\rho}) \tag{3}$$

Thus, the hazard function for the multilevel survival model with frailty is:

$$h_{ijk}(t) = \lambda \rho t^{\rho-1} \exp(\boldsymbol{x}_{ijk}^T \boldsymbol{\beta} + u_{ij} + v_j) \tag{4}$$

The survival function for observation k in patient i at hospital j is obtained from the hazard function above, as follows:

$$S_{ijk}(t) = \exp(-\exp(\eta_{ijk})\lambda t^{\rho}) \tag{5}$$

with $\eta_{ijk} = \boldsymbol{x}_{ijk}^T \boldsymbol{\beta} + u_{ij} + v_j$.

This survival function expresses the probability that the time to event for observation k in patient i at hospital j is greater then t, taking into account the effects of covariates, patient frailty, and hospital frailty. Frailty variance at the hospital level was adjusted for center-level covariates, specifically the proportion of hospital category per center (Z_j), reflecting genetic heterogeneity across hospitals, modeled as $\sigma_{v_j} = \exp(\alpha + \boldsymbol{\gamma}^T \boldsymbol{z}_j)$. Parameters α and $\boldsymbol{\gamma}^T$ were estimated using Bayesian MCMC, with priors $\alpha \sim N(0,1)$ and $\boldsymbol{\gamma}^T \sim N(0,1)$.

In the proportional hazards model, the effect of covariates on hazard is measured using the hazard ratio (HR). The hazard ratio is calculated as:

$$HR = \exp(\beta) \tag{6}$$

where β is the regression coefficient of the relevant covariate. The value HR < 1 indicates a decrease in risk, while HR > 1 indicates an increase in risk (Kleinbaum & Klein, 2005).

2.3. Estimation of Covariate-Adjusted Frailty Proportional Hazards Model Parameters

Parameter estimation in the three-level Covariate-Adjusted Frailty Proportional Hazards model was performed using a Bayesian approach. This approach was chosen because of the complex structure of the model, involving frailty at the patient and hospital levels, making the classical approach difficult to apply.

2.3.1 Likelihood

The likelihood function for all three levels of data is formulated as follows:

$$\begin{aligned} &L(\beta, \lambda, \rho, \{u_{ij}\}, \{v_j\}) \\ &= \prod_{j=1}^{J} \prod_{i=1}^{n_j} \prod_{k=1}^{K_{ij}} \left[\lambda \rho t_{ijk}^{\rho-1} \exp(\eta_{ijk})\right]^{\delta_{ijk}} \exp(-\exp(\eta_{ijk})\lambda t_{ijk}^{\rho}) \end{aligned} \tag{7}$$

In the Bayesian approach, the likelihood function is used to form a posterior distribution together with the prior.

2.3.2 Prior and Posterior Distributions

In the Bayesian approach, each model parameter is given a prior distribution. The prior used for parameters $(\boldsymbol{\beta}, \lambda, \rho, \{u_{ij}\}, \{v_j\})$ are as follows:

(1) Prior for Regression Coefficient $(\boldsymbol{\beta})$

The regression coefficient $\boldsymbol{\beta}$ is assumed to follow a multivariate normal distribution, which is:

$$\pi(\boldsymbol{\beta}) = \frac{1}{(2\pi\tau^2)^{p/2}} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right) \tag{8}$$

where p is the dimension of the parameter vector $\boldsymbol{\beta}$, which is the number of regression coefficients in that vector.

(2) Prior for Individual Frailty (u_{ij})

Individual frailty u_{ij} is assumed to be normal distribution, i.e., $u_{ij} \sim N(0, \sigma_u^2)$. The standard deviation of individual frailty σ_u follows a positive Cauchy distribution with scale 2,5 ($\sigma_u \sim \text{Cauchy}^+(0\,, 2.5)$). The prior distribution for frailty u_{ij} and σ_u can be written as:

$$\pi(u_{ij}|\sigma_u) = \prod_{i=1}^{n_j}\prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{u_{ij}^2}{2\sigma_u^2}\right) \tag{9}$$

$$\pi(\sigma_u) = \frac{2}{\pi \cdot 2{,}5\left(1 + \left(\frac{\sigma_u}{2{,}5}\right)^2\right)}, \sigma_u > 0 \tag{10}$$

(3) Prior for Cluster Frailty (v_j)

The cluster frailty v_j is also assumed to be normal distributed, i.e., $v_j \sim N(0, \sigma_{v_j}^2)$, with $\sigma_{v_j} = \exp(\alpha + \boldsymbol{\gamma}^T \boldsymbol{z}_j)$, $(\alpha \sim N(0,1), (\boldsymbol{\gamma}^T \sim N(0,1))$ The prior distribution for frailty v_j and σ_{v_j} can be written as:

$$\pi(v_j|\alpha, \gamma) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma_{v_j}^2}} \exp\left(-\frac{v_j^2}{2\sigma_{v_j}^2}\right) \tag{11}$$

with $$\pi(\alpha) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2}\right) \tag{12}$$

and $$\pi(\boldsymbol{\gamma}) = \frac{1}{(2\pi)^{q/2}} \exp\left(-\frac{1}{2}\boldsymbol{\gamma}^T\boldsymbol{\gamma}\right) \tag{13}$$

where q is the dimension of the vector $\boldsymbol{\gamma}$ (which corresponds to the number of cluster-level covariates).

(4) Prior for Scale Parameter (λ)

The scale parameter λ of the Weibull distribution is assumed to follow a Gamma distribution, i.e., $\lambda \sim \text{Gamma}(a_\lambda, b_\lambda)$.

(5) Prior for Shape Parameter (ρ)

The shape parameter ρ of the Weibull distribution is also assumed to follow the Gamma distribution, i.e., $\rho \sim \text{Gamma}(a_\rho, b_\rho)$.

Based on Bayes' Theorem (Berger, 1985), the posterior distribution of the model parameters combines the prior distribution and the likelihood function, as follows:

$$(\pi(\boldsymbol{\beta},\{u_{ij}\},\{v_j\},\lambda,\rho,\sigma_u,\alpha,\boldsymbol{\gamma}|data) \propto L(data|\boldsymbol{\beta},\{u_{ij}\},\{v_j\},\lambda,\rho)\cdot\ \pi(\boldsymbol{\beta}) \quad (14)$$
$$\propto L(data|\boldsymbol{\beta},\{u_{ij}\},\{v_j\},\lambda,\rho)\cdot\ \pi(\beta)$$

By substituting the likelihood function and prior distribution of all parameters into equation (13), the joint posterior distribution is:

$$\begin{aligned}(\pi(\boldsymbol{\beta},\{u_{ij}\},\{v_j\},\lambda,\rho,\sigma_u,\alpha,\boldsymbol{\gamma}|data)&\\ \propto \prod_{j,i,k}\left[\lambda\rho t_{ijk}^{\rho-1}\exp(\eta_{ijk})\right]^{\delta_{ijk}}&\exp\left(-\exp\left(\eta_{ijk}\lambda t_{ijk}^{\rho}\right)\right)\\ \cdot\exp\left(-\frac{1}{2\tau^2}\,\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{\beta}\right)\cdot\prod_{i,j}\frac{1}{\sqrt{\sigma_u^2}}&\exp\left(-\frac{u_{ij}^2}{2\sigma_u^2}\right)\cdot\frac{1}{1+\left(\frac{\sigma_u}{2{,}5}\right)^2}\\ \cdot\prod_{j=1}^{J}\frac{1}{\sqrt{\sigma_v^2}}&\exp\left(-\frac{v_j^2}{2\sigma_{v_j}^2}\right)\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\alpha^2}{2}\right)\cdot\pi(\boldsymbol{\gamma})\\ =\frac{1}{(2\pi)^{q/2}}&\exp\left(-\frac{1}{2}\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{\gamma}\right)\cdot\lambda^{a_\lambda-1}e^{-b_\lambda\lambda}\cdot\rho^{a_\rho-1}e^{-b_\rho\rho}\end{aligned} \quad (15)$$

Since the posterior distribution has a complex analytical form, posterior sampling is performed using the Markov Chain Monte Carlo (MCMC) numerical method.

2.3.3 Estimation Process with Markov Chain Monte Carlo (MCMC)

Model parameter estimation is performed using the Markov Chain Monte Carlo (MCMC) method. In general, the MCMC steps for posterior sampling are as follows:

1. Initialize all model parameters with initial values $\boldsymbol{\beta}^{(0)}, u_{ij}^{(0)}, v_j^{(0)}, \lambda^{(0)}, \rho^{(0)}, \sigma_u^{(0)}, \alpha^{(0)}, \boldsymbol{\gamma}^{(0)}$.
2. For each iteration t, update the parameters alternately $(\boldsymbol{\beta}^{(t)}, u_{ij}^{(t)}, v_j^{(t)}, \lambda^{(t)}, \rho^{(t)}, \sigma_u^{(t)}, \alpha^{(t)}, \boldsymbol{\gamma}^{(t)})$
3. Repeat step two for many iterations until the Markov chain reaches convergence.
4. Discard a number of initial iterations (burn-in) to eliminate the influence of initial values and use samples after burn-in for parameter inference. The final estimate is obtained from the posterior sample after burn-in.

3. MATERIAL AND METHOD

3.1. Data

This study uses secondary data from a study on chronic granulomatous disease (CGD) by Fleming and Harrington (1991) obtained from the R software in the survival package. The dataset includes 128 patients across 13 hospitals over a period of one year. Each hospital had between 4 and 26 patients. The data has a three-level structure: repeated event observations (level 1), patients (level 2), and hospitals (level 3). The outcome is time to serious infection (tstop - tstart). Data were right-censored at the end of the one-year follow-up period or upon loss to follow-up, with approximately 63% censored observations assumed to be non-informative (independent of the event risk).

3.2. Research Variables

This study involved two types of variables, independent variables and dependent variables. There are six independent variables used in this study: patient ID, center (hospital number), treatment (category of treatment, either gamma interferon (γ-INF) or placebo), sex (category of gender), age (category of patient age), hospital category (cluster-level covariate: US:NIH, US:other, Europe:Amsterdam, Europe:other) , and status (indicating censoring, where 1 = uncensored data and 0 = censored data). The dependent variable in this study is the time to recurrence of survival in infected patients. In this case, t-start indicates the start of the time interval, and t-stop indicates the end of the time interval.

Table 1. Baseline Characteristics by Treatment

Variable	γ-IFN	Placebo
Age (mean, SD)	14.3 (10.1)	15.0 (9.6)
Sex (Male, %)	81	81.5
Hospital Category		
US:NIH	15	11
US:other	31	32
Europe:Amsterdam	9	10
Europe:other	8	12

Table 1 presents baseline characteristics of CGD patients by treatment group (γ-IFN vs. placebo). The mean age is 14.3 years in the γ-IFN group and 15.0 years in the placebo group. The percentage of male patients is 81% in the γ-IFN group and 81.5% in the placebo group, with females comprising 19% and 18.5%, respectively. The distribution of hospital categories is as follows: US:NIH (23.8% in γ-IFN and 16.9% in placebo), US:other (49.2% in both groups), Europe:Amsterdam (14.3% in γ-IFN and 15.4% in placebo), and Europe:other (12.7% in γ-IFN and 18.5% in placebo).

3.3. Data Analysis Methods

The data were analyzed using the Covariate-Adjusted Frailty Proportional Hazards model for multilevel survival data. In general, the steps for data analysis are as follows:

1. Define the multilevel frailty model with Weibull baseline hazard.
2. Estimate parameters using Bayesian MCMC (50,000 iterations, 25,000 burn-in).
3. Perform posterior predictive checks and test proportional hazards using Schoenfeld residuals.
4. Interpret hazard ratios and frailty variances

4. RESULTS AND DISCUSSION

Multilevel survival data analysis was performed with two frailties (patients and hospitals) and three covariates: treatment, sex, and age. The two random frailties are patients and hospitals, with hospital frailty variance adjusted by hospital category.

A sufficiently large burn-in period (initial iterations) is essential to ensure that the sampling chain has reached its stationary distribution (Gelman et al., 2013). Furthermore, convergence diagnostics, such as the potential scale reduction factor ($\hat{R}$ statistic), must be assessed, typically requiring $\hat{R}$ values close to 1.0, to confirm the reliability of the posterior estimates across multiple chains (Brooks & Gelman, 1998).

The MCMC analysis was based on 50,000 iterations (25,000 burn-in). The posterior parameter estimates are summarized as follows:

Table 2. Parameter Estimation Results

Parameters	Estimate	95% Credible interval (2.5% - 97.5%)	R-hat
Treatment (γ-INF)	-1.0338	[-1.6435, -0.4072]	1.002
Sex (Female)	-0.1436	[-0.9063, 0.5995]	1.000
Age	-0.0329	[-0.0986, 0.0015]	1.002
Baseline Hazard (λ)	0.002221	[0.0003, 0.00705]	1.000
Shape (ρ)	1.1181	[0.8876, 1.3756]	1.000
Patient Frailty Var (σ_u)	0.7855	[0.3550, 1.2343]	1.001
Hospital Frailty Base (σ_v)	0.5289	[0.0307, 2.1591]	1.001
Alpha Hos (α_{hos})	-0.5828	[-2.3374, 1.1629]	1.000
Gamma Hos [1]	-0.1654	[-1.8702, 1.5501]	1.001
Gamma Hos [2]	-0.1962	[-1.9994, 1.7730]	1.002
Gamma Hos [3]	-0.0569	[-1.8891, 1.7472]	1,002

Based on Table 2, it can be seen that the treatment variable (γ-IFN) had a negative coefficient (-1.0338), and because the confidence interval does not exceed 0 (95% CI: [-1.643498, -0.407252]), it is significant. This means that the treatment covariate rINF-g as the treatment group is significant in reducing the risk of serious infection for CGD patients compared to the placebo, which is the control group. This can also be shown through the survival curve.

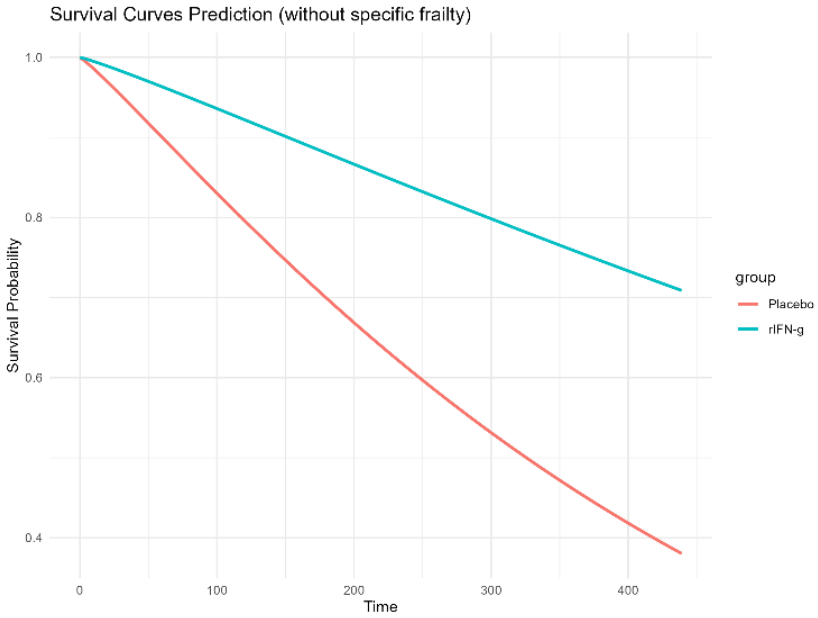

Figure 1. Survival Curve for Treatment

From Figure 1, it can be seen that at each point in time, the group receiving rINF-g treatment showed a higher probability of survival compared to the group receiving placebo treatment. This indicates that rINF-g treatment has a positive effect on the survival rate of CGD patients.

In addition, based on Table 2, it can also be seen that the confidence intervals for the covariates sex and age exceed 0, so it can be concluded that the covariates sex and age are not significant. This can be shown through the following posterior plot.

This posterior plot displays the estimated regression coefficients along with their 95% confidence intervals. From this plot, it can be observed that the treatment coefficient has a confidence interval that does not include zero and is negative, indicating that the

treatment variable has a significant effect on reducing the severity of infection in CGD patients. Conversely, the regression coefficients for sex and age, with confidence intervals that include zero, indicate that these two variables have no significant effect.

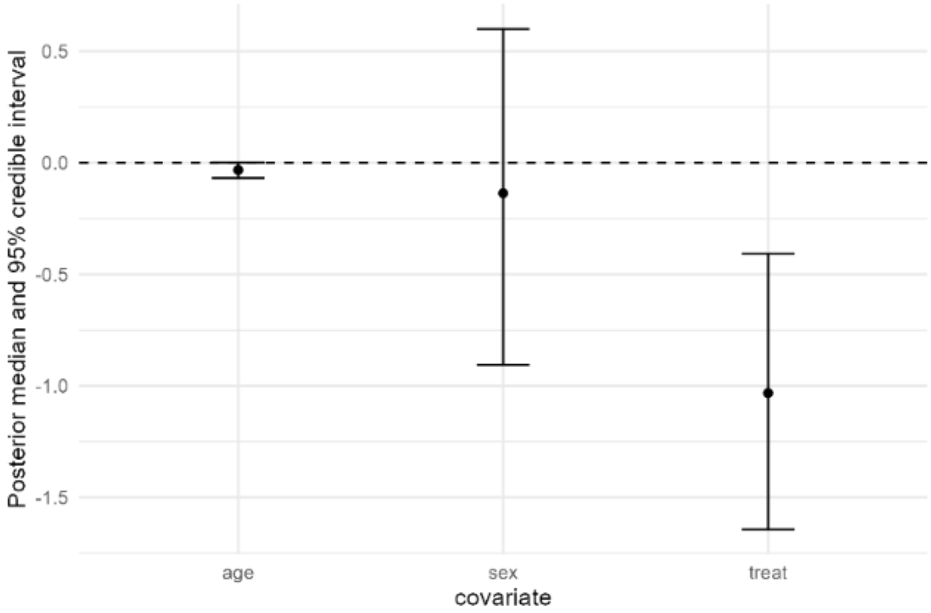

Figure 1. Posterior Plot of Treatment, Sex, and Age

From Table 2, it can also be seen that the baseline hazard parameter is very small, namely 0.002221, indicating a very low baseline hazard at the beginning of time that is close to constant. For the hazard distribution shape parameter 1.1181 > 1, this indicates that the hazard increases over time. Both patient frailty variance (σ_u: 0.7855) and hospital frailty variance (σ_v: 0.5289) were significant, confirming substantial heterogeneity between patients and hospitals.

Based on the estimation results, the regression coefficient for the treatment covariate rINF-g is -1.0338. Thus, the hazard ratio for treatment rINF-g is:\

$$HR = \exp(-1.0338) \approx 0.3556 \quad (8)$$

The HR value of 0.3556 indicates that patients receiving rINF-g treatment have a risk of serious infection in CGD patients of approximately 35.56% of the risk of patients receiving placebo treatment. In other words, rINF-g treatment reduces the risk of occurrence by 64.44% compared to placebo.

From these results, the Covariate-Adjusted Frailty Proportional Hazards method can be used as a reference when using multilevel survival data because it provides comprehensive results by analyzing the influence of random effects (frailty) within the data level or tier. The Covariate-Adjusted Frailty PH method in multilevel survival analysis allows for the handling of complex data forms such as repeated event times for each patient while considering variations at the patient and hospital levels. The posterior predictive check confirmed adequate model fit (p-value=0.95>0.05), and the proportional hazards assumption held (p-value=0.12). This model's capability to handle recurrent events offers an advantage over single-event models (Zhou et al., 2020).

5. CONCLUSION

Based on a case study of factors influencing the severity of infection in CGD patients, it was found that the rINF-g treatment variable was significant in reducing the severity of infection in CGD patients. The hazard ratio (HR) value showed that rINF-g treatment had a 64.44% effect in reducing the risk of occurrence. In addition, the random effects of patients and hospitals were also significant. This means that there is heterogeneity between patients and between hospitals.

DECLARATIONS

Ethics Statement: This study uses secondary, publicly available data from the R survival package; no new human subjects were involved, and original trial ethics approvals are assumed from the source study.
Conflicts of Interest (COI): The authors declare no conflicts of interest.
Funding: This research received no external funding.
Data Availability: The dataset is publicly available in the 'survival' package of R software.

REFERENCES

Berger, J. O. (1985). *Manual for SOA Exam MLC, Fall 2009 Edition*. New Hartford: SRBooks, Inc.

Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Carroll, K. J. (2003). On The Use and Utility of the Weibull Model in the Analysis of Survival Data. *Control Clin Trials*, *24*(6), 682–701.

Diego, I. G., Yolanda, M. G., John, L. S., Osvaldo, V., & Marcelo, B. (2025). The Multivariate Shared Truncated Normal Frailty Model with Application to Medical Data. *Scientific Report, 15*:30099.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC press.

Hox, J. (2002). *Multilevel Analysis Techniques and Applications*. New Jersey: Lawrence Erlbaum Associates, Inc.

Kiprotich, G., Gallardo, D. I., Ramos, P. L., & Augustin, T. (2025). A Shared Frailty Regression Model for Clustered Survival Data. *Statistical Methods in Medical Research*, *34*(7):1385-1412.

Kleinbaum, D. G., Klein, M. (2005). *Survival Analysis A Self-Learning Text 2nd Edition.* London: Springer.

Lee, K., Kim, S., & Park, H. (2025). Neural Network-Based Frailty Models for Complex Correlated Outcomes. *Journal of Data Science, 0*(0): 1-14.

Liu, D., Kalbfeisch, J. D., & Schaubel, D. E. (2011). A Positive Stable Frailty Model for Clustered Failure Time Data with Covariate-Dependent Frailty. *Biometrics*, *67*(1): 8-17.

Noh, M., Ha, I. D., & Lee, Y. (2006). Dispersion Frailty Models and HGLMs. *Statistic in Medicine*, *25*(8): 1341-1354.

Porndumnernsawat, P., Frank, T.D. & Ingsrisawang, L. (2025). On A Bayesian Multivariate Survival Tree Approach Based on Three Frailty Models. *Sci Rep 15*, 12017

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The Impact of Heterogenity in Individual Frailty on The Dynamics of Mortality. *Demography*, *16*(3): 439-454.

Yslas, J. (2025). Phase-Type Frailty Models: A Flexible Approach to Modeling Unobserved Heterogeneity in Survival Analysis. *Scandinavian Actuarial Journal,* 1-29.

Zhang, Y., Li, X., & Wang, J. (2020). A Bayesian Semiparametric Model for Multivariate Frailty. *Statistica Sinica*, *30*(1), 185–206.

Zhou, H., Hanson, T., Jara, A., & Zhang, J. (2015). Modeling County Level Breast Cancer Survival Data Using a Covariate-Adjusted Frailty Proportional Hazards Model. *The Annals of Applied Statistics*, *9*(1): 43-68.