BAREKENG: Journal of Mathematics and Its Applications
December 2026 Volume 20 Issue 4 Page 2755-2768
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss4pp2755-2768

SURVIVAL TIME MODELING IN HEMODIALYSIS PATIENTS USING A WEIBULL MIXTURE PROPORTIONAL HAZARDS MODEL

Rahida Rihhadatul Aisy [1], Nur Iriawan [2*], Shofi Andari [3], Mayta Rithmala [4], Yani Sumartin [5]

[1,2,3] *Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember*
Jln Teknik Mesin No.175, Surabaya, 60111, Indonesia

[4,5] *Jemursari Islamic Hospital*
Jln. Raya Jemursari No.51-57, Surabaya, 60237, Indonesia

*Corresponding author's e-mail: * **nur.iriawan@its.ac.id***

Article Info

Article History:

Received: 4th September 2025
Revised: 12th May 2026
Accepted: 16th May 2026
Available online: 24th August 2026

Keywords:

Bayesian Analysis;
hemodialysis;
Mixture Model;
Parametric Proportional Hazard.

ABSTRACT

Hemodialysis is a critical treatment for patients with end-stage chronic kidney disease (CKD). Although it is associated with a higher risk of mortality and complications compared to other renal treatment choices, many patients prefer hemodialysis as their renal replacement therapy. Traditional parametric survival models often struggle to capture the complexity of the distribution due to multimodal survival data in heterogeneous populations, such as hemodialysis patients. To overcome this, mixture models provide greater flexibility by combining several distributions to better reflect latent survival patterns. This study investigates mortality risk factors by applying a Bayesian Weibull mixture proportional hazards model to 151 hemodialysis patients from one hospital in Surabaya, Indonesia, in 2024, with data extracted from medical hospital records. The Expectation-Maximization and No-U-Turn Sampler (EM-NUTS) approach was used to estimate model parameters and latent class memberships. The Expectation-Maximization (EM) approach, originally developed for control charts and time series, was adapted for survival analysis to address latent heterogeneity. Exploratory analysis reveals multimodal survival distributions, suggesting distinct risk groups. Model selection using the Bayesian Information Criterion and the Akaike Information Criterion identified two latent groups with distinct risk profiles. In the first group, male gender and diabetes significantly increased mortality risk, while hypertension had a smaller effect. In the second group, hypertension was the dominant risk factor, with less influence from gender and diabetes. These findings emphasize the importance of accounting for population heterogeneity in the survival analysis of hemodialysis patients. The use of advanced Bayesian mixture models with EM-NUTS estimation provides robust tools for uncovering hidden subgroups and improving risk stratification, allowing for more personalized treatment strategies in CKD care.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:

R. R. Aisy, N. Iriawan, S. Andari, M. Rithmala and Y. Sumartin., "SURVIVAL TIME MODELING IN HEMODIALYSIS PATIENTS USING WEIBULL MIXTURE PROPORTIONAL HAZARDS MODEL", *BAREKENG: J. Math. & App.*, vol. 20, no. 4, pp. 2755-2768, Dec, 2026.

Copyright © 2026 Author(s)
Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The most advanced stage of chronic kidney disease (CKD), also known as end-stage renal disease (ESRD), is treated with hemodialysis, a life-sustaining procedure that is widely used. CKD occurs when the kidney's ability to filter blood is weakened [1]. A machine-based process that filters metabolic waste and excess fluids, dialysis improves quality of life and survival rates when used as a renal replacement therapy [2]. Patients frequently need therapy for the rest of their lives, which makes frequent access to specialized healthcare facilities essential. It is also categorized as a catastrophic sickness, a condition that requires repeated, long-term care and results in high medical expenses. According to this classification, CKD is an important indicator of national health strategies.

Despite being a particularly common renal replacement therapy selected by patients with CKD, hemodialysis has a higher risk of death than other treatments [2]. t is critical for understanding the survival rate and mortality risk of hemodialysis patients. The overall 1-year survival rate of hemodialysis-treated patients is 91.5% [3]. Several factors that might threaten the lives of hemodialysis patients may affect their condition, such as chronic illness, increased age, blood urea nitrogen, and gender [4],[2], [5], [6]. In this study, we also looked at the patient's creatinine level before their first hemodialysis as a potential early health predictor [7].

Earlier research on survival analysis often employed parametric, semiparametric, or nonparametric methods based on conventional statistical methods and classical distributions, such as the Weibull, Lognormal, and Gamma distributions, which are commonly observed in datasets [8]. Conventional survival models are oversimplified for a heterogeneous population. This is because several underlying causes are indicated by the heterogeneity of the disease population, suggesting that the significance of related risk factors and the course of the disease can differ across disease subgroups. These hidden subgroups in the context of a finite mixture survival model represent unobserved groups whose membership is inferred probabilistically by calculating posterior probabilities throughout the modeling process, rather than being directly observed. The estimated mixture weights (also called mixing proportions or component probabilities), which indicate the probability that the population belongs to each subgroup, compute the likelihood of each latent subgroup within the total population. This also suggests separate baseline hazard functions and covariate effects for various disease subgroups in statistical modeling [9].

To account for population heterogeneity, we used parametric approaches in this work that incorporate a mixture model. Several studies proposed a mixture of proportional-hazard models to provide additional useful information and aid the interpretation of hazard comparisons for medical trial patients [10], [11], [12]. Much more research using mixture models has obtained the best results in interpreting survival analysis in medical applications [13], [14], [15]. In this study, the number of mixture components was deliberately restricted to a maximum of three, corresponding to the short-, mid-, and long-term survival groups.

The Expectation-Maximization (EM) algorithm approach from Dempster [16] is widely recognized as an effective approach for survival analysis, particularly in the context of finite mixture models [17], [18], [19]. However, a more adaptable computing approach is required due to the analytical incompatibility of the non-closed-form log-likelihood in the maximization step. In mixture models, parameter estimation yields a large number of parameters that are challenging to estimate analytically. Numerical techniques, such as the Bayesian method, can therefore be applied to this problem because they can provide flexibility and quantify uncertainty in predictions and inferences [20], [21]. There are several techniques in Bayesian modeling for estimating model parameters. The Gibbs sampling algorithm, which includes Hamiltonian Monte Carlo (HMC), may be used. HMC is a Markov Chain Monte Carlo (MCMC) technique that reduces the random-walk behavior typical of conventional MCMC methods and increases sampling efficiency by leveraging gradient information [22], [23]. But how well HMC performs depends on the number of leapfrog steps; too few steps can result in ineffective random walks, while too many can waste computation. The No-U-Turn Sampler (NUTS) algorithm was proposed to increase sampling efficiency by automatically choosing the number of steps and stopping when a U-turn is detected [24], [25]. Although the EM-NUTS hybrid framework has been successfully implemented by Rasyid et al. [26] for robust monitoring control charts in air quality applications, its use in survival modeling remains scarcely explored. To perform robust Bayesian inference and identify latent heterogeneity and multimodal survival patterns in patient data, this study uses and modifies the EM-NUTS methodology for hemodialysis patients covered by national insurance or BPJS, and investigates potential factors that may affect their survival duration.

2. RESEARCH METHODS

2.1 Mixture Weibull Model

Supposed observation of the lifetimes y_i of a few patients for individual i, where $i = 1, 2, \ldots, n$, let u_i be an individual's potential censoring time and δ_i be the corresponding censoring indicator. The survival time consisted of $t_i = \min(y_i,u_i)$ with $\delta_i = 1$ if $y_i < u_i$ or uncensored and $\delta_i = 0$ if $y_i > u_i$ or censored observation. For a two-parameter Weibull distribution, the density function would be in Eq. (1) [11].

$$f(t;\lambda,\gamma) = \frac{\gamma}{\lambda}\left(\frac{t}{\lambda}\right)^{\gamma-1} e^{-\left(\frac{t}{\lambda}\right)^{\gamma}}, \tag{1}$$

where t is non-negative random variable for survival time, $\gamma > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. The survival and hazard functions at time t are respectively written as Eq. (2) and Eq. (3) [27], [11].

$$S(t;\lambda,\gamma) = e^{-\left(\frac{t}{\lambda}\right)^{\gamma}}, \tag{2}$$

$$h(t;\lambda,\gamma) = \frac{\gamma t^{\gamma-1}}{\lambda^{\gamma}}. \tag{3}$$

A mixture Weibull model is a weighted average for its density function shown in Eq. (4) with M groups, where $Q_i = k$ is the membership of the observation, and $k = 1, 2, \ldots, M$, where w_k is non-negative proportion / weight for the k group and $w_1 + w_2 + \ldots + w_M = 1$.

$$f(t|w) = \sum_{k=1}^{M} w_k\, f(t;\lambda_k,\gamma_k) \tag{4}$$

The survival and hazard functions of this model are presented in Eqs. (5) and (6), respectively [11].

$$S(t|w) = \sum_{k=1}^{M} w_k\, S_k(t;\lambda_k,\gamma_k), \tag{5}$$

$$h(t|w) = \sum_{k=1}^{M} v_k\, h_k(t;\lambda_k,\gamma_k), \tag{6}$$

where

$$v_k = \frac{w_k S_k(t;\lambda_k,\gamma_k)}{\sum_{k=1}^{M} w_k\, S_k(t;\lambda_k,\gamma_k)},$$

and $v_1 + v_2 + \ldots + v_M = 1$.

2.2 Proposed Weibull Mixture Proportional Hazard Model

The basic semiparametric Cox regression or proportional hazard (PH) model with the Weibull baseline function is defined as Eq. (7).

$$h(t,\boldsymbol{x}) = h_0(t)\exp(\mathbf{x}^{\intercal}\boldsymbol{\beta}), \tag{7}$$

where $\mathbf{x}_i$ is a row covariate vector, the coefficient vector $\boldsymbol{\beta}$ is the influence of the monitoring variables on the survival regression, and the Weibull baseline hazard is presented in Eq. (8).

$$h_0(t) = \sum_{k=1}^{M} v_k\, h_k(t;\lambda_k,\gamma_k) = \sum_{k=1}^{M} v_k \frac{\gamma_k t^{\gamma_k-1}}{{\lambda_k}^{\gamma_k}}, \tag{8}$$

where

$$v_k = \frac{w_k e^{-\left(\frac{t}{\lambda_k}\right)^{\gamma_k}}}{\sum_{k=1}^{M} w_k e^{-\left(\frac{t}{\lambda_k}\right)^{\gamma_k}}}.$$

Based on Eng & Hanlon [28], a right censored observation that follows PH model has density in every group presented in Eq. (9).

$$f_k(t,\delta|\mathbf{x}) = [h_{0.k}(t)\exp(\mathbf{x}^\intercal\boldsymbol{\beta}_k)]^\delta \exp[-H_0(t)\exp(\mathbf{x}^\intercal\boldsymbol{\beta}_k)], \quad (9)$$

where $H_{0.k}(t)$ is baseline cumulative hazard for the k-th class. For this research we also observe the latent class $Q = (Q_1, Q_2, \ldots, Q_n)$, where $Q_i \sim$ Multinomial(w), $Qi \in \{1, 2, \ldots, M\}$ and $q_{ik} = 1\{Qi = k\}$.

Given the mixture of the Weibull proportional hazard model and the data, the likelihood function form that considers the right censored observation is shown as Eq. (10).

$$L(\boldsymbol{\omega}|t,\boldsymbol{x}) = \prod_{i=1}^{n}\left(\prod_{k=1}^{M} w_k[h_{0.k}(t)\exp(\mathbf{x}^\intercal\boldsymbol{\beta}_k)]^{\delta_i}\exp[-H_0(t)\exp(\mathbf{x}^\intercal\boldsymbol{\beta}_k)]\right)^{q_{i.k}}, \quad (10)$$

where $\boldsymbol{\omega}$ is a set of all parameters include the mixture proportion and parameters vector with $\boldsymbol{\omega} = \{\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\beta}\}$.

2.3 Expectation-Maximization

We assume that the M mixture groups are known. The corresponding log likelihood is given by Eq. (11).

$$l(\boldsymbol{\omega}|t,\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{k=1}^{M} q_{i.k}\left(\ln w_k + \delta_i \ln h_{0.k}(t) + \delta_i\mathbf{x}_i^\mathrm{T}\boldsymbol{\beta}_k - H_{0.k}(t)\exp(\mathbf{x}_i^\mathrm{T}\boldsymbol{\beta}_k)\right). \quad (11)$$

In the E-step, we calculate the expectation of the log likelihood based on observable data and the current estimate of unknown parameters. The expectation is the posterior membership probability derived by Bayes' rule in Eq. (12).

$$\mathrm{p}(q_{i.k} = 1|t_i,\boldsymbol{x}_i,\delta_i;\boldsymbol{\omega}) \propto \frac{l(\boldsymbol{\omega}|t_i,\boldsymbol{x}_i,\delta_i)\mathrm{p}(q_{i.k}=1)}{\mathrm{p}(t_i,\delta_i|\boldsymbol{x}_i,\boldsymbol{\omega})}, \quad (12)$$

and this conditional expectation result is the target function that is maximized under the condition of mixture proportion w_k in the M-step. The maximization step would use NUTS algorithm to renew the parameters.

2.4 Parameter Estimation: NUTS Algorithm

In the maximization step, parameter updates are performed by maximizing $\mathrm{p}(\boldsymbol{\omega} \mid q, t, \mathbf{x})$ with the prior and log-posterior settings as the sampling target, as in the study by Rasyid et al. [26] using the NUTS algorithm. The log-posterior is obtained from the log-likelihood and log-prior and is divided by the log-probability of the data, which serves as the normalized constant as shown in Eq. (13).

$$\log\mathrm{p}(\boldsymbol{\omega} \mid q,t,\mathbf{x}) = l(\boldsymbol{\omega}|t,\boldsymbol{x}) + \log\mathrm{p}(\mathbf{w}) + \log\mathrm{p}(\boldsymbol{\lambda}) + \log\mathrm{p}(\boldsymbol{\gamma}) + \log\mathrm{p}(\boldsymbol{\beta}), \quad (13)$$

where $\mathbf{p}(\mathbf{w})$, $\mathbf{p}(\boldsymbol{\beta})$, $\mathbf{p}(\boldsymbol{\lambda})$, and $\mathbf{p}(\boldsymbol{\gamma})$ are prior for each parameter. The setting for parameter estimation based on log-posterior, prior distribution, and log-prior are explained on these points:

1. $\mathbf{w} \sim$ Dirichlet($\varphi_1, \ldots, \varphi_M$) with $\varphi_k = \varphi$, and $w_k^{(0)}$ in $\mathbf{w}^{(0)} = (w_1^{(0)},\ldots,w_k^{(0)},\ldots,w_M^{(0)})$ is calculated $Q_{ik} = k$ divided by number of observations. The log-posterior for parameter $\mathbf{w}$ is $\log\mathrm{p}(\mathbf{w}) = \log$ Dirichlet$(\mathbf{w}|\varphi_1,\varphi_2,\ldots,\varphi_M)$.
2. $\lambda_k \sim$ Gamma(a_λ, b_λ), where $\lambda_k^{(0)}$ is an initial value and set to a very small value and greater than 0. The log posterior parameter $\boldsymbol{\lambda}$ is $\log\mathrm{p}(\boldsymbol{\lambda}) = \sum_{k=1}^{M}\log Gamma(\boldsymbol{\lambda}|\, a_{\lambda_k}, b_{\lambda_k})$.
3. $\gamma_k \sim$ Gamma(a_γ, b_γ), where $\gamma_k^{(0)}$ is an initial value and set to a very small value and greater than 0. The log posterior parameter $\boldsymbol{\gamma}$ is $\log\mathrm{p}(\boldsymbol{\gamma}) = \sum_{k=1}^{M}\log Gamma(\boldsymbol{\gamma}|\, a_{\gamma_k}, b_{\gamma_k})$.
4. $\beta_{k.e} \sim$ Normal(μ_{ke}, $\sigma^2{}_{ke}$) with $\beta_{k.e}^{(0)}$ for $e = 1, 2, \ldots, p$, is an initial value and obtained from the *Cox* PH regression parameter. The log posterior parameter for $\boldsymbol{\beta}$ is $\log\mathrm{p}(\boldsymbol{\beta}) = \sum_{k=1}^{M}\log Normal(\beta_{e_k}|\,\mu_{e_k},\mu_{e_k}^2)$.

The NUTS algorithm is explained, with prior settings and log-posterior computed using the available membership information. For better understanding, we present the pseudocode for the NUTS algorithm in Algorithm 1.

Algorithm 1. *Pseudocode NUTS algorithm*
Input : $\boldsymbol{\omega}^{(0)}$, ε, L, and N
Output : $\boldsymbol{\omega}$
initialize $\boldsymbol{\omega} = \boldsymbol{\omega}^{(0)}$
compute gradient of log-posterior $\left(\nabla_{\omega}\log \mathbf{p}(\boldsymbol{\omega} \mid \boldsymbol{q}, \boldsymbol{t}, \mathbf{x})\right)$ using automatic differentiation
for c = 1 to C do
    sample momentum $\rho^{(0)} \sim$ Normal(0,1)
    set $(\tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\rho}}) \leftarrow (\boldsymbol{\omega}^{(c-1)}, \rho)$
    update $(\tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\rho}})$ using Leapfrog steps:
    - Half – step momentum update:

$$\tilde{\boldsymbol{\rho}}^{\left(c+\frac{1}{2}\right)} \leftarrow \tilde{\boldsymbol{\rho}} + \left(\frac{\varepsilon}{2}\right) \nabla_{\omega}\log \mathbf{p}(\boldsymbol{\omega} \mid \boldsymbol{q}, \boldsymbol{t}, \mathbf{x})$$

    - Full – step position update:

$$\tilde{\boldsymbol{\omega}}^{(c)} \leftarrow \tilde{\boldsymbol{\omega}} + \boldsymbol{\varepsilon}\tilde{\boldsymbol{\rho}}^{\left(c+\frac{1}{2}\right)}$$

    - Half – step momentum update:

$$\tilde{\boldsymbol{\rho}}^{(c+1)} \leftarrow \tilde{\boldsymbol{\rho}}^{\left(c+\frac{1}{2}\right)} + \left(\frac{\varepsilon}{2}\right) \nabla_{\omega}\log \mathbf{p}(\boldsymbol{\omega} \mid \boldsymbol{q}, \boldsymbol{t}, \mathbf{x})$$

    build binary tree
    - Perform forward and backward
      Forward step : $\tilde{\boldsymbol{\omega}}^{+}, \tilde{\boldsymbol{\rho}}^{+} \leftarrow$ Leapfrog$(\tilde{\boldsymbol{\omega}}^{+}, \tilde{\boldsymbol{\rho}}^{+}, +\boldsymbol{\varepsilon})$
      Backward step : $\tilde{\boldsymbol{\omega}}^{-}, \tilde{\boldsymbol{\rho}}^{-} \leftarrow$ Leapfrog$(\tilde{\boldsymbol{\omega}}^{-}, \tilde{\boldsymbol{\rho}}^{-}, -\boldsymbol{\varepsilon})$
    check U-turn condition by:
    $(\tilde{\boldsymbol{\omega}}^{+} - \tilde{\boldsymbol{\omega}}^{-})^{\mathbf{T}}\tilde{\boldsymbol{\rho}}^{+} < 0$ or $(\tilde{\boldsymbol{\omega}}^{+} - \tilde{\boldsymbol{\omega}}^{-})^{\mathbf{T}}\tilde{\boldsymbol{\rho}}^{-} < 0$
    calculate acceptance probability

$$\boldsymbol{\alpha} = \min\left(\mathbf{1}, \exp\left(\boldsymbol{H}\left(\tilde{\boldsymbol{\omega}}^{(0)}, \tilde{\boldsymbol{\rho}}^{(0)}\right) - \boldsymbol{H}\left(\tilde{\boldsymbol{\omega}}^{(c)}, \tilde{\boldsymbol{\rho}}^{(c)}\right)\right)\right)$$

    accept or reject $\tilde{\boldsymbol{\omega}}$ based on α
    return $\tilde{\boldsymbol{\omega}}$ collected samples
end for

In Algorithm 1, the iteration is part of a larger loop that includes the expectation step, and it will stop once the parameter values have converged, written down by $|\boldsymbol{\omega}^{(c+1)} - \boldsymbol{\omega}^{(c)}| < \tau$, where τ is the iteration tolerance threshold. In this case, the sample generated in the maximization step are 20,000 iterations with 1 chain and thinning of 10, resulting in 1,000 samples used for inference based on the maximization step.

2.5 The Fitting Mixture Groups

Defining the fitting number of groups is crucial for the mixture model. The number of groups can be explored using a possible information criterion based on the pointwise predictive log-likelihood [29]. WAIC automatically accounts for model complexity by the posterior variance and remains valid for singular models such as mixtures [30]. The WAIC formula is given in Eq. (14).

$$WAIC = -2\left(\sum_{i=1}^{N} log\left(\frac{1}{S}\sum_{s=1}^{S} p(\boldsymbol{q},\boldsymbol{t},\mathbf{x}|\boldsymbol{\omega}^{(s)})\right)\right) + 2\sum_{i=1}^{N} Var_{s=1}^{S}\left(\log p(\boldsymbol{q},\boldsymbol{t},\mathbf{x}|\boldsymbol{\omega}^{(s)})\right). \quad (14)$$

2.6 Definition and Data Source

This study was conducted in 2024 on 151 patients with CKD receiving hemodialysis at one of the hemodialysis units at Surabaya Hospital. The group of patients was traced back to the initial diagnosis, obtained from medical records and visits since 2020. Patients who received hemodialysis treatment under the National Health Program between January 1st and December 31st, 2024, were included. Exclusion visits were removed from data excluding the date of the initial diagnosis or complete creatinine information. Survival time was the amount of time from the patient's first CKD diagnosis until death. If a patient survived until the end of the observation period, the survival time was considered censored.

For the survival period analysis, 23 patients died, and the remaining patients were censored. Factors influencing survival include demographic variables such as age and gender, as well as the patient's condition, particularly the presence of comorbid conditions such as hypertension and diabetes mellitus. Additionally, early health conditions before hemodialysis, specifically blood creatinine levels, are important factors in determining the risk of death for patients.

2.7 Analysis Step

This retrospective cohort study was designed to investigate the risks that hemodialysis patients face and the factors that influence them. First, the potential combinations of the results were considered. This method will clarify how the model will ultimately be developed to estimate hidden populations based on their survival time until either censoring or death. Given their survival histories, some patients may be categorized differently, potentially leading to mixed outcomes. Model selection and determination of the optimal number of mixture components were performed by comparing results from the traditional EM and EM-NUTS algorithm across various mixture specifications, with each component evaluated separately. The WAIC, a statistical information criterion, is used to determine this. The RStan package is used for parameter estimation via the EM algorithm with a small number of mixtures for analysis [31], [32].

3. RESULTS AND DISCUSSION

Of the 151 patients, 61% were male. The patients ranged in age from 46 to 83 years, with 47% aged 60 or older. Ninety-one percent of the patients suffered from complications that worsened their condition, such as diabetes mellitus, hypertension, or possibly both. More than half of the patients received their first hemodialysis after being diagnosed with CKD for over a month; the remaining patients received treatment over the course of a month; and 11% received hemodialysis immediately after receiving a doctor's initial diagnosis, indicating that their condition is urgent and requires dialysis.

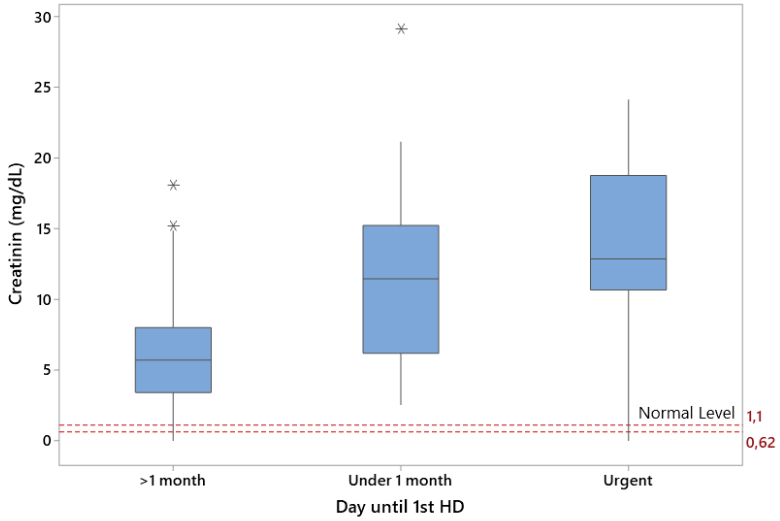


Figure 1. Creatinine Levels (mg/dL) Across Different Periods

To decide their renal condition before starting treatment, we assessed their creatinine levels. According to the boxplot in Figure 1, the patients' creatinine levels are generally more than 5 mg/dL. Severe renal dysfunction was shown by the highest creatinine levels in the urgent group of patients. Creatinine levels are moderately elevated in patients aged 1 month, while the lowest levels remain above normal in patients who receive their first hemodialysis after 1 month of diagnosis. This implies a direct correlation between the need for hemodialysis and creatinine levels.

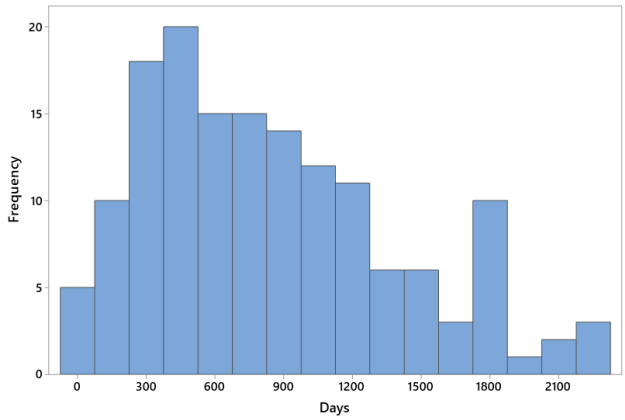
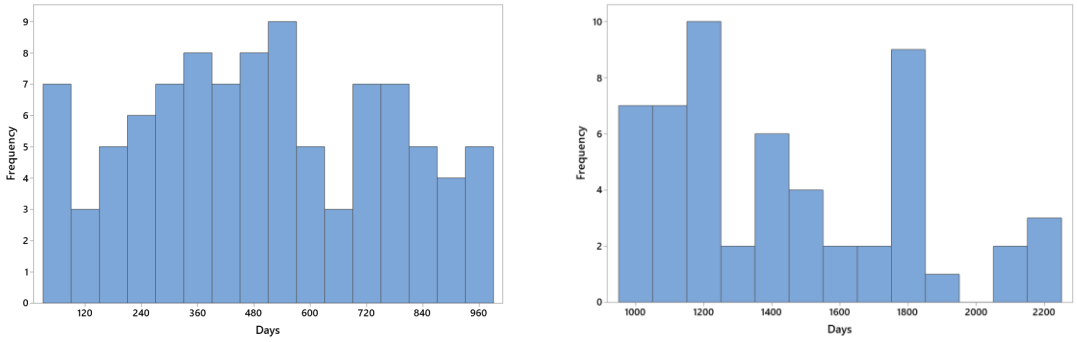


Figure 2. Histogram Of Patient Survival Times (Days)

To examine the survival time distribution, we analyzed the overall histogram shown in Figure 2 which indicates asymmetric data patterns and suggests a multimodal distribution. From earlier research, the nature of the survival time histogram can be an indicator of mixture detection, given its multimodal pattern, skewness, and heavy tails, which incorporate heterogeneity within the population [33], [34]. The histogram shows that most patients survive between 300 and 900 days, or approximately 1 to 3 years. After this peak, the frequency decreases, but there is a smaller peak at 1800 days (about 5 years), indicating a subgroup of patients with longer survival times. This may result in mixed data distribution. The component of the combination was taken into consideration in this study based on its survival time, as illustrated in Figure 3 and Figure 4 for the separate groups, respectively.



(a) (b)

Figure 3. Histogram Of Patient Survival Times (Days) For Component 1 (a) And Component 2 (b)

For the first scenario, patients are divided into two components based on survival times, as shown in Figure 3. From Figure 3 (a) presents the distribution of patients in Component 1 surviving roughly between 120 and 960 days, or about 4 months to nearly 3 years. The distribution shows greater spread but still shows a concentration of patients who survived around 400 and 600 days, or 1 to 2 years. Meanwhile, Figure 3 (b) shows patients in Component 2 survived from about 1000 and 2200 days or 3 and 5 years. This distribution is less uniform and shows fewer distinct peaks.

Figure 4 shows the second scenario, with patients divided into three groups based on survival time. (Figure 4 (a)) has a relatively uniform spread of shorter survival times between roughly 50 and 650 days, patients in Component 2 (Figure 4 (b)) have centers distribution around intermediate survival times from about 700 to 1250 days with some clustering, and patients in Component 3 (Figure 4 (c)) consists of longer

survival times ranging from 1300 to 2250 days, featuring a prominent peak near 1800 days, highlighting distinct survival patterns among the components.

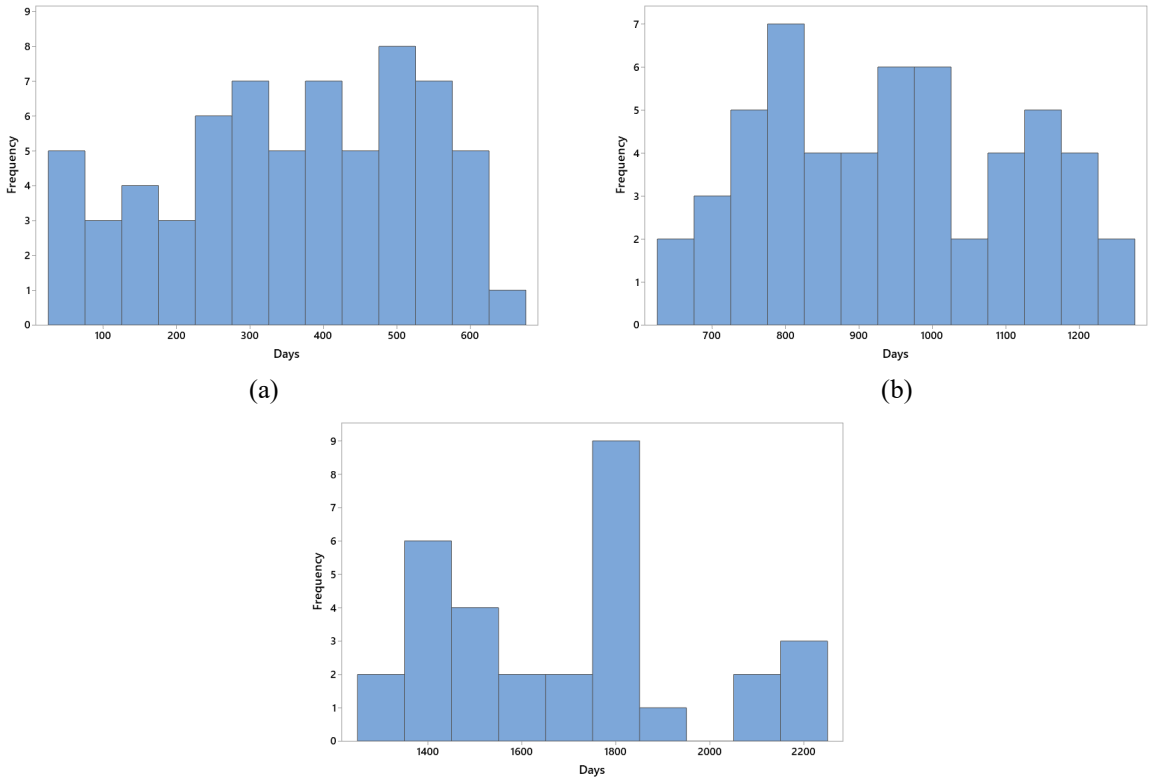


Figure 4. Histogram Of Patient Survival Times (Days) For Component 1 (a), Component 2 (b), Component 3 (c)

The goodness-of-fit between survival times and the Anderson-Darling statistics is shown in Table 1. This result indicates that the Weibull distribution is the most suitable overall distribution for the entire scenario, with the lowest value among the distributions. Therefore, the Weibull distribution was the right choice for modeling patient survival times in this dataset.

Table 1. Goodness Of Fit Anderson-Darling Statistics

Distribution	All	2 Components		3 Components			Critical Value
		1	2	1	2	3	
Weibull	**0.29**	**1.15**	**1.76**	**1.54**	**0.92**	**1.03**	2.50
Log-Normal	2.42	3.25	1.10	3.22	0.73	0.79	2.50
Log-Logistic	2.39	3.20	1.09	3.18	0.79	0.79	2.50
Gamma	1.82	2.49	1.08	2.88	0.71	0.74	2.50

In this study, parameter estimation was conducted using the EM-NUTS approach, with the prior and log-posterior setting detailed in subpart 2.4 Parameter Estimation: NUTS Algorithm. Table 2 shows the model selection based on its WAIC. The best-fitting model for the analysis is the two-component Weibull mixture proportional hazards model, estimated using the EM-NUTS approach and yielding the lowest WAIC. Hazard rates for hemodialysis patients were analyzed, and the results showed that the two Components had different survival trends; their mathematical forms are presented in Eq. (15).

$$h(t) = \sum_{k=1}^{2} v_k(t) \frac{\gamma_k t^{\gamma_k - 1}}{\lambda_k{}^{\gamma_k}} exp\left(\boldsymbol{\beta}_k^T \mathbf{x}\right), \tag{15}$$

where

$$v_k = \frac{w_k e^{-\left(\frac{t}{\lambda_k}\right)^{\gamma_k}}}{\sum_{k=1}^{2} w_k e^{-\left(\frac{t}{\lambda_k}\right)^{\gamma_k}}}.$$

Figure 5 illustrates the survival functions for the two patient components. The patient in Component 2, denoted by the blue line, had a higher survival probability than the patient in Component 1. For example, in the first year or fewer than 500 days, Component 1 patients had only a 75% survival probability; otherwise, Component 2 maintained a 95% survival rate at that time. All of the patients in Component 1 had passed away by the end of the follow-up; however, 50 and 55 percent of the patients in Component 2 were still alive. The significance of identifying and addressing factors that contribute to higher survival among hemodialysis patients is highlighted by the clear difference in the survival curves. The comparison of these survival plots highlights the diversity of the patient population and the importance of specialized clinical interventions to improve patient outcomes.

Table 2. WAIC Model Value Comparison

Model	EM	EM-NUTS
Weibull PH 2 Components	2102.829	-65.416
Weibull PH 3 Components	2486.480	-67.131

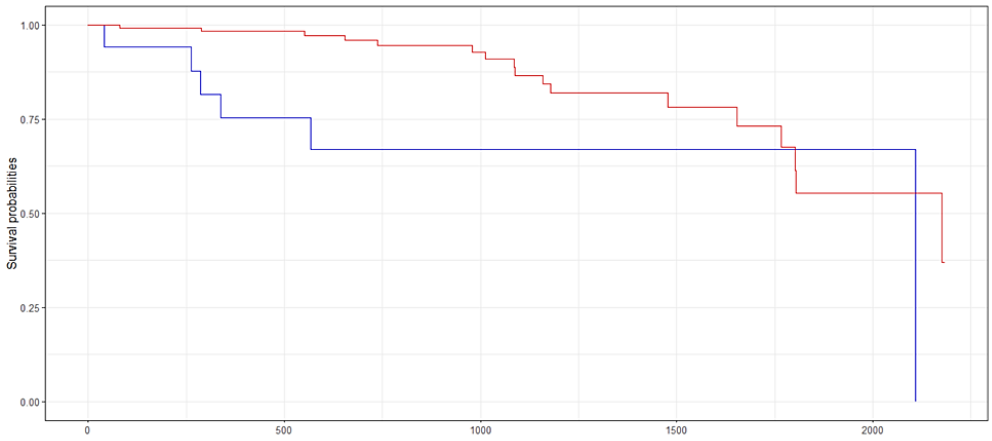


Figure 5. Survival Probability Plot (Component 1 Vs Component 2)

The investigation of risk factors for death among hemodialysis patients indicates significant variation in survival outcomes, as shown by the mixture model approach. Table 3 presents the estimated regression coefficients for several factors affecting the mortality risk of hemodialysis patients in each Component, along with the 2.5% and 97.5% credible intervals from the parameter estimation.

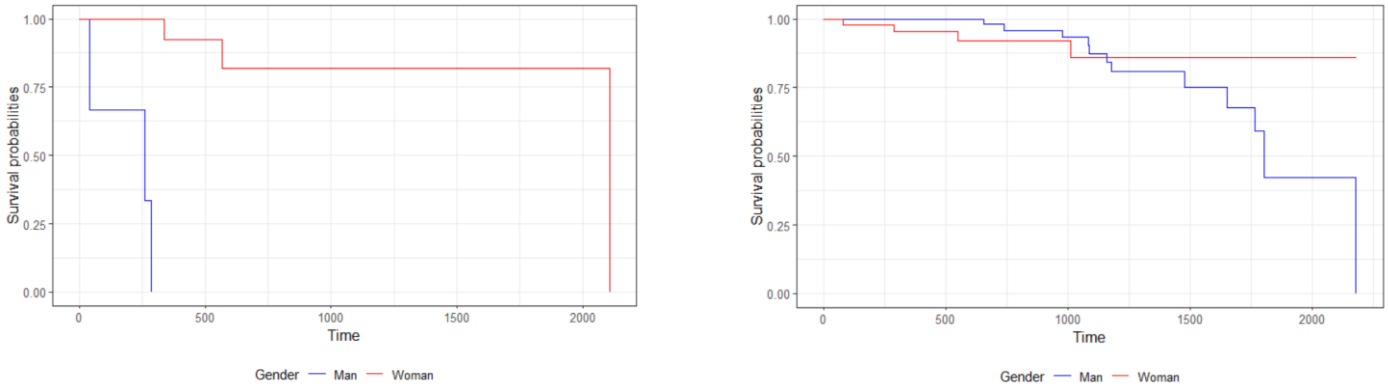
Table 3. Regression Coefficient Estimates For Weibull PH 2 Components

Variable	Mean est(β)	Exp(β)	2.5%	97.5%
		Component 1		
Gender*	2.00	7.41	1.81	2.19
Age	0.02	1.02	-0.17	0.22
Diabetes*	0.80	2.22	0.61	0.99
Hypertension*	0.28	1.32	0.08	0.47
Creatinine	0.05	1.05	-0.15	0.24
		Component 2		
Gender*	0.45	1.57	0.26	0.65
Age	0.11	1.12	-0.07	0.30
Diabetes*	0.24	1.28	0.05	0.44
Hypertension*	0.52	1.68	0.32	0.71
Creatinine	0.01	1.01	-0.19	0.20

Predictors marked with an asterisk (*) are statistically significant from credible interval

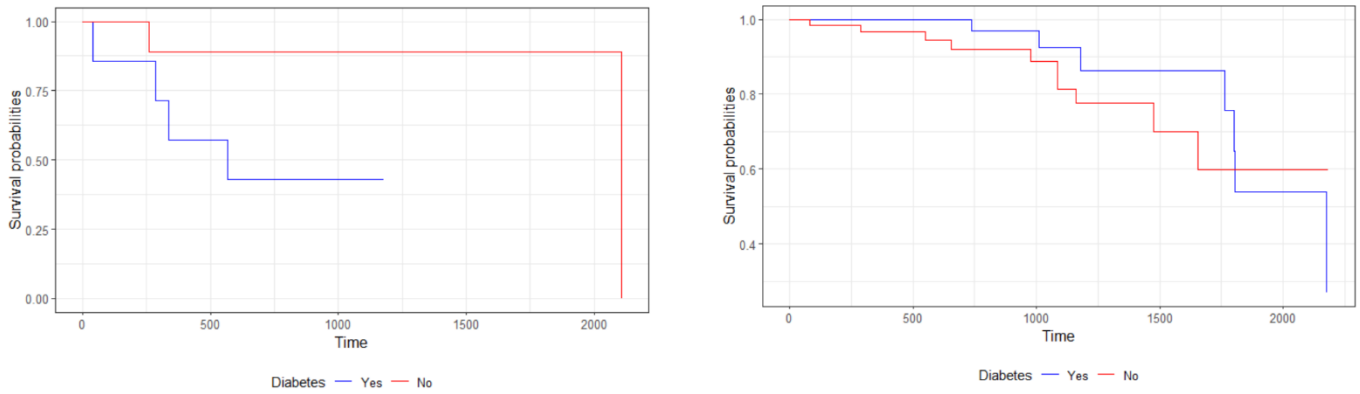
The results from Table 3 show that male patients had a 7.4 times higher risk of mortality than female patients in Component 1, as indicated by the hazard ratio (Exp(**β**) = 7.41). This pronounced gender effect is also reflected in the survival probability plots in Figure 6, where women in Component 1 consistently show much higher survival probabilities than men. On the other hand, the gender effect is less significant in Component 2, where males' risk of death is just 1.6 times that of females (Exp(**β**) = 1.57). This is in line with

prior studies, including that by Jung et al. [35], which found that although the survival advantage of females may be lower than in the general population, it still exists in the dialysis population.



(a) (b)

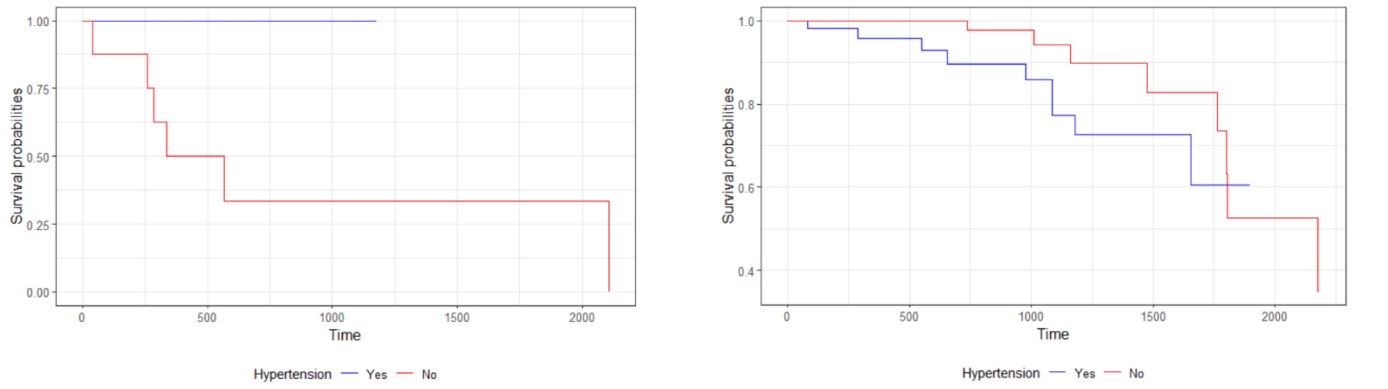
Figure 6. Comparison Of Survival Probability Plots Based on Gender In Component 1 (a) And Component **2 (b)**



(a) (b)

Figure 7. Comparison of Survival Probability Plots Based on Diabetes Status in Component 1 (a) and Component 2 (b)

Diabetes is another critical risk factor. The survival curves in Figure 7 highlight the negative effects of diabetes on long-term outcomes, showing that people without diabetes have consistently higher survival probabilities. From Table 3 in Component 1, diabetic patients have a 2.2 times higher risk of mortality compared to non-diabetic patients (Exp(β) = 2.22). On the other hand, for Component 2, the effect of diabetes is still present but less pronounced (Exp(β) = 1.28), suggesting that the impact of diabetes on mortality risk may vary across patient subgroups. This outcome was consistent with other previous research. According to a study by Vijayan et al. [36], patients who have a history of diabetes are less likely to survive hemodialysis treatment. Thus, it is important to prioritize early detection and management of therapy in individuals with chronic kidney disease and diabetes. Poor glycaemic management also leads to fluid overload and worsening cardiovascular outcomes, according to Eldehni et al. [37].



(a) (b)

Figure 8. Comparison of Survival Probability Plots Based on Hypertension Status in Component 1 (a) and Component 2 (b)

Hypertension also plays a significant role, but its impact differs between components. Based on Table 3 in Component 1, hypertension is associated with a 1.3 times higher risk of mortality (Exp(**β**) = 1.32), while in Component 2, the risk increases to 1.7 times (Exp(**β**) = 1.68). The survival plots in Figure 8 show that patients without a history of hypertension have better survival prospects in both components, but the difference is more pronounced in Component 2. According to Bansal [38], this implies that hypertension is a major risk factor in specific patient subgroups, possibly due to differences in underlying pathophysiology, including fluid overload, arterial stiffness, and sympathetic nervous system activation.

The study's use of the mixture model has enabled separation of two distinct hemodialysis patient subpopulations, each with its own risk profile. The findings show that risk factors including gender, diabetes, and high blood pressure do not have the same impact on every patient. Rather, the model shows that the most significant causes of death in Component 1 are diabetes and gender, while the most significant risk factor in Component 2 is hypertension. This difference highlights latent heterogeneity in the hemodialysis population that conventional single-population models would miss.

4. CONCLUSION

This study addresses the critical need to understand the diverse mortality risk profiles within the hemodialysis patient population by applying a mixture model estimated via the EM-NUTS algorithm. This combination approach reveals that risk factors such as gender, diabetes, and hypertension do not uniformly affect all patients but instead differ significantly across distinct subgroups. In Component 1, gender and diabetes emerged as the most influential predictors of mortality, while in Component 2, hypertension was the dominant risk factor. The use of EM-NUTS, which combines the strengths of the Expectation-Maximization approach and the No-U-Turn Sampler algorithm, enabled robust and efficient parameter estimation, including credible intervals that reflect the uncertainty in the estimates. The ability to identify these latent subpopulations is important for personalized medicine management, as it enables clinicians to treat them based on subgroup-specific risk profiles.

Clinicians should consider subgroup-specific risk profiles when assessing and treating hemodialysis patients. They should prioritize diabetes control and gender-sensitive approaches in one subgroup and aggressively manage hypertension in patients who are identified as belonging to the subgroup where it has the greatest impact. Future research could also improve results by adding variables such as hemodialysis duration and other relevant factors that may affect patient outcomes, including lifestyle factors like nutrition, exercise, and treatment regimen compliance. Also, extending the follow-up period would provide more thorough insights into the long-term consequences of treatment, leading to a better understanding of the factors influencing the risks of mortality and survival among hemodialysis patients.

To increase the accuracy and dependability of the model, future studies should address the imbalance between survivors and non-survivors from a methodological perspective. The inclusion of additional risk factors, based on variable selection, and the implementation of heterogeneous mixture models—in which each component is allowed to have a different probability distribution—have also been found to better represent subgroup-specific hazard shapes.

Author Contributions

Rahida Rihhadatul Aisy: Conceptualization, Data Curation, Project Administration, Formal Analysis, Visualization, Writing-Original Draft, Software, Editing. Nur Iriawan: Conceptualization, Software, Methodology, Supervision, Writing-Review, Editing. Shofi Andari: Conceptualization, Resources, Methodology, Supervision, Writing-Review, Editing. Mayta Rithmala: Investigation, Supervision, Validation, Writing-Review, Resources. Yani Sumartin: Investigation, Supervision, Validation, Resources. All authors discussed the results, reviewed, and approved the final version of the manuscript prior to submission.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to express sincere gratitude to Institut Teknologi Sepuluh Nopember (ITS) for providing the grant for postgraduate studies at ITS as stated in the Rector's Decree Number 20/IT2/T/HK.00.01/II/2024.

Declarations

The authors declare that they have no conflicts of interest to report in this study.

Declaration of Generative AI and AI-assisted technologies

Claude AI was used only for grammar and language editing. All scientific content and conclusions are the authors' own.

REFERENCES

[1] P. Rout and A. Aslam, "END-STAGE RENAL DISEASE IN: STATPEARLS [Internet]," StatPearls Publishing. Accessed: May 12, 2025. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK499861/

[2] M. Puspitasari, Afiatin, V. Oktaria, Y. Wardhani, and W. Wijaya, "FIVE-YEAR SURVIVAL ANALYSIS AND PREDICTORS OF MORTALITY OF ADULT HEMODIALYSIS PATIENTS IN INDONESIA: A NATIONWIDE DATABASE ANALYSIS," *Int. Urol. Nephrol.*, vol. 56, no. 11, pp. 3657–3664, Jun. 2024, doi: https://doi.org/10.1007/s11255-024-04118-1.

[3] R. Andhika *et al.*, "ONE-YEAR SURVIVAL OF END-STAGE KIDNEY DISEASE PATIENTS UNDERGOING HEMODIALYSIS IN INDONESIA," *Int. J. Nephrol. Renovasc. Dis.*, vol. Volume 18, pp. 87–101, Mar. 2025, doi: https://doi.org/10.2147/IJNRD.S508012 .

[4] N. S. Rahmi, A. Rifai, M. I. Islami, and A. A. Azifa, "COX PROPORTIONAL HAZARD AND EXPONENTIAL SURVIVAL ANALYSIS IN PATIENTS WITH END-STAGE CHRONIC KIDNEY FAILURE AT BOJONEGORO," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 3, pp. 2127–2140, Jul. 2025, doi: https://doi.org/10.30598/barekengvol19iss3pp2127-2140.

[5] B. Yaprak, N. Arslan, and H. Alatas, "MULTIPLE FACTORS INFLUENCING MORTALITY IN HEMODIALYSIS PATIENTS," *Eur Rev Med Pharmacol Sci*, vol. 27, no. 3, pp. 1095–1103, 2023.

[6] R. Msaad *et al.*, "PREDICTORS OF MORTALITY IN HEMODIALYSIS PATIENTS," *Pan African Medical Journal*, vol. 33, 2019, doi: https://doi.org/10.11604/pamj.2019.33.61.18083.

[7] S. Tanaka *et al.*, "IMPACT OF BLOOD UREA NITROGEN TO CREATININE RATIO ON MORTALITY AND MORBIDITY IN HEMODIALYSIS PATIENTS: THE Q-COHORT STUDY," *Sci. Rep.*, vol. 7, no. 1, p. 14901, Nov. 2017, doi: https://doi.org/10.1038/s41598-017-14205-2.

[8] D. D. Rivera Rosales and D. A. Tejada, "FUNDAMENTALS AND APPLICATIONS OF SURVIVAL ANALYSIS FOR HEALTH RESEARCH," *Alerta, Revista científica del Instituto Nacional de Salud*, vol. 8, no. 3, pp. 305–314, Jul. 2025, doi: https://doi.org/10.5377/alerta.v8i3.20675.

[9] T. Fei, J. J. Hanfelt, and L. Peng, "LATENT CLASS PROPORTIONAL HAZARDS REGRESSION WITH HETEROGENEOUS SURVIVAL DATA," *Stat. Interface*, vol. 17, no. 1, pp. 79–90, 2024, doi: https://doi.org/10.4310/23-SII785.

[10] J. J. Z. Liao and G. F. Liu, "A FLEXIBLE PARAMETRIC SURVIVAL MODEL FOR FITTING TIME TO EVENT DATA IN CLINICAL TRIALS," *Pharm. Stat.*, vol. 18, no. 5, pp. 555–567, Oct. 2019, doi: https://doi.org/10.1002/pst.1947.

[11] G. F. Liu and J. J. Z. Liao, "ANALYSIS OF TIME-TO-EVENT DATA USING A FLEXIBLE MIXTURE MODEL UNDER A CONSTRAINT OF PROPORTIONAL HAZARDS," *J. Biopharm. Stat.*, vol. 30, no. 5, pp. 783–796, Sep. 2020, doi: https://doi.org/10.1080/10543406.2020.1783283.

[12] M. Foulon, A. El Ghouch, and C. Legrand, "P39 - MULTIMODAL MIXTURE REGRESSION ON CENSORED DATA WITH A CURE FRACTION," *Journal of Epidemiology and Population Health*, vol. 73, p. 203070, May 2025, doi: https://doi.org/10.1016/j.jeph.2025.203070.

[13] A. Bahrampour, M. R. Baneshi, A. Karamoozian, N. S. Seyedghasemi, A. Etminan, and M. Eghbalian, "LONG-TERM SURVIVAL OF PATIENT WITH END-STAGE RENAL DISEASE USING BAYESIAN MIXTURE CURE RATE FRAILTY MODELS," *Iran. J. Public Health*, Sep. 2024, doi: https://doi.org/10.18502/ijph.v53i9.16464.

[14] Z. Shayan, V. Ebrahimi, and S. Ezzatzadegan Jahromi, "FACTORS ASSOCIATED WITH SHORT-TERM AND LONG-TERM SURVIVAL IN PATIENTS UNDERGOING HEMODIALYSIS USING SEMI- PARAMETRIC MIXTURE CURE MODELS.," *Iran. J. Kidney Dis.*, vol. 14, no. 1, pp. 44–51, Jan. 2020.

[15] R. Ghosal, M. Matabuena, and J. Zhang, "FUNCTIONAL PROPORTIONAL HAZARDS MIXTURE CURE MODEL AND ITS APPLICATION TO MODELLING THE ASSOCIATION BETWEEN CANCER MORTALITY AND PHYSICAL ACTIVITY IN NHANES 2003-2006," Feb. 2023.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "MAXIMUM LIKELIHOOD FROM INCOMPLETE DATA VIA THE *EM* ALGORITHM," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 39, no. 1, pp. 1–22, Sep. 1977, doi: https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[17] S. Chen and F. Yang, "EXPECTATION-MAXIMIZATION ALGORITHM FOR THE WEIBULL PROPORTIONAL HAZARD MODEL UNDER CURRENT STATUS DATA," *Mathematics*, vol. 11, no. 23, p. 4826, Nov. 2023, doi: https://doi.org/10.3390/math11234826 .
[18] L. Sanchez, P. L. Luaces, C. Fonte, and A. Lage, "MIXTURE SURVIVAL MODELS METHODOLOGY: AN APPLICATION TO CANCER IMMUNOTHERAPY ASSESSMENT IN CLINICAL TRIALS," *ArXiv*, 2019, doi: https://doi.org/10.48550/arXiv.1911.09765 .
[19] Y. Lee, "EXPECTATION-MAXIMIZATION ALGORITHMS FOR LEARNING A FINITE MIXTURE OF UNIVARIATE SURVIVAL TIME DISTRIBUTIONS FROM PARTIALLY SPECIFIED CLASS VALUES," Iowa State University, Digital Repository, Ames, 2013. doi: https://doi.org/10.31274/etd-180810-3372.
[20] Abdelbaset A.Sh Abdalla, Nasir Ibrahim Elmesmari, and Mohammed M. Elnazzal, "BAYESIAN INFERENCE APPROACH FOR A MIXTURE OF NORMAL REGRESSION MODELS," *World Journal of Advanced Research and Reviews*, vol. 24, no. 3, pp. 214–227, Dec. 2024, doi: https://doi.org/10.30574/wjarr.2024.24.3.3622.
[21] I. Paolucci, Y.-M. Lin, J. Albuquerque Marques Silva, K. K. Brock, and B. C. Odisio, "BAYESIAN PARAMETRIC MODELS FOR SURVIVAL PREDICTION IN MEDICAL APPLICATIONS," *BMC Med. Res. Methodol.*, vol. 23, no. 1, p. 250, Oct. 2023, doi: https://doi.org/10.1186/s12874-023-02059-4.
[22] S. Thomas and W. Tu, "LEARNING HAMILTONIAN MONTE CARLO IN R," 2020.
[23] R. M. Neal, "MCMC USING HAMILTONIAN DYNAMICS," in *Handbook of Markov Chain Monte Carlo*, New York: Chapman and Hall/CRC, 2011, pp. 113–162. doi: https://doi.org/10.1201/b10905-6.
[24] M. Nishio and A. Arakawa, "PERFORMANCE OF HAMILTONIAN MONTE CARLO AND NO-U-TURN SAMPLER FOR ESTIMATING GENETIC PARAMETERS AND BREEDING VALUES," *Genetics Selection Evolution*, vol. 51, no. 1, p. 73, Dec. 2019, doi: https://doi.org/10.1186/s12711-019-0515-1.
[25] M. D. Hoffman and A. Gelman, "THE NO-U-TURN SAMPLER: ADAPTIVELY SETTING PATH LENGTHS IN HAMILTONIAN MONTE CARLO," *Journal of Machine Learning Research*, vol. 15, pp. 1593–1623, 2014.
[26] D. A. Rasyid, N. Iriawan, and M. Mashuri, "ON THE FLEXIBLE NEO-NORMAL MSAR MSN-BURR CONTROL CHART IN AIR QUALITY MONITORING," *Air, Soil and Water Research*, vol. 17, Jan. 2024, doi: https://doi.org/10.1177/11786221241272391.
[27] D. G. Kleinbaum and M. Klein, "INTRODUCTION TO SURVIVAL ANALYSIS," 2012, pp. 1–54. doi: https://doi.org/10.1007/978-1-4419-6646-9_1.
[28] K. H. Eng and B. M. Hanlon, "DISCRETE MIXTURE MODELING TO ADDRESS GENETIC HETEROGENEITY IN TIME-TO-EVENT REGRESSION," *Bioinformatics*, vol. 30, no. 12, pp. 1690–1697, Jun. 2014, doi: https://doi.org/10.1093/bioinformatics/btu065.
[29] S. Watanabe, "A WIDELY APPLICABLE BAYESIAN INFORMATION CRITERION," *Journal of Machine Learning Research*, vol. 14, 2013, Accessed: May 12, 2025. [Online]. Available: https://www.jmlr.org/papers/volume14/watanabe13a/watanabe13a.pdf
[30] R. Costilla, I. Liu, R. Arnold, and D. Fernández, "BAYESIAN MODEL-BASED CLUSTERING FOR LONGITUDINAL ORDINAL DATA," *Comput. Stat.*, vol. 34, no. 3, pp. 1015–1038, Sep. 2019, doi: https://doi.org/10.1007/s00180-019-00872-4.
[31] J. Guo, J. Gabry, B. Goodrich, A. Johnson, S. Weber, and H. S. Badr, "R INTERFACE TO STAN,"2025, https://mc-stan.org/rstan/articles/.
[32] Stan Development Team, "RSTAN: THE R INTERFACE TO STAN," 2025, 2.32.7. Accessed: May 12, 2025. [Online]. Available: https://mc-stan.org/.
[33] S. Suddin, Gunardi, M. S. Hardianti, and F. Adi-Kusumo, "THE LOG-NORMAL-LOG-LOGISTIC MIXTURE MODEL ANALYSIS ON THE SURVIVAL DATA OF DIFFUSE LARGE B-CELL LYMPHOMA," *Engineering Letters*, vol. 33, no. 9, 2025.
[34] M. Nitivijaya, N. Iriawan, and H. Kuswanto, "BAYESIAN SURVIVAL MIXTURE MODEL ON YEARS OF SCHOOLING IN WEST PAPUA PROVINCE," in *International Conference on Science, Technology, and Humanity 2015*, 2015.
[35] H.-Y. Jung *et al.*, "SEX DISPARITIES IN MORTALITY AMONG PATIENTS WITH KIDNEY FAILURE RECEIVING DIALYSIS," *Sci. Rep.*, vol. 12, no. 1, p. 18555, Nov. 2022, doi: https://doi.org/10.1038/s41598-022-16163-w.
[36] M. Vijayan, S. Radhakrishnan, G. Abraham, M. Mathew, K. Sampathkumar, and N. P. Mancha, "DIABETIC KIDNEY DISEASE PATIENTS ON HEMODIALYSIS: A RETROSPECTIVE SURVIVAL ANALYSIS ACROSS DIFFERENT SOCIOECONOMIC GROUPS," *Clin. Kidney J.*, vol. 9, no. 6, pp. 833–838, Dec. 2016, doi: https://doi.org/10.1093/ckj/sfw069.
[37] M. T. Eldehni, L. E. Crowley, and N. M. Selby, "CHALLENGES IN MANAGEMENT OF DIABETIC PATIENT ON DIALYSIS," *Kidney and Dialysis*, vol. 2, no. 4, pp. 553–564, Oct. 2022, doi: https://doi.org/10.3390/kidneydial2040050.
[38] N. Bansal *et al.*, "HYPERTENSION IN PATIENTS TREATED WITH IN-CENTER MAINTENANCE HEMODIALYSIS: CURRENT EVIDENCE AND FUTURE OPPORTUNITIES: A SCIENTIFIC STATEMENT FROM THE AMERICAN HEART ASSOCIATION," *Hypertension*, vol. 80, no. 6, Jun. 2023, doi: https://doi.org/10.1161/HYP.0000000000000230.