

Mixed-Effects Parametric Proportional Hazard Model with Generalized Log-Logistic Baseline Distribution

Maryrose Wausi Peter^{1*}, Samuel Musili Mwalili², Anthony Kibira Wanjoya²,
Abdsalam Hassan Muse¹

¹Pan African University Institute for Basic Sciences, Technology and Innovation (PAUSTI), Nairobi, Kenya

²Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Email: *maryrose2peter@yahoo.com

How to cite this paper: Peter, M.W., Mwalili, S.M., Wanjoya, A.K. and Muse, A.H. (2023) Mixed-Effects Parametric Proportional Hazard Model with Generalized Log-Logistic Baseline Distribution. *Journal of Data Analysis and Information Processing*, 11, 81-102.

<https://doi.org/10.4236/jdaip.2023.112006>

Received: February 2, 2023

Accepted: March 11, 2023

Published: March 14, 2023

Copyright © 2023 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

Clustered survival data are widely observed in a variety of setting. Most survival models incorporate clustering and grouping of data accounting for between-cluster variability that creates correlation in order to prevent underestimate of the standard errors of the parameter estimators but do not include random effects. In this study, we developed a mixed-effect parametric proportional hazard (MEPPH) model with a generalized log-logistic distribution baseline. The parameters of the model were estimated by the application of the maximum likelihood estimation technique with an iterative optimization procedure (quasi-Newton Raphson). The developed MEPPH model's performance was evaluated using Monte Carlo simulation. The Leukemia dataset with right-censored data was used to demonstrate the model's applicability. The results revealed that all covariates, except age in PH models, were significant in all considered distributions. Age and Townsend score were significant when the GLL distribution was used in MEPPH, while sex, age and Townsend score were significant in MEPPH model when other distributions were used. Based on information criteria values, the Generalized Log-Logistic Mixed-Effects Parametric Proportional Hazard model (GLL-MEPPH) outperformed other models.

Keywords

Survival Analysis, Generalized Log-Logistic, Parametric, Proportional Hazard, Mixed-Effects, Monte Carlo, Maximum Likelihood Estimation

1. Introduction

Many different contexts have observed clustered survival data. Data analysis from various medical studies or belonging to various clusters identified by geographi-

cal or administrative regions are common today. Any survival model must consider the clustering and grouping of data accounting for between-cluster variability that creates this correlation to avoid underestimating the parameter estimators' standard deviations. In hazard regression, neglecting unobserved heterogeneity caused by clustering impacts model specification since the marginal model produced by integrating out random effects might be inconsistent with the model neglecting the random clustering [1] [2]. As a result, the hazard function and effects of covariates should be carefully interpreted without considering the possibility of unobserved covariates.

Fitting a stratified survival model that includes a frailty factor is a straightforward technique to explain clustered survival data [3]. One can more explicitly model the clustering effect by using random effects, which provide for unobserved variation or stratification between clusters and capture the hierarchical nature of the data. Mixed-effect models are a subset of regression models, particularly survival models, which include random effects. The most popular modeling method in survival analysis (Cox proportional hazards regression) [4], requires no specification of a central piece of its formulation, the baseline hazard function $h_0(t_{ij})$. This freedom from parameterizing the baseline hazard function is one of the main reasons that proportional hazards regression is widely used, however, a fully parametric proportional hazard (PH) model needs a distribution assumption [5].

Unfortunately, hierarchical modeling methods typically require such parameterization. With or without making the assumption a probability distribution for survival times, regression models play a crucial role in survival analysis. Distributional presumptions on survival times result to parametric models, while distribution-free assumptions result to semi-parametric cox or non-parametric methods [6].

A universal proportional hazards model with random effects was proposed by [7] [8] [9] for managing clustered survival data, similar to how random effects are treated in linear, generalized linear, and non-linear mixed models. The most prevalent extensions of popular regression survival models for time-to-event data are proportional regression models [10] [11]. By including random effects or Gaussian processes to the linear predictor function in hazard-based regression models, a similar class of mixed-effects models is created [12] [13].

A Bayesian semi-parametric temporally stratified PH model with spatial flaws was created by [14]. County-level frailties were used to introduce the geographical correlation, while the proportional hazard model's segmentation was used to introduce the temporal effect. A Bayesian semiparametric method to the enhanced hazard model was developed by [15], with generalization to wide spatially clustered data. The normal transformation approach made some allowance for a geographical correlation at the county level. They expanded parametric frailty models, such as Gompertz, Weibull and exponential proportional hazards (PH) models, as well as log-normal, generalized gamma and log-logistic, accelerated failure time models, in the study cited in [16] so that any number of nor-

mally distributed random effects could be used to look into baseline risk and covariate effects heterogeneity. To examine data on loblolly pine survival, [17] suggested a semi-parametric PH model with time-dependent covariates.

In a Bayesian semi-parametric framework, [18] created a complete, unified method for modeling spatial survival data that has been arbitrarily censored by adding random effects to models of proportional hazards, proportional odds, and accelerated failure time. Incorporating random effects into the linear predictor of various fixed effects hazard-based regression models in order to capture the clustering effect is what distinguishes these model classes. In a unique mixed-effects general hazard (MEGH) structure which incorporates random effects at both the hazard scale and the temporal scale, [19] established a hazard-based regression model. They considered the Power Generalized Weibull (PGW) and Log-Logistic (LL) as baseline hazard distributions.

Despite all the various advancements in incorporating random effects to regression models, especially survival models. Majority of studies have focused on non-parametric and semi-parametric models that are closed on classical distributions within the proportional hazards (PH) framework, such as Gompertz, Weibull and exponential distributions as well as gamma, log-normal, and log-logistic accelerated failure time (AFT) models. Only a few of them are closed under the proportional hazard model, and none of them is highly adaptable enough to reflect a variety of survival data [20]. These models aren't sufficiently flexible to capture non-monotone hazard rate functions. As a result, parametric survival models that can accommodate both non-monotone and monotone hazard rate functions are needed for clustered survival analysis. By parametrically modeling the baseline hazard function to calculate absolute risk metrics, this is immediately achievable.

There are several applications for the Log-Logistic (LL) and log-normal distributions in the study of survival data. They are members of the accelerated failure time family and are useful in modeling uni-modal survival data sets. Weibull hazard function is monotonic and closed under both AFT and PH at the same time. The cumulative hazard function (chf) and failure rate function of the log-logistic distribution, however, have desirable quality belonging to all classes of parametric hazard-based regression models of the survival analysis when they are generalized. A generalized log-logistic was proposed by [20] as an expansion of the closed log-logistic model that takes the PH connection into account. In the limit, it approaches the Weibull and displays log-logistic traits. With applications to right-censored healthcare data sets, [21] proposed a Bayesian inference for the generalized log-logistic proportional hazard model.

In this study, we develop a mixed effects parametric proportional hazard model with generalized log-logistic as baseline distribution that provides a reliable and flexible method for analyzing non-normal, skewed, and clustered survival data and takes into account unobserved variability or heterogeneity between clusters to account for both non-monotone and monotone hazard rate functions. The model's parameters are estimated via Maximum Likelihood Estima-

tion (MLE).

The remainder of the paper is structured as follows: Section 2, outlines the conceptualization of the mixed-effects parametric proportional hazard model (MEPPH) and PH model and discusses the baseline distribution and its sub-models. In Section 3, we provide a general framework for the likelihood function. Section 4 discusses a comprehensive simulation exercise to illustrate the model’s performance and demonstrate how to simulate using the specified MEPPH. In Section 5, we apply real data to the time-to-event settings to demonstrate the developed model. Finally, in Section 6, we briefly discuss the key findings for this study before offering some suggestions for further research and drawing a conclusion.

2. Model Formulation

We describe the formulation of the developed MEPPH model and the PH model, which is referred to as a subclass of MEPPH, in this segment. Instead of the cumulative distribution function (CDF) and probability density function (PDF), the most broadly used parametric hazard-based regression models are frequently interpreted using hazard rate function (hrf) and cumulative hazard function (chf).

2.1. Mixed Effects Parametric Proportional Hazard (MEPPH)

In survival analysis, the proportional hazard (PH) model is frequently employed. The baseline hazard distribution is left parametric and a parametric PH model with time-dependent variables is used in this study. To study the spatial heterogeneity in patient survival times, we additionally include dimensional random effects in the survival model.

Let S_{ij} , where $i = 1, 2, \dots, m$ represent the cluster and $j = 1, 2, 3, \dots, n_i$ symbolize the individual, be a sample of time-to-event interest and \mathbf{x}_{ij} = vector of covariates connected with the j th individual. Let C_{ij} be the right censoring time and $t_{ij} = \text{Min}(S_{ij}, C_{ij})$ be the observed survival time. Let $\delta_{ij} = 1(S_{ij} < C_{ij})$ be the vital status indicator. Let $N = \sum_{i=1}^m n_i$ be the sample size across m clusters.

We develop a mixed-effects parametric proportional hazard model, which is defined via the individual hazard function such that;

$$h(t_{ij} | \mathbf{x}_{ij}, b_i) = h_0(t_{ij}) \exp\{\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i\}, \tag{2.1}$$

where $h_0(\cdot)$ is a baseline hazard, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_m)^T$ and $b_i \stackrel{iid}{\sim} N(0, \sigma_b^2)$ is the random effects. The cumulative hazard rate function (chf) conforming to the hazard rate function (hrf) in Equation (2.1) is stated as follows:

$$\begin{aligned} H_{\text{MEPPH}}(t_{ij} | \mathbf{x}_{ij}, b_i) &= \int_0^{t_{ij}} h(u) du = \int_0^{t_{ij}} \frac{f(u)}{S(u)} du = \int_0^{t_{ij}} \frac{-dS(u)}{S(u)} du \\ &= -\log_e \{S(t_{ij})\} = -\log_e \left\{ \left[S_0(t_{ij}) \right]^{\exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i)} \right\} \end{aligned} \tag{2.2}$$

Here, we take the link function $\exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i)$ to be equivalent to $\exp(\eta_{ij}) = \psi(x)$. Some of the other frequent lifetime functions describing the

developed MEPPH model include;

1) The Survival function of the MEPPH. Considering the following expression, the survival function of the MEPPH is defined as;

$$\begin{aligned}
 S_{\text{MEPPH}}(t_{ij} | \mathbf{x}_{ij}, b_i) &= \exp\{-H(t_{ij} | \mathbf{x}_{ij}, b_i)\} \\
 &= \exp\left\{-\int_0^{t_{ij}} \psi(x) h_0(t_{ij}) dt_{ij}\right\} \\
 &= \exp\left\{-\psi(x) \int_0^{t_{ij}} h_0(t_{ij}) dt_{ij}\right\} \\
 &= \left[\exp\left\{-\int_0^{t_{ij}} h_0(t_{ij}) dt_{ij}\right\}\right]^{\psi(x)} \\
 &= [S_0(t_{ij})]^{\psi(x)} = [S_0(t_{ij})]^{\exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i)}
 \end{aligned} \tag{2.3}$$

2) The cumulative distribution function (cdf) of the MEPPH model, also referred as the lifetime distribution function, is denoted by

$$\begin{aligned}
 F_{\text{MEPPH}}(t_{ij} | \mathbf{x}_{ij}, b_i) &= 1 - S(t_{ij}) = 1 - \exp\{-H(t_{ij})\} \\
 &= 1 - [S_0(t_{ij})]^{\psi(x)} = 1 - [S_0(t_{ij})]^{\exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i)}
 \end{aligned} \tag{2.4}$$

3) The probability density function (pdf) of the MEPPH is calculated as follows;

$$f(t_{ij} | \mathbf{x}_{ij}, b_i) = h(t_{ij} | \mathbf{x}_{ij}, b_i) S(t_{ij} | \mathbf{x}_{ij}, b_i) = h(t_{ij} | \mathbf{x}_{ij}, b_i) \exp\{-H(t_{ij} | \mathbf{x}_{ij}, b_i)\}$$

Therefore, the pdf becomes

$$\begin{aligned}
 f_{\text{MEPPH}}(t_{ij} | \mathbf{x}_{ij}, b_i) &= h_0(t_{ij}) \psi(x) [S_0(t_{ij})]^{\psi(x)} \\
 &= \frac{f_0(t_{ij})}{S_0(t_{ij})} \exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i) [S_0(t_{ij})]^{\exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i)} \\
 &= f_0(t_{ij}) \exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i) [S_0(t_{ij})]^{\exp(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i) - 1}
 \end{aligned} \tag{2.5}$$

2.2. Proportional Hazard (PH) Model

The proportional hazards assumption states that a covariate will have an effect on the hazard by a proportionate amount that does not depend on t . By using the link function without the random effect so that $b_i = 0$ and the hazard function with the covariate vector \mathbf{x} (fixed/time dependent), the developed model (MEPPH) simplifies to PH;

$$h(t; \mathbf{x}) = h_0(t) \psi(\mathbf{x}^T \boldsymbol{\alpha}) = h_0(t) \exp(\mathbf{x}^T \boldsymbol{\alpha}) = h_0(t) e^{\mathbf{x}^T \boldsymbol{\alpha}}. \tag{2.6}$$

It is clear from Equation (2.6) that the hazard ratio comparing any two formulations of the covariates, say \mathbf{x} and \mathbf{x}^* , is

$$\frac{h(t; \mathbf{x})}{h(t; \mathbf{x}^*)} = \exp\left[\left(\mathbf{x}' - \mathbf{x}^{*'}\right) \boldsymbol{\alpha}\right] \tag{2.7}$$

which remains over time. In other words, the risk for one person is inversely correlated with the risk for any other person, with the proportionality constant being independent of time. Other important lifetime functions of PH are also given as follows:

1) Survival function, which is given as

$$S_{PH}(t) = [S_0(t)]^{\exp(\mathbf{x}^T \boldsymbol{\alpha})} \quad (2.8)$$

2) The lifespan distribution, sometimes referred to as the cumulative density function of the PH model, has the function denoted by

$$F_{PH}(t) = 1 - S(t) = 1 - \exp\{-H(t)\} = 1 - [S_0(t)]^{\exp(\mathbf{x}^T \boldsymbol{\alpha})} \quad (2.9)$$

3) The probability density function (pdf) of the PH is calculated as;

$$\begin{aligned} f_{PH}(t) &= h_0(t) \psi(x) [S_0(t)]^{\psi(x)} \\ &= \frac{f_0(t)}{S_0(t)} \exp(\mathbf{x}^T \boldsymbol{\alpha}) [S_0(t)]^{\exp(\mathbf{x}^T \boldsymbol{\alpha})} \\ &= f_0(t) \exp(\mathbf{x}^T \boldsymbol{\alpha}) [S_0(t)]^{\exp(\mathbf{x}^T \boldsymbol{\alpha}) - 1} \end{aligned} \quad (2.10)$$

By assuming that the baseline hazard function $h_0(t)$ is random and undefined, Equation (2.6) can be used to create a PH model. As a result, the popular Cox PH model [4] is obtained, which does not depend on every distributional assumptions for the dependent variables. It is also possible to construct a fully parametric PH model by presuming that $h_0(t)$ has a parametric form.

2.3. Baseline Distribution

The hazard shapes that the survival model (2.1) can apprehend depend on the choice of the parametric baseline distribution. For instance, the Weibull hazard function can increase or decrease and is closed both under AFT and PH, whereas log-normal and log-logistic hazard functions are uni-modal and not closed under PH. Other (three-parameter) distributions, such as the generalized log-logistic, Generalized Weibull (GW), Power Generalized Weibull (PGW), Exponentiated Weibull (EW) and Generalized Gamma (GG) distributions, can represent the fundamental shapes of the hazard (growing, decreasing, uni-modal, and bathtub). We considered Generalized Log-Logistic (GLL) as the parametric baseline hazard distribution based on its numerical tractability and flexibility, which is closed under PH. The generalized log-logistic distribution is a three-parameter continuous probability distribution with three-shape parameters (β, κ, γ) , is sufficiently resilient and competent of supporting a range of rate of failure functions. This is one of the characteristics that set the GLL apart from the two-parameter log-logistic distribution. The GLL also shows characteristics that are comparable to those of the log-logistic distribution. The generalized distribution additionally has the benefit of having a limit that is close to the desired Weibull distribution, which is a desirable characteristic. This model was created up by [20] and is suitable for including various hrf forms, including constant, monotonic (increasing and decreasing), and non-monotonic (unimodal and bathtub).

Assuming that the random variable T has a positive sign, probability density function (PDF) and cumulative distribution function (CDF) of the three-parameter generalized log-logistic distribution are given respectively as follows:

$$f(t; \mathbf{x}) = \frac{\beta\kappa(\kappa t)^{\beta-1}}{\left[1+(\gamma t)^\beta\right]^{\frac{\kappa\beta}{\gamma^\beta}+1}}, \quad t \geq 0, \quad \kappa, \beta, \gamma > 0, \quad (2.11)$$

$$F(t; \mathbf{x}) = \frac{\left[1+(\gamma t)^\beta\right]^{\frac{\kappa\beta}{\gamma^\beta}} - 1}{\left[1+(\gamma t)^\beta\right]^{\frac{\kappa\beta}{\gamma^\beta}}}, \quad t \geq 0, \quad \kappa, \beta, \gamma > 0, \quad (2.12)$$

where $\beta > 0; \kappa > 0; \gamma > 0$ are unknowable distributional parameters and $\mathbf{x} = (\beta, \kappa, \gamma)'$.

Some other important functions which are obtained from the GLL are:

1) Survival function $S(t; \mathbf{x})$.

$$S(t; \mathbf{x}) = 1 - F(t; \mathbf{x}) = \left[1+(\gamma t)^\beta\right]^{\frac{\kappa\beta}{\gamma^\beta}}, \quad t \geq 0, \quad \kappa, \beta, \gamma > 0. \quad (2.13)$$

2) The hazard rate function (hrf).

The hazard rate function of the GLL model is given by:

$$h(t; \mathbf{x}) = \frac{f(t; \mathbf{x})}{S(t; \mathbf{x})} = \frac{\beta\kappa(\kappa t)^{\beta-1}}{1+(\gamma t)^\beta}, \quad t \geq 0, \quad \kappa, \beta, \gamma > 0. \quad (2.14)$$

3) The reversed hazard rate function $r(t; \mathbf{x})$ which is given by:

$$r(t; \mathbf{x}) = \frac{f(t; \mathbf{x})}{F(t; \mathbf{x})} = \frac{\beta\kappa(\kappa t)^{\beta-1}}{\left[1+(\gamma t)^\beta\right]^{\frac{\kappa\beta}{\gamma^\beta}+1} - \left[1+(\gamma t)^\beta\right]}, \quad t > 0, \quad \kappa, \beta, \gamma > 0. \quad (2.15)$$

4) The cumulative hazard rate function of GLL model $H(t; \mathbf{x})$ is calculated as follows:

$$\begin{aligned} H(t; \mathbf{x}) &= \int_0^t h(u) du = \int_0^t \frac{f(u)}{S(u)} du = \int_0^t \frac{-dS(u)}{S(u)} du \\ &= -\log\{S(t; \mathbf{x})\} = \frac{\kappa\beta}{\gamma^\beta} \log\left[1+(\gamma t)^\beta\right], \quad t \geq 0, \quad \kappa, \beta, \gamma > 0. \end{aligned} \quad (2.16)$$

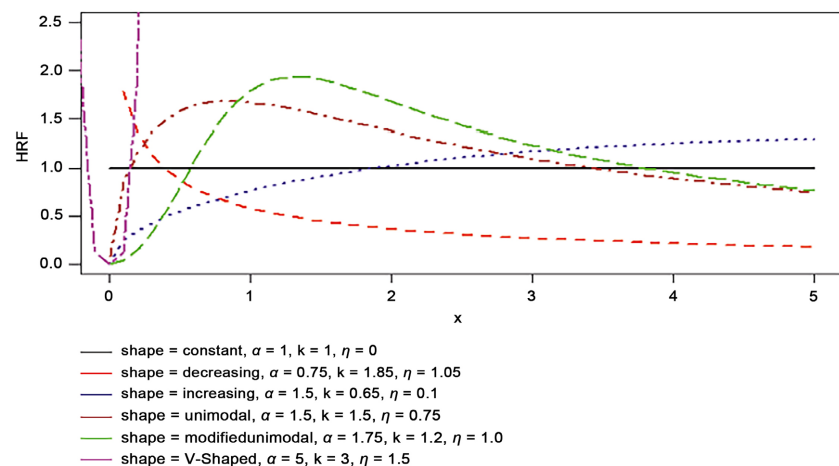


Figure 1. Different hazard rate forms of the GLL distribution with varying parameter values are represented visually.

The GLL distribution’s potential hazard rate shapes are shown in **Figure 1** for various parameter values. According to [21], it can capture the constant, monotone, and non-monotone shape.

5) The Quantile function of the GLL (which is the inverse of the CDF) which crucial in statistical and quantitative data analysis is given by:

$$\begin{aligned}
 1 - [1 + (\gamma t)^\beta]^{-\kappa^\beta / \gamma^\beta} &= p \\
 T_q = F^{-1}(q; \kappa, \beta, \gamma) &= [1 + (\gamma t)^\beta]^{-\kappa^\beta / \gamma^\beta} = 1 - p \\
 &= \frac{1}{[1 + (\gamma t)^\beta]^{\kappa^\beta / \gamma^\beta}} = 1 - p \\
 &= [1 + (\gamma t)^\beta]^{\kappa^\beta / \gamma^\beta} = \frac{1}{1 - p} \\
 &= 1 + (\gamma t)^\beta = \left(\frac{1}{1 - p}\right)^{\gamma^\beta / \kappa^\beta} \\
 &= (\gamma t)^\beta = \left(\frac{1}{1 - p}\right)^{\gamma^\beta / \kappa^\beta} - 1 \\
 &= \gamma t = \left\{ \left[\frac{1}{1 - p} \right]^{\gamma^\beta / \kappa^\beta} - 1 \right\}^{1/\beta} \\
 \therefore q &= \frac{\left\{ [1/(1 - p)]^{\gamma^\beta / \kappa^\beta} - 1 \right\}^{1/\beta}}{\gamma}
 \end{aligned} \tag{2.17}$$

The upper quartile, the median and the lower Quartile are obtained by replacing P by the values 3/4, 1/2 and 3/4 in the quartile function respectively:

Hence, the upper quartile is:

$$T_{q_3} = \frac{\left\{ 4^{\gamma^\beta / \kappa^\beta} - 1 \right\}^{1/\beta}}{\gamma}, \tag{2.18}$$

The median is:

$$T_{q_2} = \text{Median} = \frac{\left\{ 2^{\gamma^\beta / \kappa^\beta} - 1 \right\}^{1/\beta}}{\gamma}, \tag{2.19}$$

And the lower quartile is

$$T_{q_1} = \frac{\left\{ [4/3]^{\gamma^\beta / \kappa^\beta} - 1 \right\}^{1/\beta}}{\gamma}. \tag{2.20}$$

2.4. Sub Models

Other distributions which can be obtained from the baseline hazard distribution include :

- 1) Log-logistic

The hazard rate function form of a log-logistic is obtained by substituting $\kappa = \gamma$ in Equation (2.14) for the two parameters of (κ, β) in the log-logistic distribution.

The GLL hrf is

$$h(t; \mathbf{x}) = \frac{\beta \kappa (\kappa t)^{\beta-1}}{1 + (\gamma t)^\beta}, \quad t \geq 0, \quad \kappa, \beta, \gamma > 0.$$

Equating $\kappa = \gamma$, we get:

$$h(t; \mathbf{x}) = \frac{\beta k (kt)^{\beta-1}}{1 + (kt)^\beta}, \quad t \geq 0, \quad k, \beta > 0. \quad (2.21)$$

2) Weibull Distribution

It is obtained by taking $\gamma^\beta \rightarrow 0$ of Equation (2.14).

The hazard rate function of Equation (2.14) is given by

$$h(t; \mathbf{x}) = \frac{\beta \kappa (\kappa t)^{\beta-1}}{1 + (\gamma t)^\beta}, \quad t \geq 0, \quad \kappa, \beta, \gamma > 0.$$

$$h(t; \mathbf{x}) = \beta k (kt)^{\beta-1}, \quad t \geq 0, \quad k, \beta > 0. \quad (2.22)$$

3) Equation (2.14) simplifies to the hrf of standard log logistic (SLL), which is obtained by substituting $\kappa = \gamma = 1$ in the equation.

$$h(t; \mathbf{x}) = \frac{\beta t^{\beta-1}}{1 + t^\beta}, \quad t \geq 0, \quad \beta > 0. \quad (2.23)$$

4) Two-parameter Burr-X11 (BX11-2) distribution obtained by replacing $\gamma = 1$ in Equation (2.14) simplifies to the hrf of the BX11-2 distribution, which is

$$h(t; \mathbf{x}) = \frac{\beta \kappa (\kappa t)^{\beta-1}}{1 + t^\beta}, \quad t \geq 0, \quad \kappa, \beta > 0. \quad (2.24)$$

2.5. Developed Model

The generalized log-logistic distribution, which is parametric and closed under proportional hazard (PH), can be used to replace the baseline hazard in Equation (2.1) and capture good hazard rate shapes, such as constant, monotone (growing and decreasing), and non-monotone rates (unimodal and bathtub). As a result, it has been discovered that when applied to clustered survival data, parametric models perform better than their semiparametric counterparts [22]. Therefore, in order to estimate these distributional parameters effectively, higher sample sizes and greater censoring rates are often needed [19] [23] [24] [25]. The developed model has a hazard, and cumulative hazard functions, respectively expressed as follows

$$h(t_{ij} | \mathbf{x}_{ij}, b_i) = \frac{\beta \kappa (\kappa t_{ij})^{\beta-1}}{1 + (\gamma t_{ij})^\beta} \exp\{\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_i\}, \quad (2.25)$$

$$H(t_{ij} | \mathbf{x}_{ij}, b_i) = \frac{\kappa^\beta}{\gamma^\beta} \log \left[1 + (\gamma t_{ij})^\beta \right] \exp \{ \mathbf{x}_{ij}^\top \boldsymbol{\alpha} + b_i \}. \quad (2.26)$$

3. Maximum Likelihood Estimation (MLE)

Here, we discuss the conditional and marginal likelihood functions utilized by the marginal maximum likelihood estimators to calculate the parameters of the suggested model, which maximizes the likelihood of the observed data. For the purpose of characterizing the likelihood for the i^{th} cluster under the MEPPH, we will use the variables $t_i = (t_{i1}, \dots, t_{in})$ and $\mathbf{x}_i = (x_{i1}, \dots, x_{in})$ to represent the sample and the corresponding matrix of covariates, respectively. After integrating out the random effects b_i , the cluster-specific marginal probability function related to the i^{th} cluster is provided by

$$\begin{aligned} L_i(\boldsymbol{\theta}, \eta, b_i) &= \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n_i} f(t_{ij}, \delta_{ij} | \mathbf{x}_{ij}, u_i, \boldsymbol{\theta}) \right\} f(b_i | \boldsymbol{\theta}) db_i \\ &= \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n_i} [h(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i) \cdot S(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)]^{\delta_{ij}} [S(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)]^{1-\delta_{ij}} \right\} f(b_i | \boldsymbol{\theta}) db_i \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} [h(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)]^{\delta_{ij}} S(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i) \cdot f(b_i | \boldsymbol{\theta}) db_i \quad (3.1) \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} [h(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)]^{\delta_{ij}} \exp[-H(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)] \cdot f(b_i | \boldsymbol{\theta}) db_i \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} [h(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i) \exp(\eta_{ij})]^{\delta_{ij}} \exp[-[H(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)] \exp(\eta_{ij})] \cdot f(b_i | \boldsymbol{\theta}) db_i \end{aligned}$$

where $\boldsymbol{\theta}$ i = vector of distributional parameter with the baseline hazard and $\eta = (t_{ij}, \sigma_{ij}, \mathbf{x}_{ij})$ denotes the observed data, such that t_{ij} is survival time, σ_{ij} is censoring time and \mathbf{x}_{ij} is covariates. The random effects are thought to follow a multivariate normal distribution, such that;

$$f(\mathbf{u}_i | \boldsymbol{\theta}) = (2\pi|\Sigma|)^{-G/2} \exp\left(-\frac{\mathbf{u}_i^\top \Sigma^{-1} \mathbf{u}_i}{2}\right) \quad (3.2)$$

with variance-covariance matrix Σ and number of random effects G . Equation (3.1) is analytically intractable due to integrating over the random effects; hence it requires a numerical technique to evaluate. An iterative optimization process (Newton-Raphson algorithm) was used to obtain the maximum likelihood estimation $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Hence, the log-likelihood function conditional on the random effects b_i is given by

$$\begin{aligned} &\log(L_i(\boldsymbol{\theta}, \eta, b_i)) \\ &= \log \left(\prod_{j=1}^{n_i} [h(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i) \exp(\eta_{ij})]^{\delta_{ij}} \exp[-[H(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i)] \exp(\eta_{ij})] \cdot f(b_i | \boldsymbol{\theta}) \right) \\ &= \sum_{i=j}^{n_i} \delta_{ij} \log [h(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i) + \eta_{ij}] - \sum_{i=j}^{n_i} ([H(t_{ij} | \boldsymbol{\theta}, \mathbf{x}_{ij}, b_i) \exp(\eta_{ij})] + \log(f(b_i | \boldsymbol{\theta}))) \quad (3.3) \end{aligned}$$

4. Simulation Study

This section contains a full simulation study that shows how effective the developed MEPPH is. We simulated individuals survival time using the PH model. The Generalized log-logistics (GLL) distribution was taken as the baseline hazard's distribution. Four covariates were considered in the simulation study; $x_1 = \text{sex}$, $x_2 = \text{age}$, $x_3 = \text{white blood cell count}$, and $x_4 = \text{Townsend score}$ were present. Three continuous covariates (x_2 , x_3 and x_4) were generated using the standard normal distribution, while one binary covariates x_1 was generated using the Bernoulli (0.5) distribution. Our assumption was that cluster random effect b_i have a normal distribution with a mean of 0 and three standard deviations $\sigma_b = (0.2, 0.5, 1.2)$. For the GLL hazard function, we selected $\kappa = 1.0, \beta = 1.5$, and

$$\lambda = 1.0, \quad h(t; \mathbf{x}) = \frac{\beta \kappa (\kappa t)^{\beta-1}}{1 + (\gamma t)^\beta} \quad \text{and } \mathbf{X}. \quad \text{In accordance with estimates produced}$$

from the acute myeloid leukemia data application, as seen in Section 5, $\mathbf{X} = (0.7, 1.4, 1.0, 1.5)$ were selected. Using the reversed transform technique, the Exponentiated Weibull (EW) distribution was used to simulate lifetime data from the PH model framework [26].

The following are steps for executing the developed MEPPH model:

- 1) Set the initial values of the model's parameters.
- 2) Utilize the reversed transform technique to create the lifetime data by inverting the cumulative hazard rate function of the developed model.
- 3) Utilize the various estimates to evaluate the estimations' values.
- 4) Analyze the inferential properties of the estimates, taking into account the estimates, biases and mean square errors.

Using various scenarios, we first assessed MEPPH model's statistical inference and parameter estimation accuracy in the simulations. We used the following formulas to calculate the Average Biases (ABs) and Mean Squared Errors (MSEs).

$$Bias_\theta [\hat{\theta}] = E_{x|\theta} [\hat{\theta}] - \theta = E_{x|\theta} [\hat{\theta} - \theta] \quad (4.1)$$

$$MSE_\theta [\hat{\theta}] = E_{x|\theta} \left((\hat{\theta} - \theta)^2 \right) \quad (4.2)$$

We also looked into the impact of disregarding random effects. In the simulation, various situations were considered for models that included random effects and those that did not ($n = 250, 500$ and 1000). In particular, the parameter estimates, censoring rates on inference, the effect of sample size and the capacity to recover the baseline hazard shapes were all illustrated. A censoring rates of 25% for all scenarios was used and 250 replications was conducted in each scenario. In accordance with real data application in Section 5, we analyzed both the mixed hazard structure MEPPH and the hazard structure PH models in the simulations.

$$h(t_{ij} | \mathbf{x}_{ij}, b_i) = h_0(t_{ij}) \exp\{\alpha_1 \times \text{sex}_{ij} + \alpha_2 \times \text{age}_{ij} + \alpha_3 \times \text{wbc}_{ij} + \alpha_4 \times \text{tpi}_{ij} + b_i\}, \quad (4.3)$$

where the covariates $\mathbf{x}_{ij}^T = (sex_{ij}, age_{ij}, wbc_{ij}, tpi_{ij})$ are per the acute myeloid leukemia data in Section 5, and the baseline hazard we are focusing on in this article, $h_0(\cdot)$, has the parameters $\eta = (\kappa, \beta, \gamma)$.

From the simulation results based on **Tables 1-3** it can be seen that for both the GLL-PH and GLL-MEPPH models, the parameter estimations tend to true values and the MSEs approach zero as the number of samples grows. It is clear that both the GLL-MEPPH model with mixed hazard structures and the model neglecting random effects provide estimates with significantly less bias. Owing to the existence of random effects being disregarded, the bias of the model excluding them is higher for the majority of parameters. However, as stated in [19], we pointed out that the estimates produced from models with or neglecting random effects are not always comparable. The model neglecting random effects is marginal, but the model including random effects is conditional on random effects. Additionally, it can be seen from the findings that most parameter estimates are low when the variance is high, or $\sigma_b = 1.2$, as opposed to when the variance is low, or $\sigma_b = 0.2$.

5. Real Data Application

We used the developed model to examine acute myeloid leukemia survival in northwest England. 1043 patients were documented between 1982-1998, making up the data. This data set was first analyzed by [27] and is available in the R package `spBayesSurv` [18] [19] [28]. Numerous earlier analyses of this data set have recommended that there is indication of variability in survival across in this region, as shown in the non-parametric Kaplan-Meier estimators of the survival curves by district shown in **Figure 2**. These studies include [18] [19] [27] and many more.

White blood cells at diagnosis, truncated at 500 (Wbc), survival time (in years), sex (1-Male,0-Female), the Townsend score (tpi), for which higher values indicate less affluent areas, age (in years), vital status at the end of the follow-up (1-dead, 0-right censored) and administrative districts of residence (24 districts) are all used. **Figure 3** displays the TTT plots, box plot, violin plot and histogram for the data set. From the TTT plot and histogram, the hazard rate function seems to have a bathtub shape.

We fitted the MEPPH models with Generalized Log-Logistic (GLL), Power Generalized Weibull (PGW), Modified Kumaraswamy Weibull (MKW), and Weibull (W) baseline distributions and normal distribution for random effects. For all fitted models, the hazard effects x_{ij} considered are sex, age, white blood cells and Townsend score. The random effects were also defined in terms of the variable administrative districts. Equation (5.1) contains the mixed model that was fitted to the data as provided below

$$h(t_{ij} | \mathbf{x}_{ij}, b_i) = h_0(t_{ij}) \exp\{\alpha_1 \times sex_{ij} + \alpha_2 \times age_{ij} + \alpha_3 \times wbc_{ij} + \alpha_4 \times tpi_{ij} + b_i\}, \quad (5.1)$$

where the subscripts j and i denote the individual level and the region (districts) level, respectively, and measure the random effects, which are districts. We also fitted models ignoring the random effects for comparison such that

Table 1. Simulation results for the GLL-PH and GLL-MEPPH models when $\sigma_b = 0.2$.

$n = 250$						
Parameters	GLL-PH			GLL-MEPPH		
	Estimates	bias	MSE	Estimates	bias	MSE
$\kappa = 1.0$	0.9388828	-0.061	0.004	1.3338239	0.334	0.111
$\beta = 1.5$	1.4455463	-0.054	0.003	0.8110796	-0.689	0.475
$\gamma = 1.0$	0.7417767	-0.258	0.067	1.4405894	0.441	0.194
$x_1 = 0.7$	0.667347	-0.033	0.001	0.4654687	-0.235	0.055
$x_2 = 1.4$	1.2194389	-0.181	0.033	1.0893326	-0.311	0.097
$x_3 = 1.0$	0.9139290	-0.086	0.007	1.5383368	0.538	0.290
$x_4 = 1.5$	1.4758773	-0.024	0.001	2.2671719	0.767	0.588
$\sigma_b = 0.2$	-	-	-	1.7707029	1.571	2.468
$n = 500$						
$\kappa = 1.0$	0.9458801	-0.055	0.003	0.7394426	-0.260	0.067
$\beta = 1.5$	1.4543262	-0.046	0.002	0.8784832	-0.622	0.386
$\gamma = 1.0$	0.8882095	-0.112	0.012	1.4314582	0.431	0.186
$x_1 = 0.7$	0.6760823	-0.024	0.001	0.5002254	-0.200	0.040
$x_2 = 1.4$	1.3201318	-0.080	0.006	1.1615147	-0.238	0.057
$x_3 = 1.0$	0.9911217	-0.009	0.000	1.4798134	0.480	0.230
$x_4 = 1.5$	1.4843259	-0.016	0.000	1.9943590	0.494	0.244
$\sigma_b = 0.2$	-	-	-	1.5162015	1.316	1.732
$n = 1000$						
$\kappa = 1.0$	0.9890723	-0.011	0.000	1.1605406	0.161	0.025
$\beta = 1.5$	1.4717767	-0.027	0.001	0.8884446	-0.611	0.374
$\gamma = 1.0$	0.9738625	-0.026	0.001	1.3586486	0.359	0.129
$x_1 = 0.7$	0.7287477	-0.029	0.001	0.5859447	-0.114	0.013
$x_2 = 1.4$	1.4247266	-0.025	0.001	1.2776303	-0.122	0.015
$x_3 = 1.0$	1.0062952	-0.006	0.000	1.2235864	0.224	0.050
$x_4 = 1.5$	1.5056442	-0.006	0.000	1.8257287	0.326	0.106
$\sigma_b = 0.2$	-	-	-	1.4505254	1.251	1.565

Table 2. Simulation results for the GLL-PH and GLL-MEPPH models when $\sigma_b = 0.5$.

$n = 250$						
Parameters	GLL-PH			GLL-MEPPH		
	Estimates	bias	MSE	Estimates	bias	MSE
$\kappa = 1.0$	0.9388828	-0.061	0.004	1.2436554	0.244	0.059
$\beta = 1.5$	1.4455463	-0.054	0.003	0.8250604	-0.675	0.456
$\gamma = 1.0$	0.7417767	-0.258	0.067	1.3359156	0.336	0.113

Continued

$x_1 = 0.7$	0.667347	-0.033	0.001	0.4805924	-0.058	0.048
$x_2 = 1.4$	1.2194389	-0.181	0.033	1.1911160	0.209	0.044
$x_3 = 1.0$	0.9139290	-0.086	0.007	1.4365950	0.437	0.191
$x_4 = 1.5$	1.4758773	-0.024	0.001	1.8659980	0.366	0.134
$\sigma_b = 0.2$	-	-	-	1.6722802	1.172	1.374
$n = 500$						
$\kappa = 1.0$	0.9458801	-0.055	0.003	1.1631334	0.163	0.027
$\beta = 1.5$	1.4543262	-0.046	0.002	0.9494354	-0.551	0.303
$\gamma = 1.0$	0.8882095	-0.112	0.012	1.2551023	0.255	0.065
$x_1 = 0.7$	0.6760823	-0.024	0.001	0.5371689	-0.163	0.027
$x_2 = 1.4$	1.3201318	-0.080	0.006	1.2347985	-0.165	0.027
$x_3 = 1.0$	0.9911217	-0.009	0.000	1.4036142	0.404	0.163
$x_4 = 1.5$	1.4843259	-0.016	0.000	1.6327490	0.133	0.018
$x_b = 0.5$	-	-	-	1.6380454	1.138	1.295
$n = 1000$						
$\kappa = 1.0$	0.9890723	-0.011	0.000	1.108792	0.109	0.012
$\beta = 1.5$	1.4717767	-0.027	0.001	1.2538528	-0.246	0.061
$\gamma = 1.0$	0.9738625	-0.026	0.001	1.2035814	-0.204	0.041
$x_1 = 0.7$	0.7287477	-0.029	0.001	0.6423152	-0.219	0.003
$x_2 = 1.4$	1.4247266	-0.025	0.001	1.3219786	-0.078	0.006
$x_3 = 1.0$	1.0062952	-0.006	0.000	1.0863058	0.086	0.007
$x_4 = 1.5$	1.5056442	-0.006	0.000	1.6218084	0.122	0.015
$\sigma_b = 0.5$	-	-	-	1.5701139	1.070	1.145

Table 3. Simulation results for the GLL-PH and GLL-MEPPH models when $\sigma_b = 1.2$.

$n = 250$						
Parameters	GLL-PH			GLL-MEPPH		
	Estimates	bias	MSE	Estimates	bias	MSE
$\kappa = 1.0$	0.9388828	-0.061	0.004	1.2407932	0.241	0.058
$\beta = 1.5$	1.4455463	-0.054	0.003	0.9407652	-0.559	0.313
$\gamma = 1.0$	0.7417767	-0.258	0.067	1.1784751	0.178	0.032
$x_1 = 0.7$	0.667347	-0.033	0.001	0.6067050	-0.093	0.009
$x_2 = 1.4$	1.2194389	-0.181	0.033	1.2239566	-0.176	0.031
$x_3 = 1.0$	0.9139290	-0.086	0.007	1.2619245	0.262	0.069
$x_4 = 1.5$	1.4758773	-0.024	0.001	1.7076817	0.208	0.043
$\sigma_b = 1.2$	-	-	-	1.6876299	0.488	0.238

Continued

	$n = 500$					
$\kappa = 1.0$	0.9458801	-0.055	0.003	1.0726279	0.073	0.005
$\beta = 1.5$	1.4543262	-0.046	0.002	1.2622511	-0.238	0.057
$\gamma = 1.0$	0.8882095	-0.112	0.012	1.0870751	0.087	0.008
$x_1 = 0.7$	0.6760823	-0.024	0.001	0.6632815	-0.037	0.001
$x_2 = 1.4$	1.3201318	-0.080	0.006	1.3099049	-0.090	0.008
$x_3 = 1.0$	0.9911217	-0.009	0.000	1.1287101	0.129	0.017
$x_4 = 1.5$	1.4843259	-0.016	0.000	1.6037406	0.104	0.011
$\sigma_b = 1.2$	-	-	-	1.6348296	0.435	0.189
	$n = 1000$					
$\kappa = 1.0$	0.9890723	-0.011	0.000	1.0679778	0.068	0.005
$\beta = 1.5$	1.4717767	-0.027	0.001	1.3175313	-0.182	0.033
$\gamma = 1.0$	0.9738625	-0.026	0.001	0.9937257	-0.006	0.000
$x_1 = 0.7$	0.7287477	-0.029	0.001	0.7084278	0.008	0.000
$x_2 = 1.4$	1.4247266	-0.025	0.001	1.3893326	-0.011	0.000
$x_3 = 1.0$	1.0062952	-0.006	0.000	1.0113317	0.011	0.000
$x_4 = 1.5$	1.5056442	-0.006	0.000	1.5439664	0.044	0.002
$\sigma_b = 1.2$	-	-	-	1.5880305	0.388	0.151

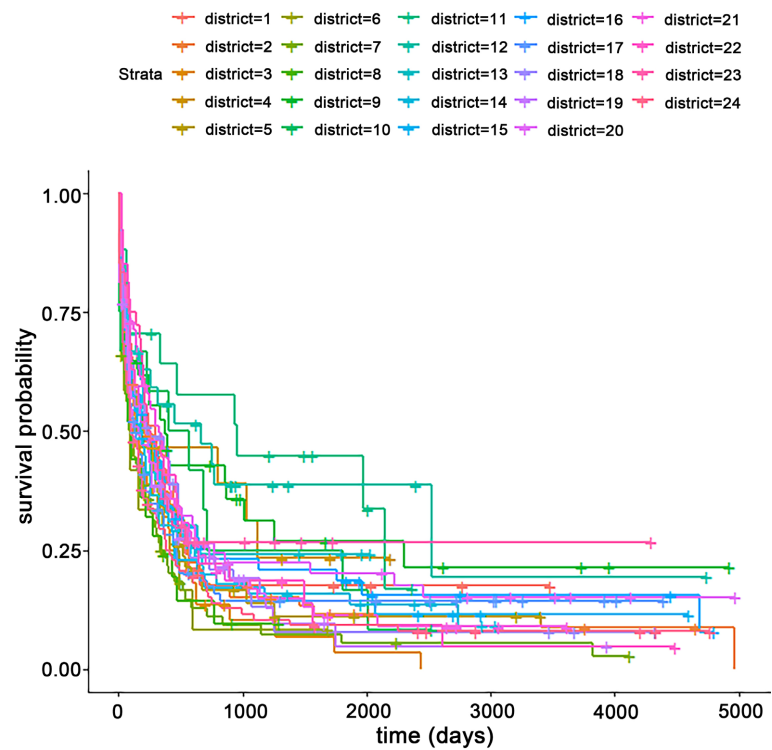


Figure 2. Data exploratory using Kaplan-Meier estimators of survival by district of acute myeloid leukemia data.

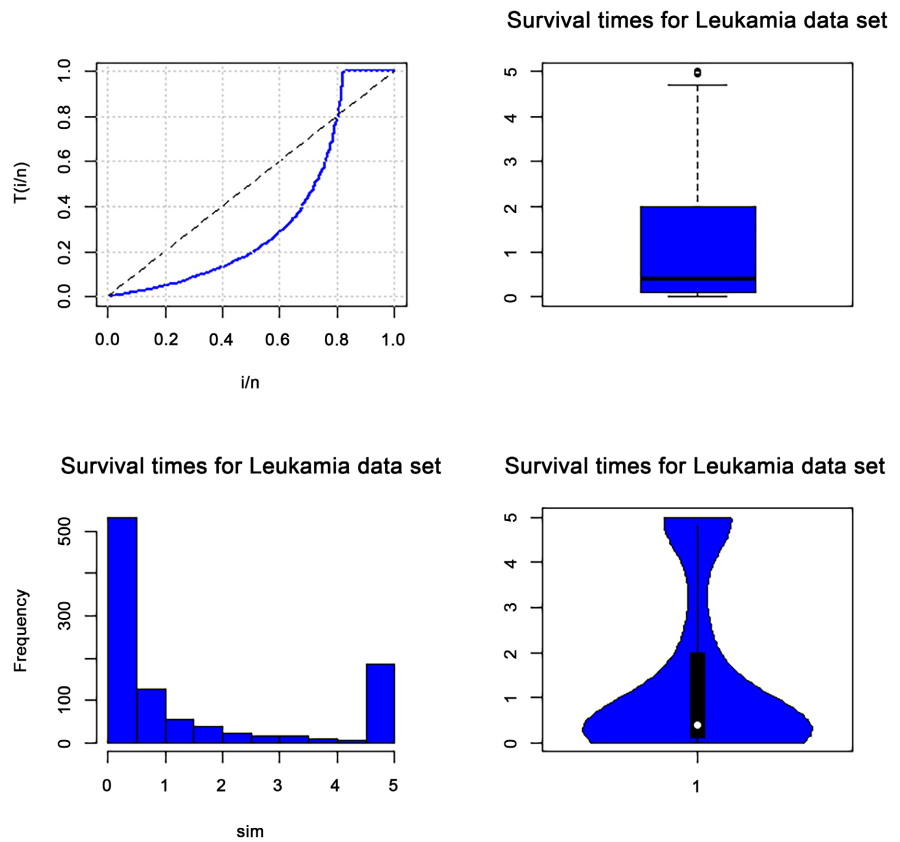


Figure 3. Plots for the survival time data of patients with acute myeloid leukemia that are non-parametric.

$$h(t_{ij} | \mathbf{x}_{ij}, b_i) = h_0(t_{ij}) \exp\{\alpha_1 \times sex_{ij} + \alpha_2 \times age_{ij} + \alpha_3 \times wbc_{ij} + \alpha_4 \times tpi_{ij}\}. \quad (5.2)$$

The results in **Table 4** show the leukemia data application on PH models for four different baseline distributions. It is clear that from the results, all covariates were significant in all PH models which was considered except age, whose p-value is greater than 0.05. However, in terms of models including the random effect as shown in **Table 5**, it is observed that age and tpi were significant when GLL-MEPPH model was used and had an influence on the survivor’s time which coincided with results obtained by [19]. This was difficult to be obtained from the PH models ignoring the random effect. However, using PGW-MEPPH, MKW-MEPPH and W-MEPPH models it was observed that sex, age and tpi were significant and had an influence to survivor’s time and white blood cells covariate had no effects.

The GLL-MEPPH, MKW-MEPPH and W-MEPPH variance component was significant with their p-value being less than 0.05, hence, GLL-MEPPH gave better results with the smallest standard error for variance component when compared to those other competing models. This finding implied that the random effect influences survival time and should be included in the model. This suggested that after taking into account the data on the observed covariates, there would be a large impact on between-cluster variability. Consequently,

Table 4. Hazard ratio, standard errors, p-values and 95% confidence intervals of fitting the GLL-PH, PGW-PH, W-PH and MKW-PH models without random effects.

Models	Parameters	HR	SE	P-Value	Lower-95%	Upper-95%
GLL-PH	κ	1.7054650	0.11248067	<0.0001	1.3680409	2.1261139
	β	0.7645979	0.04444313	<0.0001	0.7008144	0.8341865
	γ	1.0465655	0.32348986	0.8881	0.5551513	1.9729745
	x_1	1.7121844	0.03837526	<0.0001	1.5881278	1.8459317
	x_2	1.0591839	0.06773095	0.3959	0.9275101	1.2095506
	x_3	1.2441444	0.03255735	<0.0001	1.1672340	1.3261225
	x_4	1.1061834	0.03282844	0.0021	1.0372502	1.1796977
PGW-PH	κ	0.1394283	0.22319076	<0.0001	0.09002658	0.2159391
	β	0.8148975	0.05603853	0.0003	0.73013454	0.9095007
	γ	2.5696173	0.10992441	<0.0001	2.07157477	3.1873979
	x_1	1.7089774	0.03836857	<0.0001	1.58517396	1.8424500
	x_2	1.0665224	0.06780616	0.3422	0.93379871	1.2181105
	x_3	1.2426391	0.03266251	<0.0001	1.16558151	1.3247911
	x_4	1.1022306	0.03286101	0.0031	1.03347770	1.1755573
W-PH	κ	0.9217206	0.04982685	0.1019	0.8359621	1.0162768
	β	0.5752874	0.02595840	<0.0001	0.5467503	0.6053139
	x_1	1.7338492	0.03800222	<0.0001	1.6093992	1.8679226
	x_2	1.0694766	0.06769547	0.3211	0.9365885	1.2212196
	x_3	1.2376405	0.03297773	<0.0001	1.1601758	1.3202774
	x_4	1.0958195	0.03274335	0.0052	1.0277035	1.1684502
MKW-PH	κ	6.1441742	0.58892073	0.0021	1.9371947	19.4873937
	β	0.2107892	0.23550132	<0.0001	0.1328585	0.3344317
	γ	4.1779648	0.28291734	<0.0001	2.3996345	7.2741867
	x_1	1.7228612	0.03809666	<0.0001	1.5989039	1.8564285
	x_2	1.0641725	0.06770900	0.3583	0.9319187	1.2151952
	x_3	1.2430676	0.03277612	<0.0001	1.1657238	1.3255430
	x_4	1.1011999	0.03278366	0.0033	1.0326679	1.1742800

Table 5. Results from the fitted mixed effects parametric proportional hazard models to Lukemia data.

Models	Parameters	HR	SE	P-Value	Lower-95%	Upper-95%
GLL MEPPH	κ	9.6874547	0.14039722	<0.0001	7.3570440	12.7560442
	β	1.5139939	0.11897206	0.0005	1.1990987	1.9115838
	γ	0.7272335	0.04738138	<0.0001	0.6627393	0.7980038
	x_1	0.6756300	0.40721368	0.3356	0.3041501	1.5008242

Continued

	x_2	1.7132534	0.03831901	<0.0001	1.5892946	1.8468805
GLL	x_3	1.0723795	0.06780153	0.3027	0.9389355	1.2247890
MEPPH	x_4	1.2314738	0.03321032	<0.0001	1.1538690	1.3142981
	σ_b	1.1046349	0.03284730	0.0024	1.0357599	1.1780900
	κ	8.5077425	0.12330186	<0.0001	6.6812788	12.7560442
	β	0.1702107	0.24401885	<0.0001	0.1055061	1.9115838
	γ	0.7632778	0.05608177	<0.0001	0.6838262	0.7980038
PGW	x_1	2.2486757	0.11503031	<0.0001	1.7947867	1.5008242
MEPPH	x_2	1.7210799	0.03825533	<0.0001	1.5967541	1.8468805
	x_3	1.0276139	0.06738838	0.6861	0.9004692	1.2247890
	x_4	1.0627315	0.03306284	<0.0001	1.1486965	1.3142981
	σ_b	1.0627315	0.03307361	0.0658	0.9960273	1.1780900
	κ	7.7134074	0.11179124	<0.0001	6.1956860	9.6029163
	β	3.5914161	0.29195657	<0.0001	2.0265244	6.3647244
	γ	0.2661626	0.13766152	<0.0001	0.2032213	0.3485978
MKW	x_1	3.1085630	0.17393104	<0.0001	2.2105956	4.3712943
MEPPH	x_2	1.7372767	0.03855626	<0.0001	1.6108305	1.8736485
	x_3	1.0025135	0.06827398	0.9708	0.8769509	1.1460541
	x_4	1.2598252	0.03218787	<0.0001	1.1828017	1.3418645
	σ_b	1.1112844	0.03301650	0.0014	1.0416493	1.1855747
	β	11.0852263	0.16065278	<0.0001	8.0908969	15.1877158
	γ	0.9024489	0.05037524	0.0416	0.8176042	0.9960981
W	x_1	0.5730695	0.02613701	<0.0001	0.5444518	0.6031914
MEPPH	x_2	1.7244479	0.03821774	<0.0001	1.5999967	1.8585793
	x_3	1.0883148	0.06815218	0.2143	0.9522332	1.2438436
	x_4	1.2650992	0.03197148	<0.0001	1.1882570	1.3469105
	σ_b	1.0773307	0.03311348	0.0244	1.0096312	1.1495696

survival gaps between different districts are explained by a combination of complex factors, including different age and Townsend distribution, a conclusion that could have been harder to obtain with fully non-parametric methods that require data stratification.

Model Comparison

The competitor models' best-fitting features are found using specific analytical metrics. To choose the most suited ones, the values of the Hannan-Quinn Information Criterion (HQIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (CAIC) and Akaike Information Criterion (AIC), were employed.

Table 6. Model comparison without random effect.

Models	AIC	CAIC	BIC	HQIC
GLL-PH	1576.932	1577.358	1611.581	1590.074
PGW-PH	1586.483	1586.918	1621.132	1599.626
EW-PH	1580.037	1580.466	1614.686	1593.179
MKW-PH	1589.018	1589.455	1623.667	1602.16
W-PH	1632.28	1632.639	1661.979	1643.544

Table 7. Model comparison with random effect.

Models	AIC	CAIC	BIC	HQIC
GLL-MEPPH	1574.291	1574.717	1608.94	1587.433
PGW-MEPPH	1583.842	1584.277	1618.491	1596.985
EW-MEPPH	1577.396	1577.825	1612.045	1590.538
MKW-MEPPH	1586.377	1586.814	1621.026	1599.519
W-MEPPH	1629.639	1629.998	1659.338	1640.903

$$\text{AIC} = 2m - 2 \log(\hat{l}) \quad (5.3)$$

$$\text{BIC} = m \ln(n) - 2 \log(\hat{l}) \quad (5.4)$$

$$\text{CAIC} = \frac{2nm}{n-m-1} - 2 \log(\hat{l}) \quad (5.5)$$

$$\text{HQIC} = 2m \ln(\ln(n)) - 2 \log(\hat{l}) \quad (5.6)$$

where n = number of observations, m = number of parameters, and \hat{l} = likelihood function's maximum value. The AIC, BIC, CAIC, and HQIC values of the ideal model are the lowest. We observed that GLL-MEPPH and GLL-PH produced the best match and fitting when comparing all the models with and models neglecting random effects since they have the lowest values of the measured analytical tools. Additionally, as seen in **Table 6** and **Table 7**, GLL-MEPPH offered better outcomes with the minimum value of AIC compared to the model ignoring random effects.

6. Conclusion

Despite the fact that parametric survival models are gaining popularity because it has been discovered that they outperform their semiparametric counterparts when applied to clustered survival data [22]. A generalized log-logistic, an expansion of the closed log-logistic model that takes the PH connection into account, was provided by [20], in which the limit approaches the Weibull and displays log-logistic traits. The contribution for this paper is the addition of random effects to the framework used in [20] along with an adjustable parametric method that permits us to demonstrate asymptotic conclusions under usual re-

gularity constraints. Our simulation studies showed that the developed models' estimates, which are based on marginal likelihood, have good frequentist properties. Additionally, they demonstrated that ignoring random effects results in an underestimation of the standard error estimates, which leads to incorrect result interpretation of the covariates that affect survival times. The model was compared to other models using various baseline distributions, and the generalized log logistic mixed-effects proportional hazard model (GLL-MEPPH) performed better than the other models in comparison, according to the results of the information criterion.

Author Contribution

Conceptualization: MWP and SMM, methodology: MWP, SMM, AKW and AHM, software: MWP and AHM, validation MWP, SMM, AKW, data curation MWP and AHM, formula analysis MWP, SMM, AKW and AHM, resources: MWP, SMM, AKW and AHM, writing original draft preparation: MWP, SMM and AKW, writing review and editing MWP, supervision SMM and AKW, project administration MWP.

Data Availability

The real data used to illustrate the developed model is within the manuscript.

Acknowledgment

The corresponding author would like to express her appreciation to the Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI) for supporting this work.

Conflicts of Interest

All authors have read and agreed to the published version of the manuscript.

References

- [1] Hougaard, P. (1995) Frailty Models for Survival Data. *Lifetime Data Analysis*, **1**, 255-273. <https://doi.org/10.1007/BF00985760>
- [2] Omori, Y. and Johnson, R.A. (1993) The Influence of Random Effects on the Unconditional Hazard Rate and Survival Functions. *Biometrika*, **80**, 910-914. <https://doi.org/10.1093/biomet/80.4.910>
- [3] Gutierrez, R.G. (2002) Parametric Frailty and Shared Frailty Survival Models. *The Stata Journal*, **2**, 22-44. <https://doi.org/10.1177/1536867X0200200102>
- [4] 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>
- [5] Klein, J.P. and Moeschberger, M.L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*, Volume 1230. Springer, Berlin. <https://doi.org/10.1007/b97377>
- [6] Kalbfleisch, J.D. and Prentice, R.L. (2002) *The Statistical Analysis of Failure Time*

- Data. John Wiley & Sons, Hoboken. <https://doi.org/10.1002/9781118032985>
- [7] Ripatti, S., Larsen, K. and Palmgren, J. (2002) Maximum Likelihood Inference for Multivariate Frailty Models Using an Automated Monte Carlo EM Algorithm. *Life-time Data Analysis*, **8**, 349-360. <https://doi.org/10.1023/A:1020566821163>
- [8] Ripatti, S. and Palmgren, J. (2000) Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood. *Biometrics*, **56**, 1016-1022. <https://doi.org/10.1111/j.0006-341X.2000.01016.x>
- [9] Vaida, F. and Xu, R.H. (2000) Proportional Hazards Model with Random Effects. *Statistics in Medicine*, **19**, 3309-3324. [https://doi.org/10.1002/1097-0258\(20001230\)19:24<3309::AID-SIM825>3.0.CO;2-9](https://doi.org/10.1002/1097-0258(20001230)19:24<3309::AID-SIM825>3.0.CO;2-9)
- [10] Banerjee, S. and Carlin, B.P. (2003) Semiparametric Spatio-Temporal Frailty Modeling. *Environmetrics: The Official Journal of the International Environmetrics Society*, **14**, 523-535. <https://doi.org/10.1002/env.613>
- [11] Xu, R.H. (2004) Proportional Hazards Mixed Models: A Review with Applications to Twin Models. *Advances in Methodology and Statistics*, **1**, 205-212. <https://doi.org/10.51936/lmzi2020>
- [12] Austin, P.C. (2017) A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. *International Statistical Review*, **85**, 185-203. <https://doi.org/10.1111/insr.12214>
- [13] Li, Yi and Ryan, L. (2002) Modeling Spatial Survival Data Using Semiparametric Frailty Models. *Biometrics*, **58**, 287-297. <https://doi.org/10.1111/j.0006-341X.2002.00287.x>
- [14] Hanson, T.E., Jara, A. and Zhao, L.P. (2012) A Bayesian Semiparametric Temporally Stratified Proportional Hazards Model with Spatial Frailties. *Bayesian Analysis*, **7**, 147-188. <https://doi.org/10.1214/12-BA705>
- [15] Li, L., Hanson, T. and Zhang, J.J. (2015) Spatial Extended Hazard Model with Application to Prostate Cancer Survival. *Biometrics*, **71**, 313-322. <https://doi.org/10.1111/biom.12268>
- [16] Crowther, M.J., Look, M.P. and Riley, R.D. (2014) Multilevel Mixed Effects Parametric Survival Models Using Adaptive Gauss-Hermite Quadrature with Application to Recurrent Events and Individual Participant Data Meta-Analysis. *Statistics in medicine*, **33**, 3844-3858. <https://doi.org/10.1002/sim.6191>
- [17] Li, J., Hong, Y.L., Thapa, R. and Burkhart, H.E. (2015) Survival Analysis of Loblolly Pine Trees with Spatially Correlated Random Effects. *Journal of the American Statistical Association*, **110**, 486-502. <https://doi.org/10.1080/01621459.2014.995793>
- [18] Zhou, H.M. and Hanson, T. (2018) A Unified Framework for Fitting Bayesian Semiparametric Models to Arbitrarily Censored Survival Data, Including Spatially Referenced Data. *Journal of the American Statistical Association*, **113**, 571-581. <https://doi.org/10.1080/01621459.2017.1356316>
- [19] Alvarez, F.J.R. and Drikvandi, R. (2022) MEGH: A Parametric Class of General Hazard Models for Clustered Survival Data. *Statistical Methods in Medical Research*.
- [20] Khan, S.A. and Khosa, S.K. (2016) Generalized Log-Logistic Proportional Hazard Model with Applications in Survival Analysis. *Journal of Statistical Distributions and Applications*, **3**, 1-18. <https://doi.org/10.1186/s40488-016-0054-z>
- [21] Muse, H.A., Ngesa, O., Mwalili, S., Alshanbari, H.M. and El-Bagoury, A.-A. (2022) A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data. *Journal of Healthcare Engineering*, **2022**, Article ID: 2051642. <https://doi.org/10.1155/2022/2051642>

- [22] Prenen, L., Braekers, R. and Duchateau, L. (2017) Extending the Archimedean Copula Methodology to Model Multivariate Survival Data Grouped in Clusters of Variable Size. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 483-505. <https://doi.org/10.1111/rssb.12174>
- [23] Alvares, D. and Rubio, F.J. (2021) A Tractable Bayesian Joint Model for Longitudinal and Survival Data. *Statistics in Medicine*, **40**, 4213-4229. <https://doi.org/10.1002/sim.9024>
- [24] Rossell, D. and Rubio, F.J. (2022) Additive Bayesian Variable Selection under Censoring and Misspecification. *Statistical Science*, **38**, 13-29. <https://doi.org/10.1214/21-STS846>
- [25] Rubio, F.J., Remontet, L., Jewell, N.P. and Belot, A. (2019) On a General Structure for Hazard-Based Regression Models: An Application to Population-Based Cancer Research. *Statistical Methods in Medical Research*, **28**, 2404-2417. <https://doi.org/10.1177/0962280218782293>
- [26] Leemis, L.M., Shih, L.-H. and Reynertson, K. (1990) Variate Generation for Accelerated Life and Proportional Hazards Models with Time Dependent Covariates. *Statistics & Probability Letters*, **10**, 335-339. [https://doi.org/10.1016/0167-7152\(90\)90052-9](https://doi.org/10.1016/0167-7152(90)90052-9)
- [27] Henderson, R., Shimakura, S. and Gorst, D. (2002) Modeling Spatial Variation in Leukemia Survival Data. *Journal of the American Statistical Association*, **97**, 965-972. <https://doi.org/10.1198/016214502388618753>
- [28] Fahrmeir, L., Kneib, T., *et al.* (2011) Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780199533022.001.0001>