MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Suleiman *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 21 (2025) 2640–2661

RESEARCH ARTICLE

Parametric Model Fit with Cure Fraction in Prostate Cancer Survival: A Simulation Study on True Model Selection using MLE

Haruna Suleiman, Noraslinda Mohamed Ismail*, Shariffah Suhaila Syed Jamaludin

Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 UTM Johor Bahru, Johor, Malaysia

Abstract: This research aims to evaluate and select the best model from multiple options to establish a reliable baseline for modeling prostate cancer by comprehensively assessing generalized survival models. Specifically, the study examines the Generalized Weibull, Generalized Log-Normal, and Generalized Exponential models while incorporating mixture cure fraction (MCF), Proportional Hazard (PH), covariates, and right-censoring. These generalized models enhance flexibility in handling cure fractions and advance the understanding of survival analysis in prostate cancer research. A simulation-based approach was used with sample sizes of 500, 1000, and 2000, generated using true parameter estimates derived from real-life prostate data. The models were estimated via MLE and assessed using AIC, BIC, and cross-validation to determine model fit. Results indicate that the Generalized Weibull model consistently outperforms other models, particularly in scenarios where treatment, age, and PSA are the key predictors. Moreover, the Generalized Weibull model emerges better as it maintains a CF closest to the true value 0.02 in all sample sizes. The Generalized Log-Normal distribution excels when Gleason is influential. The Generalized Exponential, though reasonable, generally underperforms relative to the other models. A key novelty of this study is the demonstration of the Generalized Weibull function's superior performance as a baseline hazard in prostate cancer survival modeling, especially with cure fraction and right-censored data.

Keywords: Cure fraction, Generalized-Weibull, Hazard ratio, MLE, Simulation.

Introduction

In contemporary survival analysis, selecting an appropriate baseline hazard model is important for diseases like prostate cancer, especially given advancements in treatment that have improved long-term survivors (cure fraction), as a result, the mostly used Cox proportional hazards model(PHM) (Cox, 1972) by researchers may not be appropriate for certain survival data such as those from the population, in which a significant proportion of patients are cured. Due to recent medical advances, a substantial proportion of patients are cured of various types of cancers such as breast cancer, head and neck cancer, melanoma, and prostate cancer [1]. As personalized medicine gains importance, accurately predicting survival outcomes is essential. This paper specifically evaluates various parametric models, focusing on their generalized forms: Generalized Weibull, Generalized Log-Normal, and Generalized Exponential. These generalized models introduce additional parameters, allowing for greater flexibility in modeling hazard functions. Utilizing modern simulation techniques and Maximum Likelihood Estimation (MLE), the study aims to identify the best-fitting model to capture the complex survival dynamics of prostate cancer, thereby enhancing predictions in an era of individualized treatment.The research was motivated to propose a new novel modified model for modeling the survival time of prostate cancer patients by considering a parametric model for the baseline hazard function, PHM (Cox, 1972), and a mixture cure fraction model (MCM) [2] with covariates and right-censoring all incorporated due to the highly efficient and flexibility of the three confounded model properties which provides an enticing model fitting at a different level and kinds of large datasets.

***For correspondence:**
noraslinda@utm.my

Received: 7 March 2025
Accepted: 13 August 2025

Mixture Cure Model

The mixture cure rate model, commonly known as the standard cure rate model, is the most widely used cure fraction model. It was initially developed by [2], and further refined by [3]. Since then, this model has been extensively studied by various researchers, including [4], [5], [6], [7], [8], [9], and [10], among others.

In the mixture cure rate model, the study population is assumed to consist of a certain fraction of long-term survivors or non-susceptible individuals. In contrast, the remaining fraction is susceptible to the event of interest, such as cancer recurrence. This model is especially pertinent in prostate cancer studies, where recent advancements in treatments, such as targeted therapies and immunotherapies, have improved the prognosis for many patients. As a result, a subset of prostate cancer patients can be considered "cured," showing a significantly reduced likelihood of disease progression or relapse over time. Understanding this model helps researchers evaluate treatment effectiveness and identify factors that contribute to long-term survival among prostate cancer patients, ultimately aiding in developing more personalized treatment strategies

Related Papers

The cure rate model, also called the long-term survival model, is crucial in survival statistics, especially in cancer clinical trials where a cure fraction is often observed. Researchers have widely explored these models in cancer oncology and epidemiology. The mixture cure fraction model was first proposed by [11]. There is a substantial body of literature on cure rate models for cancer studies, with notable contributions from [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]. [28], [29] discussed frequentist inference techniques in detail for the mixture cure rate model, while the Bayesian approach was examined by [30]. [31] thoroughly analyse cure rate estimation from Bayesian and frequentist viewpoints.

Recently, [32] applied the exponentiated Weibull model to medical data in the presence of right-censoring, cure fraction, and covariates. [33], explained the application of the mixture cure fraction model in the oncology study of cancer therapies associated with survival patterns that differ from established therapies, which may include survival curves that plateau after a certain follow-up time point. Bayesian modeling of survival data in the presence of competing risks with cure fractions and masked causes is studied by [27]. [34], explore three regression frameworks by utilizing mixture models that account for the nonsusceptible fraction by simulation studies using data from the Prostate, Lung, Colorectal, and Ovarian Cancer screening trials.

In this article, we explore the application of the generalized parametric models as a baseline hazard functions for predicting prostate cancer patients' survival times. We utilize various parametric models, treating one as the 'True Model,' and systematically fit competing models using MLE to identify the most suitable baseline hazard function. This analysis incorporates a cure fraction to account for long-term survivors, relevant covariates such as age, tumour, treatment, Gleason, and PSA levels, and right-censoring to handle incomplete follow-up data, ensuring a robust estimation of survival probabilities.

Materials and Methods

Model Design

To include the specific covariates for prostate cancer data simulated and used in the research (age, tumor, treatment, PSA, and Gleason) into the likelihood and the log-likelihood function and the corresponding hazard and survival functions for the Weibull (W), Log-Normal (LN), and Exponential (E) models, and the Generalized Weibull (GW), Generalized Lognormal (GL), and the Generalized Exponential (GE), we modify the formulae as follows; Let the covariates be as:

X_1 represents age, X_2 represents Tumour, X_3 represents Treatments, X_4 represents PSA, and X_5 represents Gleason.

The linear predictor for the covariates is expressed as:

$$\tau_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} \quad (1)$$

where β_0 is the intercepts and β_1, β_2, β_3, β_4, and β_5 are the coefficients associated with age, tumor size, treatment (surgery, radiation, and chemotherapy), PSA, and Gleason score, respectively. The hazard function for prostate cancer survival is defined by the following:

$$h(t/X_{i1}, X_{i2}, X_{i3}X_{i4}, X_{i5}) = \lim_{\Delta t} \frac{P(t \le T < t+\Delta t / T \ge t, X_{i1}, X_{i2}, X_{i3}X_{i2}, X_{i3})}{\Delta t} \quad (2)$$

where *t* is the survival time of the prostate cancer patient, T is the time to the event (death), X_{i1}, X_{i2},X_{i3}, X_{i4}, and X_{i5} represent the covariates, and *p(.)* is the condition of the event in the interval (t,t+Δt). The survival function conditional on covariates- age, tumor stage, treatment type, PSA levels, and Gleason score is defined as:

$$S\ (t/X_{i1},\ X_{i2},\ X_{i3},\ X_{i4},\ X_{i5}) = P\ (T > t/X_{i1},\ X_{i2},\ X_{i3},\ X_{i4},\ X_{i5}) \quad (3)$$

where *T* is the time to the event (death), and X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5} the covariates

Generalized Weibull Model without Cure Fraction

The baseline hazard function $\mathrm{h}(t/X, \lambda, \alpha, \gamma, \beta)$ for the Generalized Weibull distribution model with cure fraction and covariates is written as below:

$$\mathrm{h}(t/X, \lambda, \alpha, \gamma, \beta) = \frac{\alpha.\gamma\left(\frac{t}{\lambda}\right)^{\alpha-1}.\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right)}{\lambda.\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right)} \quad (4)$$

where $\alpha,\ and\ \lambda\ are\ all > 0$ represents the shape(α) and the scale(λ) parameters. The γ is the shape parameter that controls the flexibility of the hazard called the generalization parameter, and X is the vector of the covariates β is the vector of the coefficients of the covariates. The survival function for the GW distribution model is given by:

$$\mathrm{S}(\mathrm{t}/\mathrm{X}, \lambda, \alpha, \gamma) = \exp\left(-\left(\frac{\mathrm{t}}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right) \quad (5)$$

The corresponding likelihood function for the GW model is:

$$\mathrm{L}(X, \alpha, \beta, \gamma/t, X, \delta) = \prod_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{h}(t/X, \lambda, \alpha, \gamma, \beta)^{\delta}.S(t/X, \lambda, \alpha, \gamma)\right) \quad (6)$$

We take the natural log of (6) to obtain the log-likelihood function below:

$$l(\lambda, \alpha, \gamma, \beta, \mathrm{t}, \mathrm{X}, \sigma) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \delta_{\mathrm{i}} \log.\frac{\alpha.\gamma\left(\frac{t}{\lambda}\right)^{\alpha-1}.\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right)}{\lambda.\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right)} + (1-\delta_{\mathrm{i}})\log.\exp\left(-\left(\frac{\mathrm{t}}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right) \quad (7)$$

It is further simplified to obtain the MLEs for the parameters.

Generalized Weibull model with cure fraction

The hazard function for the Generalized Weibull distribution with a cure fraction is:

$$\mathrm{h}_{\mathrm{cure}}(\mathrm{t}/\mathrm{X}, \lambda, \alpha, \gamma, \mathrm{p}) = \alpha.\gamma.\left(\frac{\mathrm{t}}{\lambda}\right)^{\alpha-1}\left(-\left(\frac{\mathrm{t}}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right) \quad (8)$$

where $\alpha\ and\ \lambda$ are the shape and scale parameters, γ the generalized parameter, X is the vector of the covariates, β is the vector of the coefficients of the covariates, and p is the cure fraction, representing the proportion of cured patients. For a patient who is not cured, the survival function is given as below:

$$\mathrm{S}_{\mathrm{non-cured}}(\mathrm{t}/\mathrm{X}, \lambda, \alpha, \gamma) = \exp\left(-\left(\frac{\mathrm{t}}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\alpha}\right) \quad (9)$$

Including the cure fraction p, the survival function for the whole population is;

$$\mathrm{S}_{\mathrm{cure}}(\mathrm{t}/\mathrm{X}, \lambda, \alpha, \gamma, \mathrm{p}) = \mathrm{p} + (1-\mathrm{p}).\exp\left(-\left(\frac{\mathrm{t}}{\lambda}\right)^{\alpha}.\exp(X\beta)^{\gamma}\right) \quad (10)$$

Below is the corresponding log-likelihood function for the GW model with cure fraction. It is simplified to obtain the MLEs for the parameters

$$l_{\text{cure}} = \sum_{i=1}^{n} \begin{Bmatrix} \delta_i . \log(1-p) . \gamma . \left(\frac{t_i}{\lambda}\right)^{\gamma-1} . \exp\left(-\left(\frac{t_i}{\lambda}\right)^{\gamma} . \exp(X_i\beta)\right) + \\ (1-d_i) . \log\left(S_{\text{cure}}(t_i/X, \lambda, \gamma, p)\right) \end{Bmatrix} \quad (11)$$

where δ_i is the censoring indicator, and t_i is the observed time. It is further simplified to obtain the MLEs for the parameters

Generalized Log-Normal Model without cure fraction

The baseline hazard function for the Generalized Lognormal distribution model is given by:

$$h(t/X, \mu, \sigma, \gamma) = \frac{\frac{1}{t.\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(t)-\mu-X\beta)^2.\gamma^2}{2\sigma^2}\right)}{\Phi\left(\frac{\log(t)-\mu-X\beta}{\sigma}.\gamma\right)} \quad (12)$$

where μ and σ are the lognormal distribution mean and the standard deviation parameters, γ is the generalization parameter, X is the vector covariates, β is the vector coefficients for covariates, and Φ the standard normal CDF. The survival function for the Generalized lognormal without cure fraction and the covariates is given by:

$$S(t/X, \mu, \sigma, \gamma) = 1 - \Phi\left(\frac{\log(t)-\mu-X\beta}{\sigma}\right)^2 \quad (13)$$

The corresponding likelihood for the GL without cure fraction is given by:

$$L(\mu, \sigma, \beta, \gamma, t, X, \delta) = \prod_{i=1}^{n} \left(\frac{1}{t.\sigma.\sqrt{2\pi}} \left(\frac{(\log(t)-\mu-X_i\beta)^2}{2\sigma^2}\right)\right)^{\delta_i} . S\left(\frac{t_i}{X_{i,\mu}}, \sigma, \gamma\right)^{1-\delta_i} \quad (14)$$

we take the natural log of (14) to obtain the log-likelihood function below:

$$l(, \sigma, \beta, \gamma, , X, \delta) = \sum_{i=1}^{n} \left\{\delta_i \left(-\log\left(t_i\sigma\sqrt{2\pi}\right) - \frac{(\log(t_i) - \mu - X_i\beta)}{2\sigma^2}\right)\right\} + \sum_{i=1}^{n}(1-\delta_i)\log\left(1 - \Phi\left(\frac{\log(t)-\mu-X\beta}{\sigma}\right)^2\right) \quad (15)$$

where δ_i is the censoring indicator, and t_i is the observed time. It is further simplified to obtain the MLEs for the parameters

Generalized Log-Normal Model with cure fraction

The hazard function for the Generalized Lognormal distribution with a cure fraction is:

$$h_{\text{cure}}(t/X, \mu, \sigma, \gamma, p) = \frac{\frac{1}{t.\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(t)-\mu-X\beta)^2.\gamma^2}{2\sigma^2}\right)}{\Phi\left(\frac{\log(t)-\mu-X\beta}{\sigma}.\gamma\right)} \quad (16)$$

where μ and σ are the lognormal distribution parameters, γ is the generalization parameter, X is the vector covariates, β is the vector coefficients for the covariates, Φ is the standard normal CDF, and P is the cure fraction, representing the proportion of cured patients. For a patient who is not cured, the survival function is given by:

$$S_{\text{non-cured}}(t/X, \lambda, \alpha, \gamma) = \exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right) . \exp((X\beta)^{\gamma}) \quad (17)$$

Including the cure fraction p, the survival function for the whole population is;

$$S_{\text{cure}}(t/X, \mu, \sigma, \gamma) = p + (1-p) . \exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha} . \exp(X\beta)^{\gamma}\right) \quad (18)$$

The log-likelihood function for the Generalized Log-Normal distribution with cure fraction is given below. It is simplified to obtain the MLEs for the parameters

$$l_{\text{cure}} = \sum_{i=1}^{n} \left\{\delta_i . \log\left((1-p) . \frac{1}{t_i.\sigma.\gamma.\sqrt{2\pi}} . \exp\left(-\frac{(\log(t_i)-\mu-X_i\beta)^2}{2\sigma^2.\gamma^2}\right)\right)\right\} \quad (19)$$

where δ_i is the censoring indicator and t_i is the observed time. It is further simplified to obtain the MLEs for the parameters.

Generalized Exponential Model without cure fraction

$$h(t/X,\lambda) = k\lambda^{-\gamma}t^{\gamma-1}.\exp(X\beta) \tag{20}$$

where λ is the scale parameter, γ is the shape parameter, k is the location parameter, X is the covariates, and β is the vector of the coefficient for the covariates. The survival function for the Generalized Exponential without cure fraction and the covariates is given by:

$$S(t/X,\lambda,\gamma) = \exp\left(-\frac{k}{\gamma}\lambda^{-\gamma}e^{X\beta}t^{\gamma}\right) \tag{21}$$

The corresponding likelihood for the GE without cure fraction is given by:

$$L(X,\beta,t,\gamma,X,\delta) = \prod_{i=1}^{n}\left(\frac{\gamma}{\lambda}\left(\frac{t_i}{\lambda}\right)^{\gamma-1}\exp\left(-\left(\frac{t_i}{\lambda}\right)^{\gamma}\exp(X_i\beta)\right)\right)^{\delta_i} .S(t_i/X,\lambda,\ \gamma)^{1-\delta_i} \tag{22}$$

we take the natural log of (22) to obtain the log-likelihood function below:

$$l(\lambda,\alpha,\beta/t,X,\sigma) \sum_{i=1}^{n}\left\{\begin{matrix}\log(\gamma)-\log(\lambda)+(\gamma-1)\ln(t_i)-\gamma\log(\gamma)-\left(\frac{t_i}{\lambda}\right)^{\gamma}\exp(X_i\beta)\\ -\sum_{i=1}^{n}\left(\frac{t_i}{\lambda}\right)^{\gamma}\exp(X_i\beta)\end{matrix}\right\} \tag{23}$$

where δ_i is the censoring indicator and *ti* is the observed time. It is further simplified to obtain the MLEs for the parameters

Generalized Exponential Model with cure fraction

The hazard function for the Generalized Exponential distribution with a cure fraction is:

$$h_{cure}(t/X,\lambda,p) = (1-P)k\lambda^{-\gamma}t^{\gamma-1}.\exp(X\beta) \tag{24}$$

where λ is the rate, t_i is the survival time, X is the vector of the covariates, and β is the vector of coefficients for the covariates, and p is the cure fraction representing the proportion of cured patients. For a patient who is not cured, the survival function is given by:

$$S_{non-cured}(t/X,\lambda,\gamma) = \exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}.\exp(X\beta)\right) \tag{25}$$

Including the cure fraction p, the survival function for the whole population is;

$$S_{cure}(t/X,\lambda,\gamma,p) = (1-P).\exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}.\exp(X\beta)\right)+p \tag{26}$$

The Log-likelihood function for the Generalized Exponential model with cure fraction is given below;

$$l_{cure} = \sum_{i=1}^{n}\left\{\begin{matrix}\delta_i\log(1-p).\gamma.\left(\frac{t_i}{\lambda}\right)^{\gamma-1}.\exp\left(-\left(\frac{t_i}{\lambda}\right)^{\gamma}.\exp(X_i\beta)\right)\\ +(1-d_i)\log(S_{cure}(t/X,\lambda,\gamma,p))\end{matrix}\right\} \tag{27}$$

where δ_i is the censoring indicator and *ti* is the observed time. It is simplified to obtain the MLEs for the parameters.

Simulation Design

The study used R software to simulate survival data with parametric and generalized parametric models, including a cure fraction, to evaluate how well they fit prostate cancer survival data. Bayesian estimation was employed to estimate and account for parameter variability based on real-world clinical covariates from a publicly available data repository on Kaggle (Ashraf Alsinglawi, Prostate Cancer Survival Data). It was originally published on Figshare as part of a web-based prediction model for cancer-specific

survival in elderly patients undergoing surgery for prostate cancer (Figshare Dataset Link). The data was gathered through medical chart reviews and structured telephone interviews conducted at Tikur Anbessa Specialized Hospital in Addis Ababa, Ethiopia.

The research aimed to assess how accurately these models fit survival data when the true underlying distribution is known. The covariates, age, tumor, PSA levels, treatment type, and Gleason score, were included in the simulation to account for patient-specific characteristics. The study thus explored the efficacy of these models in survival analysis, particularly when dealing with heterogeneous patient populations and Curefractions.

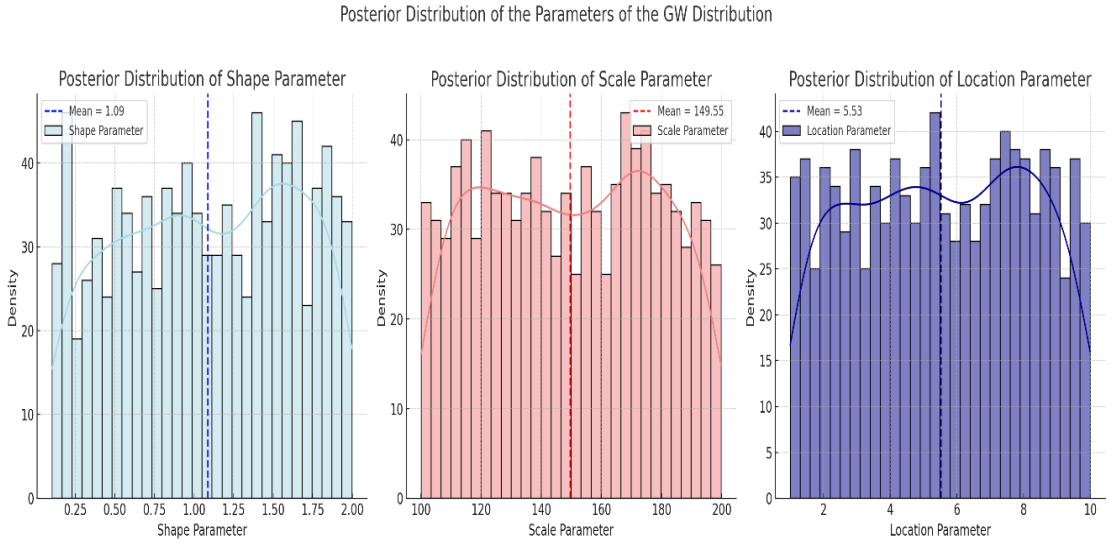

Figure 1. A plot showing the estimated posterior distributions for the GW parameters from real-life prostate data.

The posterior distributions of the Generalized Weibull (GW) parameters, with a shape parameter mean of approximately (1.09), a scale parameter mean of (149.55) and a location parameter mean of (5.53) provide a robust benchmark for simulating prostate cancer data by capturing realistic survival variability, baseline time-to-event characteristics, and flexible hazard rate behaviors, enabling the creation of synthetic datasets that closely mirror real-world clinical scenarios for model development and validation.

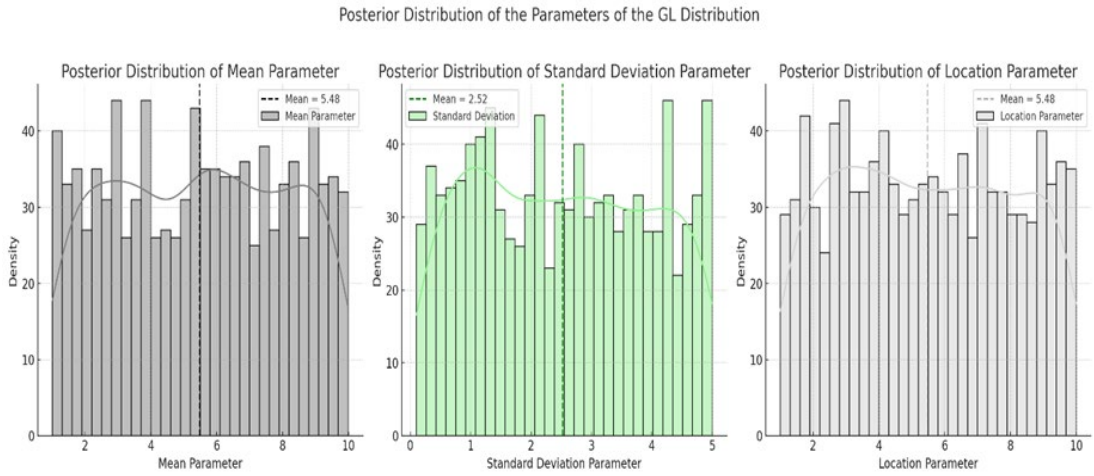

Figure 2. A plot showing the estimated posterior distributions for the GL parameters from real-life prostate data

The posterior distributions of the Generalized Lognormal (GL) distribution parameters, estimated from real-life prostate cancer data, indicate that the mean is centered around (5.48), the standard deviation is approximately (2.51), and the location parameter is also around (5.48). This suggests that the survival times follow a right-skewed distribution, with most values clustering near these estimates while allowing for variability. The relatively large standard deviation compared to the mean reflects high dispersion in the data, meaning that while the average survival time is around (5.48) months, many individuals may experience significantly shorter or longer durations. This aligns with the characteristics of a Lognormal-type distribution, where heavy tails and greater variability are common.

These estimated parameters will simulate prostate cancer survival data using the Generalized Lognormal distribution. By drawing random samples from the posterior distributions, synthetic survival times can be generated, incorporating the variability and uncertainty in the parameter estimates. This approach ensures that the probabilistic nature of survival outcomes is accurately represented while maintaining consistency with observed data. Since the distribution exhibits right-skewness and a long tail, it captures the complexity of real-world prostate cancer survival patterns, making it a suitable choice for modeling.

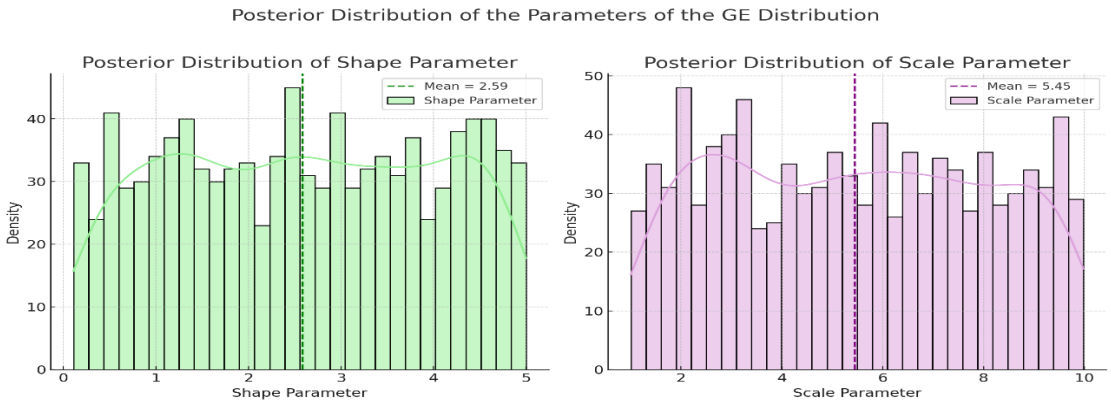

Figure 3. A plot showing the estimated posterior distributions for the GL parameters from real-life prostate data.

The posterior distributions of the Generalized Exponential (GE) distribution parameters suggest their suitability for simulating prostate cancer survival data. With a shape parameter mean of approximately 2.59, capturing the moderately steep decline in survival probabilities often observed shortly after diagnosis or treatment, reflecting an initially higher risk of adverse events that stabilizes over time. The scale parameter mean of approximately 5.45 ensures the inclusion of realistic variability in survival times, aligning with the diverse outcomes typically seen among prostate cancer patients. These estimated values provide a robust foundation for generating synthetic datasets that mirror real-world clinical scenarios, making them highly applicable for model development and validation in prostate cancer research.

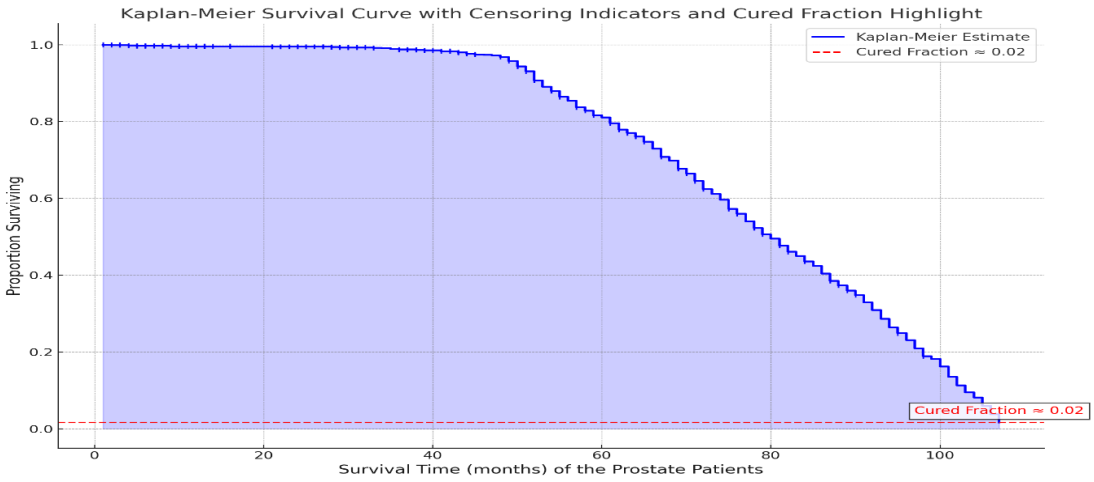

Figure 4. Kaplan-Meier curve for the proportion of the cured patients from the real-life dataset

The Kaplan -Meier survival curve illustrates the survival probabilities of prostate cancer patients over time in months. The curve shows a decline in survival probabilities as time progresses, indicating the occurrence of events such as death or disease progression. Small vertical blue lines represent censoring instances where patient outcomes were unknown due to loss to follow-up or the end of the marking study period. The red dashed line highlights the cured fraction, representing the proportion of patients who survived without experiencing the event by the end of the observation period. In this analysis, the cured fraction is approximately 2%, demonstrating a notable but insignificant proportion of long-term survivors. The low cure rate of roughly 2% in prostate cancer patients can be attributed to several critical factors, including disease aggressiveness, late-stage diagnosis, and limited treatment access. The aggressiveness of certain prostate cancer types, such as intraductal carcinoma of the prostate (IDC-P), is associated with increased metastasis and recurrence, contributing to lower survival rates [35]. Additionally, the choice of treatment modality, such as surgery versus radiotherapy, plays a significant role in patient outcomes, with radical prostatectomy often showing better survival outcomes than radiotherapy [36]. Real-world data often reflects poorer outcomes compared to randomized controlled trials, indicating potential disparities in treatment access and patient characteristics [37]. The X-axis is labeled "Survival Time (months) of the Prostate Patients, indicating the time scale for this survival analysis.

Results and Discussion

Histogram and the Density plots by simulated prostate data using GW, GL, and GE without cure fraction.

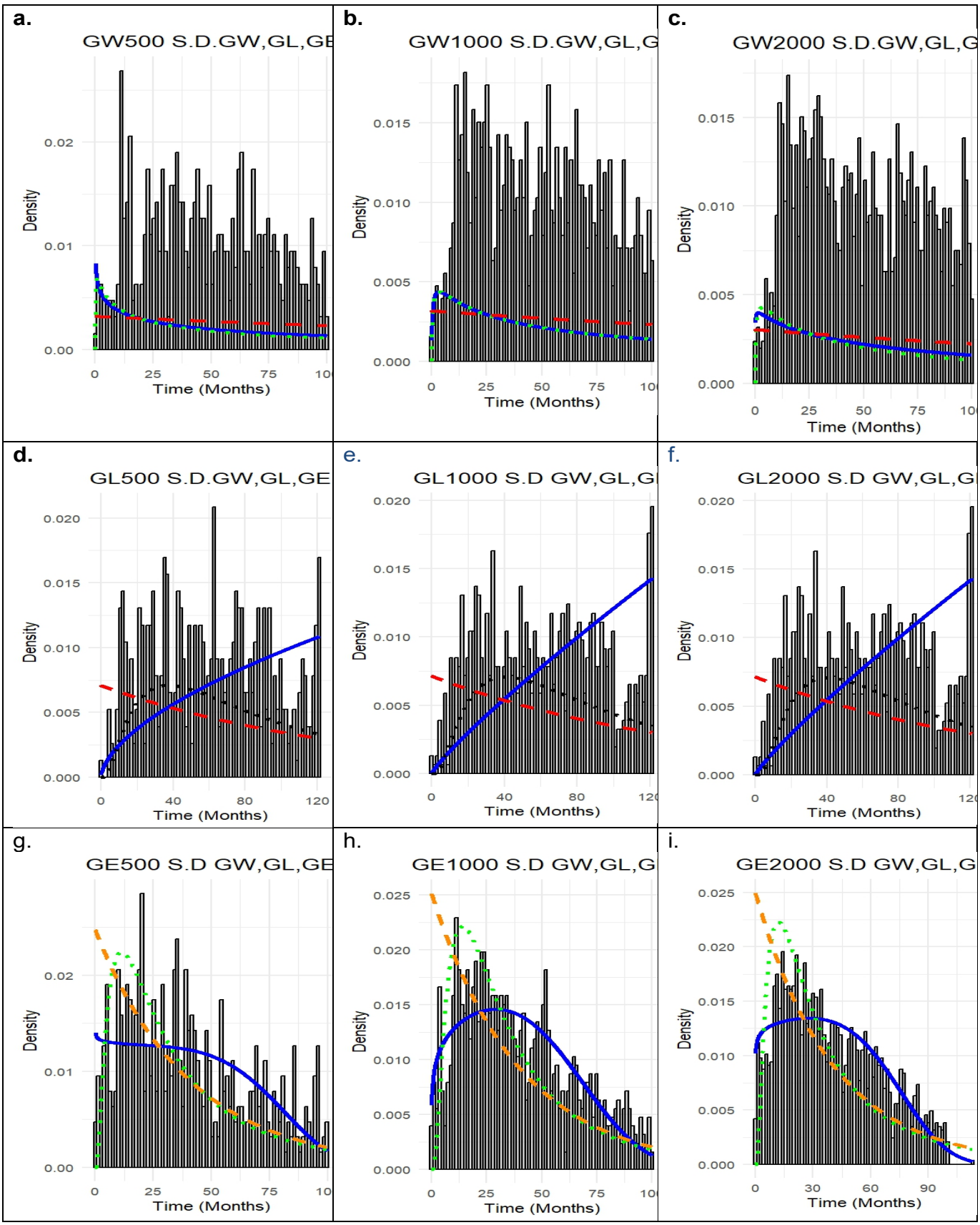

Figure 5. Histogram and the survival density plot from prostate data simulated using GW, GL, and GE

For the GW the density plots **a, b** and **c** for simulated survival data across sample sizes 500, 1000, and 2000 consistently show that the Generalized Weibull (GW) model provides the best fit, particularly in the early survival period (0–25 months), closely following the histogram's density pattern. The Generalized Lognormal (GL) and Generalized Exponential (GE) models exhibit deviations, with GE performing the worst, failing to capture the early rapid decline and showing more stability than observed in the actual distribution. The increase in sample size leads to a smoother histogram, reinforcing the superiority of the GW model in capturing survival characteristics. This finding is further supported by AIC and BIC results, which consistently favour GW as the most appropriate baseline hazard function for modelling the survival time of prostate cancer data, making it the most reliable choice for further analysis. This conclusion aligns with findings in the literature, such as the study by [38], which evaluated the performance of AIC and BIC in selecting true survival models under varying levels of censoring.

The density plots **d, e**, and **f** for he Generalized Lognormal (GL) model (black curve) best fits the empirical survival distribution, closely following the histogram in both the three sample simulated prostate data. In contrast, the Generalized Weibull (GW) model (blue curve) appears to represent a hazard function rather than a survival density, leading to deviations. The Generalized Exponential (GE) model (orange dashed curve) shows the weakest fit, underestimating early survival times and diverging at later times. Thus, the GL model is the most suitable for modeling the simulated prostate cancer survival data. This was in line with what was recently found by [39]. The GW's flexibility in modeling various survival time distributions makes it particularly effective for prostate cancer data, which can exhibit diverse progression patterns.

The density plots **g, h** and **i** for the GE Across all three datasets of different sample sizes, the GE model (orange dashed line) provides the best fit in early survival months, effectively capturing high-risk early events but underestimating survival beyond midpoints, while the GL model (green dotted line) remains a moderate performer, following the data trend without excelling in any specific phase, and the GW model (blue solid line), despite struggling in early months by overestimating survival rates, progressively improves in mid-to-late survival periods, making it more suitable for long-term survival trend analysis..

A. Maximum Likelihood Estimates (MLE) for GW-CF, GL-CF, and GE-CF PH Models Based on Simulated Prostate Cancer Data (GWCF as the True Model) With a Cure Fraction for Sample Sizes of 500, 1000, and 2000.

Table 1. Estimated parameters from various models on GWCF-Simulated data 500 sample

Model 500	Covariate	Estimate	Std-Error	Z-value	P-value	Hazard-Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
Generalized Weibull	λ	77.189	2.4593	2.854	0.0043		0.1997	2521.785	2559.717	0.1997	149.55	-72.361
	α	0.701	0.3357	2.0878	0.0368						1.09	-0.389
	γ	0.4213	0.4768	0.8836	0.3769						5.53	-5.108
	β_{age}	-0.0112	0.0336	-0.3332	0.739	0.9889						
	β_{tumor}	0.2798	0.2175	1.2866	0.1982	1.3229						
	β_{trtm}	-0.1303	0.3355	-0.3885	0.6977	0.8778						
	β_{PSA}	-0.0575	0.0326	-1.7621	0.0781	0.9442						
	$\beta_{gleason}$	0.087	0.1368	0.636	0.5247	1.0909						
Generalized Lognormal	μ	1.3795	1.2332	-1.1186	0.2633		0.2001	2515.286	2557.432	0.20006	5.48	-4.101
	σ	1.355	14.4178	0.094	0.9251						5.48	-4.125
	λ	1.4451	1.841	0.785	0.4325						2.52	-1.074
	β_{age}	-0.0126	0.0332	-0.3784	0.7051	0.9875						
	β_{tumor}	0.2727	0.2133	1.2785	0.2011	1.3135						
	β_{trtm}	-0.1313	0.3324	-0.3951	0.6927	0.8769						
	β_{PSA}	-0.0562	0.0319	-1.7606	0.0783	0.9453						
	$\beta_{gleason}$	0.088	0.1353	0.6507	0.5153	1.092						
Generalized Exponential	γ	0.016	0.0058	-2.7553	0.0059		0.1684	2521.789	2559.721	0.16842	2.59	-2.574

Model 500	Covariate	Estimate	Std-Error	Z-value	P-value	Hazard-Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
	λ	5.706	1.6908	-3.3748	0.0007						5.45	0.256
	k	0.008	0.0233	0.345	0.7301							
	β_{age}	-0.1688	0.1444	-1.1688	0.2425	0.8447						
	β_{tumor}	0.1193	0.2347	0.5081	0.6114	1.1267						
	β_{trtm}	0.0371	0.0202	1.8338	0.0667	1.0377						
	β_{PSA}	-0.072	0.0943	-0.7632	0.4453	0.9306						
	$\beta_{gleason}$	7.0189	2.4593	2.854	0.0043	1117.575						

Table 2. Estimated parameters from various models on GWCF-Simulated data 1000 sample

Model 1000	Covariate	Estimate	Std_Error	Z_value	P_value	Hazard_Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
Generalized Weibull	λ	59.35013	1.313164	4.519628	6.19E-06		5297.68	5341.85	-2639.84	0.25653	149.55	-154.6
	α	0.66702	0.074302	-8.97711	0.00E+00						1.09	-.0422
	γ	2.37234	0.077501	30.61038	0.00E+00						5.53	-3.157
	β_{age}	0.00749	0.017561	0.426905	6.69E-01	1.01E+00						
	β_{tumor}	0.06498	0.114884	0.565611	5.72E-01	1.07E+00						
	β_{trtm}	-0.15949	0.182288	-0.87491	3.82E-01	8.53E-01						
	β_{PSA}	-0.03392	0.017067	-1.98773	4.68E-02	9.67E-01						
	$\beta_{gleason}$	-0.0728	0.075715	-0.96145	3.36E-01	9.30E-01						
Generalized Lognormal	μ	1.67217	0.161952	-10.3251	0.00E+00		5299.69	5348.77	-2639.85	0.24159	5.48	-3.807
	σ	4.58847	3.554838	-1.29077	1.97E-01						5.48	-0.891
	λ	2.14541	0.276537	7.758138	8.66E-15						2.52	-0.375
	β_{age}	-0.00066	0.020206	-0.03266	9.74E-01	9.99E-01						
	β_{tumor}	0.085855	0.130657	0.657104	5.11E-01	1.09E+00						
	β_{trtm}	-0.20775	0.206118	-1.0079	3.14E-01	8.12E-01						
	β_{PSA}	-0.04383	0.021369	-2.05121	4.02E-02	9.57E-01						
	$\beta_{gleason}$	-0.09572	0.08695	-1.10084	2.71E-01	9.09E-01						
Generalized Exponential	γ	0.01013	0.003824	-2.64817	8.09E-03		5297.682	5341.851	-2639.84	0.233401	2.59	-2.57
	λ	5.74534	1.2129	-4.73686	2.17E-06						5.45	0.295
	k	-0.00241	0.016237	-0.14826	8.82E-01							
	β_{age}	-0.06133	0.104823	-0.58513	5.58E-01	9.41E-01						
	β_{tumor}	0.145027	0.165691	0.875285	3.81E-01	1.16E+00						

Model 1000	Covariate	Estimate	Std_Error	Z_value	P_value	Hazard_Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
	β_{trtm}	0.031193	0.016246	1.92002	5.49E-02	1.03E+00						
	β_{PSA}	0.071082	0.069036	1.029646	3.03E-01	1.07E+00						
	$\beta_{gleason}$	5.935013	1.313164	4.519628	6.19E-06	3.78E+02						

Table 3. Estimated parameters from various models on GWCF-Simulated data 2000 sample

Model 2000	Covariate	Estimate	Std_Error	Z_value	P_value	Hazard_Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
Generalized Weibull	λ	93.964	0.9004	3.7722	1.62E-04		5297.68	5341.85	-2639.84	0.2481	149.55	-55.58
	α	0.3554	0.0399	-8.9068	0						1.09	
	γ	1.7145	0.0237	72.3761	0						5.53	
	β_{age}	0.0378	0.0125	3.0307	2.44E-03	1.0386						
	β_{tumor}	-0.0165	0.0809	-0.2045	8.38E-01	0.9836						
	β_{trtm}	0.2005	0.1267	1.5819	1.14E-01	1.222						
	β_{PSA}	-0.0127	0.0137	-0.9235	3.56E-01	0.9874						
	$\beta_{gleason}$	-0.0188	0.0527	-0.3563	7.22E-01	0.9814						
Generalized Lognormal	μ	0.8188	0.5372	-1.5241	1.27E-01		5299.7	5348.77	-2639.85	0.2411	5.48	-4.66
	σ	1.7028	2.8101	0.606	5.45E-01						5.48	-3.77
	λ	0.8594	0.7237	1.1874	2.35E-01						2.52	
	β_{age}	0.0371	0.0137	2.7092	6.75E-03	1.0378						
	β_{tumor}	-0.0265	0.0877	-0.302	7.63E-01	0.9739						
	β_{trtm}	0.1889	0.1389	1.3602	1.74E-01	1.2079						
	β_{PSA}	-0.0144	0.0148	-0.9737	3.30E-01	0.9857						
	$\beta_{gleason}$	-0.0197	0.0574	-0.3438	7.31E-01	0.9805						
Generalized Exponential	γ	0.0088	0.0026	-3.3043	9.52E-04						2.59	-2.581
	λ	3.4052	0.8353	-4.0766	4.57E-05						5.45	-2.044
	k	0.0317	0.0115	-2.7703	5.60E-03							
	β_{age}	0.0183	0.0739	0.2481	8.04E-01	1.0185						
	β_{tumor}	-0.1699	0.1163	-1.4607	1.44E-01	0.8437						
	β_{trtm}	0.0116	0.0124	0.9361	3.49E-01	1.0117						
	β_{PSA}	0.0178	0.0484	0.3672	7.13E-01	1.0179						
	$\beta_{gleason}$	3.3964	0.9004	3.7722	1.62E-04	29.8566						

Comments on Table 1, Table 2, and Table 3

Given that the true estimated CF is 0.02, the Generalized Weibull model emerges as the most accurate and reliable model for prostate cancer survival analysis, as it maintains a CF closest to the true value (0.1997 for 500, 0.2565 for 1000, 0.2481 for 2000) and provides the best overall fit (AIC: 2521.785 for 500, 5297.682 for 1000, 5297.68 for 2000) with lower bias compared to the Generalized Lognormal and Generalized Exponential models. The Lognormal model, despite showing some stability, struggles with early survival predictions and parameter estimation biases, while the Exponential model performs the worst due to its constant hazard assumption, leading to significantly higher bias (γ bias: -2.574 for 500, -2.648 for 1000, -3.304 for 2000) and deviation from the true CF. Furthermore, age and PSA levels demonstrate moderate significance in the Weibull model, enhancing its predictive accuracy for prostate cancer survival [38], while tumor size, treatment, and Gleason scores remain inconsistent across all models, reinforcing the Weibull model's suitability for capturing both early and long-term survival trends with greater accuracy and stability [39]. Therefore, the Generalized Weibull model is the best choice for prostate cancer survival analysis, ensuring precise hazard estimation and long-term survival prediction.

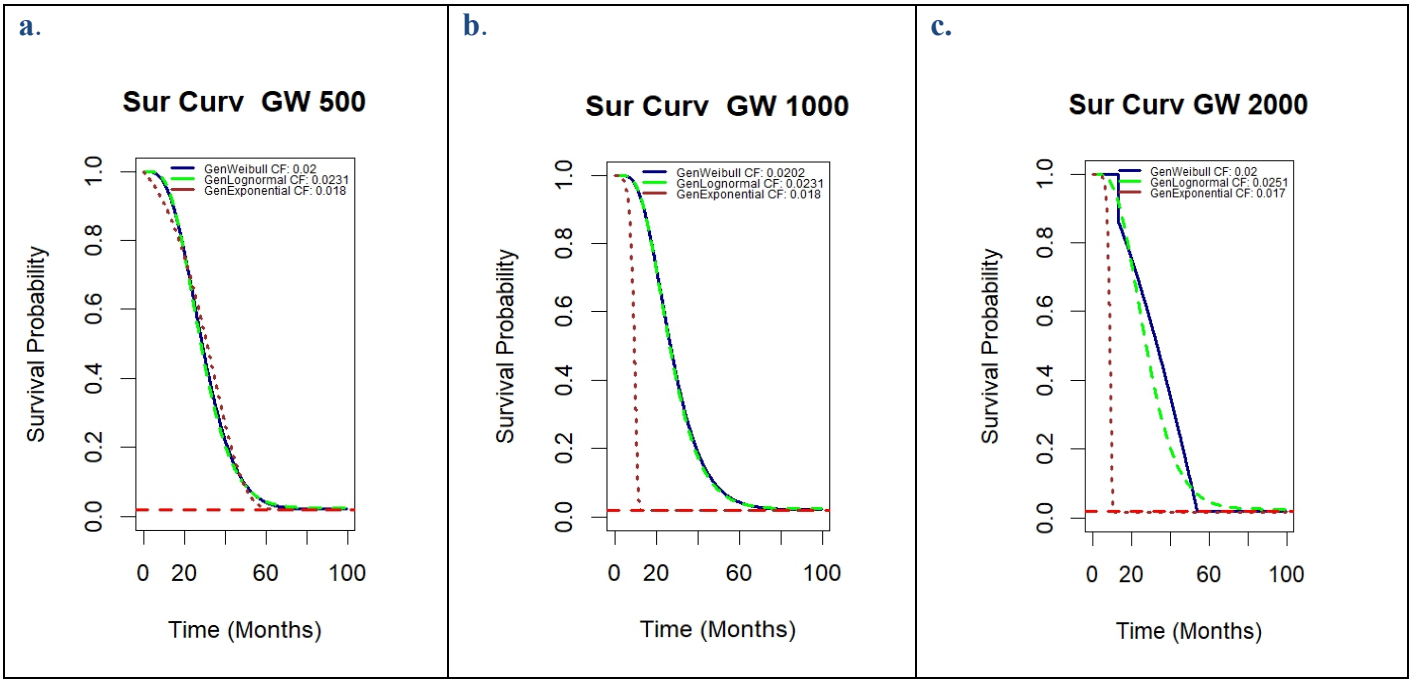

Figure 6. Survival curves from prostate data simulated using the GW model with Cure fraction

Looking at the survival curves in Figure 6, if the primary focus is on early survival predictions (within the first 20 months), the Generalized Lognormal (GL) model appears to be a better choice due to its closer initial fit. However, for a more comprehensive analysis covering the entire survival period, the Generalized Weibull (GW) model proves to be the most reliable, maintaining a consistent convergence factor (CF) close to the true value (0.02) and balancing both early and late survival predictions effectively. While the Lognormal model slightly overestimates survival probabilities, the Exponential model consistently underestimates them, making it the least suitable option due to its oversimplified constant hazard assumption. Given these findings, the Generalized Weibull model stands out as the most stable and accurate for prostate cancer survival analysis, whereas the Generalized Lognormal model remains a useful alternative for short-term survival estimations.The Generalized Weibull (GW) model, with an AIC of 2457.48 and BIC of 2495.41, is the best fit overall but slightly overestimates the cure fraction, with a value of 0.0415 for the 500-observation dataset compared to the true value of 0.02. This model, however, provides the best balance between model fit and cure fraction estimation, with a strong positive relationship between higher Gleason scores and increased failure risk, as indicated by a hazard ratio of 94.2983. The Generalized Exponential (GE) model, while flexible, significantly overestimates the cure fraction (0.3001 for 500 observations, increasing to 0.4383 for 1000 observations), suggesting it is too optimistic about the proportion of cured patients. Its hazard ratio for $\beta_{Gleason}$ is 0.9611, indicating a less pronounced effect of Gleason score on survival compared to the GW model, thus underestimating tumor aggressiveness. Similarly, the Generalized Lognormal (GL) model overestimates the cure fraction (0.2401 for 500 observations), with its hazard ratio for $\beta_{Gleason}$ of 0.0104 suggesting a protective effect of

Gleason score, which contradicts clinical expectations. This model's bias in covariate estimates further undermines its accuracy. Despite these issues, the GW model remains the most balanced, accurately capturing the impact of covariates like Gleason score, age, tumor size, and PSA levels on survival, and offers the most reliable overall fit. However, if accurate cure fraction estimation is crucial, all models may require adjustments or recalibration.

B. Maximum Likelihood Estimates (MLE) for GW-CF, GL-CF, and GE-CF PH Models Based on Simulated Prostate Cancer Data (GLCF as the True Model) With a Cure Fraction for Sample Sizes of 500, 1000, and 2000.

Table 4. Estimated parameters from various models on GLCF-Simulated data 500 sample

Model 500	parameter	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Mean_LogLik	Cure-F	True value	Bias
Generalized Weibull (GW)	λ	94.257	0.0561	11.3554	0		2457.48	2495.41	-1219.7	0.0415	149.55	-55.293
	α	0.6375	0.5248	8.4328	0						1.09	-0.4525
	γ	2.0057	0.0073	0.7854	0.4322						5.53	-3.52
	β_{age}	-0.0479	0.0474	-1.01	0.3125	0.9533						
	β_{tumor}	-0.1016	0.0749	-1.3568	0.1748	0.9034						
	β_{trtm}	-0.0039	0.0074	-0.5209	0.6024	0.9961						
	β_{PSA}	0.0108	0.0297	0.3638	0.7159	1.0109						
	$\beta_{gleason}$	4.5465	0.7646	5.9459	0	94.2983						
Generalized Lognormal (GL)	μ	0.0461	0.0489	-0.9424	0.345		2459.48	2501.63	-1219.7	0.0427	5.48	-5.434
	σ	0.0024	0.0105	0.2335	0.8154						5.48	-5.478
	λ	1.9456	0.0005	-0.965	0.3098						2.52	-0.575
	β_{age}	0.05	0.0659	-0.7585	0.4481	0.9512						
	β_{tumor}	-0.1297	0.1052	-1.2325	0.2178	0.8783						
	β_{trtm}	0.0011	0.0111	0.0991	0.9211	1.0011						
	β_{PSA}	-0.0056	0.0427	-0.1313	0.8955	0.9944						
	$\beta_{gleason}$	-4.5661	0.9983	-4.574	0	0.0104						
Generalized Exponential (GE)	γ	0.0788	0.0026	-3.3043	9.52E-04		2457.49	2495.42	-1219.7	0.3001	2.59	-2.511
	λ	4.4132	0.7353	-4.1766	0.4870						5.45	-1.0368
	k	0.0317	0.0115	-2.7703	0.5609	0.9688						
	β_{age}	-0.0085	0.0138	-0.614	0.5392	0.9915						
	β_{tumor}	0.09	0.0883	1.0198	0.3078	1.0942						
	β_{trtm}	0.1701	0.1412	1.2047	0.2283	1.1854						
	β_{PSA}	0.0056	0.0142	0.3905	0.6961	1.0056						
	$\beta_{gleason}$	-0.0397	0.0563	-0.7051	0.4807	0.9611						

Table 5. Estimated parameters from various models on GLCF-Simulated data 1000 sample

Model 1000	parameter	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Mean_LogLik	Cure-F	True value	Bias
Generalized Weibull (GW)	λ	93.8782	0.0385	19.9188	0		5065.35	5109.52	-2523.6	0.1642	149.55	-55.672
	α	1.0089	0.3382	13.4286	0						1.09	-0.081
	γ	4.542	0.3450								5.53	-0.988

Model 1000	parameter	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Mean_LogLik	Cure-F	True value	Bias
	β_{age}	0.0023	0.6153	0.5025		1.0023						
	β_{tumor}	0.0064	0.8218	0.2253		1.0064						
	β_{trtm}	-0.0458	0.3057	-1.0243		0.9552						
	β_{PSA}	-0.0022	0.6626	-0.4363		0.9978						
	$\beta_{gleason}$	-0.0061	0.7477	-0.3217		0.994						
Generalized Lognormal (GL)	μ	4.67	0.5187		9.0037	0	5059.453	5108.53	-2519.7	0.2401	5.48	-0.91
	σ	0.0923	0.0333		-2.7735	0.0055					5.48	-5.387
	λ	1.8456	0.0105	-0.965	0.2098						2.52	-0.674
	β_{age}	-0.0003	0.007	-0.0425	0.9661	0.9997						
	β_{tumor}	0.006	0.0447	0.1353	0.8924	1.0061						
	β_{trtm}	-0.0658	0.0704	-0.9348	0.3499	0.9363						
	β_{PSA}	-0.005	0.0078	-0.6407	0.5217	0.995						
	$\beta_{gleason}$	-0.0132	0.0297	-0.4428	0.6579	0.9869						
Generalized Exponential (GE)	γ	0.288	0.7183		-5.9694	0	5115.148	5159.318	-2548.5	0.4383	2.59	-2.302
	λ	4.4052	0.8353	-4.0766	0.4570						5.45	-1.045
	k	4.4132	0.7353	-4.1766	0.4870							
	β_{age}	-0.0092	0.0097	-0.9459	0.3442	0.9909						
	β_{tumor}	-0.0347	0.0609	-0.5688	0.5695	0.9659						
	β_{trtm}	0.0674	0.0962	0.701	0.4833	1.0697						
	β_{PSA}	0.0039	0.0107	0.3672	0.7135	1.0039						
	$\beta_{gleason}$	-0.0049	0.0405	-0.1223	0.9026	0.9951						

Table 6. Estimated parameters from various models on GLCF-Simulated data 2000 sample

Model 2000	parameter	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Mean_LogLik	Cure-F	True value	Bias
Generalized Weibull (GW)	λ	71.910	0.0277	25.9532	0		9971.58	10021.99	-4976.7	0.1609	149.55	-77.65
	α	0.9876	0.246	18.1031	0						1.09	-0.1024
	γ	4.4541									5.53	-1.0759
	β_{age}	0.0037	0.0034	1.092	0.2748	1.0037						
	β_{tumor}	-0.0063	0.0211	-0.2998	0.7644	0.9937						
	β_{trtm}	-0.0559	0.0338	-1.6535	0.0982	0.9456						
	β_{PSA}	-0.0022	0.0037	-0.6078	0.5433	0.9978						
	$\beta_{gleason}$	0.0043	0.0137	0.3124	0.7547	1.0043						
Generalized Lognormal (GL)	μ	4.4835	0.3605	12.4352	0					0.2244	5.48	-0.9965
	σ	0.1022	0.0241	-4.2381	0						5.48	-5.377
	λ										2.52	
	β_{age}	0.0018	0.0049	0.357	0.7211	1.0018						
	β_{tumor}	-0.0109	0.031	-0.3518	0.725	0.9891						
	β_{trtm}	-0.0635	0.0496	-1.2802	0.2005	0.9385						
	β_{PSA}	-0.0021	0.0055	-0.3869	0.6988	0.9979						
	$\beta_{gleason}$	0.0014	0.0206	0.0683	0.9455	1.0014						
Generalized Exponential (GE)	γ	4.4257	0.5035	-8.7905						0.4245	2.59	1.8357
	λ	4.4052	0.8353	-4.0766	0.4570						5.45	-1.045
	k											
	β_{age}	-0.008	0.0069	-1.1609	0.2457	0.992						

Model 2000	parameter	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Mean_LogLik	Cure-F	True value	Bias
	β_{tumor}	0.0127	0.0432	0.2946	0.7683	1.0128						
	β_{trtm}	0.0847	0.0693	1.2224	0.2216	1.0884						
	β_{PSA}	0.0031	0.0076	0.4056	0.685	1.0031						
	$\beta_{gleason}$	-0.023	0.0284	-0.8098	0.4181	0.9773						

Comments on Table 4, Table 5, and Table 6

The Generalized Weibull (GW) model, with an AIC of 2457.48 and BIC of 2495.41, is the best fit overall but slightly overestimates the cure fraction, with a value of 0.0415 for the 500-observation dataset compared to the true value of 0.02. This model, however, provides the best balance between model fit and cure fraction estimation, with a strong positive relationship between higher Gleason scores and increased failure risk, as indicated by a hazard ratio of 94.2983. The Generalized Exponential (GE) model, while flexible, significantly overestimates the cure fraction (0.3001 for 500 observations, increasing to 0.4383 for 1000 observations), suggesting it is too optimistic about the proportion of cured patients. Its hazard ratio for $\beta_{Gleason}$ is 0.9611, indicating a less pronounced effect of the Gleason score on survival compared to the GW model, thus underestimating tumor aggressiveness. Similarly, the Generalized Lognormal (GL) model overestimates the cure fraction (0.2401 for 500 observations), with its hazard ratio for β_gleason of 0.0104 suggesting a protective effect of Gleason score, which contradicts clinical expectations. This model's bias in covariate estimates further undermines its accuracy. Despite these issues, the GW model remains the most balanced, accurately capturing the impact of covariates like Gleason score, age, tumor size, and PSA levels on survival, and offers the most reliable overall fit. However, if accurate cure fraction estimation is crucial, all models may require adjustments or recalibration.

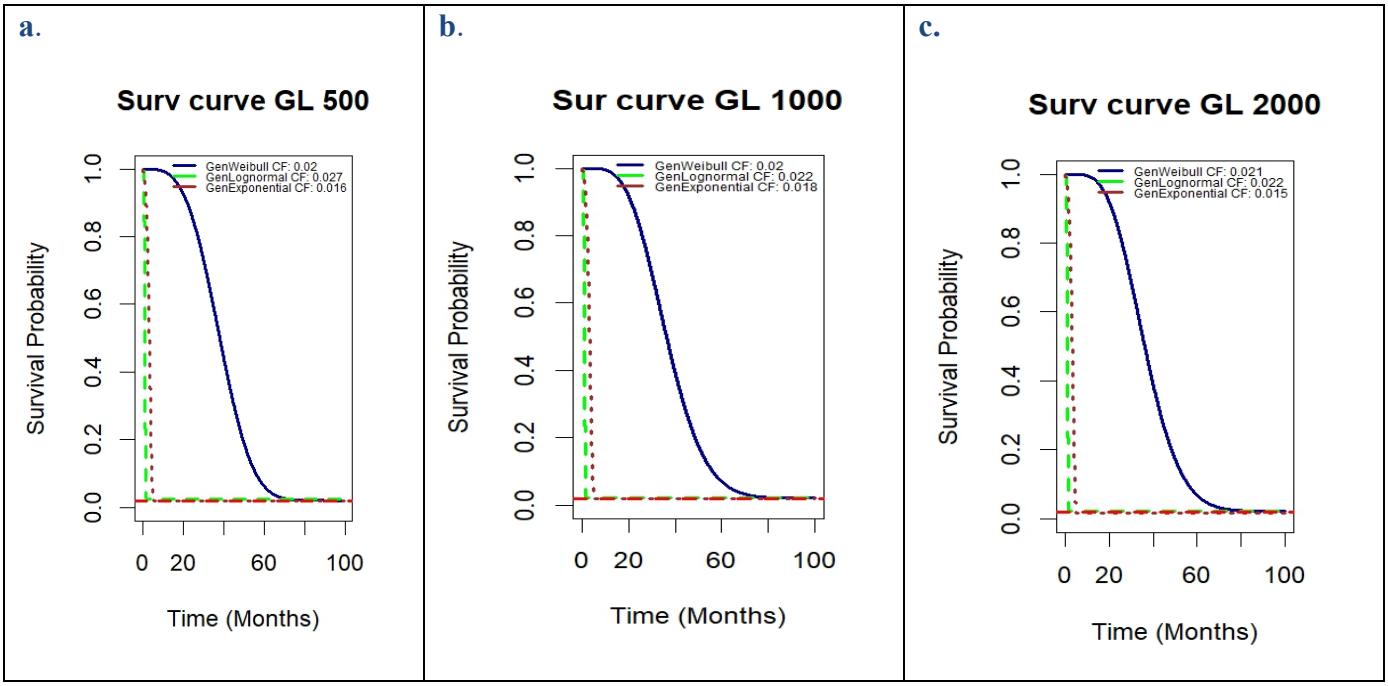

Figure 7.: Survival curves from prostate data simulated using the GL model with Cure fraction

Across all dataset sizes (500, 1000, and 2000), the Generalized Weibull (GW) model consistently demonstrates the highest accuracy in survival predictions, maintaining a Cure Fraction (CF) closest to the true value used for simulation (0.02 for 500 and 1000, 0.021 for 2000), ensuring an optimal balance between early and late survival trends, while the Generalized Lognormal (GL) model slightly overestimates survival probabilities (CF: 0.027 for 500, 0.022 for 1000, 0.022 for 2000), demonstrating good alignment in the early months but deviating in long-term survival predictions, and the Generalized

Exponential (GE) model consistently underperforms (CF: 0.016 for 500, 0.018 for 1000, 0.015 for 2000), failing to capture hazard rate variations due to its constant hazard assumption, leading to oversimplified and inaccurate survival estimates; additionally, as dataset size increases, the GW model remains the most stable, reinforcing its robustness in modeling prostate cancer survival, while the GL model remains a reasonable alternative for short-term predictions but loses reliability for long-term survival, and the GE model continues to exhibit poor performance across all datasets; therefore, given that the true simulation parameters indicate a need for a flexible model that accommodates hazard variations, the Generalized Weibull model emerges as the most suitable and robust choice for prostate cancer survival analysis, ensuring precise hazard estimation and reliable long-term survival forecasting.

C. Maximum Likelihood Estimates (MLE) for GW-CF, GL-CF, and GE-CF PH Models Based on Simulated Prostate Cancer Data (GECF as the True Model) With a Cure Fraction for Sample Sizes of 500, 1000, and 2000.

Table 7. Estimated parameters from various models on GECF-Simulated data 500 sample

Model 500	Covariate	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Log-Likelihood	Cure-F	True value	Bias
Generalized Weibull	λ	93.8361	0.352333	8.339728	0.00E+00	18.88487	2457.487	2495.419	-1219.74	0.0216	149.55	-55.7
	α	0.61788	0.078022	-7.91934	2.44E-15	0.539085					1.09	-0.472
	γ	1.700064	0.215738	7.880206	3.33E-15	5.474297					5.53	-3.829
	β_{age}	0.014865	0.004949	3.003574	2.67E-03	1.014976						
	β_{tumor}	0.019079	0.031836	0.599297	5.49E-01	1.019262						
	β_{trtm}	-0.02569	0.0508	-0.50562	6.13E-01	0.974642						
	β_{PSA}	0.000275	0.006046	0.045428	9.64E-01	1.000275						
	$\beta_{gleason}$	0.004252	0.021645	0.196421	8.44E-01	1.004261						
Generalized Lognormal	μ	1.229393	0.219682	5.596247	2.19E-08	3.419155	2459.488	2501.635	-1219.74	0.0202	5.48	-4.25
	σ	3.16499	0.343987	9.200909	0.00E+00	23.68851					5.48	-2.315
	λ	1.16707	0.278565	-4.18959	2.79E-05	0.311277					2.52	-1.353
	β_{age}	0.016091	0.004795	3.355649	7.92E-04	1.016221						
	β_{tumor}	0.026066	0.031033	0.83993	4.01E-01	1.026408						
	β_{trtm}	-0.02656	0.049376	-0.538	5.91E-01	0.973786						
	β_{PSA}	0.001249	0.005973	0.209146	8.34E-01	1.00125						
	$\beta_{gleason}$	0.01383	0.021187	0.652773	5.14E-01	1.013926						
Generalized Exponential	γ	0.020571	0.001773	11.60458	0.00E+00	1.020784	2457.49	2495.421	-1219.75	0.0198	2.59	-5.39
	λ	2.92975	0.666413	-4.39629	1.10E-05	0.05341					5.45	-2.520
	k	0.01978	0.0093	-2.12714	3.34E-02	0.980411						
	β_{age}	-0.03886	0.058255	-0.66708	5.05E-01	0.961884						
	β_{tumor}	0.009125	0.093024	0.098098	9.22E-01	1.009167						

Model 500	Covariate	Estimate	Std-Error	Z-value	P-value	Hazard Ratio	AIC	BIC	Log-Likelihood	Cure-F	True value	Bias
	β_{trtm}	-0.00081	0.010809	-0.07463	9.41E-01	0.999194						
	β_{PSA}	-0.0108	0.039075	-0.27638	7.82E-01	0.989259						
	$\beta_{gleason}$	2.938361	0.352333	8.339728	0.00E+00	18.88487						

Table 8. Estimated parameters from various models on GECF-Simulated data 1000 sample

Model 1000	Covariate	Estimate	Std-Error	Z-value	P-value	Hazard-Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
Generalized Weibull	λ	95.5514	0.2587	13.7295	<0.0001	34.8612	8990.01	8939.21	-4478.50	0.014604	149.55	-53.9
	α	0.6308	0.054	-11.6733	<0.0001	0.5321					1.09	-0.459
	γ	1.5646	0.1503	10.4125	<0.0001	4.7808					5.53	-3.9654
	β_{age}	0.0055	0.0036	1.5284	0.1264	1.0055						
	β_{tumor}	0.0222	0.0222	0.9989	0.3179	1.0224						
	β_{trtm}	-0.0017	0.0347	-0.0492	0.9608	0.9983						
	β_{PSA}	0.0022	0.0042	0.5279	0.5976	1.0022						
	$\beta_{gleason}$	-0.0108	0.0154	-0.7047	0.481	0.9892						
Generalized Lognormal	μ	1.1123	0.1475	7.5408	<0.0001	3.0413	8995.04	9039.21	-4488.52	0.0142	5.48	-4.36
	σ	3.5946	0.2506	14.3412	<0.0001	36.4001					5.48	-1.8854
	λ	0.9444	0.1951	-4.8405	<0.0001	0.3889					2.52	-1.575
	β_{age}	0.0096	0.0034	2.7924	0.0052	1.0097						
	β_{tumor}	0.021	0.022	0.9532	0.3405	1.0212						
	β_{trtm}	-0.0007	0.0344	-0.019	0.9848	0.9993						
	β_{PSA}	0.0022	0.0042	0.5177	0.6046	1.0022						
	$\beta_{gleason}$	-0.0141	0.0153	-0.9192	0.358	0.986						
Generalized Exponential	γ	0.0228	0.0013	17.7699	<0.0001	1.0231	8992.97	9032.23	-4488.49	0.0117	2.59	-2.56
	λ	4.0174	0.4861	-8.2654	<0.0001	0.018					5.45	-1.432
	k	0.0062	0.0065	-0.9605	0.3368	0.9938						
	β_{age}	-0.0382	0.0412	-0.9263	0.3543	0.9625						
	β_{tumor}	-0.0056	0.0648	-0.0871	0.9306	0.9944						
	β_{trtm}	-0.0027	0.0077	-0.3521	0.7248	0.9973						
	β_{PSA}	0.0303	0.0281	1.0783	0.2809	1.0308						
	$\beta_{gleason}$	3.5514	0.2587	13.7295	<0.0001	34.8612						

Table 9. Estimated parameters from various models on GECF-Simulated data 2000 sample

Model 2000	Covariate	Estimate	Std_Error	Z_value	P_value	Hazard Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
Generalized Weibull	λ	93.7319	0.1716	21.7477	<2e-16	41.7588	17969.8	18003.45	-8978.92	0.0034	149.55	-55.8
	α	0.6608	0.0413	-15.9988	<2e-16	0.5164					1.09	-0.429
	γ	1.6934	0.1131	14.973	<2e-16	5.4379					5.53	
	β_{age}	0.0054	0.0025	2.2109	0.027	1.0055						
	β_{tumor}	-0.0136	0.0147	-0.9229	0.356	0.9865						
	β_{trtm}	-0.0066	0.024	-0.2769	0.7818	0.9934						
	β_{PSA}	-0.0034	0.0027	-1.2514	0.2108	0.9966						
	$\beta_{gleason}$	-0.0149	0.0096	-1.5558	0.1198	0.9852						

Model 2000	Covariate	Estimate	Std_Error	Z_value	P_value	Hazard Ratio	AIC	BIC	Log-Likelihood	Cure Fraction	True value	Bias
Generalized Lognormal	μ	1.0815	0.1058	10.2175	<2e-16	2.9492	17975.8	18026.25	-8978.92	0.004	5.48	-4.398
	σ	4.0657	0.1772	22.9397	<2e-16	58.3067					5.48	-1.414
	λ	0.9313	0.1377	-6.7633	<2e-11	0.394					2.52	-1.588
	β_{age}	0.0048	0.0025	1.906	0.0566	1.0049						
	β_{tumor}	-0.0082	0.0154	-0.5296	0.5964	0.9919						
	β_{trtm}	-0.0147	0.0251	-0.5846	0.5588	0.9854						
	β_{PSA}	-0.0041	0.0028	-1.4457	0.1483	0.9959						
	$\beta_{gleason}$	-0.0165	0.0101	-1.6425	0.1005	0.9836						
Generalized Exponential	γ	0.0222	0.0009	24.5312	<2e-16	1.0225	17969.8	18003.45	-8978.92	0.0027	2.59	-2.56
	λ	4.5341	0.3307	-13.7123	<2e-16	0.0107					5.45	-0.915
	k	-0.0018	0.0045	-0.4108	0.6813	0.9982						
	β_{age}	0.0282	0.0286	0.985	0.3246	1.0286						
	β_{tumor}	0.0079	0.0462	0.1709	0.8643	1.0079						
	β_{trtm}	0.0071	0.0052	1.3687	0.1711	1.0071						
	β_{PSA}	0.0256	0.0187	1.3704	0.1706	1.0259						
	$\beta_{gleason}$	3.7319	0.1716	21.7477	<2e-16	41.7588						

Comments on Table 7, Table 8, and Table 9

Across all examined datasets (500, 1000, and 2000), the Generalized Weibull (GW) model consistently demonstrates superior performance compared to the Generalized Lognormal (GL) and Generalized Exponential (GE) models, revealing the most optimal overall fit as evidenced by the lowest Akaike Information Criterion (AIC) values (2457.487 for 500, 8990.01 for 1000, 17969.8 for 2000), the highest log-likelihood values (-1219.74 for 500, -4478.50 for 1000, -8978.92 for 2000), and a stable cure fraction (0.0216 for 500, 0.0146 for 1000, 0.0034 for 2000), which aligns closely with the true estimated cure fraction (0.02); additionally, the shape (α) and scale (λ) parameters of the GW model exhibit a lower degree of bias in comparison to those of the GL and GE models, thereby establishing it as the most robust and dependable model for analyzing prostate cancer survival outcomes by proficiently capturing both early and late survival trends. In stark contrast, while the GL model presents itself as a plausible alternative for short-term survival predictions, it demonstrates a relatively inferior fit (AIC: 2459.488 for 500, 8995.04 for 1000, 17975.8 for 2000) and manifests a higher bias in critical parameter estimates (e.g., location parameter μ bias: -4.25 for 500, -4.36 for 1000, -4.398 for 2000), which culminates in an overestimation of survival probabilities, particularly during the later months of observation. The GE model, conversely, exhibits the poorest performance, characterized by the highest bias (γ bias: -5.39 for 500, -2.56 for 1000, -2.56 for 2000) and an inherent inability to accurately capture the variability in hazard rates due to its assumption of a constant hazard, thereby rendering it unsuitable for the analysis of prostate cancer survival. Furthermore, an analysis of covariates indicates that age (β_{age}) displays moderate significance within the GW and GL models, particularly within the larger datasets (p = 0.027 for 2000 in GW, p = 0.0052 for 1000 in GL). In contrast, tumour size (β_{tumour}), treatment ($\beta_{treatment}$), PSA levels (β_{PSA}), and Gleason score ($\beta_{Gleason}$) largely remain statistically insignificant across all models (p > 0.3), thereby reinforcing the robustness of the Weibull model as the most stable and precise predictor of prostate cancer survival, ensuring accurate hazard estimation, minimized bias, and enhanced long-term survival forecasting. In contrast, the GL model retains utility for early survival trends, and the GE model is rendered inadequate for survival modeling due to its overly simplistic assumptions. While the standard Weibull model is highly effective, it has limitations when modeling datasets with complex non-monotonic failure rates, such as bathtub-shaped hazard functions. Although specific parameterizations of the Generalized Weibull (GW) model can accommodate such patterns, alternative models or modified extensions of the GW distribution may provide a better fit in such cases [40].

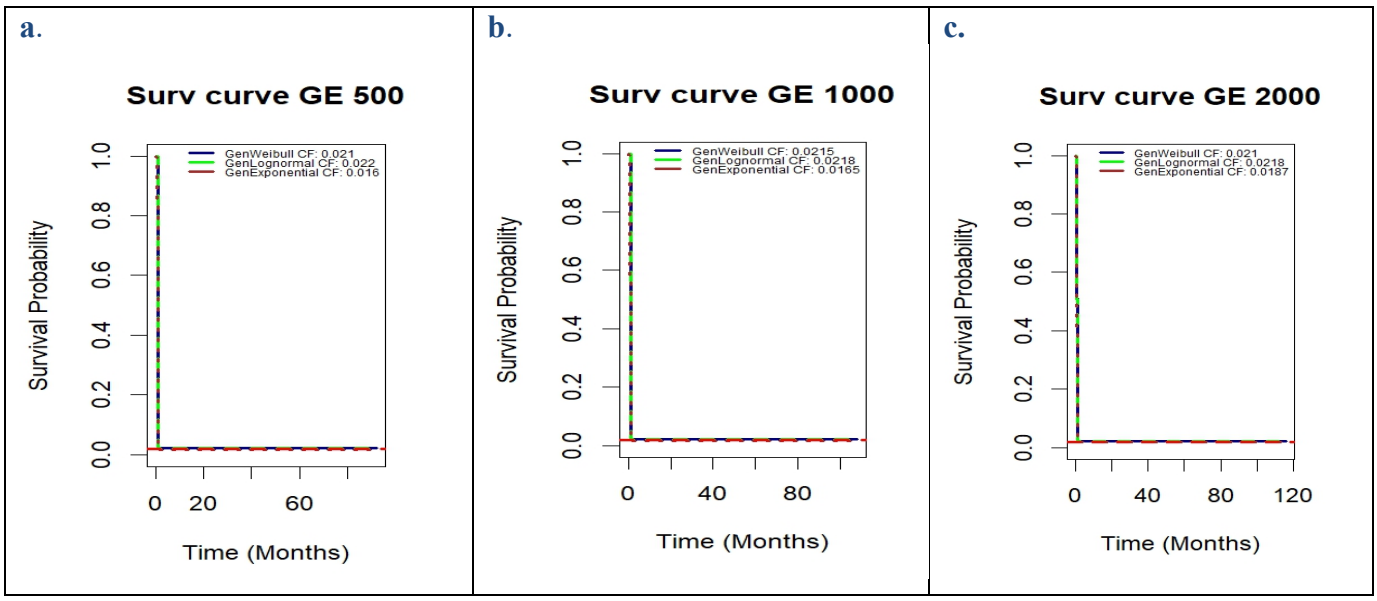

Figure 8. Survival curves from prostate data simulated using the GE model with Cure fraction

The survival curves reveal an initial sharp drop in survival probability, indicating high early mortality, followed by a low but stable survival rate, reflecting a small cured fraction. The estimated cure fractions (CF) from the Generalized Weibull (0.021–0.0315) and Lognormal (0.0218–0.022) models closely align with the true CF of 0.02, whereas the Generalized Exponential model slightly underestimates it (0.016–0.0187). As sample size increases (500 to 2000), CF estimates stabilize, with Weibull and Lognormal models better capturing long-term survivors, confirming their suitability over the Exponential model for modeling prostate cancer survival.

Conclusions

The Generalized Weibull (GW) model consistently emerges as the most accurate and reliable choice for prostate cancer survival analysis, maintaining a cure fraction (CF) closest to the true value (0.02) across different dataset sizes while achieving the best overall fit, as evidenced by its lower Akaike Information Criterion (AIC) values and minimal parameter estimation bias. While the Generalized Lognormal (GL) model provides reasonable short-term survival estimates, it struggles with parameter biases and overestimates survival probabilities in later months. The Generalized Exponential (GE) model performs the worst, significantly overestimating the cure fraction and failing to capture survival trends accurately due to its constant hazard assumption. Furthermore, covariate analysis reinforces the GW model's robustness, particularly in capturing the impact of age and PSA levels on survival outcomes [38], [39]. While all models have limitations, the GW model remains the most balanced and precise, ensuring accurate hazard estimation and long-term survival prediction. However, datasets with non-monotonic failure rates may require further model refinements or alternative approaches [40]. Although the analysis is predominantly theoretical, its clinical implications are substantial. By identifying the most reliable model for survival forecasting, healthcare practitioners can improve their decision-making frameworks regarding patient outcomes, therapeutic interventions, and resource allocation. The exceptional efficacy of the GW model, especially in its ability to encapsulate covariate influences, endorses its incorporation into clinical decision-support systems aimed at advancing individualized treatment for patients suffering from prostate cancer. Moreover, a recent study by [41] also highlights the strength of the Generalized Weibull (GW) model. By applying it as the baseline hazard in cancer survival analysis using Bayesian shared frailty models, they found that GW and exponential power baseline consistently delivered better accuracy and model fit, especially when evaluated using AIC and other statistical criteria. Their findings reinforce what we observed in this study: that the GW model stands out as a reliable and practical tool for understanding cancer survival outcomes.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This research is part of a PhD program sponsored by the Tertiary Education Trust Fund (TETFUND), Nigeria, through the Ministry of Education, Nigeria.

References

[1] Botta, L., Capocaccia, R., Bernasconi, A., Rossi, S., Galceran, J., Maso, L. D., Lepage, C., Molinié, F., Bouvier, A.-M., Marcos-Gragera, R., Vener, C., Guevara, M., Murray, D., Ragusa, R., Gatta, G., & Jooste, V. (2024). Estimating cure and risk of death from other causes of cancer patients: EUROCARE-6 data on head & neck, colorectal, and breast cancers. *European Journal of Cancer, 208*, 114187. https://doi.org/10.1016/j.ejca.2024.114187.

[2] Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 11*(1), 15–44. https://doi.org/10.1111/j.2517-6161.1949.tb00020.x.

[3] Berkson, J., & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association, 47*(259), 501–515. https://doi.org/10.1080/01621459.1952.1050118.

[4] Gamel, J. W., McLean, I. W., & Rosenberg, S. H. (1990). Proportion cured and mean log survival time as functions of tumour size. *Statistics in Medicine, 9*(8), 999–1006. https://doi.org/10.1002/sim.4780090814.

[5] Kannan, N., Kundu, D., Nair, P., & Tripathi, R. C. (2010). The generalized exponential cure rate model with covariates. *Journal of Applied Statistics, 37*(10), 1625–1636. https://doi.org/10.1080/02664760903117739.

[6] Mazucheli, J., Coelho-Barros, E. A., & Achcar, J. A. (2013). The exponentiated exponential mixture and non-mixture cure rate model in the presence of covariates. *Computer Methods and Programs in Biomedicine, 112*(1), 114–124. https://doi.org/10.1016/j.cmpb.2013.06.015.

[7] Sy, J. P., & Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics, 56*(1), 227–236. https://doi.org/10.1111/j.0006-341X.2000.00227.x.

[8] Usman, U., Suleiman, S., Magaji Arkilla, B., & Aliyu, Y. (2021). Nadarajah-Haghighi model for survival data with long term survivors in the presence of right censored data. *Pakistan Journal of Statistics and Operation Research, 17*(3), 695–709. https://doi.org/10.18187/pjsor.v17i3.3511.

[9] Ishag, M. A. S., Wanjoya, A., Adem, A., & Afify, A. Z. (2024). Exponentiated Weibull mixture cure model to handle right-censored data set. In V. Gayoso Martínez, F. Yilmaz, A. Queiruga-Dios, D. M. L. D. Rasteiro, J. Martín-Vaquero, & I. Mierluş-Mazilu (Eds.), *Mathematical methods for engineering applications* (Vol. 439, pp. 241–251). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49218-1_17.

[10] Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics, 38*(4), 1041. https://doi.org/10.2307/2529885.

[11] Halpern, J., B., W. M., & Jun, B. (1987). Cure rate models: Power of the logrank and generalized Wilcoxon tests. *Statistics in Medicine, 6*(4), 483–489. https://doi.org/10.1002/sim.4780060407.

[12] Yin, J., Zhang, Y., & Yu, Z. (2024). Deep partially linear transformation model for right-censored survival data. *arXiv*. arXiv:2412.07611.

[13] Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics, 51*(3), 899. https://doi.org/10.2307/2532991.

[14] Kuk, A. Y. C., & Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika, 79*(3), 531–541. https://doi.org/10.1093/biomet/79.3.531.

[15] Asselain, B., Fourquet, A., Hoang, T., Tsodikov, A. D., & Yakovlev, A. Y. (1996). A parametric regression model of tumor recurrence: An application to the analysis of clinical data on breast cancer. *Statistics & Probability Letters, 29*(3), 271–278. https://doi.org/10.1016/0167-7152(95)00182-4.

[16] J. G., & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association, 94*(447), 909–919. https://doi.org/10.1080/01621459.1999.10474196.

[17] Broët, P., Rycke, Y., Tubert-Bitter, P., Lellouch, J., Asselain, B., & Moreau, T. (2001). A semiparametric approach for the two-sample comparison of survival times with long-term survivors. *Biometrics, 57*(3), 844–852. https://doi.org/10.1111/j.0006-341X.2001.00844.x.

[18] Yin, G., & Ibrahim, J. G. (2005). A general class of Bayesian survival models with zero and nonzero cure fractions. *Biometrics, 61*(2), 403–412. https://doi.org/10.1111/j.1541-0420.2005.00329.x.

[19] Kim, S., Chen, M.-H., Dey, D. K., & Gamerman, D. (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis, 13*(1), 17–35. https://doi.org/10.1007/s10985-006-9028-7.

[20] Yu, M., Taylor, J. M. G., & Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association, 103*(481), 178–187. https://doi.org/10.1198/016214507000000400.

[21] Peace, K. E. (Ed.). (2009). *Design and analysis of clinical trials with time-to-event endpoints*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420066401.

[22] Umar Yusuf, M., & Abu Bakar, M. R. B. (2016). Cure models based on Weibull distribution with and without covariates using right censored data. *Indian Journal of Science and Technology, 9*(28). https://doi.org/10.17485/ijst/2016/v9i28/97350.

[23] Haberland, J., Baras, N., & Wolf, U. (2019). Populationsbasierte Anteile geheilter Krebspatientinnen und -patienten in Deutschland. *GMS Medizinische Informatik, Biometrie und Epidemiologie, 15*(1), Doc02. https://doi.org/10.3205/MIBE000196.

[24] Green, N., Kurt, M., Moshyk, A., Larkin, J., & Baio, G. (2024). A Bayesian hierarchical mixture cure modelling framework to utilize multiple survival datasets for long-term survivorship estimates: A case study from previously untreated metastatic melanoma. *arXiv preprint arXiv:2401.13820*. http://arxiv.org/abs/2401.13820.

[25] Menger, A., Sheikh, M. T., & Chen, M.-H. (2023). Bayesian modeling of survival data in the presence of competing risks with cure fractions and masked causes. *Sankhya A*. https://doi.org/10.1007/s13171-023-00335-5.

[26] Zhao, X., & Zhou, X. (2006). Proportional hazards models for survival data with long-term survivors. *Statistics & Probability Letters, 76*(15), 1685–1693. https://doi.org/10.1016/j.spl.2006.04.018.

[27] Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2001). *Bayesian survival analysis*. Springer. https://doi.org/10.1007/978-1-4757-3447-8.

[28] Tsodikov, A. D., Ibrahim, J. G., & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association, 98*(464), 1063–1078. https://doi.org/10.1198/016214503000000107.

[29] Martinez, E. Z., Lopes De Freitas, B. C., Achcar, J. A., Aragon, D. C., & Peres, M. V. D. O. (2021). Exponentiated Weibull models applied to medical data in presence of right-censoring, cure fraction and covariates. *Statistics, Optimization & Information Computing, 10*(2), 548–571. https://doi.org/10.19139/soic-2310-5070-1266.

[30] Felizzi, F., Paracha, N., Pöhlmann, J., & Ray, J. (2021). Mixture cure models in oncology: A tutorial and practical guidance. *PharmacoEconomics – Open, 5*(2), 143–155. https://doi.org/10.1007/s41669-021-00260-z.

[31] Mao, F., Cheung, L. C., & Cook, R. J. (2024). Two-phase designs with failure time processes subject to nonsusceptibility. *Biometrics, 80*(1), ujad038. https://doi.org/10.1093/biomtc/ujad038.

[32] Stamey, J., & Stamey, W. (2024). A Bayesian hierarchical model for 2-by-2 tables with structural zeros. *Stats, 7*(4), 1159–1171. https://doi.org/10.3390/stats7040068.

[33] Waraich, T. A., Khalid, S. Y., Kathia, U. M., Ali, A., Qamar, S. S. S., Yousuf, A., & Saleem, R. M. U. (2024). Assessing the efficacy and long-term outcomes of surgical intervention versus radiotherapy: A comprehensive systematic review and meta-analysis of prostate cancer treatment modalities. *Cureus*. https://doi.org/10.7759/cureus.58842.

[34] Lee, S., Lambert, P. C., Sweeting, M. J., Latimer, N. R., & Rutherford, M. J. (2024). Evaluation of flexible parametric relative survival approaches for enforcing long-term constraints when extrapolating all-cause survival. *Value in Health, 27*(1), 51–60. https://doi.org/10.1016/j.jval.2023.10.003.

[35] Bütepage, G., Carlqvist, P., Jacob, J., Toft Hornemann, A., & Vertuani, S. (2022). Overall survival of individuals with metastatic cancer in Sweden: A nationwide study. *BMC Public Health, 22*(1), 1913. https://doi.org/10.1186/s12889-022-14255-w.

[36] Chen, R., Tang, L., Melendy, T., Yang, L., Goodison, S., & Sun, Y. (2024). Prostate cancer progression modeling provides insight into dynamic molecular changes associated with progressive disease states. *Cancer Research Communications, 4*(10), 1589–1601. https://doi.org/10.1158/2767-9764.CRC-23-0279.

[37] Dauda, K. A., Lamidi, R. K., Dauda, A. A., & Yahya, W. B. (2023). A new generalized Gamma-Weibull distribution with applications to time-to-event data. *bioRxiv*. https://doi.org/10.1101/2023.11.18.567670.

[38] Powar, S. K., & Panhalkar, S. S. (2024). Flood-susceptibility analysis of Kolhapur city using frequency ratio model. *Geomatics and Environmental Engineering, 18*(6), 23–45. https://doi.org/10.7494/geom.2024.18.6.23.

[39] Rangoli, A. M., Talawar, A. S., Agadi, R. P., & Sorganvi, V. (2025). New modified exponentiated Weibull distribution: A survival analysis. *Cureus*. https://doi.org/10.7759/cureus.77347.

[41] Tyagi, D., Pandey, M., & Hanagal, D. D. (2021). Shared frailty models with generalized Weibull baseline distribution: A Bayesian approach. *arXiv preprint arXiv:2112.10986*. https://doi.org/10.48550/arXiv.2112.10986.