STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, June 2026, pp 4540–4552.
Published online in International Academic Press (www.IAPress.org)

Enhanced Personalized Clinical Treatment: Regression Models' Applications to Breast Cancer Patient Management

Isaac T. ADEDOSU[1], Dorcas M. OKEWOLE[1], John O. OLAOMI[2,*], Ayobami F. AKINTOLA[1]

[1] *Department of Mathematics and Statistics, Redeemer's University, Osun State, Nigeria*
[2] *Department of Statistics, University of South Africa, Science Campus, Florida 1710, Johannesburg, South Africa*

Abstract Incorporation of statistical predictive models into clinical practice enhances personalized treatment planning and patient management. This study compares the performance of six parametric regression models, namely Weibull, Exponential, Log-logistic, Lognormal, Gamma, and Gompertz, in identifying the prognostic factors affecting the survival of breast cancer patients. Survival models were estimated for a dataset of 686 breast cancer patients from the German Breast Cancer Study (GBCS). The following comparison criteria were used to compare the survival models: Akaike Information Criterion, Bayesian Information Criterion, Log-Likelihood, and Deviance. While all the models yielded similar outcomes, the model selection criteria indicated that the Lognormal model had the best fit for the data. All models identified the same set of significant predictors, suggesting that tumor size, tumor grade, number of lymph nodes involved, and progesterone receptor count significantly influenced survival. The findings from this study show the lognormal model to be the best fit for survival and the importance of early detection and testing to improve the prognosis of breast cancer patients.

Keywords Log-logistic, Lognormal, Gamma, Parametric models, Breast Cancer, Survival Analysis.

DOI: 10.19139/soic-2310-5070-2857

1. Introduction

Survival analysis is a statistical methodology that analyzes the time until an event of interest, such as death or disease recurrence. It allows the researcher to study and examine various factors that may affect such time-to-event outcomes and helps to compare survival rates between different groups.

As pointed out by Jager et al. [33], the Kaplan-Meier method is a well-known non-parametric method that estimates and graphically displays survival probabilities over time. In contrast with the Kaplan-Meier method, which represents survival in a simple and intuitive form, the use of parametric models is more flexible and wider in its modeling scope when conducting survival analysis. According to Clark et al. [7] and Jager et al. [33], this class of models usually assumes some known probability distribution, such as the exponential, Weibull, or log-normal distribution, for the survival times. In recent times, there has been a growing popularity of parametric models to estimate the survival of breast cancer patients. These classes of models can provide a more precise look into the underlying dynamics of survival and the influence of various prognostic factors. Examples include the work by Tasfa and Mengistie [32].

While the Kaplan-Meier method is one of the most standard nonparametric approaches for survival analysis, parametric models are increasingly being used to analyze survival as these provide much greater flexibility and comprehensiveness.

*Correspondence to: John O. OLAOMI (Email: olaomjo@unisa.ac.za). Department of Statistics, University of South Africa, Science Campus, Florida 1710, Johannesburg, South Africa.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

According to Clark et al. [7] and Jager et al. (2008), the performance of the various parametric models considered in this research has never been compared simultaneously in the analysis of survival for breast cancer. Therefore, it is necessary to compare them to identify the best fit for the breast cancer survival data. These models were: Exponential, Gamma, Weibull, Gompertz, Log-logistic, and Log-Normal distributions.

According to the WHO estimates, breast cancer affects millions of females all over the world. In 2022 alone, there were approximately 2.3 million new cases. The effect of the disease is crippling, considering not only the high prevalence but also the emotional, physical, and economic impact on the patients, the family, and the healthcare system. Different regions report discrepancies in disease incidence, as shown by Sha et al. [28] and Bellanger et al. [5], and the leading causes are genetic background, lifestyle habits, and access to health care. Management of breast cancer has improved significantly due to advances in screening, diagnosis, and therapeutic approaches, which have ensured better clinical outcomes for patients in most settings. However, significant obstacles are still faced by women in low-resource parts of the world, such as lack of access to health services, including treatment and screening, low levels of awareness of the disease, financial constraints, and sociocultural stigmas related to the disease; because of this, breast cancer is mostly diagnosed at late stages, which are not easily treatable [2].

Breast cancer is a heterogeneous disease with many molecular subtypes, symptomatically presenting different clinical outcomes that ought to be comprehensively understood for better clinical management and improvement in the prognosis of patients. In this line, Akuoko et al. [2] noted that statistical models are paramount in informing appropriate clinical decisions about care. In this context, modeling will help clinicians make better treatment decisions and estimate the probability of patient survival, thereby improving outcomes.

Understanding and predicting patient outcomes is crucial to improving survival rates and tailoring treatments to meet everyone's needs. In this respect, survival analysis is an important statistical tool that allows researchers and clinicians to analyze time-to-event data and identify key information about disease progression and treatment effectiveness.

2. Materials and methods

The dataset used in this analysis comprised 686 patients from the GBCS. It contains a range of demographic, clinical, and pathological variables, including age, tumor size, number of nodes involved, progesterone receptor status, and estrogen receptor status, where the primary outcomes of interest were survival time and event status. Survival time is measured from the date of diagnosis to the date of death, or last follow-up, and event status indicates whether the patient experienced the event of interest (e.g., death) during the study period. These variables are crucial in survival analysis since they give an insight into the different prognostic factors that influence survival outcomes among patients diagnosed with breast cancer.

- The six parametric models evaluated in this study include Exponential, Gamma, Weibull, Gompertz, Log-logistic, and Log-Normal distributions. Each of these models assumes a different form of the underlying distribution for survival time and may offer different insights into the pattern of survival among breast cancer patients. By comparing the performance of the discussed models, the study will provide insight into the strengths and limitations of each approach and guide the selection of the most appropriate model for breast cancer survival analysis.
- We did not include the semi-parametric Cox model since our interest in this paper was to compare fully parametric survival distributions that yield explicit hazard and survival functions suitable for extrapolation and personalized prediction. While the Cox model is a powerful method, it does not give a closed-form survival function and is therefore less suited for personalized prognostic calculations.

2.1. Weibull Model

The Weibull model is a flexible parametric model, allowing a fit of various hazard function shapes such as monotonically increasing, decreasing, and constant hazards.

The Weibull functions are as follows:

$$h(t) = \lambda\gamma(\lambda t)^{\gamma-1} \tag{1}$$

$$f(t) = \lambda\gamma(\lambda t)^{\gamma-1}e^{-(\lambda t)^{\gamma}} \tag{2}$$

$$S(t) = e^{-(\lambda t)^{\gamma}} \tag{3}$$

where $\lambda > 0, \gamma > 0$.
Where h is the hazard function, f is the probability density function, and S is the survival function. Also, λ is a scale parameter and γ is the shape parameter. The Weibull distribution models extreme values and exhibits a wide range of tail behaviors.
Previous studies have applied the Weibull model to analyze the impact of different prognostic variables like tumor characteristics, lymph node status, and hormone receptor profile on the survival rates of breast cancer patients [12]. Nadler & Zurbenko [23] utilized an estimation of cancer latency times, which is the period between the onset of the disease and diagnosis, using publicly available data on the survival of breast cancer cases through the Weibull model.

2.2. Exponential Model

Exponential can be viewed as a function to model the failure rate with respect to time: It is a particular case of the Weibull model, where the shape parameter γ equals 1. This leads to a constant hazard function. This distribution possesses a property known as the memoryless property and is given as:

$$P(X > a + t \mid X > a) = P(X > t) \tag{4}$$

This property implies that if X denotes the lifetime of an object of interest given that the object has lasted for 'a' unit of time and is still functioning, then the probability that it will still be functioning for another 't' unit of time is independent of the 'a' units of time [19].
The exponential functions are given below:

$$h(t) = \lambda \tag{5}$$

$$f(t) = e^{-\lambda t} \tag{6}$$

$$S(t) = e^{(-\lambda t)} \tag{7}$$

where $\lambda > 0$, h is the hazard function, f is the probability density function, and S is the survival function. The exponential model has been used to describe the distribution of survival in breast cancer patients, and it provides a parsimonious and interpretable analysis of survival [21].
The exponential model assumes a constant hazard rate over time, and this assumption may not always hold for the survival data of breast cancer patients. Their hazard function is most likely non-constant and somewhat complex.

2.3. Log-Logistic Model

The log-logistic model is a parametric model that can accommodate a wide range of hazard function shapes, including non-monotonic patterns. It assumes that the outcome variable is nonnegative and that its logarithm follows a logistic distribution. It is used in survival analysis for hazard rates that initially increase and then decrease [25]. The log-logistic functions are given below:

$$f(\alpha,\gamma) = \frac{\alpha}{x}\left(\frac{x}{\gamma}\right)^{\alpha-1}\left[1+\left(\frac{x}{\gamma}\right)^{\alpha}\right]^{-2} \tag{8}$$

$$S(\alpha,\gamma) = \frac{1}{1+\left(\frac{x}{\gamma}\right)^{\alpha}} \tag{9}$$

$$h(\alpha,\gamma) = \frac{\frac{\alpha}{x}}{\left(\frac{x}{\gamma}\right)^{\alpha}\left[1+\left(\frac{x}{\gamma}\right)^{\alpha}\right]} \tag{10}$$

where $\alpha > 0, \gamma > 0$, h is the hazard function, f is the probability density function, and S is the survival function. The use of the log-logistic model is well known in the literature. For example, Al-Shomrani et al. [3] applied it to survival data on bladder cancer patients, while Khaleeq et al. [14] applied it to survival data on ovarian cancer patients.

2.4. Lognormal Model

The lognormal model is a parametric model that assumes the logarithm of the survival time follows a normal distribution.
The log-normal functions are given below:

$$f(\mu,\sigma) = \frac{1}{t\sigma\sqrt{2\pi}} e^{\frac{-(\ln(t)-\mu)^2}{2\sigma^2}} \tag{11}$$

$$h(\mu,\sigma) = \frac{1}{t\sigma}\phi\left(\frac{\ln(t)-\mu}{\sigma}\right) \tag{12}$$

$$S(\mu,\sigma) = 1 - \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right) \tag{13}$$

where:
t is the random variable.
μ is the mean of the natural logarithm of T.
σ is the standard deviation of the natural logarithm of T
$\Phi(x)$ is the normal cumulative density function (cdf), h is the hazard function, f is the probability density function, and S is the survival function.
Another parametric distribution used in breast cancer survival analysis is the log-normal model. This model provides further flexibility in modeling the distribution of survival time, especially when the underlying hazard function has a non-monotonic pattern. Applications of the lognormal model to survival analysis can be found in Royston [26], who compared the Cox and lognormal models using breast and ovarian cancer data, and in Chapman et al. [6], who described the lognormal model's estimation of survival using colon cancer data. Further details are also available from a review of parametric models in survival analysis [31].

2.5. Gamma Model

The Gamma model is a flexible parametric model that can accommodate various hazard function shapes, including monotonic, bathtub, and unimodal patterns.
The Gamma functions are given below:

$$h(t) = \frac{\beta}{t} \tag{14}$$

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} \tag{15}$$

$$S(t) = 1 - F(t) = 1 - \int_0^t f(x)dx \tag{16}$$

where $\alpha, \beta > 0$, h is the hazard function, f is the probability density function, S is the survival function, and Γ is the Gamma function. The shape (α) and scale (β) parameters are assumed to be constant over time and observations.
The Gamma distribution is a flexible parametric model that can accommodate various shapes of the hazard function, including monotonically increasing, decreasing, and non-monotonic patterns. Based on past studies, the Gamma model has shown promising results in modeling survival in breast cancer patients.

2.6. Gompertz Model

A Gompertz model is a parametric model that assumes the hazard function increases exponentially over time. In this case, the assumption is that the probability of mortality increases exponentially with age.
The Gompertz functions are given below:

$$h(t,\lambda,\gamma) = \lambda e^{\gamma t} \tag{17}$$

$$f(t,\lambda,\lambda\gamma) = e^{\gamma t} e^{\frac{\lambda}{\gamma}(1-e^{\gamma t})} \tag{18}$$

$$S(t,\lambda,\gamma) = e^{\frac{\lambda}{\gamma}(1-e^{\gamma t})} \tag{19}$$

$\lambda > 0, \gamma \in (-\infty, +\infty)$.
where h is the hazard function, f is the probability density function, and S is the survival function.
The Gompertz model has been used to model the survival of breast cancer patients, especially in the context of evaluating the presence of a cured fraction within the patient population [27].
These models are applicable in survival analysis, and the goal of this study is therefore to identify the most applicable in the study of cancer.

2.7. Model Assumptions and Their Clinical Implications

Each of the parametric survival models entails a different structural assumption about the hazard function and about the impact of covariates on survival time. All these assumptions have important implications for interpretation and clinical relevance.
A. Weibull and Exponential Models (PH/AFT Hybrid):
The Exponential model assumes a constant hazard over time, which implies that a patient's instantaneous risk of death does not change as follow-up progresses. The Weibull distribution relaxes this by allowing hazards that either increase or decrease monotonically depending on the value of its shape parameter. Both models belong to the small class of distributions that support both proportional hazards and accelerated failure time interpretations. Thus, covariate effects can be interpreted either as hazard ratios or as acceleration factors. Their monotonic hazard structure may not fully reflect the biology of breast cancer, where risk often peaks after diagnosis and treatment.
B. Gamma and Gompertz Models (PH Framework):
The Gamma model generally produces flexible but monotone hazards, whereas the Gompertz distribution assumes an exponential increase in hazard with time. These strictly increasing hazard shapes imply a progressively worsening mortality risk, which may align with some aggressive cancers but less so for heterogeneous breast cancer cohorts where risk often stabilizes in later years. Because these models operate within a proportional hazards framework, covariate effects reflect multiplicative changes in the hazard.
C. Lognormal and Log-logistic Models (AFT Framework):
Lognormal and Log-logistic distributions do not meet the proportional hazards assumption; instead, they assume that covariates act through an accelerated failure time mechanism, stretching or compressing the survival time scale. Notably, both allow non-monotonic hazard functions that rise, peak, and then fall- a pattern consistent with breast cancer survival in many populations. For instance, a Lognormal hazard typically rises early and declines rapidly, while the Log-logistic model has a heavier tail, implying slower hazard decay at late follow-up. Several characteristics make them suitable for diseases in which early mortality risk is high, but survivors tend to stabilize over time.
Of the models considered, the Lognormal provided the best fit to the observed patterns in the data, supporting a peaked hazard that gradually declines, consistent with clinical evidence that mortality risk in breast cancer patients tends to peak in the initial post-diagnosis period before stabilizing among long-term survivors.

2.8. Model Flexibility and Alternative Parametric Families

Although this study focused on six widely used parametric survival models, more flexible families were considered conceptually, such as the Generalized Gamma distribution, a three-parameter model that contains the Weibull, Gamma, and Lognormal distributions as exceptional cases, thus providing greater flexibility, estimating its

additional parameters is more complicated and may lead to numerical instability when the dataset has a high percentage of censoring. As we aimed to compare widely used and clinically interpretable models, we restricted our current analysis to the six 'classical' distributions, while noting that the Generalized Gamma family represents an advantageous avenue for future research. Its inclusion in a follow-up study may further confirm or refine the conclusions reached by the current model comparison.

3. Model Comparison

3.1. Akaike Information Criterion (AIC)

The Akaike Information Criterion is a measure of the relative quality of statistical models for a given set of data. It balances the goodness-of-fit of the model with its complexity, penalizing models with more parameters to avoid overfitting [1]. AIC is calculated as:

$$\text{AIC} = -2\log L + 2k$$

where k is the number of parameters in the model and L is the likelihood. The lower the AIC value, the better the balance between model fits and complexity. Additionally, AIC is used to evaluate the relative quality of statistical models for a given dataset.

3.2. Bayesian Information Criterion (BIC)

The Bayesian Information Criterion is another measure of model fit. It is calculated as:

$$\text{BIC} = -2\log L + k\log(n)$$

where n is the sample size, k is the number of parameters in the model and L is the likelihood. BIC imposes a more substantial penalty on models with more parameters than AIC. As with AIC, lower BIC values indicate a better balance between model fit and complexity. BIC is similar to AIC, but it imposes a more substantial penalty for model complexity [9, 24].

3.3. Log-likelihood

The log-likelihood measures how well a model fits the observed data. It is calculated by taking twice the negative log-likelihood of the model. Lower values indicate a better fit of the model to the data. The model with the highest log-likelihood value is considered the best fit [17, 24].

3.4. Deviance

It is a statistical measure used to assess the goodness-of-fit of a model. In the context of survival analysis, deviance helps to compare how well different parametric models fit the observed survival data. The deviance is defined as:

$$\text{Deviance} = -2\log L^*$$

where L^* is the maximized value of the likelihood function for the model.
Lower Deviance indicates a better fit to the data, while higher Deviance indicates a poorer fit.

4. Results

4.1. Summary Measures

From the analysis, the mean age at diagnosis is about 53 years, with a standard deviation of about 10 years. This reflects a wide age range, from 21 to 80 years. The mean tumor size is about 29 mm, while the mean number of

affected lymph nodes is 5, ranging from 1 to 51; these numbers show a wide range of disease spread among patients. The mean progesterone receptor level is about 103, with a range from 0 to 2380, indicating a wide distribution and substantial variation in receptor levels. Similarly, the mean number of estrogen receptors is about 94. Patients develop recurrence, on average, after 1124 days from the time of diagnosis, ranging from 8 days to 2659 days, nearly 7.3 years. The mean survival time is about 1321 days, ranging from 8 days to about 7.3 years.

Most patients are menopausal (57.7%), and a significant majority of patients did not receive hormone therapy (64.1%). Also, most patients have Grade 2 tumors (64.7%), whereas fewer have Grade 1 (11.8%) or Grade 3 (23.5%) tumors. 75.1% of the patients are censored regarding death, which means their survival status has not reached the endpoint of death within the period of this study, and only 24.9% of patients have died.

Cox–Snell Residual Diagnostics

To evaluate the overall adequacy of the models, Cox–Snell residuals for the Lognormal model were calculated and compared to the theoretical Exponential (1) reference distribution. Figure 1 illustrates that the Kaplan–Meier estimate of the residual survival curve (solid line) closely tracks, for most of the range, the exponential reference line (dashed). This demonstrates that the Lognormal model fits the observed data adequately. The slight deviation in the upper tail was expected, given the high censoring proportion of 75.1%, which diminishes information at longer survival times. Overall, the Cox–Snell residual plot supports the adequacy of the Lognormal model for breast cancer survival in this cohort.

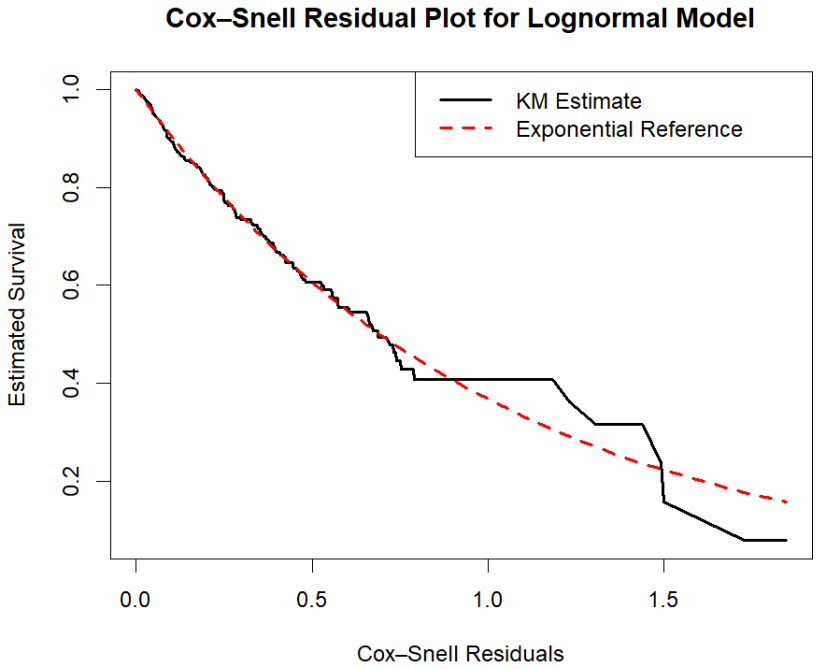

Figure 1. Cox–Snell residual plot for the Lognormal survival model.

4.2. Model Analysis

Based on AIC, BIC, Log-Likelihood, and Deviance, the Log-normal model consistently emerged as the best-fitting model for this breast cancer dataset. The Log-logistic and Gamma models also performed very strongly across most criteria and could serve as good alternatives. Based on all fit criteria, the Exponential model is consistently rated as the worst-fitting model, while the Weibull and Gompertz models are less preferred but still quite reasonable compared to the Exponential model.

Across all models, the following covariates reach statistical significance for consistently predicting the outcome of patients diagnosed with breast cancer: tumor size; tumor grade; and number of nodes involved all had an adverse effect, implying that bigger values are associated with poor outcomes; whereas the number of progesterone receptors showed a positive effect, which could be interpreted to mean that higher progesterone receptors result in better outcomes.

Tumor size was a consistently significant predictor across the different models. Larger tumor sizes were associated with worse prognosis, hence the importance of early detection. Clinically, this underscores the importance of robust screening programs that can detect tumors at earlier stages, which may be more treatable. Routine mammography and other imaging technologies are important in reducing the tumor size at diagnosis, thereby significantly improving outcomes. Tumor grade is another predictor of prognosis; higher grades are associated with a negative outcome.

Also, the number of lymph nodes involved is a prognostic factor. The more nodes involved, the more the disease spreads and the poorer the prognosis. This finding supports the use of sentinel lymph node biopsy and axillary lymph node dissection in surgical planning. Accurate assessment of nodal involvement helps stage the disease, plan adjuvant therapy, and evaluate the risk of recurrence.

An important positive predictor was the number of progesterone receptors, indicating that the greater the number of receptors, the better the prognosis. In addition, hormone receptor status has become a critical determinant in breast cancer for guiding the use of hormone therapies, such as tamoxifen or aromatase inhibitors.

Table 1. Summary Statistics of Patient Characteristics

	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Age at Diagnosis	686	21	80	53.05	10.121
Tumor Size	686	3	120	29.33	14.296
Number of Nodes	686	1	51	5.01	5.475
Number of Progesterone Receptors	686	0	2380	103.04	199.718
Number of Estrogen Receptors	686	0	1144	93.63	143.688
Time to Recurrence	686	8	2659	1124.49	642.792
Time to Death	686	8	2668	1320.61	619.197

Table 2. The Model Coefficients

	Weibull	**Exponential**	**Log-logistic**	**Lognormal**	**Gamma**	**Gompertz**
Variable	p-value	p-value	p-value	p-value	p-value	p-value
Age	0.582	0.7625	0.8441	0.95559	0.725	0.664
Menopause	0.6381	0.5975	0.4731	0.43046	0.499	0.6
Hormone	0.0855	0.2055	0.0811	0.07945	0.0845	0.0876
Size	0.0056[a]	0.0130[b]	0.0067[a]	0.0093[a]	0.0053[a]	0.0059[a]
Grade	0.0017[a]	0.0045[a]	0.0012[a]	0.0004[a]	0.0010[a]	0.0015[a]
Nodes	0.0000[a]	0.0000[a]*	0.0000[a]	0.0000[a]	0.0000[a]	0.0000[a]
Prog Recp	0.0001[a]	0.0001[a]	0.0000[a]	0.0000[a]	0.0000[a]	0.0000[a]
Estrg Recp	0.3849	0.4655	0.4752	0.35074	0.139	0.419

[a]Values are significant at least, at the 1% level. [b]Value is significant at the 5% level

Table 3. Model selection criteria of the fitted models for breast cancer data

Parametric Distributions	AIC	BIC	Log-Likelihood	Deviance
Weibull	3143.393	3188.7	-1561.7	3123.39
Exponential	3195.223	3236	-1588.6	3177.22
Log-logistic	3135.108	3180.42	-1557.6	3115.11
Lognormal	3129.144	3174.45	-1554.6	3109.14
Gamma	3136.579	3181.89	-1558.3	3116.58
Gompertz	3165.241	3210.55	-1572.6	3145.24

Table 4. Ten-fold Cross-Validation C-index for Parametric Survival Models

Parametric Distribution	C_index
Weibull	0.7349050
Exponential	0.7343449
Log-logistic	0.7313847
Lognormal	0.7342430
Gamma	0.2657871
Gompertz	0.2658146

4.3. Hazard Function Assessment of Parametric Models

To further assess the clinical plausibility of the competing parametric models, we estimated and compared the hazard functions for the Lognormal, Log-logistic, Weibull, and Gamma distributions (Figure 2). All the models had a low initial risk of death that increased over time; however, their long-term hazard shapes differed notably. The Lognormal model had a non-monotonic hazard that increased rapidly in the early follow-up period, then leveled off and finally began to decline slightly. This is biologically plausible for the progression of breast cancer, in which mortality risk typically peaks in the period following diagnosis and initial treatment and subsequently stabilizes as high-risk patients are removed from the risk set.

In contrast, the Weibull and Gamma models exhibited strictly increasing hazards over the entire follow-up, implying a monotonic acceleration of mortality risk with time. Although statistically reasonable, such monotonic acceleration is less in line with known clinical trajectories for early-stage breast cancer. The Log-logistic model exhibited a peak-and-decline structure similar to the Lognormal model, but with a heavier tail, implying a more gradual decline in hazard at later times. This further explained its weaker fit compared to the Lognormal distribution. Taken together, the analysis of the hazard shape supports the results of model selection: the Lognormal distribution not only offers the best statistical fit but also yields the most clinically interpretable hazard structure for the survival of breast cancer patients.

4.4. Consistency of Covariate Effects Across Parametric Models

Comparison of the covariate effects across the six fitted parametric models showed that some coefficients had different magnitudes and even directions, especially for hormone therapy and menopausal status. These do not represent contradictions, but stem from the different regression structures of the models. The Weibull, Exponential, and Gompertz are parameterized within the PH framework, in which the coefficients are interpreted as hazard ratios. In contrast, Lognormal, Log-logistic, and Gamma models have an accelerated failure time structure, with coefficients interpreted as time ratios. Since a positive coefficient in an AFT model indicates longer survival time while a positive PH coefficient indicates increased hazard, raw coefficients can appear to differ in sign even when they represent similar clinical effects.

To compare the interpretations, the coefficients of all PH-models were transformed into their equivalent time ratios, thereby making them directly comparable to the results from AFT-models. Most covariates, after transformation,

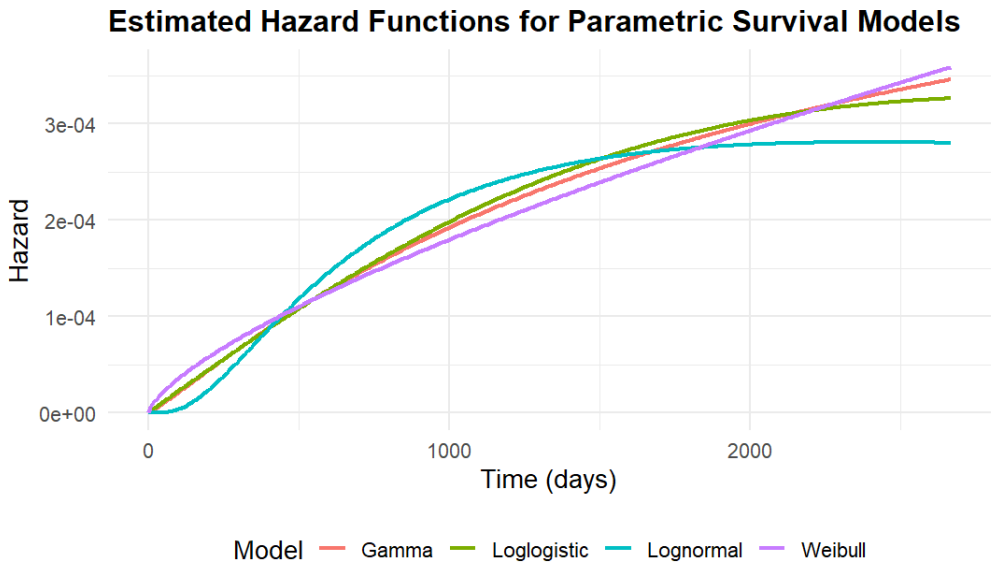

Figure 2. Estimated hazard functions for the Weibull, Lognormal, Log–logistic, and Gamma models.

had very consistent effects across models: age, tumour size, grade, and nodal involvement. Progesterone receptor status showed time ratios very near to 1.0 in all models (0.995–1.005), indicating a negligible effect.
AFT models, more appropriately capturing the non-monotonic hazard pattern, yielded stable and clinically coherent effects of hormone therapy and menopausal status. Contrasting estimates were obtained from PH-based models, which assume monotonic hazards and may thus not be appropriate for this dataset. These results confirm that the inconsistencies identified above were related to issues in the model structure, not to substantive differences in clinical interpretation.

Table 5. Comparison of Covariate Time Ratios After PH–AFT Harmonization Across Six Parametric Models

Covariate	Weibull	Exponential	Gompertz	Lognormal	Loglogistic	Gamma
Age	1.0039	1.0036	0.9948	0.9996	0.9985	1.0025
Menopause	1.0725	1.1424	0.8756	0.8832	0.8941	1.1069
Hormone therapy	0.8445	0.8084	1.3325	1.2047	1.1983	0.8419
Tumor size	1.0078	1.0118	0.9870	0.9914	0.9916	1.0081
Grade	1.3104	1.5064	0.6294	0.7200	0.7442	1.3326
Nodes	1.0327	1.0483	0.9478	0.9614	0.9618	1.0347
Progesterone receptor	0.9971	0.9955	1.0049	1.0026	1.0028	0.9973
Estrogen receptor	0.9996	0.9995	1.0006	1.0004	1.0003	0.9995

4.5. Model Validation & Performance

To validate and test the fitted parametric survival models, AIC, BIC, log-likelihood, Deviance, and out-of-sample predictive validation were employed. They provide a rigorous assessment framework that balances in-sample goodness-of-fit and predictive discrimination.
Information Criteria (AIC/BIC/Log-likelihood/ Deviance)
Of all the six models considered, Weibull, Exponential, Log-logistic, Lognormal, Gamma, and Gompertz, the Lognormal model consistently demonstrated the best in-sample fit, as indicated by the lowest AIC, BIC, and

Deviance, and the highest log-likelihood. This finding aligns with existing literature showing that Lognormal distributions often capture the heavy-tailed survival patterns characteristic of breast cancer datasets.

10-Fold Cross-Validation (Predictive C-index)

A 10-fold cross-validation procedure was also implemented to evaluate each model's ability to discriminate between patients with different survival outcomes. The Harrell's concordance index (C-index) served as a discriminatory measure.

Results in Table 4 show that the Weibull model outperforms other models, with a mean predictive C-index of 0.735, closely followed by the Lognormal (0.734) and Exponential (0.734) models. Notably, the difference between the Weibull and Lognormal models was negligible, $\Delta C \approx 0.001$, indicating equal predictive proficiency. The Gamma and Gompertz models performed very poorly, with C-indices near 0.26, indicating little discrimination ability for this dataset.

Overall, the joint evaluation from information criteria and cross-validation shows that:

- Lognormal model remains the strongest model based on both the overall fit and statistical support.
- The Weibull model yields slightly better predictive discrimination, though both models significantly outperform the rest.
- The Gamma and Gompertz distributions are inadequate to describe the survival distribution of this cohort.

This comprehensive validation enhances the methodological integrity of the study and supports the selection of the Lognormal model as a robust tool for personalized modelling of breast cancer survival.

4.6. Example of Personalized Survival Prediction Using the Lognormal Model

Assuming a postmenopausal woman 50 years of age with a tumour size of 20 mm, Grade 2 of the disease, three positive lymph nodes, a progesterone receptor count of 150, and having received hormone therapy. We want to establish the practical clinical value of the Lognormal model from the patient profile.

The Lognormal accelerated failure time model fitted yielded a median survival time of 3891 days for this patient, which is approximately 10.7 years.

This example shows how the model translates commonly available clinical data (variables) into actionable prognostic information. Such individual estimates may inform treatment discussion, facilitate follow-up planning, and enhance shared decision-making within clinical care.

4.7. Impact of Censoring on Model Performance

The dataset had a relatively high proportion of censoring of 75.1%, which was not unusual in long-term survival studies of breast cancer, where many patients were still alive at the end of follow-up [8, 15]. While high censoring may cause problems for estimating parameters in completely parametric models, studies indicate that lognormal, log-logistic, Weibull, and exponential distributions remain robust to moderate to heavy right-censoring [18, 22]. In this study, the Lognormal and Weibull models remained stable according to diagnostic assessments, cross-validation performance, and hazard-shape evaluation methods, which were consistent with other studies in survival analysis literature [16]. On the other hand, we agree that a high proportion of censoring might reduce the precision of estimates in the tail, especially in the upper tail of the survival distribution, as also discussed in other methodological papers [13]. Additional follow-up time in future studies or the use of external data may be required to validate these results further.

5. Conclusion

The Lognormal model had the lowest AIC value, indicating it is the best-fitting model according to this criterion. This result follows from the work of Gamel & Vogel [10], where the log-normal model was identified as a strong candidate among three models (Lognormal, Log-logistic, and Weibull) for survival analysis and prognostic modeling of breast cancer. Across all comparison criteria, the performance of the Lognormal model was the same,

reinforcing its position as the best-fitting model. Results from this study imply that while the lognormal model is best for survival analysis, at least for cancer data, the other models might not be misleading, since they all produced the same set of significant predictors.

The important covariates that significantly contributed to the prognosis of breast cancer included tumor size, tumor grade, number of nodes involved, and number of progesterone receptors. These have been consistently significant in several models and have played a crucial role in breast cancer outcomes. This finding agrees with the general evidence on prognostic factors in breast cancer [11, 29].

Clinical Implications and Application

The example of individualized prediction illustrates how easily the Lognormal model can be applied to obtain personalized survival estimates for newly diagnosed patients. Using the regularly available clinical variables of tumour size, grade, nodal status, and hormone receptor expression, the clinician may enter information to obtain risk-adjusted survival predictions for an individual patient. Such estimates may further inform prognosis counseling, aid in classifying patients for intensified treatment and monitoring, and provide a basis for shared decision-making.

The present study determines the critical factors for prognosis in breast cancer patients and identifies that the Lognormal model best describes the prognosis. Clinical practice will benefit by improving prognosis, treatment planning, and patient outcomes. Recommendations are that future research should be directed toward the validation and improvement of these models, the investigation of additional factors, and the development of other support tools to further enhance the care and treatment of breast cancer patients.

Note: All analysis code is available upon request for reproducibility.

Conflict of interest: There is no conflict of interest among the authors.

REFERENCES

1. H. Akaike, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, 1974.
2. C. P. Akuoko, E. Armah, T. Sarpong, D. Y. Quansah, I. Amankwaa, and D. Boateng, *Barriers to early presentation and diagnosis of breast cancer among African women living in sub-Saharan Africa*, PLoS ONE, vol. 12, no. 2, e0171024, 2017.
3. A. A. Al-Shomrani, A. I. Shawky, O. H. Arif, and M. Aslam, *Log-logistic distribution for survival data analysis using MCMC*, SpringerPlus, vol. 5, no. 1, 1774, 2016.
4. I. Ardoino, E. Biganzoli, C. Bajdik, P. Lisböa, P. Boracchi, and F. Ambrogi, *Flexible parametric modelling of the hazard function in breast cancer studies*, Journal of Applied Statistics, vol. 39, no. 7, pp. 1409–1421, 2012.
5. M. Bellanger, N. Zeinomar, P. Tehranifar, and M. Terry, *Are Global Breast Cancer Incidence and Mortality Patterns Related to Country-Specific Economic Development and Prevention Strategies?*, Journal of Global Oncology, vol. 4, pp. 1–16, 2018.
6. J. W. Chapman, C. J. O'Callaghan, N. Hu, K. Ding, G. A. Yothers, P. J. Catalano, Q. Shi, R. G. Gray, M. J. O'Connell, and D. J. Sargent, *Innovative estimation of survival using log-normal survival modelling on the ACCENT database*, British Journal of Cancer, vol. 108, no. 4, pp. 784–790, 2013, doi: 10.1038/bjc.2013.34.
7. T. G. Clark, M. Bradburn, S. Love, and D. G. Altman, *Survival Analysis Part I: Basic concepts and first analyses*, Springer Nature, vol. 89, no. 2, pp. 232–238, 2003, https://doi.org/10.1038/sj.bjc.6601118.
8. D. Collett, *Modelling Survival Data in Medical Research (3rd ed.)*, Chapman & Hall/CRC, 2015.
9. C. Cox, H. Chu, M. F. Schneider, and Á. Muñoz, *Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution*, Statistics in Medicine, vol. 26, no. 23, pp. 4352–4374, 2007.
10. J. W. Gamel and R. L. Vogel, *Comparison of parametric and non-parametric survival methods using simulated clinical data*, Statistics in Medicine, vol. 16, no. 14, pp. 1629–1643, 1997, https://doi.org/10.1002/(sici)1097-0258(19970730)16:14<1629::aid-sim594>3.0.co;2-c.
11. J. W. Gamel, R. L. Vogel, P. Valagussa, and G. Bonadonna, *Parametric survival analysis of adjuvant therapy for stage II breast cancer*, Cancer, vol. 74, no. 9, pp. 2483–2490, 1994.
12. N. Ghorbani, J. Y. Charati, K. Anvari, and N. Ghorbani, *Application of the Weibull Accelerated Failure Time Model in the Determination of Disease-Free Survival Rate of Patients with Breast Cancer*, Iran Journal of Health Sciences, vol. 4, no. 2, pp. 11–18, 2016.
13. D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time-to-Event Data (2nd ed.)*, Wiley, 2008.
14. J. Khaleeq, M. Amanullah, A. T. Abdulrahman, E. H. Hafez, and M. M. Abd El-Raouf, *Influence diagnostics in the Log-Logistic regression model with censored data*, Alexandria Engineering Journal, vol. 61, no. 3, pp. 2230–2241, 2022.
15. J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data (2nd ed.)*, Springer, 2003.
16. D. G. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text (3rd ed.)*, Springer, 2012.

17. Y. Lan and L. M. Leemis, *The logistic–exponential survival distribution*, Naval Research Logistics, vol. 55, no. 3, pp. 252–264, 2008.
18. J. F. Lawless, *Statistical Models and Methods for Lifetime Data (2nd ed.)*, Wiley, 2003.
19. J. P. Lehoczky, *Distribution, Statistical: Special and Continuous*, International Encyclopedia of the Social & Behavioral Sciences, pp. 3787–3793, 2001.
20. A. B. Mariotto, A. Noone, N. Howlader, H. Cho, G. Keel, J. Garshell, S. Woloshin, and L. M. Schwartz, *Cancer Survival: An Overview of Measures, Uses, and Interpretation*, JNCI Monographs, vol. 2014, no. 49, pp. 145–186, 2014, https://doi.org/10.1093/jncimonographs/lgu024.
21. C. M. McBride, B. W. Brown, J. Thompson, K. C. Westbrook, and C. A. Milne, *Can patients with breast cancer be cured of their disease? A sample of the M. D. Anderson hospital experience*, Cancer, vol. 51, no. 5, pp. 938–945, 1983.
22. W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, Wiley, 1998.
23. D. L. Nadler and I. G. Zurbenko, *Estimating Cancer Latency Times Using a Weibull Model*, Advances in Epidemiology, pp. 1–8, 2014.
24. G. D. Nigh, *Engelmann Spruce Site Index Models: A Comparison of Model Functions and Parameterizations*, PLoS ONE, vol. 10, no. 4, e0124079, 2015.
25. F. W. Nussbeck, *Log-Logistic Models*, In: A. C. Michalos (eds) Encyclopedia of Quality of Life and Well-Being Research, Springer, Dordrecht, 2014, https://doi.org/10.1007/978-94-007-0753-5_1691.
26. P. Royston, *The Lognormal Distribution as a Model for Survival Time in Cancer, With an Emphasis on Prognostic Factors*, Statistica Neerlandica, vol. 55, no. 1, pp. 89–104, 2001.
27. J. Scudilio, V. F. Calsavara, R. Rocha, F. Louzada, V. Tomazella, and A. S. Rodrigues, *Defective models induced by a gamma frailty term for survival data with a cured fraction*, Journal of Applied Statistics, vol. 46, no. 3, pp. 484–507, 2018.
28. R. Sha, X. Kong, X. Li, and Y. Wang, *Global burden of breast cancer and attributable risk factors in 204 countries and territories, from 1990 to 2021: results from the Global Burden of Disease Study 2021*, Biomarker Research, vol. 12, 87, 2024.
29. P. Tai, E. Yu, and D. Skarsgard, *Long-term survival rate of stages I-III small cell lung cancer patients in the SEER database: application of the lognormal model*, International Journal of Radiation Oncology, Biology, Physics, vol. 60, no. 1, S555, 2004.
30. P. Tai, E. Yu, R. Shiels, and J. Tonita, *Long-term survival rates of laryngeal cancer patients treated by radiation and surgery, radiation alone, and surgery alone: studied by lognormal and Kaplan-Meier survival methods*, BMC Cancer, vol. 5, 13, 2005, https://doi.org/10.1186/1471-2407-5-13.
31. N. Taketomi, K. Yamamoto, C. Chesneau, and T. Emura, *Parametric Distributions for Survival and Reliability Analyses, a Review and Historical Sketch*, Mathematics, vol. 10, no. 20, 3907, 2022.
32. B. Tasfa Marine and D. T. Mengistie, *Application of parametric survival analysis to women patients with breast cancer at Jimma University Medical Center*, BMC Cancer, vol. 23, 1223, 2023, https://doi.org/10.1186/s12885-023-11685-6.
33. K. J. Jager, P. van Dijk, C. Zoccali, and F. W. Dekker, *The analysis of survival data: The Kaplan–Meier method*, Kidney International, vol. 74, no. 5, pp. 560–565, 2008.