



Predictors of Length of Hospital Stay Among Breast Cancer Patients: A Survival Analysis Using Weibull Regression and Kaplan-Meier Estimates



Sule Omeiza Bashiru^{1*}, Daniel Ojochenemi Apeh² & Isaac Adaji³

^{1,3}Department of Mathematics and Statistics, Confluence University of Science and Technology, Osara, Kogi State, Nigeria

²Department of Biochemistry, Confluence University of Science and Technology, Osara, Kogi State, Nigeria

*Corresponding Author Email: bash0140@gmail.com

ABSTRACT

Hospital length of stay (LOS) is an important measure of healthcare resource use and quality of care. A total of 54 breast cancer patients were evaluated in this retrospective study to identify factors associated with LOS using Kaplan–Meier estimation and Weibull accelerated failure time (AFT) regression. The median LOS was around 10 days. The survival probability was 98.15% (95% CI: 90.34% to 99.67%) at Day 3 and fell to 0.0% at Day 35. Tumor stage, disease severity and multiple comorbidities were significant predictors of LOS ($p < 0.05$). Severity 2 and one comorbidity were not significant. Tumor stage 4 was significantly associated with LOS compared with the reference tumor category. Severity 3 was significantly different from the reference severity group, while Severity 2 was not significant. Two or three comorbidities were associated with a longer length of hospital stay, whereas one comorbidity was not significant. Age was also not significantly associated with LOS. The results indicate that tumor stage, disease severity and comorbidities are important factors associated with hospital LOS among breast cancer patients. The findings may help in identifying patients who are more likely to have longer hospital stays. This may support discharge planning and management of hospital resources. The Weibull AFT model was useful for examining the factors associated with LOS and provided interpretable estimates of hospital stay.

Keywords:

Hospital Length of Stay;
Survival analysis;
Weibull Regression;
Kaplan-Meier;
Tumor stage

INTRODUCTION

Hospital length of stay (LOS) is a highly important and relevant measure of healthcare delivery, impacting the patient, the disease burden, and also the efficiency of healthcare services. Nosocomial infection, the cost of healthcare services, scarcity of beds, and decreased patient satisfaction are among the consequences of prolonged hospital stay (Isa and Ali, 2026; Ravangard et al., 2011). Understanding what factors influence LOS is even more important in low-resource areas, where hospital beds and financial resources are limited, to facilitate effective discharge planning and clinical decision-making. Factors related to the tumor, the cancer therapy used, disease severity, comorbidities, and sociodemographic factors all affect hospitalization in oncologic patients, particularly breast cancer patients. Accurate estimation of LOS in such populations is essential to help in treatment planning, manage patient expectations, and reduce the economic burden on healthcare institutions and families.

Survival analysis has been developed as a powerful statistical tool to model time-to-event data and is unique in its ability to capture atypical data features such as right-skewed distributions and censored observations that are typical of hospital stay data (Cleves, 2008). The Kaplan-Meier estimator is a non-parametric method used to estimate the survival rate over time, while the Log-Rank test can be used to compare survival distributions between categorical groups. However, these methods are not able to handle multiple covariates at the same time. This limitation can, however, be addressed by using parametric accelerated failure time (AFT) models, which make distributional assumptions that result in clinically interpretable time ratios (Bradburn et al., 2003; George et al., 2014). The Weibull distribution is attractive because of its capability to model hazard functions that can increase or decrease over time, and hence it can be used to model many types of clinical survival data. For further discussion of the Weibull distribution and its applications, see Elgarhy et al. (2026).

Survival analysis has been demonstrated to be suitable for predicting LOS in hospitals in different healthcare settings in Nigeria. Audu and Godowoli (2024) used the Weibull regression model to determine the factors associated with LOS of patients admitted to the Emergency Department (ED) at Damaturu LGA, while Auwal et al. (2024) compared the Log-logistic model and the Exponential model in determining the factors associated with LOS of pediatric trauma patients in the ED at Asa-Akoko, using the information criterion for model comparison. Ali and Abbas (2025) applied the Weibull AFT model to hepatitis B patients in Maiduguri and found that age and comorbidities increased LOS, whereas antiviral therapy decreased it. Later, Ali and Abbas (2026) determined that the Log-Normal AFT model is more suitable than the Exponential model for this population. Okpala and Okoli (2020) compared the Weibull, Log-normal, and Log-logistic models for the survival of HIV patients with opportunistic infections in Anambra State and found the Weibull model to be the best-fitting parametric model, while Wahab (2008) used Cox regression analysis to analyse LOS among stroke patients, with hypertension and delayed presentation being important factors. Although these contributions have been made, limited research has applied survival analysis to investigate factors associated with hospital LOS among breast cancer patients, particularly the effects of tumor characteristics, disease severity, comorbidities, and age.

This study, therefore, seeks to tackle this problem by using two survival analysis tools, the Kaplan-Meier estimator and the Weibull Accelerated Failure Time (AFT) model, to identify predictors of hospital length of stay among breast cancer patients. The specific objectives are: (1) to estimate the overall survival function and median LOS using the Kaplan-Meier method; and (2) to use the Weibull AFT model to evaluate the joint effect of age, tumor stage, disease severity, and comorbidities on LOS. The results of this study will be helpful to evidence-based clinical practice as they identify patient subgroups that are at higher risk of extended hospitalization. This will support targeted interventions, enhance discharge planning, and help optimize resource allocation in breast cancer treatment.

MATERIALS AND METHODS

The study was a retrospective hospital-based study. It analyzed the length of hospital stay (LOS) among breast cancer patients at Specialist Hospital, Lokoja. The data were obtained from hospital records of breast cancer patients admitted between 2022 and 2026. Only 54 patients with confirmed breast cancer diagnosis were included. They were admitted and discharged during the

study period and had complete records. This ensured that the data were complete.

All the data were de-identified before analysis. The primary outcome was survival time (LOS). The variables included age (continuous in years), tumor stage (categorical: 2 = stage 2 and 4 = stage 4), comorbidities (categorical: 0 = none, 1 = one, 2 = two, and 3 = three), and disease severity (categorical: 1 = mild, 2 = moderate, and 3 = severe). The Kaplan-Meier method and the Weibull Accelerated Failure Time (AFT) model were used for the data analysis.

Kaplan-Meier Method

The Kaplan-Meier product-limit estimator was used to estimate the overall survival function (probability of remaining hospitalized over time). The survival function $S(t)$ is given by:

$$S(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where:

d_i = number of events (discharges) at time t_i

n_i = number of patients at risk just before time t_i

Kaplan-Meier survival probabilities were calculated at each observed discharge time. A survival curve was used to determine the median survival time (MST). The MST is the time when 50% of the patients remain hospitalized.

Weibull Accelerated Failure Time (AFT) Model

The Weibull Accelerated Failure Time (AFT) model was fitted to determine the effects of multiple predictors on the length of hospital stay. The Weibull distribution was selected because its hazard function can be increasing and decreasing. This makes it appropriate for survival data. The Weibull AFT model is given as:

$$\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \sigma W$$

where:

T = survival time (length of stay in days)

$X_1, X_2, \dots, X_k$ = covariates (Age, Tumor stage, Disease severity and Comorbidity)

$\beta_0, \beta_1, \dots, \beta_k$ = regression coefficients

σ = scale parameter

W = extreme value distribution error term

RESULTS AND DISCUSSION

Non-Parametric Kaplan-Meier Analysis

This section presents the results of the survival analysis used to determine the predictors of hospital length of stay among breast cancer patients. The results are presented below. They include the non-parametric Kaplan-Meier survival estimates, the Kaplan-Meier survival curve, and the Weibull Accelerated Failure Time (AFT) regression results.

Table 1: Kaplan–Meier Survival Estimates for Length of Hospital Stay

Time (Days)	Number at Risk	95% CI Lower	Survival Probability	95% CI Upper
3	54	0.9034	0.9815	0.9967
4	53	0.8282	0.9259	0.9701
5	48	0.6927	0.8102	0.8899
6	41	0.6617	0.7904	0.8791
7	39	0.5788	0.7094	0.8125
8	35	0.4611	0.5878	0.7037
9	29	0.3809	0.5067	0.6316
10	24	0.3524	0.4856	0.6208
11	23	0.3005	0.4222	0.5542
12	20	0.2613	0.3800	0.5150
13	18	0.2406	0.3589	0.4973
15	16	0.1882	0.2916	0.4223
17	13	0.1220	0.2019	0.3154
19	9	0.0974	0.1795	0.3071
20	8	0.0818	0.1570	0.2804
21	7	0.0668	0.1346	0.2527
22	6	0.0524	0.1122	0.2239
23	5	0.0389	0.0897	0.1934
25	4	0.0265	0.0673	0.1606
28	3	0.0154	0.0449	0.1238
35	1	0.0000	0.0000	0.0000

The Kaplan-Meier survival estimates (Table 1) demonstrated a steady decline in the probability of remaining hospitalized over time. At Day 3, the survival probability was 98.2% (95% CI: 90.3% to 99.7%), declining gradually to 92.6% (95% CI: 82.8% to 97.0%) by Day 4, and 81.0% (95% CI: 69.3% to 89.0%) by Day 5. The median length of hospital stay was approximately **10 days**, with survival probability falling below 50%

between Day 9 (50.7%) and Day 10 (48.6%). Thereafter, survival continued to decline, reaching 29.2% (95% CI: 18.8% to 42.2%) at Day 15, 15.7% (95% CI: 8.2% to 28.0%) at Day 20, and eventually 0.0% by Day 35, when all patients had been discharged. The widening confidence intervals observed at later time points reflect increased uncertainty due to the decreasing number of patients remaining at risk.

**Figure 1: Plot of the Kaplan-Meier with Median Survival**

The Kaplan-Meier survival curve for the length of hospital stay for 54 patients is displayed in Figure 1, with a median survival time of about 10 days. The curve is "stepwise declining," meaning that at each point of the curve, a number of patients were still hospitalized, with each "step" representing one or more discharges. The curve intersects the 0.5 (50%) survival probability line at around 10 days, implying that the median hospital stay was about 10 days. The number at risk table below the

curve shows that there were 54 patients at risk at the start of follow-up, and only 1 patient at risk at day 35, with a decreasing number of patients at risk over time.

Weibull Survival Regression

Table 2 presents the results of an accelerated failure time (AFT) regression model fitted to data from 54 breast cancer patients, with age, tumor stage, disease severity, and comorbidities as covariates.

Table 2: Weibull Accelerated Failure Time (AFT) Regression Analysis of Factors Associated with Length of Hospital Stay among Breast Cancer Patients

S/N	Variable	Estimate	Z	p-value
1	(Intercept)	-0.8287	-4.460	<0.01
2	Age	0.0001	0.010	0.996
3	Tumor4	0.0586	2.500	0.012
4	Severity2	0.0334	0.630	0.529
5	Severity3	0.1213	2.650	0.008
6	Comorbidities1	-0.0096	-0.180	0.856
7	Comorbidities2	0.1068	2.340	0.019
8	Comorbidities3	0.2566	2.580	<0.01
9	Log(scale)	-2.5791	-23.28	<0.01

The results of the Weibull Accelerated Failure Time (AFT) regression analysis to examine factors associated with the length of hospital stay of breast cancer patients are presented in Table 2. Results showed that tumor stage, disease severity and comorbidities were significant factors affecting length of hospital stay, while age, Severity 2 and one comorbidity category were not significantly associated with length of hospital stay. Compared with the reference tumor-stage category, Tumor stage 4 (Estimate = 0.0586, $p = 0.012$) was significantly associated with length of hospital stay. However, Severity 2 (Estimate = 0.0334, $p = 0.529$) was not significantly different from the reference group, while Severity 3 (Estimate = 0.1213, $p = 0.008$) was significantly different from the reference group, which suggests that length of hospital stay was significantly associated with disease severity. One comorbidity (Estimate = -0.0096, $p = 0.856$) was not significantly different from the no-comorbidity group, whereas two (Estimate = 0.1068, $p = 0.019$) and three (Estimate = 0.2566, $p < 0.01$) comorbidities showed statistically significant differences in length of hospital stay compared to the no-comorbidity group. The intercept (Estimate = -0.8287, $p < 0.01$) and Log(scale) (Estimate = -2.5791, $p < 0.01$) were also statistically significant.

CONCLUSION

This study used both Kaplan-Meier estimation and Weibull Accelerated Failure Time (AFT) regression to

determine the factors predicting the length of hospital stay for patients with breast cancer. The results show that tumor stage, disease severity and the number of comorbidities are important factors affecting the length of hospital stay. However, some categories of disease severity and comorbidity were not significant factors affecting LOS. Age was also not a significant factor affecting LOS. The Weibull AFT model provided clinically interpretable time ratios. It was also useful for the survival analysis of hospital stay data. These results should be tested in other centres. Other factors such as socioeconomic status, access to care and psychosocial factors should also be considered to further improve LOS prediction among breast cancer patients.

Acknowledgement

The authors gratefully acknowledge the financial support received from the Tertiary Education Trust Fund (TETFund) and Confluence University of Science and Technology, Osara, under the Institution-Based Research (IBR) Intervention Grant, with Grant ID: (CUSTECH/DRID/ADMIN/IBR/TETF/VOL.2.3)

REFERENCE

Ali, I. & Abbas, M. (2025). Application of Weibull Survival Regression for Predicting Hospital Length of Stay Among Hepatitis B patients in Maiduguri, Nigeria. *FUDMA Journal of Sciences*, 9(12), 262-268.

- Ali, I., & Abbas, M. (2026). Predicting hospital length of stay in hepatitis B patients: a comparison of survival regression models. *Journal of Statistical Sciences and Computational Intelligence*, 2(3), 739-751.
- Audu, A., & Godowoli, M. (2024). Prediction of hospital length of stay among patients admitted to the Emergency Department of General Hospital Damaturu using a Weibull regression model. *Arid Zone Journal of Basic and Applied Research*, 3(5), 47-57.
- Auwal, A. M., Sulaiman, A. I., & Isa, A. M. (2024). Comparison between Log-logistic and Exponential survival regression models in predicting hospital length of stay among pediatric trauma patients. *International Journal of Pure and Applied Science Research*, 11(13), 132-138.
- Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. *British journal of cancer*, 89(3), 431-436.
- Cleves, M. A. (2008). *An introduction to survival analysis using Stata* (3rd ed.). Stata Press.
- Elgarhy, M., Abdalla, G. S. S., Almetwally, E. M., Bashiru, S. O., Shawki, A. W., & Gemeay, A. M. (2026). Arctan inverted exponentiated Weibull model with actuarial measures and applications to insurance and inflation data. *Scientific African*, 31, e03260.
- George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of nuclear cardiology*, 21(4), 686-694.
- Isa, A. B., & Ali, M. (2026). Survival Analysis of Malaria Patients: Cox Proportional Hazards Model of Hospital Length of Stay. *International Journal of Empirical Studies and Statistical Models*, 6(5), 1-13.
- Okpala, S. T., & Okoli, C. N. (2020). Comparison on performance of the lognormal, log-logistic and Weibull distribution on survival of HIV patients with opportunistic infections in Anambra State, Nigeria. *European Journal of Statistics and Probability*, 8(2), 14-24.
- Ravangard, R., Arab, M., Rashidian, A., Akbari, S. A., Zare, A., & Zeraati, H. (2011). Comparison of the results of Cox proportional hazards model and parametric models in the study of length of stay in a tertiary teaching hospital in Tehran, Iran. *Journal of Health Administration*, 14(45), 45-56.
- Wahab, K. W. (2008). Profile of stroke in Nigerians: A prospective clinical study. *African Journal of Medicine and Medical Sciences*, 37(2), 165-169.