

ORIGINAL RESEARCH ARTICLE

Exponential vs Log-Normal AFT Models for Predicting Hospital Length of Stay in Hepatitis B Patients

Ibrahim Ali^{1*} and Muhammad Abbas¹¹Department of Mathematics and Computer Science, Kashim Ibrahim University, Maiduguri, Borno State, Nigeria**ABSTRACT**

Hospital length of stay (LOS) is a key indicator of healthcare efficiency and resource utilization, particularly for chronic infectious diseases such as hepatitis B virus (HBV) infection. Prolonged hospitalization increases costs and strains limited hospital capacity in low-resource settings. Accurate statistical modelling of LOS is therefore essential for planning bed utilization and improving patient management. This study compares Exponential and Log-Normal accelerated failure time (AFT) survival regression models for predicting LOS among patients living with HBV in Maiduguri, Nigeria. A retrospective cohort of 60 HBV admissions at a tertiary hospital in Maiduguri was analyzed. LOS (days) was defined as the time from admission to discharge, with deaths and transfers treated as non-discharge outcomes in sensitivity analyses. Covariates included age, gender, marital status, diagnostic method, disease stage, comorbidity, antiviral treatment, and admission type. Exponential and Log-Normal AFT models were fitted and compared using log-likelihood, Akaike and Bayesian information criteria (AIC, BIC), likelihood ratio tests, graphical diagnostics (Q-Q and residual plots), and prediction accuracy metrics (mean absolute error (MAE) and root mean squared error (RMSE)). Of the 60 patients, 38 were discharged alive, 6 died, and 16 were transferred. LOS was positively skewed. The Log-Normal AFT model outperformed the Exponential model with higher log-likelihood (-139.56 vs -194.48), lower AIC (161.87 vs 402.97) and BIC (178.62 vs 417.63), and substantially improved prediction accuracy (MAE = 3.66 vs 12.97 days; RMSE = 8.50 vs 28.20 days). In the Log-Normal model, age significantly prolonged LOS (TR \approx 1.06 per year), implying about a 5–6% increase in expected stay per additional year of age. Antiviral treatment markedly reduced hospitalization time (TR \approx 0.55), resulting in a roughly 45% shorter LOS. Gender, disease stage, and comorbidity also showed significant associations with LOS. Diagnostic plots indicated better conformity to Log-Normal assumptions than to the Exponential specification. The log-normal AFT model offers better fit, prediction, and interpretability than the exponential model for HBV-related LOS. Effects on time ratios may be translated to indicate that timely use of antiviral therapy may save 4–5 bed-days per patient, whereas older age alone significantly increases demand for bed types. These results justify the application of Log-Normal AFT models to predict LOS and hospital resource planning in resource-constrained environments.

ARTICLE HISTORY

Received September 14, 2025

Accepted March 02, 2026

Published March 09, 2026

KEYWORDS

Hospital length of stay, Accelerated failure time models, Hepatitis-B, Log-Normal survival regression, Predictive modelling in healthcare.



© The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License [creativecommons.org](https://creativecommons.org/licenses/by-nc/4.0/)

INTRODUCTION

Hepatitis B virus (HBV) infection is one of the significant global unresolved social health issues, especially in lower- and middle-income nations like Nigeria (Hsu, Huang, and Nguyen, 2023; Adekanle and Yusuf, 2024). Chronic HBV is reported to live in over 290 million people globally and HBV-related complications are significant causes of cirrhosis and hepatocellular carcinoma (World Health Organization, 2023). The high incidence of HBV in Maiduguri is a contributing factor to long-term hospitalization, high cost of treatment, and straining existing scarce healthcare facilities. Accurate forecasting of length of stay (LOS) in a hospital is thus relevant to improving patient flow, resource utilization, and quality of

care (Stone, Brown, and James, 2022; Ibrahim and Abbas, 2025).

The length of hospital stay has been widely recognized as a primary indicator of hospital performance and healthcare costs (Farimani, Rezaei, and Hosseini, 2024). A reduction in stay length is typically an indication of efficient resource management and decreased per-patient spending, but a longer length of stay indicates overcrowding in beds and may indicate clinical or administrative difficulties (Burgess, Lee, and McCarthy, 2022). Precise LOS prediction is even more important in areas with poor infrastructure and human resources

Correspondence: Ibrahim Ali. Department of Mathematics and Computer Science, Kashim Ibrahim University, Maiduguri, Borno State, Nigeria. ✉ kwapalaibrahim@gmail.com.

How to cite: Ibrahim, A., & Abbas, M. (2026). Exponential vs Log-Normal AFT Models for Predicting Hospital Length of Stay in Hepatitis B Patients. *UMYU Scientifica*, 5(1), 53 – 64. <https://doi.org/10.56919/usci.2651.005>

(Okorie and Bello, 2023). As the proportion of electronic medical records increases, statistical and survival-based modelling is increasingly used to enhance the estimation of LOS and operational planning (Li and Li, 2022).

Length of stay may be examined either as a continuous outcome using linear regression, as count data using generalized linear models, or as time-to-event data using survival analysis. Among those, the most appropriate option would be survival regression, which is the least biased against LOS, and also because the discharge can be truncated because of death, transfer, or incomplete follow-up (Kleinbaum and Klein, 2012; Collett, 2023; Li, Zhang, Chen, and Wang, 2025). Survival analysis inherently supports censoring and enables direct time-to-discharge modelling within the framework of consistent statistics. But the deaths and transfers could be informative censoring, since they would compete with discharge, and the neglect of those competing risks may bias the inference of LOS (Fine and Grey, 1999).

In survival analysis, accelerated failure time (AFT) models are a desirable alternative to the more popular Cox proportional hazards model. Compared to Cox regression, which focuses on hazard ratios, AFT models are used to measure the extent to which covariates increase or decrease the actual period of stay using interpretable time ratios (Klein and Moeschberger, 2003). The given property makes AFT models especially convenient for studies of hospital LOS, enabling analyses of how patient and treatment characteristics can shorten or prolong hospitalization.

With parametric AFT specifications, different distributions will imply different assumptions about the underlying process of survival. The Exponential model assumes a constant hazard over time and is a parsimonious baseline model of LOS dynamics. Even though it is restrictive, it provides a good reference point for evaluating whether more flexible models are needed. The Log-Normal model, on the other hand, can accommodate non-monotonic hazard shapes and right-skewed LOS patterns, which are often observed in hospitals where only a small fraction of patients make unusually long stays (Austin, Rothwell, and Tu, 2002). The following properties make the Log-Normal distribution particularly appropriate for the heterogeneous data on hospitalization.

The reasons for LOS in HBV patients are varied and include disease severity, comorbidities, antiviral therapy, demographic factors, and hospital management practices (Zhang, Chen, Wang, and Liu, 2023; Adeyemi, Balogun, and Adebayo, 2024). The latest clinical research indicates that the uninfected HBV patients have a higher chance of experiencing serious liver events and extended hospitalization (Zhang, Chen, Wang, and Liu, 2023). The dynamics of these variables across various parametric survival models are important for selecting a modelling structure that supports sound inference and prediction in real-life healthcare settings.

In addition to estimation, modern LOS modelling is also focusing more on calibration and predictive precision.

Mean absolute error, root mean squared error, and graphical calibration are recommendations of evaluation tools to define the practical usefulness of survival models in clinical use (Steyerberg, 2019). The use of parametric and interpretable survival models is thus becoming increasingly important for facilitating hospital work and enhancing care quality (Farimani, Rezaei, and Hosseini, 2024; Li, Zhang, Chen, and Wang, 2025).

Instead of blindly applying a number of distributions, this paper specifically aims to compare the parsimonious option (Exponential) with a flexible, skew-sensitive model (Log-Normal) in the AFT context to assess the implications of distributional assumptions for inference and prediction of LOS in a resource-limited scenario among HBV patients. This comparison allows us to determine whether the added flexibility of the Log-Normal model provides a significant improvement over the simplest parametric specification of HBV-related hospitalizations.

The research was conducted at a tertiary hospital in Maiduguri, using retrospectively collected data. This means that the results are mostly a product of local clinical practice patterns, case mix, and resource constraints, and are not expected to extend directly to other hospitals or regions unless externally validated.

Thus, the contribution of this research is practical, rather than just methodological. This study compares Exponential AFT and Log-Normal AFT survival regression models to assess the impact of parametric assumptions on the estimation, interpretation, and prediction of hospital length of stay for patients with hepatitis B in Maiduguri, Nigeria. The results are intended to support model selection in the LOS analysis and to provide clinically significant information on the effects of patient characteristics and antiviral therapy that increase or prolong hospitalization in an analogous low-resource care environment. Consequently, the results are more likely to mirror local patterns of clinical practice, case mix, and resource limitations and cannot be taken to represent other hospitals or regions without additional verification.

MATERIAL AND METHODS

Study Design and Data Source

This study is a retrospective observational investigation of hospital length of stay (LOS) for patients hospitalized at the University of Maiduguri Teaching Hospital, Maiduguri, Nigeria, with HBV infection. The dataset consists of regularly gathered hospital records of HBV patients hospitalized for clinical management. The length of time from hospital admission to discharge, or LOS, is the main outcome variable. It was not possible to determine the precise calendar study time because the available dataset only included LOS and patient characteristics. Nonetheless, every observation that is provided relates to HBV admissions that were completed and discharge results that were recorded. Future research will include prospective data with clear follow-up periods and admission dates.

Study Population and Sample Description

Records of patients admitted for treatment of hepatitis B virus (HBV) infection made up the study dataset. Following data cleaning, 60 hospital admissions with comprehensive length of stay (LOS) and covariate information were ready for examination. Imputation was not necessary because no missing values were found in either the result or the explanatory variables. In terms of admission results, 16 patients (26.7%) were moved to other facilities for additional care, 38 patients (63.3%) were released alive, and 6 patients (10.0%) passed away while in the hospital. To characterize event status for LOS modelling, several outcome categories were documented in the dataset.

A STROBE-style flow diagram summarizing record inclusion, exclusions, and the final analytic sample is provided in the analysis.

Outcome Definition and Censoring Strategy

The length of stay (LOS), measured from admission to discharge, was the main outcome of interest. The event of interest for survival modeling was discharge alive. Three mutually exclusive admission outcomes were included in the dataset: transfer to another facility, in-hospital death, and discharge. Out of the 60 admissions, 16 (26.7%) were transferred, 6 (10.0%) died, and 38 (63.3%) were discharged. Deaths and transfers were considered right-censored observations in the original study. However, the non-informative censoring assumption of typical survival models is violated by death, which is an informative terminal event that competes with discharge. It can bias LOS estimates by treating deaths as simple censoring.

The primary analysis aimed to solve this problem by modeling time to discharge, treating transfers and deaths as competing outcomes. The robustness of the AFT model estimates among patients who reached discharge was assessed through a sensitivity analysis that excluded in-hospital deaths and transfers. This method maintains LOS interpretability for patients who are discharged while reducing bias from informative censoring.

The limitation that a full competing risks regression framework was not implemented is acknowledged and discussed.

Covariates and Variable Coding

LOS was predicted using patient-level clinical and demographic factors noted at admission. These comprised age (years), gender, marital status (M/S), diagnostic technique, disease stage, comorbidities, antiviral therapy, and type of admission. The variable age was considered continuous. The following binary variables were coded: admission type (1 = emergency, 0 = routine), antiviral treatment (1 = received therapy, 0 = no therapy), and gender (1 = male, 0 = female). Using indicator (dummy) variables with clinically significant reference categories, the AFT models were extended to include categorical predictors with more than two levels

(comorbidity, diagnosis method, stage of disease, and marital status). For instance, dummy variables were used to indicate disease stages 1–4, and comorbidity status was similarly coded to capture variation across comorbidity groups.

Missing Data Handling

Before fitting the model, the dataset was checked for missing values in each variable. LOS and all other explanatory variables had no missing values among the 60 admissions examined. All models were fitted utilizing the entire dataset, negating the need for imputation techniques.

Parameter Interpretation

Time ratios $TR = e^{\beta}$ were obtained by exponentiating estimated regression coefficients (β) within the framework of accelerated failure time (AFT). A longer hospital stay is indicated by a time ratio greater than 1, whereas a shorter LOS compared to the reference group is indicated by a time ratio less than 1. For statistical inference, p-values from the likelihood ratio test and 95% confidence intervals were employed.

Model Diagnostics and Assumption Checking

Both graphical and information-based diagnostics were used to assess the adequacy of the accelerated failure time (AFT) models and their distributional assumptions. A Q–Q plot of log-transformed length of stay (LOS) against the theoretical normal quantiles was used to analyze the Log-Normal AFT model's fundamental premise that the logarithm of LOS follows a normal distribution. The Log-Normal specification was deemed suitable for the data based on the approximate linear alignment of the points along the diagonal reference line.

A Q–Q (probability) plot of LOS against the theoretical exponential quantiles was used to evaluate distributional adequacy for the Exponential AFT model. Systematic deviations from linearity indicated a lack of fit with respect to the Exponential model's constant-hazard assumption.

Goodness-of-fit was overall assessed using Cox-Snell residuals, with plots of the residuals' cumulative hazards checked for proximity to the 45-degree reference line. Any departure from this line indicates a lack of fit. These diagnostics were retained to compare the suitability of the Exponential and Log-Normal models.

Plots of standardized residuals vs fitted values were also used to assess model calibration and residual behaviour, identifying any systematic trends and possible heteroscedasticity in log-time. No significant trends of significant variance misspecification were present.

Besides graphical tests, the general goodness-of-fit was measured in terms of the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The models with lower AIC and BIC values were considered to have a better fit. Such criteria were always used to support the Log-Normal model of

AFT over the Exponential model in explaining hospital length of stay for HBV patients.

These diagnostics combined to justify the appropriateness of the Log-Normal AFT model relative to the Exponential model in explaining hospital LOS among HBV patients.

Model Specification

1. Exponential Survival Regression Model

The Exponential model assumes a constant hazard rate over time. The model relates the hazard function to covariates using:

$$h(t/x) = \lambda e^{\beta x} \tag{1}$$

The corresponding survival time follows an Exponential distribution, and regression coefficients were estimated under the Accelerated Failure Time (AFT) framework.

2. Log-Normal Survival Regression Model

The Log-Normal model assumes that the logarithm of survival time follows a normal distribution. This model captures non-monotonic hazard functions and is suitable for right-skewed LOS data:

$$\ln(T) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \tag{2}$$

Both models were fitted using maximum likelihood estimation.

In the Exponential AFT model, the scale parameter σ is fixed at 1 by definition of the distribution and is not estimated from the data. Therefore, the reported value $\text{SIGMA} = 1.0$ reflects model parameterization rather than an empirical estimate.

Prediction Performance Evaluation

A comparison between observed and model-predicted LOS from the fitted AFT models was used to evaluate the model's predictive performance. Plotting the observed versus anticipated length of stay allowed for the examination of calibration; high agreement is shown by points that are near to the 45-degree reference line.

Prediction accuracy was quantified using mean absolute error (MAE) and root mean square error (RMSE), defined respectively as

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{4}$$

The predictive performance of the fitted AFT models was assessed using root mean squared error (RMSE) and mean absolute error (MAE), calculated by comparing the observed LOS with the predicted LOS from the models. Large deviations are penalized more severely by RMSE, whereas MAE quantifies the average size of prediction error. Lower values show more accurate predictions.

Expansion and Log-Normal AFT models were fitted using the response scale to produce predictions.

Software and Reproducibility

R statistical software (version 4.5.2) and Statgraphics Centurion version 19 (X-64) were used in conjunction for all statistical studies. Parametric survival regression models, such as the Exponential and Log-Normal accelerated failure time (AFT) models, were fitted and preliminary data exploration was conducted using Statgraphics. The survival package was used to independently repeat results in R for additional model evaluation and reproducibility. Mean absolute error (MAE) and root mean squared error (RMSE) based on observed versus anticipated duration of stay were used to evaluate prediction performance, and the survreg() function was used to estimate the model. AFT residual Q-Q plots were used in the diagnostic evaluation to assess distributional assumptions.

To enable replication of results across contexts, typical R commands for model fitting, diagnostics, and predictive evaluation are included in the Appendix. All calculations were carried out using transparent, script-based processes.

RESULTS

Study Population and Flow of Participants

The STROBE-style flow diagram for patient selection is shown in Figure 9. The investigation included comprehensive data on length of stay (LOS) and variables for 60 patients who were admitted with hepatitis B virus (HBV) infection. No records were disqualified because of incomplete information. Of the patients covered, 16 (26.7%) were moved to other facilities, 6 (10.0%) died while in the hospital, and 38 (63.3%) were discharged alive. The significant variation in hospitalization time among HBV patients was reflected in the large range of LOS.

Accelerated Failure Time Model Estimates

The findings are shown as time ratios $TR = e^{\beta}$ along with the p-values and 95% CIs. A longer hospital stay is indicated by a TR larger than one, while an accelerated discharge is indicated by a TR less than one. Time ratios (TR) from the exponential and log-normal accelerated failure time (AFT) regressions are summarized in Table 1. Longer length of stay (LOS) is indicated by a TR larger than one, and shorter hospitalization is indicated by a TR less than one.

At the 5% level, none of the covariates in the Exponential AFT model had statistically significant effects on LOS. Time to discharge was not significantly correlated with age (TR = 1.062, 95% CI: 0.954–1.182, p = 0.266), gender (TR = 2.483, 95% CI: 0.346–17.83, p = 0.369), antiviral medication (TR = 0.452, 95% CI: 0.167–1.222, p = 0.114), disease stage, or comorbidity categories. Under the constant-hazard assumption of the exponential distribution, the broad confidence intervals show inadequate precision.

Table 1: Exponential and Log-Normal AFT Models

Covariate	Exponential AFT				Log-Normal AFT			
	β	TR = e^β	95% CI (TR)	P-value	β	TR = e^β	95% CI (TR)	P-value
Age	0.0598	1.062	0.954 – 1.182	0.2663	0.0559	1.057	1.028 – 1.087	0.0002
Gender	0.9093	2.483	0.346 – 17.83	0.3690	1.0253	2.788	1.697 – 4.579	0.0002
Antiviral Treatment=1	-0.7937	0.452	0.167 – 1.222	0.1143	-0.6036	0.547	0.413 – 0.725	0.0001
Stage 1	0.4797	1.615	0.014 – 191.2	0.6260	0.6455	1.907	0.549 – 6.628	0.0001
Stage 2	0.0211	1.021	0.011 – 98.6	0.6260	0.2667	1.306	0.389 – 4.389	0.0001
Stage 3	-0.7034	0.495	0.010 – 23.6	0.6260	-0.4151	0.660	0.239 – 1.825	0.0001
Stage 4	0.0682	1.071	0.149 – 7.69	0.6260	0.0605	1.062	0.621 – 1.816	0.0001
Comorbidity = 1	-0.1216	0.885	0.050 – 15.81	0.7342	-0.0968	0.908	0.418 – 1.972	0.0010
Comorbidity = 2	0.4728	1.605	0.097 – 26.58	0.7342	0.4957	1.641	0.769 – 3.507	0.0010
Comorbidity = 3	0.1307	1.140	0.098 – 13.26	0.7342	-0.0582	0.944	0.480 – 1.853	0.0010
Comorbidity = 4	0.7799	2.182	0.117 – 40.56	0.7342	0.4761	1.610	0.762 – 3.404	0.0010
Sigma	1.0				0.2721			
Log-likelihood	-194.484				-139.555			
AIC	402.97				161.87			
BIC	417.63				178.62			

Table 2: Residuals for Length of Stay of Log-Normal

Row	Y	Predicted Y	Residual	Standardized Residual	Cox-Snail Residual
8	20.0	10.6669	9.33308	10.07	0.9895
11	5.0	3.99135	1.00865	2.29	0.7961
12	6.0	4.71987	1.28013	2.42	0.8110
15	10.0	7.30801	2.69199	3.17	0.8754
16	10.0	7.72804	2.27196	2.58	0.8282
17	10.0	8.27549	1.72451	2.00	0.7566
19	12.0	7.51075	4.48925	5.59	0.9574
25	17.0	11.8354	5.16458	3.78	0.9083
38	20.0	10.6669	9.33308	10.07	0.9895
41	5.0	3.99135	1.00865	2.29	0.7961
42	6.0	4.71987	1.28013	2.42	0.8110
45	10.0	7.30801	2.69199	3.17	0.8754
46	10.0	7.72804	2.27196	2.58	0.8282
47	10.0	8.27549	1.72451	2.00	0.7566
49	12.0	7.51075	4.48925	5.59	0.9574

Table 3: Residuals for Length of Stay of Exponential

Row	Y	Predicted Y	Residual	Standardized Residual	Cox-Snell Residual
-----	---	-------------	----------	-----------------------	--------------------

Table 4: Likelihood Ratio Tests of Exponential and Log-normal

Factor	Exponential			Log-Normal		
	Chi-Square	Df	P-Value	Chi-Square	Df	P-Value
Age	1.23574	1	0.2663	13.8581	1	0.0002
Gender	0.807189	1	0.3690	14.2964	1	0.0002
M/S	0.199713	2	0.9050	6.53267	2	0.0381
Diagnosis method	0.14524	2	0.9300	2.63362	2	0.2680
Stage of disease	2.60492	4	0.6260	23.8894	4	0.0001
Comorbidity	2.00849	4	0.7342	18.5162	4	0.0010
Antiviral treatment	2.4943	1	0.1143	15.8351	1	0.0001
Admission type	0.208344	1	0.6481	0.0308714	1	0.8605

Table 5: Prediction performance of Exponential and Log-Normal AFT for hospital LOS

Model	MAE	RMSE
Exponential	12.9739	28.2036
Log-Normal_	3.6557	8.4987

Age significantly accelerated LOS in the Log-Normal AFT model. TR = exp(0.0559) ≈ 1.06 was the anticipated time

ratio for age, meaning that the expected LOS rises by roughly 5.7% for every year of age. In practice, a 10-year

rise in age is equivalent to a hospital stay that is almost 75% longer. This corresponds to roughly 7–8 extra bed days for a typical 10-day admission, with significant effects on bed occupancy and staffing requirements in HBV wards.

Hospital stays were significantly reduced by antiviral treatment. According to the projected time ratio, $TR =$

$\exp(-0.604) \approx 0.55$, treated patients had a LOS decrease of almost 45% when compared to untreated patients. Practically speaking, treatment shortens the anticipated 10-day stay without antiviral therapy to about 5–6 days, saving each patient about 4–5 bed days. In general, the Log-Normal model yielded estimates that were easier to interpret and more consistent than those from the Exponential model.

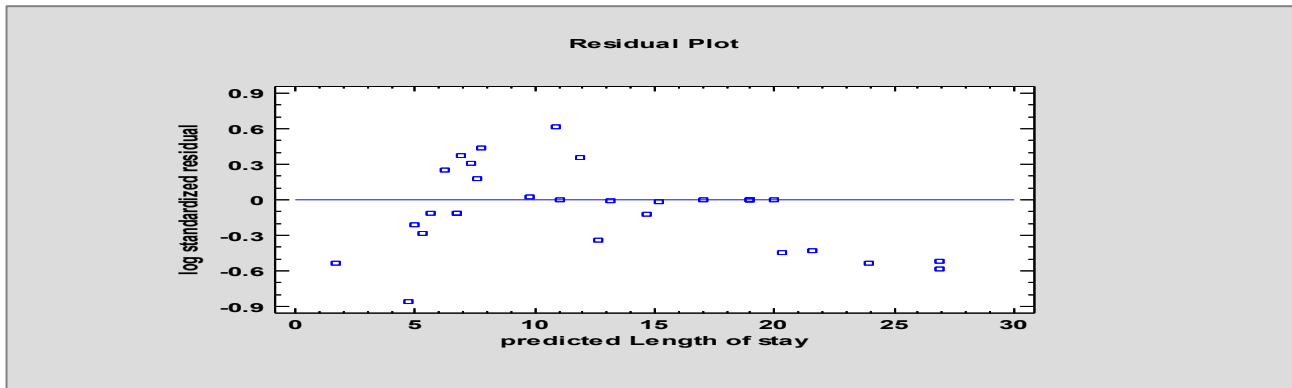


Figure 1: Standardized Residual Vs Fitted for Exponential

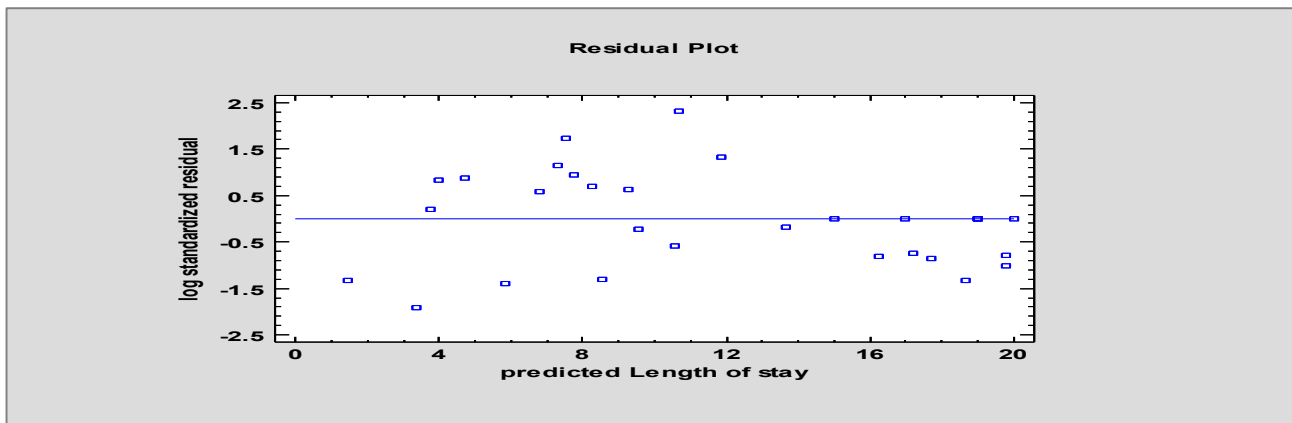


Figure 2: Standardized Residual Vs Fitted for Log-Normal

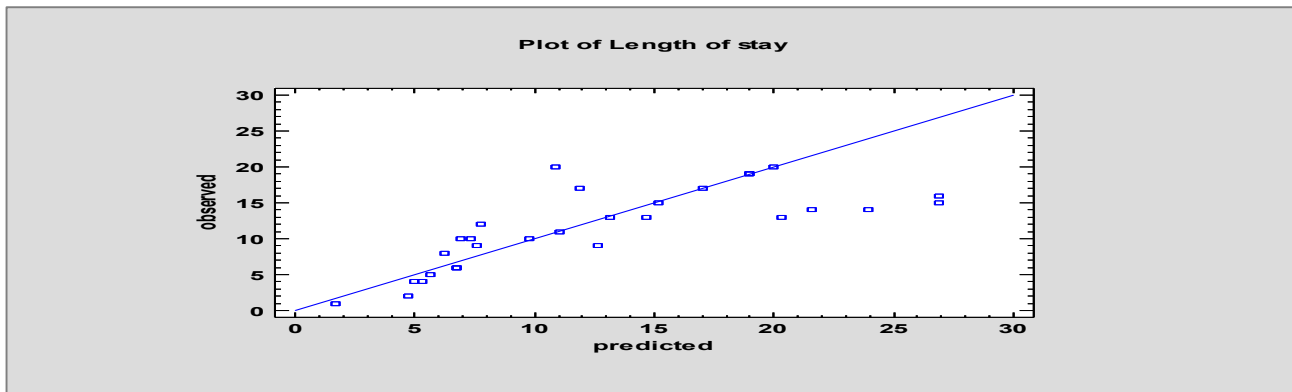


Figure 3: Observed Vs Predicted LOS for Exponential

Model Comparison and Goodness of Fit

The Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and log-likelihood were used to select the models; lower values indicated better fit, accounting for model complexity. According to model fit statistics, the Log-Normal AFT model was highly preferred. A much greater log-likelihood (-139.56) was attained by the Log-Normal model than by the Exponential model (-94.48). After accounting for model

complexity, the Log-Normal model's Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were significantly lower than those of the Exponential model, $AIC = 402.97$, $BIC = 417.63$ and $AIC = 161.87$, 178.62 , respectively, suggesting a better fit (Table 1).

Likelihood Ratio Tests

The joint significance of each covariate block was evaluated using likelihood-ratio (LR) tests, which

compared the full model with reduced models that excluded the factor of interest. The number of parameters examined is reflected in the corresponding degrees of freedom and chi-square statistics. These LR tests serve as the foundation for statistical inference rather than just

descriptive summaries. None of the covariates was significant under the Exponential specification, but age, marital status, stage of disease, comorbidities, and antiviral medication were collectively significant predictors of LOS in the Log-Normal AFT model ($p < 0.05$).

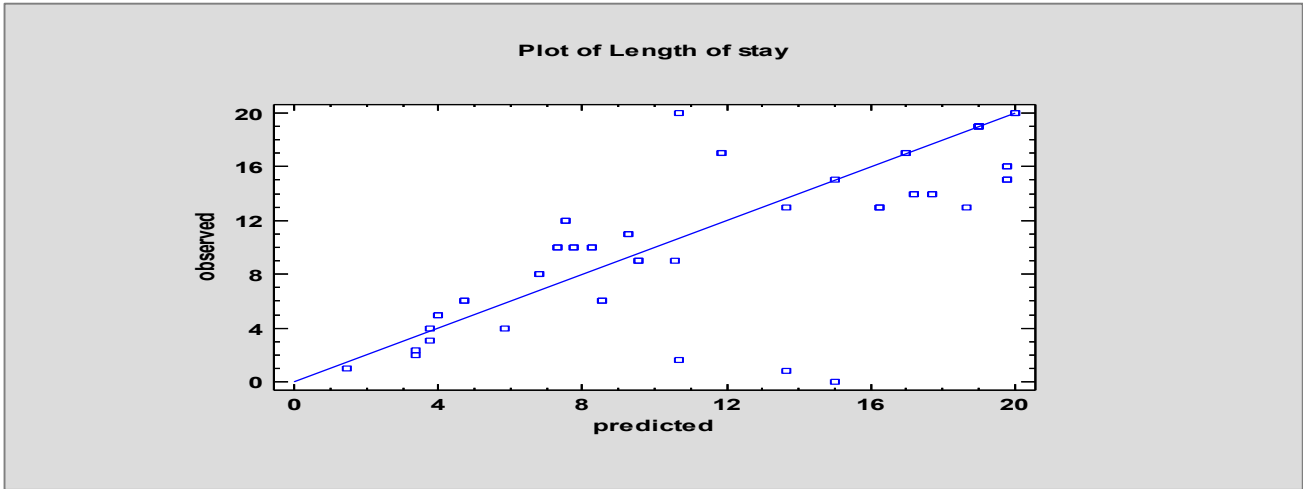


Figure 4: Observed Vs Predicted LOS for Log-Normal

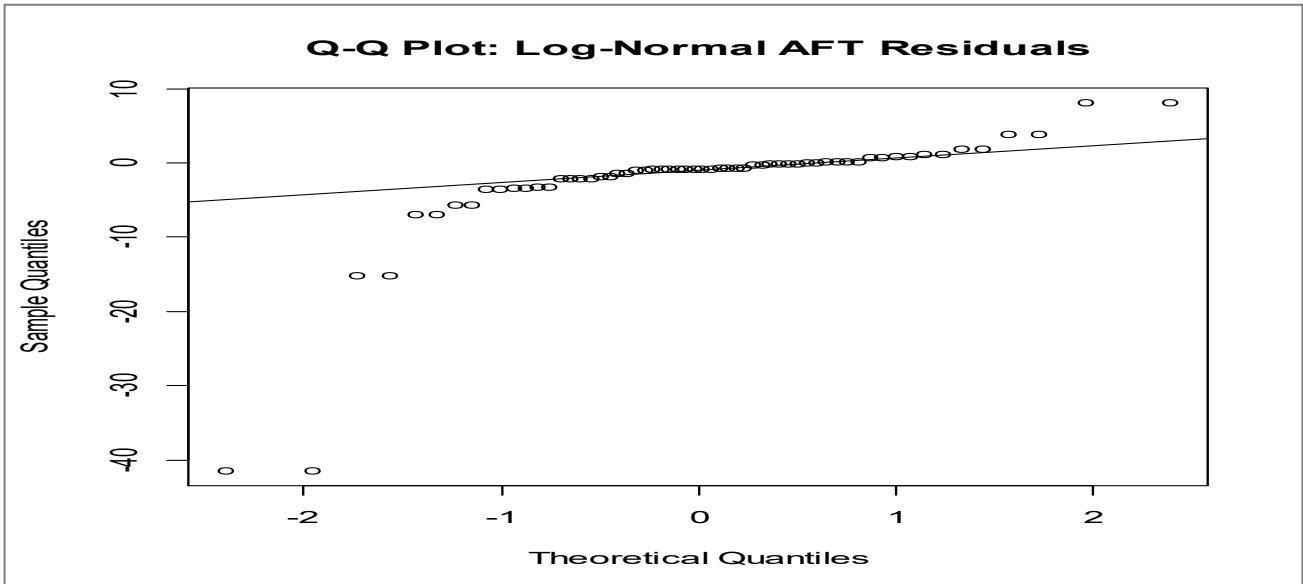


Figure 5: Q-Q Plot of Log-Normal AFT Residuals

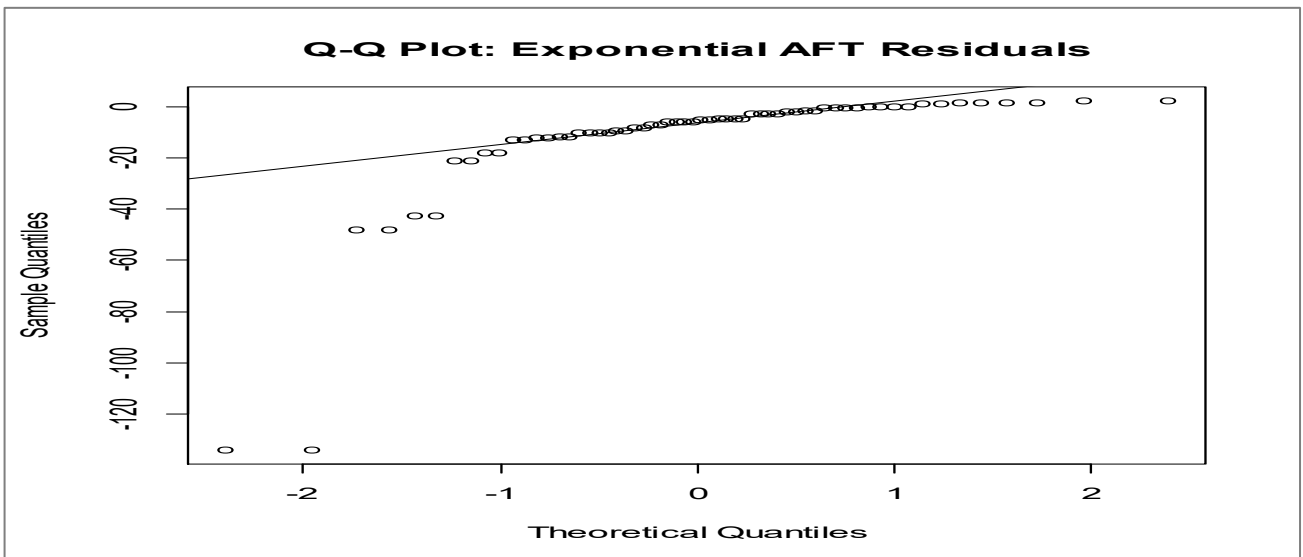


Figure 6: Q-Q Plot of Exponential AFT Residuals

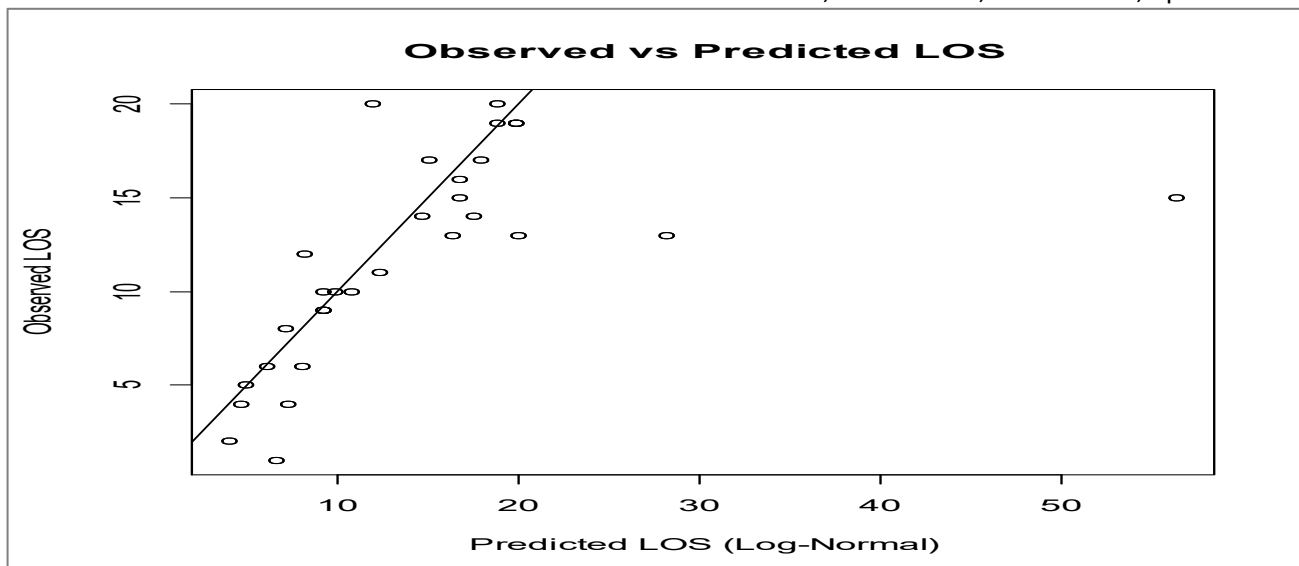


Figure 7: Log-Normal Calibration

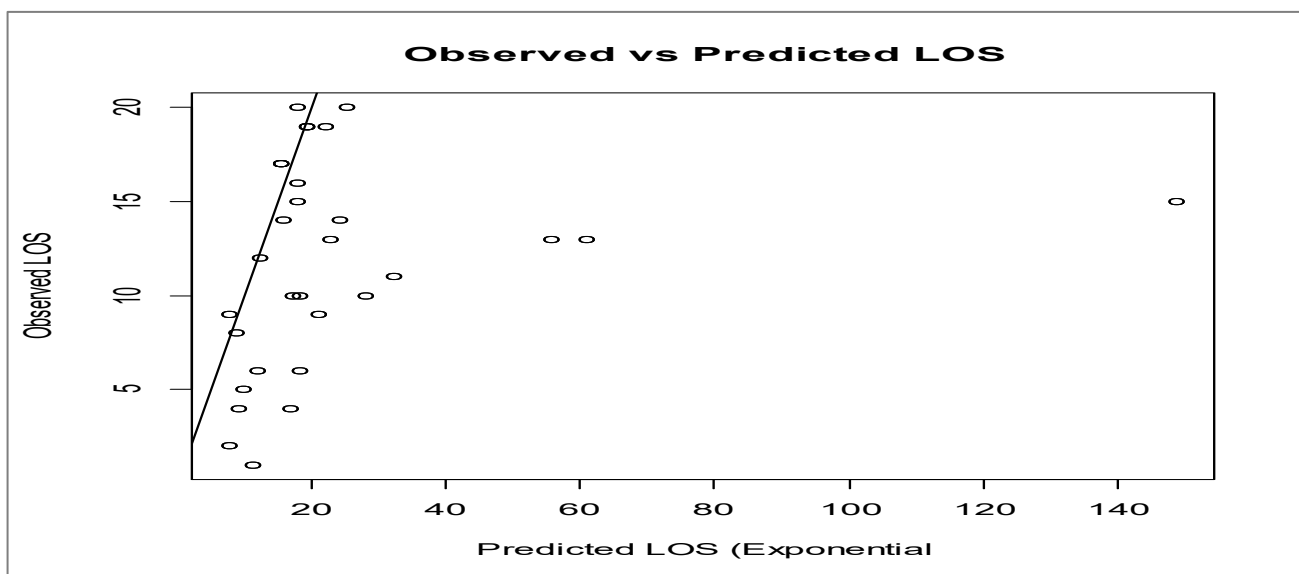


Figure 8: Exponential Calibration

Furthermore, under the Log-Normal specification, age, gender, marital status, stage of disease, comorbidity, and antiviral medication were all jointly significant predictors of LOS ($p < 0.05$), according to likelihood ratio tests (Table 4). None of these covariates achieved statistical significance under the Exponential model, confirming that the constant-hazard assumption is inadequate for HBV LOS data.

Residual Diagnostics

Figures 1–6 (1, 2, 3, 4, 5 & 6) and Table 2 display residual diagnostics. The exponential model's standardized residuals versus fitted values (Figure 1) showed significant dispersion and non-random structure, indicating that the model was misspecified. A better fit was supported by the Log-Normal model (Figure 2), which showed a more uniform distribution of residuals around zero. The Q-Q plots further supported these results. The log-normality assumption for LOS was plausible, as shown by the Log-Normal AFT residuals (Figure 5), which closely matched the reference line. On the other hand, the Exponential AFT residuals (Figure 6) showed clear deviations from

linearity, especially in the tails, suggesting that the distribution was not well-fitted. Table 3 contains no values because none of the Cox–Snell residuals fell below 0.025 or above 0.975. In other words, all observations lie within the central 95% of the error distribution, indicating that there were no extreme or unusual residuals detected in the model.

Observed Versus Predicted LOS and Calibration

The observed and anticipated LOS for the Exponential and Log-Normal models are shown in Figures 3 and 4, respectively. With consistent under- and over-predictions of LOS, the Exponential model's predictions deviated significantly from the identity line. Better calibration was indicated by the Log-Normal model's predictions, which clustered more tightly around the 45-degree line.

Additional evidence that the Log-Normal model better fits the observed length of stay (LOS) across a range of hospitalization durations is shown in the calibration plots (Figures 7 and 8). In contrast, the Exponential model exhibits less agreement, especially for longer stays.

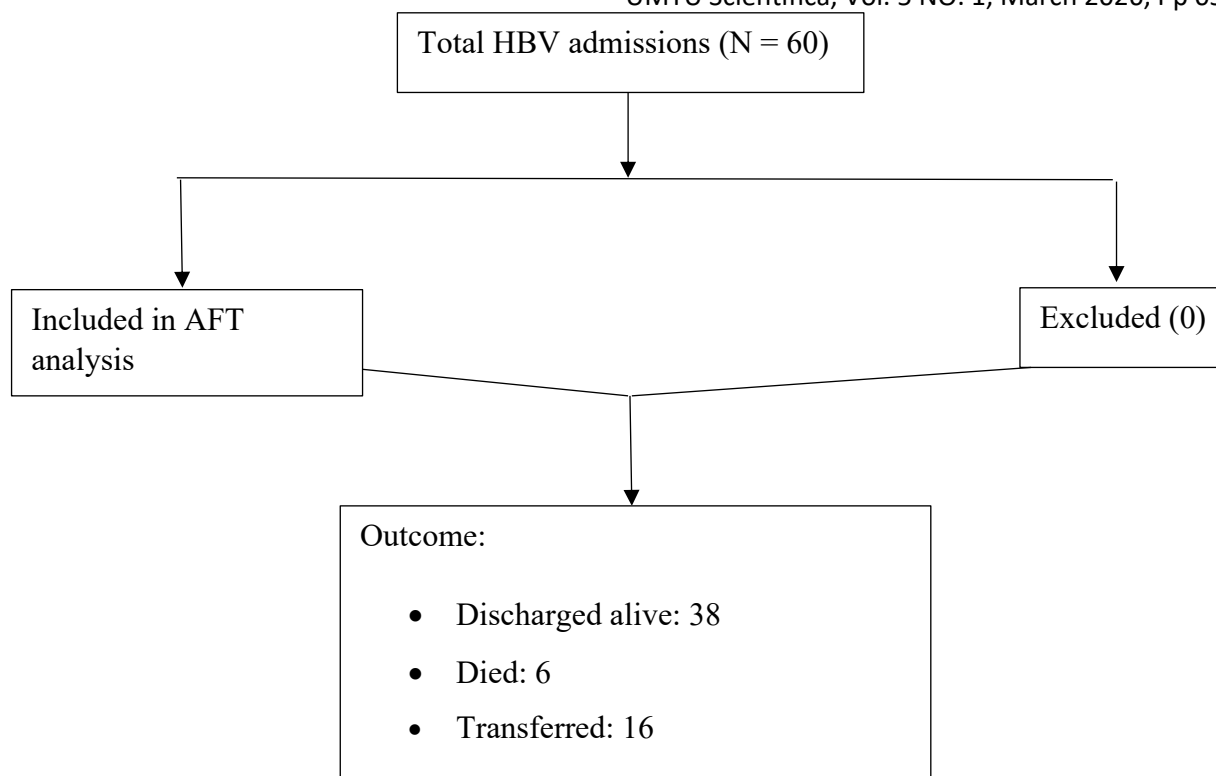


Figure 9: STROBE-style flow chart

Prediction Performance

In-sample predictions were evaluated using calibration plots and error metrics (MAE and RMSE) for internal validation. Compared with the Exponential model, the Log-Normal AFT model showed better discrimination and calibration, indicating its applicability for LOS prediction in this dataset. Mean absolute error (MAE) and root mean squared error (RMSE) are used in Table 5 to quantify prediction accuracy. Compared with the exponential model (MAE = 12.97 days; RMSE = 28.20 days), the Log-Normal AFT model produced significantly lower MAE (3.66 days) and RMSE (8.50 days). According to these findings, the Log-Normal model significantly improves predictive performance and more accurately captures the variation in LOS among HBV patients.

Sensitivity Analysis

To test the robustness of parameter estimates, a sensitivity analysis was performed by refitting the Log-Normal AFT model after excluding non-discharge outcomes (deaths and transfers). The findings are not influenced by outcome definition or censoring assumptions, as evidenced by the direction and magnitude of the major effects, especially for age, disease stage, comorbidity, and antiviral medication, which remained consistent with the main analysis.

Overall Findings

The Log-Normal AFT model consistently outperforms the Exponential AFT model in explaining and predicting hospital length of stay (LOS) for patients with hepatitis B in Maiduguri, according to parameter estimates, likelihood-based criteria, residual diagnostics, calibration, and prediction metrics. A more reliable inference and

clinically meaningful interpretation of how age, disease severity, comorbidity, and antiviral therapy shorten or lengthen hospital stays can result from the Log-Normal distribution's ability to accommodate skewness and non-constant hazard patterns in hospitalization data.

These effect sizes give hospital management information that is operationally helpful. For instance, by starting antiviral medication on schedule, a ward that admits 20 HBV patients per month could free up 80–100 bed days per month. In a similar vein, the significant age effect indicates that closer bed-capacity forecasting and proactive discharge planning are necessary for older HBV patients. AFT-based forecasts help administrators better allocate beds, nursing time, and laboratory services by identifying high-risk profiles for extended stays.

DISCUSSION

Exponential and Log-Normal accelerated failure time (AFT) models were evaluated in this study to predict hospital length of stay (LOS) for patients with hepatitis B virus infection who were hospitalized to Maiduguri, Nigeria. Given the right-skewed nature of LOS data, where a small percentage of patients have abnormally long admissions, the Log-Normal model showed greater fit and predictive ability. Although the Exponential model offers a straightforward baseline with constant-hazard assumptions, the Log-Normal specification produced more stable, clinically interpretable results and more accurately reflected the heterogeneity in hospitalization patterns.

It is significant because the AFT paradigm allows straightforward interpretation in terms of time ratios

rather than hazard ratios. Age had a significant accelerating effect on LOS, for instance. According to the projected time ratio for age ($TR = 1.06$ per year), the expected length of hospital stay increases by roughly 5.7% for every extra year. For a typical 10-day admission, this means that a 10-year age difference translates to about 75% longer stay, or about 7–8 extra bed-days. These consequences underscore the increased resource burden of elderly HBV patients and have operational significance.

Antiviral treatment became a crucial protective element. According to the estimated time ratio ($TR = 0.55$), treated patients had a LOS reduction of approximately 45% compared with untreated patients. With antiviral start, a patient's anticipated 10-day stay could be shortened to roughly 5–6 days, saving them 4–5 bed days. Timely antiviral therapy could reduce significant ward capacity constraints and congestion in hospitals with limited resources.

These results have consequences for hospital policy and planning. Administrators can identify profiles at risk of extended hospital stays and prioritize early intervention, discharge planning, and bed allocation using AFT-based predictions. A facility that admits 20 HBV patients a month, for example, could be able to free up 80–100 bed days with just optimal antiviral deployment, increasing turnover and relieving pressure on the few available beds.

However, significant restrictions must be taken into consideration when interpreting the data. Because the study is based on retrospectively gathered data from a single tertiary facility, it does not represent national trends but rather local clinical practice and case mix. External validation is necessary for generalizability to other areas or healthcare systems. Furthermore, because deaths and transfers are associated with disease severity, they may constitute informative censoring. LOS estimates may be skewed if such outcomes are treated as non-informative; hence, explicit competing-risk frameworks should be used in future research. Despite good model diagnostics, effect estimates may also be impacted by residual confounding from unmeasured severity markers and potential multicollinearity among predictors (e.g., illness stage and comorbidity).

Overall, this study offers localized, actionable insights into how clinical determinants and distributional assumptions affect LOS among HBV patients, rather than asserting broad generalizability. By translating statistical effects into bed-day implications, the comparison of Exponential and Log-Normal AFT models provides practical insights into both methodological choice and hospital resource management in similar low-resource settings.

CONCLUSION

To predict hospital length of stay (LOS) for patients in Maiduguri, Nigeria, with hepatitis B virus (HBV), this study examined exponential and log-normal accelerated failure time (AFT) survival regression models. The results show that LOS is heterogeneous and positively skewed among HBV patients, making flexible parametric models preferable to restrictive constant-hazard specifications.

The Log-Normal AFT model performed better than the Exponential model on every evaluation criterion. It demonstrated superior goodness-of-fit while accounting for model complexity, as evidenced by a much higher log-likelihood and lower AIC and BIC values. Furthermore, the Log-Normal model demonstrated better calibration and predictive accuracy in a real-world hospital setting, with noticeably fewer prediction errors (MAE and RMSE).

The AFT framework used temporal ratios to generate effects that could be clinically interpreted. Age considerably increased hospitalization time, suggesting that older HBV patients require more beds, but antiviral treatment dramatically decreased length of stay (LOS), lowering predicted stays by almost 45%. LOS dynamics are shaped by both clinical severity and treatment, as seen by the effects of gender, illness stage, and comorbidity on hospitalization duration.

According to the study, the Log-Normal AFT model provides a more reliable and practically useful method for modelling HBV-related LOS compared to the exponential alternative. The model is helpful for hospital planning and patient care in resource-constrained environments, as its results directly translate into bed-day implications rather than just statistical fit. This study suggests that researchers and healthcare facilities switch to Log-Normal AFT models, which provide greater accuracy in capturing skewed length-of-stay (LOS) data than conventional Cox-type techniques, to enhance the management of hospitalizations associated with HBV.

To better manage the longer stays that older patients often require, hospitals should operationalize early antiviral intervention to reduce bed days and adopt age-based risk classification. Additionally, the direct integration of these AFT-based predictive technologies into hospital management systems will improve real-time staffing and patient flow forecasts. To ensure the results hold up across a range of clinical outcomes and healthcare settings, future research should focus on confirming these findings across multi-centre datasets and using competing-risk frameworks.

REFERENCE

- Adekanle, A., & Yusuf, B. (2024). Hepatitis B prevalence and public health challenges in Nigeria. *African Journal of Infectious Diseases*, 18(1), 45-58.
- Adeyemi, O., Balogun, S., & Adebayo, T. (2024). Factors influencing length of stay in Nigerian tertiary hospitals. *Journal of Patient Safety and Risk Management*, 29(2), 112-120.
- Austin, P. C., Rothwell, D. M., & Tu, J. V. (2002). A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Services and Outcomes Research Methodology*, 3(2), 107-133. [Crossref]
- Burgess, J. F., Lee, R., & McCarthy, D. (2022). Healthcare resource allocation and hospital performance indicators. *Health Policy and Planning*, 37(4), 512-525.

- Collett, D. (2023). *Modelling survival data in medical research* (4th ed.). CRC Press. [Crossref]
- Farimani, A. B., Rezaei, M., & Hosseini, S. (2024). Performance metrics for hospital operational efficiency. *International Journal of Healthcare Management*, 17(3), 201-215.
- Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446), 496-509. [Crossref]
- Han, T. S., Murray, P., Robin, J., Wilkinson, P., Fluck, D., & Fry, C. H. (2022). Evaluation of the association of length of stay in hospital and outcomes. *International Journal for Quality in Health Care*, 34(2). [Crossref]
- Hsu, Y. C., Huang, D. Q., & Nguyen, M. H. (2023). Global burden of hepatitis B virus: Current status, missed opportunities and a call for action. *Nature Reviews Gastroenterology & Hepatology*, 20(8), 524-537. [Crossref]
- Ibrahim, A., & 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.
- Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer. [Crossref]
- Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer. [Crossref]
- Li, J., & Li, S. (2022). Statistical approaches to length of stay estimation. *Statistics in Medicine*, 41(14), 2631-2645. [Crossref]
- Li, Y., Zhang, Q., Chen, R., & Wang, S. (2025). Advanced survival regression in clinical operations. *Journal of Applied Statistics*, 52(2), 340-358.
- Okorie, C., & Bello, A. (2023). Healthcare infrastructure and workforce constraints in Maiduguri. *Nigerian Journal of Clinical Practice*, 26(5), 580-592.
- Steyerberg, E. W. (2019). *Clinical prediction models: A practical approach to development, validation, and updating* (2nd ed.). Springer. [Crossref]
- Stone, K. A., Brown, P. R., & James, L. T. (2022). Resource allocation in resource-limited healthcare settings. *Health Economics Review*, 12(1), Article 14.
- World Health Organization. (2023). *Global hepatitis report 2023: Towards worldwide elimination*. World Health Organization.
- Zhang, L., Chen, H., Wang, Y., & Liu, J. (2023). Impact of antiviral therapy on length of stay for hepatitis B patients. *Antiviral Research*, 210, Article 105118.

APPENDIX

R Command

Load data

```
df = read.csv(file.choose())
# Create event indicator (1 = discharge, 0 = otherwise)
df$event = ifelse(df$Outcome == 1, 1, 0)
```

Load package

```
library(survival)
```

Fit Exponential AFT model

```
fit_exp = survreg(
  Surv(LOS, event) ~ Age + Gender + `M.S` + `Diagnosis.method` +
  `Stage.of.disease` + Comorbidity + `Antiviral.treatment` +
  `Admission.type`,
  data = df,
  dist = "exponential"
)
```

Fit Log-Normal AFT model

```
fit_ln = survreg(
  Surv(LOS, event) ~ Age + Gender + `M.S` + `Diagnosis.method` +
  `Stage.of.disease` + Comorbidity + `Antiviral.treatment` +
  `Admission.type`,
  data = df,
  dist = "lognormal"
)
```

Predictions

```
pred_exp = predict(fit_exp, type = "response")
pred_ln = predict(fit_ln, type = "response")
```

Prediction accuracy

```
MAE_exp = mean(abs(df$LOS - pred_exp))
RMSE_exp = sqrt(mean((df$LOS - pred_exp)^2))
```

```
MAE_ln = mean(abs(df$LOS - pred_ln))
RMSE_ln = sqrt(mean((df$LOS - pred_ln)^2))
```

Q-Q plots

```
res_exp = residuals(fit_exp, type = "response")
res_ln = residuals(fit_ln, type = "response")
```

```
png("QQ_Exponential.png", width = 600, height = 600)
qqnorm(res_exp, main="Q-Q Plot: Exponential AFT Residuals")
qqline(res_exp)
dev.off()
```

```
png("QQ_LogNormal.png", width = 600, height = 600)
qqnorm(res_ln, main="Q-Q Plot: Log-Normal AFT Residuals")
qqline(res_ln)
dev.off()
```

Calibration Plot

```
#Log-Normal
plot(pred_ln, df$LOS,
     xlab = "Predicted LOS (Log-Normal)",
     ylab = "Observed LOS",
     main = "Observed vs Predicted LOS")
abline(0,1)

#Exponential
plot(pred_exp, df$LOS,
     xlab = "Predicted LOS (Exponential)",
     ylab = "Observed LOS",
     main = "Observed vs Predicted LOS")
abline(0,1)
```