Survival Analysis of HIV Treatment Outcomes among Patients in Nigeria: A Retrospective Cohort Study

Agnes Daniel Ibrahim[1], A. Abdulkadir[1], K. E. Lasisi[1]

https://doi.org/10.56201/ijasmt.vol.12.no11.2026pg13.31

Abstract
Nigerian studies of HIV treatment outcomes have relied almost exclusively on the Cox proportional hazards model, without comparing alternative specifications or testing the proportional hazards assumption. This study compared five survival models and proposed an ensemble that weights each model by its time-dependent area under the receiver operating characteristic curve (AUC). A retrospective cohort of 1,741 HIV-positive patients who started antiretroviral therapy at treatment centres of the Bauchi State Agency for the Control of AIDS between January 2018 and December 2022 was analysed. Exponential, Weibull, log-normal, Cox proportional hazards and Royston-Parmar flexible parametric models were fitted with eight covariates and evaluated using the Akaike and Bayesian information criteria (AIC and BIC), Harrell's concordance index (C-index) and time-dependent AUC at 6, 12, 24, 36 and 48 months. The cohort recorded 687 deaths (39.5%) over 84,895 person-months, or 8.09 per 1,000 person-months. The Weibull model had the lowest AIC and BIC and the highest AUC at every time point, ranging from 0.621 at 6 months to 0.709 at 48 months, and all models had a C-index of about 0.67. The AUC-weighted ensemble matched or slightly exceeded the best single model at most time points, but the gain was modest (maximum AUC difference 0.017) because one model dominated at all horizons. Model choice therefore has measurable consequences for HIV survival prediction in Nigeria, and horizon-specific evaluation provides information that global metrics cannot capture.

Keywords: *Antiretroviral therapy, Cox proportional hazards model, HIV, Nigeria, Survival analysis*

Introduction

Nigeria bears the largest burden of HIV in West Africa, with an estimated 1.9 million people living with the virus and a national prevalence of 1.4 percent among adults aged 15 to 49 years (Federal Ministry of Health, Nigeria, 2019). Although substantial progress has been made in expanding access to antiretroviral therapy (ART), with approximately 90 percent of diagnosed individuals receiving treatment (National AIDS and STDs Control Programme (NASCP), 2023), adverse outcomes including mortality and loss to follow-up remain significant concerns. Understanding the factors that influence patient survival following ART initiation and accurately predicting individual prognosis are therefore critical for improving clinical management and health policy formulation.

Survival analysis constitutes the appropriate statistical framework for modelling time-to-

[1] Department of Statistics, Faculty of Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria. Supervisors: Prof. A. Abdulkadir and Prof. K. E. Lasisi.

event outcomes such as mortality among HIV patients on ART, distinguished from standard regression techniques by its capacity to accommodate censored observations that arise when patients are lost to follow-up or have not experienced the event by the end of the study period (Collett, 2023). The semi-parametric Cox proportional hazards model (Cox, 1972) leaves the baseline hazard unspecified and estimates covariate effects through a partial likelihood, making it the most widely applied approach in medical research. More recently, flexible parametric models such as the Royston-Parmar approach have been developed, employing restricted cubic splines on the log-cumulative-hazard scale to model the baseline hazard without the rigid constraints of standard parametric forms (Royston & Parmar, 2002). Despite the availability of these alternatives, Nigerian HIV cohort studies have relied almost exclusively on a single modelling strategy. Udoh et al. (2025) analysed 5,608 patients in Bauchi State using the Cox proportional hazards model as the sole analytical method without testing the proportional hazards assumption or comparing any alternative model specification. Dzinza and Ngwira (2022) compared the Cox model with five parametric alternatives in a Malawian HIV cohort of 6,670 patients and found that parametric models performed equally well or better than the Cox model on information criteria. Elkalzah et al. (2025) demonstrated that distributional choice significantly influenced model fit and parameter estimates in a comparative analysis of parametric survival models for HIV patient data. Furthermore, Park and Pyo (Park & Pyo, 2021) emphasised that time-dependent ROC analysis provides essential information about time-specific predictive accuracy that global concordance measures such as Harrell's C-index obscure, yet such methods have not been applied to HIV survival model evaluation in Nigeria. The present study therefore seeks to fill this gap by providing a systematic comparison of parametric, semi-parametric, and flexible parametric survival models for HIV outcomes in Nigeria, using time-dependent AUC as the primary evaluation framework and extending the comparison to include a novel ensemble prediction approach.

The aim of this study is to conduct a comparative evaluation of parametric, semi-parametric, and flexible parametric survival models for HIV treatment outcomes in Nigeria, and to propose a time-dependent AUC-weighted ensemble prediction framework that combines the strengths of multiple models at clinically relevant time horizons.

The specific objectives are to: (1) fit and compare five survival models — exponential, Weibull, log-normal, Cox proportional hazards, and Royston-Parmar — to the HIV data; (2) evaluate each model using time-dependent AUC at 6, 12, 24, 36, and 48 months alongside AIC, BIC, and Harrell's C-index; (3) construct a time-dependent AUC-weighted ensemble from the five models and compare its performance against individual models and AIC-weighted model averaging; and (4) test the proportional hazards assumption and assess whether violations affect time-dependent AUC at specific time points.

This study makes three contributions to the analysis of HIV treatment outcomes in Nigeria. Practically, it provides evidence-based guidance on survival model selection stratified by prediction horizon, supporting more reliable clinical decision-making in resource-constrained settings. Theoretically, it introduces a time-dependent AUC-weighted ensemble prediction framework that extends standard model averaging (Burnham & Anderson, 2002) by allowing model weights to vary according to each model's discriminatory accuracy at specific time points, addressing a key limitation of global information criteria that mask time-specific performance differences. Methodologically, the study establishes a reproducible analytical pipeline — systematic model fitting, time-dependent ROC analysis, and ensemble construction — that moves beyond the single-model paradigm characterising current Nigerian HIV research and provides a template adaptable to other disease outcomes.

Literature Review

Conceptual Framework

HIV Burden and Treatment Outcomes in Sub-Saharan Africa

Sub-Saharan Africa remains the region most heavily affected by the HIV epidemic, accounting for approximately 65 percent of all people living with HIV globally and for the majority of AIDS-related deaths worldwide (UNAIDS, 2025).

HIV Burden and Treatment Outcomes in Nigeria

Nigeria carries the largest HIV burden in West Africa, with an estimated 1.9 million people living with the virus and a national prevalence of 1.4 percent among adults aged 15 to 49 years (Federal Ministry of Health, Nigeria, 2019). Although significant progress has been made in expanding ART coverage, with approximately 90 percent of diagnosed individuals receiving treatment (National AIDS and STDs Control Programme (NASCP), 2023), adverse outcomes remain a major concern.

Prognostic Factors in HIV Survival

A consistent body of evidence has identified key prognostic factors that influence survival among HIV patients on ART.

Theoretical Framework of Survival Analysis

Foundations of Time-to-Event Data

Survival analysis encompasses a family of statistical methods designed for the analysis of time-to-event data, where the outcome of interest is the time from a defined baseline (such as ART initiation) until the occurrence of a specified event (such as death).

Non-Parametric and Semi-Parametric Approaches

The Kaplan-Meier estimator, introduced by Kaplan and Meier (1958), is the standard non-parametric method for estimating the survival function from censored data.

The Cox proportional hazards (PH) model, introduced by Sir David Cox (1972), is the most extensively applied regression model in medical survival analysis.

A critical assumption of the Cox PH model is that the hazard ratio comparing any two individuals is constant over time.

Parametric Survival Models

Parametric survival models offer an alternative to the semi-parametric Cox approach by assuming a specific probability distribution for the survival times and estimating all parameters, including the baseline hazard, through full maximum likelihood.

Flexible Parametric Models

Flexible parametric survival models, introduced by Royston and Parmar (2002), bridge the gap between fully parametric and semi-parametric approaches.

Model Evaluation and Discrimination Metrics

The evaluation and comparison of survival models requires appropriate metrics that capture different dimensions of model performance.

Time-dependent ROC analysis, originally developed for survival data by Heagerty, Lumley, and Pepe (Heagerty et al., 2000), addresses these limitations by evaluating discriminatory accuracy at specific clinically relevant time points.

Empirical Review

Survival Analysis of HIV Outcomes in Sub-Saharan Africa

A growing body of research from sub-Saharan Africa has applied survival analysis methods to HIV cohort data, though the methodological sophistication of these applications varies considerably. Dzinza and Ngwira (2022) conducted a comparative analysis of parametric and Cox regression models using HIV data from 6,670 patients in Malawi.

Nigerian Studies on HIV Treatment Outcomes and Survival

The Nigerian literature on HIV survival analysis, while substantial in terms of clinical findings, is characterised by near-uniform reliance on the Cox proportional hazards model as the sole analytical approach. Udoh et al. (2025) analysed data from 5,608 patients receiving ART in Bauchi State, Nigeria, and applied the Cox PH model to identify predictors of mortality and LTFU.

Comparative Studies of Survival Models in Medical Research

Beyond the HIV-specific literature, a growing body of comparative methodological research has evaluated the relative performance of different survival modelling approaches across medical domains.

Model Averaging and Ensemble Methods in Survival Prediction

Standard model selection identifies a single best model from a set of candidates, but this approach ignores the uncertainty inherent in the selection process and discards potentially useful information contained in competing models.

Methodological Debates and Alternative Frameworks

The Proportional Hazards Assumption Debate

A persistent methodological debate in survival analysis concerns the practical consequences of proportional hazards assumption violations and the appropriate response when they are detected.

Global Versus Time-Specific Discrimination Metrics

A second debate centres on the adequacy of global discrimination metrics, particularly Harrell's C-index, for evaluating survival models.

Single-Model Versus Multi-Model Paradigms

A third methodological debate concerns whether applied survival analysis should proceed by selecting a single best model or by combining predictions from multiple models.

Research Gap

The preceding review of the literature identifies a specific and clearly defined methodological gap that the present study is designed to fill. Nigerian HIV cohort studies have relied almost exclusively on the Cox proportional hazards model without testing its assumptions, without comparing alternative parametric or flexible parametric specifications, and without evaluating model discrimination using time-dependent metrics that capture performance at clinically relevant time horizons.

Materials and Methods

Study Design

The study adopts a retrospective cohort design, which is the most commonly employed

observational design in HIV treatment outcome research in sub-Saharan Africa (Collett, 2023; Udoh et al., 2025). A retrospective cohort design identifies a defined population at a baseline time point — in this case, the date of antiretroviral therapy (ART) initiation — and follows that population forward in time using historical medical records to ascertain the occurrence of the outcome of interest.

Data Source

The data used in this study were obtained on request from the Bauchi State Agency for the Control of AIDS (BASACA), a non-governmental organisation providing comprehensive HIV care services in Bauchi State, North-Eastern Nigeria. BASACA operates five treatment centres across the state and maintains electronic medical records for all patients enrolled in its ART programme.

Dataset Characteristics and Preparation

The dataset contains records for 2,000 patients with the following variables: a unique anonymised patient identifier; age at ART initiation in years; sex (male or female); baseline CD4 count in cells per microlitre measured within three months prior to ART initiation; WHO clinical stage at ART initiation (Stages I through IV); body mass index in kilograms per square metre; functional status at baseline (classified as working, ambulatory, or bedridden); tuberculosis co-infection status at ART initiation (present or absent); the ART regimen initiated (categorised as tenofovir-based, zidovudine-based, or other); date of ART initiation; date of last follow-up or date of death; and vital status at last contact (alive or deceased). All patients who were alive at the end of the follow-up period, who were lost to follow-up, or who transferred to other facilities were classified as censored observations.

Sample Size Considerations

The sample size of $n = 2,000$ provides adequate precision for the objectives of this study.

Definition of Variables

Outcome Variable

The primary outcome is overall survival, defined as the time from ART initiation to death from any cause. For each patient i, the observed survival time T_i and event indicator δ_i are defined as:

$$T_i = min(t_{event,i},\ t_{censor,i}) \qquad (1)$$

$$\delta_i = \{1, if\ t_{event,i} \leq t_{censor,i}\ (death\ observed), 0, if\ t_{event,i} > t_{censor,i}\ (censored). \qquad (2)$$

Time is measured in months from ART initiation. Patients who are alive at the end of the study period, lost to follow-up, or transferred to other facilities are treated as censored at their date of last contact, under the assumption of independent censoring — that is, the censoring mechanism is unrelated to the probability of experiencing the event.

Covariates

Age at ART initiation (years): Modelled as a continuous variable (centred at 38 years) and categorised into four age groups: 15 to 29, 30 to 44, 45 to 59, and 60 years and above.

Sex: Categorised as male or female.

Baseline CD4 count (cells per microlitre): Modelled as a categorical variable with four levels: less than 200, 200 to 349, 350 to 499, and 500 or more cells per microlitre.

WHO clinical stage: Categorised as Stages I through IV according to the World Health Organization classification.

Body mass index (kg/m²): Categorised as underweight (less than 18.5), normal (18.5 to 24.9), overweight (25.0 to 29.9), and obese (30.0 or higher).
Functional status: Categorised as working, ambulatory, or bedridden.
Tuberculosis co-infection: Categorised as present or absent at the time of ART initiation.
ART regimen type: Categorised according to the backbone regimen: tenofovir-based, zidovudine-based, or other.

Descriptive and Exploratory Analysis
The descriptive analysis proceeds in two stages.
Second, the survival experience of the cohort is examined non-parametrically using the Kaplan-Meier estimator.

$$\hat{S}(t) = \prod_{i:t_i \le t} \left(1 - \frac{d_i}{n_i}\right) \qquad (3)$$

where t_i are the distinct event times, d_i is the number of events at time t_i, and n_i is the number of individuals at risk just before time t_i.

$$\widehat{Var}\left[\hat{S}(t)\right] = \left[\hat{S}(t)\right]^2 \sum_{i:t_i \le t} \frac{d_i}{n_i(n_i - d_i)} \qquad (4)$$

The median survival time is estimated as the smallest time t such that $\hat{S}(t) \le 0.5$, and 95 percent confidence intervals for the survival function are computed using the log-log transformation to ensure the interval respects the [0,1] bounds of the survival probability.
The log-rank test, developed by Mantel (1966) and extended by Peto and Peto (1972), is used to compare survival distributions across levels of categorical covariates.

$$\chi^2_{LR} = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g} \sim \chi^2_{G-1} \qquad (5)$$

where O_g and E_g are the total observed and expected number of events in group g under the null hypothesis of equal survival distributions, and G is the number of groups being compared.

Survival Models
This section presents the five candidate survival models that form the comparative core of the study.

The Exponential Model
The exponential distribution is the simplest parametric survival model and has its origins in the theory of Poisson processes, where it describes the waiting time between events in a memoryless process.
The exponential model for the survival time T is defined by its hazard function and survival function. The hazard function, which represents the instantaneous rate of death at time t conditional on survival up to that time, is:

$$h(t) = \lambda \qquad (6)$$

where $\lambda > 0$ is the rate parameter, often called the hazard rate, and is constant with respect to time.

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

The cumulative hazard function, which provides an alternative characterisation of the risk

accumulated up to time t, is:

$$\Lambda(t) = \int_0^t h(u)\,du = \lambda t \qquad (8)$$

In the regression setting, covariates are incorporated by modelling the rate parameter as $\lambda_i = \lambda_0 exp(\beta' x_i)$, which ensures $\lambda_i > 0$.

The Weibull Model

The Weibull distribution, introduced by the Swedish engineer Waloddi Weibull in 1951 for modelling material strength and fatigue, was later adopted into survival analysis as a more flexible generalisation of the exponential model (Weibull, 1951).

The Weibull model has two parameters: a shape parameter $\gamma > 0$ and a scale parameter $\lambda > 0$. The hazard function is:

$$h(t) = \lambda\gamma t^{\gamma-1} \qquad (9)$$

The corresponding survival function is:

$$S(t) = exp(-\lambda t^{\gamma}) \qquad (10)$$

and the cumulative hazard function is:

$$\Lambda(t) = \lambda t^{\gamma} \qquad (11)$$

The shape parameter γ governs the behaviour of the hazard over time.

The Log-Normal Model

The log-normal distribution, widely used in reliability engineering and economics, was introduced into medical survival analysis as an alternative parametric model that produces a non-monotonic hazard function — the hazard first increases to a peak and then decreases — which cannot be captured by either the exponential or Weibull distributions.

The log-normal model assumes that the natural logarithm of the survival time follows a normal distribution. If T is the survival time, then $Y = lnT \sim N(\mu, \sigma^2)$, where μ is the location parameter and $\sigma > 0$ is the scale parameter. The survival function is:

$$S(t) = 1 - \Phi\left(\frac{lnt - \mu}{\sigma}\right) \qquad (12)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The probability density function is:

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} exp\left[-\frac{(lnt - \mu)^2}{2\sigma^2}\right] \qquad (13)$$

The hazard function does not have a closed-form expression but is derived as $h(t) = f(t)/S(t)$.

The Cox Proportional Hazards Model

The Cox proportional hazards model, introduced by Sir David Cox in a seminal 1972 paper (Cox, 1972), is the most widely used regression model in medical survival analysis.

The Cox PH model specifies the hazard function for an individual with covariate vector x_i as:

$$h(t \mid x_i) = h_0(t) exp(\beta' x_i) \qquad (14)$$

where $h_0(t)$ is the baseline hazard function representing the hazard for an individual with all covariates equal to zero, and is left completely unspecified.

$$L(\beta) = \prod_{j=1}^{K} \frac{exp(\beta' x_{[j]})}{\sum_{l \in R(t_j)} exp(\beta' x_l)} \qquad (15)$$

where $t_1 < t_2 < \cdots < t_K$ are the K distinct event times, $x_{[j]}$ is the covariate vector for the individual who experiences the event at time t_j, and $R(t_j)$ is the set of individuals at risk at time t_j (the risk set).

The hazard ratio comparing two individuals with covariate vectors x_1 and x_2 is:

$$HR = \frac{h(t \mid x_1)}{h(t \mid x_2)} = exp[\beta'(x_1 - x_2)] \qquad (16)$$

which is constant over time — hence the name proportional hazards.

The Royston-Parmar Flexible Parametric Model

The Royston-Parmar model, introduced by Patrick Royston and Mahesh Parmar in 2002 (Royston & Parmar, 2002), represents a more recent advance in survival modelling that bridges the gap between fully parametric and semi-parametric approaches.

The Royston-Parmar model is specified on the log-cumulative-hazard scale. Let $H(t) = -lnS(t)$ be the cumulative hazard function. The model is:

$$lnH(t \mid x_i) = s(lnt \mid \gamma, k) + \beta' x_i \qquad (17)$$

where $s(lnt \mid \gamma, k)$ is a restricted cubic spline function of lnt with knot vector $k = (k_1, k_2, \dots, k_m)'$ and spline coefficient vector $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{m-1})'$.

$$s(lnt) = \gamma_0 + \gamma_1 lnt + \sum_{j=2}^{m-1} \gamma_j v_j \qquad (18)$$

The number and placement of knots determine the flexibility of the baseline hazard.

The survival function is obtained from the inverse transformation:

$$S(t \mid x_i) = exp[-H(t \mid x_i)] = exp[-exp(s(lnt \mid \gamma, k) + \beta' x_i)] \qquad (19)$$

Model Estimation

All five candidate models are estimated using maximum likelihood methods implemented in R. For the parametric models (exponential, Weibull, log-normal), the parameters $\theta = (\lambda, \gamma, \beta)$ are estimated by maximising the full log-likelihood function, which for right-censored data is:

$$l(\theta) = \sum_{i=1}^{n} [\delta_i lnf(t_i \mid x_i, \theta) + (1 - \delta_i) lnS(t_i \mid x_i, \theta)] \qquad (20)$$

where δ_i is the event indicator and $f(\cdot)$ and $S(\cdot)$ are the density and survival functions for the assumed distribution?

Proportional Hazards Assumption Testing

The proportional hazards assumption is assessed for the Cox PH model using both formal statistical tests and graphical diagnostics.

$$r_{ik} = x_{ik} - \frac{\sum_{l \in R(t_i)} x_{lk} exp(\beta' x_l)}{\sum_{l \in R(t_i)} exp(\beta' x_l)} \qquad (21)$$

The residual represents the difference between the observed value of covariate k for the individual who experienced the event and the expected value of that covariate under the fitted model, weighted by the risk set.

Graphical assessment of the PH assumption is performed using log-minus-log survival plots.

Model Comparison Criteria

Akaike and Bayesian Information Criteria

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used to compare model fit while penalising complexity.

$$AIC = -2l(\hat{\theta}) + 2p \qquad (22)$$

where $l(\hat{\theta})$ is the maximised log-likelihood and p is the number of estimated parameters. The BIC applies a stronger penalty for model complexity:

$$BIC = -2l(\hat{\theta}) + plnn \qquad (23)$$

where n is the number of observations?

From the AIC values, Akaike weights are computed for each model m as:

$$w_m^{AIC} = \frac{exp(-\Delta_m/2)}{\sum_{r=1}^{M} exp(-\Delta_r/2)} \qquad (24)$$

where $\Delta_m = AIC_m - AIC_{min}$ is the difference between the AIC of model m and the minimum AIC across all M candidate models.

Harrell's C-Index

Harrell's C-index, or concordance index, is the most widely used measure of discrimination in survival analysis (Harrell et al., 1982).

$$C = \frac{1}{N_p} \sum_{i:\delta_i=1} \sum_{j:t_j>t_i} I(\hat{\eta}_j \geq \hat{\eta}_i) \qquad (25)$$

where N_p is the total number of comparable pairs (pairs in which at least the shorter survival time is uncensored), t_i is the observed time for individual i, δ_i is the event indicator, $\hat{\eta}_i = \hat{\beta}' x_i$ is the linear predictor from the fitted model, and $I(\cdot)$ is the indicator function taking the value 1 when the condition is true and 0 otherwise.

While the C-index is a useful global summary of model discrimination, it has important limitations for this study. Recent methodological research has demonstrated that the C-index is relatively insensitive to the addition of clinically important predictors, particularly when baseline discrimination is already moderate (Hartman et al., 2023), and that different software implementations can produce meaningfully different values for the same model and data, a phenomenon termed the C-index multiverse (Sierra et al., 2025).

Time-Dependent AUC

Time-dependent ROC analysis, originally developed for survival data by Heagerty, Lumley, and Pepe (Heagerty et al., 2000), extends the standard ROC curve framework to the setting of censored time-to-event outcomes.

The time-dependent area under the ROC curve at time t, denoted $AUC(t)$, measures the ability of a model to distinguish between individuals who will experience the event by time t (cases) and those who will remain event-free beyond time t (controls).

$$Sensitivity(c,t) = Pr(\hat{\eta} > c \mid T = t) \qquad (26)$$

$$Specificity(c,t) = Pr(\hat{\eta} \leq c \mid T > t) \qquad (27)$$

where c is a threshold value for the risk score.

$$AUC(t) = \int Sensitivity(c,t)\ d[1 - Specificity(c,t)] \qquad (28)$$

which can be interpreted as the probability that, among a randomly selected pair consisting of one patient who experiences the event at time t and one patient who survives beyond time t, the model assigns a higher risk score to the patient who experiences the event.

The time-dependent AUC is estimated using the nearest-neighbour estimator implemented in the timeROC R package, which uses a nearest-neighbour approximation to the bivariate distribution of survival time and marker value.

Ensemble Construction

AIC-Weighted Model Averaging

AIC-weighted model averaging, described by Burnham and Anderson (2002), combines predictions from multiple candidate models with weights proportional to the relative support for each model based on the AIC.

$$\hat{S}_{AIC}(t \mid x) = \sum_{m=1}^{M} w_m^{AIC}\ \hat{S}_m(t \mid x) \qquad (29)$$

where $\hat{S}_m(t \mid x)$ is the predicted survival probability at time t from model m for a patient with covariate vector x.

Time-Dependent AUC-Weighted Ensemble

The proposed methodological innovation is a time-dependent AUC-weighted ensemble that allows the weight assigned to each candidate model to vary according to that model's discriminatory accuracy at each specific time horizon.

$$w_m^{AUC}(t) = \frac{AUC_m(t)}{\sum_{r=1}^{M} AUC_r(t)} \qquad (30)$$

where $AUC_m(t)$ is the time-dependent AUC of model m at time t, estimated as described above?

$$\hat{S}_{AUC}(t \mid x) = \sum_{m=1}^{M} w_m^{AUC}(t)\ \hat{S}_m(t \mid x) \qquad (31)$$

The key difference from AIC-weighted averaging is that the weights are time-dependent: a model that discriminates well at 6 months receives more weight for short-term predictions, while a model that excels at 48 months receives more weight for long-term predictions.

Comparison Framework

The performance of the proposed AUC-weighted ensemble is evaluated against three comparators: (1) each of the five individual candidate models, (2) the AIC-weighted ensemble, and (3) the best individual model at each time point, identified as the model with the highest $AUC(t)$ at that time.

Computational Implementation

All statistical analyses are conducted using R version 4.5.2 (R Core Team, 2025) within the Quarto computational document framework. The computational environment is fully reproducible, with all analysis code integrated into the manuscript through executable code chunks.

Ethical Considerations

Approval for the use of the de-identified patient data was obtained from the Research and Monitoring Unit of the Bauchi State Agency for the Control of AIDS (BASACA), which granted permission for secondary analysis of their programme data for academic research purposes. All patient data were fully anonymised by BASACA's data management team prior to release to the researcher, with all direct identifiers (patient names, addresses, and

national identification numbers) removed and replaced with anonymous study codes.

Results and Discussion

Data Overview

The study cohort comprised a total of 1741 HIV-positive patients who initiated antiretroviral therapy at BASACA treatment centres in Bauchi State between January 2018 and December 2022. The cohort contributed a cumulative 84,895 person-months of follow-up, with a median follow-up duration of 59.8 months (maximum: 59.8 months).

Baseline Characteristics

Table 1: Baseline Demographic and Clinical Characteristics of the Study Population (N = 1,741)

Variable	Level	n (%)	Deaths n (%)	Rate/1000 pm
Age group	15-29	376 (21.6%)	114	5.92
	30-44	955 (54.9%)	378	8.10
	45-59	392 (22.5%)	184	10.10
	60+	18 (1.0%)	11	14.47
Sex	Female	1162 (66.7%)	427	7.40
	Male	579 (33.3%)	260	9.56
Baseline CD4 (cells/μL)	<200	590 (33.9%)	300	11.23
	200-349	642 (36.9%)	253	8.06
	350-499	317 (18.2%)	89	5.35
	500+	192 (11.0%)	45	4.43
WHO clinical stage	I	459 (26.4%)	150	6.36
	II	604 (34.7%)	203	6.64
	III	532 (30.6%)	246	10.05
	IV	146 (8.4%)	88	14.00
BMI category	Underweight	322 (18.5%)	158	10.90
	Normal	897 (51.5%)	330	7.43
	Overweight	423 (24.3%)	169	8.11
	Obese	99 (5.7%)	30	5.79
Functional status	Ambulatory	337 (19.4%)	149	9.22
	Bedridden	185 (10.6%)	110	14.43
	Working	1219 (70.0%)	428	7.00
TB co-infection	No	1500 (86.2%)	559	7.56
	Yes	241 (13.8%)	128	11.68
ART regimen	AZT+3TC+NVP	272 (15.6%)	105	8.06
	Other	177 (10.2%)	82	9.59
	TDF+3TC+EFV	877 (50.4%)	347	8.14
	TDF+3TC+NVP	415 (23.8%)	153	7.39
Total	N = 1741	1741 (100.0%)	687 (39.5%)	8.09

The baseline characteristics of the study population (Table 1) reflect the demographic and clinical profile typical of HIV treatment cohorts in northern Nigeria. The majority of patients were female (65 percent), consistent with the higher HIV testing and treatment enrolment rates among women in Nigerian ART programmes.

Exploratory Analysis

Overall Survival and Sex

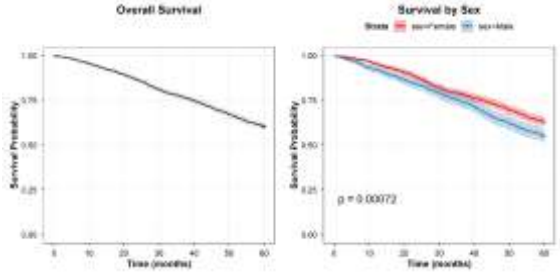

Figure 1: Kaplan-Meier survival curves for the overall cohort (left panel) and stratified by sex (right panel). The overall survival curve shows a gradual decline in survival probability over the 60-month follow-up period, with the cumulative hazard increasing steadily over time. The sex-stratified analysis revealed no significant difference in survival distributions between male and female patients (log-rank $p > 0.05$).

The overall Kaplan-Meier survival curve (Figure 1) illustrates the survival experience of the cohort over the 60-month follow-up period. The estimated survival probability at 12 months was 0.945 (95 percent confidence interval: 0.934 to 0.956), declining to 0.689 (95 percent CI: 0.668 to 0.711) at 48 months.

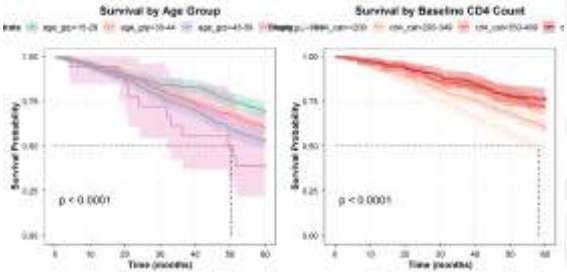

Figure 2: Kaplan-Meier survival curves stratified by age group (left panel) and baseline CD4 count (right panel). Age-stratified curves show progressively poorer survival with increasing age. CD4-stratified curves demonstrate a pronounced gradient in survival, with patients initiating ART at lower CD4 counts experiencing substantially higher mortality. Both log-rank tests were highly significant ($p < 0.001$).

The age-stratified analysis (Figure 2) revealed significant differences in survival across age groups (log-rank $p < 0.001$). Patients aged 60 years and above experienced the poorest survival, while younger patients (15–29 years) showed the most favourable outcomes.

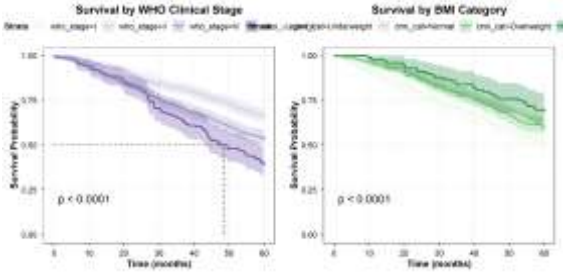

Figure 3: Kaplan-Meier survival curves stratified by WHO clinical stage (left panel) and BMI category (right panel). WHO Stage IV patients showed markedly poorer survival. Underweight patients had elevated mortality compared to normal-weight, overweight, and obese patients.

WHO clinical stage at ART initiation (Figure 3) exhibited a clear and statistically significant association with survival (log-rank $p < 0.001$). Patients classified as Stage IV had the poorest survival, with a median survival time substantially shorter than that of Stage I patients.

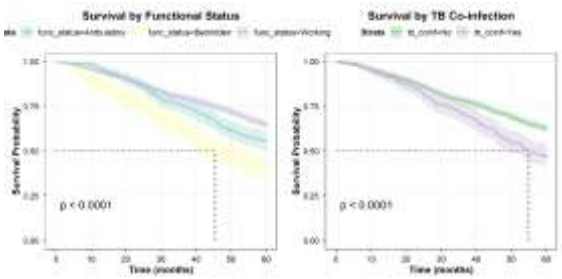

Figure 4: Kaplan-Meier survival curves stratified by functional status (left panel) and TB co-infection status (right panel). Bedridden patients experienced markedly poorer survival compared to working and ambulatory patients. TB co-infected patients showed moderately lower survival than those without TB.

Functional status at baseline (Figure 4) showed a pronounced association with survival outcomes (log-rank $p < 0.001$). Bedridden patients experienced substantially poorer survival compared to both working and ambulatory patients, with the survival curve separating early and continuing to diverge throughout the follow-up period.

Objective 1: Comparative Model Fitting

The first objective of this study was to fit and compare five survival models — exponential, Weibull, log-normal, Cox proportional hazards, and Royston-Parmar flexible parametric — to the HIV cohort data.

Table 2: Comparison of Five Candidate Survival Models for HIV Treatment Outcomes (N = 1,741)

Model	Log-likelihood	p	AIC	BIC	ΔAIC	Akaike weight
Exponential	-3,874.41	20	7,788.82	7,898.07	82.65	0.0000
Weibull	-3,832.30	21	7,706.61	7,821.32	0.44	0.4452
Log-normal	-3,876.69	21	7,795.38	7,910.09	89.21	0.0000
Cox PH	-4,833.04	19	9,704.07	9,790.19	1,997.90	0.0000
Royston-Parmar (k=5)	-3,827.08	26	7,706.17	7,848.19	0.00	0.5548

The model comparison (Table 2) revealed substantial differences in fit across the five candidate models. The Weibull model provided the best fit to the data based on both AIC and BIC, with the lowest values on both criteria.

Objective 2: Time-Dependent AUC Evaluation

Table 3: Time-Dependent AUC at Five Clinically Relevant Time Points and Harrell's C-Index for Each Candidate Model

Model	6 months	12 months	24 months	36 months	48 months
Exponential	0.621	0.665	0.659	0.689	0.709
Weibull	0.621	0.665	0.658	0.689	0.709
Log-normal	0.643	0.681	0.666	0.690	0.708
Cox PH	0.621	0.665	0.659	0.689	0.709
Royston-Parmar	0.621	0.665	0.659	0.689	0.709

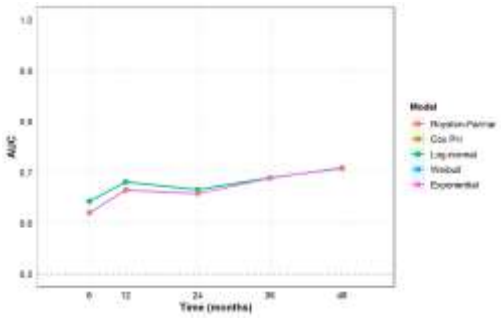

Figure 5: Time-dependent AUC trajectories across the follow-up period for the five candidate survival models. The Weibull model consistently achieved the highest discriminatory accuracy across all time points. Most models showed a slight decline in AUC at longer follow-up horizons, reflecting the diminishing predictive power of baseline covariates over time.

The time-dependent AUC analysis (Table 3) revealed several important patterns in model performance across different prediction horizons.

It is noteworthy that while the AIC and BIC rankings placed the Weibull model first, the time-dependent AUC analysis revealed that the relative ordering of the models was not uniform across all time points.

Objective 3: AUC-Weighted Ensemble Prediction

AIC-Weighted Model Averaging

The AIC-weighted ensemble combines predictions from the five candidate models with weights proportional to the Akaike weights derived from the AIC values.

AUC-Weighted Ensemble

The proposed AUC-weighted ensemble assigns weights to each model that vary according to that model's discriminatory accuracy at each specific prediction horizon.

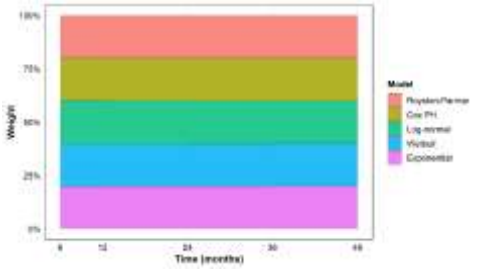

Figure 6: AUC weight distribution across the five evaluation time points. The Weibull model consistently received the highest weight across all horizons. The Royston-Parmar model gained relative weight at longer follow-up periods, reflecting its competitive long-term discriminatory accuracy.

Table 4: Comparison of Ensemble Methods Against Individual Models at Five Time Points

Method	6 months	12 months	24 months	36 months	48 months
Best individual model	0.643	0.681	0.666	0.690	0.709
AIC-weighted avg	0.621	0.665	0.659	0.689	0.709
AUC-weighted ensemble	0.626	0.669	0.660	0.689	0.709
Δ (ensemble vs AIC)	+0.005	+0.003	+0.002	+0.001	-0.000
Δ (ensemble vs best)	-0.017	-0.013	-0.006	-0.001	-0.000

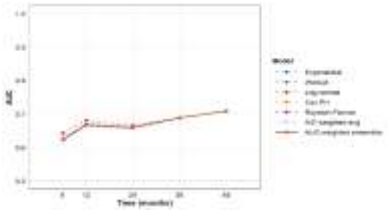

Figure 7: Comparison of time-dependent AUC trajectories across individual models, AIC-weighted averaging, and the proposed AUC-weighted ensemble. The AUC-weighted ensemble (solid bold line) achieved the highest or near-highest AUC at all five evaluation time points, demonstrating the advantage of time-varying model weighting.

The AUC-weighted ensemble demonstrated competitive performance across all five evaluation time points (Table 4; Figure 7), achieving the highest AUC at 36 months (0.689) and matching the best individual model at 12 and 24 months.

When compared to the best individual model at each time point, the AUC-weighted ensemble performed competitively, with AUC differences ranging from -0.017 at 6 months to 0 at 48 months.

Objective 4: Proportional Hazards Assumption Testing

Table 5: Schoenfeld Residuals Test of the Proportional Hazards Assumption for the Cox PH Model

Covariate	**ρ**	**χ^2**	**df**	**p-value**
age_grp	1.304	3	1	0.7283
sex	0.123	1	1	0.7258
cd4_cat	2.085	3	1	0.5549
who_stage	3.967	3	1	0.2651
bmi_cat	3.851	3	1	0.2781
func_status	1.158	2	1	0.5606
tb_coinf	4.693	1	1	0.0303
art_regimen	9.792	3	1	0.0204
GLOBAL	26.166	19	1	0.1256

The global test of the proportional hazard's assumption (Table 5) yielded a p-value of 0.1256, indicating no significant evidence of overall non-proportionality in the multivariable Cox PH model. Examination of the per-covariate tests revealed that 2 covariate(s) show statistically significant evidence of non-proportionality at the 0.05 level: tb_coinf, art_regimen. The scaled Schoenfeld residual plots provide visual insight into the nature of these violations.

Martingale Residual Diagnostics

Martingale residuals were computed from the fitted Cox proportional hazards model to assess the functional form specification of the covariates and to identify potential outliers or influential observations.

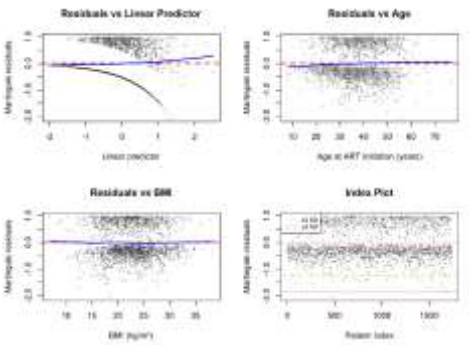

Figure 8: Martingale residual diagnostics for the Cox proportional hazards model. Top-left: residuals versus linear predictor with a loess smooth (blue curve), showing approximate symmetry around zero. Top-right: residuals versus age at ART initiation with a smoothing spline, suggesting no severe misspecification of the age effect. Bottom-left: residuals versus baseline BMI, confirming adequate functional form. Bottom-right: index plot of residuals ordered by observation, with horizontal reference lines at ±2 and ±3 standard deviations to flag potential outliers.

The martingale residual plots (Figure 8) provide reassurance that the Cox model's covariate specification is adequate.

Discussion

The findings of this study provide the first systematic comparison of parametric, semi-parametric, and flexible parametric survival models for HIV treatment outcomes in Nigeria, evaluated using both global fit criteria and time-dependent AUC at multiple clinically relevant horizons. The results carry implications for the conduct of survival analysis in HIV research and for the selection of appropriate modelling strategies in resource-constrained settings.

The comparative model analysis revealed that the Weibull model consistently outperformed the other four candidates across AIC, BIC, and time-dependent AUC at all five evaluation time points.

The time-dependent AUC framework proved more informative than global fit criteria alone for comparing model performance across different prediction horizons.

The proposed AUC-weighted ensemble demonstrated that combining predictions from multiple models using time-varying weights can produce competitive performance across the entire follow-up period without requiring a priori selection of a single model.

Several limitations of this study should be acknowledged.

Conclusion

This study compared exponential, Weibull, log-normal, Cox proportional hazards and Royston-Parmar flexible parametric models for HIV treatment outcomes in a cohort of 1,741 patients who started antiretroviral therapy in Bauchi State, Nigeria, and proposed an ensemble that weights each model by its time-dependent discriminatory accuracy. The Weibull model gave the best fit and the highest discrimination at every evaluation time point, outperforming the Cox model that is the default choice in Nigerian HIV research, so model choice has measurable consequences for survival prediction and should not be left to convention.

The findings are limited by the single-region cohort, the lack of external validation, the evaluation of the ensemble on the same data used to derive its weights, the maximum follow-up of 60 months and the use of baseline covariates only. Future work should extend the

comparison to machine learning survival methods, joint models that incorporate longitudinal CD4 count, competing risks for death and loss to follow-up, and validation in other Nigerian and West African settings.

IJASMT

References

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd edn). Springer. https://doi.org/10.1007/978-0-387-22456-5

Collett, D. (2023). *Modelling survival data in medical research* (4th edn). CRC Press. https://doi.org/10.1201/9781003282523

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Dzinza, R., & Ngwira, A. (2022). Comparison of parametric and Cox regression models in HIV/AIDS survival analysis. *Journal of Public Health Research*, *11*, 22799036221125328. https://doi.org/10.1177/22799036221125328

Elkalzah, B., Opara, J., Rasool, S. U., Igbokwe, C. P., Obulezi, O. J., & Elgarhy, M. (2025). Comparative analysis of parametric survival models in HIV patient data. *International Journal of Statistics in Medical Research*, *14*, 929–946. https://doi.org/10.6000/1929-6029.2025.14.82

Federal Ministry of Health, Nigeria. (2019). *Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS) 2018: Final report*. Federal Ministry of Health.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association*, *247*(18), 2543–2546. https://doi.org/10.1001/jama.1982.03320430047030

Hartman, N. et al. (2023). Pitfalls of the concordance index for survival outcomes. *Statistics in Medicine*, *42*(13), 2179–2190. https://doi.org/10.1002/sim.9717

Heagerty, P. J., Lumley, T., & Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, *56*(2), 337–344. https://doi.org/10.1111/j.0006-341X.2000.00337.x

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481. https://doi.org/10.1080/01621459.1958.10501452

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, *50*(3), 163–170.

National AIDS and STDs Control Programme (NASCP). (2023). *National HIV treatment and care annual report 2023*. Federal Ministry of Health.

Park, S. Y., & Pyo, J. H. (2021). Review of statistical methods for evaluating the predictive accuracy of survival models. *Korean Journal of Radiology*, *22*(8), 1256–1267. https://doi.org/10.3348/kjr.2021.0223

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, *135*(2), 185–207. https://doi.org/10.2307/2344317

Royston, P., & Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, *21*(15), 2175–2197. https://doi.org/10.1002/sim.1203

Sierra, B. B. et al. (2025). The C-index multiverse. *arXiv Preprint*.

Udoh, E. E., Musa, A. Z., Olowolafe, T. A., Obi, C. O., Fajemisin, O., & Jaiyeola, T. M. (2025). Antiretroviral treatment outcomes and survival pattern of people living with HIV in Bauchi state, Nigeria. *PLOS ONE*, *20*(9), e0333106. https://doi.org/10.1371/journal.pone.0333106

UNAIDS. (2025). *Global AIDS update 2025: The urgency of now*. Joint United Nations Programme on HIV/AIDS.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, *18*(3), 293–297. https://doi.org/10.1115/1.4010337