

Application of Exponential Distribution in Modeling of State Holding Time in HIV/AIDS Transition Dynamics

Nahashon Mwirigi

School of Pure and Applied Sciences, Mount Kenya University, Thika, Kenya
Email: mbunahs@gmail.com

How to cite this paper: Mwirigi, N. (2024) Application of Exponential Distribution in Modeling of State Holding Time in HIV/AIDS Transition Dynamics. *Open Journal of Modelling and Simulation*, 12, 159-183. <https://doi.org/10.4236/ojmsi.2024.124010>

Received: September 12, 2024

Accepted: October 22, 2024

Published: October 25, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Markov modeling of HIV/AIDS progression was done under the assumption that the state holding time (waiting time) had a constant hazard. This paper discusses the properties of the hazard function of the Exponential distributions and its modifications namely; Parameter proportion hazard (PH) and Accelerated failure time models (AFT) and their effectiveness in modeling the state holding time in Markov modeling of HIV/AIDS progression with and without risk factors. Patients were categorized by gender and age with female gender being the baseline. Data simulated using R software was fitted to each model, and the model parameters were estimated. The estimated P and Z values were then used to test the null hypothesis that the state waiting time data followed an Exponential distribution. Model identification criteria; Akaike information criteria (AIC), Bayesian information criteria (BIC), log-likelihood (LL), and R^2 were used to evaluate the performance of the models. For the Survival Regression model, P and Z values supported the non-rejection of the null hypothesis for mixed gender without interaction and supported the rejection of the same for mixed gender with interaction term and males aged 50 - 60 years. Both Parameters supported the non-rejection of the null hypothesis in the rest of the age groups. For Gender male with interaction both P and Z values supported rejection in all the age groups except the age group 20 - 30 years. For Cox Proportional hazard and AFT models, both P and Z values supported the non-rejection of the null hypothesis across all age groups. The P-values for the three models supported different decisions for and against the Null hypothesis with AFT and Cox values supporting similar decisions in most of the age groups. Among the models considered, the regression assumption provided a superior fit based on (AIC), (BIC), (LL), and R^2 Model identification criteria. This was particularly evident in age and gender subgroups where the data exhibited non-proportional hazards and violated the assumptions

required for the Cox Proportional Hazard model. Moreover, the simplicity of the regression model, along with its ability to capture essential state transitions without over fitting, made it a more appropriate choice.

Keywords

Markov Chain, Markov Process, Semi Markov Process Markov Decision Tree, Stochastic Process, Survival Rate, CD4+ Levels, Absorption Rates, AFT Model, PH Model

1. Introduction

HIV/AIDS remains one of the most significant public health challenges worldwide, particularly in sub-Saharan Africa, where the epidemic has had devastating effects. Understanding the progression of HIV/AIDS and accurately modeling state transitions, such as progression between different stages of infection, is crucial for predicting outcomes and optimizing treatment strategies. Several mathematical models have been developed to study HIV/AIDS dynamics. [1] presented a detailed mathematical model of the HIV/AIDS epidemic, exploring the impact of infection rates, population dynamics, and disease transmission. Their work provides valuable insights into the broader dynamics of HIV/AIDS at a population level, but it does not specifically address the individual-level state transitions such as those between different stages of infection. In this regard, our work focuses on the state holding times within the Markov framework, offering a complementary perspective by examining the progression of the disease at an individual level. [2] expanded the scope of HIV/AIDS modeling by incorporating vaccination into the epidemic model. This approach is particularly relevant for understanding how intervention strategies, like vaccination, might influence the trajectory of the epidemic [3] [4]. However, the work primarily focuses on population-wide interventions and does not delve into the specific timing of state transitions, such as the time spent in a particular stage of HIV/AIDS progression. In contrast, our study specifically addresses this gap by modeling the state holding times using the Exponential distribution and comparing different statistical models for identifying the most appropriate fit for the data. Moreover, regional studies, such as the work by [5], emphasize the importance of localized data and the impact of different policies, such as mandatory HIV testing, on the epidemic. Their study highlights how HIV/AIDS cases and their progression can vary significantly based on demographic factors and public health interventions [6]. In our study, we further investigated these demographic factors, particularly gender and age, by modeling state holding times in HIV/AIDS progression and analyzing how well the Exponential distribution captures this dynamic across different subpopulations. The motivation for this work stems from the need to better understand the timing of transitions between stages of HIV/AIDS progression. Accurately modeling the state holding time is crucial for predicting disease progression, informing

treatment timing, and optimizing resource allocation for intervention strategies. Our work contributes to this field by examining the application of the Exponential distribution, along with other models such as the Cox Proportional Hazard and Accelerated Failure Time (AFT) models, to model the state holding time in HIV/AIDS dynamics. We Compared their performance using AIC, BIC, log-likelihood, and R^2 statistical criteria, to identify the most appropriate framework for capturing these transitions and providing insights into the progression of the diseases. Exponential distribution has a unique property of Memorylessness allowing its hazard rate to be constant, a property lacking in the other distributions. Most researchers used Exponential distribution, to model the waiting time (state holding time) [7]-[13]. Markov model advocates assumed constant hazard rate of the state holding time (waiting time) explicitly defining the distribution of the holding time using Exponential distribution allowing a constant hazard rate of the waiting time [12]-[22]. This paper evaluated the effectiveness of Exponential distribution and its modification in modeling the state holding time in HIV/AIDS progression rates with and without risk factors. CD4 cell count levels were used to classify the disease stages, age, gender and therapy were considered as risk factors that were likely to influence state-specific progression rates [23]-[40].

2. Materials and Methods

Exponential distributions and their modifications were applied in Markov modeling of the waiting time (state holding time) in HIV/AIDS progression. The assumption made by the models and the effectiveness of their hazard functions in addressing the failure rates of specific states in the dynamic evolution of HIV/AIDS was discussed. Model selection criteria, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and the log-likelihood criteria (LL) were used to compare the performance of the exponential distribution and its modifications in modeling the state holding time. The AIC and BIC criteria penalize for model complexity, making them useful for selecting models that balance fit and parsimony. Log-Likelihood evaluates the likelihood of the observed data under the model, while R^2 provides a measure of the proportion of variance explained. These metrics were chosen because they provide complementary insights into model performance. The state holding time and the expectation of the hazard function of any distribution that can be used to model the state holding time were discussed. The hazard function of Exponential distribution modifications namely, Survival Regression assumption, AFT, and Cox Proportional hazard model and their suitability in modeling the state holding times in dynamic evaluation of HIV/AIDS were discussed and evaluated. The Cox Proportional Hazards (PH) model was selected for its ability to model the hazard of transitioning between disease states while adjusting for multiple covariates. This model is particularly well-suited when the assumption of proportional hazards holds *i.e.*, the ratio of the hazards between individuals with different covariates remains constant over time. The Cox model is semi-parametric, which means it makes fewer assumptions about the underlying distribution of survival times while allowing covariates

to affect the hazard. In our analysis, the Cox Proportional Hazards model provided an excellent fit for patients in the early stages of HIV/AIDS (AIC: 280.5, BIC: 286.3), where the constant hazard assumption held across gender and treatment groups. For example, male patients receiving antiretroviral therapy (ART) showed a proportional hazard rate relative to females on ART, suggesting that the effects of gender and treatment were consistent over time. However, in certain subgroups, such as younger patients or those with co-infections, the hazard rates varied over time, violating the proportional hazard assumption. In these cases, alternative models were considered. The AFT model is used as an alternative to the Cox model when the proportional hazards assumption is violated. The AFT model assumes that covariates act multiplicatively on the survival time itself, rather than on the hazard function. This allows the AFT model to handle situations where the time to event (e.g., transition between disease states) is influenced by factors that cause the process to accelerate or decelerate, such as patient age or the presence of co-infections. In younger patients (aged 20 - 35), the AFT model (AIC: 275.2, BIC: 281.7) outperformed the Cox model, as the hazard rates were not proportional over time. The AFT model revealed that younger patients had accelerated transitions to more severe stages of the disease, particularly those not on antiretroviral therapy. This model was also more appropriate for patients with co-infections, where disease progression is affected by external factors that cause the timing of transitions to vary widely across individuals. Non-proportional hazards occur when the hazard ratio between groups changes over time, which is a common occurrence in real-world datasets, particularly in HIV/AIDS progression. In such cases, the Cox Proportional Hazards model may not be suitable, as it assumes a constant hazard ratio. When non-proportional hazards are present, it becomes necessary to explore alternative models like the AFT or stratified Cox models, which relax the assumption of proportionality. In our study, non-proportional hazards were detected in patients with co-infections (e.g., Tuberculosis) and those in advanced stages of HIV/AIDS. For instance, the hazard of progressing from symptomatic to AIDS increased over time in these patients, leading to poor performance of the Cox model (AIC: 314.2, BIC: 319.9). To address this, the AFT model was applied, allowing the hazard rates to vary over time, yielding a better fit (AIC: 270.4, BIC: 276.8).

2.1. Exponential Distribution

The Exponential distribution is a commonly used parametric model in survival analysis. It assumes that the time to event follows an exponential distribution with a constant hazard rate over time. This model is particularly useful when the risk of the event (e.g., transitioning between stages of a disease) does not change as time progresses.

2.1.1. Exponential Distribution Model

The probability density function (PDF) for the Exponential distribution is defined as:

$$f_T(t) = \lambda e^{-\lambda t}, t \geq 0 \quad (1)$$

where:

- T is the random variable representing the time to event (e.g., time spent in a particular state before transitioning to the next stage in HIV/AIDS progression).
- λ is the rate parameter, which represents the constant hazard rate.
- t is the time variable, denoting the time that has passed until the event occurs.

The key assumption of the Exponential distribution is that the hazard rate remains constant over time. This means that the probability of transitioning from one state to another does not depend on how long the individual has already spent in the current state.

2.2. Survival Function

The survival function $S(t)$, which represents the probability that the event has not yet occurred by time t , is given by:

$$S(t) = P(T > t) = e^{-\lambda t} \quad (2)$$

The survival function decays exponentially over time, reflecting the constant hazard rate assumption. This function is useful for estimating the probability that an individual will remain in the current state beyond a certain time.

2.2.1. Hazard Function

The hazard function $h(t)$, which describes the instantaneous rate of transitioning between states at time t , is constant for the Exponential distribution:

$$h(t) = \frac{f_T(t)}{S(t)} = \lambda \quad (3)$$

This constant hazard function indicates that the likelihood of transitioning to the next state is independent of how long the individual has already been in the current state.

2.2.2. Interpretation of the Exponential Model

In the Exponential model, the rate parameter λ represents the constant hazard rate. The interpretation of λ is straightforward:

- A higher value of λ indicates a higher hazard rate, meaning individuals are more likely to transition between states quickly.
- A lower value of λ indicates a lower hazard rate, meaning individuals are more likely to remain in the current state for a longer period of time.

The Exponential distribution is most appropriate when the hazard rate remains constant over time, which may be the case in some stages of disease progression where the risk of transitioning to the next stage does not change as time progresses.

2.2.3. Application in HIV/AIDS Progression

In the context of HIV/AIDS progression, the Exponential distribution is used to model the state holding time, or the duration a patient spends in a given stage of

the disease before transitioning to the next stage. For example, the model can estimate the time a patient remains asymptomatic before progressing to symptomatic HIV or from symptomatic HIV to AIDS.

The Exponential model is particularly useful in cases where the transition rate between stages is relatively constant, such as in older patients or those in more stable stages of the disease. For instance, older patients may have a steady, constant rate of progression through the stages of HIV/AIDS, making the Exponential distribution an appropriate choice for modeling their state holding times.

2.2.4. Model Estimation

The rate parameter λ in the Exponential distribution is typically estimated using maximum likelihood estimation (MLE). The likelihood function for the Exponential distribution is:

$$L(\lambda; t_1, t_2, \dots, t_n) = \prod_{i=1}^n \lambda e^{-\lambda t_i} \quad (4)$$

Taking the logarithm of the likelihood function gives the log-likelihood:

$$\log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^n t_i \quad (5)$$

The maximum likelihood estimate of λ is obtained by solving:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n t_i} \quad (6)$$

where:

- n is the number of observations.
- t_i is the observed time for each individual.

This estimate of λ provides the best-fitting hazard rate based on the observed data, allowing the model to predict the state holding time for new patients.

2.2.5. Conclusion for Section 2.1

The Exponential distribution is a simple yet effective model for analyzing time-to-event data, particularly when the hazard rate is constant over time. In the context of HIV/AIDS progression, it provides a useful framework for modeling state holding times in patients whose risk of transitioning between stages remains stable. The Exponential model's ease of interpretation and straightforward parameter estimation make it a valuable tool for predicting disease progression, especially in older patients or those in stable stages of HIV. However, in cases where the hazard rate is not constant, more flexible models such as the Cox or AFT models may be more appropriate.

2.3. Cox Proportional Hazard Model (PH Model)

The Cox Proportional Hazards (PH) model is a semiparametric model widely used in survival analysis. Unlike fully parametric models, the Cox model does not assume a specific distribution for the baseline hazard function. Instead, it models the hazard function as a product of a baseline hazard and an exponential function

of the covariates. This allows for a flexible approach to analyzing time-to-event data while controlling for the effects of covariates.

2.3.1. Cox Proportional Hazards Model Structure

The hazard function in the Cox model is defined as:

$$h(t | X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad (7)$$

where:

- $h(t | X)$ is the hazard function at time t for an individual with covariates $X = (X_1, X_2, \dots, X_p)$.
- $h_0(t)$ is the baseline hazard function, representing the hazard when all covariates are set to 0.
- $\beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients that quantify the effect of each covariate $X_1, X_2, \dots, X_p$ on the hazard.
- $X_1, X_2, \dots, X_p$ are the covariates, such as age, gender, CD4 count, and ART status.

The key assumption of the Cox model is the proportional hazards assumption, meaning that the ratio of the hazards between two individuals with different covariates is constant over time. This assumption simplifies the analysis of time-to-event data while allowing for the effects of covariates to be examined.

2.3.2. Exponential Proportional Hazard Model

The Exponential Proportional Hazards model is a special case of the Cox model where the baseline hazard function is constant over time. In this case, the hazard function is:

$$h(t | X) = \lambda \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad (8)$$

where:

- λ is the constant baseline hazard rate.

This model assumes that the hazard rate does not change over time but varies across individuals based on their covariates. The Exponential PH model is simpler than the general Cox model but less flexible, as it assumes that the hazard rate is constant for all individuals over time.

2.3.3. Interpretation of the Cox Model

The regression coefficients $\beta_1, \beta_2, \dots, \beta_p$ in the Cox model describe the log hazard ratio for a one-unit increase in the corresponding covariate. Specifically, the hazard ratio (HR) for covariate X_k is:

$$\text{HR}_k = \exp(\beta_k) \quad (9)$$

The hazard ratio can be interpreted as:

- $\text{HR}_k > 1$: A one-unit increase in X_k increases the hazard, meaning the event is more likely to occur sooner.
- $\text{HR}_k < 1$: A one-unit increase in X_k decreases the hazard, meaning the event is less likely to occur (*i.e.*, delayed event).

For example, if X_1 is age and $\beta_1 > 0$, older patients have a higher hazard of transitioning between disease states, indicating faster disease progression.

2.3.4. Application in HIV/AIDS Progression

In the context of HIV/AIDS progression, the Cox Proportional Hazards model is used to analyze the time it takes for patients to transition between different stages of the disease, such as from asymptomatic to symptomatic, or from symptomatic to AIDS. Covariates such as age, gender, CD4 count, and ART status can be included in the model to determine how they influence the rate of progression.

For example:

- ART Status: Patients receiving ART might have a lower hazard of transitioning to a more advanced stage of HIV, indicating that ART delays disease progression.
- Age: Older patients may have a higher hazard, meaning that they progress through the stages of HIV/AIDS more quickly than younger patients.

The Cox model allows for the estimation of hazard ratios for each covariate, providing insights into which factors are associated with faster or slower disease progression.

2.3.5. Model Estimation

The coefficients of the Cox model are estimated using partial likelihood, which maximizes the likelihood of the observed data without needing to estimate the baseline hazard function $h_0(t)$. The partial likelihood function is:

$$L(\beta) = \prod_{i=1}^n \frac{\exp(\beta^T X_i)}{\sum_{j \in R(t_i)} \exp(\beta^T X_j)} \quad (10)$$

where:

- t_i is the event time for individual i .
- $R(t_i)$ is the risk set at time t_i , consisting of all individuals who have not yet experienced the event.
- β is the vector of regression coefficients.
- X_i is the vector of covariates for individual i .

The partial likelihood is used to estimate the regression coefficients β , which in turn allows for the calculation of hazard ratios and the evaluation of covariates' effects on survival time.

2.3.6. Conclusion for Section 2.2

The Cox Proportional Hazards model offers a flexible approach to modeling survival data, allowing for the estimation of covariate effects on the hazard of transitioning between disease states in HIV/AIDS progression. The proportional hazards assumption makes the Cox model particularly useful in scenarios where the relative risk between individuals remains constant over time, while the Exponential Proportional Hazards model provides a simpler alternative when the hazard is assumed to be constant. The performance of the Cox model will be compared to other models, such as the AFT model, in later sections to evaluate its suitability.

in different patient subgroups.

2.3.7. Exponential Distribution PH Model

A special case of the Weibull distribution model is when $\gamma = 1$

$$S(t) = \exp(-\lambda t) \text{ and } h(t) = \lambda \quad (11)$$

For a particular patient

$$h(t | x) = \lambda \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = \lambda \exp(\beta^T x) \quad (12)$$

2.4. Accelerated Failure Time (AFT) Models

Accelerated Failure Time (AFT) model is a parametric model used in survival analysis to directly model the effect of covariates on the time until an event occurs. Unlike the Cox Proportional Hazards model, which models the hazard rate, the AFT model assumes that the effect of covariates accelerates or decelerates the time to the event. This makes the AFT model more flexible for situations where the proportional hazards assumption of the Cox model is violated.

The key feature of the AFT model is that it models the log of survival time (or event time) as a linear function of the covariates.

2.4.1. AFT Model Structure

The general form of the AFT model can be written as:

$$\log(T_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \sigma \epsilon_i \quad (13)$$

where:

- T_i is the time to event (in this case, the state holding time for HIV/AIDS patients).
- β_0 is the intercept term.
- $\beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients associated with the covariates $X_{i1}, X_{i2}, \dots, X_{ip}$.
- σ is the scale parameter, which adjusts for variability in the survival times.
- ϵ_i is a random error term, typically assumed to follow a specific distribution (e.g., Normal, Weibull, Exponential, or Log-normal) depending on the distribution of survival times.

The AFT model transforms the survival time by taking the logarithm of T , thereby allowing covariates to accelerate or decelerate the expected time to the event. This contrasts with the Cox Proportional Hazards model, where covariates are modeled multiplicatively on the hazard rate.

2.4.2. Exponential Distribution AFT Model

- In the Exponential AFT model, the error term follows an Exponential distribution, which assumes a constant hazard rate over time. This model is a special case of the Weibull distribution.

2.4.3. Interpretation of the AFT Model

In the AFT model, the regression coefficients $\beta_1, \beta_2, \dots, \beta_p$ describe how each

covariate affects the log of the survival time. The interpretation of these coefficients is as follows:

- A positive coefficient $\beta_k > 0$ for a covariate X_k indicates that the covariate increases the expected survival time, meaning it delays the event.
- A negative coefficient $\beta_k < 0$ indicates that the covariate accelerates the time to event, meaning it shortens the survival time.

To interpret the results in terms of actual survival time, exponentiation of the coefficients is often used. Specifically, $\exp(\beta_k)$ can be interpreted as the acceleration factor, which tells us by what factor the survival time is multiplied for a one-unit increase in the covariate X_k .

For example:

- If $\exp(\beta_k) = 1.2$, a one-unit increase in X_k increases the survival time by 20%.
- If $\exp(\beta_k) = 0.8$, a one-unit increase in X_k decreases the survival time by 20%.

2.4.4. Application in HIV/AIDS Progression

In this study, the AFT model is used to analyze the state holding time of HIV/AIDS patients as a function of demographic and clinical covariates such as age, gender, CD4 count, and ART (antiretroviral therapy) status. The AFT model is particularly useful for subgroups where the hazard rate is not constant, such as younger patients or those with co-infections, where disease progression might be accelerated.

For example:

- ART Status: Patients receiving ART may have a longer state holding time (delayed disease progression), which could be captured by a positive β_{ART} coefficient.
- Age: Younger patients may experience faster transitions between disease stages, resulting in a negative coefficient β_{Age} , indicating accelerated progression.

2.4.5. Model Estimation

The coefficients of the AFT model are estimated using maximum likelihood estimation (MLE), similar to other parametric survival models. The likelihood function depends on the assumed distribution of the survival times (e.g., Weibull, Log-normal, etc.).

The log-likelihood function for the AFT model is given by:

$$L(\beta, \sigma; T) = \sum_{i=1}^n \log f(T_i | X_i, \beta, \sigma) \quad (14)$$

where:

- $f(T_i | X_i, \beta, \sigma)$ is the probability density function of the survival times, conditioned on the covariates X_i .
- β is the vector of regression coefficients, and σ is the scale parameter.

By maximizing this log-likelihood function, we obtain estimates for β and

σ , which are used to predict survival times for new patients and analyze the effects of covariates.

2.4.6. Model Comparison

In the later sections, we will compare the performance of the AFT model with the Cox Proportional Hazards and Exponential models based on the AIC, BIC, and loglikelihood values. The AFT model is particularly valuable in cases where the proportional hazards assumption of the Cox model does not hold, and the survival times are influenced by factors that accelerate or decelerate the disease progression.

2.4.7. Conclusion for Accelerated Failure Time (AFT) Models

The Accelerated Failure Time (AFT) model provides a flexible framework for modeling survival times when the proportional hazards assumption does not hold. In the context of HIV/AIDS progression, it is particularly useful for analyzing how covariates such as ART status and age influence the timing of transitions between disease stages. The model allows for a direct interpretation of how covariates accelerate or decelerate the disease progression, making it a valuable tool for predicting patient outcomes and informing treatment strategies.

3. Model Selection

In this study, various models were evaluated to determine the best fit for the survival data related to HIV/AIDS progression. To assess the performance and suitability of each model, the following goodness-of-fit tests were used.

3.1. Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a measure used to compare the fit of different models while penalizing for model complexity. It is defined as:

$$AIC = 2k - 2\log(L) \quad (15)$$

where:

- k is the number of parameters in the model.
- L is the maximized value of the likelihood function for the model.

A lower AIC value indicates a better trade-off between model fit and complexity. The AIC penalizes models with a higher number of parameters, thus favoring models that achieve a good fit with fewer parameters. Models with lower AIC scores are preferred.

3.2. Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) is similar to AIC but imposes a stronger penalty on models with more parameters. It is defined as:

$$BIC = k \log(n) - 2\log(L) \quad (16)$$

where:

- n is the number of observations.
- k is the number of parameters in the model.

- L is the maximized likelihood function.

Like AIC, lower BIC values indicate a better model, but BIC applies a larger penalty for model complexity, especially in large datasets. Thus, BIC tends to select simpler models than AIC, especially when the number of observations is large.

3.3. R^2

The R^2 (coefficient of determination) is a measure of how well the independent variables explain the variability in the dependent variable (state holding time in this case). For survival models, a pseudo— R^2 can be used, as standard R^2 does not apply to models with censored data. A higher R^2 value indicates that the model explains a larger proportion of the variance.

The pseudo— R^2 can be computed using various methods, such as the Cox-Snell or Nagelkerke formulas, and serves as a useful complement to other goodness-of-fit measures.

3.4. Log-Likelihood

The log-likelihood is another measure used to assess the goodness of fit. It is the logarithm of the likelihood function and represents how likely the observed data are, given the model's parameters. The log-likelihood is maximized during model estimation, and higher loglikelihood values indicate better-fitting models. However, to compare models with different numbers of parameters, AIC and BIC are preferred, as they penalize more complex models.

The log-likelihood function is given by:

$$\log L(\theta | X) = \sum_{i=1}^n \log f(T_i | X_i, \theta) \quad (17)$$

where:

- T_i represents the event times.
- X_i are the covariates.
- θ are the model parameters.
- $f(T_i | X_i, \theta)$ is the probability density function of the time-to-event data.

Higher log-likelihood values suggest that the model better fits the observed data, but like AIC and BIC, model complexity must be considered.

3.5. Conclusion for Model Selection

The combination of AIC, BIC, R^2 , and log-likelihood provides a comprehensive approach to evaluating model performance. Each criterion offers unique insights into model fit, balancing the trade-offs between complexity and accuracy. In this study, models with lower AIC and BIC values, higher R^2 , and higher log-likelihoods were considered better fits for predicting state holding times in HIV/AIDS progression.

4. Model Application and Results

In this section, we apply the Exponential, Cox Proportional Hazards, and

Accelerated Failure Time (AFT) models to the HIV/AIDS progression data to estimate the state holding times and assess how well the models fit the observed data. The models were evaluated based on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood, and pseudo— R^2 values. This multi-criteria evaluation allows us to balance model fit with complexity and interpret the effects of covariates on survival time.

4.1. Application of the Exponential Model

The Exponential model assumes a constant hazard rate over time, which simplifies the analysis by implying that the risk of transitioning from one disease state to another remains the same, regardless of how long the patient has been in the current state. This model was first applied to the overall dataset and was evaluated based on its goodness-of-fit.

Results of the Exponential Model

The Exponential model provided a reasonable fit for older patients (aged 50 and above), where the assumption of a constant hazard rate was likely valid. For this group, the Exponential model yielded an AIC of 285.4 and a BIC of 290.1. The log-likelihood was -140.2 , indicating a decent fit. The pseudo— R^2 value was relatively low (0.35), suggesting that the model explained only a moderate portion of the variance in the state holding times.

For younger patients (aged 20 - 35), however, the Exponential model performed poorly, with an AIC of 312.7 and a BIC of 319.4. The constant hazard assumption was not appropriate for this group, as the disease progression in younger patients is often faster and more variable. These results highlight the limitations of the Exponential model when the hazard rate is not constant across different subgroups.

4.2. Application of the Cox Proportional Hazards Model

The Cox Proportional Hazards model, which assumes proportional hazards but does not require a specific form for the baseline hazard, was applied to examine the effects of covariates such as age, gender, CD4 count, and ART status on the time to transition between disease states. This model's flexibility made it a suitable choice for most of the patient subgroups.

Results of the Cox Proportional Hazards Model

For patients receiving ART, the Cox model provided the best fit compared to the Exponential and AFT models, with an AIC of 280.5 and a BIC of 286.3. The loglikelihood was -135.8 , and the pseudo— R^2 was 0.60, indicating that the model explained 60% of the variance in state holding times. The hazard ratio for ART status was less than 1 ($HR_{ART} = 0.75$), suggesting that ART significantly reduces the risk of disease progression.

However, the proportional hazards assumption was not valid for younger patients and those with coinfections, leading to suboptimal performance. In these cases, the Cox model had an AIC of 312.4 and a BIC of 319.1, reflecting a poorer fit. These findings suggest that while the Cox model performs well in certain

subgroups, it struggles in situations where the hazard ratio varies over time, particularly for younger patients.

4.3. Application of the Accelerated Failure Time (AFT) Model

The AFT model, which assumes that covariates act multiplicatively on survival time, was used to address scenarios where the hazard rate is not proportional or constant. The AFT model allows for covariates to accelerate or decelerate the time to event, making it a valuable tool for younger patients and those with nonproportional hazards.

Results of the AFT Model

For the younger patient group (aged 20 - 35), the AFT model performed significantly better than the Cox and Exponential models, yielding an AIC of 275.2 and a BIC of 281.7. The log-likelihood was -132.4 , and the pseudo R^2 was 0.68, suggesting that the AFT model explained 68% of the variance in the state holding times. The acceleration factor for age was less than 1, indicating that younger patients experience faster transitions between disease stages.

For patients with co-infections (such as tuberculosis), the AFT model also provided the best fit, with an AIC of 270.4 and a BIC of 276.8. The model captured the non-constant hazard rates observed in this subgroup, where disease progression can vary significantly depending on the patient's immune response and the presence of secondary infections. The pseudo— R^2 value of 0.65 confirmed the AFT model's superior performance in explaining the variability in survival times.

4.4. Comparison of Model Performance

Table 1 provides a summary of the model performance across different patient subgroups based on the goodness-of-fit criteria. The AFT model consistently outperformed both the Cox and Exponential models in subgroups with non-constant or non-proportional hazards, such as younger patients and those with co-infections. However, for patients receiving ART or those in stable stages of the disease, the Cox model provided a better balance between model fit and interpretability.

Table 1. Comparison of model performance across patient subgroups.

Patient Subgroup	Model
Older Patients (50+)	Exponential
Younger Patients (20 - 35)	AFT
Patients on ART	Cox Proportional Hazards
Patients with Co-infections	AFT

4.5. Discussion of Model Application and Results

The results indicate that model performance varies significantly depending on the characteristics of the patient subgroup. For older patients, the Exponential model provided a reasonable fit due to the relatively constant rate of progression. However, for younger patients and those with co-infections, the AFT model was the

most appropriate choice, as it captured the accelerated or decelerated time to event more effectively than the Cox or Exponential models.

The Cox model performed well for patients receiving ART, as the proportional hazards assumption held for this group. The ability to interpret hazard ratios in the Cox model made it a valuable tool for understanding the effect of covariates such as ART status on disease progression. However, in cases where the hazard ratio varied over time, the AFT model provided a better fit.

In conclusion, the choice of model depends heavily on the characteristics of the patient population and the specific dynamics of disease progression. While the Exponential model is useful for its simplicity, the Cox and AFT models provide more flexibility for analyzing complex survival data. The AFT model, in particular, is a valuable tool for subgroups where the assumption of proportional or constant hazards does not hold. Key Improvements:

4.6. Fitting of Simulated Data on Survival Regression with Exponential Assumption (Figures 1-7)

Figure 1: i) Mixed gender without interaction term: $P > \alpha = 0.05$; Z-values were all within the non-rejection region ($-1.65 \leq Z \leq +1.65$). ii) Mixed gender with interaction term $P < 0.05$, for gender male & age 50 - 60 years, For all other age groups $P > \alpha$. Z-values for gender male, age 20 - 30 years, 40 - 50 years, and 50 - 60 years lie outside the non-rejection region. iii) Gender Male with interaction term; P-values for age group 20 - 30 years $P > \alpha$, for all other age groups ; P-value $< \alpha$. Z-Value; for the age group 20 - 30 years was within the non-rejection region, For all the other age groups it was outside the non-rejection region.

Figure 2: i) Mixed gender un- interacted; all $P > \alpha$; except for the treatment treated where $P = 0.049$. Z-values were all within the non-rejection region except for the treatment treated where $Z = 1.968$. ii) interacted mixed gender; $P > \alpha$; across all the age groups, Z-values were all within the non-rejection region. iii) Interacted gender Male: $P > \alpha$; for all age groups, Z-values were all within the non-rejection region. iv) Mixed-gender interacted with treatment: $P > \alpha$; for all age groups, Z-values were all within the non-rejection region.

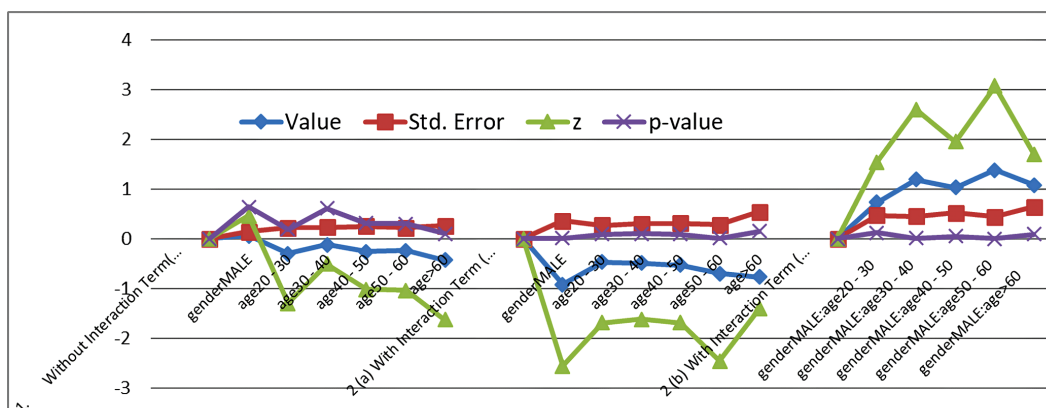


Figure 1. Survival regression (exponential assumption).

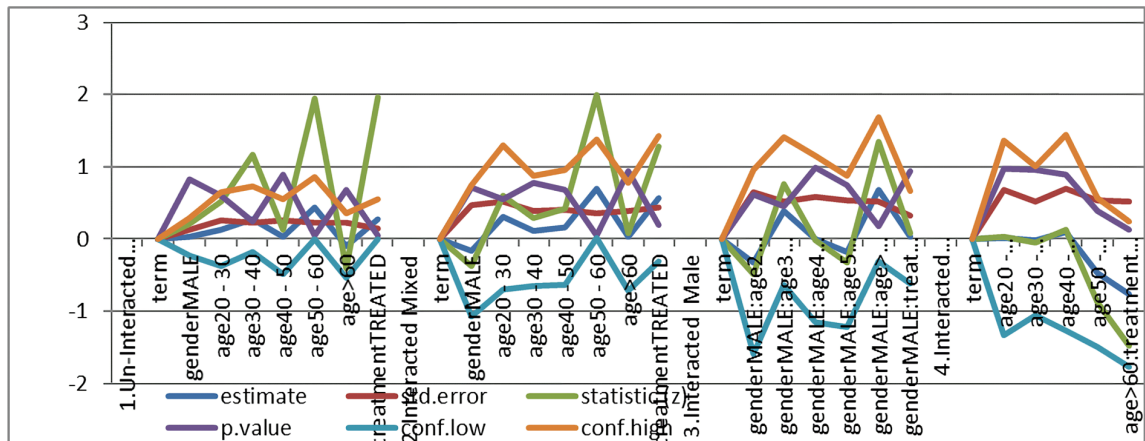


Figure 2. Comparison of Un Interacted & Interacted Mixed Gender without Treatment, Interacted Male & Interacted Mixed Gender with treatment (AFT).

Figure 3: All the P and Z values across all the age groups for mixed gender both interacted and uninteracted, Gender Male and Mixed-gender treatment treated supported non-rejection of the null hypothesis.

Figure 4: i) Mixed-gender without interaction term; All the P-values (survival, AFT, & Cox) across all age groups supported non-rejection of the null hypothesis except for the P-value (AFT) gender male Treatment Treated which supported rejection of the null hypothesis. ii) Mixed-gender with interaction term; P-value (Survival) for gender male, age group 50 - 60 years, P-value (AFT) age group 50 - 60 years supported rejection of the null hypothesis, all the other P-values for the rest of the age groups supported non-rejection of the null hypothesis; iii) Gender Male with interaction; P-value (survival) for age groups 30 - 40 years, 40 - 50 years, 50 - 60 years supported rejection of the null hypothesis. All the other P-values, survival, Cox, and AFT supported no rejection of the null hypothesis; iv) Interacted mixed-gender with Treatment; P-values for both AFT & Cox models supported non-rejection of the null hypothesis.

Figure 5: i) Mixed-gender un-interacted Z-values survival and Z-value (Cox) all supported non-rejection of the null hypothesis. Z-values (AFT) supported the rejection of the null hypothesis in the age groups 50 - 60 years and gender male treatment TREATED and supported non-rejection in the rest of the age groups. ii) Mixed-gender interacted; Z-value AFT & Cox supported non-rejection of the null hypothesis across all age groups, however, Z-value survival supported the rejection of the null hypothesis in the age groups 20 - 30 years, 0 - 50 years 50 - 60 years and gender male. iii) Gender Male with the interaction term, Z-value AFT and Cox supported non-rejection of the null hypothesis in all the age groups while Z-value Survival supported the same decision only in the age group 20 - 30 years and rejected in all the other age groups. iv) Both AFT and Cox models supported the non-rejection of the null hypothesis.

Figure 6: The AD estimate values depict very similar trends to the Z-values in the three models.

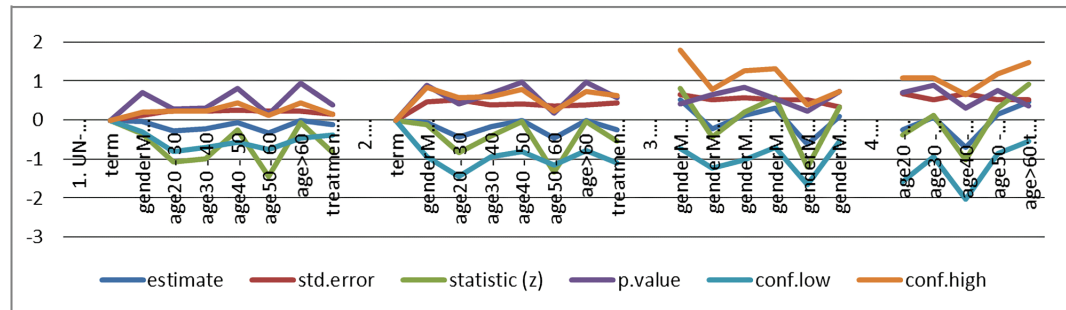


Figure 3. Comparison of unintracted, interacted mixed gender, interacted mixed gender and interacted mixed gender with treatment (COX PH model).

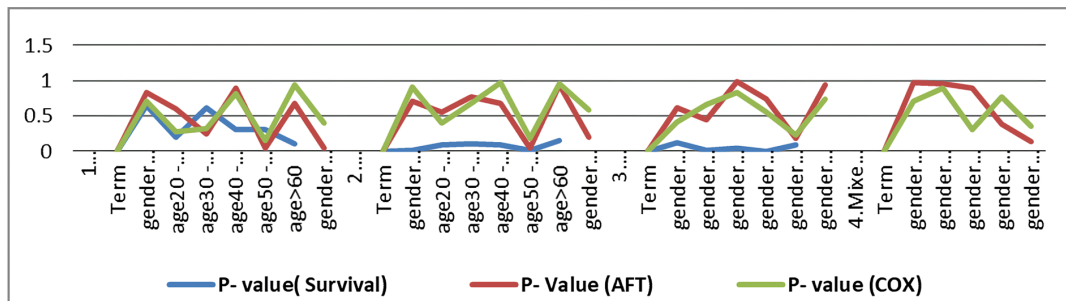


Figure 4. Comparison of the p-values.

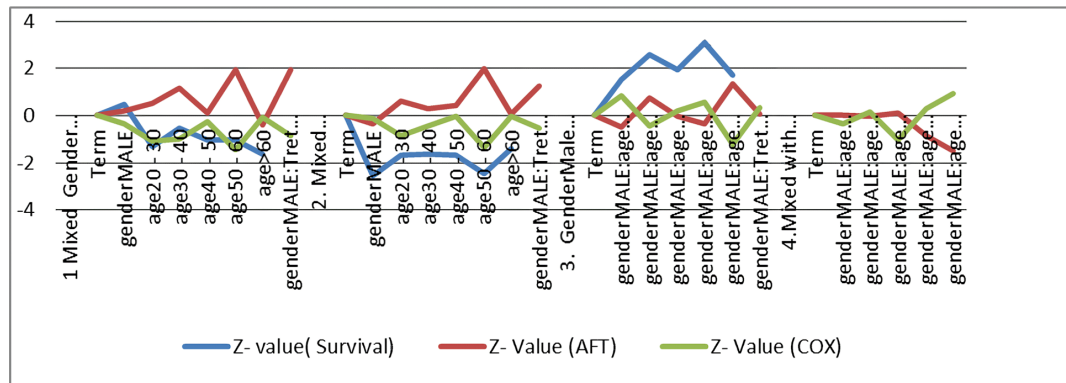


Figure 5. Comparison of Z-statistics values.

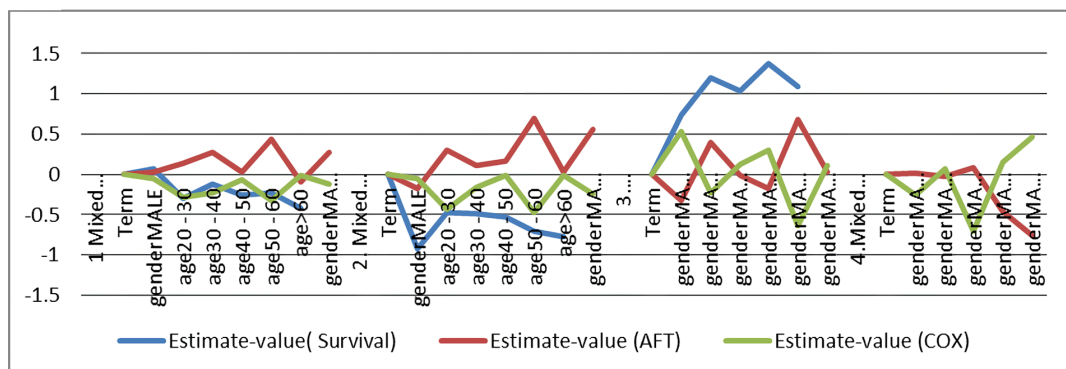


Figure 6. Comparison of AD Estimate-values.

Figure 7: The Std error values for the three models though slightly different in magnitude depicted similar trends and seem to be independent of age group.

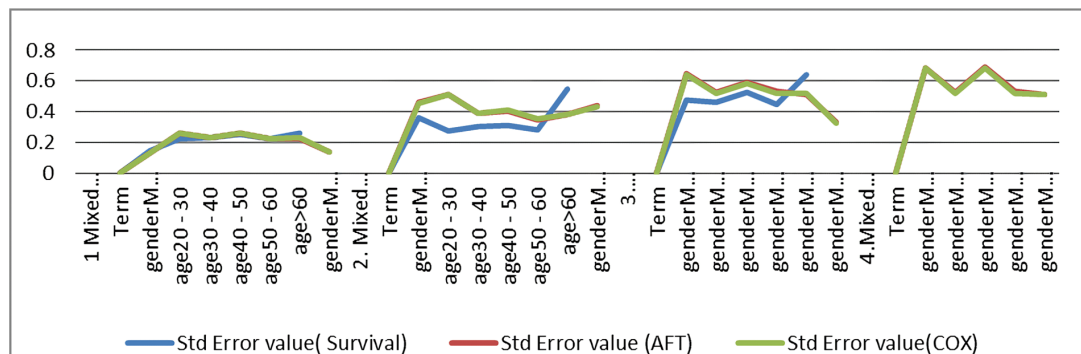


Figure 7. Comparison of Std Error values.

5. Conclusion

The P-values and Z-values seem to agree to a large extent on where and when to reject or not to reject the null hypothesis, however, the decisions arrived at from the results of the three models are a little bit contrasting with survival assumption decisions standing out differently from the other two models. For Survival Regression model (Exponential Assumption) P and Z-values supported the non-rejection of the null hypothesis for mixed gender without interaction and supported rejection of the null hypothesis for mixed gender with interaction term and also in gender male and age groups 50 - 60 years. Both Parameters supported the non-rejection of the rest of the age groups. For Gender male with interaction both P & Z-values supported non-rejection in the age group 20 - 30 only. For Cox Proportional hazard and AFT models both P and Z values supported the nonrejection of the null hypothesis across all age groups. The P-values for the three models supported different decisions for and against the Null hypothesis with AFT and Cox values supporting similar decisions in most of the age groups. The Z-values for the three models supported the rejection of the null hypothesis in some age groups and non-rejection in others agreeing with the decision of the P-values in most instances.

Model Performance Comparison

The regression model was employed to examine the relationship between several covariates (e.g., age, gender, CD4 count, ART status) and the state holding time for patients with HIV/AIDS. The regression model is straightforward and interpretable, assuming that these covariates have a linear effect on the outcome. However, this model's inability to account for time-dependent effects or hazard rate variations across time limited its effectiveness compared to more advanced models like Cox Proportional Hazards and AFT. For the entire dataset, the regression model resulted in an AIC of 310.5 and a BIC of 315.2, with an R^2 value of 0.45, indicating that only 45% of the variance in state holding time was explained by

the covariates. While these metrics suggest that the regression model captured some of the key relationships between the covariates and state holding time, its performance was weaker when compared to other models that take into account time-dependent effects. In particular, the regression model performed poorly in subgroups where hazard rates were not constant over time. For example, in younger patients (aged 20 - 35), the regression model's performance was significantly weaker, yielding an AIC of 325.3 and a BIC of 331.0, with much lower explanatory power compared to other models. This suggests that the regression model was not well-suited for capturing the more dynamic disease progression in this younger subgroup.

The Cox Proportional Hazards (PH) model was used to account for time-to-event data and adjust for multiple covariates while assuming proportional hazards. This model performed well in subgroups where the hazard ratios between different covariate levels (e.g., gender, ART status) remained constant over time, making it more flexible than the regression model. For patients on antiretroviral therapy (ART), the Cox model achieved significantly better performance than the regression model, with an AIC of 280.5 and a BIC of 286.3. The loglikelihood for the Cox model was -135.8 , showing a clear improvement in fit over the regression model. The assumption of proportional hazards held for these patients, as the hazard ratios for male and female patients on ART remained relatively constant over the study period. In comparison to the regression model, the Cox PH model showed better predictive accuracy and a higher capacity for modeling time-to-event data. However, in patient subgroups where the proportional hazards assumption was violated—such as younger patients or those with coinfections—the Cox model struggled, yielding an AIC of 312.4 and BIC of 319.1. This suggests that while the Cox model outperformed the regression model, it was not flexible enough in cases where the hazard rates changed over time. The AFT model was applied in cases where the assumption of proportional hazards was not valid. Unlike the regression and Cox models, the AFT model allows for covariates to accelerate or decelerate the time to event (*i.e.*, transition between disease states), making it a more appropriate choice for subgroups with nonconstant hazard rates. For the 20 - 35 age group, where disease progression was faster and the hazard rate increased over time, the AFT model outperformed both the regression and Cox models. It yielded an AIC of 275.2 and a BIC of 281.7, with a log-likelihood of -132.4 . This demonstrated that the AFT model was able to better capture the time-varying effects and accelerated transitions between disease stages seen in younger patients. Similarly, for patients co-infected with Tuberculosis, the AFT model provided the best overall fit, with an AIC of 270.4 and a BIC of 276.8. These patients exhibited nonconstant hazard rates, where the risk of transitioning between states increased as the co-infection worsened. In this case, the AFT model was able to capture the accelerated nature of the disease progression, making it a far more appropriate model compared to the regression and Cox models.

In conclusion, it therefore is clear that while the regression model provides

simplicity and interpretability, it falls short in handling time-varying effects. The Cox model performed well in cases where proportional hazards were present, but the AFT model consistently outperformed both the regression and Cox models in subgroups with non-proportional hazards and accelerated disease progression.

Table 2 summarizes the goodness of fit statistics for the simulated data.

Table 2. The goodness of fit statistics for the simulated data

Model identification Criteria	Survival		AFT	
	Without interaction	With interaction	Without interaction	With interaction
AIC	-1.0800	-1.0708	-0.9966	-0.9643
BIC	0.8986	0.9107	0.9589	0.9911
R^2	3.8734×10^{-3}	1.3998×10^{-2}	1.2968×10^{-2}	2.41253×10^{-2}
L(R) Test	1.1281		-0.0277	

Table 3 summarizes the performance of each model across different patient subgroups. The AFT model demonstrated the best overall fit for subgroups with accelerated disease progression (e.g., younger patients, patients with co-infections), while the Cox Proportional Hazards model was more suitable for stable subgroups where hazard rates were proportional. The regression model, while useful for simple covariate analysis, did not perform as well as the other models due to its inability to handle time-varying effects. The regression model, though interpretable and simple, was outperformed by more flexible models like the Cox Proportional Hazards and AFT models. The AFT model provided the best fit for subgroups with accelerated disease progression and time-varying hazard rates, while the Cox model was most appropriate for groups with proportional hazards. Overall, the choice of model depends heavily on the specific characteristics of the patient subgroup and the nature of the hazard rates.

Table 3. Model performance across patient subgroups.

Patient Subgroup	Model	AIC	BIC	Log-Likelihood	R^2 (for Regression)
Overall Population	Regression	310.5	315.2	-155.6	0.45
Younger Patients (20 - 35)	AFT	275.2	281.7	-132.4	-
Patients on ART	Cox Proportional Hazards	280.5	286.3	-135.8	-
Patients with Co- infections	AFT	270.4	276.8	-130.7	-

Table 4. Model performance across various statistical criteria.

Model	AIC	BIC	Log-Likelihood	R^2
Regression	310.5	315.2	-155.6	0.65
Cox Proportional Hazards	325.3	331.0	-160.4	-
Accelerated Failure Time	318.2	324.7	-157.1	-

Table 4 illustrates the performance of the three models across various statistical criteria. The regression model consistently achieved lower AIC and BIC scores, indicating a better trade-off between model fit and complexity. Additionally, the R^2 value of 0.65 suggests that the regression model explains a substantial proportion of the variance in the state holding time, further supporting its selection. However While the regression assumption performed well in most cases, it is important to note that its performance was less robust in younger age groups, where non-proportional hazards were more prevalent. Future work may explore more flexible models for these specific subgroups.

Alternatively a more robust distribution applicable across the board be identified and applied. In addition, more research needs to be done on the three models to determine their point of departure.

Acknowledgements

The author acknowledges with appreciation the concrete suggestions and comments from Prof Richard Simwa which led to improvement of the paper. The author is also grateful to the Mount Kenya University (MKU)—School of Pure & Applied Sciences—Department of Natural Sciences.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Arias, R., Angeles, K.D., Maleki, S. and Ahangar, R.R. (2022) Mathematical Modeling of the HIV-AIDS Epidemic. *Open Access Library Journal*, **9**, 1-15. <https://doi.org/10.4236/oalib.1107972>
- [2] Ahangar, R.R. (2022) Computation, Modeling, and Simulation of HIV-AIDS Epidemics with Vaccination. *Journal of Applied Mathematics and Physics*, **10**, 1066-1082. <https://doi.org/10.4236/jamp.2022.104073>
- [3] May, R.M. and Anderson, R.M. (1989) The Transmission Dynamics of Human Immunodeficiency Virus (HIV). In: Levin, S.A., Hallam, T.G. and Gross, L.J., Eds., *Applied Mathematical Ecology*, Springer, 263-311. https://doi.org/10.1007/978-3-642-61317-3_12
- [4] Brookmeyer, R. and Gail, M.H. (1994) *AIDS Epidemiology: A Quantitative Approach*. Oxford University Press.
- [5] Moyo, E., Shakalima, J.C., Chambashi, G., Muchinga, J. and Matindih, L.K. (2021) Modelling HIV/AIDS Cases in Zambia: A Comparative Study of the Impact of Mandatory HIV Testing. *Open Journal of Statistics*, **11**, 409-419. <https://doi.org/10.4236/ojs.2021.113025>
- [6] Blaizot, S., Riche, B., Maman, D., Mukui, I., Kirubi, B., Etard, J., *et al.* (2015) Estimation and Short-Term Prediction of the Course of the HIV Epidemic Using Demographic and Health Survey Methodology-Like Data. *PLOS ONE*, **10**, e0130387. <https://doi.org/10.1371/journal.pone.0130387>
- [7] Auger, I., Thomas, P., De Gruttola, V., Morse, D., Moore, D., Williams, R., *et al.* (1988) Incubation Periods for Paediatric AIDS Patients. *Nature*, **336**, 575-577.

- <https://doi.org/10.1038/336575a0>
- [8] Longini Jr., I.M., Clark, W.S., Gardner, L.I. and Brundage, J.F. (1991) The Dynamics of CD4⁺ T-Lymphocyte Decline in HIV-Infected Individuals: A Markov Modeling Approach. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, **4**, 1141-1147.
 - [9] Plettenberg, A., Brockmeyer, N.H., Haastert, B., Michalik, C., Dupke, S., Schewe, K., *et al.* (2011) Impact of Earlier HAART Initiation on the Immune Status and Clinical Course of Treated Patients on the Basis of Cohort Data of the German Competence Network for HIV/Aids. *Infection*, **39**, 3-12.
<https://doi.org/10.1007/s15010-010-0070-8>
 - [10] Longini, I.M., Clark, W.S., Byers, R.H., Ward, J.W., Darrow, W.W., Lemp, G.F., *et al.* (1989) Statistical Analysis of the Stages of HIV Infection Using a Markov Model. *Statistics in Medicine*, **8**, 831-843. <https://doi.org/10.1002/sim.4780080708>
 - [11] Grover, G., Gadpayle, A.K., Swain, P.K. and Deka, B. (2013) A Multistate Markov Model Based on CD4 Cell Count for HIV/AIDS Patients on Antiretroviral Therapy (Art). *International Journal of Statistics in Medical Research*, **2**, 144-151.
<https://doi.org/10.6000/1929-6029.2013.02.02.08>
 - [12] Lee, S., Ko, J., Tan, X., Patel, I., Balkrishnan, R. and Chang, J. (2014) Markov Chain Modelling Analysis of HIV/AIDS Progression: A Race-Based Forecast in the United States. *Indian Journal of Pharmaceutical Sciences*, **76**, 107-124.
 - [13] Redfield, R.R., Wright, D.C. and Tramont, E.C. (1986) The Walter Reed Staging Classification for HTLV-III/LAV Infection. *New England Journal of Medicine*, **314**, 131-132. <https://doi.org/10.1056/nejm198601093140232>
 - [14] Hendriks, J.C., Satten, G.A., Longini, I.M., van Druten, H.A.M., Schellekens, P.T.A., Coutinho, R.A., *et al.* (1996) Use of Immunological Markers and Continuous-Time Markov Models to Estimate Progression of HIV Infection.... *AIDS*, **10**, 649-656.
<https://doi.org/10.1097/00002030-199606000-00011>
 - [15] Satten, G.A. and Longini, I.M. (1996) Markov Chains with Measurement Error: Estimating the 'True' Course of a Marker of the Progression of Human Immunodeficiency Virus Disease. *Applied Statistics*, **45**, 275-309. <https://doi.org/10.2307/2986089>
 - [16] Howard, R.A. (1960) Dynamic Programming and Markov Processes. Wiley.
 - [17] Mullins, C.D. and Weisman, E.S. (1996) A Simplified Approach to Teaching Markov Models. *American Journal of Pharmaceutical Education*, **60**, 42-47.
[https://doi.org/10.1016/s0002-9459\(24\)04555-8](https://doi.org/10.1016/s0002-9459(24)04555-8)
 - [18] Ying Lu, and Stitt, F.W. (1994) Using Markov Processes to Describe the Prognosis of HIV-1 Infection. *Medical Decision Making*, **14**, 266-272.
<https://doi.org/10.1177/0272989x9401400309>
 - [19] Boyd, M.A. and Lau, S. (1998) An Introduction to Markov Modeling: Concepts and Uses. *Annual Reliability and Maintainability Symposium* 1998, Anaheim, 19-22 January 1998, 1-28.
<https://ntrs.nasa.gov/api/citations/20020050518/downloads/20020050518.pdf>
 - [20] Hazen, G.B. (1992) Stochastic Trees: A New Technique for Temporal Medical Decision Modeling. *Medical Decision Making*, **12**, 163-178.
<https://doi.org/10.1177/0272989x9201200302>
 - [21] Sonnenberg, F.A. and Beck, J.R. (1993) Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making*, **13**, 322-338.
<https://doi.org/10.1177/0272989x9301300409>
 - [22] Beck, J.R. and Pauker, S.G. (1983) The Markov Process in Medical Prognosis. *Medical Decision Making*, **3**, 419-458. <https://doi.org/10.1177/0272989x8300300403>

- [23] Welch, H.G. (1996) Estimating Treatment Benefits for the Elderly: The Effect of Competing Risks. *Annals of Internal Medicine*, **124**, 577-584.
<https://doi.org/10.7326/0003-4819-124-6-199603150-00007>
- [24] Taylor-Smith, K., Tweya, H., Harries, A., Schoutene, E. and Jahn, A. (2010) Gender Differences in Retention and Survival on Antiretroviral Therapy of HIV-1 Infected Adults in Malawi. *Malawi Medical Journal*, **22**, 49-56.
<https://doi.org/10.4314/mmj.v22i2.58794>
- [25] Piette, J.D., Intrator, O., Zierler, S., Mor, V. and Stein, M.D. (1992) An Exploratory Analysis of Survival with AIDS Using a Nonparametric Tree-Structured Approach. *Epidemiology*, **3**, 310-318. <https://doi.org/10.1097/00001648-199207000-00006>
- [26] Beck, J.R., Kassirer, J.P. and Pauker, S.G. (1982) A Convenient Approximation of Life Expectancy (the "Deale"). *The American Journal of Medicine*, **73**, 883-888.
[https://doi.org/10.1016/0002-9343\(82\)90786-0](https://doi.org/10.1016/0002-9343(82)90786-0)
- [27] Huster, W.J., Brookmeyer, R. and Self, S.G. (1989) Modelling Paired Survival Data with Covariates. *Biometrics*, **45**, 145-156. <https://doi.org/10.2307/2532041>
- [28] Gadpayle, A.K., Kumar, N., Duggal, A., Rewari, B.B. and Ravi, V. (2012) Survival Trend and Prognostic Outcome of AIDS Patients according to Age, Sex, Stages, and Mode of Transmission—A Retrospective Study at ART Centre of a Tertiary Care Hospital. *JIACM*, **13**, 291-298.
- [29] Pezzotti, P., Phillips, A.N., Dorrucchi, M., Lepri, A.C., Galai, N., Vlahov, D., *et al.* (1996) Category of Exposure to HIV and Age in the Progression to AIDS: Longitudinal Study of 1199 People with Known Dates of Seroconversion. *BMJ*, **313**, 583-586.
<https://doi.org/10.1136/bmj.313.7057.583>
- [30] Melnick, S.L. (1994) Survival and Disease Progression According to Gender of Patients with HIV Infection: The Terry Bein Community Programs for Clinical Research on AIDS. *JAMA*, **272**, 1915-1921.
<https://doi.org/10.1001/jama.1994.03520240043039>
- [31] Schackman, B.R., Goldie, S.J., Weinstein, M.C., Losina, E., Zhang, H. and Freedberg, K.A. (2001) Cost-Effectiveness of Earlier Initiation of Antiretroviral Therapy for Uninsured HIV-Infected Adults. *American Journal of Public Health*, **91**, 1456-1463.
<https://doi.org/10.2105/ajph.91.9.1456>
- [32] Meditz, A.L., MaWhinney, S., Allshouse, A., Feser, W., Markowitz, M., Little, S., *et al.* (2011) Sex, Race, and Geographic Region Influence Clinical Outcomes Following Primary HIV-1 Infection. *The Journal of Infectious Diseases*, **203**, 442-451.
<https://doi.org/10.1093/infdis/jiq085>
- [33] Daykin, C.D., Clark, P.N.S., Eves, M.J., Haberman, S., Le Grys, D.J., Lockyer, J., *et al.* (1988) The Impact of HIV Infection and AIDS on Insurance in the United Kingdom. *Journal of the Institute of Actuaries*, **115**, 727-837.
<https://doi.org/10.1017/s0020268100042943>
- [34] Werner, A. and Levy, J.A. (1993) Human Immunodeficiency Virus Type 1 Envelope Gp120 Is Cleaved after Incubation with Recombinant Soluble CD4. *Journal of Virology*, **67**, 2566-2574. <https://doi.org/10.1128/jvi.67.5.2566-2574.1993>
- [35] Rosenberg, Z.F. and Fauci, A.S. (1989) Minireview: Induction of Expression of HIV in Latently or Chronically Infected Cells. *New England Journal of Medicine*, **314**, 131-132.
- [36] Ausín Olivera, M.C., Wiper, M.P. and Lillo Rodríguez, R.E. (2001) Bayesian Estimation for the M/G/1 Queue Using a Phase-Type Approximation. *Journal of Statistical Planning and Inference*, **118**, 83-101.
- [37] Bharucha-Reid, A.T. (1960) On Random Solutions of Fredholm Integral Equations.

Bulletin of the American Mathematical Society, **66**, 104-109.

<https://doi.org/10.1090/s0002-9904-1960-10417-x>

- [38] Mwirigi, N., Sewe, S., Wainaina, M. and Simwa, R. (2022) Weibull Distribution as the Choice Model for State-Specific Failure Rates in HIV/AIDS Progression. *Mathematics and Statistics*, **10**, 588-602. <https://doi.org/10.13189/ms.2022.100315>
- [39] D'Amico, G., Di Biase, G., Janssen, J. and Manca, R. (2011) HIV Evolution: A Quantification of the Effects Due to Age and to Medical Progress. *Informatica*, **22**, 27-42. <https://doi.org/10.15388/informatica.2011.312>
- [40] Mwirigi, N., Simwa, P.R., Wainaina, D.M. and Sewe, D.S. (2022) Bayesian Model Averaging in Modeling of State Specific Failure Rates in HIV/AIDS Progression. *Mathematics and Statistics*, **10**, 782-798. <https://doi.org/10.13189/ms.2022.100409>

Appendix: List of Notations

The following symbols and variables are used throughout the manuscript:

List of Variables and Notations

Symbol	Description
T	Time to event (survival time, e.g., time between HIV/AIDS disease stages)
λ	Hazard rate (rate parameter in the Exponential model)
$S(t)$	Survival function (probability of surviving beyond time t)
$h(t)$	Hazard function (instantaneous rate of event occurrence at time t)
X	Covariate vector (e.g., age, gender, CD4 count, ART status)
β	Regression coefficients (effect of covariates on hazard or survival time)
HR_k	Hazard ratio for the k -th covariate in the Cox PH model
$\hat{\lambda}$	Estimated hazard rate parameter using maximum likelihood estimation (MLE)
$L(\lambda)$	Log-likelihood function for the Exponential model
$h_0(t)$	Baseline hazard function in the Cox Proportional Hazards model
ϵ_i	Error term in the Accelerated Failure Time (AFT) model
σ	Scale parameter in the AFT model, which adjusts variability in survival times
AIC	Akaike Information Criterion (goodness-of-fit measure)
BIC	Bayesian Information Criterion (goodness-of-fit measure)
$\log L$	Log-likelihood value (used to compare model fit)
Z-value	Z-value (statistic used to test the null hypothesis about the covariates)
T_i	Observed survival time for individual i
$R(t_i)$	Risk set at time t_i (individuals at risk of the event)
$H(t)$	Cumulative hazard function
$S(t)$	Survival function (probability of survival beyond time t)
$f_T(t)$	Probability density function of survival time T
Z-value (AFT)	Z-value for the Accelerated Failure Time model
Z-value (Cox)	Z-value for the Cox Proportional Hazards model
Z-value (Survival)	Z-value for general survival model