**Project Title**

# Survival Prediction in AIDS Patients Using Traditional and Machine Learning-Based Survival Models: A Comparative Study with Risk Scoring and Interactive Visualization

**Name-**

**Manthan Mahesh Mehta**

**Advisor-**

**Dr Laleh Jamshidi**

**Date-**

**05/05/2025**

**Abstract**

This thesis explores survival prediction among AIDS patients using both traditional and machine learning-based survival analysis methods. Using data from the AIDS Clinical Trials Group Study 175 (Hammer et al., 1996), we evaluate the influence of antiretroviral treatment regimens and clinical-demographic characteristics on patient survival. The study compares the predictive performance of the Cox Proportional Hazards model with machine learning approaches, including Random Survival Forests (RSF) and Gradient Boosting Machines (GBM), using the concordance index as the primary evaluation metric. While RSF demonstrated the highest predictive performance, the margin of improvement over the Cox model was minimal. Key predictors of survival such as CD8 counts, symptom status, and recent antiretroviral exposure were identified. A risk scoring system was developed based on the Cox model due to its interpretability, and an interactive R Shiny dashboard was created to visualize personalized survival estimates. This thesis underscores the potential of survival modeling for personalized prognosis in AIDS care, emphasizing both accuracy and clinical usability.

## Introduction

Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) continue to pose significant public health challenges worldwide (Deeks et al., 2015). The introduction of antiretroviral therapy (ART) in the early 1990s markedly improved the prognosis for individuals living with HIV/AIDS by suppressing viral replication and enhancing immune function (Hammer et al., 1996). However, predicting long-term survival outcomes remains a complex task, especially due to the heterogeneous nature of the disease progression and variability in treatment response across individuals (Deeks et al., 2015). Accurate survival prediction models can be instrumental in tailoring clinical decisions, guiding treatment strategies, and improving patient outcomes.

Traditional statistical approaches such as the Cox Proportional Hazards (Cox PH) model have been widely employed in survival analysis due to their interpretability and ability to model time-to-event data while accounting for covariates. Despite their effectiveness, these models rely on certain assumptions, most notably, the proportional hazards assumption—that may not always hold in clinical datasets. In contrast, machine learning-based survival models such as Random Survival Forests (RSF; Ishwaran et al., 2008) and Gradient Boosting Machines (GBM) have emerged as powerful alternatives capable of capturing complex nonlinear relationships and interactions without strict model assumptions.

This thesis explores the predictive performance of traditional and machine learning-based survival models using data from the AIDS Clinical Trials Group Study 175. The study enrolled patients diagnosed with AIDS and investigated the efficacy of different ART regimens. The dataset includes detailed clinical, demographic, and immunological variables, allowing for a comprehensive survival analysis (Hammer et al., 1996).

This thesis aims to evaluate the effects of different antiretroviral treatment regimens on patient survival using traditional survival analysis methods. It compares the predictive performance of the Cox Proportional Hazards model with machine learning models like Random Survival Forests and Gradient Boosting Machines, using the concordance index (C-index). Key predictors of survival will be identified through hazard ratios and model-derived feature importance. A patient-level risk-scoring system will be developed to estimate individualized survival probabilities. Results will be presented via an interactive R Shiny dashboard to support personalized clinical decision-making.

By integrating robust modeling techniques with interactive visual tools, this study aims to improve interpretability and utility of survival models in the context of AIDS. The findings may support clinicians in making informed, data-driven treatment decisions and provide a framework for developing similar tools in other chronic disease contexts.

**Background**

The global HIV/AIDS epidemic has transformed dramatically over the past few decades. In the early years, an HIV diagnosis was often a death sentence, with limited treatment options and poor survival outcomes(Deeks et al., 2015). The advent of antiretroviral therapy (ART), especially combination therapies introduced in the 1990s, significantly extended life expectancy and improved the quality of life for people living with HIV/AIDS (Hammer et al., 1996). Despite these advances, HIV remains a chronic condition requiring lifelong treatment and monitoring (Deeks et al., 2015). As such, predicting survival outcomes is critical for clinical decision-making, resource allocation, and personalized patient care.

Survival analysis has long been used in medical research to analyze time-to-event data, especially in the context of chronic diseases such as HIV/AIDS. The Cox Proportional Hazards (Cox PH) model, introduced in 1972, is one of the most widely used methods due to its ability to assess the effect of covariates on survival without requiring the specification of the underlying baseline hazard function (Cox, 1972). However, the Cox model assumes proportional hazards over time, which may not hold for all covariates in real-world data

(Cox, 1972). In such cases, relying solely on traditional models could result in inaccurate or biased survival estimates.

To address these limitations, researchers have increasingly turned to machine learning (ML) methods for survival prediction. Algorithms like Random Survival Forests (RSF) are capable of modeling complex, nonlinear interactions and do not rely on the proportional hazards assumption (Ishwaran et al., 2008).Gradient Boosting Machines (GBM) have also emerged as powerful alternatives, offering strong predictive performance, particularly in high-dimensional datasets, and uncovering hidden patterns that traditional models might overlook (Wang et al., 2019).

At the same time, there has been a growing interest in developing interpretable and accessible tools to visualize survival outcomes for individual patients. Risk prediction tools that integrate both traditional and machine learning (ML)-based methods, combined with interactive dashboards, can empower clinicians and researchers to explore survival probabilities dynamically based on patient-specific characteristics (Stiglic et al., 2020; Wang et al., 2019).

This study is based on data from the AIDS Clinical Trials Group (ACTG) Study 175, a landmark clinical trial conducted in the early 1990s to evaluate the efficacy of four antiretroviral regimens. The dataset includes a range of demographic, clinical, and laboratory variables collected from patients with HIV/AIDS, providing a valuable resource for evaluating survival outcomes under different treatment strategies (Hammer et al., 1996).

By combining classical survival modeling with modern ML techniques and interactive visual tools, this thesis aims to provide both predictive insights and practical solutions for improving survival prediction in AIDS patients.

## Methods

### Study Design and Data Source

This study utilized data from the AIDS Clinical Trials Group (ACTG) Study 175, a multicenter, randomized clinical trial designed to compare the effectiveness of four antiretroviral treatment regimens in patients with HIV infection. The dataset comprises 2,139 adult patients with follow-up data on survival and various baseline and follow-up characteristics (Hammer et al., 1996).

The dataset was complete and well-structured, with no missing values or inconsistencies; therefore, no data cleaning or preprocessing was required prior to analysis.

The primary outcome of interest was time to event, defined as the time (in days) from the start of treatment to death or censoring. The censoring indicator (cid) was coded as 1 if death occurred (event) and 0 if no event occurred by the end of follow-up (censored). Predictor variables included both demographic and clinical factors. Demographic variables were age, gender, and race. Clinical variables included the Karnofsky score (karnof), CD8 cell counts at baseline (cd80) and at 20±5 weeks (cd820), symptomatic status (symptom), and recent ZDV treatment (z30), which indicated whether the patient had received zidovudine within the 30 days prior to the study start (0 = no, 1 = yes). Treatment-related factors included the treatment group (trt), which categorized patients into four regimens: 0 = zidovudine (ZDV) only, 1 = zidovudine plus didanosine (ZDV + ddI), 2 = zidovudine plus zalcitabine (ZDV + Zal), and 3 = didanosine (ddI) only. Early treatment discontinuation (offtrt) indicated whether the patient stopped the assigned treatment before 96 ± 5 weeks (0 = no, 1 = yes). Both trt and offtrt were used as stratification variables due to the violation of the proportional hazards assumption in the Cox model.

Three models were used to compare predictive performance and derive individual-level risk scores. First, the Cox Proportional Hazards model was applied to assess the effect of covariates on the hazard of death. The final Cox model included only statistically significant predictors, with treatment group (trt) and off-treatment indicator (offtrt) incorporated as stratification variables to account for violations of the proportional hazards assumption, which were identified using Schoenfeld residuals. Model performance was evaluated using the Concordance Index (C-index).

Second, the Random Survival Forest (RSF) model was applied. RSF is a tree-based ensemble method specifically adapted for survival data. It can capture complex nonlinear relationships and interactions among covariates. The model was trained using 1000 trees, and variable importance was computed. The C-index was again used to assess predictive performance.

Third, a Gradient Boosting Model (GBM) for survival analysis was implemented using the gbm package with the CoxPH loss function. Cross-validation was employed to determine the optimal number of trees (best_iter), and variable importance scores were extracted. The GBM model's predictive performance was also evaluated using the C-index calculated from predicted linear predictors.

**Model Evaluation**

To compare model performance, the Concordance Index (C-index) was computed for all three models. The C-index measures the probability that, for a randomly selected pair of

patients, the one with the higher predicted risk experiences the event (e.g., death) before the other. It reflects the model's ability to correctly rank patients by risk. Higher values (closer to 1) indicate better discriminative performance.

**Risk Prediction and Visualization**

A risk scoring system was developed based on the Cox model. An interactive R Shiny dashboard was developed to allow users to input individual patient characteristics and visualize personalized survival outcomes based on the fitted Cox proportional hazards model. The input variables include age, Karnofsky score, CD8 cell counts at baseline and at approximately 20 weeks, symptomatic status, recent ZDV exposure (within 30 days), assigned treatment group, and early treatment discontinuation status. Based on these inputs, the dashboard generates a linear predictor (risk score), which represents the raw log-hazard value from the Cox model. Higher scores indicate greater risk of death (shorter survival time). Additionally, the dashboard displays a personalized survival curve with confidence bands, allowing clinicians and researchers to dynamically explore how patient-specific factors influence survival probabilities over time.

**Software and Statistical Tools**

All statistical analyses were conducted using a combination of R (version 4.3.1) and SAS (version 9.4) to ensure robust and cross-validated findings. In SAS, procedures such as PROC IMPORT were used to load the dataset from Excel, while PROC CONTENTS and PROC MEANS facilitated data exploration, descriptive statistics, and identification of missing values. Kaplan-Meier survival analysis and group comparisons were performed using PROC LIFETEST with STRATA and TEST=LOGRANK options. In R, survival analysis and machine learning modeling were carried out using the survival and survminer packages for Kaplan-Meier curves and Cox Proportional Hazards modeling, with the cox.zph() function employed to test the proportional hazards assumption. The randomForestSRC package was used to implement Random Survival Forests (RSF), and the gbm package was used with the coxph loss function to develop Gradient Boosting Machine (GBM) models. Data visualizations were generated using ggplot2, and an interactive risk prediction tool was developed using the shiny package. This hybrid analytical approach leveraged the strengths of SAS for classical statistical procedures and R for flexible, modern machine learning and interactive visualization.

**Findings:**

**Kaplan-Meier Survival Analysis**

To explore survival differences across key baseline characteristics, Kaplan-Meier survival curves were generated and stratified by categorical variables. The log-rank test was used to assess whether the survival distributions significantly differed between the groups.

**Kaplan-Meier Survival Analysis by Treatment Group**

**Figure 1** displays Kaplan-Meier curves comparing survival across four treatment groups. Patients receiving ZDV only (group 0) had the lowest survival probability over time, while those on ZDV + ddI (1), ZDV + Zal (2), or ddI only (3) showed better outcomes. The log-rank test ($p < 0.0001$) confirms a statistically significant difference in survival, suggesting that combination therapies or ddI monotherapy are more effective than ZDV monotherapy.

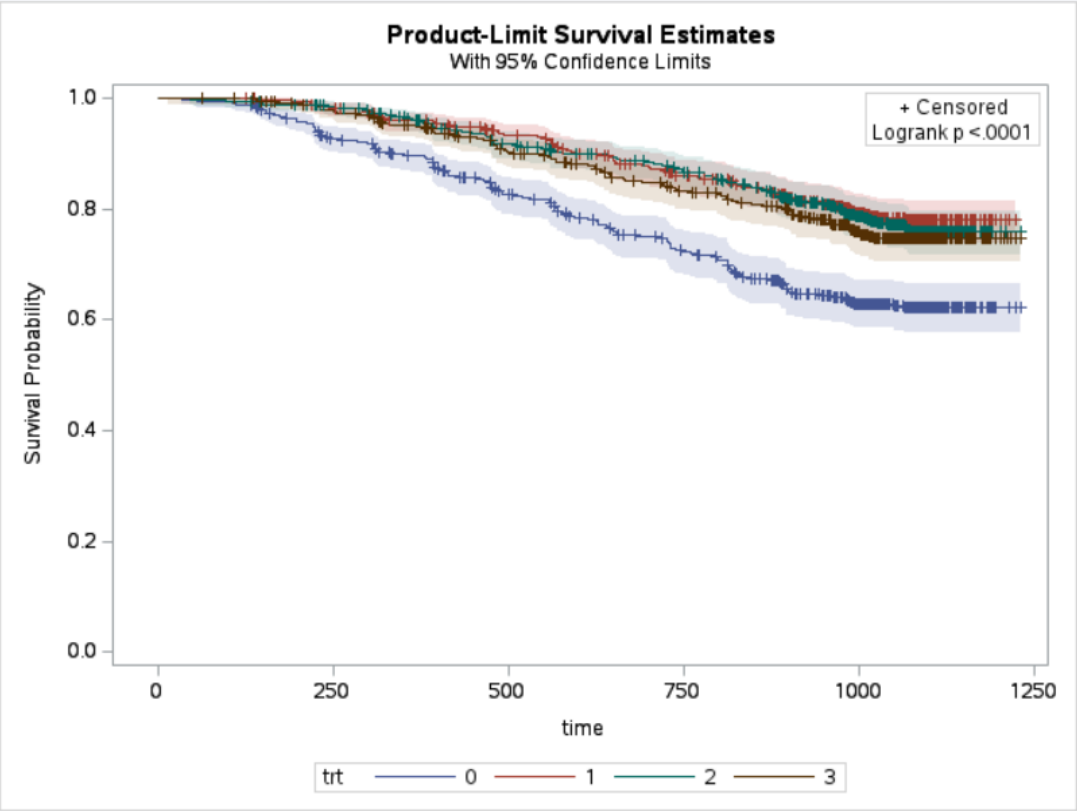


***Figure 1.*** *Kaplan-Meier survival curves by treatment group. The shaded areas represent 95% confidence intervals. "+" indicates censored observations. Group 0: ZDV only, Group 1: ZDV + ddI, Group 2: ZDV + Zal, Group 3: ddI only.*

### Kaplan-Meier Survival Analysis by Gender

**Figure 2** illustrates Kaplan-Meier survival curves stratified by gender. While males (coded as 1, red) show slightly lower survival probabilities compared to females (coded as 0, blue), the difference is not statistically significant (log-rank $p = 0.0857$). This suggests that gender may not be a strong independent predictor of survival in this cohort.

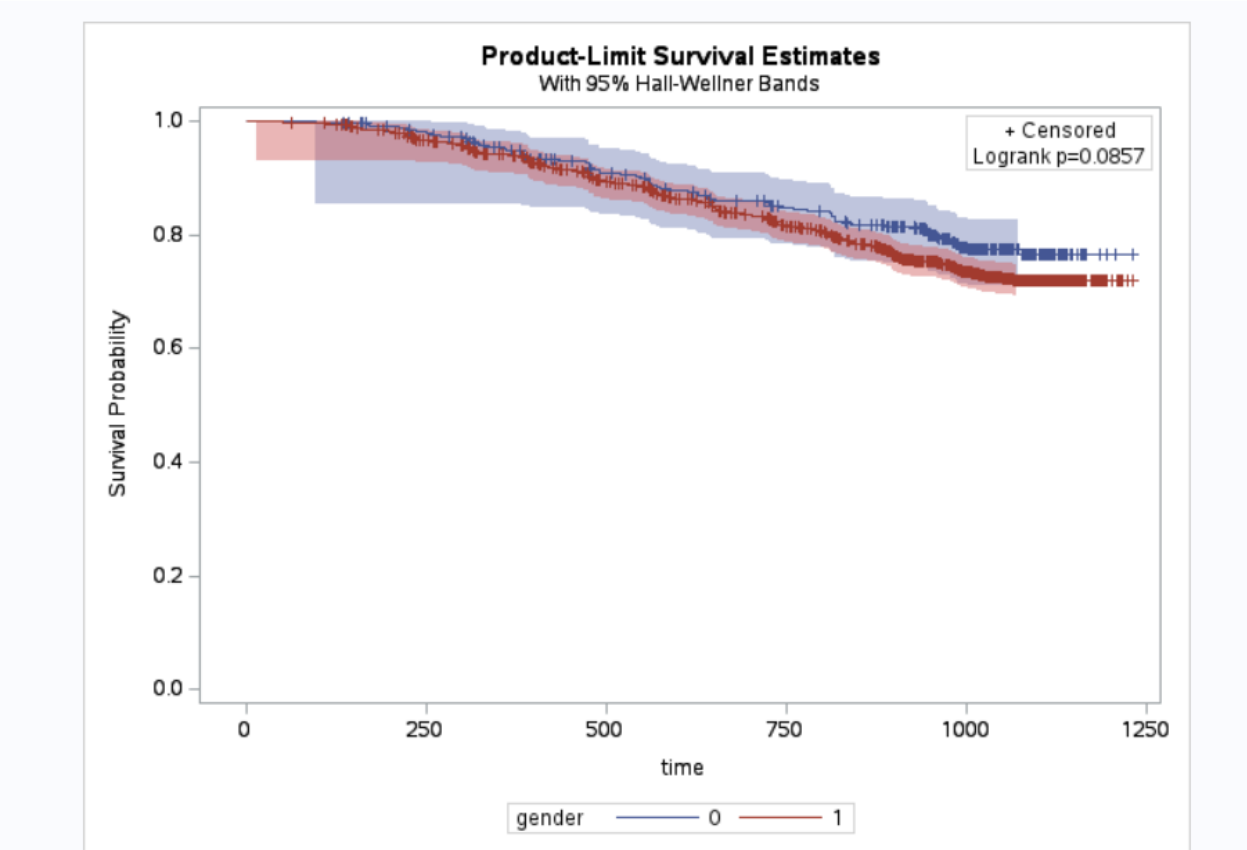


***Figure 2.*** *Kaplan-Meier survival curves by gender. The shaded areas represent 95% Hall-Wellner confidence bands. "+" indicates censored observations. Gender 0: Female, Gender 1: Male.*

### Kaplan-Meier Survival Analysis by Treatment Discontinuation

**Figure 3** displays survival curves stratified by whether patients discontinued treatment early (offtrt). Patients who stopped treatment prematurely (offtrt = 1, red) had significantly lower survival probabilities compared to those who continued treatment (offtrt = 0, blue), with a highly significant log-rank $p < 0.0001$. This suggests treatment adherence plays a critical role in patient survival.

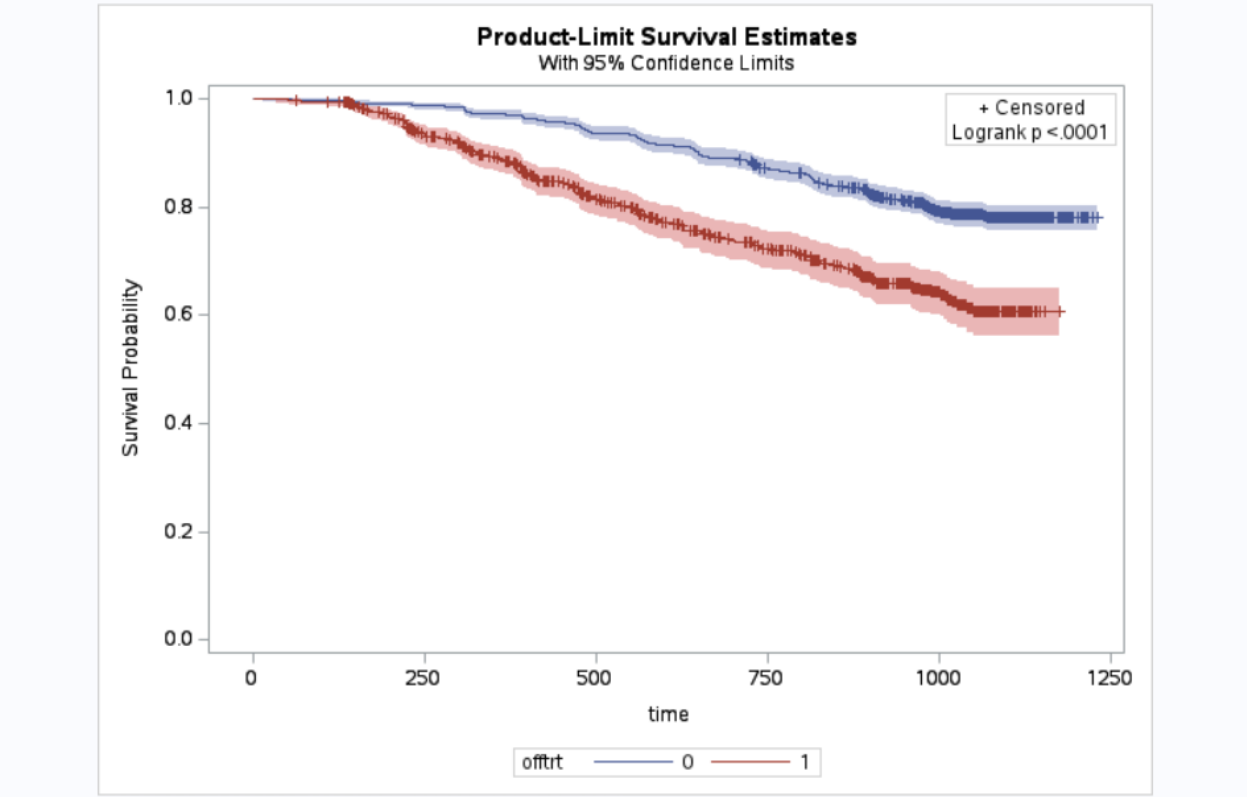


***Figure 3.*** *Kaplan-Meier survival curves by early treatment discontinuation (offtrt). Shaded areas represent 95% confidence limits. "+" indicates censored cases. offtrt 0: Continued treatment, offtrt 1: Discontinued early.*

### Kaplan-Meier Survival Analysis by Race

**Figure** 4 shows the survival curves stratified by race (0 = White, 1 = Non-White). While non-white patients (red) appear to have slightly higher survival probabilities over time compared to white patients (blue), the difference is not statistically significant (log-rank $p = 0.0619$). This suggests a possible trend, but not enough evidence to conclude racial differences in survival within this dataset.

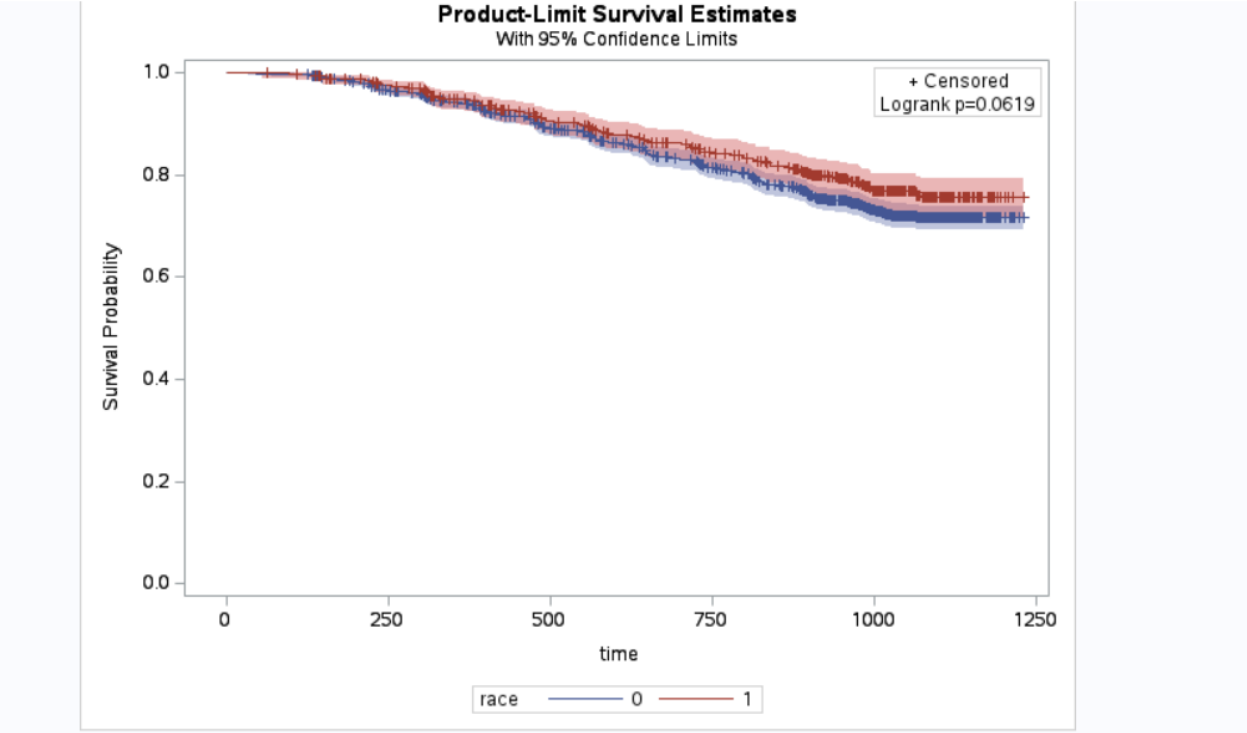


***Figure 4.*** *Kaplan-Meier survival curves by race. Shaded areas represent 95% confidence limits. "+" indicates censored observations. Race 0: White, Race 1: Non-White.*

### Kaplan-Meier Survival Analysis by Recent ZDV Use (z30)

**Figure 5** illustrates survival curves based on recent ZDV use within the past 30 days (z30: 0 = No, 1 = Yes). Patients who had ZDV exposure in the last 30 days (red) experienced significantly lower survival probabilities compared to those who did not (blue), as indicated by the log-rank $p < 0.0001$. This suggests a potential negative association between recent ZDV use and survival outcomes.

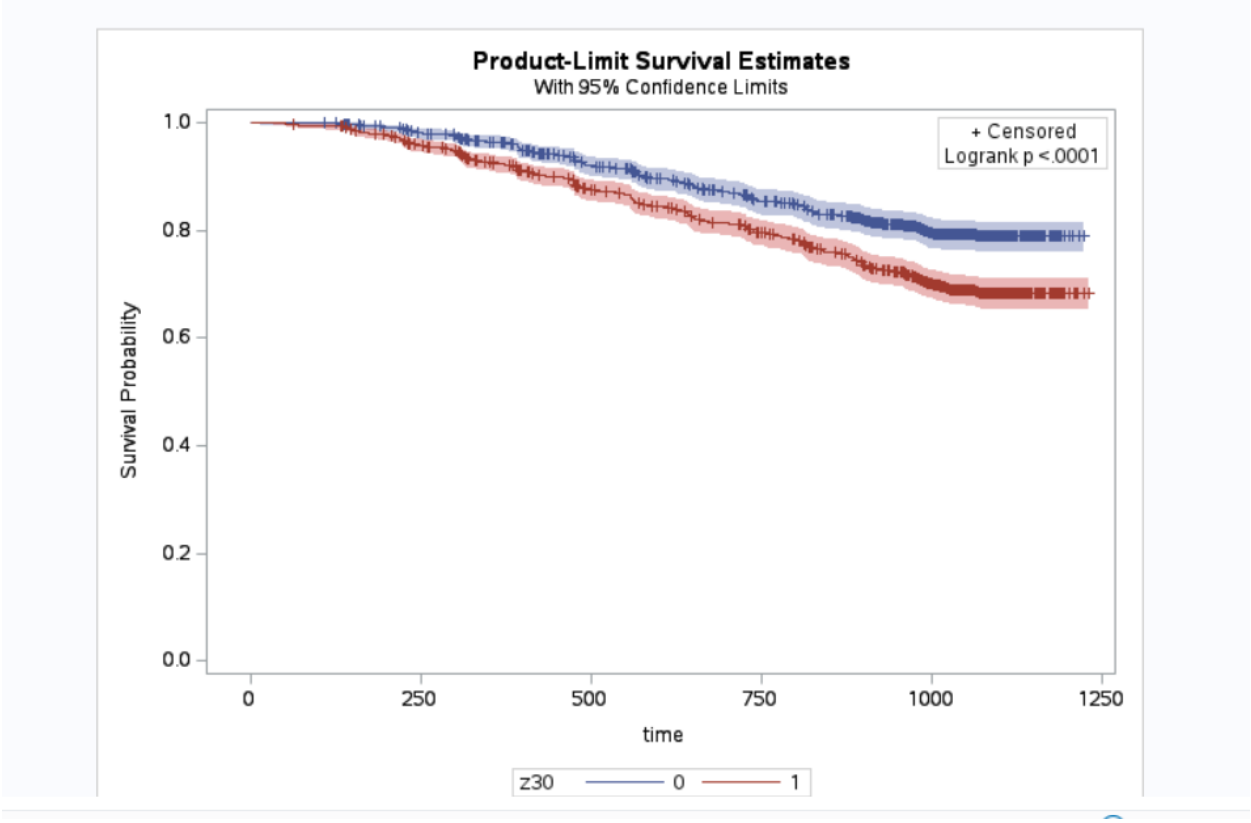


***Figure 5.*** *Kaplan-Meier survival curves stratified by z30 (ZDV use in the past 30 days). Shaded regions represent 95% confidence intervals. "+" indicates censored data. z30 = 1 (Yes), z30 = 0 (No)*

### Kaplan-Meier Survival Analysis by Antiretroviral History (str2)

**Figure 6** presents Kaplan-Meier survival curves comparing patients with no prior antiretroviral therapy (str2 = 0) to those with treatment experience (str2 = 1). Patients who were antiretroviral-naive had significantly better survival than experienced patients ($p < 0.0001$). This suggests that prior antiretroviral exposure may be associated with reduced survival.

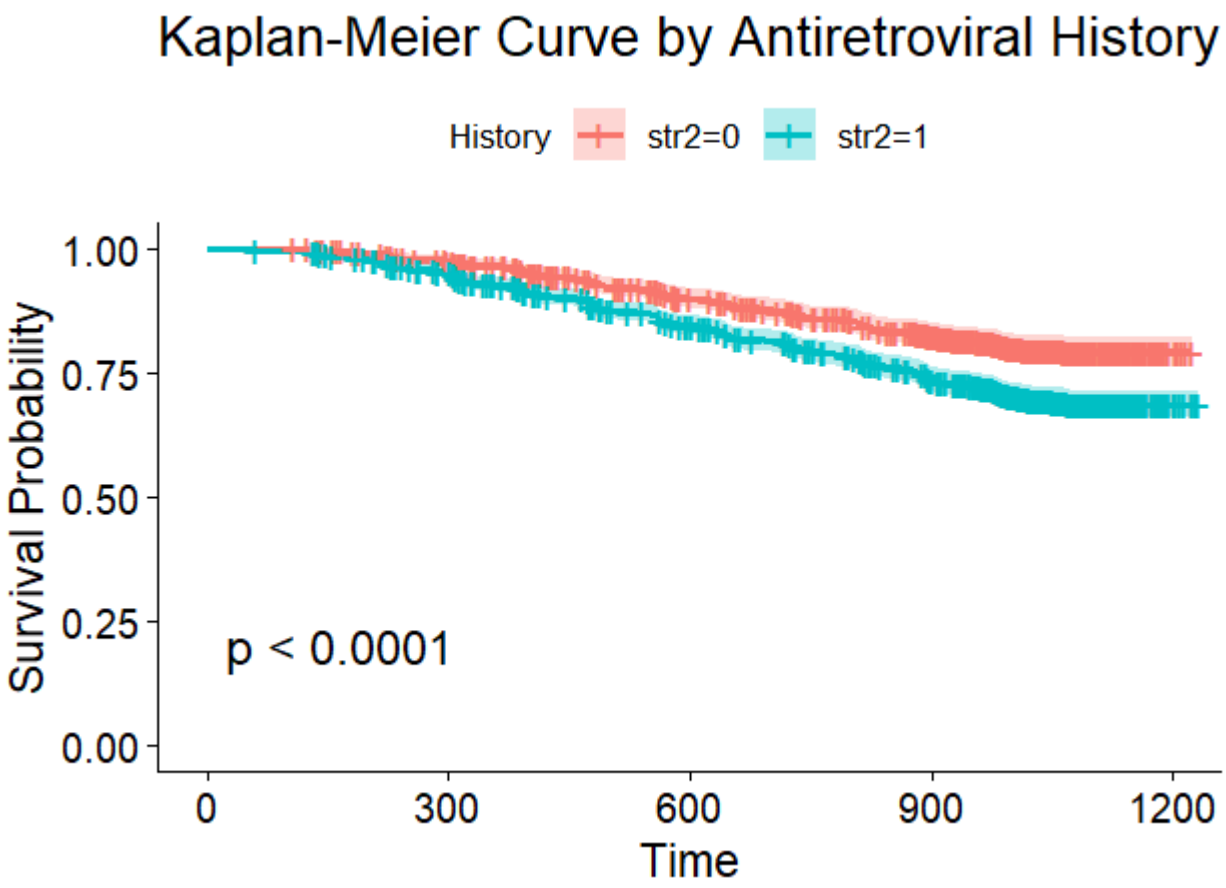


***Figure 6.*** *Kaplan-Meier survival curves stratified by antiretroviral history (str2). Shaded areas represent 95% confidence intervals. "+" indicates censored observations. str2 = 0 (Naive), str2 = 1 (Experienced).*

**Kaplan-Meier Survival Analysis by Symptom Status**

**Figure 7** displays Kaplan-Meier curves comparing survival between asymptomatic (symptom = 0) and symptomatic (symptom = 1) patients. The survival probability was significantly higher in asymptomatic patients at baseline. The log-rank test yielded a p-value $p < 0.0001$, indicating a strong association between symptom status and survival outcome.

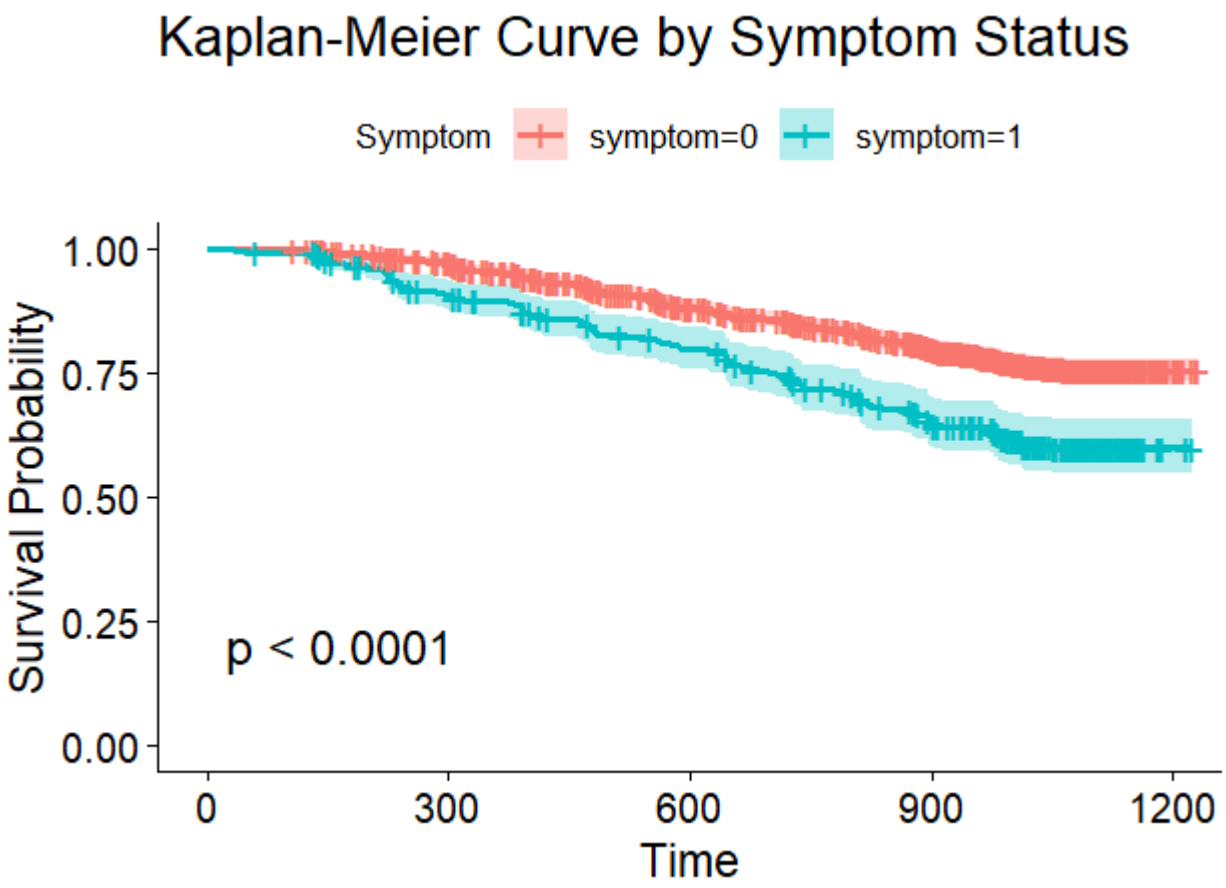


***Figure 7.*** *Kaplan-Meier survival curves by symptom status. Shaded regions represent 95% confidence intervals. "+" symbols indicate censored observations. symptom = 0 (Asymptomatic), symptom = 1 (Symptomatic)*

**Kaplan-Meier Survival Analysis by Prior Non-ZDV Therapy**

**Figure 8** presents Kaplan-Meier curves comparing patients with (opriori = 1) and without (opriori = 0) prior non-ZDV antiretroviral therapy. Patients without prior therapy showed slightly higher survival probabilities throughout the study period. However, the log-rank test yielded a p-value = 0.067, indicating that this difference was not statistically significant at the conventional 0.05 level

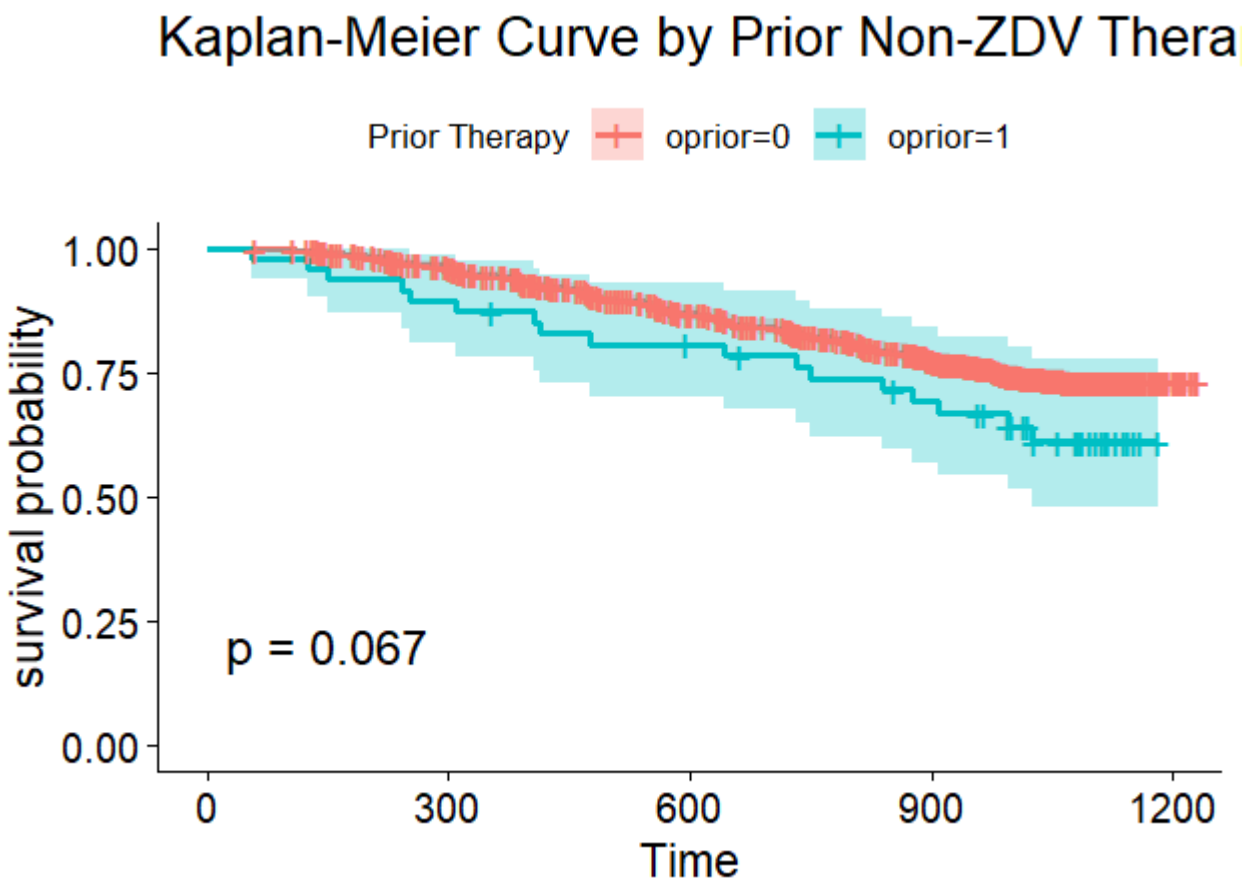


***Figure 8.*** *Kaplan-Meier survival curves by prior non-ZDV therapy (oprior). Shaded areas represent 95% confidence intervals. "+" indicates censored observations. oprior = 0 (No prior therapy), oprior = 1 (Received prior non-ZDV therapy).*

**Assessing Proportional Hazards Assumption Using Schoenfeld Residuals**

Following the Kaplan-Meier survival analyses, which provided preliminary insights into survival differences across key patient subgroups, we assessed the proportional hazards (PH) assumption—a fundamental requirement for Cox regression—using Schoenfeld residuals.

The global test was statistically significant ($p < 0.001$), indicating that at least one variable violated the proportional hazards (PH) assumption. Among the individual predictors, CD420 (CD4 count at approximately 20 weeks) showed the strongest violation ($p < 2e\text{-}16$), followed by CD40 (baseline CD4 count) and offtrt (early treatment discontinuation) with p-values of 0.00029 and 2.3e-05, respectively. Although the treatment group (trt) did not cross the conventional significance threshold ($p = 0.05007$), the borderline value warranted caution in interpretation.

Given these findings and the graphical evidence of non-proportionality from Kaplan-Meier curves, we addressed potential violations by stratifying on both trt and offtrt in the Cox model. Stratification allowed the baseline hazard functions for these variables to vary without estimating their coefficients, thus maintaining model validity while preserving the influence of other covariates

**Sensitivity Analysis: Collinearity Between str2 and z30**

Given the high correlation between z30 (ZDV use in the last 30 days) and str2 (antiretroviral history) *(r* = 0.90), we conducted a sensitivity analysis to assess potential collinearity. Two separate Cox proportional hazards models were constructed—one including z30 and the other str2—each adjusted for relevant covariates and stratified by trt and offtrt due to violations of the proportional hazards assumption.

Both z30 and str2 were statistically significant in their respective models ($p < 0.001$). However, the model containing z30 showed marginally better performance, with a higher concordance index (C-index = 0.614 vs. 0.613) and a slightly stronger model fit as indicated by the Likelihood Ratio Test ( 96.45, $p < 0.0001$ for z30 model vs. 94.55, $p < 0.0001$ for str2 model). Based on these results and clinical interpretability, z30 was retained in the final model.

**Cox Proportional Hazards Model :**

**Justification for Predictor Inclusion**

The predictors included in the final Cox proportional hazards model were selected based on multiple criteria. First, subjective judgment of importance was applied by including variables considered potentially relevant to survival outcomes based on reasoning and contextual understanding of the dataset. Second, the Cox model assumptions were assessed, and only predictors that did not violate the proportional hazards assumption, as evaluated by the Schoenfeld residual test, were retained. Third, for variables such as trt and offtrt that violated the proportional hazards assumption, stratification was used to address the issue. Lastly, a multicollinearity check was performed; due to a high correlation *(r* = 0.90) between z30 and str2, a sensitivity analysis was conducted, and z30 was retained based on better model performance.

**Interpretation Table for Cox Model Predictors**

***Table 1****Results of the final Cox proportional hazards model assessing the association between selected predictors and survival in AIDS patients. Hazard ratios (HR), 95% confidence intervals (CI), and p-values are reported. A hazard ratio >1 indicates increased risk, while <1 indicates reduced risk of the event. Variables trt and offtrt were included as stratification factors due to violation of the proportional hazards assumption.*

| | | | |
|---|---|---|---|
| **Predictor** | **HR (exp(coef))** | **95% CI (Lower – Upper)** | **p-value** |
| **Age** | 1.011 | 1.001 – 1.021 | 0.029 |
| **Weight (wtkg)** | 1.001 | 0.994 – 1.008 | 0.839 |
| **Karnofsky Score** | 0.981 | 0.967 – 0.994 | 0.006 |
| **Race** | 0.968 | 0.783 – 1.196 | 0.761 |
| **Gender** | 1.188 | 0.912 – 1.548 | 0.202 |
| **CD8 at baseline (cd80)** | 1.0006 | 1.0003 – 1.0008 | <0.001 |
| **CD8 at 20 weeks (cd820)** | 0.9995 | 0.9992 – 0.9998 | 0.002 |
| **Symptom status** | 1.675 | 1.370 – 2.047 | <0.001 |
| **Prior Non-ZDV therapy (oprior)** | 1.301 | 0.795 – 2.127 | 0.295 |
| **ZDV in last 30 days (z30)** | 1.593 | 1.325 – 1.915 | <0.001 |

Table 1 presents the results of the Cox proportional hazards model. Age was a statistically significant predictor, with each additional year associated with a 1.1% increase in the hazard of death (*HR* = 1.011, *p* = 0.029). The Karnofsky score was protective; each one-point increase corresponded to a ~2% reduction in hazard (*HR* = 0.981, *p* = 0.006). Baseline CD8 count (cd80) showed a very small but statistically significant increase in hazard, while CD8 count at 20 weeks (cd820) was slightly protective (*HR* = 0.9995, *p* = 0.002). Symptomatic patients had a 67.5% higher hazard of death compared to asymptomatic patients *(HR* = 1.675, *p* < 0.001). Recent ZDV use within the last 30 days (z30) was also associated with increased hazard *(HR* = 1.593, *p* < 0.001), suggesting a link with poorer survival outcomes.

In contrast, weight (wtkg), race, gender, prior non-ZDV therapy (oprior), and symptom status were not statistically significant predictors (*p* > 0.05), indicating no strong evidence of their independent association with survival in this model.

**Random Survival Forest (RSF) Analysis**

Two Random Survival Forest (RSF) models were trained: one using all 14 predictors and another using only the top 11 predictors, after removing the bottom three predictors (race, z30, and gender) based on their low variable importance scores. Variable importance plots for both models are shown in Figure 9 and Figure 10. In Figure 9 (full model), the top predictors were CD420 (CD4 at 20 weeks), Karnofsky score, and CD40 (CD4 at baseline), while predictors such as race, z30, and gender contributed near-zero importance. In Figure 10 (reduced model), the ranking of influential predictors shifted slightly; for example, offtrt

moved ahead of trt, and age became more prominent, suggesting that the removal of low-importance variables can affect the relative importance ranking of others by reducing noise and inter-variable competition during tree splits. Model performance also showed a modest improvement: the Out-of-Bag (OOB) Continuous Ranked Probability Score (CRPS) decreased from 82.40 to 82.17, indicating improved probabilistic accuracy; the standardized CRPS improved from 0.07680 to 0.07658, reflecting better normalized prediction performance; and the prediction error slightly decreased from 0.2440 to 0.2425, meaning the model's average error in predicting survival outcomes became marginally lower. These results indicate that simplifying the model by removing negligible predictors can slightly improve performance and sharpen variable prioritization.

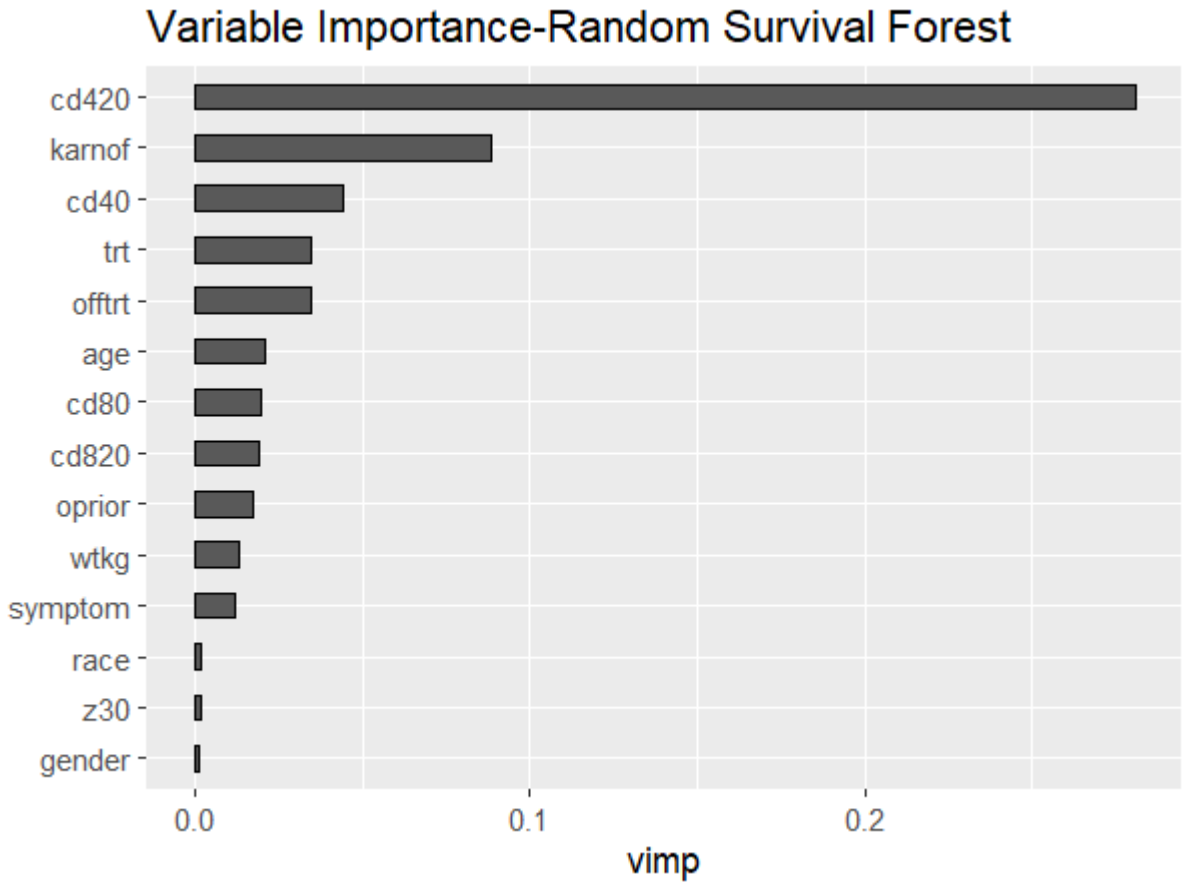


***Figure 9.*** *Variable importance plot from the full Random Survival Forest (RSF) model using all 14 predictors. CD4 count at 20 weeks (cd420), Karnofsky score (karnof), and baseline CD4 count (cd40) emerged as the most influential variables. Predictors such as race, z30 (ZDV use in the last 30 days), and gender contributed minimally to survival prediction.*

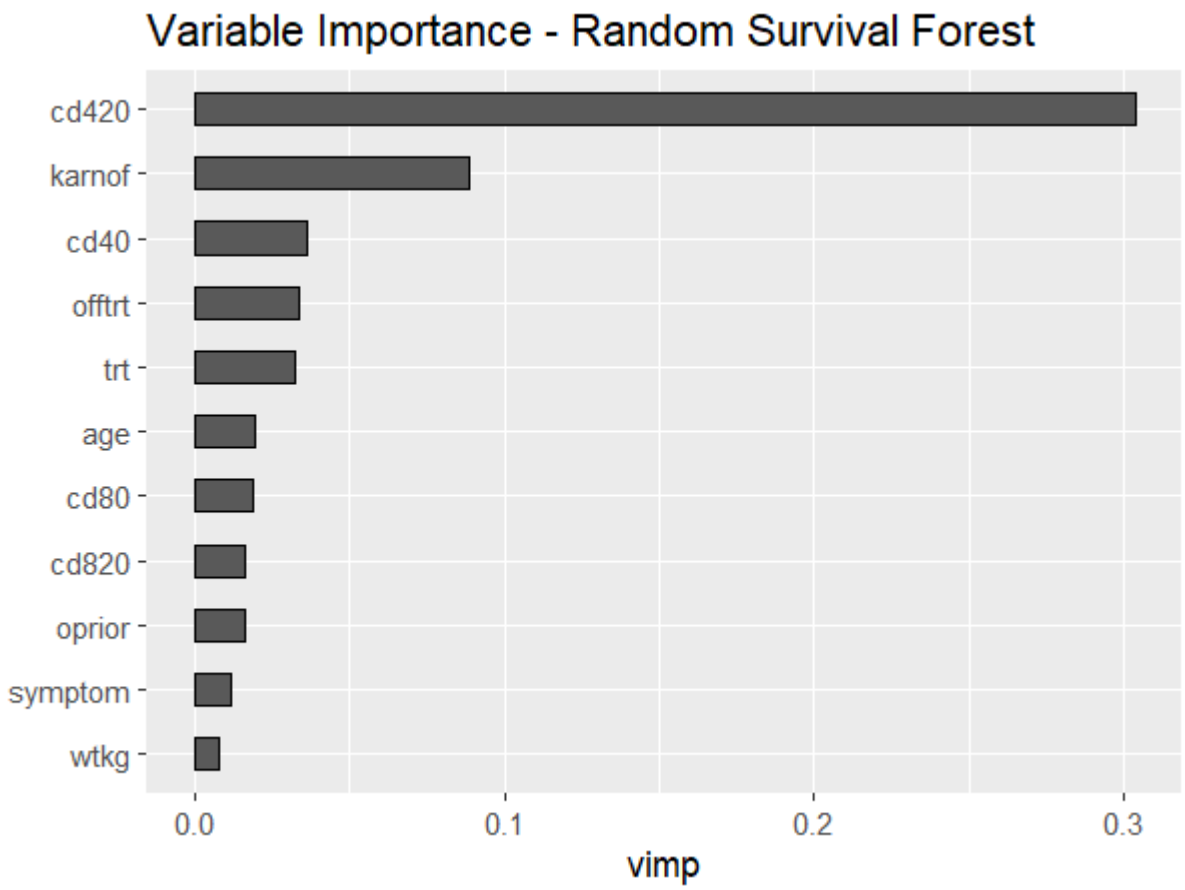


***Figure 10.*** *Variable Importance Plot for the Reduced Random Survival Forest (RSF) Model.*
*This figure displays the top 11 predictors based on variable importance in the reduced RSF model after removing gender, race, and z30 due to negligible contribution. CD420 (CD4 count at ~20 weeks) remains the most influential predictor, followed by Karnofsky score and baseline CD4 count. The ranking of some variables (e.g., offtrt, age) slightly shifted compared to the full model, highlighting changes in predictor interactions and relevance after model simplification*

**Interpretable Survival Tree from RSF**

To enhance interpretability of the Random Survival Forest (RSF) model, a representative survival tree was visualized using the most influential predictors identified through variable importance analysis. It is important to note that this is not a true decision tree used for prediction, but rather a single tree extracted from the RSF ensemble, constructed for illustrative purposes to showcase how key variables contribute to patient stratification.

As shown in Figure 11, the tree begins with a split on CD4 count at 20 weeks (cd420), highlighting its dominant influence on survival outcomes. Subsequent splits occur on variables such as age, off-treatment status (offtrt), CD8 at 20 weeks (cd820), and treatment group (trt). Each split represents a decision threshold that separates patients into subgroups with distinct risk profiles.

At the bottom of each terminal node, Kaplan-Meier survival curves depict the survival probabilities for patients in that subgroup. These curves provide insight into how combinations of clinical characteristics influence long-term survival. For example, patients with lower cd420 values and older age tended to cluster into nodes with poorer survival curves.

This tree serves as a helpful interpretive tool, offering a simplified glimpse into the more complex structure of the full RSF model, which comprises hundreds of such trees.

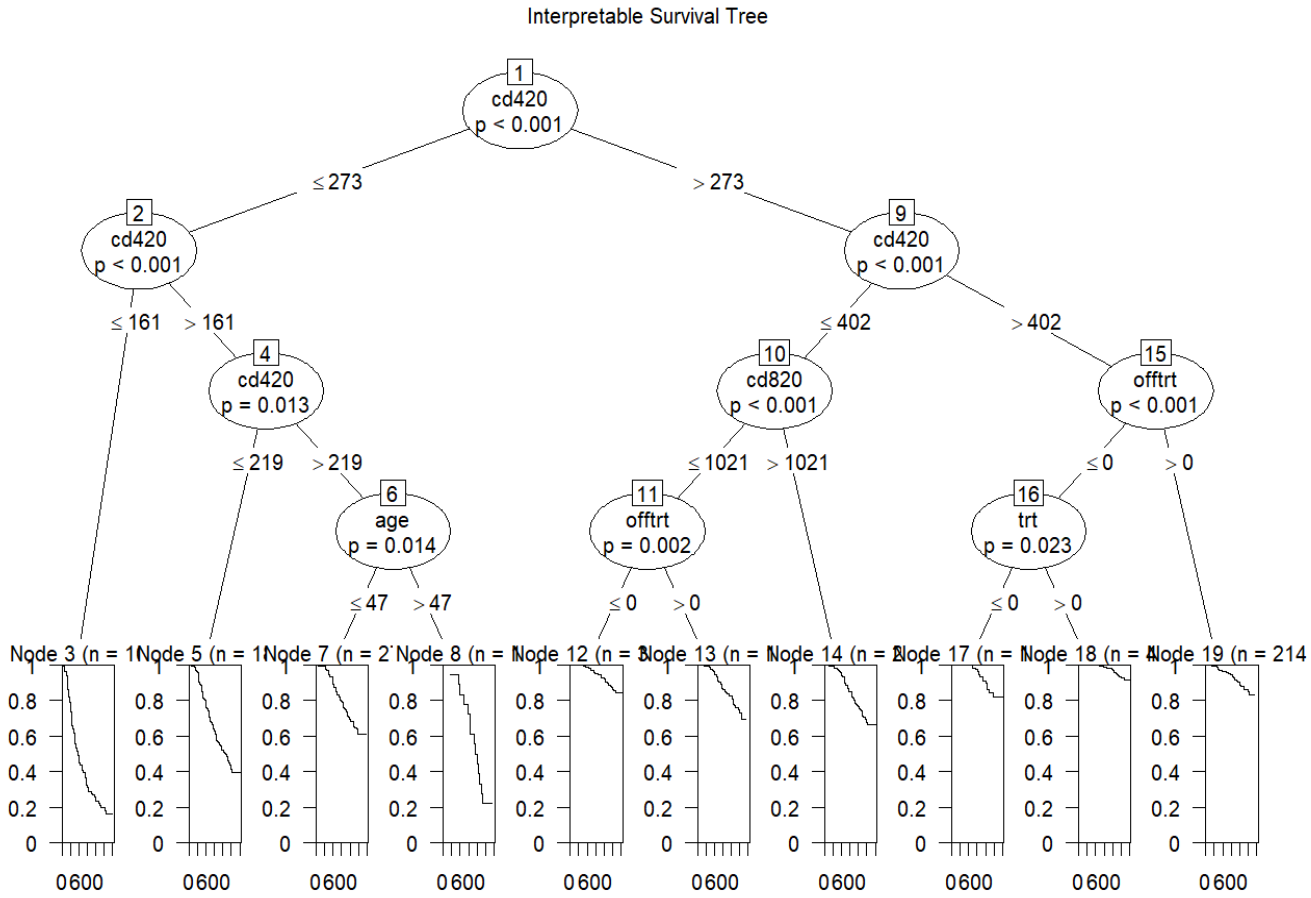


***Figure 11.*** *Illustrative survival tree extracted from the Random Survival Forest (RSF) model.*
*This tree highlights key predictors identified by RSF, including CD4 count at 20 weeks (cd420), age, CD8 at 20 weeks (cd820), off-treatment status (offtrt), and treatment group (trt). Each internal node displays the splitting variable and corresponding p-value, while terminal nodes show Kaplan-Meier survival curves for patient subgroups defined by the split path. This is not a true predictive tree but a representative visualization of how RSF partitions the data based on variable importance*.

**Gradient Boosting Model (GBM)**

Gradient Boosting is an ensemble learning technique that builds a strong predictive model by combining multiple weak learners—typically decision trees—in a sequential manner Each new tree is constructed to correct the residual errors of the previous ensemble, gradually improving prediction accuracy (Friedman et al., 2001).

For survival analysis, Gradient Boosting Machines (GBM) adapt to censored data by optimizing a loss function appropriate for time-to-event outcomes, such as the partial likelihood from the Cox model. This makes GBM capable of modeling complex non-linear relationships and interactions among predictors without the need for proportional hazards assumptions (Wang et al., 2019).

We applied GBM to evaluate its predictive utility for AIDS patient survival and to assess the relative influence of clinical and demographic variables on risk.

**Gradient Boosting Model (GBM) Tuning**

To determine the optimal number of trees (iterations) for the GBM, we evaluated the model's Cox partial deviance — a measure of model error — on both training and validation sets across different boosting iterations.

As shown in Figure 12, the training error (black line) continues to decrease, while the validation error (green line) initially decreases and then begins to rise after a point, indicating potential overfitting beyond that stage.

The vertical dashed blue line marks the point of minimum validation deviance, which occurs around 500 iterations. This point was selected as the optimal number of trees for our final GBM model to balance performance and avoid overfitting.

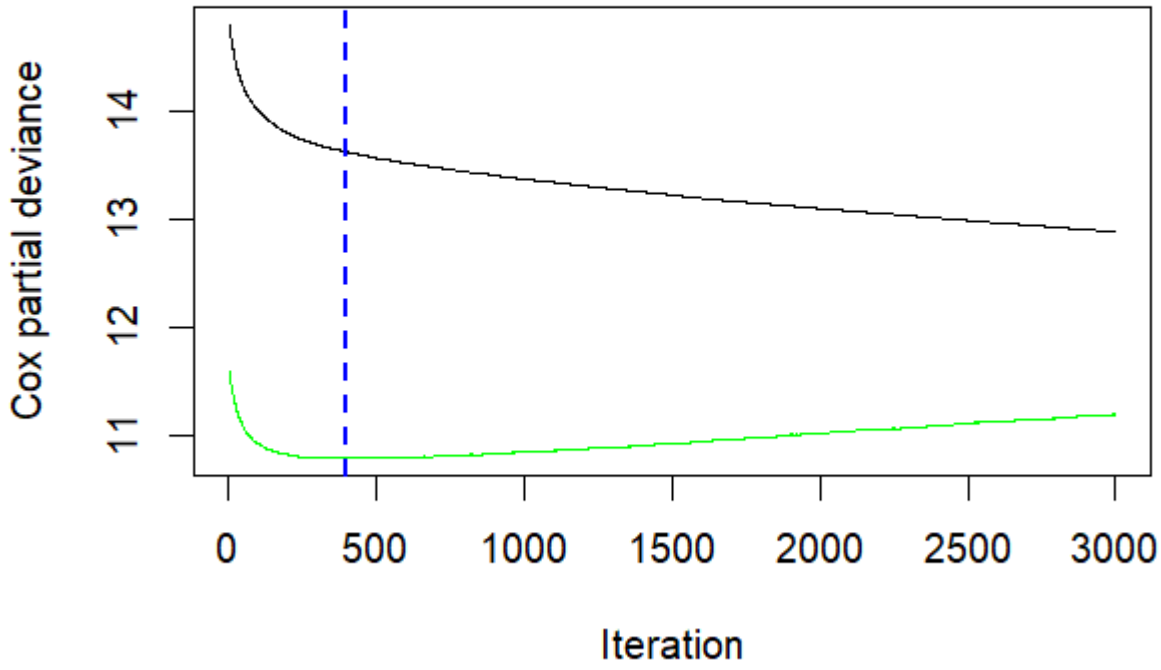


***Figure 12*** *Cox partial deviance plotted against boosting iterations in the GBM model. The green curve represents validation error, the black curve is training error, and the vertical dashed line indicates the optimal number of iterations (~500) selected to minimize overfitting and maximize generalization.*

**Variable Importance in Gradient Boosting Model**

After identifying the optimal number of trees (~500) using Cox partial deviance (Figure 12), the relative influence of each predictor was extracted from the trained Gradient Boosting Model (GBM).

The variable importance results are summarized in Table 2, where CD420 (CD4 count at ~20 weeks) demonstrated the highest influence, contributing approximately 59.4% to the model's predictive power.

Following CD420, other key predictors included CD820, CD40, age, treatment group (trt), and weight (wtkg). In contrast, predictors such as gender, race, z30, and oprior exhibited minimal impact.

These results demonstrate GBM's capability to capture complex interactions while effectively prioritizing the most informative variables.

***Table 2****Relative influence of predictors in the Gradient Boosting Model (n.trees = 500). CD420 had the highest predictive contribution, while gender and race were the least influential.*

| Predictor | Relative Influence (%) |
|---|---|
| cd420 | 59.43 |
| cd820 | 8.18 |
| cd40 | 7.63 |
| age | 5.58 |
| trt | 4.27 |
| wtkg | 4.26 |
| offtrt | 3.69 |
| cd80 | 3.59 |
| symptom | 2.03 |
| karnof | 0.54 |
| z30 | 0.50 |
| oprior | 0.24 |
| race | 0.04 |
| gender | 0.03 |

**Model Performance Comparison Based on Concordance Index (C-index):**

The Concordance Index (C-index) values for the models were as follows: Random Survival Forest (RSF) achieved a C-index of 0.616, the Cox Proportional Hazards model achieved 0.614, and the Gradient Boosting Model (GBM) achieved 0.581. These results indicate that RSF provided the highest concordance, suggesting slightly better predictive discrimination compared to the Cox model and GBM. However, the difference between RSF and Cox was minimal, while GBM demonstrated comparatively lower performance in this dataset.

**Interactive Risk Prediction Tool for AIDS Dataset Using R Shiny**

To enhance clinical interpretability and enable individualized risk assessment, an interactive R Shiny dashboard was developed based on the Cox Proportional Hazards model. While the Random Survival Forest model showed marginally higher predictive performance (C-index: 0.616) compared to the Cox model (C-index: 0.614), the Cox model was selected for the dashboard due to its greater transparency and interpretability. It provides direct estimates of hazard ratios and supports stratification, making it well-suited for clinical decision-making.

The dashboard allows users to input patient-specific values for significant predictors such as age, Karnofsky score, CD8 counts, symptom status, and recent ZDV use (z30). It

dynamically generates a linear risk score (linear predictor) representing the relative hazard and a personalized Kaplan-Meier survival curve estimating survival probability over time. This application bridges statistical modeling with practical usability and can assist researchers and clinicians in interpreting survival outcomes based on individual patient profiles.

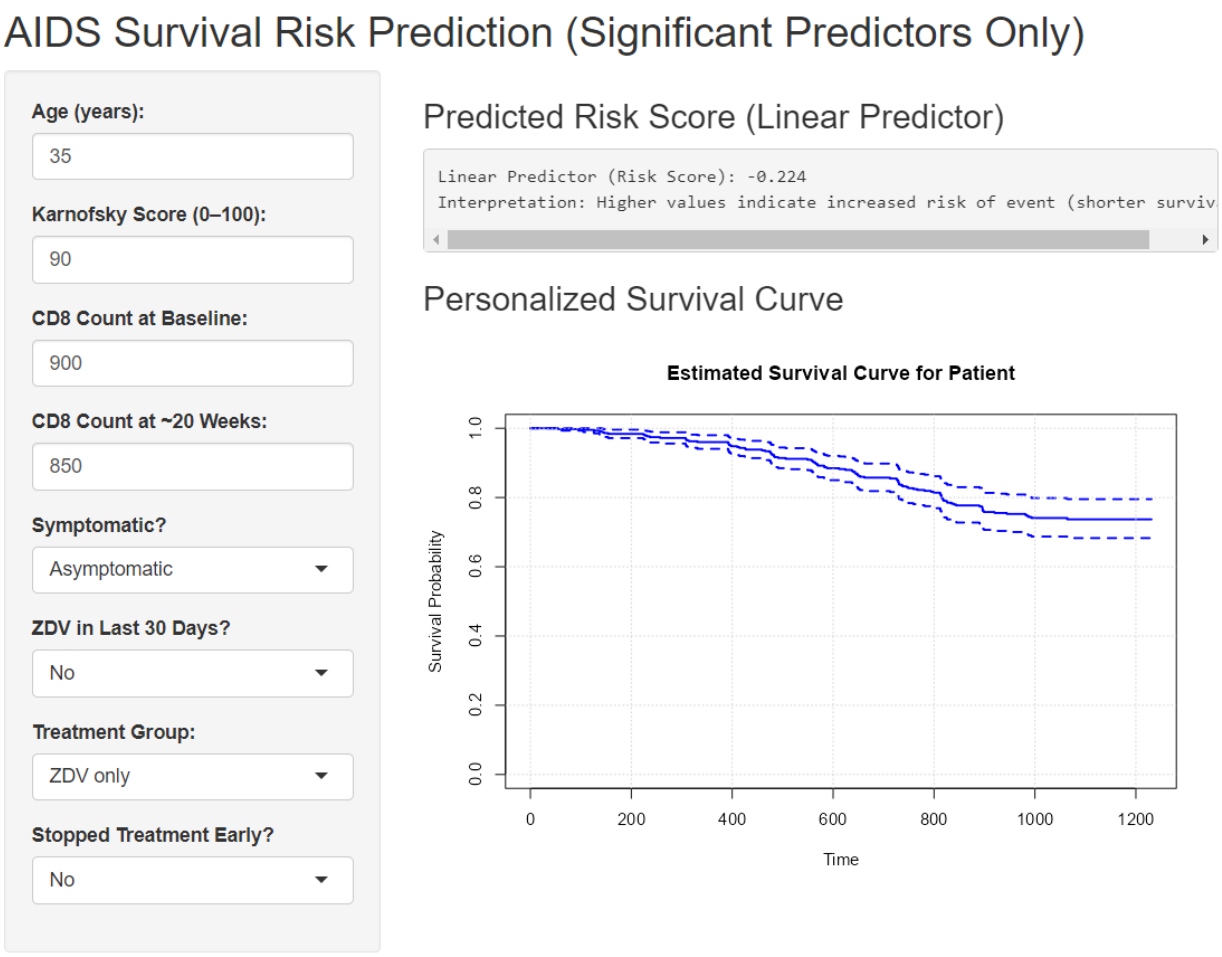


*Figure 13: R Shiny app interface for survival prediction(http://127.0.0.1:5170/)*

## Discussion

This study explored survival prediction in AIDS patients using traditional and machine learning-based survival models. Among the three models evaluated—Cox Proportional

Hazards, Random Survival Forest (RSF), and Gradient Boosting Machine (GBM)—RSF demonstrated the highest concordance index (0.616), slightly outperforming the Cox model (0.614). Although the difference was marginal, RSF showed better discrimination, likely due to its ability to model complex non-linear interactions without assuming proportional hazards.

Cox regression, despite its slightly lower C-index, offered greater interpretability and clinical transparency. It identified key statistically significant predictors, including CD8 counts at baseline and 20 weeks, symptom status, Karnofsky performance score, and recent ZDV exposure (z30). These findings align with the expected clinical relevance of immune status and symptomatic burden in AIDS progression.

Kaplan-Meier visualizations supported the inclusion of several predictors in the Cox model, and Schoenfeld residuals tests ensured that proportional hazards assumptions were met. Additionally, sensitivity analysis revealed a strong correlation between str2 and z30, justifying the use of z30 in the final Cox model. The GBM model, while powerful, underperformed relative to the other two, indicating potential limitations in capturing survival patterns in this dataset.

To improve clinical usability, an interactive R Shiny app was developed using the final Cox model. This tool allows users to input patient characteristics and visualize individualized risk scores and survival probabilities, bridging statistical insights with practical healthcare application.

**Limitations**

Despite the strengths of this analysis, several limitations must be acknowledged. The generalizability of the findings is limited, as the dataset used (AIDS Clinical Trials Group Study 175) originates from a controlled clinical trial setting and may not represent the broader population of AIDS patients, particularly across diverse geographic or socio-economic contexts. Although the Random Survival Forest (RSF) model marginally outperformed the Cox model in predictive accuracy, the overall improvement in the concordance index was small (0.002), limiting the practical impact of this difference. Additionally, the models were trained and evaluated on the same dataset without the use of an independent test set, which may lead to overfitting or optimistic performance estimates. While machine learning models such as RSF and Gradient Boosting Machines (GBM) can capture non-linearities, their complexity reduces clinical interpretability compared to the Cox model, which limits their immediate usability without interpretive tools. Furthermore, deep learning models such as DeepSurv were not implemented due to

time and resource constraints, potentially missing insights from modern neural network approaches.

**Recommendations**

Future studies should validate the developed models using external datasets from different populations to better assess their generalizability and robustness. Validation beyond the original clinical trial setting would help confirm the applicability of the findings in real-world scenarios. Incorporating time-varying covariates, such as changes in treatment status or symptom progression, could better reflect dynamic clinical realities and improve predictive accuracy. The R Shiny dashboard developed in this study could be expanded into a user-friendly decision support tool by integrating it with electronic health record systems and testing its usability among healthcare providers. In addition, exploring ensemble or hybrid approaches that combine traditional and machine learning models may offer enhanced accuracy while maintaining clinical interpretability. Simplifying predictive tools, for instance through intuitive visualizations or risk scoring systems, could make machine learning outputs more actionable and accessible to clinicians. Finally, future research should prioritize the inclusion of heterogeneous, real-world patient populations to improve the practical applicability and relevance of survival models across diverse clinical settings.

**References**

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–220. https://academic.oup.com/jrsssb/article/34/2/187/7027194

Deeks, S. G., Overbaugh, J., Phillips, A., & Buchbinder, S. (2015). HIV infection. *Nature Reviews Disease Primers, 1*(1), 15035. https://www.nature.com/articles/nrdp201535

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S., & Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine, 335*(15), 1081–1090. https://archive.ics.uci.edu/dataset/890/aids+clinical+trials+group+study+175

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics, 2*(3), 841–860. https://doi.org/10.1214/08-AOAS169

Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(5), e1379. https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1379

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys, 51*(6), 1–36. https://dl.acm.org/doi/10.1145/3214306