Identifying key factors influencing survival in HIV patients and enhancing predictions with Cox regression and machine learning models

HANNOU Fatima Zahra[a]

[a] *Université Evry Paris-Saclay, France*

Abstract

This study investigated factors influencing survival in 467 HIV-infected patients who failed or were intolerant to zidovudine (AZT) therapy, comparing the efficacy of didanosine (ddI) and zalcitabine (ddC) over an 18-month period with 1405 total observations including 412 deaths. Using Cox proportional hazards regression and Random Survival Forest models, we analyzed the association between mortality and covariates including treatment, CD4 cell count, time of observation, sex, previous opportunistic infection, and AZT intolerance/failure. Lower CD4 cell count (HR = 0.85, $p < 0.005$), time of observation (HR = 0.85, $p < 0.005$), and previous opportunistic infection (HR = 0.45, $p < 0.005$) were significantly associated with increased mortality risk, while no significant survival difference was found between ddI and ddC treatments (HR = 1.18, $p = 0.33$). The Cox model achieved a test set concordance score of 0.73, and the Random Survival Forest model achieved a test set concordance score of 0.75. These findings confirm the predictive importance of CD4 cell count and opportunistic infections in this patient population and suggest that both Cox and Random Forest models are suitable for survival prediction. Further research is needed to investigate the comparable efficacy of ddI and ddC, while emphasizing CD4 monitoring and opportunistic infection management to improve survival outcomes in HIV-infected individuals.

Keywords: HIV-infected patients, Survival analysis, Cox proportional hazards, Random Survival Forest, Survival prediction, Longitudinal study

1. Introduction

Human Immunodeficiency Virus (HIV) remains a significant global health challenge, affecting millions worldwide (1). The virus compromises the immune system, leading to Acquired Immunodeficiency Syndrome (AIDS) and ultimately increasing mortality if left untreated (2). Antiretroviral therapy (ART) has dramatically improved the future health for individuals living with HIV, yet understanding factors that influence survival remains critical for optimizing treatment strategies.

Zidovudine (AZT) was one of the first antiretroviral drugs approved for HIV treatment and has been instrumental in extending patient survival (3). However, some patients experience treatment failure or intolerance to AZT (4), necessitating the use of alternative therapies. Identifying effective alternative treatment options and understanding the factors influencing survival in this specific patient population are crucial for clinical management (5).

Didanosine (ddI) and zalcitabine (ddC) are both nucleoside reverse transcriptase inhibitors (NRTIs) that offer alternative treatment pathways for patients who cannot tolerate or no longer respond to AZT (6). These drugs work by inhibiting the viral reverse transcriptase enzyme, a critical step in HIV replication (7).

CD4+ T-lymphocyte (CD4) cell count is a well-established marker of immune function and a key indicator of HIV disease progression (8). Lower CD4 counts signify a weakened immune system, making individuals more susceptible to opportunistic infections and increasing their risk of mortality (9). Therefore, monitoring CD4 cell count is essential for assessing treatment response and guiding clinical decision-making.

This study aims to investigate the factors influencing survival in HIV-infected patients who failed or were intolerant to AZT therapy. Specifically, we aim to: 1) identify key predictors of mortality; 2) compare the efficacy of didanosine (ddI) and zalcitabine (ddC) in terms of patient survival; and 3) assess the association between CD4 cell count and the risk of death in this population. To address these objectives, we conducted a longitudinal study of HIV-infected patients and used survival analysis techniques, including Cox proportional hazards regression and Random Survival Forest models, to analyze the data and identify significant predictors of mortality.

The findings from this study will provide valuable insights into the factors that influence survival in AZT-experienced HIV patients and inform the selection of appropriate alternative treatment strategies. This research will contribute to improving the care and outcomes of individuals living with HIV.

2. Related work

Several studies have explored survival analysis and outcome in HIV-infected patients using various statistical and machine learning methods. Karimi and Safari (10) investigated the key factors affecting the survival of HIV patients using both the

Cox proportional hazards model and the random survival forest (RSF) method. Their study, based on a retrospective cohort of 769 HIV patients, highlighted the significance of injection drug use, tuberculosis (TB) status, and initial CD4 cell count as predictors of mortality. They concluded that RSF might offer advantages in handling complex relationships between risk factors compared to the Cox model. Similarly, Prosperi and Di Giambenedetto (11) compared Cox regression and RSF for predicting time to virologic failure after combination antiretroviral therapy (cART) initiation or switch. Their analysis of 2337 treatment regimens demonstrated that RSF outperformed Cox regression in predictive accuracy, suggesting its potential for improved cART monitoring and optimization. Hamidi and Tapak (12) focused on identifying predictive factors influencing AIDS progression in the presence of competing risks (death from causes other than AIDS). Using RSF, they identified age, gender, tuberculosis co-infection, and mode of HIV transmission as important predictors of AIDS progression in an Iranian cohort.

In contrast to these studies focusing primarily on survival analysis, Nisa and Mahmood (13) took a different approach by developing a prediction model for future HIV acquisition in high-risk groups. Their work utilized a broader range of machine learning techniques, including Support Vector Machines (SVMs), Neural Networks, J48 decision trees, and PART rule-based classifiers, in conjunction with feature selection methods, applied to socio-demographic, behavioral, and biological data. They achieved an accuracy of 82% in predicting HIV acquisition, demonstrating a substantial improvement (10-15%) over the individual performance of the employed classifiers. This highlights the potential of combining diverse machine-learning algorithms for enhanced prediction in this context.

While these studies offer valuable insights into HIV/AIDS progression and survival, our work contributes to the existing literature in several ways. Firstly, we focus specifically on patients who have failed or are intolerant to AZT therapy, a population with unique treatment challenges and predictive considerations. Our study directly compares the effectiveness of two alternative antiretroviral drugs, ddI and ddC, in this specific patient group. Secondly, our analysis considers the longitudinal nature of CD4 cell counts, allowing us to examine the dynamic relationship between CD4 decline and mortality risk. Thirdly, by using both Cox regression and Random Survival Forest, we provide a comparative assessment of these models' performance in predicting survival, offering insights into their respective strengths and weaknesses for this patient population.

Our approach allows for a more nuanced understanding of the predictive value of CD4 counts and the efficacy of alternative treatments in AZT-experienced patients. Moreover, our work has several potential advantages over the cited works :

- Comprehensive evaluation of both traditional statistical methods (Cox regression) and modern machine learning approaches (Random Survival Forest).
- Focus on a specific patient population (AZT failure/intolerance) with unique treatment considerations.
- Longitudinal analysis of CD4 counts, capturing the dynamic relationship with survival.
- Direct comparison of alternative treatment options (ddI vs. ddC) in the target population.

3. Dataset

3.1. Data description

The data for this study comes from a longitudinal observational study of HIV-infected patients who experienced treatment failure or intolerance to zidovudine (AZT). These patients were subsequently treated with either didanosine (ddI) or zalcitabine (ddC). The study followed these patients over an 18-month period. Each patient contributed multiple measurements over this timeframe, resulting in a total of 1405 observations from 467 unique patients. Of these observations, 412 recorded the event of interest (death), while the remaining 993 observations were censored. Censorship occurs when the event (death) is not observed during the study period, either because the patient survived beyond the 18-month timeframe or was lost to follow-up.

3.2. Variables

The analysis included several key variables. **Time (time)** is the number of months until death or censoring, and the **event indicator (death)** shows whether a patient died (1) or was censored (0). **CD4 cell count (CD4)** measures immune function and was recorded at multiple time points.

Treatment (treatment) refers to the assigned therapy, either ddI or ddC. **Sex (sex)** indicates whether the patient is male or female. **Previous infection (prev_infection)** is a yes/no variable showing if the patient had a prior AIDS diagnosis.

Reason for AZT discontinuation (AZT) explains why the patient stopped taking AZT, either due to intolerance or treatment failure. Finally, **observation time (time_obs)** is the month at which each CD4 count was measured, helping track changes over time.

3.3. Patient characteristics at baseline

A summary of the baseline characteristics of the 467 patients at the start of the study (i.e., at their first observation) is presented in Table 1. The group is predominantly male, with a majority having a prior AIDS diagnosis. Additionally, the two treatment groups (ddI and ddC) are relatively balanced in size.

3.4. Data preprocessing and data splitting

The provided dataset was complete, with no missing values (NaNs) found in any of the variables. Categorical variables, including *treatment*, *sex*, *prev_infection*, and *azt*, were encoded into numerical representations using ordinal encoding. Ordinal encoding was chosen due to the relatively small number of categories within each variable and the interpretability it offers

Table 1: Summary of patient characteristics at study start (baseline).

Variable	Category	Number of patients	Percentage (%)
Treatment	ddI	688	48.97
	ddC	717	51.03
Sex	Male	1288	91.67
	Female	117	8.33
Previous infection	AIDS	863	61.42
	No AIDS	542	38.58
AZT	Intolerance	914	65.05
	Failure	491	34.95

in the context of Cox proportional hazards models. This transformation is necessary for the statistical and machine learning models used in the analysis, which typically require numerical input.

The dataset was checked for duplicate rows, and none were found. Four patients who died after the conclusion of the 18-month study period were treated as censored observations, as their survival time exceeded the study's timeframe. The longitudinal structure of the data was preserved to leverage the repeated measurements per patient within the Cox proportional hazards model, which inherently accounts for intra-individual correlation.

The dataset was divided into outcome variable (y) and predictor variables (x). The outcome variable y was constructed to include both the time to event (*time*) and the event indicator (*death*). The predictor variables x comprised all other variables in the dataset, excluding *time* and *death*.

To evaluate the performance of the survival models, the dataset was split into 80% training and 20% testing sets using stratified random sampling based on the *death* variable. This stratification ensures a similar proportion of events (deaths) and non-events (censored observations) in both sets, which is crucial for unbiased model evaluation in survival analysis.

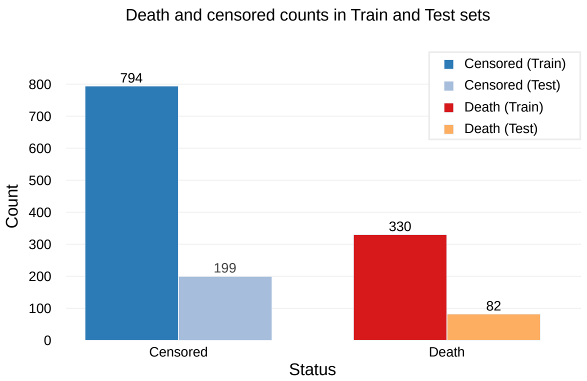

Figure 1: Death and censored counts in Train and Test sets.

As shown in Figure 1, the stratification resulted in nearly identical death rates of 29.4% in the training set and 29.2% in the testing set, confirming the effectiveness of the stratified split. The split was performed using `StratifiedShuffleSplit` from `scikit-learn` with a `random_state` of 42 for reproducibility.

4. Exploratory data analysis (EDA)

This section focuses on understanding the dataset and identifying factors that might influence the survival of HIV-infected patients who did not respond well or could not tolerate AZT therapy. We will look at the distribution of important data, and how different variables are related. This exploration will help guide further analysis and tests to understand what factors affect survival, the effectiveness of treatment, and the connection between CD4 cell count and the risk of death.

4.1. Variable distributions and correlations

The distribution of baseline CD4 cell counts is shown in Figure 2. This histogram, representing CD4 counts at study entry ($time\_obs = 0$), reveals a median count of 6.08. CD4 values range from 1 to 19.24, with a mean count of 7.13.

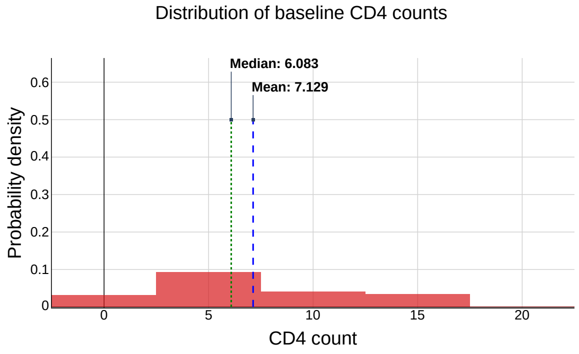

Figure 2: Distribution of CD4 at baseline.

The median time to the event (death) was 10.23 months, ranging from 0.47 to 19.07 months, while the median time to censoring was 15.53 months, ranging from 1.83 to 21.4 months. CD4 counts were recorded multiple times throughout the study, with each patient having an average of 3 measurements (median 3, range 1–5). Table 2 provides a summary of the CD4 measurement distribution over the observation periods, showing that more measurements were taken at baseline and during the earlier stages of the study.

Table 2: Number of CD4 cell count measurements at different observation times

Observation time (months)	**Category**	**Number of measurements**	**Percentage (%)**
0	Baseline	467	33.24
2	Early study	368	26.19
6	Mid-study	310	22.06
12	Late study	226	16.09
18	End of study	34	2.42
Total	All time points	1,405	100.00

The survival times and censoring patterns for a subset of 15 patients are visualized in Figure 3. A substantial proportion of patients (70.67%) were censored, meaning their event (death) was not observed during the study period. This is visually represented by the black dots extending to the right of the dashed line at 18 months. The figure underscores the need for survival analysis methods that properly account for censoring to avoid biased estimates of survival probabilities.

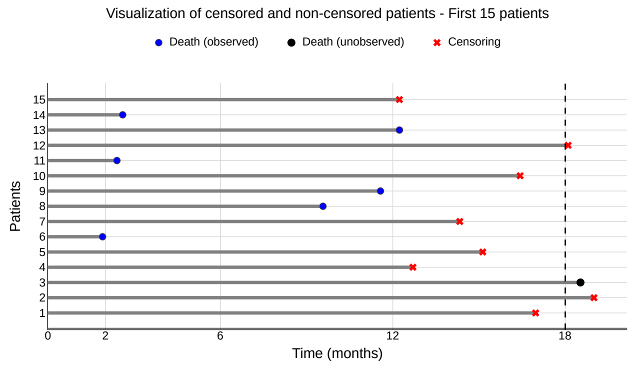

Figure 3: Visualization of survival times and censoring for a subset of patients. Blue dots indicate observed deaths, black dots represent censored observations, and red crosses mark censoring times. The vertical dashed line indicates the end of the study period (18 months).

The distributions of *time*, *CD4* count, and *observation time* are presented in Figure 4. The *time* distribution follows a roughly bell-shaped curve, peaking around 13 months, suggesting that a large number of events or censoring occurred around this time.

The *CD4* count distribution is right-skewed, with a peak near 4 and a long tail towards higher values, indicating that a significant number of patients had lower CD4 counts, while a few exhibited higher values. This is important as lower CD4 counts are often associated with increased risk of opportunistic infections and mortality in HIV patients, which we will investigate further in our survival analysis.

The *observation time* distribution reveals a progressive decline in measurements, beginning with 467 recorded observations in the initial month and diminishing to only 34 observations by the 18th month. This downward trend can be attributed to various factors, such as deaths or lack of follow-up participation.

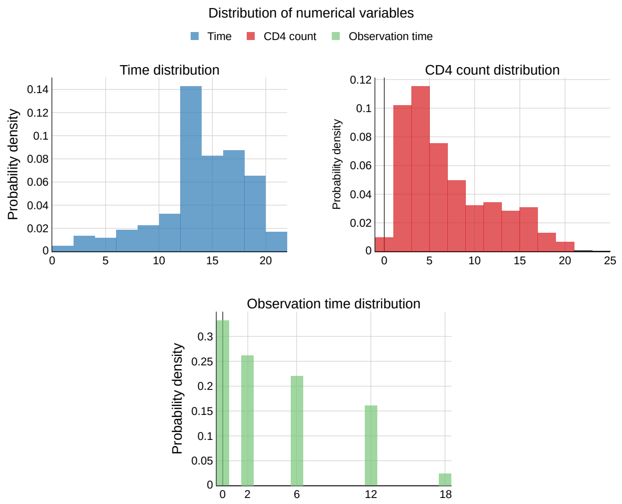

Figure 4: Distribution of numerical variables: Time to event or censoring, CD4 cell count, and observation time.

The distribution of categorical variables in the dataset is illustrated in Figure 5. The *Treatment* variable shows comparable counts for *ddC* and *ddI* cases, suggesting a relatively balanced distribution between the two treatment groups. The *Sex* variable reveals a notable imbalance, with male patients (422 cases) significantly outnumbering female patients (45 cases). This imbalance could potentially influence survival patterns and will be considered in our analysis.

Regarding the *previous infection* variable, AIDS cases are more prevalent (307) compared to noAIDS cases (160). This suggests that a majority of the patients in our study had a prior AIDS diagnosis, which is a crucial factor to consider when analyzing survival. For the *AZT* variable, intolerance cases (292) are more frequent than failure cases (175), indicating that intolerance to AZT was a more common reason for discontinuing the drug than treatment failure.

The correlation matrix, highlighting notable relationships between key variables, is depicted in Figure 6. A moderate positive correlation (0.25) is observed between *Time* and *Previous Infection*, suggesting that patients with a history of previous infections tend to have longer observation times. Additionally, a strong negative correlation (-0.63) between *Time* and *Death* indicates that longer observation periods are associated with higher mortality rates, as expected, given that surviving patients

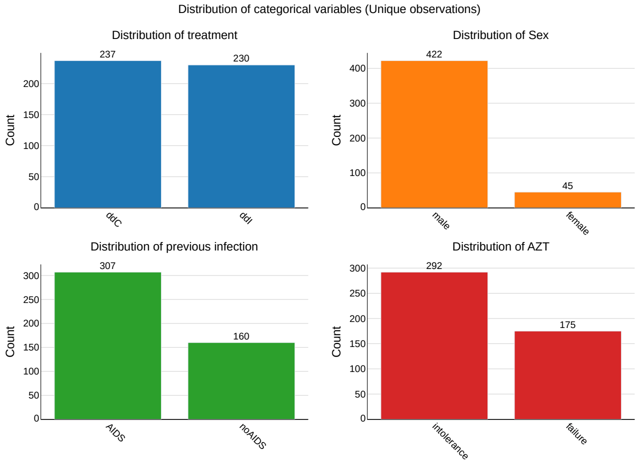

Figure 5: Distribution of categorical variables: Treatment (ddC vs. ddI), Sex (Male vs. Female), Previous Infection (AIDS vs. no AIDS), and AZT reason for stopping (Intolerance vs. Failure).

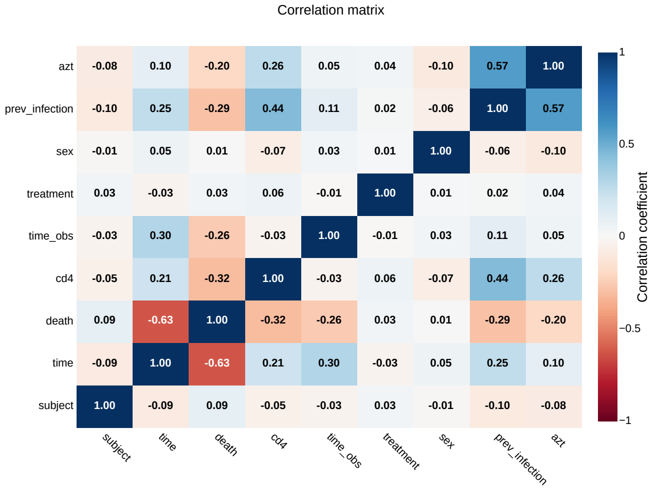

Figure 6: Correlation matrix of numerical variables.

are observed for extended durations.

The matrix also reveals insights regarding *CD4* levels. A moderate positive correlation (0.21) exists between *CD4* levels and *Time*, indicating that patients with higher CD4 counts are generally observed for longer periods, likely due to better immune function and reduced disease progression risk. In contrast, a moderate negative correlation (-0.32) between *CD4* levels and *Death* highlights the protective role of higher CD4 counts against mortality, aligning with established knowledge about their critical function in HIV infection.

The strong negative correlation (-0.63) between *Time* and *Death*, coupled with the positive correlation (0.21) between CD4 and Time, suggests that patients who survive longer tend to have higher CD4 counts. This supports the established understanding of CD4 count as a marker of immune function and a potential predictor of survival in HIV patients, a relationship we will explore further using survival analysis techniques.

4.2. Survival curves by covariates

To assess the impact of various covariates on patient survival, Kaplan-Meier survival curves were generated, stratified by treatment type, sex, previous infection status, and AZT.

The survival curves for the ddC and ddI treatment groups (Figure 7(d)) appear closely aligned, suggesting minimal differences in survival between these treatments. However, this observation requires validation through formal statistical testing.

Sex-based survival curves (Figure 7(c)) indicate comparable outcomes for males and females, implying that sex may not be a significant determinant of survival in this group. Further statistical analysis is necessary to confirm this finding.

Figure 7(b) illustrates the survival curves by previous infection status, revealing a clear difference between patients with and without a prior AIDS diagnosis. Those with a previous AIDS diagnosis consistently show lower survival probabilities, suggesting that prior infection history could be a significant predictor of survival.

Survival curves by AZT discontinuation reasons (Figure 7(a)) reveal subtle differences. Patients who stopped AZT due to intolerance demonstrate slightly better survival probabilities than those who discontinued due to treatment failure. Statistical testing will determine whether this observed difference is significant.

These Kaplan-Meier curves offer a visual overview of the relationships between various covariates and patient survival. While some trends are apparent, formal statistical analyses, detailed in the subsequent section, are essential to confirm and quantify the observed effects.

5. Methodology

5.1. Statistical analysis

Survival analysis was conducted to investigate the time to death in the study population. Two approaches were used: Cox proportional hazards regression and Random Survival Forest (RSF).

5.1.1. Cox proportional hazards regression

The Cox proportional hazards model (14) was used to estimate the hazard ratios (HRs) for each covariate, quantifying their association with mortality risk. The model assumes that the hazard function for an individual is proportional over time and can be expressed as:

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

where $h(t|X)$ is the hazard at time t given covariates X, $h_0(t)$ is the baseline hazard function, and β_i are coefficients that measure the impact of covariates X_i.

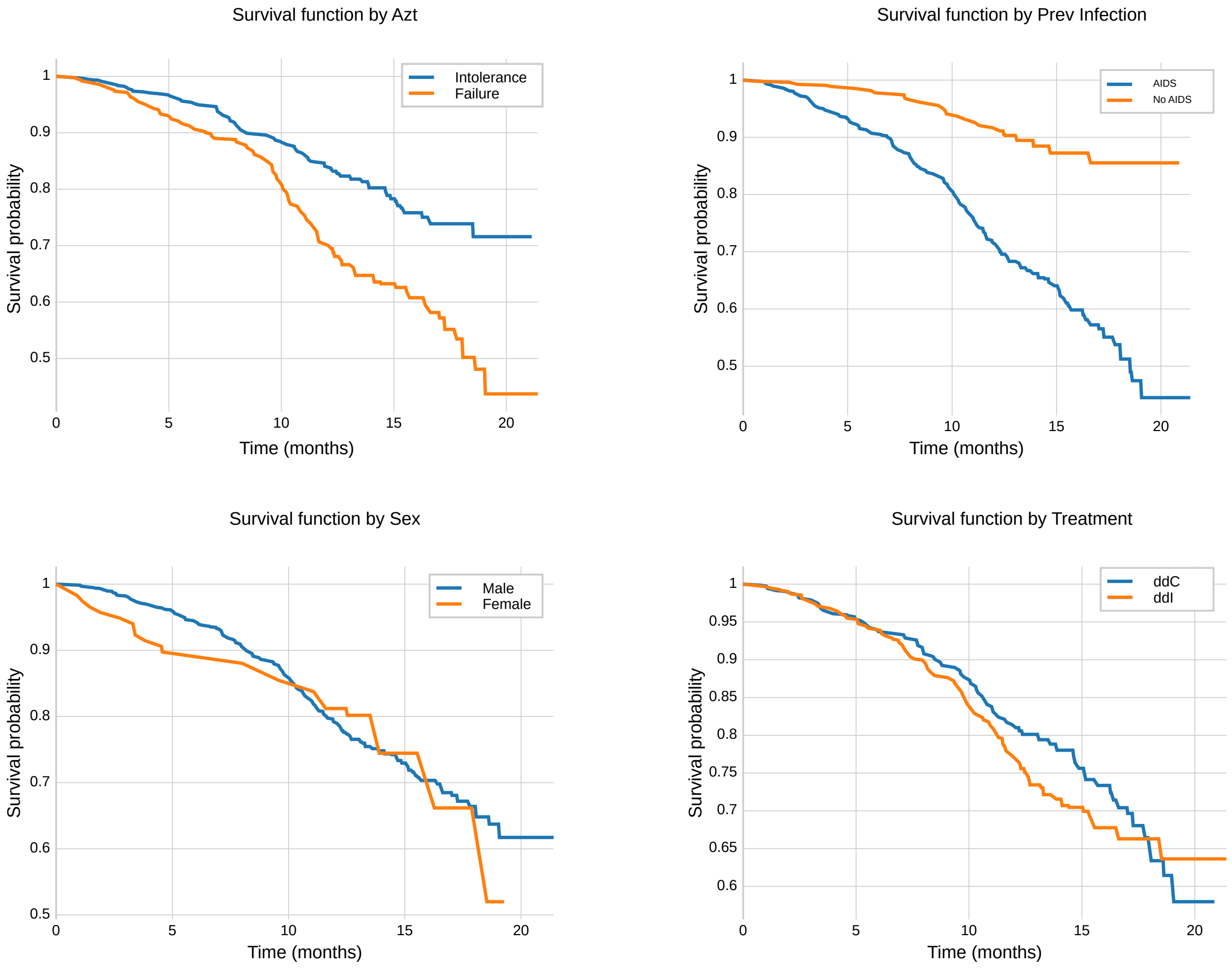

Figure 7: Kaplan-Meier Survival Curves by (a) AZT, (b) Previous opportunistic infection, (c) Sex, (d) Treatment.

5.1.2. *Random Survival Forest*

Random Survival Forest (RSF) (15) is a non-parametric ensemble method that extends the Random Forest algorithm to survival data. RSF constructs multiple decision trees, each trained on a random subset of the data and using a random subset of predictors at each node split. The ensemble prediction is obtained by averaging the survival predictions from individual trees. RSF does not require the proportional hazards assumption and can capture complex non-linear relationships between covariates and survival.

5.2. *Model evaluation*

The performance of both Cox regression and RSF models was evaluated using the concordance index (C-index) (16) and the Brier score (17). The C-index measures the probability that, for a randomly selected pair of patients, the model correctly predicts which patient experiences the event first. A C-index of 0.5 indicates random prediction, whereas a value of 1 signifies perfect prediction. Survival models generally achieve C-index values between 0.55 and 0.75 (18).

The Brier score is a measure of prediction accuracy for survival models. It assesses the difference between predicted survival probabilities and the actual observed events. A lower Brier score indicates better predictive performance.

The dataset was divided into 80% for training and 20% for testing using stratified random sampling based on the event indicator (*death*), as outlined in the data preprocessing section. Model performance was evaluated on the held-out test set to estimate its generalization capability.

5.3. *Software and packages*

The statistical analyses were performed using Python, leveraging libraries such as `lifelines` for survival analysis (including Cox regression), `sksurv` for Random Survival Forest, and `scikit-learn` for data preprocessing and model evaluation.

6. Results

The results of the survival analysis using Cox proportional hazards regression and Random Survival Forest are presented in this section.

6.1. *Cox Proportional hazards regression*

The Cox proportional hazards model revealed several significant predictors of mortality in HIV patients failing or intolerant to AZT. Table 3 summarizes the estimated hazard ratios (HRs), 95% confidence intervals, and p-values for each covariate. Lower CD4 cell count (HR = 0.85, $p < 0.005$), later observation time (HR = 0.85, $p < 0.005$), and a prior diagnosis of AIDS (HR = 0.45, $p < 0.005$) were significantly associated with an increased risk of mortality. No statistically significant difference in survival was observed between patients treated with ddI and ddC (HR = 1.18, $p = 0.33$). Sex and AZT were also not significantly associated with mortality in this model.

Table 3: Cox proportional hazards regression results.

Covariate	**HR**	**95% CI**	**p-value**
CD4 Count	0.85	(0.81, 0.90)	<0.005
Observation time	0.85	(0.83, 0.88)	<0.005
Previous infection	0.45	(0.26, 0.76)	<0.005
Treatment	1.18	(0.85, 1.63)	0.33
Sex	0.91	(0.45, 1.84)	0.80
AZT	0.89	(0.62, 1.27)	0.51

We trained two variants of the Cox proportional hazards model: one using linear relationships between covariates and another using complex non-linear relationships. Both models demonstrated strong performance on the training set, with the linear Cox model achieving a score of 0.79 and the non-linear Cox model scoring 0.72. However, on the test set, the linear Cox model significantly outperformed the non-linear Cox model, achieving a score of 0.73 compared to 0.65.

Table 4: Comparison of linear and non-linear Cox model performance.

Model variant	**Train set score**	**Test set score**
Linear Cox model	**0.79**	**0.73**
Non-linear Cox model	0.72	0.65

6.2. *Random Survival Forest*

The Random Survival Forest (RSF) model achieved a concordance index of 0.75 on the test set, slightly outperforming the Cox regression model. The variable importance scores from the RSF model, presented in Table 5, highlight CD4 count, observation time, and previous AIDS diagnosis as the most influential predictors of mortality. This is consistent with the findings from the Cox regression analysis.

The Brier scores of the linear and non-linear Cox models are compared with the RSF model in Figure 8. The comparison reveals that the Brier score curve for the Random Survival Forest (RSF) consistently remains lower than those of both the linear and non-linear Cox model variants. This indicates that the RSF model not only outperforms the non-linear Cox model but also demonstrates superior predictive accuracy compared to the linear Cox model.

Table 5: Variable importance scores from Random Survival Forest. The values range from most important (top) to less important (bottom row).

Variable	**Mean**	**Standard deviation**
CD4 Count	0.082	0.015
Observation time	0.075	0.020
Previous infection	0.029	0.011
Treatment	0.017	0.006
AZT	0.015	0.007
Sex	0.007	0.003

The consistently better performance of RSF suggests its robustness and effectiveness in handling the data, offering an advantage over traditional Cox models in terms of predictive accuracy.

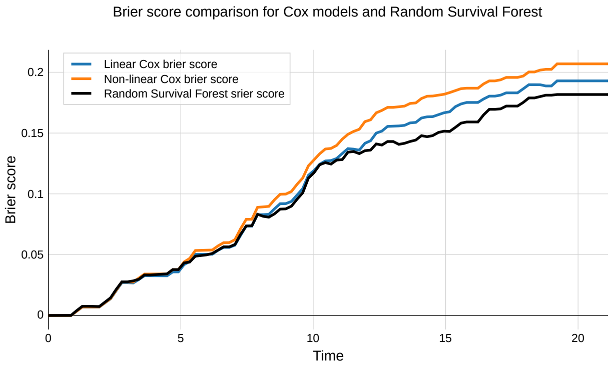

Figure 8: Brier score comparison for Cox Models and Random Survival Forest.

7. Conclusion

This study investigated the factors influencing survival in HIV-infected patients who failed or were intolerant to AZT therapy, comparing the efficacy of ddI and ddC and assessing the association between CD4 cell count and mortality. Our analysis, using both Cox proportional hazards regression and Random Survival Forest, identified lower CD4 cell count, later observation time (which reflects disease progression), and a previous AIDS diagnosis as significant predictors of increased mortality risk. Consistent with current clinical understanding, these findings highlight the importance of CD4 monitoring and the impact of opportunistic infections on outcomes in this patient population.

We found no statistically significant difference in survival between patients treated with ddI and ddC. This suggests that both treatments offer comparable efficacy in this specific patient group, although further research with larger sample sizes and longer follow-up periods is needed to confirm this observation. Additionally, neither sex nor AZT were significantly associated with mortality in our models.

The Random Survival Forest model demonstrated slightly better predictive performance compared to the Cox regression model, as evidenced by a higher concordance index and lower Brier score. This suggests that RSF, with its ability to capture non-linear relationships and interactions, may be particularly well-suited for survival prediction in this complex patient population.

Our findings reinforce the clinical importance of CD4 cell count and opportunistic infections as key predictors of survival in HIV-infected individuals who are failing or intolerant to AZT. This study highlights the utility of both Cox regression and Random Survival Forest models for survival prediction, offering valuable insights for understanding survival outcomes in this population.

Data and code availability

The data and code used in this study are available at: https://github.com/fatima-zahra-hannou/Survival_Analysis_Project

References

[1] H. D. Gayle, G. L. Hill, Global impact of human immunodeficiency virus and aids, Clinical Microbiology Reviews 14 (2) (2001) 327–335. arXiv:https://journals.asm.org/doi/pdf/10.1128/cmr.14.2.327-335.2001, doi:10.1128/cmr.14.2.327-335.2001.
URL https://journals.asm.org/doi/abs/10.1128/cmr.14.2.327-335.2001

[2] H. M. Naif, Pathogenesis of hiv infection, Infectious Disease Reports 5 (11) (2013). doi:10.4081/idr.2013.s1.e6.
URL https://www.mdpi.com/2036-7449/5/11/e6

[3] R. A. Crouch, J. D. Arras, Azt trials and tribulations, The Hastings Center Report 28 (6) (1998) 26–34.
URL http://www.jstor.org/stable/3528266

[4] M. H. S. Clair, J. L. Martin, G. Tudor-Williams, M. C. Bach, C. L. Vavro, D. M. King, P. Kellam, S. D. Kemp, B. A. Larder, Resistance to ddi and sensitivity to azt induced by a mutation in hiv-1 reverse transcriptase, Science 253 (5027) (1991) 1557–1559. arXiv:https://www.science.org/doi/pdf/10.1126/science.1716788, doi:10.1126/science.1716788.
URL https://www.science.org/doi/abs/10.1126/science.1716788

[5] H. J. Chang, LW, E. Humphreys, Optimal monitoring strategies for guiding when to switch first-line antiretroviral therapy regimens for treatment failure in adults and adolescents living with hiv in low-resource settings, Cochrane Database of Systematic Reviews (4) (2010). doi:10.1002/14651858.CD008494.
URL https://doi.org//10.1002/14651858.CD008494

[6] J. A. Sandberg, A. W. Slikkek Jr., Developmental pharmacology and toxicology of anti-hiv therapeutic agents: dideoxynucleosides, The FASEB Journal 9 (12) (1995) 1157–1163. arXiv:https://faseb.onlinelibrary.wiley.com/doi/pdf/10.1096/fasebj.9.12.7672508, doi:https://doi.org/10.1096/fasebj.9.12.7672508.
URL https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fasebj.9.12.7672508

[7] J. J. Kohler, W. Lewis, A brief overview of mechanisms of mitochondrial toxicity from nrtis, Environmental and Molecular Mutagenesis 48 (3-4) (2007) 166–172. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/em.20223, doi:https://doi.org/10.1002/em.20223.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/em.20223

[8] A. Kamat, V. Misra, E. Cassol, P. Ancuta, Z. Yan, C. Li, S. Morgello, D. Gabuzda, A plasma biomarker signature of immune activation in hiv patients on antiretroviral therapy, PLOS ONE 7 (2) (2012) 1–11. doi:10.1371/journal.pone.0030881.
URL https://doi.org/10.1371/journal.pone.0030881

[9] G. R. Seage III, E. Losina, S. J. Goldie, D. A. Paltiel, A. D. Kimmel, K. A. Freedberg, The relationship of preventable opportunistic infections, hiv-1 rna, and cd4 cell counts to chronic mortality, JAIDS Journal of Acquired Immune Deficiency Syndromes 30 (4) (2002).
URL https://journals.lww.com/jaids/fulltext/2002/08010/the_relationship_of_preventable_opportunistic.7.aspx

[10] N. Karimi, M. Safari, M. Mirzaei, A. Kassaeian, G. Roshanaei, T. Omidi, Determining the factors affecting the survival of hiv patients: Comparison of cox model and the random survival forest method, Int Electron J Med 8 (2) (2019) 124–129. arXiv:https://ddj.hums.ac.ir/PDF/iejm-84, doi:10.34172/iejm.2019.09.
URL https://ddj.hums.ac.ir/Article/iejm-84

[11] M. C. Prosperi, S. Di Giambenedetto, I. Fanti, G. Meini, B. Bruzzone, A. Callegaro, G. Penco, P. Bagnarelli, V. Micheli, E. Paolini, A. Di Biagio, V. Ghisetti, M. Di Pietro, M. Zazzi, A. De Luca, t. A. cohort, A prognostic model for estimating the time to virologic failure in hiv-1 infected patients undergoing a new combination antiretroviral therapy regimen, BMC Medical Informatics and Decision Making 11 (1) (2011) 40. doi:10.1186/1472-6947-11-40.
URL https://doi.org/10.1186/1472-6947-11-40

[12] O. Hamid, M. Tapak, J. Poorolajal, P. Amini, L. Tapak, Application of random survival forest for competing risks in prediction of cumulative incidence function for progression to aids, Epidemiology, Biostatistics, and Public Health 14 (4) (Mar. 2022). doi:10.2427/12663.
URL https://riviste.unimi.it/index.php/ebph/article/view/17473

[13] S. U. Nisa, A. Mahmood, F. S. Ujager, M. Malik, Hiv/aids predictive model using random forest based on socio-demographical, biological and behavioral data, Egyptian Informatics Journal 24 (1) (2023) 107–115. doi:https://doi.org/10.1016/j.eij.2022.12.005.
URL https://www.sciencedirect.com/science/article/pii/S1110866522000834

[14] D. R. Cox, Regression models and life-tables, Journal of the Royal Statistical Society. Series B (Methodological) 34 (2) (1972) 187–220.
URL http://www.jstor.org/stable/2985181

[15] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, M. S. Lauer, Random survival forests (2008). arXiv:arXiv:0811.1645, doi:10.1214/08-AOAS169.

[16] J. Harrell, Frank E., R. M. Califf, D. B. Pryor, K. L. Lee, R. A. Rosati, Evaluating the yield of medical tests, JAMA 247 (18) (1982) 2543–2546. arXiv:https://jamanetwork.com/journals/jama/articlepdf/372568/jama_247_18_030.pdf, doi:10.1001/jama.1982.03320430047030.
URL https://doi.org/10.1001/jama.1982.03320430047030

[17] G. W. BRIER, Verification of forecasts expressed in terms of probability, Monthly Weather Review 78 (1) (1950) 1 – 3. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
URL https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml

[18] Concordance index.
URL https://lifelines.readthedocs.io/en/latest/Survival%20Regression.html#concordance-index