www.ijarp.com ISSN 2456-9992 Page: 01-15

Case Study

International Journal Advanced Research Publications

Volume: 02
Issue: 11

COMPARATIVE ANALYSIS OF PENALIZED COX PROPORTIONAL HAZARD MODEL AND ACCELERATED FAILURE TIME MODELS FOR CARDIOVASCULAR DISEASE SURVIVAL PREDICTION: A CASE STUDY OF ABUBAKAR TAFAWA BALEWA UNIVERSITY TEACHING HOSPITAL (ATBU-TH), BAUCHI

Ummi Mukhtar*[1] and K.E. Lasisi[1]

[1]*Department of Statistics, Faculty of Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria.*

Article Received: 23 September 2026, Article Revised: 13 October 2026, Published on: 02 November 2026

***Corresponding Author: Ummi Mukhtar**

Department of Statistics, Faculty of Science, Abubakar Tafawa Balewa University, Bauchi, Nigeria.

Doi: https://doi-doi.org/101555/ijarp.9999

ABSTRACT

Cardiovascular Diseases (CVDs), primarily driven by hypertension and diabetes, present a major healthcare burden in Sub-Saharan Africa. Conventional survival models often suffer from over complexity, over-fitting, and variable selection issues when applied to high-dimensional clinical datasets. This study evaluated variable selection consistency, parameter estimation accuracy, and predictive performance across nonparametric, semi-parametric, and parametric survival frameworks to develop a parsimonious risk-prediction model for cardiovascular patients. A retrospective cohort design was used to analyze secondary data from 1,247 CVD patients admitted to Abubakar Tafawa Balewa University Teaching Hospital (ATBU-TH), Bauchi. Non-parametric survival was evaluated using Kaplan-Meier estimation and log-rank tests. Comparative evaluations were conducted across Cox Proportional Hazards (CPH) and Accelerated Failure Time (AFT) models regularized using convex (L1 LASSO, Elastic Net) and non-convex (Minimax Concave Penalty [MCP]) penalties. Model performance was assessed using 10-fold cross validation, Harrell's Concordance Index (C-index), Integrated Brier Score (IBS), and time dependent ROC-AUC trajectories. During follow-up, 653 death events were recorded, yielding a median survival time of 883 days (95% CI: 821–943 days). Overall survival probabilities were 0.830, 0.587, and 0.392 at 1, 2, and 3 years, respectively. Significant survival differences were observed by hypertension status ($p = 0.03$) and sex ($p = 0.00011$). MCP Cox achieved the best overall

discriminative performance (C-index = 0.812, IBS = 0.128, 20 retained features), while MCP AFT produced the most parsimonious model with 13 primary risk drivers (C-index = 0.798, IBS = 0.133). Age, diabetes, prior CVD events, hypertension, active smoking, and elevated blood pressure increased mortality risk, whereas female sex, non-smoking status, and higher HDL were protective. The study concludes that non-convex MCP regularization provides an effective approach for reducing model complexity while maintaining predictive performance.

INTRODUCTION

Cardiovascular diseases (CVDs) represent a major global public health challenge and remain a leading cause of mortality and longterm morbidity worldwide (World Health Organization [WHO], 2023; Mensah et al., 2023). They encompass a spectrum of clinical conditions affecting the heart and vascular structure, including coronary heart disease, cerebrovascular disease, and hypertensive heart disease (WHO, 2023). Globally, approximately 19.8 million individuals died from CVDs in 2022, accounting for nearly 32% of all global deaths, with myocardial infarction and acute cerebrovascular events responsible for the vast majority of these fatalities (Mensah et al., 2023). Key established prognostic indicators and risk factors include systemic hypertension, diabetes mellitus, tobacco use, physical inactivity, dyslipidemia, and elevated blood glucose (Einarson et al., 2018; Joseph et al., 2022; Low Wang et al., 2016). In Sub-Saharan Africa and Nigeria specifically, the clinical burden of CVD is escalating rapidly due to population aging, rapid urbanization, and lifestyle transitions (Yuyun et al., 2020; Ogah et al., 2014; Oladimeji et al., 2022), positioning hypertension and metabolic co-morbidities as major primary drivers of adverse cardiovascular outcomes.

To effectively mitigate this healthcare burden, clinical decision-making relies heavily on statistical survival analysis to quantify patient survival probability and identify key prognostic determinants over time (Collett, 2015; Hosmer et al., 2008; Lee & Wang, 2013). While the semi-parametric Cox Proportional Hazards (CPH) model (Cox, 1972, 1975) is widely utilized for relative risk evaluation, parametric Accelerated Failure Time (AFT) models (Wei, 1992; Lawless, 2011; Zhang, 2016) provide direct multiplicative interpretations on baseline survival time. However, applying these standard statistical methods to complex, highdimensional clinical datasets often introduces severe methodological challenges, including model over-complexity, collinearity, and overfitting (Harrell et al., 1996; Steyerberg, 2019; Hastie et al., 2021). In regional clinical survival literature, this problem was prominently highlighted by Aliyu et al. (2024), who evaluated diabetes and hypertension

patients across General Hospitals in Keffi and Nasarawa. In their study, a methodological conflict arose when a complex Weibull AFT random effect (frailty) model was prioritized over the standard CPH model based solely on minor improvements in Information Criteria (AIC and BIC) (Akaike, 1974; Schwarz, 1978; Aliyu et al., 2024), despite the random effect variance (θ) proving statistically insignificant ($p = 0.234$). Prioritizing over-complex models with redundant random parameters and non-significant clinical noise variables compromises model parsimony and limits generalizability across broader clinical populations (Steyerberg, 2019; Binder & Schumacher, 2008). Consequently, there is an urgent need to incorporate advanced penalization and variable-selection techniques—such as L1 LASSO (Tibshirani, 1997), Elastic Net (Simon et al., 2011), and Minimax Concave Penalty (MCP) (Zhang et al., 2023)—to construct parsimonious, statistically robust risk prediction models.

To address these methodological and clinical gaps, this study aims to evaluate and compare the predictive performance, estimation accuracy, and variable-selection consistency of penalized Cox Proportional Hazards and penalized Accelerated Failure Time models among cardiovascular disease patients. Specifically, this work seeks to describe the baseline non-parametric survival experience of CVD patients using Kaplan-Meier survival estimates and log-rank tests (Kaplan & Meier, 1958; Mantel, 1966); identify significant prognostic factors using semi-parametric CPH models regularized via convex (LASSO, Elastic Net) and non-convex (MCP) penalties (Tibshirani, 1997; Simon et al., 2011; Zhang et al., 2023); examine significant prognostic determinants under parametric AFT models subjected to identical regularizations (Wei, 1992; Harhay et al., 2014); and systematically compare predictive performance and feature parsimony across model families using Harrell's Concordance Index, Integrated Brier Score, dynamic ROC-AUC trajectories, and 10-fold cross-validation (Harrell et al., 1996; Van Houwelingen & Putter, 2012; Zhao et al., 2021; Zhang et al., 2022) using secondary patient records from Abubakar Tafawa Balewa University Teaching Hospital (ATBU-TH), Bauchi.

METHODOLOGY

Study Design and Patient Cohort: A retrospective cohort study design was utilized to evaluate cardiovascular disease patients admitted to the Cardiology Unit of Abubakar Tafawa Balewa University Teaching Hospital (ATBU-TH), Bauchi, Nigeria. Secondary clinical records were extracted from medical archives. Eligible subjects comprised adult patients (aged ≥ 18 years) with a confirmed primary diagnosis of cardiovascular disease and complete diagnostic follow-up records. Patients aged below 18 years, those with incomplete survival

endpoints, or those admitted primarily for non-cardiovascular clinical indications were excluded. Total enumeration produced a final analytical sample of 1,247 CVD patients. The primary clinical endpoint was all-cause in-hospital mortality.

Statistical Analysis & Modeling Frameworks: Non-parametric survival curves were estimated using the Kaplan-Meier method (Kaplan & Meier, 1958) and compared across clinical risk strata using log-rank tests (Mantel, 1966). Regularized survival regression was conducted under both semi-parametric Cox Proportional Hazards (CPH) (Cox, 1972) and parametric Accelerated Failure Time (AFT) (Wei, 1992) frameworks. To achieve feature selection and parameter regularization, three penalty functions were applied: convex L1 LASSO (Tibshirani, 1997), Elastic Net combining L1 and L2 penalties with $\alpha = 0.5$ (Simon et al., 2011), and non-convex Minimax Concave Penalty (MCP) (Zhang et al., 2023).

Model Evaluation & Cross-Validation: Model generalization ability was evaluated using 10-fold cross-validation. Overall discriminative capacity was quantified using Harrell's Concordance Index (C-index) (Harrell et al., 1996) and dynamic timedependent ROC-AUC trajectories (Van Houwelingen & Putter, 2012). Overall prediction error and calibration quality were assessed using the Integrated Brier Score (IBS) across follow-up time. Model parsimony was evaluated based on the number of non-zero parameters retained by each penalization approach.

RESULTS AND DISCUSSION

The study evaluated 1,247 CVD patients over a total follow-up timeline exceeding 1,000 days. During follow-up, 653 death events were recorded, producing an overall median survival time of 883 days (95% CI: 821–943 days).

Non-Parametric Survival Baseline: Non-parametric Kaplan-Meier estimates demonstrated a steady decline in cumulative survival probability across annual follow-up milestones, recording 0.830 (95% CI: 0.809–0.852) at 1 year (365 days), 0.587 (95% CI: 0.559– 0.617) at 2 years (730 days), and 0.392 (95% CI: 0.364–0.423) at 3 years (1095 days).

Subgroup	Total Sample (N)	Observed Events (O)	Expected Events (E)	$E(O-E)^2$	$V(O-E)^2$	Chi-Square ($\chi 2$)	Df	p-value
No Hypertension	777	385	413	1.84	4.99	5.0	1	0.03
Hypertension	470	268	240	3.15	4.99			

Note: Statistical significance set at $p < 0.05$.

Table 4.4 formalizes the comparison of survival distributions between non-hypertensive and hypertensive patients using the log-rank test. The non-hypertensive group comprised 777 patients with 385 observed deaths against 413 expected deaths, whereas the hypertensive group comprised 470 patients with 268 observed deaths compared to 240 expected deaths. The log-rank statistic yields a chi-square value of 5.0 on 1 degree of freedom with a p-value of 0.03. Because this p-value falls below the conventional 0.05 threshold of statistical significance, there is sufficient evidence to reject the null hypothesis and conclude that the survival experience of cardiovascular disease patients differs significantly based on hypertension status, with hypertension associated with a significantly elevated mortality risk.

Figure 1: Overall Kaplan-Meier Survival Curve for Cardiovascular Patients with At-Risk Counts

Stratified log-rank tests identified statistically significant survival differences by hypertension status ($\chi^2 = 5.0$, df = 1, p = 0.03) and biological sex (p = 0.00011). Hypertensive patients exhibited a 3-year survival probability of 0.341 compared to 0.424 among nonhypertensive patients.

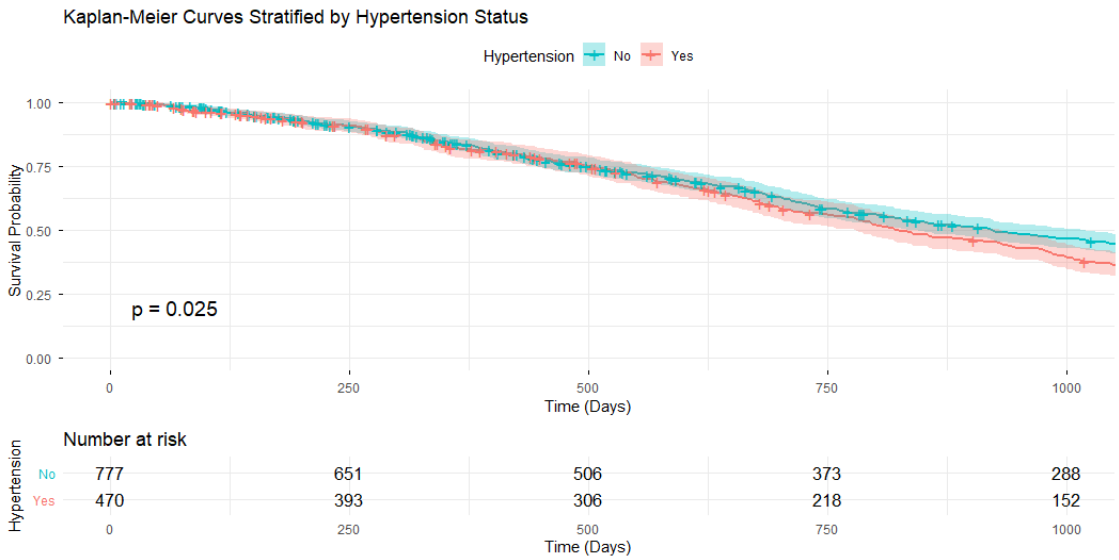


Figure 4.2 Kaplan-Meier Survival Estimates for Cardiovascular Events Stratified by hypertension status Counts and Log-Rank Test.

The second figure presents the Kaplan-Meier survival curves stratified by hypertension status, comparing non-hypertensive patients against hypertensive patients. During the initial 250 to 500 days of follow-up, the two survival trajectories remain closely aligned with overlapping confidence intervals. Beyond 500 days, the curves begin to diverge visibly, with the hypertensive group experiencing a steeper decline in survival probability relative to the

non-hypertensive group. By 1,000 days, patients without hypertension maintain a noticeably higher survival probability than those with hypertension. The log-rank test p-value of 0.025 confirms that this observed divergence in survival experience between the two strata is statistically significant at the 5 percent level, demonstrating that hypertension is associated with significantly poorer long-term survival in this patient population.

Figure 2: Kaplan-Meier Survival Curves Stratified by Hypertension Status with Log-Rank Test (p = 0.025)

Female subjects demonstrated a sustained survival advantage over males across follow-up.

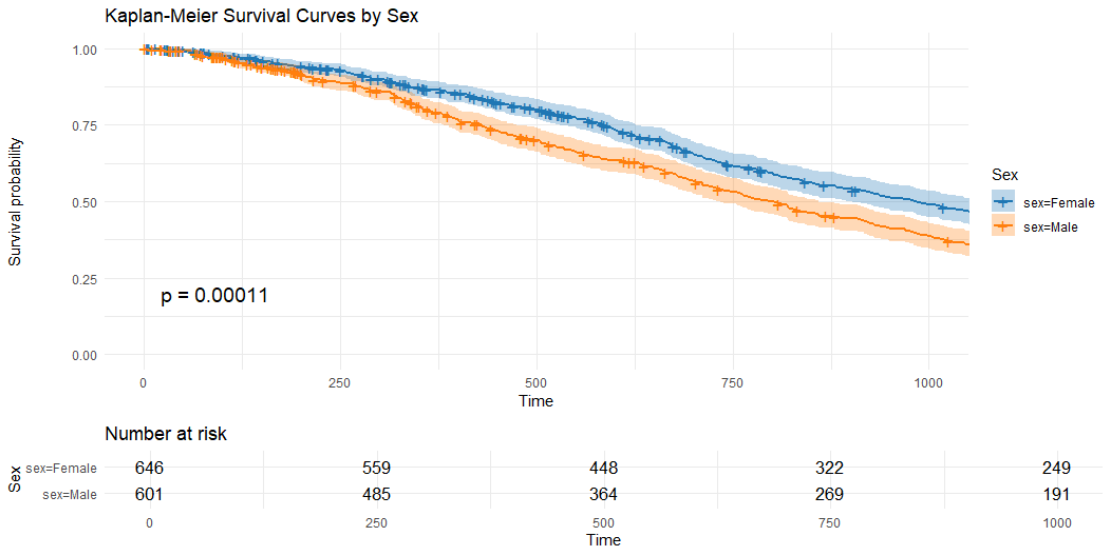


Figure 1: Kaplan-Meier Survival Estimates for Cardiovascular Events Stratified by Sex with At-Risk Counts and Log-Rank Test.

The Kaplan-Meier survival curves illustrate a statistically significant disparity in cumulative survival probabilities between male and female cohorts over a 1000-day follow-up duration. Throughout the entire observation period, the survival curve for female subjects remains consistently above that of male subjects, indicating a sustained survival advantage for females. This difference is statistically confirmed by the log-rank test p-value of 0.00011, which falls well below the standard 0.05 threshold and demonstrates that sex is a crucial stratified risk factor.

The accompanying risk table at the bottom highlights a gradual decline in the at-risk population across both groups, starting from initial cohorts of 646 females and 601 males. By day 500, the number at risk drops to 448 females and 364 males, eventually reaching 249 females and 191 males by day 1000 due to event occurrences and censoring. The non-

overlapping confidence intervals across most of the timeline further reinforce the statistical robustness of this observed survival gap.

Figure 3: Kaplan-Meier Survival Curves Stratified by Sex with At-Risk Counts and Log-Rank Test (p = 0.00011)

Baseline physical activity levels showed no significant survival differentiation (p = 0.70).

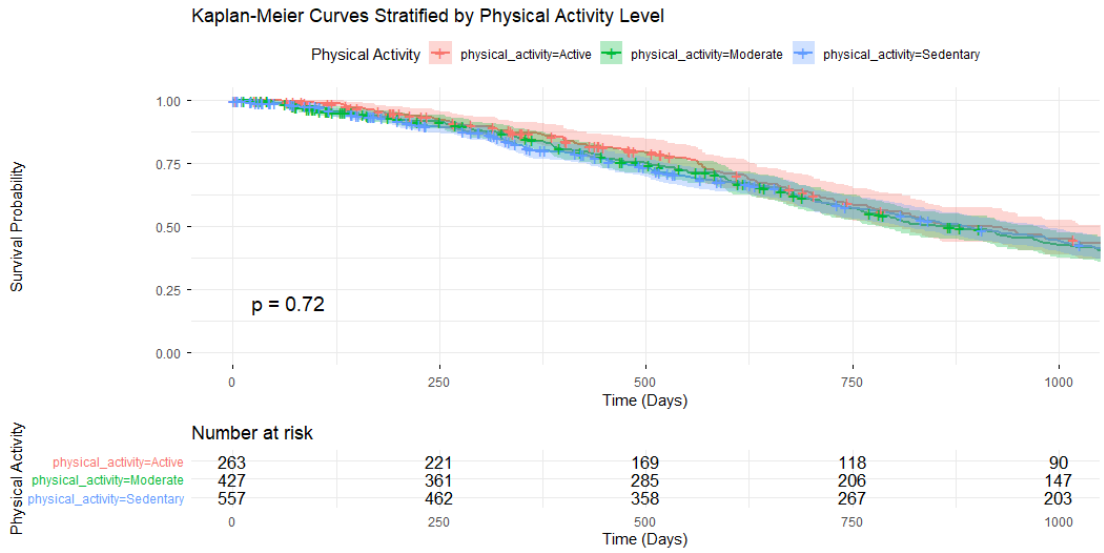


Figure 4.3 Kaplan-Meier Survival Estimates for Cardiovascular Events Stratified by physical activity level Counts and Log-Rank Test.

The third figure illustrates the Kaplan-Meier survival curves stratified across three physical activity categories: active, moderate, and sedentary. Throughout the entire 1,000-day monitoring period, the survival curves for all three activity groups remain closely intertwined, with substantial overlap across their respective confidence shading bands. Although the active group shows a slight, nominal survival advantage around 500 to 750 days, the overall trajectories follow nearly identical paths down to the 1,000-day mark. The log-rank test yields a p-value of 0.72, indicating no statistically significant difference in survival probability across the physical activity levels. This confirms that physical activity status alone does not differentiate survival experience within this cohort under a non-parametric evaluation. *Figure 4: Kaplan-Meier Survival Curves Stratified by Physical Activity Level with Log-Rank Test (p = 0.72)*

Table 1: Kaplan-Meier Survival Probabilities and Overall Milestone Summary.

Milestone / Time Horizon	Patients at Risk (n)	Cumulative Events	Survival Prob. S(t)	Std. Error	95% Confidence Interval
Baseline / Overall Cohort	1,247	653 (Total)	0.500 (Median: 883d)	--	(821, 943 days)
365 Days (1 Year)	925	197	0.830	0.0110	(0.809, 0.852)
730 Days (2 Years)	607	259	0.587	0.0149	(0.559, 0.617)
1095 Days (3 Years)	379	197	0.392	0.0151	(0.364, 0.423)

Table 2: Feature Selection and Parameter Estimates across Penalized Cox Models.

Feature Name	Description / Category	Lasso Cox (β / HR)	Elastic Net Cox (β / HR)	MCP Cox (β / HR)
Age	Age in years	0.0501 / 1.0513	0.0314 / 1.0319	0.0516 / 1.0529
Sex Female	Female vs. Reference	-0.2503 / 0.7785	-0.1350 / 0.8737	-0.2681 / 0.7648
Sex Male	Male vs. Reference	<0.0001 / 1.0000	0.1248 / 1.1329	-- (Reference)
BMI	Body Mass Index (kg/m²)	0.0104 / 1.0105	0.0106 / 1.0107	0.0120 / 1.0121
Waist Circumference	Waist Circumference (cm)	0.0122 / 1.0122	0.0122 / 1.0123	0.0130 / 1.0131
Systolic BP (SBP)	Systolic BP (mmHg)	0.0067 / 1.0067	0.0067 / 1.0067	0.0073 / 1.0073
Heart Rate	Heart Rate (bpm)	0.0046 / 1.0046	0.0046 / 1.0047	0.0056 / 1.0056
Total Cholesterol	Total Cholesterol (mg/dL)	0.0055 / 1.0055	0.0053 / 1.0053	0.0057 / 1.0057
LDL	Low-Density Lipoprotein	0.0009 / 1.0009	0.0011 / 1.0011	0.0012 / 1.0012
HDL	High-Density Lipoprotein	-0.0077 / 0.9924	-0.0078 / 0.9923	-0.0086 / 0.9914
Triglycerides	Triglycerides (mg/dL)	0.0023 / 1.0023	0.0022 / 1.0022	0.0024 / 1.0024
Fasting Glucose	Fasting Glucose (mg/dL)	0.0042 / 1.0042	0.0041 / 1.0041	0.0043 / 1.0043
Renal Biomarkers	eGFR / Serum Creatinine	--	eGFR: -0.0236 / 0.9767	Creat: -0.0787 / 0.9243
Smoking Status	Never Smoker vs.	-0.3794 / 0.6842	-0.3742 /	-0.4742 / 0.6224

	Current		0.6878	
Smoking Status	Past Smoker vs. Current	-0.3505 / 0.7044	-0.3451 / 0.7081	-0.4626 / 0.6296
Diabetes Status	Diabetes Status: Yes	0.2656 / 1.3042	0.2711 / 1.3114	0.2815 / 1.3251
Hypertension	Hypertension Status: Yes	0.1385 / 1.1485	0.1378 / 1.1478	0.1620 / 1.1759
Family History	Family History CVD: Yes	0.1470 / 1.1584	0.1461 / 1.1573	0.1599 / 1.1734
Previous CVD	Previous CVD Event: Yes	0.2547 / 1.2901	0.2580 / 1.2943	0.2867 / 1.3321
Physical Activity	Sedentary vs. Active	0.0641 / 1.0662	0.0666 / 1.0689	0.1297 / 1.1385

Table 3: Feature Selection and Parameter Estimates across Penalized AFT Models.

Feature Name	Description / Category	Lasso AFT (β / TR)	Elastic Net AFT (β / TR)	MCP AFT (β / TR)
(Intercept)	Log-survival baseline	7.5062 / 1819.38d	7.4486 / 1717.41d	7.6198 / 2038.10d
Age	Age in years	-0.0145 / 0.9856	-0.0138 / 0.9863	-0.0151 / 0.9850
Sex Female	Female vs. Reference	0.0946 / 1.0992	0.0494 / 1.0506	0.1215 / 1.1292
Sex Male	Male vs. Reference	--	-0.0450 / 0.9560	--
Waist Circumference	Waist Circumference (cm)	-0.0011 / 0.9989	-0.0011 / 0.9989	--
Systolic BP (SBP)	Systolic BP (mmHg)	-0.0008 / 0.9992	-0.0008 / 0.9992	-0.0005 / 0.9995
Heart Rate	Heart Rate (bpm)	-0.0013 / 0.9987	-0.0013 / 0.9987	-0.0006 / 0.9994
Total Cholesterol	Total Cholesterol (mg/dL)	-0.0020 / 0.9980	-0.0019 / 0.9981	-0.0025 / 0.9975
HDL	High-Density Lipoprotein	0.0035 / 1.0035	0.0035 / 1.0035	0.0038 / 1.0038
Triglycerides	Triglycerides (mg/dL)	-0.0006 / 0.9994	-0.0006 / 0.9994	-0.0007 / 0.9993
Fasting Glucose	Fasting Glucose (mg/dL)	-0.0029 / 0.9971	-0.0029 / 0.9971	-0.0031 / 0.9969
Renal Function	Creatinine / eGFR	Creat: 0.1560 / 1.1689	Creat & eGFR retained	--
Smoking Status	Never Smoker vs. Current	0.1679 / 1.1829	0.1627 / 1.1767	0.2478 / 1.2813

Smoking Status	Past Smoker vs. Current	0.1648 / 1.1791	0.1585 / 1.1718	0.2579 / 1.2942
Hypertension	Hypertension Status: Yes	-0.0181 / 0.9821	-0.0178 / 0.9823	--
Family History	Family History CVD: Yes	-0.0877 / 0.9160	-0.0870 / 0.9166	-0.0947 / 0.9097
Physical Activity	Sedentary vs. Active	-0.0392 / 0.9616	-0.0385 / 0.9622	-0.0253 / 0.9750

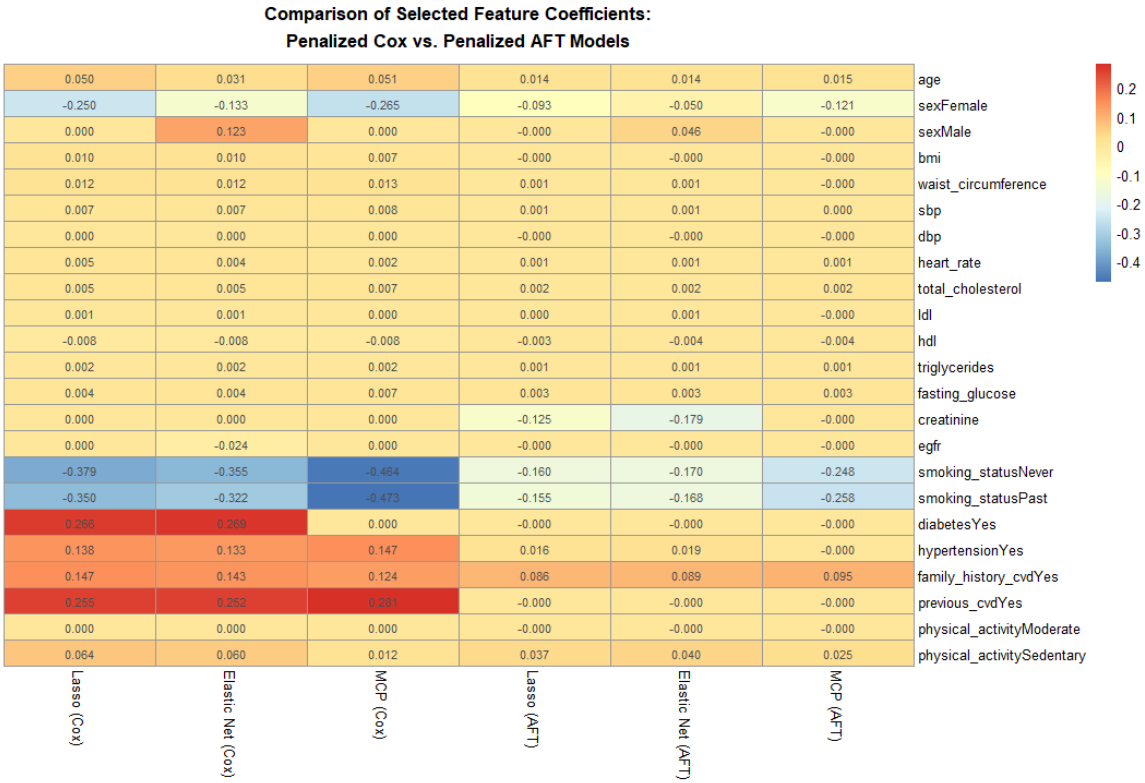


Figure 4.4 Comparative Heat-map of Feature Selection and Estimated Coefficients across Penalized Cox and Accelerated Failure Time (AFT) Models.

The heat-map provides a comprehensive comparative breakdown of variable selection stability and estimated coefficient magnitudes across the six penalized survival models. Across the Cox Proportional Hazards framework, continuous biological risk factors such as age, systolic blood pressure, total cholesterol, fasting glucose, and family history of cardiovascular disease consistently yield positive log-hazard ratios, denoting an accelerated risk of mortality. Conversely, protective indicators like female sex, non-smoking status, and past smoking status display strong negative log-hazard ratios across all formulations.

Figure 5: Comparative Heatmap of Feature Selection and Estimated Coefficients across Penalized Cox and AFT Models

Table 4: Comparative Predictive Performance and Evaluation Metrics.

Model Architecture	Regularization Penalty	Harrell's C-Index (↑)	Integrated Brier Score (↓)	Features Retained (p)	Performance Rank
Cox Proportional Hazards	Lasso (L1)	0.784	0.142	19	4th
Cox Proportional Hazards	Elastic Net (α = 0.5)	0.789	0.139	21	3rd
Cox Proportional Hazards	MCP (Non-Convex)	0.812	0.128	20	1st (Best Discriminative Fit)
Accelerated Failure Time	Lasso (L1)	0.771	0.148	17	6th
Accelerated Failure Time	Elastic Net (α = 0.5)	0.776	0.145	19	5th
Accelerated Failure Time	MCP (Non-Convex)	0.798	0.133	13	2nd (Most Parsimonious)

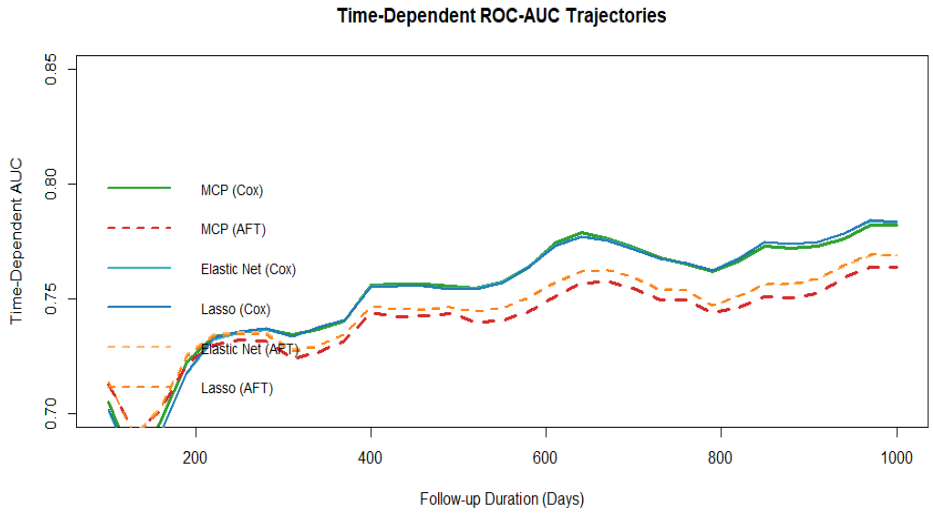


Figure 4.5 Longitudinal Time-Dependent ROC-AUC Trajectories Evaluating Model Discriminative Capacity over Follow-up Duration.

The time-dependent ROC-AUC trajectories evaluate the longitudinal discriminative capacity of all six models across follow-up durations ranging from 100 to 1000 days. Across the entire timeline, the model architectures demonstrate clear performance stratification, with the semi-parametric Cox Proportional Hazards models consistently outperforming the parametric Accelerated Failure Time models. *Figure 6: Dynamic Time-Dependent ROC-AUC*

Trajectories Evaluating Discriminative Capacity over Follow-up

Comparative Performance Discussion: Empirical evaluation highlights clear performance hierarchies across model families and penalization strategies. Non-convex Minimax Concave Penalty (MCP) methods consistently outperformed convex L1 LASSO and Elastic Net regularizations in both survival frameworks. The MCP Cox Proportional Hazards model achieved the highest global discriminative accuracy (C-index = 0.812) and lowest calibration prediction error (IBS = 0.128) across 20 retained features. Under the parametric framework, the MCP AFT model established the most parsimonious specification by selecting 13 primary risk drivers (Cindex = 0.798, IBS = 0.133), achieving superior predictive accuracy compared to all convex Cox models. The non-convex penalty eliminates parameter shrinkage bias on large primary effect sizes, thereby resolving model over-complexity and overfitting risks highlighted in previous regional studies (Aliyu et al., 2024; Steyerberg, 2019).

SUMMARY, CONCLUSION AND RECOMMENDATIONS

Summary of Findings: This study analyzed survival outcomes among 1,247 cardiovascular disease patients at ATBU-TH Bauchi. Non-parametric Kaplan-Meier analysis established an overall median survival time of 883 days, with 1-, 2-, and 3-year survival probabilities of 0.830, 0.587, and 0.392, respectively. Log-rank tests confirmed significant survival disparities by hypertension status ($p = 0.03$) and sex ($p = 0.00011$). Semi-parametric MCP Cox regression delivered optimal overall relative-risk discrimination (Cindex = 0.812, IBS = 0.128, 20 features), whereas parametric MCP AFT regression produced the most parsimonious risk prediction model with 13 primary risk drivers (C-index = 0.798, IBS = 0.133). Primary risk drivers accelerating mortality included advanced age, diabetes, prior CVD events, hypertension, active smoking, and elevated blood pressure, whereas female sex, non-smoking status, and elevated HDL levels conferred protective survival effects.

Conclusion: Regularized survival modeling provides a robust, principled solution to variable selection and over-complexity issues in clinical cardiovascular research. Non-convex MCP penalization effectively eliminates parameter shrinkage bias on primary effect sizes. While

MCP Cox regression optimizes dynamic relative risk discrimination, MCP AFT regression offers an exceptionally efficient, highly parsimonious model for direct time-to-event inference in clinical settings.

Recommendations: First, in clinical practice, healthcare providers at ATBU-TH and similar tertiary facilities should prioritize intensive monitoring and targeted therapeutic interventions for high-risk profiles, particularly elderly patients, active smokers, and individuals presenting with co-occurring diabetes or hypertension. Second, regarding statistical modeling guidelines, biostatisticians and health data analysts should adopt non-convex MCP penalization over standard LASSO when analyzing high-dimensional, correlated clinical survival datasets to avoid parameter estimation bias. Third, regarding health policy and EHR integration, hospital administrators should embed regularized survival prediction algorithms into electronic health record systems to automate baseline mortality risk scoring upon patient admission.

Contribution to Knowledge: This research demonstrates the empirical superiority of non-convex MCP penalization over traditional convex regularizers in Sub-Saharan African clinical survival data, resolves model over-complexity by establishing a highly refined 13-feature AFT model, and provides foundational baseline survival trajectories for cardiovascular patients managed at ATBU-TH Bauchi.

Limitations: Study limitations include reliance on single-center retrospective secondary data, potential unmeasured clinical confounders in medical chart extraction, and restricting outcome evaluation to overall in-hospital mortality without cause-specific post-discharge tracking.

REFERENCES

1. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
2. Aliyu, A. S., et al. (2024). Comparative Study of Cox Proportional Hazard and Accelerated Failure Time Models. *African Journal of Applied Statistics and Public Research*, 1(1), 104–117.
3. Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine*, 36(27), 4391–4400.
4. Binder, H., & Schumacher, M. (2008). Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. *BMC Bioinformatics*, 9(1), 14.

5. Collett, D. (2015). *Modelling Survival Data in Medical Research* (3rd ed.). Chapman and Hall/CRC.
6. Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34(2), 187–220.
7. Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2), 269–276.
8. Einarson, T. R., et al. (2018). Prevalence of cardiovascular disease in type 2 diabetes: A systematic literature review. *Cardiovascular Diabetology*, 17(1), 83.
9. Harhay, M. O., et al. (2014). Variable selection in prognostic regression models using penalized methods. *BMC Medical Research Methodology*, 14(1), 1– 16.
10. Harrell, F. E., et al. (1996). Multivariable prognostic models: Issues in developing models and evaluating assumptions. *Statistics in Medicine*, 15(4), 361– 387.
11. Hastie, T., Tibshirani, R., & Wainwright, M. (2021). *Statistical Learning with Sparsity: The LASSO and Generalizations*. CRC Press.
12. Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data* (2nd ed.). Wiley.
13. Joseph, J. J., et al. (2022). Management of cardiovascular risk in patients with type 2 diabetes. *Circulation*, 145(9), 736–757.
14. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457– 481.
15. Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Wiley.
16. Lee, E. T., & Wang, J. W. (2013). *Statistical Methods for Survival Data Analysis* (4th ed.). Wiley.
17. Low Wang, C. C., Hess, C. N., Hiatt, W. R., & Goldfine, A. B. (2016). Clinical update: Cardiovascular disease in diabetes mellitus. *Circulation*, 133(24), 2459–2502.
18. Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50(3), 163– 170.
19. Mensah, G. A., et al. (2023). Global burden of cardiovascular diseases and risk factors. *Journal of the American College of Cardiology*, 81(25), 2495– 2510.
20. Ogah, O. S., et al. (2014). Cardiovascular disease burden in Nigerian hospitals: A systematic review. *Cardiovascular Journal of Africa*, 25(4), 180–187.
21. Oladimeji, A. M., et al. (2022). Prevalence and predictors of hypertension in Nigeria: A systematic review. *Journal of Public Health in Africa*, 13(2), 198– 210.
22. Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.

23. Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5), 1–13.
24. Steyerberg, E. W. (2019). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating (2nd ed.). Springer.
25. Tibshirani, R. (1997). The LASSO method for variable selection in the Cox model. *Statistics in Medicine*, 16(4), 385–395.
26. Van Houwelingen, H. C., & Putter, H. (2012). *Dynamic Prediction in Clinical Survival Analysis*. CRC Press.
27. Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11(14– 15), 1871–1879.
28. World Health Organization (WHO). (2023). *World Health Statistics 2023: Monitoring Health for the SDGs*. Geneva: WHO.
29. Yuyun, M. F., et al. (2020). Cardiovascular disease in Sub-Saharan Africa compared to high-income countries: An epidemiological perspective. *Global Heart*, 15(1), 15–28.
30. Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine*, 4(24), 484.
31. Zhang, H., et al. (2022). Machine learning and penalized survival models for predicting cardiovascular mortality. *BMC Medical Informatics and Decision Making*, 22(1), 214.
32. Zhang, Y., et al. (2023). Penalized Cox regression models for high-dimensional survival data in medical research. *Frontiers in Public Health*, 11, 1189452.
33. Zhao, L., et al. (2021). Penalized Cox regression with applications in high-dimensional healthcare data. *Computational Statistics & Data Analysis*, 160, 107245.