Prognostic Factors for Mortality in Heart Failure Patients: A Cox Regression Analysis Challenging the Two-Variable Model

Rishabh Pabbi[1*] and M. Lesperance[1]

[1]Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada

rpabbi@uvic.ca, mlespera@uvic.ca .

Abstract

Background: Heart failure carries high mortality, yet optimal prognostic models remain debated. Chicco and Jurman (2020) proposed that ejection fraction and serum creatinine alone suffice for mortality prediction. We aimed to evaluate whether additional clinical variables improve prognostic accuracy.

Methods: We analyzed 299 heart failure patients from the Faisalabad Institute of Cardiology (2015). Kaplan-Meier curves with log-rank tests assessed univariate survival differences. Cox proportional hazards regression identified independent predictors. Models were compared using likelihood ratio tests and concordance indices.

Results: During median follow-up of 115 days, 96 patients (32.1%) died. In univariate analysis, serum creatinine >1.5 mg/dL was associated with 2.8-fold higher mortality (64.2% vs 22.8%, $\boldsymbol{p < 0.001}$). The final four-variable model included age (HR 1.05 per year, 95% CI 1.03–1.07), log-transformed ejection fraction (HR 0.16, 95% CI 0.08–0.29), serum creatinine (HR 1.39 per mg/dL, 95% CI 1.22–1.58), and hypertension (HR 1.54, 95% CI 1.02–2.34). This model achieved a concordance index of 0.74, significantly outperforming the two-variable literature model (C-index 0.68, $\boldsymbol{p < 0.001}$).

Conclusions: Age and hypertension provide significant independent prognostic information beyond ejection fraction and serum creatinine. A four-variable model offers superior risk stratification using routinely available clinical data.

Keywords: heart failure, prognosis, survival analysis, Cox regression, risk stratification

1 Introduction

Heart failure affects over 64 million people worldwide and carries a five-year mortality rate exceeding 50% [1]. Accurate prognostic assessment is essential for guiding treatment intensity, timing of advanced therapies such as transplantation, and informed discussions with patients about their expected outcomes.

The pathophysiology of heart failure involves complex interactions between cardiac dysfunction, neurohormonal activation, and end-organ damage. Left ventricular ejection fraction (EF) remains the cornerstone of heart failure classification, with reduced EF (<40%) indicating impaired systolic function [2]. The cardiorenal syndrome—characterized by bidirectional dysfunction between heart and kidneys—is particularly relevant, as renal impairment strongly predicts adverse outcomes through mechanisms including volume overload, electrolyte disturbances, and medication intolerance [3].

Several clinical variables have established prognostic value in heart failure. Serum creatinine reflects renal function and identifies patients with cardiorenal syndrome. Serum sodium serves as a marker of neurohormonal activation, with hyponatremia reflecting activation of the renin-angiotensin-aldosterone system [4]. Hypertension and diabetes are common comorbidities that independently contribute to cardiovascular risk.

Multiple prognostic models have been developed for heart failure, including the Seattle Heart Failure Model [5] and the MAGGIC risk score [4]. These models incorporate clinical, laboratory, and treatment variables to predict mortality. However, simpler models using fewer variables may be more practical for routine application.

Chicco and Jurman recently analyzed this dataset using machine learning techniques and concluded that ejection fraction and serum creatinine alone were sufficient for mortality prediction [6]. However, their analysis did not employ formal survival analysis methods or systematically evaluate whether additional clinical variables provide incremental prognostic information.

We hypothesized that a model incorporating age and cardiovascular comorbidities would outperform the two-variable approach. Our objectives were to: (1) identify independent prognostic factors using Cox proportional hazards regression; (2) develop a parsimonious predictive model; and (3) formally compare model performance against the published two-variable model.

2 Methods

2.1 Study Population

We analyzed the Heart Failure Clinical Records dataset from the UCI Machine Learning Repository [6, 7]. The cohort comprised 299 consecutive patients with heart failure (NYHA Class III–IV) treated at the Faisalabad Institute of Cardiology, Pakistan, between April and December 2015. All patients were followed from hospital admission until death or end of follow-up. The dataset contained no missing values.

2.2 Variables

The primary outcome was all-cause mortality during follow-up. Time-to-event was measured in days from admission to death or censoring at study end.

Predictor variables included demographics (age, sex), cardiac function (ejection fraction), cardiovascular comorbidities (hypertension, diabetes, smoking), hematologic markers (anemia, platelets), and biochemical markers (serum creatinine, serum sodium, creatinine phosphokinase).

2.3 Statistical Analysis

Baseline characteristics were summarized as means with standard deviations for continuous variables and frequencies with percentages for categorical variables. Between-group comparisons used t-tests and chi-square tests as appropriate.

Overall survival was estimated using the Kaplan-Meier method with log-log transformed 95% confidence intervals [8]. Survival differences between groups were assessed using log-rank tests, with p-values displayed on survival curves. Median follow-up was calculated using the reverse Kaplan-Meier method.

Cox proportional hazards regression was used to estimate hazard ratios (HR) with 95% confidence intervals [9]. We first performed univariate analysis for each predictor, then built multivariable models systematically: a full model (all 11 variables), clinical variables only, laboratory values only, significant univariate predictors, and the two-variable literature model (ejection fraction and serum creatinine).

Model comparison used likelihood ratio tests for nested models and Akaike Information Criterion (AIC) for non-nested comparisons. Discrimination was assessed using Harrell's concordance index (C-statistic) [10].

The proportional hazards assumption was tested using Schoenfeld residuals with the method of Grambsch and Therneau [11]. Functional form was assessed using martingale residuals, with log-transformation applied where non-linearity was detected. Influential observations were identified using deviance residuals exceeding $|2.5|$.

Random Survival Forest analysis was performed to validate variable importance using permutation-based methods [12]. This machine

learning approach provides complementary evidence for predictor selection independent of the Cox model assumptions.

All analyses were performed in R version 4.x using the survival, survminer, and randomForestSRC packages [13]. Statistical significance was set at $\alpha = 0.05$ (two-sided).

3 Results

3.1 Patient Characteristics

Table 1 presents baseline characteristics. The cohort included 299 patients with mean age 60.8 years (SD 11.9); 65% were male. Mean ejection fraction was 38% (SD 12%), indicating predominantly reduced systolic function. Comorbidities were common: 42% had diabetes, 43% had anemia, and 35% had hypertension.

Patients who died were significantly older (65.2 vs 58.8 years, $p < 0.001$), had lower ejection fraction (33.5% vs 40.3%, $p < 0.001$), higher serum creatinine (1.84 vs 1.18 mg/dL, $p < 0.001$), and lower serum sodium (135.3 vs 137.2 mEq/L, $p = 0.001$). No significant differences were observed for sex, diabetes, smoking, anemia, or CPK.

3.2 Overall Survival

During median follow-up of 115 days (IQR 51–203), 96 patients (32.1%) died. Figure 1 shows the overall Kaplan-Meier survival curve. Estimated survival was 88% at 30 days, 76% at 90 days, and 66% at 180 days. Median survival was not reached.

3.3 Stratified Survival Analysis

Figure 2 presents Kaplan-Meier curves stratified by key prognostic factors with log-rank p-values.

Serum Creatinine was the strongest univariate predictor. Patients with elevated creatinine (>1.5 mg/dL) had dramatically worse survival: 64.2% mortality versus 22.8% in those with normal values (log-rank $p < 0.001$).

Ejection Fraction showed significant prognostic value. Patients with reduced EF (<40%) had 40.1% mortality compared to 19.7% in those with preserved function (log-rank $p = 0.002$).

Age was significantly associated with outcomes. Patients older than 60 years had 38.0% mortality versus 27.2% in younger patients (log-rank $p = 0.030$).

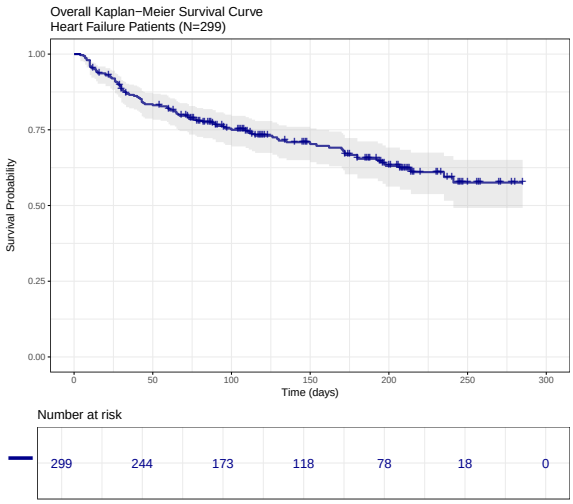


Fig. 1: Overall Kaplan-Meier survival curve for 299 heart failure patients. Shaded area represents 95% CI. Tick marks indicate censored observations. 30-day survival: 88%; 180-day survival: 66%.

Hypertension showed a trend toward worse outcomes (38.1% vs 28.9% mortality, log-rank $p = 0.073$), which became significant in multivariable analysis.

3.4 Univariate Cox Regression

Figure 3 shows hazard ratios from univariate Cox regression. Five variables achieved statistical significance ($p < 0.05$): serum creatinine (HR 1.34 per mg/dL, $p < 0.001$), ejection fraction (HR 0.96 per %, $p < 0.001$), age (HR 1.04 per year, $p < 0.001$), serum sodium (HR 0.93 per mEq/L, $p < 0.001$), and hypertension (HR 1.61, $p = 0.018$). Sex, diabetes, smoking, anemia, platelets, and CPK were not significantly associated with mortality.

3.5 Variable Importance from Random Survival Forest

Random Survival Forest analysis (Figure 4) confirmed serum creatinine and ejection fraction as the two most important predictors, with age ranking third—supporting its inclusion in our final model.

Table 1: Baseline Patient Characteristics

Characteristic	Overall (N=299)	Survivors (N=203)	Deaths (N=96)	P-value
Demographics				
Age, years	60.8 ± 11.9	58.8 ± 10.6	65.2 ± 10.1	<0.001
Male sex	194 (64.9%)	132 (65.0%)	62 (64.6%)	1.000
Cardiac Function				
Ejection fraction, %	38.1 ± 11.8	40.3 ± 10.9	33.5 ± 11.8	<0.001
EF <40%	182 (60.9%)	109 (53.7%)	73 (76.0%)	<0.001
Comorbidities				
Hypertension	105 (35.1%)	65 (32.0%)	40 (41.7%)	0.214
Diabetes	125 (41.8%)	88 (43.3%)	37 (38.5%)	1.000
Anemia	129 (43.1%)	83 (40.9%)	46 (47.9%)	0.307
Current smoker	96 (32.1%)	66 (32.5%)	30 (31.3%)	0.932
Laboratory Values				
Serum creatinine, mg/dL	1.39 ± 1.03	1.18 ± 0.65	1.84 ± 1.46	<0.001
Creatinine >1.5 mg/dL	67 (22.4%)	24 (11.8%)	43 (44.8%)	<0.001
Serum sodium, mEq/L	136.6 ± 4.4	137.2 ± 4.0	135.3 ± 5.4	0.001
CPK, mcg/L	582 ± 970	540 ± 754	670 ± 1169	0.280
Outcomes				
Follow-up, days, median	115 (51–203)	172 (96–215)	51 (23–107)	<0.001
Deaths	96 (32.1%)	—	—	—

Table 2: Comparison of Cox Regression Models

Model	Var	C-index	AIC	vs Full	vs Red
Full model	11	0.74	Ref	—	<0.001
Significant	5	0.73	+3.8	0.22	<0.001
Clinical only	6	0.71	+19.7	<0.001	0.03
Laboratory	4	0.64	+45.1	<0.001	0.84
Final	**4**	**0.74**	**+1.2**	**0.20**	**<0.001**
Reduced	2	0.68	+24.9	<0.001	—

C-index = concordance index; AIC = Akaike Information Criterion; Var = variables; Red = reduced model. P-values from likelihood ratio tests.

Table 3: Final Cox Proportional Hazards Model

Variable	HR	95% CI	P-value
Age (per year)	1.05	1.03–1.07	<0.001
Log(ejection fraction)	0.16	0.08–0.29	<0.001
Serum creatinine (per mg/dL)	1.39	1.22–1.58	<0.001
Hypertension (yes vs no)	1.54	1.02–2.34	0.040

Model performance: C-index = 0.74 (SE 0.03). Global PH test: $p = 0.195$ (assumption satisfied). HR = hazard ratio; CI = confidence interval.

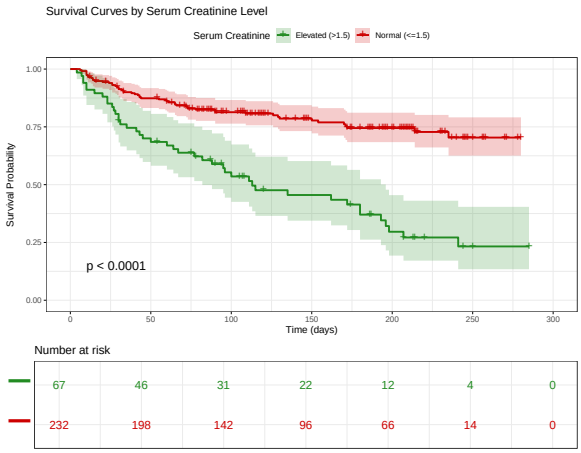
3.6 Multivariable Cox Regression

Table 2 compares the performance of different Cox regression models. The critical comparison between our model and the full 11-variable model showed no significant difference ($p = 0.22$), indicating the simpler model captures equivalent prognostic information. However, our model significantly outperformed the two-variable literature model ($p < 0.001$).

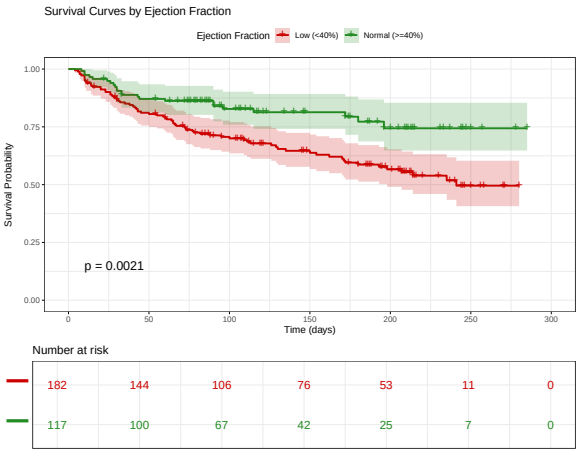
3.7 Final Model

Based on martingale residual analysis showing non-linearity for ejection fraction, we applied a log-transformation. The final four-variable model is presented in Table 3.

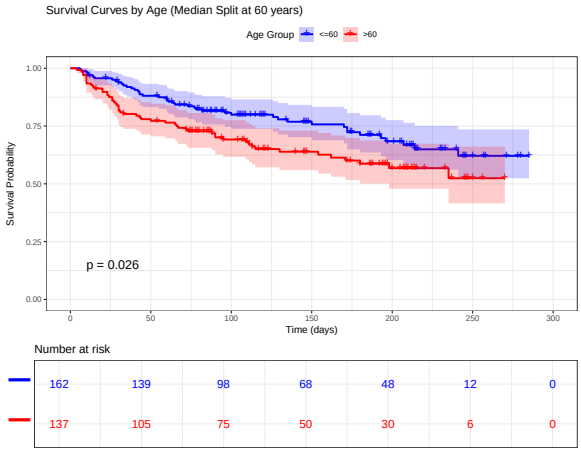
Clinical interpretation: Each 10-year increase in age was associated with 58% higher mortality risk. Doubling of ejection fraction was associated with 84% lower mortality risk. Each 1 mg/dL increase in serum creatinine was associated with 39% higher mortality risk. Presence of hypertension was associated with 54% higher mortality risk.



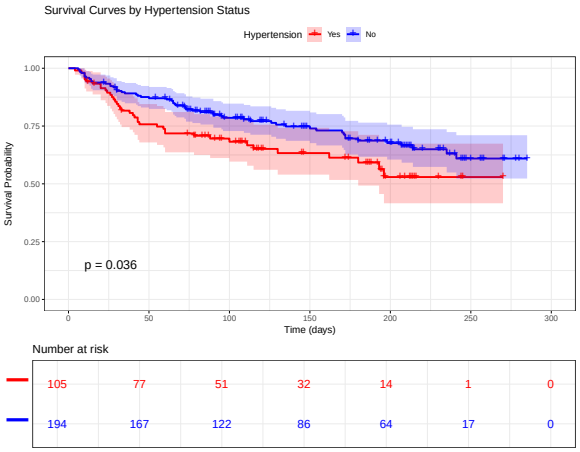
(a) Serum Creatinine ($p < 0.001$)



(b) Ejection Fraction ($p = 0.002$)



(c) Age ($p = 0.030$)



(d) Hypertension ($p = 0.073$)

Fig. 2: Kaplan-Meier survival curves stratified by key prognostic factors. (a) Serum creatinine: elevated (>1.5 mg/dL) vs normal; (b) Ejection fraction: reduced (<40%) vs preserved; (c) Age: >60 vs ≤60 years; (d) Hypertension status. Log-rank p-values shown.

3.8 Model Diagnostics

The proportional hazards assumption was tested using Schoenfeld residuals. The global test was not significant ($p = 0.195$), confirming overall model validity. Individual tests: age ($p = 0.56$), log-EF ($p = 0.031$, borderline), hypertension ($p = 0.84$), creatinine ($p = 0.34$).

Figure 5 and Figure 6 present diagnostic plots. Martingale residuals confirmed appropriate functional forms after log-transformation. Six patients (2%) had deviance residuals exceeding |2.5|, an acceptable proportion.

4 Discussion

4.1 Principal Findings

In this cohort of 299 heart failure patients, we identified four independent predictors of mortality: age, ejection fraction, serum creatinine, and hypertension. Our four-variable model achieved a concordance index of 0.74, significantly outperforming the two-variable model proposed by Chicco and Jurman [6].

The key finding is that age and hypertension provide substantial independent prognostic

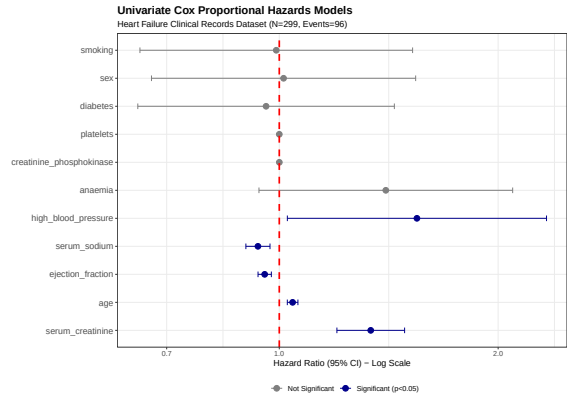


Fig. 3: Forest plot of univariate Cox proportional hazards regression results. Blue = significant ($p < 0.05$); gray = not significant. Vertical dashed line indicates HR=1.0.

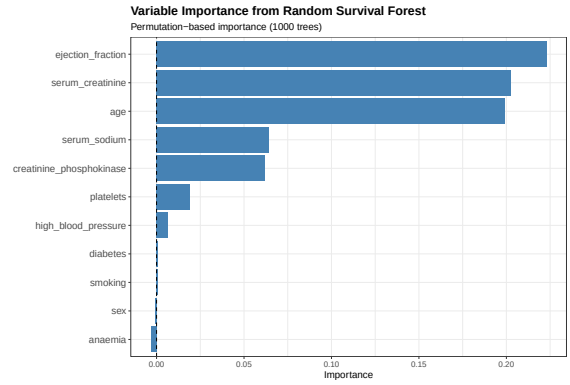


Fig. 4: Variable importance from Random Survival Forest (1000 trees). Permutation-based importance confirms top predictors.

information beyond ejection fraction and serum creatinine alone. The likelihood ratio test comparing our model to the literature model was highly significant ($p < 0.001$), and the concordance index improved from 0.68 to 0.74—representing meaningful clinical improvement in risk discrimination.

4.2 Comparison with Previous Studies

Chicco and Jurman concluded that machine learning models using only ejection fraction and serum creatinine performed comparably to models with all variables [6]. While their analysis correctly identified these as the two strongest individual

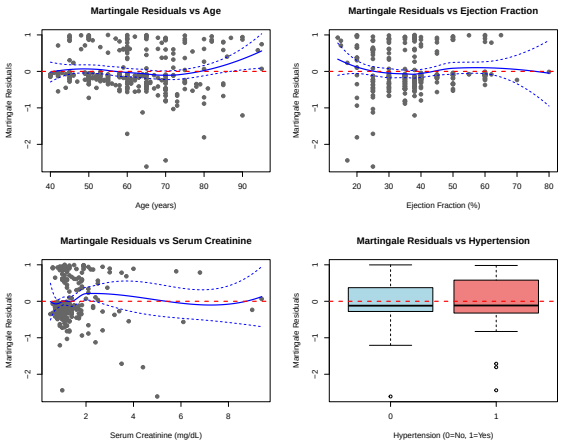


Fig. 5: Martingale residual plots assessing functional form for final model covariates. LOESS smoothing curves with 95% confidence bands shown. Smoothed curves approximately centered at zero indicate appropriate functional specification.

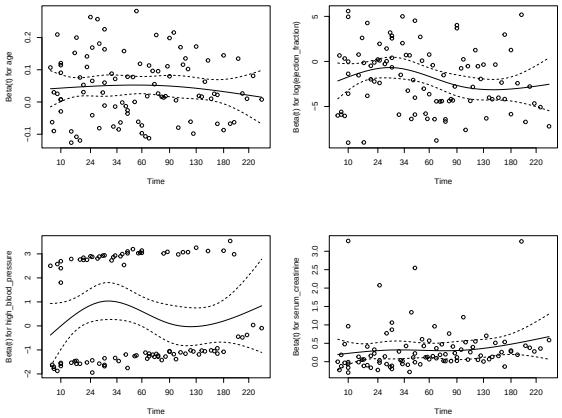


Fig. 6: Schoenfeld residual plots testing proportional hazards assumption. Smooth curves should be horizontal if assumption holds. Global test $p = 0.195$ confirms overall model validity.

predictors, our formal survival analysis reveals that this conclusion is incomplete.

Our results demonstrate that age remained independently significant (HR 1.05 per year, $p < 0.001$) after adjusting for ejection fraction and creatinine, hypertension provided additional prognostic information (HR 1.54, $p = 0.040$), and

the improvement from adding these variables was statistically significant by likelihood ratio test.

The discrepancy likely reflects differences in analytical approach. Machine learning feature importance measures univariate predictive power, whereas Cox regression estimates independent effects after mutual adjustment. Age and hypertension, while weaker than creatinine in univariate analysis, contribute non-redundant prognostic information.

Our C-index of 0.74 compares favorably with established heart failure risk scores. The MAGGIC score achieved a C-statistic of 0.74 in its derivation cohort [4], while the Seattle Heart Failure Model reported values of 0.73–0.76 [5]. Importantly, our model uses only four readily available variables compared to the 13 variables in MAGGIC or the complex algorithm of the Seattle model.

4.3 Clinical Implications

Our model enables practical bedside risk stratification using routinely available data:

High-risk profile: Age >70 years, ejection fraction <30%, serum creatinine >2.0 mg/dL, hypertension present.

Lower-risk profile: Age <60 years, ejection fraction ≥40%, serum creatinine ≤1.2 mg/dL, no hypertension.

This information could inform decisions regarding monitoring intensity, medication optimization, and timing of referral for advanced therapies.

4.4 Biological Plausibility

The four-variable model has strong biological rationale. Ejection fraction directly measures cardiac pump function. Serum creatinine reflects renal perfusion and the severity of the cardiorenal syndrome. Age captures accumulated physiologic decline and comorbidity burden. Hypertension indicates long-standing vascular disease and increased afterload.

4.5 Limitations

Several limitations warrant consideration. First, this is a single-center study from Pakistan, and findings may not generalize to other populations. Second, maximum follow-up was 285 days, precluding assessment of longer-term outcomes. Third, important prognostic biomarkers including BNP and troponin were not available. Fourth, the data were collected in 2015, before widespread use of SGLT2 inhibitors and sacubitril/valsartan [2]. Fifth, we did not perform external validation.

The borderline violation of proportional hazards for log-ejection fraction ($p = 0.031$) suggests that its prognostic effect may diminish over time, though the global test remained acceptable ($p = 0.195$).

4.6 Future Directions

External validation in diverse populations is essential before clinical implementation. Addition of natriuretic peptides could further improve discrimination. Development of a simplified integer-based risk score would facilitate bedside application.

5 Conclusions

A four-variable Cox regression model incorporating age, ejection fraction, serum creatinine, and hypertension provides superior prognostic discrimination compared to ejection fraction and serum creatinine alone. Age and hypertension contribute significant independent information that should not be ignored in heart failure risk stratification. This model offers a clinically practical tool using routinely available measurements.

Acknowledgments. This work was completed as part of STAT 559: Survival Analysis at the University of Victoria. The author thanks the course instructor (M. Lesperance) for guidance on survival analysis methodology.

Author Contributions. The author conceived the study, performed all analyses, and wrote the manuscript.

Funding. No funding was received for this study.

Data Availability. The Heart Failure Clinical Records dataset is publicly available from the UCI Machine Learning Repository, and the complete R code for all analyses is available in the author's GitHub repository.

Declarations. Conflict of Interest: The author declares no conflicts of interest.

Ethics Approval: Not applicable. This study used publicly available, de-identified data.

Appendix A Supplementary Figures

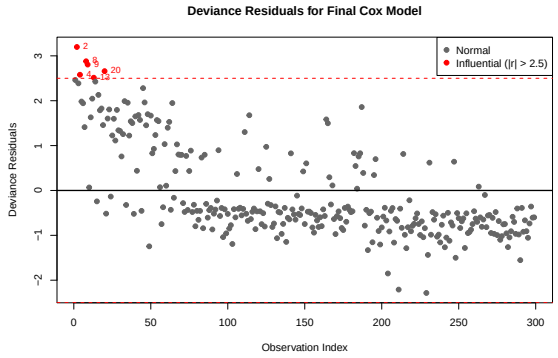


Fig. A1: Deviance residuals identifying influential observations. Reference lines at ±2.5. Six patients (2%) exceeded threshold.

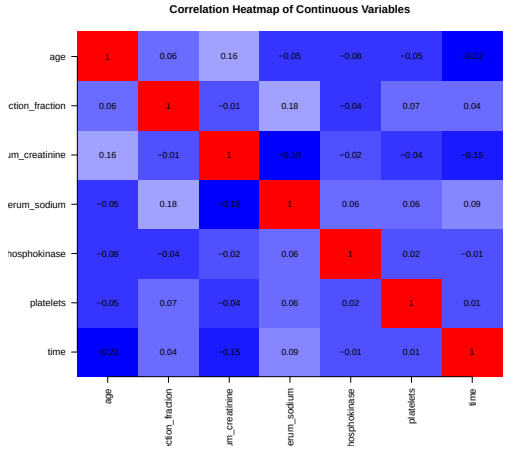


Fig. A2: Correlation heatmap. All correlations $|r| < 0.25$, indicating low multicollinearity.

References

[1] Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2016;37(27):2129–2200.

[2] McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42(36):3599–3726.

[3] Ronco C, Haapio M, House AA, et al. Cardiorenal syndrome. *J Am Coll Cardiol.* 2008;52(19):1527–1539.

[4] Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39,372 patients from 30 studies. *Eur Heart J.* 2013;34(19):1404–1413.

[5] Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation.* 2006;113(11):1424–1433.

[6] Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inform Decis Mak.* 2020;20(1):16.

[7] UCI Machine Learning Repository. Heart Failure Clinical Records Data Set. 2020. Available from: Heart Failure Clinical Records Data Set

[8] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53(282):457–481.

[9] Cox DR. Regression models and life-tables. *J R Stat Soc Series B.* 1972;34(2):187–220.

[10] Harrell FE. *Regression Modeling Strategies.* 2nd ed. Springer; 2015.

[11] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81(3):515–526.

[12] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* 2008;2(3):841–860.

[13] Moore DF. *Applied Survival Analysis Using R.* Springer; 2016.