Assaf Bitton

Statistical Modeling of Mortality Risk in Heart Failure Patients

Introduction

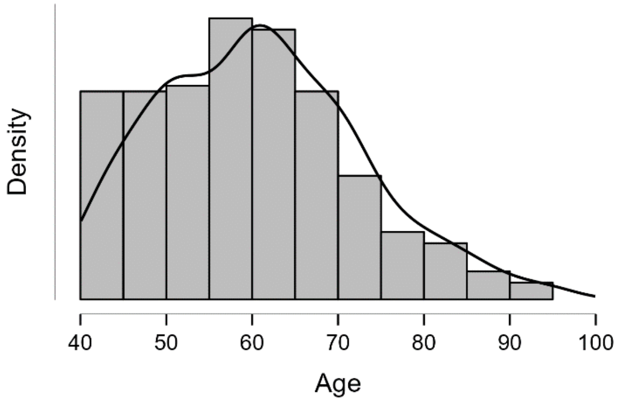
Given the extensive capabilities of predictive modeling, I decided to explore a topic that has real-world implications. My goal is to provide powerful insight into the leading cause of mortality around the world, cardiovascular disease. The report utilizes a dataset of 299 patients, which includes twelve clinical features to build a logistic regression model that predicts the likelihood of mortality, referred to as the 'Death Event'.

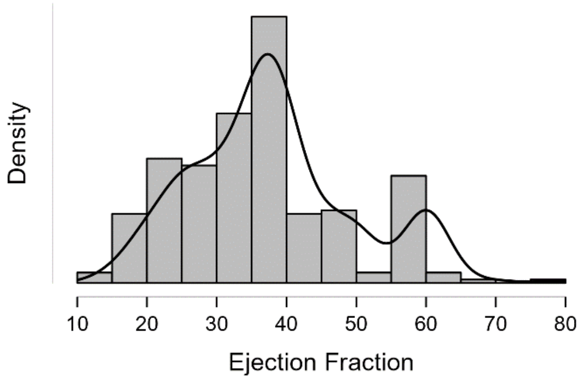
Data Description

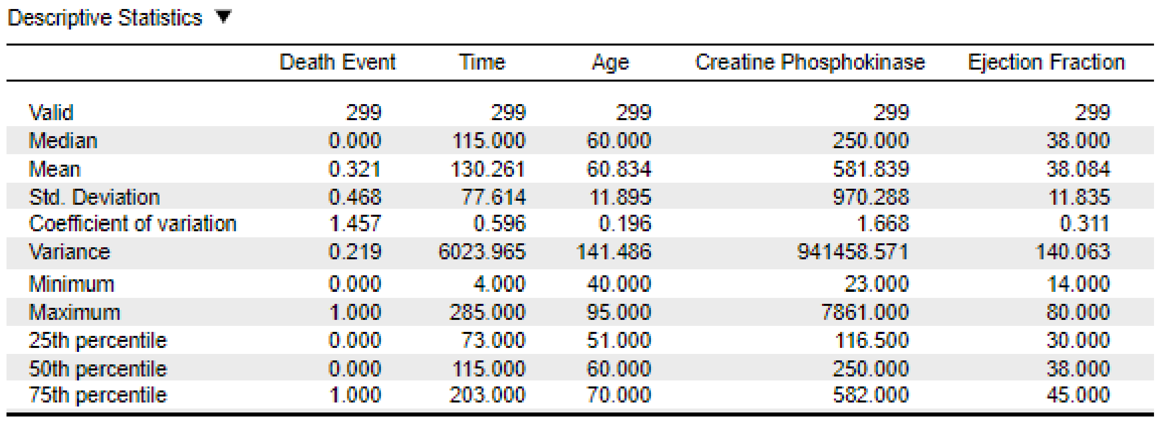
The Faisalabad Institute of Cardiology and the Allied Hospital are responsible for collecting this dataset. According to the source, all 299 patients had left ventricular systolic dysfunction, a type of heart failure where the heart's left ventricle does not contract effectively, leading to less oxygen-rich blood pumping out to the body. A description of the more complex medical terms will facilitate a better understanding of the data. Anemia is a binary variable indicating "a condition marked by a deficiency of red blood cells." Creatinine Phosphokinase is a continuous variable indicating damage to the heart muscle. Ejection Fraction is a continuous variable measuring the percentage of blood leaving the heart for every contraction. 'Platelets' is a continuous variable and is critical for blood clotting. Serum Creatinine is a continuous variable that measures creatinine levels in the blood; high levels indicate heart failure. Serum Sodium is a continuous variable that measures sodium levels in the blood; low levels indicate heart failure. For clarification, the variable 'Sex' is denoted by 1 for male and 0 for female. The remaining variables are 'Age' (continuous), 'Diabetes' (binary), 'High Blood Pressure' (binary), 'Smoking' (binary), and 'Time' (continuous).

Exploratory Data Analysis

Given the mixture of binary and continuous data in this dataset, I had to consider a variety of parameters. While mean and median could have given insight into the continuous data, it was less useful for the binary data. For example, observing the mean of 'Age' (μ = 60.834), it's apparent that on average, patients with cardiovascular disease are middle aged. In addition, 'Ejection Fraction' had similar mean and median (μ =38.084 and x ~ = 38), indicating that on average, patients pumped 38% of their blood out of their heart and into their body with every contraction. Given that both had a normal distribution, I felt that it was appropriate to use mean.



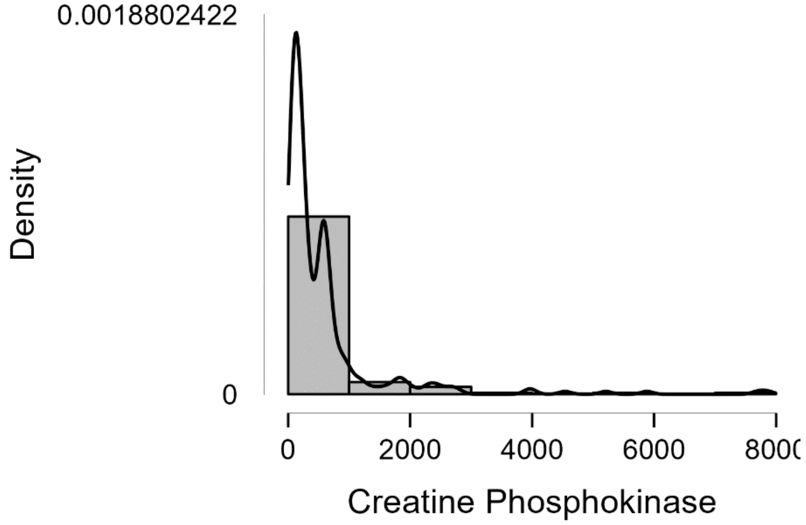




Descriptive Statistics ▼

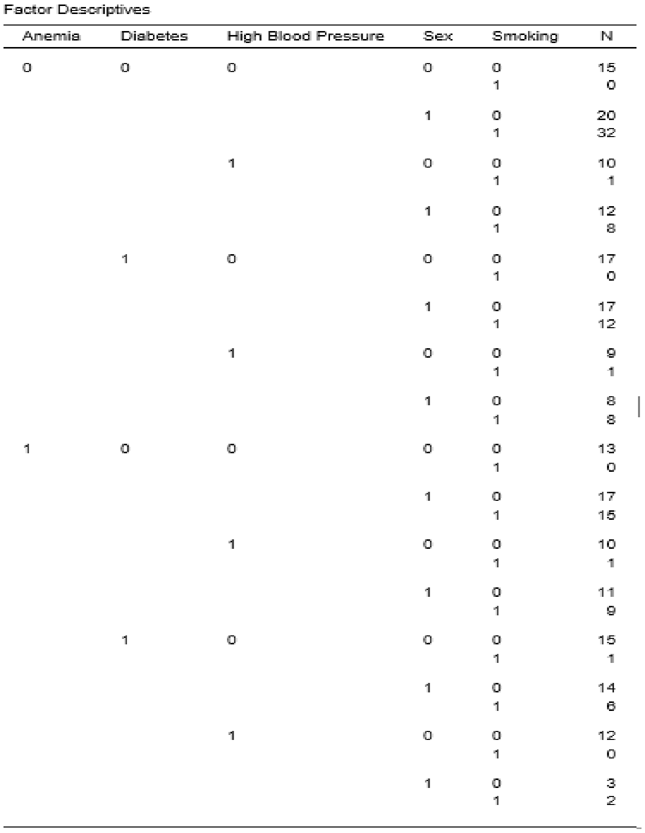
	Death Event	Time	Age	Creatine Phosphokinase	Ejection Fraction
Valid	299	299	299	299	299
Median	0.000	115.000	60.000	250.000	38.000
Mean	0.321	130.261	60.834	581.839	38.084
Std. Deviation	0.468	77.614	11.895	970.288	11.835
Coefficient of variation	1.457	0.596	0.196	1.668	0.311
Variance	0.219	6023.965	141.486	941458.571	140.063
Minimum	0.000	4.000	40.000	23.000	14.000
Maximum	1.000	285.000	95.000	7861.000	80.000
25th percentile	0.000	73.000	51.000	116.500	30.000
50th percentile	0.000	115.000	60.000	250.000	38.000
75th percentile	1.000	203.000	70.000	582.000	45.000

In the table above, the range of 'Creatine Phosphokinase' (Min = 23, Max = 7,861) is extremely wide. I took note of this extreme parameter, which led to an investigation of the variable.



Upon further investigation, I found that 'Creatine Phosphokinase' was heavily right skewed. I considered removing the outliers but given that they occurred more than a few times and given that the data came from a reputable source (low chance of incorrect data entry) I decided that these outliers were relevant to the data and should not be removed. The immediate hypothesis was that 'Creatine Phosphokinase' would be a bad predictor.

To explore the binary data, I made use of factor descriptives. Factor descriptives is a table that shows the different combinations of factors among all the observations in the data. Statisticians suggest that the more varied and numerous the combinations of factors are, the more reliable they are in a prediction model. In the table below, the factors 'Anemia,' 'Diabetes,' 'High Blood Pressure,' 'Sex,' and 'Smoking' do in fact appear across many different combinations across the observations (e.g., N > 10 for 16 of the possible 32 combinations). While there are instances of combinations that occur very few or zero times, the factor descriptives display a solid matrix of combinatory variety.



Factor Descriptives

Anemia	Diabetes	High Blood Pressure	Sex	Smoking	N
0	0	0	0	0	15
				1	0
			1	0	20
				1	32
		1	0	0	10
				1	1
			1	0	12
				1	8
	1	0	0	0	17
				1	0
			1	0	17
				1	12
		1	0	0	9
				1	1
			1	0	8
				1	8
1	0	0	0	0	13
				1	0
			1	0	17
				1	15
		1	0	0	10
				1	1
			1	0	11
				1	9
	1	0	0	0	15
				1	1
			1	0	14
				1	6
		1	0	0	12
				1	0
			1	0	3
				1	2

To gain insight into the relative variability of the data, I used the coefficient of variation. The variable with the highest coefficient was Creatine Phosphokinase (CV = 1.668), which proved to be a meaningless predictor in the regression. By contrast, Age (CV = 0.196) had the lowest coefficient of variation and proved to be a meaningful predictor. However, this parameter only allows for a basic insight into the data and was not sufficient in determining its ability to predict the outcome, given that it was in fact ‘Serum Creatinine’ (CV = 0.742) which was the strongest predictor. Furthermore, analyzing the standard deviation for the continuous data gave me similar insight, s = 970.288 for ‘Creatine Phosphokinase and s = 1.035 for ‘Serum Creatinine,’ both sample standard deviations indicating that the first variable deviated a lot from the mean while the second variable did not. Analyzing the standard deviation for the binary data indicated that the outcomes for the factors were almost equal (Approximately Equal cases of 1,0), given that the range was 0.469 < s < 0.496.

Predictive Modeling

Dependent Variable Identification

I chose ‘Death Event’ as the dependent variable for its function in predicting mortality in patients with cardiovascular disease.

Model Selection and Justification

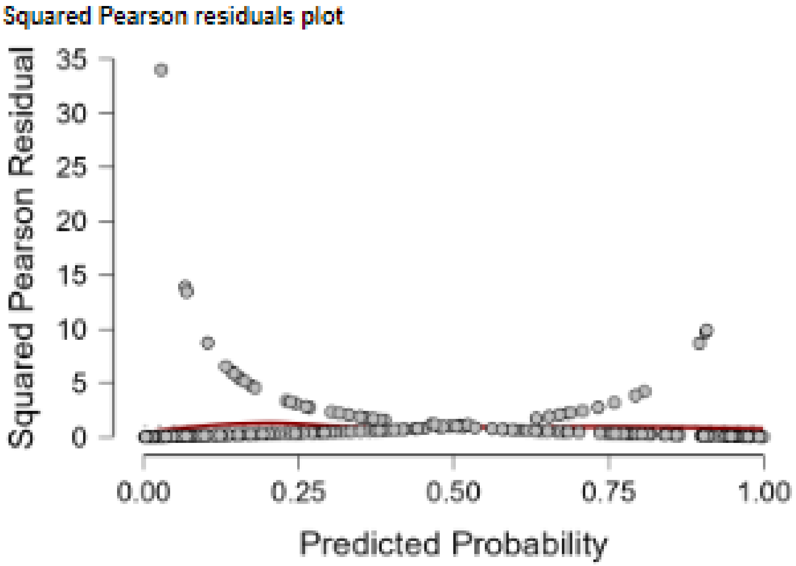
I chose logistic regression as the best model for the data given the combination of continuous and binary variables in the dataset. To adjust for the difference in magnitude across the continuous variables, I standardized the data before running the logistic regression. Below is the model summary for the logistic regression.

Model Fit Assessment

Model Summary - Death Event

Model	Deviance	AIC	BIC	df	X^2	p	McFadden R^2	Nagelkerke R^2	Tjur R^2	Cox & Snell R^2
H_0	375.349	377.349	381.049	298						
H_1	219.553	245.553	293.659	286	155.795	< .001	0.415	0.568	0.465	0.406

The rule of thumb in logistic regression is to have at least 10 events per predictor variable to accurately estimate the variable coefficients. Given a sample size of 299 and 12 predictor variables, the threshold is met. Although 286 degrees of freedom for the alternative hypothesis (H1) implies a complex model, a p value of < .001 shows that despite the model's complexity, the predictor variables have a significant relationship with the outcome variables. A pseudo-R-squared (McFadden R^2) suggests that the model accounts for about 41.5% of the variance. The remaining statistics, Nagelkerke R^2, Tjur R^2, and Cox & Snell R^2 are all pseudo-R-squared measures that assess how much of the variance the model accounts for. Given that those statistics range from 0.406 < pseudo-R^2 < 0.568, the model has moderate-to-good predictive power.



The expected value of the squared residuals is 1 and by looking at the smoother (shown by the red line) and assessing how much it lies on the expected value I determined that the model does not suffer from over dispersion. This graph is also useful for identifying outliers (note the value at the top left corner).

Model Usefulness

Confusion matrix

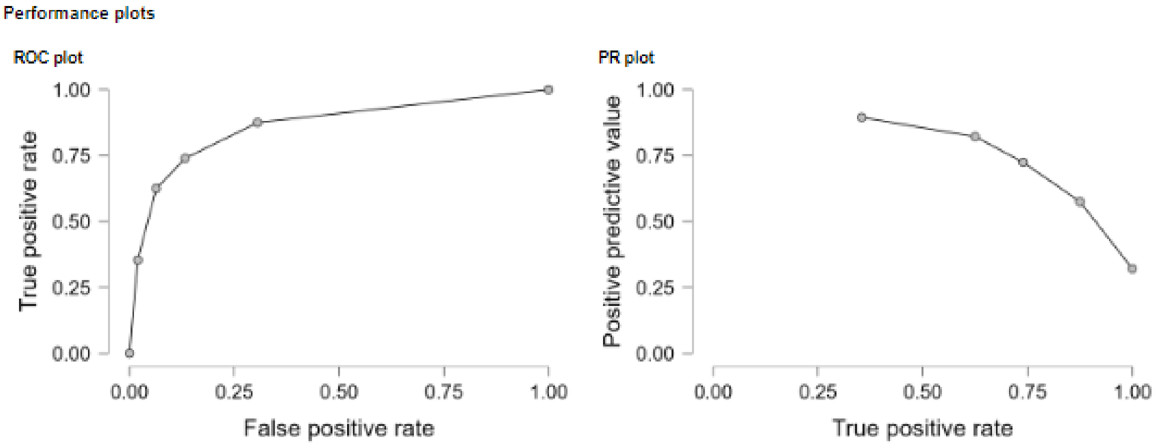
	Predicted		
Observed	0	1	% Correct
0	187	16	92.118
1	27	69	71.875
Overall % Correct			85.619

Note. The cut-off value is set to 0.5

Performance metrics

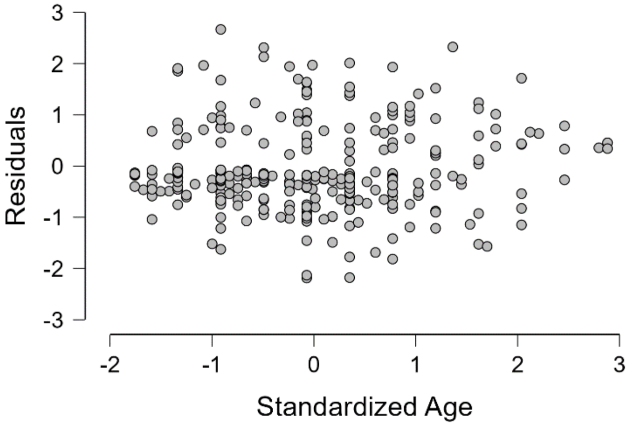
	Value
Accuracy	0.856
AUC	0.897
Sensitivity	0.719
Specificity	0.921
Precision	0.812
F-measure	0.762

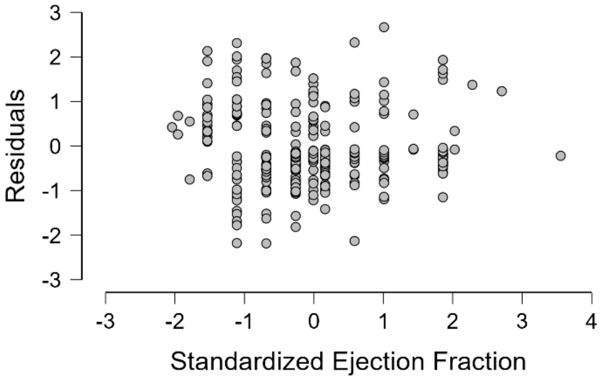
While a confusion matrix cannot determine model fit as a stand-alone statistic, it can provide insight into the accuracy, precision, recall, and specificity of the model. The model correctly predicted ‘Survival’ (True Negative) 92.118% of the time and ‘Death’ (True Positive) 71.875% of the time, with an overall 85.619% correct proportion of all true results. By looking at the chart above and to the right, the F-measure is equal to 0.762, which is a reasonable identifier for a robust model. The AUC is equal to 0.897, which is a strong indicator that the model is excellent at distinguishing between the outcome events (mortality due to heart failure).



The ROC plot shows the trade-off between the true positive rate and the false positive rate at varying thresholds. The greater the area under the curve, the larger the AUC indicator. Since the curve rises quickly, it suggests the model is effective at distinguishing between outcomes. The PR plot displays the trade-off between precision and recall at varying thresholds. A good indicator for a PR plot is for the curve to be at the upper right corner of the graph. These graphs further support the idea that the model is a good fit for the dataset.

Residual Analysis





The two graphs above are residual plots for 'Age' (standardized) and 'Ejection Fraction' (standardized). For clarity purposes, when creating residual plots with un-standardized data, the plots looked identical to the ones created with standardized data. An analysis showed that 7 out of the 7 continuous data created residual plots that were randomly and evenly distributed, indicating homoscedasticity and independence of errors.

Predictor Interaction

Please refer to the 5th page of the "Reference Section" to look at the correlation plots. As you can see, only 6 of the 21 correlation plots have some kind of visual relationship, indicating that most of the data are independent of one another. Two variables with a weak relationship are 'Age' and 'Time.' Since 'Time' is negatively correlated with 'Death Event' and 'Age' is positively correlated with 'Death Event,' it can be shown that the older the age of the patient, the less time he survives following heart failure.

Multicollinearity Diagnostics

	Tolerance	VIF
Standardized Age	0.906	1.104
Standardized Creatine Phosphokinase	0.921	1.086
Standardized Ejection Fraction	0.853	1.173
Standardized Platelets	0.957	1.045
Standardized Serum Creatinine	0.907	1.102
Standardized Serum Sodium	0.934	1.071
Standardized Time	0.868	1.152
Anemia	0.897	1.115
Diabetes	0.950	1.052
High Blood Pressure	0.941	1.063
Sex	0.724	1.381
Smoking	0.779	1.285

The graph above is the multicollinearity diagnostic which measures the relationship between variables. A VIF near 1 is low, indicating that the variables are indeed independent from one another.

Please reference “Reference Section” pages 6-9. You will see four pages of scatter plots that show the relationships between all the variables. You will see that all the graphs show that the predictor variables have little to no relationship with each other. However, only ‘Time,’ ‘Age,’ ‘Ejection Fraction,’ and ‘Serum Creatinine’ have a discernable relationship with ‘Death Event.’

Prediction Interval

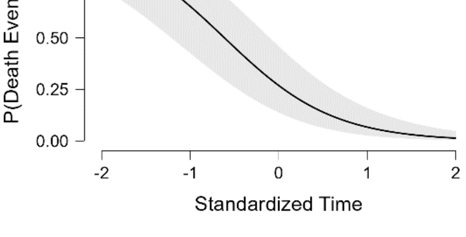
Please refer to page 1 of “Reference Section.” Using logistic regression, I have built a table that shows the estimates for the coefficients as well as built confidence intervals (in odds ratio scale) for the estimates (consider that the data was standardized before running the regression). For example, ‘Standardized Time,’ with an odds ratio of 0.196 and a 95% CI of 0.124 to 0.309 (interval not containing 1) has a strong relationship with the outcome (coefficient = -1.631). As ‘Time’ increases by one standard deviation, the odds of ‘Death Event’ decreases by approximately 80.4%. Conversely, ‘Anemia’ has an odds ratio of 0.993 and a 95% CI of 0.490 to 2.012 (interval containing 1), indicating that there is no significant relationship to ‘Death Event’. Transforming the confidence intervals into odds ratio scale allowed for an intuitive understanding of how the predictors related to the outcomes as well as providing certainty for their estimates.

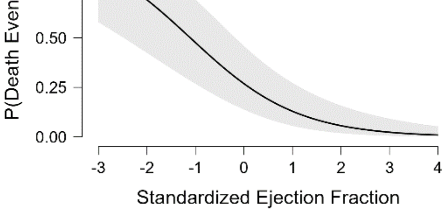
Regression (Logistic) Assumptions

Absence of multicollinearity: In the previous section “Predictor Interaction,” I showed how the model has an absence of multicollinearity by examining the VIF indicator as well as residual plots.

Lack of strongly influential outliers: In the previous section “Exploratory Data Analysis” I examined outliers and why the data should not be discarded from the dataset given the reliability of the source (The Faisalabad Institute of Cardiology and the Allied Hospital). In addition, the data involves patients that have cardiovascular disease and recently experienced heart failure. Even among a sample of patients who experienced heart failure, it would not be surprising to see multiple outliers given how intense the medical incident of a heart failure is. Lastly, the number of outliers for some data was large and clustered at the extremes, making it unreasonable to throw away such significant data to force the model to fit better.

Linearity in the logit for continuous variables: I created Estimates Plots to visually assess the linearity in the logit for continuous variables for the logistic regression. The remainder of the plots are located on page 10 of the “Reference Section.” As you can see below, there is a linear relationship between the predictors and the log odds of the outcome. The 95% confidence intervals for ‘Time’ and ‘Ejection Fraction’ (as well as ‘Serum Creatinine’ and ‘Age’) are narrow, suggesting an elevated level of certainty of the strength of the association with the outcome in the model.





Independence of errors: In the previous section “Residual Analysis” I examined how the residuals randomly distributed and evenly dispersed, giving me no reason to assume a lack of independence of errors.

Conclusion

Creating a model that removes all the variables that do not have a significant relationship with ‘Death Event’ would have created a better fit model. To avoid overfitting and unnecessary model complexity, I retained all predictors in the final specification. The tests showed that the model was useful, produced accurate results, distinguished between significant and insignificant variables, met the regression assumptions of a logistic regression, and relied on data that was reliably sourced. I concluded that it is possible to rely on ‘Time,’ ‘Ejection Fraction,’ ‘Serum Creatinine,’ and ‘Age’ to accurately predict ‘Death Event.’ Logistic regression provides a wide array of statistics that enable statisticians to make valid predictions for real-world scenarios, and those predictions are meaningful and used to mitigate harm. I am satisfied with the model and found that the research endeavor was enlightening and profound.

Reference Section

Logistic Regression

Model Summary - Death Event

Model	Deviance	AIC	BIC	df	X²	p	McFadden R²	Nagelkerke R²	Tjur R²	Cox & Snell R²
H_0	375.349	377.349	381.049	298						
H_1	219.553	245.553	293.659	286	155.795	< .001	0.415	0.568	0.465	0.406

Coefficients

					Wald Test			95% Confidence interval (odds ratio scale)	
	Estimate	Standard Error	Odds Ratio	z	Wald Statistic	df	p	Lower bound	Upper bound
(Intercept)	-1.001	0.425	0.368	-2.357	5.558	1	0.018	0.160	0.845
Standardized Age	0.562	0.187	1.754	3.001	9.006	1	0.003	1.215	2.533
Standardized Creatine Phosphokinase	0.215	0.172	1.240	1.249	1.560	1	0.212	0.885	1.737
Standardized Ejection Fraction	-0.904	0.193	0.405	-4.695	22.042	1	< .001	0.278	0.591
Standardized Platelets	-0.117	0.184	0.890	-0.635	0.403	1	0.525	0.620	1.276
Standardized Serum Creatinine	0.687	0.187	1.987	3.670	13.469	1	< .001	1.377	2.868
Standardized Serum Sodium	-0.295	0.175	0.745	-1.686	2.842	1	0.092	0.529	1.049
Standardized Time	-1.631	0.234	0.196	-6.981	48.740	1	< .001	0.124	0.309
Anemia (1)	-0.007	0.360	0.993	-0.021	4.279×10^{-4}	1	0.983	0.490	2.012
Diabetes (1)	0.145	0.351	1.156	0.413	0.171	1	0.679	0.581	2.301
High Blood Pressure (1)	-0.103	0.359	0.902	-0.286	0.082	1	0.775	0.447	1.823
Sex (1)	-0.534	0.414	0.586	-1.289	1.662	1	0.197	0.261	1.320
Smoking (1)	-0.013	0.413	0.987	-0.033	0.001	1	0.974	0.439	2.215

Note. Death Event level '1' coded as class 1.

Bootstrap Coefficients

					95% bca* Confidence interval (odds ratio scale)	
	Estimate	Bias	Standard Error	Odds Ratio	Lower bound	Upper bound
(Intercept)	-1.043	-0.065	0.507	0.352	0.137	1.015

Bootstrap Coefficients

					95% bca* Confidence interval (odds ratio scale)	
	Estimate	Bias	Standard Error	Odds Ratio	Lower bound	Upper bound
Standardized Age	0.616	0.065	0.200	1.852	1.134	2.461
Standardized Creatine Phosphokinase	0.238	0.054	0.238	1.268	0.777	1.955
Standardized Ejection Fraction	-0.980	-0.096	0.229	0.375	0.281	0.630
Standardized Platelets	-0.120	-0.009	0.210	0.887	0.573	1.333
Standardized Serum Creatinine	0.727	0.059	0.316	2.069	1.103	3.418
Standardized Serum Sodium	-0.330	-0.044	0.216	0.719	0.506	1.168
Standardized Time	-1.747	-0.143	0.291	0.174	0.127	0.343
Anemia (1)	-0.013	-0.012	0.411	0.987	0.439	2.252
Diabetes (1)	0.147	0.008	0.398	1.159	0.527	2.548
High Blood Pressure (1)	-0.139	-0.044	0.386	0.870	0.439	1.984
Sex (1)	-0.575	-0.038	0.471	0.563	0.246	1.544
Smoking (1)	-0.021	-0.006	0.458	0.979	0.413	2.525

* Bias corrected accelerated.

Note. Bootstrapping based on 5000 successful replicates.

Note. Coefficient estimate is based on the median of the bootstrap distribution.

Multicollinearity Diagnostics

	Tolerance	VIF
Standardized Age	0.906	1.104
Standardized Creatine Phosphokinase	0.921	1.086
Standardized Ejection Fraction	0.853	1.173
Standardized Platelets	0.957	1.045
Standardized Serum Creatinine	0.907	1.102
Standardized Serum Sodium	0.934	1.071
Standardized Time	0.868	1.152
Anemia	0.897	1.115
Diabetes	0.950	1.052
High Blood Pressure	0.941	1.063
Sex	0.724	1.381
Smoking	0.779	1.285

Casewise Diagnostics

Case Number	Observed	Predicted	Predicted Group	Residual	Studentized Residual	Cook's Distance
21	0	0.907	1	-0.907	-3.117	0.019
39	0	0.908	1	-0.908	-3.143	0.027

Casewise Diagnostics

Case Number	Observed	Predicted	Predicted Group	Residual	Studentized Residual	Cook's Distance
187	1	0.029	0	0.971	5.833	0.030
196	1	0.067	0	0.933	3.726	0.028
247	1	0.069	0	0.931	3.669	0.022

Factor Descriptives

Anemia	Diabetes	High Blood Pressure	Sex	Smoking	N
0	0	0	0	0	15
				1	0
			1	0	20
				1	32
		1	0	0	10
				1	1
			1	0	12
				1	8
	1	0	0	0	17
				1	0
			1	0	17
				1	12
		1	0	0	9
				1	1
			1	0	8
				1	8
1	0	0	0	0	13
				1	0
			1	0	17
				1	15
		1	0	0	10
				1	1
			1	0	11
				1	9
	1	0	0	0	15
				1	1
			1	0	14
				1	6
		1	0	0	12
				1	0
			1	0	3
				1	2

Performance Diagnostics

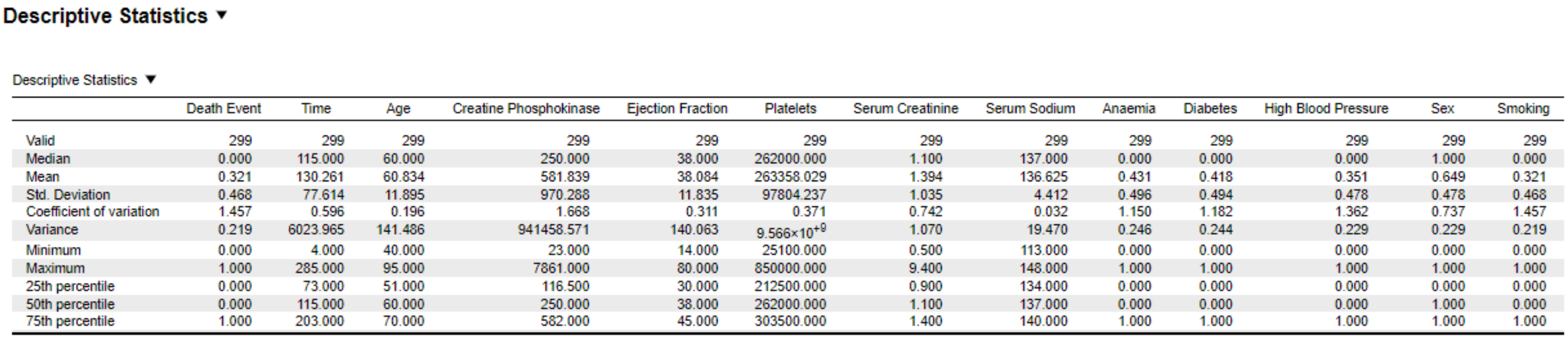
Confusion matrix

	Predicted		
Observed	0	1	% Correct
0	187	16	92.118
1	27	69	71.875
Overall % Correct			85.619

Note. The cut-off value is set to 0.5

Performance metrics

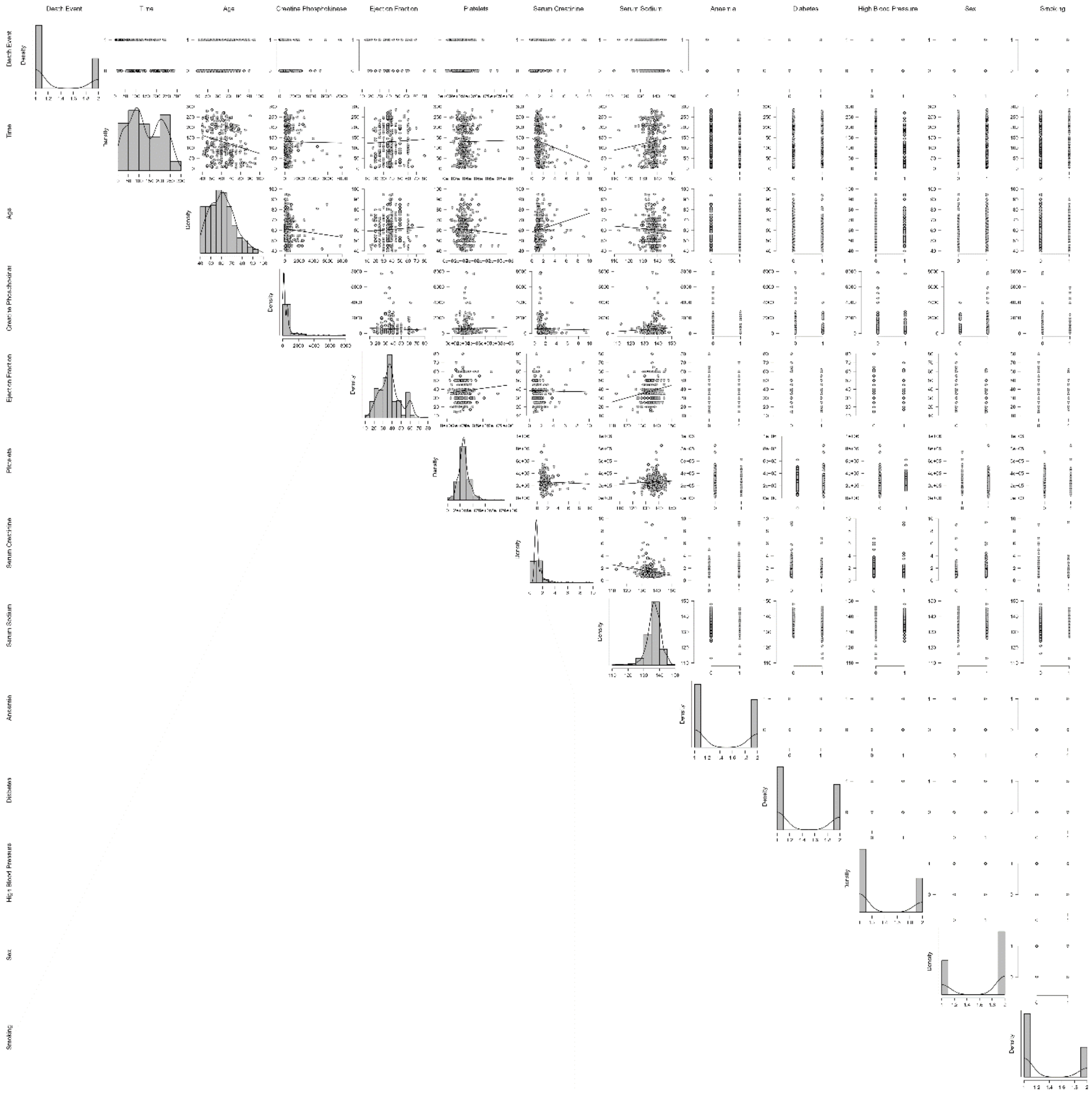
	Value
Accuracy	0.856
AUC	0.897
Sensitivity	0.719
Specificity	0.921
Precision	0.812
F-measure	0.762



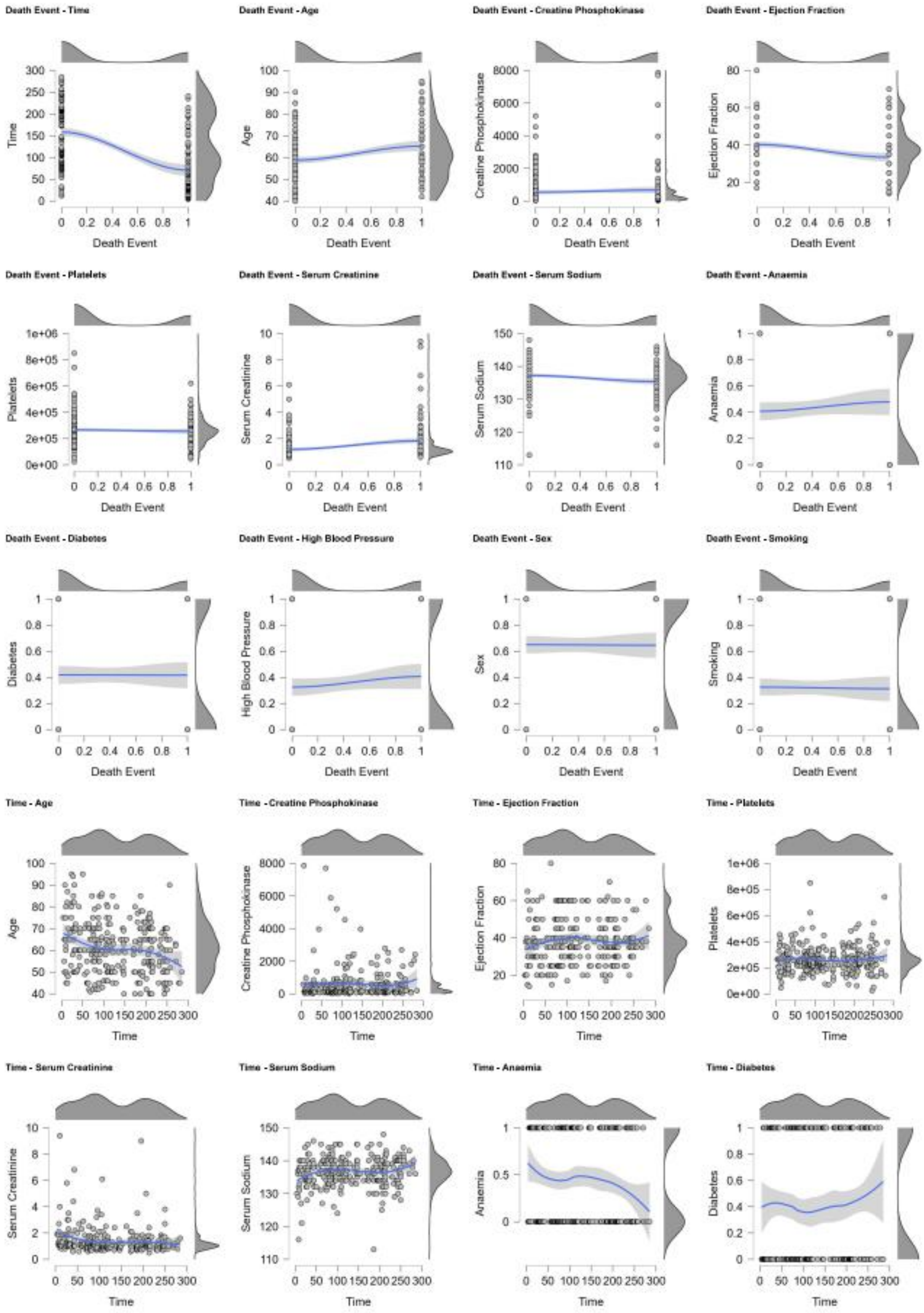
Descriptive Statistics ▼

Descriptive Statistics ▼

	Death Event	Time	Age	Creatine Phosphokinase	Ejection Fraction	Platelets	Serum Creatinine	Serum Sodium	Anaemia	Diabetes	High Blood Pressure	Sex	Smoking
Valid	299	299	299	299	299	299	299	299	299	299	299	299	299
Median	0.000	115.000	60.000	250.000	38.000	262000.000	1.100	137.000	0.000	0.000	0.000	1.000	0.000
Mean	0.321	130.261	60.834	581.839	38.084	263358.029	1.394	136.625	0.431	0.418	0.351	0.649	0.321
Std. Deviation	0.468	77.614	11.895	970.288	11.835	97804.237	1.035	4.412	0.496	0.494	0.478	0.478	0.468
Coefficient of variation	1.457	0.596	0.196	1.668	0.311	0.371	0.742	0.032	1.150	1.182	1.362	0.737	1.457
Variance	0.219	6023.965	141.486	941458.571	140.063	$9.566\times10^{+9}$	1.070	19.470	0.246	0.244	0.229	0.229	0.219
Minimum	0.000	4.000	40.000	23.000	14.000	25100.000	0.500	113.000	0.000	0.000	0.000	0.000	0.000
Maximum	1.000	285.000	95.000	7861.000	80.000	850000.000	9.400	148.000	1.000	1.000	1.000	1.000	1.000
25th percentile	0.000	73.000	51.000	116.500	30.000	212500.000	0.900	134.000	0.000	0.000	0.000	0.000	0.000
50th percentile	0.000	115.000	60.000	250.000	38.000	262000.000	1.100	137.000	0.000	0.000	0.000	1.000	0.000
75th percentile	1.000	203.000	70.000	582.000	45.000	303500.000	1.400	140.000	1.000	1.000	1.000	1.000	1.000


Death Event
Time
Age
Creatine Phosphokinase
Ejection Fraction
Platelets
Serum Creatinine
Serum Sodium
Anaemia
Diabetes
High Blood Pressure
Sex
Smoking
Density

Scatter Plots



Death Event - Time

Death Event - Age

Death Event - Creatine Phosphokinase

Death Event - Ejection Fraction

Death Event - Platelets

Death Event - Serum Creatinine

Death Event - Serum Sodium

Death Event - Anaemia

Death Event - Diabetes

Death Event - High Blood Pressure

Death Event - Sex

Death Event - Smoking

Time - Age

Time - Creatine Phosphokinase

Time - Ejection Fraction

Time - Platelets

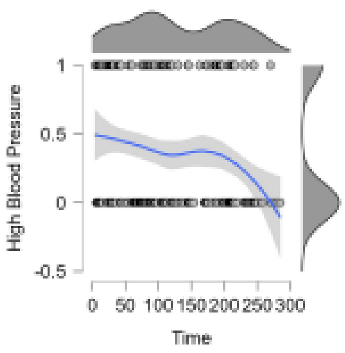
Time - Serum Creatinine

Time - Serum Sodium

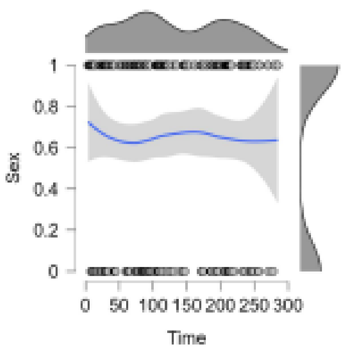
Time - Anaemia

Time - Diabetes

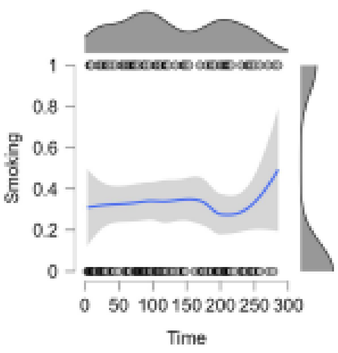
Time - High Blood Pressure



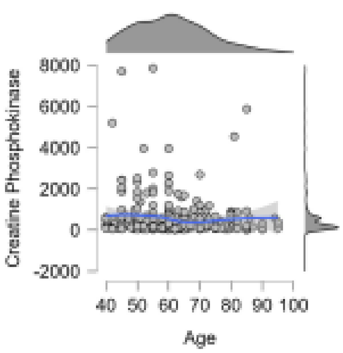
Time - Sex



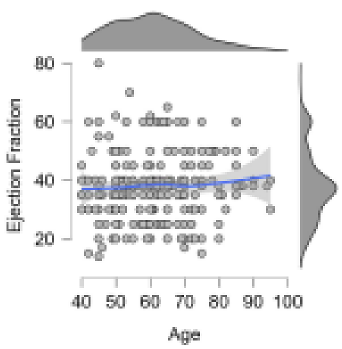
Time - Smoking



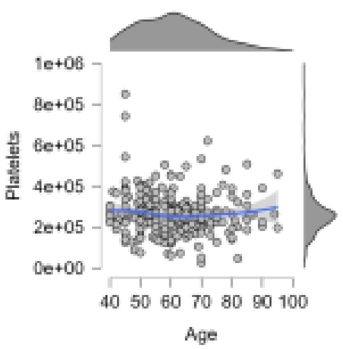
Age - Creatine Phosphokinase



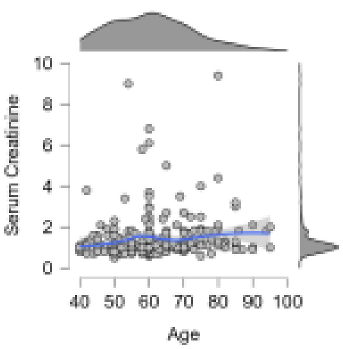
Age - Ejection Fraction



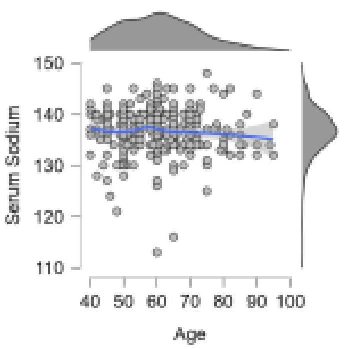
Age - Platelets



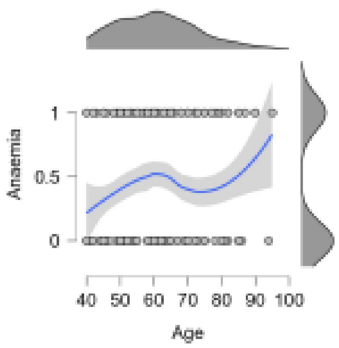
Age - Serum Creatinine



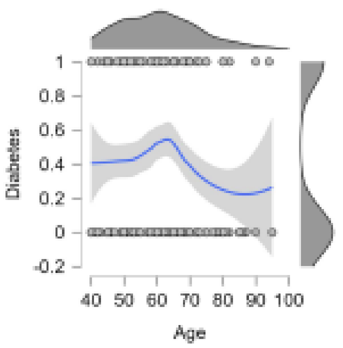
Age - Serum Sodium



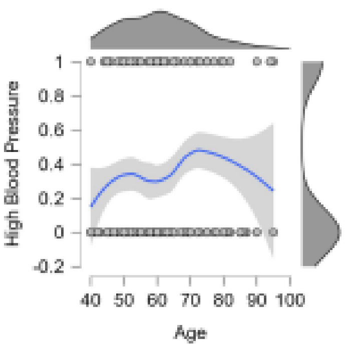
Age - Anaemia



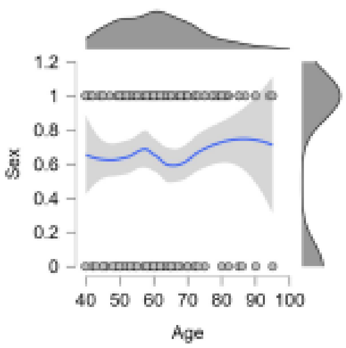
Age - Diabetes



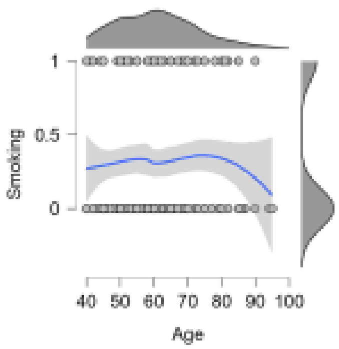
Age - High Blood Pressure



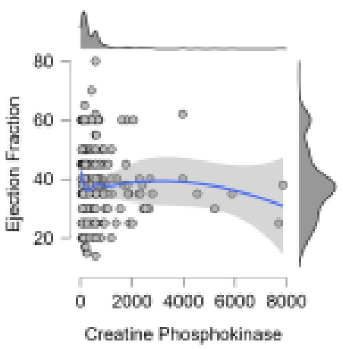
Age - Sex



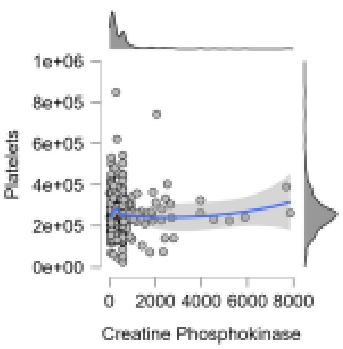
Age - Smoking



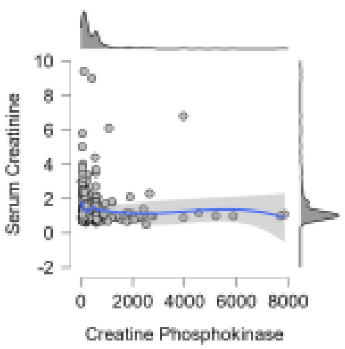
Creatine Phosphokinase - Ejection Fraction



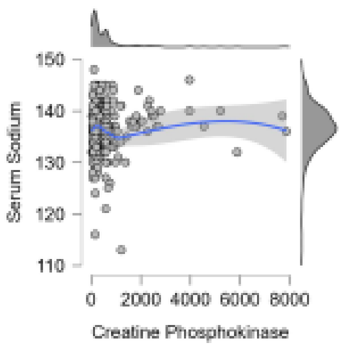
Creatine Phosphokinase - Platelets



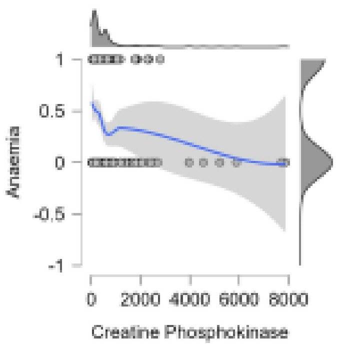
Creatine Phosphokinase - Serum Creatinine



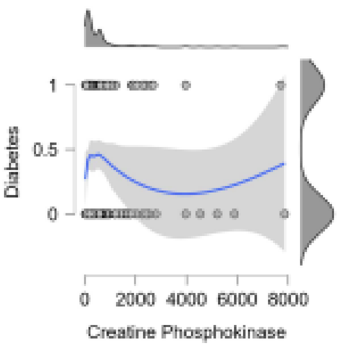
Creatine Phosphokinase - Serum Sodium



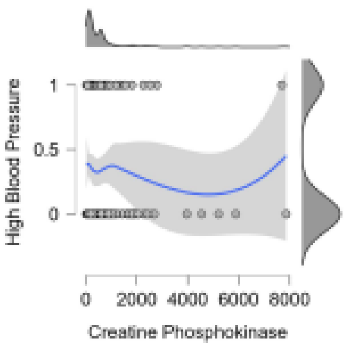
Creatine Phosphokinase - Anaemia

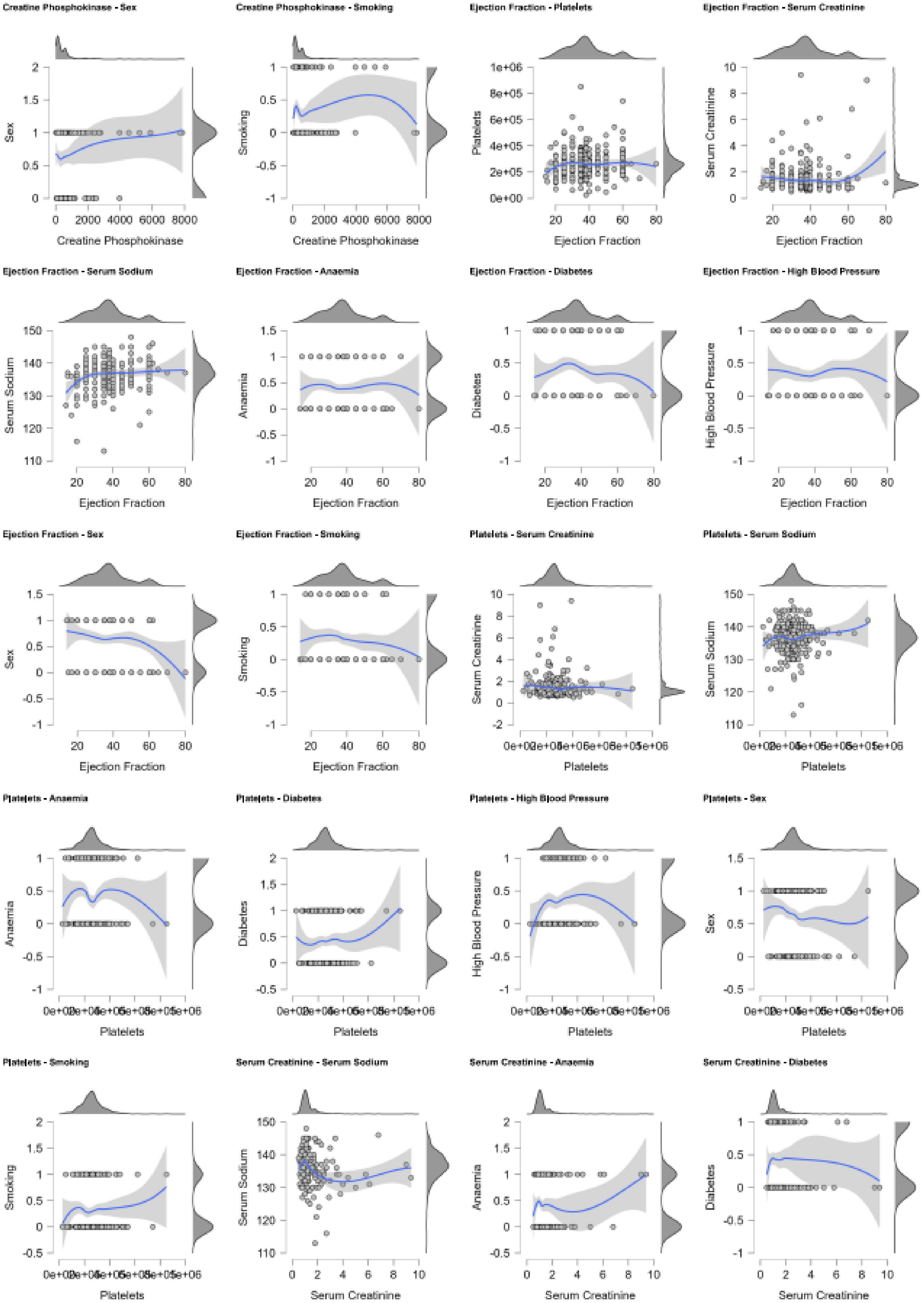


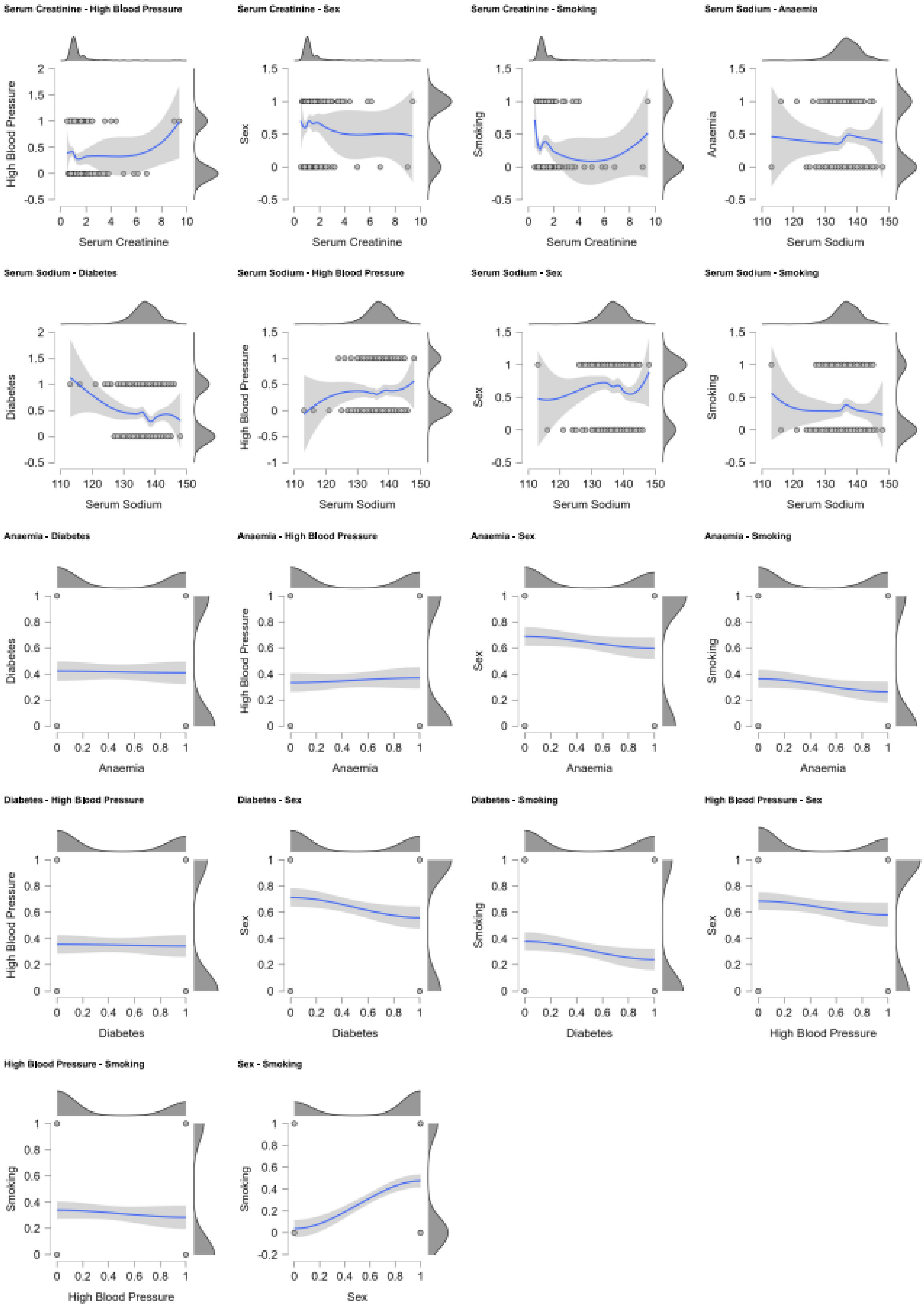
Creatine Phosphokinase - Diabetes



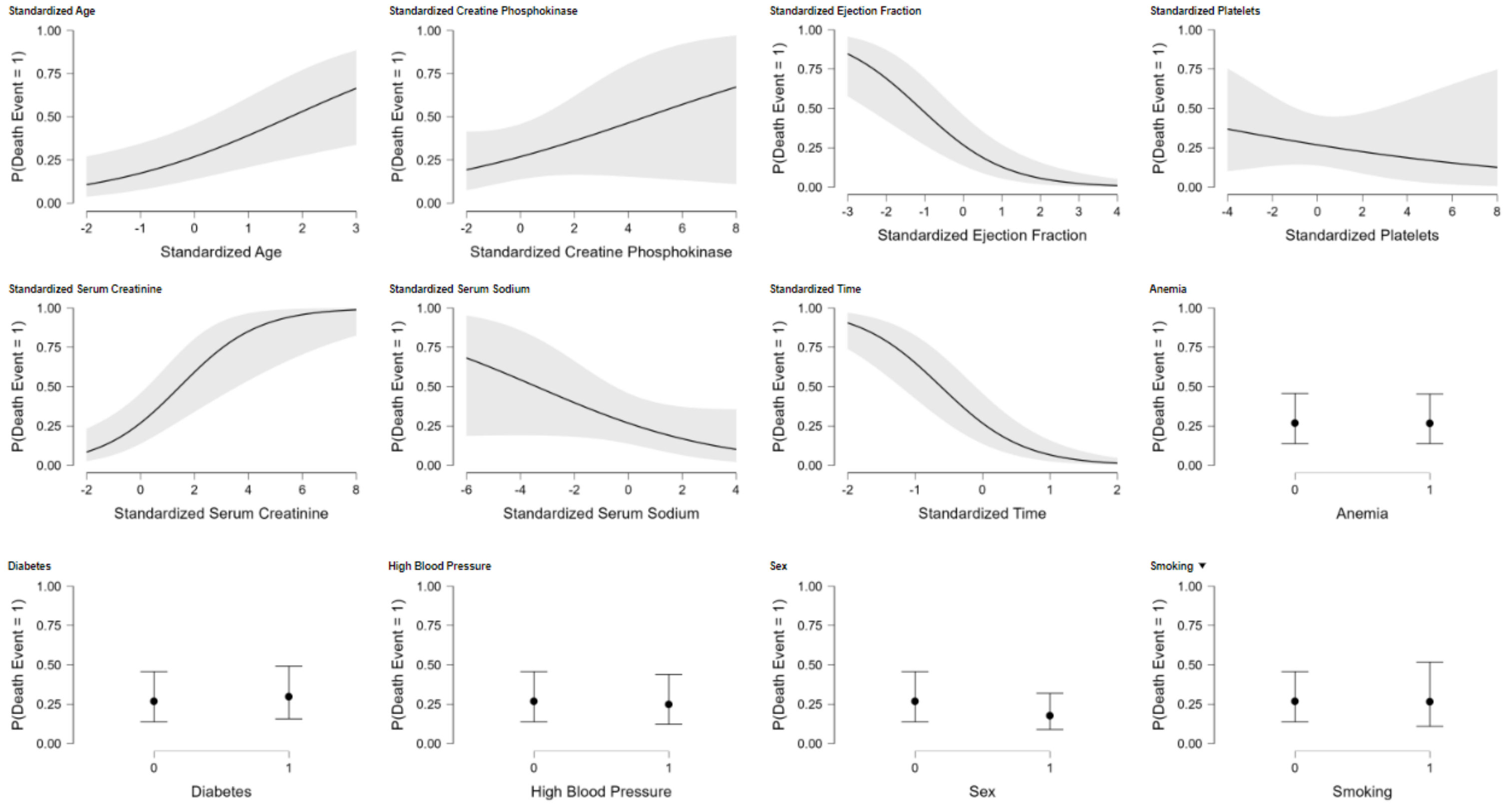
Creatine Phosphokinase - High Blood Pressure



Creatine Phosphokinase - Sex
Sex
Creatine Phosphokinase
Creatine Phosphokinase - Smoking
Smoking
Creatine Phosphokinase
Ejection Fraction - Platelets
Platelets
Ejection Fraction
Ejection Fraction - Serum Creatinine
Serum Creatinine
Ejection Fraction
Ejection Fraction - Serum Sodium
Serum Sodium
Ejection Fraction
Ejection Fraction - Anaemia
Anaemia
Ejection Fraction
Ejection Fraction - Diabetes
Diabetes
Ejection Fraction
Ejection Fraction - High Blood Pressure
High Blood Pressure
Ejection Fraction
Ejection Fraction - Sex
Sex
Ejection Fraction
Ejection Fraction - Smoking
Smoking
Ejection Fraction
Platelets - Serum Creatinine
Serum Creatinine
Platelets
Platelets - Serum Sodium
Serum Sodium
Platelets
Platelets - Anaemia
Anaemia
Platelets
Platelets - Diabetes
Diabetes
Platelets
Platelets - High Blood Pressure
High Blood Pressure
Platelets
Platelets - Sex
Sex
Platelets
Platelets - Smoking
Smoking
Platelets
Serum Creatinine - Serum Sodium
Serum Sodium
Serum Creatinine
Serum Creatinine - Anaemia
Anaemia
Serum Creatinine
Serum Creatinine - Diabetes
Diabetes
Serum Creatinine


Serum Creatinine - High Blood Pressure
High Blood Pressure
Serum Creatinine
Serum Creatinine - Sex
Sex
Serum Creatinine
Serum Creatinine - Smoking
Smoking
Serum Creatinine
Serum Sodium - Anaemia
Anaemia
Serum Sodium
Serum Sodium - Diabetes
Diabetes
Serum Sodium
Serum Sodium - High Blood Pressure
High Blood Pressure
Serum Sodium
Serum Sodium - Sex
Sex
Serum Sodium
Serum Sodium - Smoking
Smoking
Serum Sodium
Anaemia - Diabetes
Diabetes
Anaemia
Anaemia - High Blood Pressure
High Blood Pressure
Anaemia
Anaemia - Sex
Sex
Anaemia
Anaemia - Smoking
Smoking
Anaemia
Diabetes - High Blood Pressure
High Blood Pressure
Diabetes
Diabetes - Sex
Sex
Diabetes
Diabetes - Smoking
Smoking
Diabetes
High Blood Pressure - Sex
Sex
High Blood Pressure
High Blood Pressure - Smoking
Smoking
High Blood Pressure
Sex - Smoking
Smoking
Sex

Estimates plots ▼



Standardized Age

Standardized Creatine Phosphokinase

Standardized Ejection Fraction

Standardized Platelets

Standardized Serum Creatinine

Standardized Serum Sodium

Standardized Time

Anemia

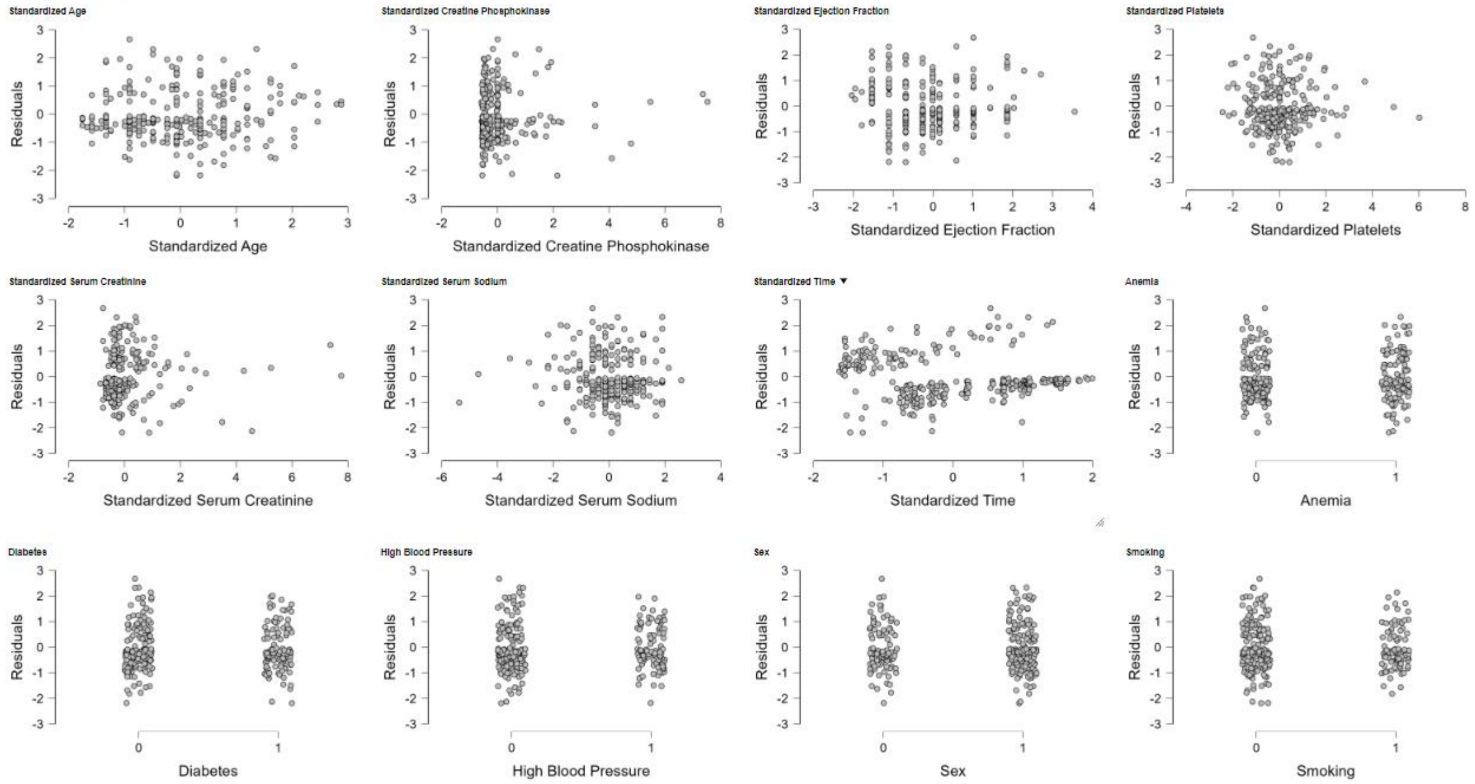
Diabetes

High Blood Pressure

Sex

Smoking ▼

Predictor - residuals plots ▼



Standardized Age

Standardized Creatine Phosphokinase

Standardized Ejection Fraction

Standardized Platelets

Standardized Serum Creatinine

Standardized Serum Sodium

Standardized Time ▼

Anemia

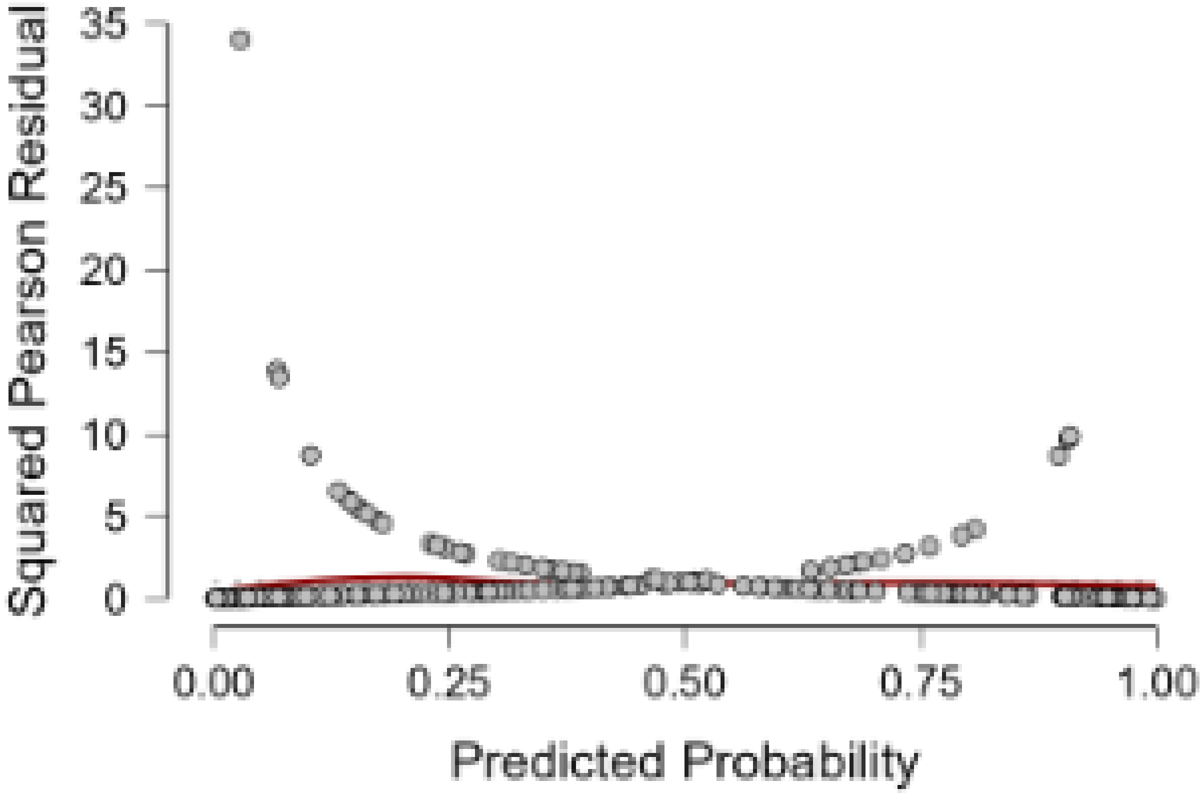
Diabetes

High Blood Pressure

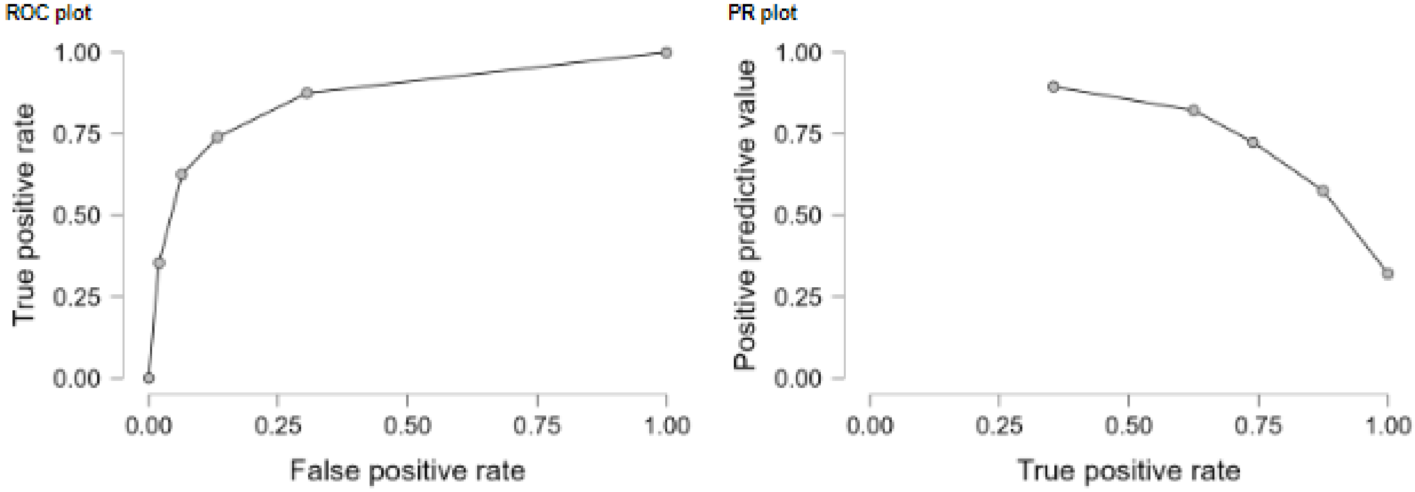
Sex

Smoking

Squared Pearson residuals plot



Performance plots



ROC plot

PR plot