Group Name: Cappoo
Group Members: Fiona Lee, Jasmine Wang

Introduction

Heart failure (HF) is a serious chronic condition, which occurs when heart muscles are too weak to pump enough oxygen and blood to other organs and can sometimes lead to death[1]. Therefore, it is reasonable to suspect some body conditions are associated with symptoms and mortality in HF, such as older adults or people with high blood pressure may more likely die from HF. Apart from these, there may also be other factors that directly or indirectly lead to death of HF. To gain a better understanding, we analyzed a data set of HF patients to find some reliable predictors for mortality in HF.

Data Analyses and Results

Data preparation

Before analysis, we organized the data and checked for duplicates, weird outliers and different data types. The data contained 270 unique patient indexes, each data entry had a value for each of the 14 variables, including 8 continuous and 6 categorical variables. All values were greater than 0, and each categorical variable had a value, either 0 or 1.

Univariate data analysis

In the Death Event variable, 88 out of 270 patients (32.6%) died. From the Sex variable, we found that there were more male patients than female patients (33.7% female, 66.3% male) in the data, so we proposed that men are more prone to HF than women. To decide further analyses, we used the Shapiro-Wilk test and found that none of the continuous variables were normally distributed ($p < .05$).

Bivariate data analyses

We obtained Pearson correlation between each variable and Death Event, and then used Mann-Whitney U test and Chi-square test to test significance and focused on 5 variables that showed a significant effect on Death Event, which were: Time ($r(268) = –.55$, $p < .001$), Serum Sodium ($r(268) = –.17$, $p < .01$), Ejection Fraction ($r(268) = –.28$, $p < .001$), Age ($r(268) = .28$, $p < .001$), and Serum Creatinine ($r(268) = .30$, $p < .001$). These 5 variables were candidates for predicting Death Event.

To evaluate the interactions between other variables and these 5 candidates, we used Mann-Whitney U test on categorical variables and significance test for Pearson correlation between two continuous variables. We found that patients who had anemia (Anaemia = 1) and died from HF had lower values in Time than those who did not have anemia ($U = 7815.0$, $p = .035$). In addition, male patients (Sex = 1) showed lower levels in Ejection Fraction than female patients (Sex = 0) ($U = 6760.5$, $p = .01$). Patients with high blood pressure (High Blood Pressure = 1) generally had a higher age ($U = 6863.0$, $p = .01$) and less time in follow up period ($U = 6760.5$, $p = .006$).

Besides the impact other variables had on these 5 candidates, we also found interactions among themselves on Death Event (Table 1).

Modeling

We also built a Random Forest Model to conduct feature selection among all variables and found 4 significant variables. Interestingly, except for Serum Sodium, we obtained exactly the same variables as in the exploratory data analysis (EDA) described above. We referred back to the p-value of Serum Sodium in EDA and found it relatively higher than other 4 significant variables. Therefore, we suspected

that this difference might be due to the small sample size of this dataset, which could generate an illusion of significance with a small *p*-value in EDA.

		Time	Serum Sodium	Serum Creatinine
Age	Pearson correlation	–0.22***	–	0.162**
	p–value	< .001	–	0.008
Ejection Fraction	Pearson correlation	–	0.186**	–
	p–value	–	0.002	–
Serum Creatinine	Pearson correlation	–0.169**	–0.185**	–
	p–value	0.005	0.002	–

Table 1. Significant interactions between the 5 candidates on Death Event. * p < .05, ** p < .01, *** p < .001

Conclusion

Overall, we found that measurements of serum sodium and serum creatinine from blood test, ejection fraction measured through appropriate instruments, a patient's age and the length of follow-up period can help in predicting a HF patient's death. However, missing Serum Sodium in feature selection may reveal its relatively lower reliability than other 4 variables. In addition, having anaemia and high blood pressure can affect the correlations of age and length of follow-up period with death, which potentially increase the possibility of death indirectly. Also, male patients may be more prone to death than female patients in HF, associated with their lower percentages of ejection fraction.

Apart from those showing significantly direct or indirect relationships with Death Event, other variables may be indicators of HF, but the data suggested that their influences on the Death Event of patients were not higher than by chance. For example, creatinine phosphokinase is an enzyme produced when heart muscles fail, while it gets circulated out of the body from the kidney and may not have a long-lasting effect on body functions. For another instance, platelets are found to be abnormal in patients with HF, but the abnormalities can be caused by either HF or comorbidities of HFs. It is therefore hard to determine a strong relationship between platelets and death of HF patients, given only 12 variables without the cause of death.

We found several limitations in the provided dataset during our analyses, which stopped us from identifying clear relationships and making well-supported conclusions. First, as mentioned above, excluded variables and small sample size can be a barrier of identifying predictors of death in HF patients. Second, death of these HF patients can be caused by HF or other comorbidities, and we thus can hardly conclude the relationships between death and HF indicators. Third, we have only one-time measurements without information about any further changes in patients' behaviors or body conditions. Future studies can focus on (a) how regular measurements would provide more details and information that help form clear relationships between variables and mortality in HF patients, and (b) application of this model in a clinical setting where given regular measurements obtaining, time is no longer an appropriate predictor variable.

1. American Heart Association. (2017, May 31). *What is Heart Failure?* https://www.heart.org/en/health-topics/heart-failure