



LONDON
SCHOOL *of*
HYGIENE
& TROPICAL
MEDICINE



RISK SCORE MODEL FOR HEART FAILURE
PATIENTS

Candidate #: 101007

Page Count: 36

Submitted in part fulfilment of the requirements for the degree of

MSc in Medical Statistics

Tuesday 12 noon – 6th September 2011

Acknowledgements

I am greatly thankful to Tim Collier who doubled as my personal tutor during the course of study and as a supervisor on this project.

I am also indebted to the members of the Eplerenone in Mild Patients Hospitalisation and Survival Study in Heart Failure (EMPHASIS-HF) study group from where the data analysed and described in this report was sourced.

My sincere appreciation goes to the KEMRI/Wellcome trust programme for providing financial support during the course of my studies.

Table of Contents

Abstract.....	iv
1 Introduction.....	1
1.1 Heart failure.....	1
1.2 Risk scores.....	2
1.3 The EMPHASIS-HF study.....	2
1.4 Objectives.....	3
1.5 Plan of report.....	3
2 Baseline descriptive statistics.....	5
2.1 Introduction.....	5
2.2 Methods.....	5
2.3 Results.....	5
2.4 Conclusion.....	7
3 Univariable Cox models.....	8
3.1 Introduction.....	8
3.2 Methods.....	8
3.3 Results.....	9
3.4 Discussion.....	13
3.5 Conclusion.....	13
4 Multivariable Cox models.....	14
4.1 Introduction.....	14
4.2 Methods.....	14
4.3 Results.....	15
4.4 Discussion.....	16
4.5 Conclusion.....	17
5 Missing data.....	18
5.1 Introduction.....	18
5.2 Methods.....	18
5.3 Results.....	18
5.4 Discussion.....	20
5.5 Conclusion.....	21
6 Risk Score.....	22
6.1 Introduction.....	22
6.2 Methods.....	22

6.3	Results.....	22
6.4	Discussion	26
6.5	Conclusion.....	27
7	Discussion	28
7.1	Summary of results.....	28
7.2	Comparison with other studies.....	28
7.3	Strengths and weaknesses.....	30
7.4	Further analyses	31
7.5	Conclusion.....	31
	References	32
	Appendix.....	35

Abstract

Heart failure remains a significant cause of mortality and morbidity in many parts of the world with predictions pointing towards rising figures in the future. A wide range of predictors of heart failure have already been identified. Novel factors have also added to the pool recently. Consequently, parsimonious prognostic models for death and hospitalization for heart failure become important.

The current study used data from a clinical trial of Eplerenone in patients with systolic heart failure and mild symptoms to; identify baseline characteristics that are associated with all cause mortality or hospitalisation for heart failure, build a multivariable risk model for this outcome, refine and test the predictive power of the risk model. Two models were developed through stepwise variable selection and Cox proportional hazards regression; one including lab predictors and another excluding them.

The final model that included the lab predictors had sixteen predictors in total, while that excluding them had fifteen predictors in total. The discriminative abilities of the models were moderately strong (Harrell's c-indices of 0.71 and 0.68 for the model with and without lab related predictors respectively).

The prognostic models developed have good predictive value and discriminative ability and can be applied in predicting mortality or heart failure hospitalisation in a wide range of settings, either in the absence or presence of already established lab predictors.

1 Introduction

1.1 Heart failure

The principal causes of heart failure are diseases that damage the heart. These include coronary heart disease which remains a significant cause of mortality and morbidity in many parts of the world. In the early 1990s coronary heart disease was the leading cause of death in the United States, claiming about 490,000 lives in 1990 alone.¹ An estimated 82 million adult Americans, about 1 in 3, have one or more types of cardiovascular disease with about half of the number aged above 60 years. Mortality data in 2007 showed that 813 804 (33.6% of all deaths) deaths were from cardiovascular diseases. On average more than 2200 Americans die of cardiovascular disease each day, an average of 1 death every 39 seconds.²

On the global picture, heart disease and stroke kill nearly 17 million people worldwide with developing countries as a group experiencing an increase in cases of cardiovascular diseases. Low and middle income countries are contributing about 80% of global cardiovascular-related deaths and 87% of cardiovascular-related disabilities. Predictions for next two decades as at 2005 are tripling of ischemic heart disease and stroke mortality in Latin America, the Middle East and sub-Saharan Africa.³

Diabetes, cigarette smoking, blood pressure and total cholesterol have been associated with heart failure.⁴⁻⁶ Preoperative myocardial injury (PMI) and coronary artery bypass grafting (CABG) have also been previously reported as risk factors for heart failure, other factors include: Increased age, female sex, diabetes, previous myocardial infarction, dyspnoea, preoperative atrial fibrillation, left ventricular dysfunction, and triple vessel disease.⁷ Novel metabolic variables reflecting insulin resistance and dyslipidemia, together with a low beta-carotene level, have also been found to predict heart failure independently of some established risk factors.⁸

With a wide range of variables already identified as risk factors for heart failure and with novel factors adding to the pool, models that use a limited number of factors but which are adequate in predicting risk of heart failure become important.

1.2 Risk scores

Risk scores have increasingly been used to quantify the amount risk that an individual is predisposed to given their values on a set of prognostic factors. They are useful because they are derived from models that take into account the contribution of each of the risk factors jointly. Several models relating to risk scores in heart failure have been reported.⁹⁻¹⁶ Some of the studies concentrated on patients with advanced functional limitations (New York Heart Association [NYHA] functional class IV symptoms)⁹. Other studies have developed risk scores for specifically predicting cardiac mortality^{10,11} and some have further narrowed down to predicting in-hospital mortality¹² or right ventricular failure.¹³

The current work develops risk score models for predicting a composite of all cause mortality or heart failure hospitalisation from older patients with systolic heart failure and mild symptoms (NYHA functional class II symptoms).

1.3 The EMPHASIS-HF study

The data used to develop the prognostic models comes from a clinical trial of Eplerenone in Mild Patients Hospitalisation and Survival Study in Heart Failure (EMPHASIS-HF) whose results have already been published.¹⁷

The study recruited subjects who were at least 55 years of age, with systolic heart failure and mild symptoms (NYHA functional class II symptoms)¹⁸, having an ejection fraction of no more than 30% or, if more but less than 35%, a QRS duration of >130 msec on electrocardiography. The eligibility criteria further included treatment with an angiotensin-converting-enzyme (ACE) inhibitor, an angiotensin-receptor blocker (ARB) or both together with a beta-blocker at the recommended dose or maximal tolerated dose.

Subjects were excluded if they had; acute myocardial infarction, NYHA class III and IV heart failure, estimated glomerular filtration rate (eGFR) of less than 30 ml/min/1.73m², a serum potassium level exceeding 5.0 mmol per liter, a need for potassium diuretic and any other clinically significant co-morbidity.

Subjects were randomized into the study if they had been hospitalized for cardiovascular reasons within 6 months prior to the screening visit date. 400 patients were also randomized based on their plasma level of B-type natriuretic peptide (BNP) set to be at least 250 pg per millimeter, or on the level of N-terminal pro-BNP (at least 500 and 750 pg per milliliter for men and women respectively).

A total of 2737 patients were recruited into the study from 278 centers in 29 countries, with 1364 assigned to treatment arm. The median follow up time among all patients was 21 months with an accumulated follow up time of 4783 person-years.

In the study treatment with eplerenone was associated with a reduction in the rate of death from any cause or heart failure hospitalization.

1.4 Objectives

The objectives of the current project are:

- I. To identify baseline patient characteristics that are associated with all cause mortality or hospitalisation for heart failure.
- II. To build a multivariable risk model for all cause mortality or hospitalisation for heart failure.
- III. To refine and test the predictive power of the risk model.

1.5 Plan of report

The rest of the report is organised as follows:

Chapter 2 contains the baseline descriptive statistics, where the distribution of each candidate variable has been summarised by whether the primary outcome (all cause death or hospitalisation for heart failure) was experienced or not.

Chapter 3 reports the results of univariable Cox proportional hazard regression models which aimed to identify important predictors of the primary outcome for further consideration in the multivariable models.

Chapter 4 presents the final prognostic models arrived at through a forward stepwise procedure. Adjusted hazard ratios associated with the predictors in the final models are shown. Two prognostic models are presented; one which includes some lab measured predictor variables and one which excludes them.

Chapter 5 deals with missing values. The models from chapter 4 are re-estimated after multiple imputation of missing values and the new coefficients are compared to those estimated from complete case analysis.

Chapter 6 contains the results of classification of patients into risk groups according to their risk scores, which are linear predictors from the prognostic models. The discriminative power of the models and their goodness of fit are presented in the chapter.

Chapter 7 concludes the report with a discussion of the study's results in comparison to other related work, its strengths and weaknesses and further analyses that can be done.

2 Baseline descriptive statistics

2.1 Introduction

This chapter explored the distribution of each candidate variable for inclusion in the final multivariable model, by whether the outcome of interest was experienced or not. Descriptive statistics were used to explore if there were any important differences at baseline between those who experienced the outcome and those who did not.

2.2 Methods

Continuous variables were summarized through means and standard deviations. Categorical variables were summarized using proportions. Missing values were also quantified and reported for the variables with most frequent numbers.

2.3 Results

Table 1: Baseline characteristics by whether outcome of interest was experienced

Variable	Without event n=2091 Mean \pm SD or n (%)	With event n= 646 Mean \pm SD or n (%)
Treatment: Eplerenone vs. Placebo	1094 (52.3)	270 (41.8)
Demographic and Social characteristics		
Female vs. Male	485 (23.2)	125 (19.4)
Race		
White	1747 (83.6)	521 (80.7)
Black	50 (2.4)	17 (2.6)
Asian	230 (11.0)	86 (13.3)
Other	64 (3.1)	22 (3.4)
Smoking status		
Never	938 (44.9)	235 (44.1)
Current	221 (10.6)	72 (11.2)
Ex	932 (44.2)	289 (44.7)
Age (years)	68.2 \pm 7.6	70.0 \pm 7.7
Clinical characteristics		
Weight (kg)	80.1 \pm 16.7	76.6 \pm 17.4
Height (cm)	169.6 \pm 9.4	168.4 \pm 9.6
Waist circumference (cm)	99.7 \pm 13.3	97.7 \pm 13.8
Body Mass Index (kg/m ²)	27.7 \pm 4.9	26.8 \pm 4.9
Diastolic blood pressure (mmHg)	75.0 \pm 10.2	73.5 \pm 10.3
Systolic blood pressure (mmHg)	125.1 \pm 16.8	121.1 \pm 16.9
Pulse pressure (mmHg)	50.1 \pm 13.2	47.6 \pm 12.8
Heart rate (bpm)	72.5 \pm 15.1	76.5 \pm 16.5

Heart/Kidney/Liver functionality

Left ventricular ejection fraction (%)	26.3 ± 4.6	25.6 ± 4.7
QRS duration (msec)	120.5 ± 47.1	122.4 ± 39.0
QRS duration >130 msec	651 (31.7)	220 (34.9)
eGFR (ml/min/1.73m ²)	72.7 ± 21.8	64.7 ± 20.2
Albumin (g/DL)	4.1 ± 0.5	4.0 ± 0.5
Alanine transaminase (IU/L)	27.1 ± 16.0	27.8 ± 18.7
Aspartate transaminase (IU/L)	26.3 ± 14.2	26.9 ± 13.1
Creatinine (mg/DL)	1.1 ± 0.3	1.2 ± 0.3
Haemoglobin (g/DL)	13.9 ± 1.6	13.5 ± 1.6
Potassium (MEQ/L)	4.3 ± 0.4	4.3 ± 0.4
Sodium (mg/DL)	140.2 ± 3.9	139.8 ± 4.2
Total bilirubin (mg/DL)	0.8 ± 0.4	0.9 ± 0.5

Medical History (Yes/No)

Hospitalization for heart failure	1037 (49.6)	403 (62.5)
Duration of heart failure (years)	4.4 ± 5.6	5.5 ± 6.1
Myocardial infarction	1015 (48.5)	366 (56.7)
Stroke	185 (8.8)	77 (11.9)
PCI	455 (21.8)	141 (21.9)
Pacemaker	306 (14.7)	87 (13.6)
Diabetes mellitus	606 (29.0)	253 (39.2)
COPD	278 (13.3)	113 (17.5)
Cancer	69 (3.3)	25 (3.9)
CABG	361 (17.3)	155 (24.0)
Asthma	79 (3.8)	18 (2.8)
Angina	912 (43.6)	277 (42.9)
Atrial fibrillation or flutter	625 (29.9)	319 (34.0)
Hypertension	1388 (66.4)	431 (66.8)
ICD	285 (13.6)	77 (11.9)
Left bundle-branch block present	493 (23.6)	195 (30.2)

Medication at randomization visit (Yes/No)

ACE inhibitor	1683 (80.5)	518 (80.2)
ARB	427 (20.4)	132 (20.4)
Beta-blocker	1874 (89.6)	547 (84.7)

Principal cause of heart failure

Ischesmic heart disease	1398 (66.9)	488 (75.5)
Non-ischesmic heart disease	691 (33.1)	155 (24.0)
Unknown	2 (0.1)	3 (0.5)

*COPD, chronic obstructive pulmonary disease; CABG, coronary artery bypass graft; PCI, percutaneous coronary-artery; ICD, Implantable Cardioverter-Defibrillator ; ACE, angiotensin-converting enzyme; ARB, angiotensin-receptor blocker; eGFR, estimated glomerular filtration rate.

A total of 646 out of 2737 subjects in the study experienced the outcome of interest (hospitalization for heart failure or death). The person-years of observation was 665.96 and 3818.18 in those who did and did not experience the outcome respectively.

There was a higher proportion of subjects who had been treated with Eplerenone among those who did not experience the outcome (52.3%) compared to those who did (41.8%). The proportion of females among those with and without outcome was 19.4% and 23.2%

respectively. Those who experienced the outcome were slightly older, with an average age of 70 years compared to 68 years in those without the outcome at the end of follow-up.

On the clinical characteristics, subjects with the primary outcome had on average a higher waist circumference, a higher body mass index, a higher systolic blood pressure and a lower heart rate. The average left ventricular ejection fraction, estimated glomerular filtration rate and haemoglobin level was higher among subjects who did not experience the outcome.

Looking at prior medical history, there was a higher prevalence of; previous heart failure hospitalization, myocardial infarction, stroke, diabetes mellitus, chronic obstructive pulmonary disease (COPD), coronary artery bypass graft (CABG), atrial fibrillation and left-bundle branch within the subjects who were later hospitalized for heart failure or died.

Considering the cause of heart failure, there was a higher proportion of heart failure hospitalisations or all cause deaths from Ischemic heart disease in those who experienced the outcome of interest compared to those who did not (Table 1).

Missing data

Data recording was reasonably complete, the variables with most frequent rate of missing values included; waist circumference, 6.6% of values missing; alanine transaminase (5.8%); aspartate transaminase (5.2%); left ventricular ejection fraction (5.3%); total bilirubin levels (4.8%) ; sodium levels (2.6%) and haemoglobin levels (2.5%).

2.4 Conclusion

646 (23.6%) subjects either had a hospitalization for heart failure or died during follow-up. The subjects who experienced the outcome of interest and those who did not differed in a number of factors at baseline. The main differences were observed in treatment arm and previous medical history, where the relative frequency of previous hospitalization, Myocardial infarction, stroke and diabetes mellitus was higher among those who experienced the outcome of interest. A higher proportion of those who did not experience the outcome of interest were treated with Eplerenone.

3 Univariable Cox models

3.1 Introduction

One of the aims of the project was to build a multivariable model that well balances between good prediction of the risk of outcome and simplicity. Since the study had a cohort design, with a median follow-up time of 21 months, techniques that account for the varying observation times between subjects by using rates and their ratios as measures of effect were used. A strategy for selecting the variables for inclusion in the multivariable model was also critical. This section presents crude hazard ratios associated with each of the candidate variables to tell between strong and weak predictors of death or heart failure hospitalisation as a step towards building multivariable models.

3.2 Methods

Survival analysis techniques were used to analyze the data. The association of each explanatory variable with the outcome of interest was assessed by fitting a univariable Cox proportional hazards model. Before fitting each of the Cox PH model the proportionality assumption was checked informally, for categorical variables, by graphical techniques through the log-cumulative hazards in Nelson-Aalen plots. After fitting the model, the proportionality assumption was formally tested on the basis of the Schoenfeld residuals. However, any violation at this stage were ignored.

Martingale residuals from a model without covariates were plotted against each of the continuous variables and assessed to judge whether transformation was required before entering them in the model.

The continuous variables significantly associated with the outcome of interest were each categorized into 6 quantiles and their relationship with outcome across these quantiles graphed as hazard ratios and corresponding 95% confidence intervals. Kaplan-Meier graphs for the categorical variables that showed significant association with outcome were also plotted. P-values from the likelihood ratio test were reported for each variable. The effects of continuous variables with small changes in hazard per unit increase in their value were reported for every 10-unit increase. Hazard ratios were computed from complete cases.

3.3 Results

Table 2: Univariable Hazard ratios from Cox PH regression.

Variable	Crude hazard ratio	95% CI	P-value*
Treatment: Eplerenone vs. Placebo	0.67	(0.58, 0.79)	<0.0001
Demographic and Social characteristics			
Sex: Female vs. Male	0.83	(0.68, 1.01)	0.052
Race			
White	Ref.		
Black	1.58	(0.97, 2.56)	0.0017
Asian	1.48	(1.18, 1.86)	
Other	1.51	(0.99, 2.32)	
Smoking status			
Never	Ref.		
Current	1.11	(0.86, 1.44)	0.67
Ex	1.05	(0.90, 1.24)	
Age (years)	1.03	(1.01, 1.04)	<0.0001
Clinical characteristics			
Weight (per 10 kg)	0.88	(0.84, 0.92)	<0.0001
Height (per 10 cm)	0.88	(0.81, 0.96)	0.0026
Waist circumference (per 10 cm)	0.90	(0.85, 0.96)	0.0007
Body Mass Index (kg/m ²)	0.96	(0.94, 0.98)	<0.0001
Diastolic blood pressure (per 10 mmHg)	0.85	(0.79, 0.92)	<0.0001
Systolic blood pressure (per 10 mmHg)	0.87	(0.83, 0.91)	<0.0001
Pulse pressure (per 10 mmHg)	0.87	(0.82, 0.93)	<0.0001
Heart rate (per 10 b.p.m)	1.14	(1.09, 1.19)	<0.0001
Heart/Kidney/Liver functionality			
Left ventricular ejection fraction (%)	0.97	(0.96, 0.99)	0.0009
QRS duration (per 10 msec)	1.01	(1.00, 1.03)	0.081
QRS duration >130 msec	0.81	(0.69, 0.95)	0.013
eGFR (per 10 ml/min/1.73m ²)	0.84	(0.80, 0.87)	<0.0001
Albumin (g/DL)	0.57	(0.49, 0.67)	<0.0001
Alanine transaminase (per 10 IU/L)	1.02	(0.97, 1.07)	0.40
Aspartate transaminase (per 10 IU/L)	1.04	(0.99, 1.10)	0.16
Creatinine (mg/DL)	2.48	(1.99, 3.10)	<0.0001
Haemoglobin (g/DL)	0.86	(0.81, 0.90)	<0.0001
Potassium (MEQ/L)	0.94	(0.79, 1.13)	0.54
Sodium (mg/DL)	0.97	(0.95, 0.99)	0.0006
Total bilirubin (mg/DL)	1.47	(1.27, 1.71)	<0.0001
Medical History (Yes/No)			
Hospitalization for heart failure	1.72	(1.47, 2.02)	<0.0001
Duration of heart failure (years)	1.03	(1.02, 1.04)	<0.0001
Myocardial infarction	1.39	(1.19, 1.63)	<0.0001
Stroke	1.38	(1.09, 1.75)	0.011
PCI	1.06	(0.88, 1.28)	0.52
Pacemaker	1.08	(0.86, 1.35)	0.52

Diabetes mellitus	1.57	(1.34, 1.84)	<0.0001
COPD	1.40	(1.14, 1.71)	0.0020
Cancer	1.12	(0.75, 1.67)	0.59
CABG	1.54	(1.29, 1.85)	<0.0001
Asthma	0.81	(0.51, 1.30)	0.37
Angina	0.93	(0.80, 1.09)	0.37
Atrial fibrillation or flutter	1.13	(0.96, 1.33)	0.16
Hypertension	1.01	(0.86, 1.19)	0.89
ICD	1.24	(0.97, 1.57)	0.089
Left bundle-branch block present	1.32	(1.12, 1.56)	0.0015
Medication at randomization visit (Yes/No)			
ACE inhibitor	0.85	(0.70, 1.03)	0.098
ARB	1.05	(0.87, 1.27)	0.64
Beta-blocker	0.67	(0.54, 0.83)	0.0004
Principal cause of heart failure			
Nonischemic heart disease	Ref.		
Ischemic heart disease	1.42	(1.18, 1.70)	0.0001

* Likelihood ratio test against a model without covariates. COPD, chronic obstructive pulmonary disease; CABG, coronary artery bypass graft; PCI, percutaneous coronary-artery; ICD, Implantable Cardioverter-Defibrillator ; ACE, angiotensin-converting enzyme; ARB, angiotensin-receptor blocker; eGFR, estimated glomerular filtration rate.

Treatment with eplerenone and increased; weight, body mass index, systolic blood pressure, diastolic blood pressure, pulse pressure, estimated glomerular filtration rate, albumin level and haemoglobin level were each strongly ($p < 0.0001$ for all) associated with reduced rates of hospitalization for heart failure or death. Treatment was associated with a 33% reduction in rate of death or heart failure hospitalisation (HR=0.67 95% CI: 0.58, 0.79; $p < 0.0001$) (Table 2).

An increased left ventricular ejection fraction together with larger waist circumference, higher height, QRS duration of at least 130 msec, use of beta blockers at randomisation and higher levels of sodium were also associated with reduced hazard of the primary outcome.

Factors that were strongly ($p < 0.0001$) related to increased hazard of outcome included; older age, higher heart rate, higher levels of creatinine and total bilirubin, previous heart failure hospitalisation, previous myocardial infarction, diabetes mellitus and presence of coronary artery bypass graft.

Variables that were also associated with increased risk of death or heart failure hospitalisation were; race other than white, previous stroke, Chronic obstructive pulmonary disease, presence of left bundle-branch block and an aetiology of Ischemic heart disease.

The female sex had reduced rates of outcome compared to the male but the significance was borderline ($p=0.052$) (Table 2).

Figures 1 and 2 illustrate the effects of some variables significantly associated with outcome.

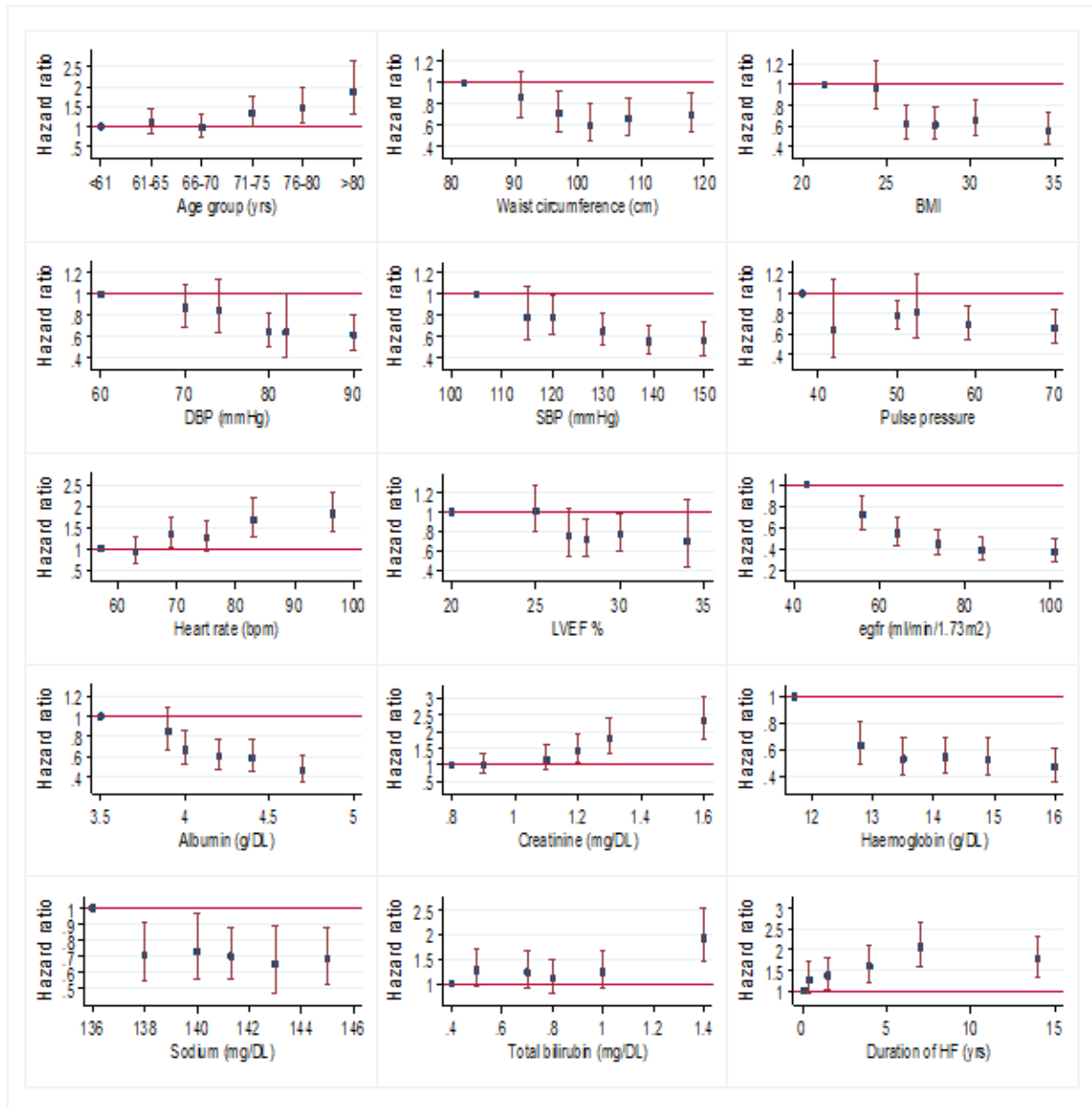


Figure 1: Relationships between continuous variables significantly associated and outcome; Unadjusted HRs and 95% CIs across 6 quantiles of the variable.

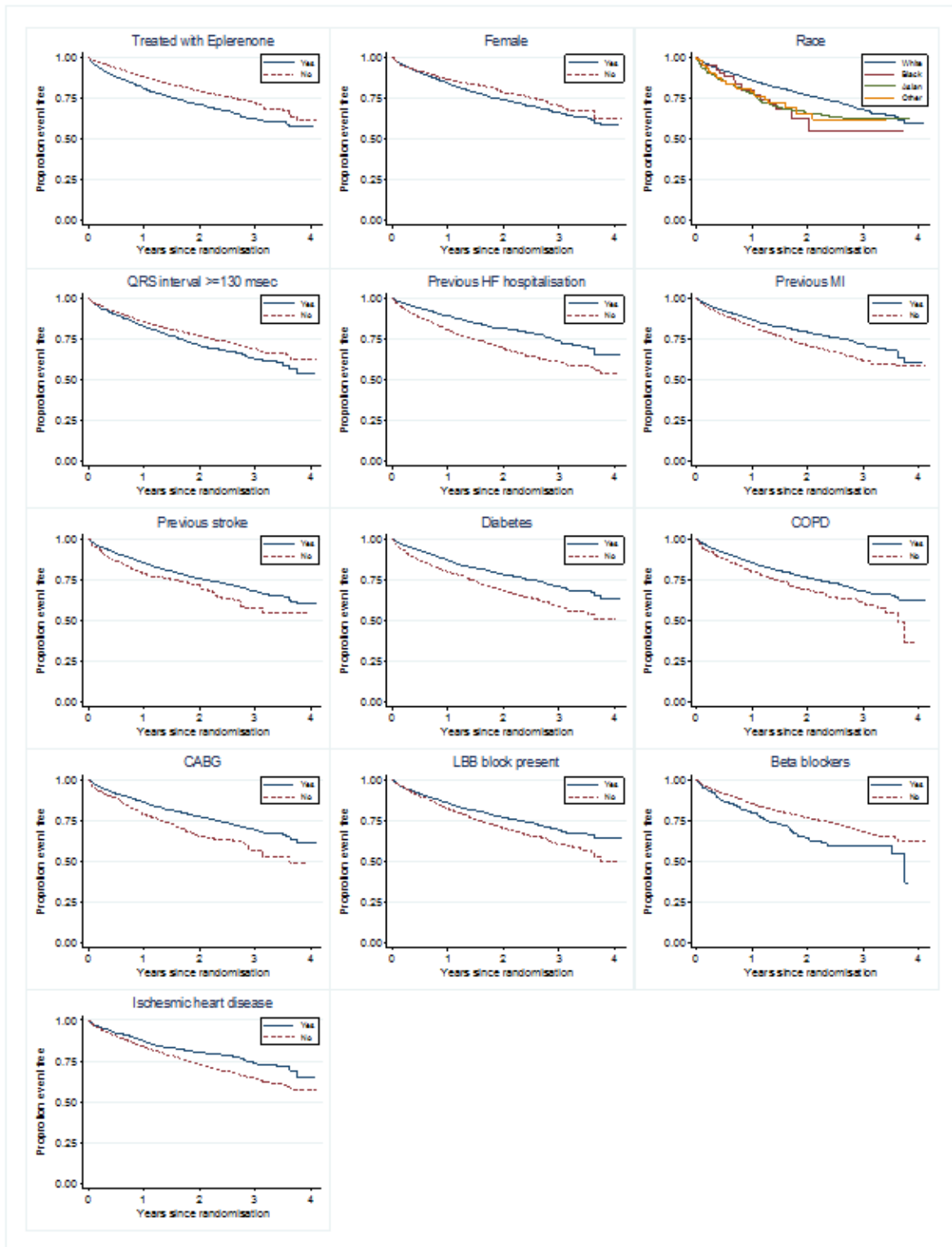


Figure 2: Kaplan-Meier estimates of the rates of death or hospitalisation for heart failure.

3.4 Discussion

A number of variables were identified to be associated with death or heart failure hospitalisation from the univariable analysis. These variables were taken as candidates for inclusion in the next stage of analysis, multivariable models. Among them is age where older subjects were at a higher risk of experiencing the outcome of interest. Height and Weight were also associated with outcome but were not included in the candidates for multivariable models as they are components of the body mass index, which was included as a candidate. There was weak evidence of an association of sex with outcome ($p=0.052$), this factor was nonetheless be included in the next stage of analysis.

Treatment, race and levels of sodium, potassium and albumin were observed to violate the proportional hazards assumption. Since the departure from proportionality could partly arise when other important variable(s) are omitted, the violation of the proportional hazard assumption was not much of a concern at the stage.

3.5 Conclusion

Treatment with Eplerenone was associated with reduced rates of hospitalisation for heart failure and death. Of the demographic and social characteristics, the white race had lower hazard compared to either Black, Asian or other races, increasing age was associated with increased hazard of death or heart failure hospitalization. Higher values of all of the clinical characteristics considered were associated with reduced hazard, except a higher heart rate which was associated with increased hazard. Previous history of heart failure and Myocardial infarction were some of risk factors for subsequent heart failure hospitalisation or death, from previous medical history.

4 Multivariable Cox models

4.1 Introduction

The risk of heart failure hospitalization or death in an individual is dependent on several factors, some of which act in the same direction and others which act in opposite directions as regards inclining or protecting an individual from the outcome. It is therefore necessary to consider all the important predictors simultaneously in defining an individual's risk.

Multivariable Cox proportional hazard regression models were used to bring factors found to be individually related to outcome together for the assessment of their adjusted effects, and whether these effects remained relevant.

4.2 Methods

Two models were developed. The first model was developed from the set of all the variables which were associated with hospitalization for heart failure or death in the univariable analysis, these were from; demographic characteristics; clinical characteristics; factors that measure heart, kidney and liver functionality; previous medical history; ongoing medication and principal cause of heart failure. The second model was built from the same set of predictors but excluding those that were not readily obtainable as they required lab processes to measure, such predictors included the levels of; albumin, creatinine, haemoglobin, sodium, total bilirubin and estimated glomerular filtration rate - since it includes creatinine levels in its computation.

A forward stepwise model selection procedure was applied to arrive at the final models. The inclusion criteria was set at $P < 0.05$. Since the data was from a trial where the study drug eplerenone was significantly associated with reduced heart failure hospitalization and cardiovascular death, a term for treatment was included without being subjected to the stepwise selection criteria, in all models. The demographic variables age and sex were also included a priori.

Complete records on the set of the candidate variables were used in the stepwise selection process. After a model was selected, a Cox PH model was fitted using the complete records on the variables in the model. The resulting hazard ratios reported in this chapter.

4.3 Results

Table 3: Final model (model 1) for heart failure hospitalization and all cause death. Includes laboratory data.

Variable	Adjusted HR*	95% CI	P-value
Treatment: Eplerenone vs. placebo	0.63	(0.53, 0.74)	<0.0001
Age (years)	1.01	(1.00, 1.02)	0.1539
Sex: Female vs. Male	0.76	(0.61, 0.95)	0.0151
eGFR (per 10 ml/min/1.73m ²)	0.87	(0.83, 0.91)	<0.0001
Previous hospitalization for heart failure: Yes/No	1.51	(1.27, 1.79)	<0.0001
Heart rate (per 10 b.p.m)	1.14	(1.08, 1.20)	<0.0001
Myocardial Infarction : Yes/No	1.41	(1.18, 1.67)	0.0001
Albumin (g/DL)	0.73	(0.61, 0.86)	0.0003
Diabetes Mellitus : Yes/No	1.46	(1.23, 1.73)	<0.0001
Body Mass Index (kg/m ²)	0.97	(0.96, 0.99)	0.0060
Race: Black	2.20	(1.33, 3.65)	0.0023
Total bilirubin (mg/DL)	1.45	(1.24, 1.70)	<0.0001
Haemoglobin (g/DL)	0.89	(0.84, 0.94)	<0.0001
Duration of heart failure (years)	1.02	(1.01, 1.03)	0.0061
Systolic blood pressure (per 10 mmHg)	0.93	(0.88, 0.98)	0.0058
COPD: Yes/No	1.34	(1.08, 1.67)	0.0068
Left bundle-branch block: Yes/No	1.22	(1.02, 1.46)	0.0275

+COPD, chronic obstructive pulmonary disease; eGFR, estimated glomerular filtration rate.* each hazard ratio is adjusted for the rest of the factors in the table. Number of observation used:2458

A unit increase in age was associated with 1% increase in the rate of the primary outcome (Adjusted HR 1.01; 95%CI: 1.00, 1.02). Other variables associated with increased risk of the primary outcome in the model were; previous hospitalisation, which was associated with a 51% increase in hazard of primary outcome; heart rate; previous myocardial infarction; diabetes mellitus; black race; total bilirubin levels; duration of heart failure; COPD and presence of a left bundle-branch block. Females were estimated to have a 24% lower hazard compared to males. Eplerenone treatment, higher estimated glomerular filtration rate, higher albumin levels, higher body mass index, higher haemoglobin levels and higher systolic blood pressure were all associated reduced rates of outcome (Table 3).

Table 4: Final model (model 2) for heart failure hospitalization and all cause death. Excludes laboratory data

Variable	Adjusted HR*	95% CI	P-value
Treatment: Eplerenone vs. placebo	0.65	(0.55, 0.76)	<0.0001
Age (years)	1.03	(1.01, 1.04)	<0.0001
Sex: Female vs. Male	0.82	(0.67, 1.02)	0.0748
Previous hospitalization for heart failure: Yes/No	1.54	(1.30, 1.84)	<0.0001
Heart rate (per 10 b.p.m)	1.16	(1.10, 1.22)	<0.0001
Diabetes Mellitus : Yes/No	1.64	(1.38, 1.94)	<0.0001
Systolic blood pressure, SBP (mmHg)	0.90	(0.85, 0.94)	<0.0001
Myocardial Infarction : Yes/No	1.35	(1.13, 1.62)	0.0012
Waist circumference (per 10 cm)	0.90	(0.84, 0.96)	0.0019
Race: Black	2.17	(1.31, 3.62)	0.0027
Duration of heart failure (years)	1.02	(1.00, 1.03)	0.0254
CABG: Yes/No	1.31	(1.07, 1.62)	0.0107
Beta-blocker : Yes/No	0.79	(0.62, 1.01)	0.0629
Left ventricular ejection fraction (%)	0.98	(0.96, 1.00)	0.0216
COPD: Yes/No	1.25	(1.00, 1.56)	0.0551
Left bundle-branch block present	1.20	(1.00, 1.44)	0.0521

+COPD, chronic obstructive pulmonary disease; CABG, coronary artery bypass graft. *each hazard ratio is adjusted for the rest of the factors in the table. Number of observations used:2402

The lab related variables that are in model 1 but were excluded from model 2 are; estimated glomerular filtration rate, albumin levels, total bilirubin levels and haemoglobin levels. Body mass index was not selected from the candidate variables by the stepwise procedure and was therefore present in model 1 but not model 2.

The variables that were not selected in model 1 but are in model 2 are; waist circumference, CABG, left ventricular ejection fraction and indicator of the use of beta-blockers (Table 4).

The adjusted hazards ratios in model 1 and the corresponding ones in model 2 for the variables that are common in both the models are similar.

4.4 Discussion

Forward stepwise selection has been used together with Cox proportional hazards regression to arrive at two predictive models of death or heart failure hospitalisation. The first model, built from both laboratory and non-laboratory measured variables, has seventeen predictors. These come from demographic characteristics, clinical characteristics such as BMI and blood pressure, measures of vital organ functionality such as albumin and total

bilirubin levels, and largely from previous medical history where previous hospitalisation for heart failure, previous myocardial infarction, diabetes and duration of heart failure were some of the important predictors. The second model, which excluded the variables that are measured through laboratory procedures and therefore intended to be predictive from easily measurable quantities, has sixteen predictors in total.

The adjusted effect of age was stronger ($p < 0.0001$) in model 2, which excluded lab variables, than the effect in model 1 ($p = 0.154$). This can mainly be attributed to the exclusion of the estimated glomerular filtration rate in the second model, a measure of kidney functionality which includes age in its formulation.

The rate of death or hospitalisation for heart failure has been estimated to reduce with increase in either body mass index or systolic blood pressure. The effect of these two factors are in the opposite direction of what would be expected. Higher BMI among heart failure patients may be associated with reduced rates of wasting leading to a better prognosis. The reason for the reduced rate of outcome with increased systolic blood pressure is not obvious but may relate to a protective mechanism from factors influencing cardiac output.

The p-values associated with the variables COPD, the use of a beta-blocker and the presence of the left bundle-branch block in the second model (Model 2) are slightly above the set p-value (0.05) for inclusion of a variable stepwise selection procedure. This is because the model was refitted with more cases following the selection of variables.

Inferences from these models are based on the proportionality assumption implied by fitting Cox PH regression models. The proportional hazards assumption was tested globally for the two predictive models and there was no evidence of violation of the assumption.

4.5 Conclusion

Factors that are important in predicting the risk of death of heart failure hospitalisation are varied and act either to reduce the risk or project it. Those identified to be important in predicting the risk of the primary outcome are sixteen, excluding treatment group which was added in consideration that the data are from a clinical trial. When levels of albumin, total bilirubin and haemoglobin are left out of the set of candidate variables the resulting prognostic model contains fifteen predictors - excluding treatment group - with the use of beta-blockers, presence of coronary artery bypass graft, left ventricular ejection fraction and waist circumference becoming more prognostic in the opportunity.

5 Missing data

5.1 Introduction

The adjusted hazard ratios quantifying the level of association between variables and primary outcome in models 1 and 2 have been estimated from complete data on the set of variables in each of the models. This means that the information provided by the non-missing values in the omitted records was ignored. Ignoring part of the data may in some situations affect the estimates of hazard materially. In this chapter multiple imputation was used to impute the missing values in order to enable the use of all the patient records. A comparison is made between the estimates obtained by complete case analysis and the estimates after imputation.

5.2 Methods

For each of the two prognostic models the coefficients were re-estimated after imputing missing values within the variables in each model. The missing values were assumed to be missing at random. The imputation procedure created 10 separate datasets each with a different set of imputed values for the missing data. Continuous variables were imputed through multivariable linear regression. Missing categorical variables were predicted from other variables in the model by using logistic regression.

To obtain the coefficient estimates for each of the two models, they were fitted in each of the 10 imputed datasets and then results combined using Rubin's rules.

Results were then compared with those obtained through complete case analysis for any fundamental differences in estimated coefficients.

5.3 Results

The estimated hazard ratios associated with each predictor following multiple imputation of missing values are not materially different from the corresponding hazard ratios estimated in complete data analysis. Comparing the hazard ratios from tables 3 and 5, the largest differences between corresponding hazard ratios are observed for the indicators of diabetes, race and COPD. These differences are 0.09, 0.09 and 0.06 respectively. The differences in other variables do not exceed 0.03.

The 95% confidence intervals are also not fundamentally different after multiple imputation.

Table 5: Final model (model 1) for heart failure hospitalization and all cause death. Includes laboratory data. Estimates after multiple imputation.

Variable	Adjusted HR*	95% CI	P-value
Treatment: Eplerenone vs. placebo	0.62	(0.53, 0.72)	<0.001
Age (years)	1.01	(1.00, 1.02)	0.047
Sex: Female vs. Male	0.74	(0.60, 0.91)	0.005
eGFR (per 10 ml/min/1.73m ²)	0.88	(0.85, 0.92)	<0.001
Previous hospitalization for heart failure: Yes/No	1.48	(1.25, 1.74)	<0.001
Heart rate (per 10 b.p.m)	1.14	(1.09, 1.20)	<0.001
Myocardial Infarction : Yes/No	1.39	(1.18, 1.64)	<0.001
Albumin (g/DL)	0.72	(0.61, 0.85)	<0.001
Diabetes Mellitus : Yes/No	1.55	(1.32, 1.82)	<0.001
Body Mass Index (kg/m ²)	0.97	(0.96, 0.99)	0.003
Race: Black	2.11	(1.29, 3.45)	0.003
Total bilirubin (mg/DL)	1.44	(1.23, 1.69)	<0.001
Haemoglobin (g/DL)	0.89	(0.84, 0.93)	<0.001
Duration of heart failure (years)	1.02	(1.00, 1.03)	0.010
Systolic blood pressure (per 10 mmHg)	0.92	(0.88, 0.97)	0.001
COPD: Yes/No	1.28	(1.04, 1.58)	0.017
Left bundle-branch block: Yes/No	1.19	(1.01, 1.41)	0.043

+COPD, chronic obstructive pulmonary disease; eGFR, estimated glomerular filtration rate. *each hazard ratio is adjusted for the rest of the factors in the table. Number of observation used:2737

In the model that excludes laboratory variables, the estimated effects of each of the variable in the model were also close to those obtained from complete case analysis (Table 4 and 6). The differences in corresponding hazard ratios estimated through complete case analysis and after missing value imputation did not exceed 0.04. This is with the exception of corresponding hazard ratios associated with myocardial infarction and race where the differences were 0.07 and 0.23 respectively. The 95% confidence intervals for the estimates with and without imputation were also not materially different.

Table 6: Final model (model 2) for heart failure hospitalization and all cause death. Excludes laboratory data. Estimates after multiple imputation.

Variable	Adjusted HR*	95% CI	P-value
Treatment: Eplerenone vs. placebo	0.65	(0.55, 0.76)	<0.001
Age (years)	1.03	(1.02, 1.04)	<0.001
Sex: Female vs. Male	0.80	(0.65, 0.98)	0.028
Previous hospitalization for heart failure: Yes/No	1.53	(1.30, 1.80)	<0.001
Heart rate (per 10 b.p.m)	1.15	(1.10, 1.21)	<0.001
Diabetes Mellitus : Yes/No	1.65	(1.40, 1.94)	<0.001
Systolic blood pressure, SBP (mmHg)	0.89	(0.85, 0.93)	<0.001
Myocardial Infarction : Yes/No	1.28	(1.08, 1.52)	0.004
Waist circumference (per 10 cm)	0.90	(0.84, 0.96)	0.001
Race: Black	1.94	(1.19, 3.18)	0.008
Duration of heart failure (years)	1.02	(1.00, 1.03)	0.022
CABG: Yes/No	1.30	(1.07, 1.58)	0.007
Beta-blocker : Yes/No	0.75	(0.60, 0.94)	0.013
Left ventricular ejection fraction (%)	0.98	(0.96, 1.00)	0.021
COPD: Yes/No	1.25	(1.01, 1.53)	0.039
Left bundle-branch block present	1.21	(1.02, 1.43)	0.030

+COPD, chronic obstructive pulmonary disease; CABG, coronary artery bypass graft.*each hazard ratio is adjusted for the rest of the factors in the table. Number of observations used:2737

5.4 Discussion

Some variables that were selected for inclusion in the prognostic models had missing values. Effort was made, through imputation of missing values, to recover the information that was otherwise ignored by the complete case approach in the estimation of model coefficients. Minimal or no changes have been observed between the estimated hazard ratios obtained through complete case analysis or those following multiple imputation.

Theoretically, materially different estimates would have been observed as a consequence of bias, if the complete cases were not representative of the ignored incomplete cases. The p-values associated with each coefficient were observed to reduce following imputation. For instance, the p-value associated with age in model 1 dropped from 0.154 to 0.047 before and after imputation. The changes in the p-values are however not directly important than changes in the coefficients in the assignment of risk scores to patients.

5.5 Conclusion

Using complete case analysis did not result in fundamentally different estimates of effect of the prognostic variables in the two models developed. The estimated coefficient from the complete case analysis can be used to assign risk scores to subjects without marked bias.

6 Risk Score

6.1 Introduction

Having arrived at prognostic models for death or heart failure hospitalisation the use of the models in computing risk scores for subjects based on their covariate values is described in this chapter. The goodness of fit of the models are also examined and a validation procedure used to check the models' adequacy.

6.2 Methods

A risk score for each individual was calculated by a linear combination of an individual's covariate values and the corresponding coefficients in the prognostic model. Two year survival probabilities of the subjects were computed from the estimated 2-year baseline survival probability and the linear predictor from the prognostic model. The relation between an individual's risk score and their 2-year survival was explored by graphical methods.

Risk scores were divided into quintiles. The discriminative potential of each model was illustrated through Kaplan-Meier plots relating time to death or heart failure hospitalisation within each quintile of the risk scores. Bar plots of observed and expected proportion of outcomes within a two year period in each quintile were plotted to assess goodness of fit of each of the two prognostic model developed.

The discriminative power of the models were assessed through the Harrell's C-indices. The bootstrapping re-sampling technique was used to draw repeated samples. The Harrell's c-index was computed for each of these samples and its sampling distribution summarised as a way of assessing the internal validity of the models. One hundred bootstrap samples were used for each model.

6.3 Results

Subjects were divided into five risk groups labelled 1 to 5 with individuals in the risk group 1 having the lowest risk scores and those in risk group 5 having the highest risk scores.

For the model including the lab variables (model 1) the mean risk score ranged from -4.3 to -2.2 in the lowest risk group to the highest risk group respectively. The rate of primary outcome in the lowest risk group was 4.67 per hundred person-years of observation. The rate in the highest risk group was 38.1 per hundred person-years of observation (Table 5).

The hazard ratio relating the rate of death or heart failure hospitalisation in the lowest risk group to that in the second lowest risk group was 1.66 (95% CI; 1.13, 2.43), and was about 8 times higher when comparing the lowest and the highest risk groups (Table 5).

Table 5: Hazard ratios by quintiles of risk score from model 1

Risk group	Mean risk score	No. of events	Rate per 100 pyo	KM 2-year survival probability	Hazard Ratio	95% CI	p-value
1	- 4.3	43	4.67	0.91	Ref.	-	-
2	- 3.7	68	7.79	0.86	1.66	(1.13, 2.43)	0.009
3	- 3.3	96	11.01	0.81	2.35	(1.64, 3.36)	<0.001
4	- 2.8	155	20.27	0.67	4.26	(3.04, 5.97)	<0.001
5	- 2.2	225	38.06	0.48	7.83	(5.65, 10.86)	<0.001

pyo, person years of observation; KM, Kaplan-Meier; CI, Confidence Interval

For the model excluding the lab data (model 2), the mean risk score ranged from -0.8 to 1.1 in the lowest to the highest risk group similarly labelled 1 to 5. The rate of death or heart failure hospitalisation varied from 5.7 per hundred person-years of observation in the lowest risk group to 34.6 per hundred person-years of observation in the highest risk group.

The hazard ratios quantifying the risk of the primary outcome in higher risk groups compared to the lowest risk group ranged from 1.42 (95% CI; 0.99, 2.05) to 5.87 (95% CI; 4.31, 7.99) in the second lowest risk group to the highest risk groups respectively (Table 6). These hazard ratios were lower than the corresponding ones from the model including lab data.

Table 6: Hazard ratios by quintiles of risk score from model 2

Risk group	Mean risk score	No. of events	Rate per 100 pyo	KM 2-year survival probability	Hazard Ratio	95% CI	p-value
1	- 0.8	50	5.70	0.89	Ref.	-	-
2	- 0.2	70	8.12	0.87	1.42	(0.99, 2.05)	0.056
3	0.1	102	12.87	0.76	2.24	(1.60, 3.14)	<0.001
4	0.5	135	18.08	0.72	3.13	(2.26, 4.33)	<0.001
5	1.1	211	34.56	0.49	5.87	(4.31, 7.99)	<0.001

pyo, person years of observation; KM, Kaplan-Meier; CI, Confidence Interval

Figure 3 illustrates the relationship between risk scores and probability of death or heart failure within 2 years of follow-up.

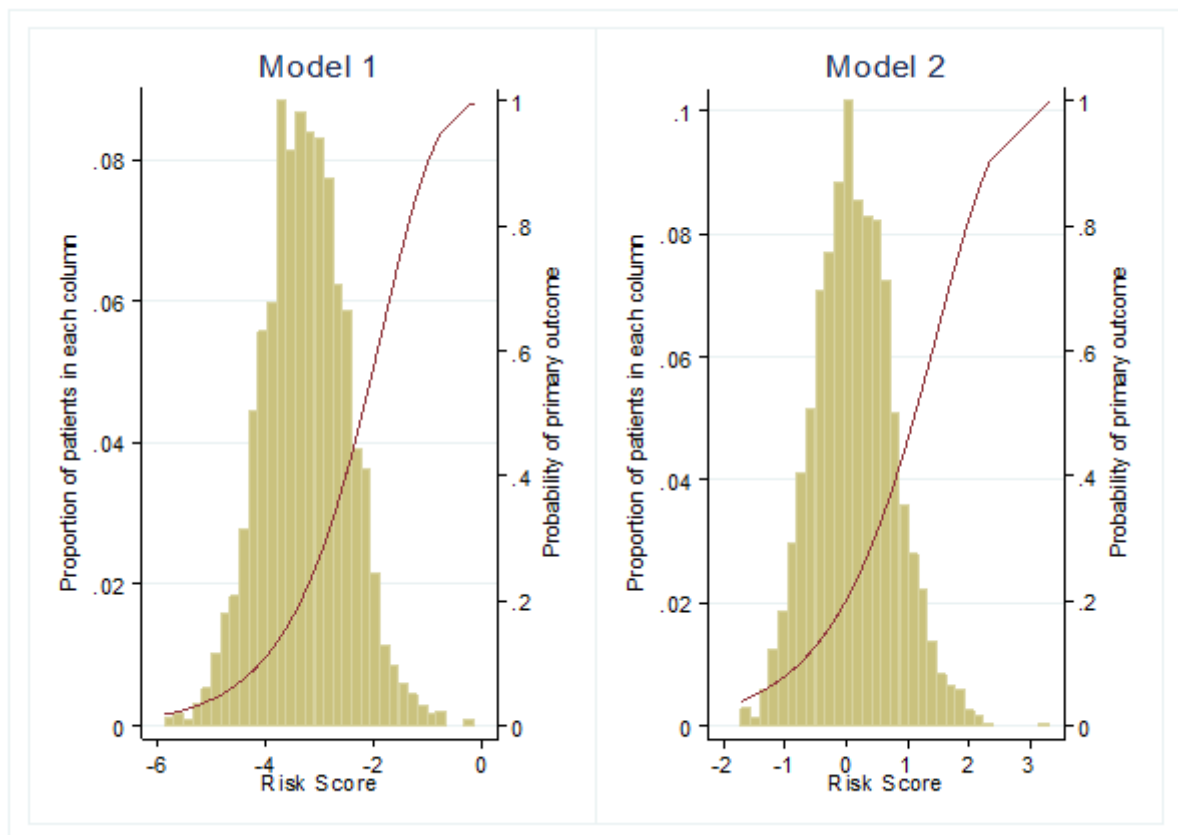


Figure 3: Relationship between risk score and probability of death of heart failure hospitalisation within 2 years. Left panel for model including lab data and right panel for model excluding lab data.

Figure 4 shows the survival experiences of subjects in the five risk groups for each of the two prognostic models. The individual graphs are well separated from each other, suggesting a good discriminative potential, with wider gaps between higher risk groups. There is however a crossing of the survival curves corresponding to the second and third risk groups derived from the model excluding lab variables in later years; between 3 and 4 years of follow up.

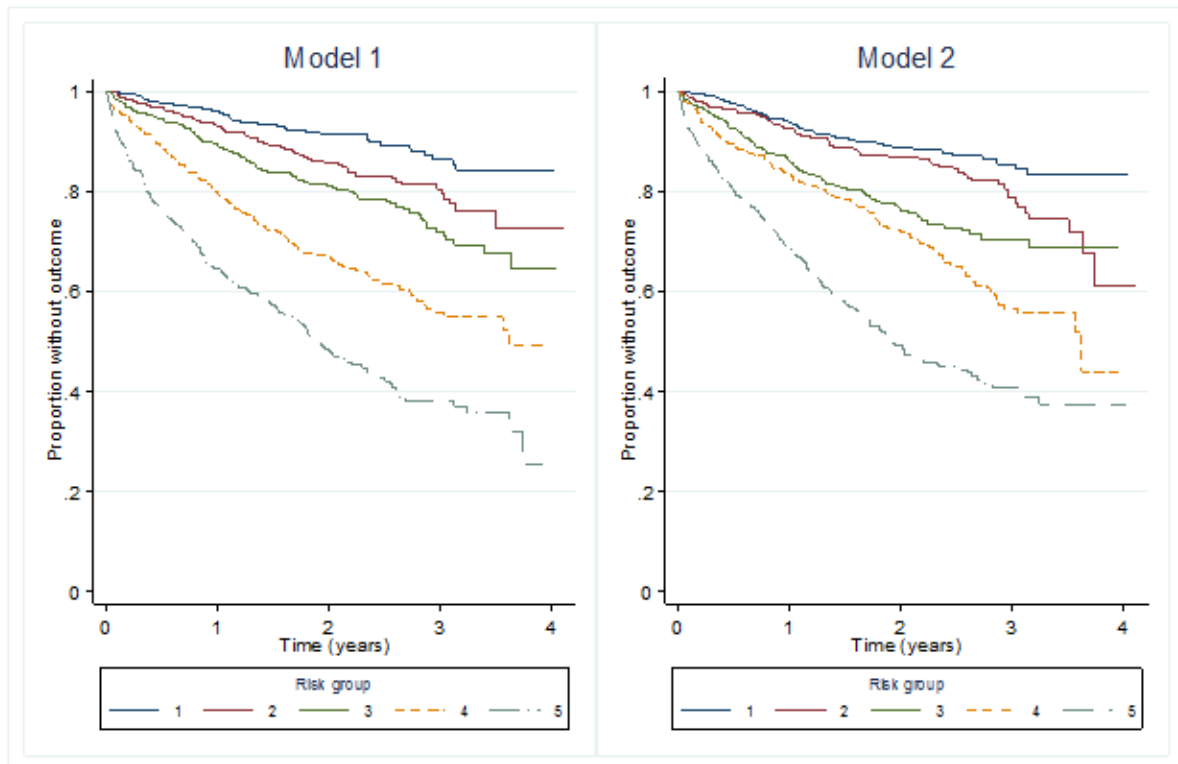


Figure 4: Kaplan-Meier plots by quintiles of risk score. Left panel for model including lab data and right panel for model excluding lab data.

Figure 5 plots the observed and predicted survival probabilities related to death or heart failure hospitalisation within two years of follow up within each risk group. The observed probabilities are from risk group specific Kaplan-Meier survival function at two years. The predicted probabilities are the average of the risk of outcome within 2 years of individuals in a given risk group. The risk of outcome are 2-year survival estimates from Cox models (model 1 and 2). The figure shows very close values of the predicted and observed probabilities within each risk group, apart from predictions from the second model in the highest risk group where the predicted survival probability was much higher. It also shows a steady decrease in the probability of survival across higher risk groups but the gradient is more pronounced across risk groups defined by the model with lab data (Model 1).

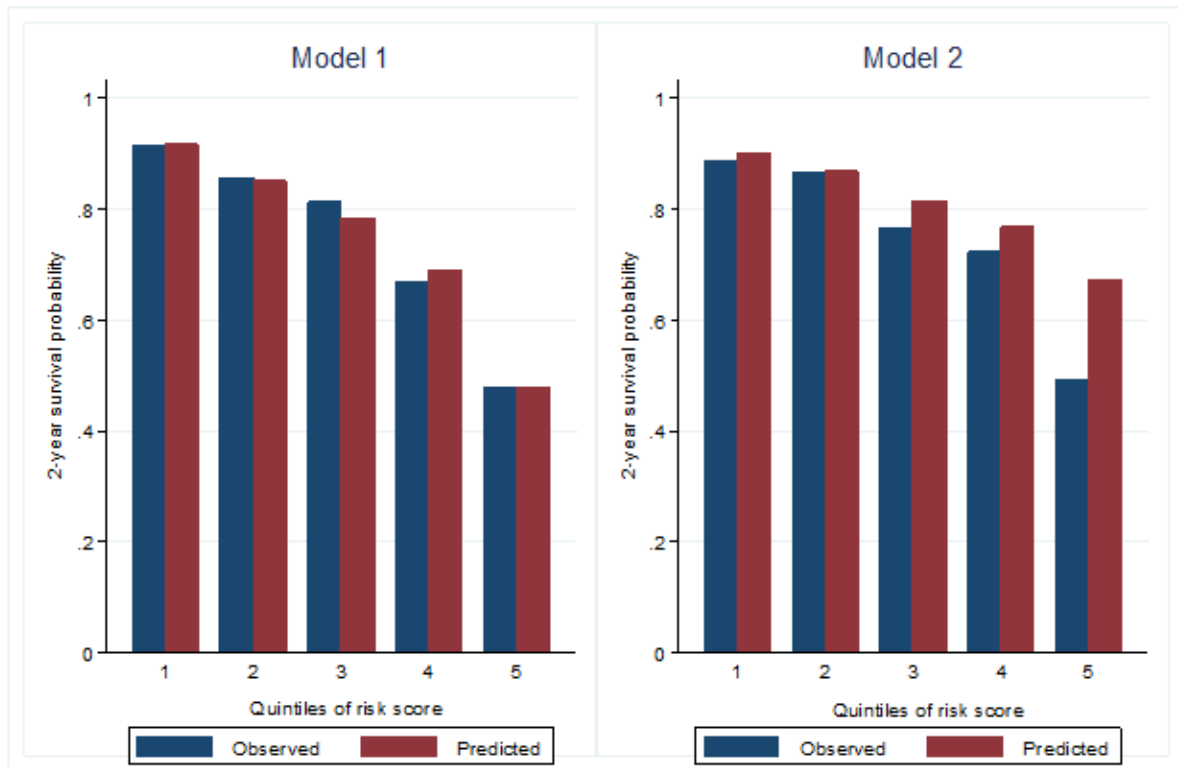


Figure 5: Observed and expected probabilities of 2-year death and heart failure hospitalisation by quintiles of risk score. Left panel for model including lab data and right panel for model excluding lab data.

The discrimination abilities of the two models were moderately strong with, the Harrell's C-statistics were 0.71 and 0.68 for models 1 and 2 respectively. The predictive discrimination power estimated by internal validation using bootstrap samples were 0.72 (range: 0.69-0.75) and 0.69 (range: 0.66-0.72) for the model including lab predictors and that excluding lab predictor respectively.

6.4 Discussion

The two prognostic models of death and heart failure hospitalisation have been used to define five risk groups based on the risk scores from the linear predictors of the models. These risk groups have been summarised in this chapter in terms of the survival experiences within each of them. The performance of the models based on how well they predict the risk of death or heart failure hospitalisation within 2-years of follow up has also been assessed. The model including lab variables had slightly higher discriminative power compared the one excluding the lab variables. The higher ability is expected due to the additional predictive capacity offered by the lab related predictors. However, the slight difference suggest that the additional value of including the estimated glomerular filtration

rate, albumin, total bilirubin and haemoglobin levels –the lab variables selected in model 1- in the prognostic model can be still be achieved without much loss in prognostic ability by including factors such as the left ventricular ejection fraction, CABG and the use of beta-blockers instead. This is mainly important in patient management in cases where lab processing might not be feasible.

6.5 Conclusion

The prognostic model which includes age, sex, race, estimated glomerular filtration rate, indicators of previous heart failure hospitalisation, myocardial infarction, diabetes mellitus and chronic obstructive pulmonary disease, heart rate, systolic blood pressure, duration of heart failure, body mass index, and levels of albumin, total bilirubin and haemoglobin has a good predictive value and discriminative ability.

The prognostic model which does not rely of lab predictors does not perform worse compared to that which includes them and shows a comparable discriminative ability. However, its predictive ability among subjects classified in the high risk group is somewhat poor.

7 Discussion

7.1 Summary of results

The aim of the study was to derive parsimonious prognostic models for predicting death and heart failure hospitalisation from a set of candidate variables. The study has defined two such models using a variable selection technique and Cox proportional hazards regression. One of the models included lab related predictors and the other purposely excluded them with the intention of being prognostic from easily obtainable predictors.

The model including the lab related variables has 16 predictors in total while that excluding them has 15 predictors in total, excluding treatment arm indicator. Treatment with eplerenone is not counted among the prognostic factors since it was included to acknowledge the fact that the data came from a clinical trial where the treatment had an effect on outcome.

The discriminative abilities of the models were moderately strong (Harrell's c-indices of 0.71 and 0.68 for the model with and without lab related predictors respectively). The rate of death or heart failure hospitalisation across risk groups defined by each of the two models were well varied. The rates across the risk groups defined by the model including the lab predictors were 4.67 and 38.1 per hundred person-years of observation in the lowest and the highest risk groups respectively. The corresponding rates across risk groups defined by the model excluding the lab predictors were 5.7 and 34.6 per hundred person-years of observation respectively. This reflects an 8-fold and 6-fold difference in rates of the primary outcome between the highest and lowest risk groups defined by the prognostic models.

The observed and predicted probabilities of death or heart failure hospitalisation within two years of follow-up by quintiles of risk score were quite close in each of the models. This is with the exception of the highest risk group in the model excluding the lab predictors, where there was a possible over prediction.

7.2 Comparison with other studies

Some of the variables identified as important predictors of death or heart failure hospitalisation in the prognostic models have previously been identified in other studies. Previous hospitalisation for heart failure, heart rate, myocardial infarction and diabetes mellitus were among the factors strongly associated with the primary outcome in both the prognostic models developed.

Previous hospitalisation for heart failure was independently associated with 51% and 54% increased risk of subsequent death or hospitalisation for heart failure in the model with and without the lab predictors respectively. A history of previous hospitalisation for heart failure has also been shown to be predictive but more so when the time since the previous hospitalisation is not longer than six months. A strong effect of diabetes of projecting the risk of heart failure hospitalisation and death has also been reported.¹⁶

Increasing age has been associated with the increased risk of heart failure hospitalisation in several studies.^{9,10} In the model including the lab predictors the effect of age was not pronounced and was not statistically significant. However, in the model excluding the lab predictors the effect of age was stronger ($p < 0.0001$), with a unit increase in age associated with a 3% increase in the hazard of death or heart failure hospitalisation. The gain in strength of age as a predictor in the model excluding the lab predictors can be attributed to the consequential absence of the lab related variable, estimated glomerular filtration rate, which by computation includes age.

Being white was associated with reduced rate of death or heart failure hospitalisation in comparison to other races. The black race was associated with a doubling in the risk of death or heart failure hospitalisation in both the models derived in the current study. Evidence from previous studies have pointed towards an earlier progression into worsening systolic dysfunction among blacks with heart failure.¹⁹⁻²¹ A previous report hypothesises that mechanism variations in sodium intake and handling, potassium and calcium intake, fasting insulin levels, plasma renin activity and urinary kallikrein excretion could potentially identify whether the neurohormonal milieu in heart failure with renal impairment may differ by race. Such differences could explain mortality risks by race.²² The mechanisms have been proposed as causal factors for race differences in hypertension.^{23,24}

An elevated body mass index (BMI) has been associated with increased risk of heart failure in patients without heart failure at baseline.²⁵⁻²⁹ However, among patients with heart failure, elevated BMI at earlier stages of follow-up were associated with a better prognosis.³⁰⁻³² A reduction in the rate of death or heart failure hospitalisation was also associated with raised BMI at baseline in the current study. Higher BMI among heart failure patients may be associated with reduced rates of wasting leading to a better prognosis.³³ Part of the association of elevated BMI with a favourable prognosis among heart failure patients may be mediated by adiponectin and N-terminal pro brain natriuretic peptide (NT-proBNP). A high adiponectin level was a predictor of mortality in a study where adiponectin level was positively associated with NT-proBNP among patients with established chronic heart failure, and both biomarkers were negatively associated with BMI.³⁴

The rate of death or hospitalisation for heart failure has been estimated to reduce with increase in systolic blood pressure in both the prognostic models developed in the current work. Similar finding on the direction of effect of systolic blood pressure has been observed elsewhere.³⁵ The reason for the reduced rate of outcome with increased systolic blood pressure is not clear but is posited to be due to acute protective mechanism from factors influencing cardiac output or vascular tone that are yet to be elucidated.³⁵

7.3 Strengths and weaknesses

The cases in the data from which the models are developed were recruited from rather diverse backgrounds in terms of race and geography. The subjects were sourced from a total of 29 countries worldwide in 278 centers and included the White, Black, Asian and others races. The results could therefore be applicable to the older population across the wide range of countries.

The models arrived at from the current work might not be well prognostic in all patients with mild symptoms. The data from which the models were built was further restricted by design to people aged above 55 years of age and in majority of cases having an ejection fraction above 30%. As the additional selection criteria are associated with increased risk of heart failure, the models might perform poorly among younger individuals with mild symptoms.

When data are not missing completely at random, the complete cases are more likely to be similar than the underlying population. The resulting residual variance is then smaller - meaning more variables are likely to be selected than should be - since the standard errors of the coefficients would be smaller.³⁶ Novel variable selection methods for the multiply imputed data that jointly fit models on the imputed data leading to a consistent selection of variables across all imputed datasets have been recently developed.^{36,37}

The current analysis did not use the novel methods of stepwise variable selection when some variables have missing values, but used complete cases instead. However, the rates of death or heart failure hospitalization were computed - for 5 most frequently missing variables - for subjects who did and did not have missing values in each of the variables and were found to be similar, suggesting no underlying differences prevailed between the subjects with and without missing values in the concerned variables. This similarity was important as it lowered the chances that any of the variables selected managed through the selection criteria as a consequence of other candidate variables having missing values.

Multiple imputation was performed after model selection on the variables that managed through the selection process but had missing values; the re-estimated hazard ratios after imputation were not markedly different from those estimated from complete case analysis.

7.4 Further analyses

The plots of hazard ratios across 6 quantiles of continuous variables in the univariable analysis showed that some variables would possibly fit better in the model if categorised (Figure 1). Age for instance had a modest impact on the rate of death or hospitalisation for heart failure until after age 70 years where it was associated with up to a 2-fold increase in hazard. A body mass index of above 25 kg/m² was associated with a patent drop in hazard than the hazard observed in lower quantiles of the BMI. Other variables that suggested some non-linearity in their effect included haemoglobin and sodium levels. An analysis that includes these variables after categorisation would be important to assess whether doing so would improve the discriminative and predictive abilities of the current models in this study.

The application of novel stepwise procedures for variable selection with multiply imputed data to see asses the sensitivity of the selection procedure to the missing values in the data would be appropriate.

The models developed in the current study were only validated internally through bootstrap re-sampling. Though the results of the validation process suggested a good precision, external validation of the models would be a better way of objectively testing their validity.

7.5 Conclusion

In conclusion, the prognostic models developed have good predictive value and discriminative ability and can be applied in predicting mortality or heart failure hospitalisation in a wide range of settings, either in the absence or presence of already established lab predictors. The models should however be used with caution in younger populations as they have not been externally validated in such populations where their performance could be poor.


References

1. McGovern PG, Pankow JS, Shahar E, et al. Recent trends in acute coronary heart disease--mortality, morbidity, medical care, and risk factors. The Minnesota Heart Survey Investigators. *The New England journal of medicine*. 1996;334(14):884-90.
2. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics--2011 update: a report from the American Heart Association. *Circulation*. 2011;123(4):e18-e209.
3. Callow AD. Cardiovascular disease 2005--the global picture. *Vascular pharmacology*. 2006;45(5):302-7.
4. T G, WP C, MC H, Kannel WB, TR D. Diabetes, blood lipids, and the role of obesity in coronary heart disease risk for women. *Ann Intern Med*. 1977;87:393-397.
5. Brand FN, Abbott RD, Kannel WB. Diabetes, intermittent claudication, and risk of cardiovascular events. The Framingham Study. *Diabetes*. 1989;38(4):504-509.
6. Gordon T, Kannel WB. Multiple risk functions for predicting coronary heart disease: The concept, accuracy, and application. *American Heart Journal*. 1982;103(6):1031-1039.
7. Steuer J, Granath F, Faire U de, Ekbom A, Stähle E. Increased risk of heart failure as a consequence of perioperative myocardial injury after coronary artery bypass grafting. *Heart (British Cardiac Society)*. 2005;91(6):754-8.
8. Ingelsson E, Arnlöv J, Sundström J, et al. Novel metabolic risk factors for heart failure. *Journal of the American College of Cardiology*. 2005;46(11):2054-60.
9. O'Connor CM, Hasselblad V, Mehta RH, et al. Triage after hospitalization with advanced heart failure: the ESCAPE (Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness) risk model and discharge score. *Journal of the American College of Cardiology*. 2010;55(9):872-8.
10. Vazquez R, Bayes-Genis A, Cygankiewicz I, et al. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure. *European heart journal*. 2009;30(9):1088-96.
11. Jones RC, Francis GS, Lauer MS. Predictors of mortality in patients with heart failure and preserved systolic function in the Digitalis Investigation Group trial. *Journal of the American College of Cardiology*. 2004;44(5):1025-9.
12. Singh M, Gersh BJ, Li S, et al. Mayo Clinic Risk Score for percutaneous coronary intervention predicts in-hospital mortality in patients undergoing coronary artery bypass graft surgery. *Circulation*. 2008;117(3):356-62.
13. Matthews JC, Koelling TM, Pagani FD, Aaronson KD. The right ventricular failure risk score a pre-operative tool for assessing the risk of right ventricular failure in left ventricular assist device candidates. *Journal of the American College of Cardiology*. 2008;51(22):2163-72.
14. Adlam D, Silcocks P, Sparrow N. Using BNP to develop a risk score for heart failure in primary care. *European heart journal*. 2005;26(11):1086-93.

15. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837-47.
16. Pocock SJ, Wang D, Pfeffer M a, et al. Predictors of mortality and morbidity in patients with chronic heart failure. *European heart journal*. 2006;27(1):65-75.
17. Zannad F, McMurray JJV, Krum H, et al. Eplerenone in patients with Systolic Heart Failure and Mild Symptoms. *The New England Journal of Medicine*. 2011;364(1):11-21.
18. Zannad F, McMurray JJV, Drexler H, et al. Rationale and design of the Eplerenone in Mild Patients Hospitalization And Survival Study in Heart Failure (EMPHASIS-HF). *European journal of heart failure*. 2010;12(6):617-22.
19. Yancy C. Heart failure in blacks: etiologic and epidemiologic differences. *Curr Cardiol Rep*. 2001;3:191-197.
20. Dries D, Exner D, Gersh B, et al. Racial differences in the outcome of left ventricular dysfunction. *N Engl J Med*. 1999;340:609-616.
21. East M, Jollis J, Nelson C, Marks D, Peterson E. The influence of left ventricular hypertrophy on survival in patients with coronary artery disease: do race and gender matter? *J Am Coll Cardiol*. 2003;41:949-954.
22. Smith GL, Shlipak MG, Havranek EP, et al. Race and renal impairment in heart failure: mortality in blacks versus whites. *Circulation*. 2005;111(10):1270-7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15769768> [Accessed August 18, 2011].
23. He J, Klag M, Appel L, Charleston J, Whelton P. The renin-angiotensin system and blood pressure: differences between blacks and whites. *Am J Hypertens*. 1999;12:555-562.
24. Richardson A, Piepho R. Effect of race on hypertension and antihypertensive therapy. *Int J Clin Pharmacol Ther*. 2000;38:75-79.
25. Ingelsson E, Sundstrom J, Arnlov J, Zethelius B, Lind L. Insulin resistance and risk of congestive heart failure. *JAMA*. 2005;294:334-341.
26. Kenchaiah S, Evans J, Levy D, et al. Obesity and the risk of heart failure. *N Engl J Med*. 2002;347:305-313.
27. Wilhelmsen L, Rosengren A, Eriksson H, Lappas G. Heart failure in the general population of men: morbidity, risk factors and prognosis. *J Intern Med*. 2001;249:253-261.
28. Johansson S, Wallander M, Ruigomez A, Garcia Rodriguez L. Incidence of newly diagnosed heart failure in UK general practice. *Eur J Heart Fail*. 2001;3:225-231.
29. Kenchaiah S, Sesso HD, Gaziano JM. Body mass index and vigorous physical activity and the risk of heart failure among men. *Circulation*. 2009;119(1):44-52.
30. Horwich T, Fonarow G. The impact of obesity on survival in patients with heart failure. *Heart Fail Monit*. 2002;3:8-14.
31. Lavie C, Osman A, Milani R, Mehra M. Body composition and prognosis in chronic systolic heart failure: the obesity paradox. *Am J Cardiol*. 2003;91:891-894.

32. Davos C, Doehner W, Rauchhaus M, et al. Body mass and survival in patients with chronic heart failure without cachexia: the importance of obesity. *J Card Fail.* 2003;91:29-35.
33. Poehlman E, Scheffers J, Gottlieb S, Fisher M, Vaitekevicius P. Increased resting metabolic rate in patients with congestive heart failure. *Ann Intern Med.* 1994;121:860-862.
34. Kistorp C, Faber J, Galatius S, et al. Plasma adiponectin, body mass index, and mortality in patients with chronic heart failure. *Circulation.* 2005;112(12):1756-62.
35. Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D. Predicting Mortality Among Patients Hospitalized for Heart Failure. *JAMA.* 2003;290(19):2581-2587.
36. Lachenbruch P a. Variable selection when missing values are present: a case study. *Statistical methods in medical research.* 2010.
37. Wood A, White I, Royston P. How should variable selection be performed with multiply imputed data? *Statistics in Medicine.* 2008;27(17):3227-46.

Appendix

<p>London School of Hygiene & Tropical Medicine (University of London) Combined Academic, Risk assessment and Ethics (CARE) approval form for MSc Project Reports</p>	
<p><i>*This form must be completed electronically. For detailed guidance, please refer to the Project Handbook for your course.</i></p>	

SECTION 1 – STUDENT AND COURSE INFORMATION	
MSc DETAILS AND DEADLINES (deadlines to be communicated by Course Director)	
Academic Year	2010-11
MSc course (and stream, where applicable)	Medical Statistics
Deadline for Supervisor approval	Friday 25 March 2011
Deadline for Course Director approval	Friday 25 March 2011
Deadline for submission to Ethics Committee	Friday 25 March 2011
Target for approved form to be passed to TSO	Friday 27 May 2011
STUDENT AND SUPERVISOR DETAILS (to be completed by student)	
Full name of student	
Student email address	
Year of study (part-time students only)	<input type="checkbox"/> First Year <input type="checkbox"/> Second Year
Supervisor name	Tim Collier
Supervisor email address	Timothy.Collier@lshtm.ac.uk
Supervisor institution/organisation	LSHTM
Supervisor status (at time of this version of the form being completed)	<input checked="" type="checkbox"/> Confirmed <input type="checkbox"/> Provisional <input type="checkbox"/> Still to be identified
Name of personal tutor (where Supervisor is still to be identified)	Tim Collier

SECTION 2 – APPROVAL AND SUBMISSION STATUS	
<p><i>*Students please note: It is a requirement of your LSHTM degree that you obtain all required approvals <u>before</u> beginning your project work. Your Supervisor and Course Director must specifically give Risk Assessment approval. Ethics approval must also be obtained where necessary (answers in Section 5 will help determine if this is required or not).</i></p>	

STUDENT DECLARATION (to be completed for all projects)	
<p>I agree to conduct my project on the basis set out in this form, and to consult staff (initially, my Supervisor) if making any subsequent changes – especially any that would affect the information given with respect to ethics approval.</p>	<input checked="" type="checkbox"/>
<p>I agree to comply with the relevant safety requirements, and will submit a separate request for LSHTM travel insurance where relevant.</p>	<input checked="" type="checkbox"/>
<p><i>*Where seeking ethics approval for a study involving human subjects, please also attach copies of any information sheets, consent forms, and other relevant documents.</i></p>	
Date of declaration	8 March 2011

**Please save the electronic file of this CARE form in the format
“[MSc title]_[Year of Submission]_[Surname]_[Forename]_CARE”**

You will also be required to submit a copy of this CARE form with your final written-up project. This should be anonymised, i.e. with your name and email address removed.

STAFF APPROVAL	
<p>*Staff please note: Sections 3 and 4 of the form should be completed by the student before you give approval. Rather than 'sign' this form, you should email the student and explicitly confirm approval, e.g. stating "In my role as supervisor, I approve the attached form". The student is then responsible for updating the form and passing it on to any other staff.</p> <p>However if you would answer 'no' to any of the 'Yes/No' questions below, or disagree with any of the statements given, or have any other concerns, then you should not give approval. Instead, please contact the student immediately to inform them of your concerns and discuss changes which they may need to make before you may be willing to give approval.</p> <p>Please also be aware that in the exceptional case of a request to undertake a project in a country or region to which the Foreign & Commonwealth Office advise against travel, the student would need to fill out a separate form which will then need further School-level approval by the Safety Manager and Secretary & Registrar.</p>	

SUPERVISOR'S APPROVAL (required for all projects – this approval should be given first)	
Supervisor has agreed that Section 3 of this form is a reasonable summary of the proposed project.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Supervisor has agreed that responses in Section 4 of this form address the main risks connected with a project of this nature.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Name of Supervisor (if not yet identified, personal tutor <u>or</u> Course Director should approve)	Tim Collier
Date of approval	11 March 2011

COURSE DIRECTOR'S APPROVAL (required for all projects – should follow Supervisor approval)	
Course Director has agreed that the proposed project's academic content, set out at Section 3 of this form, is suitable for this MSc.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Course Director has agreed that responses in Section 4 of this form address the main risks connected with a project of this nature.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Name of Course Director (or nominee)	Tim Collier
Date of approval	11 March 2011

FACULTY SAFETY SUPERVISOR'S APPROVAL (only required if project involves working with pathogenic organisms, human blood or radiochemicals – should follow Supervisor approval)	
Faculty Safety Supervisor has agreed that the proposed project, as set out in this form and particularly Section 4, may proceed.	<input type="checkbox"/> Yes <input type="checkbox"/> No
Name of Faculty Safety Supervisor (or nominee)	
Date of approval	

ETHICS APPROVAL (required for all projects involving human subjects or human data, except for public domain data that cannot enable the identification of living people – NB that Supervisor approval must have been received before the application is submitted to the Ethics Committee)	
The Ethics Committee has approved the project proposal set out on this form.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Date of approval	11/04/2011
Ethics Committee application number assigned	010/112

SECTION 3 – APPLICATION FOR ACADEMIC APPROVAL

**All students should complete all sub-sections (3.1, 3.2 and 3.3). If particular questions are not applicable to you then please write 'N/A'.*

3.1 PROJECT OUTLINE (should not normally exceed 750 words total)

Proposed project title: (should not normally exceed 20 words)

Risk Score Model for Heart Failure Patients

Proposed project type:

**See course-specific section of Project Handbook for details of project types permitted for each MSc. Be aware that restrictions may apply for individual courses.*

Analysis of an existing dataset

Proposed project length:

**For almost all students, this will be 'Standard'. Long and extended projects are only available for certain ITD courses; they have a different schedule and allow a slightly greater word count.*

Standard Long Extended

Background: (about 200 words)

**Indicate why this topic is of interest or relevance.*

**If the project involves work with a specific organisation please give details.*

**Please give any other details specifically relevant for consideration by the Ethics Committee, e.g. related to purpose.*

Identifying patients at high risk of cardiovascular mortality and morbidity will be useful in helping to decide on treatment strategy. Cardiovascular conditions are a significant contributor to mortality and hospital admission. Identification of patient at increased risk from a set of easily obtainable clinical characteristics remains a challenge despite its importance in patient management.

Hypothesis: (about 30 words, where applicable)

Baseline characteristics will help discriminate between patients at lower and higher risk of subsequent cardiovascular events and mortality.

Overall aim of project: (about 30 words)

To produce a risk score model for cardiovascular mortality/morbidity in patients with New York Heart Association class II chronic systolic heart failure.

Specific objectives of project: (about 70 words)

To identify baseline patient characteristics (including clinical, demographic, anthropometric measurement) that are associated with cardiovascular mortality/morbidity.

To build a multivariable risk model for cardiovascular mortality/morbidity.

To refine and test the predictive power of the risk model.

Proposed methods: (about 200 words)

Please summarise methods, and include **any relevant details for consideration by the Ethics Committee such as numbers of participants and procedures to be performed.*

The proposed project will use data from a randomized placebo controlled trial of eplerenone in 2737 subjects with New York Heart Association class II chronic systolic heart failure. The primary endpoint in the trial was a composite of cardiovascular mortality or hospitalisation for worsening heart failure. The current project will apply a

range of statistical methods, e.g Cox regression, to the data to develop risk scores.

References: (max 150 words)

**List any key references which will shape the project, including for methods to be used. It should not normally be necessary to quote more than 5 references.*

Faiez Z, John J.V.M, Henry K, et al. Eplerenone in patients with Systolic Heart Failure and Mild Symptoms. N Engl J Med 2011;364:11-21

Prior work: (only where relevant; max 100 words)

**Indicate any previous work you have done related to this project topic, including student work, professional work, or publications.*

3.2 FEASIBILITY (about 100 words total – but can write more or write less if appropriate)

What could cause this project to fail, i.e. prevent you from achieving your objectives?

**Please indicate any aspects of your proposed approach which could potentially experience difficulties, e.g. delays with permissions, data collection or storage problems, lack of sufficient comparable information, etc. You may also wish to mention any wider matters which could affect your project, e.g. civil unrest, natural disasters, transport availability.*

There is no foreseeable reason why this project might fail. The study has been conducted and the data have been collected and the data provider has agreed to provide the data.

What alternative plans do you have in case you encounter any of the potential problems you have identified?

If the data provider fails to provide the data there is a list of alternative projects.

3.3 INTELLECTUAL PROPERTY, COPYRIGHT AND OTHER PERMISSIONS

**Please also see Section 5.2 regarding any specific data rights limitations arising from local ethics or research governance requirements*

If you expect to use existing data, how will you obtain it and what permissions will be required?

The data provider has agreed to send the data. I will be using the data with their permission.

Having considered whether intellectual property rights (IPR) or copyright issues may affect your project, will any specific agreements be required?

**Please tick all boxes that apply, and attach copies of any forms/agreements (even if in draft).*

- No specific IPR, Copyright or permissions issues should apply to this project (student retains Copyright and related IPR by default, in line with LSHTM registration declaration)
- IPR to be retained by LSHTM (specific LSHTM form to be completed)
- Copyright to be transferred to LSHTM (specific LSHTM form to be completed)
- IPR, Copyright or other agreements/permissions required with external parties/organisations

Please give any further relevant details about IPR, copyright or other

permissions.

The project will require the student to sign a confidentiality agreement. The data remains the property of Pfizer and should be returned at the end of the project and any copies deleted.

SECTION 4 – APPLICATION FOR RISK ASSESSMENT APPROVAL

**All students should answer all questions in sub-section 4.1; this will make clear which of the subsequent sub-sections you need to complete.*

Ensuring safety during project work is the responsibility of each individual student, and not of LSHTM or LSHTM staff. **Please see the Project Handbook for further guidance.*

4.1 TYPE OF RISK (to be completed by all students)

Where will the project be carried out? (please tick all that apply)

**Note that work away from LSHTM or outside the UK means any form of work for your project, not just primary data collection. Some courses may have specific restrictions on this.*

- All work will take place either at LSHTM, in libraries in the UK, or at my personal residence in the UK.** [If so, you do not need to complete either section 4.2 or section 4.3]
- Some work will take place in the UK that is away from LSHTM sites in London, is non-Library-based, and is not at my personal residence.** [If so, section 4.2 on 'Work away from LSHTM' must be completed]
- Some work will take place at my personal residence outside the UK** [If so, section 4.3 on 'Work outside the UK' must be completed]
- Some work will take place outside the UK that is not at my personal residence** [If so, both sections 4.2 and 4.3 on 'Work away from LSHTM' and 'Work outside the UK' must be completed]

Will the project involve working with or handling any of the following materials?

- Pathogenic organisms** Yes No
- Human blood** Yes No
- Radiochemicals** Yes No

[If 'Yes' to any of the above, Sections 4.4 and 4.5 must be completed]

Are any other potentially hazardous activities likely to be carried out during the project?

- Yes No

[If 'Yes', Section 4.5 must be completed]

Do any special requirements (e.g. disability-related issues) or other concerns need to be taken into account for either you as a student, study participants or colleagues?

- Yes No

[If 'Yes', Section 4.6 must be completed]

4.2 WORK AWAY FROM LSHTM (to be completed if any work will be done away from LSHTM, other than at your home or at libraries elsewhere in the UK)

<p>Will the project be based in an established hospital, college, research institute, NGO headquarters, field station or other institutional site? If 'Yes', please give the name and location of the site(s); describe approximately what proportions of your project will be spent there; and state name and role of person who has confirmed willingness to support you at each site (indicating extent of correspondence, especially what they have confirmed in writing).</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<p>Will you have an 'external supervisor', co-supervisor or other main advisor, or be working with any specific organisation(s), during your work away from LSHTM? If 'Yes', please indicate the name, role, contact details, and level of support that any such external advisors are expected to provide, and give details about any organisations you will be working with.</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<p>Will the project involve personal visits, interviews or interactions with people in their homes, workplaces, community settings or similar? If 'Yes', please give details, including approximately what proportion of your project this will involve.</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<p>Will the project involve lone/isolated work or significant travel? If 'Yes', please give details, including approximately what proportion of your project this will involve, and state how you can be contacted while working or travelling.</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
<p>What arrangements are proposed for contact with your main supervisor while you are away from LSHTM? Indicate expected ease and frequency of contact, and communication methods to be used.</p>	
<p>N/A</p>	
<p>Please tick to confirm:</p>	<p><input checked="" type="checkbox"/> I have read the LSHTM Code of Practice on off-site work.</p>

4.3 WORK OUTSIDE THE UK (to be completed if any work will be done outside the UK)

<p>What form of project work will be undertaken outside the UK? (please tick all that apply)</p>	
<p> <input type="checkbox"/> Work at my family home or personal residence only <input type="checkbox"/> Work at an established hospital, college, research institute, NGO headquarters, field station or other institutional site <input type="checkbox"/> Work away from my personal residence or an established site <i>*Note that for either the second or third options, you should also have completed Section 4.2.</i> </p>	
<p>Name the country/countries and region(s) in which work will be undertaken:</p>	
<p>Country or countries: N/A Region(s) : N/A</p>	
<p>Do the Foreign & Commonwealth Office's (FCO) Travel Advice</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No

<p>Notices (www.fco.gov.uk/en/travelling-and-living-overseas/travel-advice-by-country/) advise against travel to the regions(s), country or countries involved?</p> <p><i>*Note that if 'Yes', the School will not normally permit such travel for project work. In exceptional circumstances only, requests may be considered by the Safety Committee and require approval by the Safety Manager and Secretary & Registrar.</i></p>	No
<p>Please tick to confirm:</p>	<p><input type="checkbox"/> I understand that LSHTM travel insurance is required for any international travel as part of my project.</p> <p><i>*Travel insurance can be applied for using a separate form.</i></p>

<p>4.4 WORK WITH HAZARDOUS SUBSTANCES (to be completed if the project involves any work with pathogenic organisms, human blood or radiochemicals – NB that this will require approval by the Faculty Safety Supervisor)</p>	
<p>Name the organism or organisms to be used:</p>	
<p>N/A</p>	
<p>Identify all potential routes of infection:</p>	
<p>N/A</p>	
<p>Name the radiochemical or radiochemicals to be used:</p>	
<p>N/A</p>	
<p>List laboratories where work with pathogens or radioisotopes will be carried out:</p>	
<p>N/A</p>	
<p>List disinfectants to be used, and describe arrangements for disposal of used material:</p>	
<p>N/A</p>	
<p>Will or might Health Surveillance be required for you or any staff working with you? If 'Yes', please give details.</p>	<p><input type="checkbox"/> Yes <input type="checkbox"/> No</p>
<p>N/A</p>	

<p>4.5 PRECAUTIONS AGAINST HAZARDS (to be completed if any potentially hazardous activities are likely to be carried out during the project. Refer to Project Handbook and School safety documentation for further information. Faculty Safety Supervisor's approval may be further requested where felt appropriate by project Supervisor.)</p>	
<p>Indicate any procedures, activities or aspects of the proposed project which may entail hazards (including work with hazardous substances as per Section 4.4, or anything else relevant). Please set distinct hazards out separately, in a numbered list.</p>	
<p>N/A</p>	
<p>Indicate the precautions you will take to prevent or mitigate such potential hazards. Please number these to refer to the specific hazards identified in the preceding question.</p>	
<p>N/A</p>	

4.6 SPECIAL REQUIREMENTS (to be completed if the project involves any special requirements, e.g. disability-related issues, or other concerns that need to be taken into account for either you as a student, study participants or colleagues)

What special requirements or concerns need to be taken into account?	
N/A	
Do these need to be considered in planning arrangements? If 'Yes', please give details.	<input type="checkbox"/> Yes <input type="checkbox"/> No
N/A	
Do these impact on supervision arrangements? If 'Yes', please give details.	<input type="checkbox"/> Yes <input type="checkbox"/> No
N/A	
Does the project location need to be considered in relation to these? If 'Yes', please give details.	<input type="checkbox"/> Yes <input type="checkbox"/> No
N/A	
Do arrangements for access to specialist medical treatment need to be considered? If 'Yes', please give details.	<input type="checkbox"/> Yes <input type="checkbox"/> No
N/A	

SECTION 5 – APPLICATION FOR ETHICS APPROVAL

All students should **answer all questions in sub-sections 5.1 and 5.2. Answers to 5.1 will make clear whether approval by the LSHTM Ethics Committee is necessary, and which later sub-sections you may need to complete. Section 5.2 covers any external approvals required.*

Before completing this part of the form, please read the Ethics Approval Policy & Procedure plus guidance notes at <http://intra.lshtm.ac.uk/reference/forms/ethicsstuds.html> . This describes what to do next if formal LSHTM ethics approval is required. NB that supervisor approval must be obtained **before an application is submitted to the Ethics Committee.*

5.1 SCOPE OF STUDY (to be completed by all students)

Which of the following applies to your project? (please tick one option only)

**Note – the term 'human data' includes any documentary data, datasets or biological samples.*

Project does not involve any human subjects or any human data. [If so, formal LSHTM ethics approval is not required and you do not need to complete Sections 5.3 or 5.4]

Project involves human data, but all this human data is fully in the public domain. [If so, formal LSHTM ethics approval is not required and you do not need to complete Sections 5.3 or 5.4]

**Public domain human data must be: available to any member of the public without special permission; to which access is not restricted in any way; and which does not*

enable the identification of living people, either directly or by linking to other data.

Project involves some non-public-domain human data, all of which was previously collected in another study or studies. [If so, formal LSHTM ethics approval is required and Section 5.3 must be completed]

Project involves some additional collection of data, further to an ongoing or previously completed study or studies. [If so, formal LSHTM ethics approval is required and Section 5.4 must be completed]

Project is a completely new study which will involve human subjects or human data. [If so, formal LSHTM ethics approval is required and Section 5.4 must be completed]

5.2 LOCAL ETHICS APPROVAL OR RESEARCH GOVERNANCE APPROVAL (to be completed by all students)

** As well as approval from the LSHTM Ethics Committee, projects may require specific approval from other involved or responsible bodies. For example, in the UK you may need specific authorisation to work in an NHS facility, or to work with vulnerable groups such as patients or children. Outside the UK a wide range of requirements may apply e.g. from local or national Ethics Committees, government departments etc. **Students must investigate all potential local approval required for your project work. Failure to check or gain any necessary external approval may invalidate LSHTM approval.***

Is local approval required for the work being done (whether this approval is still to be obtained, or has already been granted)?

**This should include any forms of ethics approval, research governance approval or other specific permissions that may apply.*

Yes No

If 'Yes', give details of local approval to be obtained (this must be in place before commencing fieldwork) or which has already been granted.

**Please name all bodies whose approval is required, or indicate where work is expected to take place using permissions already granted for a 'parent' project. Where approval has already been granted, quote approval reference numbers and if possible give web links to documents.*

If 'No', explain why formal local approval is not required, and describe any less formal permissions, invitations or support you are being given for this work.

**If you will be working away from LSHTM with human subjects or human data, but cannot identify a local Ethics Committee or believe that no formal approval is required, then please give details and explain what you have done to check this. In such cases, if you do not have formal approval you should always demonstrate appropriate local support, such as correspondence with local government officials or an involved Non-Governmental Organisation.*

The data I will be analysing come from a completed randomized controlled trial for which local ethics approval was received for each participating centre.

For data to be used or collected in the project, will any specific data rights permissions be required or usage limitations apply?

Yes No

The data belong to Pfizer and will be returned at the end of the project and any copies deleted.

5.3 PROJECTS USING ONLY PREVIOUSLY-COLLECTED HUMAN DATA (to be completed if project involves non-public-domain human data, datasets or biological samples previously collected in another study or studies; if collecting any new data,

complete Section 5.4 instead)

*Further guidance is given at <http://intra.lshtm.ac.uk/reference/forms/ethicsstuds.html>

Summary of purpose and methods of the original study or studies: (max 100 words)

The purpose of the original study was to evaluate the effects of eleprenone in patients with chronic systolic heart failure and mild symptoms. 2737 patients with New York Association class II heart failure were randomised to receive placebo or eleprenone in addition to recommended therapy, in a double-blind trial.

Give details of all approvals under which the original study or studies took place:

**Please quote names of Ethics Committees and approval reference numbers (required if previous approval was from LSHTM); if possible give web link to original study application.*

The data come from a multi-centre, international RCT conducted by a leading pharmaceutical company. The trial was approved by each centre's ethics committee. Due to the large number of centres (278 centres in 29 countries) involved we cannot provide full ethics approval details. The clinicalTrials.gov number is NCT00232180. All patients provided written informed consent.

Proposed study: Ensure that the project outline given in Section 3.1 states the purpose, methods and procedures of the new work to be done in your project, and describes how this builds on the previous study or studies (for which participants will already have been recruited, data or samples collected, and procedures performed). Do not reproduce here.

Will your analyses be for purposes not covered by the original application detailed above? If 'Yes', indicate how you will obtain (i) permission to use the data from the principal investigator responsible for each original study; and (ii) retrospective consent, where appropriate, from the participants in each original study.

Yes No

Does the project involve analysis of documentary information and/or data already collected from or about human subjects? If 'Yes', specify analyses briefly.

Yes No

Baseline data including, but not limited to, patient demographics, anthropometrics, vital signs and prior medical history will be used in Cox regression models (and other statistical models) to identify predictors of cardiovascular mortality and hospitalisation for heart failure. These will then be used to develop a risk score for prediction purposes.

Does the project involve laboratory analysis of human biological samples already collected, or new or additional analysis of stored samples? If 'Yes', specify the laboratory analyses or tests to be performed.

Yes No

Specify how confidentiality will be maintained. When small numbers are involved, indicate how possible identification of individuals will be avoided.

The data have been fully anonymised. It would be impossible to identify any individual patient.

5.4 PROJECTS COLLECTING ANY NEW HUMAN DATA (to be completed if project involves collection of human data, datasets or human biological samples – either as a completely new study, or collecting additional data further to an ongoing or previously

completed study)

*Further guidance is given at <http://intra.lshtm.ac.uk/reference/forms/ethicsstuds.html>

Proposed study: Ensure that the project outline given in Section 3.1 contains sufficient detail (inc. purpose, methods, procedures for both new data collection and any work building on previous studies), so as to allow the Ethics Committee to make an informed decision without reference to other documents. Do not reproduce here.

Is your project a randomised trial?

Yes
 No

Will any human biological samples be collected? If 'Yes', specify details.

Yes
 No

Will any human biological material be stored at LSHTM for more than 24 hours? If 'Yes', specify which samples and how they will be stored.

*Further guidance is given at

http://intra.lshtm.ac.uk/safety/safety_manual_appendix_3_human_tissue_act.pdf

Yes
 No

Specify the number - with scientific justification for sample size - age, gender, source and method of recruiting subjects for the study.

State the location and likely duration of new or additional human data collection, and the extent to which this will be carried out by you alone, or in collaboration with others, or by others.

State the potential distress, discomfort or hazards, and their likelihood, to which research subjects may be exposed (these may include physical, biological and/or psychological hazards). What precautions are being taken to control and modify these hazards?

Specify how confidentiality will be maintained. When small numbers are involved, indicate how possible identification of individuals will be avoided.

State the manner in which consent will be obtained from subjects and supply copies of the information sheet and consent form.

- Written consent is normally required. Where not possible, explain why and confirm that a record of those giving verbal consent will be kept.
- Where appropriate, please state if and how the information and consent form will be translated into local language(s).

As well as collecting new data, will your project also make use of any human data or biological samples collected in a previous study or studies?

If 'Yes', summarise the purpose and methods of the original study or studies - for which participants will already have been recruited, data or samples collected, and

Yes
 No

procedures performed. (max 100 words)	