Personalisation of Heart Failure Care using Clinical Trial Data

Dr Carly Adamson
BMSc, MBChB, MRCP (UK)

Submitted for the fulfilment of the requirements for the degree of Doctor of Philosophy

BHF Glasgow Cardiovascular Research Centre
College of Medical, Veterinary and Life Sciences
Faculty of Medicine
University of Glasgow

April 2023

Abstract

Heart failure is a common, debilitating and life limiting disease, resulting in a large burden for both the individual patient and healthcare provision. Therefore, optimisation of treatments for these patients is of prime importance. Heart failure with reduced ejection fraction has a large evidence base for effective treatments, and more recently effective treatments have started to be identified for those with preserved ejection fraction. The effectiveness of these treatments is calculated at a population level, and there is a great deal of interest to try and identify if different patients may benefit more from certain treatments. In addition, we wish to understand more about different phenotypes in heart failure, to help understand what the patient might expect for the trajectory of their illness and potentially develop targeted treatments. To explore these issues further, this thesis presents several approaches using heart failure clinical trial data to try and further understand the patient journey and explore how treatment may be delivered in a more personalised fashion.

The first analyses look at the patterns of heart failure hospitalisations, including the timing of admissions, and the relationship with different modes of death. This was examined in both heart failure with preserved and reduced ejection fraction. The accepted trajectory of recurrent admissions falling closer together over time was confirmed, and admissions closer together were linked to a higher risk of cardiovascular death, particularly due to progressive pump failure. Sudden death did appear to be truly sudden and not strongly linked to hospitalisations.

The next approach was to perform latent class analysis to try and identify clusters of patients, or phenotypes, within heart failure with preserved and reduced ejection fraction separately using a data driven method. Phenotypes were identified with consistency across different data and using different approaches. These phenotypes were clinically recognisable. Identifying phenotypes in this way may be a route to looking for differential responses to treatments.

Lastly, supervised machine learning methods were used to predict outcomes in patients with heart failure and reduced ejection fraction. These techniques

provide more analytical flexibility, but did not show performance benefit compared with prognostic models based on survival analysis methods. Overall, the predictive abilities were modest.

In conclusion, several avenues were explored to help understand the patient journey in heart failure, aiming to give more detail about the expected patient trajectory and exploring methods to examine for differential treatment responses in phenotypes of patients in heart failure.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Professor Pardeep Jhund, firstly for the opportunity to complete a period of research in the Heart Failure group and all your support and guidance throughout. Thank you in your backing and advice in my development as a researcher over the past few years.

Thank you to my co-supervisor Professor John McMurray for the excellent opportunities to complete interesting and varied work during my time in the research group. I was always inspired by your comments, discussions and feedback on my findings which has helped me develop as a flexible critical thinker.

I am extremely grateful to the British Heart Foundation for funding my research post at the University of Glasgow, and for the support and guidance of everyone in the CORE team.

Thank you to all my colleagues in the Heart Failure research group for your help but most importantly your friendship and making the office an enjoyable and stimulating place to work!

Lastly, thanks to my family. To my husband Jonny for helping me maintain my motivation and often feeding me, as well as helping make the working from home stage of the Covid pandemic more bearable. To my mum and dad, who have inspired and supported me throughout my career. And to my wider family and friends (and pets), thank you for always being there for me.

Author's Declaration

The work described in this thesis was performed during my employment as a Clinical Research Fellow at the Institute of Cardiovascular and Medical Sciences at the BHF Cardiovascular Research Centre, Glasgow supported by the British Heart Foundation Centre of Research Excellence. I was supervised by Professor Pardeep Jhund and Professor John McMurray.

I confirm that this thesis has been composed by me solely and that it has not been submitted for any other degree at the University of Glasgow or any other institution.

Dr Carly Adamson

April 2023

Publications during course of study but not included in this work

1. Adamson C, Kondo T, Jhund PS, de Boer RA, Cabrera Honorio JW, Claggett B, Desai AS, Alcocer Gamba MA, al Habeeb W, Hernandez AF, Inzucchi SE, Kosiborod MN, Lam CSP, Langkilde AM, Lindholm D, Bachus E, Litwin SE, Martinez F, Petersson M, Shah SJ, Vaduganathan M, Nguyen Vinh P, Wilderäng U, Solomon SD, McMurray JJV. Dapagliflozin for heart failure according to body mass index: the DELIVER trial. Eur Heart J 2022;43:4406–4417.

2. Adamson C, Cowan LM, de Boer RA, Diez M, Drożdż J, Dukát A, Inzucchi SE, Køber L, Kosiborod MN, Ljungman CEA, Martinez FA, Ponikowski P, Sabatine MS, Lindholm D, Bengtsson O, Boulton DW, Greasley PJ, Langkilde AM, Sjöstrand M, Solomon SD, McMurray JJV, Jhund PS. Liver tests and outcomes in heart failure with reduced ejection fraction: findings from DAPA-HF. Eur J Heart Fail 2022;24:1856-1868.

3. Adamson C, Welsh P, Docherty KF, de Boer RA, Diez M, Drożdż J, Dukát A, Inzucchi SE, Køber L, Kosiborod MN, Ljungman CEA, Martinez FA, Ponikowski P, Sabatine MS, Morrow DA, Lindholm D, Hammarstedt A, Boulton DW, Greasley PJ, Langkilde AM, Solomon SD, Sattar N, McMurray JJV, Jhund PS. IGFBP-7 and Outcomes in Heart Failure With Reduced Ejection Fraction: Findings From DAPA-HF. JACC Heart Fail 2022; 20:S2213-1779(22)00571-6

4. Adamson C, Docherty KF, Heerspink HJL, de Boer RA, Damman K, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Petrie MC, Ponikowski P, Sabatine MS, Schou M, Solomon SD, Verma S, Bengtsson O, Langkilde AM, Sjöstrand M, Vaduganathan M, Jhund PS, McMurray JJV. Initial Decline ('dip') in Estimated Glomerular Filtration Rate Following Initiation of Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction: Insights from DAPA-HF. Circulation 2022;146:438-449

5. Adamson C, Jhund PS, Docherty KF, Bělohlávek J, Chiang CE, Diez M, Drożdż J, Dukát A, Howlett J, Ljungman CEA, Petrie MC, Schou M,

Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Ponikowski P, Sabatine MS, Solomon SD, Bengtsson O, Langkilde AM, Lindholm D, Sjöstrand M, McMurray JJV. Efficacy of dapagliflozin in heart failure with reduced ejection fraction according to body mass index. Eur J Heart Fail 2021;23:1662–1672.

Presentations related to this work

1. Adamson C, Abraham W, Desai A, Dickstein K, Kober L, McMurray JJV, Packer M, Rouleau J, Solomon S, Zile M, Jhund PS. Patterns Of Recurrent Heart Failure Hospitalisations In Relation To Cardiovascular Death In Heart Failure With Reduced Ejection Fraction. *J Card Fail* 2022;28:S112. Oral presentation at Heart Failure Society of America Annual Scientific Meeting 2021. Denver, USA (virtual attendance), 2021

Presentations during study not related to this work

2. Adamson C, Welsh P, Morrow DA, Docherty KF, Hammarstedt A, Inzucchi SE, Kober L, Kosiborod MN, Martinez FA, Ponikowski P, Sabatine MS, Solomon SD, Sattar N, Jhund PS, McMurray JJV. Outcomes related to IGFBP-7 in patients with heart failure and reduced ejection fraction and effects of dapagliflozin: findings from DAPA-HF. European Heart Journal (2022) 43 (Supplement), 913. Oral presentation at European Society of Cardiology (ESC) Congress 2022. Barcelona, Spain, August 2022.

3. Adamson C, Butt JH, Rouleau J, Abraham W, Desai A, Dickstein K, Kober L, Lefkowitz MP, Packer M, Petrie MC, Swedberg K, Solomon SD, Zile M, Jhund PS, McMurray JJV. Alkaline phosphatase and bilirubin combined are a powerful predictor of outcome in patients with heart failure and reduced ejection fraction: an analysis of the ATMOSPHERE and PARADIGM-HF trials. European Heart Journal (2022) 43 (Supplement), 871. Poster presentation at European Society of Cardiology (ESC) Congress 2022. Barcelona, Spain, August 2022.

Definitions/Abbreviations

ACEI	Angiotensin converting enzyme inhibitors
AF	Atrial fibrillation
AIC	Akaike information criteria
ANN	Artificial neural networks
ANOVA	Analysis of variance
ARNI	Angiotensin receptor-neprilysin inhibitor
ASIAN-HF	Asian Sudden Cardiac Death in Heart Failure
ATMOSPHERE	Aliskiren Trial to Minimize Outcomes in Patients with Heart Failure Trial
BEST	ß-blocker Evaluation of Survival Trial
BIC	Bayesian information criteria
BMI	Body mass index
BNP	B-type natriuretic peptide
CHARM	Candesartan in Heart failure: Assessment of Reduction in Mortality and morbidity
CHECK-HF	Chronisch Hartfalen ESC-richtlijn Cardiologische praktijk Kwaliteitsproject HartFalen
COPD	Chronic obstructive pulmonary disease

CRT	Cardiac resynchronisation therapy
CV	Cardiovascular
ECG	Electrocardiogram
eGFR	Estimated glomerular filtration rate
EMPHASIS-HF	Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure
EPHESUS	Eplerenone Post–Acute Myocardial Infarction Heart Failure Efficacy and Survival Study
GSEM	Generalised structural equation modelling
GWTG	Get With The Guidelines
HF	Heart failure
HF-ACTION	Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training
HFH	Heart failure hospitalisation
HFMREF	Heart failure with mildly reduced ejection fraction
HFPEF	Heart failure with preserved ejection fraction
HFREF	Heart failure with reduced ejection fraction
HR	Hazard ratio
I-PRESERVE	Irbesartan in Heart Failure with Preserved Ejection Fraction Study

ICD	Implantable cardioverter defibrillator
IV	Intravenous
KCCQ	Kansas City Cardiomyopathy Questionnaire
KCCQ CSS	KCCQ Clinical Summary Score
LBBB	Left bundle branch block
LCA	Latent class analysis
LVEF	Left ventricular ejection fraction
MAGGIC	Meta-Analysis Global Group in Chronic Heart Failure
MRA	Mineralocorticoid receptor antagonist
NSAIDS	Non steroidal anti inflammatory drugs
NT-proBNP	N-terminal pro-B type natriuretic peptide
PARADIGM-HF	Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure
PARAGON-HF	Prospective Comparison of ARNI with ARB Global Outcomes in HF with Preserved Ejection Fraction
PREDICT-HF	PARADIGM Risk of Events and Death in the Contemporary Treatment of Heart Failure
RAAS	Renin-angiotensin-aldosterone system
ROC AUC	Area under the receiver operating characteristic curve
SGLT2i	Sodium-glucose co-transporter 2 inhibitor

SwedeHF	Swedish Heart Failure registry
TOPCAT	Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist

Chapter 1 Literature review

1.1 Introduction to Heart Failure

1.1.1 Definition and epidemiology

Heart failure is a syndrome caused by failure of the heart to pump blood efficiently enough to meet the peripheral tissues metabolic requirements causing typical signs and symptoms including breathlessness, ankle swelling, fatigue, elevated jugular venous pressure and pulmonary crepitations.[1] This can be caused by any abnormality in structure or function of the heart, with left ventricular systolic function being an important distinction with implications for treatment options. The diagnosis of heart failure is only that of the syndrome described, and evaluation for the underlying aetiology is required to understand the pathology involved and again the associated implications for treatment options. A core investigation for making a diagnosis of heart failure is the finding of elevated natriuretic peptides, which is found universally through different subgroups and aetiologies of heart failure.

Sub-classification of heart failure is primarily dependent on the left ventricular ejection fraction, as determined by echocardiography or cardiac magnetic resonance imaging, which is associated with differing aetiologies and pathologies underlying the syndrome of heart failure. Reduced ejection fraction (HFREF) is defined as an ejection fraction of ≤40%, and the vast weight of evidence for heart failure treatment is in this population. The other main subgroup is heart failure with preserved ejection fraction (HFPEF), usually defined as an ejection fraction ≥50%. These patients have other cardiac abnormalities, such as dilated atria, diastolic dysfunction and raised filling pressures, as well as elevation in natriuretic peptides. More recently, the subgroup of mildly reduced ejection fraction (HFMREF) with an ejection fraction between 41 and 49% have been defined, with some evidence of similarities in response to treatments as those with reduced ejection fraction.[2]

Heart failure is common and, despite stable or declining incidence rates, the absolute number of patients with heart failure and overall prevalence is increasing, likely reflecting an aging population and improved survival with

modern treatments.[3,4] It is thought the prevalence of heart failure is 1-2% of adults, although as these estimates only include diagnosed cases the likely true total is higher.[5–8] The care of patients with heart failure is costly, and accounts for 1 million bed days per year in the NHS and 2% of the NHS budget.[9] Heart failure is associated with a low quality of life and high symptom burden, and high mortality rates across the world.[4] Therefore any increased understanding of how our patients might be expected to behave in future in terms of hospitalisations can be helpful in service planning and provision.

1.1.2 Aetiology

Common causes of heart failure in the developed world include ischaemic heart disease, hypertension, arrhythmia, idiopathic, toxins (including alcohol and chemotherapy) and genetic causes, with less common causes including valve disease, infections (including viral infections and Chagas disease), congenital heart disease, metabolic causes such as haemochromatosis, infiltrative disease such as amyloid and pericardial diseases. In relation to this research, data regarding aetiology is often gathered in a less granular way, for example ischaemic or non-ischaemic aetiology, and data on some relevant comorbidities will also be included. However, some aetiologies are underrepresented in clinical trial data due to exclusion criteria and the countries that enrolled patients in the trials. This is important when we consider phenotyping of patients, as phenotypes created by any machine learning method can only include the data included in the model.

1.1.3 Classification of heart failure

Classification of heart failure into phenotypes has largely been based on left ventricular ejection fraction. Most trials for original treatments in heart failure enrolled patients with an ejection fraction of 40% or below and this phenotype has the largest evidence base for effective treatments, these patients are designated as heart failure with reduced ejection fraction (HFREF). Many patients with heart failure have a normal ejection fraction (50% or above) with evidence of other structural or functional abnormalities in cardiac function and are classified as heart failure with preserved ejection fraction (HFPEF). More recently, the area between these phenotypes including an ejection fraction of

41 to 49% have been classified as heart failure with mildly reduced ejection fraction (HFMREF), who behave similarly to patients with HFREF.[1] Although these cut-points are useful to understand phenotypes of patients and which treatments they might benefit from, they remain somewhat arbitrary, and it is important to consider that ejection fraction is a continuous variable. Although these are recognised as important subdivisions in heart failure, in reality there are many more facets of clinical differences that create the recognisable phenotypes that are encountered in patients with heart failure. Treatments are now being identified that are effective across the spectrum of ejection fraction, and perhaps we need more sophisticated ways to identify different phenotypes to help us understand their likely trajectory of illness or likely response to treatment.[10]

1.1.4 Diagnosis of heart failure

A diagnosis of chronic heart failure can be made when there is a combination of typical clinical signs and symptoms, elevation in natriuretic peptides and objective evidence of cardiac dysfunction. Suspicion of heart failure is raised when there are risk factors for development of the disease, for example prior myocardial infarction, hypertension, diabetes or a family history of heart failure. An abnormal electrocardiogram may be the trigger for further investigation. The next stage in the diagnostic algorithm from the European Society of Cardiology is measurement of natriuretic peptide level, either N-terminal pro-B type natriuretic peptide (NT-proBNP) or B-type natriuretic peptide (BNP).[1] The primary stimulus for secretion of these hormones is through myocyte stretch and neurohormonal activation. If natriuretic peptide levels are elevated, or not available and the clinical suspicion of heart failure is high, an echocardiograph is required to examine for evidence of cardiac dysfunction and to group patients into heart failure with preserved, mildly reduced or reduced ejection fraction.

Further investigations to elucidate aetiology include coronary angiography (invasive or by computed tomography scan), prolonged electrography monitoring for arrhythmia, more detailed imaging by magnetic resonance scan, genetic testing or on rare occasions by myocardial biopsy.

1.1.5 Management

The major goals in treatment of heart failure are reduction in mortality, prevention of recurrent hospitalisations due to worsening heart failure and improvement in clinical status, functional capacity and quality of life.[1] All medications should be up titrated to the maximum tolerated dose.

1.1.5.1 Heart failure with reduced ejection fraction

Most of the evidence for pharmacological therapies in heart failure is in patients with heart failure and reduced ejection fraction, for which there are four core treatments that all patients should be established on if tolerated. Angiotensin converting enzyme inhibitors (ACEI)/angiotensin receptor-neprilysin inhibitor (ARNI), beta blockers and mineralocorticoid receptor antagonists (MRA) act by modulating the renin-angiotensin-aldosterone system (RAAS) and sympathetic nervous system. The fourth treatment is sodium-glucose co-transporter 2 inhibitors (SGLT2i).

ACEI act on the angiotensin converting enzyme in the lungs, preventing the conversion of angiotensin 1 to angiotensin 2. Angiotensin 2 has many actions throughout the body which in a physiological condition increase blood pressure and improve perfusion in response to low renal perfusion pressure; these actions include peripheral vasoconstriction, sodium and water retention through direct action on the kidneys and by stimulating aldosterone secretion and stimulating vasopressin release. In heart failure, there is over-activation of these pathways, resulting in inappropriate systemic vasoconstriction and sodium/water retention contributing to the cycle of worsening cardiac function. ACEI were one of the first established treatments in heart failure to reduce mortality, with angiotensin receptor blockers being shown to be an appropriate alternative in those who could not tolerate ACEI due to cough.[11-13] Natriuretic peptide hormones act in opposition to the RAAS system, causing vasodilation, natriuresis and diuresis, and is degraded through the action of neprilysin. By inhibiting neprilysin as well as the angiotensin receptor, the action of the natriuretic peptides is prolonged providing additional mortality and morbidity benefits compared to ACE inhibition in isolation.[14]

Beta blockers have the largest evidence base of benefit in heart failure and act through inhibition of beta-adrenergic stimulation of the heart therefore decreasing inotropy and chronotropy.[15-17] Mineralocorticoid receptor antagonists also act on the RAAS system and provide benefit in addition to ACEI.[18,19]

SGLT2i were first developed as treatment for diabetes and act by reducing glucose reabsorption in the kidneys therefore increase the excretion of excess glucose in the urine. Cardiovascular outcome studies in diabetic patients suggested reduction in heart failure hospitalisation in diabetic patients therefore these drugs were studied in patients with heart failure with and without diabetes and found to provide morbidity and mortality benefit.[20-24]

Other pharmacological treatments include: loop diuretics, primarily to reduce symptom burden; ivabradine for patients in sinus rhythm where rate is not controlled to 70 beats per minute or less with beta blocker or in patients with contraindication to beta blocker therapy; soluble guanylate cyclase receptor stimulator (vericiguat) in patients who have had worsening HF despite optimised therapies to reduce risk of hospitalisation; hydralazine and isosorbide dinitrate in patients self-identified as black with ejection fraction 35% or below on optimised medical therapy or if ACEI, ARB or ARNI are not tolerated; and digoxin in patients who remain symptomatic despite optimised medical therapy and remain in sinus rhythm to reduce hospitalisation.

Non pharmacological therapies include device therapies such as cardiac resynchronisation therapy and implanted cardioverter-defibrillator. Many patients with heart failure die suddenly with modes of death including brady- and tachy-arrhythmia as well as acute vascular events, and implantation of a defibrillation device can allow treatment of these arrhythmias. If a patient survives a ventricular tachyarrhythmia causing haemodynamic instability, a device can be inserted as a secondary prevention intervention. Otherwise, primary prevention in patients who have never had an arrhythmic event is considered when the LVEF remains low despite optimised pharmacological therapy with ongoing symptoms, and the strength of evidence depends on the aetiology of heart failure being ischaemic or non-ischaemic. Cardiac resynchronisation therapy can be considered in patients with ongoing symptoms,

reduced ejection fraction and prolonged QRS duration on electrocardiogram testing despite optimised pharmacological therapies.[1]

1.1.5.2 Heart failure with mildly reduced ejection fraction

There are not specific trials of patients in this range of ejection fraction therefore there are not strong recommendations for pharmacological therapies in these patients, but the same drug treatments should be considered as for HFREF based on some subgroup analyses and likely clinical comorbidity and indications for treatment. Device therapy is not recommended in guidelines.

1.1.5.3 Heart failure with preserved ejection fraction

Many treatments found to be effective in heart failure with reduced ejection fraction have been trialled in patients with HFPEF with neutral results (including ACEI, ARB, MRA and ARNI).[25–29] There is evidence from the TOPCAT trial including patients enrolled in the Americas that MRAs may reduce CV death and HF hospitalisation in HFPEF (perhaps reflecting patients with 'true' HFPEF) and pooled analysis of the PARADIGM-HF and PARAGON-HF studies suggested ARNI are beneficial in higher ejection fraction.[30] It should be acknowledged that many of these patients will have a separate indication for these treatments due to co-existing comorbidity.

More recently, SGLT2i have been studied in patients with heart failure with preserved and mildly reduced ejection fraction and in two drugs of the class have been found to reduce the incidence of the primary endpoint of worsening heart failure events and CV death.[31,32] These findings are not yet reflected in guidelines.

HF is a heterogenous condition. Treatments that have proven efficacy are assessed at a population level, and there may be underlying patterns of differential responses between phenotype groups that are not fully explored. This is particularly of interest in HFPEF where effective treatments have remained more elusive.

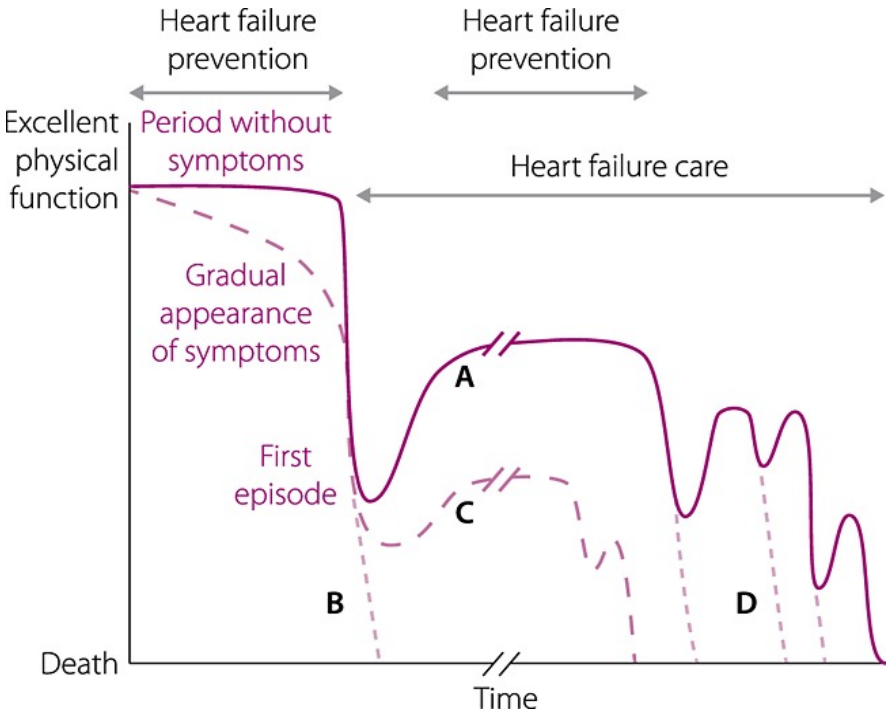
1.1.6 Prognosis

With improving treatments in heart failure, prognosis has improved over time but remains poor, with a 5-year age adjusted mortality in a population-based cohort study between 1996 and 2000 of 52%.[33] Improvements in survival have been primarily seen in patients with HFREF rather than HFPEF, likely reflecting the improved therapies available over the last few decades.[34] Another difference between HFREF and HFPEF is the different proportion of patients with cardiovascular compared with non-cardiovascular deaths; more individuals with HFREF died of cardiovascular rather than non-cardiovascular causes whereas the opposite was true of patients with HFPEF.[35]

1.1.6.1 The patient journey in heart failure

The typical trajectory of functional status in a patient diagnosed with heart failure is often described graphically with several deteriorations, incomplete recovery of function and accelerating frequency of deterioration until death with sudden deaths occurring at any point along this trajectory (Figure 1-1).[36] These deteriorations often result in hospital admissions, which can act as warning flags to the treating physician for risk of further deterioration and death.

Figure 1-1. The heart failure patient journey
Reproduced from Cowie et al., 2014 with permission. [36] Typical progression of acute heart failure, showing a range of clinical courses. A, good recovery after first episode followed by stable period of variable length; B, first episode not survived; C, poor recovery after first episode followed by deterioration; D, ongoing deterioration with intermittent crises and unpredictable death point.



It would be of clear utility to the treating physician and patient to have a good understanding of which patients are at higher risk of recurrent hospitalisation or at higher risk of death following an admission, however, determining which patients are at particularly high risk of readmission remains difficult.[37-39] In addition, it is recognised that the greater burden of other comorbidity in patients with heart failure and preserved ejection fraction will likely result in high rates of hospitalisation for non-cardiovascular reasons therefore an understanding of how much heart failure hospitalisations are responsible for morbidity for these patients is important.

1.2 Machine learning in heart failure populations

Machine learning is an umbrella term for different analytical techniques that aim to utilise data to gain insight into a quantity of interest and involves aspects of statistics, computer science and artificial intelligence. Machine learning differs from some statistical analytical techniques with increased flexibility and less requirement for prior assumptions of the data and is well suited to analysis of

big data. These attributes make machine learning an area of research interest given the increasing amounts of data available in many areas, including healthcare. At a very high level, machine learning can be subdivided into supervised and unsupervised techniques.

In supervised machine learning, data has a known 'label'. The label is the ground truth answer to the question the machine learning method is going to try and address, and will be the output of the machine learning method. For example, a label might be whether the patient is alive at 1 year after diagnosis, or could be a label of a diagnosis assigned to a chest x-ray image. The aim of the machine learning technique is to map the inputs (i.e. variables or factors) to the output. Inputs can be a variety of different types of data, including spreadsheets of values for different laboratory results, demographic data about the patient, binary variables for presence/absence of a comorbidity or a data representation of a chest x-ray image. Performance of the model can be assessed by how well the predicted label using the machine learning model matches the ground truth, or pre-existing label.

The aim of unsupervised machine learning is to find unknown patterns or clusters within data. For these analyses, data is not labelled with any known outcome or grouping variable, and the aim of the analysis is to first identify the number of groups and then to describe them.

1.2.1 Unsupervised machine learning in heart failure

In the field of heart failure, unsupervised machine learning of analysis is most commonly used to identify phenotypically similar groups of patients using data collected during clinical trials or extracted from routine electronic health records. This emulates the way clinicians recognise clinical symptoms, characteristics and diagnoses that commonly appear together, and learn how different types of patients might behave. However, each individual is limited by their own experience and exposures; machine learning may provide a tool to describe a group of patients or indeed identify new, or less apparent, phenotype groups . This could be particularly useful for more junior clinicians and non-specialists who do not have the same wealth of individual experience encountered in clinical work as an aid to understanding the patient's prognosis

and behaviour. This could ultimately improve personalisation of heart failure care if physicians can have a better understanding of the likely patient course and prognosis depending on the phenotype subgroup.

As well as possible benefits in our ability to describe the phenotype of the patient, and explore their likely prognosis, another area of interest is to look for difference in treatment response in different phenotype groups. Finding effective treatments for patients with heart failure and preserved ejection fraction continues to be elusive, this may be due to the highly heterogeneous population. Perhaps by effectively identifying phenotypes of patients, therapies may be targeted to patients in whom it is believed the treatments may be more effective – again another avenue to increasing personalisation of care.

In this thesis, phenotype groups are identified using latent class analysis as an unsupervised machine learning technique to define subgroups in heart failure populations. However, an overview of prior research in different clustering techniques will be reviewed for both HFPEF and HFREF prior to discussing latent class analysis specifically.

1.2.1.1 Machine learning identified subgroups in heart failure with reduced ejection fraction

The most commonly utilised techniques for clustering include K-means clustering and hierarchical clustering. Karwart et al.[40] utilised combined data from trials of beta blocker therapy in HFREF resulting in a large dataset of 15659 patients, with which they performed cluster analysis for patients with atrial fibrillation and sinus rhythm separately. Data was pre-processed for dimensionality reduction, meaning the number of variables were reduced by creating new inputs combining information from several variables while minimising loss of information. In this study, variational autoencoders were used for dimensionality reduction as they can be used for mixed data types including factor and numeric variables. Other more simple examples of dimensionality reduction techniques include principle component analysis. In this very large population, five phenotypes were identified in those with atrial fibrillation and six for those in sinus rhythm, with some signal of differing responses to beta blocker therapy in subgroups.

Most studies have involved smaller populations, with several aiming to find phenotypes of patients with differing response to cardiac resynchronisation therapy using clinical trial data and data collected in routine practice, again with some signal of differing response to treatment.[41,42] Others have described phenotype groups using finite mixture model based clustering, identifying three phenotype subgroups with differing risk of adverse outcome.[43]

Cluster analysis has been carried out in patients enrolled in the HF-ACTION trial (Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training) using a large number of variables.[44] Four clusters were identified, with differing responses to exercise therapy. Two groups had predominant ischaemic cardiomyopathy, with the main difference being degree of angina. Interestingly the left ventricular ejection fraction did not vary significantly between groups, this being a commonly used measure to describe patients with heart failure.

1.2.1.2 Machine learning identified subgroups in heart failure with preserved ejection fraction

The greater level of heterogeneity in patients with HFPEF has resulted in a larger pool of research using clustering techniques to help understand different phenotypes in this population. Inputs to clustering models range include data from echocardiography, clinical examination, laboratory results and biomarkers, with range of phenotype subgroups varying from three to six.[45-54] Although each clustering study uses different types of input variables, there are some subgroup descriptions that appear quite consistently, including: obese patients with high levels of comorbidity such as hypertension and diabetes and often younger age; older patients with atrial fibrillation; patients with high burden of chronic obstructive pulmonary disease (COPD). [46,47,52,53] Use of biomarkers for clustering suggested some groups of high levels of inflammation, and use of echocardiographic measurements identified groups of patients with specific abnormalities such as stiff right ventricle or left ventricular relaxation abnormality.[50,51,54]

1.2.1.3 Machine learning identified subgroups in heart failure with mixed or unspecified ejection fraction

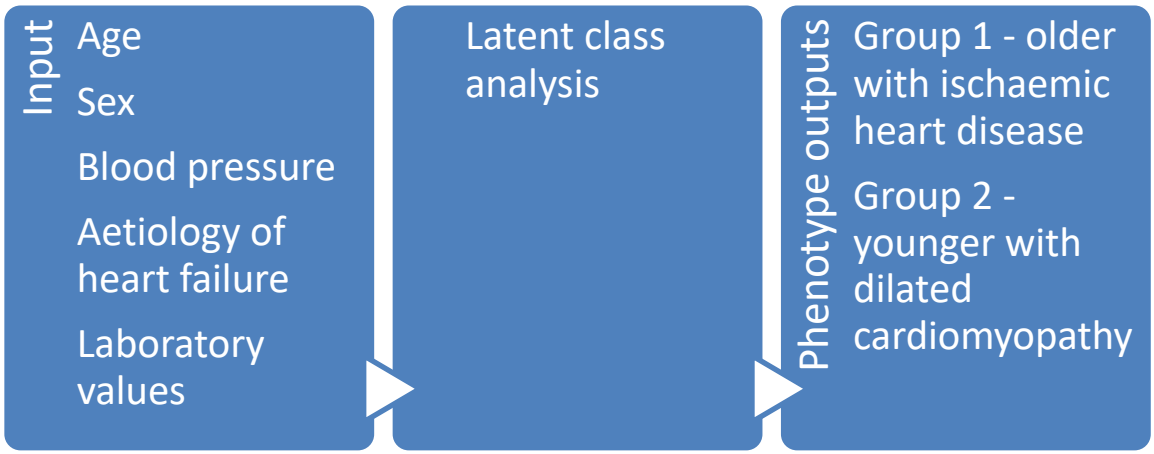
Given left ventricular ejection fraction is a measure of left ventricular contractile function, and division into HFPEF and HFREF could be seen as an arbitrary split, it might be expected that phenotypes would be found across the spectrum of ejection fraction in heart failure. Gevaert et al. [55] explored this using hierarchical clustering in patients hospitalised for heart failure, identifying 6 subgroups with key defining features including: atrial fibrillation; coronary artery disease; COPD; obstructive sleep apnoea. This was in a relatively small group of patients, with no confirmation of similar phenotypes in other populations.

Others have used supervised machine learning techniques to distinguish between HFPEF and HFREF i.e., to phenotype broadly into these two groups using baseline data, however more advanced machine learning techniques such as random forests, boosted trees or support vector machine did not provide significant benefit over logistic regression models.[56]

1.2.2 Latent class analysis

One unsupervised machine learning method that can be utilised with the aim of identifying phenotypic subgroups is latent class analysis. Latent class analysis is an approach aiming to use multiple measured observations, hypothesised to be an expression of the underlying patient subgroup, to work backward to identify and describe underlying phenotypic subgroups (Figure 1-2).[57]

Figure 1-2. Illustration of the process of latent class analysis
Simplified illustration of a latent class analysis. Input variables are hypothesised to express the underlying phenotype group. Latent class analysis aims to best describe the unmeasured phenotype group using the measured input variables. Each individual data point can then be allocated to the best fitting phenotype group.



1.2.2.1 Latent class analysis in heart failure with reduced ejection fraction

Latent class analysis has been utilised in evaluation of the EMPHASIS-HF trial (Eplerenone in Mild Patients Hospitalisation and Survival Study in Heart Failure) with validation in the EPHESUS trial (Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study).[58] Subgroups included: patients with high prevalence of hypertension and diabetes; patients with low weight, anaemia and lower potassium; patients with low weight, anaemia, poor renal function, previous revascularisation and higher potassium; and male patients with higher potassium. The second and third subgroup were observed to have a poorer response to eplerenone treatment.

Retrospective analysis of patients with non-ischaemic aetiology of HFREF in the ß-blocker Evaluation of Survival Trial (BEST) using latent class identified groups with differing outcomes and response to treatment.[59] Two separate analyses were conducted – the first using variables thought to be related to heart failure pathogenesis and the other to markers of heart failure severity and progression. The first analysis generated better differentiated groups, summarised as: non-Caucasian, males, hypertension, atrial fibrillation; middle age, female, anaemia, obesity, diabetes, hypertension and hyperlipidaemia; middle age, female, Caucasian, hyperlipidaemia, anaemia, left bundle branch block; younger patients, non-Caucasian, obesity, anaemia, less risk factors for cardiac disease; older, Caucasian, atrial fibrillation, mitral and aortic valve disease.

1.2.2.2 Latent class analysis in heart failure with preserved ejection fraction

Latent class analysis has been used to help understand different phenotypes in heart failure with preserved ejection fraction, making use of both registry and clinical trial data.[60-62] Some phenotypes appear consistently through different populations and analytical techniques, while others appear more specific to the population studied.

Latent class analysis has been utilised to explore phenotype groups in the Swedish Heart Failure Registry, with validation of results using data from patients in the CHECK-HF (Chronisch Hartfalen ESC-richtlijn Cardiologische praktijk Kwaliteitsproject HartFalen) registry by applying the clustering model and comparing the sizes of clusters and the median probabilities of group membership.[60] Subgroups included: young patients with low comorbidity; patients with atrial fibrillation and hypertension; older patients with atrial fibrillation and cardiovascular comorbidity; patients with hypertension and diabetes; and patients with poor renal function, hypertension and ischaemic heart disease. There was variation in prognosis between these subgroups in age and sex adjusted analysis, with those in cluster 3 and 5 having worst outcomes, as well as variation in treatments for heart failure.

Another latent class analysis in heart failure with preserved ejection fraction was carried out in the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist (TOPCAT) trial, which randomised patients with heart failure and preserved ejection fraction to spironolactone or placebo. [61] Three phenotypes were described: younger patients with low levels of comorbidity; older patients with atrial fibrillation; and multimorbidity patients with diabetes, obesity and renal impairment. There was variation in prognosis and response to spironolactone, with greater benefit seen with spironolactone in group 3.

Kao et al. used data from patients enrolled in the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE) and Candesartan in Heart failure: Assessment of Reduction in Mortality and morbidity (CHARM)-Preserved trials to perform latent class analysis.[62] Development of the latent classes was in the I-PRESERVE trial and identified six subgroups: younger, male patients with

higher levels of alcohol intake; younger female patients with more anaemia; multimorbid patients with obesity, diabetes and hyperlipidaemia; a class made mainly of female patients with middling levels of comorbidity; males with atrial fibrillation and coronary artery disease; and older patients with low BMI, high levels of atrial fibrillation, valvular disease, renal dysfunction and anaemia. Differing prognosis was found between latent class subgroups.

Two analyses by another research group have looked at latent class analysis of patients with heart failure and preserved ejection fraction in the context of hospitalisation.[63,64] In the first, four phenotypes were identified: arrhythmia triggering (predominantly atrial fibrillation); hypertensive and worse echocardiographic features of HFPEF; systemic congestion; and infection triggered hospitalisation.[63] In the second analysis, they looked to limit the number of variables required to accurately assign patients to these latent classes to make the method more appealing to use in a clinical setting and were able to reduce from 32 to 16 variables with consistent grouping results.[64]

1.2.2.3 Latent class analysis in heart failure with mixed or unspecified ejection fraction

Analysis of a mixture of HFPEF and HFREF patients in Asian regions enrolled in the Asian Sudden Cardiac Death in Heart Failure (ASIAN-HF) registry found five subgroups.[65] Subgroups had geographical differences in prevalence and differences in the primary outcome of all-cause mortality and HF hospitalisation. The subgroup of lean diabetic patients was novel in this analysis, found predominantly in Southeast Asia. Other groups included a metabolic syndrome and a subgroup with young patients and low rates of ischaemic aetiology.

1.2.3 Supervised machine learning in heart failure

Supervised machine learning, where the aim is to develop an algorithm to map features (or variables) to a known output and assess performance when applied to new data, has also gained popularity in heart failure research. This approach can be used to develop methods to make a diagnosis from an image, for example diagnose heart failure from an electrocardiogram or for automated interpretation of an echocardiogram.[66,67] In terms of personalisation of heart failure care, there is great interest in whether machine learning methods may

provide more accurate prediction of adverse events as these methods benefit from increased flexibility compared with traditional statistical methods utilised to create heart failure risk models.[68–70]

1.2.3.1 Established risk models in heart failure

As highlighted, rates of mortality and morbidity are high in heart failure but it remains difficult to predict individual patient prognosis.[71,72] It is important to understand which patients are at higher risk to guide decision making around intensification of treatments, monitoring requirements or decisions about end of life care. There are several individual factors that are associated with prognosis in heart failure, such as NYHA class and ejection fraction, but using each individual marker in isolation does not well inform us of the overall risk.[73,74] Using multiple variables together in a regression model can allow risk to be more accurately estimated. There are many published heart failure risk models, which can be used as benchmarks to compare the performance of supervised machine learning models for adverse outcomes in heart failure. All the models below have online calculators to support the use of the risk prediction model in clinical practice.

MAGGIC risk score

The Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) risk score was developed using an international database including multiple cohort studies and patient level data of 39372 participants. [69] Patients with both preserved and reduced ejection fraction were included. Thirty-one candidate variables were available at baseline which were used in Poisson regression models with forward stepwise variable selection to include the most powerful predictive variables in the final models. Thirteen independent predictor variables were identified. An integer score was created from the risk coefficients which are summed to give the 1- or 3-year probability of death.

Get With The Guidelines-Heart Failure risk score

The Get With The Guidelines (GWTG) score utilised data from 39783 patients participating in the GWTG registry across 198 hospitals and was built to predict

in-hospital mortality.[75] A multivariable logistic regression model was used to assess candidate variables. Like the MAGGIC risk score, an integer score was calculated from coefficients in the regression model. The C-statistic for discrimination was 0.75 in both derivation and validation data.

PREDICT-HF risk score

The PARADIGM Risk of Events and Death in the Contemporary Treatment of Heart Failure (PREDICT-HF) models for the composite of HF hospitalisation and CV death, for CV death alone and for all-cause mortality were created using data from the Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure (PARADIGM-HF) trial, with validation using patients from the Aliskiren Trial to Minimize Outcomes in Patients with Heart Failure Trial (ATMOSPHERE) and the Swedish Heart Failure registry (SwedeHF).[68] An important addition in this model was natriuretic peptide levels. Sixty-three variables were considered, and variables were selected using a Cox proportional hazards model with stepwise selection. The C-statistics for the composite outcome in the PARADIGM-HF data at 1 and 2 years were 0.74 (95%CI 0.71–0.76) and 0.71 (95%CI 0.70-0.75) respectively, and in validation in ATMOSPHERE the C-statistics were 0.71 (95% CI, 0.69-0.72) and 0.70 (95%CI 0.68-0.71) respectively.

Barcelona Bio-Heart Failure Risk Calculator (BCN Bio-HF Calculator)

The BCN Bio-HF calculator uses a combination of clinical data, routine laboratory test and biomarkers to risk stratify patients with heart failure.[76] The enrolled patients were ambulatory patients in a multidisciplinary heart failure unit. Patients were required to have at least one heart failure hospitalisation and/or reduced left ventricular ejection fraction, meaning there are patients with both HFREF and HFPEF. 23 candidate variables were considered. A Cox regression analysis was utilised to create the model, with outcomes of 1, 2 and 3 year mortality. Generalisation was tested using multiple cross validated samples, with no separate validation cohort, giving an average C statistic of 0.79.

1.2.3.2 Common supervised learning techniques in heart failure analysis

On review of current supervised machine learning models, commonly used techniques included penalised linear regression, K-nearest neighbours, random forests, support vector machines, gradient boosted trees and artificial neural networks.

1.2.3.3 Supervised machine learning for outcome prediction in heart failure with reduced ejection fraction

Many studies utilise several different machine learning techniques in the same population and compare predictive performance between methods, with no consistent signal that one method outperforms others.[77–84] Often, logistic regression was found to have similar performance to more complex machine learning models. Outcomes assessed include: all-cause mortality; response to cardiac resynchronisation therapy (10% increase in left ventricular ejection fraction at 1 year); composite of hospitalisation for heart failure and mortality; and readmission to hospital within 30 days of discharge. Populations studied include clinical trial data, routine data collected from electronic records, and registry data, and the size of population ranges from a few hundred patients to several thousand. Almost all include a split of the data into training and testing groups appropriately, but few have external validation in fresh, unconnected data meaning the results may not be generalisable to other populations. The most common metric used to assess model performance is area under the receiver operating characteristic curve, generated by plotting sensitivity and specificity at different thresholds, equivalent to the C-statistic in a binary outcome.

1.2.3.4 Supervised machine learning for outcome prediction in heart failure with preserved ejection fraction

Similar approaches have been taken for prediction of outcome in patients with heart failure and preserved ejection fraction.[45,85–87] Populations were primarily patients in clinical trial data, with outcomes examined including composite of heart failure hospitalisation or mortality. There were much fewer studies in this area, but in one analysis of data from patients enrolled in the TOPCAT trial, random forest was found to be the best performing model for all-cause mortality

and heart failure hospitalisation, with modest C statistics (0.72 and 0.76 respectively).

1.2.3.5 Supervised machine learning for outcome prediction in heart failure with mixed or unspecified ejection fraction

The majority of studies examining risk prediction using machine learning methods either do not specify whether the population had preserved or reduced ejection fraction or include a mixture of both.[38,88-120] This is particularly common when data has been extracted from routine data collected in electronic records, where ejection fraction may not be recorded. There is not consistency in which method provided the best performing model across these studies, including whether machine learning outperformed logistic regression in prediction of a binary outcome. Gradient boosting algorithms performed best in several analyses, but prediction performance overall remains modest. [88,89,93,95,97,100,110-113]. In addition, not all analyses report model performance statistics in a separate test dataset, meaning performance measures may be erroneously inflated, and even fewer had external validation in another cohort of patients raising questions about generalisability.

1.2.4 Summary

Prognosis in heart failure remains challenging to quantify, despite multiple carefully developed risk prediction models. There is therefore a need to try new techniques to explore where progress can be made to improve prognostication. Supervised machine learning presents an opportunity to try new more flexible modelling to explore whether the well-established techniques can be improved upon. Machine learning has become more accessible to researchers across different fields as programmes which require less specialist coding knowledge become available. This creates the opportunity to explore the use of machine learning in clinical trial data in heart failure. It should be acknowledged that there is a fast rise in the number of publications relating to machine learning techniques with similar aims in different populations and settings. The complexity of analysis possible with machine learning can bring challenges in terms of critical analysis of results. The most common issue is lack of external validation of risk models, so even those with apparent strong performance measures might be erroneously inflated due to overfitting in the training data

and lack of generalisability. In addition, users of prediction models may be less comfortable with the use of machine learning due to the 'black box' processing involved, meaning we often do not know what the prognostic decision is dependent on. As a general principle, a more parsimonious solution is preferred therefore if a simpler model has similar performance capability this would be the preferred option. Despite these acknowledged issues, there is value in exploring these approaches further.

Another potential data driven method to better understand the prognosis and treatment responses of patients is phenotyping. Although there have been phenotype groups identified by different methods, one key question is consistency identification of groups. Another is exploring groups in different clinical trial datasets to explore for any other differences in treatment effects across phenotype groups. This could be explored further by looking at latent class analysis in the PARADIGM-HF and PARAGON-HF trials to examine both HFREF and HFPEF.

We widely accept the patient trajectory as previously described, with some deaths occurring suddenly with other patients having repeated deteriorations in functional status, mainly through our clinical experience of looking after these patients. However, the risk of cardiovascular death associated with clustered admissions has not been fully quantified and using clinical trial data we have accurate recordings of serial admissions and outcomes, including cause of death. This provides the data required to explore this further and aim to quantify the increased risk associated with hospitalisations.

In conclusion, machine learning is being utilised in a variety of settings in heart failure care, including supervised and unsupervised techniques. There is a large and growing interest in this area given the great flexibility in methods to address wide ranging aims. Many have found machine learning to perform better than other statistical approaches, but this certainly not always the case therefore critical review of methods and results is vital in this expanding field.

1.3 Aims and Objectives

1.3.1 Aims

Following review of the literature, the following aims were developed for this thesis:

- To describe the contemporary patient journey in patients with HFREF and HFPEF and explore the relationship between time between HF admissions and risk of cardiovascular death.

- To explore whether consistent phenotype groups can be identified using different latent class analysis techniques in different sets of data.

- To explore the ability of supervised machine learning techniques to predict adverse outcome in HFREF.

1.3.2 Objectives

These aims will be explored by addressing the following objectives:

- To use data from PARADIGM-HF and ATMOSPHERE trials to describe the time between adjacent HF hospitalisations in HFREF patients.

- To use data from the PARAGON-HF trial to describe the time between adjacent HF hospitalisation in HFPEF patients.

- To use time updated Cox regression models to describe how the risk of cardiovascular death is affected by admissions occurring at short intervals in both HFREF and HFPEF.

- To develop a latent class model using data from PARADIGM-HF and validated in ATMOSPHERE trials for patients with HFREF and compare with previously identified classes.

- To use the same variables as prior latent class analyses in HFREF in data from the PARAGON-HF trial to examine if consistent groups are identified using two different latent class analysis (LCA) approaches.

- To create machine learning models to predict cardiovascular death using data in the PARADIGM-HF trial and the variables used to create the PREDICT-HF prognostic model.

Chapter 2 Methods

2.1 Summary of datasets

This research made use of multiple large trials for various analyses aiming to explore greater personalisation of care for patients with heart failure. In heart failure with reduced ejection fraction, the included trials were ATMOSPHERE and PARADIGM-HF. In heart failure with preserved ejection fraction the trial was the Prospective Comparison of ARNI with ARB Global Outcomes in HF with Preserved Ejection Fraction (PARAGON-HF) trial.

2.1.1 ATMOSPHERE

The trial aimed to examine the effect of blockade of the renin-angiotensin-aldosterone system using a direct renin inhibitor (aliskiren) either alone or in combination with an angiotensin converting enzyme (enalapril) on cardiovascular death or heart failure hospitalisation in patients with HFREF, with single agent enalapril as comparator.

2.1.1.1 Population

The Aliskiren Trial to Minimize OutcomeS in Patients with Heart failure (ATMOSPHERE) trial recruited patients globally with heart failure and reduced ejection fraction.[121] Patients were recruited in 41 countries in around 800 centres. The protocol was developed by the Executive Committee of the trial with collaboration with clinical scientists from the sponsoring company Novartis. Approval was granted from Ethic Review Committees at each participating centre and all patients provided written informed consent.

Inclusion criteria for the trial where outpatients with: New York Heart Association class II-IV symptoms; age over 18 years; left ventricular ejection fraction of ≤ 35%; elevation of natriuretic peptide levels (BNP ≥150pg/ml or NT-proBNP >600pg/ml at screening visit OR BNP ≥100pg/ml or NT-proBNP >400pg/ml with an unplanned hospitalisation for heart failure within the preceding 12 months); be on a stable dose of ACE inhibitor for at least four weeks; and beta blocker (unless contraindicated or not tolerated) for at least four weeks.

Exclusion criteria included: history of hypersensitivity or allergy to the trial drug or other ACE inhibitors; patients on both angiotensin receptor blocker and aldosterone antagonist in addition to trial drug; current decompensation of heart failure; symptomatic hypotension or blood pressure less than 95mmHg systolic at baseline; recent (within 3 month) acute coronary syndrome, stroke, transient ischaemic attack, vascular surgery, percutaneous coronary intervention; coronary or carotid disease likely to need intervention within six months; right heart failure due to severe pulmonary disease; previous peripartum or chemotherapy induced cardiomyopathy; certain history of cardiac arrhythmias; recent implantation of cardiac resynchronisation therapy; significant valve disease; elevated serum potassium; chronic use of NSAIDS; current use of cyclosporine or itraconazole; and recent use of a direct renin inhibitor or intravenous vasodilators or inotropic drugs.

2.1.1.2 Trial intervention

There were three trial arms; enalapril alone, aliskiren alone and combination therapy with enalapril and aliskiren. The trial was performed as a randomised, double-blind trial.

Prior to randomisation, patients entered an active run-in period. Patients were switched from their baseline ACE inhibitor to enalapril, and the dose updated to the maximum tolerated then aliskiren added as a combination therapy, while monitoring for hyperkalaemia, symptomatic hypotension, renal dysfunction or symptoms of postural hypotension. Randomisation to the trial drug was stratified based on maximum tolerated dose then further up-titration was carried out as tolerated.

Follow up was event driven, with an initial plan for follow up to continue until 2318 events. During the trial, safety concern of treatment with aliskiren in patients with diabetes meant diabetic patients were switched to conventional therapy during the trial.[122,123] Diabetic patients were censored at the date this protocol amendment was carried out.

2.1.1.3 Analysis

The primary outcomes of the trial were to investigate whether combination therapy was superior to enalapril monotherapy in delaying time to first heart failure hospitalisation or cardiovascular death and to investigate whether aliskiren monotherapy was superior or non-inferior to enalapril monotherapy.

The primary analysis was a time to first event analysis (heart failure hospitalisation or cardiovascular death) using a Cox proportional hazards model. The model was stratified by high or low dose tolerance in run-in period and NYHA class, and baseline BNP was included as a covariate.

2.1.1.4 Primary results

8835 patients entered the run-in period, 1771 did not fulfil criteria for randomisation or were enrolled in sites subsequently closed due to violations of Good Clinical Practice guidelines. 7016 patients were randomised and included in the intention to treat analysis.[124] The median follow up was 36.6 months.

The overall result of the trial was neutral. There was no significance in the hazard of the primary outcome with combination therapy compared with enalapril monotherapy, with a hazard ratio of 0.93 (95% confidence interval 0.85-1.03, p =0.17). The hazard ratio for aliskiren monotherapy compared with enalapril monotherapy was 0.99 (95% confidence interval 0.90-1.10, p = 0.91 for superiority). Aliskiren monotherapy did not fulfil the prespecified requirement for non-inferiority compared with enalapril monotherapy. Patients randomised to combination therapy (compared to enalapril monotherapy) had more symptomatic hypotension, more renal impairment and more hyperkalaemia.

2.1.2 PARADIGM-HF

The Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF) investigated the addition of neprilysin inhibition to ACE inhibition in patients with heart failure and reduced ejection fraction.[14,125] The trial was an active comparator (enalapril), randomised, double blind trial with patients randomised to either

sacubitril valsartan or enalapril. The trial was approved by ethics centres at each participating centre, and all patients provided written informed consent.

2.1.2.1 Population

The recruitment criteria for the PARADIGM-HF trial were very similar to that for ATMOSPHERE. Entry criteria included: ambulant patients with New York heart association class II-IV symptoms; age 18 and over; LVEF ≤ 35% (originally ≤ 40% changed by protocol amendment); elevated BNP or NT-proBNP at the same levels as for ATMOSPHERE; on stable dose of ACE inhibitor or ARB for at least 4 weeks; on stable dose beta blocker (unless contraindicated or not tolerated) for at least 4 weeks.[125]

Exclusion criteria included: hypersensitivity to the trial drug or similar classes of drug; previous intolerance of ACE inhibitor or ARB at recommended target doses; history of angioedema; acute decompensated heart failure; symptomatic hypotension; reduced estimated glomerular filtration rate to < 30mL/min/1.73m^2; elevated serum potassium; acute coronary syndrome, stroke, transient ischaemic attack, major cardiovascular surgery, percutaneous coronary intervention or carotid angioplasty within 3 months; coronary or carotid disease likely to require intervention within six months; left ventricular assist device or heart transplantation; severe pulmonary disease; peri-partum or chemotherapy induced cardiomyopathy; certain ventricular arrhythmias; significant valvular disease; diseases altering absorption, distribution, metabolism or excretion of trial drug; or presence of other disease with life expectancy expected to be less than 5 years.

2.1.2.2 Trial intervention

Prior to randomisation, patients had a single-blind run-in period to assess whether they were able to tolerate both trial drugs with up titration with monitoring for hypotension, renal dysfunction and hyperkalaemia. Patients were then randomised 1:1 to sacubitril valsartan or enalapril with further up titration as tolerated.

2.1.2.3 Analysis

The primary endpoint was to investigate the effect of sacubitril valsartan, as compared to active comparator enalapril, on delaying time to first heart failure hospitalisation or cardiovascular death.

The trial was event driven and was stopped early by the data safety and monitoring committee due to overwhelming evidence of benefit with sacubitril valsartan. Results were analysed by intention to treat principle. Time to event data was analysed using Kaplan-Meier estimates and Cox proportional hazard models.

2.1.2.4 Primary results

In total, 10521 patients entered the run-in period, 2079 did not fulfil criteria for randomisation, another 43 were erroneously randomised or recruited in sites closed for violation of Good Clinical Practice guidelines. In total 4187 patients were randomised to receive sacubitril valsartan and 4212 to receive enalapril. The median follow up was 27 months.

The primary result of the PARADIGM-HF was positive. The hazard ratio for the primary composite outcome of time to first heart failure hospitalisation or cardiovascular death for sacubitril valsartan as compared to enalapril was 0.80 (95% confidence interval 0.73-0.87, $p < 0.001$). There were no safety concerns. Less patients randomised to sacubitril valsartan stopped the trial medication. Symptomatic hypotension was more common with sacubitril valsartan, but this did not lead to an excess in trial drug discontinuation. Renal dysfunction was less common with sacubitril valsartan.

2.1.3 PARAGON-HF

The PARAGON-HF trial investigated the addition of neprilysin inhibition to angiotensin receptor blockade in patients with heart failure with preserved ejection fraction.[29] There were similarities in trial design with PARADIGM-HF for patients with HFREF. PARAGON-HF was a randomised, double-blind, active controlled event driven trial comparing effect of sacubitril valsartan with

valsartan on heart failure hospitalisation or cardiovascular death in patients with HFPEF.

2.1.3.1 Population

Inclusion criteria for the trial included: age of 50 years or over; left ventricular ejection fraction of ≥ 45%; New York heart association functional class II-IV; on diuretic therapy for at least 30 days prior to screening; elevation of natriuretic peptides (>200 pg/ml if patient had been hospitalised for heart failure in the last 9 months or >300 pg/ml if no recent hospitalisation, triple these values if the patient was in atrial fibrillation); and evidence of structural heart disease including left ventricular hypertrophy or left atrial enlargement. [126]

Exclusion criteria included prior left ventricular ejection fraction of <40%; any alternative diagnosis that may explain the patient's symptoms and systolic blood pressure of <100 mmHg or ≥180mmHg (or >150mmHg unless on three antihypertensive medications).

2.1.3.2 Trial intervention

Patients had a pre-trial single blind run-in period to assess ability to tolerate both trial medications with up titration, with monitoring of systolic blood pressure, renal function, and potassium levels. Patients were then randomised 1:1 to either sacubitril valsartan or valsartan and followed up until a target number of events occurred or 26 months after randomisation of the last patient, whichever occurred later.

2.1.3.3 Analysis

The primary outcome was the effect of sacubitril valsartan as compared to valsartan on rate of hospitalisation for heart failure (total events- first and recurrent) and cardiovascular death. To model recurrent events, a semi-parametric proportional rates model (the LWYY method) was used.[127]

2.1.3.4 Primary results

A total of 4822 patients were randomised to receive either sacubitril valsartan or valsartan, with 26 patients excluded from analysis due to sites with violations of Good Clinical Practice guidelines.

PARAGON-HF was a neutral trial. The rate ratio for the primary composite outcome of total heart failure hospitalisation and cardiovascular death was 0.87 (95% confidence interval 0.75-1.01, p = 0.06).[29]

2.2 Analysis of the patient journey

For this analysis examining time between admissions, data from PARADIGM-HF and ATMOSPHERE trials were included to examine HFREF, and data from PARAGON-HF to examine HFPEF.

2.2.1 Identification of in trial hospitalisations

Dates of admission to hospital for worsening heart failure were extracted from the trial data. HF hospitalisations from the PARADIGM-HF trial and PARAGON-HF trial were adjudicated by a clinical events committee according to standardised definitions.[128] Hospitalisations on the same day as the patient died were excluded. In PARAGON-HF, non-HF hospitalisations were also analysed, information about these was collected at each trial visit but events were not adjudicated.

2.2.2 Statistical analysis

2.2.2.1 Descriptive statistics

Baseline characteristics were compared between patients with no HF hospitalisation, one HF hospitalisation and two or more HF hospitalisations. Normally disturbed continuous variables were summarised using mean and standard deviation, non-normally distributed continuous variables were summarised using median and interquartile range and categorical variables were summarised by counts and percentages. Differences between groups were examined using ANOVA for continuous variables, chi-square test for categorical variables and Kruskal-Wallis for non-normally distributed continuous variables.

2.2.2.2 Survival analysis and time-updating variables

This analysis was based on survival analysis techniques; therefore, an overview of the statistical approaches is presented here.

Survival analysis includes data on the amount of time each patient is observed in the trial, ending with either the event of interest or a censoring event. This includes more data than simply including whether the event of interest occurred at any point. Data of this structure is often called survival data, or time-to-event data.[129] Censoring is a key component of survival analysis and marks the end of the follow up time where the event of interest has not occurred, or the patient is no longer followed up for another reason. The censoring event may be the end of trial follow up, loss to follow up or withdrawing of consent. This type of censoring is "right-censoring". These patients do not provide information about when the event occurs as follow up ended before the event could be observed, but still provide useful information for the time they are involved in the trial without experiencing an event.

The survival function [*S(t)*] is given as the probability of a patient being event free at follow up time (t). The Kaplan-Meier (KM) estimate is commonly used to estimate the survival function and to display this graphically. The survival function can be expressed as:

$$S(t) = \frac{number\ of\ individuals\ surviving\ longer\ than\ t}{total\ number\ of\ individuals\ studied}$$

This formula only includes patients who are uncensored, or at risk, at time t.

The hazard function is another important concept in survival analysis, which gives the instantaneous rate of occurrence of event of interest in patients still at risk of the event at time t, which is commonly derived from the Cox proportional hazards regression model.[129] The underlying assumption of this model is that baseline hazard remains consistent over time and allows covariates in the model to influence the hazard. Each covariate in the model has a corresponding coefficient which describes the relationship between the covariate and the hazard of event.

The formula can be expressed as:

$$h(t) = h0(t)\exp(\beta X)$$

where h0 is the baseline hazard, β is the coefficient and X is a covariate. The hazard ratio (HR) for a covariate is calculated by the exponentiation of the β coefficient. This means the HR from a Cox regression model can be interpreted as the extent by which a covariate is associated with the rate of the outcome of interest, holding other covariates constant. A HR of 1 denotes no association between the covariate and rate of outcome, a HR <1 denotes a reduction in rate of the outcome associated with the covariate and a HR >1 denotes an increase in rate of the outcome associated with the covariate.

In most Cox-regression models, the value of the covariate of interest is included as the value measured at baseline, be that as a continuous value (e.g., the level of a biomarker at baseline) or as a categorical variable (e.g., presence or absence of diabetes at baseline). However, clearly these values can change over time, which is where time updating covariable analysis can be used for more accurate modelling. This is the method used in this analysis to examine the effect of number of days between HF hospitalisation and risk of cardiovascular death.

Time varying covariates can be used in several different ways. At a simple level, a binary variable can be used to signify the development of a feature of interest, such as a comorbidity, during the course of follow up. Each person in the analysis has the follow up time split into before the development of the feature of interest, and the period after the feature develops (if that patient does develop that feature). The association between development of the feature of interest and the outcome in the time to event analysis can be described.

Another way in which this method can be used is as a continually updating variable. An example may be results from a blood test that is measured repeatedly through follow up. A time updating covariable can change at each visit date to the new recording of the blood test. This approach was used to include the number of days between hospital admissions as a time updating variable in a Cox regression model. The time updating variable started as the

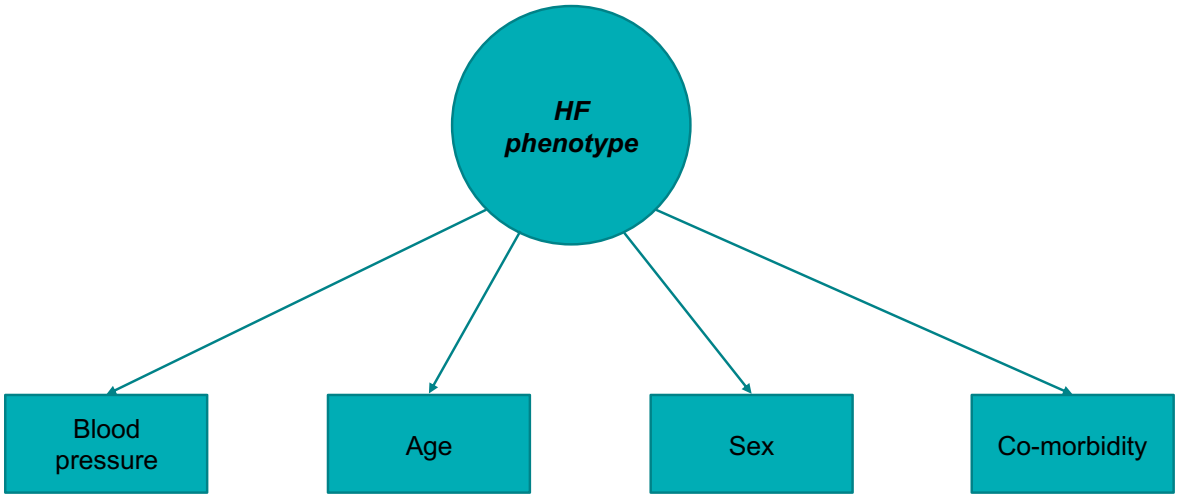
value of the number of days between the first hospitalisation and second, giving a value for baseline risk. This was heavily influenced by how long before randomisation the first hospitalisation took place, giving a reflection of how stable the patient was. At each subsequent hospitalisation, the value for the variable updated to be the number of days from the prior hospitalisation to the current admission. Patients with zero or one heart failure hospitalisations could not be included in this analysis as it is not possible to assign an appropriate baseline value for the variable. Further details are given in additional methods in Chapter 3.

2.3 Latent class analysis (unsupervised machine learning)

Latent class analysis (LCA) is a form of unsupervised machine learning, meaning the aim is to find patterns within the data without the data being labelled with an outcome. More specifically, latent class analysis aims to divide data in subgroups, called latent classes, which describe an unobserved and unmeasured construct in the data. In this context, the latent construct being investigated is the phenotype of patients with heart failure. Latent class analysis can be viewed as a form of cluster analysis.

As the data is not prelabelled with a phenotype group, and the underlying construct of the phenotype group is not known, we assume variables included in the data (for example age, sex or presence of comorbidity) reflect the underlying latent class construct. A visual summary of this concept is given in Figure 2-1. These variables can be used as indicators to which latent class group the data belongs too. As there is a degree of measurement error, the true class membership remains unknown and data is allocated to the best fitting latent class group. Patients can only be allocated to one latent class, and all data measurements are included in a latent class.

Figure 2-1. Visual illustration of underlying concept of latent class analysis
A visual description of the unobserved latent class influencing other variables that can be used to define and describe the underlying latent class.



The solution to a latent class analysis includes two main components. The first is the latent class prevalences, or the probability of membership in each class. This is sometimes known as the gamma parameter and gives an indication of the size of each latent class subgroup. The second set of parameters are the probability of reporting an indicator variable based on membership of that class, called item-response probabilities. For example, class A may report a 90% probability of reporting female sex while class B might report a 5% probability of reporting female sex.

2.3.1 LCA with categorical indicator variables

Latent class analysis uses categorical indicator variables to build the latent class construct. This is in contrast to latent profile analysis, which can be used to describe latent class models using continuous indicator variables. I will use the language latent class analysis to include both types of indicator variables and will specify the type of variables utilised in each analysis. Given latent class with categorical indicators is the most established, I will discuss this technique first.

Analyses of latent class groups was carried out using Stata software (StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC). Latent class analysis using categorical indicator variables was carried out using the LCA Stata Plugin from Penn State university.[130]

When the included indicator variables are categorical variables, the item response probabilities for each indicator will be a probability between 0 and 1, giving the probability that a person belonging to that latent class will report that indicator variable. An item response probability near 1 can be interpreted that persons belonging to that group are very likely to report that indicator variable, whereas near 0 means they are very unlikely to report the indicator variable.

Item response parameters close to the extremes indicate that the variable is strongly related to the latent class subgroup and is useful in discriminating between classes to create highly homogenous subgroups. A variable with middling item response parameters is less useful in discriminating between groups and can lead to groups with lower homogeneity.

Another important concept is the degree of separation between latent classes. Latent classes are highly separated if they have high or low item response probabilities in different indicator variables, essentially suggesting different indicator variables are strongly associated with each underlying latent class.

2.3.2 LCA with continuous indicator variables

Analysis using a mixture of continuous and categorical variables was completed using the inbuilt generalised structural equation modelling capabilities of Stata. Item response parameters for continuous variables are given as the expected mean value in each LCA subgroup. To avoid issues with differing scales of continuous variables, each was standardised to a mean of zero and standard deviation of one prior to entering the model. The parameter means are allowed to vary across different LCA subgroups. If there is difficulty in model identification, variances can be constrained to be equal across the subgroups, greatly lessening the number of ‘unknowns’ in the model; this approach is the default in Stata inbuilt programming.

2.3.3 Determining the optimum number of LCA groups

With latent class analysis, the first problem is identifying the number of subgroups which best fit and describe the data, with the number of classes being unknown at the start of the analysis. Often there is not a clearly defined best

fitting model, and several aspects of the model need to be considered and a decision made as to which model to present.

The most important attribute in accurately describing the latent construct of LCA subgroups is selection of appropriate indicator variables that give a strong reflection of the underlying subgroups. In this context, variables for each analysis were selected clinically as factors that are known to affect prognosis, or that are clinically used to separate important subgroups (such as ischaemic aetiology of heart failure). In the analysis of latent class subgroups in HFPEF, variables were selected that had been used in prior LCA analyses and results compared with the analysis carried out in a different population and using different LCA techniques, namely continuous and categorised variables.

Firstly, latent class models with a range of number of subgroups (2 to 7) are fit, and model fit criteria compared. Each LCA model is fit using maximum likelihood estimation. The likelihood function gives the likelihood of the observed data given the model being fit as a function of all parameter estimates. The best parameter estimates are those which maximise the likelihood. Problems with under-identification of the model can occur if the number of 'unknown' parameters that are to be estimated come close to or exceed the numbers of 'knowns' i.e., the data itself. Therefore, under-identification becomes more problematic with greater number of indicator variables or greater number of latent classes as both of these increase the number of parameters to be estimated.

2.3.3.1 Relative model fit

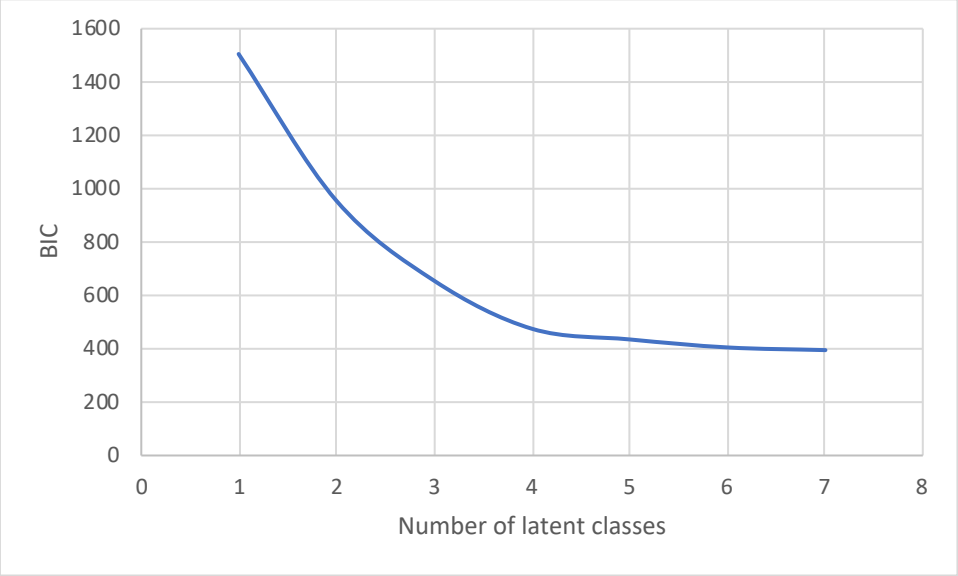
Although methods exist to examine absolute model fit in some cases of latent class analysis, this is not available when continuous variables are included in the programming languages utilised for this analysis. Therefore, model fit criteria were assessed in terms of relative model fit to compare LCA models with differing number of latent classes to inform which best described the data.

Information criteria were used to describe the models. Both Akaike information criteria (AIC) and Bayesian information criteria (BIC) were considered. Information criteria use penalised log-likelihood, which adds a penalty for

increasing complexity of the models to find a balance between model fit and parsimony. Models with smaller AIC or BIC indicate better model fit.

In models containing only categorical indicator variables, it is not uncommon that a clear minimum of the information criteria is seen, then both increase again with increasing number of LCA classes. When continuous variables are included, commonly the information criteria will continue to fall with increasing number of latent classes, however at a point increasing number of latent classes will only provide small decreases in information criteria. In these instances, a judgement can be made, for better model parsimony, where increased number of classes does not result in significant improvement in information criteria and this number of classes can be selected. An illustration is given in Figure 2-2.

Figure 2-2. BIC to identify appropriate number of latent class groups
An example of information criteria for latent class models with increasing number of subgroups. Although BIC continues to fall with increasing number of classes, there is only marginal improvement after 4 classes. Therefore, for model parsimony a 4-class solution may be selected.



2.3.3.2 Multiple starting seeds

One indication of lack of identification of the model is that different starting estimates for parameters lead to different model solutions, with differing likelihoods. This suggests the model solution is finding differing local minima, meaning the model result is less reliable as it does not consistently come to the same global minimum solution. The Penn State university program reports the percentage of 100 random starting seeds that come to the same solution which

can be used to inform the best model selection. There is no specific cut-off for the percentage of seeds coming to the same solution to be satisfied that the solution is stable but should be a factor of the decision of the number of latent classes.

2.3.3.3 Proportion of patients assigned to each class

A key output from the latent class model is the proportion of the data assigned to each latent class, or the gamma parameter. A group of too small a size, for example <5% of the population, is unlikely to be a genuine group of interest and LCA solutions with very small latent classes can be rejected. However, this is not an absolute rule, and if a smaller class is associated with a model with good fit statistical criteria and the small group is conceptually plausible the result can be accepted.

2.3.3.4 Entropy

Entropy is a measure of how well classes are delineated and gives a value between zero and one. Values approaching one indicate a clear delineation of latent classes.[131] Entropy should not be used to choose the solution to the latent class problem, but when a solution is identified from information criteria, latent class size and stability of solution entropy can be used to further evaluate the model. It is a simple measure of how easy it is to tell the latent subgroups apart from each other. A model can perform well statistically but still have a relatively low (poor) entropy and can be partly viewed as a measurement of how well indicators define the latent class groups. At lower entropy, the confidence in which patients can be allocated to the most likely LCA subgroup using modal probability, falls. In general, an entropy level of >0.8 is considered to be good, and entropy <0.6 indicates some difficulty in separating the latent class groups.[132]

Another way to examine certainty of class membership, and how well the classes are delineated, is to examine the mean posterior probability of class membership for each class. For example, for patients allocated to class A, examine the mean posterior probability of membership of class A, B, C and D. A good solution would have a high average posterior probability for class A (i.e.,

close to 1) and low posterior probability for other class membership (i.e., close to 0). This would support clear delineation between classes.

2.3.3.5 Allocating patient to most likely LCA group

The probability of an individual belonging to each latent class can be calculated using Bayes theorem using the combination of the parameter estimates for each LCA class and the individual responses to the indicator variables. The probability of membership on each class is calculated, termed the posterior probability, and the patient allocated to the class with the highest posterior probability.

2.3.3.6 Describing the latent class analysis solution

Once the final parameters have been determined by evaluating the aspects of the model as described above, these parameters can be used to describe the groups. The most descriptive indicator variables were highlighted for each parameter estimate, for example categorical indicators with values closest to zero and one. These form the basis by which to describe each latent class model in a human-readable way.

Further descriptive statistics for variables not included in the construction of the latent class groups can be summarised by assigning each patient to the most likely latent class and describing using percentages or mean and median as appropriate, with comparison between groups using tests such as chi-square or ANOVA. However, it is important to recognise this modal probability assignment does not account for the uncertainty of class membership therefore may not describe the group accurately.

2.3.4 Distal outcomes in LCA groups

Once LCA groups have been defined, examining outcomes in each latent class group is of interest. The simplest approach is to allocate each patient to the latent class they are most likely to belong to, then use the latent class variable in a Cox time to event analysis. Benefits of this approach include the ability to perform a time to event analysis. The main disadvantage to the approach is the degree of uncertainty of class membership is lost when the patient is allocated to their most likely latent class, meaning some relationships between latent

class group and outcome can be lost. For example, a patient with a 90% probability of belonging to the latent class contributes the same amount of data as someone with less certain class membership, for example 60%. A relationship between the latent class and outcome can be lost due to this uncertainty in class membership. This approach is known as maximum probability assignment, or modal probability assignment.

Another approach to prediction of a distal outcome is to account for classification error using specialised weights, often termed the "BCH" approach. This can be implemented in Stata using the Penn State plugin, which means the approach is limited to categorical indicator variables. The outcome of interest is analysed using logistic regression, and time to event analyses are not possible.

In my analysis, I focus on modal probability assignment and Cox regression for time to event analysis, accepting this is a conservative approach and some relationships between the latent class group and outcome may be lost due to uncertainty in class membership. However, this approach is applicable in analyses using continuous and categorical indicator variables in the software I have available, and benefits from time to event analysis rather than logistic regression for the outcome of interest.

2.4 Supervised machine learning

Supervised machine learning uses data with a known label, or outcome, and aims to map the input variables, or features, to the known outcome. An example might be chest x-ray images as an input with an attached known diagnosis as the output. The machine learning algorithm could then be trained to predict the diagnosis by learning patterns in the input images. Another example could be data regarding a patient with heart failure, such as age, sex, laboratory results and echocardiography parameters, with the outcome being mortality at two years of follow up. A machine learning model could be trained to predict the probability of mortality and could also be used to try and understand relationships between predictor features and the output (mortality).

Ultimately, the aim of all prediction models is to take readily available inputs (or variables) and use them in a function to generate an output, in the case of

risk prediction models the probability of adverse outcome. The true function relating the inputs to outputs is unknown and the prediction model aims to estimate the true function using the observed data. The observed data includes random error and may not include all important variables (for example, some aspect of the patient's physiology that has not been [or cannot be] measured), therefore a prediction model can never fully estimate the true function.

This is closely related to the concept of over-fitting. If a model becomes overfit, it may be able to predict outcome very accurately in the data used to develop the model as it has been fit to the underlying true function and the random error incorporated in the data. As a result, performance statistics may appear extremely positive when evaluated in the training data, but when the model is applied to fresh data model performance declines. Overfitting is a particular hazard in machine learning where the high degree of flexibility needs to be controlled to create a generalisable model.

2.4.1 Common supervised learning techniques used for prediction of outcome

In machine learning methods for event prediction, generally the outcome is a binary (i.e., event or no event at 1 year follow up) and the methods utilised are therefore classification models. The alternative is regression models, where the outcome is a numeric value (a simple example of this might be prediction of weight given inputs of height, sex and physical activity level). This can be contrasted with survival model, where time to event as well as event occurrence or censoring is modelled. General principles of each machine learning method used in this research is given in turn.

Logistic regression

Logistic regression is an extension of a linear regression which gives a probability of the outcome lying in each classification (i.e., probability of event or no event with "event" coded as 1 and "no event" coded as 0). Multiple *linear* regression gives a numerical output based on the equation:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C \ldots$$

where Y is the output, β_0 is the constant, A, B and C are input variables or predictors and β_{1-3} are coefficients for each predictor. The output is a numeric value. The formula for a logistic regression model, giving the probability of a binary outcome, is:

$$p = \exp(\beta_0 + \beta_1 A + \beta_2 B + \beta_3 C \ldots)$$

where p is the probability of the outcome. The output of this model is a value between 0 and 1 for the probability of "event". Usually, probability of 0.5 or above would be classified as "event" and below 0.5 be classified as "no event", however this threshold can be altered depending on the relative importance of false positives and false negatives in the context of the model.

Penalised logistic regression

Penalised regression techniques aim to penalise a model for high numbers of predictor variables. This can be desirable to lessen the impact of variables that have large variances or are highly correlated, or to lessen the number of predictors to make an easier to interpret and use model. These models include a regularisation parameter which controls the relationship between model fit and the penalty term. There are different functions that can be used for regularisation. In ridge regression, regression coefficients further from 0 are penalised favouring smaller coefficients, known as shrinkage. Coefficients can approach zero but can never be shrunk to zero. Lasso regression is similar but uses a different penalty function which will allow coefficients to be shrunk to exactly 0. Elastic net regression uses a mixture of ridge and lasso penalty formulas. The optimum value of the penalty and the optimum mixture of ridge and lasso regression is not known prior to fitting the model and must be identified by tuning the model, as discussed later.

K-nearest neighbours

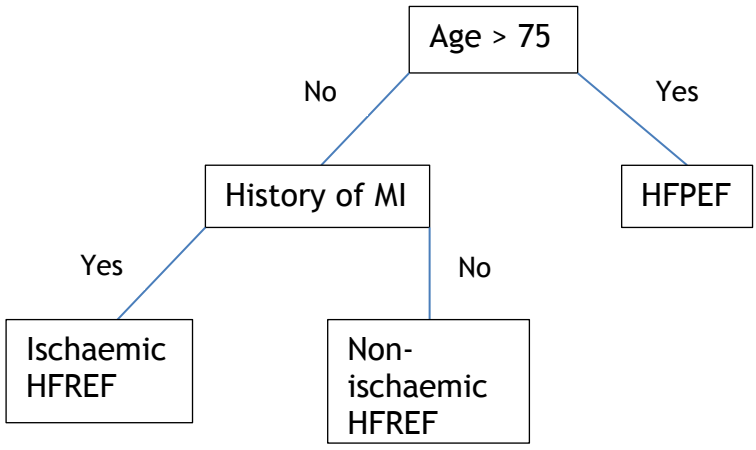
K-nearest neighbours (KNN) can be used to classify points into groups, whether that be binary or multiple groups. The number of neighbours must be decided (*k*), this is usually done iteratively in a tuning process. Each data point has a location within the feature space, dependent on the input variables, and has a

known label. A new datapoint is analysed, if *k* = 5 then the classification of the 5 nearest data points is reviewed and the datapoint allocated to the most commonly occurring class, defining a boundary between classifications. A small value of *k* will lead to a very flexible decision boundary with low bias but high variance, and be prone to overfitting, however a very large value of *k* will have more linear decision boundaries with high bias and low variance.

Random forest

There are several tree-based machine learning methods, with random forests being a commonly used method. Random forests are based on decision trees, which include multiple decision points which ultimately will allocate a patient to the most likely class, as illustrated simply in Figure 2-3.

Figure 2-3. Example decision tree
A simplified illustration of a decision tree.



An issue with using a single decision tree is there is a large amount of variance depending on which sample of data is used to grow the decision tree. For example, if the data were split in half and a tree grown using data in each half, the resultant trees are likely to be very different from each other. There are several ways to circumvent this issue, including bootstrap aggregation. In this method, multiple bootstrap samples (multiple samples of data taken from the master data with resampling) are generated then decision trees grown in each,

with the final output being an average of all the bootstrapped trees. Taking an average result like this reduces variance.

Random forests are a further development on bagged trees. Again, multiple trees are grown, the primary difference being only a limited selection of predictors are considered at each split in the decision tree, the number of which is a tuneable parameter. If there is a strong predictor and for every tree all predictors are considered, the strong predictor will decide an early split in all the decision trees, resulting in highly correlated trees. By controlling the number of predictors considered at each split, this problem is reduced and there is less correlation between trees, resulting in less variable predictions.

Random forests have several benefits in that they do not require data pre-processing (such as normalising continuous variables to a mean of zero and standard deviation of one) and are not prone to overfitting.

Gradient boosting

Gradient boosted trees are a further development on random forests, where the result of each tree depends on the results of the trees grown prior and is classified as an ensemble method. In essence, many weakly performing trees are combined in an iterative fashion, with each iteration trying to improve on the performance of the previous model under regulation of the loss function. This method can often out-perform random forests.

Artificial neural network

Artificial neural networks (ANN) aim to mimic the function of a neuron in the nervous system. There are several inputs to the perceptron (the computational neuron) and an activation function, these are processed in the perceptron and an output generated with an output transformation dependent on whether it is a classification or regression problem. Neural networks can have several layers (or "hidden layers") of perceptrons, in this research single layer neural networks are utilised. To prevent overfitting, a penalty term is applied, usually termed back-propagation in ANN. ANN are very flexible and can derive complex relationships in a model with high numbers of input variables, making them a valuable tool.

2.4.2 Outcome

In the analysis of data from the PARADIGM-HF trial, the outcome of interest is cardiovascular death at 2 years follow up to allow comparison with the PREDICT-HF model. Cardiovascular death is a relevant outcome in heart failure, occurs commonly enough to model and is modifiable by various treatments. The machine learning methods utilised, namely penalised logistic regression, random forest, gradient boosted trees and a single layer neural network, aim to predict a binary outcome i.e., presence or absence of the adverse outcome. Therefore, all models were classification models, rather than regression. To increase interpretability, and relevance in the clinical setting, the occurrence of each outcome was calculated after set times of follow up rather than whether the outcome occurred at all during the trial. The predicted risk of experiencing an outcome at one year or five years rather than over the median trial follow up is easier to interpret for the physician and patient.

2.4.3 Training and test data

Supervised machine learning techniques use the input variables to learn to predict the labelled output or seek to infer information about the relationships between the input features and the output. Machine learning techniques have advantages in behaving flexibly without the required constraints of some statistical analysis techniques. This has clear benefits, however, can also have negative implications. Any machine learning method is trying to define a function that transforms the input to the output. However, the true underlying data which is used to define the function also includes an unknown amount of error. Therefore, an algorithm trained on the data can also flexibly model including the degree of error, and ultimately be over-fit to the training data. This can be reflected in over-optimistic estimates of model performance if this is evaluated in the same data on which the algorithm was trained. A core component of any supervised machine learning technique is once the method is optimised in the training data, best practice is that it be applied to unseen "test" data and model performance evaluated in this analysis.

For each type of model, the same train/test split was used. An 80:20 split was used for training and testing data. The split in data was stratified by the

outcome variable of interest to ensure there was an equal spread of positive and negative cases in the test and train data. The same training and test split was used for all model development.

2.4.4 Hyperparameter tuning

Each machine learning method has different hyperparameters which control an aspect of how the machine learning algorithm works. For example, in a random forest classification supervised learning problem, one hyperparameter is deciding the minimum number of data points to be included in a terminal node. If this is too small, the model would be prone to overfitting. If too large, the model will not discriminate well between different outputs as many observations are included together in the terminal node. The optimum value for these parameters then needs to be searched for by repeatedly running the algorithm and assessing model performance using cross-validated or bootstrapped samples of data. Once the optimum parameters which give the best model performance are identified, these are applied to create the final model. Machine learning packages provide helper functions to generate a range of possible hyperparameter combinations and iteratively assess each. Regular grids were used initially for exploration with further evaluation of models carried out using space filling grids generated using a maximum entropy design with 20 candidate values.

Ten-fold cross validation was used to tune hyperparameters. Ten subgroups of the training data are formed. Different values for the hyperparameters are used to train models in nine of the samples, using the last sample to test the model and provide model performance measures. This is repeated iteratively, with the fold used to assess model performance changing at each iteration (Figure 2-4).

Figure 2-4. Five fold cross validation
A simplified diagram illustrating fivefold cross validation used for hyperparameter tuning in machine learning models. The model is trained in 4 folds of the data, and model performance assessed in the other fold. The testing fold changes with each iteration.

TRAIN	TRAIN	TRAIN	TRAIN	TEST
TRAIN	TRAIN	TRAIN	TEST	TRAIN
TRAIN	TRAIN	TEST	TRAIN	TRAIN
TRAIN	TEST	TRAIN	TRAIN	TRAIN
TEST	TRAIN	TRAIN	TRAIN	TRAIN

2.4.5 Assessment of model performance

Classification analysis is interested in mapping the input features to a categorical or binary output variable. Analysis of model performance can use metrics including accuracy, sensitivity and specificity. Sensitivity and specificity can be displayed succinctly in the confusion matrix. Firstly, the algorithm is applied to new data, and the predicted classification is compared to the true classification.

Accuracy is simply the percentage of predictions that are in the correct category, with a higher accuracy (closer to 100) indicating better model performance. Accuracy is less reliable if the outcome is imbalanced, for example if the number of “Yes” responses greatly outnumber the number of “No” responses. Even if the model predicted every outcome as “Yes”, which is clearly a very poorly performing model, the accuracy for the prediction of “Yes” would still be very high, given at baseline many more of the correct responses are yes.

A confusion matrix compares the true responses with the predictions from the model (Figure 2-5). From this, sensitivity and specificity can be calculated. Sensitivity, or the true positive rate, is the rate that the event of interest is correctly predicted calculated as;

$$\text{true positive} / (\text{true positive} + \text{false negative})$$

Specificity, or true negative rate, is the rate at which non-events are correctly predicted, given as;

$$\text{true negative} / (\text{true negative} + \text{false positive})$$

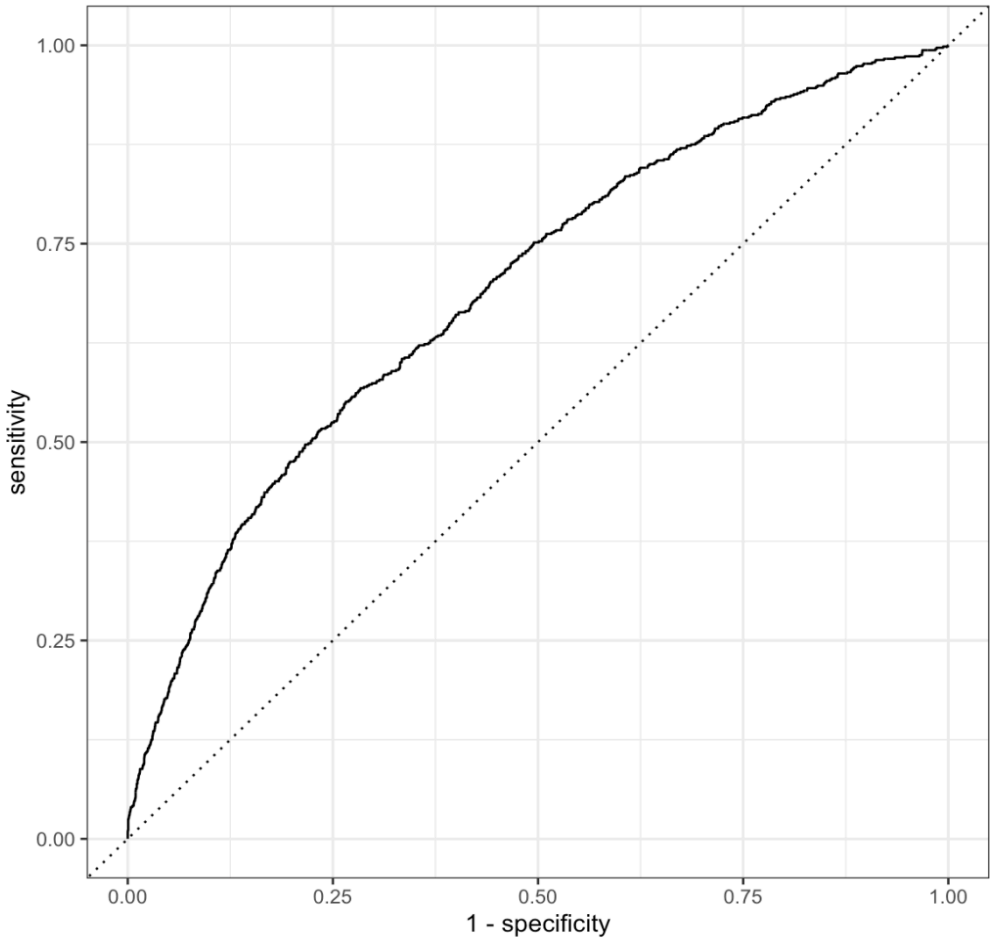
A very well performing model would have both sensitivity and specificity approaching 1.

Figure 2-5. Confusion matrix
An example of a confusion matrix. The true classification is compared with the predictions from the model. Sensitivity and specificity can be calculated from the confusion matrix.

	Truth	
Prediction	No	Yes
No	True negative	False negative
Yes	False positive	True positive

Model discrimination is a measure of how well the model is able to separate patients who have an event from those who did not. This is commonly described using the concordance (C) statistic for binary outcomes. A value of 0.5 equates to a model no better at discrimination than a 50:50 "play of chance", and a value of 1 equates to perfect discrimination. For a binary outcome, the C statistic equates to the area under the receiver operating characteristic curve (ROC AUC). The ROC AUC is plotted as the sensitivity against 1-specificity over the range of cut points for the probability of the outcome. An example of a ROC AUC is plotted in Figure 2-6.

Figure 2-6. Example ROC AUC plot
An example of a ROC AUC plot. Sensitivity is plotted on the y-axis with 1-specificity on the x-axis. The area below the curve equates the ROC AUC and for a binary outcome is equivalent to the C statistic.



2.5 Statistical software packages

For analysis of time between heart failure hospitalisations using a time updating variable, Stata software (version 17) (StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC) was used.

Latent class analysis using categorical indicator variables was carried out using the LCA Stata Plugin from Penn State university within Stata software.[130] LCA using mixture of categorical and continuous variables was using inbuilt Stata generalised structural equation modelling.

Supervised machine learning techniques were performed using R software (version 4.2.1) and R studio (version 2022.07.1).[133] The main packages used for machine learning were in the *tidymodels* package.[134] For neural network

analysis, *keras* package using Python was utilised within R using the *tidymodels* framework.

Chapter 3 Identifying the features of the patient journey in HFREF

3.1 Introduction

The aim of this analysis is to explore the contemporary patient journey with regards recurrent heart failure hospitalisations in patients with heart failure and reduced ejection fraction. The accepted trajectory is of recurrent declines, which might be associated with a hospitalisation for heart failure, with declines remaining unpredictable but occurring closer together as the patient approaches death with sudden death occurring at any point along the trajectory.[36] The associated risk of cardiovascular events following discharge is increased, and there is a trend to be more hospitalisations toward the end of life.[135,136] The degree of increased risk after a hospitalisation with acute heart failure can be difficult to quantify, and there are risk stratification tools available to specifically assess risk of mortality at seven days following admission to the emergency department.[137] In addition, what is less understood is the impact of the proximity of adjacent hospitalisations and the degree of increased risk in the post hospitalisation period, therefore this analysis aims to address this question.

Another aspect of this issue is the type of death that is more common after a hospitalisation. Cardiovascular death is a common cause of death in patients with heart failure, particularly those with HFREF and lower ejection fraction, and sudden death accounts for a large proportion of these deaths.[138,139] Established treatments for heart failure reduce cardiovascular mortality and specifically sudden cardiac deaths.[140–142] This is reflected in the finding that rates of sudden death in patients with heart failure enrolled in clinical trials as continued to decline over time.[143] It would therefore be interesting to look at the patterns of repeated hospitalisations and in patients with different modes of death to examine whether sudden deaths are truly ‘sudden’ or if a relationship between hospitalisations and sudden death becomes apparent. This could have implications in treatment choices for patients, particularly device therapies.

This analysis aims to explore these issues using contemporary, geographically diverse data from clinical trials and explore the associated risk of cardiovascular

death with admissions which are close together chronologically. The data utilised is from the PARADIGM-HF and ATMOSPHERE studies.

3.2 Methods

3.2.1 Identification of hospitalisations, worsening outpatient events and patient reported symptom scores

3.2.1.1 In trial heart failure hospitalisations

Information on in-trial HF hospitalisations (HFH) was extracted from adjudicated endpoints and hospitalisation records. Hospitalisations where the patient died on the same day as admission were excluded. Discharge date was imputed if not available using the average days of hospitalisation per region as per the trial statistical analysis plan. Other causes of hospitalisation were not recorded as a trial endpoint therefore the focus in this analysis is on HF hospitalisation only.

The main outcome examined was CV death with breakdown of different causes of death (pump failure, sudden and other CV deaths). Secondary outcomes considered are all cause mortality and non-CV mortality.

In a sensitivity analysis, HFH where the patient did not survive to discharge were excluded, regardless of length of admission.

3.2.1.2 In trial outpatient worsening events

To further enrich the data, in trial outpatient worsening events were added to the hospitalisation data. These were defined as an emergency room visit for HF or intravenous (IV) treatment for heart failure.

3.2.1.3 Pre-trial hospitalisation

Patients in both trials had date of most recent HFH before enrolment recorded at screening. To facilitate analysis of time between admissions, date of prior HFH were added to in-trial HFH if recorded with sufficient accuracy (i.e., day, month and year) as well as hospitalisations occurring between screening and randomisation. In the analysis of time between admissions, patients were included if they had at least two heart failure hospitalisations.

This analysis was repeated excluding hospitalisation where the patient died during the admission in a sensitivity analysis.

3.2.1.4 KCCQ

Although declines in function are primarily assessed through marked declines resulting in hospitalisation in this analysis, there are also patient reported symptom scores measured repeatedly during study follow up that have been analysed to further expand and explore the trajectory of functional status, split by both mode of death and number of hospitalisations.

The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a self-administered questionnaire validated in HF with both reduced and preserved ejection fraction which measures quality of life over several domains including symptoms, physical limitations, social limitations, quality of life and self-efficacy.[144-147] Symptom burden and functional status was examined using the Kansas City Cardiomyopathy Questionnaire Clinical Summary Score (KCCQ-CSS). In PARADIGM-HF KCCQ was measured at randomisation visit, then at 4, 8, 12, 24, 36 and 48 months. In ATMOSPHERE it was measured at randomisation, 4, 8, 12 and 24 months.

3.2.2 Statistical analysis

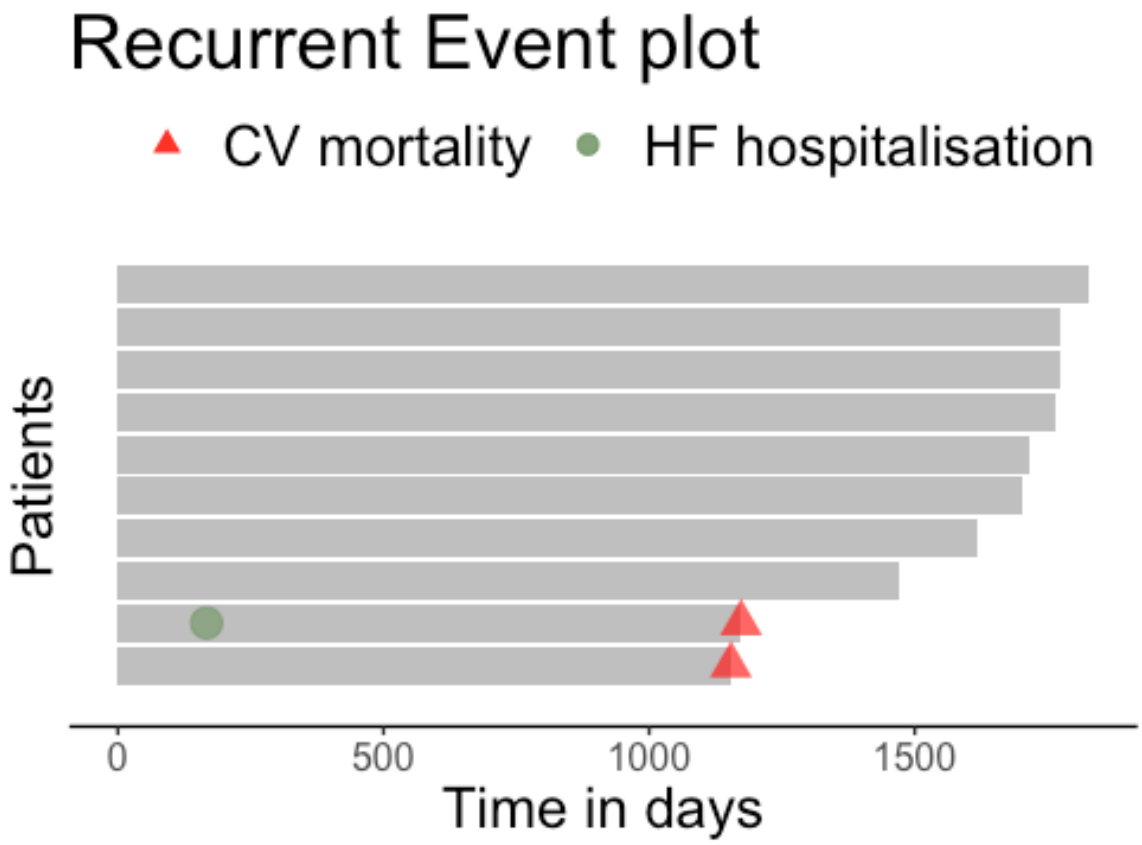
3.2.2.1 In trial hospitalisations

Patient demographics comparing patients with no HFH, 1 HFH or more than 1 HF hospitalisation during trial follow up were compared using ANOVA for continuous variables, chi square test for categorical variables and Kruskal-Wallis for non-normally distributed continuous variables. The number of HFH per patient was calculated was cross tabulated with the cause of CV death. The Lin, Wei, Ying and Yang (LWYY) method was used to calculate rates for recurrent HF hospitalisation and calculated in groups of patients with different causes of CV death.[127]

Recurrent event plots to visualize patterns of HFH were plotted using *reReg* package in R.[133] In these plots, each individual patient has a follow up time illustrated with a grey line. The time at which a hospitalisation happens is

marked with a green circle. At the end of follow up time, the occurrence of CV death is marked with a red triangle. Due to the large number of patients included, the grey follow up timelines appear merged. The density of green circles can be visually appreciated, as well as the pattern at which they are occurring. A simplified example of 10 random patients is given in Figure 3-1.

Figure 3-1. Example of recurrent event plot
A recurrent event plot for 10 random patients. Each grey line is the duration of follow up for one patient. A green circle marks the timing of a heart failure hospitalisation. If the follow up ends with a CV mortality, this is marked with a red triangle.



In a sensitivity analysis, the above analysis was repeated excluding HFH where the patient died in hospital regardless of length of admission. The above analysis was also repeated including outpatient worsening HF events.

3.2.2.2 Time between admissions analysis

Patients were included in this analysis if they had at least 2 recorded HFH (including previously described pre trial hospitalisations). The median number of days between each HFH hospitalisation was cross tabulated with cause of CV death.

Time between admissions was examined as a time updating covariate. Patients entered the analysis on the date of their second hospitalisation. The time varying covariate value started as the time between first and second HFH. At each HFH, the variable updated to the time between the prior admission to current admission. The time updated variable was entered in a restricted cubic spline analysis using 5 knots in a Cox regression for cardiovascular (CV) mortality, all-cause mortality and non-CV mortality.

3.2.2.3 KCCQ

A mixed model for repeated measurement was used to examine change in KCCQ-CSS over time (adjusted for randomized treatment, number of HF hospitalisations, and interaction of number of HF hospitalisations and visit, with a random intercept and slope per patient).

3.3 Results

3.3.1 In-trial heart failure hospitalisations

Of the 15415 patients enrolled, 2518 had at least 1 hospitalisation after randomization and between them these 2518 participants accrued a total of 4318 admissions. There were 2872 CV deaths which accounted for 83% of all deaths. Of the 2872 CV deaths, 1332 (46%) occurred suddenly, and 735 (26%) were due to worsening heart failure. The maximum number of HF hospitalisations experienced by a patient was 18.

Table 3-1 gives baseline demographic information for patients with no HF hospitalisations, 1 hospitalisation and >1 hospitalisation. Patients with recurrent HFH have several differences including older age, greater proportion patients of black race, lower systolic BP, higher heart rate, lower ejection fraction, higher NT-proBNP, worse NYHA scores, greater prevalence of comorbidity including diabetes, atrial fibrillation and more patients had a history of HFH prior to trial enrolment. More patients who had recurrent admissions were prescribed a diuretic and digoxin, while less were prescribed beta blockers. More patients with recurrent HFH had ICDs or CRTs implanted.

Table 3-1. Baseline data by number of HFH (HFREF)
Baseline demographic table by number of in follow-up heart failure hospitalisations. Continuous variables are expressed as mean ± standard deviation, or median [interquartile range], as appropriate.

	No HFH N=12,897	1 HFH N=1,603	≥2 HFH N=915	p-value
Age (years)	63.4±11.5	64.1±11.8	64.8±11.9	<0.001
Female sex (%)	2,897 (22.5)	317 (19.8)	143 (15.6)	<0.001
Race or ethnic group (%)				<0.001
Caucasian	8,427 (65.5)	1,088 (68.2)	621 (68.5)	
Black	421 (3.3)	69 (4.3)	47 (5.2)	
Asian	2,755 (21.4)	326 (20.4)	192 (21.2)	
Other	1,255 (9.8)	113 (7.1)	47 (5.2)	
Region (%)				<0.001
North America	588 (4.6)	112 (7.0)	79 (8.6)	
Latin America	2,225 (17.3)	221 (13.8)	106 (11.6)	
Western Europe	3,193 (24.8)	454 (28.3)	255 (27.9)	
Central Europe	4,027 (31.2)	478 (29.8)	265 (29.0)	
Asia/Pacific	2,864 (22.2)	338 (21.1)	210 (23.0)	
Systolic blood pressure (mmHg)	122.7±16.6	121.5±17.7	120.3±17.2	<0.001
Heart rate (beats/min)	71.8±12.1	73.1±13.0	73.4±13.0	<0.001
Body mass index (kg/m^2)	27.7±5.4	28.2±5.6	28.2±5.9	<0.001
eGFR (mL/min/1.73m^2)	33.2±40.1	37.4±39.5	39.1±38.8	<0.001
Ischaemic cardiomyopathy (%)	6,811 (52.8)	807 (50.3)	504 (55.1)	0.057
Left ventricular ejection fraction (%)	29.2±5.9	28.2±6.1	27.7±6.3	<0.001
Median NT pro BNP (pg/ml)	1337.0 [739.0-2575.0]	1796.0 [973.0-3720.0]	2168.5 [1095.0-4417.0]	<0.001
NYHA class III/IV (%)	4,427 (34.4)	675 (42.1)	405 (44.3)	<0.001
Hypertension (%)	8,562 (66.4)	1,106 (69.0)	604 (66.0)	0.10
Diabetes (%)	3,945 (30.6)	567 (35.4)	339 (37.0)	<0.001
Atrial fibrillation (%)	4,456 (34.6)	650 (40.5)	375 (41.0)	<0.001
Hospitalisation for heart failure (%)	7,634 (59.2)	1,111 (69.3)	717 (78.4)	<0.001
Loop diuretic (%)	10,130 (78.5)	1,386 (86.5)	820 (89.6)	<0.001
Digitalis (%)	2,130 (29.6)	259 (33.0)	150 (36.6)	0.002
Beta-blocker (%)	11,955 (92.7)	1,459 (91.0)	829 (90.6)	0.006
Mineralocorticoid antagonist (%)	6,069 (47.1)	762 (47.5)	442 (48.3)	0.73
Implantable cardioverter-defibrillator (%)	985 (13.7)	166 (21.1)	92 (22.4)	<0.001
Cardiac resynchronization therapy (%)	720 (5.6)	143 (8.9)	104 (11.4)	<0.001
Number of HF hospitalisations	0.0 (0.0-0.0)	1.0 (1.0-1.0)	2.0 (2.0-3.0)	<0.001
CV death (%)	1,882 (14.6)	542 (33.8)	448 (49.0)	<0.001

The number of HFH was cross-tabulated with cause of CV death (Table 3-2 and Figure 3-2). Due to small numbers of patients with more than 5 admissions these were combined into one category. Of patients that die of progressive heart failure ("pump failure"), 77% had at least 1 HFH during trial follow up. In contrast, 15% of those with sudden death had at least 1 admission, 27% of those with other types of CV death and 12% of those who did not have CV death. The pattern was similar in a sensitivity analysis where hospitalisations where the patient died the same admission were excluded, however the proportion of patients with pump failure death who had at least 1 HFH falls to 56% (Table 3-3). Examining this another way, of the 1882 patients with no HFH who have CV death, 59% die of sudden CV death and 8.8% die of pump failure; of 448 patients with two or more HFH who die of CV causes, the proportion having sudden CV death falls to 15%, with 65% dying of progressive heart failure. When outpatient HF worsening events are added to the hospitalisation data, the results were similar (Table 3-3).

Table 3-2. Mode of CV death and number of HFH (HFREF)
Mode of CV death cross tabulated with number of HF hospitalisations during trial follow up. Percentages are given within each column.

Number of HFH	Pump failure (n=735)	Sudden death (n=1332)	Other CV death (n=805)	Non-CV death (n=569)	Alive/ Censored (n=11974)	Total
0	166 (22.6%)	1,127 (84.6%)	589 (73.2%)	477 (83.8%)	10,538 (88.0%)	12,897 (83.7%)
1	278 (37.8%)	137 (10.3%)	127 (15.8%)	63 (11.1%)	998 (8.3%)	1,603 (10.4%)
2	150 (20.4%)	39 (2.9%)	51 (6.3%)	18 (3.2%)	254 (2.1%)	512 (3.3%)
3	62 (8.4%)	15 (1.1%)	15 (1.9%)	8 (1.4%)	102 (0.9%)	202 (1.3%)
4	29 (4.0%)	7 (0.5%)	7 (0.9%)	2 (0.4%)	43 (0.4%)	88 (0.6%)
5 or more	50 (6.8%)	7 (0.5%)	16 (2.0%)	1 (0.2%)	39 (0.3%)	113 (0.7%)

Figure 3-2. Proportion of patients experiencing multiple HFH (HFREF)
Proportion of patients who experienced 0, 1, 2, 3, 4 or 5 or more HF hospitalisations by mode of CV death (%).

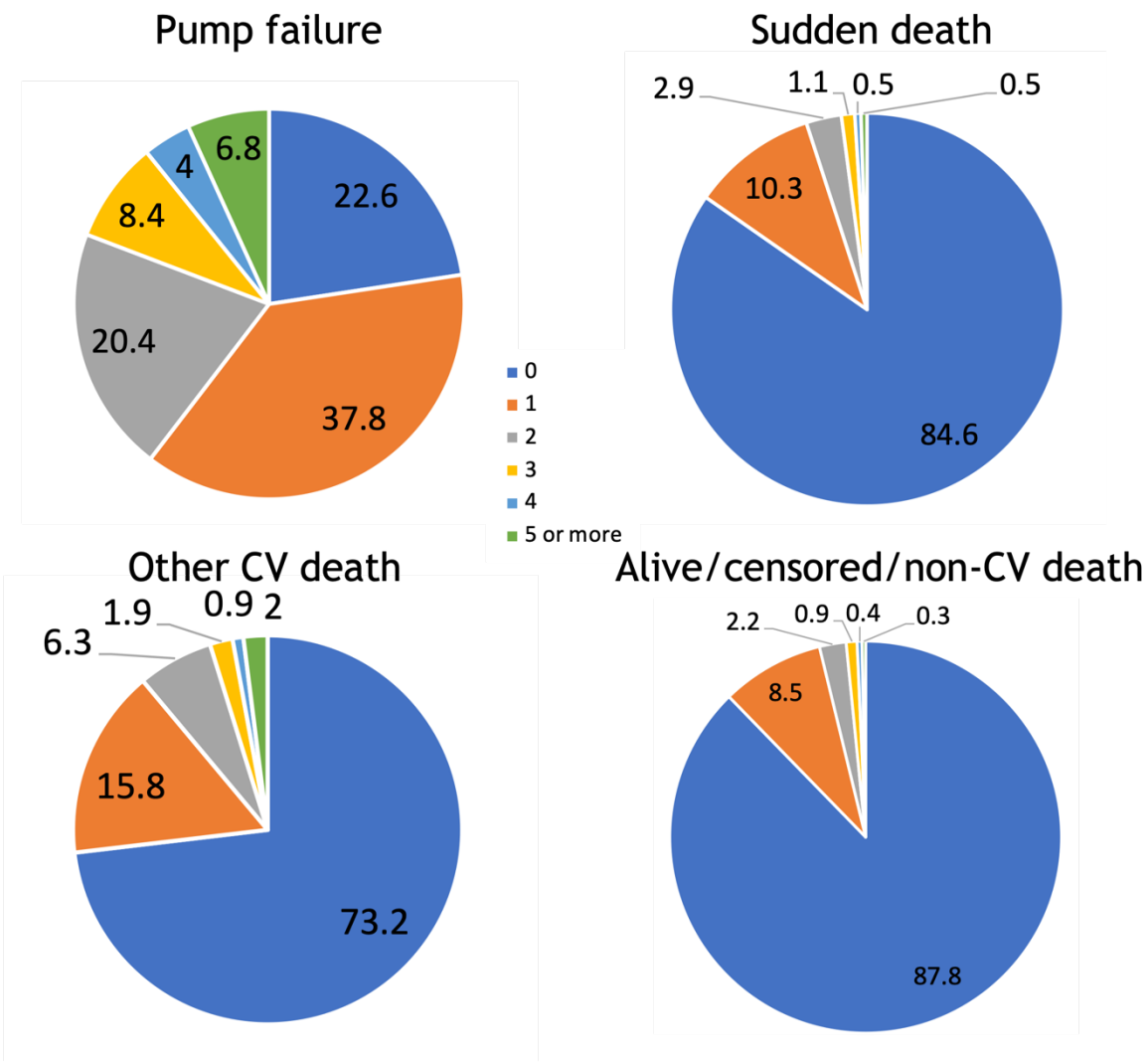


Table 3-3. CV death and HFH in sensitivity analyses (HFREF)
Cross tabulation of cause of CV death and number of HF hospitalisations in sensitivity analysis with [A] admissions removed where patient was not discharged alive and [B] with outpatient worsening events added to all HF hospitalisations [B].

A					
Number of HFH	Pump failure (n = 735)	Sudden death (n = 1332)	Other CV death (n = 805)	Alive/ Censored (n = 12543)	Total
0	322 (43.8%)	1128 (84.7%)	605 (75.2%)	11022 (87.9%)	13077 (84.8%)
1	213 (29.0%)	136 (10.2%)	118 (14.7%)	1055 (8.4%)	1522 (9.9%)
2	91 (12.4%)	39 (2.9%)	44 (5.5%)	272 (2.2)	446 (2.9)
3	51 (2.0%)	15 (1.1%)	15 (1.9%)	109 (0.9%)	190 (1.2%)
4	15 (2.0%)	7 (0.5%)	8 (1.0%)	45 (0.4%)	75 (0.5%)
5 or more	43 (5.9%)	7 (0.5%)	15 (1.9%)	40 (0.3%)	105 (0.7%)
B					
Number of HF events	Pump failure (n = 735)	Sudden death (n = 1332)	Other CV death (n = 805)	Alive/ Censored (n = 12543)	Total
0	122 (16.6%)	1085 (81.5%)	546 (67.8%)	10715 (85.4%)	12468 (80.9%)
1	277 (37.7%)	156 (11.7%)	148 (18.4%)	1163 (9.3%)	1744 (11.3%)
2	152 (20.7%)	50 (3.8%)	57 (7.1%)	381 (3%)	640 (4.2%)
3	76 (10.3%)	20 (1.5%)	26 (3.2%)	152 (1.2%)	274 (1.8%)
4	46 (6.3%)	11 (0.8%)	6 (0.8%)	68 (0.5%)	131 (0.9%)
5 or more	62 (8.4%)	10 (0.8%)	22 (2.7%)	64 (0.5%)	158 (1%)

Rates of HFH in patients with different modes of CV death are given in Table 3-4. Rates of HFH were highest in patients who went on to die of progressive heart failure. The rate for sudden death, other CV death and non-CV death was lower, with the lowest rate in patients who were alive or censored at trial end. The pattern was similar in the sensitivity analysis excluding HF admissions where the patients did not survive to discharge (Table 3-4) with a reduction in rate in the patients with progressive heart failure as the cause of death. Rates of events

increased with the addition of outpatient worsening events, but the pattern was similar across groups by type of CV death (Table 3-4).

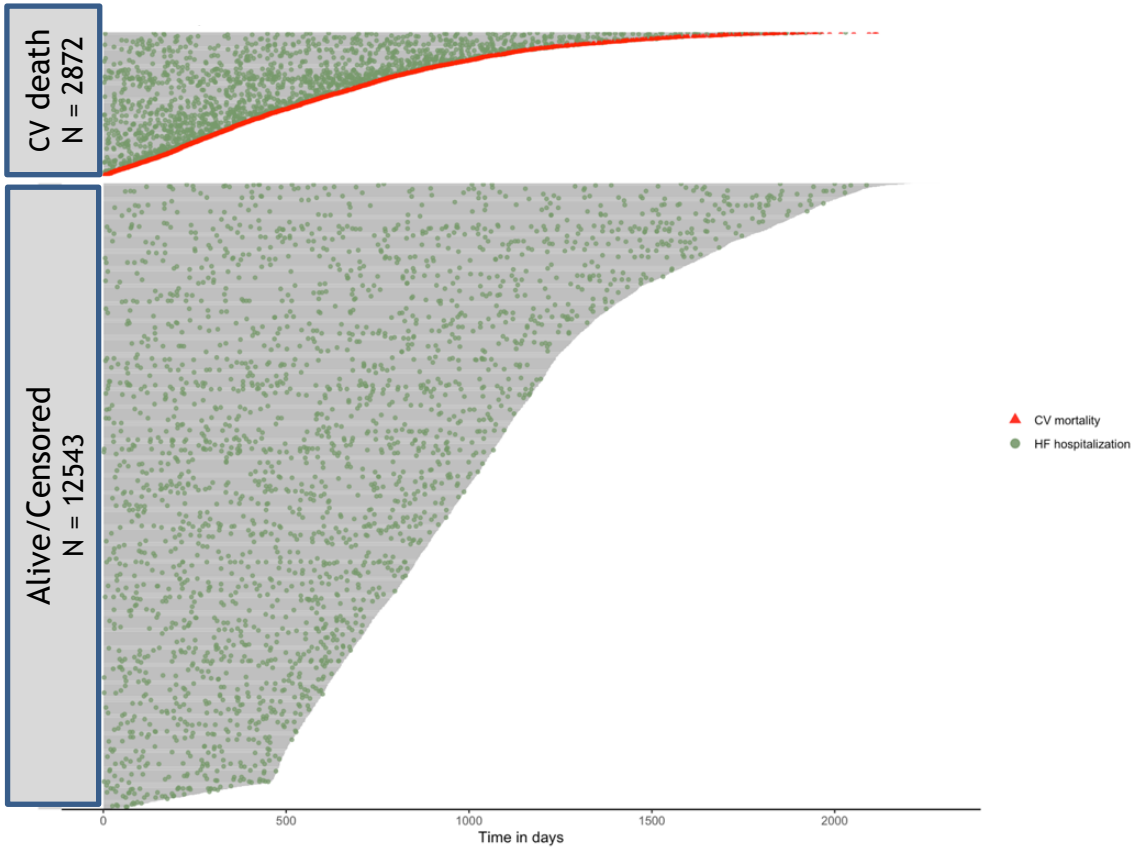
Table 3-4. Rate of HFH by mode of CV death (HFREF)
Rate of heart failure hospitalisations by different causes of CV death. *Admissions excluded where the patient was not discharged alive. § ER visits for HF and IV treatment for HF included. † Includes non-CV deaths, alive at trial end and otherwise censored patients.

Mode of CV death	Rate (per 100 patient years)	95% confidence interval
Pump failure death	92.6	87.5 – 98.0
Sudden death	16.9	15.2 – 18.8
Other CV death	24.7	22.4 – 27.3
Non-CV death	13.2	11.2 – 15.6
Alive/censored at trial end	6.5	6.3 – 6.8
Sensitivity analysis *		
Pump failure death	67.9	63.5 – 72.5
Sudden death	16.8	15.1 – 18.7
Other CV death	23.2	21.0 – 25.7
Alive/censored at trial end †	6.7	6.4 – 7.0
Non hospitalized HF events added to all HF hospitalisations §		
Pump failure death	110.2	104.7 – 116.1
Sudden death	21.6	19.6 – 23.7
Other CV death	31.0	28.4 – 33.8
Alive/censored at trial end †	8.7	8.4 – 9.0

Recurrent event plots stratified by occurrence of CV death suggest a greater number of HF hospitalisations in patients who go on die of CV causes and is suggestive of a greater density of hospitalisations close to the time of death (Figure 3-3). Examining different types of CV death separately suggests greatest density of HFH in patients who die of progressive heart failure with a concentration of admissions near the time of death (Figure 3-3). The appearance is similar in the sensitivity analysis (Figure 3-4).

Figure 3-3. Recurrent event plots (HFREF)
Recurrent event plots split by (A) occurrence of CV death and (B) by difference modes of CV death.

A



B

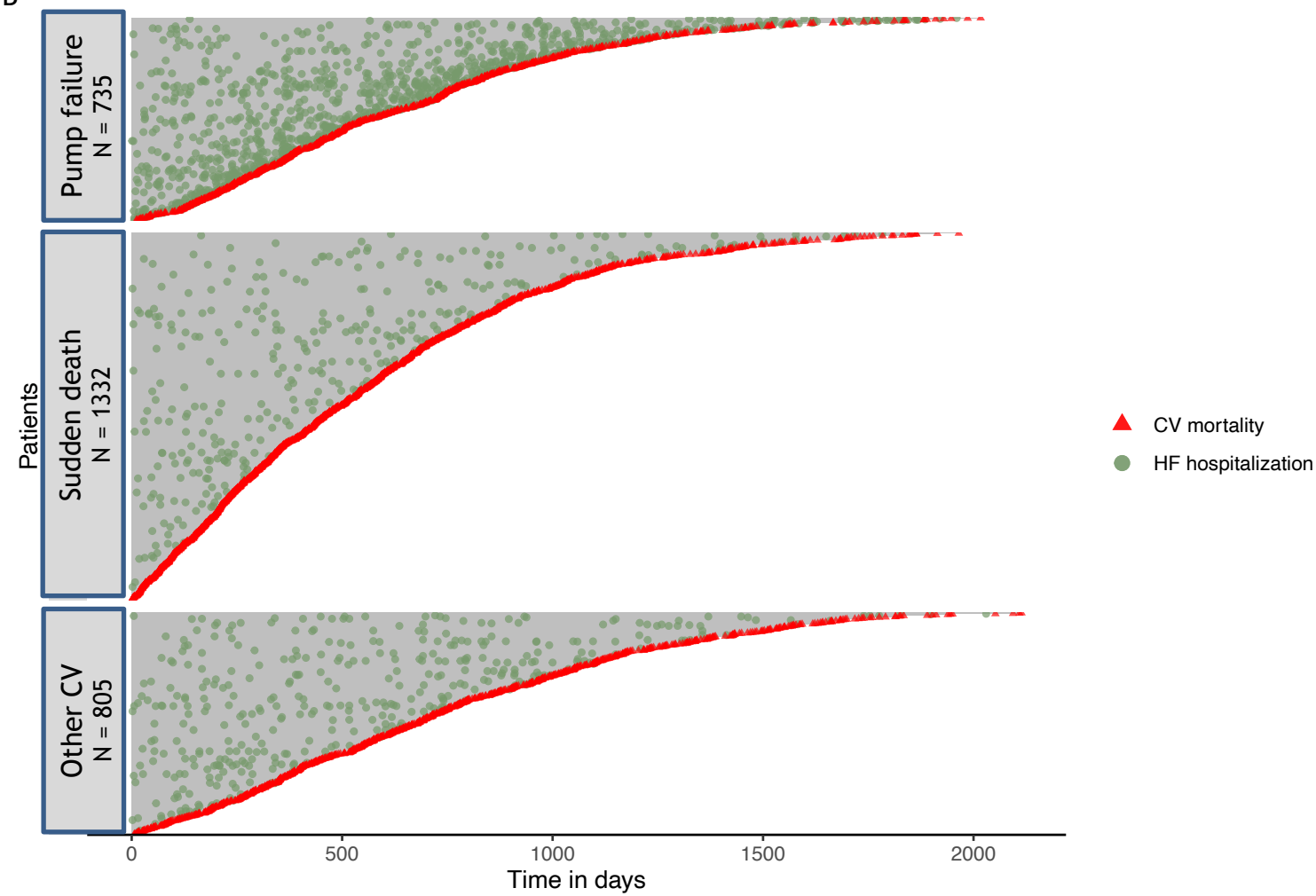
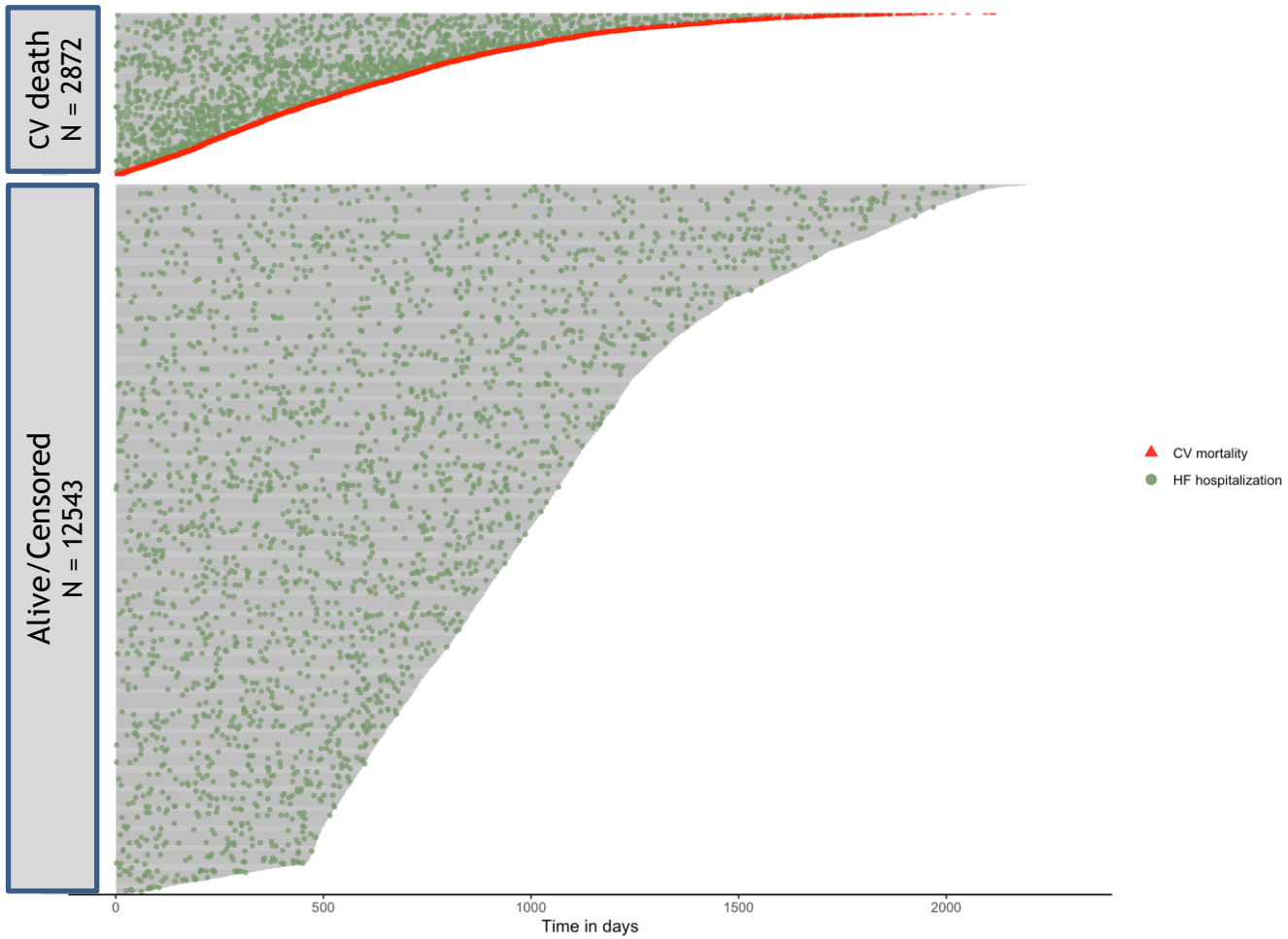

Pump failure
N = 735
Sudden death
N = 1332
Other CV
N = 805
Patients
0
500
1000
1500
2000
Time in days
CV mortality
HF hospitalization

Figure 3-4. Recurrent event plot in sensitivity analysis (HFREF)
Recurrent event plot for the sensitivity analysis.



3.3.2 Analysis of days between heart failure hospitalisations

After identifying dates of HFH prior to trial randomisation then selecting patients with at least 2 known HF admissions gave a population of 1928 patients with 5424 admissions. The demographics of these patients are given in Table 3-5. They are similar to patients in the main analysis who have at least 1 hospitalisation during trial follow up.

Table 3-5. Baseline data in patients with at least 2 HFH (HFREF)
Patient demographics in those with at least 2 HF admission identified (including pre-randomisation admissions). Data are presented as mean ± SD or median (IQR) for continuous measures, and n (%) for categorical measures.

	Total N=1,928
Age (years)	64.0±11.7
Female sex (%)	328 (17.0)
Race or ethnic group (%)	
Caucasian	1,285 (67.1)
Black	89 (4.6)
Asian	428 (22.4)
Other	112 (5.9)
Region (%)	
North America	148 (7.7)
Latin America	219 (11.4)
Western Europe	509 (26.4)
Central Europe	602 (31.2)
Asia/Pacific	450 (23.3)
Systolic blood pressure (mmHg)	120.9±17.4
Heart rate (beats/min)	73.9±13.1
Body mass index (kg/m^2)	28.3±5.8
eGFR (mL/min/1.73m^2)	36.9±38.9
Ischaemic cardiomyopathy (%)	1,010 (52.4)
Left ventricular ejection fraction (%)	28.0±6.3
Median NT pro BNP (pg/ml)	1988 [1042-4305]
NYHA class group III/IV (%)	878 (45.5)
Hypertension (%)	1,314 (68.2)
Diabetes mellitus (%)	734 (38.1)
Atrial fibrillation (%)	819 (42.5)
Loop diuretic (%)	1,718 (89.1)
Digitalis (%)	731 (37.9)
Beta-blocker (%)	1,739 (90.2)
Mineralocorticoid antagonist (%)	980 (50.8)
Implantable cardioverter-defibrillator (%)	407 (21.1)
Cardiac resynchronization therapy (%)	199 (10.3)

The median number of days between each HFH hospitalisation was cross tabulated with cause of CV death (Table 3-6). In all groups, including patients who do not experience CV death, time between subsequent admissions progressively shortens. In patients who go on to die of progressive heart failure, more have recurrent admissions and time between admissions is shorter than all other groups.

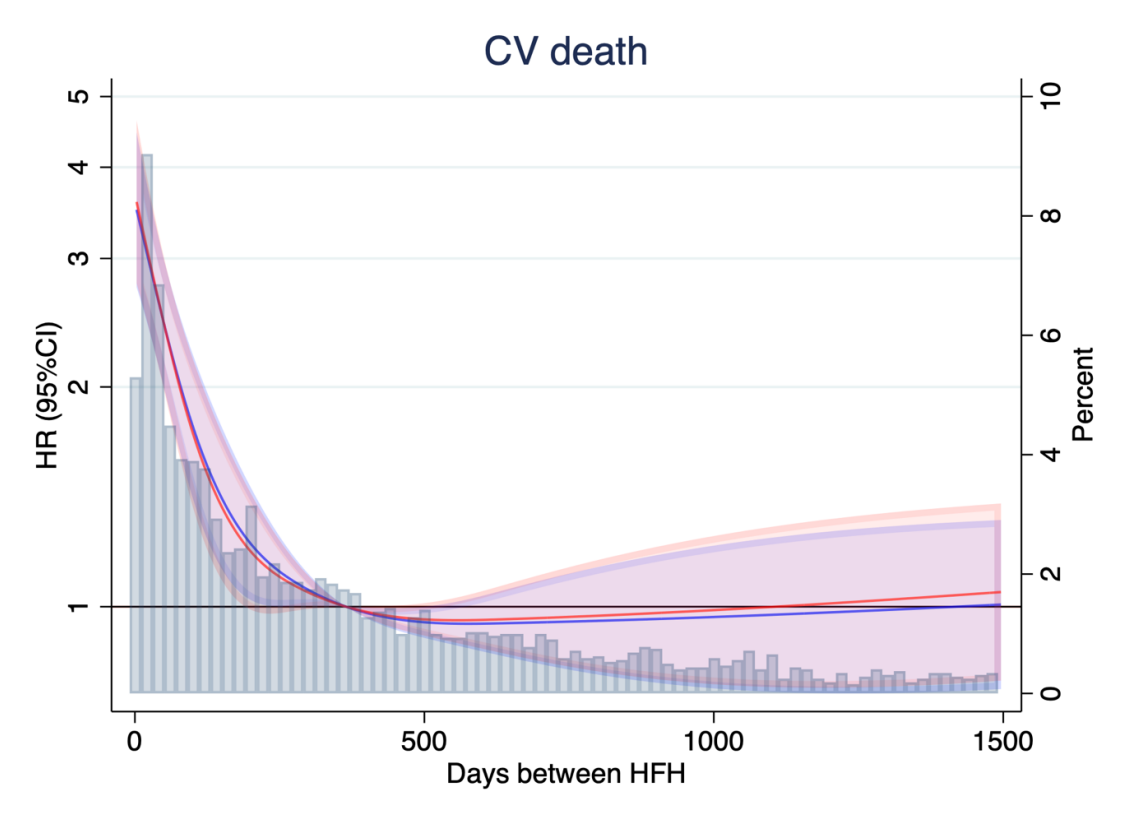
Table 3-6. Median days between adjacent HFH in different modes of CV death (HFREF) Median number of days between subsequent HF hospitalisations stratified by mode of CV death.

		Days from HFH 1-2	Days from HFH 2-3	Days from HFH 3-4	Days from HFH 4-5
Pump failure death	Number (%)	469	239 (51)	118 (25)	72 (16)
N = 469	Median (IQR)	547 (234-1058)	84 (28-197)	49 (22-133)	42.5 (17-134)
Sudden death	Number (%)	167	63 (38)	22 (13)	11 (7)
N = 167	Median (IQR)	487 (212-895)	112 (34-285)	72.5 (43-98)	89 (42-343)
Other CV death	Number (%)	174	72 (41)	36 (21)	21 (12)
N = 174	Median (IQR)	532.5 (282-966)	115 (39-280)	106.5 (34-211)	32 (19-84)
No CV death	Number (%)	1118	390 (35)	173 (15)	83 (7)
N = 1118	Median (IQR)	613 (304-1156)	176 (66-340)	119 (38-291)	98 (27-216)

The number of days between HF hospitalisations was considered as a time updating variable in a Cox regression, with the value updated at each HF admission. Outcomes examined were CV death, all-cause mortality and non-CV mortality. The HR for an increase in 1 month in time between admissions (adjusted for randomised treatment only) for cardiovascular death was 0.99 (95% CI 0.99 – 1.00, p = 0.001). The time updating variable was examined as a restricted cubic spline (Figure 3-5). The baseline spline has adjustment for randomised treatment only. The adjusted spline includes adjustment for randomised treatment, (log transformed) NT-proBNP, age, sex, race, region,

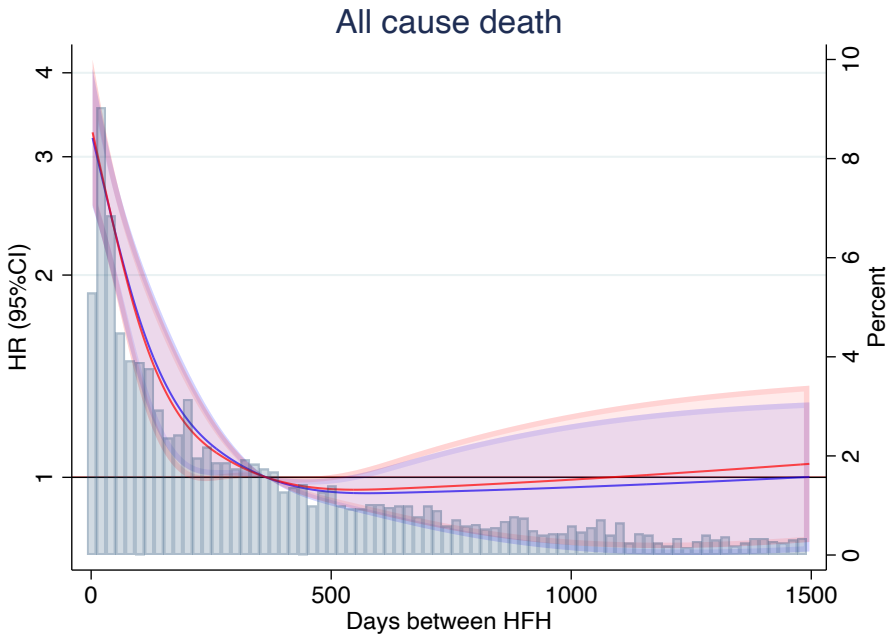
systolic blood pressure, heart rate, body mass index, eGFR, ejection fraction, diabetes and atrial fibrillation. The shorter the time between admissions, the greater the hazard of CV death, with a plateau in risk at around 1 year (the reference point for the spline). This is unchanged with adjustment for the variables listed above. The pattern was similar for all-cause mortality, however the degree of increased risk at was attenuated; there was no significant relationship between time between HF hospitalisation and non-CV mortality (Figure 3-6).

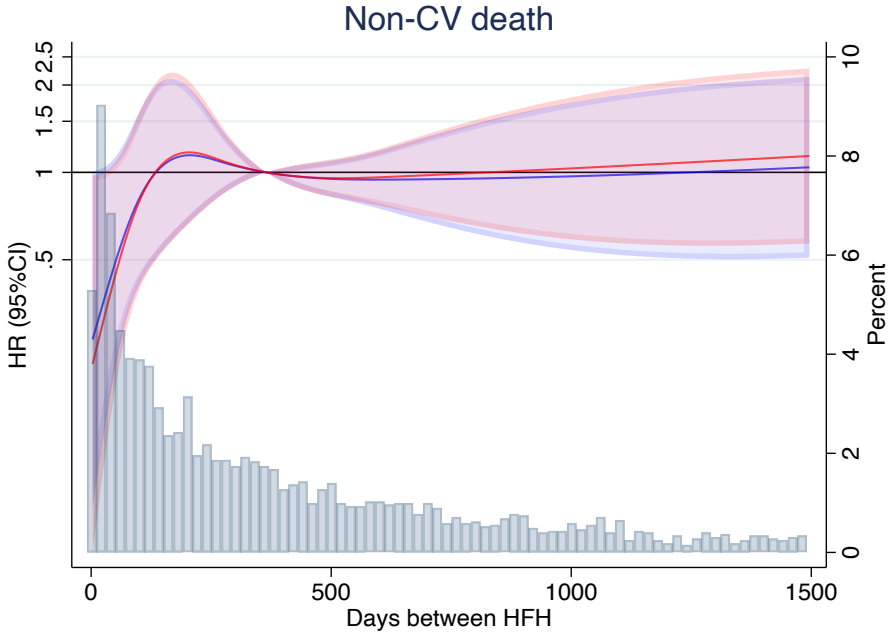
Figure 3-5. Restricted cubic spline of number of days between HFH and risk of CV death (HFREF)
Restricted cubic spline examining days between HFH as a time varying covariate (includes pre-randomisation HF hospitalisations) with the relative HR for cardiovascular death on the left axis. The bar chart (right axis) gives the percent of patients with the corresponding number of days between admissions to show distribution.



———	adjusted for randomised treatment only
———	adjusted for randomised treatment, (log transformed)NT-proBNP, age, sex, race, region, systolic blood pressure, heart rate, body mass index, eGFR, ejection fraction, history of diabetes and atrial fibrillation

Figure 3-6. Restricted cubic splines for all cause mortality and non-CV mortality (HFREF) Restricted cubic spline examining days between HFH as a time varying covariate (includes pre-randomisation HF hospitalisations) and hazard of all-cause mortality and non-CV mortality.





———	adjusted randomised treatment only
———	adjusted for randomised treatment, (log transformed)NT-proBNP, age, sex, race, region, systolic blood pressure, heart rate, body mass index, eGFR, ejection fraction, history of diabetes and atrial fibrillation

3.3.3 KCCQ

KCCQ-CSS over time in groups of patients with different number of HF hospitalisations is given in Figure 3-7. Patients with higher number of hospitalisations have a greater decline in KCCQ over follow up (Table 3-7).

Figure 3-7. Change in KCCQ score over time according to number of HFH (HFREF)
Change in KCCQ-CSS over time in groups of patients by number of HF hospitalisations during the trial.

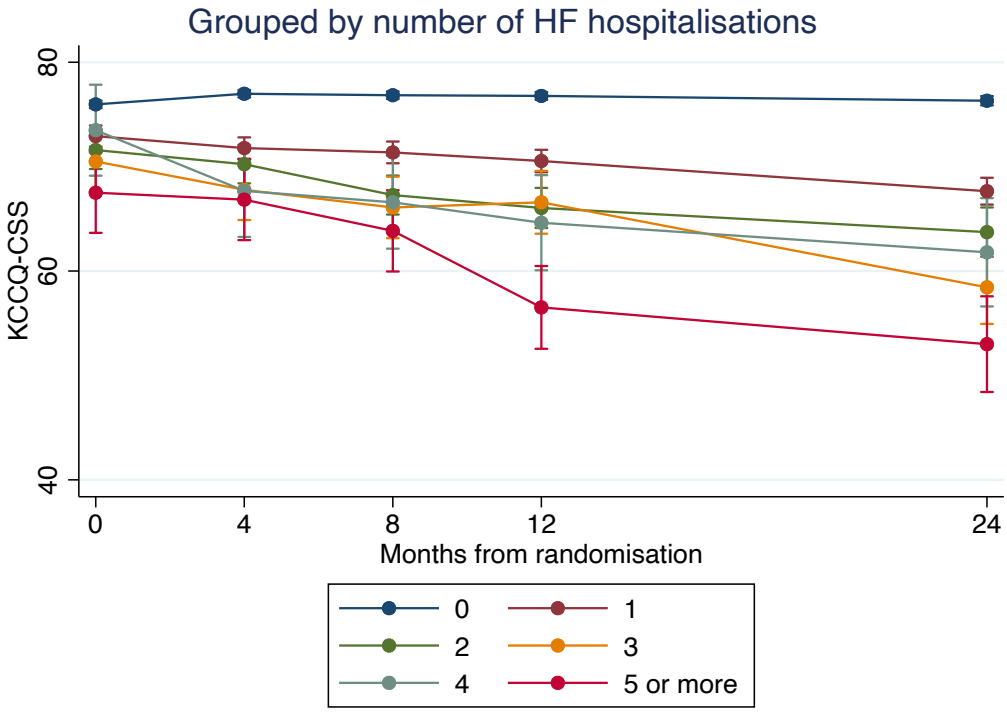


Table 3-7 Slope of change in KCCQ by number of HFH (HFREF)
Slope in KCCQ-CSS over time in patients grouped by number of HFH.

No. HF admissions	Slope (95%CI)	P for difference between slopes
0	0.01 (-0.01-0.02)	<0.001
1	-0.21 (-0.26--0.17)	
2	-0.36 (-0.45--0.28)	
3	-0.46 (-0.58--0.34)	
4	-0.46 (-0.64--0.28)	
5 or more	-0.52 (-0.66--0.38)	

Change in KCCQ-CSS in groups of patients by different causes of CV death are given in Figure 3-8 and Table 3-8. Patients who go on to have a CV death attributed to pump failure death have a greater decline in KCCQ score during

follow up. Patients who do not have CV death have the highest and most stable KCCQ score. The slope of decline in sudden death is less than that for other CV deaths.

Figure 3-8. KCCQ over time by mode of CV death (HFREF)
KCCQ-CSS over duration of follow up in patients split by mode of CV death.

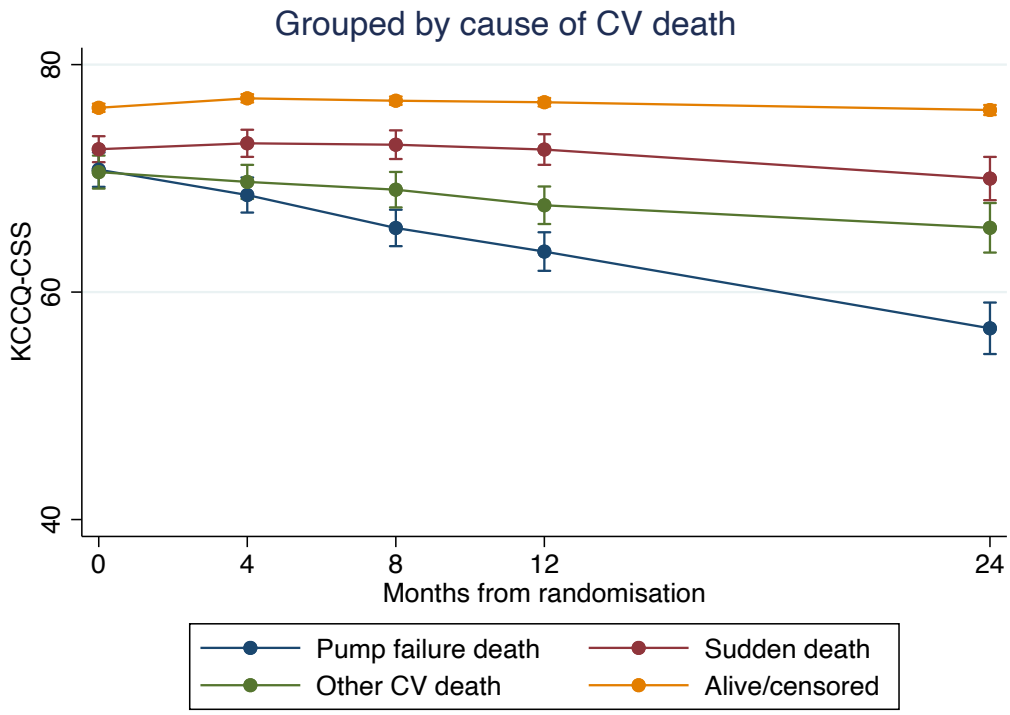


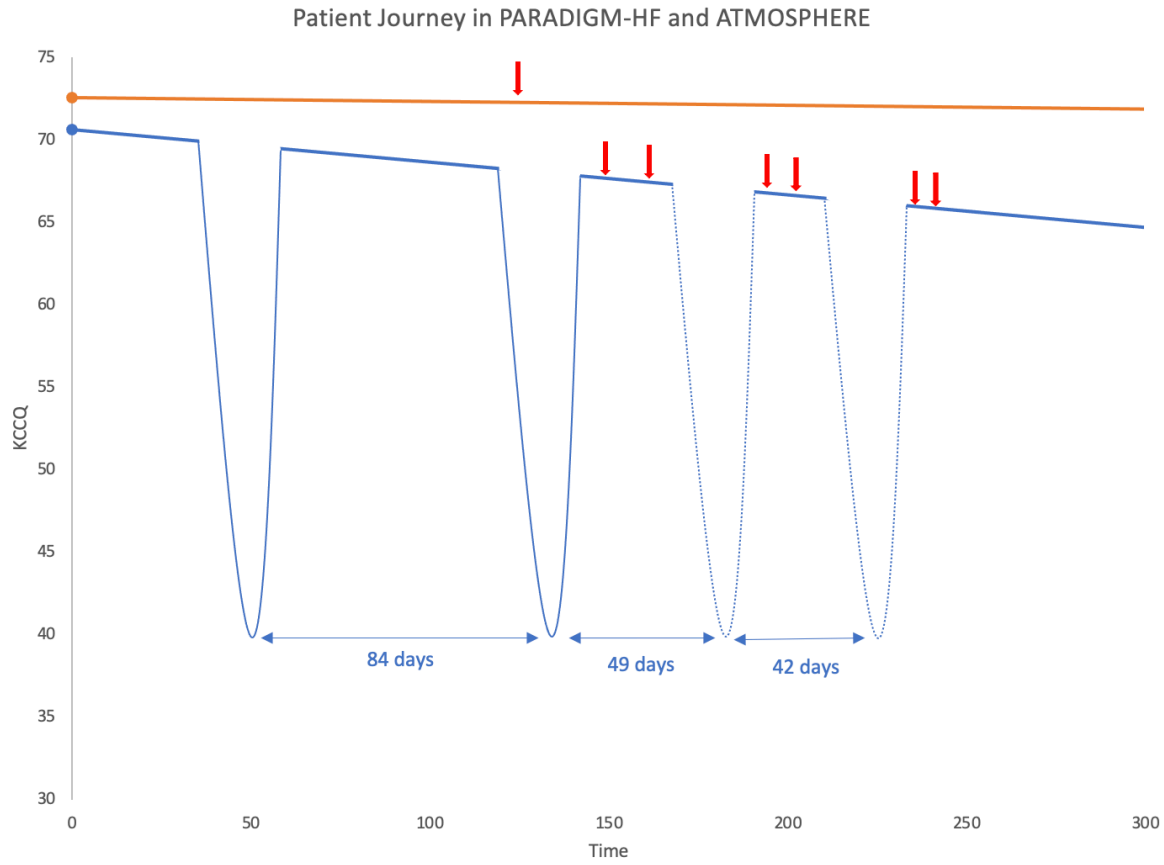
Table 3-8. Slope in KCCQ over time in patients grouped by mode of death (HFREF)
Slope in KCCQ-CSS over time in patients grouped by mode of CV death.

Mode of CV death	Slope (95%CI)	P for difference in slopes
Pump failure death	-0.59 (-0.67--0.51)	<0.001
Sudden death	-0.07 (-0.14--0.01)	
Other CV death	-0.21 (-0.29--0.14)	
No CV death	-0.01 (-0.03-0)	

3.4 Discussion

Despite improvements in heart failure care and reduction in 30-day mortality and readmission rates over time, recurrent hospital admissions are a significant burden.[148] This analysis shows that the accepted patient trajectory in HFREF is true for individuals who die from progressive worsening of heart failure but not for sudden death where only a minority of patients experience preceding HF hospitalisation. With increasing numbers of HFH, patients were more likely to die of progressive heart failure and less likely to have a sudden CV death. Patients with pump failure as the cause of death had a steeper decline in symptom scores over follow up. These results are summarised graphically in Figure 3-9.

Figure 3-9. Visual summary of the patient journey in patients with HFREF
The KCCQ symptom score declines more steeply in patients with pump failure death (blue line). The trajectory is less steep in patients with sudden death (orange). Heart failure hospitalisations are depicted as worsening in KCCQ score and occur more commonly in patients with pump failure death. These trend to fall closer together with recurrent hospitalisations. Red arrows depict patient deaths, which can occur at any stage in patients who die suddenly but are more common (higher risk) after heart failure hospitalisation.



For patients admitted with decompensation of heart failure, particularly clustered admissions, a priority should be up-titration of HF therapies. Conventional sequencing of HF treatment titration in a euvolemic patient may take several months.[1,149] More recently quicker sequencing has been suggested with the focus on initiation of multiple therapies in the first four weeks with up titration of doses later, alternatively with aim to initiate evidence based oral medication prior to discharge in patients hospitalised with heart failure.[150-152] These results support the view that HF hospitalisation is a significant and concerning prognostic sign for patients, and in addition to core HF treatments further optimisation aiming to reduce HF hospitalisation including vericiguat, ivabradine, digoxin, hydralazine and isosorbide dinitrate should be considered as well as interventional and device approaches dependent on the cause of HF.[153-156]

Some healthcare systems apply a penalty if a patient is readmitted within 30 days of a HF hospitalisation which naturally places greater interest on how to reduce 30 day readmission rates.[157–161] Potential strategies have included telemonitoring, improving information sharing between primary and secondary care, comprehensive discharge planning with post discharge support and specialized multidisciplinary follow up.[158–161] While no comment is made in this analysis regarding whether readmissions were avoidable, a more patient centred metric may be to flag readmissions over a longer time frame, perhaps even up to six months, as being of particular concern and broaden efforts to enhance outpatient care and reduce rehospitalisation in this population.

Recurrent admissions appear to only highlight a select population of at-risk patients. Almost half of CV deaths are attributed to sudden death, and most of these patients did not have a HF hospitalisation during the trial follow up. Although recurrent admissions, and particularly clustered admissions, can bring a more unwell patient with heart failure to the attention of the treating team it is not as strong an indicator for risk of sudden CV death emphasising the need to optimise medical treatments and device therapies promptly in apparently stable outpatient HF patients.

In the sensitivity analysis, with admissions excluded where the patient was not discharged alive, the number of patients with prior admissions fell, confirming that many patients that die of progressive heart failure are admitted for treatment prior to death. Despite this, when these admissions are excluded the rate of HFH during the trial remains significantly higher than other types of CV death.

Patterns in change in KCCQ score mirror patterns the burden of HF hospitalisations. In patients with higher numbers of HF hospitalisations, patient reported symptom burden increases over time. The change in KCCQ score is notably different when examining groups by type of CV death, with a mirroring of greater symptom burden and greater number of hospitalisations in the pump failure group. Patients with sudden CV death had a lesser decline in symptom score.

3.4.1 Limitations

This study has several limitations. There is difficulty in separating the risk associated with a HF admission at any time and the specific property of the number of days between admissions given the magnitude of increased risk associated with an admission. Information about hospitalisation prior to trial enrolment may not have been recorded accurately.

3.4.2 Conclusions

In conclusion, these findings support the conventional view of the patient journey with heart failure, with decompensations and admissions occurring with increasing frequency as the patient approaches death and with sudden death occurring at any stage. Clustered admissions should be regarded as a marker for a period of increased risk, and taken as an opportunity to optimise HF medications, including those that reduce HF admissions, or to start conversations with the patient about their end-of-life care wishes. Recurrent hospitalisations are not strongly linked to sudden cardiovascular death further reinforcing the need to promptly optimise treatments and device therapy in outpatient HF patients.

Chapter 4 Identifying the features of the patient journey in HFPEF

4.1 Introduction

Heart failure with preserved ejection fraction (HFPEF) is a common disease associated with substantial morbidity and mortality.[162] It is more common in the older population as compared with heart failure with reduced ejection fraction (HFREF), and reflective of this is often associated with a high degree of comorbidity.[163] The number of HF hospitalisations has been shown to be similar in HFPEF and HFREF, with the associated significant burden for both the patient and service provision.[164]

When we consider HFREF and HFPEF and broad phenotypes of heart failure, we consider how they might behave similarly in terms of patterns of hospitalisation and risk of death following these hospitalisations. The proportion of patients with HFPEF that die from cardiovascular causes, sudden cardiovascular death and progressive heart failure is slightly lower than patients with HFREF, with a higher proportion of non-cardiovascular deaths.[165] This likely reflects the higher degree of other co-morbidity, therefore competing risk of other modes of death is higher. However, cardiovascular deaths remain the mode of death in the majority of patients with HFPEF, being the mode of death of 51-60% of deaths in epidemiological studies and around 70% in clinical trials.[165] Therefore, understanding similarities and differences in the patient trajectory between HFREF and HFPEF is of value to help understand risk and prioritise interventions and monitoring.

As discussed in Chapter 2, treatment of HFPEF remains challenging. Recently, SGLT2i have been shown to reduce the combined risk of cardiovascular death and hospitalisation for heart failure in HFPEF, largely driven by reduction in HF hospitalisation.[31,32] Patients with HFPEF may not be under routine outpatient cardiology follow up, as most management (e.g. optimisation of blood pressure control and weight loss) is carried out in primary care. Therefore, if the period after hospitalisation is associated with higher risk of CV death, a hospitalisation episode should be prioritised as a time to add an SGLT2i for a patient with HFPEF. Given the burden of HF hospitalisation in patients with HFPEF and

evidence of treatments that effect rates of HF hospitalisation, understanding patterns of HF hospitalisation and the relationship between clustered admissions and the risk of CV death is of value.

Patients with HFPEF in general have a higher burden of other comorbidity, therefore it could be hypothesised that they would have significant burden of all cause hospitalisation. The proportion of HF hospitalisation compared with all cause hospitalisations can give an indication of how valuable it is to reduce one type of hospitalisation for these multimorbid patients.

To help understand these issues, the number and timing of both HF hospitalisations and all cause hospitalisations in the PARAGON-HF trial in patients with HFPEF were examined to understand the pattern of hospitalisations and the relationship between timing of HF admissions and risk of CV death.[29]

4.2 Methods

4.2.1 Trial population

The design and results of PARAGON-HF have been described in more detail in Chapter 2.[29,126] In summary, patients with heart failure and preserved ejection fraction were randomised to either sacubitril/valsartan or valsartan after an active run-in phase to assess tolerance to both treatments. Key inclusion criteria included left ventricular ejection fraction (LVEF) ≥ 45%, NYHA class II to IV, elevated NT-proBNP (value dependent on history of HF hospitalisation within past 9 months and presence of AF), and evidence of structural heart disease. Exclusion criteria included any previous LVEF <40%; recent acute coronary syndrome, cardiac surgery or percutaneous coronary intervention; acute decompensated heart failure at the time of screening; intolerance of either trial medications or history of angioedema; blood pressure >180mmHg or <110mmHg; eGFR <30 mL/min/1.73m^2; and serum potassium level >5.2mmol/L.

4.2.2 Identification of hospitalisations and patient reported symptom scores

4.2.2.1 In trial heart failure hospitalisations

Information on HF hospitalisations was collected as a trial endpoint. Hospitalisations where the patient died on the same day as admission were excluded. Data on any cause hospitalisation were also collected during follow up but not adjudicated as an endpoint.

The main outcome examined was CV death, with all cause death and non-CV death analysed as secondary outcomes. Mode of CV death was broken down further into heart failure (or "pump failure"), sudden CV death and other.

4.2.2.2 Pre-trial HF hospitalisation

The most recent date of HF hospitalisation was recorded as part of the patient screening visit. To facilitate analysis of time between admissions, date of prior HFH were added to in-trial HFH if recorded with sufficient accuracy (i.e., day, month and year) as well as hospitalisations occurring between screening and randomisation. In the analysis of time between admissions, patients were included if they had at least 2 known heart failure hospitalisations.

4.2.2.3 KCCQ

Symptom burden and functional status was examined using the KCCQ score including use of validated translations. The self-administered assessment was undertaken during treatment run-in, at randomisation, and follow up weeks 16, 32, 48 then every 48 weeks until end of trial.

4.2.3 Statistical analysis

4.2.3.1 In trial hospitalisations

Patient characteristics were examined in patients with no HFH, 1 HFH or more than 1 HFH and compared using ANOVA for continuous variables, chi square test for categorical variables and Kruskal-Wallis for non-normally distributed continuous variables. The number of HFH per patient was calculated and cross tabulated with the cause of death. The Lin, Wei, Ying and Yang (LWYY) method

was used to calculate rates for recurrent HF hospitalisation in groups of patients with different modes of death.[127] Recurrent event plots were used to visualise patterns of HFH, constructed in the same way as in Chapter 3.

4.2.3.2 Time between admissions analysis

Patients were included in this analysis if they had at least 2 recorded HFH (including previously described pre-trial hospitalisations). The median number of days between each HFH hospitalisation was cross tabulated with cause CV death.

Time between admissions was examined as a time updating covariate. Patients entered the analysis on the date of their second hospitalisation. The time varying covariate value started as the time between first and second HFH. At each HFH, the variable updated to the time between the prior admission to active admission. The time updated variable was entered in a restricted cubic spline analysis using 5 knots in a Cox regression for CV mortality, all-cause mortality and non-CV mortality. The baseline spline has adjustment for randomised treatment only. The adjusted spline includes adjustment for randomised treatment, (log transformed) NT-proBNP, age, sex, race, region, systolic blood pressure, heart rate, body mass index, eGFR, ejection fraction, diabetes and atrial fibrillation.

4.2.3.3 KCCQ

A mixed model for repeated measurement was used to examine change in KCCQ over time (adjusted for randomised treatment, number of HF hospitalisations, and interaction of number of HF hospitalisations and visit, with a random intercept and slope per patient). This was repeated with number of hospitalisations replaced with mode of CV death.

4.3 Results

4.3.1 HF hospitalisations

Of the 4796 patients randomised in the trial, 838 (17%) patients had a hospital admission with heart failure during follow up. In total, there were 1487 HF

hospitalisations. 691 patients died, with 416 deemed to be due to cardiovascular causes meaning 60% of all deaths were due to cardiovascular causes. Of the 416 cardiovascular deaths, 154 (37%) were due to sudden cardiac death and 118 (28.6%) were due to progressive heart failure, or "pump failure". The maximum number of HF hospitalisations experienced by a patient was 18.

Table 4-1 describes the baseline characteristics split by number of HF hospitalisations. Patients with multiple hospitalisations were older, more likely to be from North America, had higher systolic blood pressure and heart rate, higher body mass index, lower eGFR, higher NT-proBNP, were more likely to be diabetic, had greater burden of atrial fibrillation, were more likely to have a history of HF hospitalisation, were more likely to be treated with a diuretic and more likely to be prescribed a beta blocker. Ejection fraction was not significantly different between groups.

Table 4-1. Baseline data by number of HFH (HFPEF)
Baseline demographic table by number of heart failure hospitalisations during trial follow up. Continuous variables are expressed as mean ± standard deviation, or median [interquartile range], as appropriate.

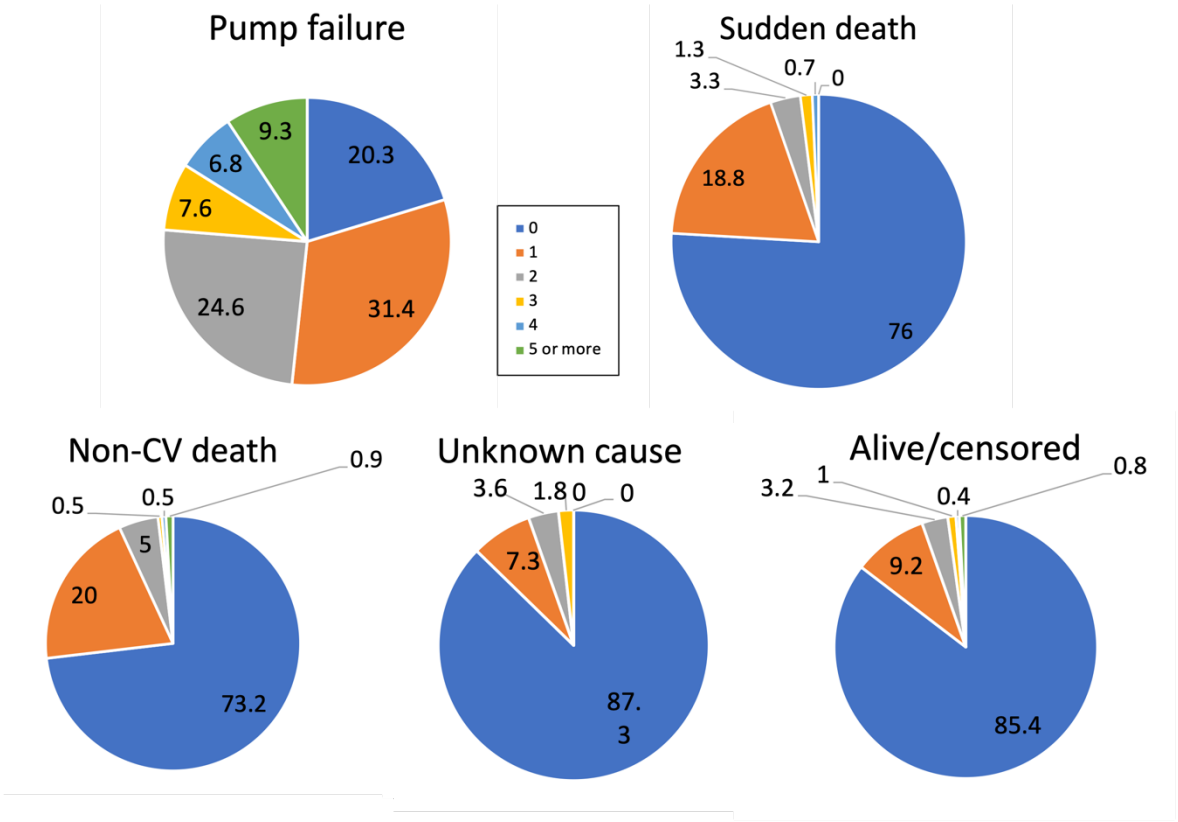
	No HFH N=3,958	1 HFH N=517	≥2 HFH N=321	p-value
Age (years)	72.5±8.4	73.8±8.6	74.3±8.6	<0.001
Female sex (%)	2,062 (52.1)	261 (50.5)	156 (48.6)	0.41
Race (%)				0.001
Asian	499 (12.6)	71 (13.7)	37 (11.5)	
Black or African American	74 (1.9)	13 (2.5)	15 (4.7)	
Other	164 (4.1)	11 (2.1)	5 (1.6)	
White	3,221 (81.4)	422 (81.6)	264 (82.2)	
Region (%)				<0.001
Asia/Pacific and other	618 (15.6)	86 (16.6)	58 (18.1)	
Central Europe	1,490 (37.6)	162 (31.3)	63 (19.6)	
Latin America	334 (8.4)	25 (4.8)	11 (3.4)	
North America	380 (9.6)	86 (16.6)	93 (29.0)	
Western Europe	1,136 (28.7)	158 (30.6)	96 (29.9)	
Systolic BP (mmHg)	130.4±15.2	130.4±15.6	132.7±18.2	0.035
Heart rate (bpm)	70.3±12.2	71.2±12.5	71.4±13.0	0.098
Body mass index (kg/m^2)	30.1±4.9	30.3±5.3	31.2±5.2	<0.001
Creatinine (mg/dL)	1.1±0.3	1.1±0.3	1.2±0.3	<0.001
Estimated glomerular filtration rate (mL/min/1.73m^2)	63.2±19.1	60.3±18.9	58.0±19.1	<0.001
Ischaemic aetiology (%)	1,392 (35.2)	201 (39.0)	130 (40.5)	0.051
Ejection fraction (%)	57.6±7.9	57.4±7.8	56.8±7.8	0.20
NT-proBNP (pg/mL)	861.5 (451.0-1523.5)	1096.0 (536.0-1982.0)	1403.5 (634.0-2558.0)	<0.001
NYHA class III/IV (%)	737 (18.6)	135 (26.1)	79 (24.6)	<0.001
Past medical history				
Hypertension (%)	3,780 (95.5)	496 (95.9)	308 (96.0)	0.85
Diabetes (%)	1,632 (41.2)	242 (46.8)	188 (58.6)	<0.001
Atrial fibrillation (%)	1,249 (31.7)	187 (36.3)	116 (36.2)	0.035
Stroke (%)	403 (10.2)	68 (13.2)	37 (11.6)	0.096
Prior HF hospitalisation (%)	1,738 (43.9)	324 (62.7)	244 (76.0)	<0.001
Myocardial infarction (%)	872 (22.0)	130 (25.1)	81 (25.2)	0.14
Treatments				
Loop diuretic (%)	3,018 (87.1)	438 (84.9)	318 (98.8)	<0.001
ACEi/ARB (%)	3,446 (87.1)	432 (83.6)	261 (81.3)	0.002
MRA (%)	984 (24.9)	161 (31.1)	94 (29.3)	0.003
Beta blocker (%)	3,149 (79.6)	411 (79.5)	261 (81.3)	0.75

The number of HFH was tabulated with mode of death. Due to small numbers, patients with 5 or more admissions were combined into one group. 59% of patients who died of CV causes had no HF admissions, compared with 73% who died of non-CV death and 87% in those with unknown cause of death. Splitting further by mode of CV death, patients who died of progressive HF ("pump failure") generally had at least 1 HFH during the trial with only 20% having no admissions, while of patients dying suddenly 76% had no in trial HF hospitalisations. Of patients alive or censored, 85% had no HF hospitalisations (Table 4-2, Figure 4-1). Examining this another way, of the 245 patients with no HFH who have CV death, 48% die of suddenly and 10% die of pump failure; of 81 patients with two or more HFH who die of CV causes, the proportion having sudden CV death falls to 10%, with 70% dying of progressive heart failure.

Table 4-2. Mode of CV death and number of HFH (HFPEF)
Mode of death cross tabulated with number of HF hospitalisations during trial follow up. Percentages are given within each column.

No. HF admissions	CV death	Pump failure	Sudden death	Non-CV death	Unknown cause	Alive/censored	Total
0	245 (58.9%)	24 (20.3%)	117 (76.0%)	161 (73.2%)	48 (87.3%)	3504 (85.4%)	3958 (82.5%)
1	90 (21.6%)	37 (31.4%)	29 (18.8%)	44 (20.0%)	4 (7.3%)	378 (9.2%)	516 (10.8%)
2	45 (10.8%)	29 (24.6%)	5 (3.3%)	11 (5.0%)	2 (3.6%)	131 (3.2%)	189 (3.9%)
3	13 (3.1%)	9 (7.6%)	2 (1.3%)	1 (0.5%)	1 (1.8%)	41 (1.0%)	56 (1.2%)
4	12 (2.9%)	8 (6.8%)	1 (0.7%)	1 (0.5%)	0 (0%)	18 (0.4%)	31 (0.7%)
5 or more	11 (2.6%)	11 (9.3%)	0	2 (0.9%)	0 (0%)	33 (0.8%)	46 (1.0%)

Figure 4-1. Proportion of patients experiencing multiple HFH (HFPEF)
Proportion of patients who experienced 0, 1, 2, 3, 4 or 5 or more HF hospitalisations by mode of CV death (%).



Rates of HFH by different modes of death showed a higher rate of hospitalisations in patients who die of CV causes (48.3 [95%CI 43.3 - 53.7]) compared with non-CV causes (19.7 [95%CI 15.9 - 24.4]) (Table 4-3). The highest rate of HFH was found in patients who died of pump failure at 99.5 (95% CI 87.2 - 113.6).

Table 4-3. Rate of HFH by mode of CV death (HFPEF)
Rates of HF hospitalisations by different causes of death.

Mode of CV death	Rate (per 100 patient years)	95% confidence interval
CV	48.3	43.4 - 53.7
Pump failure	99.5	87.2 - 113.6
Sudden death	19.4	14.6 - 25.6
Non-CV	19.7	15.9 - 24.4
Unknown	9.5	5.3 - 17.2
Alive/censored at trial end	8.4	7.9 - 8.9

Examining recurrent event plots, there was a higher density of HF hospitalisations in patients who went on to die of cardiovascular causes compared to those who were alive or censored for cardiovascular death (Figure 4-2). Further breaking down different types of CV death, patients with pump failure death had a higher concentration of HF hospitalisations compared to those who died suddenly or from other cardiovascular causes (Figure 4-3).

Figure 4-2. Recurrent event plots (HFPEF)
Recurrent event plots split by occurrence of CV death.

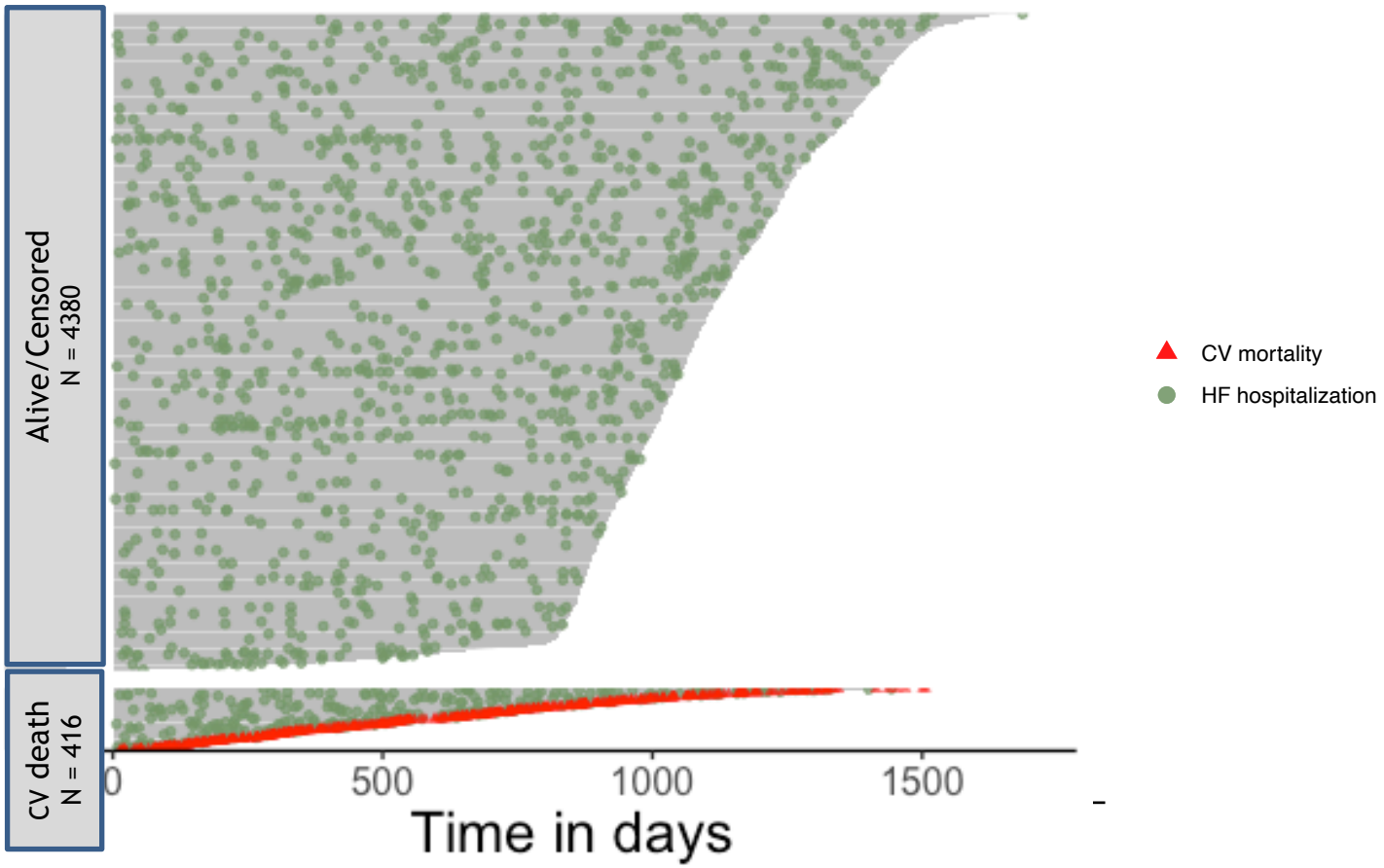
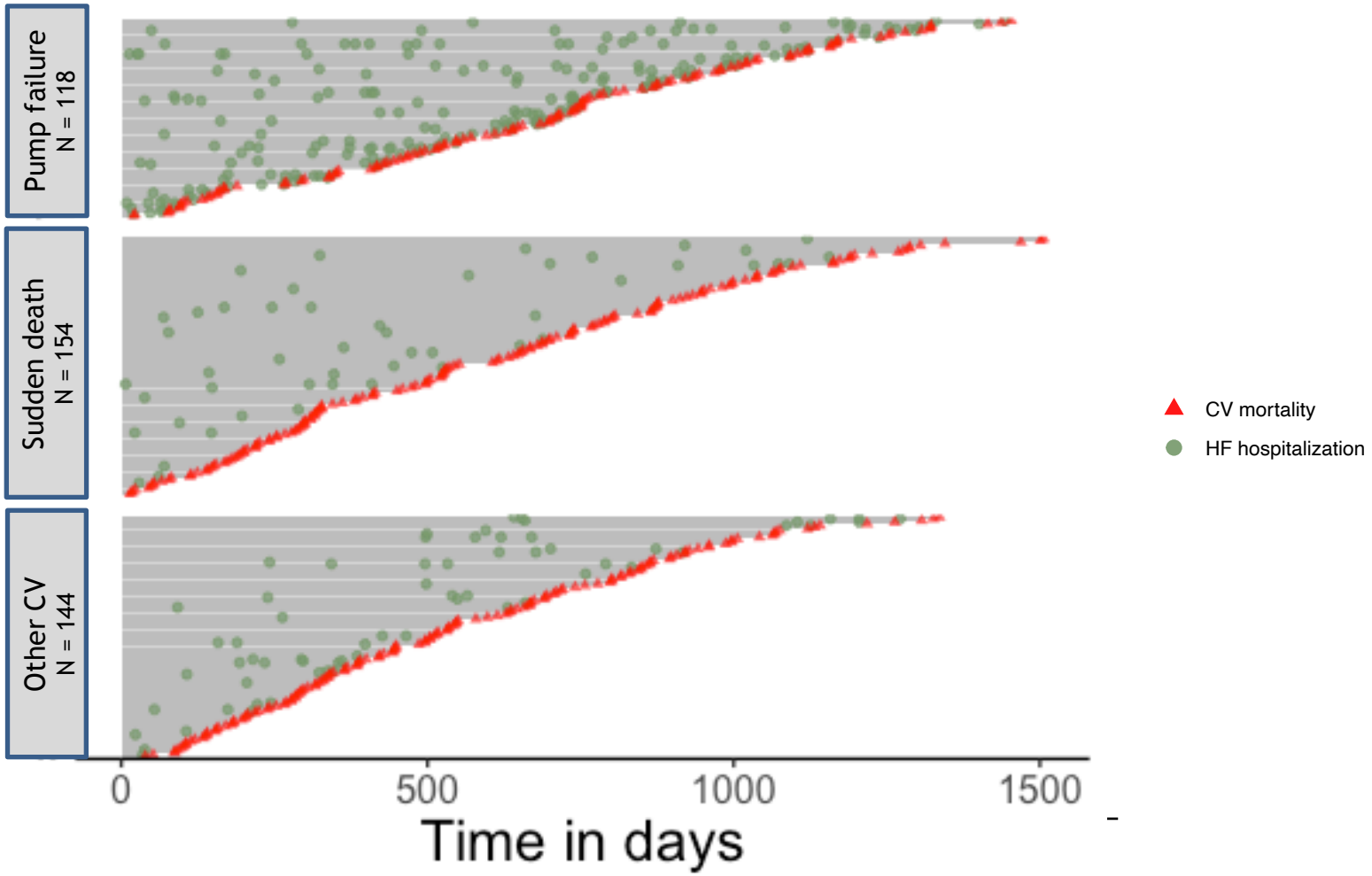


Figure 4-3. Recurrent event plots in different modes of CV death (HFPEF)
Recurrent event plots split by difference modes of death.



4.3.1.1 Analysis of days between heart failure hospitalisations

For analysis of time between admissions, pre-randomisation hospitalisations were added and patients with at least 2 recorded HF hospitalisations were included, giving a population of 643 patients with 1859 hospitalisations. The baseline characteristics are given in Table 4-4.

Table 4-4. Baseline data in patients with at least 2 HFH (HFPEF)
Baseline characteristics in patients with at least 2 recorded HFH for analysis of days between HF admissions. Data are presented as mean ± SD or median (IQR) for continuous measures, and n (%) for categorical measures.

	Total N=643
Age (years)	73.6±8.8
Female sex (%)	320 (49.8)
Race (%)	
Asian	85 (13.2)
Black or African American	26 (4.0)
Other	13 (2.0)
White	519 (80.7)
Region (%)	
Asia/Pacific and other	119 (18.5)
Central Europe	175 (27.2)
Latin America	25 (3.9)
North America	141 (21.9)
Western Europe	183 (28.5)
Systolic Blood Pressure (mmHg)	131.5±17.0
Heart rate (bpm)	71.5±12.8
Body mass index (kg/m^2)	30.7±5.3
Creatinine (mg/dL)	1.1±0.3
Estimated glomerular filtration rate (mL/min/1.73m^2)	59.7±19.3
Ischaemic aetiology (%)	260 (40.4)
Ejection fraction (%)	56.9±7.8
NT-proBNP (pg/mL)	1216.5 (577.0-2180.0)
NYHA class III/IV (%)	170 (26.4)
Past medical history	
Hypertension (%)	616 (95.8)
Diabetes (%)	347 (54.0)
Atrial fibrillation (%)	242 (37.8)
Stroke (%)	84 (13.1)
Prior HF hospitalisation (%)	566 (88.0)
Myocardial infarction (%)	156 (24.3)
Treatments	
Loop diuretic (%)	582 (90.5)
ACEi/ARB (%)	532 (82.7)
MRA (%)	203 (31.6)
Beta blocker (%)	517 (80.4)

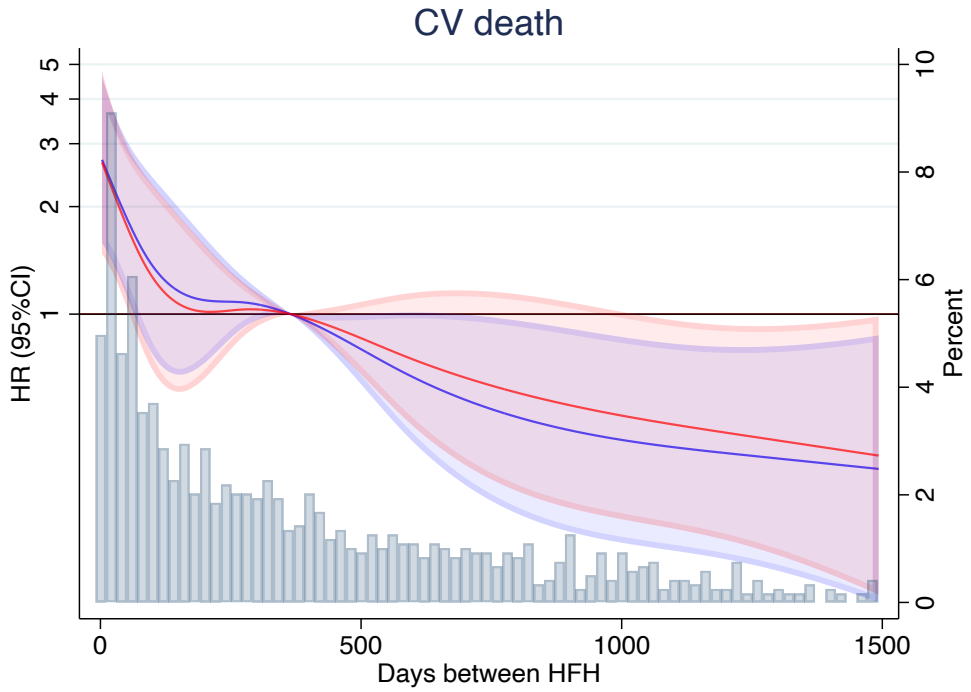
The median number of days between subsequent HF hospitalisations was cross tabulated with cause of death. For patients who died of CV causes, the time between hospitalisations tended to shorten with recurrent admissions. For patients who died of other causes or were censored, the time between admissions also trended downwards but was much longer than those who died of CV causes with around a 100-day difference between groups (Table 4-5). This was not further split into mode of CV death, as in the analysis of HFREF, due to smaller numbers.

Table 4-5. Median days between adjacent HFH in CV death (HFPEF)
Median days between subsequent HF hospitalisations in patients with CV death during follow up and those censored for CV death (other cause death, end of follow up or censor for other reasons).

		Days from HFH 1-2	Days from HFH 2-3	Days from HFH 3-4	Days from HFH 4-5
CV death	Number	138	68	34	21
N = 138	Median (IQR)	414 (217-772)	82 (33-237)	58 (26–157)	47 (18–130)
Censored for CV death	Number	505	199	88	48
N = 505	Median (IQR)	577 (273–952)	176 (57-394)	155 (55-257)	133 (23–282)

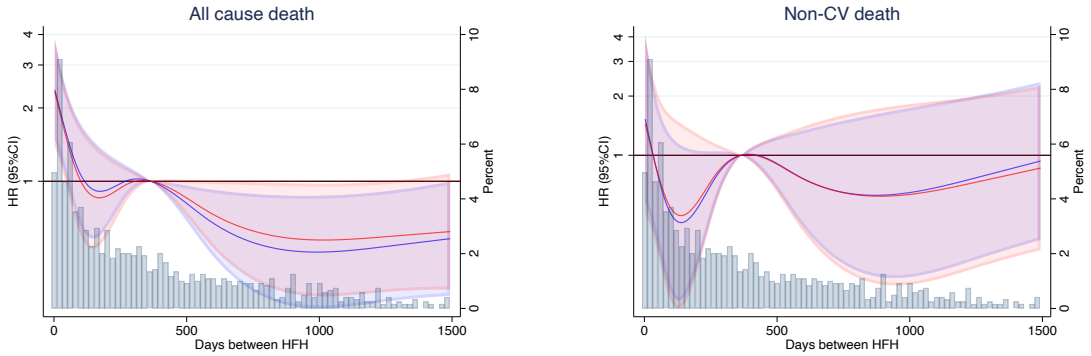
The number of days between HF hospitalisations was considered as a time updating variable in a Cox regression, with the value updated at each HF admission. Outcomes examined were CV death, all-cause mortality and non-CV mortality. The time updating variable was examined as a restricted cubic spline (Figure 4-4). At a shorter time between admissions, there is an increased hazard of cardiovascular death; the 95% confidence interval first falls above 1 at 64 days. The shorter the time between admissions, the greater the hazard of CV death, with a plateau in risk at around 1 year (the reference point for the spline). This is similar with adjustment for the variables listed above. The pattern was similar for all-cause mortality, however the degree of increased risk was attenuated; there was no significant relationship between time between HF hospitalisation and non-CV mortality (Figure 4-5).

Figure 4-4. Restricted cubic spline of number of days between HFH and risk of CV death (HFPEF)
Restricted cubic spline examining days between HFH as a time varying covariate (includes pre-randomisation HF hospitalisations) with the relative HR for cardiovascular death on the left axis. The bar chart (right axis) gives the percent of patients with the corresponding number of days between admissions to show distribution.



———	adjusted for randomised treatment only
———	adjusted for randomised treatment, (log transformed)NT-proBNP, age, sex, race, region, systolic BP, HR, BMI, eGFR, ejection fraction, history of diabetes and AF

Figure 4-5. Restricted cubic splines for all cause mortality and non-CV mortality (HFPEF) Restricted cubic spline examining days between HFH as a time varying covariate (includes pre-randomisation HF hospitalisations) and hazard of all-cause mortality and non-CV mortality.



	adjusted for randomised treatment only
	adjusted for randomised treatment, (log transformed)NT-proBNP, age, sex, race, region, systolic BP, HR, BMI, eGFR, ejection fraction, history of diabetes and AF

1.1.1.1 KCCQ

KCCQ over the course of follow up in groups of patients defined by number of heart failure hospitalisations is given in Figure 4-6 and Table 4-6. Patients with greater number of HF hospitalisations had a greater decline in KCCQ score over time. KCCQ grouped by cause of death is given in Figure 4-7 and Table 4-7, patients who survived or were censored had a higher baseline KCCQ with lesser decline compared with all causes of death, where the decline was similar in all groups.

Figure 4-6. Change in KCCQ score over time according to number of HFH (HFPEF)
Change in KCCQ-CSS over time in groups of patients by number of HF hospitalisations during the trial.

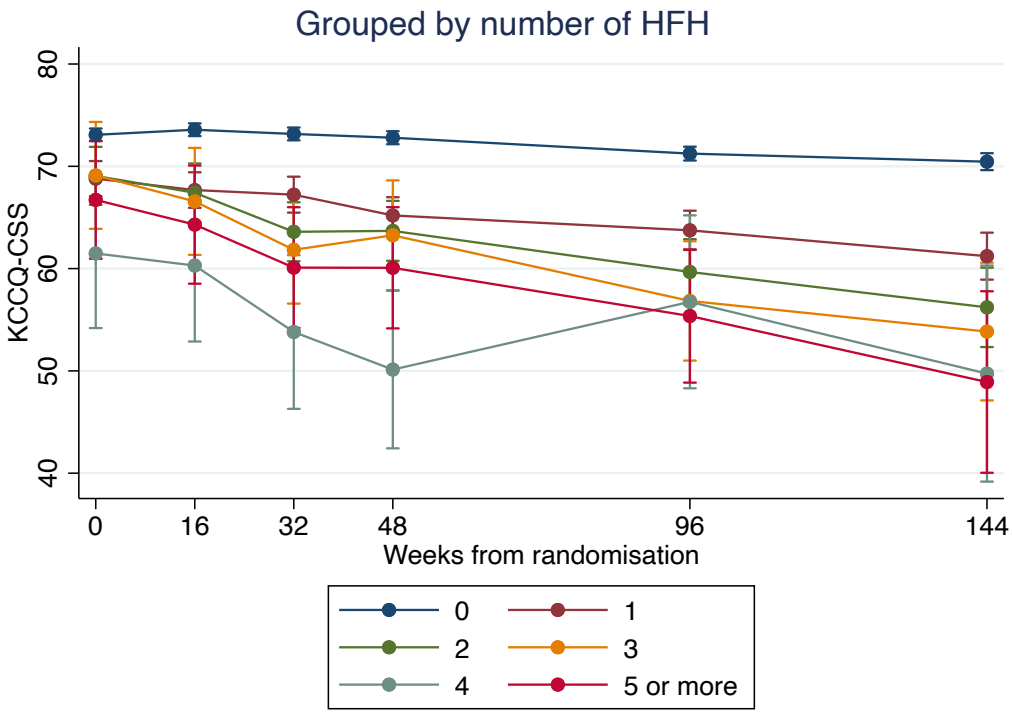


Table 4-6. Slope of change in KCCQ by number of HFH (HFPEF)
Slope in KCCQ over time in patients grouped by number of HFH.

No. HF admissions	Slope (95%CI)	P for difference between slopes
0	-0.04 (-0.04, -0.03)	<0.001
1	-0.09 (-0.11, -0.07)	
2	-0.15 (-0.19, -0.12)	
3	-0.18 (-0.25, -0.12)	
4	-0.12 (-0.22, -0.02)	
5 or more	-0.20 (-0.28, -0.12)	

Figure 4-7. KCCQ over time by cause of death (HFPEF)
Change in KCCQ-CSS over time in groups of patients by mode of death.

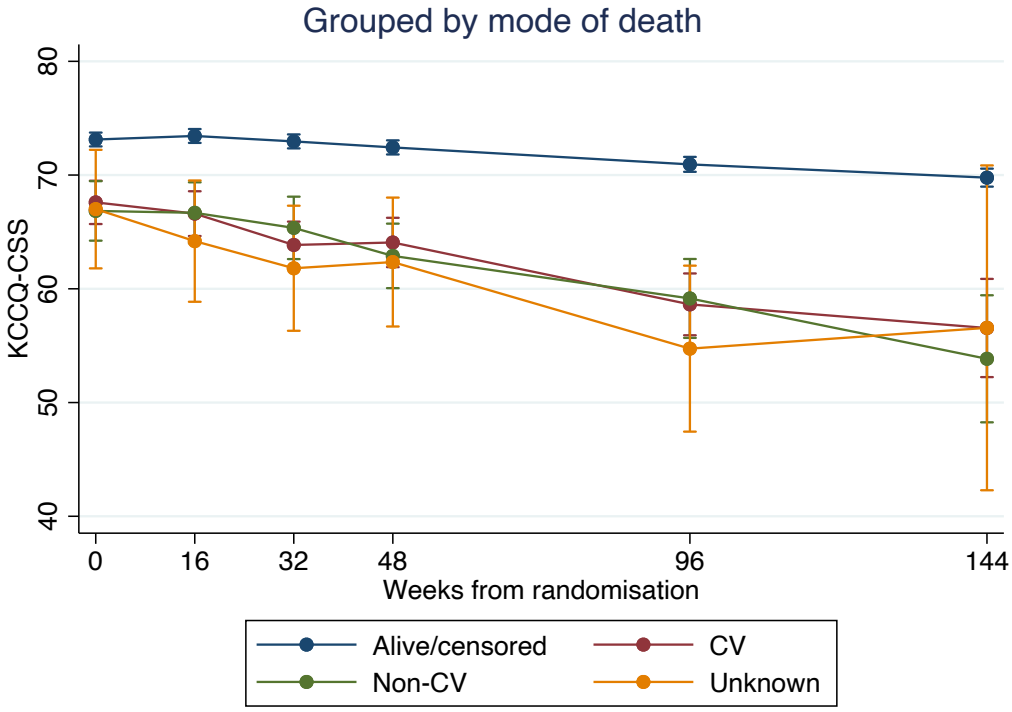


Table 4-7. Slope in KCCQ over time in patients grouped by mode of death (HFPEF)
Slope in KCCQ over time in patients grouped by mode of death.

Mode of death	Slope (95%CI)	P for difference between slopes
Alive/censored	-0.04 (-0.05, 0.04)	<0.001
CV death	-0.15 (-0.18, -0.11)	
Non-CV	-0.15 (-0.20, -0.10)	
Unknown	-0.18 (-0.28, 0.08)	

4.3.2 All cause hospitalisations

When all hospitalisations are included, 2648 patients have a hospitalisation, with 4800 total admissions.

Baseline characteristics split by number of all cause hospitalisations is given in Table 4-8. Like HF hospitalisations, patients with multiple hospitalisations were older and were more likely to be from North America. They have lower eGFR, higher ejection fraction and higher NT-proBNP. Those with more admissions were more likely to be diabetic.

Table 4-8. Baseline data by number of all cause hospitalisations (HFPEF)
Baseline demographics by number of all cause hospitalisation during trial follow up. Data are presented as mean ± SD or median (IQR) for continuous measures, and n (%) for categorical measures.

	0 hospital admissions N=2,148	1 hospital admission N=1,546	≥ 2 hospital admissions N=1,102	p-value
Age (years)	71.3±8.5	73.7±8.4	74.2±8.0	<0.001
Female sex (%)	1,103 (51.4)	825 (53.4)	551 (50.0)	0.21
Race (%)				<0.001
Asian	251 (11.7)	202 (13.1)	154 (14.0)	
Black or African American	39 (1.8)	30 (1.9)	33 (3.0)	
Other	105 (4.9)	51 (3.3)	24 (2.2)	
White	1,753 (81.6)	1,263 (81.7)	891 (80.9)	
Region (%)				<0.001
Asia/Pacific and other	304 (14.2)	247 (16.0)	211 (19.1)	
Central Europe	941 (43.8)	506 (32.7)	268 (24.3)	
Latin America	224 (10.4)	107 (6.9)	39 (3.5)	
North America	150 (7.0)	196 (12.7)	213 (19.3)	
Western Europe	529 (24.6)	490 (31.7)	371 (33.7)	
Systolic Blood Pressure (mmHg)	130.8±14.6	130.6±15.7	130.1±16.8	0.52
Heart rate (bpm)	70.0±11.9	70.9±12.5	70.6±12.7	0.081
Body mass index (kg/m^2)	30.1±4.9	30.2±5.1	30.5±5.0	0.062
Creatinine (mg/dL)	1.1±0.3	1.1±0.3	1.1±0.3	<0.001
Estimated glomerular filtration rate (mL/min/1.73m2)	64.8±18.9	61.5±19.3	59.8±18.7	<0.001
Ischaemic aetiology (%)	774 (36.0)	533 (34.5)	416 (37.7)	0.23
Ejection fraction (%)	57.3±7.9	57.6±7.9	58.0±7.8	0.049
NT-proBNP (pg/mL)	828.0 [440.0-1496.0]	978.0 [508.0-1770.0]	938.0 [485.0-1650.0]	<0.001
NYHA class III/IV (%)	360 (16.8)	349 (22.6)	242 (22.0)	<0.001
Past medical history				
Hypertension (%)	2,044 (95.2)	1,482 (95.9)	1,058 (96.0)	0.43
Diabetes (%)	823 (38.3)	699 (45.2)	540 (49.0)	<0.001
Atrial fibrillation (%)	668 (31.2)	526 (34.2)	358 (32.7)	0.16
Stroke (%)	203 (9.5)	174 (11.3)	131 (11.9)	0.058
Prior HF hospitalisation (%)	912 (42.5)	790 (51.1)	604 (54.8)	<0.001
Myocardial infarction (%)	475 (22.1)	348 (22.5)	260 (23.6)	0.63
Treatments				
Loop diuretic (%)	1,559 (74.4)	1,252 (81.0)	901 (81.8)	<0.001
ACEi/ARB (%)	1,906 (88.7)	1,313 (84.9)	920 (83.5)	<0.001
MRA (%)	535 (24.9)	430 (27.8)	274 (24.9)	0.097
Beta blocker (%)	1,707 (79.5)	1,242 (80.3)	872 (79.1)	0.71

Tabulating the number of hospitalisations with cause of death, patients with non-CV death mostly had at least one admission, with only 5% having no admissions. In the CV death group 26% had no hospitalisations, 25% in the unknown cause of death group and 49% in the alive/censored patients (Table 4-9). Further breaking down CV death into pump failure and sudden death, 0.9% of patients with pump failure death had no hospital admissions as compared with 55% of those with sudden CV death.

Table 4-9 Mode of CV death and number of all cause hospitalisations (HFPEF)
Mode of death cross tabulated with number of all cause hospitalisations during trial follow up.

Number of hospitalisations	CV death	Pump Failure	Sudden death	Non-CV death	Unknown cause of death	Alive/ censored	Total
0	108 (25.96%)	1 (0.9%)	84 (54.6%)	12 (5.45%)	14 (25.45%)	2014 (49.06%)	2148 (44.79%)
1	184 (44.23%)	57 (48.3%)	48 (31.2%)	104 (47.27%)	28 (50.91%)	1230 (29.96%)	1546 (32.24%)
2	70 (16.83%)	31 (26.3%)	11 (7.1%)	55 (25%)	10 (18.18%)	483 (11.77%)	618 (12.89%)
3	23 (5.53%)	10 (8.5%)	6 (3.9%)	18 (8.18%)	1 (1.82%)	186 (4.53%)	228 (4.75%)
4	18 (4.33%)	8 (6.8%)	5 (3.3%)	15 (6.82%)	1 (1.82%)	86 (2.1%)	120 (2.5%)
5 or more	13 (3.12%)	11 (9.3%)	0	16 (7.27%)	1 (1.82%)	106 (2.58%)	136 (2.84%)

Rates of all cause hospitalisation by different causes of death showed an increased rate of hospitalisations in patients who die of non-CV causes (99.0 [95%CI 90.0 – 108.9]) compared with CV causes (78.5 [95%CI 72.1 – 85.3]) (Table 4-10). Within CV death, the highest rate was in patients with pump failure deaths at 111.7 (95%CI 98.6 – 126.5).

Table 4-10. Rate of all cause hospitalisations by mode of CV death (HFPEF)
Rates of all cause hospitalisations by different modes of death.

All cause hospitalisations		
CV	78.5	72.1 - 85.3
Pump failure	111.7	98.6 – 126.5
Sudden death	42.7	35.3 – 51.5
Non-CV	99.0	90.0 - 108.9
Unknown	53.7	41.8 – 68.8
Alive/censored at trial end	29.9	28.9 – 30.8

4.4 Discussion

The proportion of deaths attributed to cardiovascular causes in the PARAGON-HF trial was similar to that of previous clinical trials and epidemiological studies of HFPEF at around 60%.[165,166] The burden of hospitalisation for patients in HFPEF is high, with 20% having an admission due to heart failure over the course of follow up and 55% having a hospitalisation for any cause. This is slightly different in the EMPEROR-Preserved trial, where 10% of patients had a hospitalisation for heart failure during trial follow up and 90% had a hospitalisation for any reason.[31] The overall rate of HF hospitalisation was 10.7 per 100 patient years in the PARAGON-HF trial. This is comparable to rates in the DELIVER trial; rates for total HFH and CV death are given as 15.3 and 11.8 per 100 years in the placebo and dapagliflozin arms respectively.[167] This reflects the high degree of morbidity in these patients.

Patients with more HF admissions had markers of higher risk for cardiovascular mortality , including higher NT-proBNP, older age, lower eGFR and had greater degree of comorbidity including diabetes and atrial fibrillation. They were more likely to have had a heart failure hospitalisation prior to trial enrolment.

Interestingly, ejection fraction did not vary significantly between groups. Patterns were similar when all cause hospitalisation was considered, with older age and greater comorbidity as well as higher BNP being seen in patients with multiple all-cause admissions.

A greater proportion of patients who died of CV causes had HF hospitalisation during the trial than those who died of non-CV causes (41% and 27% respectively). When examining all cause hospitalisations, a greater proportion of patients with non-CV death had a hospitalisation for any reason (95% for non-CV death, 74% with CV death). This reflects that although HFPEF patients tend to be multimorbid, and are likely to have hospitalisations for many reasons, there remains an underlying link between HF hospitalisations and risk of cardiovascular death. Cardiovascular death remains a common mode of death in patients with HFREF, with 60% percent of deaths in the PARAGON-HF trial being attributed to cardiovascular causes.[166]

Similar to the proportions of patients experiencing events, examining rates of hospitalisations gives a similar picture. Rates of HF hospitalisations were highest by a large margin in the patients who died of CV causes. Further subdivision showed the highest rate was in patients who died of progressive pump failure. Rates of HF hospitalisation were similar in patients with sudden CV death and non-CV death. Looking at all cause hospitalisations, rates were highest in those with non-CV death and were slightly lower in patients with cardiovascular death. Patients who died of pump failure also had high rates of all cause hospitalisation, but this largely reflects high rates of hospitalisation for HF (111.7 all-cause hospitalisation per 100 patient years, compared with 99.5 per 100 patient years for heart failure hospitalisations).

The risk of cardiovascular death was higher in the period immediately following a heart failure hospitalisation. Due to the competing risk of cardiovascular death, the comparative risk of non-CV death appears unchanged or lower in the post HF hospitalisation period.

When examining types of CV death in more detail, patients who died of progressive heart failure were more likely to have had HF hospitalisations during the trial than those who died suddenly. This would be reflective of the accepted

trajectory of patients with heart failure, with recurrent declines in function, often resulting in hospitalisation, and incomplete recovery before death with sudden death occurring at any point.[36] The number of patients with each cause of death is quite small in this post-hoc analysis however it does suggest patients with HFPEF behave similarly to patients with HFREF and this accepted trajectory is true in this contemporary analysis.

The change in KCCQ, and trajectory in symptom burden, was also similar in pattern to that of the HFREF analysis. Patients with HFPEF had a lower baseline KCCQ overall with patients with increasing number of hospitalisations for heart failure having a more rapid decline in KCCQ score.

Recently, SGLT2i have been shown to be effective treatments in heart failure and mildly reduced and preserved ejection fraction in reduction in heart failure hospitalisations and cardiovascular death as a composite endpoint, and have included patients randomised during a hospitalisation with a good safety profile.[31,32] In subgroup analysis, initiation during or shortly after a hospitalisation has been shown to be both safe and effective.[168] Therefore although treatment options are more limited in patients with heart failure and preserved ejection fraction, the risk of cardiovascular death is increased after a hospitalisation for heart failure and therefore it may be an opportune time to initiate treatment with an SGLT2i.

In summary, the trajectory with regards symptom score and heart failure hospitalisations over time is overall similar between HFREF and HFPEF. Although patients with HFPEF have high burden of both all cause hospitalisation and all cause mortality, the relationship between heart failure hospitalisation and risk of CV death remains closely linked. An admission with HF should alert the physician to a period of higher risk of outcome and treatments, such as SGLT2i can be considered for optimisation.

Chapter 5 Latent class analysis in heart failure with reduced ejection fraction

5.1 Introduction

Chapters 3 and 4 have examined one approach to better understanding prognosis in patients with heart failure through examining patterns of heart failure hospitalisations and the risk of cardiovascular death, including by different modes of cardiovascular death. Overall, the results regarding the relationship between timings of heart failure hospitalisation and CV death were similar in HFREF and HFPEF. This adds an extra layer of understanding patient risk and prognostication by describing the average patient journey from the clinical trial data.

Ejection fraction is only one factor of the patient phenotype that determines prognosis and likely response to treatments and, given the very different evidence base for treatments in between these populations, remains an important distinction. However, the heterogeneity of patients within these broad groups is well recognised clinically. As summarised in Chapter 1 there are several approaches aiming to find clusters of similar patients in a data-driven fashion and describe their characteristics, prognosis and responses to treatments.

There appear to be several phenotypes that appear consistently in previous machine learning unsupervised clustering techniques utilising different types of data, including clinical trial and routine databases. The focus in this analysis is latent class analysis, the method for which is described in more detail in Chapter 2. Previous LCA in HFREF have shown some success in both identifying groups and some signal to different treatment responses.[58,59] To increase the utility of this approach, finding consistent groups in analysis of different data, and including contemporary and geographically diverse data, is key.

This analysis uses patient level data from the PARADIGM-HF and ATMOSPHERE trials to determine if latent class analysis, a cluster analysis based approach, can identify phenotypic subgroups of patients with HFREF. These subgroups will

then be examined for any differential treatment effects using modal class assignment.

5.2 Methods

5.2.1 Phenotype identification

Patients were characterised according to 14 prospectively selected clinical features: age, sex, race, ischaemic aetiology of heart failure, duration of heart failure, left bundle branch block (LBBB) on ECG, estimated glomerular filtration rate, body mass index (BMI), clinical history of hypertension, diabetes, hyperlipidaemia, valvular heart disease, atrial fibrillation and presence of anaemia (haematocrit < 39% for men and < 35% for women). Haematocrit was selected rather than haemoglobin to allow comparison with other LCA analysis and in keeping with established reference ranges.[59,169] Variables were considered as continuous where possible (age, BMI and eGFR) with heart failure duration split into three categories (<1 year, 1-5 years, >5 years). Otherwise variables were categorical or binary.

Latent class analysis (LCA) was performed using Stata with the generalised structural equation modelling function (Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC). Latent class definitions were derived using maximum-likelihood estimation to identify the most common patterns of the 14 variables for a range of 1-7 subgroups. The optimal number of subgroups was determined using Information Criterion and reproducibility of maximum likelihood estimation using different random seeds.

Probabilities of membership in each LCA group were used to determine the most likely subgroup for each patient. The median probability of patients being allocated to each class was calculated, as well as the proportion of patients who had a >50% and >70% chance of being allocated to the LCA class to evaluate the discrimination between groups.

5.2.2 Association between HFREF phenotype and outcomes

The primary composite outcome was analysed according to the intention-to-treat principle using Kaplan-Meier estimates and Cox proportional-hazards

models. Phenotype was treated as a categorical covariate in a model adjusted for treatment to explore differences in outcome by phenotype group. Rates of events within each phenotype group according to randomised treatment were examined and interaction between treatment effect and subgroup investigated.

5.2.3 External validation (ATMOSPHERE)

HFREF phenotype classification derived from PARADIGM-HF were applied to ATMOSPHERE subjects. Associations between phenotype, outcome, and interaction with treatment group were analysed using Kaplan-Meier estimates and Cox proportional hazards models as in PARADIGM-HF.

5.3 Results

5.3.1 Developing the latent class model

Latent class models were created for one to seven classes. A model for seven classes could not be identified and was disregarded. AIC was considered for each model, with a 6 class solution found to have lowest AIC (Figure 5-1). The process of developing the model was repeated using 10 random seeds and the log-likelihoods compared. Several different log-likelihoods were identified for different starting seeds, suggesting difficulty in identifying the global minimum and a less stable solution. Therefore the 5 class model was examined in the same manner. 10 random starting seeds came to the same log likelihood therefore the 5 class model was taken forward. The probability of patients belonging to each class was evaluated, and each patient allocated to the class with the highest probability. The median probability of group allocation was calculated in each phenotype, as well as the proportion with >50% and >70% change of being allocated to each group, to assess the ability of the latent class model to discriminate between subgroups (Table 5-1).

Figure 5-1. AIC for different LCA models in PARADIGM-HF
AIC of latent class models with 1 – 7 subgroups in PARADIGM-HF.

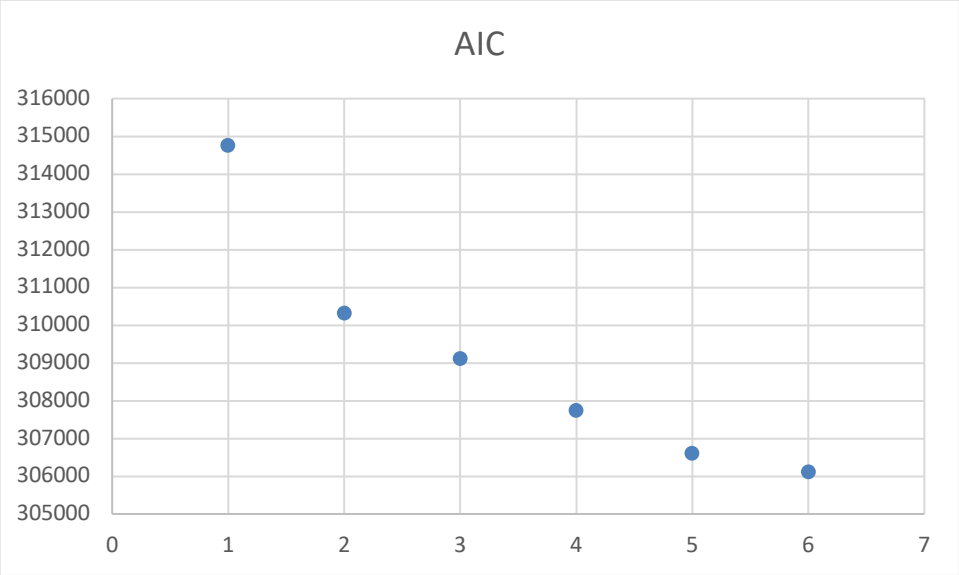


Table 5-1. Probability of class membership in PARADIGM-HF
Probability of class membership, PARADIGM-HF.

	1	2	3	4	5
Number	2220	1209	848	3026	1096
Number >50% probability of belonging to class (%)	2047 (92.2)	1127 (93.2)	764 (90.1)	2895 (95.7)	1027 (93.7)
Number >70% probability of belonging to class	1406 (63.3)	848 (70.1)	576 (67.9)	2389 (78.9)	845 (77.1)
Median probability of belonging to class [IQR]	0.790 [0.619-0.931]	0.859 [0.660-0.969]	0.848 [0.631-0.968]	0.892 [0.743-0.944]	0.880 [0.718-0.959]

5.3.2 Patient characteristics within LCA groups

Subgroups are summarised in Table 5-2 and described below.

5.3.2.1 Phenotype 1 –Non-ischaemic cardiomyopathy in older patients

Phenotype 1 subjects were older (70.0±8.3 years) and most likely to be female (31.2%). The majority of subjects had a non-ischaemic aetiology (66.7%). Phenotype 1 had the highest rate of BBB (25.8%). Rates of diabetes (21.8%) were lower than most other groups whereas valvular heart disease (10.1%) and atrial fibrillation (45.0%) were more common than other phenotypes. Phenotype 1 subjects had the highest NT-proBNP (2018 [1124-4066] pg/mL).

5.3.2.2 Phenotype 2 – Idiopathic dilated cardiomyopathy

Phenotype 2 were by far the youngest (age 48.2±9.1 years). The rate of non-ischaemic cardiomyopathy was the highest (83.3%) of any phenotype. Phenotype 2 had low rates of hypertension (40.5%), diabetes (12.8%), valvular heart disease (4.1%), and COPD (3.1%) with the highest eGFR (84.0 [72.0-99.0] ml/min/1.73 m^2). Patients were less likely to have NYHA class III/IV symptoms (14.9%).

5.3.2.3 Phenotype 3– Metabolic cardiomyopathy of obesity

Phenotype 3 was intermediate in age (59±8.9), primarily Caucasian (84.8%) and similar rates of ischaemic (46.2%) and non-ischaemic aetiology. BMI was the highest of any phenotype (38.5±4.8), and rate of hypertension (92.6%) and diabetes (59.9%) were highest of any phenotype, although proportion taking lipid lowering therapies was lower (56.6%). Phenotype 3 had significant prevalence of atrial fibrillation (41.4%) and had the greatest proportion of patients with COPD (17.5%). Subjects were most likely to have had a prior HF hospitalisation (69.1%) and the highest proportion of NYHA class III/IV (36.7%).

5.3.2.4 Phenotype 4 – Ischaemic cardiomyopathy in older patients

Phenotype 4 subjects were older (age 67.9±8.6 years) and predominantly Caucasian (89.7%). There was a high rate of ischaemic cardiomyopathy (94.6%), and the longest overall duration of HF (81% > 1 year). Proportion with hypertension and diabetes were moderately high (79.6% and 42% respectively) and the majority of patients were taking lipid lowering therapies (93.7%).

5.3.2.5 Phenotype 5 – Ischaemic cardiomyopathy, Asian subtype

Phenotype 5 was had predominantly ischaemic CM (76.7%). It was comprised almost entirely of Asian subjects (90.5%) despite Asian subjects comprised only 18.0% of the trial population. Phenotype 5 subjects had a comparatively high rate of diabetes mellitus (44.6%) and high rates of lipid lowering therapy (64.4%) with a relatively low rate of hypertension (50.6%) and the lowest BMI (23.5±3.6 kg/m^2). Otherwise, they had the lowest rates of atrial fibrillation (4.4%), COPD (4.7%) and LBBB (4.7%) but had among the highest rates of anaemia (39.3%).

Duration of HF was relatively low (88.6% < 5 years). Interestingly, the rate of device therapy was much lower than all other phenotypes.

Table 5-2. Baseline characteristics of phenotype groups in PARADIGM-HF
Baseline characteristics and outcomes by subgroup, PARADIGM-HF.

	1	2	3	4	5	P=value
Number	2,220	1,209	848	3,026	1,096	
LCA variables						
Age (years)	70.0 (8.3)	48.2 (9.1)	59.1 (8.9)	67.9 (8.6)	60.8 (9.0)	<0.001
Female sex	692 (31.2%)	229 (18.9%)	256 (30.2%)	432 (14.3%)	223 (20.3%)	<0.001
Race						
Caucasian	1,690 (76.1%)	421 (34.8%)	719 (84.8%)	2,714 (89.7%)	0 (0.0%)	<0.001
Black	98 (4.4%)	209 (17.3%)	75 (8.8%)	31 (1.0%)	15 (1.4%)	
Asian	97 (4.4%)	384 (31.8%)	6 (0.7%)	30 (1.0%)	992 (90.5%)	
Other	335 (15.1%)	195 (16.1%)	48 (5.7%)	251 (8.3%)	89 (8.1%)	
Ischaemic aetiology (%)	739 (33.3%)	202 (16.7%)	392 (46.2%)	2,862 (94.6%)	841 (76.7%)	<0.001
Duration of HF						
<1 year	599 (27.0%)	586 (48.5%)	210 (24.8%)	575 (19.0%)	553 (50.5%)	<0.001
1-5 years	895 (40.3%)	451 (37.3%)	346 (40.8%)	1,122 (37.1%)	418 (38.1%)	
>5 years	726 (32.7%)	172 (14.2%)	292 (34.4%)	1,329 (43.9%)	125 (11.4%)	
Body Mass Index (kg/m^2)	27.0 (3.9)	27.0 (4.5)	38.5 (4.8)	28.3 (3.9)	23.5 (3.6)	<0.001
Hypertension	1,700 (76.6%)	490 (40.5%)	785 (92.6%)	2,410 (79.6%)	555 (50.6%)	<0.001
Diabetes	483 (21.8%)	155 (12.8%)	508 (59.9%)	1,272 (42.0%)	489 (44.6%)	<0.001
Lipid lowering therapy	403 (18.2%)	305 (25.2%)	480 (56.6%)	2,835 (93.7%)	706 (64.4%)	<0.001
Valvular heart disease	225 (10.1%)	49 (4.1%)	46 (5.4%)	259 (8.6%)	25 (2.3%)	<0.001
Atrial fibrillation	986 (45.0%)	157 (13.3%)	344 (41.4%)	502 (16.8%)	47 (4.4%)	<0.001
COPD	346 (15.6%)	37 (3.1%)	148 (17.5%)	498 (16.5%)	51 (4.7%)	<0.001
LBBB	573 (25.8%)	179 (14.8%)	133 (15.7%)	596 (19.7%)	172 (15.7%)	<0.001
eGFR ($mL/min/1.73m^2$)	62 (51-73)	84 (72-99)	70 (58-83)	62 (51-74)	69 (56-81)	<0.001
Anaemia	398 (17.9%)	134 (11.1%)	72 (8.5%)	630 (20.8%)	431 (39.3%)	<0.001

Table 5-2 continued.

Non LCA variables						
Ejection fraction (%)	31 (26-35)	28 (24-32.6)	32 (27-35)	30 (25-35)	29 (24-33)	<0.001
NYHA class III/IV	660 (29.8%)	180 (14.9%)	311 (36.7%)	785 (26.0%)	142 (13.0%)	<0.001
Prior HF hospitalisation	1,375 (61.9%)	811 (67.1%)	586 (69.1%)	1,854 (61.3%)	648 (59.1%)	<0.001
Heart rate (bpm)	73 (12)	74 (12)	76 (13)	70 (11)	75 (11)	<0.001
Systolic blood pressure (mmHg)	123 (15)	117 (14)	125 (16)	122 (15)	117 (15)	<0.001
NT-proBNP (pg/ml)	2018 (1124-4066)	1463 (845-3108)	1219.5 (722-2254)	1497.5 (840.5-2909.5)	1778 (933-3750)	<0.001
Treatments						
Beta blocker	2,032 (91.5%)	1,138 (94.1%)	804 (94.8%)	2,869 (94.8%)	968 (88.3%)	<0.001
MRA	1,187 (53.5%)	816 (67.5%)	504 (59.4%)	1,598 (52.8%)	566 (51.6%)	<0.001
Digoxin	783 (35.3%)	495 (40.9%)	274 (32.3%)	553 (18.3%)	434 (39.6%)	<0.001
Diuretic	1,832 (82.5%)	968 (80.1%)	766 (90.3%)	2,380 (78.7%)	792 (72.3%)	<0.001
ICD	221 (10.0%)	107 (8.9%)	148 (17.5%)	742 (24.5%)	25 (2.3%)	<0.001
CRT	140 (6.3%)	55 (4.5%)	57 (6.7%)	298 (9.8%)	24 (2.2%)	<0.001
Endpoints						
Composite primary outcome	519 (23.4%)	261 (21.6%)	219 (25.8%)	757 (25.0%)	275 (25.1%)	0.088
All cause death	434 (19.5%)	194 (16.0%)	124 (14.6%)	571 (18.9%)	223 (20.3%)	0.001
CV death	338 (15.2%)	168 (13.9%)	102 (12.0%)	442 (14.6%)	201 (18.3%)	0.002
HF hospitalisation	263 (11.8%)	132 (10.9%)	139 (16.4%)	432 (14.3%)	119 (10.9%)	<0.001

5.3.3 Validation

The classification developed in PARADIGM-HF was applied to patients in ATMOSPHERE. Patients were allocated to the class with the maximum predicted probability of class membership. The probabilities of belonging to each class

were lower when applied to ATMOSPHERE suggesting a lesser ability to discriminate between different latent classes (Table 5-3).

Table 5-3. Probability of latent class membership, ATMOSPHERE
Probability of class membership, PARADIGM-HF criteria applied to ATMOSPHERE.

	1	2	3	4	5
Number	2648	1287	699	1211	1171
Number >50% probability of belonging to class (%)	2365 (89.3)	1163 (90.4)	612 (87.6)	1003 (82.8)	1098 (93.8)
Number >70% probability of belonging to class	1503 (56.8)	826 (64.2)	444 (63.5)	505 (41.7)	829 (70.8)
Median probability of belonging to class [IQR]	0.741 [0.583-0.874]	0.773 [0.615-950]	0.805 [0.605-0.950]	0.653 [0.534-0.821]	0.840 [0.665-0.939]

Baseline characteristics by LCA class in PARADIGM-HF were overall very similar to baseline characteristics in ATMOSPHERE when validated (Table 5-4). Proportionally, group sizes were similar.

Table 5-4. Baseline characteristics of phenotype groups in ATMOSPHERE
Baseline characteristics and outcomes by subgroup, PARADIGM-HF coefficients applied to ATMOSPHERE.

	1	2	3	4	5	p-value
Number	2,648	1,287	699	1,211	1,171	
LCA variables						
Age	70.4 (8.1)	48.5 (9.7)	59.6 (8.7)	66.9 (8.2)	61.5 (9.6)	<0.001
Female sex	685 (25.9%)	196 (15.2%)	184 (26.3%)	172 (14.2%)	288 (24.6%)	<0.001
Race						
Caucasian	2,255 (85.2%)	565 (43.9%)	642 (91.8%)	1,130 (93.3%)	0 (0.0%)	<0.001
Black	27 (1.0%)	50 (3.9%)	19 (2.7%)	8 (0.7%)	5 (0.4%)	
Asian	114 (4.3%)	537 (41.7%)	4 (0.6%)	12 (1.0%)	1,097 (93.7%)	
Other	252 (9.5%)	135 (10.5%)	34 (4.9%)	61 (5.0%)	69 (5.9%)	
Ischaemic aetiology	886 (33.5%)	664 (51.6%)	251 (35.9%)	680 (56.2%)	605 (51.7%)	<0.001
Duration of HF						
<1 year	690 (26.1%)	654 (50.9%)	178 (25.5%)	236 (19.5%)	603 (51.5%)	<0.001
1-5 years	1,050 (39.7%)	443 (34.5%)	283 (40.5%)	406 (33.6%)	411 (35.1%)	
>5 years	908 (34.3%)	188 (14.6%)	238 (34.0%)	568 (46.9%)	156 (13.3%)	
BMI (kg/m^2)	26.8 (3.8)	26.1 (4.3)	37.3 (4.1)	28.4 (3.6)	23.0 (3.4)	<0.001
Hypertension	1,859 (70.2%)	436 (33.9%)	628 (89.8%)	898 (74.2%)	511 (43.6%)	<0.001
Diabetes	489 (18.5%)	134 (10.4%)	368 (52.6%)	538 (44.4%)	415 (35.4%)	<0.001
Hyperlipidaemia	1,087 (41.0%)	448 (34.8%)	441 (63.1%)	1,125 (92.9%)	618 (52.8%)	<0.001
AF	1,471 (55.6%)	210 (16.3%)	309 (44.2%)	302 (24.9%)	98 (8.4%)	<0.001
COPD	382 (14.4%)	43 (3.3%)	117 (16.7%)	193 (15.9%)	52 (4.4%)	<0.001
LBBB	1,020 (38.5%)	330 (25.6%)	221 (31.6%)	442 (36.5%)	304 (26.0%)	<0.001
eGFR	65.0 (53.0-76.0)	93.0 (79.0-107.0)	74.0 (63.0-87.0)	68.0 (56.0-80.0)	72.0 (59.0-85.0)	<0.001
Anaemia	556 (21.0%)	169 (13.1%)	76 (10.9%)	274 (22.6%)	543 (46.4%)	<0.001

Table 5-4 continued

Non LCA variables						
LVEF (%)	30 (25-34)	28 (23-32)	30 (26-34)	30 (25-33)	30 (25-33)	<0.001
NYHA III/IV	899 (34.0%)	259 (20.1%)	284 (40.6%)	320 (26.4%)	241 (20.6%)	
Heart rate	82.66352 (16.72452)	87.07226 (18.70538)	88.14735 (16.61531)	80.25021 (14.50019)	86.75747 (16.53381)	<0.001
Systolic BP (mmHg)	126.224 (17.8481)	117.761 (17.1953)	128.792 (18.0632)	125.985 (18.4846)	118.816 (17.2478)	<0.001
NT-proBNP (pg/ml)	1522 (861-2705)	859 (444-1674)	917.5 (508-1682)	1085 (616-2064)	1191 (608.5-2379.5)	<0.001
Treatments						
Beta-blocker	2,416 (91.2%)	1,197 (93.0%)	662 (94.7%)	1,139 (94.1%)	1,018 (86.9%)	<0.001
Spironolactone	821 (31.0%)	471 (36.6%)	255 (36.5%)	363 (30.0%)	397 (33.9%)	<0.001
Digoxin	821 (31.0%)	491 (38.2%)	192 (27.5%)	274 (22.6%)	464 (39.6%)	<0.001
Diuretic	2,114 (79.8%)	975 (75.8%)	606 (86.7%)	974 (80.4%)	929 (79.3%)	<0.001
ICD	388 (14.7%)	138 (10.7%)	132 (18.9%)	292 (24.1%)	35 (3.0%)	<0.001
CRT	175 (6.6%)	31 (2.4%)	44 (6.3%)	118 (9.7%)	25 (2.1%)	<0.001
Outcomes						
Composite primary outcome	994 (37.5%)	393 (30.5%)	227 (32.5%)	377 (31.1%)	378 (32.3%)	<0.001
All cause death	871 (32.9%)	298 (23.2%)	160 (22.9%)	279 (23.0%)	287 (24.5%)	<0.001
CV death	720 (27.2%)	269 (20.9%)	137 (19.6%)	228 (18.8%)	267 (22.8%)	<0.001
HF hospitalisation	525 (19.8%)	217 (16.9%)	153 (21.9%)	241 (19.9%)	188 (16.1%)	0.003

5.3.4 Outcomes

In the LCA identified in PARADIGM-HF there was no significant difference in the primary composite outcome between the phenotype subgroups [adjusted for randomised treatment] (Figure 5-2 and Table 5-5). When validated in

ATMOSPHERE class 2 (idiopathic dilated cardiomyopathy) had the lowest rate of primary composite outcome with the other groups having similar rates of event (Figure 5-3 and Table 5-6).

Figure 5-2. Primary endpoint by phenotype group in PARADIGM-HF
Kaplan-Meier curves for the primary endpoint by phenotype subgroups, PARADIGM-HF.

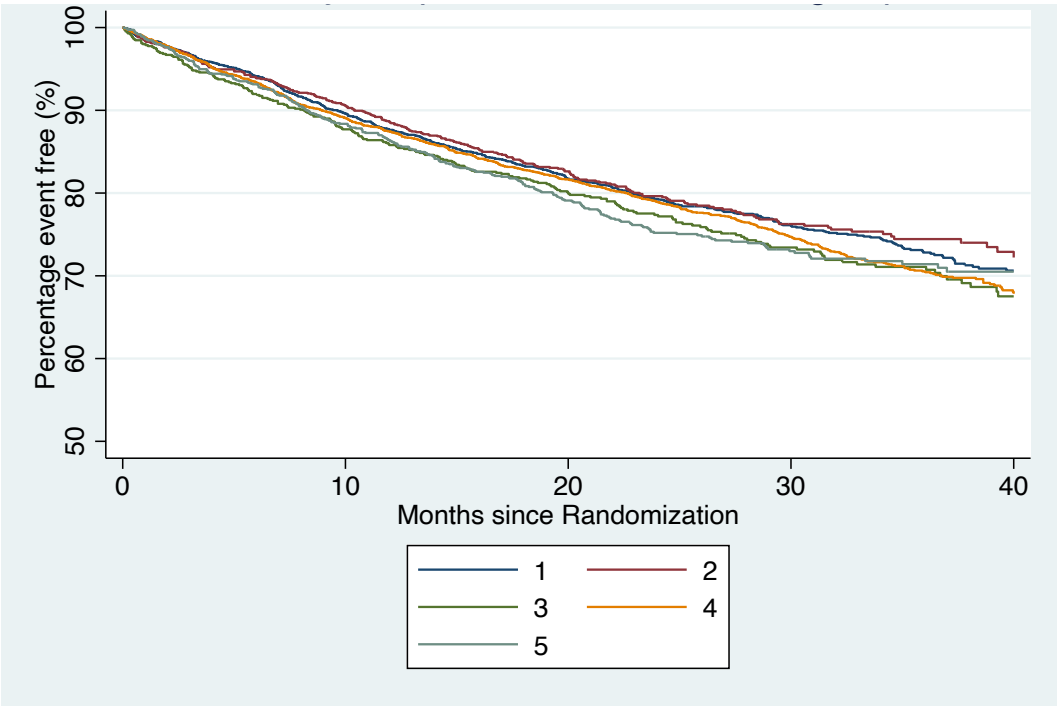


Table 5-5. Primary endpoint by phenotype group in PARADIGM-HF
Number of events and hazard ratio for the primary outcome in LCA subgroups for PARADIGM-HF. HR is adjusted for the randomised treatment.

Subgroup	Total	Events	%	HR	95% CI
1	2220	519	23.4	1.05	0.91 – 1.22
2	1209	261	21.6	REFERENCE	REFERENCE
3	848	219	25.8	1.17	0.98 - 1.40
4	3026	757	25.2	1.13	0.98 - 1.30
5	1096	275	25.1	1.17	0.99 – 1.39

Figure 5-3. Primary endpoint by phenotype group in ATMOSPHERE
Kaplan-Meier curves for the primary endpoint in phenotype subgroups, PARADIGM definitions applied to ATMOSPHERE.

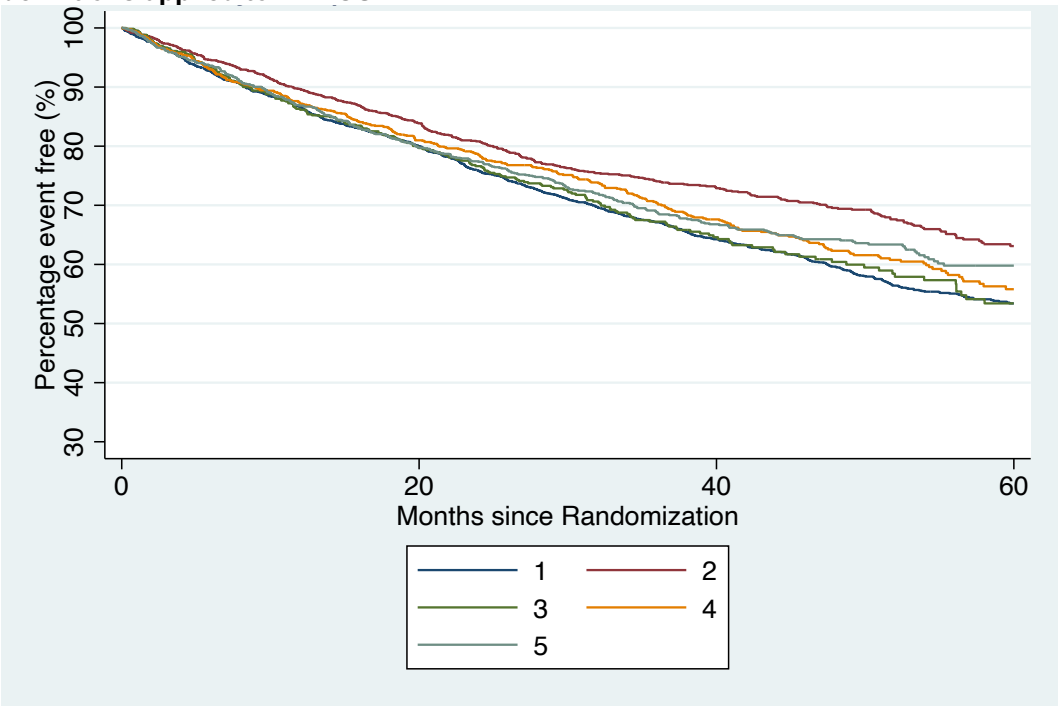


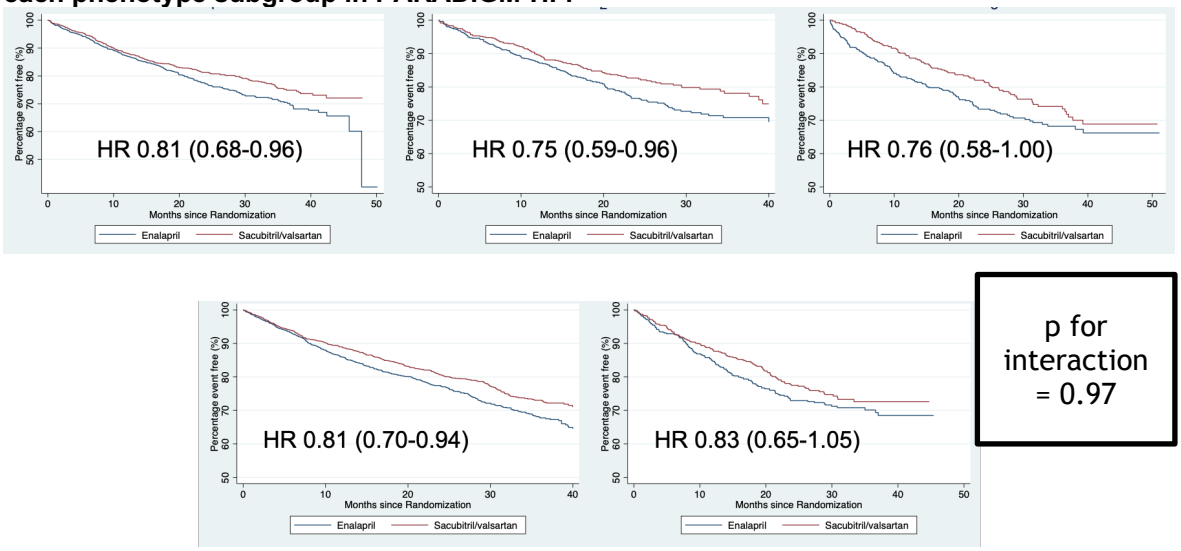
Table 5-6. Primary endpoint by phenotype group in ATMOSPHERE
Number of events and hazard ratio for the primary outcome in LCA subgroups for ATMOSPHERE. HR is adjusted for the randomised treatment.

Subgroup	Total	Events	%	HR (95%CI)
1	2648	994	37.5	1.36 (1.21-1.53)
2	1287	393	30.5	REFERENCE
3	699	227	32.5	1.33 (1.13-1.56)
4	1211	377	31.1	1.21 (1.05-1.39)
5	1171	378	32.3	1.20 (1.04-1.38)

5.3.5 Randomised treatment effect

Treatment effect of sacubitril valsartan as compared with enalapril was consistent across phenotype subgroups (p for interaction 0.97) (Figure 5-4).

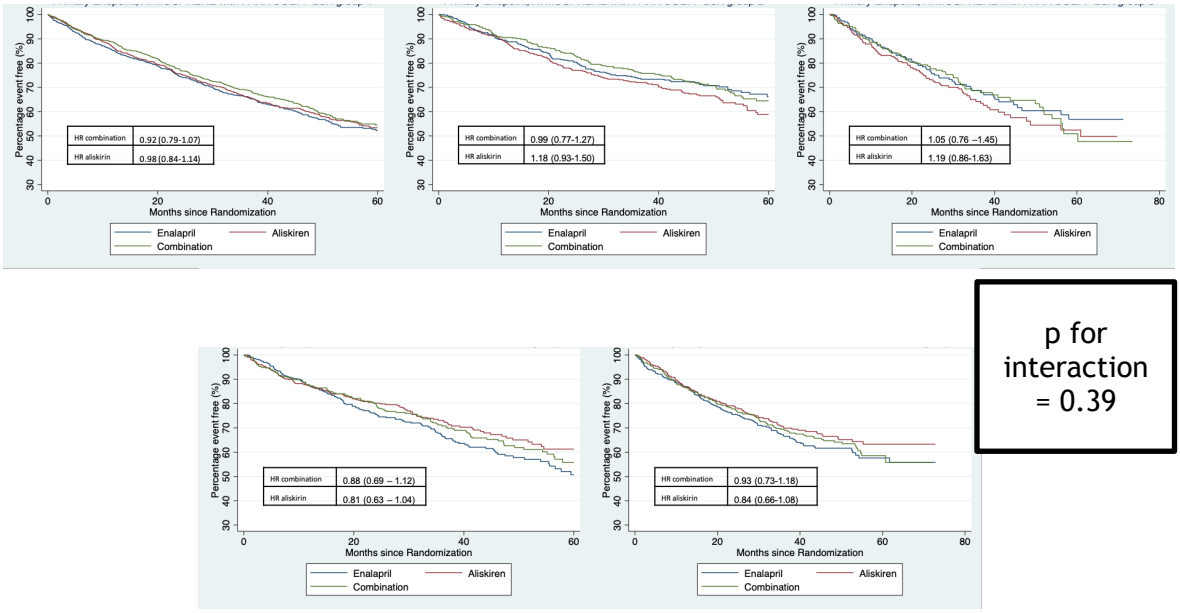
Figure 5-4. Treatment effect in each phenotype group in PARADIGM-HF
Kaplan-Meier curves and hazard ratios for sacubitril-valsartan as compared to enalapril in each phenotype subgroup in PARADIGM-HF.



p for interaction = 0.97

In ATMOSPHERE there was no significant variation in treatment effect for either combination vs. enalapril or aliskiren vs. enalapril by LCA group (p for interaction 0.39) (Figure 5-5).

Figure 5-5. Treatment effect in each phenotype group in ATMOSPHERE
Kaplan-Meier curves and hazard ratios for aliskiren as compared to enalapril and combination therapy as compared with enalapril alone in each phenotype subgroup in ATMOSPHERE.



p for interaction = 0.39

5.4 Discussion

In this chapter, I have identified multiple latent classes within these large global clinical trials in HFREF, representing groups of patients with clusters of similarities across a range of commonly measured variables. This means patients are classified into groups based on multiple different variables with no one feature determining class. The complex multi-levelled nature of this clustering adds additional information about the patient above single variable classifications such as ischaemic and non-ischaemic heart failure.

Several of these latent groups are recognisable as frequently encountered clinical phenotypes. These include younger patients with features suggestive of idiopathic dilated cardiomyopathy, older patients with ischaemic aetiology and obese patients with multimorbidity. Although in this study there was no significant difference in prognosis or treatment response identified, it is valuable that unsupervised latent class analysis was able to identify clinically recognisable phenotypes. Further work may be able to further differentiate phenotypes and explore deeper into prognostic differences and treatment responses.

Retrospective analysis of patients with non-ischaemic aetiology of HFREF in the ß-blocker Evaluation of Survival Trial (BEST) using latent class analysis has previously been able to identify groups with differing outcomes and response to treatment.[59] This analysis cannot be directly compared to the current study given the different patient populations. However, there are some similarities between group 3 of the current study and two groups in the BEST analysis including patients with high BMI, hypertension, hyperlipidaemia and diabetes.

Latent class analysis has been utilised in evaluation of the EMPHASIS-HF trial with validation in the EPHESUS trial.[58] Four subgroups were identified. The subgroup with the worst outcome (composite CV mortality and heart failure hospitalisation) shared similarities to group 4 of this study including older age, higher proportion male patients and prevalence of ischaemic aetiology. Variation in response to eplerenone was demonstrated between subgroups, with two subgroups experiencing less events, however this only remained true for one subgroup when tested in validation data.

Latent class analysis of the ASIAN-HF registry found five subgroups. Subgroups had geographical differences in prevalence and differences in the primary outcome of all-cause mortality and HF hospitalisation.[65] The subgroup of lean diabetic patients was novel, found predominantly in Southeast Asia. The presence of these characteristics may influence group 5 of this trial, as this subgroup consists of majority Asian patients and has the lowest BMI of any subgroups and the second highest rate of diabetes. Other groups appear similar to those found in this study, for example a metabolic group and a subgroup with young patients and low rates of ischaemic aetiology (similar to 2).

Latent class analysis provides the tools to use data driven techniques to identify phenotypes, and reassuringly there appear to be both clinically recognisable and reproducible across different patient populations. Strengths of this analysis include the number of patients included in creation and validation of the latent classes and the global recruitment of patients to the trials. These phenotypes can help us understand our patients with heart failure and contribute towards more personalised care.

Although there have not yet been any consistent signals in terms of different treatment responses in latent classes, LCA remains a potential pathway to identify this. As discussed in Chapter 2, the signal for any difference in treatment effect can be hard to detect when patients are split into the latent classes as there is always a degree of uncertainty of class membership. Therefore, it is unlikely study protocols would be made incorporating allocation to latent class in the randomisation process until improved techniques for linking latent classes with distal outcomes are developed. Another avenue to using these defined phenotypes to look for responses in phenotypes would be to use the phenotype descriptions to inform recruitment criteria to a study to specifically recruit a certain phenotype hypothesised to respond preferentially to the treatment.

5.4.1 Limitations

Success and accuracy of latent class analysis is dependent on selection of variables for analysis. In this study, clinically pertinent and readily recorded variables were selected but it is possible other variables may provide greater

discrimination between groups. Other variables that were not recorded, for example exercise capacity, may improve performance of the model. Given the greater recruitment of men to the trials, all latent class groups have a higher proportion of men than women and differences in behaviour of male compared with female patients may be missed. Sex disaggregated analysis is limited by the number of patients but might provide an interesting avenue of exploration. Similarly, it may be of interest to analyse separately by aetiology of HF, as was done in analysis of the BEST trial, however this is restricted by reduced patient numbers for analysis.

Chapter 6 Latent class analysis in heart failure with preserved ejection fraction

6.1 Introduction

In Chapter 5, a new latent class analysis was carried out in PARADIGM-HF with validation in the ATMOSPHERE trial, identifying clinically recognised phenotypes with groups that were similar to those identified in previous latent class analyses. HFPEF is thought to have a wider variation in phenotypes, particularly given the high degree of comorbidity associated with the condition. Identifying these phenotypes is a potential avenue to identify effective treatments in HFPEF which remains challenging. The lack of effective treatments is assessed in the broad category of heart failure with preserved ejection fraction, but perhaps some subgroups respond either to treatments we already have available or new treatments might have differential responses.

There have been several previous latent class analyses of patient populations with heart failure and preserved ejection fraction.[60–62] These have identified several consistent phenotypes in different data. There are some slight differences in approaches used, in addition most are 'traditional' latent class analyses using only categorical/factor variables rather than continuous versions of the variable where possible. Each analysis based the construction of the latent classes on different observed variables.

I then became interested in whether consistent groups could be identified by using different input variables in the same patient population. Latent class analysis depends on the input variables selected therefore it was hypothesised that the core phenotypes that appear throughout different analyses could be identified using different input variables. In addition, these strong phenotypes should be identified using both categorical and continuous variables, providing reassurance that phenotypes are genuine clinical phenotypes and not random patterns within the data being identified by the data analysis process itself.

This analysis uses patient level data from the PARAGON-HF trial to generate latent class analysis solutions. The aims of this analysis are to investigate whether identified groups appear consistently using different variable selection

based on those used in previous trials, and examining whether these remain consistent when slightly different techniques are used – namely categorising continuous variables or using the continuous variable.

6.2 Methods

6.2.1 Analysis of prior latent class analysis results

For each prior analysis, the variables entered into the latent class analysis were identified, and where continuous variables were categorised the cut-point for defining groups were extracted. A summary of the variables, and categorisation, used in each prior trial is given in Table 6-1.

Table 6-1. Summary of variables used in creation of latent classes in each prior research

1	2	3
Uijl et al.[60]	Cohen et al.[61]	Kao et al.[62]
Sex	Sex	Sex
Age (years)	Age (years)	Age (years)
<65	<65	<70
65-74	65-74	70-80
75-85	75-85	>80
>85	>85	
Body mass index (kg/m^2)	Obesity (BMI ≥ 30 kg/m^2)	Body mass index (kg/m^2)
<25		<18.5
25-30		18.5-25
>30		25-30
		>30
eGFR (mL/min/1.73m^2)	Chronic kidney disease (eGFR <60 mL/min/1.73m^2)	eGFR (mL/min/1.73m^2)
>60		>90
30-60		60-90
<30		30-60
		15-30
Atrial fibrillation	Atrial fibrillation	Atrial fibrillation
Ischaemic heart disease		Coronary artery disease
Diabetes	Diabetes	Diabetes
NYHA I/II vs III/IV	NYHA I/II vs III/IV	
Chronic obstructive pulmonary disease	Race	Dyslipidaemia
Hypertension	Asian	Alcohol
	Black or African American	Haematocrit (%)
	Other	>0.5
	White	0.4-0.5
		0.3-0.4
		0.2-0.3

6.2.2 Labelling latent class groups in prior research

As latent classes are built and defined using all the input variables, it can become difficult to succinctly describe each group. The description of the group also becomes inherently linked to the input variables, adding to the challenge of comparing identified groups between different analyses. In each published analysis, the identified latent class groups were reviewed and descriptions summarised to be as concise as possible, while not removing key distinguishing features. A summary is presented in Table 6-2.

Table 6-2. Description of core characteristics of latent classes in prior research

Model	Uijl et al.	Cohen et al.	Kao et al.
Model number	1	2	3
Population	Swedish HF registry and CHECK-HF	TOPCAT	I-PRESERVE & CHARM
Descriptors	Younger, low comorbidity	Younger, low comorbidity	Male, alcohol, younger
	AF, HTN	Older, more women, AF, CKD	Female, anaemia, younger
	Older, AF	Diabetes, obesity	Obesity, diabetes, hyperlipidaemia
	HTN, diabetes, obesity		Women
	IHD, poor renal function, older		Male, AF, CAD
			Older women, AF, renal dysfunction, anaemia

6.2.3 Validating Model 1 in PARAGON-HF trial data

To evaluate whether groups identified in model 1 appeared consistently in the PARAGON-HF population, latent class analysis was repeated using the same variables as used in the prior analysis. Firstly, analysis using categorisation of continuous variables with the same cut-points was carried out using the LCA Stata Plugin from Penn State university.[130] The same categories were used as had been done previously. Secondly, the same variables were used but continuous variables were preserved and analysed using the generalised structural equation modelling (GSEM) capabilities inbuilt in Stata.

6.2.3.1 LCA using categorised variables

Solutions for LCA with 1-7 groups were defined using maximum likelihood estimation, and relative model fit compared using AIC and BIC (Figure 6-1). Lower BIC and AIC indicate better model fit, and the preferred number of latent classes, in this case a 4 or 5 class solution appeared best. However, the number of random starting seeds associated with a 4 and 5 class solution was low (20% and 40% respectively, compared with 71% for 3 class solution), suggesting the 4 and 5 class solutions did not identify a global minima and the solution could be unstable. Beyond 3 classes, there was minimal benefit in BIC with addition of extra classes, and 3 classes offered a more stable solution. Therefore, the 3-class solution was selected. The median posterior probability of belonging to the allocated class was reasonably high (all ~80%), giving further support to this solution (Table 6-3).

Figure 6-1. Model fit statistics for Model 1 with categorical variables
AIC and BIC for latent class solution using categorised variables for Model 1.

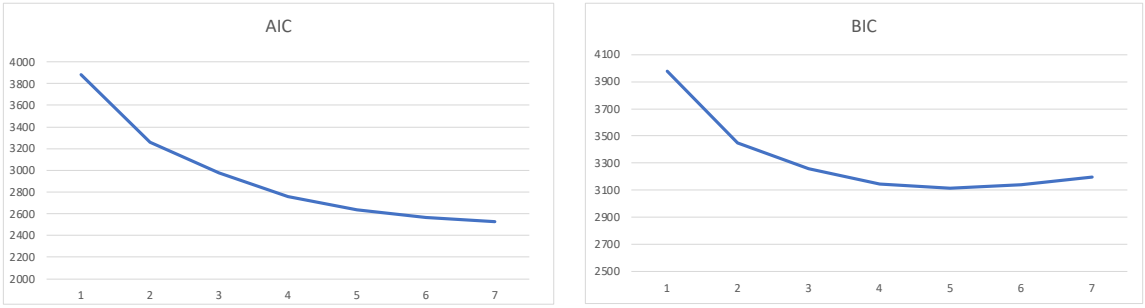


Table 6-3. Median posterior probability of class membership for Model 1 with categorical variables
Median posterior probability of class membership. The diagonal gives the median posterior probability of class membership for the patients allocated to that class.

	Probability class 1	Probability class 2	Probability class 3
Class 1	0.793 [0.63 - 0.911]	0.104 [0.0355 - 0.257]	0.028 [0.00782 - 0.122]
Class 2	0.0806 [0.0186 - 0.218]	0.789 [0.63 - 0.881]	0.0696 [0.0333 - 0.17]
Class 3	0.0247 [0.00489 - 0.0877]	0.123 [0.00164 - 0.229]	0.825 [0.634 - 0.952]

The model was then described using the item response probabilities. The most homogenous variables in each group (i.e., with item response probabilities

closest to 0 or 1) were highlighted to aid interpretation, and used to create a description of each group (Table 6-4).

Table 6-4. Phenotype characteristics for Model 1 with categorical variables
Description of groups in model 1 (categorised variables) using item response probabilities, with variables most different between groups highlighted to aid interpretation.

Class	1	2	3
Descriptors	Male Young Ischaemic Good renal function	Female Obesity Hypertension COPD Diabetes	Older Atrial fibrillation Low BMI
Proportion	0.28	0.38	0.34
Female sex	0.37	0.60	0.55
Age			
<65	0.48	0.08	0.02
65-75	0.46	0.43	0.22
75-85	0.07	0.46	0.59
>85	0.0001	0.03	0.16
NYHA III/IV	0.15	0.26	0.16
BMI			
<25	0.13	0.0006	0.36
25-30	0.35	0.20	0.53
>30	0.52	0.80	0.11
eGFR			
>60	0.82	0.35	0.43
30-60	0.18	0.64	0.56
<30	0.002	0.02	0.01
Ischaemic aetiology	0.58	0.40	0.37
AF	0.17	0.37	0.40
HTN	0.95	0.98	0.92
COPD	0.11	0.17	0.12
Diabetes	0.49	0.54	0.26

Outcomes of time to cardiovascular death and time to first heart failure hospitalisation were examined in the 3 phenotype groups. The lowest rate for both outcomes was in phenotype 1 (younger patients with predominant ischaemic aetiology). The hazard ratio, adjusted for randomised treatment, for phenotype 3 (older patients, high proportion AF and low BMI) compared to phenotype 1 for cardiovascular death was 1.57 [95%CI 1.20-2.05] (Table 6-5).

There was a significant interaction between randomised treatment and phenotype group for the outcome of first heart failure hospitalisation. Patients

in phenotype 2 (greater proportion female, with multimorbidity) appeared to benefit more from treatment with sacubitril valsartan (p for interaction 0.01) Table 6-5).

Table 6-5. Outcomes according to phenotype group for Model 1 with categorical variables Outcomes according to phenotype group for model 1 (categorical variables) and effect of randomised treatment in each phenotype subgroup. [1] adjusted for randomised treatment

Phenotype	1	2	3	interaction p-value
Cardiovascular death				
Rate	2.26 (1.83 - 2.8)	3.12 (2.68 - 3.62)	3.47 (2.97 - 4.04)	
Hazard ratio[1]	REF	1.42 (1.08-1.85)	1.57 (1.20-2.05)	
Treatment effect in each phenotype group	1.49 (0.96 - 2.30)	0.92 (0.68-1.25)	0.79 (0.58-1.07)	0.06
1st HFH				
Rate	5.2 (4.49 - 6.01)	7.57 (6.84 - 8.38)	6.73 (6 - 7.56)	
Hazard ratio[1]	REF	1.34 (1.11-1.60)	1.19 (0.99-1.44)	
Treatment effect in each phenotype group	1.03 (0.77-1.38)	0.72 (0.58-0.88)	1.11 (0.88-1.40)	0.01

6.2.3.2 LCA using continuous variables

LCA models using continuous variables were fit using the GSEM command inbuilt in Stata software. Age, body mass index and eGFR were entered as continuous variables after being standardised to a mean of 0 and standard deviation of 1. Models were evaluated for 1-7 classes to compare model fit, with initial models fit with relaxation of convergence rules using the 'nonrtolerance' option to allow convergence in models with greater number of latent classes. However, final fitted models used default convergence criteria. A five-class solution appeared best by model fit criteria (Figure 6-2), however there were several different solutions with random starting seeds, and some starting seeds would not allow the model to converge therefore this solution appeared unstable. Given the 'elbow' at the three-class solution, where additional classes result in minor benefit in BIC, and the finding that a 3 class solution fit best using the simpler categorical analysis, the 3 class solution was explored further. A 3-class solution came to the same solution with 10 starting seeds with good median posterior

probability of class membership (Table 6-6), therefore a 3 class solution was selected.

Figure 6-2. Model fit statistics for Model 1 with continuous variables
Model fit criteria for Model 1 using continuous variables.

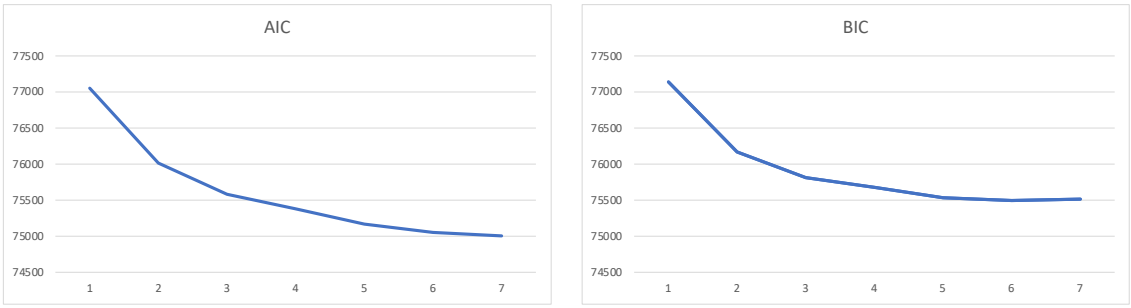


Table 6-6. Median posterior probability of class membership for Model 1 with continuous variables
Median posterior probability of class membership. The diagonal gives the median posterior probability of class membership for the patient allocated to that class.

	Probability class 1	Probability class 2	Probability class 3
Class 1	0.795 [0.636 - 0.917]	0.0208 [0.00361 - 0.107]	0.105 [0.0294 - 0.256]
Class 2	0.0632 [0.0136 - 0.193]	0.835 [0.636 - 0.956]	0.0335 [0.00557 - 0.139]
Class 3	0.0645 [0.0167 - 0.194]	0.00547 [0.000872 - 0.0365]	0.895 [0.738 - 0.972]

The model was then described using the item response probabilities as described previously (Table 6-7).

Table 6-7. Phenotype characteristics for Model 1 with continuous variables
Description of groups in model 1 (continuous variables) using item response probabilities, with variables most different between groups highlighted to aid interpretation. [1]Continuous variable item response are on a standardised scale (mean of 0 and standard deviation of 1); values above 1 infer the mean value in that phenotype is higher than the population mean.

Phenotype	1	2	3
Descriptors	**Female Hypertension Diabetes Obesity**	**Male Ischaemic Good renal function**	**Atrial fibrillation Older**
Proportion	0.33	0.19	0.49
Female sex	0.58	0.32	0.55
NYHA III/IV	0.27	0.15	0.17
Ischaemic aetiology	0.43	0.59	0.40
Atrial fibrillation	0.32	0.16	0.39
Hypertension	0.99	0.94	0.94
COPD	0.17	0.11	0.13
Diabetes	0.59	0.45	0.32
Age[1]	-0.18	-1.12	0.54
BMI[1]	0.84	-0.08	-0.53
eGFR[1]	-0.25	1.07	-0.24

Rates of cardiovascular death and time to first heart failure hospitalisation were calculated in each phenotype group. Phenotype 2, younger males with high proportion ischaemic aetiology, had the lowest rates of both outcomes. Phenotype 3 (older, atrial fibrillation) had the highest rate of CV death with a HR of 1.86 [95%CI 1.37-2.52] as compared with phenotype 2. Phenotype 1 had the highest rate of heart failure hospitalisation with a HR of 1.64 (95%CI 130-2.07) compared with phenotype 2. There was no interaction between randomised treatment and phenotype group on the occurrence of either outcome (Table 6-8).

Table 6-8. Outcomes according to phenotype group for Model 1 with continuous variables
Outcomes according to phenotype group for model 1 (continuous variables) and effect of randomised treatment in each phenotype subgroup. [1] adjusted for randomised treatment

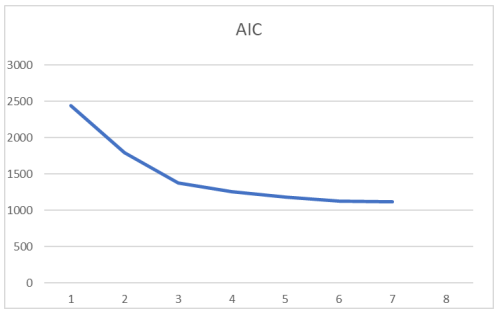
Phenotype	1	2	3	interaction p-value
Cardiovascular death				
Rate	2.54 (2.11 - 3.05)	2.04 (1.55 - 2.68)	3.65 (3.23 - 4.14)	
Hazard ratio[1]	1.28 (0.91 - 1.78)	REF	1.86 (1.37 - 2.52)	
Treatment effect in each phenotype group	1.05 (0.73-1.52)	1.15 (0.66-2.00)	0.89 (0.69-1.13)	0.59
1st HFH				
Rate	7.55 (6.75 - 8.45)	4.16 (3.42 - 5.08)	6.97 (6.35 - 7.66)	
Hazard ratio[1]	1.64 (1.3 - 2.07)	REF	1.5 (1.2 - 1.88)	
Treatment effect in each phenotype group	0.79 (0.63-0.99)	1.08 (0.72-1.30)	0.95 (0.78-1.14)	0.30

6.2.4 Validating Model 2 in PARAGON-HF trial data

6.2.4.1 LCA using categorised variables

Solutions for LCA with 1-7 groups were fitted, and model fit criteria compared (Figure 6-3). There was an 'elbow' at 3-4 classes, with BIC not improving beyond this point. The number of starting seeds coming to the same solution was higher for the 4-class solution compared to 3 (96% and 61% respectively), therefore a 4 class solution was selected as this appeared more reliable and stable. The median posterior probability of class membership was high (Table 6-9).

Figure 6-3. Model fit statistics for Model 2 with categorical variables
AIC and BIC for latent class solutions using categorised variables for Model 2.



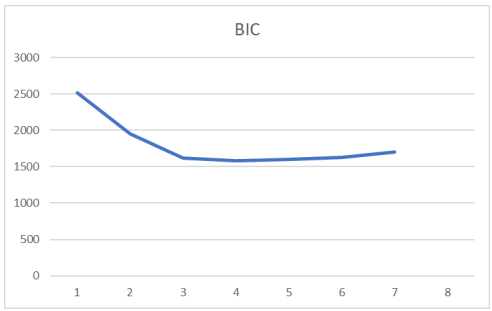


Table 6-9. Median posterior probability of class membership for Model 2 with categorical variables

	Probability class 1	Probability class 2	Probability class 3	Probability class 4
Class 1	0.787 [0.652 - 0.899]	0.0335 [0.0212 - 0.0577]	0.000104 [0.0000286 - 0.000301]	0.15 [0.0516 - 0.325]
Class 2	0.0771 [0.0379 - 0.126]	0.883 [0.793 - 0.93]	0.0212 [0.00625 - 0.0761]	0.000596 [0.000223 - 0.00173]
Class 3	0.00163 [0.0000742 - 0.022]	0.115 [0.00237 - 0.285]	0.881 [0.688 - 0.988]	0.00622 [0.00169 - 0.0113]
Class 4	0.147 [0.000757 - 0.258]	0.00518 [0.000118 - 0.0128]	0.00163 [0.001 - 0.00251]	0.837 [0.725 - 0.996]

The model was described using item response probabilities, with the most homogenous variables highlighted to aid interpretation (Table 6-10).

Table 6-10. Phenotype characteristics for Model 2 with categorical variables Description of groups in model 2 (categorised variables) using item response probabilities, with variables most different between groups highlighted to aid interpretation.

Class	1	2	3	4
Descriptors	**Female White race High symptom burden**	**Older Atrial fibrillation**	**Male Younger Asian race Non-obese**	**Diabetes Obesity**
Gamma	0.27	0.34	0.15	0.25
Female sex	0.63	0.54	0.37	0.46
Age				
<65	0.0006	0.001	0.49	0.39
65-75	0.36	0.23	0.47	0.51
75-85	0.58	0.62	0.04	0.10
>85	0.06	0.15	0.0007	0.0004
Race				
Asian	0.02	0.18	0.32	0.05
Black or African American	0.02	0.01	0.02	0.03
Other	0.03	0.04	0.02	0.05
White	0.92	0.77	0.63	0.87
Diabetes	0.48	0.32	0.40	0.55
AF	0.37	0.40	0.18	0.26
Obese	0.87	0.04	0.004	0.99
NYHA 3/4	0.28	0.17	0.15	0.18
CKD	0.71	0.57	0.23	0.30

Risk of outcomes were compared in each phenotype group. The lowest rate of cardiovascular death was for phenotype 4 (diabetes and obesity). The hazard ratio for phenotype 2 (older, AF) with phenotype 4 as reference was 1.89 [95%CI 1.43 - 2.5). For heart failure hospitalisation, the lowest rate was for phenotype 3 (male, younger, Asian race). The hazard ratio for phenotype 1, with 3 as reference, was 1.72 [95%CI 1.33 – 2.22]. There was no significant interaction between phenotype group and randomised treatment on the occurrence of either outcome (Table 6-11).

Table 6-11. Outcomes according to phenotype group for Model 2 with categorical variables Outcomes according to phenotype group for model 2 (categorical variables) and effect of randomised treatment in each phenotype subgroup. [1] adjusted for randomised treatment

Phenotype	1	2	3	4	interaction p-value
Cardiovascular death					
Rate	3.02 (2.49 - 3.65)	3.74 (3.23 - 4.32)	2.88 (2.23 - 3.72)	2.03 (1.6 - 2.56)	
Hazard ratio[1]	1.53 (1.12 - 2.07)	1.89 (1.43 - 2.5)	1.39 (0.98 - 1.98)	REF	
Treatment effect in each phenotype group	0.88 (0.60 - 1.30)	0.78 (0.58-1.05)	1.52 (0.90-2.58)	1.26 (0.79-2.03)	0.13
1st HFH					
Rate	7.97 (7.04 - 9.02)	7.03 (6.29 - 7.86)	4.34 (3.51 - 5.38)	6.14 (5.33 - 7.07)	
Hazard ratio[1]	1.72 (1.33 - 2.22)	1.55 (1.21 - 1.98)	REF	1.45 (1.12 - 1.88)	
Treatment effect in each phenotype group	0.84 (0.66-1.08)	0.99 (0.79-1.24)	1.06 (0.69-1.63)	0.81 (0.61-1.07)	0.49

6.2.4.2 LCA using continuous variables

LCA solutions were found for 1-7 classes using the same variables, but treated as continuous variables where possible, using the inbuilt STATA generalised structural equation modelling capabilities. A 7-class solution could not be identified and was abandoned. There was a loss of benefit after a 4-class solution (Figure 6-4). The 4-class solution was found to be stable with 17 of 20 random starting seeds coming to the same solution therefore the 4 class solution was taken forward. There was a fair to good median posterior probability of class membership (Table 6-12).

Figure 6-4. Model fit statistics for Model 2 with continuous variables
Model fit criteria for Model 2 using continuous variables.

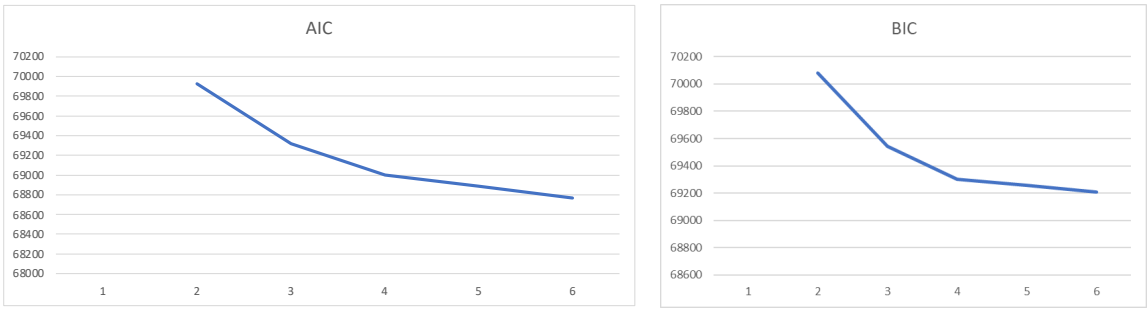


Table 6-12. Median posterior probability of class membership for Model 2 with continuous variables

	Probability class 1	Probability class 2	Probability class 3	Probability class 4
Class 1	0.77 [0.611 - 0.9]	0.0268 [0.00408 - 0.113]	0.126 [0.0336 - 0.272]	0.0014 [0.000234 - 0.00996]
Class 2	0.0373 [0.00337 - 0.155]	0.793 [0.601 - 0.948]	0.00472 [0.000347 - 0.0302]	0.0495 [0.00851 - 0.185]
Class 3	0.0457 [0.00908 - 0.149]	0.00121 [0.00017 - 0.00825]	0.787 [0.624 - 0.907]	0.0662 [0.0128 - 0.22]
Class 4	0.000481 [0.0000517 - 0.0038]	0.00594 [0.000918 - 0.0422]	0.0492 [0.00846 - 0.174]	0.892 [0.713 - 0.973]

Classes were described as previously, with most homogenous characteristics used to create a group description (Table 6-13).

Table 6-13. Phenotype characteristics for Model 2 with continuous variables
Description of groups in model 2 (continuous variables) using item response probabilities, with variables most different between groups highlighted to aid interpretation. [1]Continuous variable item response are on a standardised scale (mean of 0 and standard deviation of 1); values above 1 infer the mean value in that phenotype is higher than the population mean.

	1	2	3	4
Descriptors	**Diabetic Younger Obese**	**Male Asian race Low BMI**	**Female High symptom burden White race Worse renal function**	**Older Atrial fibrillation Low BMI**
Proportion	0.14	0.13	0.31	0.41
Female sex	0.43	0.35	0.61	0.54
Diabetes	0.59	0.37	0.52	0.32
Age	0.22	0.18	0.34	0.39
NYHA III/IV	0.18	0.15	0.28	0.16
Race				
Asian	0.03	0.35	0.01	0.17
Black or African American	0.05	0.01	0.03	0.01
Other	0.06	0.02	0.03	0.04
White	0.87	0.62	0.93	0.77
Age[1]	-1.09	-1.08	0.16	0.60
Body Mass Index[1]	0.92	-0.65	0.75	-0.67
eGFR[1]	0.63	0.85	-0.36	-0.22

The rate of CV death was highest in phenotype 4 (older, AF). The HR for phenotype 4 compared with phenotype 1 (lowest rate of event) was 2.26 [95%CI 1.57-3.26]. The highest rate of heart failure hospitalisation was in phenotype 3 (female, high symptom burden, poor renal function) with a hazard ratio compared to phenotype 2 (lowest rate) of 2.09 [95%CI 1.57-2.80]. There was no interaction between treatment and phenotype group on occurrence of either outcome (Table 6-14).

Table 6-14. Outcomes according to phenotype group for Model 2 with continuous variables
Outcomes according to phenotype group for model 1 (continuous variables) and effect of randomised treatment in each phenotype subgroup. [1]adjusted for randomised treatment

Phenotype	1	2	3	4	interaction p-value
Cardiovascular death					
Rate	1.77 (1.27 - 2.48)	2.48 (1.85 - 3.34)	2.6 (2.16 - 3.13)	3.86 (3.39 - 4.4)	
Hazard ratio[1]	REF	1.38 (0.87 - 2.17)	1.52 (1.03 - 2.23)	2.26 (1.57 - 3.26)	
Treatment effect in each phenotype group	1.20 (0.61 - 2.37)	1.30 (0.71 - 2.36)	0.86 (0.59 - 1.24)	0.93 (0.72 - 1.21)	0.66
1st HFH					
Rate	6.58 (5.48 - 7.92)	3.51 (2.72 - 4.53)	7.44 (6.63 - 8.35)	7.04 (6.36 - 7.79)	
Hazard ratio[1]	2.05 (1.48 - 2.83)	REF	2.09 (1.57 - 2.8)	1.95 (1.48 - 2.58)	
Treatment effect in each phenotype group	1.00 (0.69 - 1.44)	1.04 (0.62 - 1.74)	0.78 (0.62 - 0.98)	0.96 (0.79 - 1.18)	0.46

6.2.5 Validating Model 3 in PARAGON-HF trial data

6.2.5.1 LCA using categorised variables

Model fit criteria were compared for LCA solutions of 1-7 classes. There was a loss in benefit from 4 classes onward (Figure 6-5). The number of seeds coming to the same solution was higher for the 5-class solution than for a 4 class solution (99% and 51% respectively), therefore a 5 class solution was selected. The median posterior probability of class membership was fair (ranging from 0.69 to 0.82) (Table 6-15).

Figure 6-5. Model fit statistics for Model 3 with categorical variables
AIC and BIC for latent class solution using categorised variables for Model 2.

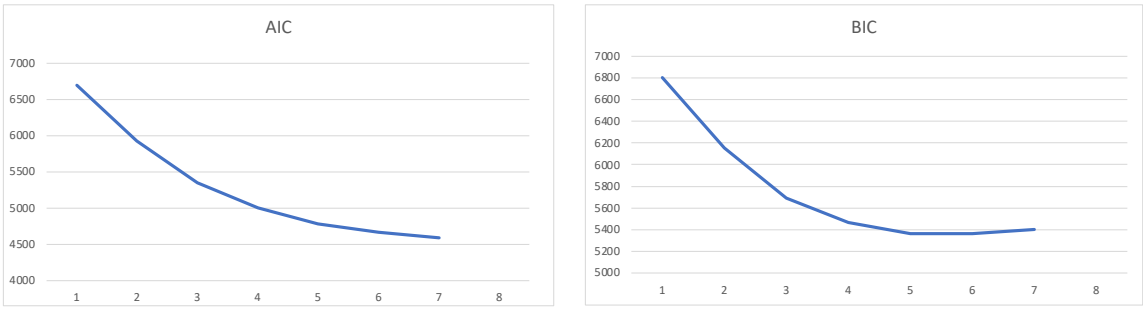


Table 6-15. Median posterior probability of class membership for Model 3 with categorical variables

	Probability class 1	Probability class 2	Probability class 3	Probability class 4	Probability class 5
1	0.722 [0.573 - 0.862]	0.0109 [0.0014 - 0.0661]	0.0368 [0.00674 - 0.147]	0.0198 [0.00562 - 0.113]	0.000414 [0.000097 - 0.0638]
2	0.0226 [0.00524 - 0.0909]	0.718 [0.59 - 0.864]	0.00556 [0.001 - 0.0349]	0.0641 [0.0161 - 0.171]	0.000281 [0.0000746 - 0.08]
3	0.0489 [0.0116 - 0.159]	0.000224 [0.0000146 - 0.00319]	0.693 [0.555 - 0.83]	0.01 [0.00204 - 0.0674]	0.0723 [0.000437 - 0.22]
4	0.0272 [0.00506 - 0.0926]	0.0191 [0.00252 - 0.093]	0.0175 [0.00399 - 0.0879]	0.828 [0.626 - 0.932]	0.000704 [0.000335 - 0.00146]
5	0.0219 [0.00381 - 0.0922]	0.00578 [0.000574 - 0.05]	0.0292 [0.00989 - 0.117]	0.0183 [0.00737 - 0.0429]	0.784 [0.628 - 0.902]

The five-class solution is described in Table 6-16 using item response probabilities.

Table 6-16. Phenotype characteristics for Model 3 with categorical variables Description of groups in model 3 (categorised variables) using item response probabilities, with variables most different between groups highlighted to aid interpretation.

Phenotype	1	2	3	4	5
Descriptors	**AF Alcohol High haemato-crit**	**Younger CAD**	**Oldest Low BMI Low haemato-crit**	**Male Obese CAD Dyslipid-aemia Diabetes**	**Female Obese Diabetes Low haemato-crit**
Proportion	0.19	0.15	0.20	0.20	0.27
Female sex	0.36	0.32	0.60	0.07	0.99
Age					
<60	0.32	0.80	0.07	0.32	0.28
60-70	0.52	0.20	0.41	0.48	0.52
>70	0.16	0.0006	0.52	0.20	0.20
BMI					
<18.5	<0.00001	0.004	0.007	<0.00001	0.004
18.5-25	0.15	0.14	0.43	0.05	0.05
25-30	0.37	0.38	0.50	0.31	0.24
>30	0.48	0.48	0.64	0.64	0.71
AF	0.64	0.08	0.32	0.33	0.24
CAD	0.13	0.70	0.42	0.68	0.36
DM	0.15	0.48	0.26	0.73	0.50
Dyslipidaemia	0.34	0.67	0.47	0.83	0.68
Alcohol	0.70	0.59	0.46	0.67	0.31
CKD					
eGFR >90	0.09	0.32	0.05	0.03	0.03
60-90	0.52	0.64	0.32	0.32	0.38
30-60	0.39	0.04	0.62	0.64	0.58
15-30	0.003	<0.00001	0.01	0.01	0.02
Haematocrit					
>0.5	0.16	0.09	0.003	0.08	0.01
0.4-0.5	0.78	0.77	0.54	0.60	0.53
0.3-0.4	0.07	0.14	0.44	0.32	0.45
<0.3	<0.00001	0.002	0.01	0.002	<0.00001

The highest rate of CV death was in phenotype 4 (male, multimorbidity including diabetes, dyslipidaemia and obesity). Compared to the lowest rate phenotype 1 (AF, high alcohol use, high haematocrit), the HR for CV death was 1.74 (95%CI 1.24-2.44). The highest and lowest rates for heart failure hospitalisation were

the same as CV death, with a HR of 1.64 (95%CI 1.29-2.09) between groups. There was no significant interaction between phenotype and randomised treatment on the occurrence of any outcome (Table 6-17).

Table 6-17. Outcomes according to phenotype group for Model 3 with categorical variables Outcomes according to phenotype group for model 3 (categorical variables) and effect of randomised treatment in each phenotype subgroup. [1] adjusted for randomised treatment

Phenotype	1	2	3	4	5	interaction p-value
Cardiovascular death						
Rate	2.21 (1.69 - 2.9)	2.56 (1.97 - 3.33)	3.81 (3.12 - 4.65)	3.86(3.16 - 4.71)	2.69(2.24 - 3.23)	
Hazard ratio[1]	REF	1.13 (0.77 - 1.65)	1.7(1.21 - 2.38)	1.74(1.24 - 2.44)	1.21(0.87 - 1.67)	
Treatment effect in each phenotype group	0.99 (0.57 - 1.69)	1.43 (0.84 - 2.44)	0.87 (0.58 - 1.30)	0.79 (0.53 - 1.18)	0.96 (0.66 - 1.39)	0.48
1st HFH						
Rate	4.66 (3.85 - 5.65)	5.39 (4.47 - 6.51)	8.09 (7 - 9.34)	8.72 (7.57 - 10.04)	6.36 (5.61 - 7.2)	
Hazard ratio[1]	REF	1.22 (0.93 - 1.61)	1.61 (1.26 - 2.05)	1.64 (1.29 - 2.09)	1.33 (1.05 - 1.67)	
Treatment effect in each phenotype group	0.96 (0.65 - 1.41)	1.11 (0.76 - 1.62)	1.07 (0.80 - 1.43)	0.94 (0.71 - 1.24)	0.70 (0.55 - 0.91)	0.17

6.2.5.2 LCA using continuous variables

The variables used for model 3 were handled as continuous variables where possible (age, body mass index, estimated glomerular filtration rate and haematocrit) and latent class solutions for 2-7 classes fit using the generalised structural equation modelling capabilities in STATA. There was a loss of benefit after 5 classes, therefore a 5-class solution was selected. This was a stable solution with 97% seeds achieving the same solution (Figure 6-6). Median posterior probability of class membership was good, ranging from 0.75 to 0.80 (Table 6-18).

Figure 6-6. Model fit statistics for Model 3 with continuous variables
AIC and BIC for latent class solution using continuous variables for Model 3.

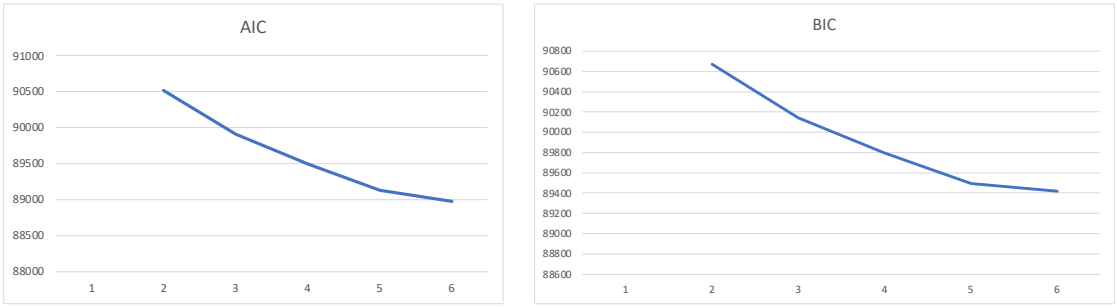


Table 6-18. Median posterior probability of class membership for Model 3 with continuous variables

	Probability class 1	Probability class 2	Probability class 3	Probability class 4	Probability class 5
1	0.786 [0.59 - 0.917]	0.000000121 [0.0000000633 - 0.000000271]	0.0253 [0.0055 - 0.0937]	0.0132 [0.00234 - 0.0752]	0.0558 [0.0138 - 0.169]
2	0.0169 [0.00636 - 0.0483]	0.759 [0.619 - 0.888]	0.00397 [0.000854 - 0.0185]	0.00453 [0.000585 - 0.0345]	0.104 [0.0323 - 0.251]
3	0.0253 [0.00604 - 0.0931]	0.0000000226 [0.00000000693 - 0.000000106]	0.749 [0.559 - 0.893]	0.0196 [0.0032 - 0.105]	0.0512 [0.011 - 0.159]
4	0.0136 [0.00275 - 0.0757]	0.000000043 [0.00000000609 - 0.0477]	0.0284 [0.00614 - 0.104]	0.8 [0.601 - 0.939]	0.00503 [0.000724 - 0.0367]
5	0.00803 [0.00174 - 0.0479]	0.0556 [0.000000102 - 0.22]	0.0132 [0.00214 - 0.0587]	0.000889 [0.000136 - 0.0073]	0.764 [0.604 - 0.905]

This model was described using item response probabilities, with key features highlighted and used to create the phenotype descriptions (Table 6-19).

Table 6-19. Phenotype characteristics for Model 3 with continuous variables
Description of groups in model 3 (continuous variables) using item response probabilities, with variables most different between groups highlighted to aid interpretation. [1]Continuous variable item response are on a standardised scale (mean of 0 and standard deviation of 1); values above 1 infer the mean value in that phenotype is higher than the population mean.

	1	2	3	4	5
Descriptors	**Male CAD Diabetes Dyslipidaemia**	**Female High BMI Low haemato-crit**	**AF Alcohol Younger High haemato-crit**	**Young Not AF Good renal function**	**Older Low BMI Female**
Proportion	0.18	0.24	0.14	0.140	0.31
Female sex	0.07	1.00	0.17	0.35	0.63
AF	0.27	0.23	0.64	0.08	0.40
CAD	0.74	0.39	0.23	0.63	0.33
Diabetes	0.75	0.54	0.23	0.48	0.23
Dyslipidaemia	0.87	0.72	0.47	0.63	0.42
Alcohol	0.67	0.31	0.72	0.56	0.50
Age[1]	0.04	0.005	-0.33	-1.15	0.64
BMI[1]	0.20	0.54	0.23	0.02	-0.63
eGFR[1]	-0.32	-0.31	0.22	1.26	-0.26
Haematocrit[1]	-0.07	-0.40	1.05	0.29	-0.26

The highest rate of CV death was in phenotype 1 (male, diabetes, dyslipidaemia, poor renal function) with the lowest rate in phenotype 4 (young, good renal function). The corresponding hazard ratio for phenotype 1 compared to phenotype 4 was 2.12 (95%CI 1.45-3.09). The highest rate for HF hospitalisation was phenotype 1, with lowest rate in phenotype 3 (younger, AF, alcohol). The corresponding hazard ratio was 1.95 (95%CI 1.48-2.57) (Table 6-20). There is a suggestion of differential treatment effect between phenotype groups for the HF hospitalisation outcome (p for interaction 0.03) with the phenotype with high proportion of female patients gaining more benefit (phenotype 2).

Table 6-20. Outcomes according to phenotype group for Model 3 with continuous variables Outcomes according to phenotype group for model 3 (categorical variables) and effect of randomised treatment in each phenotype subgroup. [1] adjusted for randomised treatment

Phenotype	1	2	3	4	5	interaction p-value
Cardiovascular death						
Rate	4.15(3.4 - 5.06)	2.83 (2.32 - 3.43)	2.26 (1.65 - 3.08)	2.03 (1.48 - 2.78)	3.27 (2.77 - 3.86)	
Hazard ratio[1]	2.12 (1.45 - 3.09)	1.43 (0.99 - 2.08)	1.15 (0.74 - 1.79)	REF	1.66 (1.16 - 2.39)	
Treatment effect in each phenotype group	1.17 (0.79-1.75)	1.07 (0.73-1.59)	0.66 (0.35-1.25)	1.22 (0.65-2.29)	0.81 (0.58-1.13)	0.42
1st HFH						
Rate	9.31 (8.09 - 10.72)	7.06 (6.2 - 8.05)	4.2 (3.33 - 5.31)	4.57 (3.67 - 5.68)	6.82 (6.05 - 7.69)	
Hazard ratio[1]	1.95 (1.48 - 2.57)	1.65 (1.26 - 2.15)	REF	1.13 (0.82 - 1.55)	1.51 (1.16 - 1.97)	
Treatment effect in each phenotype group	0.91 (0.69-1.21)	0.68 (0.52-0.88)	0.71 (0.441.15)	1.06 (0.68-1.64)	1.16 (0.91-1.48)	0.03

6.3 Results

6.3.1 Comparison between identified groups using categorised versus continuous data

6.3.1.1 Model 1

The variables for model 1 included variables describing patient demographics (age and sex), indicating severity of heart failure by symptom burden (NYHA class), and recording comorbidity in both cardiovascular disease as well as renal, metabolic, and respiratory illness. eGFR was the only laboratory value used.

Both methods best fitted with a three-class solution. The median probability of class membership was marginally higher for the model using continuous variables (range 0.80-0.90 for continuous variables, 0.79-0.83 with categorical variables).

Both models strongly identified a phenotype of older patients with high prevalence of atrial fibrillation. These phenotypes had similar crude rates of cardiovascular death (3.47 [95%CI 2.97-4.04] events per 100 patient years in the

categorical analysis: 3.65 [95%CI 3.23-4.14] in continuous analysis. This phenotype had the middle rate for HF hospitalisation in both analyses.

The class with the next highest median posterior probability of class membership (indicating good discrimination between classes) was similar in both studies. This phenotype was young male patients, with a higher number with ischaemic aetiology of heart failure and good renal function. They had the lowest rate of both outcomes examined in both analyses.

The third class had similar characteristics in both analyses. These patients were more likely to be female and had high comorbidity burden in terms of hypertension, obesity, diabetes and COPD.

Treatment effect of sacubitril-valsartan was examined in each phenotype group in both analyses. The point estimate for the treatment effect on CV death was lowest in the older & AF phenotype. The point estimate for the HF hospitalisation outcome was lowest in female patients with multimorbidity in both analyses. The test for interaction between phenotype and randomised treatment was positive for the HF hospitalisation outcome in the categorical analysis only.

6.3.1.2 Model 2

This model includes more variables describing the patient demographics (age, sex and race), severity of HF symptoms (NYHA class) and comorbidity including diabetes, obesity, renal dysfunction and atrial fibrillation.

Both methods found a four-class solution best described the data. The median posterior probability was higher (better) in the categorical analysis (range 0.79-0.88 for categorical analysis, 0.77-0.89 for continuous analysis).

The class with the highest median probability, indicating a well differentiated class, was older patients with atrial fibrillation. These patients had the highest crude rate of CV death in both analyses. They had the second highest rate of HF hospitalisation in both analyses.

Both analyses identified a phenotype of younger male patients, more likely to be of Asian race and with low body mass index. The crude rate of CV death for this phenotype was similar in both analyses. The risk of CV death, compared with the other phenotypes, was middling in both analyses. This phenotype had the lowest crude rate of HF hospitalisation in both analyses.

Another identified class in both analyses were younger patients with diabetes and obesity. This phenotype had the lowest crude rate of CV death in both analyses and middling risk (compared to other phenotypes) of HFH.

The last identified class were predominantly female patients of White race with higher (worse) NYHA functional class scores. Renal function tended to also be lower.

The lowest point estimate for effect of randomised treatment on CV death was in the older/AF phenotype in the categorical analysis and for females with high symptom burden in continuous analysis. The lowest point estimate for HF hospitalisation was in diabetic/obese phenotype in categorical analysis.

6.3.1.3 Model 3

Model 3 variables include those describing patient demographics, comorbidity (both cardiovascular and other significant diseases), history of alcohol intake and laboratory measurement of renal function and haematocrit. Measurement of severity of heart failure is not included.

The latent class solution using both techniques came to a 5-class solution, with slightly higher (better) posterior probability of class membership in the continuous variable solution.

In the categorical analysis, the class with the highest posterior probability of class membership, indicating a well-defined group, was the phenotype of male patients with obesity, coronary artery disease, dyslipidaemia and diabetes (Group 4). In the continuous analysis, a similar phenotype was identified and also had good posterior probability of class membership. This group had the

highest crude rates of cardiovascular death and first heart failure hospitalisation in both the categorical and continuous analysis.

The phenotype with lowest rates of events in the categorical analysis were distinguished by high prevalence of atrial fibrillation and of alcohol use, of middling age, with a similar finding in the continuous analysis.

There was no significant interaction between phenotype group and randomised treatment on the occurrence of cardiovascular death. There was a significant p for interaction in the continuous variable analysis of model three for the HF hospitalisation outcome. This should be interpreted with caution given the multiple testing in this analysis (p value for significance has not been adjusted for this). The group that gained greater benefit were the female group with high body mass index (phenotype 2). Those with less benefit were older patients with low BMI (phenotype 5).

6.3.2 Identification of consistent subgroups

Within each model, the continuous and categorical type of analysis found similar consistent phenotype groups. Some phenotypes were identified with use of different variables across models.

The first recurring phenotype is that of older patients with high prevalence of atrial fibrillation and low BMI. This was clearly identified in models 1 and 2 using both categorical and continuous techniques. In model 3 (categorical variables), that has more phenotype groups, it is likely these patients are split between two phenotype groups, the first being high prevalence of AF (phenotype 1) and the other being older patients with low BMI (phenotype 3). This group had the highest crude rate of cardiovascular death in model 1 and 2, but not the highest in first heart failure hospitalisation.

The second consistently identified group is that of patients with multiple comorbidities including diabetes, dyslipidaemia and obesity. How this group is described clearly varies by the included baseline variables, but this group appears to be identified consistently. In model one, they are described with hypertension, diabetes and obesity. In model two, diabetes and obesity are

included. In model 3, there are two groups with obesity, dyslipidaemia and diabetes. In model one and two, the phenotype is split fairly evenly between males and females. In model 3, there are two similar groups with comorbidity, the main difference between them being one is predominantly female and the other being predominantly male. These phenotype groups had relatively high rates of heart failure hospitalisation throughout the models.

Another group that is often identified is younger patients with ischaemic aetiology of heart failure who are younger and predominantly male and generally have good renal function. In model 2, there is no input variable to describe coronary disease, but there is a group of younger male patients identified. These groups tended to have lower rates of both outcomes.

6.4 Discussion

Identifying groups of phenotypically similar groups of patients has many potential benefits. For the patient and treating physician, understanding different phenotype may enable clearer communication about likely prognosis. An area of great interest is precision medicine, potentially tailoring treatments to patients where the treatment is more likely to be efficacious and well tolerated. Developing this knowledge would require data driven methods to define patient phenotypes, rather than relying on the impression of an individual researcher or treating physician. What is reassuring in this analysis is that phenotypes that appeared consistently are clinically recognisable, suggesting this unsupervised method of clustering patients is effective at describing these patients. In addition, variations in technique lead to similar clustering results. This is important, as the categorisation approach is simpler to implement and therefore is achievable and approachable by a wider range of researchers.

An interesting difference between the models is the selection of types of variables utilised to generate the latent class solution. For example, in model 2 which includes NYHA score, in the output there is a phenotype which is predominately distinguished by high symptom burden and high proportion of female sex. This phenotype has high rates of heart failure hospitalisation, which is perhaps not surprising. The phenotype group itself does not add much to our understanding of the pathophysiology underlying the condition. Perhaps high

NYHA score, as a reflection of severity and advanced state of illness, could be used alongside phenotype group to help understand risk rather than as a variable to define the phenotype groups themselves.

Exploring differences in rates of outcomes across phenotype groups had some expected results and some that are more interesting. It is not surprising that the phenotype groups with more older patients had higher rates of mortality outcomes, and this analysis cannot distinguish the relative importance of different variables used to make the phenotype groups and the risk of outcome and age may have a large effect on the rate of mortality outcomes. However, it is interesting that younger patients with high morbidity burden had high rates of heart failure hospitalisation, both important to the individual and for healthcare provision and planning. These patients are likely to live with this chronic disease for a long time, therefore interventions which result in reduced hospitalisations is of great value.

The consistently identified phenotype groups share some similarities with phenotypes identified in patients with heart failure with reduced ejection fraction in Chapter 5, particularly those with multimorbidity/metabolic cardiomyopathy (diabetes, hypertension, obesity), predominantly ischaemic aetiology, and older patients with high burden of atrial fibrillation. Although there is a focus on the differences between HFPEF and HFREF, perhaps there should also be a greater focus on similarities throughout the ejection fraction spectrum. This is of particular importance now we have treatments which have been shown to be effective across the spectrum of ejection fraction.[31,32]

A question remains as to how to investigate potential differences in treatment effect across phenotype groups. Any phenotype group allocation has a degree of uncertainty allocated to it. This is minimised as far as possible when creating the latent class solution, but when applied to new data there will continue to be unmeasured uncertainty in class allocation. Therefore, any true difference in treatment response between phenotype groups will be diluted by uncertainty in class membership. There is ongoing research interest in exploring relationship between latent class and distal outcomes, but these do not generally incorporate treatment effect.[170,171] Therefore a currently applicable approach would require subjects to be allocated on modal assignment, or most likely class

allocation, and would require a higher number of participants to detect a true difference in treatment response between phenotype groups. This may not be easily applicable in a clinical trial setting, but may be possible using routinely collected data, particularly if variables required to create phenotype groups are kept reasonably simple.

In conclusion, this analysis reinforces that consistent latent class groups can be identified using a variety of input variables and with slight differences in analytical techniques. The identified phenotypes are clinically logical and recognisable. A data driven method that can identify these phenotypes opens doors for further personalisation of heart failure care in terms of prognostication and in the future potentially personalised treatment strategies.

Chapter 7 Supervised machine learning in HFREF

7.1 Introduction

Two broad approaches to personalisation of treatment of patients with heart failure have been presented in this thesis so far. In the first, covered in Chapters 3 and 4, I have extracted additional information about future risk of adverse events following a heart failure hospitalisation and mapped the average patient journey for patients with HFREF and HFPEF. This additional information equips us to prioritise treatments and have better informed discussions with patients about how their disease might progress. Overall, there were more similarities between HFREF and HFPEF than differences. HFPEF and HFREF are only very broad phenotypes of heart failure, therefore in chapter 5 and 6 I used latent class analysis to define phenotypes in patients with HFPEF and HFREF, showing some consistent results in terms of the groups identified in different analyses in different data. The general risk of different outcomes can be described for these different phenotypes, and perhaps in the future phenotype specific treatments may be identified. This approach still describes the average trajectory for similar patients within a cohort, rather than making a specific prediction for an individual patient. In this last analysis, I use supervised machine learning methods to see if improved individual prediction for cardiovascular death can be made compared with previously developed risk prediction models.

Supervised machine learning methods use data inputs with a known, or labelled, output and create a function to predict the output from the input. This function is then applied to new data to assess the predictive performance of the model. In general, machine learning techniques can perform well in prediction of events but less well in causality as the function to predict the output from input maps both direct and indirect effects of the inputs on the outputs. The general process of supervised machine learning is described further in Chapter 2.

There are various types of supervised machine learning techniques which can be utilised, with commonly used methods in heart failure research including penalised linear regression, K-nearest neighbours, random forests, gradient

boosted trees and artificial neural networks. The basis for these techniques is given in Chapter 2. These models benefit from greater flexibility compared with statistical techniques that were used to create commonly used risk scores such as Poisson regression, multivariable logistic regression and Cox regression models.[68,69,75] The usual requirements for analysis, such as the requirement for normal distribution, are relaxed. Supervised machine learning methods are also well suited to analysis of 'big data', or data with many participants and multiple measured variables.

As discussed in Chapter 1, there is no approach that appears to be consistently superior in event prediction in heart failure, and often supervised machine learning performance is similar to more simple logistic regression approach.

To explore the utility of supervised machine learning in a large clinical trial dataset, an analysis of the PARADIGM-HF trial was undertaken.[14,172] This is the same data that was used to create the PREDICT-HF model and can be used as a comparison in performance of a model built using a Cox regression model and each supervised machine learning approach.[68]

7.2 Methods

7.2.1 Trial population

The Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF) investigated the addition of neprilysin inhibition to ACE inhibition in patients with heart failure and reduced ejection fraction. A total of 8399 patients were randomised. As this was a positive trial, the randomised treatment allocation was included as an input in these analyses. A more detailed description of the trial is given in Chapter 2.

7.2.2 Software

For these analyses, R was used with the *tidymodels* suite of packages designed to facilitate machine learning analyses.[133,173]

7.2.3 Inputs

In order to provide a fair comparison of the techniques when compared with the PREDICT-HF model, all variables considered in the creation of the PREDICT-HF model were considered in the machine learning methods. Many variables were considered for PREDICT-HF but not used in the final model. As machine learning can generally handle big data, all the considered variables were included. This was to mainly compare techniques, as the use of a very large number of variables in the input would make this a less usable risk prediction tool in the clinical setting. However, filtering the variables would result in an unfair comparison of techniques and may result in loss of benefit of the machine learning model which can tolerate a very high number of input variables. A complete case analysis approach was utilised rather than an imputation technique as an acceptable number of datapoints were included in the complete case analysis approach (7395). The included variables are listed in Table 7-1.

Table 7-1. Input variables to the machine learning models
Variables in the PARADIGM-HF dataset considered as inputs in supervised machine learning analyses.

Variable	Type of variable
Age	Numeric
Sex	Factor
Race	Factor
Region	Factor
Body mass index	Numeric
Weight	Numeric
Hip: waist ratio	Numeric
Smoking history	Factor
Alcohol history	Factor
Randomised treatment	Factor
Past medical history	
Chronic obstructive pulmonary disease	Factor
Stroke or transient ischaemic attack	Factor
Valvular heart disease	Factor
Atrial fibrillation	Factor
Hypertension	Factor
Diabetes	Factor
Myocardial infarction	Factor
Previous percutaneous coronary intervention	Factor
Previous coronary artery bypass graft	Factor
Peripheral arterial disease	Factor
Asthma	Factor
History of cancer	Factor

Renal disease	Factor
Anaemia	Factor
Heart failure characteristics	
LBBB on ECG	Factor
QRS duration	Numeric
Systolic blood pressure	Numeric
Diastolic blood pressure	Numeric
Pulse pressure	Numeric
Heart rate	Numeric
Left ventricular ejection fraction	Numeric
New York Heart Association class	Factor
Aetiology of heart failure	Factor
Prior history of heart failure hospitalisation	Factor
Rales	Factor
Peripheral oedema	Factor
Dyspnoea on effort	Factor
Fatigue	Factor
Paroxysmal nocturnal dyspnoea	Factor
Third heart sound	Factor
Laboratory variables	
N-Terminal Pro–B-Type Natriuretic Peptide	Numeric
Estimated glomerular filtration rate	Numeric
Serum creatinine	Numeric
Albumin	Numeric
Haemoglobin	Numeric
LDL cholesterol	Numeric
HDL cholesterol	Numeric
Triglycerides	Numeric
Total cholesterol	Numeric
Uric acid	Numeric
Sodium	Numeric
Potassium	Numeric
BUN	Numeric
Chloride	Numeric
Total bilirubin	Numeric
Alkaline phosphatase	Numeric
Total protein	Numeric
Calcium	Numeric
Platelet count	Numeric
Red blood cells	Numeric
Haematocrit	Numeric
Neutrophils	Numeric
Basophils	Numeric
Eosinophils	Numeric
Lymphocytes	Numeric
Monocytes	Numeric
Treatments	
Aldosterone antagonist	Factor
Beta blocker	Factor

Digoxin	Factor
Aspirin	Factor
Anticoagulant	Factor
Diuretic	Factor
Lipid lowering	Factor
Implantable cardioverter defibrillator	Factor
Cardiac resynchronisation therapy	Factor

7.2.4 Outcome

The outcome of the assessed machine learning approaches utilised here is a binary outcome of "event" or "no event". Rather than using the binary censor for event/no event over the full trial follow up, a binary marker for event or no event was created for the 2 year follow up time point i.e., whether the patient had an event before 2 years or was censored. The outcome of interest was selected as cardiovascular death, again allowing comparison with the PREDICT-HF risk prediction model.

7.2.5 Train/test split

The same training and test split was used for all the machine learning approaches, controlled by using the same random seed. The data was split in an 80:20 fashion, 80% of data used for training and model development and 20% used for testing each model. The split was stratified by the endpoint to ensure a proportional split of events and non-events in the training and test data.

A ten fold cross validation approach was taken to split the training data for finding optimum hyperparameters in model tuning. This again was stratified by the endpoint variable. Cross validation for hyperparameter tuning is explained in more detail in Chapter 2. In short, 9 of the folds are used to create the model using a range of hyperparameters and the model is tested on the 10^{th} fold. This is repeated multiple times, using a different fold of the data for testing each time.

7.2.6 Hyperparameter tuning

For each model, there are different hyperparameters that need to be tuned to optimise the performance of the machine learning model. Tuning methods aim

to minimise the test error of the method and bring the machine learning function as close as possible to the true (unknown) function that relates the inputs to outputs, while avoiding overfitting to the training data. To facilitate tuning, tuning grids including a range of each hyperparameter were generated and the model fit iteratively over the cross validated folds using different combinations of the hyperparameters to find the best performing combination.

7.2.7 Regularised logistic regression

The first model tested in the PARADIGM-HF data was a regularised logistic regression model. Pre-processing steps included creating dummy variables for factor variables, removing variables with only a single value or zero variance (this is possible for some splits in the data), and numeric predictors normalised to a mean of zero and standard deviation of 1.

The tuning parameters in this analysis were the penalty term and the mixture term. The penalty term, also called the lambda value, describes how much the coefficient for the variable is shrunk. The mixture term determines how much of each of a lasso or ridge regression model is used. A mixture term of 0 is a model that uses pure ridge regression, while a mixture term of 1 uses pure lasso regression. Another term for models that use a mixture of lasso and ridge regression is elastic net. Hyperparameter tuning was carried out using a range of 20 penalty measures and 6 mixture terms in a tuning grid. Hyperparameter tuning was performed using the 10 cross validated folds of the training data, and the best fitting hyperparameters were selected based on the highest area under the receiver operator curve. The hyperparameter values corresponding to the highest AUC were then applied to generate the final model which was fit to all the training data and tested on the test data, giving the final AUC value.

7.2.8 Random forest

Random forest analysis was the next model assessed. Random forests do not require any data pre-processing. The first tuning parameter is the number of randomly selected parameters to be considered at each split in the decision tree. The second is the minimum node size, or the smallest number of datapoints at the terminal nodes of the random forest. Another hyperparameter

that can be controlled is the number of random trees to be grown, but this was not tuned. A regular tuning grid with 5 levels for each hyperparameter, generated using the automated processes in the *tidymodels* suite, identified that lower numbers of both hyperparameters provided a higher ROC AUC. Therefore, a manual tuning grid was created using five values over the lower range of both hyperparameters. The best performing combination of hyperparameters was applied to the whole training data then tested on the test data to give a final ROC AUC.

7.2.9 Single layer neural network

For the neural network analysis, R and R studio was used to run the *keras* neural network software in Python. Pre-processing steps were to create dummy variables for the factor input variables and continuous variables were normalised to a mean of zero and standard deviation of 1 after log transformation to normalise distribution if appropriate. The hyperparameters to be tuned are the number of units in the hidden model, the number of training iterations or 'epochs', and the dropout or proportion of model parameters randomly set to zero during model training. Again, the best performing hyperparameters in testing using the cross validated folds were applied and the final ROC AUC calculated in the test data.

7.2.10 Gradient boosted machine

Pre-processing for gradient boosted machine was to create dummy variables for factor variables as the method requires numeric input variables and remove variables with zero variance (or only one value). There are multiple tuning parameters for gradient boosted machines. Some overlap with random forests, namely the minimum size of the terminal node of the tree and the number of trees grown to make the random forest. There is a tuning variable for the depth, or complexity, of each tree. In gradient boosted trees, each new tree depends on the result of the previous trees and other parameters control the learning rate of that process. These tuning parameters are the learning rate and loss reduction.

7.3 Results

7.3.1 Regularised logistic regression

In the tuning step, iteratively fitting the model with different hyperparameters using the 10-fold cross validated data folds, the best performing hyperparameters were compared using the ROC AUC. The best fitting model had a penalty term of 0.0298 and a mixture term of 0.05, meaning it was a greater proportion of ridge regression in the elastic net model.

The final fitted model had an accuracy of 0.88 and a ROC AUC of 0.67. The confusion matrix is shown in Table 7-2. It should be noted that confusion matrices in this analysis are less informative, as there is an imbalance in the outcome. In a standard confusion matrix, the prediction for event/no event will be based on a cut point of 50% probability. When the event is imbalanced, it may be beneficial to move the decision point to predict "event" if the probability is at a lower level, for example 20% probability, to create a greater proportion predicted as an event. This lessens the chance of a false negative when the outcome is important not to miss. However, to compare models, the focus in this analysis is comparing the ROC AUC which gives a measure of the ability to discriminate between positive and negative cases over a range of cut points and is more easily interpretable.

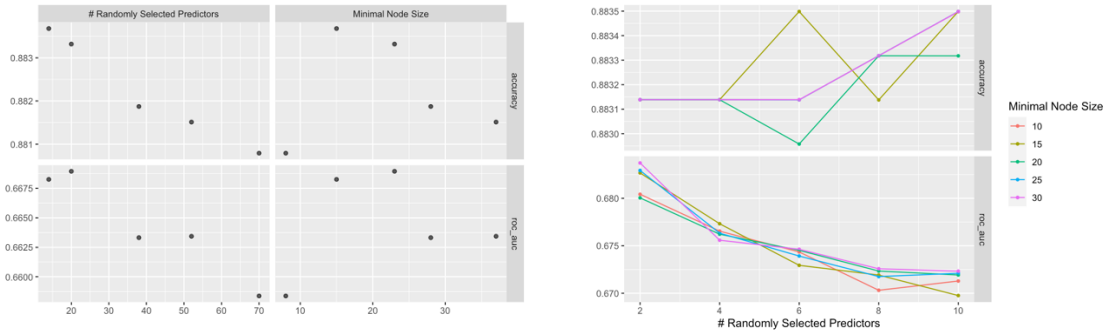
Table 7-2. Confusion matrix for penalised logistic regression model

	Truth	
Prediction	no event	event
no event	1625	215
event	8	2

7.3.2 Random forest

The tuning stage for the random forest model suggested lower randomly selected parameters came to a higher ROC AUC solution, as shown in Figure 7-1. A manual tuning grid with values between 2 and 10 for number of randomly selected parameters and a range of 10 to 30 for the minimal node size was created and tuning repeated, the results of which are given in the right column of Figure 7-1. The number of randomly selected parameters to give the best fitting model was 2 with a minimal node size of 30. This gave an ROC AUC of 0.68.

Figure 7-1. Tuning hyperparameters of the random forest model

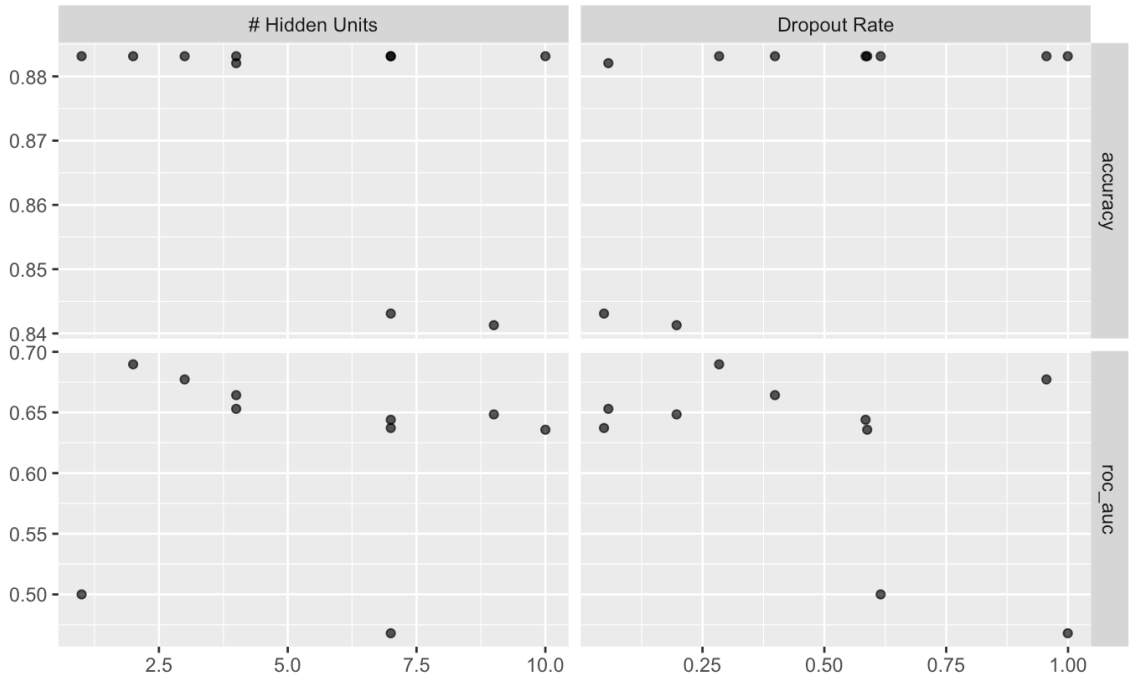


7.3.3 Single layer neural network

To create the tuning network for the neural network, a space filling parameter grid was utilised. The initial tuning results are shown in Figure 7-2. The number

of epochs was not tuned but was selected at 100 to provide sufficient iterations without excessive computational time. By comparing the ROC AUC results, the best fitting model had 2 hidden units and a dropout rate of 0.284. This was fit to the final model and gave a ROC AUC of 0.68.

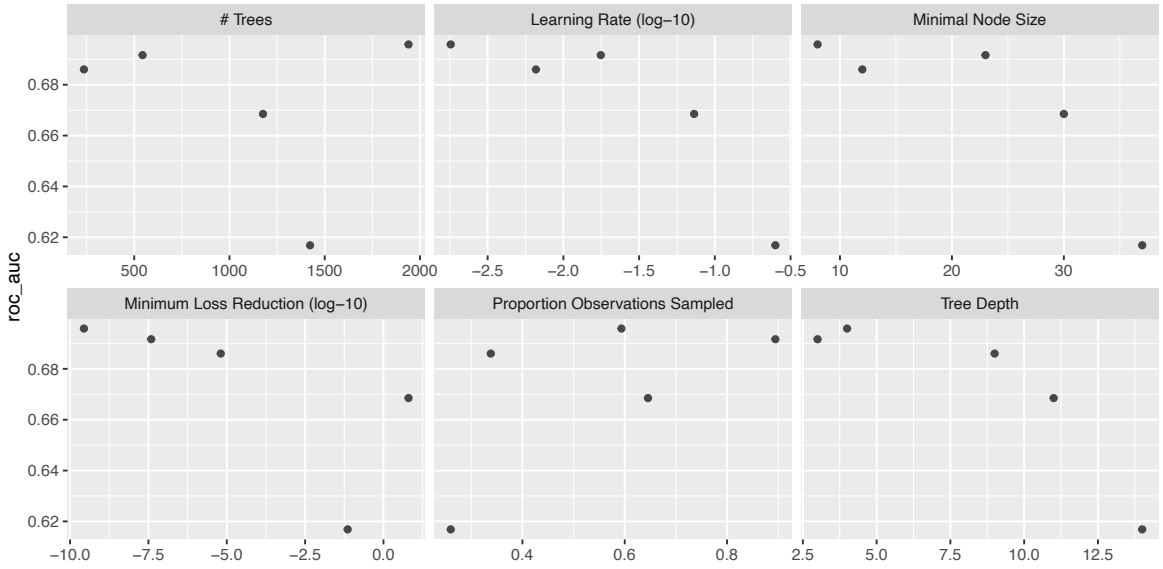
Figure 7-2. Tuning hyperparameters of the single layer neural network model



7.3.4 Gradient boosted machine

Given the larger number of tuneable hyperparameters for the gradient boosted tree model, a tuning grid created using default settings in the *tidymodels* suite was utilised. The best fitting model had 1940 trees, minimal node size of 8, tree depth of 4, learn rate of 0.0018, loss reduction of 2.76e-10 and sample size of 0.59. This is displayed graphically in Figure 7-3. The final ROC AUC in the test data was 0.69.

Figure 7-3 Tuning hyperparameters of the gradient boosted machine model
Hyperparameter tuning using grid search for the gradient boosted machine model. Each panel demonstrates the ROC-AUC results for one of the tuning parameters.



7.3.5 Comparison of model fit statistics

The main comparison between models in this study was using the ROC AUC, which are given in Table 7-3. There is little difference in the performance of the different approaches in predicting the outcome of cardiovascular death at 2 years follow up as a binary event, with a minor improvement in ROC AUC using a gradient boosted machine.

Table 7-3. Comparison of model fit statistics from different machine learning models

Model	ROC AUC
Regularised logistic regression	0.67
Random forest	0.68
Single layer neural network	0.68
Gradient boosted trees	0.69
PREDICT-HF	0.71 0.70 in external validation

7.4 Discussion

As has been found in several other studies, there was little performance benefit in prediction for a binary outcome using the type of data collected in a clinical trial using more complex machine learning methods. Although the input included a lot of detail about the patient demographics, past medical history, heart failure severity and characteristics and laboratory data, this was not sufficient to improve prediction despite more flexible modelling techniques. There are likely several reasons underlying this finding.

Firstly, these analyses use only a binary outcome of event or no event at 2 years of follow up. In the PREDICT-HF prognostic model, survival data was utilised. This provides more granular data about survival time rather than a binary variable to indicate event or no event at a set time of follow up. The PREDICT-HF prediction model was able to achieve a higher C-statistic of 0.71. Although there are more advanced machine learning techniques that include the use of survival data, this is not yet available in a format approachable by a clinical researcher and requires more advanced programming experience.[174] This is an

area that is undergoing development in the *tidymodels* packages and may soon be achievable by a clinical researcher without advanced programming background.[175]

Consideration was given to repeating the models for different outcomes at different time points, however provisional outcomes for a 1 year endpoint were similar using basic models to those presented here. Overall, the weight of evidence would suggest that, for this type of data, there is unlikely to be great improvements in prognostic performance over traditional risk models. It should also be noted the large processing requirements required for each model. Given the lack of suggestion of improved performance with these models, investigation of additional models was deemed futile. In addition, this model was not truly validated in fresh unrelated data. This means that the C-statistics are probably still over-estimated due to some overfitting to the training data i.e., fit specifically to the patients in the PARADIGM-HF trial. Although some external validation would have been possible in the ATMOSPHERE trial, not all of the input variables would have been available. Overall, there was no benefit to be expected from additional external validation due to the baseline modest performance of the models. It should be reiterated that if the model was found to be better performing and worthy of further assessment it would be important to find a patient population in which to externally validate the model to be sure of consistent performance in new populations.

Machine learning techniques are excellent when there is big data, or data with multiple measurements and multiple variables. It may be that this data is not granular enough to be able to fully benefit from the flexibility of the analytical techniques to achieve improved prediction. For example, recording of co-morbidity is binary, for the presence of absence of disease. However, it is clear that a patient with well controlled diabetes on a single pharmacological agent is expected to behave quite differently from a patient with poorly controlled diabetes despite multiple pharmacological agents. In reality, this level of data input would not be practical to obtain for a clinical trial, as it would be extremely time consuming and therefore incur great expense. It is also not the main aim of a trial data collection, as randomisation should control for the many differences we find between patients to allow detection of a treatment effect.

It would also be extremely impractical to have a risk score where all this data had to be manually inputted by the treating physician to generate a risk score as the effort and expense involved in collecting the vast amount of data would be an insurmountable barrier. However, it could be imagined that in the future in a fully integrated electronic health record, this data could be automatically withdrawn to generate a risk score. Clearly this has ethical implications surrounding access to the depth of potentially sensitive patient level data to both develop and use such a model. Natural language processing, or extracting data from free text documents, could be a further avenue to extract useful data from free text of medical records.[176]

Another area that is not addressed in this type of analysis is features of the patient that are more nebulous or subjective, such as frailty. We do have tools to help the physician quantify frailty in patients[177–180], but it is possible machine learning techniques may be able to identify generally unmeasured features from other sources, for example in imaging. Machine learning has been used in imaging to help automate some of the processes in image analysis in a conventional sense (for example, calculating the left ventricular ejection fraction from an echocardiogram), but has also been used to predict outcomes.[180] As it is not a requirement to "tell" the machine learning algorithm what it should be looking for, it is possible that the machine learning method may look both at the automated measurements of ejection fraction and other features of the imaging that relate to outcome. Hypothetically, the algorithm may note the amount of adipose tissue, muscle mass or bone density from a CT scan as well as the expected measurements of left ventricular size and function. This again raises one of the core issues of machine learning that can limit use in this way – often the algorithm results in a 'black box' where the link between the input and output is not human interpretable which can make it unacceptable to the user or patient.

In conclusion, the use of these machine learning approaches at a fairly basic level, accessible to a clinical researcher with interest in this area, did not provide better predictive ability than an established prognostic model utilising survival analysis in a Cox regression model. Accurate event prediction in patients with heart failure remains challenging, and there are several avenues in

which machine learning may be further explored in this field to help achieve this elusive goal. This will likely require close collaborative work between data scientists and clinicians with some fundamental understanding of the underlying techniques.

Chapter 8 Final discussion

Throughout this thesis, I have demonstrated several different approaches using data analysis to try and personalise the care of patients with heart failure. These have involved a wide range of analytical techniques to examine the question from different perspectives.

Firstly, I used a data driven analysis aiming to examine whether the accepted trajectory of the patient journey for patients with heart failure holds true for contemporary patients with heart failure, both with reduced and preserved ejection fraction. In the analysis of heart failure with reduced ejection fraction, the lack of relationship between heart failure hospitalisations and sudden death was confirmed, supporting that they are truly 'sudden', while there was a strong relationship between heart failure hospitalisations and risk of overall cardiovascular death as well as specifically pump failure death. These findings can be acted on in two distinct ways. A patient being admitted to hospital with heart failure does highlight a time of increased risk of adverse outcome and should be taken as an opportunity to optimise their heart failure pharmacological treatment and consider device therapies if appropriate, or indeed make advanced care plans if they appear to be approaching end of life. They may benefit from a period of enhanced outpatient monitoring. In contrast, finding that a patient with heart failure appears to be in a stable condition in terms of lack of hospitalisations when being reviewed in outpatient clinic does not predict the risk of sudden cardiovascular death, a major cause of cardiovascular death in this population. Therefore, for patients in both settings, there should be continued up titration of medical therapies to the maximum tolerated, as per guidelines. It could be argued that this does not provide an avenue for personalisation of care, and this might be true in terms of which treatments should be offered, but it can help the patient understand what their own 'trajectory' might look like and can facilitate discussions with their doctor about the aims and objectives of treatment.

In the paired analysis of the patient trajectory in a heart failure with preserved ejection fraction population, there were similar findings. Although patients with HFPEF tend to be older and to have multiple comorbidities, and therefore are at risk of hospitalisation for many causes other than heart failure, there remains a

strong link between heart failure hospitalisation and cardiovascular death. This is important, as although treatment with SGLT2i has been found to be effective in heart failure and preserved ejection fraction, there is apprehension that this is driven by a reduction in hospitalisation events rather than mortality outcomes. The link between the two types of events, and reduction in rates of both events with treatment, would provide reassurance that these are globally effective. In addition, although these patients did have a high burden of all cause hospitalisation, a considerable proportion of these were due to heart failure and around 60% of deaths were attributed to cardiovascular causes. The patterns of repeated hospitalisation, pump failure deaths and sudden deaths mirrored what was found in the analysis of heart failure with reduced ejection fraction patients, therefore the same arguments could be made that all patients should be considered for optimisation of therapy with an SGLT2i regardless of setting. Again, this data might be found useful to help a patient understand the likely trajectory of their illness.

It is important to recognise that heart failure hospitalisation is only one expression of a worsening heart failure event. In the analysis of HFREF, a sensitivity analysis was included for heart failure hospitalisations and worsening outpatient HF events (defined as emergency room visit for heart failure or intravenous treatment for heart failure) with consistent results. In general, there is no consensus in the definition of a worsening heart failure event and they are not routinely included in clinical trial endpoints. Emergency department visits and observation stays have been reported to represent approximately one half of worsening heart failure events in routine electronic health record data.[181] There is increased interest in reproducibly evaluating and recording these events to learn from them.[182,183] A greater understanding of outpatient events where patients require intensification of treatments would greatly enrich the type of analysis carried out in this work with hospitalisation data.

The next approach was to try a clustering technique, namely latent class analysis, to identify phenotypically similar groups of patients with heart failure and reduced ejection fraction. This data driven approach was able to identify clinically recognisable phenotypes, and some phenotypes were similar to those

identified in previous analyses using different data. Finding both recognisable and consistent groups provides reassurance that genuine phenotype groups are being identified rather than being an artificial finding by modelling error or overfitting of the model to the analysed data. These analyses could be used to help understand the patient's risk of heart failure and mortality outcomes. An avenue for further exploration is to look for differential treatment responses, as was carried out in the analysis of the EMPHASIS-HF and EPHESUS trials.[58]

A major limitation in this type of approach is the degree of uncertainty in allocating any one patient to a phenotype group. Each patient is allocated to the most likely class, be that if they have a 90% probability of being in that class or a 50% probability if they do not closely fit the common characteristics of any phenotype. It would then be extremely difficult to identify if there were a differential treatment effect as this would be diluted by the uncertainty in class membership. It is recognised that allocating a patient to a group if there is a specific causal event (for example, a gene mutation) results in a phenotype that is easy to differentiate. This is in contrast to many chronic diseases where multiple factors build together (for example, low impact gene mutations and environmental exposures) to create a less clearly defined phenotype group.[184] Ultimately, phenotyping may be most useful to identify the core, most commonly seen groups first and understand their prognosis and response to treatment, while accepting patients on the fringes of groups are not as well served with this approach. To generate truly personalised care for the individual, perhaps an approach where a patient can be categorised in a multifaceted manner in a multidimensional space without need to allocate to one phenotype group over another may prove most effective. There is interest in the analysis of distal outcomes together with degree of certainty or weighting of allocation to a latent class phenotype therefore this is something that may have the potential to be modelled more accurately in the future to be able to detect a signal of differential treatment response.

In a similar analysis in heart failure with preserved ejection fraction, the persistence of latent class phenotype groups was assessed by using variables selected in prior analyses to examine whether these were identified again using data from the PARAGON-HF trial. Two slightly different approaches to latent

class analysis were tested, again for consistency. Several phenotypes appeared reliably, with similar outcomes and importantly were clinically recognisable. This would support the hypothesis that these are genuinely occurring phenotype groups in heart failure with preserved ejection fraction that can be identified using a data driven technique.

Although no significant difference in treatment effect was found in these latent class analyses, it is an important method to bear in mind when we consider the development for specific treatments for heart failure of different causes. For example, some specific aetiologies of heart failure now have directed treatments available, such as cardiac amyloidosis.[185] As amyloid can affect other body systems, a phenotyping approach including a wider range of data about patients may be able to identify patients who had a higher likelihood of having amyloid heart disease to facilitate targeting of further investigation. More patients are being diagnosed with transthyretin cardiac amyloidosis over time, and at an earlier stage allowing earlier treatment.[186] Identification of latent class phenotypes that have a higher proportion of patients with amyloidosis could aid in targeted investigation and support earlier treatment. In addition, patients within a latent class group do not always seem to fall in a group of a specific aetiology. Hypothetically, finding of a particular phenotype of patient with heart failure may provide a route to understanding yet unknown aetiologies and pathologies of heart failure. Three mechanistically distinct phenotypes of dilated cardiomyopathy have been identified using a combination of clinical, genetic, cardiac MRI and proteomic assessments analysed using machine learning techniques by Tayal et al.[187] Although the number of patients was relatively small (426 in the derivation cohort), differences in prognosis were identified between phenotype groups and these groups are proposed as potential targets for targeted interventions. Approaches allowing analyses of deep and complex data about patients to help our understanding of pathology, prognosis and treatment options provide exciting opportunities to further personalisation of care in heart failure.

Another valuable aspect of phenotyping could be identifying groups of people who are at high risk of developing heart failure and consider screening. It has been suggested that patients with diabetes, particularly with other risk factors

such as hypertension and obesity, should be screened for heart failure using natriuretic peptides or high sensitivity troponin given the high prevalence of heart failure in patients with diabetes of up to 22%.[188] The phenotype of heart failure patients with comorbidities including diabetes, obesity, hypertension and hyperlipidaemia (summarised in Chapter 5 as metabolic cardiomyopathy of obesity) is found consistently through HFREF and HFPEF analyses and perhaps patients with these clusters of comorbidity without known heart failure should be a particular target for screening. Health infrastructure improvements and significant funding would be required to screen all patients with diabetes for heart failure, therefore initially screening for patients fitting this phenotype group could be the initial focus. The same could be considered for older patients with atrial fibrillation, although the effect of AF on raising natriuretic peptide levels could mean an optimum cut-off becomes more difficult to define.[189,190]

The question remains as to how to manage patients with elevation in natriuretic peptide levels and at risk of heart failure, without the syndrome of heart failure. In the HOMAGE trial[191] patients at risk of developing heart failure due to known coronary disease or with risk factors for coronary disease with modest elevation in NT-proBNP were randomised to placebo or spironolactone. Over follow up, there were changes in type 1 collagen synthesis and degradation in those randomised to spironolactone which was accompanied by favourable changes in left atrial size and left ventricular systolic function. In the PONTIAC trial[192], patients with diabetes with increased risk of developing heart failure as identified by elevated NT-proBNP were randomised to standard care or addition of intensified CV risk modification with uptitration of beta blocker and RAS inhibitor. Patients in the intensified treatment group had reduced rates of hospitalisation or death due to cardiac disease. These results would both suggest there may be a role in treating “pre-HF” with conventional HFREF treatments which requires further investigation.

Lastly, I created a range of supervised machine learning models to predict cardiovascular death at two years as a binary outcome using data from the PARADIGM-HF trial. I was not able to identify any particularly well performing models, with all models having only modest C-statistics and overall performance was poorer than the PREDICT-HF prognostic model. There was a lot of available

data for these patients, but perhaps not enough granularity to realise the benefit of increased flexibility provided by machine learning methods. This is consistent with other studies that have compared machine learning methods with other statistical analytical techniques to predict outcome. Sometimes the machine learning methods provide slight performance improvement, but this is marginal and often more basic methods can have similar or better performance. Perhaps this type of data is not how machine learning will be utilised in heart failure research going forward. Machine learning has been effective in analysis of electrocardiograms and imaging, and perhaps combining this with clinical data may provide the additional granular data to allow for more accurate prediction of outcome, as discussed in Chapter 7. It has already been demonstrated that a combination of ECG traces with some basic demographic data could be used to build an effective machine learning algorithm for diagnosis of clinically significant structural heart disease.[193] Deep learning has also been utilised to analyse cardiac MRI imaging in tangent with a polygenic score from genome-wide association studies with respect to prediction of thoracic aortic aneurysm and aortic stenosis.[194] A similar principle could be applied for risk prediction models in heart failure using a variety of different sources of input data. A further potential source of gaining more detailed patient information could be by analysing unstructured data included in patient records, for example free text entries, using machine learning capabilities such as natural language processing.[195]

Another potential future avenue for machine learning techniques could be in using data from devices such as smart watches to gain more granular data about patient physiology. Wearable devices have been used to identify atrial fibrillation in undiagnosed population and to measure amount of physical activity and are likely to be capable of more in depth data collection.[196,197] Medical devices are capable of measuring various markers of congestion and the use of haemodynamic monitoring can be associated with improved clinical outcomes.[198] This additional depth of data generated using devices could be well suited to machine learning analytical techniques. This links with the general move within all branches of healthcare delivery towards telemedicine driven by the Covid-19 pandemic, perhaps the use of mobile phone apps and mobile diagnostic devices

may increase, requiring new approaches to handle, sort and understand the vast quantity of data generated.

A growing area of research is genomic profiling, which also results in generation of large volumes of data.[199-201] Although conventional statistical approaches can be used in these settings, it poses another option to optimise the use of more flexible machine learning methods to both identify patterns, or clusters, in data and to link data with outcomes. Finding genetic loci may also be able to find clusters of patients with different treatment response, as has been demonstrated by finding genetic loci related to discontinuation of ACEI due to cough and polygenic response predictor score associated with differential response to beta blocker therapy.[202,203]

As discussed in Chapter 7, there are many hurdles before these rich data sources could be used in complex machine learning predictive modelling. Firstly, ethically gaining access to the vast quantities of data required is challenging, either through direct patient consent or robust anonymisation techniques. It is extremely important to include a broad range of patients in terms of race, sex, age, and so on to ensure the resulting model does not inherit any biases because of unequally distributed input data. Next, the computational processing power required could quickly become prohibitive, potentially requiring the use of cloud services to facilitate which adds complexity around data security. Another challenge is the 'black box' processing of machine learning. Often it is not possible to understand how the model creates the prediction in the same way we can with conventional statistical approaches, and we therefore cannot easily explain the model. This can result in healthcare providers and the general public finding the model unacceptable for use.

8.1 Limitations

This work was all carried out using clinical trial data, therefore findings may not be applicable to patients with heart failure in the general population due to the selection of patients for trial participation. This is particularly important in latent class analysis, as other phenotypes may be prevalent in the population but excluded due to inclusion and exclusion criteria. In general, older patients, minorities and female patients are underrepresented with little improvement in

these trends over time.[204] Therefore these analyses likely only capture some prevalent phenotypes of heart failure, and use of routine data may identify other clusters with potential benefit for patient care in these patients.

Supervised machine learning analysis was carried out within a framework of models designed to make machine learning accessible to more researchers, therefore better performance may have been achieved if a more advanced programmer was able to manipulate the models more extensively. However, given the lack of consistent signal of machine learning outperforming statistical methods in this field it would be surprising if with further optimisation a significant improvement was made.

Although clinical trial data is rich and of good population size, some analyses of 'big data' involve many more data points and many more variables. Therefore although a lot of data was utilised, some data scientists would consider this to still be on a relatively small scale.

8.2 Future work

This work has found consistent results in phenotyping patients using data driven techniques. It is unlikely future randomised trials will include stratification of randomised treatment based on phenotype group due to the issue of underlying uncertainty in class membership. However, clustering could remain as a post-hoc technique to look for signal of differential treatment effect. An alternative would be to try and target a particular phenotype that has been previously identified by latent class analysis in a trial by shaping the inclusion criteria to reflect the phenotype group of interest.

Although in my analyses there was an early split between HFREF and HFPEF, perhaps a valuable area for further exploration would be phenotyping patients with heart failure regardless of ejection fraction. The results of my analyses did suggest similar groups were identified in both populations. There is a move recently to think of heart failure more as a continuous spectrum rather than by cut-offs for ejection fraction. This reflects the underlying issue with identifying groups of patients by one characteristic alone, as in reality phenotypes are much more complex than this. It could be argued that phenotyping should then be

repeated using all patients with heart failure together, regardless of ejection fraction, to understand phenotypes that exist across the spectrum of ejection fraction. Ejection fraction could provide one input into such a phenotyping analysis and be part of a more complex multilevel description of a patient with heart failure rather than being the primary difference defining heart failure groups.

The issue of separating patients into HFREF and HFPEF would also hold true in the analysis of HF hospitalisations. This was mainly driven by the expectation that these patients perhaps behave differently, however given results suggestive of more similarity than difference it would be interesting to repeat the analysis amalgamating patients with heart failure regardless of ejection fraction.

Supervised machine learning using the type of data collected in the process of clinical trials is less likely to make steps forward in terms of predicting adverse outcomes in heart failure. We do have many other types of data about patients available to us, such as ECGs and imaging, which paired with clinical data might be an avenue to try and improve event prediction.

8.3 Conclusion

This thesis presents several approaches to using clinical trial data to personalise the care of patients with heart failure. I have explored the patient journey in terms of heart failure hospitalisation in greater depth in both HFREF and HFPEF, defined phenotypes of patients in heart failure and examined the associated prognosis and developed machine learning predictive models to try and better understand the individual's risk of cardiovascular death. The first two approaches help understand the average risks of adverse outcomes at a group level, for example in recently hospitalised patients or in certain HF phenotypes. The elusive area remains improved risk prediction at an individual level which was not achieved using supervised machine learning in this data. Supervised machine learning remains an area of exciting potential for further research if other rich sources of data such as imaging or genomics are used as input variables.

References

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, Burri H, Butler J, Čelutkienė J, Chioncel O, Cleland JG, Coats AJ, Crespo-Leiro MG, Farmakis D, Gilard M, Heymans S, Hoes AW, Jaarsma T, Jankowska EA, Lainscak M, Lam CSP, Lyon AR, McMurray JJV, Mebazaa A, Mindham R, Muneretto C, Francesco Piepoli M, Price S, Rosano GMC, Ruschitzka F, Kathrine Skibelund A, ESC Scientific Document Group. ESC Scientific Document Group. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2022;42:3599–3726.

2. Bozkurt B, Coats AJS, Tsutsui H, Abdelhamid CM, Adamopoulos S, Albert N, Anker SD, Atherton J, Böhm M, Butler J, Drazner MH, Michael Felker G, Filippatos G, Fiuzat M, Fonarow GC, Gomez-Mesa JE, Heidenreich P, Imamura T, Jankowska EA, Januzzi J, Khazanie P, Kinugawa K, Lam CSP, Matsue Y, Metra M, Ohtani T, Francesco Piepoli M, Ponikowski P, Rosano GMC, Sakata Y, Seferović P, Starling RC, Teerlink JR, Vardeny O, Yamamoto K, Yancy C, Zhang J, Zieroth S. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. *Eur J Heart Fail*. 2021;23:352–380.

3. Conrad N, Judge A, Tran J, Mohseni H, Hedgecott D, Crespillo AP, Allison M, Hemingway H, Cleland JG, McMurray JJV, Rahimi K. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet*. 2018;391:572–580.

4. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. *Card Fail Rev*. 2017;3:7–11.

5. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries

and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392:1789–1858.

6. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, Djousse L, Elkind MSV, Ferguson JF, Fornage M, Khan SS, Kissela BM, Knutson KL, Kwan TW, Lackland DT, Lewis TT, Lichtman JH, Longenecker CT, Loop MS, Lutsey PL, Martin SS, Matsushita K, Moran AE, Mussolino ME, Perak AM, Rosamond WD, Roth GA, Sampson UKA, Satou GM, Schroeder EB, Shah SH, Shay CM, Spartano NL, Stokes A, Tirschwell DL, VanWagner LB, Tsao CW, Wong SS, Heard DG. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. *Circulation*. 2020;141:E139–E596.

7. Roger VL. Epidemiology of heart failure. *Circ Res*. 2013;113:646–659.

8. van Riet EES, Hoes AW, Limburg A, Landman MAJ, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. *Eur J Heart Fail*. 2014;16:772–777.

9. NICOR | About Heart Failure [Internet]. [cited 2022 Oct 2];Available from: https://www.nicor.org.uk/national-cardiac-audit-programme/about-heart-failure/

10. Jhund PS, Kondo T, Butt JH, Docherty KF, Claggett BL, Desai AS, Vaduganathan M, Gasparyan SB, Bengtsson O, Lindholm D, Petersson M, Langkilde AM, de Boer RA, DeMets D, Hernandez AF, Inzucchi SE, Kosiborod MN, Køber L, Lam CSP, Martinez FA, Sabatine MS, Shah SJ, Solomon SD, McMurray JJV. Dapagliflozin across the range of ejection fraction in patients with heart failure: a patient-level, pooled meta-analysis of DAPA-HF and DELIVER. *Nat Med*. 2022;28:1956–1964.

11. CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North

Scandinavian Enalapril Survival Study (CONSENSUS). *N Engl J Med*. 1987;316:1429–1435.

12. SOLVD Investigators, Yusuf S, Pitt B, Davis CE, Hood WB, Cohn HN. Effect of Enalapril on Survival in Patients with Reduced Left Ventricular Ejection Fractions and Congestive Heart Failure. *N Engl J Med*. 1991;325:293–302.

13. Granger CB, McMurray JJV, Yusuf S, Held P, Michelson EL, Olofsson B, Östergren J, Pfeffer MA, Swedberg K. Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function intolerant to angiotensin-converting-enzyme inhibitors: the CHARM-Alternative trial. *Lancet*. 2003;362:772–776.

14. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Solomon SD, Swedberg K, Zile MR. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371:993–1004.

15. MERT-HF Study Group. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure (MERIT-HF). *Lancet*. 1999;353:2001–2007.

16. CIBIS-II Investigators and Committee. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): A randomised trial. *Lancet*. 1999;353:9–13.

17. Packer M, Fowler MB, Roecker EB, Coats AJS, Katus HA, Krum H, Mohacsi P, Rouleau JL, Tendera M, Staiger C, Holcslaw TL, Amann-Zalan I, DeMets DL. Effect of carvedilol on the morbidity of patients with severe chronic heart failure: results of the carvedilol prospective randomized cumulative survival (COPERNICUS) study. *Circulation*. 2002;106:2194–2199.

18. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999;341:709–717.

19. Zannad F, McMurray JJV, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pocock SJ, Pitt B. Eplerenone in patients with systolic heart failure and mild symptoms. *N Engl J Med*. 2011;364:11–21.

20. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, Shaw W, Law G, Desai M, Matthews DR. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med*. 2017;377:644–657.

21. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med*. 2015;373:2117–2128.

22. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, Silverman MG, Zelniker TA, Kuder JF, Murphy SA, Bhatt DL, Leiter LA, McGuire DK, Wilding JPH, Ruff CT, Gause-Nilsson IAM, Fredriksson M, Johansson PA, Langkilde A-M, Sabatine MS. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2019;380:347–357.

23. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Ponikowski P, Sabatine MS, Anand IS, Bělohlávek J, Böhm M, Chiang C-E, Chopra VK, Boer RA de, Desai AS, Diez M, Drozdz J, Dukát A, Ge J, Howlett JG, Katova T, Kitakaze M, Ljungman CEA, Merkely B, Nicolau JC, O'Meara E, Petrie MC, Vinh PN, Schou M, Tereshchenko S, Verma S, Held C, DeMets DL, Docherty KF, Jhund PS, Bengtsson O, Sjöstrand M, Langkilde A-M, de Boer RA, Desai AS, Diez M, Drozdz J, Dukát A, Ge J, Howlett JG, Katova T, Kitakaze M, Ljungman CEA, Merkely B, Nicolau JC, O'Meara E, Petrie MC, Vinh PN, Schou M, Tereshchenko S, Verma S, Held C, DeMets DL, Docherty KF, Jhund PS, Bengtsson O, Sjöstrand M, Langkilde A-M, DAPA-HF Trial Committees and Investigators. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2019;381:1995–2008.

24. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, Januzzi J, Verma S, Tsutsui H, Brueckmann M, Jamal W, Kimura K, Schnee

J, Zeller C, Cotton D, Bocchi E, Böhm M, Choi D-J, Chopra V, Chuquiure E, Giannetti N, Janssens S, Zhang J, Gonzalez Juanatey JR, Kaul S, Brunner-La Rocca H-P, Merkely B, Nicholls SJ, Perrone S, Pina I, Ponikowski P, Sattar N, Senni M, Seronde M-F, Spinar J, Squire I, Taddei S, Wanner C, Zannad F, EMPEROR-REDUCED Trial Investigators. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med*. 2020;383:1413–1424.

25. Cleland JGF, Tendera M, Adamus J, Freemantle N, Polonski L, Taylor J, PEP-CHF Investigators. The perindopril in elderly people with chronic heart failure (PEP-CHF) study. *Eur Heart J*. 2006;27:2338–2345.

26. Yusuf S, Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJV, Michelson EL, Olofsson B, Östergren J. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. *Lancet*. 2003;362:777–781.

27. Massie BM, Carson PE, McMurray JJ, Komajda M, McKelvie R, Zile MR, Anderson S, Donovan M, Iverson E, Staiger C, Ptaszynska A, I-PRESERVE Investigators. Irbesartan in Patients with Heart Failure and Preserved Ejection Fraction. *N Engl J Med*. 2008;359:2456–2467.

28. Pitt B, Pfeffer MA, Assmann SF, Boineau R, Anand IS, Claggett B, Clausell N, Desai AS, Diaz R, Fleg JL, Gordeev I, Harty B, Heitner JF, Kenwood CT, Lewis EF, O'Meara E, Probstfield JL, Shaburishvili T, Shah SJ, Solomon SD, Sweitzer NK, Yang S, McKinlay SM, TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med*. 2014;370:1383–1392.

29. Solomon SD, McMurray JJV, Anand IS, Ge J, Lam CSP, Maggioni AP, Martinez F, Packer M, Pfeffer MA, Pieske B, Redfield MM, Rouleau JL, van Veldhuisen DJ, Zannad F, Zile MR, Desai AS, Claggett B, Jhund PS, Boytsov SA, Comin-Colet J, Cleland J, Düngen H-D, Goncalvesova E, Katova T, Kerr Saraiva JF, Lelonek M, Merkely B, Senni M, Shah SJ, Zhou J, Rizkala AR, Gong J, Shi VC, Lefkowitz MP, PARAGON-HF Investigators and Committees.

Angiotensin–Neprilysin Inhibition in Heart Failure with Preserved Ejection Fraction. *N Engl J Med*. 2019;381:1609–1620.

30. Solomon SD, Vaduganathan M, L. Claggett B, Packer M, Zile M, Swedberg K, Rouleau J, A. Pfeffer M, Desai A, H. Lund L, Kober L, Anand I, Sweitzer N, Linssen G, Merkely B, Luis Arango J, Vinereanu D, Chen CH, Senni M, Sibulo A, Boytsov S, Shi V, Rizkala A, Lefkowitz M, McMurray JJV. Sacubitril/Valsartan Across the Spectrum of Ejection Fraction in Heart Failure. *Circulation*. 2020;141:352–361.

31. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, Rocca H-PB, Choi D-J, Chopra V, Chuquiure-Valenzuela E, Giannetti N, Gomez-Mesa JE, Janssens S, Januzzi JL, Gonzalez-Juanatey JR, Merkely B, Nicholls SJ, Perrone S V., Piña IL, Ponikowski P, Senni M, Sim D, Spinar J, Squire I, Taddei S, Tsutsui H, Verma S, Vinereanu D, Zhang J, Carson P, Lam CSP, Marx N, Zeller C, Sattar N, Jamal W, Schnaidt S, Schnee JM, Brueckmann M, Pocock SJ, Zannad F, Packer M, EMPEROR-Preserved Trial Investigators. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. *N Engl J Med*. 2021;385:1451–1461.

32. Solomon SD, McMurray JJV, Claggett B, de Boer RA, DeMets D, Hernandez AF, Inzucchi SE, Kosiborod MN, Lam CSP, Martinez F, Shah SJ, Desai AS, Jhund PS, Belohlavek J, Chiang C-E, Borleffs CJW, Comin-Colet J, Dobreanu D, Drozdz J, Fang JC, Alcocer-Gamba MA, al Habeeb W, Han Y, Cabrera Honorio JW, Janssens SP, Katova T, Kitakaze M, Merkely B, O'Meara E, Saraiva JFK, Tereshchenko SN, Thierer J, Vaduganathan M, Vardeny O, Verma S, Pham VN, Wilderäng U, Zaozerska N, Bachus E, Lindholm D, Petersson M, Langkilde AM, DELIVER Trial Committees and Investigators. Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction. *N Engl J Med*. 2022;387:1089–1098.

33. Roger VL, Weston SA, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ. Trends in heart failure incidence and survival in a community-based population. *JAMA*. 2004;292:344–350.

34. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med*. 2006;355:251–259.

35. Tsao CW, Lyass A, Enserro D, Larson MG, Ho JE, Kizer JR, Gottdiener JS, Psaty BM, Vasan RS. Temporal Trends in the Incidence of and Mortality Associated With Heart Failure With Preserved and Reduced Ejection Fraction. *JACC Heart Fail*. 2018;6:678–685.

36. Cowie MR, Anker SD, Cleland JGF, Felker GM, Filippatos G, Jaarsma T, Jourdain P, Knight E, Massie B, Ponikowski P, López-Sendón J. Improving care for patients with acute heart failure: before, during and after hospitalization. *ESC Heart Fail*. 2014;1:110–145.

37. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S. Risk prediction models for hospital readmission: A systematic review. JAMA. 2011;306:1688–1698.

38. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, Bhatt DL, Fonarow GC, Laskey WK. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. *JAMA Cardiol*. 2017;2:204–209.

39. Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Dwivedi G. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. *ESC Heart Fail*. 2019;6:428–435.

40. Karwath A, Bunting KV, Gill SK, Tica O, Pendleton S, Aziz F, Barsky AD, Chernbumroong S, Duan J, Mobley AR, Cardoso VR, Slater L, Williams JA, Bruce EJ, Wang X, Flather MD, Coats AJS, Gkoutos GV, Kotecha D. Redefining ß-blocker response in heart failure patients with sinus rhythm and atrial fibrillation: a machine learning cluster analysis. *Lancet*. 2021;398:1427–1435.

41. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, Pouleur AC, Knappe D, Biering-Sorensen T, Kutyifa V, Moss A, Stein K, Solomon SD, Bijnens B, Biering-Sørensen T, Kutyifa V, Moss A, Stein K, Solomon SD, Bijnens B. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. 2019;21:74–85.

42. Riolet C, Menet A, Verdun S, Altes A, Appert L, Guyomar Y, Delelis F, Ennezat PV, Guerbaai RA, Graux P, Tribouilloy C, Marechaux S. Clinical and prognostic implications of phenomapping in patients with heart failure receiving cardiac resynchronization therapy. *Arch Cardiovasc Dis*. 2021;114:197–210.

43. Segar MW, Patel KV, Ayers C, Basit M, Wilson Tang WH, Willett D, Berry J, Grodin JL, Pandey A. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail*. 2020;22:148–158.

44. Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL, Kitzman DW, Lee KL, O'Connor CM, Felker GM. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol*. 2014;64:1765–1774.

45. Chirinos JA, Orlenko A, Zhao L, Basso MD, Cvijic ME, Li Z, Spires TE, Yarde M, Wang Z, Seiffert DA, Prenner S, Zamani P, Bhattacharya P, Kumar A, Margulies KB, Car BD, Gordon DA, Moore JH, Cappola TP. Multiple Plasma Biomarkers for Risk Stratification in Patients With Heart Failure and Preserved Ejection Fraction. *J Am Coll Cardiol*. 2020;75:1281–1295.

46. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan LM, Shah SJ, Linde CM, Donal E, Daubert JC, Mälarstig A, Ziemek D, Lund L. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart*. 2020;106.

47. Gu J, Pan JA, Lin H, Zhang JF, Wang CQ. Characteristics, prognosis and treatment response in distinct phenogroups of heart failure with preserved ejection fraction. *Int J Cardiol*. 2021;323:148–154.

48. Casebeer A, Horter L, Hayden J, Simmons J, Evers T. Phenotypic clustering of heart failure with preserved ejection fraction reveals different rates of hospitalization. *J Cardiovasc Med (Hagerstown)*. 2021;22:45–52.

49. Nouraei H, Rabkin SW. A new approach to the clinical subclassification of heart failure with preserved ejection fraction. *Int J Cardiol*. 2021;331:138–143.

50. Stienen S, Ferreira JP, Kobayashi M, Preud'homme G, Dobre D, Machu JL, Duarte K, Bresso E, Devignes MD, López N, Girerd N, Aakhus S, Ambrosio G, Brunner-La Rocca HP, Fontes-Carvalho R, Fraser AG, van Heerebeek L, Heymans S, de Keulenaer G, Marino P, McDonald K, Mebazaa A, Papp Z, Raddino R, Tschöpe C, Paulus WJ, Zannad F, Rossignol P. Enhanced clinical phenotyping by mechanistic bioprofiling in heart failure with preserved ejection fraction: insights from the MEDIA-DHF study (The Metabolic Road to Diastolic Heart Failure). *Biomarkers*. 2020;25:201–211.

51. Harada D, Asanoi H, Noto T, Takagawa J. Different Pathophysiology and Outcomes of Heart Failure With Preserved Ejection Fraction Stratified by K-Means Clustering. *Front Cardiovasc Med*. 2020;7:607760.

52. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131:269–279.

53. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, Dickstein K, Filippatos G, Zannad F, Marco M, Ng L, Samani N, van Veldhuisen DJ, Lang C, Lam CS, Voors AA. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. *Eur J Heart Fail*. 2021;23:983–991.

54. Sabbah MS, Fayyaz AU, de Denus S, Felker GM, Borlaug BA, Dasari S, Carter RE, Redfield MM. Obese-Inflammatory Phenotypes in Heart Failure With Preserved Ejection Fraction. *Circ Heart Fail*. 2020;13:217–228.

55. Gevaert AB, Tibebu S, Mamas MA, Ravindra NG, Lee SF, Ahmad T, Ko DT, Januzzi JL, van Spall HGC. Clinical phenogroups are more effective than left ventricular ejection fraction categories in stratifying heart failure outcomes. *ESC Heart Fail*. 2021;8:2741–2754.

56. Austin PC, Tu J v., Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J Clin Epidemiol*. 2013;66:398–407.

57. Lanza ST, Rhoades BL. Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment. *Prevention Science*. 2013;14:157–168.

58. Ferreira JP, Duarte K, McMurray JJV, Pitt B, van Veldhuisen DJ, Vincent J, Ahmad T, Tromp J, Rossignol P, Zannad F. Data-driven approach to identify subgroups of heart failure with reduced ejection fraction patients with different prognoses and aldosterone antagonist response patterns. *Circ Heart Fail*. 2018;11:e004926.

59. Kao DP, Wagner BD, Robertson AD, Bristow MR, Lowes BD. A Personalized BEST: Characterization of Latent Clinical Classes of Nonischemic Heart Failure That Predict Outcomes and Response to Bucindolol. *PLoS One*. 2012;7:e48184.

60. Uijl A, Savarese G, Vaartjes I, Dahlström U, Brugts JJ, Linssen GCM, Empel V, Brunner-La Rocca H, Asselbergs FW, Lund LH, Hoes AW, Koudstaal S. Identification of distinct phenotypic clusters in heart failure with preserved ejection fraction. *Eur J Heart Fail*. 2021;ejhf.2169.

61. Cohen JB, Schrauben SJ, Zhao L, Basso MD, Cvijic ME, Li Z, Yarde M, Wang Z, Bhattacharya PT, Chirinos DA, Prenner S, Zamani P, Seiffert

DA, Car BD, Gordon DA, Margulies K, Cappola T, Chirinos JA. Clinical Phenogroups in Heart Failure With Preserved Ejection Fraction: Detailed Phenotypes, Prognosis, and Response to Spironolactone. *JACC Heart Fail*. 2020;8:172–184.

62. Kao DP, Lewsey JD, Anand IS, Massie BM, Zile MR, Carson PE, McKelvie RS, Komajda M, McMurray JJV, Lindenfeld J. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. *Eur J Heart Fail*. 2015;17:925–35.

63. Sotomi Y, Hikoso S, Komukai S, Sato T, Oeun B, Kitamura T, Nakagawa A, Nakatani D, Mizuno H, Okada K, Dohi T, Sunaga A, Kida H, Seo M, Yano M, Hayashi T, Nakagawa Y, Tamaki S, Ohtani T, Yasumura Y, Yamada T, Sakata Y. Phenotyping of acute decompensated heart failure with preserved ejection fraction. *Heart*. 2022;0:heartjnl-2021-320270.

64. Sotomi Y, Sato T, Hikoso S, Komukai S, Oeun B, Kitamura T, Nakatani D, Mizuno H, Okada K, Dohi T, Sunaga A, Kida H, Seo M, Yano M, Hayashi T, Nakagawa A, Nakagawa Y, Tamaki S, Ohtani T, Yasumura Y, Yamada T, Sakata Y. Minimal subphenotyping model for acute heart failure with preserved ejection fraction. *ESC Heart Fail*. 2022;9:2738–2746.

65. Tromp J, Tay WT, Ouwerkerk W, Teng T-HK, Yap J, MacDonald MR, Leineweber K, McMurray JJ v., Zile MR, Anand IS, Lam CSP, Richards AM, Lam CSP, Authors A-H. Multimorbidity in patients with heart failure from 11 Asian regions: A prospective cohort study using the ASIAN-HF registry. *PLoS Med*. 2018;15:e1002541.

66. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021;18:465–478.

67. Ghorbani A, Ouyang D, Abid A, He B, Chen JH, Harrington RA, Liang DH, Ashley EA, Zou JY. Deep learning interpretation of echocardiograms. *NPJ Digit Med*. 2020;3:10.

68. Simpson J, Jhund PS, Lund LH, Padmanabhan S, Claggett BL, Shen L, Petrie MC, Abraham WT, Desai AS, Dickstein K, Køber L, Packer M, Rouleau JL, Mueller-Velten G, Solomon SD, Swedberg K, Zile MR, McMurray JJV. Prognostic Models Derived in PARADIGM-HF and Validated in ATMOSPHERE and the Swedish Heart Failure Registry to Predict Mortality and Morbidity in Chronic Heart Failure. *JAMA Cardiol*. 2020;5:432–441.

69. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, Doughty RN. Predicting survival in heart failure: A risk score based on 39 372 patients from 30 studies. *Eur Heart J*. 2013;34:1404–1413.

70. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006;113:1424–1433.

71. Ferrero P, Iacovoni A, D'Elia E, Vaduganathan M, Gavazzi A, Senni M. Prognostic scores in heart failure - Critical appraisal and practical use. *Int J Cardiol*. 2015;188:1–9.

72. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, Woodward M, Patel A, McMurray J, MacMahon S. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail*. 2014;2:440–446.

73. Solomon SD, Anavekar N, Skali H, McMurray JJV, Swedberg K, Yusuf S, Granger CB, Michelson EL, Wang D, Pocock S, Pfeffer MA. Influence of Ejection Fraction on Cardiovascular Outcomes in a Broad Spectrum of Heart Failure Patients. *Circulation*. 2005;112:3738–3744.

74. Arnold J, Liu P, Howlett J, Ignaszewski A, Leblanc M, Kaan A, Pearce C, Sinclair L, Pearce S, Prentice C, Shochat M, Shotan A, Kazatsker M, Asif A, Shochat I, Dahan I, Frimerman A, Levy Y, Blondheim D, Meisel Hillel S, Rywik T, Koziarek J, Piotrowski W, Leszek P, Korewicki J, Zielinski T, Rozentryt P, Rywik H, Broda G. Ten year survival by NYHA functional

class in heart failure outpatients referred to specialized multidisciplinary heart failure clinics 1999 to 2011. *Eur Heart J*. 2013;34:P1505–P1505.

75. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, Fonarow GC, Masoudi FA. A validated risk score for in-hospital mortality in patients with heart failure from the American heart association get with the guidelines program. *Circ Cardiovasc Qual Outcomes*. 2010;3:25–32.

76. Lupón J, de Antonio M, Vila J, Peñafiel J, Galán A, Zamora E, Urrutia A, Bayes-Genis A. Development of a Novel Heart Failure Risk Tool: The Barcelona Bio-Heart Failure Risk Calculator (BCN Bio-HF Calculator). *PLoS One*. 2014;9:e85466.

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

78. Liang LW, Shimada YJ. Endotyping in Heart Failure - Identifying Mechanistically Meaningful Subtypes of Disease. *Circ J*. 2021;85:1407–1415.

79. Tokodi M, Schwertner WR, Kovács A, Tosér Z, Staub L, Sárkány A, Lakatos BK, Behon A, Boros AM, Perge P, Kutyifa V, Széplaki G, Gellér L, Merkely B, Kosztin A. Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. *Eur Heart J*. 2020;41:1747–1756.

80. Feeny AK, Rickard J, Patel D, Toro S, Trulock KM, Park CJ, Labarbera MA, Varma N, Niebauer MJ, Sinha S, Gorodeski EZ, Grimm RA, Ji X, Barnard J, Madabhushi A, Spragg DD, Chung MK. Machine Learning Prediction of Response to Cardiac Resynchronization Therapy: Improvement Versus Current Guidelines. *Circ Arrhythm Electrophysiol*. 2019;12:e007316.

81. Wang Z, Chen X, Tan X, Yang L, Kannapur K, Vincent JL, Kessler GN, Ru B, Yang M. Using Deep Learning to Identify High-Risk Patients with

Heart Failure with Reduced Ejection Fraction. *J Health Econ Outcomes Res*. 2021;8:6–13.

82. Tokodi M, Behon A, Merkel ED, Kovács A, Tősér Z, Sárkány A, Csákvári M, Lakatos BK, Schwertner WR, Kosztin A, Merkely B. Sex-Specific Patterns of Mortality Predictors Among Patients Undergoing Cardiac Resynchronization Therapy: A Machine Learning Approach. *Front Cardiovasc Med*. 2021;8:611055.

83. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and Machine Learning for Heart Failure Survival Analysis. *Stud Health Technol Inform*. 2015;216:40–44.

84. Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, Demets DL, Field ME, Eckhardt LL, Page CD. Machine Learning Algorithm Predicts Cardiac Resynchronization Therapy Outcomes: Lessons From the COMPANION Trial. *Circ Arrhythm Electrophysiol*. 2018;11:e005499.

85. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. *JACC Heart Fail*. 2020;8:12–21.

86. Howell SJ, Stivland T, Stein K, Ellenbogen KA, Tereshchenko LG. Using Machine-Learning for Prediction of the Response to Cardiac Resynchronization Therapy: The SMART-AV Study. *JACC Clin Electrophysiol*. 2021;7:1505–1515.

87. Zhou L, Guo Z, Wang B, Wu Y, Li Z, Yao H, Fang R, Yang H, Cao H, Cui Y. Risk Prediction in Patients With Heart Failure With Preserved Ejection Fraction Using Gene Expression Data and Machine Learning. *Front Genet*. 2021;12:652315.

88. Li F, Xin H, Zhang J, Fu M, Zhou J, Lian Z. Prediction model of in-hospital mortality in intensive care unit patients with heart failure:

machine learning-based, retrospective analysis of the MIMIC-III database. *BMJ Open*. 2021;11:e044779.

89. Jing L, Ulloa Cerna AE, Good CW, Sauers NM, Schneider G, Hartzel DN, Leader JB, Kirchner HL, Hu Y, Riviello DM, Stough J v., Gazes S, Haggerty A, Raghunath S, Carry BJ, Haggerty CM, Fornwalt BK. A Machine Learning Approach to Management of Heart Failure Populations. *JACC Heart Fail*. 2020;8:578–587.

90. Kim W, Park JJ, Lee HY, Kim KH, Yoo BS, Kang SM, Baek SH, Jeon ES, Kim JJ, Cho MC, Chae SC, Oh BH, Kook W, Choi DJ. Predicting survival in heart failure: a risk score based on machine-learning and change point algorithm. *Clinical Research in Cardiology*. 2021;110:1321–1333.

91. Sarijaloo FB, Park J, Zhong X, Wokhlu A. Predicting 90 day acute heart failure readmission and death using machine learning-supported decision analysis. *Clin Cardiol*. 2021;44:230–237.

92. Negassa A, Ahmed S, Zolty R, Patel SR. Prediction Model Using Machine Learning for Mortality in Patients with Heart Failure. *Am J Cardiol*. 2021;153:86–93.

93. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, Han Q, Zhang Y. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med*. 2021;137:104813.

94. van der Galiën OP, Hoekstra RC, Gürgöze MT, Manintveld OC, van den Bunt MR, Veenman CJ, Boersma E. Prediction of long-term hospitalisation and all-cause mortality in patients with chronic heart failure on Dutch claims data: a machine learning approach. *BMC Med Inform Decis Mak*. 2021;21:1–13.

95. Wang Q, Li B, Chen K, Yu F, Su H, Hu K, Liu Z, Wu G, Yan J, Su G. Machine learning-based risk prediction of malignant arrhythmia in hospitalized patients with heart failure. *ESC Heart Fail*. 2021;8:5363–5371.

96. Tohyama T, Ide T, Ikeda M, Kaku H, Enzan N, Matsushima S, Funakoshi K, Kishimoto J, Todaka K, Tsutsui H. Machine learning-based model for predicting 1 year mortality of hospitalized patients with heart failure. *ESC Heart Fail*. 2021;8:4077–4085.

97. König S, Pellissier V, Hohenstein S, Bernal A, Ueberham L, Meier-Hellmann A, Kuhlen R, Hindricks G, Bollmann A. Machine learning algorithms for claims data-based prediction of in-hospital mortality in patients with heart failure. *ESC Heart Fail*. 2021;8:3026–3036.

98. Lorenzoni G, Sabato SS, Lanera C, Bottigliengo D, Minto C, Ocagli H, de Paolis P, Gregori D, Iliceto S, Pisanò F. Comparison of Machine Learning Techniques for Prediction of Hospitalization in Heart Failure Patients. *J Clin Med*. 2019;8.

99. Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Chow BJ, Dwivedi G. Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death. *PLoS One*. 2019;14:e0218760.

100. Ju C, Zhou J, Lee S, Tan MS, Liu T, Bazoukis G, Jeevaratnam K, Chan EWY, Wong ICK, Wei L, Zhang Q, Tse G. Derivation of an electronic frailty index for predicting short-term mortality in heart failure: a machine learning approach. *ESC Heart Fail*. 2021;8:2837–2845.

101. Lewis M, Elad G, Beladev M, Maor G, Radinsky K, Hermann D, Litani Y, Geller T, Pines JM, Shapiro N l., Figueroa JF. Comparison of deep learning with traditional models to predict preventable acute care use and spending among heart failure patients. *Sci Rep*. 2021;11:1164.

102. Najafi-Vosough R, Faradmal J, Hosseini SK, Moghimbeigi A, Mahjub H. Predicting Hospital Readmission in Heart Failure Patients in Iran: A Comparison of Various Machine Learning Methods. *Healthc Inform Res*. 2021;27:307–314.

103. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlström U, O'Connor CM, Felker GM, Desai NR, Dahlstrom U, O'Connor CM, Felker GM, Desai NR. Machine Learning Methods Improve Prognostication, Identify Clinically Distinct Phenotypes, and Detect Heterogeneity in Response to Therapy in a Large Cohort of Heart Failure Patients. *J Am Heart Assoc*. 2018;7.

104. Wang Z, Zhu Y, Li D, Yin Y, Zhang J. Feature rearrangement based deep learning system for predicting heart failure mortality. *Comput Methods Programs Biomed*. 2020;191:105383.

105. Mahajan SM, Ghani R. Using Ensemble Machine Learning Methods for Predicting Risk of Readmission for Heart Failure. *Stud Health Technol Inform*. 2019;264:243–247.

106. Greenberg B, Adler E, Campagnari C, Yagil A. A machine learning risk score predicts mortality across the spectrum of left ventricular ejection fraction. *Eur J Heart Fail*. 2021;23:995–999.

107. Stampehl M, Friedman HS, Navaratnam P, Russo P, Park S, Obi EN. Risk assessment of post-discharge mortality among recently hospitalized Medicare heart failure patients with reduced or preserved ejection fraction. *Curr Med Res Opin*. 2020;36:179–188.

108. Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. *Sci Rep*. 2019;9.

109. Park J, Zhong X, Babaie Sarijaloo F, Wokhlu A. Tailored risk assessment of 90-day acute heart failure readmission or all-cause death to heart failure with preserved versus reduced ejection fraction. *Clin Cardiol*. 2022;

110. Luo C, Zhu Y, Zhu Z, Li R, Chen G, Wang Z. A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure. *J Transl Med*. 2022;20.

111. Sharma V, Kulkarni V, McAlister F, Eurich D, Keshwani S, Simpson SH, Voaklander D, Samanani S. Predicting 30-Day Readmissions in Patients With Heart Failure Using Administrative Data: A Machine Learning Approach. *J Card Fail*. 2021;28:710–722.

112. Beecy AN, Gummalla M, Sholle E, Xu Z, Zhang Y, Michalak K, Dolan K, Hussain Y, Lee BC, Zhang Y, Goyal P, Campion TR, Shaw LJ, Baskaran L, Al'Aref SJ. Utilizing electronic health data and machine learning for the prediction of 30-day unplanned readmission or all-cause mortality in heart failure. *Cardiovasc Digit Health J*. 2020;1:71–79.

113. Wang K, Tian J, Zheng C, Yang H, Ren J, Li C, Han Q, Zhang Y. Improving Risk Identification of Adverse Outcomes in Chronic Heart Failure Using SMOTE+ENN and Machine Learning. *Risk Manag Healthc Policy*. 2021;14:2453–2463.

114. Park J, Hwang I-C, Yoon YE, Park J-B, Park J-H, Cho G-Y. Predicting Long-Term Mortality in Patients with Acute Heart Failure Using Machine Learning. *J Card Fail*. 2022;28:1078–1087.

115. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, Zhu W, Sama I, Tadel M, Campagnari C, Greenberg B, Yagil A. Improving risk prediction in heart failure using machine learning. *Eur J Heart Fail*. 2020;22:139–147.

116. Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Dwivedi G. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. *ESC Heart Fail*. 2019;6:428–435.

117. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li S-XX, Negahban SN, Krumholz HM. Analysis of Machine Learning Techniques for Heart Failure Readmissions. *Circ Cardiovasc Qual Outcomes*. 2016;9:629–640.

118. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes. 2020;3:e1918962.

119. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, Hisamitsu T, Kojima G, Felsted J, Kakarmath S, Kvedar J, Jethwani K. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak*. 2018;18:44.

120. Kwon JM, Kim K-HKHKH, Jeon K-HH, Lee SE, Lee H-YY, Cho H-JJ, Choi JO, Jeon E-SS, Kim M-SS, Kim J-JJ, Hwang K-KK, Chae SC, Baek SH, Kang S-MM, Choi D-JJ, Yoo B-SS, Kim K-HKHKH, Park H-YY, Cho M-CC, Oh B-HH. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PLoS One*. 2019;14:e0219302.

121. Krum H, Massie B, Abraham WT, Dickstein K, Kober L, McMurray JJV, Desai A, Gimpelewicz C, Kandra A, Reimund B, Rattunde H, Armbrecht J. Direct renin inhibition in addition to or as an alternative to angiotensin converting enzyme inhibition in patients with chronic systolic heart failure: rationale and design of the Aliskiren Trial to Minimize OutcomeS in Patients with HEart failuRE (ATMOSPHERE) study. *Eur J Heart Fail*. 2011;13:107–114.

122. Parving H-H, Brenner BM, McMurray JJV, de Zeeuw D, Haffner SM, Solomon SD, Chaturvedi N, Persson F, Desai AS, Nicolaides M, Richard A, Xiang Z, Brunel P, Pfeffer MA. Cardiorenal End Points in a Trial of Aliskiren for Type 2 Diabetes. *N Engl J Med*. 2012;367:2204–2213.

123. Gheorghiade M, Böhm M, Greene SJ, Fonarow GC, Lewis EF, Zannad F, Solomon SD, Baschiera F, Botha J, Hua TA, Gimpelewicz CR, Jaumont X, Lesogor A, Maggioni AP. Effect of aliskiren on postdischarge mortality and heart failure readmissions among patients hospitalized for heart failure: the ASTRONAUT randomized trial. *JAMA*. 2013;309:1125–1135.

124. McMurray JJV v, Krum H, Abraham WT, Dickstein K, K\ober L v, Desai AS, Solomon SD, Greenlaw N, Ali MA, Chiang Y, Shao Q, Tarnesby G, Massie BM, Investigators AC, Køber L v., Desai AS, Solomon SD, Greenlaw N, Ali MA, Chiang Y, Shao Q, Tarnesby G, Massie BM. Aliskiren, Enalapril, or Aliskiren and Enalapril in Heart Failure. *N Engl J Med*. 2016;374:1521–1532.

125. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau J, Shi VC, Solomon SD, Swedberg K, Zile MR. Dual angiotensin receptor and neprilysin inhibition as an alternative to angiotensin-converting enzyme inhibition in patients with chronic systolic heart failure: Rationale for and design of the Prospective comparison of ARNI with ACEI to Determine Impact. *Eur J Heart Fail*. 2013;15:1062–1073.

126. Solomon SD, Rizkala AR, Gong J, Wang W, Anand IS, Ge J, Lam CSP, Maggioni AP, Martinez F, Packer M, Pfeffer MA, Pieske B, Redfield MM, Rouleau JL, van Veldhuisen DJ, Zannad F, Zile MR, Desai AS, Shi VC, Lefkowitz MP, McMurray JJV. Angiotensin Receptor Neprilysin Inhibition in Heart Failure With Preserved Ejection Fraction: Rationale and Design of the PARAGON-HF Trial. *JACC Heart Fail*. 2017;5:471–482.

127. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc Series B Stat Methodol*. 2000;62:711–730.

128. Mogensen UM, Gong J, Jhund PS, Shen L, Kober L, Desai AS, Lefkowitz MP, Packer M, Rouleau JL, Solomon SD, Claggett BL, Swedberg K, Zile MR, Mueller-Velten G, McMurray JJV. Effect of sacubitril/valsartan on recurrent events in the Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF). *Eur J Heart Fail*. 2018;20:760–768.

129. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Wiley-Blackwell; 20032003.

130. Lanza ST, Dziak JJ, Huang L, Wagner AT, Collins LM. LCA Stata Plugin Users' Guide Version 1.2.1.

131. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification 1996 13:2*. 1996;13:195–212.

132. Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*. 2020:287–311.

133. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2019;Available from: https://www.r-project.org/

134. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles [Internet]. 2020;Available from: https://www.tidymodels.org

135. Chun S, Tu J v., Wijeysundera HC, Austin PC, Wang X, Levy D, Lee DS. Lifetime analysis of hospitalizations and survival of patients newly admitted with heart failure. *Circ Heart Fail*. 2012;5:414–421.

136. Huusko J, Tuominen S, Studer R, Corda S, Proudfoot C, Lassenius M, Ukkonen H. Recurrent hospitalizations are associated with increased mortality across the ejection fraction range in heart failure. *ESC Heart Fail*. 2020;7:2406–2417.

137. Lee DS, Stitt A, Austin PC, Stukel TA, Schull MJ, Chong A, Newton GE, Lee JS, Tu J v. Prediction of heart failure mortality in emergent care: a cohort study. *Ann Intern Med*. 2012;156:767–775.

138. Lee DS, Gona P, Albano I, Larson MG, Benjamin EJ, Levy D, Kannel WB, Vasan RS. A Systematic Assessment of Causes of Death after Heart Failure Onset in the Community: Impact of Age at Death, Time Period, and Left Ventricular Systolic Dysfunction. *Circ Heart Fail*. 2011;4:36.

139. Packer M. What causes sudden death in patients with chronic heart failure and a reduced ejection fraction? *Eur Heart J*. 2020;41:1757–1763.

140. Rohde LE, Chatterjee NA, Vaduganathan M, Claggett B, Packer M, Desai AS, Zile M, Rouleau J, Swedberg K, Lefkowitz M, Shi V, McMurray JJV, Solomon SD. Sacubitril/Valsartan and Sudden Cardiac Death According to Implantable Cardioverter-Defibrillator Use and Heart Failure Cause: A PARADIGM-HF Analysis. *JACC Heart Fail*. 2020;8:844–855.

141. Desai AS, Jhund PS, Claggett BL, Vaduganathan M, Miao ZM, Kondo T, Barkoudah E, Brahimi A, Connolly E, Finn P, Lang NN, Mc Causland FR, McGrath M, Petrie MC, McMurray JJV, Solomon SD. Effect of Dapagliflozin on Cause-Specific Mortality in Patients With Heart Failure Across the Spectrum of Ejection Fraction: A Participant-Level Pooled Analysis of DAPA-HF and DELIVER. *JAMA Cardiol*. 2022;7:1227–1234.

142. Al-Gobari M, Khatib C el, Pillon F, Gueyffier F. Beta-blockers for the prevention of sudden cardiac death in heart failure patients: a meta-analysis of randomized controlled trials. *BMC Cardiovasc Disord*. 2013;13:52.

143. Shen L, Jhund PS, Petrie MC, Claggett BL, Barlera S, Cleland JGF, Dargie HJ, Granger CB, Kjekshus J, Køber L, Latini R, Maggioni AP, Packer M, Pitt B, Solomon SD, Swedberg K, Tavazzi L, Wikstrand J, Zannad F, Zile MR, McMurray JJV. Declining Risk of Sudden Death in Heart Failure. *N Engl J Med*. 2017;377:41–51.

144. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City cardiomyopathy questionnaire: A new health status measure for heart failure. *J Am Coll Cardiol*. 2000;35:1245–1255.

145. Huang W, Teng TK, Tay WT, Richards AM, Kadam U, Lawson CA, Shimizu W, Loh SY, Anand I, Lam CSP. Patient-reported outcomes in heart failure with preserved vs. reduced ejection fraction: focus on physical independence. *ESC Heart Fail*. 2020;7:2051–2062.

146. Joseph SM, Novak E, Arnold SV, Jones PG, Khattak H, Platts AE, Davila-Roman VG, Mann DL, Spertus JA. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure

with preserved and reduced ejection fraction. *Circ Heart Fail*. 2013;6:1139–1146.

147. Spertus JA, Jones PG, Sandhu AT, Arnold SV. Interpreting the Kansas City Cardiomyopathy Questionnaire in Clinical Trials and Clinical Care: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2020;76:2379–2390.

148. Labrosciano C, Horton D, Air T, Tavella R, Beltrame JF, Zeitz CJ, Krumholz HM, Adams RJT, Scott IA, Gallagher M, Hossain S, Hariharaputhiran S, Ranasinghe I. Frequency, trends and institutional variation in 30-day all-cause mortality and unplanned readmissions following hospitalisation for heart failure in Australia and New Zealand. *Eur J Heart Fail*. 2021;23:31–40.

149. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, Hollenberg SM, Lindenfeld JA, Masoudi FA, McBride PE, Peterson PN, Stevenson LW, Westlake C. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation. 2017;136:e137–e161.

150. McMurray JJV, Packer M. How Should We Sequence the Treatments for Heart Failure and a Reduced Ejection Fraction? A Redefinition of Evidence-Based Medicine. *Circulation*. 2020;143:875–877.

151. Gattis WA, O'Connor CM. Predischarge initiation of carvedilol in patients hospitalized for decompensated heart failure. *American Journal of Cardiology*. 2004;93:74–76.

152. Velazquez EJ, Morrow DA, DeVore AD, Duffy CI, Ambrosy AP, McCague K, Rocha R, Braunwald E. Angiotensin–Neprilysin Inhibition in Acute Decompensated Heart Failure. *N Engl J Med*. 2019;380:539–548.

153. Armstrong PW, Pieske B, Anstrom KJ, Ezekowitz J, Hernandez AF, Butler J, Lam CSP, Ponikowski P, Voors AA, Jia G, McNulty SE, Patel MJ, Roessig L, Koglin J, O'Connor CM. Vericiguat in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2020;160:1883–93.

154. Swedberg K, Komajda M, Böhm M, Borer JS, Ford I, Dubost-Brama A, Lerebours G, Tavazzi L. Ivabradine and outcomes in chronic heart failure (SHIFT): A randomised placebo-controlled study. *Lancet*. 2010;376:875–885.

155. Taylor AL, Ziesche S, Yancy C, Carson P, D'Agostino R, Ferdinand K, Taylor M, Adams K, Sabolinski M, Worcel M, Cohn JN. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N Engl J Med*. 2004;351:2049–2057.

156. Cohn JN, Archibald DG, Ziesche S, Franciosa JA, Harston WE, Tristani FE, Dunkman WB, Jacobs W, Francis GS, Flohr KH, Goldman S, Cobb FR, Shan PM, Saunders R, Fletcher RD, Loeb HS, Hughes VC, Baker B. Effect of vasodilator therapy on mortality in chronic congestive heart failure. Results of a Veterans Administration Cooperative Study. *N Engl J Med*. 1986;314:1547–1552.

157. Goldgrab D, Balakumaran K, Kim MJ, Tabtabai SR. Updates in heart failure 30-day readmission prevention. Heart Fail Rev. 2019;24:177–187.

158. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, de Marco T, Escarce JJ, Evangelista LS, Hanna B, Ganiats TG, Greenberg BH, Greenfield S, Kaplan SH, Kimchi A, Liu H, Lombardo D, Mangione CM, Sadeghi B, Sadeghi B, Sarrafzadeh M, Tong K, Fonarow GC, Davidson B, Ghasemzadeh H, Gropper M, Mourad M, Ahmadpour A, Davila W, Engel S, Jacolbia R, Lee H, Linares L, Michel E, Weyrich MS, Zellmer E, Esbati-Mashayekhi L, Haddad E, Haskins M, Larson T, Pratt K, Ansorie H, Aoki K, Baron R, Brinker E, Carroll M, Contasti A, Fekete A, Guzman V, Larsen L, Martinez L, Myers S, Schimmel M, Schnell-Heringer A, Taylor A, Tooley T, van den Brande G, Zaharias E, Billimek J, Castaneda R, Reyes MED, Fine D, Lo T, Luu X, Ochoa S, Perez M, Rincon D, Sillas F, Uy V, Wang

E, Xu H, Yala S, Yan T. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure the better effectiveness after transition-heart failure (BEAT-HF) randomized clinical trial. *JAMA Intern Med*. 2016;176:310–318.

159. Bradley EH, Curry L, Horwitz LI, Sipsma H, Wang Y, Walsh MN, Goldmann D, White N, Piña IL, Krumholz HM. Hospital strategies associated with 30-day readmission rates for patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2013;6:444–450.

160. Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive Discharge Planning with Postdischarge Support for Older Patients with Congestive Heart Failure: A Meta-analysis. J Am Med Assoc. 2004;291:1358–1367.

161. McAlister FA, Stewart S, Ferrua S, McMurray JJV. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: A systematic review of randomized trials. *J Am Coll Cardiol*. 2004;44:810–819.

162. Redfield MM. Heart Failure with Preserved Ejection Fraction. *N Engl J Med*. 2016;375:1868–1877.

163. Coats AJS. Ageing, demographics, and heart failure. *Eur Heart J Suppl*. 2019;21:L4.

164. Goyal P, Almarzooq ZI, Horn EM, Karas MG, Sobol I, Swaminathan R v., Feldman DN, Minutello RM, Singh HS, Bergman GW, Wong SC, Kim LK. Characteristics of Hospitalizations for Heart Failure with Preserved Ejection Fraction. *Am J Med*. 2016;129:635.e15-635.e26.

165. Chan MMY, Lam CSP. How do patients with heart failure with preserved ejection fraction die? *Eur J Heart Fail*. 2013;15:604–613.

166. Desai AS, Vaduganathan M, Cleland JG, Claggett BL, Barkoudah E, Finn P, McCausland FR, Yilmaz MB, Lefkowitz M, Shi V, Pfeffer MA,

McMurray JJV, Solomon SD. Mode of Death in Patients with Heart Failure and Preserved Ejection Fraction: Insights from PARAGON-HF Trial. *Circ Heart Fail*. 2021;14:1283–1290.

167. Jhund PS, Claggett BL, Talebi A, Butt JH, Gasparyan SB, Wei L-J, McCaw ZR, Wilderäng U, Bengtsson O, Desai AS, Petersson M, Maria Langkilde A, de Boer RA, Hernandez AF, Inzucchi SE, Kosiborod MN, P Lam CS, Martinez FA, Shah SJ, Vaduganathan M, Solomon SD, V McMurray JJ, Author C, Heart B. Effect of Dapagliflozin on Total Heart Failure Events in Patients With Heart Failure With Mildly Reduced or Preserved Ejection Fraction: A Prespecified Analysis of the DELIVER Trial. *JAMA Cardiol*. 2023;8:554–563.

168. Cunningham JW, Vaduganathan M, Claggett BL, Kulac IJ, Desai AS, Jhund PS, de Boer RA, DeMets D, Hernandez AF, Inzucchi SE, Kosiborod MN, Lam CSP, Martinez F, Shah SJ, McGrath MM, O'Meara E, Wilderäng U, Lindholm D, Petersson M, Langkilde AM, McMurray JJV, Solomon SD. Dapagliflozin in Patients Recently Hospitalized With Heart Failure and Mildly Reduced or Preserved Ejection Fraction. *J Am Coll Cardiol*. 2022;80:1302–1310.

169. Kratz A, Ferraro M, Sluss PM, Lewandrowski KB. Normal Reference Laboratory Values. *New England Journal of Medicine*. 2004;351:1548–1563.

170. Bakk Z, Vermunt JK. Robustness of Stepwise Latent Class Modeling With Continuous Distal Outcomes. *Struct Equ Modeling*. 2016;23:20–31.

171. Dziak JJ, Bray BC, Zhang J, Zhang M, Lanza ST. Comparing the performance of improved classify-analyze approaches for distal outcomes in latent profile analysis. *Methodology*. 2016;12:107–116.

172. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau J, Shi VC, Solomon SD, Swedberg K, Zile MR. Dual angiotensin receptor and neprilysin inhibition as an alternative to angiotensin-converting enzyme inhibition in patients with chronic systolic heart failure: rationale for and design of the Prospective comparison of ARNI with ACEI

to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF). *Eur J Heart Fail*. 2013;15:1062–1073.

173. Kuhn M, Silge J. Tidy Modeling with R: A Framework for Modeling in the Tidyverse. O'Reilly Media, Inc.; 2022.

174. Spooner A, Chen E, Sowmya A, Sachdev P, Kochan NA, Trollor J, Brodaty H. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific Reports 2020 10:1*. 2020;10:1–10.

175. Survival Analysis in tidymodels [Internet]. [cited 2023 Jan 6];Available from: https://www.tidyverse.org/blog/2021/11/survival-analysis-parsnip-adjacent/

176. Ruan X, Li Y, Jin X, Deng P, Xu J, Li N, Li X, Liu Y, Hu Y, Xie J, Wu Y, Long D, He W, Yuan D, Guo Y, Li H, Huang H, Yang S, Han M, Zhuang B, Qian J, Cao Z, Zhang X, Xiao J, Xu L. Health-adjusted life expectancy (HALE) in Chongqing, China, 2017: An artificial intelligence and big data method estimating the burden of disease at city level. *Lancet Reg Health West Pac*. 2021;9:100110.

177. Butt JH, Dewan P, Jhund PS, Anand IS, Atar D, Ge J, Desai AS, Echeverria LE, Køber L, Lam CSP, Maggioni AP, Martinez F, Packer M, Rouleau JL, Sim D, van Veldhuisen DJ, Vrtovec B, Zannad F, Zile MR, Gong J, Lefkowitz MP, Rizkala AR, Solomon SD, McMurray JJV. Sacubitril/Valsartan and Frailty in Patients With Heart Failure and Preserved Ejection Fraction. *J Am Coll Cardiol*. 2022;80:1130–1143.

178. Moorhouse P, Rockwood K. Frailty and its quantitative clinical evaluation. *J R Coll Physicians Edinb*. 2012;42:333–340.

179. Dewan P, Jackson A, Jhund PS, Shen L, Ferreira JP, Petrie MC, Abraham WT, Desai AS, Dickstein K, Køber L, Packer M, Rouleau JL, Solomon SD, Swedberg K, Zile MR, McMurray JJV. The prevalence and importance of frailty in heart failure with reduced ejection fraction - an

analysis of PARADIGM-HF and ATMOSPHERE. *Eur J Heart Fail*. 2020;22:2123–2133.

180. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, Marwick TH. Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2019;73:1317.

181. Ambrosy AP, Parikh R v., Sung SH, Tan TC, Narayanan A, Masson R, Lam PQ, Kheder K, Iwahashi A, Hardwick AB, Fitzpatrick JK, Avula HR, Selby VN, Ku IA, Shen X, Sanghera N, Cristino J, Go AS. Analysis of Worsening Heart Failure Events in an Integrated Health Care System. *J Am Coll Cardiol*. 2022;80:111–122.

182. Greene SJ, Bauersachs J, Brugts JJ, Ezekowitz JA, Lam CSP, Lund LH, Ponikowski P, Voors AA, Zannad F, Zieroth S, Butler J. Worsening Heart Failure: Nomenclature, Epidemiology, and Future Directions: JACC Review Topic of the Week. *J Am Coll Cardiol*. 2023;81:413–424.

183. Okumura N, Jhund PS, Gong J, Lefkowitz MP, Rizkala AR, Rouleau JL, Shi VC, Swedberg K, Zile MR, Solomon SD, Packer M, McMurray JJV. Importance of Clinical Worsening of Heart Failure Treated in the Outpatient Setting: Evidence From the Prospective Comparison of ARNI With ACEI to Determine Impact on Global Mortality and Morbidity in Heart Failure Trial (PARADIGM-HF). *Circulation*. 2016;133:2254–2262.

184. McCarthy MI. Painting a new picture of personalised medicine for diabetes. *Diabetologia*. 2017;60:793–799.

185. Adam RD, Coriu D, Jercan A, Bădeliță S, Popescu BA, Damy T, Jurcuț R. Progress and challenges in the treatment of cardiac amyloidosis: a review of the literature. *ESC Heart Fail*. 2021;8:2380–2396.

186. Ioannou A, Patel RK, Razvi Y, Porcari A, Sinagra G, Venneri L, Bandera F, Masi A, Williams GE, O'Beara S, Ganesananthan S, Massa P, Knight D, Martinez-Naharro A, Kotecha T, Chacko L, Brown J, Rauf MU, Manisty C, Moon J, Lachmann H, Wechelakar A, Petrie A, Whelan C,

Hawkins PN, Gillmore JD, Fontana M. Impact of Earlier Diagnosis in Cardiac ATTR Amyloidosis Over the Course of 20 Years. *Circulation*. 2022;146:1657–1670.

187. Tayal U, Verdonschot JAJ, Hazebroek MR, Howard J, Gregson J, Newsome S, Gulati A, Pua CJ, Halliday BP, Lota AS, Buchan RJ, Whiffin N, Kanapeckaite L, Baruah R, Jarman JWE, O'Regan DP, Barton PJR, Ware JS, Pennell DJ, Adriaans BP, Bekkers SCAM, Donovan J, Frenneaux M, Cooper LT, Januzzi JL, Cleland JGF, Cook SA, Deo RC, Heymans SRB, Prasad SK. Precision Phenotyping of Dilated Cardiomyopathy Using Multidimensional Data. *J Am Coll Cardiol*. 2022;79:2219–2232.

188. Pop-Busui R, Januzzi JL, Bruemmer D, Butalia S, Green JB, Horton WB, Knight C, Levi M, Rasouli N, Richardson CR. Heart Failure: An Underappreciated Complication of Diabetes. A Consensus Report of the American Diabetes Association. *Diabetes Care*. 2022;45:1670–1690.

189. Brady PF, Chua W, Nehaj F, Connolly DL, Khashaba A, Purmah YJV, Ul-Qamar MJ, Thomas MR, Varma C, Schnabel RB, Zeller T, Fabritz L, Kirchhof PF. Interactions Between Atrial Fibrillation and Natriuretic Peptide in Predicting Heart Failure Hospitalization or Cardiovascular Death. *J Am Heart Assoc*. 2022;11:22833.

190. van Doorn S, Geersing GJ, Kievit RF, van Mourik Y, Bertens LC, van Riet EES, Boonman-De Winter LJ, Moons KGM, Hoes AW, Rutten FH. Opportunistic screening for heart failure with natriuretic peptides in patients with atrial fibrillation: a meta-analysis of individual participant data of four screening studies. *Heart*. 2018;104:1236–1237.

191. Cleland JGF, Ferreira JP, Mariottoni B, Pellicori P, Cuthbert J, Verdonschot JAJ, Petutschnigg J, Ahmed FZ, Cosmi F, Brunner La Rocca HP, Mamas MA, Clark AL, Edelmann F, Pieske B, Khan J, McDonald K, Rouet P, Staessen JA, Mujaj B, González A, Diez J, Hazebroek M, Heymans S, Latini R, Grojean S, Pizard A, Girerd N, Rossignol P, Collier TJ, Zannad F. The effect of spironolactone on cardiovascular function and markers of fibrosis in people at increased risk of developing heart failure: the heart

'OMics' in AGEing (HOMAGE) randomized clinical trial. *Eur Heart J*. 2021;42:684–696.

192. Huelsmann M, Neuhold S, Resl M, Strunk G, Brath H, Francesconi C, Adlbrecht C, Prager R, Luger A, Pacher R, Clodi M. PONTIAC (NT-proBNP Selected PreventiOn of cardiac eveNts in a populaTion of dIabetic patients without A history of Cardiac disease): A Prospective Randomized Controlled Trial. *J Am Coll Cardiol*. 2013;62:1365–1372.

193. Ulloa-Cerna AE, Jing L, Pfeifer JM, Raghunath S, Ruhl JA, Rocha DB, Leader JB, Zimmerman N, Lee G, Steinhubl SR, Good CW, Haggerty CM, Fornwalt BK, Chen R. RECHOmmend: An ECG-Based Machine Learning Approach for Identifying Patients at Increased Risk of Undiagnosed Structural Heart Disease Detectable by Echocardiography. *Circulation*. 2022;146:36–47.

194. Nekoui M, Pirruccello JP, di Achille P, Choi SH, Friedman SN, Nauffal V, Ng K, Batra P, Ho JE, Philippakis AA, Lubitz SA, Lindsay ME, Ellinor PT. Spatially Distinct Genetic Determinants of Aortic Dimensions Influence Risks of Aneurysm and Stenosis. *J Am Coll Cardiol*. 2022;80:486–497.

195. Wehbe RM. Unleashing the Power of Machine Learning to Predict Myocardial Recovery after Left Ventricular Assist Device: A Call for the Inclusion of Unstructured Data Sources in Heart Failure Registries. *Circ Heart Fail*. 2022;15:E009278.

196. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, Pagoto S, McConnell M v., Pantelopoulos A, Foulkes AS. Detection of Atrial Fibrillation in a Large Population Using Wearable Devices: The Fitbit Heart Study. *Circulation*. 2022;146:1415–1424.

197. Ho FK, Zhou Z, Petermann-Rocha F, Para-Soto S, Boonpor J, Welsh P, Gill JMR, Gray SR, Sattar N, Pell JP, Celis-Morales C. Association between Device-Measured Physical Activity and Incident Heart Failure: A

Prospective Cohort Study of 94 739 UK Biobank Participants. *Circulation*. 2022;146:883–891.

198. Zito A, Princi G, Romiti GF, Galli M, Basili S, Liuzzo G, Sanna T, Restivo A, Ciliberti G, Trani C, Burzotta F, Cesario A, Savarese G, Crea F, D'Amario D. Device-based remote monitoring strategies for congestion-guided management of patients with heart failure: a systematic review and meta-analysis. *Eur J Heart Fail*. 2022;24:2333–2341.

199. O'Sullivan JW, Ashley EA, Elliott PM. Polygenic risk scores for the prediction of cardiometabolic disease. *Eur Heart J*. 2023;44:89–99.

200. Marston NA, Garfinkel AC, Kamanu FK, Melloni GM, Roselli C, Jarolim P, Berg DD, Bhatt DL, Bonaca MP, Cannon CP, Giugliano RP, O'Donoghue ML, Raz I, Scirica BM, Braunwald E, Morrow DA, Ellinor PT, Lubitz SA, Sabatine MS, Ruff CT. A polygenic risk score predicts atrial fibrillation in cardiovascular disease. *Eur Heart J*. 2023;44:221–231.

201. Sandhu RK, Dron JS, Liu Y, Moorthy MV, Chatterjee NA, Ellinor PT, Chasman DI, Cook NR, Khera A v., Albert CM. Polygenic Risk Score Predicts Sudden Death in Patients With Coronary Disease and Preserved Systolic Function. *J Am Coll Cardiol*. 2022;80:873–883.

202. Ghouse J, Tragante V, Muhammad A, Ahlberg G, Skov MW, Roden DM, Jonsdottir I, Andreasen L, Lundegaard PR, Trudsø LC, Banasik K, Brunak S, Ostrowski SR, consortium eMERGE, Torp-Pedersen C, Pedersen O v, Sørensen E, Køber L, Iversen K, Thorsteinsdottir U, Thorgeirsson G, Ullum H, Gudbjartsson DF, Mosley JD, Holm H, Stefansson K, Bundgaard H, Olesen MS. Polygenic risk score for ACE-inhibitor-associated cough based on the discovery of new genetic loci. *Eur Heart J*. 2022;43:4707–4718.

203. Lanfear DE, Luzum JA, She R, Gui H, Donahue MP, O'connor CM, Adams KF, Sanders-Van Wijk S, Zeld N, Maeder MT, Sabbah HN, Kraus WE, Brunner-Larocca HP, Li J, Williams LK. Polygenic Score for ß-Blocker Survival Benefit in European Ancestry Patients With Reduced Ejection Fraction Heart Failure. *Circ Heart Fail*. 2020;13:E007012.

204. Heiat A, Gross CP, Krumholz HM. Representation of the Elderly, Women, and Minorities in Heart Failure Clinical Trials. *Arch Intern Med*. 2002;162:1682–1688.

Bibliography

1. Harrell , FE. *Regression Modeling Strategies*. Cham: Springer International Publishing; 2015.

2. Steyerberg EW. *Clinical Prediction Models*. Cham: Springer International Publishing; 2019.

3. Collins LM, Lanza ST. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley and Sons Inc.; 2010.

4. Kuhn M, Johnson K. *Applied Predictive Modeling*. Springer New York; 2013.

5. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer US; 2021.