Potential of joint modelling of longitudinal observations and time-to-event data to improve prognosis in chronic heart failure studies

Ryan James Field
BSc (Hons), MSc

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

School of Health and Wellbeing
College of Medical, Veterinary & Life Sciences

September 2023

Abstract

Background

Heart failure is a clinical syndrome with inter-relationships between numerous biochemical, physical, and imaging characteristics, and adverse outcomes. Patients with heart failure typically have poor prognosis, and previous prognostic models are limited to using only baseline measurements of these characteristics. In recent years there has been a rise in interest and usage of joint modelling. Joint modelling seeks to combine two or more models, typically containing longitudinal observations and time-to-event (survival) data. The purpose of which is to reduce bias and increase efficiency, allowing for these repeat measurements of longitudinal observations whilst accounting for correlation and measurement error. The inter-relationships within heart failure and properties of joint modelling make heart failure an excellent candidate for joint modelling. It is therefore the aim of this thesis to explore the use of joint modelling in heart failure and to see whether it can be used to improve prognosis.

Methods

This research comprised of a systematic review paired with an exemplar to introduce joint modelling and how joint models are currently being applied to heart failure; whilst also illustrating how joint modelling can be applied to clinical trial data. Following this, seven joint models were fit under a Bayesian framework, using data from a randomised control trial, and validated with data from different randomised control trials. These joint models were then compared to models fitted using current standards of prognostic model methodology to evaluate and assess how joint modelling can potentially improve model performance. Finally, a web application was developed to illustrate how these joint models can translate into real world applications.

Results

On average the joint models performed better, in a statistical sense, than the traditional models (considered the current standard for prognosis) and performed adequately when validated with data from another randomised control trial. The web application effectively shows how these joint models can be used in practice and highlights the potential of the dynamic nature of joint models when used in a prognostic setting.

Conclusion

This thesis illustrates how joint modelling can improve on the current standard of prognostic models, adding repeated measurements and allowing for dynamic predictions over time, whilst outperforming the traditional models. However, with limitations around the use of latent parameters such as random effects, and the novel nature of these models with their limited use, it may be prudent to wait until these types of models mature, are evaluated further, and the statistical packages used to fit these models mature before implementing them in clinical practice.

Table of Contents

Potential of joint modelling of longitudinal observations and time-to-event data to improve prognosis in chronic heart failure studies ..... *i*

Abstract ..... *ii*

List of Tables ..... *viii*

List of Figures ..... *x*

Acknowledgements ..... *xiii*

Author's Declaration ..... *xiv*

Contribution Statements ..... *xiv*

COVID-19 Impact Statement ..... *xv*

Additional Publications ..... *xv*

Definitions/Abbreviations ..... *xvi*

Chapter 1 Introduction *19*

1.1 Foreword **19**

1.2 Heart Failure **19**

1.2.1 Definition 19
1.2.2 Chronic and Acute Heart Failure 20
1.2.3 Functional Classification 20
1.2.4 Diagnosis of Chronic Heart Failure 21
1.2.5 Left Ventricular Ejection Fraction 22
1.2.6 Epidemiology 23
1.2.7 Prognosis 23
1.2.8 Co-Morbidities 23
1.2.9 Treatment for Chronic Heart Failure 24
1.2.10 Natriuretic Peptides in Heart Failure 25
1.2.11 Renal Function in Heart Failure 26
1.2.12 Prognostic Models in Heart Failure 27

1.3 Joint Modelling Longitudinal Repeat Measures and Time to Event Outcomes **41**

1.3.1 Time-to-event (Survival) analysis 41
1.3.2 Repeated Measures in Survival analysis 42
1.3.3 Exogenous and Endogenous Covariates 43
1.3.4 Biomarkers as Time-Varying Covariates 44
1.3.5 Formation of Joint Models 44
1.3.6 Alpha Parameterisation 51
1.3.7 Slope Parameterisation 52
1.3.8 Cumulative Effects (area) Parameterisation 53

1.4 Research Aim and Specific Research Questions **55**

Chapter 2 Joint Modelling of longitudinal processes and time-to-event outcomes in heart failure: systematic review and exemplar examining the relationship between serum digoxin levels and mortality. *56*

2.1 Foreword **56**

2.2 Title, authorship, and publication details **57**

2.3 Abstract **57**

2.3.1 Background 57
2.3.2 Methods 57
2.3.3 Results 57

2.3.4 Conclusion 57
2.3.5 Keywords 58

2.4 Background 58

2.5 Methods 60
2.5.1 Systematic Review: Joint Modelling Applications within Heart Failure 60
2.5.2 Exemplar: Joint Modelling of Serum Digoxin Concentration and All-Cause Mortality 60

2.6 Results 64
2.6.1 Systematic Review: Use of Joint Modelling in Heart Failure 64
2.6.2 Association between serum digoxin concentration and mortality 75

2.7 Discussion 82

2.8 Conclusions 88

2.9 Declarations 89
2.9.1 Ethical approval and consent to participants 89
2.9.2 Consent for Publication 89

2.10 Availability of Data and Materials 89

2.11 Competing Interests 89

2.12 Funding 89

2.13 Authors' Contributions 90

2.14 Acknowledgements 90

2.15 Supplementary Material 91
2.15.1 Supplementary Results 93

Chapter 3 Data Source 102

3.1 Foreword 102

3.2 Primary Data Source 102

3.3 Validation Data Source 108

Chapter 4 Joint Modelling of NT-ProBNP and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure 115

4.1 Introduction 115

4.2 Methods 116
4.2.1 Data Source 116
4.2.2 Statistical Analysis 116
4.2.3 Data Cleaning and Processing 117
4.2.4 Covariate Selection 117
4.2.5 Missing Data 118
4.2.6 Longitudinal Sub Model 118
4.2.7 Survival Sub Model 121
4.2.8 Joint Models 123
4.2.9 Model Performance: Prognostic Accuracy, Fit and Calibration 126
4.2.10 External Validation 127
4.2.11 Comparative Analysis 128

4.3 Results 129
4.3.1 Baseline Characteristics 129
4.3.2 Number of Measurements of NT-ProBNP 131
4.3.3 Longitudinal Profile of NT-ProBNP 131
4.3.4 Longitudinal Outcome from Joint Models 133
4.3.5 Survival Outcomes from Joint Model 137
4.3.6 Joint Model Performance 143
4.3.7 Dynamic Predictions from Joint Models 144

4.3.8 External Validation of Joint Models....145
4.3.9 Comparative Analysis....152

4.4 Discussion....153

Chapter 5 Joint Modelling of eGFR and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure....158

5.1 Foreword....158

5.2 Introduction....158

5.3 Methods....159
5.3.1 Data Source....159
5.3.2 Statistical Analysis....159
5.3.3 Data Cleaning and Processing....160
5.3.4 Covariate Selection....160
5.3.5 Missing Data....160
5.3.6 Longitudinal Sub Model....161
5.3.7 Survival Sub Model....163
5.3.8 Joint Models....164
5.3.9 Model Performance: Prognostic Accuracy, Fit and Calibration....167
5.3.10 External validation....167
5.3.11 Comparative Analysis....168

5.4 Results....169
5.4.1 Baseline Characteristics....169
5.4.2 Number of Measurements of eGFR....171
5.4.3 Longitudinal Profile of eGFR....171
5.4.4 Longitudinal Outcome from Joint Models....173
5.4.5 Survival outcomes from Joint Models....177
5.4.6 Joint Model Performance....183
5.4.7 Dynamic Predictions from Joint Models....184
5.4.8 External Validation....185
5.4.9 Comparative Analysis....192

5.5 Discussion....194

Chapter 6 Multivariate Joint Modelling of NT-ProBNP and eGFR, and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure 199

6.1 Foreword....199

6.2 Introduction....199

6.3 Methods....200
6.3.1 Data Source....200
6.3.2 Statistical Analysis....200
6.3.3 Data Cleaning and Processing....201
6.3.4 Covariate Selection....201
6.3.5 Missing Data....202
6.3.6 Longitudinal Sub Models....202
6.3.7 Longitudinal Sub Model One – NT-ProBNP....202
6.3.8 Longitudinal Sub Model Two – eGFR....205
6.3.9 Survival Sub Model....207
6.3.10 Joint Models....208
6.3.11 Model Performance: Prognostic Accuracy, Fit and Calibration....211
6.3.12 External Validation....212
6.3.13 Comparative Analysis....212

6.4 Results....214
6.4.1 Baseline Characteristics....214
6.4.2 Number of Measurements of NT-ProBNP and eGFR....217

6.4.3 Longitudinal Profiles of NT-ProBNP and eGFR 217
6.4.4 Longitudinal Outcomes from Joint Model 220
6.4.5 Survival Outcomes from Joint Model 224
6.4.6 Joint Model Performance 225
6.4.7 Dynamic Predictions from Joint Model 229
6.4.8 External Validation 232
6.4.9 Comparative Analysis 235

6.5 Discussion 236

Chapter 7 Real-World Application of Prognostic Joint Models Illustrated Using a Web Application 242

7.1 Introduction 242

7.2 Methods 243
7.2.1 Choice of Joint Model 243
7.2.2 R Shiny Application 243

7.3 Results 245
7.3.1 Baseline Measurements Tab 249

7.4 Discussion 250

Chapter 8 Discussion 253

8.1 Foreword 253

8.2 Overview of chapters 253
8.2.1 Chapter 2 Joint Modelling of longitudinal processes and time-to-event outcomes in heart failure: systematic review and exemplar examining the relationship between serum digoxin levels and mortality. 253
8.2.2 Chapter 3 Data Source 254
8.2.3 Chapter 4 Joint Modelling of NT-ProBNP and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure 254
8.2.4 Chapter 5 Joint Modelling of eGFR and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure 255
8.2.5 Chapter 6 Multivariate Joint Modelling of NT-ProBNP and eGFR, and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure 257
8.2.6 Chapter 7 Real-World Application of Prognostic Joint Models Illustrated Using a Web Application 259

8.3 Strengths and limitations of this research 260

8.4 Future work 263

8.5 Conclusion 265

Appendices 266

List of References 272

List of Tables

Table 1 The definition of the four classes of NYHA classification adapted from The Criteria Committee of the New York Heart Association. 20
Table 2 The three Phenotypes based on LVEF....22
Table 3 Level of Discriminatory Ability for Risk Prediction Models Adapted from Hosmer et al....29
Table 4 Level of Discriminatory Ability for Risk Prediction Models Adapted from Di Tanna et al....29
Table 5 C-Statistics (95% CI) at one and two years from the Cox PH Prognostic Models from Predict-HF adapted from the PREDICT-HF Study 33
Table 6 AUC-ROC (C-Statistic) (95% CI) for All Datasets at 1 Year Adapted from. Levy et al. 37
Table 7 AUC-ROC (C-Statistic) (95% CI) at 1 Year Comparison against ADHERE and Toronto Models Adapted from Levy et al. 37
Table 8 Actual VS Predicted Survival (± SD) for 1, 2 and 3 Years for each Dataset Adapted from Levy et al....37
Table 9 Formulation of JMs Included in Exemplar....62
Table 10 Summary of Longitudinal Data of Included Studies....67
Table 11 Survival End Points of Included Studies 69
Table 12 Baseline Characteristics According to Patient Status (Dead / Censored) and Overall Sample (n = 2012) 76
Table 13 Event Summary of JMs Represented as Hazard Ratios....80
Table 14 Model Equations for All Models....91
Table 15 Summary of Included Studies Ordered by Year 97
Table 16 Breakdown of Composite Components of Included Studies 99
Table 17 Coefficients from the Longitudinal Sub model from JM2 99
Table 18 Performance Summary of Joint Models, Cox PH and Extended Cox PH models. 100
Table 19 Hazard Ratios and Standard Errors from Cox Models and JM2....101
Table 20 Inclusion and Exclusion Criteria for the PARADIGM-HF RCT Adapted from McMurray et al....105
Table 21 Inclusion and Exclusion Criteria for the ATMOSPHERE RCT Adapted from McMurray et al....111
Table 22 Formulae for Joint Models 125
Table 23 Baseline Characteristics of Included Patients 130
Table 24 Number of Repeat Measurements of NT-ProBNP for Included Patients 131
Table 25 Parameter estimates (95% CI) from the Longitudinal Outcome of the Joint Models for meaningfully interpretable variables. 135
Table 26 Hazard Ratios (95% CI) for Variables in the Three Joint Models 138
Table 27 Marginal and Conditional Performance Statistics of Joint Models....139
Table 28 Hazard Ratio (95% CI) and P-Values from the Cox PH Models 150
Table 29 Formulae for Joint Models 166
Table 30 Baseline Characteristics of Included Patients. 169
Table 31 Number of Repeat Measurements of eGFR for Included Patients 171
Table 32 Parameter Estimates (95% CI) from the Longitudinal Outcomes of the Joint Models for the Meaningfully Interpretable Variables 175
Table 33 Hazard Ratios (95% CI) from the Survival Component of the Three Joint Models 178

Table 34 Marginal and Conditional Performance Statistics of Joint Models.....179
Table 35 Hazard Ratio (95% CI) and P-Values from the Cox PH Models 190
Table 36 Baseline Characteristics of Included Patients 215
Table 37 Number of Repeat Measurements of NT-ProBNP and eGFR for Included Patients 217
Table 38 Parameter Estimates (95% CI) from The Longitudinal Outcome for NT-ProBNP from The Joint Model for Meaningfully Interpretable Variables.....222
Table 39 Parameter Estimates (95% CI) from The Longitudinal Outcome for eGFR from The Joint Model for Meaningfully Interpretable Variables 222
Table 40 Hazard Ratios (95% CI) for Variables in the Multivariate Joint Model 225
Table 41 Marginal and Conditional Performance Statistics from the Multivariate Joint Model 225
Table 42 Hazard Ratio (95% CI) and P-Values from the Cox PH Models 233

List of Figures

Figure 1 Diagnostic algorithm for heart failure adapted from McDonagh et al. 21
Figure 2 Graphical Representation of an Extended Cox PH and Joint Model Showing How the Hazard Function Evolves Over Time in Comparison with the Longitudinal Outcome - Adapted from Rizopoulos D. 43
Figure 3 Graphical Representation of an LME Showing the Subject Specific Longitudinal Trajectories of two Subjects and the Marginal Longitudinal Trajectory - Adapted from Rizopoulos D 46
Figure 4 Graphical Representation of a Joint Model Showing How the Hazard Function Evolves Over Time in Comparison with the Longitudinal Outcome with the Value and Slope Parameterisations - Adapted from Rizopoulos D. 53
Figure 5 Graphical Representation of a Joint Model Showing How the Hazard Function Evolves Over Time in Comparison with the Longitudinal Outcome with the Area Parameterisation - Adapted from Rizopoulos D. 54
Figure 6 Prisma 2020 Flow Chart 65
Figure 7 Average Estimated Biomarker Pattern, Combined with Individual Biomarker Measurements. During Follow-up in Patients with and without the Primary Endpoint from the van Vark et al. Study 71
Figure 8 Dynamic survival probabilities with 95% CI based on various measurements of NT- ProBNP for a patient whose values fell. 73
Figure 9 Average Trajectories of SDC by Patient Status as Predicted by JM 2 with Observed Values and Trajectories of SDC 78
Figure 10 Individual Patient Trajectory of sqrt SDC and Survival Probability from Randomly Sampled Patient as Predicted by JM 2 81
Figure 11 Full search strategy 95
Figure 12 Overall Study design and timeframe between study visits occurring during the active run-in and double-blind periods for the PARADIGM RCT adapted from McMurray et al. 107
Figure 13 Study design of the ATMOSPHERE RCT including run-in period and stratification adapted from McMurray et al. 114
Figure 14 Longitudinal Profile of log NT-ProBNP for 42 Randomly Sampled Patients 132
Figure 15 Average Trajectories of log NT-ProBNP by Treatment Group with Observed Values and Trajectories of NT-ProBNP 136
Figure 16 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months 139
Figure 17 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months 140
Figure 18 Calibration Curves for Joint Models at Month 12 using Longitudinal Data at Month 0 140
Figure 19 Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12 141
Figure 20 Dynamic Predictions from Joint Model 4 for a Randomly Sampled Patient with Measurements of NT-ProBNP and Corresponding Survival Probability 142
Figure 21 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs and Brier Scores for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months using Data from the ATMOSPHERE Trial 147

Figure 22 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months 148
Figure 23 External Validation Calibration Curves for Joint Models at Month 12 using Longitudinal Data At Month 0.....149
Figure 24 External Validation Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12.....149
Figure 25 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 12 151
Figure 26 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 24 151
Figure 27 Longitudinal Profiles of log eGFR.....172
Figure 28 Average Trajectories of log eGFR by Treatment Group with Observed Values and Trajectories of log eGFR Scaled for Readability.....176
Figure 29 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months.....179
Figure 30 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months.....180
Figure 31 Calibration Curves for Joint Models at Month 12 using Longitudinal Data at Month 0.....180
Figure 32 Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12 181
Figure 33 Dynamic Predictions from Joint Model 1 for a Randomly Sampled Patient with Measurements of eGFR and Corresponding Survival Probability.....182
Figure 34 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs and Brier Scores for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months using Data from the ATMOSPHERE Trial 187
Figure 35 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months 188
Figure 36 External Validation Calibration Curves for Joint Models at Month 12 using Longitudinal Data At Month 0.....189
Figure 37 External Validation Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12.....189
Figure 38 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 12 191
Figure 39 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 24 191
Figure 40 Baseline NT-ProBNP against eGFR by Treatment with Loess Curves and Pearson Correlation Coefficient.....216
Figure 41 Longitudinal Profile of log NT-ProBNP for 42 Randomly Sampled Patients.....218
Figure 42 Longitudinal Profile of log eGFR for 42 Randomly Sampled Patients.....219
Figure 43 Average Trajectories of log NT-ProBNP by Treatment Group from the multivariate joint model with Observed Values and Trajectories on NT-ProBNP.....223
Figure 44 Average Trajectories of log eGFR by treatment group from the multivariate joint model with Observed Values and Trajectories Scaled for Readability. 223
Figure 45 Multivariate Joint Model Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0.....226

Figure 46 Multivariate Joint Model Time-Varying ROC, AUC and Brier Score at Month 24 Using Longitudinal Data up until Month 12 226
Figure 47 Multivariate Joint Model Calibration Curve at Month 12 using Longitudinal Data at Month 0.....227
Figure 48 Multivariate Joint Model Calibration Curve at Month 24 using Longitudinal Data up until Month 12 227
Figure 49 Dynamic Predictions from Multivariate Joint Model for a Randomly Sampled Patient with Measurements of eGFR and NT-ProBNP, and Corresponding Survival Probability 228
Figure 50 Multivariate Joint Model - External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial 230
Figure 51 Multivariate Joint Model - External Validation Time-Varying ROC, AUC, and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial.....230
Figure 52 Multivariate Joint Model - External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial 231
Figure 53 Multivariate Joint Model External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial 231
Figure 54 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 12 234
Figure 55 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 24 234
Figure 56 Web Application User Interface for Baseline Measurements Tab 246
Figure 57 Web Application User Interface for the Repeat Measurements Tab 247
Figure 58 Web Application User Interface for the Debug Tab 248

Acknowledgements

I would first like to express my gratitude to both of my supervisors Professor Jim Lewsey and Professor Pardeep Jhund without whom this would not be possible. Their knowledge, expertise, experience, encouragement, advisement, support, and patience were invaluable throughout my PhD journey.

I am also grateful to Carly Adamson for her help as a second reviewer for my systematic review as published and included in this thesis.

I would also like to thank the MVLS DTP Programme for funding this PhD, providing me with the opportunity to undertake this PhD without added financial burden.

I am thankful to both my parents for their continued support throughout my PhD and academic journey. I am also grateful to my HEHTA colleagues who have provided, motivation, support, and friendship throughout my PhD.

Finally, I would like to thank my partner Ellen and cat Romeo, who supported me and gave me the strength to complete this PhD, especially during the COVID-19 Pandemic.

Author's Declaration

"I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution."

Printed Name: Ryan James Field

Contribution Statements

Chapter 2

I was responsible for the conceptualisation of the research aims with input from both my supervisors (Prof. Jim Lewsey and Prof. Pardeep Jhund). Data for the systematic review was curated by myself and co-author Carly Adamson. Data for the exemplar was solely curated by me. Formal analysis was conducted by me with contribution from co-author Carly Adamson for the full text screening of the systematic review. Investigation, methodology, project management, visualisation, writing of the first draft, and validation was carried out by myself. The manuscript was revised with input from all authors (Carly Adamson, Prof. Pardeep Jhund and Prof. Jim Lewsey).

COVID-19 Impact Statement

Much of this PhD was undertaken during the COVID-19 Pandemic. In my view, this this had limited impact on this thesis. However, it should be acknowledged that due to my involvement with the Scottish Covid Response Consortium, there were extra time pressures during the following time period, June 2020 to June 2022. My research during the COVID-19 pandemic made important contributions to collaborative projects and led to the following publications:

Bouttell, J. et al. (2020) Evidence Review: Assessment of COVID-19 in Primary Care: The identification of symptoms, clinical signs, characteristics, and comorbidities in adults, children and young people, which may indicate a higher risk of progression to severe disease. Other. Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh.

Quinn, T. J. et al. (2021) Following the Science? Comparison of methodological and reporting quality of covid-19 and other research from the first wave of the pandemic. BMC Medicine, 19, 46. (doi: 10.1186/s12916-021-01920-x) (PMID:33618741) (PMCID:PMC7899793)

Mitchell, S. N. et al. (2022) FAIR Data Pipeline: provenance-driven data management for traceable scientific workflows. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 380(2233), 20210300. (doi: 10.1098/rsta.2021.0300) (PMID:35965468) (PMCID:PMC9376726)

Additional Publications

While not directly related to this thesis, I was also a co-author on the following publication during my PhD:

Jani, B. D. , McQueenie, R., Nicholl, B. I. , Field, R. , Hanlon, P. , Gallacher, K. I. , Mair, F. S. and Lewsey, J. (2021) Association between patterns of alcohol consumption (beverage type, frequency and consumption with food) and risk of adverse health outcomes: a prospective cohort study. BMC Medicine, 19, 8. (doi: 10.1186/s12916-020-01878-2) (PMID:33430840) (PMCID:PMC7802201)

Definitions/Abbreviations

ACE-I: Angiotensin-converting enzyme inhibitor

ACE: Angiotensin-converting enzyme

Active Controlled: A trial design whereby the comparator of the investigational drug is an already established treatment rather than placebo

ADHERE: The Acute Decompensated Heart Failure National Registry

AF: Atrial Fibrillation

AHA: American Heart Association

AIC: Akaike Information Criterion

AIDS: Acquired Immune Deficiency Syndrome

ALTITUDE: Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Disease Endpoints

ARB: Angiotensin Receptor Blockers

ARNI: Angiotensin Receptor-Neprilysin Inhibitor

ASTRONAUT: Aliskiren Trial on Acute Heart Failure Outcomes

ATMOSPHERE: The Aliskiren Trial to Minimize Outcomes in Patients with Heart Failure

AUC: Area under the ROC Curve

b.i.d: bis in die (twice a day)

BIOLINCC: Biologic Specimen and Data Repository Information and Coordinating Center

Biomarker: Biological Marker

BMI: Body Mass Index

BNP: B-type natriuretic peptide

C-Index: Concordance Index - A generalisation of the AUC, AKA. AUC-ROC

CA125: Cancer Antigen 125

CAD: Coronary Artery Disease

CI: Confidence interval

CKD: Chronic Kidney Disease

COPD: Chronic Obstructive Pulmonary Disease

Cox PH: Cox Proportional Hazards

CRP: C-Reactive Protein

CRT-D: Cardiac Resynchronization Therapy Devices

CV: Cardiovascular

DIC: Deviance Information Criterion

DIG: The Effect of Digoxin on Mortality and Morbidity in Patients with Heart Failure

Double Blind: Neither researcher nor patient know the assigned intervention

Double Dummy: If a treatment in a trial requires different treatment forms (e.g., multiple dosage) the comparator arm will contain a placebo of these extra treatment forms to enable proper blinding.

ECG / EKG: Electrocardiogram

EDV: End-diastolic Volume

eGFR: Estimated Glomerular Filtration Rate

ELITE2: Evaluation of Losartan in the Elderly

Endpoint: Event or Outcome of interest

ESC: European Society of Cardiology

Event Driven: A trial design whereby the power needed for the trial is derived from the number of events

GAL-3: Galectin-3

GDF-15: Growth Differentiation Factor 15

GFR Glomerular Filtration Rate

GLMM: Generalized Linear Mixed Model

HF: Heart Failure

HFmrEF: Heart Failure with mid-range Ejection Fraction / Heart Failure with mildly reduced Ejection Fraction

HFpEF: Heart Failure with Preserved Ejection Fraction

HFrEF: Heart Failure with reduced Ejection Fraction

HIV: Human Immunodeficiency Virus

HR: Hazard Ratio

HsTnT: High Sensitivity Troponin T

ICD: Cardioverter-Defibrillators

IN-CHF Italian Heart Failure Registry

IQR Interquartile range

JM: Joint Model

KCCQ-OS: Kansas City Cardiomyopathy Questionnaire Overall Summary Score

LCZ / LCZ696: Sacubitril/Valsartan

LDH: Lactase Dehydrogenase

LME: Linear Mixed Effects

LMM: Linear mixed-effects models

LOCF: Last Observation Carried Forward

LPML: log pseudo-marginal likelihood

LVAD: Left Ventricular Assist Device

LVED: Left ventricular end-diastolic diameter

LVEF: Left Ventricular Ejection Fraction

LVES: Left ventricular end-systolic diameter

MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure

MCMC: Markov chain Monte Carlo

MI: Myocardial Infarction

miRNA: Micro ribonucleic acid

MNAR: Missing Not at Random

MRA: Mineralocorticoid Receptor Antagonists

MRI: Magnetic Resonance Imaging

NT-ProBNP: N-Terminal Pro-Brain Natriuretic Peptide

NYHA: New York Heart Association

PARADIGM-HF: Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure

Parallel Group: Patients are randomised to one or more groups and continue in this group throughout the study

PRAISE1: Prospective Randomized Amlodipine Survival Evaluation

PRISMA: Preferred Reporting Items for Systematic Reviews

PROBAST: Prediction model Risk Of Bias ASsessment Tool

QOL: Quality of Life

QRS: A combination of the Q, R and S waves from an Electrocardiogram

RCT: Randomized Control Trial

RDW: Red blood cell Distribution Width
RENAISSANCE: Randomized Enbrel North American Strategy to Study Antagonism of Cytokines
ROB: Risk of Bias
ROC: Receiver operating characteristic
SBP: Systolic Blood Pressure
SD: Standard Deviation
SDC: Serum Digoxin Concentration
SF-36: 36 Item Short Form Survey
Sham Adjustment: When one arm of the trial requires an adjustment, the other arm(s) also appear to be adjuster with a placebo to enable proper blinding
SHFM: Seattle Heart Failure Model
Sqrt: Square Root
ST2: Suppression of Tumorigenicity 2
SV: Stoke Volume
SwedeHF: Swedish Heart Failure Registry
UI: User Interface
UW: University of Washington
Val-HeFT: Valsartan Heart Failure Trial

Chapter 1 Introduction

1.1 Foreword

The purpose of this chapter is to introduce the background surrounding both heart failure and joint modelling to provide context and a foundation for the aims of this thesis.

1.2 Heart Failure

1.2.1 Definition

The European Society of Cardiology (ESC) provides the following definition of heart failure:

"Heart failure is not a single pathological diagnosis, but a clinical syndrome consisting of cardinal symptoms (e.g. breathlessness, ankle swelling, and fatigue) that may be accompanied by signs (e.g. elevated jugular venous pressure, pulmonary crackles, and peripheral oedema). It is due to a structural and/or functional abnormality of the heart that results in elevated intracardiac pressures and/or inadequate cardiac output at rest and/or during exercise." [1]

While the American Heart Association (AHA) defines heart failure as:

"A complex clinical syndrome with symptoms and signs that result from any structural or functional impairment of ventricular filling or ejection of blood. " [2]

While these definitions differ, they agree that heart failure is a clinical syndrome with symptoms and possible signs resulting from structural and / or functional abnormalities in the heart.

1.2.2 Chronic and Acute Heart Failure

Heart failure can either present over time with a history of signs and / or symptoms of cardiac dysfunction known as chronic heart failure; or with a rapid or progressive onset of signs and / or symptoms acute enough for the patient to present themselves for urgent medical treatment, known as acute heart failure. Whilst both chronic and acute are types of heart failure, they have distinct characteristics such as a difference in mortality and rehospitalisation rates and management [1]. This thesis will focus mainly on chronic heart failure.

1.2.3 Functional Classification

The New York Heart Association (NYHA) classification is still widely used as one of the most straightforward functional classifications of heart failure. NYHA categorises heart failure into one of four classes based solely on symptoms as shown in Table 1.

Table 1 The definition of the four classes of NYHA classification adapted from The Criteria Committee of the New York Heart Association.

Class	Functional Capacity
I	Patients with cardiac disease, but without resulting limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea, or anginal pain.
II	Patients with cardiac disease resulting in slight limitation of physical activity. They are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea, or anginal pain.
III	Patients with marked limitation of physical activity. They are comfortable at rest. Less than ordinary activity causes fatigue, palpitation, dyspnea, or anginal pain.
IV	Patients with cardiac disease resulting in inability to carry on any physical activity without discomfort. Symptoms of heart failure or of the anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.

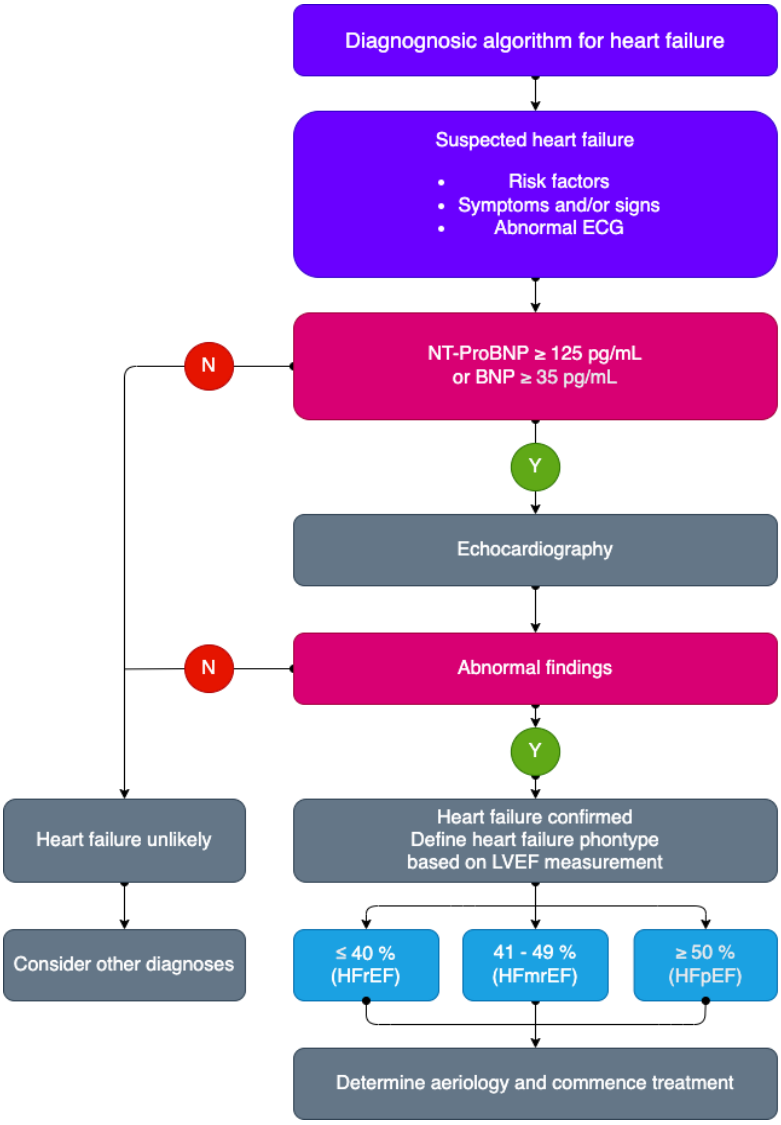
[3]

Whilst still widely used NYHA classification relies exclusively on symptoms and newer evidence suggests it may be outdated and that there are numerous prognostic indicators of heart failure which are superior [4].

1.2.4 Diagnosis of Chronic Heart Failure

Whilst the presence of signs and or symptoms can be indicative of heart failure, they alone are not accurate enough to provide a diagnosis of heart failure. Instead, the ESC provides an algorithm for diagnosis of chronic heart failure, illustrated in Figure 1 starting with suspected heart failure and including tests for natriuretic peptides and echocardiography which not only provides a confirmatory diagnosis but categorisation of heart failure by left ventricular ejection fraction.

Figure 1 Diagnostic algorithm for heart failure adapted from McDonagh et al.



[1]

1.2.5 Left Ventricular Ejection Fraction

Left Ventricular Ejection Fraction (LVEF) is widely used to phenotype heart failure. LVEF is expressed as a fraction of the stroke volume (SV) and the end diastolic volume (EDV) calculated as a percentage:

$$LVEF = \frac{SV}{EDV} \times 100$$

[5]

Both the ESC and AHA recommend echocardiography for ascertaining a patient's LVEF [1], [2]. However, LVEF can be determined from a variety of techniques including magnetic resonance imaging (MRI) [6], computed tomography [7] and nuclear cardiac imaging [8].

Previously heart failure had two distinct phenotypes: Heart Failure with Reduced Ejection Fraction HFrEF (LVEF ≤40%) or Heart Failure with Preserved Ejection Fraction HFpEF (≥40%). However more recently both the AHA and ESC have recognised a third phenotype: Heart Failure with mid-range ejection fraction / Heart Failure with mildly reduced ejection fraction (HFmrEF). With this addition HFpEF now only encapsulates patients with a LVEF of ≥50% [1]. All three phenotypes are outlined in Table 2. This thesis will primarily focus on patients with HFrEF.

Table 2 The three Phenotypes based on LVEF

Phenotype	LVEF %
HFrEF	≤40%
HFmrEF	41-49%
HFpEF	≥50%

1.2.6 Epidemiology

In Europe the prevalence of heart failure is estimated at 1-2 % of adults. However, this estimate is derived from studies which typically only include persons with diagnosed or recognised heart failure so the true prevalence may be higher [1]. While prevalence may be high due to an ageing population age-adjusted incidence in developed countries may be decreasing, suggesting better management of cardiovascular disease [1].

Common causes of heart failure include Coronary Artery Disease (CAD), hypertension, valve disease, arrhythmia, cardiomyopathy, congenital heart disease, infective, drug induced, infiltrative, storage disorders, endomyocardial disease, pericardial disease, metabolic, and neuromuscular disease. However, aetiology can vary with region; with CAD and hypertension being the most common cause in developed countries [9].

1.2.7 Prognosis

Prognosis for patients with heart failure is typically poor with a considerably reduced quality of life. Although prognosis has improved in more recent years, this improvement is restricted to patients with HFrEF [1]. Mortality and overall prognosis can be dependent on factors such as sex and ejection fraction with females having a greater survival over males [10].

1.2.8 Co-Morbidities

Heart Failure is often accompanied by both cardiovascular and non-cardiovascular co-morbidities. Co-morbidities can be dependent on the aetiology; common cardiovascular co-morbidities include arrhythmias such as atrial fibrillation, chronic coronary syndromes such as CAD, valvular heart disease including aortic stenosis, aortic regurgitation both mitral and tricuspid, hypertension, and stroke [1].

Non cardiovascular co-morbidities include anaemia and iron deficiency, arthritis, cachexia, cancer, depression, diabetes, electrolyte disorders such as hypokalaemia, hyponatraemia and hypochloraemia, erectile disfunction, frailty, gout, hyperlipidaemia, infection, kidney disfunction, lipid-modifying therapy, lung disease, obesity, sarcopenia, sleep-disordered breathing, and thyroid disorders [1].

It is important to consider co-morbidities and underlying aetiology as these need to be considered for overall disease management [11]. Co-morbidities are also considered in selection criteria of studies such as clinical trials and need to be considered for generalisability of studies [12].

1.2.9 Treatment for Chronic Heart Failure

Treatment and management of chronic heart failure will depend on the underlying aetiology and LVEF phenotypes. Evidence suggests that only patients with HFrEF have been shown to respond to treatment lowering mortality and reducing morbidity [1].

1.2.9.1 Treatment of patients with HFrEF

Patients with HFrEF may benefit from pharmacotherapy which is recommended by the ESC prior to invasive treatment such as device therapy. The goal of treatment for patients with HFrEF is threefold: reducing mortality, preventing recurring hospitalisations due to heart failure (HF), and improvement in patient status i.e., clinical status, function capacity and quality of life (QOL) [1].

Guidance on treatment for HFrEF is dependent on aetiology. The ESC provides guidelines for treatment of HFrEF starting with pharmacotherapy including angiotensin-converting

enzyme inhibitor (ACE-I) / angiotensin receptor-neprilysin inhibitor (ARNI), beta blocker, mineralocorticoid receptor antagonist (MRA), dapagliflozin/empagliflozin and loop diuretics. Following pharmacotherapy, the use of implantable devices such as Cardioverter-Defibrillators (ICD) or pacemakers. After consideration of device therapy other therapies can be considered, but the level of evidence for these therapies and their efficacy is lacking or conflicting [1].

1.2.10 Natriuretic Peptides in Heart Failure

As previously mentioned, natriuretic peptides are used in the diagnosis of heart failure, either BNP (Brain Natriuretic Peptide) or NT-ProBNP (N-terminal pro-B type natriuretic peptide). Both BNP and NT-ProBNP are circulating bio-makers. Originally identified in porcupine brain [13], Both BNP and NT-ProBNP are now acknowledged to be circulating markers, produced mainly in the cardiac ventricles as a result of pressure overload and ventricular stretch, with small amounts being produced by the atrium [14], [15].

Plasma levels of B-Type Natriuretic peptides in heart failure patients increase in levels proportionate to the level of systolic and diastolic myocardial dysfunction [15]–[17]. NT-ProBNP has a longer half-life than BNP; as a result, plasma levels of NT-ProBNP are generally higher and may provide better diagnostic indicators [16], [18]. Levels of B-Type Natriuretic peptides are known to rise with age [19], [20]. Evidence suggests that lower levels of both NT-ProBNP and BNP are found in patients with obesity; however, the mechanisms of these findings are unknown [21]. Latest ESC guidelines for heart failure suggest an NT-ProBNP level $\geq$ 125 pg/mL or a BNP level $\geq$ 35 pg/mL are suggestive of heart failure [1].

BNP and NT-ProBNP are generally considered interchangeable, which was recently illustrated by Rorth et al. using data from an RCT [22]. In the same study the authors observed differences in levels of BNP and NT-ProBNP in patients with atrial fibrillation and also showed that levels increased with both age and declining renal function [22].

1.2.11 Renal Function in Heart Failure

There is a known and well documented relationship between heart failure and renal decline [23]–[26]. CKD is also a known risk factor for heart failure which is also directly linked to renal decline [27], [28]. These factors make renal function in heart failure an important consideration in prognosis, as renal decline can lead to adverse events such as hospitalisation and death [25], [29] .

Renal function is measured using Glomerular Filtration Rate (GFR) which is considered "the volume of water filtered out of the plasma through glomerular capillary walls into Bowman's capsules per unit of time" [30] as defined as in Equation 1.

Equation 1 Equation for Measuring Glomerular Filtration Rate

$$GFR\ (ml/min)\ = \frac{Urine\ concentration\ \times Urine\ Flow}{Plasma\ Concentration}$$

[31]

The measurement of GFR in practice is considered inconvenient, costly, and technically difficult as it requires injection of a marker such as inulin, and then measuring how effectively the marker is filtered by the kidneys [32]. Therefore, the estimated Glomerular Filtration Rate (eGFR) is preferred. Which can be estimated using several validated equations; one of the most popular was developed by Levy et al. Noted in Equation 2, this formula uses serum creatinine, age, sex and if race is black to estimate the GFR in mL/min/1.73 m^2 [33], [34].

Equation 2 Formula for eGFR by Levy et al.

$$eGFR = 186 \times Serum\ creatinine\ (mg/dL)^{-1.154} \times Age^{-0.203} \times 0.742\ (if\ Female) \times 1.212\ (if\ Black)$$

[33]

Serum creatinine, age, sex, and race have all been shown to be associated with GFR, with age and creatinine having a negative effect on GFR. Both black race and the male sex increase GFR when compared against the female sex and nonblack races [34].

The 'normal range' of eGFR is that greater than 60 mL/min/1.73m^2; however, the 'normal' range of creatinine is 0.8 - 1.4 mg/dL for adult men and 0.6 – 1.1 mg/dL for adult woman and due to the effects of age, race and sex a person could have a 'normal' range of eGFR but an 'abnormal' range of creatinine and vice versa [34].

1.2.12 Prognostic Models in Heart Failure

In general, a prognostic model can be defined as a model fit with multiple covariates to make predictions about a specific outcome of interest, dependent on a specified state of health for a stated period of time. Prognostic models can be applied to continuous, dichotomous and nominal / ordinal outcome measures. Models can also use continuous outcomes to predict the expected outcome of interest for an individual. Models for dichotomous outcomes predict the risk of the outcome of interest for an individual by a specified time; these models are also known as risk prediction models. The so called 'holy grail' of prognostic models are models that can predict subject specific outcomes, providing personalised predictions [35].

Risk prediction prognostic models can be assessed using metrics measuring fit , discrimination and calibration. Goodness of fit references how well the model fits the data,

providing an indication of overall performance of the model. This is typically assessed using an information criterion such as Akaike Information Criterion (AIC) for frequentist models and Deviance Information Criterion (DIC) for Bayesian models, the former based on prediction error and the latter deviance at the posterior mean [36]. Discrimination is defined as how well a model's predictions discriminate between subjects who had the outcome of interest and those who did not. Discrimination is typically assessed using the area under the curve of the Receiver Operating Characteristic (ROC) which plots sensitivity against specificity. Finally, calibration is used to examine the predicted and observed risks and should be examined across the range of predicted risks. Calibration can be assessed visually using a calibration curve which plots predicted risk vs observed risk [37], [38].

Risk prediction models in heart failure have previously been reviewed in a systematic review by Rahimi et al. identifying 64 models from 48 studies predicting death, hospitalisation for heart failure and a composite of death and hospitalisation for heart failure. Reporting a variety of data sources including patient data from medical records, prospectively collected data, and data from clinical trials. Using data from 1985 up until 2010 [36].

Di Tanna et al. in a more recent review developed upon the review by Rahimi et al. identifying 58 models from 40 studies predicting all-cause mortality, hospitalisation for heart failure, cardiovascular death, and composite endpoints. Data from these studies were noted as originating from longitudinal, experimental, and retrospective studies. The authors also used Prediction model Risk Of Bias ASsessment Tool (PROBAST) to assess the Risk of Bias (ROB) identifying only seven studies meeting the criteria to be considered suitable with minimal ROB, suggesting reporting around analysis, calibration, discrimination, and missing data needed improvement [37].

Di Tanna et al. based their categorisation of the level of discrimination on Hosmer et al., However, shown in Table 3 and Table 4 the latest version of this text differs [38]. It is important to consider these differences when describing discrimination. The Di Tanna et al. definition only encompasses a range from <0.60 - ≥ 0.70 with only three categories whereas the Hosmer et al. definition ranges from 0.5 to ≥ 0.9 with five categories [37], [38].

Table 3 Level of Discriminatory Ability for Risk Prediction Models Adapted from Hosmer et al.

AUC-ROC / C-Statistic	Level of Discrimination
= 0.5	No Discrimination (Flip of a Coin)
> 0.5 < 0.7	Poor Discrimination (Slightly Better than a Flip of a Coin)
≥ 0.7 < 0.8	Acceptable Discrimination
≥ 0.8 < 0.9	Excellent Discrimination
≥ 0.9	Outstanding Discrimination

[38]

Table 4 Level of Discriminatory Ability for Risk Prediction Models Adapted from Di Tanna et al.

AUC-ROC / C-Statistic	Level of Discrimination
< 0.60	Poor Discrimination
≥ 0.6 < 0.7	Moderate Discrimination
≥ 0.7	Good Discrimination

[37]

A key issue raised in these reviews was calibration and discrimination. In more recent years, more focus has been put on the use of both calibration and discrimination when assessing prognostic models [39], [40]. Calibration is as important as discrimination to ensure that the model gives accurate predictions while not over or underestimating those predictions. Model performance must therefore be assessed on both discrimination and calibration. In their

earlier review Rahimi et al. only reported the use of c-statistic, whereas Di Tanna et al. reported both c-statistic and whether calibration was assessed. The models from the Di Tanna et al. review ranged from 0.61-0.84 with regards to c-statistic. However, as previously stated only seven studies met the PROBAST criteria for low ROB.

As well as performance, it is also important to consider validation. Validation can either be internal or external with internal using techniques such as bootstrapping or cross fold validation and external using new data from another source to validate the model [41]. The method of validation may be limited by available data. The method of validation may be limited by available data. However, each type has a purpose in the development of prognostic models. Internal validation can be used to assess the optimism of the model because of overfitting. While external validation can be used to assess the generalisability of the model [38].

Further to the Di Tanna et al. review, a new prognostic model was published by Simpson et al., a study entitled 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 (PREDICT-HF) [42]. However, as it was published after both reviews it was not included in either. Being one of the latest prognostic models it warrants separate review.

PREDICT-HF used data from the Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure (PARADIGM-HF) RCT [43] and validated models with data from the Aliskiren Trial to Minimize Outcomes in Patients with Heart Failure (ATMOSPHERE) RCT [44] and the Swedish Heart Failure Registry (SwedeHF) [45]. The authors used a complete case

analysis and used a stepwise approach to perform a multivariable analysis fitting a Cox Proportionate Hazards (Cox PH) model for each outcome. These outcomes included the primary composite endpoint of the PARADIGM-HF trial (death from cardiovascular causes and first hospitalisation for heart failure), all-cause mortality, and death from cardiovascular causes.

Each prognostic model had different variables dependent on the outcomes of the stepwise analysis; These variables are shown in Appendix Table 1 along with the hazard ratios associated with these variables.

The calibration of the models was assessed at 1 and 2 years, using baseline survival, the Cox PH model was used to predict survival, and then the quintiles from the predicted survival was compared against the observed survival.

The models were validated with the ATMOSPHERE RCT, applying the ATMOSPHERE RCT data to the model and then comparing the c-statistics at one and two years against the original model. Using the same process, the all-cause mortality model was also validated using data from the SwedeHF cohort. The authors also conducted a sensitivity analysis using BNP over NT-ProBNP, once again comparing the c-statistic at one and two years.

Finally, the authors also compared the models against the commonly known risk prediction models within heart failure including MAGGIC, SEATTLE and EMPHASIS-HF. The authors also provided a web-based calculator for risk prediction using PREDICT-HF model to aid in translation from research to clinical practice [45].[42]

Appendix Table 1 Illustrates the hazard ratios from each of the Cox PH models, with demographics around race and region having the highest HRs suggesting demographics influence negative outcomes in heart failure.

The C-Statistics illustrated in Table 5 show the model fitted with the original data, using the composite endpoint performed best at both one and two years with a C-Statistic of 0.74 (0.71 – 0.76) and 0.71 (0.70 – 0.75) respectively. Whilst at the lower end of the acceptable discrimination, the validation with the ATMOSPHERE trial data and use of calibration add a certain level of confidence to the results. With an external validation C-Statistic of 0.71 (0.69 – 0.72) and 0.70 (0.68 – 0.71) at one and two years respectively, the model appears to perform only slightly worse with this external data. While the all-cause mortality model performed worse than the other two models both at one and two years with respect to C-Statistic scoring a C-Statistic of 0.71 (0.69 – 0.74) and 0.70 (0.67 – 0.72) respectively. However, the C-Statistics when validated with the SwedeHF cohort on the all-cause mortality model showed a reasonable improvement at one and two years with a C-Statistic of 0.79 (0.75 – 0.81) and 0.78 (0.75 – 0.80) respectively, moving the model into the upper end of the acceptable criteria for a risk prediction model.

Table 5 C-Statistics (95% CI) at one and two years from the Cox PH Prognostic Models from Predict-HF adapted from the PREDICT-HF Study

	Composite		CV Death		All-Cause Mortality	
	1 Year	2 Years	1 Year	2 Years	1 Year	2 Years
C-Statistic with Original Data	0.74 (0.71 - 0.76)	0.71 (0.70 - 0.75)	0.73 (0.71 -0.75)	0.71 (0.69 - 0.73)	0.71 (0.69 - 0.74)	0.70 (0.67 - 0.72)
C-Statistic with Atmosphere Data (Validation)	0.71 (0.69 - 0.72)	0.70 (0.68 - 0.71)	0.71 (0.69 - 0.74)	0.70 (0.69-072)	0.71 (0.69 - 0.74)	0.70 (0.68 - 0.72)
C-Statistic with SwedeHF Data (Validation)					0.79 (0.75 – 0.81)	0.78 (0.75 – 0.80)

[42]

Calibration of the model suggests it is well calibrated with the original data, with only minimal deviations between the predicted and observed quintiles, with the largest being the 5th quintile from the cardiovascular death model at two years.

The authors list limitations of their study, primarily around the use of RCT data for their data source, and possible selection bias, including the inclusion of only patients with HFrEF. However, they address this concern in part by their validation of the all-cause mortality using the SwedeHF cohort [46]–[48].

The, Di Tanna et al. review also stated that due to the time limitations restricted on their search both The Seattle Heart Failure Model [46] and the Meta-Analysis Global Group in Chronic heart failure (MAGGIC) [47] risk prediction models were excluded even though they have both been used to inform modern clinical guidelines [1], [2], [48], [49]. Whilst these models were evaluated in Rahimi et al., it is important to highlight them here, especially considering they were both used as comparison in the PREDICT-HF study.

The earlier of the studies the Seattle Heart Failure Model published in 2006 and used data from six patient cohorts. The Prospective Randomized Amlodipine Survival Evaluation (PRAISE1) study which included 1125 patients (the primary data), The Randomized Enbrel North American Strategy to Study Antagonism of Cytokines (RENAISSANCE) RCT, Valsartan Heart Failure Trial (Val-HeFT), Evaluation of Losartan in the Elderly (ELITE2) RCT, The University of Washington (UW) prospective study and the Italian Heart Failure Registry (IN-CHF).

A stepwise approach was used to fit a Cox PH model. Included variables and their corresponding hazard ratios are shown in Appendix Table 2 and Appendix Table 3 the latter estimated from meta-analysis and clinical trials.

Missing data was imputed using median values, with the exception of medication and devices where patients with missing data were assigned no device / medication. [49]
The authors used a scoring technique known as the Seattle Heart Failure Model (SHFM) Score, which multiplied the 13 coefficients from the multivariable model and then summed them. This score was then applied to the validation data sets to obtain survival estimates at 1-5 years for the individual patients from these data. Model accuracy was determined by comparison of the mean predicted survival against the recorded survival at 1-3 years.

The model's discriminatory capability was assessed using the AUC-ROC (C-Statistic) at one year for each dataset and the combined datasets as shown in Table 6. With combination of the PRAISE1, Val-HeFT, IN-CHF and UW allowing for comparison with the ADHERE and Toronto Models as shown in Table 7.

The authors also provided a "web based calculator" to aid in the clinical usage of the model [46].

The Hazard ratios from Appendix Table 2 illustrate that NYHA Class has the largest effect of the demographics on survival. Statin usage had the largest effect of the medications, suggesting a preventative effect. Finally, cholesterol has the highest effect on survival out of the laboratory measures.

The variables derived from clinical trials and meta-analysis are shown in Appendix Table 3. They suggest that all medication and devices had preventative effects on survival, except for Biventricular pacemaker which had no effect when the patient was already on the device.

The ROCs from the model and validation data sets shown in Table 6 suggest the model performed well with respect to discrimination. With ROCs ranging from 0.682 to 0.810, with the lowest discrimination falling in the higher end of the poor discrimination category and the highest discrimination falling in the lower end of the excellent discrimination category.

The comparison ROCs shown in Table 7 show that the Seattle model outperforms both ADHERE and Toronto Models when using data combined from four datasets (PRAISE1, Val-HeFT, IN-CHF & UW).

Calibration of the model was generally favourable as illustrated in Table 8 with some underestimation of mortality illustrated most at two and three years in the lower risk patients. Within the ELITE2 dataset it was observed that there was an overestimation of ~2% as highlighted at one year. The largest dataset (Val-HeFT) illustrated the model accurately predicted survival at one and two years.

The authors acknowledge the limitations of the inclusion of medication and devices from clinical trials and meta-analysis i.e., published literature and state that this may make these variables less generalisable. They also state other limitations around generalizability specifically for patients who are hospitalised, and patients with chronic comorbidities such as cancer, dementia, cirrhosis and renal failure [50].

Table 6 AUC-ROC (C-Statistic) (95% CI) for All Datasets at 1 Year Adapted from. Levy et al.

AUC-ROC	PRAISE (Derivation Dataset)	ELITE2	UW	RENAISSANCE	VAL-HeFT	IN-CHF	All 6 datasets
1 Year	0.725 (0.69 – 0.76)	0.682 (0.65 – 0.73)	0.810 (0.72 - 0.90)	0.682 (0.63 -0.73)	0.694 (0.68 – 0.72)	0.749 (0.70 – 0.80)	0.729 (0.714 – 0.744)

[46]

Table 7 AUC-ROC (C-Statistic) (95% CI) at 1 Year Comparison against ADHERE and Toronto Models Adapted from Levy et al.

AUC-ROC	Seattle (PRAISE1, Val-HeFT, IN-CHF & UW)	ADHERE	Toronto
1 Year	0.75 (0.73 – 0.77)	0.59 (0.57 – 0.61)	0.68 (0.66 – 0.70)

[46]

Table 8 Actual VS Predicted Survival (± SD) for 1, 2 and 3 Years for each Dataset Adapted from Levy et al.

	PRAISE1 (Derivation Dataset)	ELITE2	UW	RENAISSANCE	Val-HeFT	IN-CHF
1-Year Survival						
Actual	74.3±1.4%	88.5±0.6%	86.5±2.8%	83.3±1.4%	91.0±0.4%	86.7±1.2%
Predicted	73.4±0.5%	90.5±0.1%	86.5±1.0%	83.8±0.5%	90.9±0.1%	89.6±0.4%
2-Year Survival						
Actual	56.0±1.8%	80.0±1.0%	79.7±3.3%	65.4±4.6%	81.6±0.6%	
Predicted	56.7±0.6%	82.4±0.2%	76.5±1.6%	72.3±0.7%	83.3±0.2%	
3-Year Survival						
Actual			71.8±3.7%		71.7±1.3%	
Predicted			68.6±1.8%		76.8±0.2%	
SD: Standard Deviation						

[49]

The later model, the MAGGIC Model developed in 2013 [50], used data from 30 studies. Using individual patient data from 30 cohorts, including 24 observational registries and six randomised control trials. The authors fit a multivariable piecewise Poisson regression model along with an integer score for predicting a patient's risk of death within 3 years ranging from 0 (the best possible with a predicted risk of death at 3 years of 0.039) to 50 (the worse possible, with a risk of death at three years of 0.985).

The authors used multiple imputation techniques to handle missing data creating 25 imputed datasets, fitting the model on each dataset and then pooling the models based on the imputed datasets.

The final model included 13 variables, as described in Appendix Table 4. The calibration of the model was assessed using a comparison of predicted vs observed mortality at three years for 6 risk groups, grouped by integer score). The model results we stratified into two groups by ejection fraction, with the first group containing 21442 patients with an ejection fraction of less than 40, of which 8900 patients experienced death. The second group contained 17930 patients with an ejection fraction greater than or equal to 40, and of which 6951 patients experienced death.

The authors stated they believed there was no good generalisable dataset for external validation, but internal validation may be sufficient. It should however be noted that the MAGGIC risk score has since been validated on a prospective registry of patients with HFpEF by Rich et al. in 2018 [47].

The discriminatory ability of the model was assessed using plots for the distribution of the risk score for all patients and the association with the risk of death at three years with their

respective 95% CI. Along with a plot of the cumulative risk of death over 3 years in the previously mentioned 6 groups and corresponding 95% confidence intervals.
The authors provided an “easy-to-use” website to allow clinical usage of the MAGGIC Score [50].

The Rate Ratios from the MAGGIC models shown in Appendix Table 4, show that for each stratification of ejection fraction all variables are significant. With only the use of ACE-I/ARBs in the ≥ 40 % ejection fraction group of patients being considered statistically insignificant. Both groups have similar rate ratios for each of the variable with mostly marginal differences between the two. The largest rate ratio being NYHA Class IV for both groups. Within both groups the usage of beta blockers has one of the most preventative rates of mortality with it being more preventative in the > 40 % ejection fraction group.

The predicted vs observed mortality at three years within the six groups shows that the model is well calibrated at three years with only marginal differences between the observed and predicted mortality, with groups 1-3 overestimating the mortality rate and groups 4 and 5 underestimating the mortality rate and group 6 only underestimating the mortality rate by an almost unobservable amount.

The authors state the model has a powerful discriminatory ability of the model to predict a patient’s risk of mortality at 3 years and an excellent good-ness of fit.

One of the major limitations described by the authors was that of the missing data, because of the large number of included studies and subsequently patients, this was inevitable.

While the MAGGIC model did not include natriuretic peptides; The Rich et al. study illustrated a statistically significant improvement in the prognostic capability of the MAGGIC model when BNP was added.

All three of the highlighted models assessed calibration, however there was not a consistency between the models of how this was assessed, suggesting the need for better consistency with respect to calibration.

Both PREDICT-HF and Seattle used external validation whilst the authors of MAGGIC suggested external validation would be harder due to the generalisability of the model, however the MAGGIC model has since been validated in a HFpEF cohort [47].

PREDICT-HF was the only study to originally use natriuretic peptides in the form of NT-ProBNP.

Finally, all three models provided a web interface to aid in the clinical application of the models; this highlights the need for translation of prognostic models in heart failure to clinical practice.

1.3 Joint Modelling Longitudinal Repeat Measures and Time to Event Outcomes

Briefly put, the term joint model is used to describe a model which combines two or more statistical models for a particular purpose such as efficiency, reduction of bias or to overcome a limitation of a singular model [51]–[53].

Joint models of longitudinal repeated measurements and time to event outcomes which shall hereby be referred to as joint models have been used in clinical research as early as the 1990s with initial studies investigating the relationship between CD4 cell count and survival within human immunodeficiency virus (HIV) / acquired immunodeficiency syndrome (AIDS) studies [54], [55]. Models of this kind, jointly model a longitudinal process such as the profile of a biomarker and a time-to-event process, such as time to death. This joint modelling is often facilitated through such methods as shared random effects, that 'link' the two models [51], [56].

To introduce joint modelling, it is important to first introduce the individual components of a joint model, namely the time-to-event and repeated measurements components.

1.3.1 Time-to-event (Survival) analysis

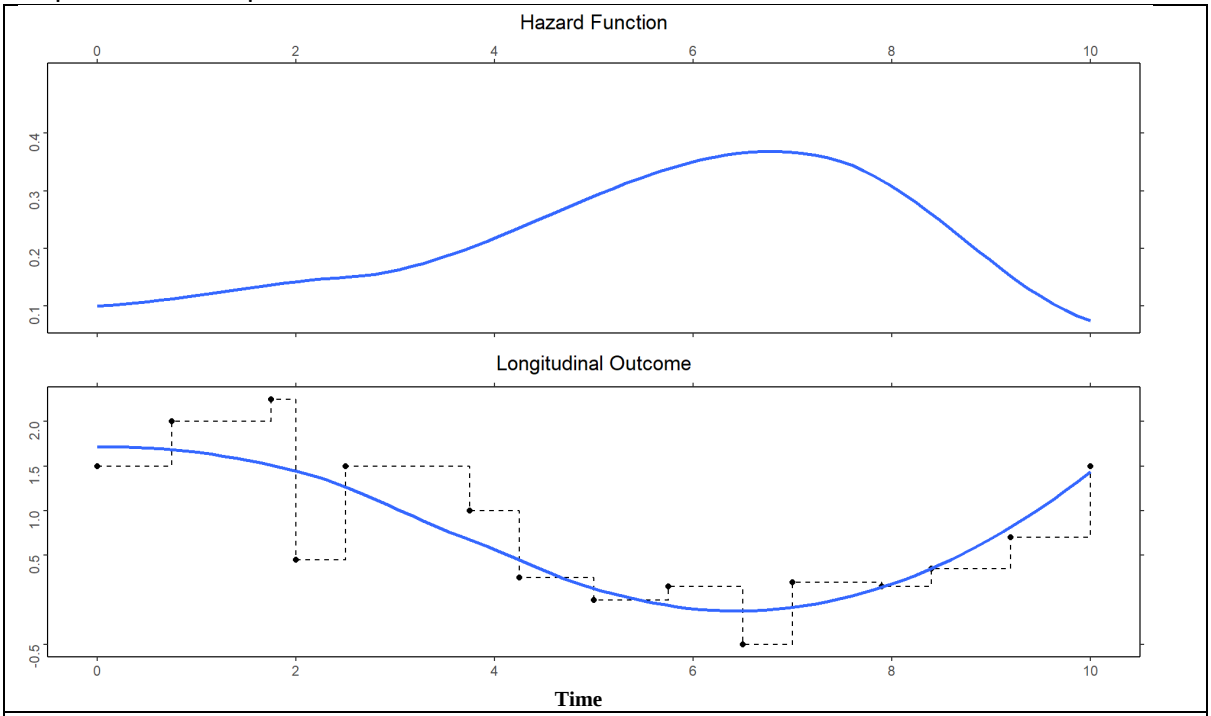
Time-to-event or survival analysis is the analysis of data from a formally specified time point up until an event of interest or endpoint such as death. A common use for this kind of analysis is in clinical trials where data is stratified by treatment arm to detect a treatment effect on mortality or another endpoint [57].

Survival data is typically subject to censoring. Censoring occurs when the patient does not experience the event of interest, is lost to follow-up and their status is unknown or the patient is excluded from the study for other reasons [57].

1.3.2 Repeated Measures in Survival analysis

Repeated measures such as biomarkers are often routinely collected during clinical trials and can add prognostic value to survival analysis [58]–[60]. Limiting these measurements to a single time point such as the Last Observation Carried forward (LOCF) i.e., landmarks, can lead to loss of information and bias [61]. However, it is important to consider the nature of the repeated measures. Models such as the extended Cox PH (i.e., time-varying Cox PH) model allow for repeated measures in the form of time-varying covariates. However, the extended Cox PH model is only suitable for exogenous covariates and not endogenous covariates due to the underlying methodology and the step function approximation it relies on. Illustrated in Figure 2, the dashed line in the bottom panel represents the previously mentioned step function approximation which assumes that a time dependent covariate remains constant between visits and is measured without error and only changes at follow-up visits [52].

Figure 2 Graphical Representation of an Extended Cox PH and Joint Model Showing How the Hazard Function Evolves Over Time in Comparison with the Longitudinal Outcome - Adapted from Rizopoulos D.



The top panel illustrates the hazard function (blue line), with the bottom panel illustrating the longitudinal trajectory, with the black dots indicating an observed measurement, the dashed line indicating the step function of a extended Cox PH model and the blue solid line representing the joint model approximation as defined by $m_i(t)$ (true unobserved outcome of the longitudinal process).

[52]

1.3.3 Exogenous and Endogenous Covariates

When including time-varying covariates in models, it is important to know whether the covariate is exogenous or endogenous to inform the modelling approach. An exogenous covariate is a variable that is "outside the model" meaning while it may influence the endogenous variables, it itself is not influenced by endogenous variables or the model. Endogenous variables are therefore variables which are influenced by the model [62]; for example, in a model where there are environmental factors such as the weather and its effect on mental health, the weather would be a time-varying exogenous covariate. Which while it may affect the model, the model will not affect the weather. A common example of an endogenous covariate is a biomarker i.e., biological marker; as for a biomarker to be collected, the patient needs to be alive and thus in a survival model the outcome of the model affects the biomarker. Special consideration needs to be taken when including

endogenous covariates in a model to account for how the model effects the covariate [52], [62].

1.3.4 Biomarkers as Time-Varying Covariates

A biomarker is not only an endogenous time-varying covariate but also subject to measurement error and biological variances. These properties make them hard to model as the model must be able to capture them effectively to minimise bias and improve model fit [52], [59].

1.3.5 Formation of Joint Models

A joint model has two elemental components, a longitudinal component and a survival component; these components are often referred to as sub models. The longitudinal sub model models the longitudinal response (typically a biomarker). This model typically takes the form of a linear mixed effects model (LME) which is a synonym for a Linear Mixed Model (LMM) and contains both fixed and random effects [63]–[65]. The general form of these models is shown in Equation 3.

Equation 3 General form of an LME

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i, \\ b_i \sim N\{0, D), \\ \varepsilon_i(t) \sim N(0, \sigma^2 I_{ni}), \end{cases}$$

[52], [63], [64]

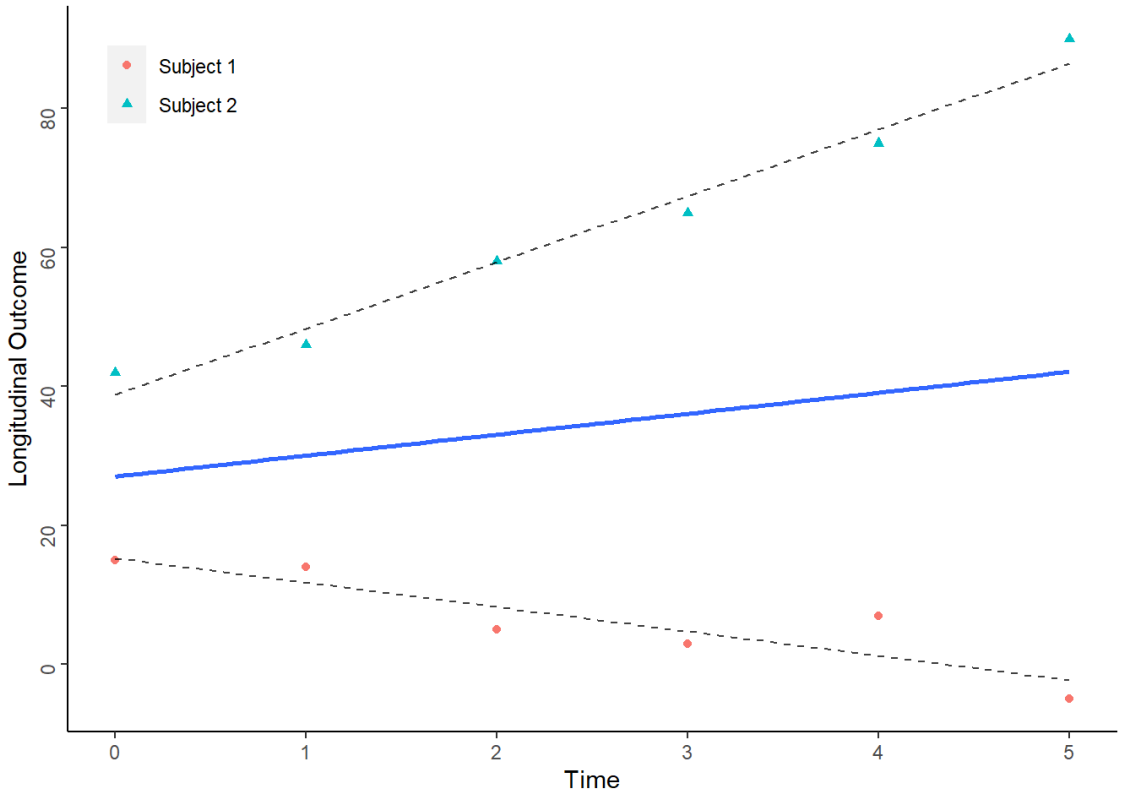
With y_i denoting the longitudinal response for subject i. Both X_i and Z_i are matrices of known design for the fixed effects (β) and random effects (b) regression coefficients respectively. I_{ni} represents the n_i-dimensional identity matrix. The model assumes that the random effects follow a normal distribution and have a mean of zero with a variance-covariance matrix D. The model also assumes the random effects are independent of the error terms ε_i (which are assumed to follow a normal distribution with a mean of 0, with a

variance $\sigma^2 I_{ni}$). The fixed effects β can be interpreted in a manner akin to linear regression, with the assumption of p covariates within the matrix of X, the coefficient β_j where $j = 1, \ldots, p$ represents the change in the average longitudinal response y_i when all other predictors are held constant. The interpretation of the random effects b_i are that of how ith subject deviates from the population in terms of a subset of their regression parameters.

Common random effects include random intercepts and random slopes. Random intercepts assume that for each subject with a subset of regression parameters their intercept deviates from the population. Random slopes assume for each subject with a subset of regression parameters their slopes i.e., the effect of time deviates from the population. In joint modelling, typically both random intercepts and random slopes will be used together [52], [66].

Figure 3 illustrates a graphical representation of an LME and highlights the use of both random intercepts and slopes, with both subjects deviating from the marginal longitudinal trajectory at both intercept and slope.

Figure 3 Graphical Representation of an LME Showing the Subject Specific Longitudinal Trajectories of two Subjects and the Marginal Longitudinal Trajectory - Adapted from Rizopoulos D



The points represent longitudinal measurements (hypothetical). The dashed lines represents the fitted longitudinal trajectories (subject-specific). The solid blue line represents the marginal or population average longitudinal trajectories.

[52]

The survival component typically takes the form of a relative risk model also known as the proportional hazards model. These proportional hazards models assume a multiplicative effect of the covariates on the hazard of an event [52]. They can be formulated as shown in Equation 4.

Equation 4 Formula for a Basic Relative Risk (Proportional Hazards) Model

$$h_i(t \mid w_i) = \lim_{dt \to 0} \Pr(t \leq T^* < t + dt \mid T^* \geq t, w_i) / dt$$

$$= h_0(t)\exp\{\gamma^\top w_i\},$$

[52]

With the vector $w_i^\top = (w_{i1}, \ldots, w_{ip})$ representing covariates assumed to be associated with the hazard of each subject. Whilst γ represents the corresponding vector of regression coefficients. The function expressed as $h_0(t)$ is known as the baseline hazard or risk function for a subject having $\gamma^\top w_i = 0$. $\gamma^\top w_i$ may or may not include an intercept γ_0 dependent on whether the baseline risk function has a constant scale [52].

Equation 5 Formula for Relative Risk Model Written on a Log Scale

$$\log h^i(t \mid w_i) = \log h_0(t) + \gamma_1 w_{i1} + \gamma_2 w_{i2} + \cdots + \gamma_p w_{ip},$$

[52]

When written on a log scale as shown in Equation 5, It can be observed that the regression coefficient for γ_j, for the predictor w_j, represents the log hazard change at any fixed point in time t if w_j is increased by one unit when all other predictors remain constant. Equally, $\exp(\gamma_j)$ represents the hazard ratio for a one unit change in w_{ij} at any time point t [52].

Up until this point these models are independent of each other. To define a joint model, both the longitudinal and survival processes need to be linked or joined. For this purpose, a joint distribution is used, based on the principle that the time-independent random effects noted as b_i encompasses both the survival and longitudinal processes. This implies that the random effects not only account for the correlation between the repeated measures, providing conditional independence, but also the association between the event and longitudinal outcomes. There are also additional assumptions surrounding the decision of participants to withdraw from the study or present for a longitudinal measurement. The assumption being that this decision depends on the observed previous history (both baseline covariates and longitudinal measurements) but does not depend on the underlying latent / unobserved characteristics associated with prognosis [52].

To define a joint model, the two independent models are redefined as sub models of the joint model.

For the survival sub model, a new term is introduced to measure the association of the level of the longitudinal marker and the risk of an event. This new term needs to account for the special features of the longitudinal marker as described earlier. Therefore, this new term represents the unobserved and true value of the longitudinal outcome at time point t and is represented as the term $m_i(t)$. It is important to note that this new term is different than that of the observed value at time point t for the ith subject known as $\gamma_i(t)$, as the observed observation is deemed to be contaminated with measurement error at the time point t.

Equation 6 Relative Risk Model Reformulated to Include $\boldsymbol{M_i(t)}$

$$h_i(t \mid M_i(t), w_i) = \lim_{dt \to 0} \Pr(t \leq T_i^* < t + dt \mid T_i^* \geq t, M_i(t), w_i) \,/\, dt$$

$$= h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\}, \quad t < 0,$$

[52]

To include this new term and therefore the association of $m_i(t)$ and the risk of an event, the relative model as defined in Equation 4 can be redefined as shown in Equation 6, quantifying the association of the term $m_i(t)$ and the risk of an event. In this model $h_0(.)$ represents the baseline hazard function, and the term $M_i(t) = \{m_i(s), 0 \leq s < t\}$ represents the true and unobserved longitudinal process up until time point t. w_i in this model is specified as a vector of covariates at baseline with corresponding vector of regression coefficients known as γ. In this model rather than the use of the observed event time for the ith subject represented as T_i, the true event time for the ith subject represented by $T_i^*$is used. Like γ, the α parameter provides the quantification of the effect of the underlying longitudinal outcome on the risk of an event. The parameter γ when exponentiated and expressed as exp (γ_j) represents the hazard ratio for a unit change in w_{ij} at any time point t. The α parameter on the other hand has a slightly different representation, in that when exponentiated and expressed as exp (α) represents the hazard ratio relative to a unit increase in $m_i(t)$ at the same time point *t*. [52]

In joint modelling, the specification of the baseline hazard function is of importance and requires appropriate assumptions of the distribution. Depending on the method used to fit the joint model, various distribution options are available. The baseline hazard can be specified under the assumption of a commonly used parametric distribution such as the Weibull distribution or through use of a parametric yet flexible specifications such as a piecewise-constant or regression splines such as B-splines [52].

The longitudinal process represented in the survival sub model as $m_i(t)$, represents the true value of the covariate at time point t for the underlying longitudinal variable. To capture the longitudinal process in its entirety, accounting for the measurement error and the intermittency of the collection of the longitudinal marker at set time points known as $t_{ij}, m_i(t)$ needs to be appropriately estimated and the complete longitudinal history $M_i(t)$ constructed for each subject. To accomplish this an LME can be used assuming a normally distributed outcome.

Equation 7 Linear Mixed Effects Model

$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_i(t), \\ m_i(t) = x_i^{\top}(t)\beta + z_i^{\top}(t)b_i, \\ b_i \sim N(0, D), \quad \varepsilon_i(t) \sim N(0, \sigma^2), \end{cases}$$

[52]

Equation 7 shows the formulation of the linear mixed effect model, in this model the design vectors $x_i(t)$ and $z_i(t)$ for fixed (β) and random effects (b_i) respectively along with the error terms $\varepsilon_i(t)$ are all time dependent. The error terms are assumed to be normally distributed with a mean zero and variance σ^2 and both mutually independent of each other and independent of the random effects. The LME accounts for measurement error by assuming that $y_i(t)$ the observed level of the longitudinal outcome is equal to $m_i(t)$ the true level of the longitudinal outcome and the error term. Through use of the time structure within the definitions of the design vectors for the fixed and random effects

$x_i(t)$ and $z_i(t)$ respectively, along with the subject-specific random effects, the construction of $M_i(t)$ for each subject is possible. Figure 2 illustrates the concept of how the true level of the marker (solid blue line of the bottom panel) is associated with the hazard of an event (the hazard function in the top panel). As previously stated, it also demonstrates the step function of an extended Cox PH Model (dashed line) and how this might be an unrealistic approximation of the longitudinal process [52].

As the longitudinal process is captured within $M_i(t)$, it is important that $M_i(t)$ provides an accurate estimation of the longitudinal process. It is therefore important to consider more elaborate structures of time in both design vectors for the fixed $x_i(t)$ and random $z_i(t)$ effects which could mean the addition of interaction terms or the consideration of non-linearity by the inclusion of natural cubic splines [52].

Equation 8 Conditional Distribution of the Time-To-Dropout

$$\boldsymbol{p(T_i^o \mid y_i^o, y_i^m; \theta)} \quad = \quad \int \boldsymbol{p(T_i^* \mid b_i; \theta)\, p(b_i \mid y_i^o, y_i^m; \theta)\, db_i}$$

[52]

An important note regarding the use of an LME for the longitudinal outcome, is that it makes assumptions regarding the vector for the complete longitudinal response. One such assumption is that the responses would have happened even after the event or censoring. While this may be seen as problematic, when both the observed y_i^o and missing y_i^m parts of longitudinal response vector are taken into consideration, the dropout mechanism can be derived and as such is defined as the conditional distribution of the time-to-dropout given both the observed and missing parts of the longitudinal response vector. Simplified, this is shown in Equation 8, in which it can be observed that the time-to-drop out is dependent on the missing longitudinal response vector by means of the posterior distribution of the random effects. Essentially, this means that a joint model corresponds to a Missing Not at Random (MNAR) missing data mechanism, a key component to this being the random effects.

Equation 9 Formulation of Joint Model

$$\begin{cases} \boldsymbol{y_i(t)} & = & \boldsymbol{X_i^{\top}(t)\beta + z_i^{\top}(t)b_i + \varepsilon_i(t)} \\ \boldsymbol{h_i(t)} & = & \boldsymbol{h_0(t)\, exp[\gamma^{\top} + w_i + \alpha\{x_i^{\top}(t)\beta + z_i^{\top}(t)b_i\}]}, \end{cases} \quad [52]$$

With the definitions of both sub models, the joint model is therefore defined as shown in Equation 9 where both models share the same random effects.

The estimation technique used in joint models depends on the approach chosen. The two common approaches implemented in popular statistics software such as R are the frequentist approach using a maximum likelihood method, along with a Bayesian approach using Markov Chain Monte Carlo (MCMC) algorithms, where maximum likelihood models are predicated on the maximisation of the log likelihood correspondent to both outcomes [67]. The latter samples from both the random effects and the posterior conditional distributions to fit the joint models [68]. Both approaches use an iterative approach and are therefore dependent on the model complexity, and can be both time and computationally expensive, especially when compared to more traditional models such as (extended) Cox PH models.

1.3.6 Alpha Parameterisation

As described previously, in the basic joint model, the alpha parameter in the survival outcome is represented by $m_i(t)$ which corresponds to the true level of the longitudinal outcome at time point t, however joint models allow for multiple representations of the alpha parameter(s) [52].

1.3.7 Slope Parameterisation

An additional parameterisation is the slope parameterisation represented as $m_i'(t)$, which corresponds to the rate of change of the longitudinal outcome at time point t. The $m_i'(t)$ parameter is estimated using the derivative of the fixed and random effects of the longitudinal outcome with respect to time. The $m_i'(t)$ take the form as shown in Equation 10. [55], [70][52], [67]

Equation 10 Formula of the slope parameterisation $\boldsymbol{m_i'(t)}$

$$m_i'(t) = \frac{d}{dt} m_i(t) = \frac{d}{dt}\{x_i^\top(t)\beta + z_i^\top b_i\}.$$

[52]

Typically, both the slope and value parameters are included together; the basic formula for the survival component including both parameters are shown in Equation 11. Figure 4 provides a graphical representation of both parameterisations when included in a joint model.

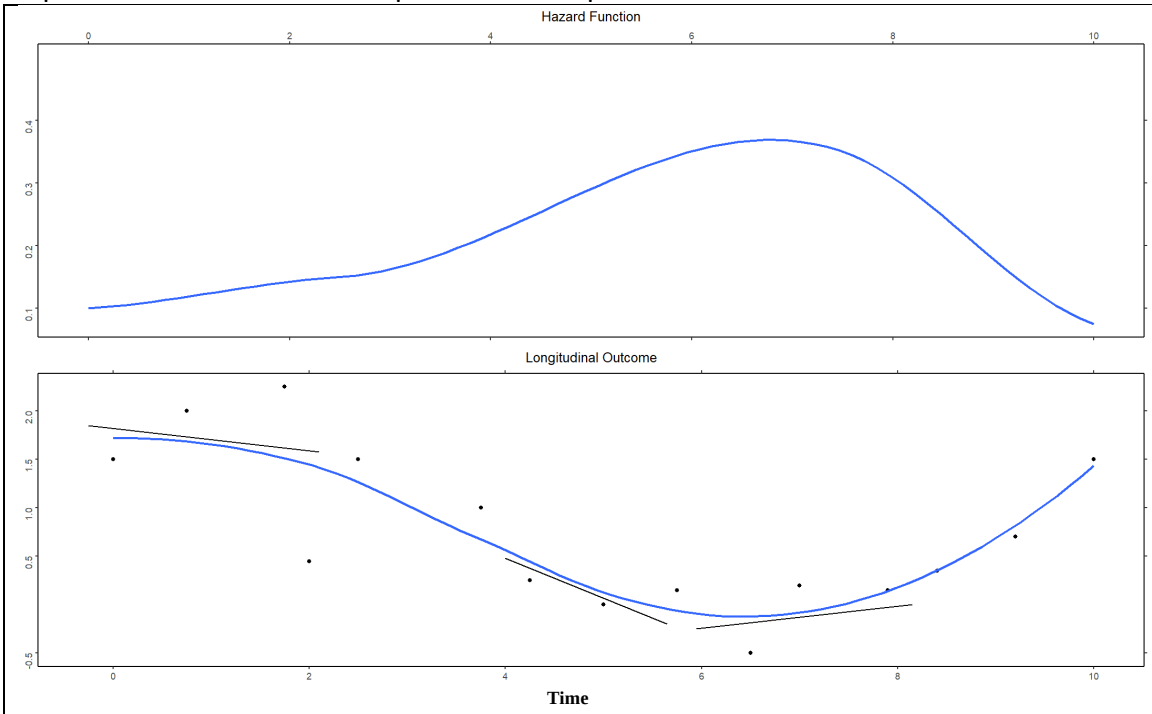
Equation 11 Formula for Survival Component of a Joint Model Including Both Value and Slope Parameterisations

$$h_i(t) = h_0(t)\exp\{\gamma^\top w_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\},$$

[52]

The $m_i'(t)$ parameter when exponentiated corresponds to the hazard ratio of a unit increase in the slope of the longitudinal outcome at any time point. When both $m_i(t)$ and $m_i'(t)$ are both included in the model, $m_i'(t)$ when exponentiated corresponds to the hazard of a one unit increase in the slope of the longitudinal outcome at time point t for patients having the same level of the true longitudinal value at the same time point. The other components such as $\gamma^\top w_i$ have the same interpretation as in Equation 6 [52] .

Figure 4 Graphical Representation of a Joint Model Showing How the Hazard Function Evolves Over Time in Comparison with the Longitudinal Outcome with the Value and Slope Parameterisations - Adapted from Rizopoulos D.



The top panel illustrates the hazard function (blue line), with the bottom panel illustrating the longitudinal trajectory, with the black dots indicating an observed measurement, the short black lines representing the slope of the longitudinal trajectory and the blue solid line representing the joint model approximation as defined by $m_i(t)$ (true unobserved outcome of the longitudinal process). In this figure the hazard directly corresponds to the longitudinal process, with both the value and slope effecting the hazard function e.g. with a smaller value, the hazard increased.

[52]

1.3.8 Cumulative Effects (area) Parameterisation

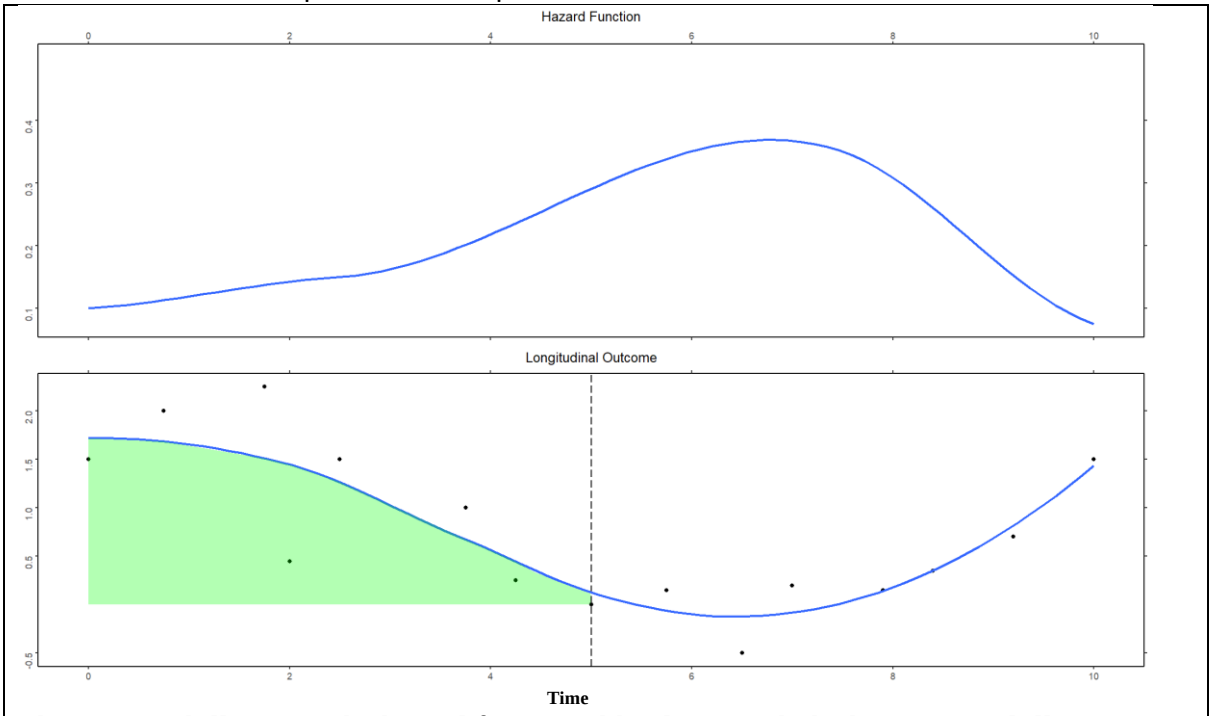
The cumulative effect parameterisation which is also known as the area parameterisation corresponds to the whole area under the longitudinal trajectory ($M_i(t)$) up until time point t whist also accounting for the length of time of the observation period [69]. This parameterisation takes the form of a function as shown in Equation 12. Alternatives of this "area" parameterisation do not allow for the accounting of length of time of the observation period [67], [68].

Equation 12 Formula for the Survival Outcome Including the Area Parameterisation.

$$h_i(t) = \exp\left\{\gamma^\top w_i + \alpha \frac{\int_0^t m_i(s)\, ds}{t}\right\}$$

[69]

Figure 5 Graphical Representation of a Joint Model Showing How the Hazard Function Evolves Over Time in Comparison with the Longitudinal Outcome with the Area Parameterisation - Adapted from Rizopoulos D.



The top panel illustrates the hazard function (blue line), with the bottom panel illustrating the longitudinal trajectory and the black dots indicating an observed measurement, the green area representing the cumulative effects parameterisation (area) and the dashed line represents the time point of interest and the blue solid line representing the joint model approximation as defined by $m_i(t)$ (true unobserved outcome of the longitudinal process). In this representation, the green area effects the hazard up until the time point of interest.

Figure 5 shows a graphical representation of a joint model and the area parameter up until time point t, where the area is based on the trajectory of the true unobserved longitudinal outcome. The area parameter when exponentiated can be interpreted as the hazard for an event per one unit increase in the area under the profile of the longitudinal outcome [52], [69].

1.4 Research Aim and Specific Research Questions

The aim of this thesis is to explore the use of joint modelling and how it can be applied to prognostic modelling within heart failure, in order to answer the research questions:

1) ‘Can joint modelling enhance the methodological toolkit and have utility in the development of prognostic models within heart failure?’.

2) ‘Can prognostic models fitted with joint modelling outperform current standard prognostic models within heart failure?’

Chapter 2 Joint Modelling of longitudinal processes and time-to-event outcomes in heart failure: systematic review and exemplar examining the relationship between serum digoxin levels and mortality.

2.1 Foreword

This section provides an introduction to how joint modelling is currently being applied to studies of heart failure. This builds a foundation for the research aims of this thesis.

This introduction is in the format of a systematic review but is also paired with and exemplar to provide the reader with the basic principles of how a joint model can be fit using frequentist method using R statistical software.

While this thesis focuses primarily on chronic heart failure with reduced ejection fraction, no such limitations were placed on the search criteria in this systematic review. This decision was made to provide a better overall picture on how joint modelling is being applied within heart failure. In order to not only evaluate more of the evidence base but to use that evidence base to aid in the formulation of the research aims and process.

As stated previously the exemplar helps to provide the reader with the basic principles of how to fit a joint model. While this was achieved using a frequentist method the application is transferable to other methods such as Bayesian joint models.

The exemplar uses commonly known and available data from a historic RCT to illustrate how this relatively new technique (joint modelling) can be applied to data from older clinical trials.

2.2 Title, authorship, and publication details

Field, R.J., Adamson, C., Jhund, P. *et al.* Joint modelling of longitudinal processes and time-to-event outcomes in heart failure: systematic review and exemplar examining the relationship between serum digoxin levels and mortality. *BMC Med Res Methodol* **23**, 94 (2023). https://doi.org/10.1186/s12874-023-01918-4

2.3 Abstract

2.3.1 Background

Joint modelling combines two or more statistical models to reduce bias and increase efficiency. As the use of joint modelling increases it is important to understand how and why it is being applied to heart failure research.

2.3.2 Methods

A systematic review of major medical databases of studies which used joint modelling within heart failure alongside an exemplar; joint modelling repeat measurements of serum digoxin with all-cause mortality using data from the Effect of Digoxin on Mortality and Morbidity in Patients with Heart Failure (DIG) trial.

2.3.3 Results

Overall, 28 studies were included that used joint models, 25 (89%) used data from cohort studies, the remaining 3 (11%) using data from clinical trials. 21 (75%) of the studies used biomarkers and the remaining studies used imaging parameters and functional parameters. The exemplar findings show that a per unit increase of square root serum digoxin is associated with the hazard of all-cause mortality increasing by 1.77 (1.34–2.33) times when adjusting for clinically relevant covariates.

2.3.4 Conclusion

Recently, there has been a rise in publications of joint modelling being applied to heart failure. Where appropriate, joint models should be preferred over traditional models

allowing for the inclusion of repeated measures while accounting for the biological nature of biomarkers and measurement error.

2.3.5 Keywords

Joint Modelling, Heart Failure, Shared Parameter Models, Systematic Review, Digoxin, Mortality

2.4 Background

Heart failure is a condition where there are well documented inter-relationships between numerous physical, biochemical and imaging characteristics and outcomes. Many studies tend to examine these associations with outcomes using data from one point in time such as randomization in a trial or the start of a cohort study. This fails to account for changes in characteristics over time. Just as baseline values may be associated with outcomes, changes in variables are also associated with changes in outcomes e.g., falling levels of natriuretic peptides are associated with lower mortality. However, analysing the association between changes in variables and outcomes is often performed with traditional time to event models using change values and starting follow up for outcomes once change has occurred. More recently joint modelling, combining two or more statistical models to increase efficiency and reduce bias, has gained favour in the literature as a method of dealing with this issue. The most common type of joint modelling within medicine is the joint modelling of repeat measure longitudinal data (e.g., repeated measures of biomarkers over time) and time-to-event data (i.e., survival data) which are linked through an association structure via shared random effects [51], [55], [56], [70]. This seeks to improve efficiency and reduce bias in respect of treatment effect, censoring and mortality when compared against traditional models. Joint models (JMs) of this type are formed of two sub models: a longitudinal model such as a linear mixed effect (LME) model (which allows the modelling of

longitudinal changes in biomarkers or other characteristics like blood pressure) and a survival model such as a Cox Proportional Hazards (Cox PH) model to model the outcome e.g., mortality. The LME model allows for both fixed and random effects accounting for non-independence of repeated measures from the same patient, whilst also allowing for unevenly spaced measurement occasions, biological variances, and measurement error [52]. The survival model allows for covariates and typically includes an association parameter representing the association between the longitudinal and survival process [51], [52], [56], [63], [71]. JMs which use data from randomized controlled trials (RCTs) can also model the overall treatment effect as well as the treatment effect on both the longitudinal and survival models [51] i.e., the effect on the characteristic and the effect on the outcome. This is analogous to other JMs which are used in the cardiovascular literature, for example in recurrent events analyses where models that examine the effect of a treatment on recurrent hospitalisations while also estimating a treatment effect for a terminal event such as death [72].

Given the increasing use of JMs, the aim of this paper is to review the application of joint modelling in heart failure and to provide guidance on how to assess and interpret results of joint modelling. To achieve this, we conduct a systematic review to identify and critically review current applications of joint modelling within the heart failure population and then present a critical summary of how joint modelling can be applied to heart failure data sets with use of an illustrative example. We examine the association between changes in serum digoxin levels and mortality in the Effect of Digoxin on Mortality and Morbidity in Patients with Heart Failure (DIG) trial as prior studies have tried to examine the association between digoxin levels and outcomes and suggested that higher levels at one month following randomization may be associated with higher mortality [73] .

2.5 Methods

2.5.1 Systematic Review: Joint Modelling Applications within Heart Failure

Our systematic review was conducted following the Preferred Reporting Items for Systematic Reviews (PRISMA) framework [74] and the protocol is registered with PROSPERO, registration number: CRD42020210056. The aim of the review was to identify journal articles which employed joint modelling on an adult heart failure population to review how joint modelling was being applied to heart failure.

2.5.1.1 Searches

Our search strategy is provided in the Figure 11. Medline, Embase, Scopus and Google Scholar were searched, with the last search being conducted on 10th December 2021.

2.5.1.2 Screening

Articles were screened by two reviewers and full text was accessed for relevant articles. To capture all available articles no date limit was set and only English language articles were included. Only full text journal articles where joint modelling was applied to an adult heart failure population were considered for inclusion. Data were extracted by two reviewers.

2.5.2 Exemplar: Joint Modelling of Serum Digoxin Concentration and All-Cause Mortality

To demonstrate applications of JMs on heart failure data, the ‘The Effect of Digoxin on Mortality and Morbidity in Patients with Heart Failure’ (DIG) [75] trial was used. The

dataset was obtained from the Biologic Specimen and Data Repository Information and Coordinating Center (BIOLINCC) under application #9257.

2.5.2.1 Statistical Methods

Only data from patients on the treatment arm with at least one measurement of serum digoxin concentration (SDC) was used. SDC measurements were right skewed and therefore a square root (sqrt) transformation was applied. For this illustrative example, only patients with no missing covariates were included.

The *JM* Package was used to fit all joint models, this package allows the fitting of joint models of longitudinal and time-to-event data in R under a maximum likelihood approach. [67].

Time must be modelled on the same scale for both models, and was modelled in the form of months (28 day calendar month) since randomisation; for SDC, time was taken as the specimen time. While the *JM* package allows for non-linear effects of time; for simplicity and ease of interpretation only linear terms were included.

The *JM* package requires an LME as fitted by the '*LME*' function from the *nlme* package for the longitudinal sub-model [67], [68]. For this example, both an unadjusted and adjusted LME model were fitted. With all models using sqrt SDC as the response variable. The unadjusted model included random intercepts as random effects. The adjusted model included the main effects of: estimated Glomerular Filtration Rate (eGFR), patient reported self-adherence, hours since last dose of the study drug and dose as fixed effects and included random intercepts and slopes for random effects. Full model equations for the LMEs and all other models are included in Table 14.

Time-to-Event models for the JM package can be fit using either the '*coxph*' or '*survreg*' functions from the *Survival* package. For simplicity, Cox PH models fitted by the '*coxph*' function were used. Like the LME both an unadjusted and adjusted model were fitted. The adjusted model containing the covariates of age, sex, ejection fraction, New York Heart Association class, history of hypertension, ischemic etiology of heart failure and body mass index. These covariates were selected on the basis of clinical relevance and prior knowledge of factors associated with outcomes in heart failure. The outcome examined was all-cause mortality.

Four JMs were constructed from both the unadjusted and adjusted LMEs and Cox PH Models as previously defined. Table 9 summarises the formulation of the JMs.

Table 9 Formulation of JMs Included in Exemplar

JM	LME	Cox PH	Time dependent parameter	Time dependent slope parameter
1	Unadjusted	Unadjusted	Y	N
2	Adjusted*	Adjusted†	Y	N
3	Adjusted*	Adjusted†	N	Y
4	Adjusted*	Adjusted†	Y	Y

* Adjusted for Estimated Glomerular Filtration Rate (eGFR), patient reported self-adherence, hours since last dose of the study drug and dose
† Adjusted for age, sex, ejection fraction, New York Heart Association (NYHA) class, history of hypertension, ischemic etiology of heart failure and Body Mass Index (BMI)

As an additional analysis the JMs were compared against traditional models. The traditional models were Cox PH models using first and last measurements of sqrt SDC as a covariate and an extended Cox PH model including sqrt SDC as a time-varying covariate. All models were adjusted for the same clinical covariates as the adjusted time-to-event models from the JMs. The model fit of the JMs were compared against each other using the Akaike Information Criterion (AIC). Likewise, the model fit of the traditional models were compared against each other using AIC. All models were compared for performance

using a discrimination index: c-index for the traditional models and a dynamic discrimination index for the JMs. The dynamic discrimination index was obtained using the function ‘dynCJM’ from the JM package. Based on the time dependent discrimination index proposed by Antolini et al. the dynamic discrimination index in this context provides a single statistic to summarise the discrimination power of the model over the follow-up time and is calculated from a weighted average of time-dependent AUCs which is comparable to the well-known c-index. Like the c-index it does not take into account censoring [76], [77]. The parameter estimates and standard errors from the model were also compared. One hundred bootstrap samples were used to internally validate the comparison of the discrimination index.

For descriptive purposes, categorical variables are represented as percentages, continuous variables are represented as median (IQR). JM association parameters are represented as hazard ratios (HRs) and 95% confidence intervals (CI). A time dependent association parameter is the hazard of all-cause mortality per one unit increase of sqrt SDC at any time point. A time dependent slope association parameter is the hazard of all-cause mortality per one standard deviation increase in the slope of sqrt SDC at a time point (known as the instantaneous or current slope). A p-value of less than 0.05 is considered statistically significant.

All statistical analysis was conducted using R Version 4.0 [78] and *JM* package version 1.4-8 [67].

Ethical approval was not required for this systematic review and exemplar. All methods were carried out in accordance with relevant guidelines and regulations.

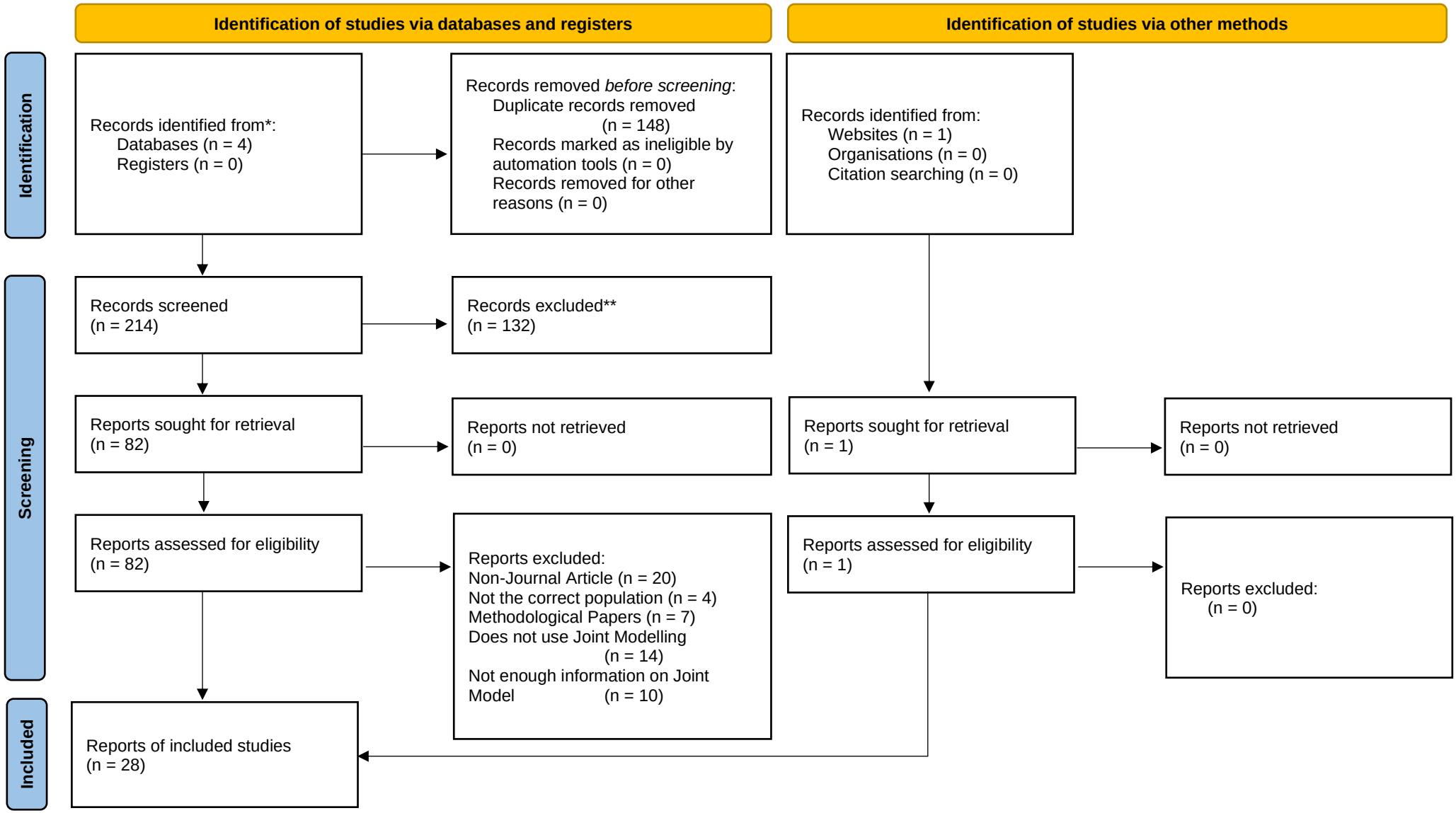
2.6 Results

2.6.1 Systematic Review: Use of Joint Modelling in Heart Failure

Figure 6 shows the PRISMA flow diagram, with 28 studies meeting the criteria for inclusion. Table 15 outlines the data sources of the 28 studies which met the inclusion criteria, the earliest included study being published in 2014 and between 4-7 studies being published per year from 2017 to the last search (10th December 2021).
From the included studies, 25 (89%) used data from cohort studies and the remaining 3 (11%) studies used data from clinical trials. It is worth noting that 10 (36%) of the cohort studies used data from the Bio SHiFT study [79], likely because of the study design with its focus on repeated measurements of biomarkers.

From the studies, 22 (78%) exclusively included patients with heart failure, 6 of which specified patients with heart failure with reduced ejection fraction (HFrEF). The remaining 6 (22%) studies exclusively included patients with acutely decompensated heart failure. There were also studies which further selected patients on specific characteristics such as patients implanted with cardiac devices such as Cardioverter-Defibrillators (ICD) or Cardiac Resynchronization Therapy Devices (CRT-D), patients with advanced heart failure and patients who had undergone transcatheter mitral-valve repair.

Figure 6 Prisma 2020 Flow Chart



2.6.1.1 Rationale

The most common rationale for using joint modelling was to assess the association of a biomarker with the hazard of an event. Other rationale included: joint modelling as a sensitivity analysis, reduction of bias due to censoring / mortality, comparison of prognostic models e.g., Weibull survival models and JMs, personalised prognostication using JMs, accounting for measurement error, different follow-up times and efficiency through combining data (i.e., smaller standard errors [52], [80]).

2.6.1.2 Longitudinal data

Table 10 summarises the longitudinal data used in the included studies; 21 (75%) studies used biomarkers with the most common biomarker being N-Terminal Pro-Brain Natriuretic Peptide (NT-ProBNP). Some studies included multiple biomarkers in longitudinal sub-models, and some used multiple JMs of different biomarkers. The remaining studies used imaging parameters such as Left Ventricular Ejection Fraction (LVEF) and functional parameters such as health status, physical activity and depression for their longitudinal data. All but two studies specified their longitudinal sub-models as a linear mixed effects model.

Table 10 Summary of Longitudinal Data of Included Studies

Paper	Primary Longitudinal Data Type	Longitudinal Data	List of Biomarkers
Abebaw et al., 2021 [81]	Biomarker	Pulse Rate	Pulse Rate
Alvarez-Alvarez et al., 2021 [82]	Imaging Parameters	Left Ventricular Ejection Fraction	
Arnold et al., 2019 [83]	Functional Parameters	Health Status: KCCQ-OS	
Belay et al., 2021 [84]	Biomarker	Pulse Rate	Pulse Rate
Biegus et al., 2019 [85]	Biomarker	Meld-XI	Creatinine, Bilirubin
Bouwens et al., 2019 [86]	Biomarker	Cardiac Remodelling Biomarkers	ST2, Gal-3, Gal-4, GDF-15, MMP-2, MMP-3, MMP-9, TIMP-4, PLC, AP-N, CASP3, CTSD, CTSZ, CSTB, NT-ProBNP
Bouwens et al., 2020 [87]	Biomarker	Cell Adhesion Circulation Biomarkers	SELP, SELE, CDH5, ICAM-2, PECAM-1, C1qR, CHI3L1, CNTN1, EPHB4, Ep-CAM, ITGB2, JAM-A
Bouwens et al., 2020 [88]	Biomarker	Multiple Biomarkers	CCL15, CC16, CCL24, CXCL16, FAS, IL-1RT1, IL-1RT2, IL-17RA, IL-18BP, IL2-RA, IL-6RA, LTBR, TNF-R1, TNF-R2, TNFRSF10C, TNFRSF14, TNFSF13B
Brankovic et al., 2017 [89]	Biomarker	Renal Markers	Creatinine, eGFR, CysC, KIM-1, NAG, NAGL
Canepa et al., 2020 [90]	Biomarker	Multiple Biomarkers	SBP, Heart Rate, Haemoglobin, Creatinine, Uric Acid
Castelvecchio et al., 2018 [91]	Biomarker	Natriuretic Peptides	NT-ProBNP
Freedland et al., 2021 [92]	Functional Parameters	Depression: PHQ-9	
Hurst et al., 2019 [93]	Biomarker	Serum Lactate Dehydrogenase	LDH
Kelly et al., 2020 [94]	Functional Parameters	Physical Activity reported by ICD or CRT-D (Accelerometer Measurement of >25mg)	
Klimczak-Tomaniak et al., 2020 [95]	Biomarker	Macrophage and Neutrophil Related Biomarkers	M130(CD163), TRAP, GRN, SPON1, PGLYRP1, TFPI
Liu et al., 2018 [96]	Biomarker	Growth-Differentiation Factor	GDF-15
Nunez et al., 2014 [80]	Biomarker	Red Blood Cell Distribution Width	RDW

Nunez et al., 2017 [97]	Biomarker	Carbohydrate Antigen, Natriuretic Peptide	NT-ProBNP, CA125
Schreuder et al., 2021 [98]	Biomarker	Multiple Biomarkers	NT-ProBNP, HsTNT, CRP, Creatinine, eGFR, CysC, NAG, KIM-1
van Boven et al., 2017 [99]	Biomarker	MicroRNAs	miR-1254, miR-22-3p, miR-423-5p, miR-486-5p, miR-320a, miR-345-5p, miR-378a-3p
van Boven et al., 2018 [100]	Biomarker	Multiple Biomarkers	NT-ProBNP, HsTNT, CRP
van den Berg et al., 2019 [101]	Biomarker	Fibrinolysis Factors	PAI-1, tPA, uPA, suPAR
van den Berg et al., 2019 [102]	Imaging Parameters	Echocardiographic Parameters	
van den Berge et al., 2021 [103]	Imaging Parameters	Remodelling Parameters: LVEF, LVED, LVES	
van Vark et al., 2017 [104]	Biomarker	Galectin-3	Gal-3
van Vark et al., 2017 [105]	Biomarker	ST2	ST2
Veen et al., 2021 [103]	Imaging Parameters	Tricuspid regurgitation	
Zhang et al., 2018	Biomarker	Natriuretic Peptides	NT-ProBNP

AP-N: aminopeptidase-N; CASP3: caspase-3, CSTB: cystatin-B, CTSD: cathepsin D, CTSZ: cathepsin Z; eGFR: estimated glomerular filtration rate, Gal-3: galectin-3, Gal-4: galectin-4, GDF-15: growth differentiation factor 15, HsTnT: highly sensitive cardia45c troponin T, MMP-2, 3, and 9: matrix metalloproteinase 2, 3, and 9, NT-proBNP: N-terminal pro–B-type natriuretic peptide, PLC: perlecan, ST2: suppression of tumorigenicity-2, TIMP-4: tissue inhibitor metalloproteinase 4, C1qR Complement component: C1q receptor, CDH5: Cadherin 5, CHI3L1: Chitinase-3-like protein 1, CNTN1: Contactin-1, Ep-CAM: Epithelial cell adhesion molecule, EPHB4: Ephrin type-B receptor 4, ICAM-2 Intercellular adhesion: molecule-2, ITGB2: Integrin beta-2, JAM-A: Junctional adhesion molecule A, PECAM-1: Platelet endothelial cell adhesion molecule 1, SELE: E-selectin, SELP: P-selectin, CCL15: C-C motif chemokine 15, CCL16: C-C motif chemokine 16, CCL24: C-C motif chemokine 24, CXCL16: C-X-C motif chemokine 16, FAS: tumour necrosis factor receptor superfamily member 6, IL-18BP: interleukin-18-binding protein, IL-17RA: interleukin-17 receptor A, IL2-RA: interleukin-2 receptor subunit alpha, IL-6RA: interleukin-6 receptor subunit alpha, IL-1RT1: interleukin-1 receptor type 1, IL-1RT2: interleukin-1 receptor type 2, LTBR: lymphotoxin b receptor, TNF-R1: tumour necrosis factor receptor 1, TNF-R2: tumour necrosis factor receptor 2, TNFRSF14: tumour necrosis factor receptor superfamily member 14, TNFRSF10C: tumour necrosis factor receptor superfamily member 10C, TNFSF13B: tumour necrosis factor ligand superfamily member 13B, CysC: cystatin C, estimated glomerular filtration rate: eGFR, NAG: N-acetyl-beta-D-glucosaminidase, KIM-1: kidney injury molecule, NGAL plasma and urinary neutrophil gelatinase-associated lipocalin, SBP: Systolic Blood Pressure, NT-ProBNP: N-terminal pro-B-type natriuretic peptide, PHQ-9: Patient Health Questionnaire-9, CD163 (M130): scavenger receptor cysteine-rich type 1 protein M130, TRAP: tartrate-resistant acid phosphatase type 5, GRN: granulins, SPON1: spondin-1, PGLYRP1: peptidoglycan recognition protein 1, TFPI : tissue factor pathway inhibitor, GDF-15: Growth-differentiation factor-15, RDW: Red Blood Cell Distribution Width, CA125 Carbohydrate Antigen 125, miRs: microRNAs, CRP: C-reactive protein, PAI-1: plasminogen activator inhibitor 1, tPA: tissue-type plasminogen activator, uPA: urokinase-type plasminogen activator, suPAR: soluble urokinase plasminogen activator surface receptor, LVEF: Left ventricular ejection fraction, LVED: left ventricular end-diastolic diameter, LVES: left ventricular end-systolic diameter.

2.6.1.3 Time-to-Event (Survival data)

Many studies included multiple events for their survival data through use of a composite outcome or multiple JMs. Table 11 shows that composite outcomes were the most common, but the events of composite outcomes varied by patient population as shown in Table 16. The second most common event was all-cause mortality. Most models utilised Cox PH models for their survival sub-models with only two studies specifying a parametric Weibull model.

Table 11 Survival End Points of Included Studies

Endpoint	Overall (N=41)
All-Cause Mortality	12 (29.3%)
Cardiovascular Mortality	4 (9.8%)
Components of composite endpoint	2 (4.9%)
Components of the composite endpoint, MI, PCI, CABG, CVA and all-cause mortality	2 (4.9%)
Composite	17 (41.5%)
Default from Treatment	1 (2.4%)
Development of Anaemia	1 (2.4%)
Some studies included multiple JMs with different end points, so the total number of endpoints (41) is more than the number of included studies (27).	

2.6.1.4 Missing Data

Common joint modelling packages such as *JM* and *JMBayes* allow for both uneven spacing and missing longitudinal measurements. Both these packages require all covariates from both longitudinal and survival sub-models to be complete. This common limitation resulted in 13 (46%) of the included studies using imputation methods to complete missing data.

2.6.1.5 JM

All included studies modelled their JMs with R. The two most common packages being *JM* and *JMBayes*, with 7 (25%) studies using the *JM* package and 10 (36%) using *JMBayes and* another 3 (10%) studies specified both packages. Three studies used custom code, one study used the *joineRML* package and the remaining 3 (10%) did not specify the package used.

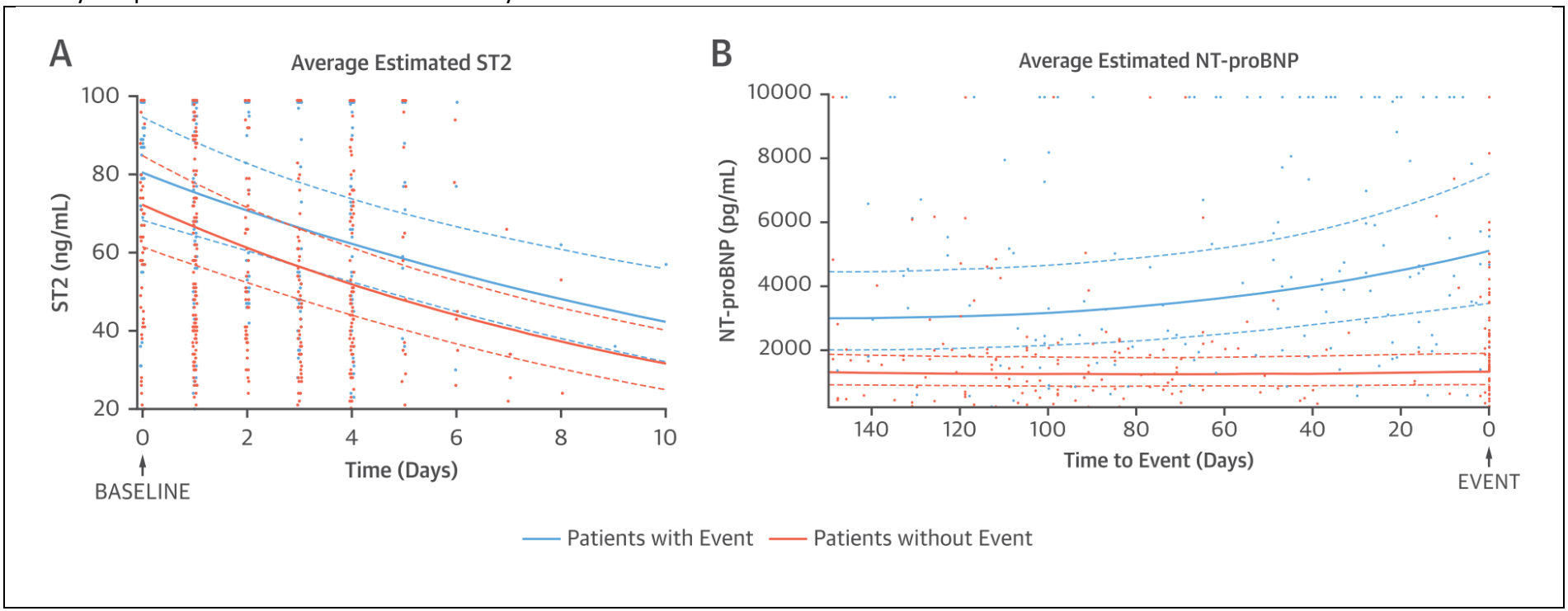
The R Packages used show both use of frequentist and Bayesian analysis. While the frequentist method use the maximum likelihood approach and is more comparable to more traditional models the Bayesian approach typically relies on Markov chain Monte Carlo (MCMC) sampling algorithms and may improve analysis by using related historical information and allowing for more flexible estimation [106].

2.6.1.6 Presentation of Results

Generally, the results from the JMs included the hazard ratio of the associated longitudinal outcome of interest on the time-to-event outcome; this was typically either the association of the value of the longitudinal outcome or the slope of the outcome on the time-to-event model.

The longitudinal sub-model is often presented as a coefficient or a graph of the average change in the longitudinal outcome of interest over time, these graphs are commonly split into groups of subjects e.g., those who did or did not experience the time-to-event outcome of interest. An example of this is illustrated in Figure 7 from the Vark et al. study [105].

Figure 7 Average Estimated Biomarker Pattern, Combined with Individual Biomarker Measurements. During Follow-up in Patients with and without the Primary Endpoint from the van Vark et al. Study

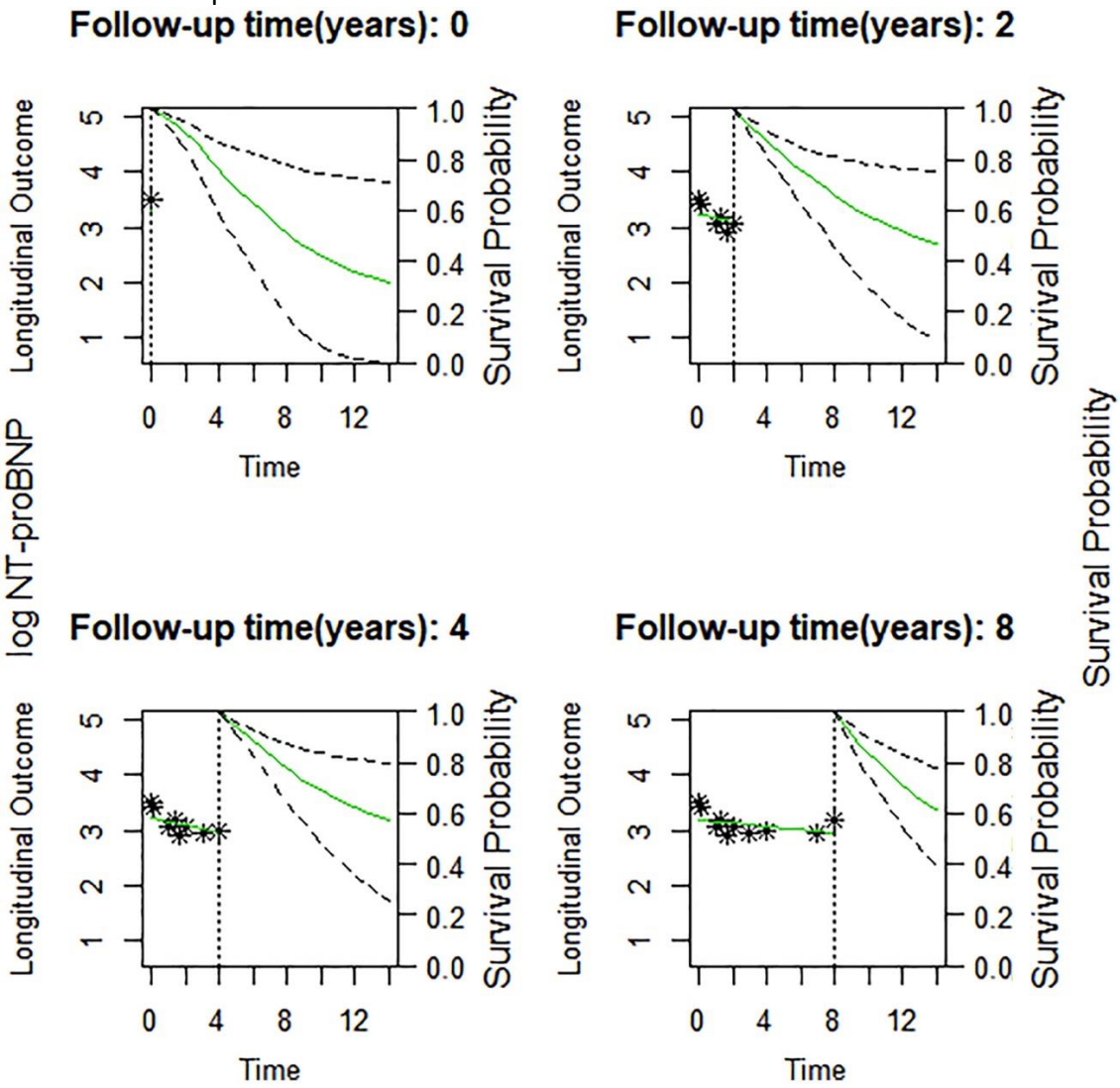


Reprinted with permission from van Vark et al. © 2017 The American College of Cardiology Foundation [105]

Commonly used packages in R provide capability to visually represent a trajectory of the longitudinal measure and the resulting changes in survival probability as shown in Figure 8 taken from Zhang et al. where the trajectory of the longitudinal measures (NT-proBNP) is plotted on the left and the survival probability with 95% confidence intervals are plotted on the right. This is useful when looking at individual patient trajectories, in their example Zhang et al. show how the probability of survival changes in response to changes of the trajectory of NT-proBNP and a narrowing of the confidence intervals can be observed with the increase of measurements of NT-proBNP [107]. These such plots while useful were not common amongst the included studies.

Given this is a relatively new approach to analysing longitudinal data simultaneously with survival data studies often compare results from JMs against more traditional models such as a Cox PH model with only a singular measurement of the variable of interest.

Figure 8 Dynamic survival probabilities with 95% CI based on various measurements of NT- ProBNP for a patient whose values fell.



Reprinted with permission from Zhang et al. © 2018 Elsevier B.V. All rights reserved. [107].

2.6.1.7 Joint Modelling Outcomes

Generally, the JMs of the included studies performed favourably in terms of improving prognostication and identifying associations with adverse events. The Bio SHiFT study being the most common data source explored a variety of biomarkers and imaging parameters including cell adhesion circulating bio markers, fibrinolysis factors, renal markers, echocardiographic parameters, Micro Ribonucleic Acid (MiRNA's), cardiac remodelling bio markers and macrophage and neutrophil related Biomarkers highlighting a variety of biomarkers and imaging parameters that were associated with the adverse events

[86]–[89], [95], [98]–[102]. Whilst many of the tested biomarkers and parameters produced positive results, Van den berg et al. suggested that repeated measures of imaging parameters such as LVEF do not add any more value than single parameters due to the lack of change in those parameters over the observation period [102]. While this suggestion may be true for the population of the Bio SHiFT study, both Alvarez-Alvarez et al. and Van den Berge suggested that for other heart failure populations repeated measurements of echocardiographic parameters such as LVEF can be useful, with Alvarez-Alvarez et al. investigating these parameters in a chronic heart failure population after CRT and Van den Berge exploring these parameters in an acute heart failure population [82], [103].

One of the key biomarkers which was explored in many studies was NT-ProBNP with Zhang et al., Castelvecchio et al. and Van Boven exploring its association with adverse events in a chronic heart failure population [91], [99], [107]. While these studies demonstrated the association between NT-ProBNP and adverse events, the Zhang et al. study suggested that the most recent value of NT-ProBNP had a similar predictive value as the serial measurements, but similar to the Van den Berg et al. study this may simply be due to the lack of change in values of NT-ProBNP within the study population and may not be generalisable [107].

Other key biomarkers which appeared in multiple studies were High sensitivity Troponin T (HsTnT), C-Reactive Protein (CRP), Cancer Antigen 125 (CA125), creatinine, Suppression of Tumorigenicity 2 (ST2), Galectin-3 (GAL-3) and Growth Differentiation Factor 15 (GDF-15) [85], [96]–[98], [100], [104], [105], [108], indicating that repeat measures of these markers are of interest within heart failure populations. Additionally, other less frequent markers included Lactase Dehydrogenase Trends (LDH) [93], Red blood cell Distribution Width (RDW) [80] and ambulatory markers such as Systolic Blood Pressure (SBP), heart rate and haemoglobin [81], [84], [90].

Along with biomarkers, functional parameters were also of interest, with Kelly et al. taking a novel approach investigating the association of physical activity as reported by implanted devices i.e., by ICD or CRT-D [94] and Arnold et al. using the joint modelling of health status in the form of Kansas City Cardiomyopathy Questionnaire Overall Summary Score (KCCQ-OS) score and all-cause mortality as a sensitivity analysis to illustrate how censoring attenuated health status with respect to treatment effect [83].

Along with the echocardiographic parameters mentioned above imaging parameters were used in a total of four studies, using many of the parameters obtained from imaging such as LVEF, Left Ventricular End-Diastolic diameter (LVED), Left Ventricular End-Systolic diameter (LVES) and tricuspid regurgitation [103], [109].

2.6.2 Association between serum digoxin concentration and mortality

2.6.2.1 Baseline Characteristics

Table 12 shows the baseline characteristics of included patients (n=2012), with a median age of 64 years, 22% of patients being women, median ejection fraction was 29% and 35% of patients died.

Table 12 Baseline Characteristics According to Patient Status (Dead / Censored) and Overall Sample (n = 2012)

	Dead (n=713)	**Alive/ Censored (n=1299)**	**Overall (n=2012)**
Age			
Median (IQR)	66.0 (13.0)	63.0 (13.0)	64.0 (13.0)
Sex			
FEMALE	151 (21.2%)	300 (23.1%)	451 (22.4%)
MALE	562 (78.8%)	999 (76.9%)	1561 (77.6%)
Ejection Fraction			
Median (IQR)	25.0 (13.0)	30.0 (13.0)	29.0 (13.0)
NYHA Class			
Class I	75 (10.5%)	215 (16.6%)	290 (14.4%)
Class II	345 (48.4%)	741 (57.0%)	1086 (54.0%)
Class III	265 (37.2%)	325 (25.0%)	590 (29.3%)
Class IV	28 (3.9%)	18 (1.4%)	46 (2.3%)
History of Hypertension			
FALSE	376 (52.7%)	727 (56.0%)	1103 (54.8%)
TRUE	337 (47.3%)	572 (44.0%)	909 (45.2%)
Ischemic HF			
ISCHEMIC	505 (70.8%)	939 (72.3%)	1444 (71.8%)
NON-ISCHEMIC	208 (29.2%)	360 (27.7%)	568 (28.2%)
BMI			
Median (IQR)	25.9 (5.91)	26.6 (5.97)	26.4 (5.76)

2.6.2.2 JM – Longitudinal Data sqrt SDC over time

The coefficients from the longitudinal sub-model of JM 2 (Table 17) for time -0.004 (-0.005 - -0.002) suggests that the sqrt root SDC decreases by 0.004 per month after adjusting for covariates. Figure 9 shows a representation of the predicted, and therefore adjusted, average trajectories of SDC over time from JM 2. Which is divided into patients who died during follow-up and those who did not, it suggests that patients who died had on average higher levels of SDC.

Figure 9 Average Trajectories of SDC by Patient Status as Predicted by JM 2 with Observed Values and Trajectories of SDC

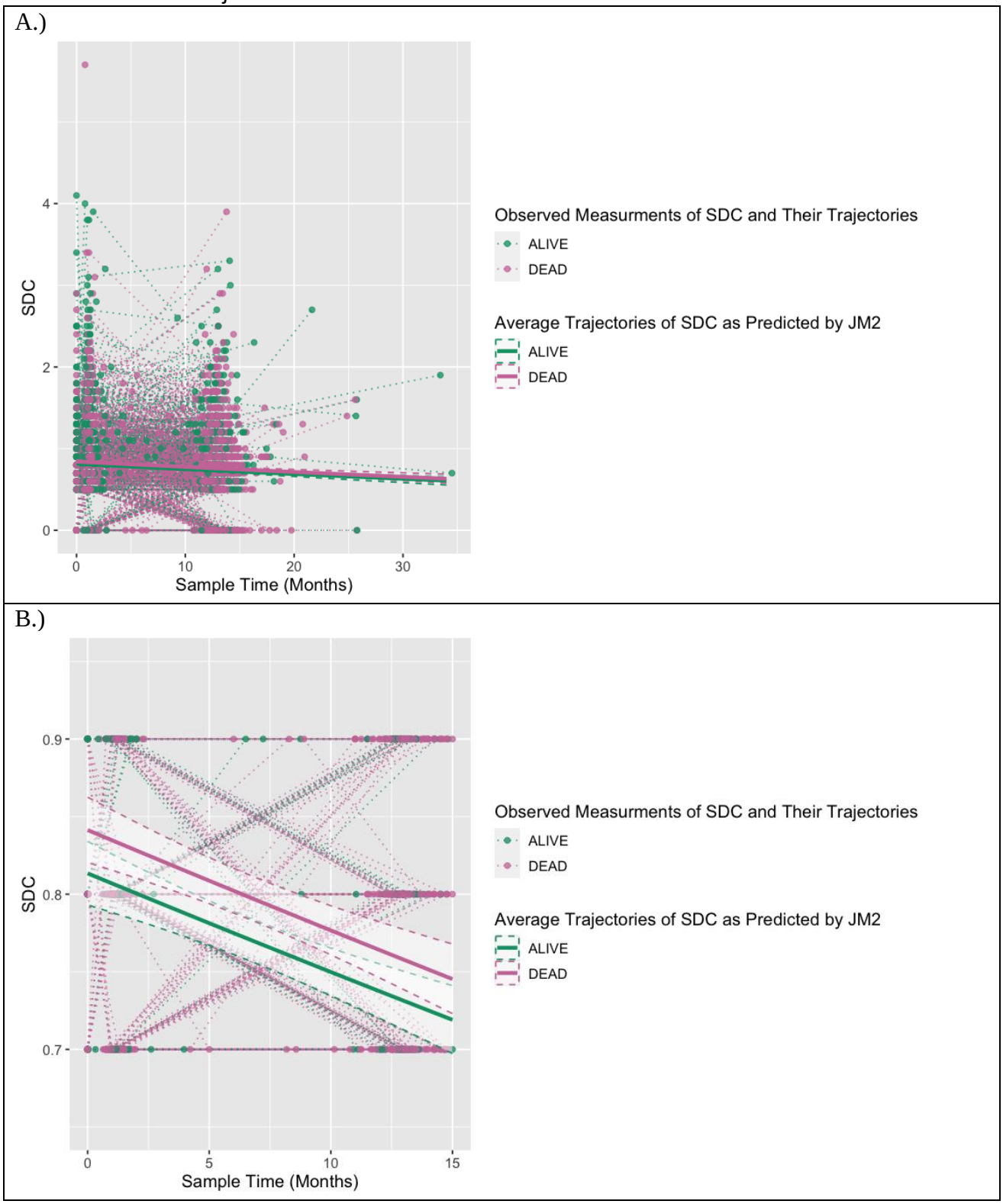


Figure 9a: Average Trajectories of SDC by Patient Status as Predicted by JM 2 with Observed Values and Trajectories of SDC on original axis scale, Figure 9b: Average Trajectories of SDC by Patient Status as Predicted by JM 2 with Observed Values and Trajectories of SDC on scaled axis scaled for readability. Average Trajectories were predicted using JM2 for patients whose status were either alive or dead based on mean and mode characteristics (covariates) of each stratum of patients (Alive or Dead).

2.6.2.3 JM – Hazard Ratios

Table 13 illustrates the results from the survival sub-models of the JMs in terms of a HR (95% CI) and p-value. All the JMs are time dependent relative risk models with a baseline risk function and as such the HRs can be interpreted similar to HRs from a proportional hazards model such as a Cox PH model. Focusing on the time dependent association parameter, the unadjusted model (JM 1) with a HR of 5.32 (3.07 – 9.22) suggesting a 5-fold increase in the hazard of all-cause mortality per unit increase of sqrt SDC. This association is attenuated when adjusted for clinical covariates in JM 2 with a HR of 1.77 (1.34 - 2.33). The time dependent slope parameter of JM 3 is above the significance threshold (p-value 0.092) indicating insufficient evidence to establish an association between the slope of sqrt SDC and all-cause mortality. Neither the HR for the time dependent parameter or the HR for the time dependent slope parameter of JM are above the significance threshold (p-values of 0.427 and 0.13, respectively) suggesting insufficient evidence to establish an association with either value or slope when the model is adjusted for both. The interpretation of the time dependent parameter of this model would be the hazard of all-cause mortality per unit increase of sqrt SDC for patients having the same slope. The interpretation of the time dependent slope parameter of this model would be the hazard of all-cause mortality per one SD increase in slope for patients having the same level of sqrt SDC.

Table 13 Event Summary of JMs Represented as Hazard Ratios

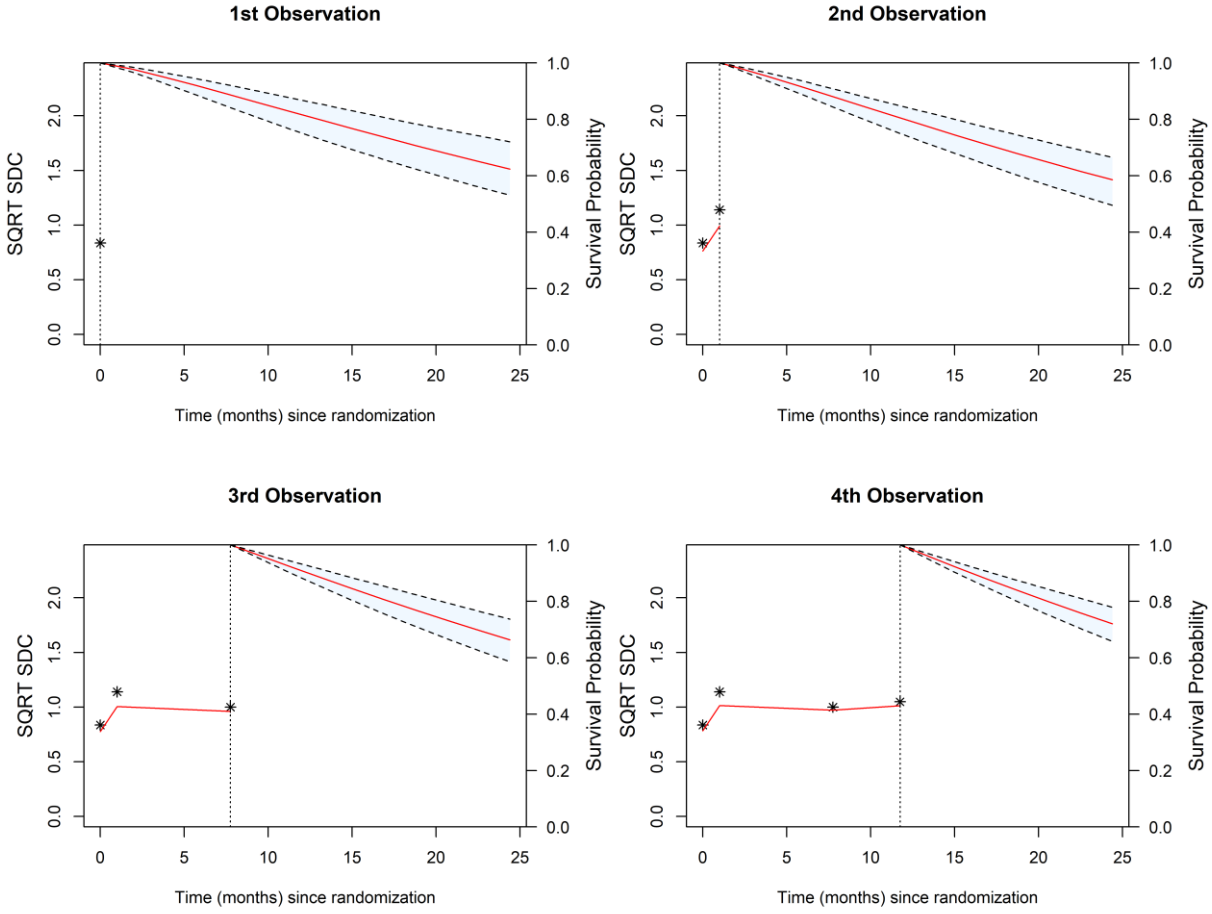
	JM 1 (Unadjusted)		JM 2 (Adjusted)		JM 3 (Adjusted Time Dependent Slopes)		JM 4 (Time Dependent and Time Dependent slopes)	
Variable	**HR**	**P Value**	**HR**	**P Value**	**HR**	**P Value**		
Age			1.02 (1.01-1.03)	<0.001	1.02 (1.01-1.03)	<0.001	1.02 (1.01-1.03)	<0.001
Male			1.19 (0.99-1.43)	0.062	1.2 (0.99-1.45)	0.064	1.22 (1.01-1.47)	0.042
Ejection Fraction %			0.97 (0.96-0.98)	<0.001	0.97 (0.96-0.98)	<0.001	0.97 (0.96-0.98)	<0.001
NYHA Class II			1.22 (0.95-1.57)	0.115	1.2 (0.93-1.55)	0.171	1.22 (0.94-1.57)	0.131
NYHA Class III			1.66 (1.28-2.16)	<0.001	1.63 (1.25-2.14)	<0.001	1.66 (1.27-2.17)	<0.001
NYHA Class IV			2.26 (1.45-3.53)	<0.001	2.3 (1.43-3.72)	0.001	2.36 (1.48-3.76)	<0.001
History of Hypotension			1.15 (0.99-1.34)	0.07	1.16 (0.99-1.36)	0.06	1.16 (0.99-1.35)	0.063
Non-Ischemic HF			1.07 (0.91-1.26)	0.428	1.07 (0.9-1.27)	0.455	1.08 (0.91-1.28)	0.386
BMI			0.98 (0.97-1)	0.036	0.98 (0.97-1)	0.061	0.99 (0.97-1)	0.113
Association (sqrt SDC)	5.32 (3.07-9.22)	<0.001	1.77 (1.34-2.33)	<0.001			1.33 (0.66-2.65)	0.427
Association Slope *					1.24 (0.97-1.59)	0.092	1.17 (0.96-1.42)	0.13

HR of all parameters except Association Slope reported as hazard of all-cause mortality per one unit increase at any point in time.
* HR of Association Slope reported as hazard of all-cause mortality per one standard deviation increase in the slope of sqrt

2.6.2.4 JM – individual patient trajectories.

Figure 10 shows the individual patient trajectories of a patient randomly selected (from patients with at least four measurements of sqrt SDC) at four different time points. These plots contain the longitudinal measurements of sqrt SDC as fitted by the JM on the left and the survival probability on the right, the dashed line indicating the last point the patient was known to be alive and the start of the survival curve, this point changes with each added measurement and as a result the survival curves are not directly comparable. However, these plots demonstrate how the measurements of sqrt SDC effects the survival probability and how the confidence intervals change over time with more measurements.

Figure 10 Individual Patient Trajectory of sqrt SDC and Survival Probability from Randomly Sampled Patient as Predicted by JM 2



2.6.2.5 Additional (Comparative) analysis

Results from the comparative analysis suggest that JM 2 had the best overall performance of the JMs with the lowest AIC, highest log likelihood and joint highest discrimination index of all models. The lower HR from the extended Cox PH model suggests that it underestimated the HR of the sqrt SDC parameter likely due to the nature of SDC as a time-varying covariate; SDC being an endogenous biological covariate subject to measurement error, biological variances, being able to change between measurements and finally requiring the subject to be alive at measurement. Underestimation of the association parameter has been previously demonstrated in simulation studies [110].

Internal validation of the discrimination index using 100 bootstrap samples showed that JM2 had a mean discrimination index of 0.66 (range 0.6 - 0.72) and outperformed the extended Cox PH model which had a mean discrimination index of 0.65 (range 0.62 – 0.67) 71% of the time with respect to the discrimination index, and 66% of the time when compared to Cox PH last measurement model which had a mean discrimination index of 0.65 (range 0.62 - 0.68). The full additional analysis is available in section 2.15.

2.7 Discussion

In 2016, a systematic review by Sudell et al. showed an increase in use of JMs of longitudinal and time-to-event data over time. However, only 3 identified studies used ‘heart related’ data; the most common applications were to cancer and HIV/AIDS studies [111]. Developing on their search strategy, we identified 28 studies by systematic review applying joint modelling within an adult heart failure population. and with use of an illustrative example have shown how to fit an interpret a JM. We have also shown how a JM approach can be used to examine the association between a biochemical test and outcomes in patients with heart failure.

Open-source software packages available in R such as *JM* and *JMBayes* make joint modelling more accessible reflected by the 20 (71%) studies using these packages. While these packages limit the JMs by way of underlying methodology [67], [68], if this methodology is not suitable custom code may be written as illustrated by the Hurst et al study [93]. Both the JM and JMBayes packages also contain limitations around missing data in covariates. While the packages allow for missing longitudinal data, they do not allow for missing covariates in the sub models used to build the JMs. This results in the need to either use a complete case with regards to the covariates as shown in our exemplar or use imputation techniques such as multiple imputation as highlighted in the included studies.

Due to the clinical nature of the included studies we found that studies often lacked details on the formulation of the JMs, e.g, the baseline risk function. Whilst this information could usually be derived by considering the packages used to fit the JMs, this information may be useful for reproducibility. We also identified that there was a lot of ambiguity around the origin of figures; whether or not they came from JMs or the individual components of the JMs e.g., a linear mixed effects model, modelled independently of the JM. We therefore suggest the need for clarity and transparency of the presentation of results from JMs.

It is also important that the results are easily understandable to a general audience. For example, the HR of a time dependent association parameter is intuitive but the HR of a time dependent slope parameter less so. Clinicians will often consider trends of biomarkers in day-to-day decision making so understanding these association parameters are key to relating them to clinical practice.

One driving motivation of the use of JMs was utilisation of repeated measures to inform prognosis and the comparison against a single measure. Most studies investigated biomarkers such as NT-ProBNP, CA125 and renal markers. However, JMs were not only

limited to biomarkers; with such studies as van den Berg et al. investigating echocardiographic parameters [102], Arnold et al. focusing on health status [83] and Kelly et al. exploring physical activity as reported by an implanted device [94]. The use of these data illustrates the robustness of joint modelling. Another key rationale was the use of joint modelling to reduce bias due to censoring and mortality. Bias of this nature often occurs because subjects who are sicker are more likely to experience the event of interest or withdraw from the study earlier than those who are healthier leading to fewer longitudinal measurements [56]. To overcome this joint modelling provides a framework that acknowledges the underlying relationship between the longitudinal and event process through the use of a joint distribution [52]. The Arnold et al. study illustrates this bias visually highlighting how censoring likely attenuated heath status with respect to treatment effect [83]. Further, we have highlighted the use of joint models to handle missing not at random data through use of a joint distribution [52].

Only three studies used data from RCTs [83], [85], [90], as previously mentioned joint modelling can be used to reduce bias with respect to treatment effect. Whilst this highlights a potential gap in the literature it should be noted that during screening, we identified numerous studies using joint modelling as a sensitivity analysis with results consistent to those from separate longitudinal and survival models but were excluded from review as not enough details about the models were included for full appraisal.

Compared to cancer studies, there was a lack of focus on quality-of-life data with only one study including quality of life in the form of a Kansas City Cardiomyopathy Questionnaire and SF-36 scores [83], and one which included depression by means of patient health questionnaire 9 scores [92]. Whereas joint modelling with quality of life is much more prevalent in cancer studies [51], [83]. This highlights another area which may be of interest to future studies using joint modelling in heart failure.

The Bio-Shift Cohort made up 36% of studies primary data, illustrating how a study can be developed to fully use the capabilities of joint modelling; with frequent blood sampling and measurements of endpoints the study leads the way for further larger studies of this nature [79].

From the studies using data from the Bio-Shift Cohort we identified 3 studies which only selected baseline and the last two measurements closest to the endpoint [86], [101], [112]. While justified to investigate the trajectories before and after an event it should be noted that this kind of analysis could lead to bias and should only be conducted with proper justification.

Many of the included studies demonstrated how repeated measures added value with respect to both prognostication and model fit. The outcomes of the JMs illustrate how joint modelling can improve on traditional models and highlights the use of joint modelling to assess associations of various biomarkers, imaging parameters and functional parameters, and adverse outcomes as well as provide dynamic predictions. However, there were studies which stated repeat measurements did not add prognostic value or improve model fit. For example, Van den Berg et al. stated that repeated measurements of echocardiographic parameters were associated with adverse events but did not add prognostic value due to the lack of change in measurements over time [102]. Whist this may be true for the bio-shift cohort, both van den Berge et al. and Alvarez-Alvarez et al. illustrated that given the right context these repeat measurements can still add value [82], [103]. This highlights an important caveat regarding JMs, in that the cost may not outweigh the benefit of the JMs; whilst biomarkers are routinely collected at little added cost other parameters may be costly to collect and an understanding of the temporal patterns of these parameters prior to joint modelling is advisable.

Our exemplar shows how joint modelling can be applied to older studies in order to maximise information from data that was sometimes collected but unused in expensive clinical trials. We highlighted how they compare to traditional models and how they can compete and improve on these models while also providing new clinical insight. The HR of the extended Cox PH model when compared to JM2 and the last measurement Cox PH Model suggests that it underestimated the association parameter, as previously stated likely due to the nature of SDC as a covariate. Whilst the extended Cox PH allows for repeated measures it does not account for measurement error, biological variance or that SDC may vary between time points or after the last observed measurement; this underestimation has previously been demonstrated in in simulation studies [110]. Joint modelling while allowing for repeated measurements of SDC can handle the biological endogenous nature of SDC providing better inferences [52]. Our results suggest that higher values of SDC rather than the slope is associated with higher mortality in patients with heart failure. Our work extends the findings of prior studies that have tried to examine this association with landmark methods which do not perform as well as shown by our exemplar analysis. The implications of this finding are that for patients, their SDC level should be kept as low as possible while still maintaining adequate dosing and in patients with high SDC consideration may have to be given to reducing the dose. There is however an issue of reverse causality, sicker patients may be prescribed higher doses and consequently have higher SDC. However, higher SDC could still act as an indicator of risk and should alert clinicians to reassess the patient and consider other therapies for their heart failure.

Our exemplar only included patients who had no missing covariate values. While this is satisfactory for an exemplar, it can lead to loss of information and possible bias in research studies and the best practice may be to use multiple imputation [113].However, it should

be noted that multiple imputation requires pooling for valid inference, which may cause issues with computational complexity and the need for pooling for dynamic predictions.

Our internal validation using 100 bootstrap samples showed JM2 outperformed the extended Cox PH and Cox PH last measurement models with respect to the discrimination index most of the time (71% and 65% respectively) within the bootstrap samples providing validation to the prognostic performance of the JM. However, the range of the discrimination indices of JM2 is wider than both other models suggesting more variability of discrimination with the joint model within the bootstrap samples. While our exemplar used a dynamic discrimination index for prognostic comparison against traditional models, we found that there was little consistency in methods used to compare JMs against each other and traditional models, highlighting a need for consistency when evaluating JMs. We would suggest that any model specifications or parameters are clearly described to allow any comparisons to be made in future research.

JM 3 in our exemplar included a slope parameter corresponding to the rate of change in sqrt SDC at a time point, known as the instantaneous or current slope. As previously stated, this parameter can be difficult to interpret and as such the *JMBayes2* package offers the use of other slope parameters such as delta change i.e., change in the last month / year prior to the time point. This parameter should be easier to interpret and is likely to be more prognostic than currently used slope parameters [114].

Both our review and exemplar highlight the various output and figures that can be produced from a JM and show how powerful joint modelling can be, with applications for prognostication, research of the association of repeat measurements of biomarkers and an endpoint, sensitivity analysis and more.

While joint modelling has a variety of uses, it may be most beneficial in the presence of informative censoring or dropout, when incorporating time-varying exogenous covariates such as biomarkers into survival models, and for prognostic modelling where dynamic predictions are useful. However Joint modelling can be computationally complex and take longer to fit than traditional models. It should also be noted that inferences may only be valid where the joint model has been correctly specified both with respect to the sub models and baseline hazard function. Joint models are also only valid when conducted on the same population, this is to say that both the longitudinal and time-to-event responses need to come from the same group of subjects. Joint models may not provide better prognostic inference where there is limited variability in the repeated measures such as shown by van den Berg et al. [102].

Our exemplar has some limitations such as the limited number of repeat measurements. This may have affected the power to estimate the slope association parameter and overall accuracy of the model.

2.8 Conclusions

In conclusion, this hybrid systematic review with exemplar highlights the rise in the use of JMs within heart failure, and our exemplar illustrates how JMs can be fitted to existing datasets adding value by utilising information from the repeated measures collected. This highlights why JMs are an increasingly popular alternative to traditional models such as Cox PH and Extended Cox PH.

2.9 Declarations

2.9.1 Ethical approval and consent to participants

Ethics approval was not needed for this systematic review and exemplar, a waiver was obtained from the University of Glasgow College of Medical, Veterinary and Life Sciences Ethics Committee. All methods were carried out in accordance with relevant guidelines and regulations.

2.9.2 Consent for Publication

Not applicable

2.10 Availability of Data and Materials

The DIG study is available from the National Heart, Lung and Blood Institute: Biological Specimen and Data Repository Information and Coordinating Center (NIH BIOLINCC): https://biolincc.nhlbi.nih.gov/studies/dig/

2.11 Competing Interests

The authors declare that they have no competing interests.

2.12 Funding

Ryan Field – Funded by University of Glasgow MVLS Doctoral Training Programme

Carly Adamson – Supported by British Heart Foundation Centre of Research Excellence Grant RE/18/6/34217

The systematic review and exemplar was carried out independently with no involvement or participation from the funders

2.13 Authors' Contributions

Conception and aims of the research was developed by RF, PJ and JL. Screening and Full text analysis was carried out by RF and CA. Analysis of the DIG dataset was carried out by RF with input from all authors (RF, CA, PJ, JL). The first draft was written by RF. All authors reviewed edited and commented on all versions of the manuscript. All authors read and approved the final version of the analysis.

2.14 Acknowledgements

The authors would like to thank the National Heart, Lung and Blood Institute: Biological Specimen and Data Repository Information and Coordinating Center for providing access to the DIG dataset.

2.15 Supplementary Material

Table 14 Model Equations for All Models

LME Models	
Unadjusted LME	$\begin{cases}\sqrt{SDC}_{ij} = \beta_0 + \beta_1 Specimen\ Time_{ij} + b_{i0} + \varepsilon_{ij}, \\ b_i \sim \mathrm{N}(0, D), \quad \varepsilon_{ij} \sim N(0, \sigma^2)\end{cases}$
Adjusted LME	$\begin{cases}\sqrt{SDC}_{ij} = \beta_0 + \beta_1 Specimen\ Time_{ij} + \beta_2 eGFR_{ij} + \beta_3 Self\ Adherence_{ij} + \beta_4 Hours\ Since\ Last\ Dose_{ij} + \beta_5 Dose_{ij} + b_{i0} + b_{i1} Specimen\ Time_{ij} + \varepsilon_{ij}, \\ b_i \sim \mathrm{N}(0, D), \quad \varepsilon_{ij} \sim N(0, \sigma^2)\end{cases}$
Cox PH Models	
Unadjusted Cox PH	$h_i(t) = h_0(t)$
Adjusted PH	$h_i(t) = h_0(t) \exp(\gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i)$
Joint Models	
JM1	$\begin{cases} y_i(t) = m_i(t) + \varepsilon_{i(t)} \\ \quad = \beta_0 + \beta_1 Specimen\ Time + b_{i0} + \varepsilon_{ij}, \quad \varepsilon_i(t) \sim N(0, \sigma^2) \\ h_i(t) = h_0(t) \exp\{\alpha m_i(t)\}\end{cases}$
JM2	$\begin{cases} y_i(t) = m_i(t) + \varepsilon_{i(t)} \\ = \beta_0 + \beta_1 Specimen\ Time + \beta_2 eGFR + \beta_3 Self\ Adherence + \beta_4 Hours\ Since\ Last\ Does + \beta_5 Dose\ b_{i0} + b_{i1} t + \varepsilon_{ij}, \quad \varepsilon_i(t) \sim N(0, \sigma^2), \\ h_i(t) = h_0(t) \exp\left\{\begin{matrix}\gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \\ \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i + \alpha m_i(t)\end{matrix}\right\},\end{cases}$

JM3	$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_{i\,(t)} \\ = \beta_0 + \beta_1 Specimen\ Time + \beta_2 eGFR + \beta_3 Self\ Adherence + \beta_4 Hours\ Since\ Last\ Does + \beta_5 Dose\ b_{i0} + b_{i1}t + \varepsilon_{ij}, \quad \varepsilon_i(t) \sim N(0, \sigma^2), \\ h_i(t) = h_0(t) \exp\left\{ \begin{matrix} \gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \\ \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i + \alpha m_i'(t) \end{matrix} \right\}, \end{cases}$$
JM4	$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_{i\,(t)} \\ = \beta_0 + \beta_1 Specimen\ Time + \beta_2 eGFR + \beta_3 Self\ Adherence + \beta_4 Hours\ Since\ Last\ Does + \beta_5 Dose\ b_{i0} + b_{i1}t + \varepsilon_{ij}, \quad \varepsilon_i(t) \sim N(0, \sigma^2), \\ h_i(t) = h_0(t) \exp\left\{ \begin{matrix} \gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \\ \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t) \end{matrix} \right\}, \end{cases}$$
Comparative (Traditional) Models	
Cox PH First Measurement	$$h_i(t) = h_0(t) \exp(\gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i + \gamma_{10} First\ Sqrt\ SDC_i)$$
Cox PH Last Measurement	$$h_i(t) = h_0(t) \exp(\gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i + \gamma_{10} Last\ Sqrt\ SDC_i)$$
Extended Cox PH	$$h_i(t) = h_0(t) \exp(\gamma_1 Age_i + \gamma_2 Male_i + \gamma_3 Ejection\ Fraction_i + \gamma_4 NYHA\ Class\ II_i + \gamma_5 NYHA\ Class\ III_i + \gamma_6 NYHA\ Class\ IV_i + \gamma_7 History\ of\ Hypertension_i + \gamma_8 Non-Ischemic\ HF_i + \gamma_9 BMI_i + \alpha SQRT\ SDC_i(t))$$

2.15.1 Supplementary Results

2.15.1.1 Additional (Comparative) analysis

Table 18 shows the performance characteristics of the JMs and traditional models. JM 2 had the best overall performance, with the lowest AIC (7273.89) and highest log likelihood (-3613.94) of the JMs and joint highest discrimination index of all models. The extended Cox PH model performed better than the traditional models with the lowest AIC (10217.12) and highest log likelihood (-5188.38). JM 1 had the worse overall performance of the JMs with the highest AIC, lowest log likelihood and lowest discrimination index. The two traditional models performed similarly to each other, with equal log likelihood and marginal differences in AIC and the last measurement model having a slightly better discrimination index of (0.65). This discrimination index of 0.65 shared by JM 2, the last measurement and extended Cox PH models suggest that the last measurement of SDC has a similar prognostic performance to repeated measurements of SDC.

Table 19 shows the hazard ratios from the Cox PH models and JM 2. It shows a marginal difference between the hazard ratio of Sqrt SDC from the last measurement model (1.78) and the time dependent association parameter JM 2 (1.77). Both the extended Cox and first measurement model had lower hazard ratios for Sqrt SDC of 1.73 and 1.52 respectively. The extended Cox model appears to have underestimated the HR of the sqrt SDC parameter, likely due to the nature of SDC as a covariate; SDC is a biological endogenous covariate, which is to say it has special properties that need to be considered and respected. These properties include: the need for the patient to be alive for the measurement, the measurement is subject to biological variances and measurement error, the measurement may only be observed at specific time points but may vary between these time points. This underestimation of the association parameter has previously been demonstrated in simulation studies such as those by Sweeting and Thompson [110]. Using one hundred bootstrap samples, the discrimination index of JM2, the Extended Cox PH and Cox PH last

measurement models were compared to internally validated the results of the discrimination index. From the 100 bootstrap samples JM 2 outperformed the extended Cox PH Model 71% of the time with respect to the discrimination index and 66% of the time when compared to the Cox PH Last Measurement model. The discrimination indices from JM2 within the 100 bootstrap samples had a mean of 0.66, median of 0.66 and range of 0.6 – 0.72, whilst the extended Cox PH model had a mean of 0.65, median of 0.65 and range of 0.62 – 0.67 and the Cox PH last measurement model having a mean of 0.65, median of 0.65 and range of (0.62 – 0.68).

Figure 11 Full search strategy

Embase (OVID)

1. (joint adj3 model*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
2. ((heart* or cardi* or myocard*) adj2 (failure* or incompet* or insufficien* or decompensat*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
3. exp heart failure/
4. 2 or 3
5. 1 and 4

Embase (OVID)

1. (joint adj3 model*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
2. ((heart* or cardi* or myocard*) adj2 (failure* or incompet* or insufficien* or decompensat*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
3. exp heart failure/
4. 2 or 3
5. 1 and 4

Medline (OVID)

1. (joint adj3 model*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
2. ((heart* or cardi* or myocard*) adj2 (failure* or incompet* or insufficien* or decompensat*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
3. exp heart failure/
4. 2 or 3
5. 1 and 4

Medline (OVID)

1. (joint adj3 model*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
2. ((heart* or cardi* or myocard*) adj2 (failure* or incompet* or insufficien* or decompensat*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
3. exp heart failure/
4. 2 or 3
5. 1 and 4

Scopus

(TITLE-ABS-KEY (joint W/3 model*) AND TITLE-ABS-KEY ((heart* OR cardi* OR myocard*) W/2 (failure* OR incompet* OR insuffcien* OR decompensat*)))

Scopus

(TITLE-ABS-KEY (joint W/3 model*) AND TITLE-ABS-KEY ((heart* OR cardi* OR myocard*) W/2 (failure* OR incompet* OR insuffcien* OR decompensat*)))

Google Scholar (Through Publish or Perish)

Title Words: "heart failure"

Keywords: ("heart failure" AND "Joint Model")

Google Scholar (Through Publish or Perish)

Title Words: "heart failure"

Keywords: ("heart failure" AND "Joint Model")

Table 15 Summary of Included Studies Ordered by Year

Paper	Year	Primary Data Source (Study)	Primary Study Type	Included Study Population Size	Type of HF	Specific Characteristics
Nunez et al., 2014 [115]	2014	Custom COHORT	COHORT	1702	Acute	
Nunez et al., 2017 [116]	2017	Custom COHORT	COHORT	946	Acute	
Brankovic et al., 2017 [117]	2017	Bio SHiFT	COHORT	263	Chronic	
van Vark et al., 2017 [104]	2017	TRIUMPH	COHORT	475	Acute	
van Vark et al., 2017 [105]	2017	TRIUMPH	COHORT	475	Acute	
van Boven et al., 2017 [99]	2017	Bio SHiFT	COHORT	263	Chronic	
Zhang et al., 2018 [107]	2018	Custom COHORT	COHORT	1998	Chronic	
Castelvecchio et al., 2018 [118]	2018	Biomarker Plus	COHORT	143	Chronic	Ischemic
Liu et al., 2018 [119]	2018	Custom COHORT	COHORT	173	Chronic	Ischemic, HFrEF
van Boven et al., 2018 [100]	2018	Bio SHiFT	COHORT	263	Chronic	
Hurst et al., 2019 [120]	2019	Custom COHORT	COHORT	323	Chronic	Advanced Heart Failure
Arnold et al., 2019 [121]	2019	COAPT	RCT	611	Chronic	HFrEF, Patients randomised to Transcatheter Mitral-Valve Repair vs standard therapy
Biegus et al., 2019 [122]	2019	RELAX-AHF	RCT	1120	Acute	
van den Berg et al., 2019 [123]	2019	Bio SHiFT	COHORT	263	Chronic	
van den Berg et al., 2019 [102]	2019	Bio SHiFT	COHORT	106	Chronic	HFrEF
Bouwens et al., 2019 [101]	2019	Bio SHiFT	COHORT	263	Chronic	
Kelly et al., 2020 [124]	2020	Boston Scientific ALTITUDE registry	COHORT	20927	Chronic	With ICD / CRT-D

Paper	Year	Primary Data Source (Study)	Primary Study Type	Included Study Population Size	Type of HF	Specific Characteristics
Bouwens et al., 2020 [86]	2020	Bio SHiFT	COHORT	263	Chronic	
Bouwens et al., 2020 [88]	2020	Bio SHiFT	COHORT	250	Chronic	
Klimczak-Tomaniak et al., 2020 [125]	2020	Bio SHiFT	COHORT	263	Chronic	
Canepa et al., 2020 [126]	2020	GISSI-HF	RCT	5469	Chronic	HFrEF
Veen et al., 2021 [127]	2021	EUROMACS	COHORT	2496	Chronic	After LVAD Implantation
van den Berge et al., 2021 [123]	2021	Custom COHORT	COHORT	111	Acute	New Onset HF
Alvarez-Alvarez et al., 2021 [128]	2021	Custom COHORT	COHORT	328	Chronic	After Cardiac Resynchronization therapy
Schreuder et al., 2021 [129]	2021	Bio SHiFT	COHORT	250	Chronic	
Abebaw et al., 2021 [81]	2021	Custom COHORT	COHORT	302	Chronic	
Belay et al., 2021 [84]	2021	Custom COHORT	COHORT	271	Chronic	
Freedland et al., 2021 [130]	2021	Custom COHORT	COHORT	400	Chronic	

HFrEF: Heart Failure Reduce Ejection Fraction, ICD: Implantable Cardioverter-Defibrillator, CRT-D: implantable cardiac resynchronization therapy (CRT) defibrillator, LVAD: left ventricular assist device.

Table 16 Breakdown of Composite Components of Included Studies

Composite Components
All-Cause Mortality and HF Hospitalisation
Cardiac Death, Heart Transplantation, Left Ventricular Assist Device Implantation and Hospitalisation for the Management of acute or Worsened HF
Suspected Pump Thrombosis and Confirmed Pump Thrombosis
LV assist device implantation, heart transplantation, or all-cause mortality
Cardiac Death, Cardiac Transplant, LVAD Implantation and Hospitalisation for the management of Acute or Worsened HF
All-Cause Mortality and Readmission for HF
Hospitalisation for the Management of Acute or Worsened Heart Failure, LVAD Implantation, Cardiac Transplantation and Cardiac Death
MACE: Major Adverse Cardiac Events (All Cause Death, MI, and First HF Rehospitalization)

Table 17 Coefficients from the Longitudinal Sub model from JM2

Variable	**Value**	**Standard Error**	**p-Value**
Intercept	0.9798	0.0153	<0.0001
Specimen Time (Months)	-0.0036	0.0006	<0.000
eGFR	-0.0049	0.0002	<0.0001
Self-Adherence: None	-0.0570	0.0122	<0.0001
Self-Adherence: Some	-0.1498	0.0221	<0.0001
Number of Hours Since Dose	-0.0016	0.0001	<0.0001
Dose	1.1306	0.0503	<0.0001

Table 18 Performance Summary of Joint Models, Cox PH and Extended Cox PH models.

Measure	Joint Model 1 (Basic)	Joint Model 2 (Adjusted)	Joint Model 3 (Adjusted Time Dependent Slopes)	Joint Model 4 (Adjusted Time Dependent and Time Dependent Slopes)	Cox PH Model 1 (First Measurement)	Cox PH Model 2 (Last Measurement)	Cox PH Model 3 (Extended Cox)
AIC	10889.64	7273.89	10352.85	10355.42	10291.65	10277.86	10217.12
Log Likelihood	-5437.82	-3613.94	-5153.43	-5153.71	-5220.84	-5220.84	-5188.38
Discrimination Index	0.56	0.65	0.65	0.65	0.64	0.65	0.65

Table 19 Hazard Ratios and Standard Errors from Cox Models and JM2

	First Measurement Model			Last Measurement Model			Extended Cox Model			Joint Model 2		
Variable	**HR**	**SE**	**p-value**	**HR**	**SE**	**p-value**	**HR**	**SE**	**p-value**	**HR**	**SE**	**p-value**
Age	1.02	0.004	<0.001	1.02	0.004	<0.001	1.02	0.004	<0.001	1.02	0.004	<0.001
Male	1.18	0.094	0.079	1.19	0.094	0.06	1.18	0.094	0.079	1.19	0.094	0.062
Ejection Fraction %	0.97	0.004	<0.001	0.97	0.004	<0.001	0.97	0.004	<0.001	0.97	0.004	<0.001
NYHA Class II	1.21	0.128	0.14	1.21	0.128	0.138	1.20	0.128	0.157	1.22	0.128	0.115
NYHA Class III	1.65	0.134	<0.001	1.63	0.134	<0.001	1.62	0.134	<0.001	1.66	0.134	<0.001
NYHA Class IV	2.22	0.226	<0.001	2.27	0.226	<0.001	2.23	0.226	<0.001	2.26	0.227	<0.001
History of Hypotension	1.15	0.077	0.066	1.16	0.077	0.06	1.16	0.077	0.059	1.15	0.077	0.07
Non-Ischemic HF	1.07	0.085	0.456	1.06	0.085	0.499	1.05	0.085	0.529	1.07	0.085	0.428
BMI	0.98	0.008	0.033	0.98	0.008	0.033	0.98	0.008	0.038	0.98	0.008	0.036
Sqrt SDC	1.52	0.114	<0.001	1.78	0.113	<0.001	1.73	0.114	<0.001			
Time dependent association parameter for sqrt SDC*										1.77	0.140	<0.001

*The time dependent association parameter for sqrt SDC gives the hazard of all-cause mortality per 1 unit increase in square root transformed SDC at any time point. This HR can be compared with the HR for sqrt SDC from the Cox models.

Chapter 3 Data Source

3.1 Foreword

The aim of this chapter is to introduce the data source that will be used to construct joint models throughout the remaining chapters. Each of the chapters will briefly explain the data source. However, to allow for alternative format and continuity throughout the thesis, this chapter will introduce the data source in depth.

Both the primary and validation data sources were chosen as they are both large, multi-national randomised control trials which collected biomarker data alongside survival data making them suitable for joint modelling. Both trials are similar in design and sponsored by the same company (Novartis). At the time of this analysis, they were still two of the largest clinical trials within heart failure trialling the latest drugs, with the primary data source still being referenced by the ESC Heart Failure Guidelines [1].

3.2 Primary Data Source

Data for all models was obtained from the Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure (PARADIGM-HF) clinical trial. The PARADIGM-HF clinical trial was a two-armed, double-blind, parallel-group, event-driven, randomised control trial which randomised 8,442 chronic heart failure patients with an ejection fraction of ≤ 40% (HFrEF) and a NYHA class of II, III or IV to receive either 200mg dose of LCZ696 (sacubitril/valsartan) twice daily or 10mg dose of enalapril twice daily, in addition to recommended therapy. The study's primary outcome was a composite of hospitalisation for heart failure and death from cardiovascular causes. However, the trial was designed to identify any treatment difference in the rate of deaths by cardiovascular causes.
The full trial design, inclusion and exclusion criteria have been described in detail elsewhere and is shown in Table 20 [134]–[136].[131]–[133] Briefly, other inclusion

criteria for the trial included a minimum age of 18 years for informed consent, a BNP ≥ 100 picograms per millilitre ($pgml^{-1}$) or NT-ProBNP ≥ 400 $pgml^{-1}$ if the patient had a hospitalisation for HF within the last year, or a BNP ≥ 150 $pgml^{-1}$ or NT-ProBNP ≥ 600 $pgml^{-1}$ without a hospitalisation for HF. Patients needed to be receiving a stable dose of an ACE inhibitor or an ARB for a minimum of four weeks prior to screening, with a dose equivalent to enalapril 10 mg/day. Unless not tolerated or contraindicated, the patient should have also been receiving a stable dose of beta-blockers in the four weeks prior to screening.

Exclusion criteria included an eGFR < 30 millilitres per minute per 1.73 metre squared of body surface area (mL/min/1.73M^2) at visits 1, 3 and 5 or a reduction of ARBs, 5% from visits 1 to 3 or visits 1 and 5. Symptomatic hypertension or a systolic blood pressure < 100 mmHg at visit 1 or < 95 mmHg at visit 3 or 5. A Potassium level > 5.2 mmol/L at visit 1 or > 5.4 mmol/L at visit 3 or 5. Contraindication or history of hypersensitivity/allergy to ACE inhibitors, ARBs or similar drugs.

The stopping criteria for the trial were:

1. When a pre-specified number of patients (2410) had experienced the composite endpoint.
2. A statistically significant result in efficacy in the interim analysis (based on the Lan-deMets alpha spending function applied to the actual number of patients who experienced the primary outcome).
3. As the result of a critical safety concern at any of the interim safety analysis.

Each ethics committee of the 1,043 study centres approved the trial, and all participants gave written informed consent.

The full trial timeline is illustrated in Figure **12**. Prior to randomisation, the PARADIGM-HF trial had a two-week, single-blind, run-in period at visit 2 where the majority of eligible patients received a treatment of enalapril 10 mg twice daily. This run-in period was followed by single-blind treatment of LCZ696 100mg twice daily for two weeks and up titrated to 200mg twice daily for another two to four weeks. Other medications for HF were continued during these run-in periods except ACE inhibiters and ARB. The primary purpose of this run-in period was to ensure that patients that underwent randomisation could tolerate the target doses of the study drugs and maximising the possibility of those patients to tolerate the drugs over the long follow-up period [134].

Double-blind randomisation happened at visit 5; 8,399 patients were randomised at a 1:1 ratio of either LCZ696 200mg twice daily (sacubitril/valsartan) or enalapril 10mg twice daily, with doses being adjusted for tolerability. Forty-three patients were excluded from the trial either because they were enrolled to sites which closed due to serious infractions of Good Clinical Practice or were randomised as a result of an error.

For the first four months of the double-blind period, patients were followed up every two to eight weeks and then every four months, with further unscheduled visits when deemed necessary under the advisement of the investigator.

As per protocol and the prior specification of rules for interval efficacy analysis, the trial was stopped early as the threshold for the beneficial use of LCZ696 was crossed. The trial was stopped with a median follow-up time of 27 months. LCZ696 showed a significant reduction in primary outcome (CV mortality or HF hospitalisation) with a HR of 0.80 95% CI (0.76 - 0.93), whilst also showing a significant reduction in all-cause mortality with a HR of 0.84 95% CI (0.76 – 0.93) [43].

Table 20 Inclusion and Exclusion Criteria for the PARADIGM-HF RCT Adapted from McMurray et al.

Inclusion Criteria:

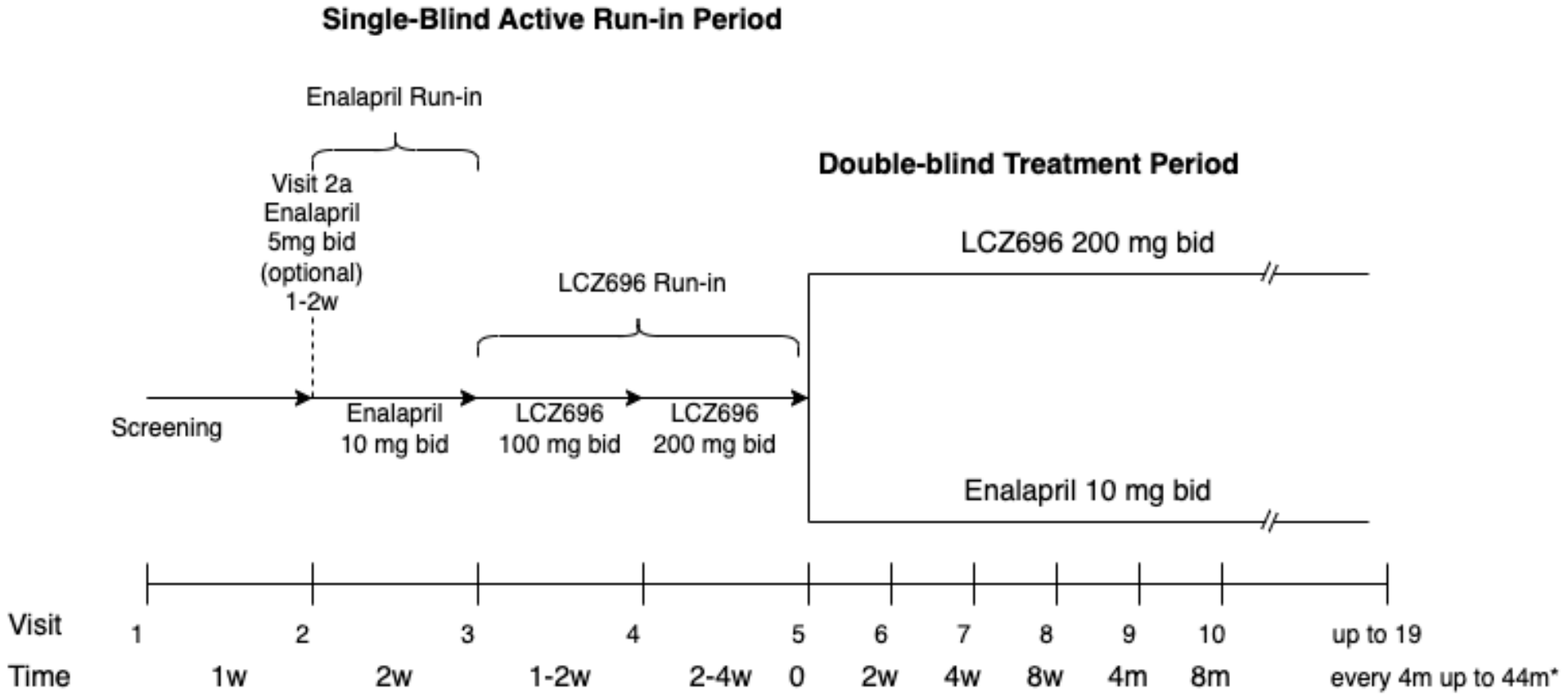
1. Patients must give written informed consent before any assessment is performed.
2. Outpatients ≥ 18 years of age, male or female.
3. Patients with a diagnosis of CHF NYHA class II-IV and reduced ejection fraction:
    - LVEF ≤ 40% at Visit 1 (any local measurement, made within the past 6 months using echocardiography, MUGA, CT scanning, MRI or ventricular angiography is acceptable, provided no subsequent measurement above 40%)
    - BNP ≥ 150 pg/ml (NT-proBNP ≥ 600 pg/ml) at Visit 1 OR BNP ≥ 100 pg/mL (NTproBNP ≥ 400 pg/ml) and a hospitalization for HF within the last 12 months
4. Patients must be on an ACEI or an ARB at a stable dose of at least enalapril 10 mg/d or equivalent for at least 4 weeks before Visit 1
    - For this protocol doses of other ACEIs considered to be equivalent to enalapril 10 mg/d include captopril 100 mg/d, cilazapril 2.5 mg/d, fosinopril 20 mg/d, lisinopril 10 mg/d, moexipril 7.5 mg/d, perindopril 4 mg/d, quinapril 20 mg/d, ramipril 5 mg/d, trandolapril 2 mg/d, and zofenopril 30 mg/d.
    - For this protocol doses of ARBs considered to be equivalent to enalapril 10 mg/d include candesartan 16 mg/d, eprosartan 400 mg/d, irbesartan 150 mg/d, losartan 50 mg/d, olmesartan 10 mg/d, telmisartan 40 mg/d, and valsartan 160 mg/d.
5. Patients must be treated with a β-blocker, unless contraindicated or not tolerated, at a stable dose for at least 4 weeks prior to Visit 1 (reason should be documented for patients not on CHF target doses per local guidelines, or in absence of that medication).

Exclusion Criteria:

1. Use of other investigational drugs at the time of enrollment, or within 30 days or 5 halflives of enrollment, whichever is longer
2. History of hypersensitivity or allergy to any of the study drugs, drugs of similar chemical classes, ACEIs, ARBs, or NEP inhibitors as well as known or suspected contraindications to the study drugs
3. Previous history of intolerance to recommended target doses of ACEIs or ARBs
4. Known history of angioedema
5. Requirement of treatment with both ACEIs and ARBs
6. Current acute decompensated HF (exacerbation of chronic HF manifested by signs and symptoms that may require intravenous therapy)
7. Symptomatic hypotension and/or a SBP < 100 mmHg at Visit 1 (screening) or < 95 mmHg at Visit 3 or at Visit 5 (randomization)
8. Estimated GFR < 30 mL/min/1.73m2 as measured by the simplified MDRD formula at Visit 1 (screening), Visit 3 (end of enalapril run-in), or Visit 5 (end of LCZ696 run-in and randomization) or > 25% decline in eGFR between Visit 1 and Visit 3 or between Visit 1 and Visit 5
9. Serum potassium > 5.2 mmol/L at Visit 1 (screening) or > 5.4 mmol/L at Visit 3 or Visit 5 (randomization)
10. Acute coronary syndrome, stroke, transient ischemic attack, cardiac, carotid or other major CV surgery, percutaneous coronary intervention (PCI) or carotid angioplasty within the 3 months prior to Visit 1
11. Coronary or carotid artery disease likely to require surgical or percutaneous intervention within the 6 months after Visit 1
12. Implantation of a cardiac resynchronization therapy device (CRTD) within 3 months prior Visit 1 or intent to implant a CRTD

13. History of heart transplant or on a transplant list or with left ventricular assistance device (LVAD)
14. History of severe pulmonary disease
15. Diagnosis of peripartum or chemotherapy induced cardiomyopathy within the 12 months prior to Visit 1
16. Documented untreated ventricular arrhythmia with syncopal episodes within the 3 months prior to Visit 1
17. Symptomatic bradycardia or second or third degree heart block without a pacemaker
18. Presence of hemodynamically significant mitral and/or aortic valve disease, except mitral regurgitation secondary to left ventricular dilatation
19. Presence of other hemodynamically significant obstructive lesions of left ventricular outflow tract, including aortic and sub-aortic stenosis
20. Any surgical or medical condition which might significantly alter the absorption, distribution, metabolism, or excretion of study drugs, including but not limited to any of the following:
    - History of active inflammatory bowel disease during the 12 months before Visit 1.
    - Current duodenal or gastric ulcers during the 3 months prior to Visit 1
    - Evidence of hepatic disease as determined by any one of the following: AST or ALT values exceeding 2 x ULN at Visit 1, history of hepatic encephalopathy, history of esophageal varices, or history of portacaval shunt
    - Active treatment with cholestyramine or colestipol resins
21. Presence of any other disease with a life expectancy of < 5 years
22. Pregnant or nursing (lactating) women, where pregnancy is defined as the state of a female after conception and until the termination of gestation, confirmed by a positive hCG laboratory test (> 5 mIU/mL)
23. Women of child-bearing potential, defined as all women physiologically capable of becoming pregnant, including women whose career, lifestyle, or sexual orientation precludes intercourse with a male partner and women whose partners have been sterilized by vasectomy or other means, UNLESS they are using two birth control methods. The two methods can be a double barrier method (if accepted by the local regulatory authority and ethics committee) or a barrier method plus a hormonal method
    - Adequate barrier methods of contraception include: diaphragm, condom (by the partner), intrauterine device (copper or hormonal), sponge or spermicide. Hormonal contraceptives include any marketed contraceptive agent that includes an estrogen and/or a progesterone agent.
    - Reliable contraception should be maintained throughout the study and for 7 days after study drug discontinuation.
    - Women are considered post-menopausal and not of child bearing potential if they have had 12 months of natural (spontaneous) amenorrhea with an appropriate clinical profile (e.g. age appropriate, history of vasomotor symptoms) or six months of spontaneous amenorrhea with serum FSH levels > 40 mIU/mL [for US only: and estradiol < 20 pg/mL] or have had surgical bilateral oophorectomy (with or without hysterectomy) at least six weeks ago. In the case of oophorectomy alone, only when the reproductive status of the woman has been confirmed by follow up hormone level assessment.

Figure 12 Overall Study design and timeframe between study visits occurring during the active run-in and double-blind periods for the PARADIGM RCT adapted from McMurray et al.



w: week, m: month. * Projected trial duration, actual duration of the trial was event driven.

[135]

3.3 Validation Data Source

Data to be used for validation purposes for all joint models was obtained from The Aliskiren Trial to Minimize Outcomes in Patients with Heart Failure (ATMOSPHERE) trial [44]. The ATMOSPHERE trial was a multicentre, double-blind, double-dummy, active-controlled, parallel-group, randomised control trial which randomised 7,016 patients with chronic heart failure with an ejection fraction of ≤ 35% and an NYHA class of II, III or IV at a 1:1:1 ratio to receive either enalapril at a dose of 5 or 10 mg twice a day, aliskiren at a dose of 150mg once a day, or a combination of enalapril (at a dose of 5 or 10mg twice daily) and aliskiren (at a dose of 150mg once a day). The dosage of aliskiren in both the aliskiren and combination group was increased to 300mg two weeks after undergoing randomisation, with the enalapril arm receiving a sham adjustment. The study's primary outcome was a composite of first hospitalization for heart failure and death from cardiovascular causes.

The full trial design, including inclusion and exclusion criteria, has been described elsewhere and are shown in Table 21 [137], [139], [140].[134], [136], [137] Put briefly, other inclusion criteria of the trial included a minimum age of 18 years, a BNP concentration ≥ 150 $pgml^{-1}$ (or an NT-proBNP concentration ≥ 600 $pgml^{-1}$), unless hospitalised for heart failure in the previous 12 months, in which the criteria was lowered to a BNP concentration ≥ 100 $pgml^{-1}$ (or an NT-ProBNP concentration ≥ 400 $pgml^{-1}$). At the time of enrolment, patients needed to have been receiving a beta blocker and a stable dose of an ACE inhibitor equivalent to a dose of 10 mg of enalapril daily.

Exclusion criteria for the trial included a systolic blood pressure < 95 mmHg at screening or < 90 mmHg at randomisation, symptomatic hypotension, an eGFR < 40 mL/min/1.73M^2 at screening or < 35 mL/min/1.73M^2 at randomisation, or if the patient

experienced a decline of > 25% in their eGFR between screening and randomisation. Other criteria also included contraindication or a history of intolerance to ACE inhibitors and a potassium level ≥ 5.0 mmol/L at screening or ≥ 5.2 mmol/L at randomisation.

Stopping criteria for the trial were:

1. A prespecified number of patients (2318) experiencing the composite endpoint.
2. A statistically significant result in efficacy, with a significance level of 0.005 (one-sided).
3. The data monitoring committee recommend stopping due to safety concerns.

Each ethics committee of the 789 centres approved the protocol, and all participants provided written informed consent.

The full trial timeline is illustrated in Figure 13. Prior to the randomisation of patients, the trial included a two-part, single-blind, run-in period. Patients who were already taking ACE inhibitors at a dose equal to 20mg of enalapril daily prior to the trial were eligible to bypass the first part of the run-in period and proceed to the second part directly. During the first part of the run-in period, patients received one to four weeks of enalapril twice daily at a dose of 5mg. Patients with an acceptable level of adverse events then proceeded to receive 10mg of enalapril twice daily for two to four weeks. At the end of the first part of the run-in period, patients were stratified into two groups - those who received 10mg twice daily (high-dose stratum) and those who did not (low-dose stratum). During the second part of the run-in period, patients received a 150mg dose of aliskiren once daily in addition to the enalapril. The purpose of this run-in period was to evaluate the tolerability of the study drugs and to ensure a maximum number of patients could be randomised, using stratified randomisation [44].

Double-blind randomisation happened after the run-in period was completed at visit 4. In total, 7,064 patients underwent randomisation at a 1:1:1 ratio to receive either enalapril at a dose of 5 or 10mg twice daily, aliskiren at a dose of 150mg once a day, or a combination of enalapril (5 or 10mg twice daily) and aliskiren (150mg once a day). As mentioned previously, the dosage of aliskiren was increased to 300mg once a day in both arms receiving it, with a sham increase in the enalapril arm at two weeks.

During the first four months of the double-blind period, patients were followed up every two to eight weeks and then every four months after. Dosage could be reduced if the patient could not tolerate the target dosage.

Following the outcome of the early termination of the Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Disease Endpoints (ALTITUDE) due to ineffectiveness and safety concerns and the ensuing findings that patients treated with aliskiren who had diabetes had worse outcomes than those treated with a placebo in the Aliskiren Trial on Acute Heart Failure Outcomes (ASTRONAUT). A protocol amendment was implemented worldwide in 2013 leading to patients in the ATMOSPHERE trial who had diabetes at baseline or developed it during the trial to be censored and the study drug discontinued and treated with conventional therapy instead.

The trial ended in July 2015 with a median follow-up time of 36.6 months. For each treatment group, there was no significant difference in treatment effect detected with a HR of 0.93 95% CI (0.85 – 1.03) for the primary composite outcome when comparing the combination therapy against enalapril. A HR of 0.99 95% CI (0.9 – 1.10) suggested there was also no significant difference in treatment effect on the primary composite outcome when comparing aliskiren and enalapril [44].

Table 21 Inclusion and Exclusion Criteria for the ATMOSPHERE RCT Adapted from McMurray et al.

Inclusion Criteria

1. Outpatients ≥ 18 years of age, male or female.
2. Patients with a diagnosis of chronic heart failure (NYHA Class II – IV):
    - LVEF ≤ 35% at visit 1 (local measurement, within the past 6 months assessed by echocardiography, MUGA, CT scan, MRI or ventricular angiography)
    - Elevated BNP at visit 1: BNP ≥ 150 pg/ml (according to local measurement).
    - OR
    - BNP ≥ 100 pg/ml (according to local measurement) and unplanned hospitalization for HF within the last 12 months prior to visit 1.
3. Patients must be treated with an ACE inhibitor at a stable dose (enalapril 10 mg daily at least or any other ACE inhibitor, e.g., ramipril, quinapril, lisinopril, fosinopril, perindopril, trandolapril; for at least 4 weeks prior to visit 1.
4. Patients must be treated with a beta blocker, unless contraindicated or not tolerated, at a stable dose for at least 4 weeks prior to visit 1 (for patients not on target dose, according to local guidelines, or in absence of that medication, the reason should be documented).
5. Written informed consent to participate in the study and ability to comply with all requirements.

Exclusion Criteria

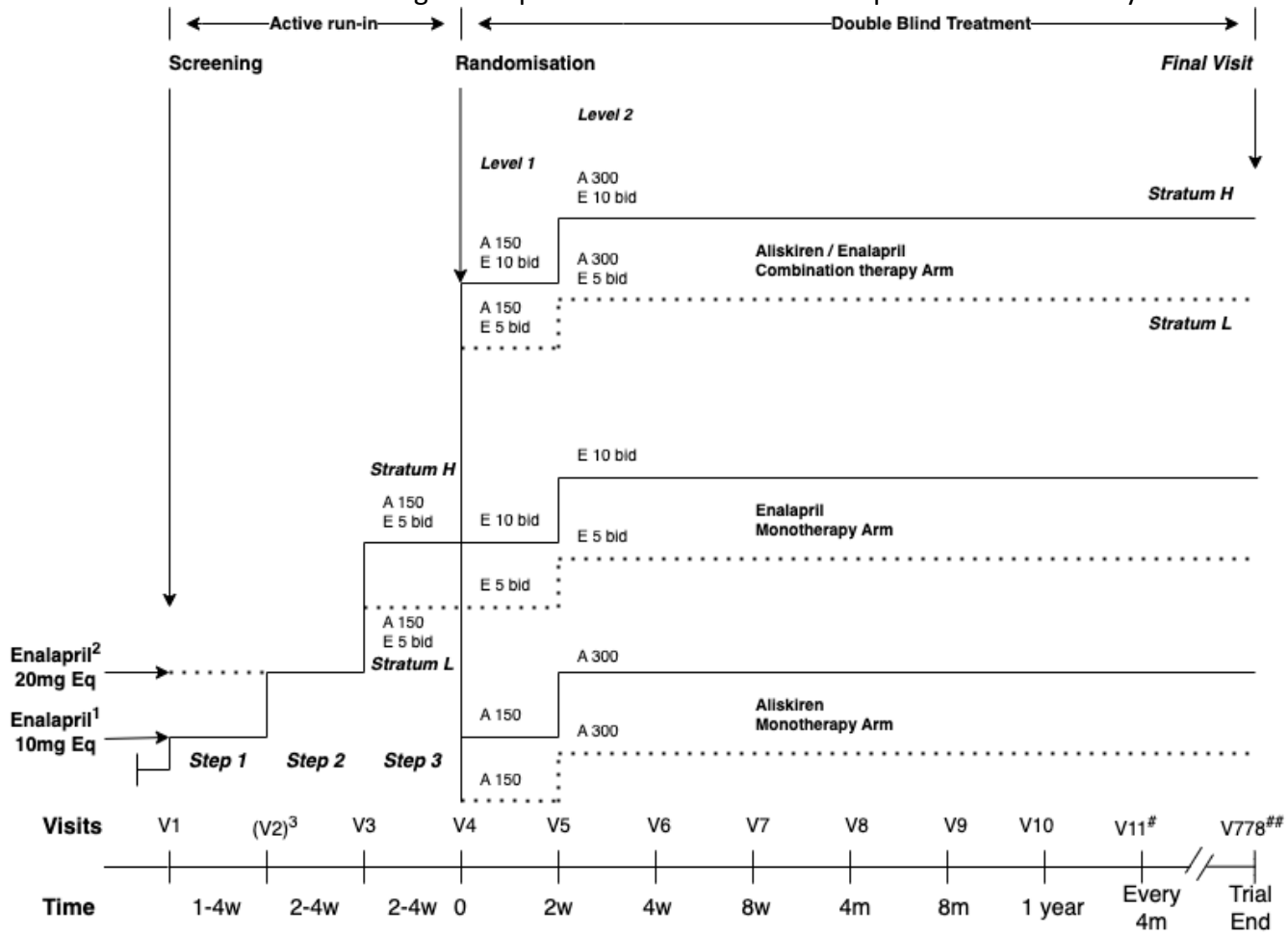
1. History of hypersensitivity to any of the study drugs including history or allergy to ACEi's as well as known or suspected contraindications to the study drugs or previous history of intolerance to high doses of ACEi's during up-titration process.
2. Patients treated concomitantly with both ARB and aldosterone antagonist in addition to study drug at visit 1.
3. Current acute decompensated HF (defined as an acute exacerbation of a chronic heart failure status manifested by typical signs and symptoms of HF like dyspnea, fatigue etc, that may require IV therapy with diuretics, vasodilators and/or inotropic drugs).
4. Symptomatic hypotension and/or less than 95 mmHg SBP at visit 1 and/or less than 90 mmHg SBP at visit 4.
5. Acute coronary syndrome, stroke, transient ischemic attack, cardiac, carotid or major vascular surgery, percutaneous coronary intervention (PCI) or carotid angioplasty, within the past 3 months prior to visit 1.
6. Coronary or carotid artery disease likely to require surgical or percutaneous intervention within the 6 months after visit 1.
7. Right heart failure due to severe pulmonary disease.
8. Diagnosis of peripartum or chemotherapy induced cardiomyopathy within the 12 months prior to visit 1.
9. Patients with a history of heart transplant or who are on a transplant list or with LVAD (left ventricular assistance device).
10. Documented ventricular arrhythmia with syncopal episodes within past 3 months, prior to visit 1, that is untreated.
11. Documented history of ventricular tachycardia or ventricular fibrillation without ICD producing significant hemodynamic consequences or considered life-threatening within the 3 months prior to visit 1.

12. Treatment with Vaughn Williams Type Ic anti-arrhythmic agents.
13. Symptomatic bradycardia, or second or third degree heart block without a pacemaker.
14. Implantation of a CRT (cardiac resynchronization therapy) device within the prior 3 months from visit 1 or intent to implant a CRT device.
15. Presence of hemodynamically significant mitral and /or aortic valve disease, except mitral regurgitation secondary to left ventricular dilatation.
16. Presence of hemodynamically significant obstructive lesions of left ventricular outflow tract, including aortic stenosis.
17. Any surgical or medical condition which might significantly alter the absorption, distribution, metabolism, or excretion of study drugs including, but not limited to, any of the following:
    - Any history of pancreatic injury, pancreatitis or evidence of impaired pancreatic function/injury as indicated by, e.g. abnormal lipase or amylase.
    - Primary liver disease considered to be life threatening.
    - Renal disease likely to be life threatening or eGFR < 40 ml/min/1.73m2 as calculated by the MDRD formula at visit 1 or eGFR < 35 ml/min/1.73m2 as calculated by the MDRD formula at visit 4 or decrease of eGFR of more than 25% from visit 1 to visit 4. (according to local laboratory measurement).
    - Current duodenal or gastric ulcers, or gastrointestinal/rectal bleeding during the 3 months prior to visit 1.
    - Current treatment with cholestyramine and colestipol resins at visit 1.
18. Serum potassium ≥ 5.0 mmol/L at visit 1 or ≥ 5.2 mmol/L at visit 4 (according to local laboratory measurement).
19. History or presence of any other diseases (i.e., including malignancies) with a life expectancy of < 5 years.
20. Current double-blind treatment in HF trials.
21. Participation in an investigational drug study at the time of enrolment or within the past 30 days or 5 half-lives of enrolment whichever is longer.
22. Any surgical or medical condition that in the opinion of the investigator would jeopardize the evaluation of efficacy or safety.
23. History of noncompliance to medical regimens and patients who are considered potentially unreliable.
24. Pregnant or nursing (lactating) women, where pregnancy is defined as the state of a female after conception and until the termination of gestation, confirmed by a positive Human Chorionic Gonadotropin (hCG) laboratory test (> 5 mIU/ml).
25. Women of child-bearing potential (WOCBP), defined as all women physiologically capable of becoming pregnant, including women whose career, lifestyle, or sexual orientation precludes intercourse with a male partner and women whose partners have been sterilized by vasectomy or other means, UNLESS they meet the following definition of post-menopausal: 12 months of natural (spontaneous) amenorrhea or 6 months of spontaneous amenorrhea with serum Follicle Stimulating Hormone (FSH) levels > 40 mIU/m or 6 weeks post-surgical bilateral oophorectomy with or without hysterectomy OR are using one or more of the following acceptable methods of contraception: surgical sterilization (e.g., bilateral tubal ligation), hormonal contraception (implantable, patch, and oral), and double-barrier methods (if accepted by local regulatory authority and ethics committee). Reliable contraception should be maintained throughout the study and for 7 days after study drug discontinuation.

26. Chronic long-term requirement for NSAIDs (high dose) or COX2 inhibitors, with the exception of aspirin at doses used for CV prophylaxis (≤ 325 mg o.d.).
27. Current treatment with cyclosporine at visit 1.
28. Treatment with any of the following drugs within the past 4 weeks prior to visit 1:
    - Direct renin inhibitor including aliskiren.
    - Intravenous vasodilators and/or intravenous inotropic drugs.

[134]

Figure 13 Study design of the ATMOSPHERE RCT including run-in period and stratification adapted from McMurray et al.



A 150: aliskiren 150mg, A 300: aliskiren 300mg, E 5 bid: enalapril 5 mg b.i.d, E 10 bid: enalapril 10 mg b.i.d. [#] Subsequent visits labelled sequentially from 11. [##]Visit 778: final visit per patient scheduled upon decision to end study. [1]Patients receiving a dose of ACEi of enalapril 10mg daily or any other ACE inhibitor with an equivalent dose with start at step 2, patients starting at step 1 will be up titrated to maximum recommended dose of enalapril at visit 2 where applicable. [2]Patients receiving a dose of ACEi of enalapril 20mg or any other ACE inhibitor with an equivalent dose will start immediately on step 2 and skip visit 2. 3 Visit 2 is only applicable for patients starting with enalapril 10mg daily at V1.

[134]

Chapter 4 Joint Modelling of NT-ProBNP and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure

4.1 Introduction

Heart failure is a clinical syndrome with well-documented interrelationships between patient characteristics and outcomes, including biomarkers and other physical characteristics. Given the rise in awareness and use of joint modelling within heart failure, it is informative to understand how joint modelling can be used for prognostic model development in heart failure. Joint models allow for covariates with repeated measurements, while accounting for correlation and correcting for measurement error, as well as allowing for dynamic and subject-specific predictions [51], [52]. Clinical trials within heart failure often collect repeat measurements of biomarkers such as NT-ProBNP as well as time-to-event data used in survival analysis, making them ideal for joint modelling. Currently, there are no identified prognostic models within the heart failure literature using joint modelling. Therefore, the aim of this research is to illustrate and critically appraise how joint modelling can be used for prognostic model development in heart failure. Further, using data from clinical trials, I will compare the performance of prognostic models that employ joint modelling with models that use current recommended approaches.

The main objective of this research is to ascertain what value prognostic models using joint models can provide over current recommended approaches.

4.2 Methods

4.2.1 Data Source

As previously mentioned in Chapter 3, data was obtained from the PARADIGM-HF trial in which 8,399 patients were randomised to receive either LCZ696 at a dose of 200mg twice daily or enalapril at a dose of 10mg twice daily at a 1:1 ratio in a double-blind fashion.

4.2.2 Statistical Analysis

To address the aims of this research, specifically how joint models can be applied to heart failure with the possibility of providing better prognostic models, data were analysed, processed and cleaned prior to the fitting of joint models. For variable selection, a clinical approach using prior clinical knowledge and literature, was preferred over a stepwise approach. The dangers of a stepwise approach are well known. For example, they could lead to a known clinically important variable to be 'unselected' if statistical power is low or, alternatively, lead to overfitting [141]. While a sample size calculation is available for joint models, the formula is not suitable for this application as it cannot account for an interaction between treatment and time [138]. All joint models were fitted using the R package *JMbayes2*; while this package is still in development, it is based on the previously available R packages from the same author (*JM* and *JMbayes*) [67], [68], [114]. The *JMbayes2* package uses a fully parametric Bayesian approach using Markov chain Monte Carlo (MCMC) sampling to fit joint models of longitudinal and time-to-event outcomes. While this package is still in development, it was chosen over the previous packages as it allows for better parameterisations of the longitudinal model within the survival model, for example the area parameter as discussed in Section 1.3.8, as well as calibration plots which are key performance indicators for prognostic models [39], [40]. The basic joint model is comprised

of a longitudinal component and a time-to-event (survival) component (often referred to as sub models), which are first fit and then supplied to the '*jm*' function within the *JMbayes2* package to produce a joint model.

4.2.3 Data Cleaning and Processing

As stated, data were analysed and cleaned prior to analysis; this included analysis of variables and their distributions, recoding of categorical variables to factors and recoding of the original composite event variable so that 1 indicated an event and not whether the patient was censored. In addition, the various functions to fit the LME, Cox PH and joint models require data to be in a specified format(s), this meant ensuring the same group of patients were given to functions for the LME and Cox PH models and merging baseline covariates with the repeated measurements for the LME. Additionally, the new data argument for the various performance metrics required all covariates to be present with repeat measurements in a singular dataset.

4.2.4 Covariate Selection

Both the longitudinal and survival components used in a joint model allow for covariates; for both, clinically relevant covariates were selected. These covariates were selected based on prior research and clinical knowledge, both as described in Chapter 1 and in conjunction with my clinical supervisor. Neither the PARADIGM nor the ATMOSPHERE trials were missing any covariates which were identified as being clinically relevant by prior research and clinical knowledge.

4.2.5 Missing Data

While joint models allow for uneven and missing longitudinal measurements as explained previously in Chapter 2. All covariates for both sub models must be complete. Data were incomplete for 66 patients; due to this being less than 1% of the data available, a complete case analysis was chosen. Excluding those 66 patients with missing data.

4.2.6 Longitudinal Sub Model

JMBayes2 requires the longitudinal sub model to be either a Generalised Linear Mixed Model (GLMM) fitted by the function 'mixed_model' from the 'GLMMadaptive' package or a Linear Mixed Effects Model (LME) fitted by the function 'lme' from the 'nlme' package. Because this analysis is focused on a continuous outcome, an LME was required. An LME contains a response variable and allows for fixed and random effects.

4.2.6.1 NT-ProBNP

The response variable for the LME was NT-ProBNP; NT-ProBNP is known to be associated with clinical outcomes in heart failure research and has previously been used in prognostic models [36], [42]. As values of NT-ProBNP are known to have a wide range and were observed to be skewed, a natural logarithm transformation was applied to the NT-ProBNP values. As stated in Section 1.3, joint models make assumptions around normality in the response variable specifically with respect to the random effects therefore necessitating this transformation [52].

At least one measurement of NT-ProBNP was available for all patients who underwent randomisation; however, because the trial included a run-in period and due to the trial design, 6,218 (75%) patients only had a value of NT-ProBNP at the screening visit, because the screening visit is before both run-in and randomisation (the start point for the LME) these values would not otherwise be included, in the interest of maximising the use of data the decision to include the values for these 6218 patients and recode them as month zero (the start of the LME).

4.2.6.2 Time (Fixed Effect)

LMEs allow for the uneven spacing (across time) of values of the response variable, and because the sample date of the value of NT-ProBNP was available, time was modelled in the form of months since randomisation. Calculated from the difference between the sample date and the randomisation date, using a 28-day calendar month. Sample date was preferred over scheduled visit date to provide a more accurate model; taking advantage of the ability to use uneven measurements across time usage of specimen time allow accurate estimation of the true value of log NT-ProBNP by the LME, which in turn provides a better estimation of the longitudinal trajectory of NT-ProBNP for the joint models.

The association of the values of the natural logarithm-transformed NT-ProBNP, hereby known as log NT-ProBNP, and time were observed to be non-linear; therefore, the decision was made to include time non-linearly using restricted natural cubic splines; this was achieved using the '*ns*' function of the '*splines*' package. Due to computational difficulties with convergence, likely due to the imbalance of data at the start of the study, the boundary knots of the spline were manually specified using quantile values.

4.2.6.3 Covariates (Remaining Fixed Effects)

As previously stated, covariates were specified using prior clinical knowledge, these included age, atrial fibrillation, BMI and treatment effect. Typically, because of randomisation, treatment effect is only modelled as the effect of treatment over time [69]. However, to account for the use of screening values, the decision was made to include both the main effect of treatment and the effect of treatment over time. An intercept term was included by default in the LME, this intercept is the mean predicted value of NT-ProBNP when all other variables are zero.

4.2.6.4 Random Effects

The LME was fit using both random intercepts and random slopes as random effects. The random slopes included non-linear effect of time using the same restricted natural cubic splines described earlier.

4.2.6.5 Model Formulation

Equation 13 denotes the formulation of the LME sub model.

Equation 13 Formulation of the LME sub model

The intercept term of the model is represented by $\boldsymbol{\beta_0}$, with the fixed effects being represented by $\boldsymbol{\beta_n}$ and the

$$\begin{cases} \log NTProBNP_{ij} = \beta_0 + \beta_1 f_1(Specimen\ Time_{ij}) + \beta_2 f_2(Specimen\ Time_{ij}) + \beta_3 f_3(Specimen\ Time_{ij}) \\ + \beta_4 Treatment\ LCZ_{ij} \\ + \beta_5 f_1(Specimen\ Time_{ij})\ Treatment\ LCZ_{ij} + \beta_6 f_2(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} + \beta_7 f_3(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} \\ + \beta_8 Age_{ij} + \beta_9 Atrial\ Fibrillation\ Yes_{ij} + \beta_{10} BMI_{ij} \\ + b_{i0} + b_{i1} f_1 Specimen\ Time_{ij} + b_{i2} f_1 Specimen\ Time_{ij} + b_{i3} f_1 Specimen\ Time_{ij} + \varepsilon_{ij}, \\ b_i \sim \mathrm{N}(0, D),\ \varepsilon_{ij} \sim N(0, \sigma^2) \end{cases}$$

function $\boldsymbol{f_n}$ representing the natural cubic spline. The random effects consisting of the random intercept $\boldsymbol{b_{i0}}$ and random slopes $\boldsymbol{b_{in} f_n}$ (which incorporates the natural cubic spline) are assumed to be normally distributed with a mean of $\mathbf{0}$ and a variance co-variance matrix $\boldsymbol{D}$. The error term represented by is assumed to be normally distributed with a variance of $\boldsymbol{\sigma^2}$.

4.2.6.6 Model Fitting

To fit the LME the formatted data was passed to the *LME* function, specifying the response variable as log NT-ProBNP, the time variable using the *ns* function to specify a restricted cubic natural spline, the interaction of the natural cubic spline and treatment, as well as the other covariates using the *fixed* argument. The random effects were specified as random intercepts and slope using the *random* argument. To aid in convergence the optimiser was manually specified as *optim*, which is a general-purpose optimization method based on Nelder–Mead, quasi-Newton and conjugate-gradient algorithms, this is known from experience to aid in convergence over the default optimisation method and is specified using the *lmeControl* argument [139].

4.2.7 Survival Sub Model

The survival model for the *JMbayes2* package needs to be a survival model fitted using either the 'survreg' or 'coxph' functions of the survival package, the former allowing for parametric survival models and the latter allowing for Cox PH models. To avoid incorrect specification

of the distribution, a Cox PH model was preferred over a parametric model. This decision is supported by the literature, as illustrated in Chapter 2 where 82% of studies using joint modelling in heart failure opted to use a Cox PH model as their survival sub model.

4.2.7.1 Covariates

As previously stated, covariates were selected on prior clinical knowledge, and in the case of the survival sub model, the covariates were the same as the original trial in its survival analysis. These covariates included treatment, age, sex, region, BMI, eGFR, ejection fraction, NYHA classification, whether the patient had diabetes, SBP, heart rate, whether the patient had a history of atrial fibrillation, hospitalisation for heart failure, myocardial infarction, or stroke.

4.2.7.2 Event (End Point)

The primary outcome of the original trial was used for the event of interest in the survival sub model; this was a composite of first hospitalisation for heart failure and death from cardiovascular causes.

4.2.7.3 Model Formulation

Equation 14 denotes the formulation of the survival sub model.

Equation 14 Formulation of the Survival Sub Model

$$h_i(t) = h_0(t) \exp\,(\,\gamma_1 Treatment\ LCZ_i + \gamma_2 Age_i + \gamma_3 Male_i + \gamma_4 Region\ Latin\ America_i + \gamma_5 Region\ North\ America_i + \gamma_6 Region\ Asia, Pacific\ and\ Other_{\ i} + \gamma_7 Region\ Western\ Europe_i + \gamma_8 BMI_i + \gamma_9 eGFR_i + \gamma_{10} Ejection\ Fraction_i + \gamma_{11} NYHA\ Class\ II_i + \gamma_{12} NYHA\ Class\ III_i + \gamma_{13} NYHA\ Class\ IV_i + \gamma_{14} Diabetes\ Yes_i + \gamma_{15} SBP_i + \gamma_{16} Heart\ Rate_i + \gamma_{17} Atrial\ Fibrillation\ Yes_i + \gamma_{18} Ischemic\ Heart\ Failure\ Yes_i + \gamma_{19} Prior\ Hospitalisation\ for\ Heart\ Failure\ Yes_i + \gamma_{20} Prior\ Miocardial\ Infarction\ Yes_i + \gamma_{21} Prior\ Stroke\ Yes_i)$$

$h_i(t)$ represents the hazard of the composite event at time point t. With $h_0(t)$ representing the baseline hazard and γ_n representing the covariates.

4.2.7.4 Model Fitting

To fit the Cox PH model the cleaned and formatted data was passed to the *coxph* function from the *survival* package using the *data* argument. The model was specified in the format of the month of when the event occurred, or the patient was censored, and a status indicator of whether the event occurred; as well as the other previously specified covariates using the *formula* argument.

4.2.8 Joint Models

As joint modelling allows for multiple alpha parameterisations, three joint models were fit to explore how best to represent the association between the biomarker and the survival outcome. Described in section 1.3.6, and chosen for their suitability for a prognostic model; these models included the previously specified sub models, the LME and Cox PH models. The first joint model (Joint Model 1) included a value parameterisation; this measures the association between the value of log NT-ProBNP and the composite endpoint (event). The second joint model (Joint Model 2) included both value and slope parameterisations measuring the association of both the value of log NT-ProBNP and the slope, i.e., rate of change of NT-ProBNP and the composite endpoint. The third and final model (Joint Model

3) included the area parameterisation; this measures the association of the area under the longitudinal profile of NT-ProBNP and the composite endpoint. All models were fit using the *jm* function from the package *JMbayes2,* which fits a joint model with a piecewise hazard function using quadratic B-splines with ten hazard segments.

The formulae for the joint models are shown in Table 22, As all the models are only differentiated by their alpha parameters the models are represented by the base model and their individual alpha parameters.

4.2.8.1 Model fitting

To fit the joint models, the fitted LME and Cox PH models were passed to the *jm* function, using the *Mixed_objects* and Surv_*object* arguments respectively; along with the type of parameter for the model specified in the *functional_forms* argument.

Table 22 Formulae for Joint Models

Joint Model 1 (Value)	$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_{i(t)} \\ = \beta_0 + \beta_1 f_1(\text{Specimen Time}) + \beta_2 f_2(\text{Specimen Time}) + \beta_3 f_3(\text{Specimen Time}) \\ + \beta_4 \text{Treatment LCZ} \\ + \beta_5 f_1(\text{Specimen Time})\text{Treatment LCZ} + \beta_6 f_2(\text{Specimen Time})\text{Treatment LCZ} + \beta_7 f_3(\text{Specimen Time})\text{Treatment LCZ} \\ + \beta_8 \text{Age} + \beta_9 \text{Atrial Fibrillation Yes} + \beta_{10} \text{BMI} \\ + b_{i0} + b_{i1} f_1 \text{Specimen Time} + b_{i2} f_1 \text{Specimen Time} + b_{i3} f_1 \text{Specimen Time} + \varepsilon_i, \quad \varepsilon_i(t) \sim N(0, \sigma^2), \\ h_i(t) = h_0(t)\, exp\left\{ \begin{array}{c} \gamma_1 \text{Treatment LCZ}_i + \gamma_2 \text{Age}_i + \gamma_3 \text{Male}_i \\ + \gamma_4 \text{Region Latin America}_i + \gamma_5 \text{Region North America}_i + \gamma_6 \text{Region Asia, Pacific and Other}_i \\ + \gamma_7 \text{Region Western Europe}_i + \gamma_8 \text{BMI}_i + \gamma_9 \text{eGFR}_i + \gamma_{10} \text{Ejection Fraction}_i \\ + \gamma_{11} \text{NYHA Class II}_i + \gamma_{12} \text{NYHA Class III}_i + \gamma_{13} \text{NYHA Class IV}_i \\ + \gamma_{14} \text{Diabetes Yes}_i + \gamma_{15} \text{SBP}_i + \gamma_{16} \text{Heart Rate}_i + \gamma_{17} \text{Atrial Fibrillation Yes}_i \\ + \gamma_{18} \text{Ischemic Heart Failure Yes}_i + \gamma_{19} \text{Prior Hospitalisation for Heart Failure Yes}_i \\ + \gamma_{20} \text{Prior Miocardial Infarction Yes}_i + \gamma_{21} \text{Prior Stroke Yes}_i + \alpha m_i(t) \\ + \{\text{Alpha Parameter}(s)\} \end{array} \right\}, \\ log\, h_0(t) = \gamma h_0, 0 + \sum_{q=0}^{Q} \gamma h_{0,q} B_q(t, v), \end{cases}$$
Joint Model 1 Alpha Parameter (Value)	$\alpha m_i(t)$
Joint Model 2 Alpha Parameter (Value + Slope)	$\alpha m_i(t) + \alpha_2 m_i'(t)$
Joint Model 3 Alpha Parameter (Area)	$\alpha \frac{\int_0^t m_i(s)\, ds}{t}$

The longitudinal outcome at time point t is represented by $y_i(t)$, and comprises of the true and unobserved value of log NT-ProBNP at time point t $\left(m_i(t)\right)$, and the error term $\varepsilon_{i(t)}$. This error term is assumed normally distributed with a mean of 0 and a variance of σ^2. The intercept term for the longitudinal outcome is represented by β_0. The remaining β parameters represent the coefficients of the covariates, where the natural cubic spline is represented by the function f_n. b_{in} represent the random effects. The hazard of the composite outcome at time point t is represented by $h_i(t)$. The baseline hazard function of the model is represented by $h_0(t)$ which is comprised of a piecewise hazard function, which includes quadratic B-splines and 10 baseline hazard segments. With $B_q(t, v)$ relating to the q-th basis function of the B-spline with knots v_1- v_{10}, and a vector of spline coefficients (γh_0). The remaining γ_n parameters being the covariates of the survival component. The slope parameter $m_i'(t)$ corresponds to the rate of change of the longitudinal outcome at time point t and is estimated using the derivative of the fixed and random effects of the longitudinal outcome with respect to time. The area alpha parameter represented by a function corresponds to the area under the whole longitudinal trajectory accounting for the observation period.

4.2.9 Model Performance: Prognostic Accuracy, Fit and Calibration

The *JMbayes2* package provides measures of performance including time-varying receiver operating characteristics (ROCs) using the function *tvROC*, time-varying AUCs using the function *tvAUC*, a time-varying Brier score using the function *tvBrier* and time-varying calibration curves using the function *calibration_plot*. In the current context these time-varying ROCs and AUCs are an extension of traditional ROCs and AUCs and provide a metric for discrimination given information on the longitudinal marker up until a specified time point and predicting the survival outcomes at a given, later time point [140], [141]. Similar to this, both the time-varying Brier score and time-varying calibration curves extend on the traditional Brier score and calibration curve in a similar way in order to provide accuracy and calibration measures. These measures were used to assess the joint models for prognostic accuracy and fit. For all measures, two time points were selected based on commonly used time points in heart failure prognostic models; these were at month 12 (one year), using only longitudinal data at baseline (time 0), and month 24 (two years), using longitudinal data up until month 12 (one year). The joint models were also compared against each other using these measures along with the deviance information criterion (DIC), log pseudo-marginal likelihood (LPML) for both marginal and conditional formulations, as well as hazard ratios and corresponding (CIs). The DIC can be interpreted like any other information criterion in that a smaller value indicates a better fit. The LPML can be interpreted like the log likelihood of a frequentist model in that a larger value indicates a better fit. For both DIC and LPML, the marginal value relates to fit for the overall population, whereas the conditional value relates to the subject-specific or individual fit.

4.2.10 External Validation

As previously mentioned in Chapter 3, external validation was performed using data obtained from the ATMOSPHERE trial, which randomised 7,016 patients to either enalapril at a dose of 5 or 10 mg twice a day, aliskiren at a dose of 150mg once a day or a combination of enalapril (at a dose of 5 or 10mg twice daily) and aliskiren (at a dose of 150mg once a day) at a 1:1:1 ratio in a double-blind, double-dummy fashion.

Data were recoded to the same format as the data from the PARADIGM trial, including the transformation of NT-ProBNP using a natural logarithm, use of the same covariates, same, endpoint, and replicating the sample month format (time variable) for date of log NT-ProBNP measurement. Treatment was recoded as enalapril, which included the enalapril arm, and LCZ, which included the other two arms.

The time-varying accuracy and fit measures described previously allow the use of new data to validate the joint models. The formatted data from the ATMOSPHERE trial were provided to the time-varying AUC, time-varying ROC, time-varying Brier score and time-varying calibration curve functions using the *newdata* argument. This argument allows the predictions from the joint model to be made on new data in this case the ATMOSPHERE data. Following this, the measures were computed at the same time points as used for the previous model fit and accuracy measures. Finally, these measures were then assessed and compared with the model fit and accuracy measure stated previously.

4.2.11 Comparative Analysis

To illustrate the differences in the joint models and current recommended prognostic models, the joint models were compared against Cox PH models - a landmark Cox PH model, which included the same covariates at the survival sub model and included the last available value of NT-ProBNP (LOCF), and an extended (time-varying) Cox PH model using these same covariates and including NT-ProBNP as a time-varying covariate. These models were fit using the coxph function from the *survival* package, with the LOCF model using only the last observation of log NT-ProBNP, and the time-varying Cox PH model using the time-varying measurements of log NT-ProBNP. A time-varying Cox PH model requires the data to be split into time periods based on the sample dates of the measurements of log NT-ProBNP, with the status of an event reflecting whether the event occurred within that period of time.

The *JMbayes2* package only allows a limited number of methods for accuracy and performance measures. For this reason, the decision was made to compare the models using the hazard ratios and confidence intervals (CIs) and the use of time-varying ROCs and AUCs under the formulation defined by Heagerty *et al*. and implemented using the *risksetROC* function from the *risksetROC* package in R [142], [143]. For both Cox PH models, time-varying ROCs and corresponding AUCs were obtained at 12 and 24 months.

For descriptive statistics, continuous variables are summarised with medians and quartiles [Q1, Q3] and the distribution of categorical variables expressed as percentages. Association parameters from joint models are expressed as hazards ratios and corresponding 95% (CIs). A time-dependent association parameter is the hazard of the composite event per one unit increase in log NT-ProBNP at any time point. A time-dependent slope association parameter is the hazard of the composite event per one unit increase in the slope of NT-ProBNP at any

time point. A time-dependent area association parameter is the hazard of the composite event per one unit increase of the area of the longitudinal profile of log NT-ProBNP at any time point.

P-Values less than 0.05 will be considered statistically significant. All statistical analysis was performed using R Version 4.0 [144] and *JMbayes* package version 0.1-81 [114].

4.3 Results

4.3.1 Baseline Characteristics

Table 23 shows the baseline characteristics of included patients (N = 8333), with a median age of 64 years, 22% of patients being female and the majority of patients being enrolled within Central and Western Europe. As reported in the main trial, there were only minimal differences in distributions of baseline characteristics between the arms of the trial.

Table 23 Baseline Characteristics of Included Patients

Characteristic	LCZ (N=4154)	Enalapril (N=4179)	Overall (N=8333)
Age, years			
Median [Q1, Q3]	64.0 [57.0, 72.0]	64.0 [57.0, 72.0]	64.0 [57.0, 72.0]
Sex			
Female	874 (21.0%)	949 (22.7%)	1823 (21.9%)
Male	3280 (79.0%)	3230 (77.3%)	6510 (78.1%)
Region			
Central Europe	1388 (33.4%)	1425 (34.1%)	2813 (33.8%)
Latin America	705 (17.0%)	713 (17.1%)	1418 (17.0%)
North American	307 (7.4%)	288 (6.9%)	595 (7.1%)
Pacific Asia/Pacific and Other	743 (17.9%)	736 (17.6%)	1479 (17.7%)
Western Europe	1011 (24.3%)	1017 (24.3%)	2028 (24.3%)
BMI kg/m^2			
Median [Q1, Q3]	27.5 [24.4, 31.2]	27.5 [24.5, 31.2]	27.5 [24.4, 31.2]
eGFR mL/min/1.73 m^2			
Median [Q1, Q3]	66.0 [54.0, 79.0]	66.0 [54.0, 79.0]	66.0 [54.0, 79.0]
Ejection Fraction %			
Median [Q1, Q3]	30.0 [25.0, 34.0]	30.0 [25.0, 34.4]	30.0 [25.0, 34.1]
NYHA Class			
Class 1	180 (4.3%)	207 (5.0%)	387 (4.6%)
Class 2	2974 (71.6%)	2903 (69.5%)	5877 (70.5%)
Class 3	967 (23.3%)	1042 (24.9%)	2009 (24.1%)
Class 4	33 (0.8%)	27 (0.6%)	60 (0.7%)
Diabetes			
No	2714 (65.3%)	2731 (65.4%)	5445 (65.3%)
Yes	1440 (34.7%)	1448 (34.6%)	2888 (34.7%)
SBP mmHg			
Median [Q1, Q3]	120 [110, 130]	120 [110, 130]	120 [110, 130]
Heart Rate beats per minute			
Median [Q1, Q3]	71.0 [64.0, 80.0]	72.0 [64.0, 80.0]	71.0 [64.0, 80.0]
Prior History of Atrial Fibrillation			
No	2652 (63.8%)	2617 (62.6%)	5269 (63.2%)
Yes	1502 (36.2%)	1562 (37.4%)	3064 (36.8%)
Ischemic Heart Failure			
No	1668 (40.2%)	1674 (40.1%)	3342 (40.1%)
Yes	2486 (59.8%)	2505 (59.9%)	4991 (59.9%)

Characteristic	LCZ (N=4154)	Enalapril (N=4179)	Overall (N=8333)
Prior History of Hospitalisation for Heart Failure			
No	1567 (37.7%)	1533 (36.7%)	3100 (37.2%)
Yes	2587 (62.3%)	2646 (63.3%)	5233 (62.8%)
Prior History of Myocardial Infarction			
No	2353 (56.6%)	2379 (56.9%)	4732 (56.8%)
Yes	1801 (43.4%)	1800 (43.1%)	3601 (43.2%)
Prior History of Stroke			
No	3800 (91.5%)	3811 (91.2%)	7611 (91.3%)
Yes	354 (8.5%)	368 (8.8%)	722 (8.7%)
NT-ProBNP pg/mL			
Median [Q1, Q3]	1420 [779, 2880]	1450 [790, 2940]	1440 [784, 2910]

4.3.2 Number of Measurements of NT-ProBNP

Table 24 shows the number of repeated measurements of NT-ProBNP, with the majority of patients (76%) having a singular measurement and the remaining patients (24%) having two or more measurements.

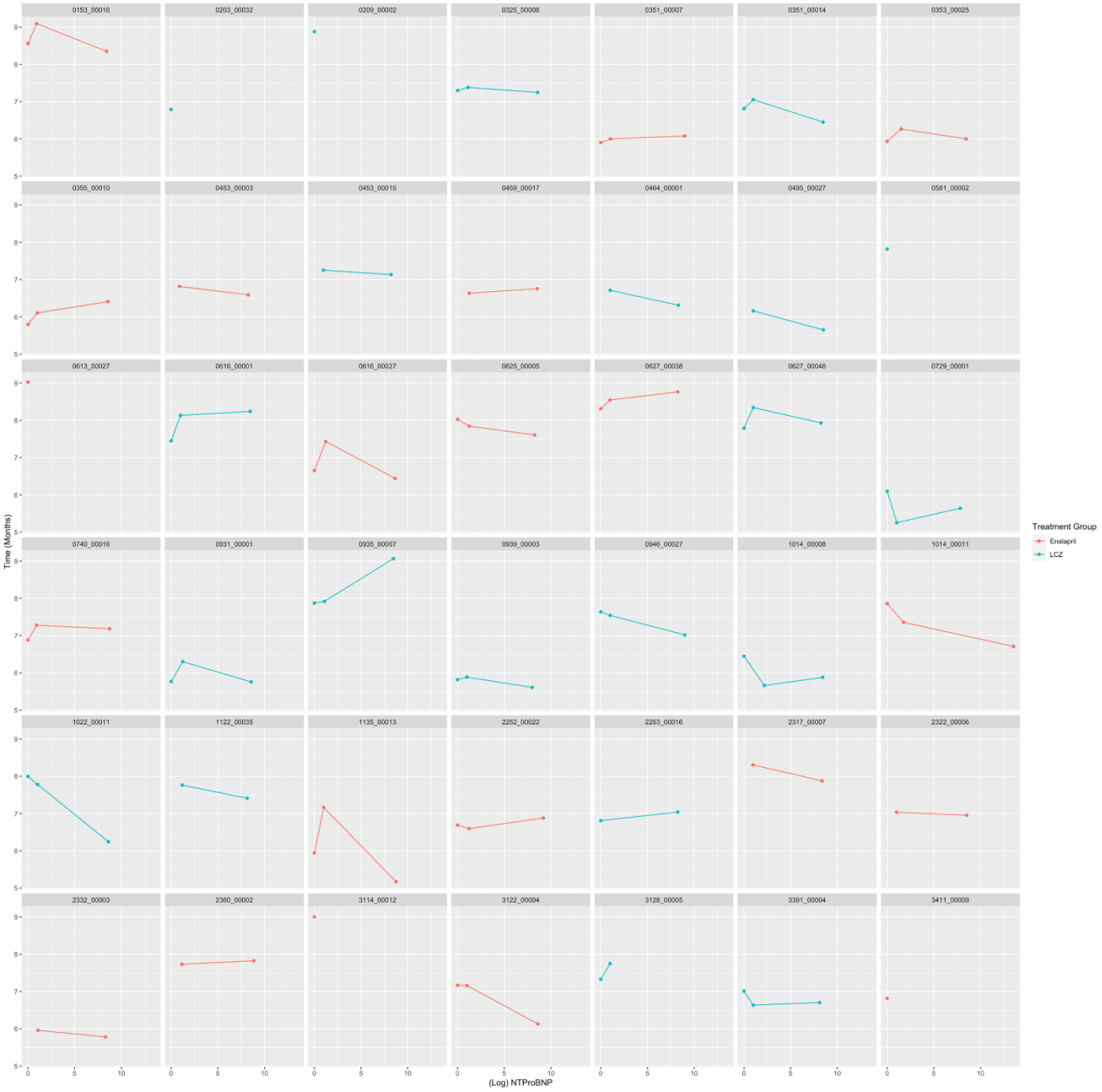
Table 24 Number of Repeat Measurements of NT-ProBNP for Included Patients

1	2	3	4
6336 (76%)	882 (11%)	1109 (13%)	6 (<1%)

4.3.3 Longitudinal Profile of NT-ProBNP

Figure 14 shows the longitudinal profile of log NT-ProBNP of 42 randomly sampled patients (stratified by number of measurements), illustrating some patients (e.g., 1135_00013) who showed non-linearity of NT-ProBNP over time, and a wide range of log NT-ProBNP taking into consideration the transformation into the log scale. With the log scale in mind, the non-linearity would be amplified on the original scale.

Figure 14 Longitudinal Profile of log NT-ProBNP for 42 Randomly Sampled Patients



4.3.4 Longitudinal Outcome from Joint Models

Table 25 shows the parameter estimates from the longitudinal outcome of the joint models for all meaningfully interpretable variables; because natural splines are not interpretable as coefficients, both the natural spline function of sample month (time) and the interaction of treatment with the natural spline function have been excluded. All variables were considered statistically significant, with treatment of LCZ and BMI having a negative effect on log NT-ProBNP, and having a prior history of atrial fibrillation and age having a positive effect on log NT-ProBNP. While statistically significant, both age and BMI show only a small effect on log NT-ProBNP, however as NT-ProBNP is represented on a log scale and time is in months, scale may be an influencing factor on effect size. There are only small differences in the variables between the models; for the main effect of treatment with LCZ the coefficient ranged from 0.220 to 0.235 showing only a maximum point difference of 0.015. The coefficient for age did not show a point difference between models, however the 95% CIs showed a 0.001 between Joint Model 1 and Joint Models 2 and Joint Model 3, this is reflected by the minimally higher differences in P-Values with Joint Model 1 and Joint Model 3 having a higher P-Value of 0.003 and 0.002 respectively when compared to Joint Model 2 with a P-Value of 0.001. The coefficient for patients with atrial fibrillation shows only a difference in 0.001 between Joint Model 3 and Joint Models 1 and Joint Model 3, with Joint Model 1 having a slightly higher P-Value of 0.007 compared to 0.001 of Joint Model 1 and Joint Model 3. The coefficient for BMI shows no difference in point estimate or P-Value between the joint models and only a small difference in 95% CIs between the models. These differences suggest that there may only be small differences between the joint model for the longitudinal outcome, but that the confidence and significance may vary, however scale should be taken into consideration when interpreting these coefficients.

While the coefficients of the natural splines can't be meaningfully interpreted because they only represent parts of a function, Figure 15 illustrates the predicted average trajectories for each treatment by each joint model. These trajectories were predicted using the median and mode characteristics for each stratum of patients, i.e., by treatment assignment, illustrating the change of log NT-ProBNP over time. All models show similar trajectories, with Joint Model 2 having overlapping 95% CIs past month eight. On average, the figures suggest that the LCZ treatment group had lower levels of NT-ProBNP, and each treatment group had a change in mean log NT-ProBNP around month three, with the enalapril group showing an increase in the mean and the LCZ group showing a decrease.

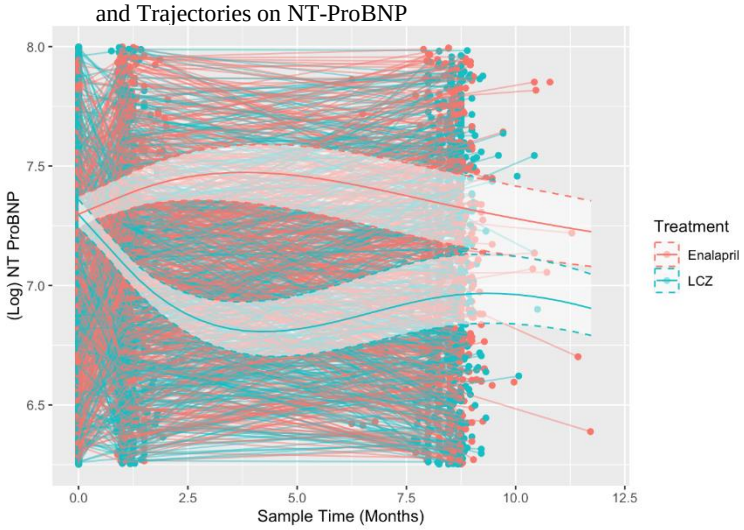
Table 25 Parameter estimates (95% CI) from the Longitudinal Outcome of the Joint Models for meaningfully interpretable variables.

	Joint Model 1 Value		Joint Model 2 Value + Slope		Joint Model 3 Area	
Variable	**Estimate (95% CI)**	**P-Value**	**Estimate (95% CI)**	**P-Value**	**Estimate (95% CI)**	**P-Value**
Treatment - LCZ	-0.226 (-0.321 – -0.014)	< 0.001	-0.220 (-0.314 – -0.134)	< 0.001	-0.235 (-0.310 – -0.156)	< 0.001
Age	0.004 (0.001 – 0.007)	0.003	0.004 (0.001 – -0.006)	0.001	0.004 (0.001 – 0.006)	0.002
Atrial Fibrillation - Yes	0.253 (0.192 – 0.313)	0.007	0.253 (0.195 – 0.311)	< 0.001	0.254 (0.194 – 0.310)	< 0.001
BMI	-0.043 (-0.048 – -0.037)	< 0.001	-0.043 (-0.048 – -0.038)	< 0.001	-0.043 (- 0.048 – 0.038)	< 0.001

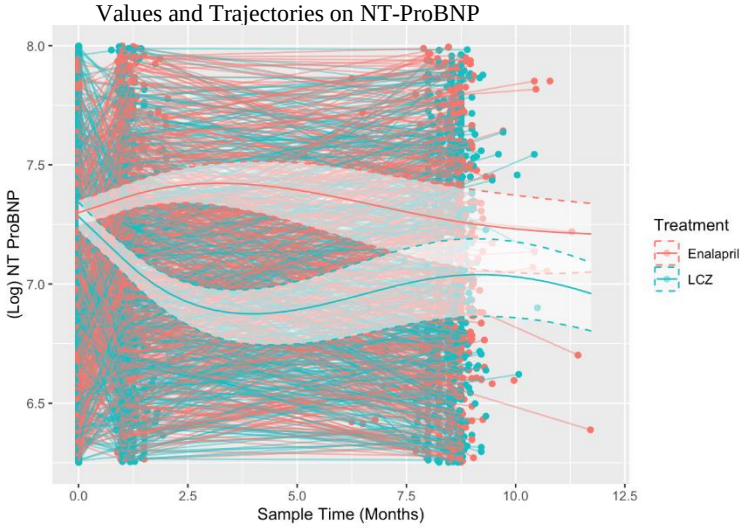
The parameter estimates for both the natural splines of the sample month and the interaction of the natural splines with treatment are excluded as the values cannot be meaningfully interpreted but are instead illustrated in Figure 15.

Figure 15 Average Trajectories of log NT-ProBNP by Treatment Group with Observed Values and Trajectories of NT-ProBNP

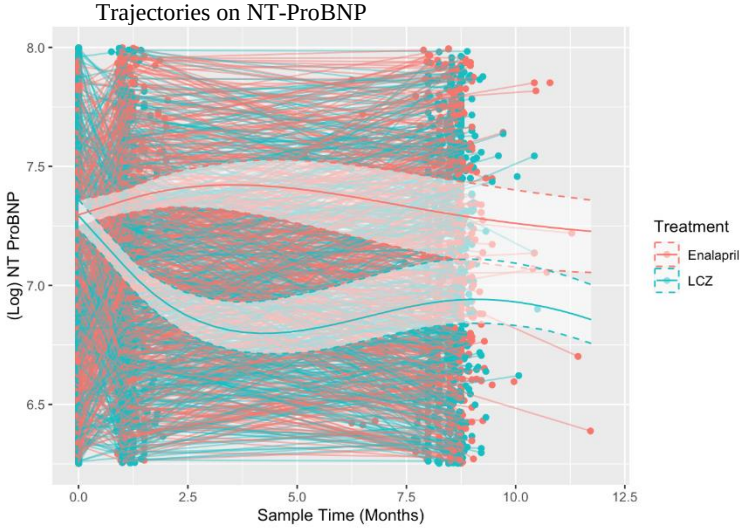
A.) Average Trajectories of log NT-ProBNP by Treatment Group from Joint Model 1 (Value) with Observed Values and Trajectories on NT-ProBNP



B.) Average Trajectories of log NT-ProBNP by Treatment Group from Joint Model 2 (Value + Slope) with Observed Values and Trajectories on NT-ProBNP



C.) Average Trajectories of log NT-ProBNP by Treatment Group from Joint Model 3 (Area) with Observed Values and Trajectories on NT-ProBNP



Average trajectories were predicted by the respective joint models, using the mode and median characteristics of each stratum of patients (treatment group).

4.3.5 Survival Outcomes from Joint Model

Table 26 shows the hazard ratios with 95% (CIs) and p-values for the different covariates in the three joint models, illustrating only minor differences. One covariate of note is that of the treatment effect on the survival outcome. With a range of 1.01-1.06 suggesting a limited treatment effect on survival. The survival treatment effect is also considered statistically insignificant in each of the models. This is not to say that there is no treatment effect, as this treatment effect has already been illustrated in the main trial. Instead, it suggests that the treatment effect is captured within the longitudinal process and that the parameterisation of the longitudinal process, i.e., the α or association parameter is a mediator for the treatment effect. Indeed, the association parameter estimates suggest an association between log NT-ProBNP and the composite outcome (first hospitalisation for heart failure and death from cardiovascular causes). The association parameter estimate for Joint Model 1 (Value) suggests that per unit increase of log NT-ProBNP, the hazard of the composite event almost doubles, with a HR of 1.93 (1.73 -2.13). When the model is adjusted for both slope and values, as in Joint Model 2, the HR attenuates slightly, with a HR of 1.8 (1.44 - 2.05), suggesting a 1.8 times increase in hazard of the composite endpoint per unit increase of log NT-ProBNP for patients who had the same slopes. The slope parameter estimate of this model is not considered statistically significant. The association parameter estimate of Joint Model 3 (Area) is similar to Joint Model 1 (Value), with a HR of 1.9 (1.76 - 2.05), suggesting that per unit increase in the area under the trajectory of log NT-ProBNP, the hazard of the composite endpoint increases by 1.9 times.

Table 26 Hazard Ratios (95% CI) for Variables in the Three Joint Models

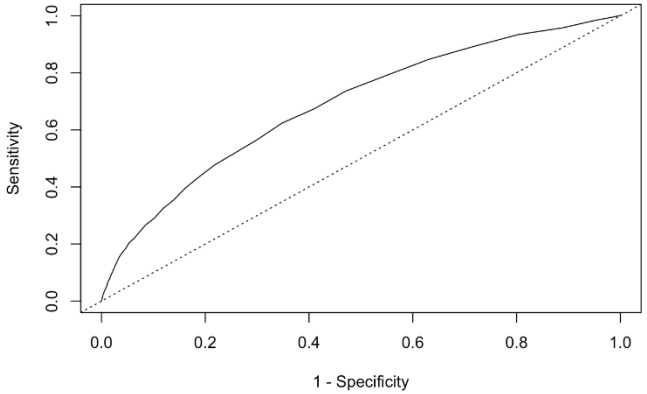
	Joint Model 1 - Value		Joint Model 2 – Value + Slope		Joint Model 3 - Area	
Variable	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**
Treatment - LCZ	1.01 (0.84 - 1.23)	0.934	1.04 (0.87 - 1.28)	0.728	1.06 (0.87 - 1.25)	0.514
Age	1 (1 - 1.01)	0.956	1 (1 - 1.01)	0.894	1 (1 - 1.01)	0.928
Sex - Male	1.28 (1.14 - 1.44)	< 0.001	1.28 (1.14 - 1.43)	< 0.001	1.28 (1.15 - 1.44)	< 0.001
Region - Latin America	0.99 (0.85 - 1.15)	0.852	1 (0.86 - 1.16)	0.991	0.98 (0.85 - 1.15)	0.821
Region - North American	1.03 (0.86 - 1.22)	0.749	1.03 (0.86 - 1.24)	0.726	1.03 (0.87 - 1.23)	0.719
Region - Pacific Asia/Pacific and Other	1.09 (0.93 - 1.26)	0.261	1.1 (0.95 - 1.28)	0.194	1.08 (0.94 - 1.26)	0.295
Region - Western Europe	0.92 (0.81 - 1.04)	0.18	0.92 (0.81 - 1.05)	0.21	0.92 (0.81 - 1.04)	0.193
BMI	1.02 (1.01 - 1.03)	< 0.001	1.02 (1.01 - 1.03)	< 0.001	1.02 (1.01 - 1.03)	< 0.001
eGFR	1 (0.99 - 1)	0.004	1 (0.99 - 1)	0.002	1 (0.99 - 1)	0.004
Ejection Fraction %	0.98 (0.97 - 0.99)	< 0.001	0.98 (0.97 - 0.99)	< 0.001	0.98 (0.97 - 0.99)	< 0.001
NYHA Class 2	1.23 (0.96 - 1.56)	0.105	1.24 (0.96 - 1.6)	0.093	1.25 (0.96 - 1.61)	0.084
NYHA Class 3	1.56 (1.2 - 2.01)	0.001	1.6 (1.23 - 2.08)	< 0.001	1.59 (1.19 - 2.07)	0.003
NYHA Class 4	1.67 (1.02 - 2.74)	0.042	1.71 (1.02 - 2.79)	0.041	1.72 (1.04 - 2.84)	0.033
Diabetes - Yes	1.37 (1.25 - 1.5)	< 0.001	1.37 (1.26 - 1.5)	< 0.001	1.38 (1.25 - 1.5)	< 0.001
SBP	1 (1 - 1)	0.294	1 (1 - 1)	0.34	1 (1 - 1)	0.304
Heart Rate	1.01 (1 - 1.01)	0.002	1.01 (1 - 1.01)	< 0.001	1.01 (1 - 1.01)	0.002
History of Atrial fibrillation – Yes	1.05 (0.94 - 1.16)	0.394	1.06 (0.95 - 1.19)	0.279	1.05 (0.95 - 1.18)	0.346
Ischemic Heart Failure - Yes	0.99 (0.87 - 1.12)	0.874	0.98 (0.87 - 1.12)	0.784	1 (0.88 - 1.13)	0.95
Prior Hospitalisation for Heart Failure - Yes	1.38 (1.26 - 1.52)	< 0.001	1.38 (1.25 - 1.51)	< 0.001	1.38 (1.26 - 1.53)	< 0.001
Prior History of Myocardial Infarction - Yes	1.2 (1.07 - 1.36)	0.002	1.2 (1.07 - 1.36)	0.006	1.2 (1.06 - 1.36)	0.001
Prior History of Stroke - Yes	1.1 (0.95 - 1.27)	0.185	1.11 (0.95 - 1.28)	0.178	1.11 (0.95 - 1.27)	0.18
Value of log NT-ProBNP	1.93 (1.73 - 2.13)	< 0.001	1.8 (1.44 - 2.05)	< 0.001		
Slope of log NT-ProBNP			0.5 (0.08 - 2.45)	0.479		
Area of log NT-ProBNP					1.9 (1.76 - 2.05)	< 0.001

Table 27 Marginal and Conditional Performance Statistics of Joint Models

	Marginal			Conditional		
Statistic	Joint Model 1 Value	Joint Model 2 Value + Slope	Joint Model 3 Area	Joint Model 1 Value	Joint Model 2 Value + Slope	Joint Model 3 Area
DIC	43419.22	47301.98	39253.35	70926.80	56671.85	71299.12
LPML	-25624.85	-26060.13	-23622.82	-43948.67	-36000.15	-44637.90

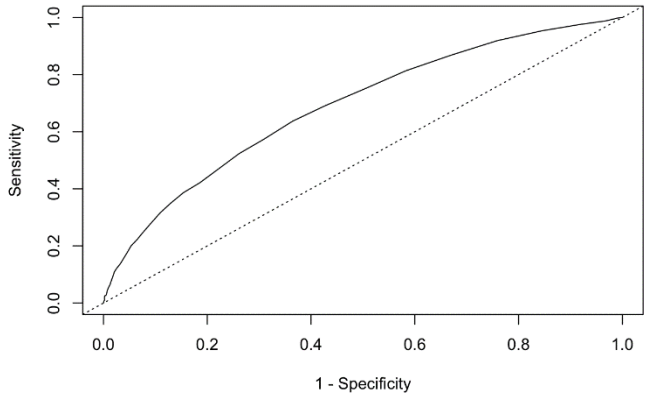
Figure 16 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months

A.) Joint Model 1 – Value: Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



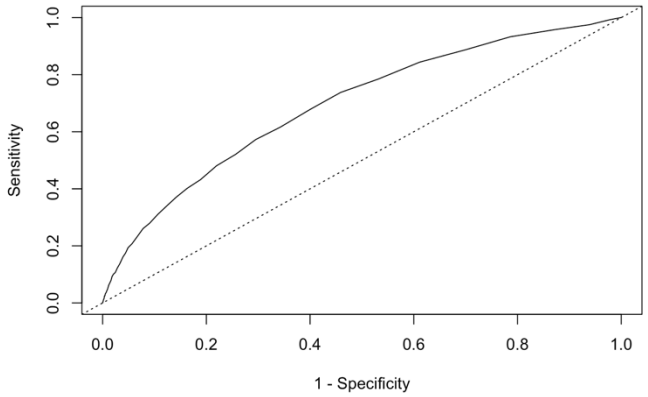
Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

B.) Joint Model 2 - Value and Slope: Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

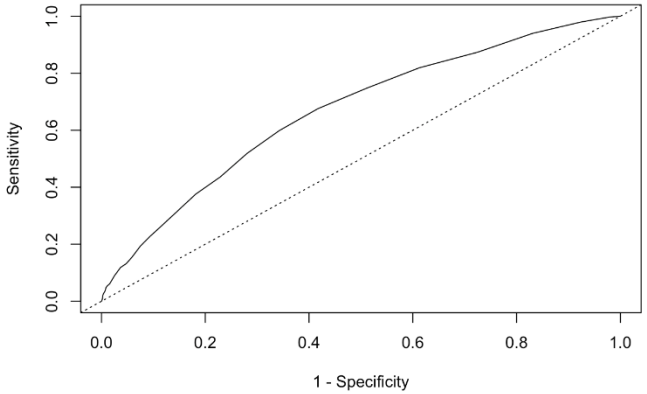
C.) Joint Model 3 – Area: Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

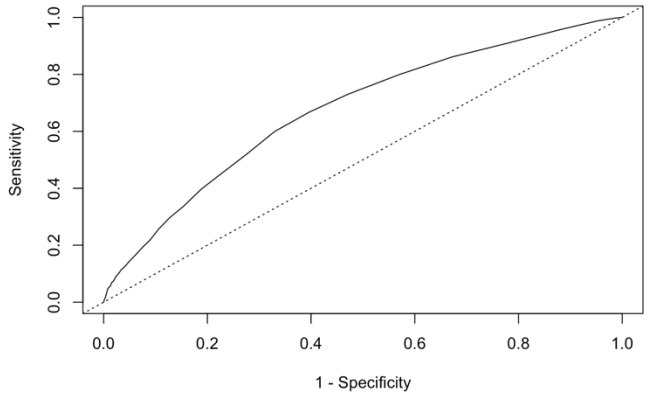
Figure 17 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months

A.) Joint Model 1 – Value: Time-Varying ROC, AUC and Brier Score at Month 24 Using Longitudinal Data up until Month 12



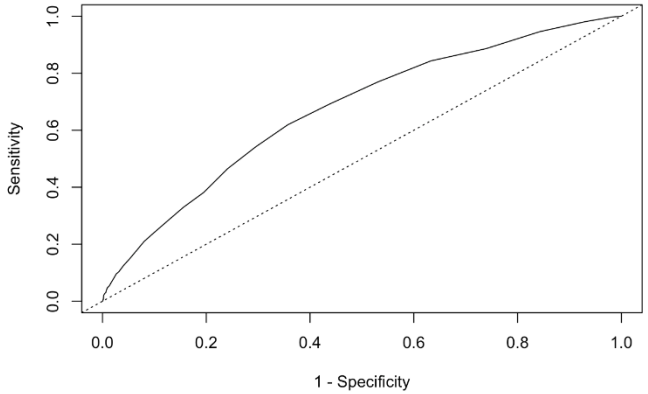
Time-Varying AUC: 0.67
Time-Varying Brier Score: 0.08

B.) Joint Model 2 - Value and Slope: Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data up until Month 12



Time-Varying AUC: 0.67
Time-Varying Brier Score: 0.08

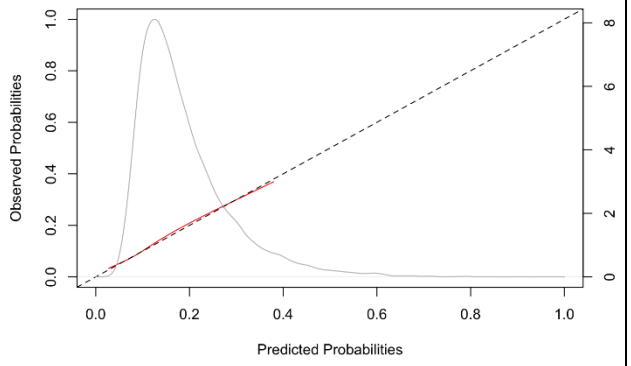
C.) Joint Model 3 – Area: Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data up until Month 12



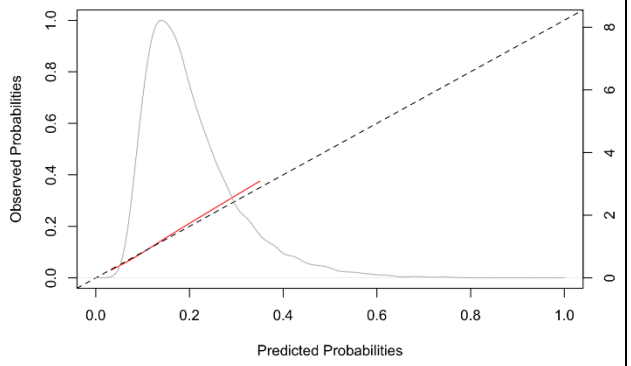
Time-Varying AUC: 0.67
Time-Varying Brier Score: 0.08

Figure 18 Calibration Curves for Joint Models at Month 12 using Longitudinal Data at Month 0

A.) Joint Model 1 – Value: Calibration Curve at Month 12 using Longitudinal Data at Month 0



B.) Joint Model 2 – Value and Slope: Calibration Curve at Month 12 using Longitudinal Data at Month 0



C.) Joint Model 3 – Area: Calibration Curve at Month 12 using Longitudinal Data at Month 0

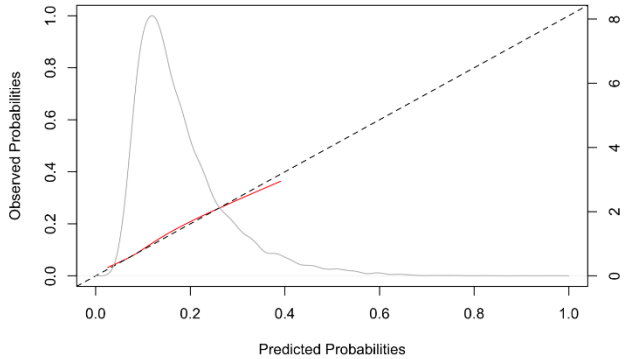
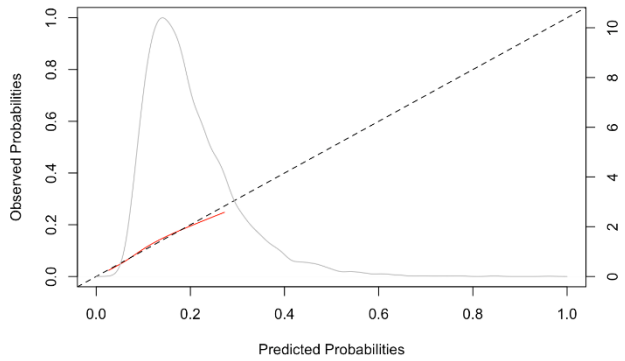
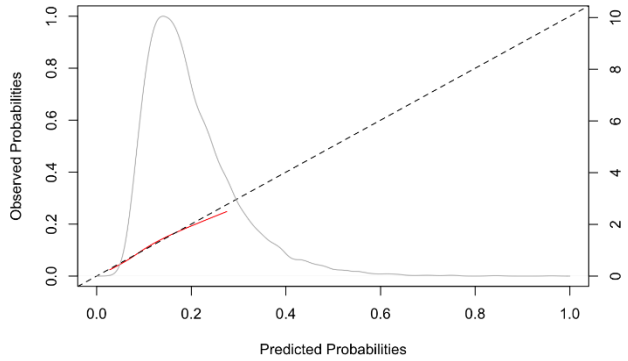


Figure 19 Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12

A.) Joint Model 1 – Value: Calibration Curve at Month 24 using Longitudinal Data up until Month 12



B.) Joint Model 2 – Value and Slope: Calibration Curve at Month 24 using Longitudinal Data up until Month 12



C.) Joint Model 3 – Area: Calibration Curve at Month 24 using Longitudinal Data up until Month 12

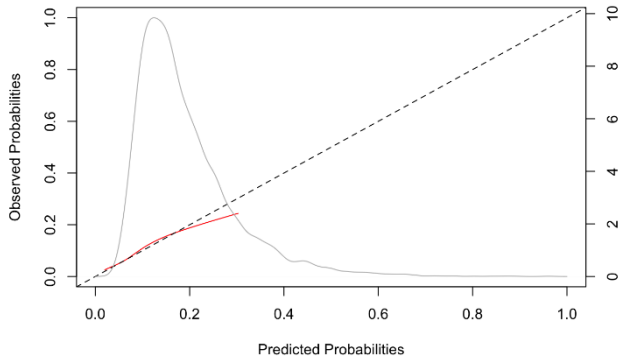
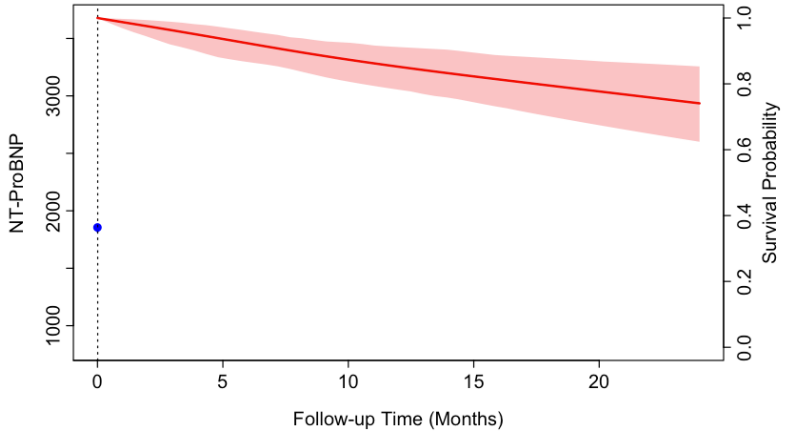


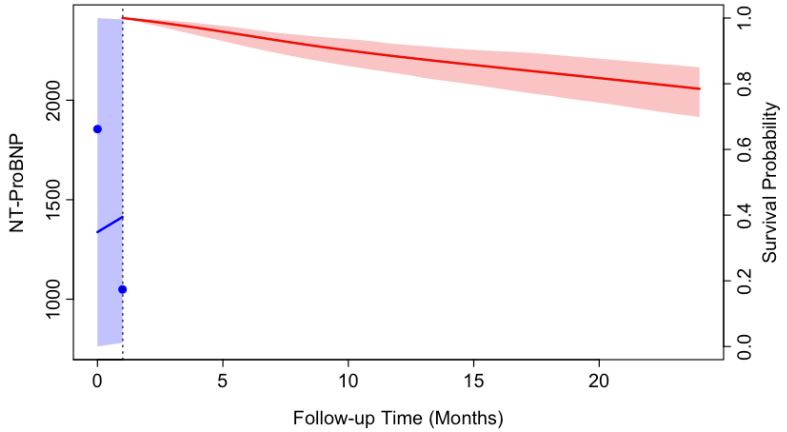
Figure 20 Dynamic Predictions from Joint Model 4 for a Randomly Sampled Patient with Measurements of NT-ProBNP and Corresponding Survival Probability

A.) Observation 1, NT-ProBNP Measurement at Month 0 and Corresponding Survival Probability



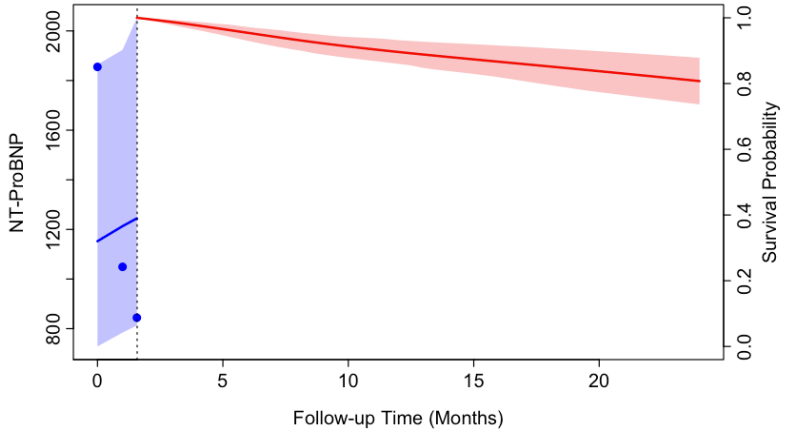
Survival at 24 Months: 0.74 (0.85 - 0.62)

B.) Observation 2, NT-ProBNP Measurements at Months 0 and 1 and Corresponding Survival Probability



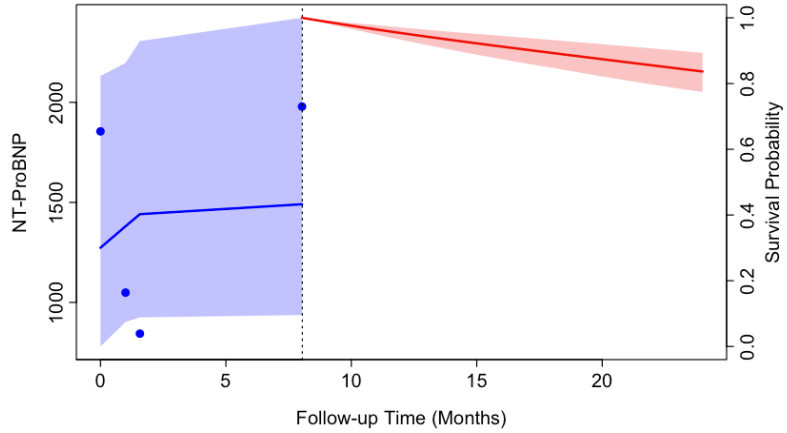
Survival at 24 Months: 0.78 (0.85 - 0.7)

C.) Observation 3, NT-ProBNP Measurements at Months 0, 1 and 1.5 and Corresponding Survival Probability



Survival at 24 Months: 0.81 (0.88 - 0.74)

D.) Observation 4, NT-ProBNP Measurements at Months 0, 1, 1.5 and 8 and Corresponding Survival Probability



Survival at 24 Months: 0.84 (0.89 - 0.77)

NT-ProBNP is represented by the blue line with confidence interval, the dashed line represents the last time the patient was known to be alive, and the red line represents the survival probability with confidence interval.

4.3.6 Joint Model Performance

Table 27 shows the marginal and conditional DICs and LPMLs for each of the joint models. For marginal fit, both the DIC and LPML suggest that Joint Model 3 (Area) provides the best fit, with the highest LPML (-23622.82) and lowest DIC (39253.35). This suggests that Joint Model 3 (Area) provides a better fit for the data for the overall population. These results also suggest that Joint Model 2 (Value and Slope) performed worst out of the joint models with respect to marginal fit, with a marginal LPML of -26060.13 and LPML of 47301.98. On the other hand, the DICs and LPMLs for the conditional fit suggest that Joint Model 2 provides the best fit for individuals with a DIC of 56671.85 and LPML of -36000.15. The differences in DICs and LPMLs between the joint models for the conditional fit are greater than those differences in the marginal fit.

Figure 16 shows the time-varying ROCs and corresponding AUCs and Brier scores for the joint models at month 12 using data at time point 0 (baseline). Each of the models perform similarly with a time-varying AUC of 0.69, a time-varying Brier score of 0.10 and only minor differences in time-varying ROC curves. The time-varying AUCs at month 12 suggests that the joint models' performance is below the general guidelines used for acceptable discrimination [16] at 12 months, instead falling at the top end of the poor discrimination category. Similarly, as seen in Figure 17, each model performs equally at 24 months using longitudinal data up until month 12 with respect to time-varying AUC and time-varying Brier score, with a time-varying AUC of 0.67 and a time-varying Brier score of 0.08, suggesting a marginal loss in accuracy but a better overall discrimination. The time-varying AUC at 24 months using longitudinal data until month 12 is again below the general guidelines for acceptable discrimination.

While there is not a widely accepted good or bad range for the Brier score, generally, the lower the score the better, and it can be used to compare model accuracy with the same data. As such, the joint models appear to perform equally as well as each other at 12 months, using baseline data with a Brier score of 0.10 at the lower end of possible Brier scores. The models also perform equally at 24 months, using data up until 12 months with a better time-varying Brier score of 0.8.

The calibration curves for the joint models at month 12 using data at baseline, as shown in Figure 18, show Joint Model 1 (Value) as being well calibrated, whereas Joint Model 2 (Value and Slope) exhibits minor signs of poor fit near the higher end of the predicted probabilities. Inversely, Joint Model 3 (Area) shows minor signs of issues with fit nearer the higher end of predicted possibilities. This suggests that Joint Model 1 (Value) is the better calibrated model at month 12 using baseline data; however, this is only one time point. At 24 months using longitudinal data up until month 12 (shown in Figure 19), Joint Model 1 (Value) shows a slight issue with fit at the higher end of the predictive probabilities, as does Joint Model 2 (Value and Slope). Joint model 3 (Area) has a worse calibration than itself at 12 months using longitudinal data at baseline and the other joint models at month 24 using longitudinal data up until month 12, suggesting a larger number of issues with calibration at the higher end of the predicted probabilities.

4.3.7 Dynamic Predictions from Joint Models

Figure 20 illustrates the dynamic predictions of a randomly sampled patient by Joint Model 2 (Value and Slope) chosen for its better subject specific (conditional fit); at four different time points. This shows the observed levels and predicted trajectories of NT-ProBNP on the original scale. While the 95% CIs on the longitudinal side are quite wide, this is possibly due

to scale and range of the observed measurements of NT-ProBNP. The 95% CIs on the survival side narrow with added measurements, likely as there is more data available. The survival probability increases with the measurements, as the predicted trajectory of NT-ProBNP stabilises and the patient has survived for longer. The trajectory suggests an increase in predicted NT-ProBNP until month three; however, the relative increase of the predicted trajectory of NT-ProBNP does not appear to have as much effect on survival, possibly due to the patient surviving longer at each measurement and the relatively stable predicted profile of NT-ProBNP.

4.3.8 External Validation of Joint Models

Figure 21 illustrates the external validation time-varying ROCs, AUCs and Brier scores, using data from the ATMOSPHERE trial to provide external validation. The time-varying ROCs at 12 months using data from baseline are very similar. Joint Model 2 (Value and Slope) showing minor differences, which is highlighted in the corresponding time-varying AUC. With Joint Model 2 (Value and Slope) having a lower time-varying AUC of 0.68, minimally lower than Joint Model 1 (Value) and Joint Model 3 (Area) with a time-varying AUC of 0.69, this once again suggests the joint models perform just below the acceptable level of discrimination. The time-varying Brier score is similar to the original data, with each model performing equally with a score of 0.10, which is at the lower end of possible Brier scores, at 24 months using longitudinal data up until month 12 from the ATMOSPHERE trial. Figure 22 shows an improvement for time-varying AUCs for both Joint Model 2 (Value and Slope) and Joint Model 3 (Area), with all models showing an improvement in time-varying Brier score of 0.09, suggesting the models perform better at month 24 using data from the ATMOSPHERE trial and its longitudinal data until 12 months.

Inversely, the time-varying calibration curves at 12 months using ATMOSPHERE data at baseline show deviation from the ideal validation curve shown in Figure 23, with Joint Model 1 (Value) and Joint Model 3 (Area) having similar curves and Joint Model 2 (Value and Slope) showing slightly more deviation; however, this deviation is minimal. The time-varying calibration curves at month 24 using longitudinal data up until 12 months illustrate further deviation from the ideal fit, with Joint Model 1 (Value) and Joint Model 2 (Value and Slope) performing similarly and Joint Model 3 (Area) performing better.

Figure 21 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs and Brier Scores for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months using Data from the ATMOSPHERE Trial

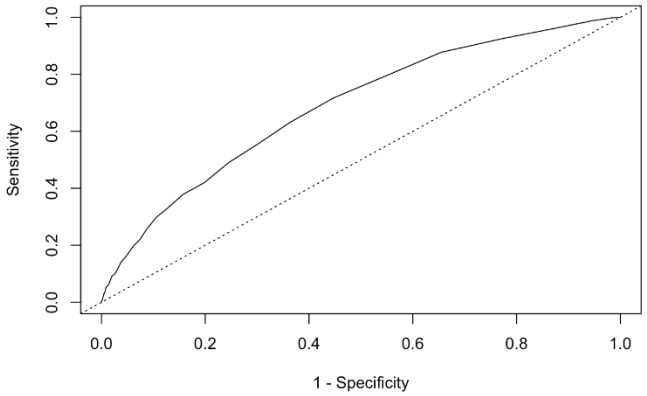
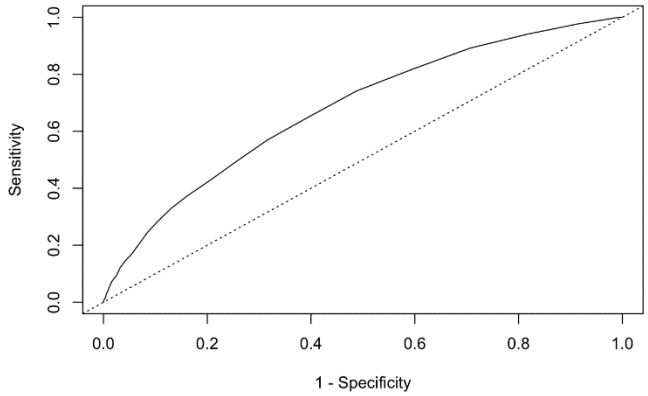
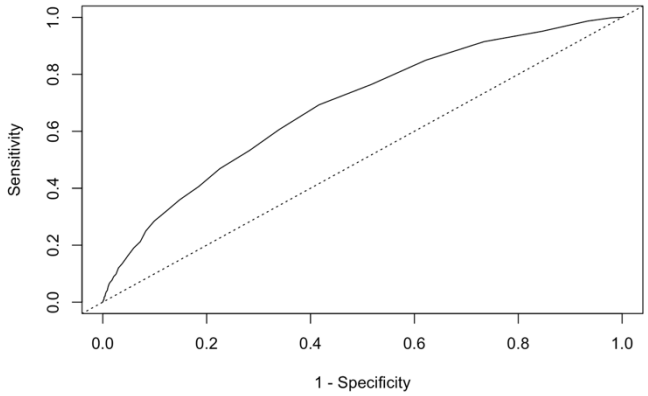
A.) Joint Model 1 – Value: External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial	B.) Joint Model 2 - Value and Slope: External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial	C.) Joint Model 3 – Area: External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial
		
Time-Varying AUC: 0.69 Time-Varying Brier Score: 0.10	Time-Varying AUC: 0.68 Time-Varying Brier Score: 0.10	Time-Varying AUC: 0.69 Time-Varying Brier Score: 0.10

Figure 22 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months

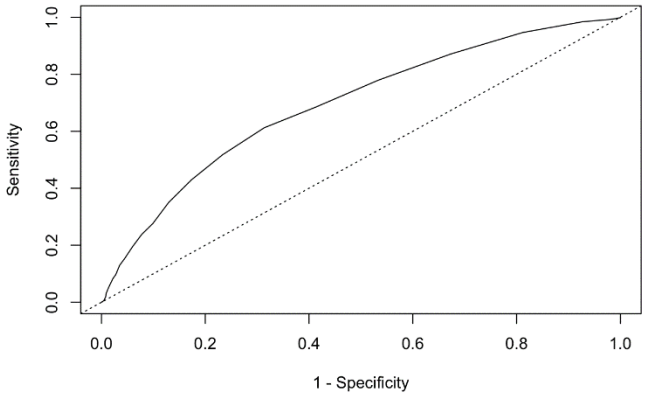
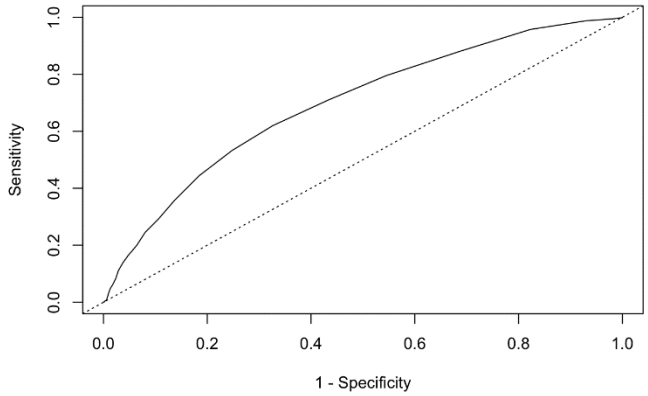
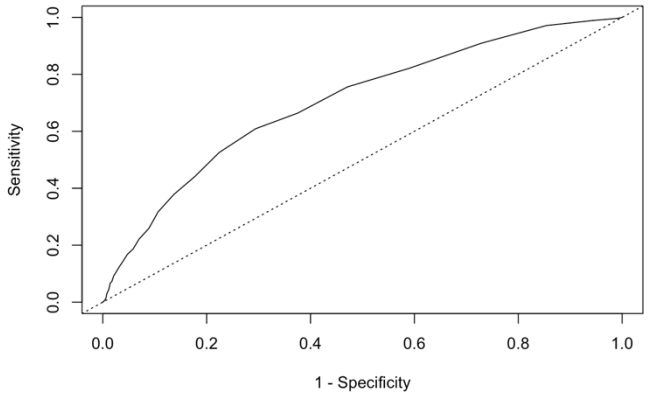
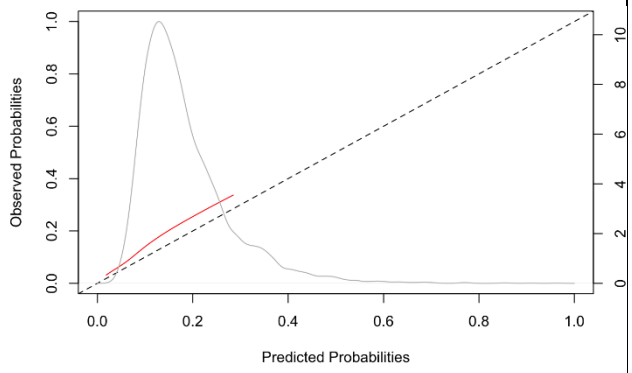
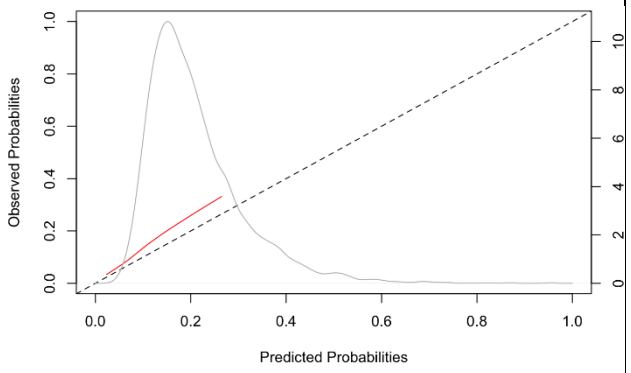
A.) Joint Model 1 – Value: External Validation Time-Varying ROC, AUC, and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial	B.) Joint Model 2 - Value and Slope: External Validation Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial	C.) Joint Model 3 – Area: External Validation Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial
		
Time-Varying AUC: 0.69 Time-Varying Brier Score: 0.09	Time-Varying AUC: 0.70 Time-Varying Brier Score: 0.09	Time-Varying AUC: 0.70 Time-Varying Brier Score: 0.09

Figure 23 External Validation Calibration Curves for Joint Models at Month 12 using Longitudinal Data At Month 0

A.) Joint Model 1 – Value: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



B.) Joint Model 2 – Value and Slope: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



C.) Joint Model 3 – Area: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial

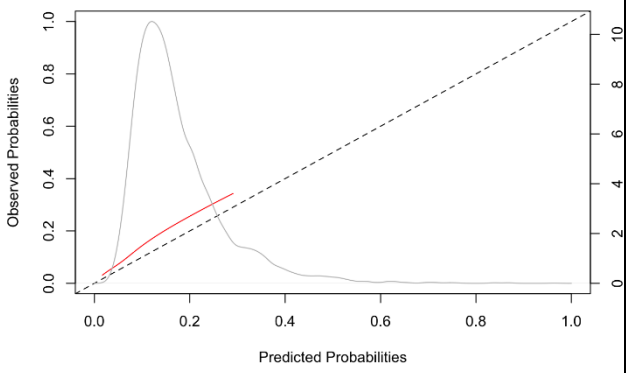
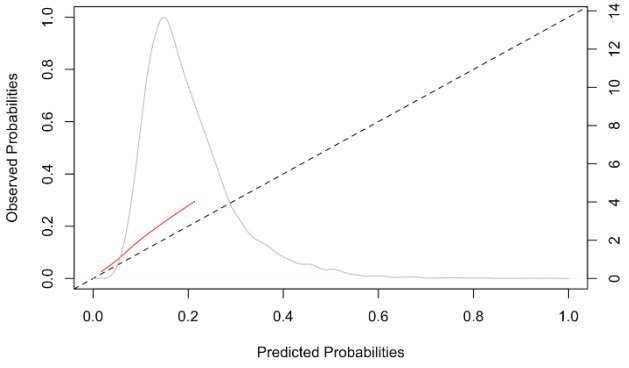
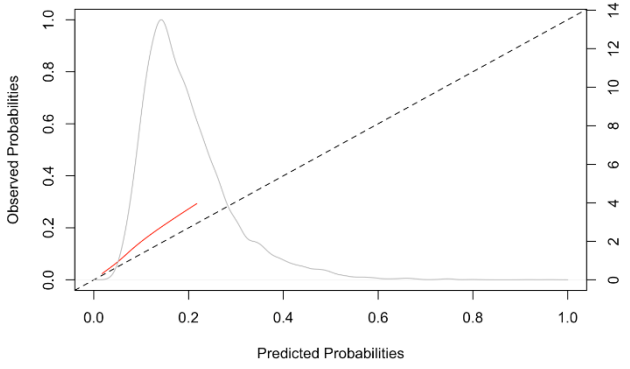


Figure 24 External Validation Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12

A.) Joint Model 1 – Value: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



B.) Joint Model 2 – Value and Slope: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



C.) Joint Model 3 – Area: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial

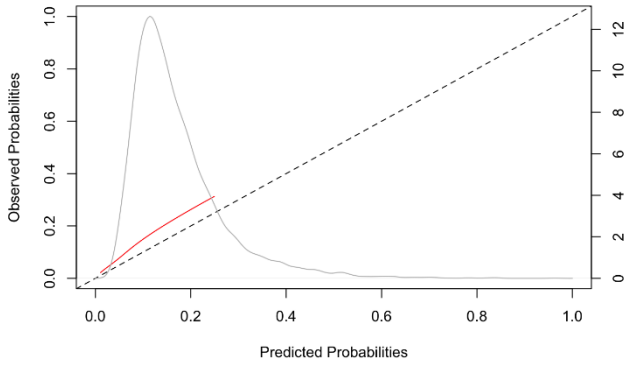


Table 28 Hazard Ratio (95% CI) and P-Values from the Cox PH Models

	Cox PH Last Measurement		Extended Cox PH	
Variable	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**
Treatment - LCZ	0.83 (0.76-0.91)	<0.001	0.83 (0.76-0.91)	<0.001
Age	1 (1-1.01)	0.865	1 (1-1.01)	0.924
Sex - Male	1.28 (1.14-1.44)	<0.001	1.27 (1.13-1.43)	<0.001
Region - Latin America	0.98 (0.85-1.14)	0.794	1.02 (0.88-1.18)	0.831
Region - North American	1.04 (0.87-1.24)	0.683	0.98 (0.82-1.18)	0.851
Region - Pacific Asia/Pacific and Other	1.08 (0.93-1.25)	0.337	1.12 (0.96-1.3)	0.145
Region - Western Europe	0.92 (0.81-1.04)	0.192	0.9 (0.79-1.03)	0.125
BMI	1.02 (1.01-1.03)	0.001	1.02 (1.01-1.03)	0.001
eGFR	1 (0.99-1)	0.005	1 (0.99-1)	0.004
Ejection Fraction %	0.98 (0.97-0.99)	<0.001	0.98 (0.97-0.99)	<0.001
NYHA Class 2	1.21 (0.95-1.55)	0.124	1.17 (0.92-1.5)	0.2
NYHA Class 3	1.53 (1.18-1.98)	0.001	1.5 (1.16-1.94)	0.002
NYHA Class 4	1.71 (1.04-2.81)	0.035	1.78 (1.08-2.93)	0.023
Diabetes - Yes	1.37 (1.25-1.5)	<0.001	1.36 (1.24-1.49)	<0.001
SBP	1 (1-1)	0.309	1 (0.99-1)	0.201
Heart Rate	1.01 (1-1.01)	0.002	1.01 (1-1.01)	0.005
History of Atrial fibrillation – Yes	1.09 (0.99-1.21)	0.08	1.08 (0.97-1.19)	0.153
Ischemic Heart Failure - Yes	1 (0.88-1.13)	0.977	0.98 (0.86-1.11)	0.768
Prior Hospitalisation for Heart Failure - Yes	1.37 (1.25-1.51)	<0.001	1.35 (1.22-1.49)	<0.001
Prior History of Myocardial Infarction - Yes	1.21 (1.07-1.36)	0.002	1.22 (1.08-1.38)	0.001
Prior History of Stroke - Yes	1.1 (0.95-1.28)	0.183	1.1 (0.95-1.27)	0.223
Log NT-ProBNP	1.61 (1.54-1.69)	<0.001	1.63 (1.55-1.71)	<0.001

Figure 25 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 12

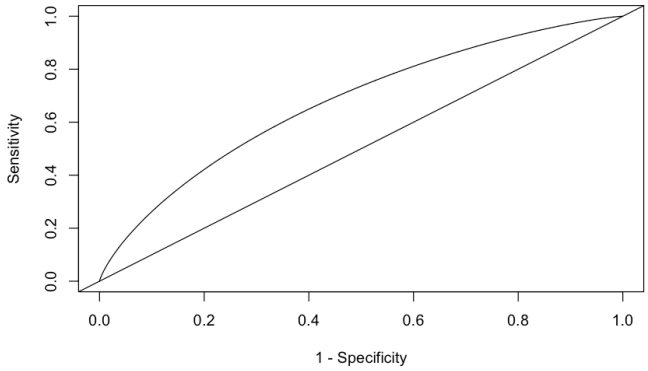
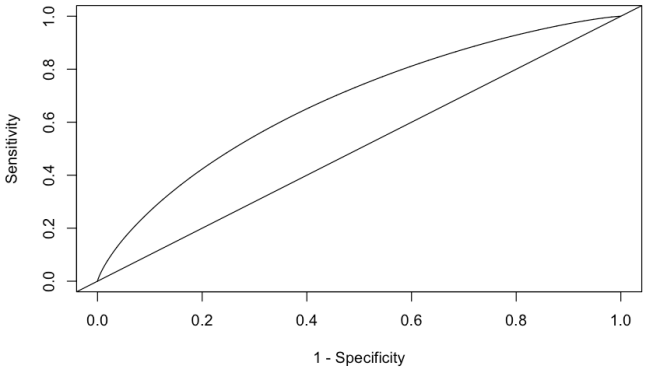
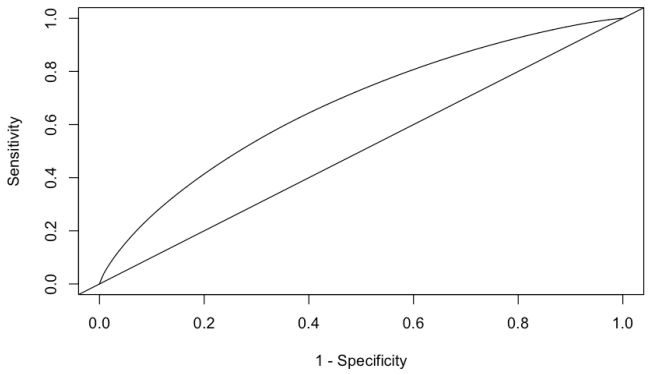
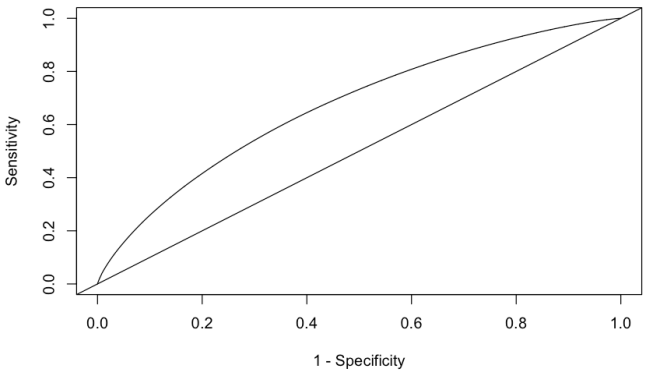
A.) Time-Varying ROC and AUC at 12 Months for the Cox PH Last Measurement Model	B.) Time-Varying ROC and AUC at 12 Months for the Extended Cox PH Model
	
Time-Varying AUC at 12 Months: 0.67	Time-Varying AUC at 12 Months: 0.67

Figure 26 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 24

A.) Time-Varying ROC and AUC at 24 Months for the Cox PH Last Measurement Model	B.) Time-Varying ROC and AUC at 24 Months for the Extended Cox PH Model
	
Time-Varying AUC at 24 Months: 0.67	Time-Varying AUC at 24 Months: 0.67

4.3.9 Comparative Analysis

Table 28 shows the HRs and corresponding 95% CI and P-Values for both Cox PH models, the last measurement model, and the extended Cox PH model. Only minor differences in HRs between the models are observed, but the extended Cox PH has a slightly higher HR for Log NT-ProBNP, suggesting that a unit increase of Log NT-ProBNP increases the hazard of the composite event increases by 1.63 times, whereas the last measurement model suggests a slightly lower association, with the hazard of the composite event increasing by 1.61 times per unit increase in Log NT-ProBNP. With both still suggesting a strong association of Log NT-ProBNP and the composite endpoint. There are also only minor differences in most HRs when comparing the Cox PH models against the joint models, with the only major differences in treatment with LCZ and Log NT-ProBNP. The treatment effect of LCZ is considered statistically significant in the Cox PH models, suggesting a treatment with LCZ lowers the hazard of the composite event by 0.83 times. The joint models also give a bigger hazard ratio for Log NT-ProBNP (between 1.8 and 1.93). Both could possibly be explained by the direct treatment effect on Log NT-ProBNP and how Log NT-ProBNP is likely to be a mediator, as explained previously.

The time-varying ROCs and time-varying AUCs of the two Cox PH models shown in Figure 25 and Figure 26 suggest that the models perform equally with each other with respect to discrimination at 12 and 24 months. However, when the time-varying AUCs are compared against the joint models, the Cox PH models perform worse at 12 months, with an AUC of 0.67 for both models compared to the 0.69 of the joint models, and equally with the joint models at 24 months, with an AUC of 0.67.

4.4 Discussion

With the recent rise in awareness and use of joint modelling in heart failure clinical research, it is important to explore their use, not only to assess the association of a biomarker with an endpoint, but how and why they may be used for prognostic models and if and how they improve on the current methodology used in prognostic models. To do this, three joint models were fit to two clinical trial data sets using different parameterisations of association parameter and compared against each other and two traditional models.

Results from all the models suggest an association with Log NT-ProBNP and the composite endpoint of first hospitalisation for heart failure and death from cardiovascular causes. Whilst this association has been demonstrated in the past, these joint models add to the evidence base and suggest that the treatment effect of the active drug (LCZ) from the PARADIGM trial is mediated through the biomarker NT-ProBNP. All the joint models show a statistically insignificant treatment effect on the survival outcome of the model, whilst showing a treatment effect on the longitudinal side. Unfortunately, because the focus of this work was to build and assess a prognostic model and due to the non-linearity of measurements of NT-ProBNP, it is not possible to determine the overall treatment effect as illustrated by J. Ibrahim et al., as this overall treatment effect requires treatment coefficients from both the survival and longitudinal components of the models [51], and as the longitudinal model included natural splines and an interaction of these splines with treatment, these coefficients are not directly interpretable and therefore not applicable for this application.

Joint Model 1 (Value) and Joint Model 3 (Area) both suggest a near double increase in the hazard of the composite event per unit increase in their respective association parameter estimates. However, due to the previously mentioned treatment effect on NT-ProBNP, the

treatment effect must be taken into consideration when interpreting these results, making it more difficult to interpret.

The average longitudinal profile of NT-ProBNP, illustrated in Figure 15, shows that for the average patient of each treatment there is a difference in treatment over time on NT-ProBNP. However, because joint models include random effects and therefore, subject-specific trajectories, these results vary depending on the specific patient.

The longitudinal coefficients from the joint model suggest that all the chosen clinical covariates are considered statistically significant, with both treatment and BMI having a negative effect on NT-ProBNP and age and a history of atrial fibrillation having a positive effect on NT-ProBNP, with only minor differences between the joint models.
The survival outcomes from the joint models show only minimal difference in the hazard ratios apart from the association parameters, as discussed previously.

Joint Model 3 (Area) performed the best with regards to marginal fit with an LPML of -23622.82 and DIC of 39253.35. However, Joint Model 2 (Value and Slope) performed the best with regards to conditional fit with an LPML of -36000.35 and DIC of 56671.85. Illustrating an improvement of 19% for the LPML and 20% in DIC when compared to the conditional fit of Joint Model 3. When comparing the marginal fit, Joint Model 3 has an improvement of 9% in LPML and 17% in DIC to Joint Model 2 (Value and Slope). For a prognostic model focused on personalised/subject-specific prognosis, the trade-off in marginal fit against the improvement in conditional fit may mean that Joint Model 2 (Value and Slope) is the better performing model.

This better conditional fit is of note, as while the association slope parameter estimate of this model is considered statistically insignificant, possibly due to the limited number of repeated measurements, the better condition fit may suggest that some patient specific slopes are important. This gives warrant to the use of a clinical approach rather than a stepwise approach to model selection, as a stepwise approach may have excluded this model due to the apparent statistical insignificance of the slope parameter.

In terms of prognostic performance with regards to the time-varying ROCs, time-varying AUCs and time-varying Brier scores, both at 12 months using longitudinal data at month 0 and 24 months using longitudinal data at month 12 suggest that the joint models performed equally, with time-varying AUCs of 0.69 at 12 months using longitudinal data at month 0 and 0.67 at 24 months using longitudinal data up until month 12. While this discrimination is just short of the acceptable range, the Brier score and calibration curves suggest that the model is well calibrated at these two time points, the Brier score encapsulating discrimination and calibration to provide an accuracy metric and may provide a better overview of the model performance than the time-varying AUCs.

Validation with data from the ATMOSPHERE trial suggests that the models perform as well with external data, at 12 months using longitudinal data at month 0, with respect to discrimination, with only Joint Model 2 (Value and Slope) performing marginally worse with a time-varying AUC of 0.68 compared to the time-varying AUC of 0.69 with the original data. The calibration curves suggest a worse calibration but not by much, as reflected in the time-varying Brier score being the same (0.10) at this time point. The time-varying AUCs at month 24 using longitudinal data up until month 12 suggest that Joint Models 2 and 3 perform marginally better with respect to prognosis, with a time-varying AUC of 0.70,

whereas Joint Model 1 (Value) performs equally with a time-varying AUC of 0.69. The calibration of the models appears worse, with the Brier score for all joint models increasing from 0.8 to 0.9. Overall, the validation with the ATMOSPHERE data suggests the models perform as well with external data. This performance with external data may suggest that the joint models provide a good fit for the validation data. However, it may also be due to similarities in the inclusion and exclusion criteria of the two trials.

The comparative analysis with the two Cox PH models suggests the joint models may outperform the Cox PH models at 12 months and perform equally at 24 months; however, because of the way the joint models predict outcomes based on longitudinal data, this may be a case of comparing apples with oranges, and the time-varying AUCs may not be comparable.

This analysis highlights the difficulties in comparing the Cox PH models against the joint models and therefore, is limited in its ability to say which is better.

Other prediction models such as PREDICT-HF have used the same patient population and outcomes as the joint models presented in this study [45]. They assess their models using the C-statistic at 12 and 24 months and while these are the same time points, the authors did not report how they calculated these statistics at these time points. Importantly, the C-statistic when calculated at specified time points has been shown to be problematic as it can fail to account for the event status at these specified points. Which may lead to overestimation of the discriminatory ability of the model [145]. Therefore comparison between the two metrics (time-varying AUC / C-statistic) while possible is unlikely to be meaningful.

Overall, the joint models perform adequately and may provide good prognostic models. The dynamic and subject-specific predictions are where joint models are more useful over the Cox PH models, with the joint models allowing for both. The dynamic predictions allowing for the use previous and repeat measurements of NT-ProBNP adding the potential to better inform prognosis. However, the subject-specific predictions are based on the random effects which are, by definition, latent variables, meaning they cannot be measured and rely on the premise that these unseen properties provide enough information to provide accurate subject-specific predictions.

This analysis has some limitations, such as the use of screening values of NT-ProBNP for patients missing values of NT-ProBNP at randomisation. There is also an imbalance of data, as the number of patients with more than one measurement of NT-ProBNP is less than the number of included patients. The use of the sample date rather than the scheduled visit date breaks the original intention-to-treat design of the original trial and therefore, may introduce bias. Another possible issue is that of inclusion and exclusion criteria of the original trial may not be as translatable to the general population; for example, the data is limited to patients with HFrEF, however this is a limitation of the data itself and therefore out of the scope of this current research. It is also worth mentioning that the *JMBayes2* package, whilst based on the *JMBayes* package, is still in development and only provides limited accuracy measures, which, while appropriate for joint models, makes it difficult to compare against other types of models. Finally, whilst the JMbayes2 package provides a function for calibration curves these curves do not include (CIs) which would provide for a better interpretation of calibration.

Chapter 5 Joint Modelling of eGFR and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure

5.1 Foreword

This chapter is like that of Chapter 4 and therefore takes the same style and may repeat on what has previously been stated. This repetition has been rephrased but included to allow this chapter to be read independently of Chapter 4.

5.2 Introduction

There are interrelationships which are well documented between outcomes and patient characteristics, within the clinical syndrome of heart failure. These characteristics include physical characteristics and biomarkers. With the previously mentioned rise in use and awareness of joint modelling in heart failure and lack of existing prognostic models using joint modelling in heart failure; it is informative to understand how joint modelling could be applied to heart failure and how it may improve on current prognostic modelling techniques. As previously mentioned in Chapter 1, joint models allow for repeated measurements for covariates, whist also correcting for measurement error and accounting for correlation between measurement occasions; and also allow for subject-specific and dynamic predictions [51], [52]. Repeat measurements of biomarkers including composite markers such as eGFR are often collected during clinical trials along with time-to-event data, which is used within survival analysis. Except for the prognostic model developed in Chapter 4, there are no known prognostic models which employ joint modelling identified within the heart failure literature. Therefore, the aim of this chapter like the previous, is to demonstrate the application of prognostic models using joint models in heart failure and to critically appraise

them. Additionally, I will use data from clinical trials to compare the performance of joint models for prognostication against the currently recommended approaches.

5.3 Methods

5.3.1 Data Source

As noted in Chapter 3, data from the clinical trial known as PARADIGM-HF was used. PARADIGM-HF randomised patients 8,399 patients at a 1:1 ratio in a double-blind fashion to receive either enalapril at a dose of 10mg twice daily or LCZ696 at a dose of 200mg twice daily.

5.3.2 Statistical Analysis

To address the aforementioned research aims, data were analysed, processed and cleaned before the fitting of joint models. No formal sample size calculation was applied due to limitations of the available formula [138]. The R package *JMbayes2* was used to fit all joint models. As previously stated, this package is still in active development, and is based on the R packages *JM* and *JMbayes* [67], [114] by the same author. As stated earlier the JMbayes2 package fits joint models using a fully parametric Bayesian approach using MCMC sampling. Although this package is still in development, it was used as it includes the ability to produce calibration plots, which are important indicators of performance of prognostic models [39], [40]. The package also allows for more flexible functional form representation of numerical covariate effects as previously mentioned.

The basic joint model is comprised of two components: a longitudinal and time-to-event component; these are fit individually and then provided to the *jm* function within the *JMbayes2* package and refit as a joint model.

5.3.3 Data Cleaning and Processing

Data needed to be analysed and cleaned prior to the analysis. This comprised the analysis of variables and their distributions to ensure there were no outliers, duplicated or improbable values, as well as the manipulation of categorical variables to ensure they were in the correct format and easily human readable. This also included the recoding of the event indicator to ensure that 1 indicated the participant experienced the event. Additionally, the data was reformatted to be compatible with the various functions used to fit the sub models and models. This included making sure the same group of patients was provided to both sub models as well as merging baseline covariates with the longitudinal data. The performance metrics required all covariates and the repeated measurements to be in one dataset.

5.3.4 Covariate Selection

Both components of the joint model; longitudinal and survival allow for modelling of covariate effects. For the same reasons stated in Chapter 4, for both components, covariates were chosen based on prior clinical knowledge as described in Chapter 1 and in conjunction with my clinical supervisor. No covariates identified through these means were missing from either dataset.

5.3.5 Missing Data

Joint models allow for uneven and missing longitudinal measurements, however in each of the sub models (longitudinal and survival) all covariates must be complete. For 55 patients' data were incomplete for covariates; because this is less than 1% of the available data a complete case analysis was chosen.

5.3.6 Longitudinal Sub Model

The longitudinal sub model needs to be either an LME fitted by the function *'lme'* from the *'nlme'* package or a GLMM fitted by the *'mixed_model'* function from the *'GLMMadaptive'* package as required by the *JMbayes2* package. Due to the use of a continuous response variable an LME was required, allowing for both fixed and random effects.

5.3.6.1 eGFR

eGFR was chosen as the response variable for the LME, as explained in Chapter 1, declining renal function has a well-documented relationship with heart failure and for that reason it was chosen as the response variable in this analysis. eGFR was observed to be skewed and therefore was transformed using the natural logarithm. This transformation is required due to assumptions of normality in the random effects [52].

5.3.6.2 Time (Fixed Effect)

The response variable of an LME can be unevenly spaced (with respect to time). Sample date was available for eGFR and used to calculate time in months since randomisation using a 28-day calendar month. This was preferred to scheduled date to provide a more accurate model. Evaluation of longitudinal profiles of log eGFR from randomly sampled patients showed that some of their profiles exhibited signs of non-linearity over time. Therefore, it was decided to include time non-linearly through the use of restricted natural cubic splines; the 'ns' function from the 'splines' package was used. Due to computational complexity the boundary knots were manually specified using quantile values.

5.3.6.3 Covariates (Remaining Fixed Effects)

Stated previously, the remaining covariates were selected using prior clinical knowledge, which included treatment effect, age, sex, atrial fibrillation, SBP, NYHA and the presence of diabetes. As patients underwent randomisation, treatment was only included as the effect of treatment over time (as modelled by the natural splines) as per common guidance [69]. By default, the LME included an intercept term (the predicted mean when all other variables are zero).

5.3.6.4 Random Effects

For the random effects of the LME both a random intercept, and random slopes were included. With the random slopes incorporating the nonlinear effect of time modelled by the previously mentioned restricted natural cubic spline.

5.3.6.5 Model Formulation

The formula for the LME is denoted in Equation 15.

Equation 15 Formula for LME sub model

$$\left\{\begin{array}{c} \log eGFR_{ij} = \beta_0 + \beta_1 f_1(Specimen\ Time_{ij}) + \beta_2 f_2(Specimen\ Time_{ij}) + \beta_3 f_3(Specimen\ Time_{ij}) \\ + \beta_4 f_1(Specimen\ Time_{ij})\ Treatment\ LCZ_{ij} + \beta_5 f_2(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} + \beta_6 f_3(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} \\ + \beta_7 Age_{ij} + \beta_8 Sex_{ij} + \beta_9 Atrial\ Fibrillation\ Yes_{ij} + \beta_{10} SBP_{ij} + \beta_{11} Diabetes\ Yes_{ij} \\ \beta_{12} NYHA\ Class\ II_{ij} + \beta_{13} NYHA\ Class\ III_{ij} + \beta_{14} NYHA\ Class\ IV_{ij} \\ + b_{i0} + b_{i1} f_1 Specimen\ Time_{ij} + b_{i2} f_1 Specimen\ Time_{ij} + b_{i3} f_1 Specimen\ Time_{ij} + \varepsilon_{ij}, \\ b_i \sim \mathrm{N}(0, D), \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \end{array}\right\}$$

With β_0 representing the intercept term, and β_n representing the fixed effects and the natural cubic spline represented by the function f_n. Both random effects the random intercept b_{i0} and random slopes $b_{in}f_n$ (incorporating the natural cubic spline) are assumed normally distributed with a mean 0 and variance-covariance matrix D. The error term ε_{ij} is assumed to be normally distributed with a variance of σ^2.

5.3.6.6 Model Fitting

In order to fit the LME, the formatted data was provided to the *lme* function; the response variable was specified as log eGFR, and time was specified using the ns function specifying a restricted natural cubic spline. This time variable along with the interaction of it with treatment and the remaining covariates were supplied to the *lme* function using the fixed argument. The random slopes and intercepts were supplied via the random argument. The optimisation algorithm was manually specified as *optim*, which is a general-purpose optimising method which is based on Nelder-Mead, quasi-newton and conjugate-gradient algorithm, specified using the lmeControl argument, this optimiser is known to aid convergence [139].

5.3.7 Survival Sub Model

The *JMbayes2* package requires that the survival model be one fitted by either the function '*coxph*' or '*survreg*' from the *survival* package. The former allowing the fitting of a Cox PH model and the latter allowing a parametric survival model. A Cox PH model was preferred to avoid incorrect specification of the distribution.

5.3.7.1 Covariates

As previously stated, covariates were chosen based on prior clinical knowledge and were the same as the original trial's survival analysis, with the addition of log NT-ProBNP. Covariates included treatment, age, sex, region, BMI, eGFR, ejection fraction, NYHA classification, whether the patient had diabetes, SBP, heart rate, whether the patient had a history of atrial fibrillation, hospitalisation for heart failure, MI, or stroke.

5.3.7.2 Event (End Point)

The endpoint or event of interest was the primary outcome of the original trial. A composite of death for cardiovascular causes and first hospitalisation for heart failure.

5.3.7.3 Model Formulation

The formula for the survival sub model is denoted in Equation 16.

Equation 16 Formula of the Survival Sub model

$$h_i(t) = h_0(t) \exp (\gamma_1 Treatment\ LCZ_i + \gamma_2 Age_i + \gamma_3 Male_i + \gamma_4 Region\ Latin\ America_i + \gamma_5 Region\ North\ America_i + \gamma_6 Region\ Asia, Pacific\ and\ Other_{\ i} + \gamma_7 Region\ Western\ Europe_i + \gamma_8 BMI_i + \gamma_9 Log\ NT\text{-}ProBNP_i + \gamma_{10} Ejection\ Fraction_i + \gamma_{11} NYHA\ Class\ II_i + \gamma_{12} NYHA\ Class\ III_i + \gamma_{13} NYHA\ Class\ IV_i + \gamma_{14} Diabetes\ Yes_i + \gamma_{15} SBP_i + \gamma_{16} Heart\ Rate_i + \gamma_{17} Atrial\ Fibrillation\ Yes_i + \gamma_{18} Ischemic\ Heart\ Failure\ Yes_i + \gamma_{19} Prior\ Hospitalisation\ for\ Heart\ Failure\ Yes_i + \gamma_{20} Prior\ Miocardial\ Infarction\ Yes_i + \gamma_{21} Prior\ Stroke\ Yes_i)$$

The hazard of the composite event at time point t is represented by $h_i(t)$. The baseline hazard represented by $h_0(t)$, with γ_n representing the covariates.

5.3.7.4 Model Fitting

To fit the Cox PH model, the formatted and cleaned data was passed to the *coxph* function from the survival package using the data argument. The model specified as time to the month at which the patient was censored, or the composite event occurred, along with a status indicator of whether the patient experienced the composite event. This was specified along with the previously detailed covariates using the *formula* argument.

5.3.8 Joint Models

Joint Modelling allows for different alpha parameterisations i.e., the association parameter in the survival sub-model. Three models were fit with different parameterisations, chosen for their suitability for prognostic models; these models included the aforementioned sub-models both the Cox PH model and the LME model. Joint Model 1 included a value alpha parameterisation. Joint Model 2 included a slope and value alpha parameterisation. Finally,

Joint Model 3 included an area alpha parameterisation. The value alpha parameterisation measures the association with the value of log eGFR and the composite event. The slope alpha parameterisation measures the association of the rate of change of log eGFR and the composite end point. The area parameterisation measures the association of the area under the trajectory of log eGFR and the composite endpoint.

All models were fit using the *JMbayes2* package through use of the *jm* function fitting a joint model with a piecewise baseline hazard function, with quadratic B-splines and 10 baseline hazard segments.

The formulae for all the joint models are shown in Table 29, each joint model contains the same longitudinal and survival components, except for the alpha parameter which is therefore listed separately for each model.

5.3.8.1 Model Fitting

To fit the joint models, the previously fit LME was passed to the *jm* function using the *Mixed_object* argument, similarly the Cox PH model was passed using the *Surv_objects* argument. Additionally, the parameterisation of the longitudinal component within the survival component was specified using the *functional_forms* argument.

Table 29 Formulae for Joint Models

Base Joint Model	$$\begin{cases} y_i(t) = m_i(t) + \varepsilon_{i\,(t)} \\ = \beta_0 + \beta_1 f_1(Specimen\ Time) + \beta_2 f_2(Specimen\ Time) + \beta_3 f_3(Specimen\ Time) \\ + \beta_4 f_1(Specimen\ Time) Treatment\ LCZ + \beta_5 f_2(Specimen\ Time) Treatment\ LCZ + \beta_6 f_3(Specimen\ Time) Treatment\ LCZ \\ + \beta_7 Age + \beta_8 Sex + \beta_9 Atrial\ Fibrillation\ Yes + \beta_{10} BMI \\ + \beta_{11} NYHA\ Class\ II + \beta_{12} NYHA\ Class\ III + \beta_{13} NYHA\ Class\ IV \\ + b_{i0} + b_{i1} f_1 Specimen\ Time + b_{i2} f_1 Specimen\ Time + b_{i3} f_1 Specimen\ Time + \varepsilon_i, \quad \varepsilon_i(t) \sim N(0, \sigma^2), \\ h_i(t) = h_0(t)\, exp \left\{ \begin{matrix} \gamma_1 Treatment\ LCZ_i + \gamma_2 Age_i + \gamma_3 Male_i + \gamma_4 Region\ Latin\ America_i + \gamma_5 Region\ North\ America_i \\ + \gamma_7 Region\ Western\ Europe_i + \gamma_8 BMI_i + \gamma_9 Log\ NT\text{-}ProBNP_i + \gamma_{10} Ejection\ Fraction_i \\ + \gamma_{11} NYHA\ Class\ II_i + \gamma_{12} NYHA\ Class\ III_i + \gamma_{13} NYHA\ Class\ IV_i \\ + \gamma_{14} Diabetes\ Yes_i + \gamma_{15} SBP_i + \gamma_{16} Heart\ Rate_i + \gamma_{17} Atrial\ Fibrillation\ Yes_i + \gamma_{18} Ischemic\ Heart\ Failure\ Yes_i \\ + \gamma_{19} Prior\ Hospitalisation\ for\ Heart\ Failure\ Yes_i + \gamma_{20} Prior\ Miocardial\ Infarction\ Yes_i + \gamma_{21} Prior\ Stroke\ Yes_i \\ + \{Alpha\ Parameter(s)\} \end{matrix} \right\}, \\ log\ h_0(t) = \gamma h_0, 0 + \sum_{q=0}^{Q} \gamma h_{0,q}\, B_q(t, v), \end{cases}$$
Joint Model 1 Alpha Parameter (Value)	$\alpha m_i(t)$
Joint Model 2 Alpha Parameter (Value + Slope)	$\alpha m_i(t) + \alpha_2 m_i'(t)$
Joint Model 3 Alpha Parameter (Area)	$\alpha \frac{\int_0^t m_i(s)\, ds}{t}$

Where $y_i(t)$ is the is the longitudinal outcome at time point t, comprising of the true and unobserved value of log eGFR at time point t $(m_i(t))$ and error term $\varepsilon_{i\,(t)}$. β_0 is the intercept term for the longitudinal outcome. The β parameters in the longitudinal models are the coefficients for the covariates, with the natural cubic spline represented as the functions f_n. The Random effects of the longitudinal outcome are represented by b_{in}. The error term of the longitudinal outcome is assumed to be normally distributed with a mean of 0 and a variance of σ^2. $h_i(t)$ is the hazard of the composite event at time point t. With $h_0(t)$ representing the baseline hazard function, comprised of a piecewise hazard function, with quadratic B-splines and 10 baseline hazard segments. Where $B_q(t, v)$ relates to the q-th basis function of the B-spline with knots v_1- v_{10} and γh_0 being a vector of spline coefficients. The covariates of the survival components are represented by the remaining γ_n**.** The $m_i'(t)$ slope parameter corresponds to the rate of change of the longitudinal outcome at time point t estimated using the derivative of the fixed and random effects of the longitudinal outcome with respect to time. The area parameter corresponds to the whole area under the longitudinal trajectory, which accounting for the observation period, and is represented as a function.

5.3.9 Model Performance: Prognostic Accuracy, Fit and Calibration

This analysis takes advantage of the functions available in *JMbayes2*, these include time-varying ROCs and AUCs produced by the functions *tvROC* and *tvAUC* respectively, as well as the time-varying Brier score and time-varying calibration curves, produced by the functions *tvBrier* and *calibration_plot* respectively. These metrics extend on the original metrics using the longitudinal data up until a specified time point and predicting the future survival outcome [52], [142] and are interpreted like the original metrics however they provide metrics at specific time points. These measures were used to assess the prognostic accuracy and fit. Each of the measures was calculated at month 12, first using covariates only measured at baseline, and then using longitudinal covariate data up until 12 months.

All three of the joint models were compared against each other using LPML and DIC as well as comparing hazard ratios and 95% CIs. LPML being like the log likelihood of a frequentist model and can be interpreted as such, with a larger value indicating a better fit. A DIC is interpreted like any other information criterion with a lower value indicating a better fit. For both metrics, the conditional value relates to subject specific fit whereas the marginal value relates to the population level fit.

5.3.10 External validation

As stated in Chapter 3, model validation was achieved using data obtained from the ATMOSPHERE randomised control trial. In which 7016 patients were randomised at a 1:1:1 ratio to receive either a combination of enalapril (twice daily at a dose of 5 or 10mg) and aliskiren (once daily at a dose of 150mg), or enalapril twice daily at a dose of 5 or 10mg, or aliskiren once daily at a dose of 150mg). The trial used a double-blind, double dummy method.

Data from the ATMOSPHERE trial needed to be recoded into the same format as the data from the PARADIGM trial including the transformation of both eGFR and NT-ProBNP using the natural logarithm. Time was modelled on the same scale (months since randomisation using a 28-day calendar month). Treatment was recoded into enalapril and LCZ the former containing the enalapril arm and the latter containing the remaining arms.

To evaluate the models using the validation data, the time-varying metrics (AUC, ROC, Brier score and calibration curves) were used, supplying the ATMOSTSPHERE data to the *newdata* argument of the functions. This argument allows the validation using a new dataset, which is used for predictions based on the original joint models. As with the original data, the measures were calculated at month 12 using baseline data and month 24 using longitudinal data up until time point 12.

5.3.11 Comparative Analysis

The joint models were compared against Cox PH Models (the current standard for prognostic models) to illustrate how joint models differ. Two Cox PH models were fit, one using the LOCF method and one using an extended Cox PH model (also known as the time-varying Cox PH model). Due to the previously mentioned limited performance metrics the comparison of the joint models with the Cox PH models was done using time-varying AUCs and ROCs and by comparison of the hazard ratios and corresponding CIs. The time-varying AUCs and ROCs for the Cox PH models were calculated using the function *risksetROC* from the *risksetROC* package which is defined by the formula by Heagerty et al., [142], [143].

Descriptive statistics for the distribution of categorical variables are expressed as percentages, with continuous variables being summarised with medians [Q1, Q3]. The association parameters from the joint models are represented by hazard ratios with corresponding 95% CI.

The time dependent association value parameter represents the hazard of a composite event given a unit increase in log eGFR at any given time point. The time dependent association slope parameter represents the hazard of the composite event given a one unit increase in the slope of the trajectory of log eGFR at any given time point. The association area parameter represents the hazard of the composite event given a unit increase in the area under the longitudinal profile of log eGFR at any given time point.

P-Values are considered statistically significant if below the threshold of 0.05. All statistical analysis was performed using *JMbayes2* package version 0.1-81 [114] and R version 4.0 [144].

5.4 Results

5.4.1 Baseline Characteristics

The baseline characteristics of the 8344 included patients are shown in Table 30, with a median age of 64 years and 22% of patients being female. As with the main trial there are only minimal differences between the arms for the distributions of baseline characteristics.

Table 30 Baseline Characteristics of Included Patients.

Characteristic	**LCZ (N=4158)**	**Enalapril (N=4186)**	**Overall (N=8344)**
Age, years			
Median [Q1, Q3]	64.0 [57.0, 72.0]	64.0 [57.0, 72.0]	64.0 [57.0, 72.0]
Sex			
Female	872 (21.0%)	949 (22.7%)	1821 (21.8%)
Male	3286 (79.0%)	3237 (77.3%)	6523 (78.2%)
Region			
Central Europe	1390 (33.4%)	1429 (34.1%)	2819 (33.8%)
Latin America	708 (17.0%)	711 (17.0%)	1419 (17.0%)
North American	307 (7.4%)	290 (6.9%)	597 (7.2%)
Pacific Asia/Pacific and Other	739 (17.8%)	735 (17.6%)	1474 (17.7%)

Characteristic	LCZ (N=4158)	Enalapril (N=4186)	Overall (N=8344)
Western Europe	1014 (24.4%)	1021 (24.4%)	2035 (24.4%)
BMI kg/m²			
Median [Q1, Q3]	27.5 [24.4, 31.2]	27.5 [24.5, 31.2]	27.5 [24.4, 31.2]
Ejection Fraction %			
Median [Q1, Q3]	30.0 [25.0, 34.0]	30.0 [25.0, 34.5]	30.0 [25.0, 34.1]
NYHA Class			
Class II	179 (4.3%)	207 (4.9%)	386 (4.6%)
Class II	2981 (71.7%)	2906 (69.4%)	5887 (70.6%)
Class III	967 (23.3%)	1046 (25.0%)	2013 (24.1%)
Class IV	31 (0.7%)	27 (0.6%)	58 (0.7%)
Diabetes			
No	2715 (65.3%)	2736 (65.4%)	5451 (65.3%)
Yes	1443 (34.7%)	1450 (34.6%)	2893 (34.7%)
SBP mmHg			
Median [Q1, Q3]	120 [110, 130]	120 [110, 130]	120 [110, 130]
Heart Rate beats per minute			
Median [Q1, Q3]	71.0 [64.0, 80.0]	72.0 [64.0, 80.0]	71.0 [64.0, 80.0]
Prior History of Atrial Fibrillation			
No	2652 (63.8%)	2622 (62.6%)	5274 (63.2%)
Yes	1506 (36.2%)	1564 (37.4%)	3070 (36.8%)
Ischemic Heart Failure			
No	1667 (40.1%)	1675 (40.0%)	3342 (40.1%)
Yes	2491 (59.9%)	2511 (60.0%)	5002 (59.9%)
Prior History of Hospitalisation for Heart Failure			
No	1565 (37.6%)	1536 (36.7%)	3101 (37.2%)
Yes	2593 (62.4%)	2650 (63.3%)	5243 (62.8%)
Prior History of Myocardial Infarction			
No	2350 (56.5%)	2381 (56.9%)	4731 (56.7%)
Yes	1808 (43.5%)	1805 (43.1%)	3613 (43.3%)
Prior History of Stroke			
No	3804 (91.5%)	3817 (91.2%)	7621 (91.3%)
Yes	354 (8.5%)	369 (8.8%)	723 (8.7%)
NT-ProBNP pg/mL			
Median [Q1, Q3]	1630 [887, 3160]	1600 [888, 3310]	1610 [888, 3230]
eGFR mL/min/1.73 m2			
Median [Q1, Q3]	66.0 [54.0, 79.0]	66.0 [53.0, 79.0]	66.0 [54.0, 79.0]

5.4.2 Number of Measurements of eGFR

Table 31 shows the number of repeat measurements of eGFR for the included patients. With the majority of patient having between 6-10 measurements (42%) and 11-15 measurements (43%).

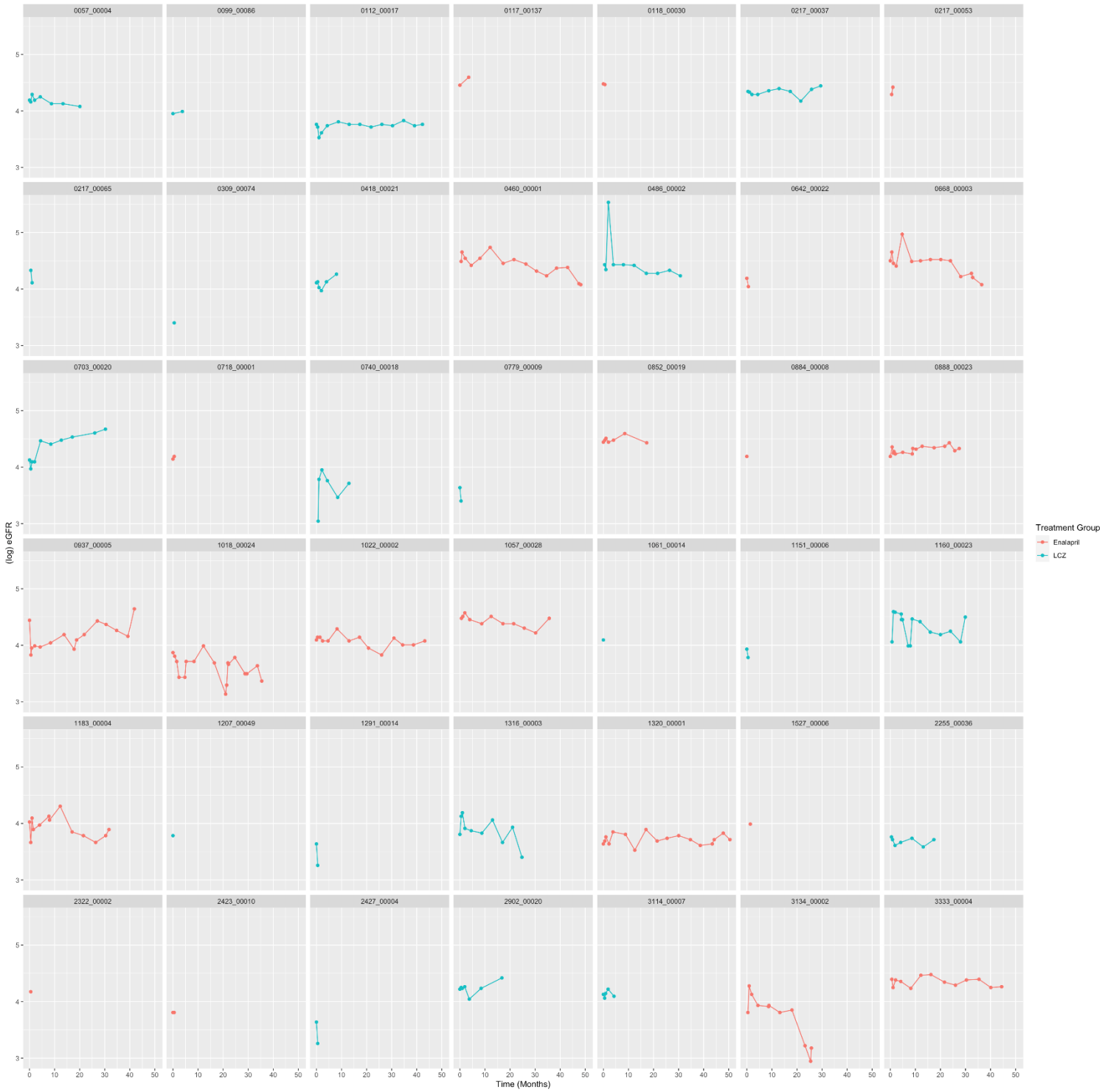
Table 31 Number of Repeat Measurements of eGFR for Included Patients

1-5	6-10	11-15	16-20	21+
736 (9%)	3520 (42%)	3611 (43%)	432 (5%)	45 (<1%)

5.4.3 Longitudinal Profile of eGFR

Figure 27 illustrates the longitudinal profiles of 42 randomly sampled patients (stratified by number of measurements), these profiles indicate some patients such as 1183_0004 showed signs of non-linearity in values of log eGFR over time. As eGFR has been transformed into a log scale, the non-linearity of eGFR in these profiles over time would be amplified on the original scale.

Figure 27 Longitudinal Profiles of log eGFR



5.4.4 Longitudinal Outcome from Joint Models

Parameter estimates from the longitudinal outcome from the joint models are shown in Table 32 for all meaningfully interpretable variables. Variables with natural splines were excluded as they cannot be meaningfully interpreted directly from the parameter estimates, instead they are illustrated graphically in Figure 28. All variables apart from the NYHA Class variable are considered significantly significant with a P-Value of less than 0.001. The variables age, having atrial fibrillation and having diabetes all have a small negative effect on log eGFR, with being of the male sex, and a unit increase of SBP both having positive effects on log eGFR. It should be noted that a higher eGFR is considered better, and therefore a negative effect would be worse for the patient. It is important to highlight that eGFR is on a log scale and time is modelled in months therefore this should be taken into consideration when interpreting these values, and as such the scale of the effects may be higher than the model appears to represent them as.

There are only small differences between the models with respect to both coefficients and 95 % CIs, for example, age has only a difference of 0.001 in 95% CIs between the models, and being of the male sex, has only a difference of 0.001 in the parameter estimate and 0.001 in 95% CI for joint model two compared to Joint Model 1 and Joint Model 3. There are only differences in p-values for the NYHA class variables, however none of these variables are considered statistically significant.

The coefficients of the natural splines cannot be meaningfully interpreted as they only represent parts of the spline function. However, the natural splines can be graphically represented as illustrated in Figure 28 which illustrates the average predicted trajectories of log eGFR by each of the joint models, with the y axis for log eGFR being scaled for

readability. All the models show similar average trajectories of log eGFR with eGFR declining on average in each treatment group, and a larger treatment effect as time progresses. However the overlapping 95% CIs suggest a lack of certainty around a difference in treatment effect on log eGFR over time on average.

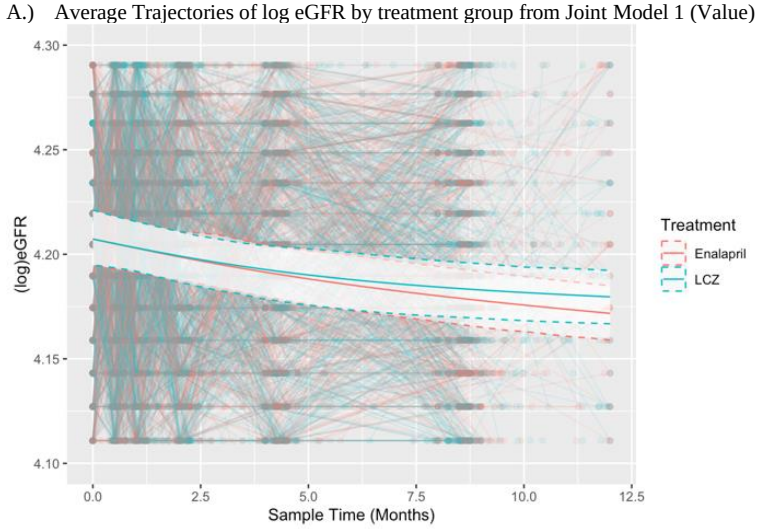
Table 32 Parameter Estimates (95% CI) from the Longitudinal Outcomes of the Joint Models for the Meaningfully Interpretable Variables

	Joint Model 1 Value		Joint Model 2 Value + Slope		Joint Model 3 Area	
Variable	**Estimate (95% CI)**	**P-Value**	**Estimate (95% CI)**	**P-Value**	**Estimate (95% CI)**	**P-Value**
Age	-0.010 (-0.011 - -0.009)	< 0.001	-0.010 (-0.011 - -0.009)	< 0.001	-0.010 (-0.011 - -0.009)	< 0.001
Sex - Male	0.041 (0.025 - 0.057)	< 0.001	0.040 (0.024 - 0.057)	< 0.001	0.041 (0.025 - 0.057)	< 0.001
Atrial Fibrillation - Yes	-0.030 (-0.045 - -0.016)	< 0.001	-0.030 (-0.045 - -0.016)	< 0.001	-0.030 (-0.045 - -0.016)	< 0.001
SBP	0.001 (0.001 - 0.002)	< 0.001	0.001 (0.001 - 0.002)	< 0.001	0.001 (0.001 - 0.002)	< 0.001
NYHA Class II	0.001 (-0.03 - 0.032)	0.940	0.001 (-0.03 - 0.033)	0.934	0.001 (-0.03 - 0.032)	0.934
NYHA Class III	-0.011 (-0.045 - 0.023)	0.523	-0.011 (-0.045 - 0.023)	0.507	-0.011 (-0.045 - 0.023)	0.525
NYHA Class IV	0.020 (-0.069 - 0.111)	0.666	0.020 (-0.069 - 0.110)	0.658	0.020 (-0.069 - 0.111)	0.667
Diabetes - Yes	-0.049 (-0.063 - -0.034)	< 0.001	-0.048 (-0.062 - -0.034)	< 0.001	-0.049 (-0.063 - -0.035)	< 0.001

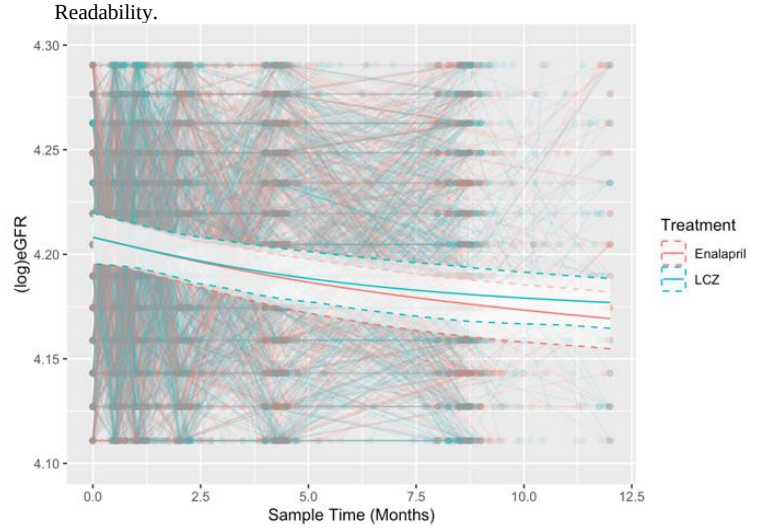
Parameter estimates for variables including natural splines are excluded as they cannot be meaningfully interpreted, instead they are illustrated in Figure 28.

Figure 28 Average Trajectories of log eGFR by Treatment Group with Observed Values and Trajectories of log eGFR Scaled for Readability

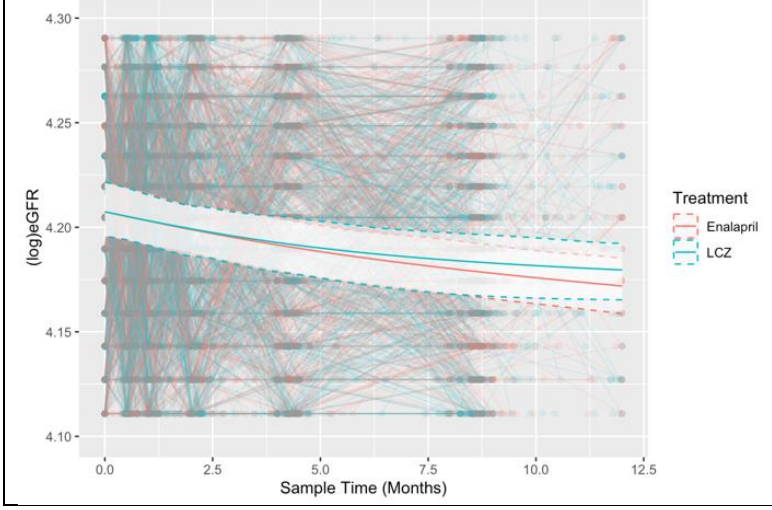
A.) Average Trajectories of log eGFR by treatment group from Joint Model 1 (Value) with Observed Values and Trajectories Scaled for Readability.



B.) Average Trajectories of log eGFR by treatment group from Joint Model 2 (Value and Slope) with Observed Values and Trajectories Scaled for Readability.



C.) Average Trajectories of log eGFR by treatment group from Joint Model 3 (Area) with Observed Values and Trajectories Scaled for Readability.



The average trajectories we predicted using the respective joint models, using the mode and median characteristics of each stratum of patients (the treatment group). The y axis log(eGFR has been scalled for readability).

5.4.5 Survival outcomes from Joint Models

Table 33 shows the HRs with 95% CIs and corresponding P-Values for the three joint models. There are only minor differences in the hazard ratios between the models, with the greatest difference being the log NT-ProBNP and the log eGFR parameters. With Joint Model 2 having a lower hazard ratio at 1.45 when compared to the 1.48 from Joint Model 1 and 1.49 from Joint Model 3. The log eGFR parameters also have differences, with the value parameter of Joint Model 1 having a lower HR of 0.42 (0.36 - 0.5) than both the value parameter of Joint Model 2 with a HR of 0.49 (0.41 – 0.58) and Joint Model 3 with an area parameter with a HR of 0.48 (0.41 – 0.58). This suggests that the hazard of the composite event decreases by 0.42 times per unit increase in log eGFR at any time point when using Joint Model 1. Whereas Joint Model 2 suggests the hazard of the composite event decreases by 0.49 times per unit increase in log eGFR at any time point for patients with the same slope. Finally, Joint Model 3 suggests a decrease in the hazard of the composite event of 0.48 per unit increase in area under the longitudinal profile of log eGFR at any time point. It can be observed that the slope parameter, while significant is exceptionally small and therefore uninterpretable, this means while the slope may have a significant effect it is almost unmeasurable. All alpha parameters are below the significance threshold suggesting log eGFR is significantly associated with the hazard of the composite event.

Except for Joint Model 2, the joint models share the same significant parameters, with the history of atrial fibrillation being significant in Joint Model 2. A variable of interest that is not considered significant is age as clinically a unit increase in age should have a significant effect on the hazard of adverse outcomes such as the composite outcome. One possible reason for age not being considered statistically significant is the inclusion of it in the longitudinal sub-model.

Table 33 Hazard Ratios (95% CI) from the Survival Component of the Three Joint Models

	Joint Model 1 (Value)		Joint Model 2 (Value and Slope)		Joint Model 3 (Area)	
Variable	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**
Treatment LCZ	0.81 (0.73 - 0.9)	< 0.001	0.82 (0.74 - 0.92)	< 0.001	0.81 (0.73 - 0.89)	< 0.001
Age	1 (0.99 - 1)	0.055	1 (0.99 - 1)	0.197	1 (0.99 - 1)	0.22
Sex – Male	1.34 (1.2 - 1.49)	< 0.001	1.32 (1.18 - 1.48)	< 0.001	1.32 (1.18 - 1.49)	< 0.001
Region Latin America	1.03 (0.9 - 1.2)	0.668	1.02 (0.88 - 1.18)	0.828	1.04 (0.91 - 1.21)	0.54
Region North American	0.92 (0.77 - 1.1)	0.357	0.9 (0.76 - 1.06)	0.21	0.94 (0.79 - 1.11)	0.485
Region Pacific Asia/Pacific and Other	1.14 (0.98 - 1.31)	0.091	1.13 (0.97 - 1.31)	0.105	1.15 (1 - 1.32)	0.054
Region Western Europe	0.87 (0.77 - 0.98)	0.022	0.86 (0.75 - 0.98)	0.025	0.88 (0.78 - 1)	0.047
BMI	1.01 (1 - 1.02)	0.01	1.01 (1 - 1.02)	0.016	1.01 (1 - 1.02)	0.007
Ejection Fraction	0.98 (0.98 - 0.99)	< 0.001	0.98 (0.98 - 0.99)	< 0.001	0.98 (0.98 - 0.99)	< 0.001
NYHA Class II	1.24 (0.97 - 1.63)	0.081	1.25 (0.99 - 1.62)	0.061	1.24 (0.98 - 1.58)	0.075
NYHA Class III	1.58 (1.24 - 2.13)	< 0.001	1.6 (1.25 - 2.11)	< 0.001	1.59 (1.22 - 2.05)	< 0.001
NYHA Class IV	1.91 (1.14 - 3.18)	0.016	1.93 (1.13 - 3.15)	0.012	1.89 (1.12 - 3.07)	0.024
Diabetes – Yes	1.32 (1.2 - 1.45)	< 0.001	1.29 (1.17 - 1.41)	< 0.001	1.34 (1.22 - 1.47)	< 0.001
SBP	1 (1 - 1)	0.218	1 (0.99 - 1)	0.114	1 (1 - 1)	0.228
Heart Rate	1.01 (1 - 1.01)	0.003	1.01 (1 - 1.01)	0.002	1.01 (1 - 1.01)	0.002
History of Atrial fibrillation – Yes	1.1 (1 - 1.21)	0.056	1.12 (1.02 - 1.23)	0.02	1.09 (0.99 - 1.21)	0.081
Ischemic Heart Failure – Yes	0.99 (0.88 - 1.12)	0.89	0.99 (0.87 - 1.12)	0.823	0.99 (0.87 - 1.13)	0.894
Prior Hospitalisation for Heart Failure - Yes	1.37 (1.24 - 1.51)	< 0.001	1.39 (1.26 - 1.54)	< 0.001	1.37 (1.25 - 1.51)	< 0.001
Prior History of Myocardial Infarction - Yes	1.19 (1.05 - 1.34)	0.003	1.21 (1.07 - 1.36)	0.001	1.19 (1.05 - 1.35)	0.006
Prior History of Stroke - Yes	1.08 (0.93 - 1.25)	0.313	1.1 (0.95 - 1.26)	0.211	1.09 (0.94 - 1.25)	0.255
Log NT-ProBNP	1.48 (1.41 - 1.54)	< 0.001	1.45 (1.39 - 1.52)	< 0.001	1.49 (1.42 - 1.57)	< 0.001
Value of log eGFR	0.42 (0.36 - 0.5)	< 0.001	0.49 (0.41 - 0.58)	< 0.001		
Slope of log eGFR			< 0.01*	< 0.001		
Area of log eGFR					0.48 (0.41 - 0.58)	< 0.001

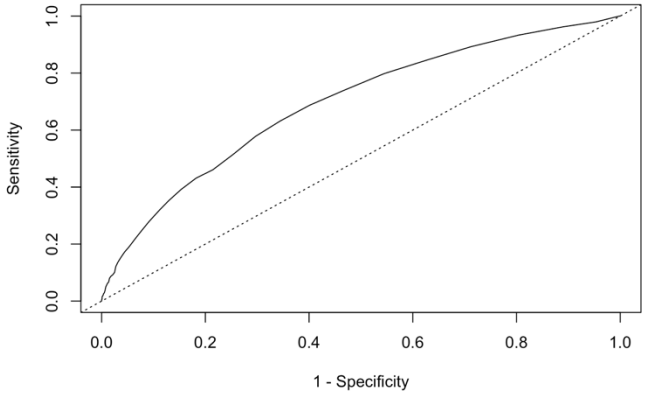
* The HR from the model was less than 0.01 when rounded to two decimal places.

Table 34 Marginal and Conditional Performance Statistics of Joint Models

	Marginal			Conditional		
Statistic	Joint Model 1 Value	Joint Model 2 Value + Slope	Joint Model 3 Area	Joint Model 1 Value	Joint Model 2 Value + Slope	Joint Model 3 Area
DIC	-41264.18	-41353.59	-41227.16	-68820.58	-69086.27	-68867.06
LPML	20684.51	20806.83	20683.34	27927.59	28052.39	27996.73

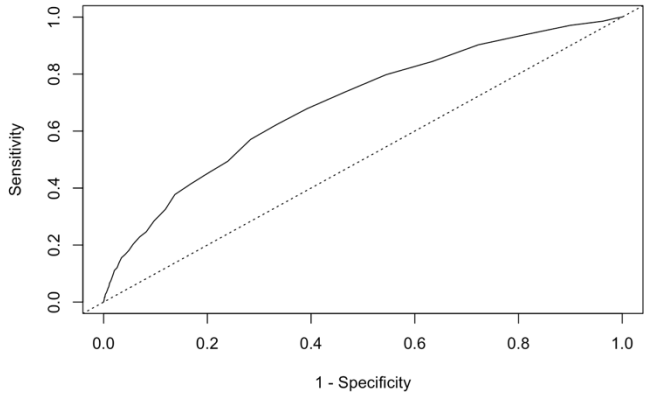
Figure 29 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months

A.) Joint Model 1 – Value: Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



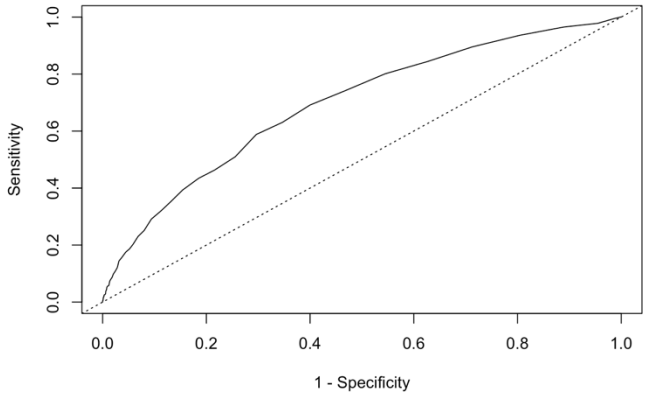
Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

B.) Joint Model 2 - Value and Slope: Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



Time-Varying AUC: 0.70
Time-Varying Brier Score: 0.10

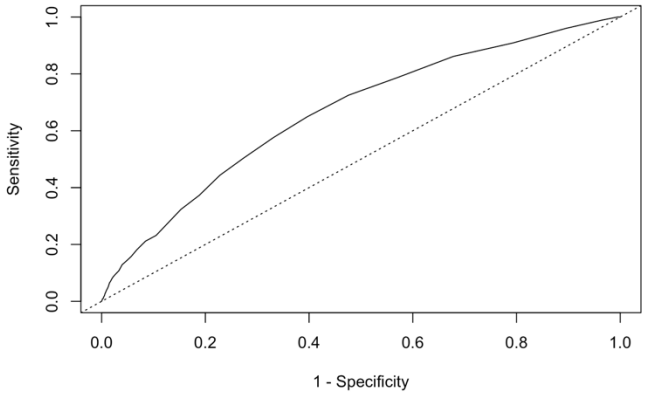
C.) Joint Model 3 – Area: Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

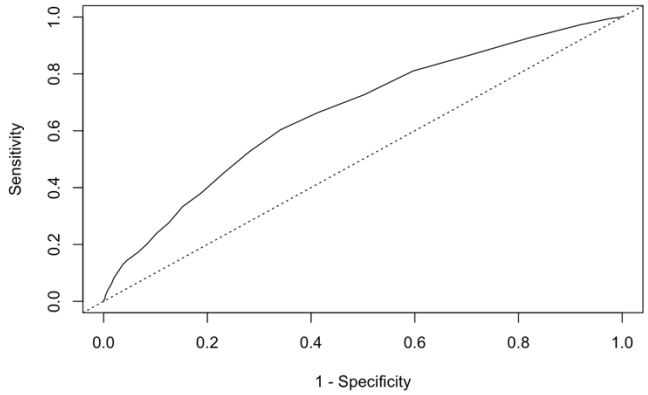
Figure 30 Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months

A.) Joint Model 1 – Value: Time-Varying ROC, AUC and Brier Score at Month 24 Using Longitudinal Data up until Month 12



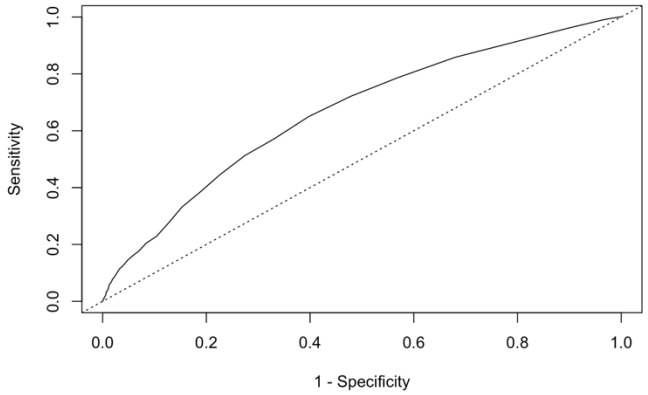
Time-Varying AUC: 0.66
Time-Varying Brier Score: 0.08

B.) Joint Model 2 - Value and Slope: Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data up until Month 12



Time-Varying AUC: 0.66
Time-Varying Brier Score: 0.08

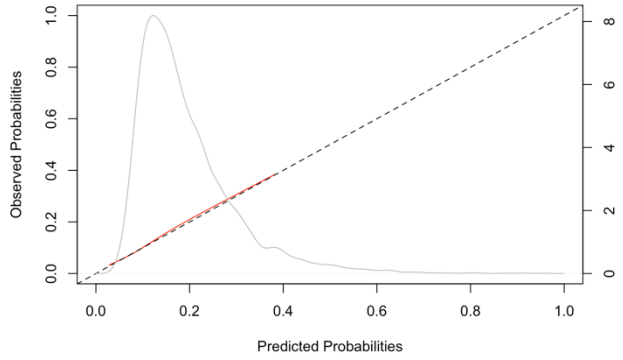
C.) Joint Model 3 – Area: Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data up until Month 12



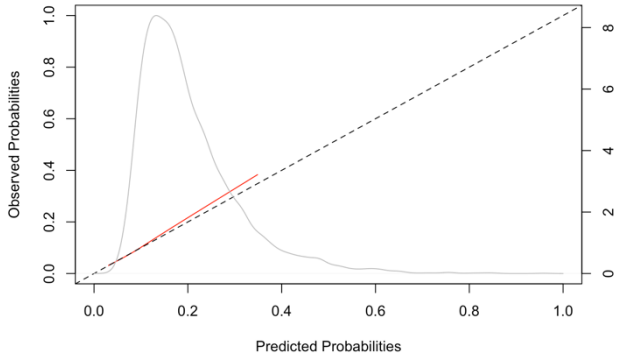
Time-Varying AUC: 0.66
Time-Varying Brier Score: 0.08

Figure 31 Calibration Curves for Joint Models at Month 12 using Longitudinal Data at Month 0

A.) Joint Model 1 – Value: Calibration Curve at Month 12 using Longitudinal Data at Month 0



B.) Joint Model 2 – Value and Slope: Calibration Curve at Month 12 using Longitudinal Data at Month 0



C.) Joint Model 3 – Area: Calibration Curve at Month 12 using Longitudinal Data at Month 0

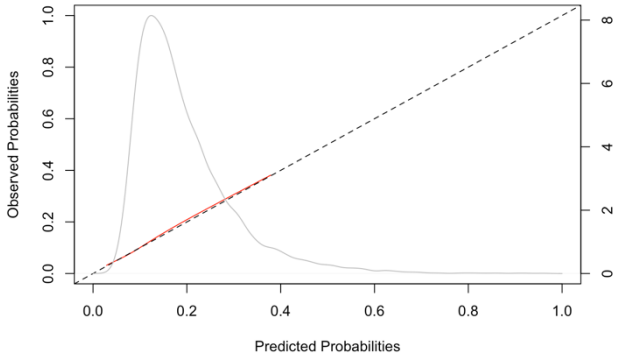
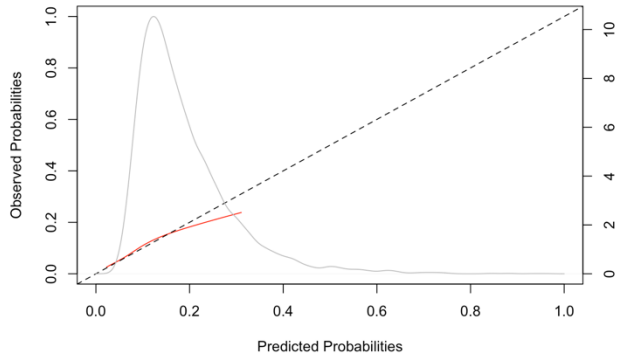
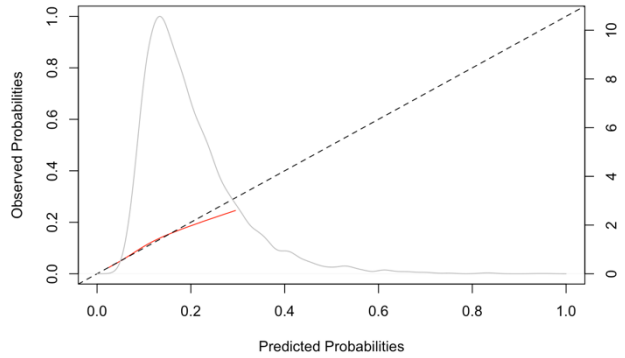


Figure 32 Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12

A.) Joint Model 1 – Value: Calibration Curve at Month 24 using Longitudinal Data up until Month 12



B.) Joint Model 2 – Value and Slope: Calibration Curve at Month 24 using Longitudinal Data up until Month 12



C.) Joint Model 3 – Area: Calibration Curve at Month 24 using Longitudinal Data up until Month 12

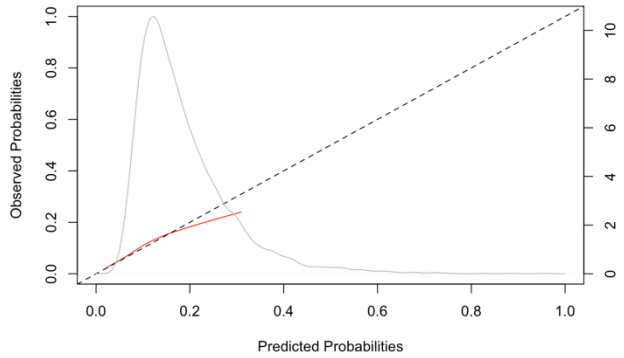
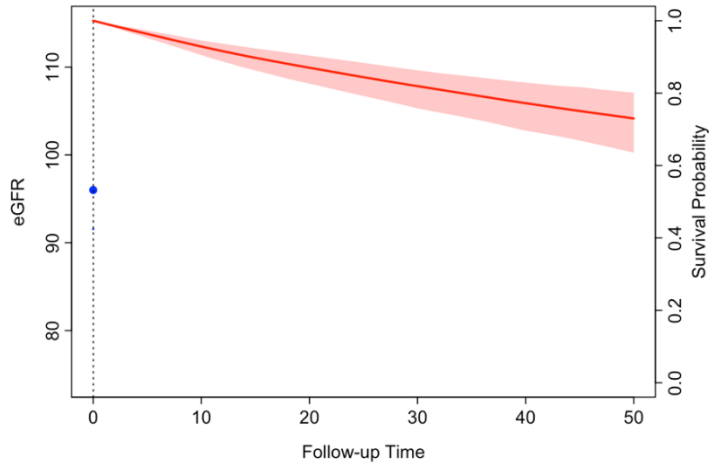


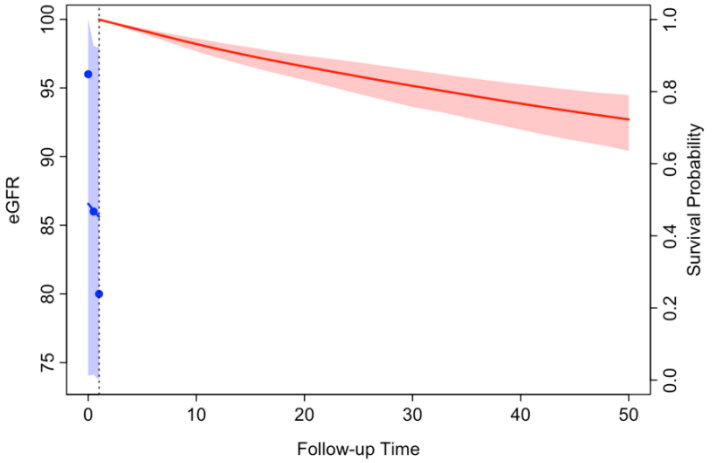
Figure 33 Dynamic Predictions from Joint Model 1 for a Randomly Sampled Patient with Measurements of eGFR and Corresponding Survival Probability

A.) Observation 1, eGFR Measurement at Month 0 with Corresponding Survival Probability



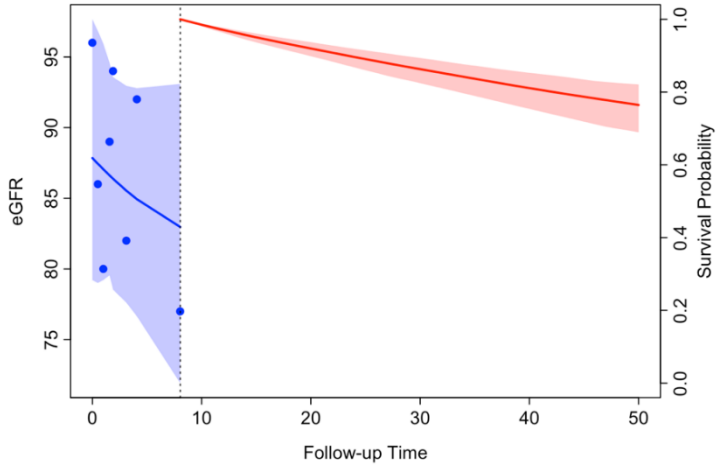
Survival at 24 Months: 0.85 (0.89 - 0.80)

B.) Observation 3, eGFR Measurements at Months 0, 0.5 and 1 with Corresponding Survival Probability



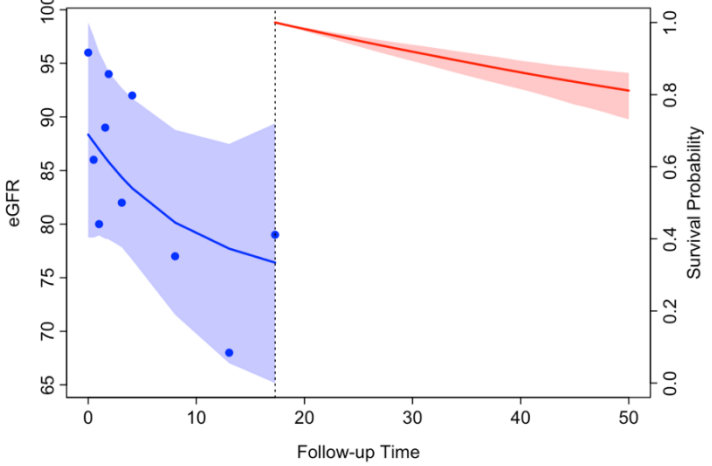
Survival at 24 Months: 0.85 (0.88 - 0.80)

C.) Observation 8, eGFR Measurements at Months 0, 0.5, 1, 1.5, 2, 3, 4 and 8 with Corresponding Survival Probability



Survival at 24 Months: 0.90 (0.92 - 0.87)

D.) Observation 10, eGFR Measurements at Months 0, 0.5, 1, 1.5, 2, 3, 4, 8, 13, and 17 with Corresponding Survival Probability



Survival at 24 Months: 0.96 (0.97 - 0.94)

eGFR is represented by the blue line with 95% CI, the dashed line represents the last time the patient was known to be alive, and the red line represents the survival probability with 95% CI.

5.4.6 Joint Model Performance

The marginal and conditional DICs and LPMLs for each of the joint models are shown in Table 34. Both DIC and LPML for Joint Model 2 suggest that this joint model performs best with regards to marginal fit, with an LPML of 2086.83 and a DIC of -41353.59 which are the highest and lowest respectively out of all the joint models. Similarly, the DIC and LPML for Joint Model 2 suggests that this model also performs best with regards to conditional fit, having the lowest DIC of -69086.27 and highest LPML of 28052.39.

The time Varying ROCs, AUCs and Brier scores for the Joint Models at month 12 using longitudinal data at baseline are shown in Figure 29. Joint Model 1 and Joint Model 3 perform equally with respect to time-varying AUC and Brier score with time-varying AUCs of 0.69 and time-varying Brier scores of 0.10, whereas Joint Model 2 performs slightly better with a time-varying AUC of 0.70 but the same time-varying Brier score of 0.10. The models only show minor variance in the time-varying ROCs. The time-varying AUCs of Joint Model 1 and Joint Model 3 suggest the models perform just below the acceptable threshold for acceptable discrimination instead falling into the top end of the poor discrimination category [38]. However, the time-varying ROC of Joint Model 2 suggests it is within the acceptable threshold for a prognostic model.

The time-varying ROCs AUCs and Brier scores shown in Figure 30 show similar performance between the joint models at month 24 using longitudinal data up until month 12 with Joint Model 2 and 3 performing the same, with a time-varying AUC of 0.67 and a time-varying Brier score of 0.08, with Joint Model 1 having a time-varying AUC of 0.66 and the same time-varying Brier score of 0.08. These metrics suggest a drop in discrimination with respect to time-varying AUC from the 12-month metrics using longitudinal data at baseline

but an increase with respect to Brier score in which the Brier score improved by 0.02. The time-varying AUC suggesting the discrimination of the models at 24 months using longitudinal data up until month 12 are still in the poor discrimination category.

The calibration curves shown in Figure 31 for the joint models at month 12 using baseline longitudinal data suggest that Joint Model 1 and Joint Model 3 are well calibrated at this time point, whilst Joint Model 2 shows signs of deviation nearer the higher end of the probabilities, suggesting minor fitment issues.

Figure 32 shows the calibration curves for all the joint models at month 24 using longitudinal data up until 12 months, with all models showing issues around fit nearer the higher probabilities suggesting issues with the fit of the models. All models performed worse in fit than at the previous time point of 12 months using baseline data.

5.4.7 Dynamic Predictions from Joint Models

Dynamic predictions from a randomly sampled patient at 4 points in time from Joint Model 1 (chosen for its better conditional fit) can be seen in Figure 33. Showing both observed measurements of eGFR and predicted trajectory of eGFR on the left, and the corresponding predicted survival probability on the right. From these dynamic predictions it can be observed that the confidence intervals are narrower where there are more measurements closer to the predicted trajectory. The first two time points (month 0 and 1) the patient has the same predicted probability of survival of 0.85 at 24 months. However, this increases to a survival probability of 0.90 at 24 months at month 8, and whilst eGFR did decrease the patient had survived longer. The survival probability at 24 months at time point 17 increases to 0.96 once again likely to the patient surviving longer, as well as predicted trajectory of eGFR appears to be stabilising nearer the timepoint of interest.

5.4.8 External Validation

Figure 34 and Figure 35 show the time-varying ROCs, AUCs and Brier scores for the joint models at 12 months using baseline longitudinal data and 24 months using longitudinal data up until time point 12 using data from the ATMOSPHERE trial respectively. For both time points there is no difference in time-varying ROCs or AUCs between the models. With the models having a time-varying AUC of 0.67 and Brier score of 0.10 at month 12 using baseline data and 0.65 and 0.09 respectively at month 24 using longitudinal data up until time point 12. The time-varying ROCs only show minor differences between models. These results suggest that the model performance was worse at 24 months using longitudinal data at time point 12 than it was at month 12 using baseline data from the ATMOSPHERE validation dataset with both the validation and original data. The difference is only 0.02 for the time-varying AUC of Joint Model 1 and Joint Model 2 at both time points. Both between time points and with the original data at the same time points, but this is enough to move the models from the upper limit of the poor discrimination category and further away from the acceptable category. The difference in time-varying AUC for Joint Model 3 is 0.02 between time points and between the original data at month 24 using longitudinal data up until month 12. However, the difference at month 12 using baseline data is 0.03 from the original data; enough to move the model from the acceptable discrimination category into the poor discrimination category. The Brier scores suggest the models perform slightly better at the second time point with a lower Brier score of 0.9 but this is only a minor improvement and is possibly due to the amount of data available, as both the time-varying AUC and calibration curves suggest a worse prognostic performance. The models have the same Brier score at month 12 using baseline data from ATMOSPHERE as the original data and a 0.01 decrease in time-varying Brier score at month 24 using data up until month 12, suggesting a worse discrimination at the second time point compared to the original data, and while the time-

varying Brier score at the first time point may suggest the models perform equally with the original data and the ATMOSPHERE data, this is contraindicated by time-varying AUCs and calibration curves. The calibration curves at 12 and 24 months using the ATMOSPHERE data illustrated in Figure 36 and Figure 37 respectively show similar fit between the models. With Joint Model 2 appearing to have the best fit at 12 months, and minor differences at between the models at 24 months. However, they show a worse fit when compared to the original data. With the models showing signs of underprediction, which are exacerbated for the higher risk patients.

Figure 34 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs and Brier Scores for Joint Models using Longitudinal Data at Month 0 and Predicting Survival Probability at 12 Months using Data from the ATMOSPHERE Trial

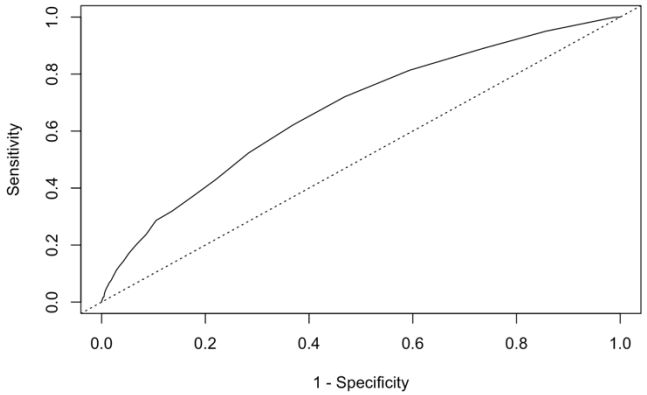
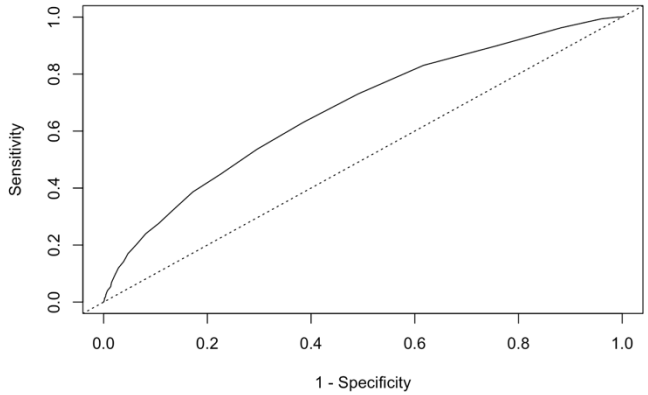
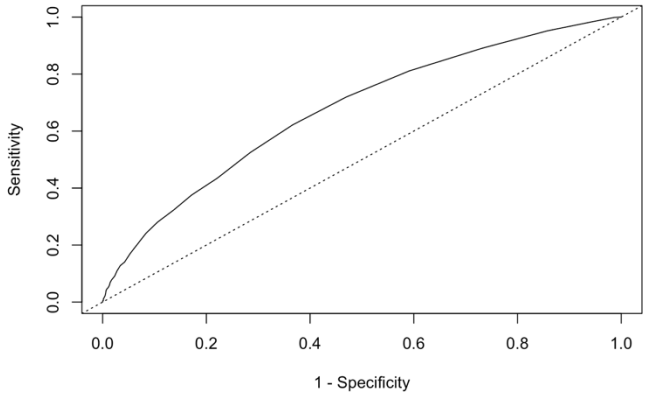
A.) Joint Model 1 – Value: External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial	B.) Joint Model 2 - Value and Slope: External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial	C.) Joint Model 3 – Area: External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial
		
Time-Varying AUC: 0.67 Time-Varying Brier Score: 0.10	Time-Varying AUC: 0.67 Time-Varying Brier Score: 0.10	Time-Varying AUC: 0.67 Time-Varying Brier Score: 0.10

Figure 35 External Validation Time-Varying ROC Curves and Corresponding Time-Varying AUCs for Joint Models using Longitudinal Data up until Month 12 and Predicting Survival Probability at 24 Months

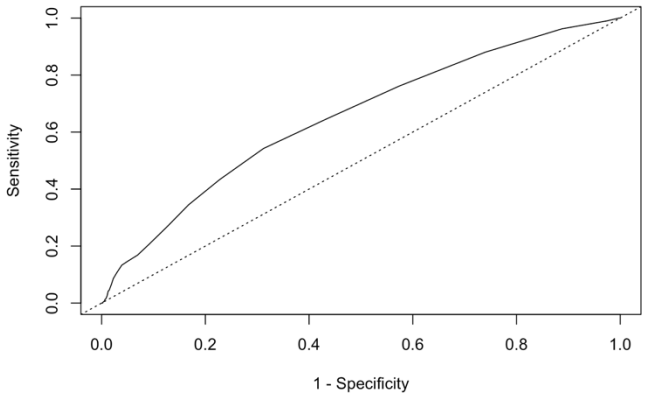
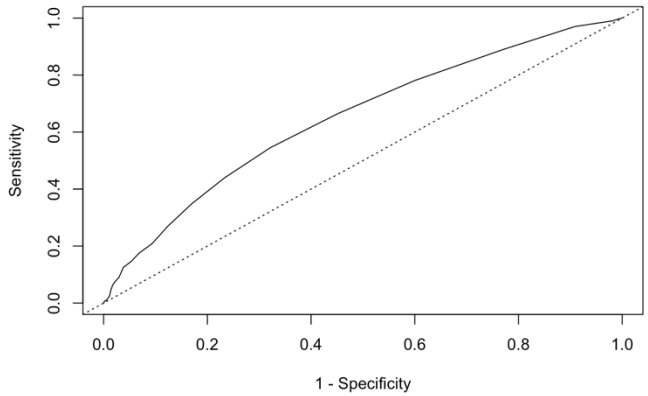
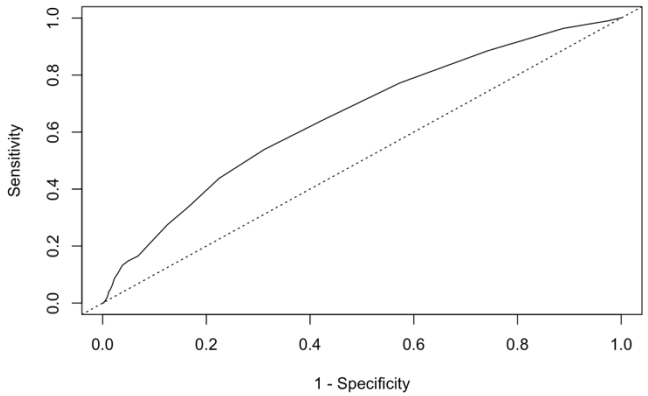
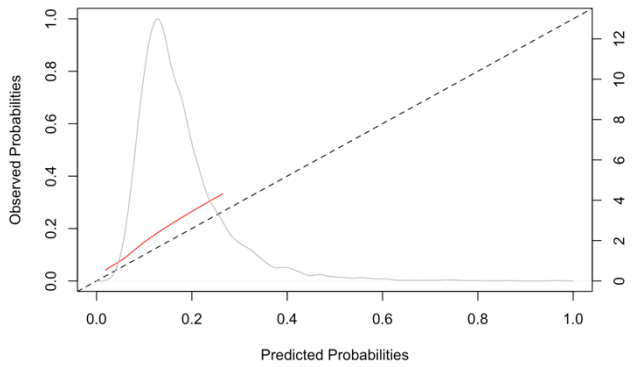
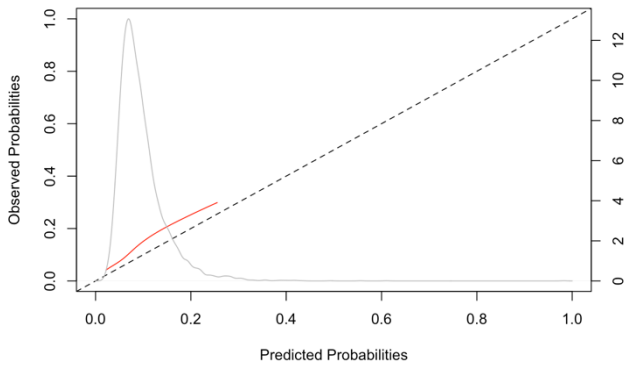
A.) Joint Model 1 – Value: External Validation Time-Varying ROC, AUC, and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial	B.) Joint Model 2 - Value and Slope: External Validation Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial	C.) Joint Model 3 – Area: External Validation Time-Varying ROC, AUC and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial
		
Time-Varying AUC: 0.65 Time-Varying Brier Score: 0.09	Time-Varying AUC: 0.65 Time-Varying Brier Score: 0.09	Time-Varying AUC: 0.65 Time-Varying Brier Score: 0.09

Figure 36 External Validation Calibration Curves for Joint Models at Month 12 using Longitudinal Data At Month 0

A.) Joint Model 1 – Value: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



B.) Joint Model 2 – Value and Slope: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



C.) Joint Model 3 – Area: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial

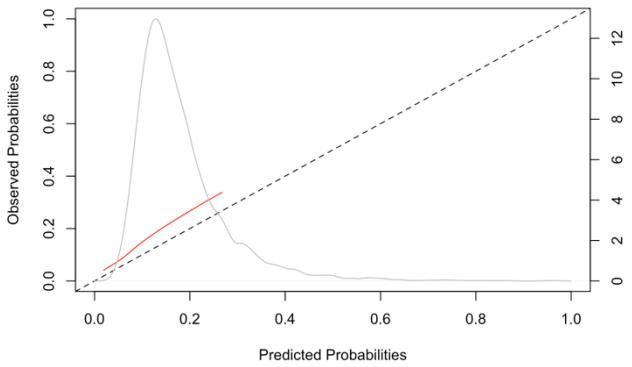
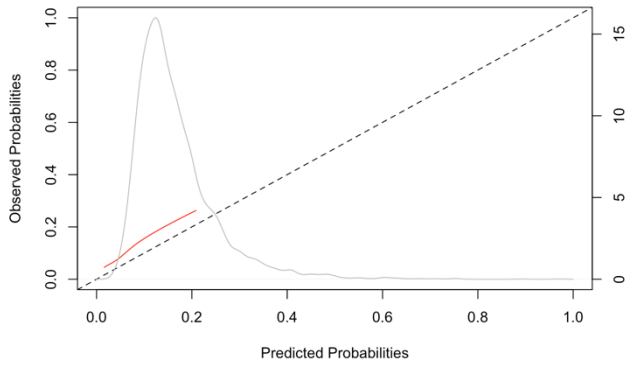
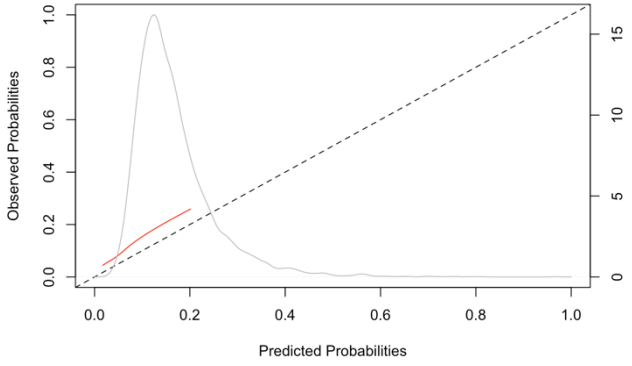


Figure 37 External Validation Calibration Curves for Joint Models at Month 24 using Longitudinal Data up until Month 12

A.) Joint Model 1 – Value: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



B.) Joint Model 2 – Value and Slope: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



C.) Joint Model 3 – Area: External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial

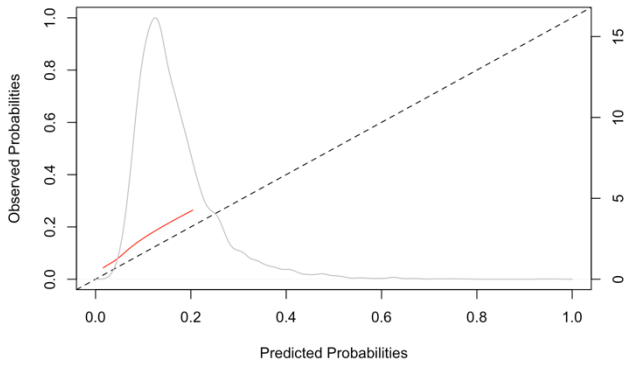
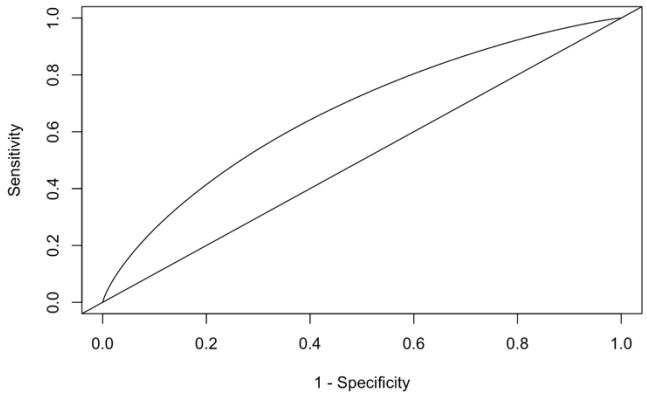


Table 35 Hazard Ratio (95% CI) and P-Values from the Cox PH Models

	Cox PH Last Measurement		Extended Cox PH	
Variable	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**
Treatment - LCZ	0.81 (0.75-0.89)	<0.001	0.82 (0.72-0.92)	0.001
Age	1 (0.99-1)	0.516	1 (0.99-1)	0.468
Sex – Male	1.29 (1.15-1.45)	<0.001	1.35 (1.15-1.59)	<0.001
Region - Latin America	1.04 (0.9-1.2)	0.621	1.28 (1.05-1.55)	0.013
Region - North American	0.93 (0.78-1.11)	0.437	0.75 (0.57-0.97)	0.032
Region - Pacific Asia/Pacific and Other	1.15 (0.99-1.33)	0.068	1.15 (0.94-1.41)	0.166
Region - Western Europe	0.88 (0.78-1)	0.056	0.84 (0.7-1)	0.05
BMI	1.01 (1-1.02)	0.004	0.99 (0.98-1.01)	0.462
Ejection Fraction %	0.98 (0.97-0.99)	<0.001	0.98 (0.97-0.99)	<0.001
NYHA Class II	1.24 (0.97-1.59)	0.09	1.06 (0.77-1.46)	0.722
NYHA Class III	1.58 (1.22-2.05)	0.001	1.6 (1.14-2.25)	0.006
NYHA Class IV	1.92 (1.16-3.15)	0.011	1.13 (0.5-2.56)	0.762
Diabetes - Yes	1.33 (1.21-1.46)	<0.001	1.16 (1.02-1.32)	0.023
SBP	1 (0.99-1)	0.201	1 (0.99-1)	0.348
Heart Rate	1.01 (1-1.01)	0.002	1 (1-1.01)	0.204
History of Atrial fibrillation – Yes	1.11 (1-1.22)	0.04	1.02 (0.89-1.17)	0.77
Ischemic Heart Failure - Yes	0.99 (0.88-1.12)	0.9	1.09 (0.92-1.29)	0.339
Prior Hospitalisation for Heart Failure - Yes	1.38 (1.26-1.52)	<0.001	1.04 (0.92-1.18)	0.53
Prior History of Myocardial Infarction - Yes	1.2 (1.07-1.36)	0.003	1.17 (0.99-1.37)	0.067
Prior History of Stroke - Yes	1.11 (0.96-1.28)	0.173	1.1 (0.9-1.34)	0.362
Log NT-ProBNP	1.5 (1.43-1.58)	<0.001	1.45 (1.36-1.55)	<0.001
Log eGFR	0.59 (0.52-0.67)	<0.001	0.5 (0.42-0.61)	<0.001

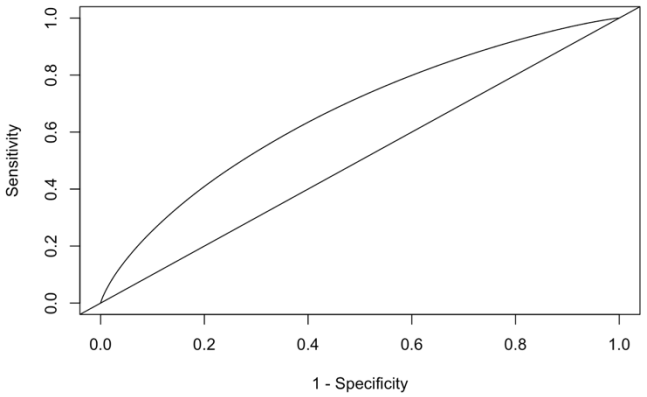
Figure 38 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 12

A.) Time-Varying ROC and AUC at 12 Months for the Cox PH Last Measurement Model



Time-Varying AUC at 12 Months: 0.67

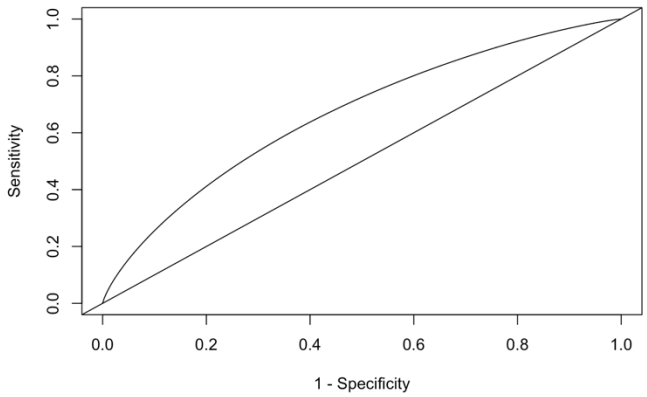
B.) Time-Varying ROC and AUC at 12 Months for the Extended Cox PH Model



Time-Varying AUC at 12 Months: 0.66

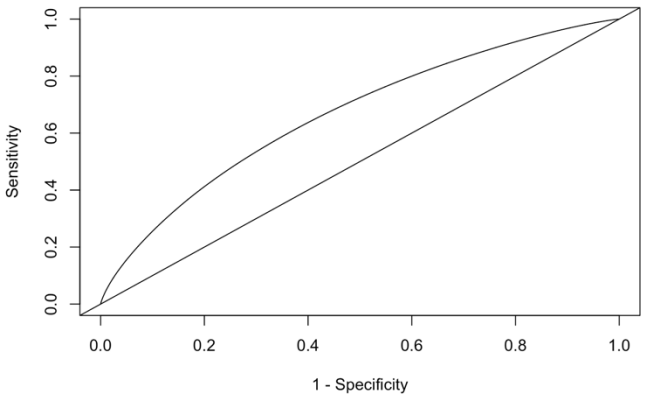
Figure 39 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 24

A.) Time-Varying ROC and AUC at 24 Months for the Cox PH Last Measurement Model



Time-Varying AUC at 24 Months: 0.67

B.) Time-Varying ROC and AUC at 24 Months for the Extended Cox PH Model



Time-Varying AUC at 24 Months: 0.66

5.4.9 Comparative Analysis

The HRs, 95% CIs and corresponding P-Values from the comparative Cox PH models are shown in Table 35. There are some differences between the models, with a minor difference of 0.01 for the treatment effect suggesting the models agree on a hazard of 0.81-0.82 times of the composite event for patients receiving the LCZ treatment, which is also in agreement with the joint models. Age per year increase in baseline age follows the joint models showing no significant effect. On the other hand, there are significant differences in region, with the Extended Cox PH model suggesting a significant association with patients, from the Latin America, North America, and Western Europe, whereas the Last Measurement Cox PH model suggests no significant differences in any region with Western Europe missing the significance cut off by 0.006. Both models differ from the joint models which suggest a significant difference from the Western Europe region, but the hazard ratios from the last measurement model for the region variables match more closely to those from the joint models. Again, there are differences between the two Cox PH models when it comes to the NYHA class variables, with differences in patients with NYHA class II and NYHA class IV of 0.18 and 0.79 respectively. There are also differences in significance of NYHA class III and NYHA class IV with both classes being considered statistically significant in the Last Measurement Cox PH model, and not in the Extended Cox PH Model. As with region the HRs from the Last Measurement Cox PH model for the NYHA Class variables more closely match the joint models than the Extended Cox PH model. Along with the previously mentioned variables, diabetes, having a history of atrial fibrillation, having a history of prior hospitalisation for heart failure, and a prior history of myocardial infarction are all other variables with differences in both HR and statistical significance, With differences ranging from HRs from 0.03 for a prior history of myocardial infarction and 0.38 for prior history of heart failure and as with previous variables the Last Measurement Cox PH model matches

more closely to the joint models with regards to HRs than the Extended Cox PH model. However, there are minimal differences between the HRs of Log NT-ProBNP and Log eGFR of 0.05 and 0.09 respectively with no difference in P-Value. These differences and the fact the Last Measurement Cox PH model is more closely matched with the joint models, suggest there may be a loss of power with the added repeated measurements of log eGFR in the Extended Cox PH Model. The log eGFR HRs in both Cox PH models are higher than the joint models showing a difference between the Cox PH Models and the joint models of upwards of 0.08.

The time-varying ROCs and AUCs shown in Figure 38 and Figure 39 show only a minor difference in AUC between the two Cox PH models at month 12 with the Extended Cox PH performing slightly worse with and AUC of 0.66 compared with 0.67 for the Last Measurement Cox PH model. This suggests only minimal differences in prognostic accuracy between the Cox PH models and between the two time points of the Extended Cox PH model. When compared with the joint models, the joint models perform better at month 12 with an AUC for Joint Model 1 and Joint Model 3 of 0.69 and Joint Model 2 of 0.70 compared to the 0.67 of the Last Measurement Cox PH Model and 0.66 of the Extended Cox PH model. However, the Last Measurement Cox PH Model outperforms all models at time point 12 with a time-varying AUC of 0.67 compared to 0.66 of all other models. This suggests that the joint models may be prognostically more accurate at 12 months than the Cox PH models, but that the Last Measurement Cox PH model may perform slightly better at 24 months.

5.5 Discussion

There is a need to not only explore the use of joint modelling for assessing the association of a biomarker and an endpoint. But to understand how and why joint models might be useful as prognostic models. As well as to explorer whether they can improve on the current methodology, considering the rise in usage and awareness of joint models. For this purpose, three joint models were fit using different parameterisations of the association parameter, using data from two clinical trials. Then compared against both themselves and two traditional models.

All models suggest a statistically significant association between log eGFR and the composite event of cardiovascular death and first hospitalisation for heart failure. This further adds to the evidence base, suggesting a link between renal decline and adverse events in patients with heart failure.

The HR for treatment effect from the survival component from the joint models as well as the Cox PH models were within 0.02 of the original trial, suggesting the models agree with the original findings that patients who were assigned to the LCZ treatment were significantly less likely to experience the composite outcome. However, as with Chapter 4 it was not possible to derive the overall treatment effect from the joint models because of the inclusion of the interaction of treatment and the natural cubic spline of time in the longitudinal component. Both Joint Model 2 and Joint Model 3 suggest nearly a 50% decrease in the hazard of the composite outcome per unit increase of log eGFR / unit increase in the area under the trajectory of log eGFR respectively. Joint Model 1 suggests a lesser decrease of 42% per unit increase in log eGFR. However, these values may not indicate the true treatment effect as

they cannot account for any treatment effect on the longitudinal component as previously discussed.

Average trajectories of log eGFR shown in Figure 28 show little difference in the treatment effect for the average patient over time. However, due to the use of subject specific trajectories using random effects these results may vary between patients.

The coefficients from the longitudinal component, show that all the clinically selected variables except for NYHA class are considered statistically significant with age, the presence of atrial fibrillation and diabetes having a statistically significant negative effect on log eGFR, whereas being male, or having a unit increase in SBP have a statistically significant positive effect on eGFR with only minor differences in coefficients between the joint models.

The marginal and conditional performance metrics indicate that the joint models perform similarly to each other with Joint Model 2 performing best with regards to both marginal and conditional fit; with Joint Model 3 performing the worst out of the models, with respect to marginal fit and Joint Model 1 performing worse with respect of conditional fit.

With respect to model prognostic performance, all models performed similarly except for Joint Model 2 at month 12 using baseline data in which it outperformed the other models with a time-varying AUC of 0.70. However, the model performed the same with respect to the time-varying Brier score, scoring 0.08 equal to the other models at this time point. Joint Model 1 and Joint Model 3 miss the cut-off for acceptable performance by 0.01 at month 12 using baseline data and by 0.03 at month 24 using longitudinal data up until month 12.

However, the time-varying Brier scores at both time points and calibration curves (more so at the first time point) suggest that the models may perform better overall than the time-varying AUCs suggest. Joint Model 2, with a time-varying AUC of 0.70 at month 12 using baseline data enters the lower end of the acceptable category for prognosis.

When validated with the ATMOSPHERE dataset the joint models, perform worse at both time points, with a decrease in time-varying AUC of 0.02 at both time points (month 12 using baseline data and month 24 using longitudinal data up until month 12), this is except for Joint Model 2 which saw a decrease in time-varying AUC of 0.03 at the first time point. However, the time-varying Brier score is the same at the first time point for all joint models (0.10) when using the ATMOSPHERE data to that of the original data. With an increase in time-varying Brier score at the second time point of 0.01 from the original data. The calibration curves also suggest a slightly worse fit with the ATMOSPHERE validation data set at both time points. This validation suggests that overall, the models perform slightly worse with the validation data, which may suggest these models are less generalisable.

The comparison with the Cox PH models suggests that the Cox PH model perform worse at 12 months with a difference in AUC of 0.02 in all models except Joint Model 2 with a difference of 0.03. On the other hand, the Last Measurement Cox PH model performs better than the joint models by 0.01 at month 24, whereas the Extended Cox PH model performs equally with Joint Model 1 at month 12 but worse than Joint Model 2 and Joint Model 3.

The Extended Cox PH exhibits issues with the HRs and P-Values when compared against both the Joint Models and the Last Measurement Cox PH Model. This may be explained by a lack of power due to the sheer number of repeat measurements, possibly suggesting the

Extended Cox PH model is not suited to larger datasets with respect to repeated measurements.

While the joint models performed better than the Cox PH Models, the prognostic performance may not be considered adequate, especially at 24 months with a time-varying AUC of 0.66 on the original data. When considering the worse performance with external data, this may suggest that log eGFR could be a less suited biomarker for a prognostic joint model within heart failure, if the model is to be used at 24 months. The calibration with the ATMOSPHERE data suggests that the model does not perform well with this external dataset. While this may suggest the model is less generalisable, this may be an issue with the recoding of the treatment arms of the ATMOSPHERE trial and not the model itself, therefore this should be taken into consideration when interpreting these results. This analysis may suggest that the extended Cox PH models may not perform as well with larger time-varying covariates as suggested by the differences in hazard ratios and significance levels of the covariates in the survival component.

This analysis has some limitations. For example, the use of sample data breaks the intention to treat design of the original trial which has the potential to introduce bias. The use of randomised control trials while providing an accurate source of data, it limits the population to the original trials including the limitation of patients to those with HFrEF. This may limit the generalisability of the models; however, being a limitation of the data itself, this is out of the scope of the current research. The JMbayes2 package is still in development and has limited accuracy measures, making it difficult to compare the joint models against other models. There is also the issue that because of the way the time-varying ROCs and time-varying AUCs are calculated, they may not be comparable to the time-varying ROCs and

AUCs of the Cox PH Models, meaning that they may be comparing apples to oranges. Finally, the calibration curves lack confidence intervals which may aid in interpretation of the curves.

Chapter 6 Multivariate Joint Modelling of NT-ProBNP and eGFR, and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure

6.1 Foreword

This chapter builds on both Chapter 4 and Chapter 5. As a result, this chapter is styled similarly and may repeat previous statements. This repetition has been included but rephrased to ensure this chapter can be read independently.

6.2 Introduction

The well documented interrelationships between patient characteristics and outcomes within the clinical syndrome of heart failure have been mentioned previously. For context, these characteristics can include biomarkers and physical characteristics. Along with this relationship, the rise in awareness as well as the use of joint modelling within the area of heart failure makes it informative to understand how joint models can be applied for prognosis in heart failure. As previously stated, joint models allow for the repeated measurements of covariates whilst accounting for correlation and measurement error and allowing dynamic and subject specific predictions [51], [52], [146]. The traditional joint models can be extended further to allow for multiple longitudinal responses. This may allow for more accurate predictions over traditional joint models. As modelling multiple longitudinal responses together through shared random effects, adds the potential to further account for individual variability and therefore improve predictions [52], [69], [146]. RCTs often collect multiple biomarkers and time-to-event information making them an excellent candidate for prognostic models using multivariate joint modelling. Except for the prognostic models developed in Chapter 4 and Chapter 5, there are no known prognostic models using

joint modelling or multivariate joint modelling within heart failure. This presents a novel opportunity to apply multivariate joint models to heart failure data and evaluate their value to the field. Given this opportunity, and the potential to produce better predictions, the aim of this chapter is to demonstrate the application of multivariate joint models with the purpose of prognostication within heart failure, and critically appraise them. To this end, I will use data from RCTs to compare the performance of multivariate joint models to the current recommended approaches.

The specific aim of this chapter is therefore to evaluate the value a multivariate joint model can add to prognostic models compared to the current recommended approaches.

6.3 Methods

6.3.1 Data Source

As mentioned in Chapter 3, data from the PARADIGM RCT was used. This was a double-blind randomised control trial which assigned 8,399 patients to either LCZ696 at a dose of 200mg twice daily or enalapril at a dose of 10mg twice daily.

6.3.2 Statistical Analysis

To address the previously mentioned aims, data needed to be analysed, processed, and cleaned prior to the fitting of the multivariate joint models. No known sample size calculation is available for multivariate joint models, so was not possible for this analysis. To fit the multivariate joint models the R Package *JMbayes2* was used. This package is still in development and is based on two packages by the same author, *JM* and *JMbayes*. *JMbayes2* uses a fully parametric approach to fit joint models including multivariate joint models using MCMC sampling. While the package is still in development, it allows for more functional

forms, and calibration curves which were not present in the previous packages, with the latter being an important tool for measuring the performance of prognostic models [39], [40].

The multivariate extension of joint models is comprised of two components, the time-to-event and the longitudinal component, the latter encompassing two (or more) outcomes. These components are fit individually (the longitudinal outcomes fitted independently of each other), and then are refitted into a joint model using the *JM* function from the package *JMbayes2*.

6.3.3 Data Cleaning and Processing

Data were analysed and cleaned prior to analysis as previously stated. Including the analysis of variables and their distributions to examine for outliers, duplicate and improbable values, and manipulation of categorical variables ensuring they were human readable and in the correct format. Alongside this, data for the functions to fit the sub models i.e., LMEs and Cox PH models needed to be in a specified format. This meant ensuring the same group of patients were provided to these functions, and merging baseline covariates with the repeated measurements. In addition to this, the various performance metrics required all covariates to be present in the dataset provided to the *newdata* argument of the functions. Finally, both longitudinal outcomes needed to be for the same patients at the same time points.

6.3.4 Covariate Selection

Both components and all three sub-models (the two longitudinal and one time-to-event) allow for covariates. For all three of these models, covariates were selected on a clinical basis for the same reasons as Chapter 4, using prior knowledge and in conjunction with my clinical supervisor. No identified clinically relevant covariate were missing from either dataset.

6.3.5 Missing Data

Multivariate joint models, whilst allowing for missing longitudinal measurements, the *JMbayes2* package requires all covariates to be present and the longitudinal measurements be at the same time point. As a result of these limitations, data were incomplete for 167 patients. As this is less than 2% of available patients, a complete case analysis was preferred, excluding those 167 patients with incomplete data.

6.3.6 Longitudinal Sub Models

JMbayes2 requires the longitudinal sub models to be either an LME fitted by the function *'lme'* from the *'nlme'* package or a GLMM fitted by the function *'mixed_model'* from the *'GLMMadaptive'* package. As both models used continuous outcomes, an LME was required. These LMEs contain a single outcome and allows for fixed and random effects.

6.3.7 Longitudinal Sub Model One – NT-ProBNP

The longitudinal component is broken up into two sub models, each of which is an LME. The first is for the NT-ProBNP biomarker, which will be the response variable. As previously stated, and illustrated in Chapter 1, NT-ProBNP is associated with clinical outcomes within heart failure. It has also been previously used in prognostic models including the models demonstrated in Chapter 1 [36], [42]. Due to the previously mentioned assumptions around normality and the wide range that NT-ProBNP is known to have, it was necessary to transform the measurements of NT-ProBNP; this was achieved by using the natural logarithm.

While all patients who underwent randomisation had at least one measurement of NT-ProBNP, due to the trial design, 75% of patients had only a single measurement of NT-ProBNP at the screening visit. The decision was made to use these measurements from the screening visit and carry them forward to the randomisation visit in the interest of maximising data.

6.3.7.1 Time (Fixed effects)

The response variable of an LME is allowed to be unevenly spaced with respect to time. However, as stated previously in this chapter the '*jm'* function used to fit multivariate joint models requires both longitudinal measurements to be coded at the same point in time. While measurements of NT-ProBNP and eGFR were collected at scheduled visit times, the actual dates of the measurements varied between the two biomarkers; as such, when the data was merged and there were differences in sample dates, one date had to be chosen over the other. While scheduled visit dates could have been used, as Chapter 4 and Chapter 5 both used sample dates, for continuity one sample date was chosen over the other and used.

For the purpose of this analysis, the sample date from the measurements of eGFR were chosen. This sample date was in the form of months (28-day calendar month) since randomisation.

Similar to Chapter 4, some values of log NT-ProBNP, were observed to be non-linear over time and for this reason, time was included using a natural cubic spline, using the *'ns'* function from the *'splines'* package. The boundary knots where manually specified using quantile values to avoid convergence issues, again likely the result of the imbalance of data at the start of the study.

6.3.7.2 Covariates (Remaining Fixed Effects)

Covariates for the LMEs were specified using prior clinical knowledge. These variables included age, atrial fibrillation, BMI, and treatment effect. Due to the use of screening values, treatment effect was included as both the main effect of treatment and the effect of treatment over time. By default, an intercept term was included in the LME, an intercept term is the mean predicted value of log NT-ProBNP while all other variables are held constant at zero.

6.3.7.3 Random Effects

Both random intercepts and random slopes (including the non-linear effect of time described previously) were included as random effects in the LME.

6.3.7.4 Model Formulation

The model formulation for the log NT-ProBNP shown in Equation 17.

Equation 17 Formulation of the LME sub model for log NT-ProBNP

$$\begin{cases} \log NTProBNP_{ij} = \beta_0 + \beta_1 f_1(Specimen\ Time_{ij}) + \beta_2 f_2(Specimen\ Time_{ij}) + \beta_3 f_3(Specimen\ Time_{ij}) \\ + \beta_4 Treatment\ LCZ_{ij} \\ + \beta_5 f_1(Specimen\ Time_{ij})\ Treatment\ LCZ_{ij} + \beta_6 f_2(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} + \beta_7 f_3(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} \\ + \beta_8 Age_{ij} + \beta_9 Atrial\ Fibrillation\ Yes_{ij} + \beta_{10} BMI_{ij} \\ + b_{i0} + b_{i1} f_1 Specimen\ Time_{ij} + b_{i2} f_1 Specimen\ Time_{ij} + b_{i3} f_1 Specimen\ Time_{ij} + \varepsilon_{ij}, \\ b_i \sim \mathrm{N}(0, D), \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \end{cases}$$

Where β_0 is the intercept term, the remaining β_n are the fixed effects including the natural cubic spline as a function f_n of the specimen time. b_{in} are the random effects including both the random intercept b_{i0} and the $b_{in} f_n$ representing the random slopes incorporating the natural cubic spline. Where the random effects are assumed to be normally distributed with a mean of 0 with a variance-covariance matrix D. ε_{ij} represents the error term, which is assumed to have a mean of 0 and a variance of σ^2.

6.3.7.5 Model fitting

The LME was fit in the same way as Chapter 4, with the response variable (log NT-ProBNP), restricted natural cubic spline as specified using the *ns* function, interaction of this spline with treatment and the remaining specified covariates to the fixed parameter of the *lme* function.

The random slopes and intercepts were also specified using the random parameter of the function, and the optimiser set to the same general-purpose optimization method based on Nelder–Mead, quasi-Newton and conjugate-gradient algorithms, *optim* using the *lmeControl* argument [139].

6.3.8 Longitudinal Sub Model Two – eGFR

For the same reasons as Chapter 5, eGFR was chosen for the second sub model, in which eGFR was used as the response variable. These reasons include the well documented relationship with declining renal function in heart failure. As with Chapter 5, the values of eGFR were observed to be skewed and for this reason eGFR was transformed using the natural logarithm. This transformation was necessary to satisfy the normality assumption of the random effects [52].

6.3.8.1 Time (Fixed Effect)

Due to the limitations of the JMbayes2 packages, as previously stated the measurements of each longitudinal process need to be at the same time points. For this reason, sample date was chosen, and time modelled in the form of month from randomisation (28-day calendar month). While this limitation is present, there is still the option of modelling time within the sub model using natural cubic splines. As some of the longitudinal profiles of log eGFR showed signs of non-linearity over time, time within the LME was modelled non-linearly using the previously mentioned ‘ns’ function and as with the previous model, the boundary knots were manually specified using quantile values.

6.3.8.2 Covariates (Remaining Fixed Effects)

As with the previous model, the remaining covariates were selected based on prior clinical knowledge and included treatment effect, age, sex, atrial fibrillation, SBP, NYHA and the presence of diabetes. Only the effect of treatment over time (modelled using the natural cubic spline) was included in the sub model due to randomisation, as is standard practice [69]. As with the earlier model, by default an intercept term was included, which is the predicted mean when all other covariates are zero.

6.3.8.3 Random Effects

Like that of the previous model, the random effects included a random intercept, and random slopes which included the non-linear effect of time as modelled by the restricted cubic spline.

6.3.8.4 Model Formulation

The model formulation for this sub model is noted in Equation 18.

Equation 18 Formulation of the LME sub model for log eGFR

$$\left\{\begin{array}{c} \log eGFR_{ij} = \beta_0 + \beta_1 f_1(Specimen\ Time_{ij}) + \beta_2 f_2(Specimen\ Time_{ij}) + \beta_3 f_3(Specimen\ Time_{ij}) \\ + \beta_4 f_1(Specimen\ Time_{ij})\ Treatment\ LCZ_{ij} + \beta_5 f_2(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} + \beta_6 f_3(Specimen\ Time_{ij}) Treatment\ LCZ_{ij} \\ + \beta_7 Age_{ij} + \beta_8 Sex_{ij} + \beta_9 Atrial\ Fibrillation\ Yes_{ij} + \beta_{10} SBP_{ij} + \beta_{11} Diabetes\ Yes_{ij} \\ \beta_{12} NYHA\ Class\ II_{ij} + \beta_{13} NYHA\ Class\ III_{ij} + \beta_{14} NYHA\ Class\ IV_{ij} \\ + b_{i0} + b_{i1} f_1 Specimen\ Time_{ij} + b_{i2} f_1 Specimen\ Time_{ij} + b_{i3} f_1 Specimen\ Time_{ij} + \varepsilon_{ij}, \\ b_i \sim \mathrm{N}(0, D), \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \end{array}\right\}$$

The intercept term is represented by β_0 and the remaining fixed effects represented by β_n with the natural cubic spline being represented by the function f_n. The random effects represented by b_{in} incorporating the random intercept b_{i0} and random slopes $b_{in}f_n$ (including the natural cubic spline). These random effects are assumed to be normally distributed with a mean 0 and a variance-covariance matrix D. The error term represented by ε_{ij} is assumed to have a mean of 0 with a variance of σ^2.

6.3.8.5 Model Fitting

To fit the second LME, the same process was used as described in Chapter 5. The response variable of log eGFR, previously specified covariates, restricted natural cubic spline (fit using the ns function), and its interaction with treatment being passed to the *fixed* parameter of the *lme* function. Along with the random intercepts and slopes being passed to the *random* parameter, and the setting of the general-purpose optimization method based on Nelder–Mead, quasi-Newton and conjugate-gradient algorithms *optim* using the *lmeContol* parameter to aid convergence.

6.3.9 Survival Sub Model

The survival sub model needs to be either a Cox PH model or a semiparametric survival model fitted by the functions '*coxph*' and '*survreg*' of the `*survival*` package respectively. The survival sub model for this analysis is comprised of a Cox PH model, chosen to avoid misspecification of the baseline hazard function.

6.3.9.1 Covariates

The covariates were based on prior clinical knowledge and were the same as the original trials' survival analysis. These covariates included age, BMI, ejection fraction, heart rate, NYHA classification, region, sex, SBP, treatment, whether the patient had diabetes, whether the patient had a history of atrial fibrillation, hospitalisation for heart failure, myocardial infarction, or stroke.

6.3.9.2 Event (End Point)

The endpoint was a composite outcome of death for cardiovascular causes and first hospitalisation for heart failure. This was also the primary outcome of the original trial.

6.3.9.3 Model Formulation

The formula for the survival sub model is noted in Equation 19.

Equation 19 Formular of the Survival Sub model

$$h_i(t) = h_0(t) \exp (\gamma_1 Treatment\ LCZ_i + \gamma_2 Age_i + \gamma_3 Male_i + \gamma_4 Region\ Latin\ America_i + \gamma_5 Region\ North\ America_i + \gamma_6 Region\ Asia, Pacific\ and\ Other_i + \gamma_7 Region\ Western\ Europe_i + \gamma_8 BMI_i + \gamma_9 Ejection\ Fraction_i + \gamma_{10} NYHA\ Class\ II_i + \gamma_{11} NYHA\ Class\ III_i + \gamma_{12} NYHA\ Class\ IV_i + \gamma_{13} Diabetes\ Yes_i + \gamma_{14} SBP_i + \gamma_{15} Heart\ Rate_i + \gamma_{16} Atrial\ Fibrillation\ Yes_i + \gamma_{17} Ischemic\ Heart\ Failure\ Yes_i + \gamma_{18} Prior\ Hospitalisation\ for\ Heart\ Failure\ Yes_i + \gamma_{19} Prior\ Miocardial\ Infarction\ Yes_i + \gamma_{20} Prior\ Stroke\ Yes_i$$

Where $h_i(t)$ represents the hazard of the composite event at time point t. $h_0(t)$ represents the baseline hazard and the covariates are represented by γ_n.

6.3.9.4 Model Fitting

Like Chapter 4 and Chapter 5 the Cox PH model was fit using time to the composite end point or at the point the patient was censored in months. This was passed along with an indicated of whether the patient experienced the composite event and the remaining covariates using the formula argument of the *coxph* function from the *survival* package.

6.3.10 Joint Models

While joint modelling allows for different alpha parameterisations; because the longitudinal outcomes for both eGFR and NT-ProBNP have been modelled in Chapter 4 and Chapter 5 respectively, for this analysis the functional forms were chosen based on the best performing functional forms from the previous analyses. Performance was judged on the performance metrics for conditional fit, as conditional fit is more likely produce a better prognostic model, especially if the goal was personalised predictions. This meant that only one joint model was fit within this analysis, including both the value and slope parameters for log NT-ProBNP and log eGFR. These parameters have been previously defined but briefly, the value parameter(s) measures the association of the value of each longitudinal outcome and the

hazard of the composite event, and the slope parameter measures the association of the rate of change of each longitudinal outcome.

The joint model was fit by passing the two LMEs and the Cox PH models to the '*jm'* function and specifying the functional forms using the '*functional_forms'* argument of the function. The joint model was fit using a piecewise baseline hazard function, with quadratic B-splines and 10 baseline hazard segments.

The formula for the multivariate joint model is shown in Equation 20.

Equation 20 Formula for the multivariate joint model

$$
\begin{cases}
\begin{gathered}
y_{i1}(t) = m_{i1}(t) + \varepsilon_{i1\,(t)} \\
= \beta_0 + \beta_1 f_1(Specimen\ Time) + \beta_2 f_2(Specimen\ Time) + \beta_3 f_3(Specimen\ Time) \\
+ \beta_4 Treatment\ LCZ \\
+ \beta_5 f_1(Specimen\ Time) Treatment\ LCZ + \beta_6 f_2(Specimen\ Time) Treatment\ LCZ + \beta_7 f_3(Specimen\ Time) Treatment\ LCZ \\
+ \beta_8 Age + \beta_9 Atrial\ Fibrillation\ Yes + \beta_{10} BMI \\
+ b_{i0} + b_{i1} f_1 Specimen\ Time + b_{i2} f_1 Specimen\ Time + b_{i3} f_1 Specimen\ Time + \varepsilon_{i1}, \quad \varepsilon_{i1}(t) \sim N(0, \sigma^2),
\end{gathered} \\
\\
\begin{gathered}
y_{i2}(t) = m_{i2}(t) + \varepsilon_{i2\,(t)} \\
= \beta_0 + \beta_1 f_1(Specimen\ Time) + \beta_2 f_2(Specimen\ Time) + \beta_3 f_3(Specimen\ Time) \\
+ \beta_4 f_1(Specimen\ Time) Treatment\ LCZ + \beta_5 f_2(Specimen\ Time) Treatment\ LCZ + \beta_6 f_3(Specimen\ Time) Treatment\ LCZ \\
+ \beta_7 Age + \beta_8 Sex + \beta_9 Atrial\ Fibrillation\ Yes + \beta_{10} BMI \\
+ \beta_{11} NYHA\ Class\ II + \beta_{12} NYHA\ Class\ III + \beta_{13} NYHA\ Class\ IV \\
+ b_{i0} + b_{i1} f_1 Specimen\ Time + b_{i2} f_1 Specimen\ Time + b_{i3} f_1 Specimen\ Time + \varepsilon_i, \quad \varepsilon_{i2}(t) \sim N(0, \sigma^2),
\end{gathered} \\
\\
h_i(t) = h_0(t)\, exp \left\{ \begin{gathered}
\gamma_1 Treatment\ LCZ_i + \gamma_2 Age_i + \gamma_3 Male_i + \gamma_4 Region\ Latin\ America_i + \gamma_5 Region\ North\ America_i \\
+ \gamma_7 Region\ Western\ Europe_i + \gamma_8 BMI_i + \gamma_9 Log\ NT\text{-}ProBNP_i + \gamma_{10} Ejection\ Fraction_i \\
+ \gamma_{11} NYHA\ Class\ II_i + \gamma_{12} NYHA\ Class\ III_i + \gamma_{13} NYHA\ Class\ IV_i \\
+ \gamma_{14} Diabetes\ Yes_i + \gamma_{15} SBP_i + \gamma_{16} Heart\ Rate_i + \gamma_{17} Atrial\ Fibrillation\ Yes_i + \gamma_{18} Ischemic\ Heart\ Failure\ Yes_i \\
+ \gamma_{19} Prior\ Hospitalisation\ for\ Heart\ Failure\ Yes_i + \gamma_{20} Prior\ Miocardial\ Infarction\ Yes_i + \gamma_{21} Prior\ Stroke\ Yes_i \\
+ \alpha m_{i1}(t) + \alpha_2 m'_{i1}(t) + \alpha m_{i2}(t) + \alpha_2 m'_{i1}(t)
\end{gathered} \right\}, \\
\\
log\ h_0(t) = \gamma h_0, 0 + \sum_{q=0}^{Q} \gamma h_{0,q}\, B_q(t, v),
\end{cases}
$$

Where $y_{i1}(t)$ and $y_{i2}(t)$ are the longitudinal outcomes at time point t, composed of the corresponding true and unobserved values of the makers, NT-ProBNP and eGFR at time point t ($m_{i1}(t)$, $m_{i2}(t)$) and their corresponding error terms $\varepsilon_{i1\,(t)}$ and $\varepsilon_{i2\,(t)}$,each longitudinal outcome having an intercept term β_0 and corresponding coefficients for the fixed effects β_n. The natural cubic splines represented as by the functions f_n. The error terms for both longitudinal outcomes are assumed to be normally distributed with a mean 0 and variance σ^2. The random effects in the longitudinal outcomes are represented by b_{i0} for the random intercept and $b_{in} f_n$ for the random slopes incorporating the natural cubic spline of time as the functions f_n. The hazard of the composite event being represented by $h_i(t)$, comprising of the baseline hazard function $h_0(t)$ and coefficients for the covariates γ_n. The baseline hazard function being comprised of a piecewise hazard function, with quadratic B-splines, with 10 baseline hazard sections. In which $B_q(t, v)$ corresponds to the q-th basis function of the B-spline with knots v_1-v_{10} and γh_0 being a vector of spline coefficients.

6.3.10.1 Model Fitting

The multivariate joint model was fit using the *jm* function from the jmBayes2 package. To fit the model, both longitudinal outcomes fit as the earlier described LMEs were passed to the *Mixed_objects* parameter. The Cox PH model was passed to the *Surv_object* argument and the parameterisation of each of the longitudinal outcomes in the survival outcome was passed using the *functional_forms* argument.

6.3.11 Model Performance: Prognostic Accuracy, Fit and Calibration

Taking advantage of the functions available in the joint modelling package JMbayes2, this analysis uses time-varying AUCs and ROCs using the functions *tvAUC* and *tvROC* respectively along with the time-varying Brier score and time-varying calibration curves produced by the functions *tvBrier* and *calibration_plot* respectively. Although previously discussed, briefly, these metrics extend on the traditional metrics providing time-dependent metrics by using longitudinal data up until a given time point and then predicting the survival outcome at a specified time-point [69], [142]. These metrics are interpreted similar to the original metrics with the exception that they are assessed at specified time points. These time points were defined as month 12 using longitudinal data at baseline and month 24 using longitudinal data up until month 12.

The marginal (population level) and conditional (subject specific) fit were compared using the LPML and DIC, the LPML being similar to the log likelihood of a frequentist model, a larger value indicating a better fit, the DIC being similar to any other information criterion, a smaller value indicating a better fit.

6.3.12 External Validation

As previously stated in Chapter 3, external validation was conducted using data from the ATMOSPHERE RCT. This trial randomised 7016 patients to receive either enalapril at dose of 5 or 10mg twice daily, aliskiren at a dose of 150mg once a day or a combination of both enalapril (at 5 or 10mg twice daily) and aliskiren (at a dose of 150mg once a day). The study was conducted in a double-blind double dummy fashion.

Data from the ATMOSPHERE RCT was required to be formatted in the same way as the data from the original trial, to achieve this, both biomarkers NT-ProBNP and eGFR were log transformed. Time was modelled in the same fashion, specimen time in months from randomisation (28 calendar month). The treatment arms were recoded to both enalapril and LCZ, the former containing the enalapril arm and the latter the remaining arms.

Evaluation of the models using the data from the ATMOSPHERE RCT was achieved by comparison of the time-varying metrics (AUC, ROC, Brier score, and calibration curves). These metrics were calculated using the fitted multivariate joint model specifying the ATMOSPHERE data into the *newdata* argument of the time-varying metric functions. As with the original data, these scores were calculated at 12 months using baseline data, and 24 months using data up until month 12. These metrics were then compared against those obtained using the original data.

6.3.13 Comparative Analysis

As a comparative analysis, the joint models were compared against the current standard for prognostic models (Cox PH models). To achieve this, two Cox PH models were fit, one using the LOCF method and one using an extended Cox PH model. These two models were then

compared against the joint model using time-varying AUCs and ROCs, obtained from the *risksetROC* function from the *risksetROC* package (which uses the formula defined by Heagerty et al., [142], [143]) along with a comparison of the hazard ratios and Cis.

Descriptive statistics for the distributions of continuous variables are described as medians (Q1, Q3), and described as percentages for categorical variables. The time dependent association parameters from the joint model are expressed as an HR with corresponding 95% CI, indicating the hazard of the composite event per unit increase of the corresponding longitudinal marker (Log NT-ProBNP / Log eGFR) at any timepoint. Likewise, the association slope parameter is expressed as a HR and 95% CI, indicating the hazard of the composite event per unit increase in slope of the corresponding longitudinal marker (log NT-ProBNP / log eGFR) at any time point.

P-Values below the threshold of 0.05 are considered statistically significant. All statistical analysis was conducted using R Version 4.0 [144] and JMbayes2 package version 0.1-81 [114].

6.4 Results

6.4.1 Baseline Characteristics

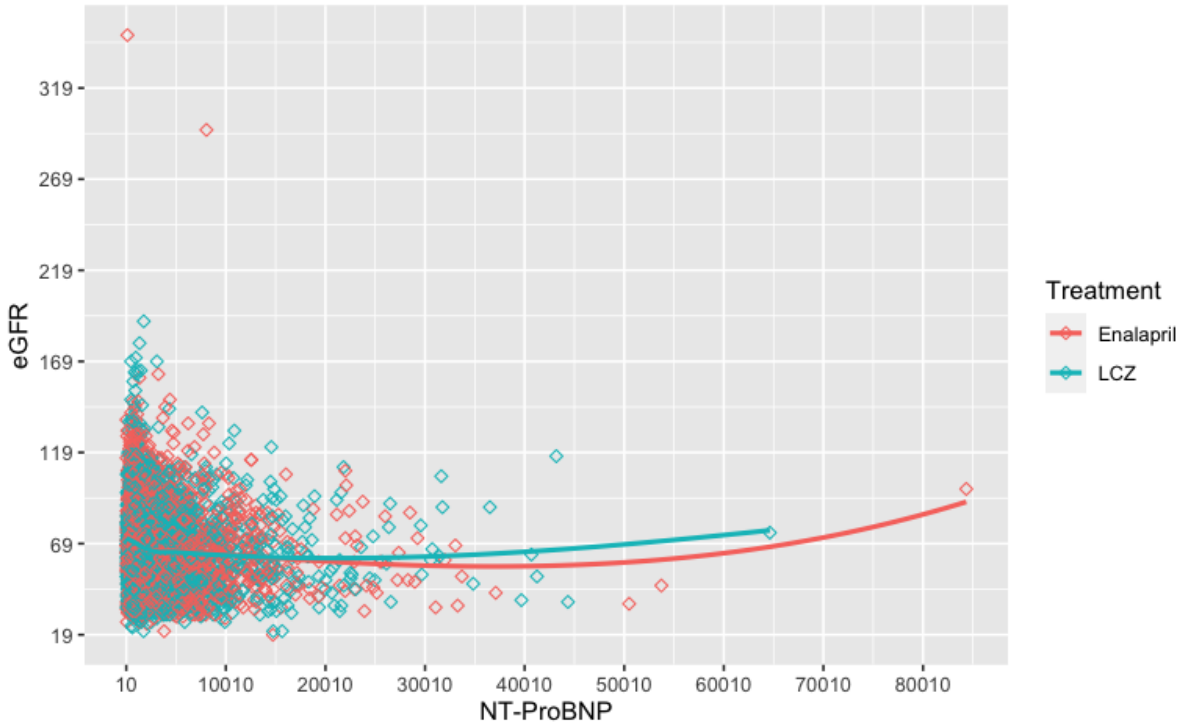
Baseline characteristics of the included 8232 patients are shown in Table 36. With 22% of patients being female and having a median age of 64 years. As reported in the original trial, the differences between the arms for the distributions of baseline characteristics are minimal. Figure 40 illustrates the baseline measurements of both biomarkers by treatment. To assist in showing the co-variation in the two biomarkers, loess curves for each treatment group have been added to the scatterplot. The visual inspection and the Pearson correlation coefficient of -0.09 indicate minimal correlation between the two biomarkers at baseline.

Table 36 Baseline Characteristics of Included Patients

Characteristic	LCZ (N=4096)	Enalapril (N=4136)	Overall (N=8232)
Age, years			
Median [Q1, Q3]	64.0 [57.0, 72.0]	64.0 [57.0, 72.0]	64.0 [57.0, 72.0]
Sex			
Female	862 (21.0%)	937 (22.7%)	1799 (21.9%)
Male	3234 (79.0%)	3199 (77.3%)	6433 (78.1%)
Region			
Central Europe	1371 (33.5%)	1409 (34.1%)	2780 (33.8%)
Latin America	691 (16.9%)	697 (16.9%)	1388 (16.9%)
North American	304 (7.4%)	288 (7.0%)	592 (7.2%)
Pacific Asia/Pacific and Other	735 (17.9%)	732 (17.7%)	1467 (17.8%)
Western Europe	995 (24.3%)	1010 (24.4%)	2005 (24.4%)
BMI kg/m² Ejection Fraction %			
Median [Q1, Q3]	27.5 [24.4, 31.2]	27.5 [24.5, 31.2]	27.5 [24.4, 31.2]
Ejection Fraction %			
Median [Q1, Q3]	30.0 [25.0, 34.0]	30.0 [25.0, 34.3]	30.0 [25.0, 34.0]
NYHA Class			
Class II	178 (4.3%)	206 (5.0%)	384 (4.7%)
Class II	2936 (71.7%)	2869 (69.4%)	5805 (70.5%)
Class III	952 (23.2%)	1034 (25.0%)	1986 (24.1%)
Class IV	30 (0.7%)	27 (0.7%)	57 (0.7%)
Diabetes			
No	2672 (65.2%)	2699 (65.3%)	5371 (65.2%)
Yes	1424 (34.8%)	1437 (34.7%)	2861 (34.8%)
SBP mmHg			
Median [Q1, Q3]	120 [110, 130]	120 [110, 130]	120 [110, 130]
Heart Rate beats per minute			
Median [Q1, Q3]	71.0 [64.0, 80.0]	72.0 [64.0, 80.0]	71.0 [64.0, 80.0]
Prior History of Atrial Fibrillation			
No	2611 (63.7%)	2590 (62.6%)	5201 (63.2%)
Yes	1485 (36.3%)	1546 (37.4%)	3031 (36.8%)
Ischemic Heart Failure			
No	1646 (40.2%)	1652 (39.9%)	3298 (40.1%)
Yes	2450 (59.8%)	2484 (60.1%)	4934 (59.9%)

Characteristic	LCZ (N=4096)	Enalapril (N=4136)	Overall (N=8232)
Prior History of Hospitalisation for Heart Failure			
No	1537 (37.5%)	1520 (36.8%)	3057 (37.1%)
Yes	2559 (62.5%)	2616 (63.2%)	5175 (62.9%)
Prior History of Myocardial Infarction			
No	2320 (56.6%)	2350 (56.8%)	4670 (56.7%)
Yes	1776 (43.4%)	1786 (43.2%)	3562 (43.3%)
Prior History of Stroke			
No	3746 (91.5%)	3770 (91.2%)	7516 (91.3%)
Yes	350 (8.5%)	366 (8.8%)	716 (8.7%)
NT-ProBNP			
Median [Q1, Q3]	1420 [776, 2880]	1450 [790, 2930]	1440 [782, 2910]
eGFR			
Median [Q1, Q3]	66.0 [54.0, 79.0]	66.0 [53.0, 79.0]	66.0 [54.0, 79.0]

Figure 40 Baseline NT-ProBNP against eGFR by Treatment with Loess Curves and Pearson Correlation Coefficient



Pearson Correlation Coefficent: -0.09

6.4.2 Number of Measurements of NT-ProBNP and eGFR

Table 37 shows the number of repeat measurements of NT-ProBNP and eGFR. With the majority of patients having a single measurement (75%). With a minimum number of one measurement and a maximum of 4 measurements.

Table 37 Number of Repeat Measurements of NT-ProBNP and eGFR for Included Patients

1	2	3	4
6174 (75%)	215 (3%)	857 (10%)	977 (12%)

6.4.3 Longitudinal Profiles of NT-ProBNP and eGFR

The longitudinal profiles of log NT-ProBNP and log eGFR for 42 randomly sampled patients (stratified by number of measurements) are illustrated in Figure 41 and Figure 42.

6.4.3.1 NT-ProBNP

Some of the patients' longitudinal profiles of NT-ProBNP shown in Figure 41, such as 0333_00014, show signs on non-linearity of (log) NT-ProBNP over time. As this is on a log scale, the non-linearity would be amplified on the original scale. The figure also shows the wide range of NT-ProBNP taking into consideration the log scale.

6.4.3.2 eGFR

Like those of (log) NT-ProBNP, the longitudinal profiles of (log) eGFR over time shown in Figure 42, illustrate that some patients, such as 1122_00023, exhibit signs of non-linearity and this would also be amplified on the original scale.

Figure 41 Longitudinal Profile of log NT-ProBNP for 42 Randomly Sampled Patients

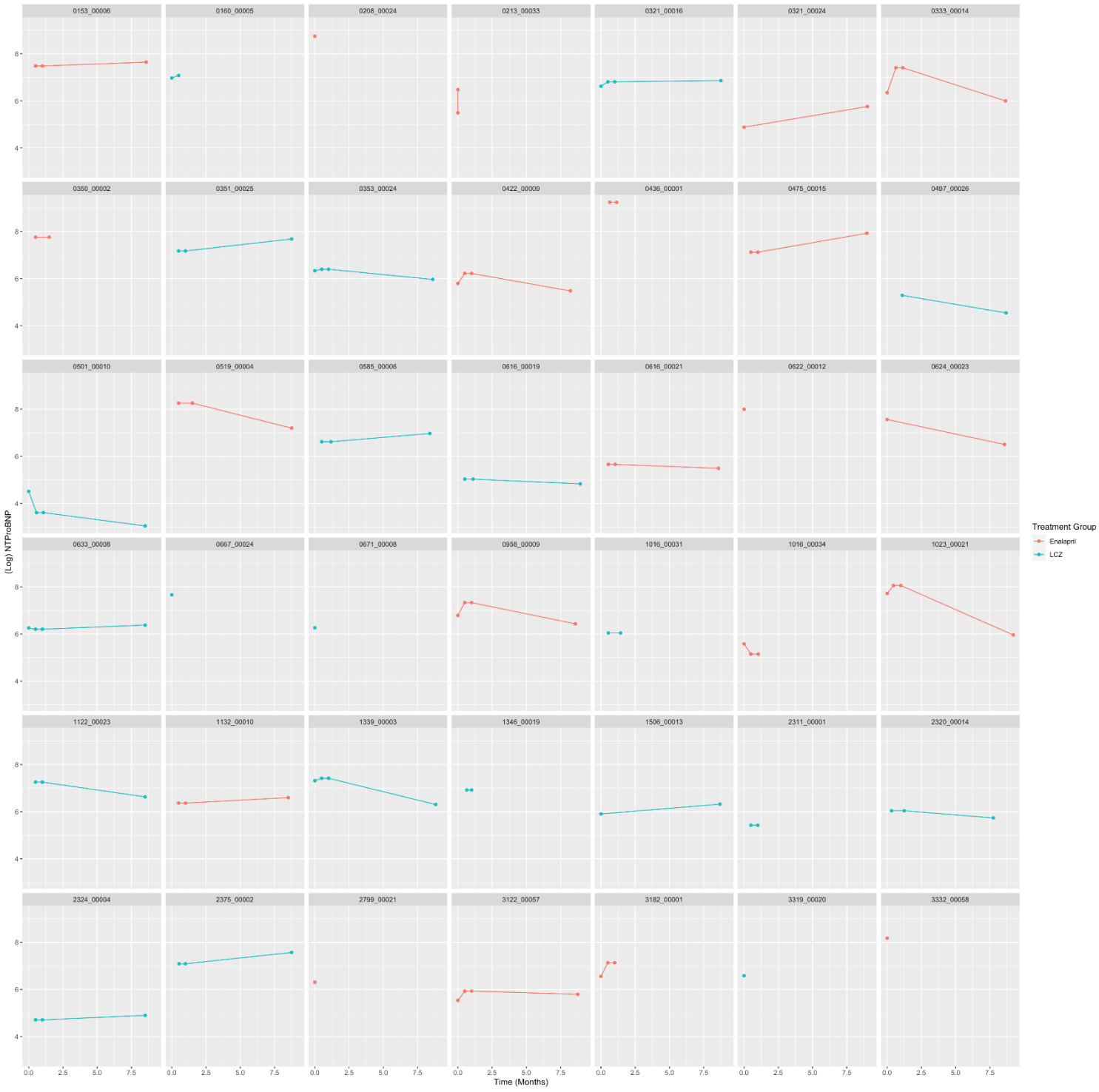
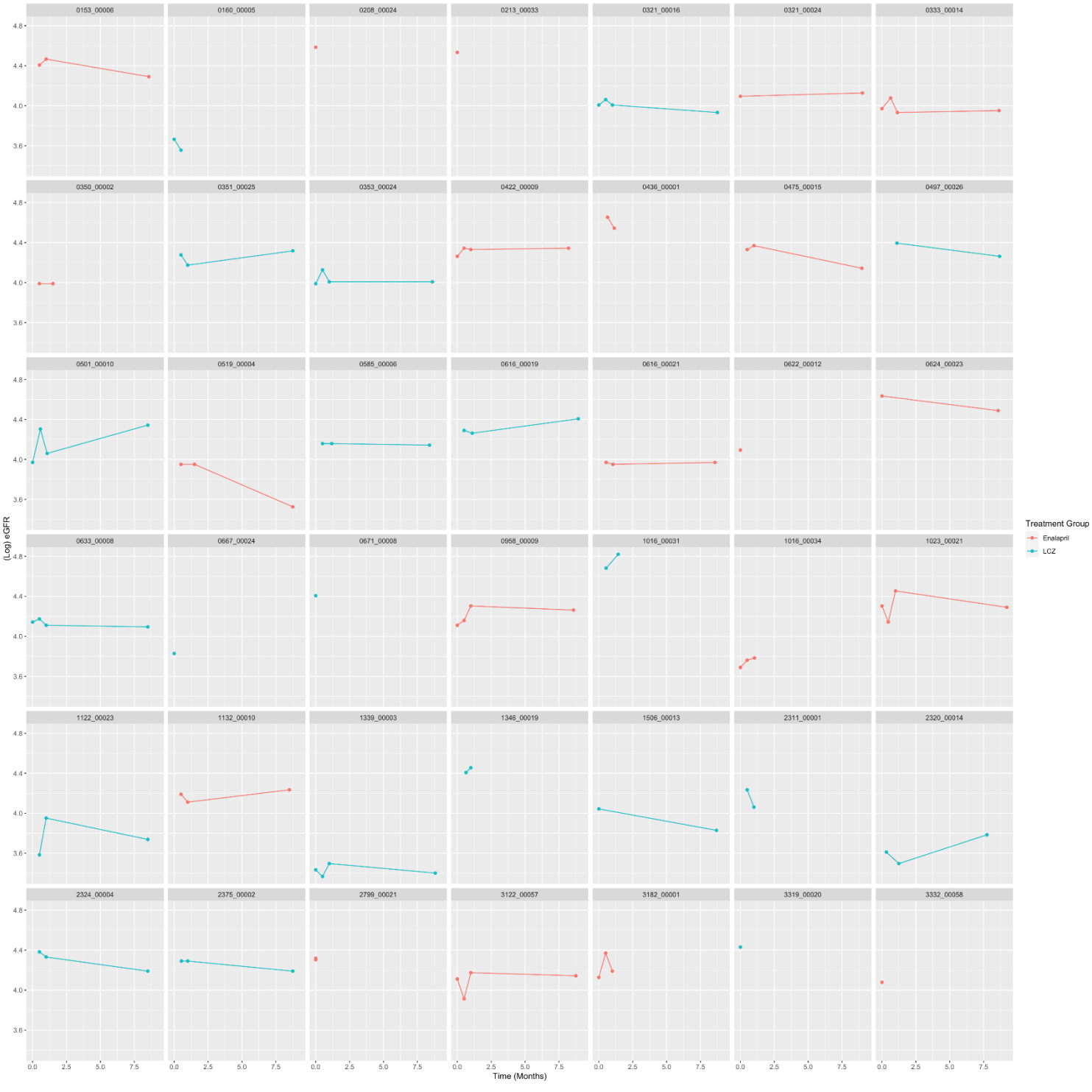


Figure 42 Longitudinal Profile of log eGFR for 42 Randomly Sampled Patients



6.4.4 Longitudinal Outcomes from Joint Model

Parameter estimates for all meaningfully interpretable variables from the longitudinal outcomes from the joint model are shown in Table 38 and Table 39. Variables which included a natural cubic spline have been omitted as they are not directly interpretable instead, they are visually represented in Figure 43 and Figure 44.

6.4.4.1 NT-ProBNP

Illustrated in Table 38 are the meaningfully interpretable parameters, all the parameter estimates for the log NT-ProBNP longitudinal outcome are considered statistically significant. Both treatment with LCZ and BMI have a negative effect on log NT-ProBNP while age and the presence of atrial fibrillation have a positive effect on log NT-ProBNP. Whilst significant, all the parameter estimates show only small effects on log NT-ProBNP with age having the smallest effect of 0.003 and treatment having the largest effect of -0.262. However it should be noted that time is represented in months and NT-ProBNP is being represented on a log scale, both of which could influence the effect size. While parameters which included natural splines are not able to be meaningfully interpreted as they represent parts of a function. They can be graphically illustrated such as in Figure 43 which shows the average predicted trajectories of log NT-ProBNP for each treatment group for the average patient (using median and mode characteristics) from the multivariate joint model, along with the observed trajectories of NT-ProBNP. This figure suggests that the LCZ treatment group had a lower log NT-ProBNP on average, with a change in log NT-ProBNP at month 3 for both treatments, but with a lack of certainty with overlapping 95% CIs after month eight.

6.4.4.2 eGFR

Meaningfully interpretable parameters are illustrated in Table 39, all parameter estimates for the log eGFR longitudinal outcome apart from NYHA class are considered statistically significant, with age, the presence of atrial fibrillation and the presence of diabetes having a negative effect on log eGFR. Male sex and SBP have a positive effect on eGFR. As with the log NT-ProBNP longitudinal outcome, the effect size for all variables is quite small, with the male sex having the highest effect size of 0.041 and SBP having the smallest 0.001, but once again like the NT-ProBNP outcome, eGFR is also represented on a log scale and time is represented in months which could have also affected the parameter effect sizes. As with the NT-ProBNP longitudinal outcome, the parameters with natural cubic splines cannot be meaningfully interpreted. Figure 44 however, shows the trajectories of log eGFR for the average patient (predicted using the median and mode characteristics) in each treatment group from the multivariate joint model, along with the observed trajectories, scaled for readability. This figure suggests that on average there was a decrease in eGFR until month five and then a slight increase in eGFR. However, the trajectories along with the overlapping 95% CIs suggest a lack of difference in treatment effect on these average trajectories. While the trajectories of log NT-ProBNP appear similar to those from Figure 15 of Chapter 4, the trajectories of log eGFR appear to differ from Figure 28 of Chapter 5, possibly as a result of the removal of values of log eGFR to satisfy the need to have values of both biomarkers at the same time.

Table 38 Parameter Estimates (95% CI) from The Longitudinal Outcome for NT-ProBNP from The Joint Model for Meaningfully Interpretable Variables

Variable	Estimate (95% CI)	P-Value
Treatment – LCZ	-0.262 (-0.397 - -0.135)	< 0.001
Age	0.003 (0.001 - 0.005)	0.016
Atrial Fibrillation – Yes	0.258 (0.202 – 0.312)	< 0.001
BMI	-0.045 (-0.050 - -0.040)	< 0.001

The parameter estimates for variables which include the natural cubic spline have been omitted as they are not directly interpretable but instead are illustrated by Figure 43.

Table 39 Parameter Estimates (95% CI) from The Longitudinal Outcome for eGFR from The Joint Model for Meaningfully Interpretable Variables

Variable	Estimate (95% CI)	P-Value
Age	-0.010 (-0.011 - -0.009)	< 0.001
Sex - Male	0.041 (0.023 – 0.059)	< 0.001
Atrial Fibrillation - Yes	-0.033 (-0.049 - -0.017)	< 0.001
SBP	0.001 (0.001 – 0.002)	< 0.001
NYHA Class II	-0.002 (-0.034 – 0.033)	0.889
NYHA Class III	-0.003 (-0.041 – 0.035)	0.872
NYHA Class IV	0.020 (-0.078 - 0.117)	0.676
Diabetes - Yes	-0.037 (-0.052 - -0.021)	< 0.001

The parameter estimates for variables which include the natural cubic spline have been omitted as they are not directly interpretable but instead are illustrated by Figure 44.

Figure 43 Average Trajectories of log NT-ProBNP by Treatment Group from the multivariate joint model with Observed Values and Trajectories on NT-ProBNP

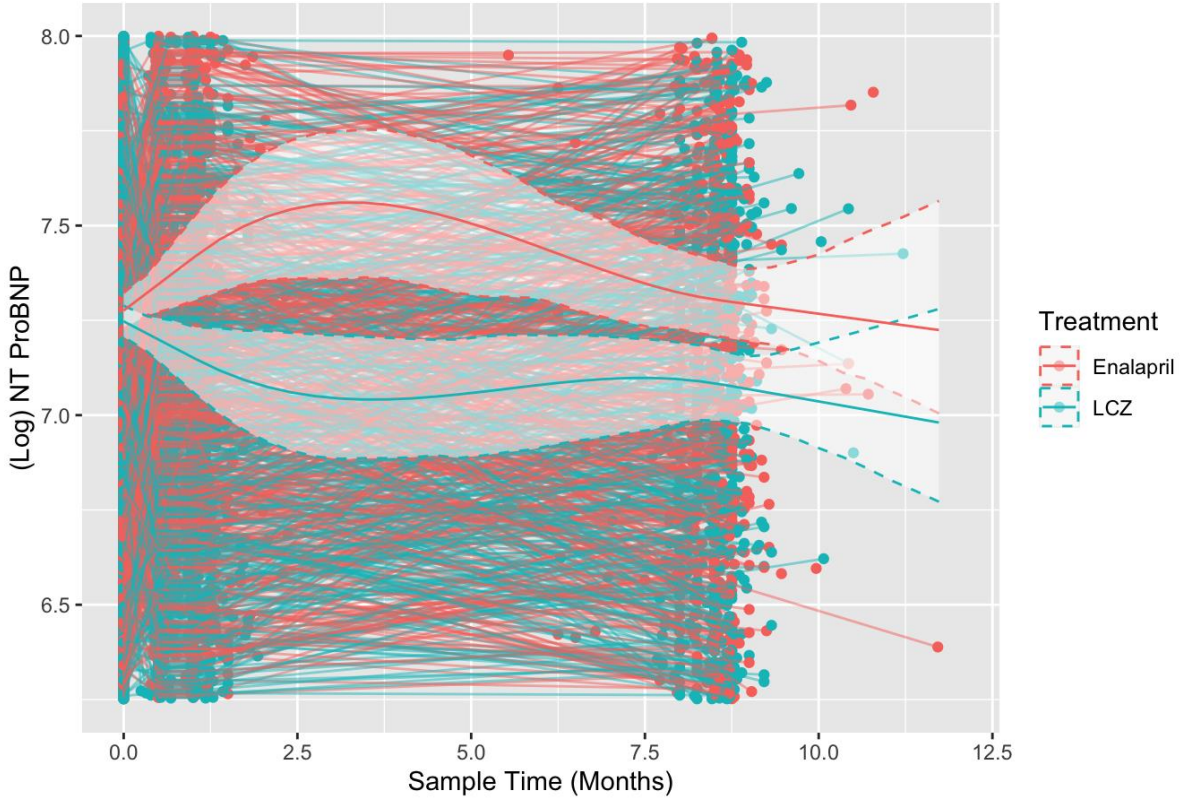
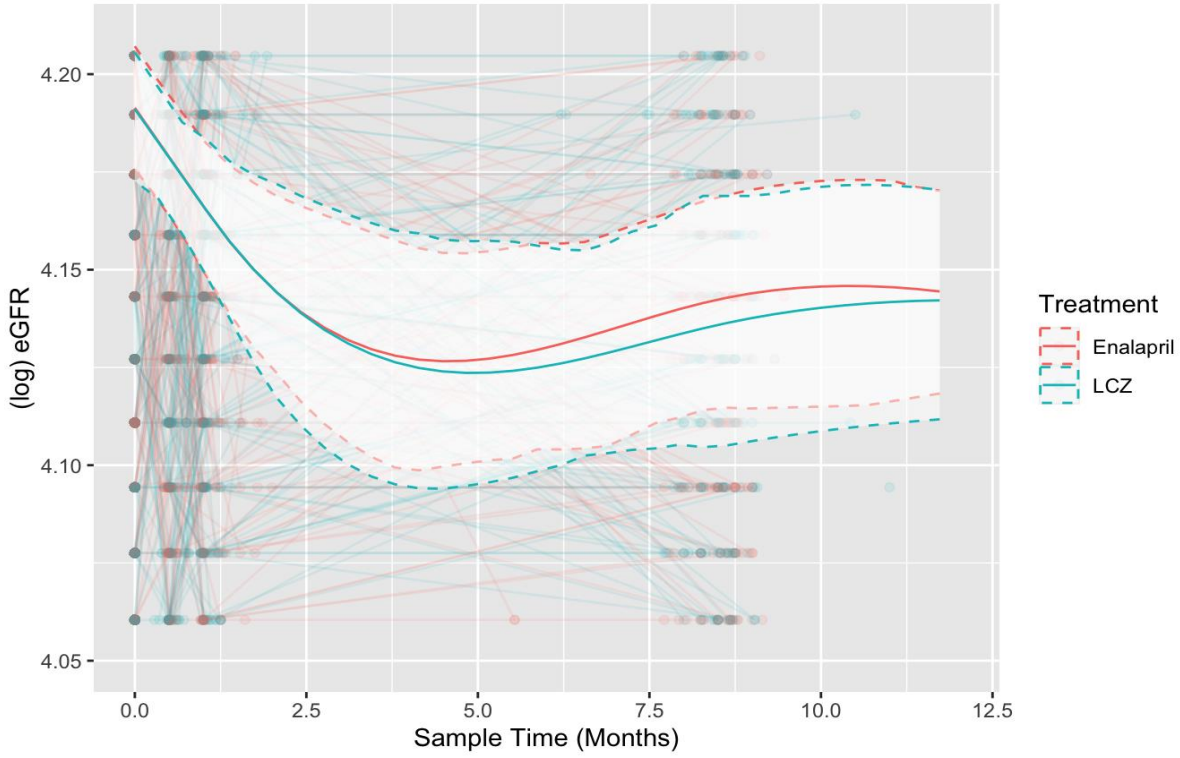


Figure 44 Average Trajectories of log eGFR by treatment group from the multivariate joint model with Observed Values and Trajectories Scaled for Readability.



6.4.5 Survival Outcomes from Joint Model

Table 40 illustrates the HRs with 95% CIs and P-Values for the variables of the survival component of the multivariate joint model. From the variables, only male sex, BMI, ejection fraction (%), NYHA Classes II-IV, heart rate, prior hospitalisation for heart failure, value of log NT-ProBNP and value of log eGFR meet the threshold to be considered statistically significant. Which suggests that the remaining variables may not be statistically significant. From the four alpha parameters (value of log NT-ProBNP, slope of NT-ProBNP, value of log eGFR and slope of log eGFR), both the slope of NT-ProBNP and slope of NT-ProBNP fail to meet the significance threshold. This may suggest that the slope parameter of NT-ProBNP may not be important to the model, however it has been previously shown in Chapter 5, to provide a better conditional fit in joint models with the same data. Similarly, the slope parameter of log eGFR is also considered to be statistically insignificant, however this may be due to the limited number of repeated measurements of log eGFR in this model. The HR for the value parameter for the log NT-ProBNP longitudinal outcome suggests that per unit increase in log NT-ProBNP, the hazard of the composite event increases by 1.68 times. The HR for the value parameter for the log eGFR longitudinal outcome suggests that per unit increase in log eGFR the hazard of the composite event decreases by 0.73 times.

The HR for treatment effect suggests that there is no significant treatment effect with LCZ, however this is likely due to treatment being mediated through the NT-ProBNP longitudinal outcome.

Table 40 Hazard Ratios (95% CI) for Variables in the Multivariate Joint Model

Variable	HR (95% CI)	P-Value
Treatment - LCZ	0.94 (0.77 - 1.14)	0.497
Age	1 (0.99 - 1.01)	0.722
Sex - Male	1.29 (1.13 - 1.48)	< 0.000
Region - Latin America	1 (0.85 - 1.18)	0.998
Region - North American	1.07 (0.88 - 1.31)	0.525
Region - Pacific Asia/Pacific and Other	1.12 (0.95 - 1.32)	0.175
Region - Western Europe	0.94 (0.81 - 1.08)	0.367
BMI	1.02 (1 - 1.03)	0.008
Ejection Fraction %	0.98 (0.97 - 0.99)	< 0.000
NYHA Class 2	1.3 (1 - 1.71)	0.052
NYHA Class 3	1.69 (1.29 - 2.25)	< 0.000
NYHA Class 4	1.93 (1.1 - 3.31)	0.024
Diabetes - Yes	1.42 (1.28 - 1.57)	< 0.000
SBP	1 (0.99 - 1)	0.186
Heart Rate	1.01 (1 - 1.01)	0.001
History of Atrial fibrillation – Yes	1.1 (0.98 - 1.24)	0.105
Ischemic Heart Failure - Yes	1 (0.87 - 1.16)	0.998
Prior Hospitalisation for Heart Failure - Yes	1.45 (1.3 - 1.63)	< 0.000
Prior History of Myocardial Infarction - Yes	1.19 (1.04 - 1.37)	0.013
Prior History of Stroke - Yes	1.13 (0.96 - 1.34)	0.157
Value of Log NT-ProBNP	1.68 (1.54 - 1.85)	< 0.000
Slope of Log NT-ProBNP	1.58 (0.73 - 3.31)	0.232
Value of Log eGFR	0.73 (0.55 - 0.98)	0.036
Slope of Log eGFR	0.03 (< 0.00 - 97.71)	0.339

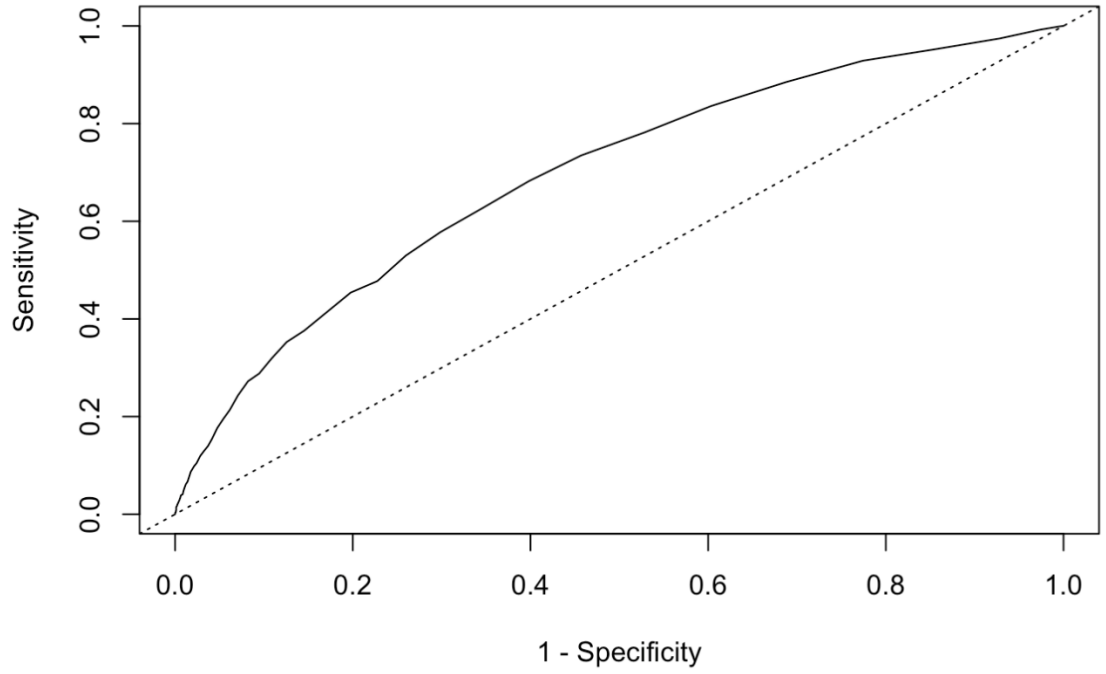
Table 41 Marginal and Conditional Performance Statistics from the Multivariate Joint Model

Statistic	Marginal	Conditional
DIC	40506.50	83133.98
LPML	-24769.74	-51799.76

6.4.6 Joint Model Performance

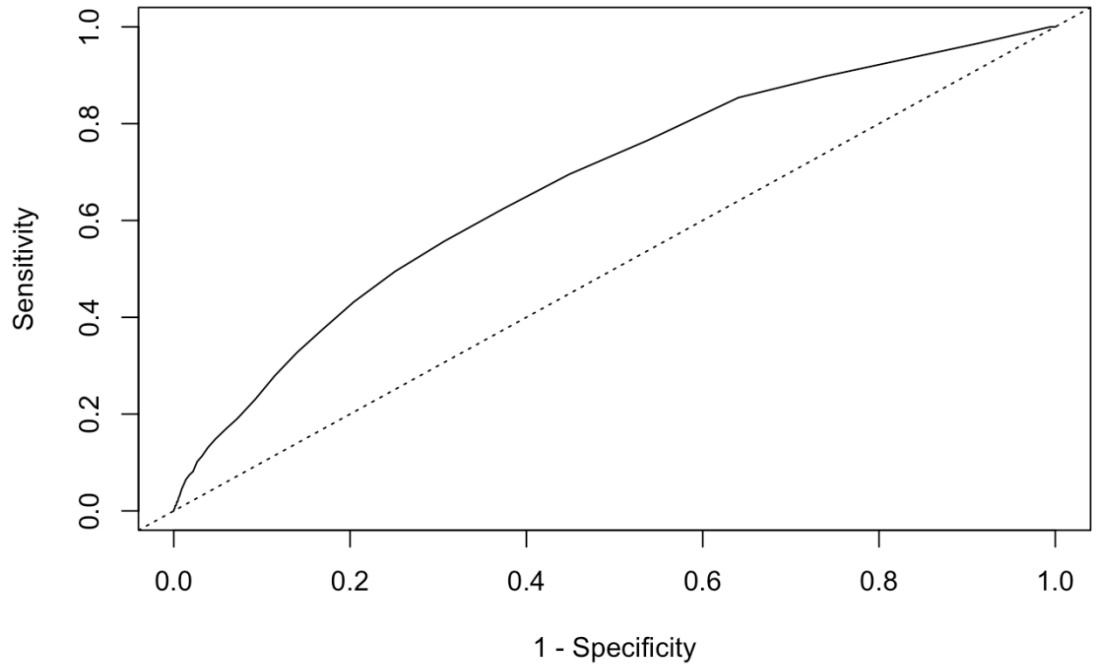
Table 41 Shows the marginal and condition performance statistics, which suggest that for both DIC and LPML statistics the marginal fit is better.

Figure 45 Multivariate Joint Model Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0



Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

Figure 46 Multivariate Joint Model Time-Varying ROC, AUC and Brier Score at Month 24 Using Longitudinal Data up until Month 12



Time-Varying AUC: 0.67
Time-Varying Brier Score: 0.08

Figure 47 Multivariate Joint Model Calibration Curve at Month 12 using Longitudinal Data at Month 0

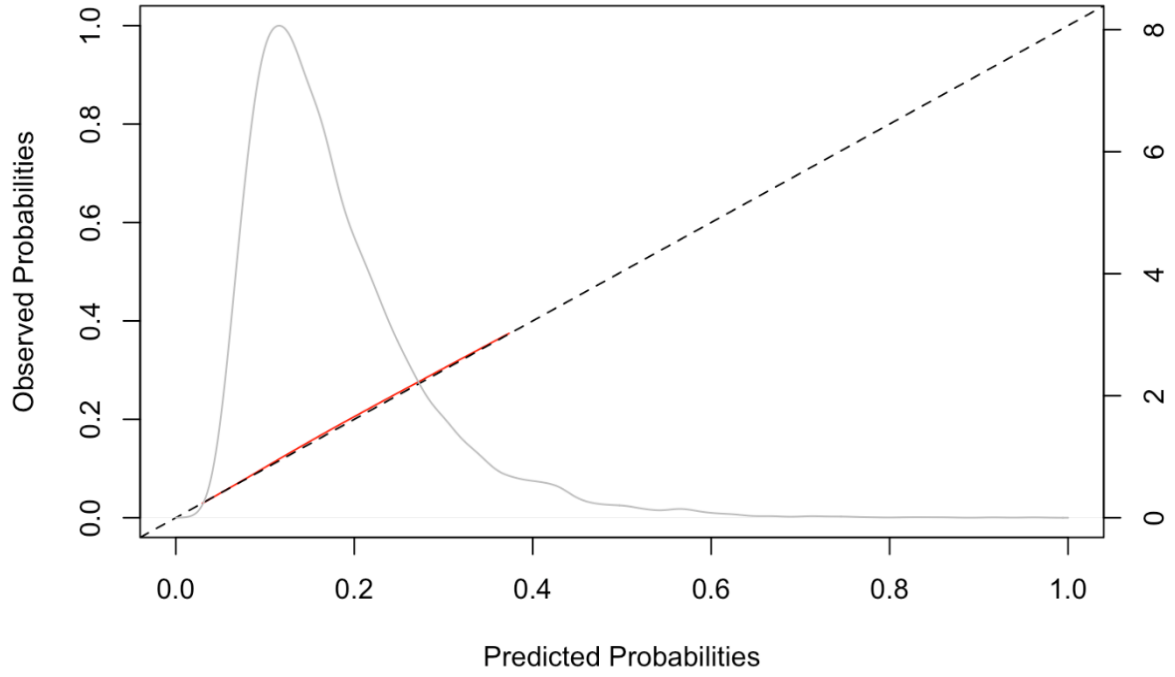


Figure 48 Multivariate Joint Model Calibration Curve at Month 24 using Longitudinal Data up until Month 12

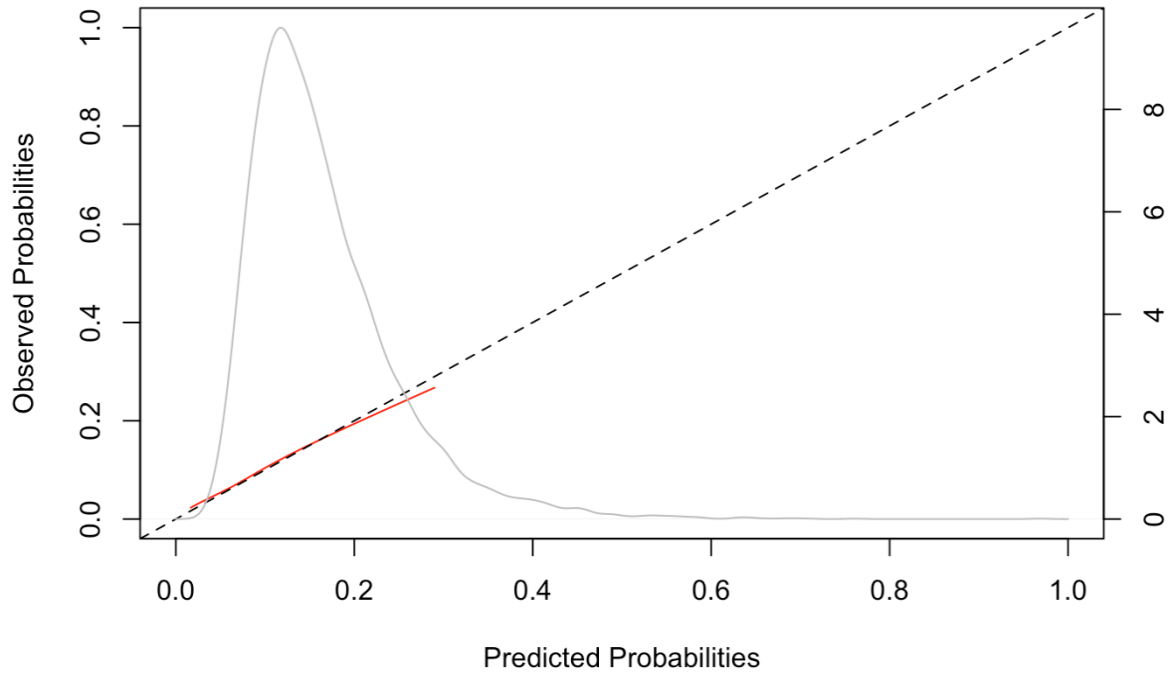
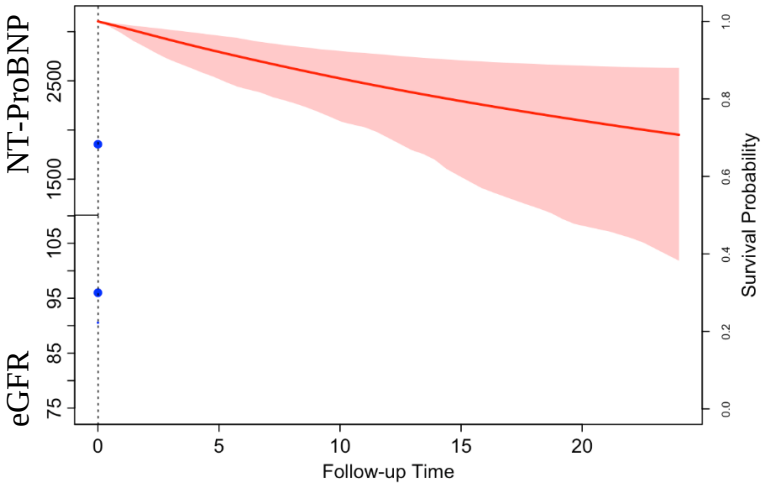


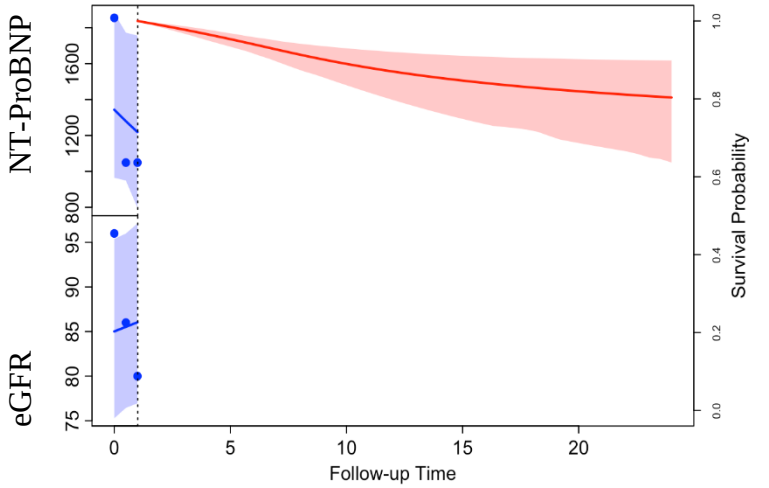
Figure 49 Dynamic Predictions from Multivariate Joint Model for a Randomly Sampled Patient with Measurements of eGFR and NT-ProBNP, and Corresponding Survival Probability

A.) Observation 1, eGFR and NT-ProBNP Measurements at Month 0 with Corresponding Survival Probability



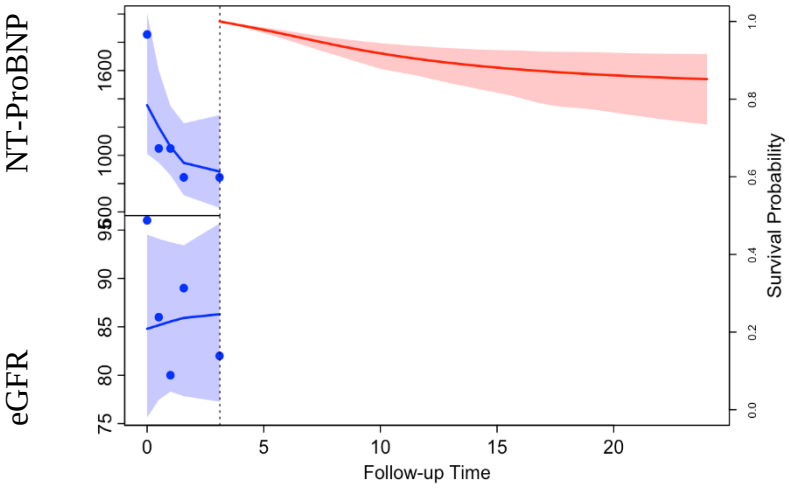
Survival at 24 months: 0.73 (0.87 – 0.44)

B.) Observation 3, eGFR and NT-ProBNP Measurements at Months 0, 0.5 and 1 with Corresponding Survival Probability



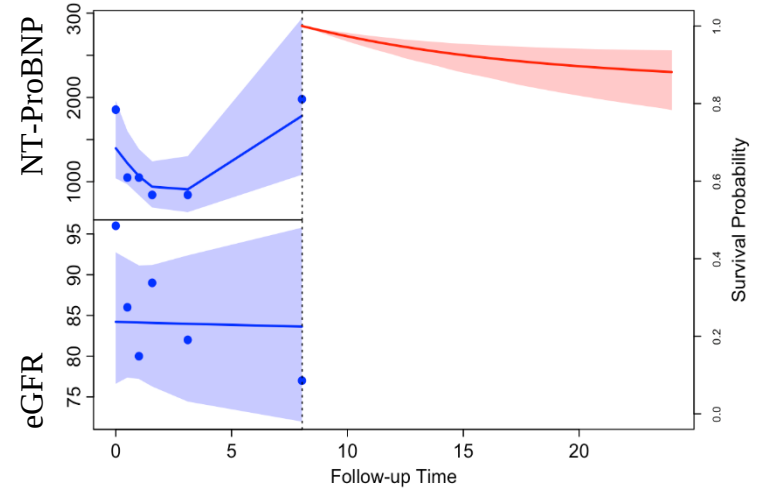
Survival at 24 months: 0.80 (0.90 – 0.64)

C.) Observation 5, eGFR and NT-ProBNP Measurements at Months 0, 0.5, 1, 1.5, and 3 with Corresponding Survival Probability



Survival at 24 months: 0.85 (0.92 – 0.73)

D.) Observation 6, eGFR and NT-ProBNP Measurements at Months 0, 0.5, 1, 1.5, 3 and 8 with Corresponding Survival Probability



Survival at 24 months: 0.88 (0.94 – 0.78)

eGFR and NT-ProBNP are represented by the blue lines on the left with 95% CI, the dashed line represents the last known point the patient was alive, and the red line indicates survival probability with 95% CI.

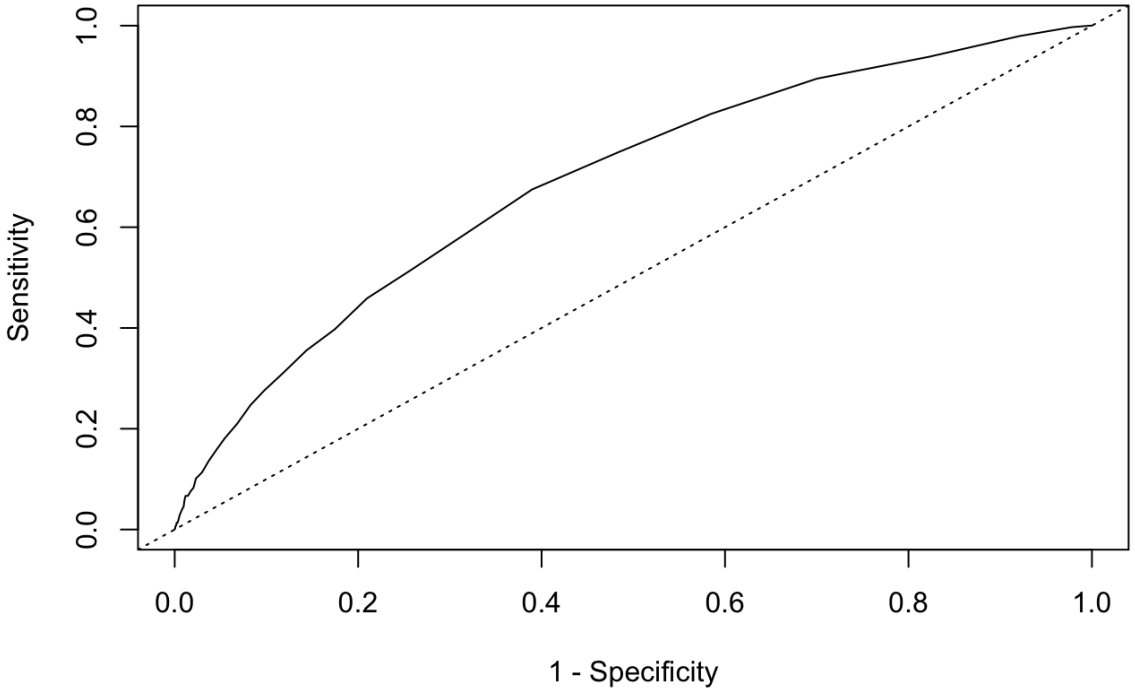
Figure 45 and Figure 46 show the time-varying ROCs, AUCs and Brier scores for the multivariate joint model at month 12 using baseline data and month 12 using longitudinal data up until month 24 respectively. The time-varying AUCs suggest that the model performs better at month 12 (using baseline longitudinal data) than it does at month 24 (using longitudinal data until month 12), with time-varying AUCs of 0.69 and 0.67 respectively. However, the time-varying Brier score suggests the opposite with the model performing better at month 12 (using baseline longitudinal data) than it does at month 12 (using longitudinal data until month 12) with time-varying Brier scores of 0.10 and 0.08 respectively.

The time-varying calibration curves for both time points suggest that the model is well calibrated with only minor deviation towards the predicted probabilities at the higher end of the scale at the second time point.

6.4.7 Dynamic Predictions from Joint Model

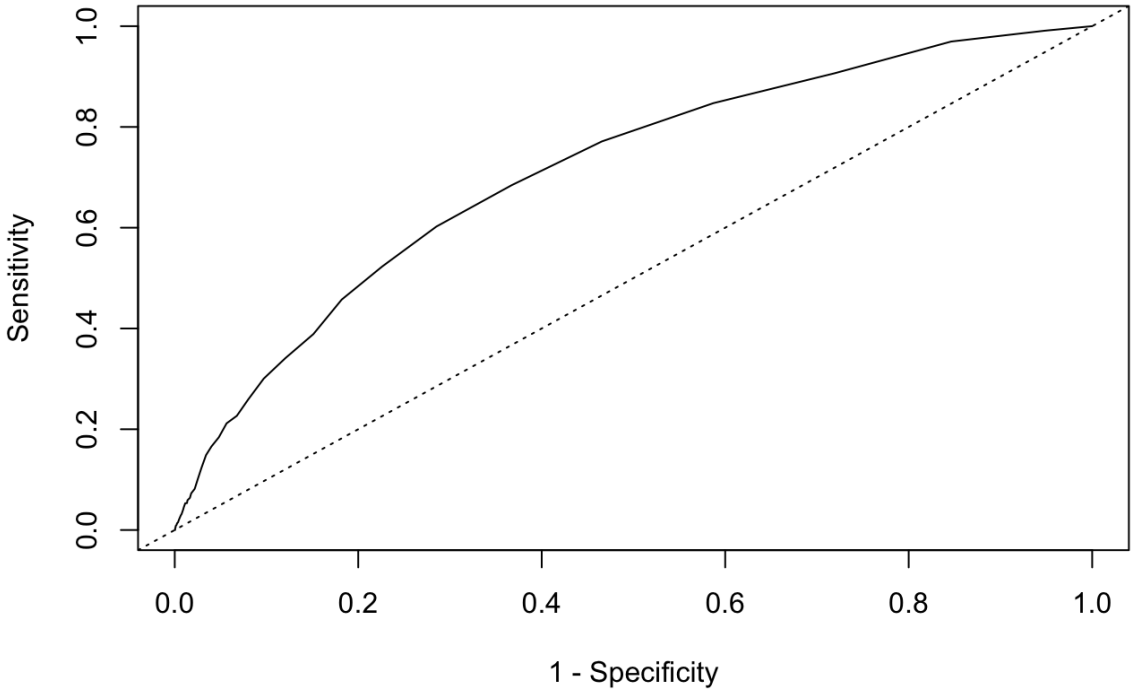
Figure 49 illustrates the dynamic predictions from a randomly sampled patient for four time points from the multivariate joint model. These predictions show the observed measurements of eGFR and NT-ProBNP with the predicted trajectories from the model, along with the corresponding survival curves. Overall survival at 24 months increased with more measurements, however, this may be due to the patient surviving longer. Overall, the 95% CIs for the survival curves are wide with the narrowest being at observation 5, this may be explained by the closeness of the observed measurements to the predicted trajectory and the closeness of the measurements themselves.

Figure 50 Multivariate Joint Model - External Validation Time-Varying ROC, AUC and Brier Score at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



Time-Varying AUC: 0.69
Time-Varying Brier Score: 0.10

Figure 51 Multivariate Joint Model - External Validation Time-Varying ROC, AUC, and Brier Score at Month 24 using Longitudinal Data until Month 12 Data from the ATMOSPHERE Trial



Time-Varying AUC: 0.71
Time-Varying Brier Score: 0.90

Figure 52 Multivariate Joint Model - External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial

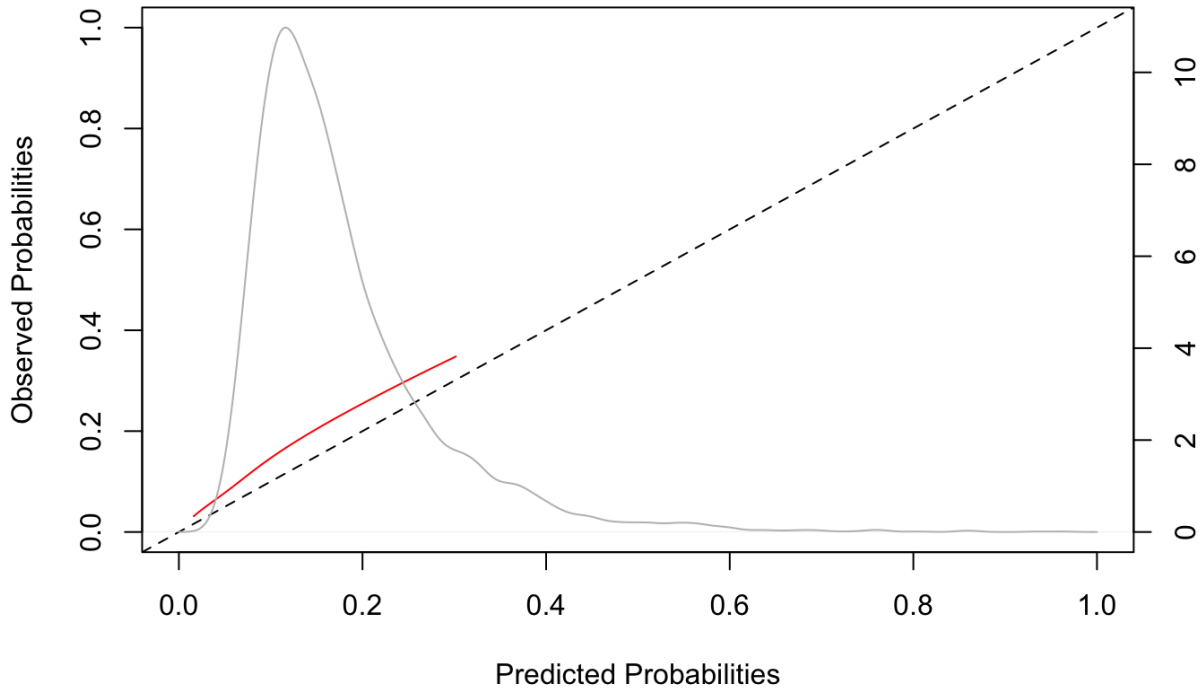
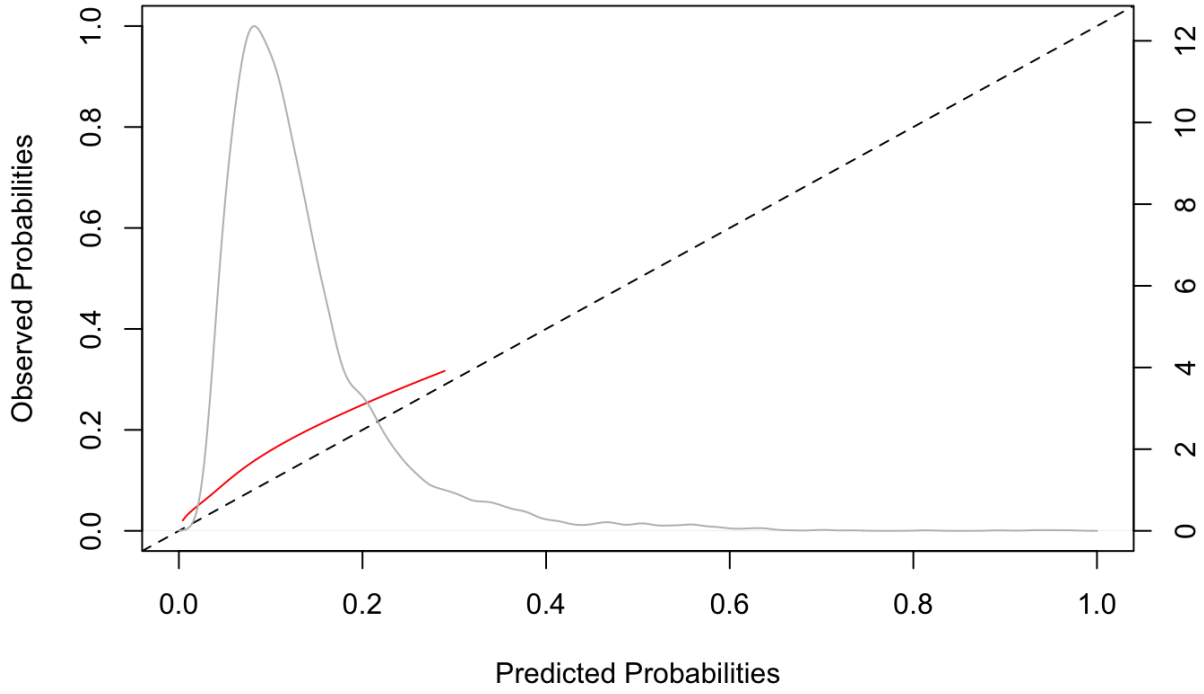


Figure 53 Multivariate Joint Model External Validation Calibration Curve at Month 12 using Longitudinal Data at Month 0 Data from the ATMOSPHERE Trial



6.4.8 External Validation

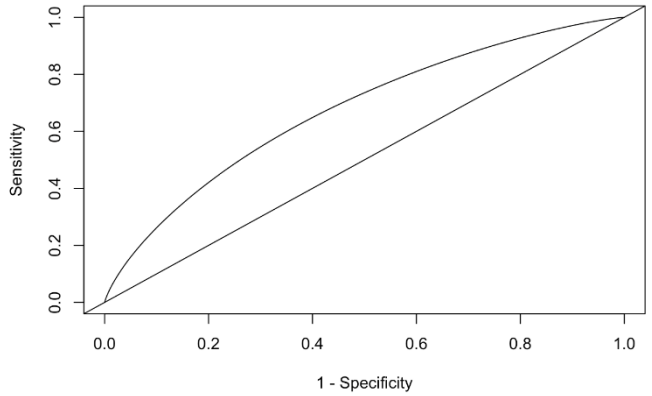
The time-varying AUC's ROC's and Brier scores for the multivariate joint model at month 12 (using baseline longitudinal data) and month 12 (using longitudinal data up until time point 12) are shown in Figure 50 and Figure 51 respectively. The time-varying AUC's suggest the model performs the same with the ATMOSPHERE data at the month 12 (using baseline longitudinal data), however the time-varying AUC at month 12 (using longitudinal data up until time point 12) suggests that the model performs better with the ATMOSPHERE data than it does with the original data suggesting 0.05 point increase in time-varying AUC, moving the model into the acceptable category for discrimination with a score of 0.71. However, the time-varying Brier score and calibration curves suggest that the calibration is worse with the ATMOSPHERE data, with the calibration curves showing more deviation towards the observed probabilities with a greater deviation in the middle of the curve. While the time-varying Brier score remains the same at month 12 (using baseline data), it shows an increase of 0.01 to 0.09 at time point 24 (using longitudinal data up until time point 12), suggesting a small drop in overall performance.

Table 42 Hazard Ratio (95% CI) and P-Values from the Cox PH Models

	Cox PH Last Measurement		Extended Cox PH	
Variable	**HR (95% CI)**	**P-Value**	**HR (95% CI)**	**P-Value**
Treatment - LCZ	0.83 (0.76-0.91)	<0.001	0.83 (0.76-0.91)	<0.001
Age	1 (1-1)	0.967	1 (1-1)	0.99
Sex – Male	1.29 (1.15-1.45)	<0.001	1.29 (1.14-1.45)	<0.001
Region - Latin America	0.99 (0.85-1.15)	0.871	1.02 (0.88-1.19)	0.762
Region - North American	1.03 (0.86-1.22)	0.779	0.97 (0.81-1.17)	0.779
Region - Pacific Asia/Pacific and Other	1.08 (0.93-1.25)	0.31	1.12 (0.97-1.31)	0.132
Region - Western Europe	0.92 (0.81-1.05)	0.2	0.9 (0.79-1.03)	0.129
BMI	1.02 (1.01-1.03)	0.001	1.02 (1.01-1.03)	0.001
Ejection Fraction %	0.98 (0.97-0.99)	<0.001	0.98 (0.97-0.99)	<0.001
NYHA Class II	1.23 (0.96-1.57)	0.1	1.19 (0.93-1.53)	0.162
NYHA Class III	1.53 (1.18-1.98)	0.001	1.5 (1.16-1.95)	0.002
NYHA Class IV	1.79 (1.08-2.94)	0.023	1.87 (1.13-3.08)	0.014
Diabetes - Yes	1.36 (1.24-1.5)	<0.001	1.36 (1.24-1.49)	<0.001
SBP	1 (1-1)	0.287	1 (0.99-1)	0.169
Heart Rate	1.01 (1-1.01)	0.002	1.01 (1-1.01)	0.005
History of Atrial fibrillation – Yes	1.09 (0.99-1.2)	0.093	1.07 (0.97-1.19)	0.177
Ischemic Heart Failure – Yes	1.01 (0.89-1.14)	0.888	0.99 (0.87-1.13)	0.915
Prior Hospitalisation for Heart Failure - Yes	1.37 (1.25-1.51)	<0.001	1.35 (1.22-1.49)	<0.001
Prior History of Myocardial Infarction - Yes	1.2 (1.06-1.36)	0.003	1.21 (1.07-1.37)	0.002
Prior History of Stroke - Yes	1.09 (0.94-1.26)	0.238	1.09 (0.94-1.26)	0.263
Log NT-ProBNP	1.6 (1.53-1.68)	<0.001	1.61 (1.54-1.69)	<0.001
Log eGFR	0.74 (0.63-0.87)	<0.001	0.73 (0.62-0.87)	<0.001

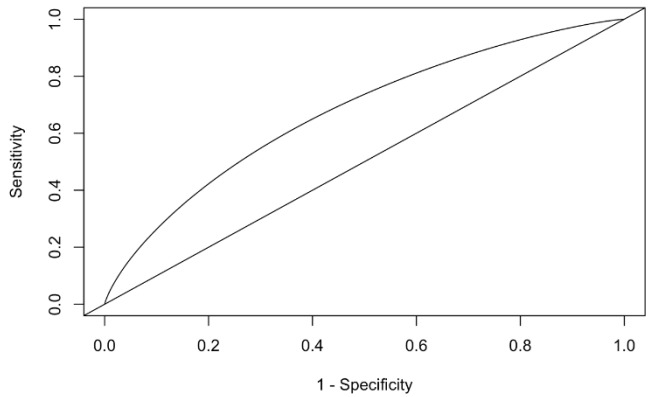
Figure 54 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 12

A.) Time-Varying ROC and AUC at 12 Months for the Cox PH Last Measurement Model



Time-Varying AUC at 12 Months: 0.67

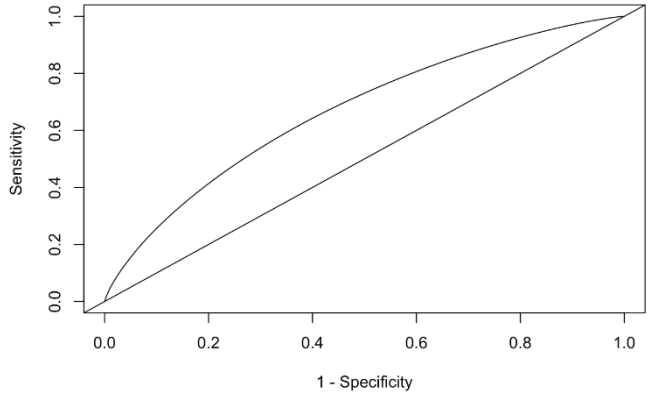
B.) Time-Varying ROC and AUC at 12 Months for the Extended Cox PH Model



Time-Varying AUC at 12 Months: 0.67

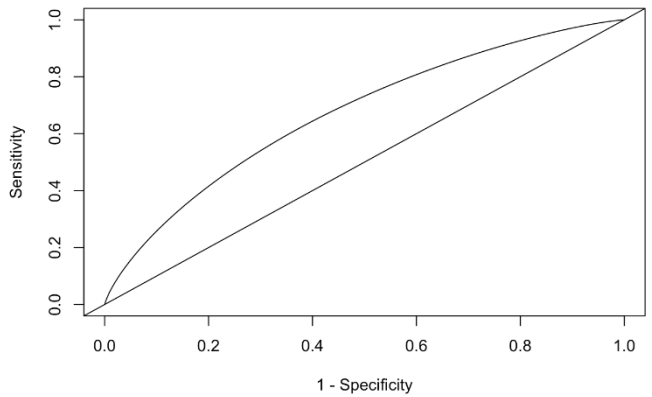
Figure 55 Time-Varying ROCs and Time-Varying AUCs for the Cox PH Models at Month 24

A.) Time-Varying ROC and AUC at 24 Months for the Cox PH Last Measurement Model



Time-Varying AUC at 24 Months: 0.67

B.) Time-Varying ROC and AUC at 24 Months for the Extended Cox PH Model



Time-Varying AUC at 24 Months: 0.67

6.4.9 Comparative Analysis

Table 42 shows the HRs (95%CI) and P-Values for both Cox PH models, illustrating differences between both the Cox PH model and the multivariate joint model. Between the two Cox PH models, there are marginal differences in both P-Values and HRs for the region parameters. However, all parameters suggest region is statistically insignificant in all models.

There are also differences between the two Cox PH models with regards to NYHA, the largest being NYHA Class IV, with the extended Cox PH model being closer to the multivariate joint model. However, there is no difference in significance between any of the models.

Finally, there is a difference of 0.07 and 0.06 when comparing Log NT-ProBNP of the last measurement Cox PH and the extended Cox PH model against the multivariate value of Log NT-ProBNP respectively, suggesting a difference between the models. However similar to Chapter 4 there is a significant treatment effect for both Cox PH models and not in the multivariate models, which again may suggest that the treatment effect is mediated by NT-ProBNP.

The time-varying ROCs and corresponding AUCs shown in Figure 54 suggest that there is no difference in prognostic performance between the two Cox PH models at either time point, with and time-varying AUC at both of 0.67. Comparison with the multivariate model suggest the Cox PH models perform worse at 12 months with an AUC of 0.69 for the multivariate model. However, the AUC at 24 month is the same for all models.

6.5 Discussion

With the rise in both awareness and usage of joint modelling, along with the need to explore joint modelling for the purpose of assessing the association of biomarkers and an endpoint, there is also a need to understand how joint modelling may be useful in the context of prognostic models, and if it can improve on the current methodology. To this end, a multivariate joint model was fit and using data from two clinical trials, this model was evaluated and compared against traditional models fit with the same data.

The multivariate JM suggests that there is a significant association between both log NT-ProBNP and log eGFR and the primary composite endpoint of cardiovascular death and first hospitalisation for heart failure. This further adds to the evidence base, suggesting a link between both biomarkers and adverse events for patients with heart failure.

Like Chapter 4, the HR for the treatment effect on the survival outcome suggests that there is no significant direct treatment effect on survival, suggesting instead that again the treatment effect is being mediated through NT-ProBNP. However, due to the inclusion of natural splines and the interaction of them with treatment, it is not possible to determine the overall treatment effect.

The multivariate model showed that both the value parameters for log NT-ProBNP and log eGFR were considered statistically significant, suggesting that per unit increase in log NT-ProBNP the hazard of the composite outcome increases by 1.68 times (P-Value <0.000), and per unit increase in log eGFR the hazard of the composite event decreases by 27%. It should be noted that due to the possibility that the treatment effect is being mediated through NT-

ProBNP, the treatment effect on Log NT-ProBNP should be considered when interpreting this parameter.

As with Chapter 4, the slope parameter of log NT-ProBNP does not meet the significance threshold, however as previously reported in Chapter 4, this parameter improved the subject specific fit, so while it may not be significant at a population level, it may add value for subject specific predictions. Similarly, the slope of eGFR is measurable at two decimal places, unlike Chapter 5. However, it is considered insignificant and has wide confidence intervals, as previously stated likely due to the limited repeated measurements of log eGFR in this model, compared to the previous chapter.

The average longitudinal trajectories illustrated in Figure 43 and Figure 44 suggest that for the average patient (using mode and median baseline characteristics) there is a difference in treatment effect over time on log NT-ProBNP. However, no such difference is illustrated with log eGFR. Whilst no difference between the treatment groups for log eGFR are illustrated, the predicted profile of log eGFR is shown still illustrated over time. It should be noted however that as the multivariate model is based around random effects, and as such the trajectories may vary by patient.

All directly interpretable parameters in the log NT-ProBNP longitudinal outcome are considered statistically significant, with treatment and BMI having a negative effect and age and the presence of atrial fibrillation having a positive effect on log NT-ProBNP. These results are supported by prior research, BMI has previously been shown to have an inverse relationship with NT-ProBNP in patients with chronic heart failure, while both age and atrial fibrillation have been shown to have positive relationship with NT-ProBNP [151]. While

these effects are small < 0.1, it should be noted that NT-ProBNP is represented on a log scale and that time is represented in months, and this should be taken into consideration when interpreting the coefficients.

In the log eGFR longitudinal outcome, all parameters but those for the NYHA Classes are considered statistically significant. With age, the presence of atrial fibrillation and diabetes having a negative effect, and the male sex and SBP having a positive effect. These results are supported by the literature as age, the presence of atrial fibrillation and diabetes have all been previously shown to have an inverse relationship with eGFR. While a positive relationship between the male sex, SBP and eGFR has also been previously demonstrated [34], [152]–[156]. As with the log NT-ProBNP outcome, the effect sizes are small < 0.1 however eGFR is represented on a log scale and time as months, so this need to be considered when interpreting the coefficients. It should also be noted that higher eGFR values are considered better so a negative effect of a coefficient may be positive for the patient.

For both DIC and LPML the model suggest that the marginal (population level) fit is better than the conditional (subject specific) fit. This may indicate that the model may not be as suitable for subject specific predictions as it is for population level coefficients and HRs, as a result the model may be more suitable to measuring the association of the biomarkers and the composite outcome.

The time-varying AUCs suggest that the model performs better with regards to prognostic fit at 12 months (using baseline data) than it does at 24 months (using longitudinal data at 12 months), however the time-varying Brier score suggest the model performs better at 24 months when calibration is taken into consideration. The time-varying AUCs at both time

points suggest that the model missed the threshold for acceptable performance when assessed at these two points using the respective longitudinal data.

External calibration with the ATMOSPHERE data suggests similar prognostic performance to when using the original data at month 12 (using baseline data) when assessing the time-varying AUC with the model, again missing the threshold for acceptable performance. However, the model with the ATMOSPHERE data shows an improvement over the original data at month 12 (using longitudinal data at month 12) of 0.5, moving the model into the acceptable range for prognostic performance. However, when calibration is also considered using the time-varying Brier score, the model shows a decrease in performance of 0.1 in Brier score at the 24-month time point. The performance with external data suggests that the model may provide a good fit for the validation data and that the data may be a better fit at 24 months (using longitudinal data at month 12) than the original data, possibly due to the amount of data available, however it may simply be due to the selection of patients used in the predictions used to calculate the time-varying AUCs.

The comparative analysis suggests that the joint model performed better than the Cox PH models at month 12 and similarly at month 24 with respect to time-varying AUC, suggesting that the joint models may be better with regards to performance at the first time point. While the time-varying AUCs use the same underlying methodology there are differences between the way the time-varying AUCs are calculated, meaning they should be interpreted with caution.

Altogether, the multivariate joint model performed adequately, and could potentially provide a good prognostic model. The multivariate model has the advantage of allowing for multiple

longitudinal responses and both subject specific and dynamic predictions over the current standard Cox PH models as well as allowing for repeated measurements, whilst handling the special properties of biomarkers. However, when compared to the two univariate joint models of Chapter 4 and Chapter 5 the multivariate models show little or no advantage, with respect to time-vary AUCs and Brier scores, suggesting with this data, there may not be an advantage of the multivariate joint model, possibly due to the limitations around the repeated measurements. It should also be acknowledged that both these advantages, are based on the premise of random effects which are latent and therefore unobservable. This needs to be considered when determining the suitability of the methodology of joint models as capturing these random effects relies on the correct specification of the longitudinal process as described in Chapter 1. Incorrect specification of the longitudinal process may result in incorrect specification of the random effects therefore affecting the validity of the subject specific predictions that rely on the random effects, leading to less accurate predictions for new patients.

There are some limitations of this analysis, including that of the use of the screening values of NT-ProBNP as well as the need to have the biomarkers taken at the same date. In this analysis, while the visits were scheduled for the specified dates, the actual samples of both biomarkers were not always collected on the same date and due to this limitation, required the sample dates to be joint using the prespecified date and use of one of the sample dates over the other. There is also an in balance of data at the beginning of the longitudinal data, paired with the use of screening values of NT-ProBNP at time zero; this could potentially introduce bias. It should also be noted that the inclusion and exclusion criteria of the trial may not be completely generalisable and limits the model to HFrEF, however this is a limitation of the data itself and therefore outside the scope of this research. While the results of the

multivariate joint model are close to those of the LOCF Cox PH model, the *JMbayes2* package whilst based on previous packages (JM and JMbayes) is still in development and therefore this analysis is limited to the functionality of the package and its subsequent accuracy measures. On this point, it can be noted that the calibration curves do not include confidence intervals which would further enhance the interpretation of calibration.

Chapter 7 Real-World Application of Prognostic Joint Models Illustrated Using a Web Application

7.1 Introduction

In Chapter 4, Chapter 5 and Chapter 6 joint models were fit for the purpose of evaluating the suitability of joint models as prognostic models. However, one key factor in prognostic modelling is the usability of the models in real-world application such as clinical practice. While Chapter 4, Chapter 5 and Chapter 6 showed how the joint models performed and the various outputs of the joint models, these are only useful in prognosis if they can be easily generated for new patients. However, generation of these outputs typically requires knowledge of the R code and packages. As illustrated in Chapter 1 prognostic models are often accompanied by user-friendly interfaces such as web applications, which allow the models to be used in real world application. In the case of prognostic modelling, real world application could be that of a clinician / patient consultation. The choice to use R for the joint model presents a unique opportunity to make use of R Shiny, an add-on to R to easily create responsive web applications with minimal effort (requiring less code), whilst allowing for the use of R functionality within the application. Therefore, the purpose of the chapter is to illustrate how a prognostic joint model could be used in real-world application through the use of an R Shiny web application.

7.2 Methods

7.2.1 Choice of Joint Model

For ease of use of the web application a single joint model was selected to be used in the web application. To do this the models were considered on their reported performance characteristics. Both joint model 2 from Chapter 4 (log NT-ProBNP and a composite endpoint of death from cardiovascular causes and first hospitalisation for heart failure using the value and slope of log NT-ProBNP as alpha parameters) and the multivariate joint model from Chapter 6 (log NT-ProBNP and Log eGFR, and a composite endpoint of death from cardiovascular causes and first hospitalisation for heart failure using the value and slope of log NT-ProBNP, and value of log eGFR as alpha parameters) performed similarly with regards to time-varying AUC and time-varying Brier score. Either of these two models would have been suitable, however the multivariate model requires an additional biomarker to be collected repeatedly at the same time point, and for this reason joint model 2 from Chapter 4 was chosen for use in the web application. While this model was chosen for this application, it should be noted that further evaluation of the models may be prudent prior to the use of this application in clinical practice.

7.2.2 R Shiny Application

An R Shiny application requires two components, a User Interface (UI) and a Server, these components form the basis of the application, with the UI providing the functionality that the user will interact with and the server component containing the code that interacts with the data and model.

7.2.2.1 UI

The user interface needs to contain all the user inputs and outputs. For the chosen joint model, the user input was separated into two categories, one for baseline measurements and one for repeat measurements. These categories form a tab within the application; these tabs are essentially pages within the application that users can interact with. Both tabs contain the same output, a dynamically generated figure showing the repeated measurements of log NT-ProBNP on one side and survival probability on the other, this figure is like the dynamic predictions from Figure 20, Figure 33, and Figure 49, with the exception that there is only a single graph containing all time points, which is dynamically updated dependent on the user input instead of multiple graphs for each measurement point that have been added. Along with this graph each tab also contains a point estimate of the survival probability at a user-selected time point. The UI also contains a third tab named ‘debug’ which simply shows the user the formatted data.

7.2.2.2 Server

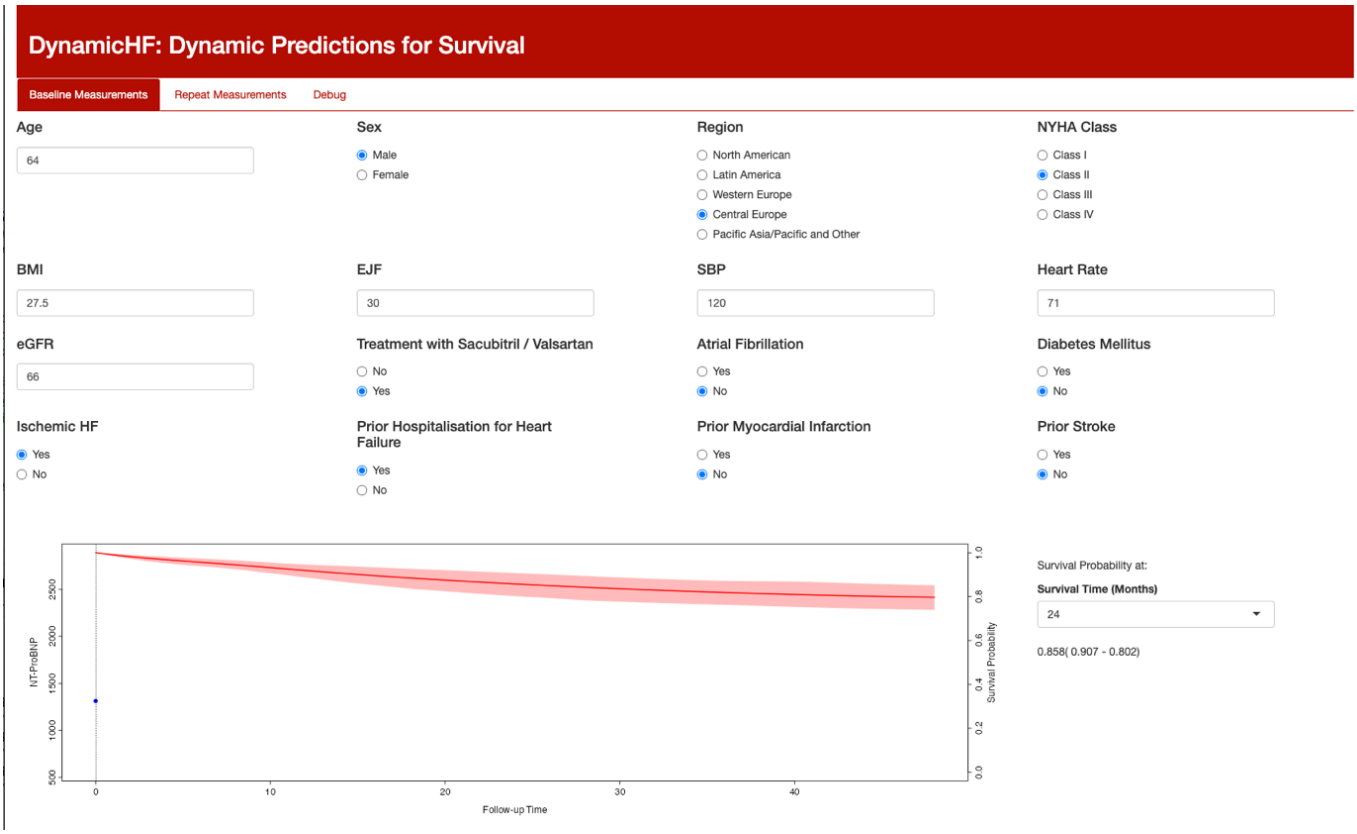
The server contains the code for the processing of the user input and for interacting with the joint model. The server code makes use of R shiny reactive functionality (which allows the user interface to be updated automatically based on the user input, without the need to press any additional button e.g., an update button). The server code contains processing and formatting scripts to take the user input and make it readable by the R function of the *JMbayes2* package. The code makes use of the *predict* and *plot* functions in *JMbayes2*, formatting the user input into the correct format for the new data argument of the *predict* function.

7.3 Results

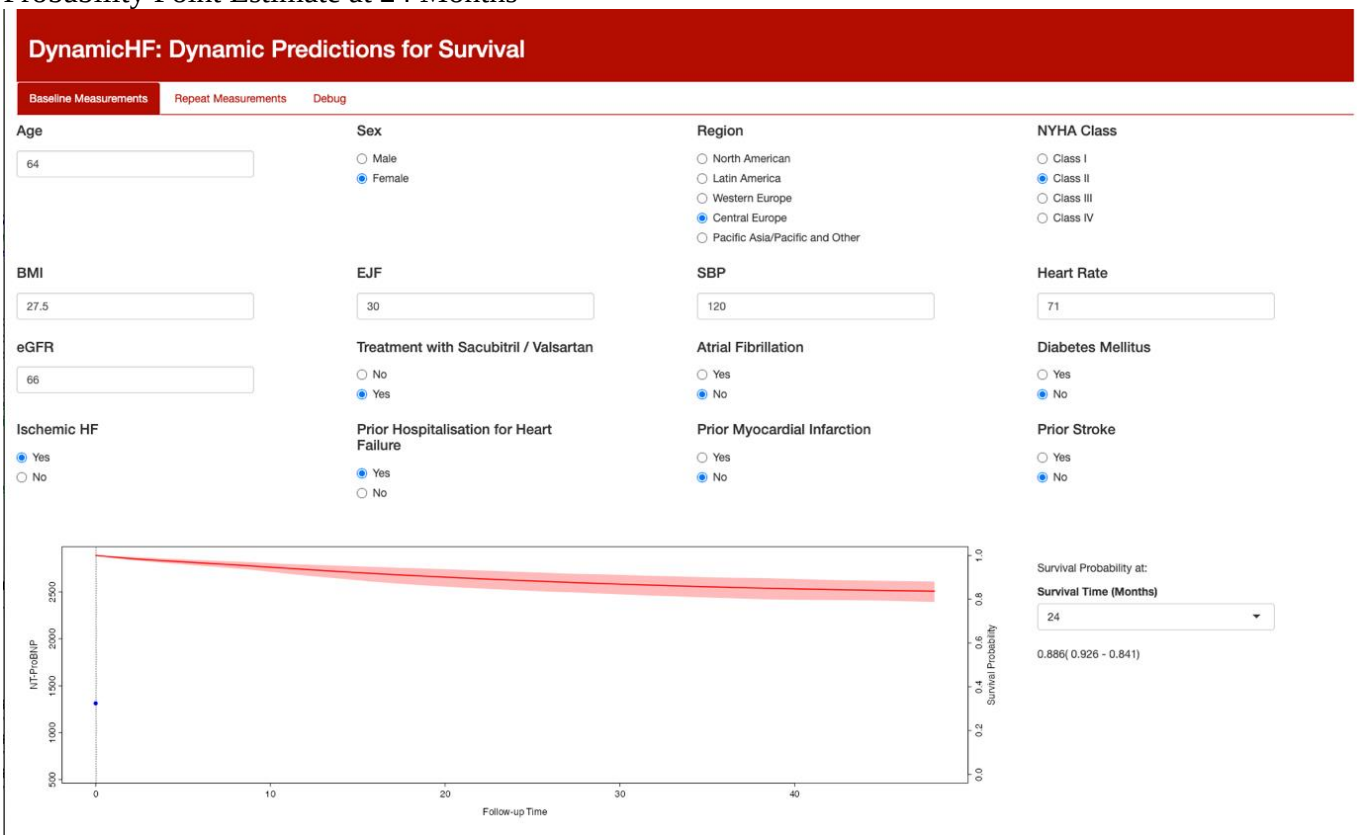
Figure 56, Figure 57 and Figure 58 show screenshots of the three tabs of the web application, demonstrating its functionality.

Figure 56 Web Application User Interface for Baseline Measurements Tab

A.) Baseline Measurement Tab of Web Application with Male Sex Selected, Showing Survival Probability Point Estimate at 24 Months



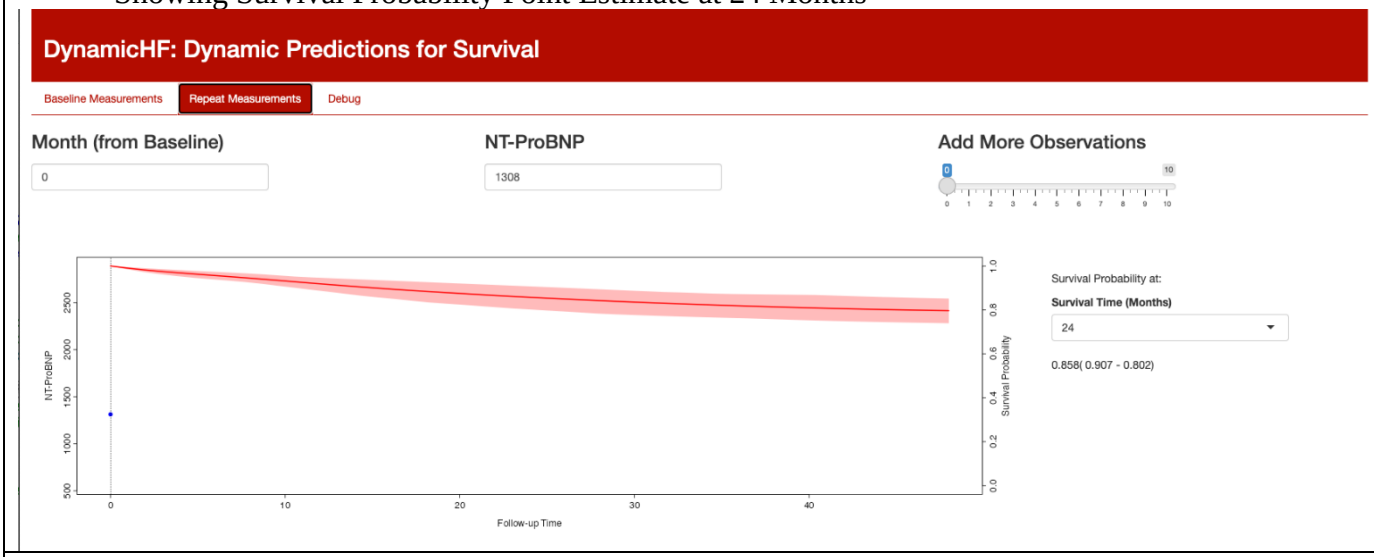
B.) Baseline Measurement Tab of Web Application with Female Sex Selected, Showing Survival Probability Point Estimate at 24 Months



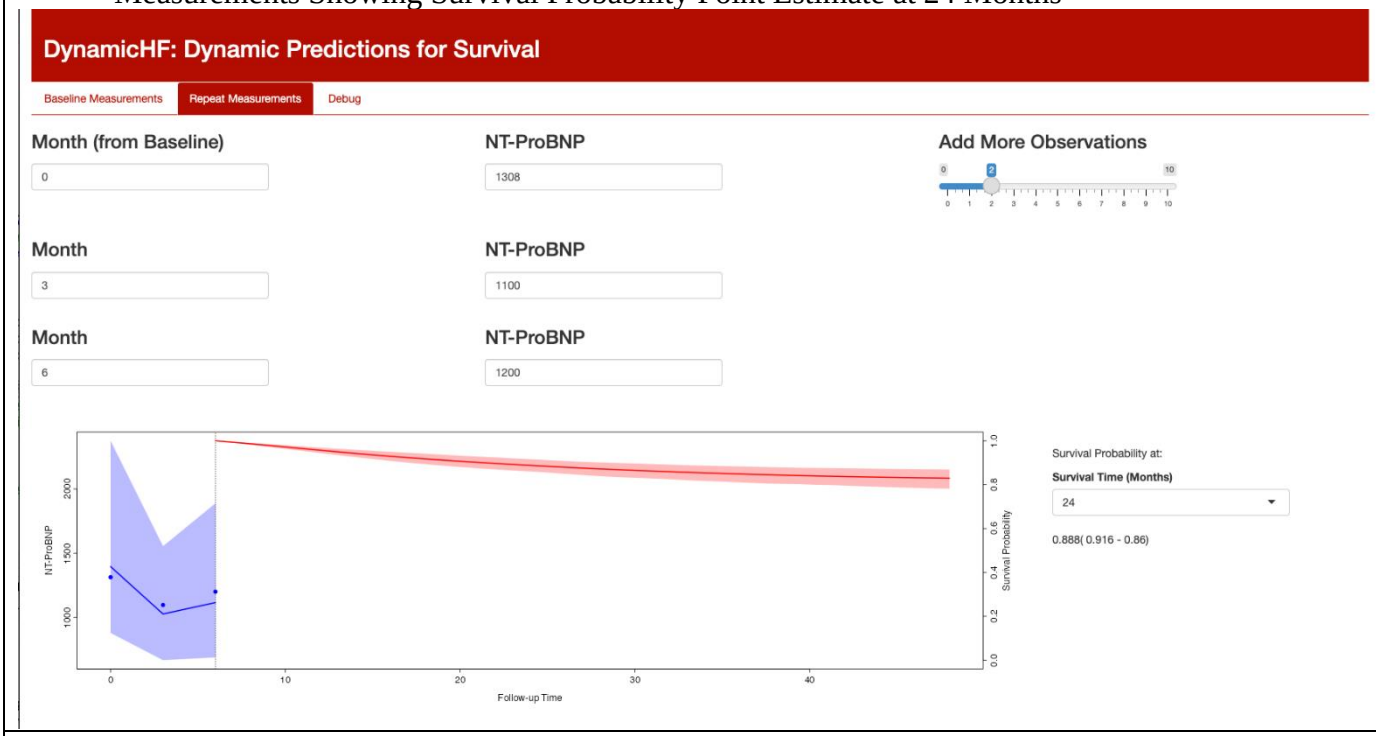
All other variables are set to the mode / median values with the exception of treatment which has been preset to Treatment with Sacubitril / Valsartan: Yes, the first measurement of NT-ProBNP (Baseline) has been set to the median value for NT-ProBNP (1308) as in Figure 57A.

Figure 57 Web Application User Interface for the Repeat Measurements Tab

A.) Repeat Measurement Tab with Single Measurement of NT-ProBNP Set to Median Measurement Showing Survival Probability Point Estimate at 24 Months



B.) Repeat Measurements Tab with 3 Measurements of NT-ProBNP Set to Median and Random Measurements Showing Survival Probability Point Estimate at 24 Months



The baseline characteristics are the same as selected in Figure 56A

Figure 58 Web Application User Interface for the Debug Tab

Repeat and baseline measurements are the same as Figure 56A and Figure 57B.

7.3.1 Baseline Measurements Tab

Figure 56A illustrates the user interface for the baseline measurements tab of the web application. From this figure it can be seen how all the baseline measurements can be set by the user. The web application fills in default values for the baseline measurements based on median and mode values from the original dataset (PARADIGM); except for Treatment with Sacubitril / Valsartan which is set to ‘yes’ by default. This tab also shows a graph for the dynamic prediction. This graph has been previously described but briefly contains the measurements of NT-ProBNP on the left and corresponding survival curve on the right with 95% CI, the dashed line is the last point the patient was known to be alive (last measurement of NT-ProBNP). This graph is updated whenever the user changes any of the baseline measurements. It includes by default one measurement of NT-ProBNP at month 0 (considered to be baseline), this measurement can be seen in Figure 57A as being set to the median NT-ProBNP value of 1308. Accompanying the dynamic prediction graph is a point estimate of the survival probability, which has been set to 24 months.

Figure 56B shows the same tab, but with sex changed from Male to Female; with this change you can see the change in survival probability at 24 months has changed from 0.858 (0.907 – 0.802) to 0.886 (0.926 -0.841) a change of 2.8%, this illustrates a change at baseline between males and females with the same measurement of NT-ProBNP.

Figure 57A illustrates the repeat measurements tab of the web application. With one measurement prefilled, based on the median NT-ProBNP value from the original (PARADIGM) data. This tab has a selector where the user can change the number of observations, depending on the number of measurements available. It also contains the same dynamic prediction graph from the baseline measurements tab and survival

probability point estimate which has been set to 24 months and has the same point estimate as Figure 56A of 0.858 (0.907 – 0.802).

Figure 57B shows the repeat measurements tab with two additional randomly selected measurements, by default however, the spacing between measurements is set to 3 months but can be adjusted by the user. In this figure the dynamic prediction graph has been updated to reflect the additional measurements. The left side of the graph now contains the observed measurements as blue dots and the predicted trajectory from the joint model with the corresponding 95% CIs. It can also be seen that the point estimate for survival probability at 24 months has been updated to 0.888 (0.916 – 0.860) an increase on the previous value of 3%. Possibly due to the patient surviving longer and the relatively stable profile of NT-ProBNP.

Figure 58 shows the debug tab, the purpose of which is to allow the user to confirm that the data the web application is using is correct and has not been corrupted during formatting.

7.4 Discussion

This chapter has explored how joint models can be translated into real-world applications using a web application.

The web application allows for a user such as a clinician to obtain dynamic predictions of survival (the probability of death by cardiovascular causes and first hospitalisation for heart failure) based on the data available, including baseline characteristics and repeat measurements of NT-ProBNP. Alongside this, it adds the potential to see what future survival probabilities may be, say, if a patient’s baseline characteristics and/or the trajectory of NT-ProBNP were to change, making it a powerful tool.

The dynamic and responsive nature of the web application makes it both user-friendly, and potentially more valuable than current web-based prognostic tools.

While this application may not be suitable for use by individual patients without a clinician, due to the requirement of access to biomarker measurements, if the models were fitted with ambulatory measurements rather than biomarkers which rely on blood tests, the application use could be expanded to use by the public.

This application has some limitations: while it is possible to deploy the application to the web, it relies on the pre-fit joint model, meaning the model needs to be correctly specified and validated for the tool to be of use. It also means the tool has the same limitations as the original joint model, detailed in Chapter 4. These limitations include the use of screening values as baseline values when fitting the joint model, as well as the assumptions made around the random effects, and how the joint model bases its predictions on these effects which are unobserved.

While the application is user-friendly the clinician would need to know how to interpret the results and effectively convey them to the patient; while the graph is simple, an extension of this web application would be a user guide and possibly a better legend for the graph, however this is out of the scope of this PhD.

Furthermore, this application would need to be tested thoroughly prior to release for use in clinical practice. This may entail alpha testing by clinical practitioners to ensure it is fit for purpose, and any feedback received would be used to improve the application. Then the application could be beta tested within a clinician / patient environment and once again feedback would be used to improve the application.

Overall, the web application performs as it should, adding a powerful tool to translate prognostic joint models into clinical practice. It would also be simple to update or change the joint model so that the application could be improved or adapted for other uses.

Chapter 8 Discussion

8.1 Foreword

The purpose of this chapter is to discuss the work of this thesis and provide a conclusion. While many of the previous chapters have provided their own discussions, this chapter will link these discussions together whilst also answering the research questions ‘how can joint modelling improve on the methodology of the current prognostic models within heart failure’ and ‘can prognostic models fitted with joint modelling outperform current standard prognostic models within heart failure?’.

8.2 Overview of chapters

8.2.1 Chapter 2 Joint Modelling of longitudinal processes and time-to-event outcomes in heart failure: systematic review and exemplar examining the relationship between serum digoxin levels and mortality.

Chapter 2 Introduces how joint modelling has been applied within heart failure, through means of a systematic review, paired with and exemplar. It provides the reader with both the current uses of joint modelling and introduces how joint modelling can be applied to data from patients with heart failure. This chapter helps to provide background and helps to answer the aim to explore the use of joint modelling in heart failure. This chapter highlighted the gaps around the application of joint models in heart failure, with specific regards to the lack of prognostic models within heart failure which use joint modelling and the number of studies which used data from RCTs. The exemplar illustrates how joint modelling can improve on alternative models such as Cox PH models using LOCF and Extended Cox PH with respect to fit (as measured by a discrimination index).

8.2.2 Chapter 3 Data Source

The purpose of Chapter 3 is to introduce the data sources used in the remaining joint models of this thesis. The data, obtained from two RCTs (PARADIGM and ATMOSPHERE) is summarised, and used in the later chapters in the fitting of joint models with the purpose of exploring how these joint models can improve on the existing current standard of prognostic models.

8.2.3 Chapter 4 Joint Modelling of NT-ProBNP and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure

Chapter 4 introduces the first joint models that were fit with the intention of evaluating the models as prognostic models within this thesis. Whilst also comparing them to alternative more traditional models including the current standard (Cox PH models). The models used data from the PARADIGM RCT. Using the common and well documented biomarker NT-ProBNP and a composite endpoint of negative outcomes in heart failure, death from cardiovascular causes and first hospitalisation for heart failure. Three joint models where fit, using different parameterisations of log NT-ProBNP in the survival component. These models were then compared against the current standard of prognostic model the LOCF Cox PH model and alternative to joint models the extended Cox PH model. Finally, the models were validated using data from the ATMOSPHERE clinical trial to see how well the models performed with external data.

The joint models suggest that the value and slope parameterisation of log NT-ProBNP performed the best with regards to subject specific predictions according to the available summary measures. While the joint models missed the threshold for being considered ‘acceptable’ with respect to prognostic performance at both 12 and 24 months, they appear to have outperformed the Cox PH models at 12 months. Validation of the joint models with

an external data set shows that the models performed better at month 24 than with the original data. Whilst also meeting the threshold for ‘acceptable performance’ at this time point.

Overall, this chapter highlights how joint models can be applied as prognostic models in heart failure using a well-established biomarker and common adverse outcomes. Whilst also illustrating how joint models compare to the current standard for prognostic models, and an alternative to joint modelling.

8.2.4 Chapter 5 Joint Modelling of eGFR and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure

In Chapter 5 a further three joint models were fit, once again using data from the PARADIGM clinical trial, and using an alternative biomarker eGFR. A surrogate for renal function to which there is a documented link between renal decline and adverse outcomes in heart failure. The same endpoint, a composite of cardiovascular death and first hospitalisation for heart failure was used, again as it is indicative of adverse outcomes in heart failure. Each of the joint models had a different representation of log eGFR in the survival outcome. Like Chapter 4 the joint models were compared against both the current standard of prognostic model the LOCF Cox PH model and an alternative to joint modelling the Extended Cox PH model. The models were also validated with external data from the ATMOSPHERE RCT.

The joint models suggest an association of log eGFR and the composite endpoint of death from cardiovascular causes and first hospitalisation for heart failure. The comparison of the joint models with different representations of log eGFR in the survival component, suggests the model which included both the value and slope of log eGFR performed best

with regards to both marginal and conditional fit based on the available summary measures. However, there appeared to be issues surrounding the model which included both the value and slope of log eGFR with the slope parameter being estimated at <0.001. These issues may indicate that the slope parameter may either be too small to estimate or may have been incorrectly estimated, in which case this may affect the validity of the predictions of the model.

The comparison of the joint models against the Cox PH models, suggest that the joint models are more like the LOCF Cox PH model than the Extended Cox PH models, and may suggest that due to the large number of repeat measurements, the Extended Cox PH model may have had issues estimating the other covariates.

With regards to prognostic performance, the joint model containing the value and slope parameterisation of log eGFR performed better than the other joint models at 12 months, meeting the acceptable threshold of 0.70 with regards to the time-varying AUC. However, all joint models performed equally at 24 months missing the acceptable threshold for the time-varying AUC. The comparison against the LOCF Cox PH model and the Extended Cox PH model suggest that the joint models outperform these models at both time points. The external validation with the ATMOSPHERE data, suggest that the joint models performed worse at both 12 and 24 months with regards to time-varying AUC.

When compared to the models from Chapter 4, the time-varying AUCs suggest that Joint Model 2 (Value + Slope of eGFR) outperforms all the NT-ProBNP joint models from Chapter 4 at month 12, however all the NT-ProBNP models marginally outperform the eGFR models at month 24. Suggesting that at 12 months the value and slope of eGFR provides a better model in terms of prognostic performance, whereas any parameterisation of NT-ProBNP performs better at 24 months.

Like Chapter 4 this chapter adds further to the evidence base, illustrating how another established biomarkers in heart failure can be jointly modelled with common adverse outcomes to produce prognostic models and how these models compare to the current established standard of prognostic model and an alternative to joint modelling.

8.2.5 Chapter 6 Multivariate Joint Modelling of NT-ProBNP and eGFR, and a Composite Endpoint of Death from Cardiovascular Causes and First Hospitalisation for Heart Failure

Chapter 6 builds on both Chapter 4 and Chapter 5 by fitting a single multivariate joint model containing two longitudinal outcomes from the PARADIGM RCT. One of log NT-ProBNP and one of log eGFR. Modelled with the same composite end point of death from cardiovascular causes and hospitalisation for first heart failure, this composite endpoint representing adverse outcomes in heart failure. This multivariate joint model was fit using the best performing parameterisations for NT-ProBNP and eGFR as defined by the best subject specific fit in Chapter 4 and Chapter 5 which was both the value and slope of log NT-ProBNP and the value of eGFR. Like the aforementioned chapters, the multivariate model was compared against the current best practice for prognostic models a LOCF Cox PH model as well as an alternative model to joint models, the Extended Cox PH model. The model was validated, again using data from the ATMOSPHERE RCT.

The multivariate joint model suggested an association between the both log NT-ProBNP and eGFR and the composite endpoint. Further strengthening the evidence base. Along with this association, like Chapter 4 the multivariate joint model suggested that treatment effect was being mediated through NT-ProBNP as the treatment effect in the survival outcome did not meet the significance threshold.

Similar to Chapter 4 the slope parameter for NT-ProBNP was also below the significance threshold however, this was included as it had been demonstrated to improve the subject specific fit.

Unlike Chapter 5 the slope parameter of eGFR is measurable at two decimal places, however, has wide confidence intervals and is considered statistically insignificant. Likely due to the limited number of repeated measurements of eGFR as a result of the requirement of both biomarkers to be present at each time point.

The average predicted longitudinal profiles of both biomarkers were similar to Chapter 4 and Chapter 5, but both had wider confidence intervals, possibly to the amount of data lost as a result of the limitation to only allow the biomarkers to be at the same time points.

The time-varying AUCs are similar to those of Chapter 4, missing the threshold for acceptable performance at 12 and 24 months, with a drop of 0.2 from 12 to 24 months.

External validation showed an increase in time-varying AUC of 0.2 at month 12 this exceeding the threshold for acceptable performance, but with a higher Brier score at month 24 than month 12 suggesting a worse performance when calibration is accounted for.

The HRs from the comparative analysis suggest that the multivariate joint model is more similar to the LOCF Cox PH model, with the HRs being closer in value to this model. The extended Cox PH model also suggests there is not significant association between log eGFR and the composite endpoint which contradicts both the multivariate joint model and the LOCF model.

While the multivariate model shows an improvement in time-varying AUC over the models in Chapter 5 at month 24, they perform equally to the models in Chapter 4, suggesting that modelling with both eGFR and NT-ProBNP outcomes are an improvement on eGFR alone. However, given the issue that both biomarkers must be collected simultaneously, the multivariate model may be best suited when the biomarkers are routinely collected together, in clinical practice.

Overall, this chapter illustrates how a multivariate joint model containing two established biomarkers in heart failure can be modelled jointly with common adverse outcomes to produce a prognostic model. As well as evaluating how that joint model performs compared to the current standard and an alternative to joint models.

8.2.6 Chapter 7 Real-World Application of Prognostic Joint Models Illustrated Using a Web Application

In Chapter 7, an interactive web application was developed to illustrate how a prognostic joint model could be used in real world setting to provide prognosis based on baseline characteristic and repeated measurements of a biomarker.
This chapter seeks to demonstrate what a web application for prognosis using joint modelling is capable of and how an end user such as a clinician would interact with it.

Using the R Shiny framework an interactive web application was developed, allowing for dynamic predictions which are updated on input by the user. The web application is both reactive in that it updates automatically without the need for further interaction such as an update button, while also providing a user-friendly interface.

Overall, this chapter demonstrates how a joint model could be used in practice to provide prognostic information using both baseline characteristics and repeated measurements through use of an interactive web application.

8.3 Strengths and limitations of this research

This research has both strengths and limitations. This research has identified that joint modelling is currently being applied to data from heart failure studies, using both primary and secondary data. Although only three studies were identified to be using data from randomised control trials, as stated possibly due to the standard of reporting around joint models, identifying the need for better standards for the reporting of this type of models.

This research is not only novel in nature, in that it is one of the first pieces of research to explore the use of joint models as prognostic models, but also uses external validation to strengthen the evaluation of the joint models' performance.

Overall, the joint models presented in this thesis were shown to have outperformed both the current standard of prognostic models and alternative models using time-varying AUCs. However, this should be interpreted with caution as while it is possible to obtain a time-varying AUC from both joint models and Cox PH model, using the same underlying methodology, joint models have other characteristics such as a joint distribution and random effects. Both of which need to be taken into consideration when interpreting the results of the time-varying AUCs obtained from the different types of models, as this may be a case of comparing apples to oranges.

Existing prognostic models such as 'Predict HF' may appear to outperform the joint models presented in this thesis, but as the existing prognostic models only report AUC-

ROC and not time-varying AUCs and due to the previously mentioned limitations, the performance of the models cannot and should not be compared. This research was limited to the available metrics for joint models as presented in this thesis such as the time-varying AUC and time-varying Brier Score. While the metrics are appropriate for joint models and this limitation may be due to the underlying methodology of joint models, it has identified a need for further research into both the existing metrics and into new more generalisable metrics especially for comparison with more traditional models. For example, the dynamic discrimination index used in Chapter 2 provides an overall summary of discriminatory performance of a joint model which could be compared to a C-Statistic of a traditional model, however this has been removed in JMBayes2 with no alternative available [117].

While joint prognostic models were fit and evaluated, due to the recommendation that calibration is assessed alongside prognostic performance, an early release of the R package *JMBayes2* was used, and while the HRs from the joint models were similar to those from the Cox PH models, this limitation must be acknowledged.

Although, in the data sets used in this thesis, the estimates of the HRs of the Cox PH Models and the LOCF Cox models were similar to those from the joint models, the former models are limited to a singular measurement of the biomarker of interest. This is typically the last observed measurement of the biomarker (LOCF), assuming that that the measurements from patients whose baseline measurements were used are similar to those whose measurements were captured at a later time. In other situations, this could lead to greater bias and underestimation of the biomarker parameter in the model. On the other hand, joint modelling uses a linear mixed effect model to estimate the true value of the biomarker over time, which should overcome this limitation and therefore may make joint models superior, especially when there are fewer later measurements of the biomarker of interest.

This research also highlighted that while alternative models such as the extended Cox PH model allow for repeat measurements, the special properties of the time-varying measure such as the correlation between measurements and measurement error need to be accounted for. While this has been previously established, this research further identifies a potential issue with extended Cox PH models, in that when a large number or repeated measurements are added the results from the extended Cox PH model appeared to have exhibited issues in its estimation of parameters, illustrating significant differences compared to both the joint models and the LOCF Cox PH model. However, as this is based on a limited number of models, further research is needed to validate this claim.

From a clinical perspective, the dynamic and subject specific prediction capabilities in JMs could allow for more personalised predictions. With the illustration of a web-based application, this also allows these dynamic and subject specific predictions to be used in a real world setting such a clinician / patient consultation. However, once again these subject specific predictions are based on the latent random effects which needs to be considered when interpreting their results. The application also needs to be tested and evaluated by clinicians. Therefore, further research is required prior to this application being used in clinical practice.

There are also limitations surrounding the original data from the PARADIGM HF trial. While at least one measurement of NT-ProBNP was collected for all participants. Due to joint modelling requiring a baseline, and a run-in period during the trial, only values at screening were available for many patients, and while in the interest of maximising use of data these patients were included in both Chapter 4 and Chapter 6 this may have introduced bias in the analysis. Along with bias, the limited number of repeat measurements of NT-ProBNP may have not been the best example for illustrating the benefits of joint modelling, and more repeat measures may have improved model fit.

While the composite endpoint of the PARADIGM HF trial used in this research encapsulates common adverse outcomes, this research may have benefited from the use of a singular outcome and the comparison of the individual outcomes of the composite outcome. This could be considered for further research.

As well as limitations around the PARADIGM HF trial, there were limitations using the ATMOSPHERE HF trial for validation. While the trials are similar and contain the same covariates, there are differences in treatment effects between the active arms of the two trials. The recoding of these arms to fit the model parameters for validation could have led to bias, and underestimation of the treatment effect. A solution to this may be to use the equations from the original PARADIGM HF model to apply predictions to patients in the control arm(s) of the ATMOSPHERE HF data and then adjust these for the observed treatment effects on these patients.

8.4 Future work

As previously stated, there are a number of areas that this research can be improved and developed on. Future work should include further models evaluating the individual endpoints of the primary data and then the comparison of these models. Alternative models could evaluate competing risks, which are events ‘competing to be first’ that when occur prevent or alter the risk of other events from happening such as cardiovascular death and non-cardiovascular death [157]. With alternative data, models for reoccurring events could also be possible, such as hospitalisation for heart failure to evaluate their feasibility and usefulness in joint prognostic models. The use of different primary data could also be used to extend the population beyond that of patients with HFrEF, and evaluate the issues previously mentioned around the slope parameterisations. It would also be beneficial to use

registry data as further external validation to fully evaluate the generalisability of the models. A simulation study may be necessary to fully evaluate the aforementioned issues around the parameter estimates from the extended Cox PH models and the comparison of the Joint Models with the Cox PH models. Such a study could be used to compare the two types of Cox PH models against joint models using simulated data and evaluate which model had the least bias and greatest accuracy. Finally, further development of the web application, including adding documentation and gaining user feedback to ensure the application is fully fit for purpose.

8.5 Conclusion

This research sought to answer the following questions:

‘Can joint modelling enhance the methodological toolkit and have utility in the development of prognostic models within heart failure?’.

‘Can prognostic models fitted with joint modelling outperform current standard prognostic models within heart failure?’

With these questions in mind, I conclude that joint modelling can be used for prognostic models within heart failure, however at this time it is not possible to say with certainty whether in terms of model performance they are an improvement over the current gold standards of prognostic modelling. Therefore, further development and research of joint models is required, alongside research into the comparison of these models with other prognostic modelling approaches.

Appendices

All R Code for the models and R Shiny application is available at:

https://github.com/RyanJField/PhD-Thesis-Code

Appendix Table 1 Variables and Associated Hazard Ratios (95% CI) from the Three Cox PH Models adapted from the PREDICT-HF Study

Variable	Composite Model	Cardiovascular Death Model	All-cause Mortality model
Baseline			
Age (per 10 y >60 y)		1.15 (1.06-1.25)	1.20 (1.11-1.30)
BMI (for every kg/m2 decrease less than 30 kg/m2)			1.02 (1.00-1.04)
Bundle branch block	1.16 (1.05-1.28)	1.20 (1.06-1.35)	
Diabetes mellitus	1.36 (1.24-1.50)	1.26 (1.12-1.42)	1.26 (1.12-1.41)
Ejection fraction (per 5% decrease <40%)	1.09 (1.05-1.13)	1.10 (1.05-1.15)	1.08 (1.03-1.12)
HF duration 1-5 y	1.37 (1.21-1.54)	1.21 (1.04-1.40)	1.25 (1.09-1.44)
HF duration >5 y	1.50 (1.31-1.70)	1.36 (1.16-1.60)	1.35 (1.16-1.56)
Male	1.22 (1.08-1.38)	1.37 (1.16-1.60)	1.34 (1.16-1.55)
No previous PCI		1.21 (1.03-1.41)	1.29 (1.12-1.50)
Not prescribed β-blocker	1.34 (1.14-1.58)	1.36 (1.12-1.65)	1.27 (1.06-1.53)
Not prescribed sacubitril/valsartan	1.24 (1.13-1.36)	1.25 (1.11-1.40)	1.18 (1.07-1.31)
NYHA III/IV	1.24 (1.11-1.38)	1.32 (1.16-1.51)	1.30 (1.15-1.47)
Peripheral arterial disease	1.28 (1.11-1.47)	1.41 (1.18-1.68)	1.36 (1.16-1.60)
Prior HF hospitalization	1.35 (1.21-1.51)		
Prior MI	1.12 (1.02-1.23)	1.23 (1.08-1.39)	1.15 (1.03-1.29)
Race/ethnicity			
Asian	1.43 (1.23-1.66)	1.89 (1.57-2.30)	1.42 (1.18-1.71)
Black	1.67 (1.36-2.04)	1.34 (1.03-1.78)	
Region			
Central Europe	1.34 (1.18-1.52)	1.63 (1.37-1.92)	1.42 (1.23-1.65)
Latin America	1.50 (1.21-1.85)	1.84 (1.53-2.21)	1.69 (1.43-2.00)
Systolic BP (per 10 mm Hg decrease <120 mm Hg)		1.10 (1.01-1.19)	1.09 (1.01-1.17)

Valvular heart disease	1.33 (1.14-1.55)		
Interaction terms			
Prior hospitalization for HF from Latin America	1.41 (0.78-2.50)		
Laboratory Values			
% Monocytes for every % greater than 7%			1.04 (1.02-1.07)
Absolute lymphocytes (for every 1000/mL decrease less than 2500/mL)	1.06 (1.02-1.09)		
Absolute neutrophils (for every 1000/mL increase below 6000/mL)	1.07 (1.03-1.11)		1.10 (1.05-1.15)
Albumin (for every 0.1 g/dL decrease less than 4.2 g/dL)	1.05 (1.03-1.07)	1.05 (1.02-1.08)	1.06 (1.03-1.08)
AST (For every 598.8 U/L increase greater than 1796.4 U/L)			1.07 (1.01-1.13)
Chloride (for every 1 mEq/L decrease 100 mEq/L)			1.06 (1.03-1.09)
Haemoglobin (for every 1 g/dL decrease less than 14 g/dL)	1.08 (1.03-1.13)	1.10 (1.04-1.17)	1.11 (1.05-1.17)
LDL (for every 38.61 mg/dL increase greater than 115.83 mg/dL)	1.15 (1.04-1.26)		1.19 (1.07-1.33)
Potassium (for every 0.1 mEq/L decrease less than 4 mEq/L)	1.07 (1.03-1.10)	1.09 (1.05-1.13)	1.05 (1.02-1.10)
Total bilirubin (for every 0.29 mg/dL increase greater than 0.58 mg/dL)	1.11 (1.08-1.15)	1.10 (1.06-1.14)	1.08 (1.04-1.13)
Total cholesterol (for every 38.61 mg/dL increase greater than 115.83 mg/dL)		1.09 (1.03-1.16)	
Triglycerides (for every 88.5 mg/dL decrease less than 221.24 mg/dL)			1.12 (1.01-1.24)
Urea (for every 2.8 mg/dL increase greater than 14.01 mg/dL)	1.02 (1.00-1.04)	1.03 (1.01-1.05)	1.02 (1.00-1.04)
Uric acid (for every 0.84 mg/dL increase greater than 6.72 mg/dL)	1.08 (1.05-1.11)	1.07 (1.03-1.11)	1.07 (1.04-1.11)
Natriuretic peptides			
NTproBNP category	1.34 (1.28-1.40)	1.40 (1.32-1.48)	1.33 (1.24-1.38)

[42]

Appendix Table 2 Hazard Ratios (95% CI) from the Seattle Prognostic Model Adapted from Levy et al.

Variable	Hazard Ratio (95% CI)
Demographics	
Age (decade)*	1.090 (0.985–1.205)
Gender (male)	1.089 (0.839–1.414)
NYHA class	1.600 (1.019–2.511)
100/Ejection fraction	1.030 (1.010–1.050)
Ischemic etiology	1.354 (1.074–1.707)
SBP, 10 mm Hg* (for SBP <160 mm Hg)	0.877 (0.823–0.935)
Medications	
Diuretic dose, mg/kg per day	1.178 (1.097–1.266)
Allopurinol use	1.571 (1.170–2.109)
Statin use	0.63 (0.410–0.978)
Laboratory	
If sodium <138, 138-sodium	1.050 (1.005–1.097)
100/Cholesterol,*dL/mg	2.206 (1.045–4.656)
If haemoglobin <16, 16-haemoglobin	1.124 (1.053–1.200)
If haemoglobin >16, haemoglobin-16	1.336 (1.010–1.767)
% Lymphocytes,* each 5% (for lymphocytes <47%)*	0.897 (0.846–0.951)
Uric acid, mg/dL (for uric acid >3.4)*	1.064 (1.022–1.108)
* Continuous Variable	

[61]

Appendix Table 3 Additional Hazard Ratios from the Seattle Prognostic Model using Hazard Ratios estimated from Meta-analysis or Clinical Trials Adapted from Levy et al.

Variable	Hazard Ratio: When Medication / Device Added	Hazard Ratio: When Patient on Medication / Device
Medications		
ACE inhibitor	0.77	0.77
13-Blocker	0.66	0.66
Angiotensin receptor blocker	0.87	0.85
K-sparing diuretic	0.70	0.74
Statin	0.78	0.63
Devices		
Biventricular pacemaker	0.74	1.00
Implantable cardioverter-defibrillator	0.74	0.73
Biventricular implantable cardioverter-defibrillator	0.64	0.79
Left ventricular assist device	0.52	N/A
Hazard Ratios estimated from meta-analysis or clinical trials, NA: Data not available		

[46]

Appendix Table 4 Rate Ratios (95% CI) from the MAGGIC Model Stratified by Ejection Fraction Adapted from Pocock et al.

	Ejection Fraction < 40%	Ejection Fraction ≥ 40%
Variable	**Rate Ratio (95% CI)**	**Rate Ratio (95% CI)**
Age (per 10 years)	1.407 (1.375 - 1.439)	1.589 (1.536 - 1.643)
Male	1.101 (1.044 - 1.161)	1.113 (1.053 - 1.177)
BMI (per 1 kg/m2 increase up to 30 kg/m2)	0.970 (0.961 - 0.978)	0.960 (0.951- 0.969)
Current smoker	1.154 (1.091 - 1.222)	1.174 (1.095 - 1.258)
SBP (per 10 mmHg increase)	0.936 (0.924 - 0.948)	0.982 (0.968 - 0.998)
Diabetes	1.421 (1.347 - 1.499)	1.401 (1.311 - 1.498)
NYHA Class		
I	0.828 (0.744 - 0.922)	0.756 (0.682 - 0.838)
II	1.000	1.000
III	1.372 (1.303 - 1.445)	1.458 (1.361 - 1.561)
IV	1.640 (1.503 - 1.790)	1.756 (1.599 - 1.928)
Ejection fraction (Per 5% increase)	0.915 (0.902 - 0.928)	
COPD	1.191 (1.096 - 1.295)	1.284 (1.181 - 1.396)
HF duration > 18 months	1.191 (1.127 - 1.259)	1.166 (1.088 - 1.250)
Creatinine (Per 10 µmol/L up to 350 µmol/L)	1.041 (1.035 - 1.046)	1.035 (1.029 - 1.041)
Beta-blocker	0.736 (0.694 - 0.781)	0.798 (0.746 - 0.855)
ACE-I/ARB	0.834 (0.770 - 0.905)	0.938 (0.842 - 1.044)

BMI: body mass index, SBP: systolic blood pressure, NYHA: New York Heart Association, COPD: chronic obstructive pulmonary disease, HF: heart failure, ACE-I: angiotensin-converting enzyme inhibitor, ARB: angiotensin-receptor blockers.

[16]

List of References

[1] T. A. McDonagh *et al.*, “2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failureDeveloped by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC,” *Eur Heart J*, vol. 42, no. 36, pp. 3599–3726, Sep. 2021, doi: 10.1093/EURHEARTJ/EHAB368.

[2] P. A. Heidenreich *et al.*, “2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines,” *Circulation*, Apr. 2022, doi: 10.1161/CIR.0000000000001063.

[3] The Criteria Committee of the New York Heart Association, “Functional Capacity and Objective Assesment,” in *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*, 9th ed., M. Dolgin, Ed., Little, Brown and Co., 1994, pp. 253–256.

[4] C. Caraballo *et al.*, “Clinical Implications of the New York Heart Association Classification,” *J Am Heart Assoc*, vol. 8, no. 23, Dec. 2019, doi: 10.1161/JAHA.119.014240.

[5] E. H. SONNENBLICK, “Correlation of Myocardial Ultrastructure and Function,” vol. 38, no. 1, pp. 29–44, 1968, doi: doi:10.1161/01.CIR.38.1.29.

[6] R. B. Rehr, C. R. Malloy, N. G. Filipchuk, and R. M. Peshock, “Left ventricular volumes measured by MR imaging,” *Radiology*, vol. 156, no. 3, pp. 717–719, 1985, doi: 10.1148/RADIOLOGY.156.3.4023232.

[7] S. H. Orakzai, R. H. Orakzai, K. Nasir, and M. J. Budoff, “Assessment of Cardiac Function Using Multidetector Row Computed Tomography,” *J Comput Assist Tomogr*, vol. 30, no. 4, 2006, [Online]. Available: https://journals.lww.com/jcat/Fulltext/2006/07000/Assessment_of_Cardiac_Function_Using_Multidetector.1.aspx

[8] J. R. Corbett *et al.*, “Equilibrium radionuclide angiocardiography,” 2006, doi: 10.1016/j.nuclcard.2006.08.007.

[9] S. L. James *et al.*, “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*, vol. 392, no. 10159, pp. 1789–1858, Nov. 2018, doi: 10.1016/S0140-6736(18)32279-7.

[10] J. Motiejūnaitė *et al.*, “The association of long-term outcome and biological sex in patients with acute heart failure from different geographic regions,” *Eur Heart J*, vol. 41, no. 13, pp. 1357–1364, Apr. 2020, doi: 10.1093/EURHEARTJ/EHAA071.

[11] C. A. Lawson *et al.*, “Comorbidity health pathways in heart failure patients: A sequences-of-regressions analysis using cross-sectional data from 10,575 patients in the Swedish Heart Failure Registry,” *PLoS Med*, vol. 15, no. 3, Mar. 2018, doi: 10.1371/JOURNAL.PMED.1002540.

[12] M. S. Khan *et al.*, “Trends in prevalence of comorbidities in heart failure clinical trials,” *Eur J Heart Fail*, vol. 22, no. 6, pp. 1032–1042, Jun. 2020, doi: 10.1002/EJHF.1818.

[13] T. Sudoh, K. Kangawa, N. Minamino, and H. Matsuo, “A new natriuretic peptide in porcine brain,” *Nature 1988 332:6159*, vol. 332, no. 6159, pp. 78–81, 1988, doi: 10.1038/332078a0.

[14] Y. Ogawa *et al.*, “Natriuretic peptides as cardiac hormones in normotensive and spontaneously hypertensive rats: The ventricle is a major site of synthesis

and secretion of brain natriuretic peptide," *Circ Res*, vol. 69, no. 2, pp. 491–500, 1991, doi: 10.1161/01.RES.69.2.491.

[15] Z. Abassi, T. Karram, S. Ellaham, J. Winaver, and A. Hoffman, "Implications of the natriuretic peptide system in the pathogenesis of heart failure: diagnostic and therapeutic importance," *Pharmacol Ther*, vol. 102, no. 3, pp. 223–241, Jun. 2004, doi: 10.1016/j.pharmthera.2004.04.004.

[16] C. Hall, "NT-ProBNP: The Mechanism Behind the Marker," *J Card Fail*, vol. 11, no. 5, pp. S81–S83, Jun. 2005, doi: 10.1016/J.CARDFAIL.2005.04.019.

[17] J. Magga, M. Marttila, P. Mäintymaa, O. Vuolteenaho, and H. Ruskoaho, "Brain natriuretic peptide in plasma, atria, and ventricles of vasopressin- and phenylephrine-infused conscious rats," *Endocrinology*, vol. 134, no. 6, pp. 2505–2515, 1994, doi: 10.1210/ENDO.134.6.8194476.

[18] C. J. Pemberton, M. L. Johnson, T. G. Yandle, and E. A. Espiner, "Deconvolution analysis of cardiac natriuretic peptides during acute volume overload," *Hypertension*, vol. 36, no. 3, pp. 355–359, 2000, doi: 10.1161/01.HYP.36.3.355.

[19] A. Maisel, "B-type natriuretic peptide levels: Diagnostic and therapeutic potential," *Cardiovasc Toxicol*, vol. 1, no. 2, pp. 159–164, 2001, doi: 10.1385/CT:1:2:159/METRICS.

[20] M. M. Redfield, R. J. Rodeheffer, S. J. Jacobsen, D. W. Mahoney, K. R. Bailey, and J. C. Burnett, "Plasma brain natriuretic peptide to detect preclinical ventricular systolic or diastolic dysfunction: a community-based study.," *Circulation*, vol. 109, no. 25, pp. 3176–3181, 2004, doi: 10.1161/01.CIR.0000130845.38133.8F.

[21] T. J. Wang *et al.*, "Impact of Obesity on Plasma Natriuretic Peptide Levels," *Circulation*, vol. 109, no. 5, pp. 594–600, Feb. 2004, doi: 10.1161/01.CIR.0000112582.16683.EA.

[22] R. Rørth *et al.*, "Comparison of bnp and nt-probnp in patients with heart failure and reduced ejection fraction," *Circ Heart Fail*, 2020, doi: 10.1161/CIRCHEARTFAILURE.119.006541.

[23] A. Michel, M. Martín-Pérez, A. Ruigómez, and L. A. García Rodríguez, "Incidence and risk factors for severe renal impairment after first diagnosis of heart failure: A cohort and nested case–control study in UK general practice," *Int J Cardiol*, vol. 207, pp. 252–257, Mar. 2016, doi: 10.1016/J.IJCARD.2016.01.167.

[24] W. Sato *et al.*, "Impaired renal function is a major determinant of left ventricular diastolic dysfunction: assessment by stress myocardial perfusion imaging," *Ann Nucl Med*, vol. 27, no. 8, p. 729, Oct. 2013, doi: 10.1007/S12149-013-0739-Z.

[25] B. Waldum *et al.*, "Renal Function in Outpatients With Chronic Heart Failure," *J Card Fail*, vol. 16, no. 5, pp. 374–380, May 2010, doi: 10.1016/J.CARDFAIL.2010.01.001.

[26] R. De Silva *et al.*, "Incidence of renal dysfunction over 6 months in patients with chronic heart failure due to left ventricular systolic dysfunction: contributing factors and relationship to prognosis," *Eur Heart J*, vol. 27, no. 5, pp. 569–581, Mar. 2006, doi: 10.1093/EURHEARTJ/EHI696.

[27] D. E. Weiner *et al.*, "Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies," *J Am Soc Nephrol*, vol. 15, no. 5, pp. 1307–1315, May 2004, doi: 10.1097/01.ASN.0000123691.46138.E2.

[28] M. C. Foster *et al.*, "Cardiovascular risk factor burden, treatment, and control among adults with chronic kidney disease in the United States," *Am Heart J*, vol. 166, no. 1, pp. 150-156.e1, Jul. 2013, doi: 10.1016/J.AHJ.2013.03.016.

[29] A. Al-Naher, D. Wright, M. A. J. Devonald, and M. Pirmohamed, “Renal function monitoring in heart failure – what is the optimal frequency? A narrative review,” *Br J Clin Pharmacol*, vol. 84, no. 1, p. 5, Jan. 2018, doi: 10.1111/BCP.13434.

[30] S. Smarick and T. C. Hallowell, “Urine output,” *Small Animal Critical Care Medicine, Second Edition*, pp. 1001–1004, Jan. 2014, doi: 10.1016/B978-1-4557-0306-7.00192-6.

[31] A. S. Levey *et al.*, “K/DOQI clinical practice guidelines for chronic kidney disease: Evaluation, classification, and stratification,” *American Journal of Kidney Diseases*, vol. 39, no. 2 SUPPL. 1, pp. i-ii+S1-S266, 2002, doi: 10.2/JQUERY.MIN.JS.

[32] E. J. Lamb, C. R. V. Tomson, and P. J. Roderick, “Estimating kidney function in adults using formulae,” *http://dx.doi.org/10.1258/0004563054889936*, vol. 42, no. 5, pp. 321–345, Aug. 2016, doi: 10.1258/0004563054889936.

[33] U. E. N. C. for E. Assessment, “A simplified equation to predict glomerular filtration rate from serum creatinine,” Mar. 2009.

[34] R. Jin, G. L. Grunkemeier, J. R. Brown, and A. P. Furnary, “Estimated Glomerular Filtration Rate and Renal Function,” *Annals of Thoracic Surgery*, vol. 86, no. 1, pp. 1–3, Jul. 2008, doi: 10.1016/j.athoracsur.2008.05.007.

[35] R. D. Riley *et al.*, “Prognostic model research,” in *Prognosis Research in Health Care*, Oxford University Press, 2019, pp. 139–187. doi: 10.1093/MED/9780198796619.003.0008.

[36] K. Rahimi *et al.*, “Risk prediction in patients with heart failure: A systematic review and analysis,” *JACC Heart Fail*, vol. 2, no. 5, pp. 440–446, Oct. 2014, doi: 10.1016/j.jchf.2014.04.008.

[37] G. L. Di Tanna, H. Wirtz, K. L. Burrows, and G. Globe, “Evaluating risk prediction models for adults with heart failure: A systematic literature review,” *PLoS One*, vol. 15, no. 1, p. e0224135, Jan. 2020, doi: 10.1371/JOURNAL.PONE.0224135.

[38] D. Hosmer, S. Lemeshow, and R. Sturdivant, “Applied Logistic Regression,” in *Applied Logistic Regression*, Third., Wiley, 2013, pp. 173–182.

[39] R. F. Wolff *et al.*, “PROBAST: A tool to assess the risk of bias and applicability of prediction model studies,” *Ann Intern Med*, vol. 170, no. 1, pp. 51–58, Jan. 2019, doi: 10.7326/M18-1376/SUPPL_FILE/M18-1376_SUPPLEMENT.PDF.

[40] K. G. M. Moons *et al.*, “Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration,” *https://doi.org/10.7326/M14-0698*, vol. 162, no. 1, pp. W1–W73, Jan. 2015, doi: 10.7326/M14-0698.

[41] D. G. Altman, Y. Vergouwe, P. Royston, and K. G. M. Moons, “Prognosis and prognostic research: validating a prognostic model,” *BMJ*, vol. 338, no. 7708, pp. 1432–1435, May 2009, doi: 10.1136/BMJ.B605.

[42] J. Simpson *et al.*, “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 Supplemental content,” *JAMA Cardiol*, vol. 5, no. 4, pp. 432–441, 2020, doi: 10.1001/jamacardio.2019.5850.

[43] J. J. V Mcmurray *et al.*, “Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure,” *n engl j med*, vol. 11, pp. 993–1004, 2014, doi: 10.1056/NEJMoa1409077.

[44] J. J. V. McMurray *et al.*, “Aliskiren, Enalapril, or Aliskiren and Enalapril in Heart Failure,” *New England Journal of Medicine*, vol. 374, no. 16, pp.

1521–1532, Apr. 2016, doi: 10.1056/NEJMOA1514859/SUPPL_FILE/NEJMOA1514859_DISCLOSURES.PDF.

[45] G. Savarese, P. Vasko, Å. Jonsson, M. Edner, U. Dahlström, and L. H. Lund, "The Swedish Heart Failure Registry: a living, ongoing quality assurance and research in heart failure," *Ups J Med Sci*, vol. 124, no. 1, p. 65, Jan. 2019, doi: 10.1080/03009734.2018.1490831.

[46] W. C. Levy *et al.*, "The Seattle Heart Failure Model," *Circulation*, vol. 113, no. 11, pp. 1424–1433, Mar. 2006, doi: 10.1161/CIRCULATIONAHA.105.584102.

[47] J. D. Rich, J. Burns, B. H. Freed, M. S. Maurer, D. Burkhoff, and S. J. Shah, "Meta-analysis global group in chronic (MAGGIC) heart failure risk score: Validation of a simple tool for the prediction of morbidity and mortality in heart failure with preserved ejection fraction," *J Am Heart Assoc*, vol. 7, no. 20, Oct. 2018, doi: 10.1161/JAHA.118.009594.

[48] C. W. Yancy *et al.*, "2013 ACCF/AHA guideline for the management of heart failure: A report of the American college of cardiology foundation/american heart association task force on practice guidelines," *J Am Coll Cardiol*, vol. 62, no. 16, pp. e147–e239, Oct. 2013, doi: 10.1016/j.jacc.2013.05.019.

[49] P. Ponikowski *et al.*, "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of," *Eur Heart J*, vol. 37, no. 27, pp. 2129–2200, 2016, doi: 10.1093/eurheartj/ehw128 %J European Heart Journal.

[50] S. J. Pocock *et al.*, "Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies," *Eur Heart J*, vol. 34, no. 19, pp. 1404–1413, May 2013, doi: 10.1093/EURHEARTJ/EHS337.

[51] J. G. Ibrahim, H. Chu, and L. M. Chen, "Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data," *Journal of Clinical Oncology*, vol. 28, no. 16, pp. 2796–2801, Jun. 2010, doi: 10.1200/JCO.2009.25.0654.

[52] D. Rizopoulos, *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. 2012.

[53] G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona, "Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues," *BMC Med Res Methodol*, vol. 16, no. 1, p. 117, 2016, doi: 10.1186/s12874-016-0212-5.

[54] V. De Gruttola and X. M. Tu, "Modelling progression of CD4-lymphocyte count and its relationship to survival time," *Biometrics*, vol. 50, no. 4, pp. 1003–1014, 1994.

[55] A. A. Tsiatis, V. Degruttola, and M. S. Wulfsohn, "Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS," *J Am Stat Assoc*, vol. 90, no. 429, pp. 27–37, 1995, doi: 10.1080/01621459.1995.10476485.

[56] A. A. Tsiatis and M. %J S. S. Davidian, "Joint modeling of longitudinal and time-to-event data: an overview," *Stat Sin*, vol. 14, pp. 809–834, 2004.

[57] D. Collett, *Modelling survival data in medical research, third edition*. CRC Press, 2015.

[58] J. E. Dancey *et al.*, "Guidelines for the Development and Incorporation of Biomarker Studies in Early Clinical Trials of Novel Agents On behalf of the

Biomarkers Task Force of the NCI Investigational Drug Steering Committee", doi: 10.1158/1078-0432.CCR-09-2167.

[59] WHO Task Group on Biomarkers and Risk Assessment: Concepts and Principles., A. Robinson, United Nations Environment Programme., International Labour Organisation., World Health Organization., and International Program on Chemical Safety., "Biomarkers and risk assessment : concepts and principles.," p. 82, 1993.

[60] Step, "ICH guideline E16 on genomic biomarkers related to drug response: context, structure and format of qualification submissions," 2010, Accessed: Jun. 16, 2022. [Online]. Available: www.ema.europa.eu

[61] J. M. Lachin, "Fallacies of Last Observation Carried Forward," *Clin Trials*, vol. 13, no. 2, p. 161, Apr. 2016, doi: 10.1177/1740774515602688.

[62] G. H. Fisher, "Endogenous and Exogenous Investment in Macro-Economic Models," *Rev Econ Stat*, vol. 35, no. 3, p. 211, Aug. 1953, doi: 10.2307/1925919.

[63] N. M. Laird and J. H. Ware, "Random-Effects Models for Longitudinal Data," *Biometrics*, vol. 38, no. 4, p. 963, Dec. 1982, doi: 10.2307/2529876.

[64] D. A. Harville, "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *J Am Stat Assoc*, vol. 72, no. 358, p. 320, Jun. 1977, doi: 10.2307/2286796.

[65] G. Molenberghs and G. Verbeke, *Linear Mixed Models for Longitudinal Data*. in Springer Series in Statistics. New York, NY: Springer New York, 2000. doi: 10.1007/978-1-4419-0300-6.

[66] B. T. West, K. B. Welch, and A. T. Gałecki, *LINEAR MIXED MODELS - A Practical Guide Using Statistical Software*, Second. CRC Press, 2014.

[67] D. Rizopoulos, "JM: An R package for the joint modelling of longitudinal and time-to-event data," *J Stat Softw*, vol. 35, no. 9, pp. 1–33, Jul. 2010, doi: 10.18637/jss.v035.i09.

[68] D. Rizopoulos, "The R package jmbayes for fitting joint models for longitudinal and time-to-event data using MCMC," *J Stat Softw*, vol. 72, no. 1, pp. 1–46, Jan. 2016, doi: 10.18637/jss.v072.i07.

[69] D. Rizopoulos, "Joint Modeling of Longitudinal and Time-to-Event Data with Applications in R," Rotterdam (Online): Erasmus University, Aug. 2021.

[70] G. Papageorgiou, K. Mauff, A. Tomer, and D. Rizopoulos, "An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes," *Annu Rev Stat Appl*, vol. 6, no. 1, pp. 223–240, 2019, doi: 10.1146/annurev-statistics-030718-105048.

[71] "Analysis of Longitudinal Data - Peter Diggle, Department of Mathematics and Statistics Peter J Diggle, Peter J.. Diggle, Patrick Heagerty, Kung-Yee Liang, Patrick J. Heagerty, Scott Zeger, Both at Biostatistics Department Scott Zeger - Google Books." Accessed: Nov. 19, 2020. [Online]. Available: https://books.google.co.uk/books?hl=en&lr=&id=JCwSDAAAQBAJ&oi=fnd&pg=PP1&ots=jV_8ZKExJK&sig=zzSCaZg3xgFzF8twHBhHsmdNQIg&redir_esc=y#v=onepage&q&f=false

[72] J. K. Rogers, S. J. Pocock, J. J. V McMurray, and ..., "Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved," *... of heart failure*, 2014, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.29

[73] S. S. Rathore, J. P. Curtis, Y. Wang, M. R. Bristow, and H. M. Krumholz, "Association of serum digoxin concentration and outcomes in patients with heart failure," *JAMA*, vol. 289, no. 7, pp. 871–878, Feb. 2003, doi: 10.1001/JAMA.289.7.871.

[74] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, doi: 10.1136/bmj.n71.

[75] R. Garg, R. Gorlin, T. Smith, S. Yusuf, and The Digitalis Investigation Group, “The Effect of Digoxin on Mortality and Morbidity in Patients with Heart Failure,” *New England Journal of Medicine*, vol. 336, no. 8, pp. 525–533, Feb. 1997, doi: 10.1056/nejm199702203360801.

[76] L. Antolini, P. Boracchi, and E. Biganzoli, “A time-dependent discrimination index for survival data,” *Stat Med*, vol. 24, no. 24, pp. 3927–3944, Dec. 2005, doi: 10.1002/SIM.2427.

[77] “dynCJM function | R Documentation.” Accessed: Sep. 28, 2020. [Online]. Available: https://www.rdocumentation.org/packages/JM/versions/1.4-8/topics/dynCJM

[78] R Core Team, “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria, 2019.

[79] “The Role of Biomarkers and Echocardiography in Prediction of Prognosis of Chronic Heart Failure Patients - Full Text View - ClinicalTrials.gov.” Accessed: Oct. 12, 2020. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT01851538

[80] J. Nunez *et al.*, “Red blood cell distribution width is longitudinally associated with mortality and anemia in heart failure patients,” *Circulation Journal*, vol. 78, no. 2, pp. 410–418, 2014, doi: http://dx.doi.org/10.1253/circj.CJ-13-0630.

[81] Y. Abebaw, K. Mohammed, A. Aragaw, and B. Melese, “Joint Modeling of Longitudinal Pulse Rate and Time-to-Default from Treatment of Congestive Heart Failure Patients,” *Research Reports in Clinical Cardiology*, vol. 12, pp. 41–52, Oct. 2021, doi: 10.2147/RRCC.S326229.

[82] B. Alvarez-Alvarez *et al.*, “Long-term cardiac reverse remodeling after cardiac resynchronization therapy,” *J Arrhythm*, vol. 37, p. 653, 2021, doi: 10.1002/joa3.12527.

[83] S. V Arnold *et al.*, “Health Status After Transcatheter Mitral-Valve Repair in Heart Failure and Secondary Mitral Regurgitation: COAPT Trial,” *J Am Coll Cardiol*, vol. 73, no. 17, pp. 2123–2132, 2019, doi: https://doi.org/10.1016/j.jacc.2019.02.010.

[84] A. T. Belay, D. B. Belay, S. G. Gebremichael, and S. B. Agegn, “Congestive Heart Failure Patients’ Pulse Rate Progression and Time to Death at Debre Tabor Referral Hospital, Ethiopia,” *Adv Public Health*, vol. 2021, pp. 1–8, Nov. 2021, doi: 10.1155/2021/9550628.

[85] J. Biegus, B. Demissei, D. Postmus, G. Cotter, and ..., “Hepatorenal dysfunction identifies high-risk patients with acute heart failure: insights from the RELAX-AHF trial,” *... Heart Failure*, 2019, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ehf2.12477

[86] E. Bouwens *et al.*, “Temporal patterns of 14 blood biomarker candidates of cardiac remodeling in relation to prognosis of patients with chronic heart failure-the Bio-SHiFT study,” *J Am Heart Assoc*, vol. 8, no. 4, 2019, doi: http://dx.doi.org/10.1161/JAHA.118.009555.

[87] E. Bouwens *et al.*, “Circulating biomarkers of cell adhesion predict clinical outcome in patients with chronic heart failure,” *J Clin Med*, vol. 9, no. 1, 2020, doi: http://dx.doi.org/10.3390/jcm9010195.

[88] E. Bouwens *et al.*, “Serially Measured Cytokines and Cytokine Receptors in Relation to Clinical Outcome in Patients With Stable Heart Failure,” *Canadian Journal of Cardiology*, vol. 36, pp. 1587–1591, 2020, doi: 10.1016/j.cjca.2020.08.010.

[89] M. Brankovic *et al.*, “Patient-specific evolution of renal function in chronic heart failure patients dynamically predicts clinical outcome in the Bio-SHiFT study”, doi: 10.1016/j.kint.2017.09.013.

[90] M. Canepa, G. Siri, M. Puntoni, R. Latini, L. Tavazzi, and A. P. Maggioni, “Testing longitudinal data for prognostication in ambulatory heart failure patients with reduced ejection fraction. A proof of principle from the GISSI-HF database,” *Int J Cardiol*, 2020, doi: http://dx.doi.org/10.1016/j.ijcard.2020.03.064.

[91] S. Castelvecchio *et al.*, “Longitudinal profile of NT-proBNP levels in ischemic heart failure patients undergoing surgical ventricular reconstruction: The Biomarker Plus study,” *Int J Cardiol*, vol. 260, pp. 24–30, 2018, doi: http://dx.doi.org/10.1016/j.ijcard.2018.02.084.

[92] K. E. Freedland, B. C. Steinmeyer, R. M. Carney, J. A. Skala, L. Chen, and M. W. Rich, “Depression and Hospital Readmissions in Patients with Heart Failure,” *American Journal of Cardiology*, vol. 0, no. 0, Dec. 2021, doi: 10.1016/J.AMJCARD.2021.10.024.

[93] T. E. Hurst *et al.*, “Dynamic prediction of left ventricular assist device pump thrombosis based on lactate dehydrogenase trends,” *ESC Heart Fail*, vol. 6, no. 5, pp. 1005–1014, 2019, doi: 10.1002/ehf2.12473.

[94] J. P. Kelly *et al.*, “Association of Implantable Device Measured Physical Activity With Hospitalization for Heart Failure,” *JACC Heart Fail*, vol. 04, 2020, doi: http://dx.doi.org/10.1016/j.jchf.2019.10.009.

[95] D. Klimczak-Tomaniak *et al.*, “Temporal patterns of macrophage- and neutrophil-related markers are associated with clinical outcome in heart failure patients,” *ESC Heart Fail*, 2020, doi: http://dx.doi.org/10.1002/ehf2.12678.

[96] J. X. Liu *et al.*, “Repeated measurement of growth-differentiation factor-15 in Chinese Han patients with post-myocardial infarction chronic heart failure,” *Journal of Geriatric Cardiology*, vol. 15, no. 10, pp. 618–627, 2018, doi: http://dx.doi.org/10.11909/j.issn.1671-5411.2018.10.002.

[97] J. Nunez *et al.*, “Long-term serial kinetics of N-terminal pro B-type natriuretic peptide and carbohydrate antigen 125 for mortality risk prediction following acute heart failure,” *Eur Heart J Acute Cardiovasc Care*, vol. 6, no. 8, pp. 685–696, 2017, doi: http://dx.doi.org/10.1177/2048872616649757.

[98] M. M. Schreuder *et al.*, “Sex-specific temporal evolution of circulating biomarkers in patients with chronic heart failure with reduced ejection fraction,” 2021, doi: 10.1016/j.ijcard.2021.04.061.

[99] N. van Boven *et al.*, “Serially measured circulating miR-22-3p is a biomarker for adverse clinical outcome in patients with chronic heart failure: The Bio-SHiFT study,” *Int J Cardiol*, vol. 235, pp. 124–132, 2017, doi: http://dx.doi.org/10.1016/j.ijcard.2017.02.078.

[100] N. van Boven, L. C. Battes, K. M. Akkerhuis, and ..., *Toward personalized risk assessment in patients with chronic heart failure: detailed temporal patterns of NT-proBNP, troponin T, and CRP in the Bio-SHiFT …*. Elsevier, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0002870317303368

[101] V. J. Van Den Berg *et al.*, “Longitudinally Measured Fibrinolysis Factors are Strong Predictors of Clinical Outcome in Patients with Chronic Heart Failure: The Bio-SHiFT Study,” *Thromb Haemost*, vol. 119, no. 12, pp. 1947–1955, 2019, doi: http://dx.doi.org/10.1055/s-0039-1696973.

[102] V. J. van den Berg *et al.*, “Repeated Echocardiograms Do Not Provide Incremental Prognostic Value to Single Echocardiographic Assessment in Minimally Symptomatic Patients with Chronic Heart Failure: Results of the

Bio-SHiFT Study," *Journal of the American Society of Echocardiography*, vol. 32, no. 8, pp. 1000–1009, 2019, doi: http://dx.doi.org/10.1016/j.echo.2019.04.419.

[103] J. C. Van Den Berge *et al.*, "Left ventricular remodelling and prognosis after discharge in new-onset acute heart failure with reduced ejection fraction," 2021, doi: 10.1002/ehf2.13299.

[104] L. C. van Vark *et al.*, "Prognostic value of serial galectin-3 measurements in patients with acute heart failure," *J Am Heart Assoc*, vol. 6, no. 12, 2017, doi: http://dx.doi.org/10.1161/JAHA.116.003700.

[105] L. C. van Vark *et al.*, "Prognostic Value of Serial ST2 Measurements in Patients With Acute Heart Failure," *J Am Coll Cardiol*, vol. 70, no. 19, pp. 2378–2388, 2017, doi: http://dx.doi.org/10.1016/j.jacc.2017.09.026.

[106] M. Alsefri, M. Sudell, M. García-Fiñana, and R. Kolamunnage-Dona, "Bayesian joint modelling of longitudinal and time to event data: A methodological review," *BMC Med Res Methodol*, vol. 20, no. 1, pp. 1–17, Apr. 2020, doi: 10.1186/S12874-020-00976-2/TABLES/5.

[107] J. Zhang, P. Pellicori, D. Pan, R. Dierckx, A. L. Clark, and ..., *Dynamic risk stratification using serial measurements of plasma concentrations of natriuretic peptides in patients with heart failure*. Elsevier, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167527317374296

[108] M. Canepa, G. Siri, M. Puntoni, R. Latini, L. Tavazzi, and A. P. Maggioni, "Testing longitudinal data for prognostication in ambulatory heart failure patients with reduced ejection fraction. A proof of principle from the GISSI-HF database," *Int J Cardiol*, 2020, doi: http://dx.doi.org/10.1016/j.ijcard.2020.03.064.

[109] K. M. Veen *et al.*, "Clinical impact and 'natural' course of uncorrected tricuspid regurgitation after implantation of a left ventricular assist device: An analysis of the European Registry for Patients with Mechanical Circulatory Support (EUROMACS)," *European Journal of Cardio-thoracic Surgery*, vol. 59, no. 1, pp. 207–216, Jan. 2021, doi: 10.1093/ejcts/ezaa294.

[110] M. J. Sweeting and S. G. Thompson, "Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture," *Biom J*, vol. 53, no. 5, pp. 750–763, Sep. 2011, doi: 10.1002/BIMJ.201100052.

[111] M. Sudell, R. Kolamunnage-Dona, and C. Tudur-Smith, "Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis," *BMC Med Res Methodol*, vol. 16, no. 1, p. 168, 2016, doi: 10.1186/s12874-016-0272-6.

[112] E. Bouwens *et al.*, "Circulating biomarkers of cell adhesion predict clinical outcome in patients with chronic heart failure," *J Clin Med*, vol. 9, no. 1, 2020, doi: http://dx.doi.org/10.3390/jcm9010195.

[113] K. J. M. Janssen *et al.*, "Missing covariate data in medical research: To impute is better than to ignore," *J Clin Epidemiol*, vol. 63, no. 7, pp. 721–727, Jul. 2010, doi: 10.1016/J.JCLINEPI.2009.12.008.

[114] P. Miranda, "Extended Joint Models for Longitudinal and Time-to-Event Data [R package JMbayes2 version 0.1-8]," Sep. 2021, Accessed: Oct. 22, 2021. [Online]. Available: https://CRAN.R-project.org/package=JMbayes2

[115] J. Nunez *et al.*, "Red blood cell distribution width is longitudinally associated with mortality and anemia in heart failure patients," *Circulation Journal*, vol. 78, no. 2, pp. 410–418, 2014, doi: http://dx.doi.org/10.1253/circj.CJ-13-0630.

[116] J. Nunez *et al.*, "Long-term serial kinetics of N-terminal pro B-type natriuretic peptide and carbohydrate antigen 125 for mortality risk prediction

following acute heart failure," *Eur Heart J Acute Cardiovasc Care*, vol. 6, no. 8, pp. 685–696, 2017, doi: http://dx.doi.org/10.1177/2048872616649757.

[117] M. Brankovic *et al.*, "Patient-specific evolution of renal function in chronic heart failure patients dynamically predicts clinical outcome in the Bio-SHiFT study", doi: 10.1016/j.kint.2017.09.013.

[118] S. Castelvecchio *et al.*, "Longitudinal profile of NT-proBNP levels in ischemic heart failure patients undergoing surgical ventricular reconstruction: The Biomarker Plus study," *Int J Cardiol*, vol. 260, pp. 24–30, 2018, doi: http://dx.doi.org/10.1016/j.ijcard.2018.02.084.

[119] J. X. Liu *et al.*, "Repeated measurement of growth-differentiation factor-15 in Chinese Han patients with post-myocardial infarction chronic heart failure," *Journal of Geriatric Cardiology*, vol. 15, no. 10, pp. 618–627, 2018, doi: http://dx.doi.org/10.11909/j.issn.1671-5411.2018.10.002.

[120] T. E. Hurst *et al.*, "Dynamic prediction of left ventricular assist device pump thrombosis based on lactate dehydrogenase trends," *ESC Heart Fail*, vol. 6, no. 5, pp. 1005–1014, 2019, doi: 10.1002/ehf2.12473.

[121] S. V Arnold *et al.*, "Health Status After Transcatheter Mitral-Valve Repair in Heart Failure and Secondary Mitral Regurgitation: COAPT Trial," *J Am Coll Cardiol*, vol. 73, no. 17, pp. 2123–2132, 2019, doi: https://doi.org/10.1016/j.jacc.2019.02.010.

[122] J. Biegus, B. Demissei, D. Postmus, G. Cotter, and ..., "Hepatorenal dysfunction identifies high-risk patients with acute heart failure: insights from the RELAX-AHF trial," *... Heart Failure*, 2019, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ehf2.12477

[123] J. C. Van Den Berge *et al.*, "Left ventricular remodelling and prognosis after discharge in new-onset acute heart failure with reduced ejection fraction," 2021, doi: 10.1002/ehf2.13299.

[124] J. P. Kelly *et al.*, "Association of Implantable Device Measured Physical Activity With Hospitalization for Heart Failure," *JACC Heart Fail*, vol. 04, 2020, doi: http://dx.doi.org/10.1016/j.jchf.2019.10.009.

[125] D. Klimczak-Tomaniak *et al.*, "Temporal patterns of macrophage- and neutrophil-related markers are associated with clinical outcome in heart failure patients," *ESC Heart Fail*, 2020, doi: http://dx.doi.org/10.1002/ehf2.12678.

[126] M. Canepa, G. Siri, M. Puntoni, R. Latini, L. Tavazzi, and A. P. Maggioni, "Testing longitudinal data for prognostication in ambulatory heart failure patients with reduced ejection fraction. A proof of principle from the GISSI-HF database," *Int J Cardiol*, 2020, doi: http://dx.doi.org/10.1016/j.ijcard.2020.03.064.

[127] K. M. Veen *et al.*, "Clinical impact and 'natural' course of uncorrected tricuspid regurgitation after implantation of a left ventricular assist device: An analysis of the European Registry for Patients with Mechanical Circulatory Support (EUROMACS)," *European Journal of Cardio-thoracic Surgery*, vol. 59, no. 1, pp. 207–216, Jan. 2021, doi: 10.1093/ejcts/ezaa294.

[128] B. Alvarez-Alvarez *et al.*, "Long-term cardiac reverse remodeling after cardiac resynchronization therapy," *J Arrhythm*, vol. 37, p. 653, 2021, doi: 10.1002/joa3.12527.

[129] M. M. Schreuder *et al.*, "Sex-specific temporal evolution of circulating biomarkers in patients with chronic heart failure with reduced ejection fraction," 2021, doi: 10.1016/j.ijcard.2021.04.061.

[130] K. E. Freedland, B. C. Steinmeyer, R. M. Carney, J. A. Skala, L. Chen, and M. W. Rich, "Depression and Hospital Readmissions in Patients with Heart

Failure," *American Journal of Cardiology*, vol. 0, no. 0, Dec. 2021, doi: 10.1016/J.AMJCARD.2021.10.024.

[131] J. J. V Mcmurray *et al.*, "Baseline characteristics and treatment of patients in Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF)," *Eur J Heart Fail*, vol. 16, pp. 817–825, 2014, doi: 10.1002/ejhf.115.

[132] S. D. Solomon *et al.*, "Angiotensin–Neprilysin Inhibition in Heart Failure with Preserved Ejection Fraction," *New England Journal of Medicine*, vol. 381, no. 17, pp. 1609–1620, Oct. 2019, doi: 10.1056/NEJMOA1908655/SUPPL_FILE/NEJMOA1908655_DATA-SHARING.PDF.

[133] S. D. Solomon, A. R. Rizkala, J. Gong, and ..., "Angiotensin receptor neprilysin inhibition in heart failure with preserved ejection fraction: rationale and design of the PARAGON-HF trial," *... : Heart Failure*, 2017, [Online]. Available: http://heartfailure.onlinejacc.org/content/5/7/471.short

[134] ATMOSPHERE Committees Investigators, "Protocol for: McMurray JJV, Krum H, Abraham WT, et al. Aliskiren, enalapril, or aliskiren and enalapril in heart failure.," *N Engl J Med*, Apr. 2016, doi: 10.1056/NEJMoa1514859.

[135] PARADIGM-HF Investigators and Committees, "Protocol for: McMurray JJV, Packer M, Desai AS, et al. Angiotensin–neprilysin inhibition versus enalapril in heart failure.," *New England Journal of Medicine*, vol. 371, no. 11, pp. 993–1004, Sep. 2014, doi: 10.1056/NEJMoa1409077.

[136] H. Krum *et al.*, "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*, vol. 13, no. 1, pp. 107–114, Jan. 2011, doi: 10.1093/EURJHF/HFQ212.

[137] H. Krum *et al.*, "The Aliskiren Trial to Minimize OutcomeS in Patients with HEart failure trial (ATMOSPHERE): revised statistical analysis plan and baseline characteristics," *Eur J Heart Fail*, vol. 17, no. 10, pp. 1075–1083, Oct. 2015, doi: 10.1002/EJHF.408.

[138] L. M. Chen, J. G. Ibrahim, and H. Chu, "Sample size and power determination in joint modeling of longitudinal and survival data," vol. 30, no. 18, pp. 2295–2309, 2011, doi: 10.1002/sim.4263.

[139] "optim function - RDocumentation." Accessed: Mar. 07, 2023. [Online]. Available: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/optim

[140] D. Rizopoulos, "Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data," *Biometrics*, vol. 67, no. 3, pp. 819–829, Sep. 2011, doi: 10.1111/J.1541-0420.2010.01546.X.

[141] P. J. Heagerty and Y. Zheng, "Survival model predictive accuracy and ROC curves," *Biometrics*, vol. 61, no. 1, pp. 92–105, Mar. 2005, doi: 10.1111/J.0006-341X.2005.030814.X.

[142] P. J. Heagerty and Y. Zheng, "Survival Model Predictive Accuracy and ROC Curves," *Biometrics*, vol. 61, no. 1, pp. 92–105, Mar. 2005, doi: 10.1111/J.0006-341X.2005.030814.X.

[143] A. J. Patrick Heagerty, P. Saha-Chaudhuri, and M. Paramita Saha-Chaudhuri, "Title Riskset ROC Curve Estimation from Censored Survival Data," 2022.

[144] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2019.

[145] P. Blanche, M. W. Kattan, and T. A. Gerds, “The c-index is not proper for the evaluation of t-year predicted risks,” *Biostatistics*, vol. 20, no. 2, pp. 347–357, Apr. 2019, doi: 10.1093/BIOSTATISTICS/KXY006.

[146] J. D. Long and J. A. Mills, “Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington’s disease,” *BMC Med Res Methodol*, vol. 18, no. 1, pp. 1–15, Nov. 2018, doi: 10.1186/S12874-018-0592-9/FIGURES/5.