The pharmaco-epidemiology of loop diuretic dispensing and its relationship to the diagnosis of heart failure and to prognosis

Jocelyn Marie Friday
MSc, BSc (Hons)

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

Schools of Health & Wellbeing and Cardiovascular & Metabolic Health
College of Medical, Veterinary and Life Sciences
University of Glasgow

October 2023

Abstract

Heart failure is a major and growing public health problem associated with poor patient outcomes, including reduced quality of life and high hospitalisation and mortality rates. It is a complex clinical syndrome rather than a single disease, which lacks a practical, universal, and standardised definition. Currently, the definition relies on the identification of symptoms and signs of cardiac dysfunction, such as ankle swelling and breathlessness, which are neither specific nor objective. Many patients are only diagnosed once their symptoms and signs are severe enough to require hospitalisation. Pathophysiologically, heart failure can be defined by the presence of salt and water retention, also known as congestion, associated with cardiac dysfunction. Within the United Kingdom, the pharmacological class of loop diuretics is used primarily for the treatment of congestion due to cardiac dysfunction. The aim of this thesis is to investigate the pharmacoepidemiology of loop diuretic dispensing and its relationship to the diagnosis of heart failure, with a particular focus on patient outcomes.

The first analysis describes the prevalence of repeated loop diuretic dispensing and/or diagnosis of heart failure within the NHS Greater Glasgow & Clyde Health Board population on 1st January 2012, including patient outcomes over the following five years. This research is thought to be the first population-level investigation into the prevalence of repeated loop diuretic dispensing and its prognostic significance in patients with and without a diagnosis of heart failure. The analysis found that an estimated 3.2% of the population received repeated loop diuretic dispensing, while only 1.3% of the population had a diagnosis of heart failure. Hospitalisation rates were higher in those with a loop diuretic (0.99 admissions per patient-year at risk for those with only repeated loop diuretic dispensing and 1.51 admissions per patient-year at risk for those with both) than those with only a diagnosis of heart failure (0.93 admissions patient-year at risk). All-cause mortality followed a similar pattern; adjusting for age, sex, socioeconomic deprivation and comorbidity status, the 5-year hazard ratio and (95% confidence interval) were 1.8 (1.8 - 1.9) for those with those only repeated loop diuretic dispensing and 2.3 (2.2 - 2.4) for those with both, while only 1.2 (2.2 - 2.4) for those with only a diagnosis of heart failure, implying that the presence of repeated loop diuretic dispensing is a marker of serious disease.

The second analysis stepped backwards in 'patient-time' to describe the pattern of hospitalisations in the year leading up to the initiation of loop diuretic dispensing or an incident diagnosis of

heart failure using network graphs. While the precursors to heart failure are known, this research is thought to be the first to report the common patterns in events leading up to the initiation of loop diuretics. While there was little difference in comorbidity and medication levels 24 months prior, in the year leading up to the initiation, those who received a diagnosis of heart failure were more likely to be admitted for well-recognised contributors to the condition, including ischaemic heart disease in particular, but also atrial fibrillation/flutter and valve disease. In contrast, these patterns were not often seen in those who were only initiated on a loop diuretic, instead with a focus on admissions for non-specific symptoms and signs, most commonly unspecified chest pain.

The third analysis starts where the second leaves off. It assesses the prognostic relationship between the initiation of loop diuretic and diagnosis of heart failure on mortality and whether the sequence of these events matters using semi-Markov multi-state modes, a flexible model for use on longitudinal time data where there is an event-related dependence on outcomes. Those on repeated loop diuretic dispensing without a diagnosis of heart failure were majority women (62%). Many with evidence of left atrial dilation (53%), while those with a diagnosis of heart failure without a repeat loop diuretic were majority men (63%). Many had a history of myocardial infarction (51%). Hospitalisations and mortality were higher in those with a repeat loop diuretic (within the first year per patient-year at risk: hospitalisation, 1.44; mortality, 0.20) compared to those with a diagnosis of heart failure without a repeat loop diuretic (within the first year per patient-year at risk: hospitalisation, 1.47; mortality, 0.14). Rates were higher still in those with both loop diuretic and heart failure (where both events occurred together within the first year per patient-year at risk: hospitalisation, 1.74; mortality, 0.16; or where the diagnosis of HF preceded the initiation of loop diuretic, within the first year per patient-year at risk: hospitalisation, 1.68; mortality, 0.20), with the highest being in those who initiated the loop diuretic in advance of receiving a diagnosis of heart failure (within the first year per patient-year at risk: hospitalisation, 2.26; mortality, 0.28).

The fourth and final analysis subsets the population to investigate the mortality of the 24,921 patients with ischaemic heart disease according to whether or not they have had a repeat loop diuretic and/or diagnosis of heart failure; of whom, 3,806 had only repeat loop diuretic, 2,384 had only a diagnosis of heart failure, and 3,531 had both. This analysis found that after adjusting for age, sex, and other prognostic markers, mortality was associated with the repeat loop diuretic regardless of the patient's heart failure status. Those with a repeat loop diuretic without a diagnosis of heart failure experienced substantially higher rates of cardiovascular (an estimated 15%) and all-cause mortality (47%) than those with a diagnosis of heart failure without a repeat loop diuretic (an estimated 8% cardiovascular and 19% all-cause mortality), while rates were highest for those with both (an estimated 25% cardiovascular and 57% all-cause mortality).

In conclusion, these analyses found that many more patients are repeatedly treated with loop di-

uretic than ever receive a diagnosis of heart failure. These patients are at a high risk of hospitalisation and death, and based on their characteristics, many probably have undiagnosed heart failure. From a public health and epidemiological perspective, the current definition of heart failure likely underestimates the true burden on the healthcare system. From the patient's perspective, with the efficacy of angiotensin receptor-neprilysin inhibitor, sodium-glucose co-transporter-2 inhibitors, and mineralocorticoid receptor antagonistss, a missed diagnosis means a missed opportunity to improve the patient's outcome and quality of life, regardless of their heart failure phenotype. Even more alarming, if these patients are receiving the loop diuretic inappropriately, the loop diuretic is likely causing these increased hospitalisation and mortality rates. If the loop diuretic can be safely withdrawn, other medications with diuretic properties exist which have good safety profiles. Ultimately, further research is required to determine the optimal strategy for managing these patients.

Contents

List of Tables

List of Figures

Abbreviations & Acronyms

95% CI 95% confidence interval.

ACaDME Acute, Cancer, Deaths and Mental Health.

ACEi angiotensin-converting enzyme inhibitor.

ACM all-cause mortality.

AF atrial fibrillation.

AF/AFL atrial fibrillation/flutter.

AFL atrial flutter.

AFT accelerated failure time.

AR aortic regurgitation.

ARB angiotensin receptor blocker.

ARNi angiotensin receptor-neprilysin inhibitor.

AS aortic stenosis.

BMI body mass index.

BNF British National Formulary.

BNP B-type natriuretic peptide.

BSA body surface area.

CABG coronary artery bypass surgery.

CAD coronary artery disease.

CCB calcium channel blocker.

CHI Community Health Index.

CIF cumulative incidence function.

CKD chronic kidney disease.

CKD-EPI Chronic Kidney Disease Epidemiology Collaboration.

COD cause of death.

COPD chronic obstructive pulmonary disease.

Cox PH Cox proportional hazards.

CPRD Clinical Practice Research Datalink.

CRT cardiac resynchronisation therapy.

CSV comma-separated value.

CV cardiovascular.

Diabetes+ diabetes mellitus or concurrent hypoglycaemic agents.

DM diabetes mellitus.

DOB date of birth.

DOD date of death.

Dx diagnosis of.

ECG electrocardiogram.

echo echocardiogram.

eGFR estimated glomerular filtration rate.

EHRs electronic health records.

EP electrophysiology/electrophysiological.

EPR electronic patient records.

ESC European Society of Cardiology.

GI gastrointestinal.

GP general practitioner.

GP LES General Practice Local Enhanced Services.

H/o history of.

Hb haemoglobin.

HF heart failure.

HFmrEF heart failure with mildly reduced ejection fraction.

HFpEF heart failure with preserved ejection fraction.

HFrEF heart failure with reduced ejection fraction.

HFsnEF heart failure with supra-normal ejection fraction.

HR hazard ratio.

ICD implantable cardioverter defibrillator.

ICD-10 International Classification of Diseases, 10th Revision.

IDMS isotope dilution mass spectrometry.

IHD ischaemic heart disease.

IVC inferior vena cava.

KM Kaplan-Meier survival estimator.

LA left atrium/atrial.

LAA left atrial area.

LAD left atrial diameter.

LAV left atrial volume.

LD loop diuretic.

LES Local Enhanced Services.

LV left ventricular/ventricle.

LVEF left ventricular ejection fraction.

LVSD left ventricular systolic dysfunction.

MAR missing at random.

MCAR missing completely at random.

MDRD Modification of Diet in Renal Disease Study Equation.

MI myocardial infarction.

MNAR missing not at random.

MR mitral regurgitation.

MRA mineralocorticoid receptor antagonists.

NHS National Health Service.

NHS GG&C NHS Greater Glasgow & Clyde Health Board.

NRS National Records Scotland.

NSAIDs non-steroidal anti-inflammatory drugs.

NT-proBNP N-terminal pro-B-type natriuretic peptide.

OPCS-4 Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4.

PAD peripheral arterial disease.

PC primary care.

PCI percutaneous coronary intervention.

PH proportional hazards.

PHS Public Health Scotland.

PIS Prescribing Information System.

Pulm Dis pulmonary disease.

QTc corrected QT interval.

RAA right atrial area.

RAAS renin-angiotensin-aldosterone system.

RAASi renin-angiotensin-aldosterone system inhibitors.

RAV right atrial volume.

RCT randomised controlled trial.

RV right ventricular/ventricle.

Rx prescription of.

S&S symptoms and signs.

SCI Diabetes Scottish Care Information-Diabetes Collaboration.

SCI Store Scottish Care Information Store.

SGLT2i sodium-glucose co-transporter-2 inhibitor.

SIMD Scottish Index of Multiple Deprivation.

SMR Scottish Morbidity Records.

SMR00 Scottish Morbidity Records - Outpatient Attendance.

SMR01 Scottish Morbidity Records - General/Acute Inpatient and Day Case.

SMR04 Scottish Morbidity Records - Mental Health Inpatient and Day Case.

SQL Structured Query Language.

SVT supraventricular tachycardia.

T2DM type-2 diabetes mellitus.

TAPSE tricuspid annular plane systolic excursion.

Thiazides+ thiazides and thiazide related diuretics.

TR tricuspid regurgitation.

UK United Kingdom.

UTI urinary tract infections.

WHO World Health Organisation.

WoS West of Scotland.

Special Terms

Both: HF First The grouping assigned when a patient started a repeat loop diuretic prescription more than 30 days after receiving a diagnosis of heart failure.

Both: LD + HF The grouping assigned when a patient has both a repeat prescription of a loop diuretic and a diagnosis of heart failure where the ordering of events cannot be determined.

Both: LD First The grouping assigned when a patient received a diagnosis of heart failure more than 30 days after starting a repeat loop diuretic prescription.

Both: Together The grouping assigned when a patient started a repeat loop diuretic prescription within a 30-day window of receiving a diagnosis of heart failure.

HF Death as First Record The grouping assigned when a patient was diagnosed with heart failure during a fatal hospitalisation did not have a pre-existing, repeat loop diuretic prescription.

HF Only The grouping assigned when a patient has a diagnosis of heart failure in the absence of a repeat loop diuretic prescription.

LD Only The grouping assigned when a patient has a repeat loop diuretic prescription in the absence of a diagnosis of heart failure.

Neither The grouping assigned when a patient has a diagnosis of coronary artery disease or peripheral arterial disease, or has been prescribed an angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, mineralocorticoid receptor antagonists, beta-blocker, or a non-repeat loop diuretic prescription in the absence of a diagnosis of heart failure and a repeat loop diuretic prescription.

Neither ($\geq$60 yrs) The grouping assigned when a patient has a diagnosis of coronary artery disease or peripheral arterial disease, or has been prescribed an angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, mineralocorticoid receptor antagonists, beta-blocker, or non-repeat prescription of a loop diuretic in the absence of a diagnosis of

heart failure and repeat loop diuretic prescription and is at least 60 years of age at the time of assignment.

Neither (18-59 yrs) The grouping assigned when a patient has a diagnosis of coronary artery disease (CAD) or peripheral arterial disease (PAD), or has been prescribed an angiotensin-converting enzyme inhibitor (ACEi), angiotensin receptor blocker (ARB), mineralocorticoid receptor antagonists (MRA), beta-blocker, or non-repeat prescription of a loop diuretic in the absence of a diagnosis of heart failure and repeat loop diuretic prescription and is between 18 and 59 years of age at the time of assignment.

No LD/No HF The grouping assigned when a patient has a prevalent diagnosis of ischaemic heart disease in the absence of a diagnosis of heart failure and a repeat loop diuretic prescription.

Acknowledgements

> No one achieves success does so without acknowledging the help of others. The wise and confident acknowledge this help with gratitude.
>
> *Alfred North Whitehead*
> *Mathematician and philosopher*

While this thesis is the culmination of many years of hard work, not to say that I am wise, but I am confident that it would not have been possible without the support of a diverse group of people and organisations.

The following research would not have been possible without the considerable help, support, and efforts of the NHS Greater Glasgow & Clyde Health Board Safe Haven team. I owe many thanks to the staff, past and present, who were involved in the collection, processing, linking, anonymising, and releasing NHS data. In particular, I would like to thank Neil Hillen, Alison Hamilton, Charlie Mayor, and Beth Bruce.

Thanks are also due to the Robertson Centre for Biostatistics staff for being a wonderful department to work within, especially to Richard Papworth for introducing me to the world of Safe Haven analysis, providing sense checking, code review, and the idea for Chapter 4.

The work contained within this thesis was enriched and supported by my five-member multidisciplinary supervisory team. As my primary supervisor and insight into the world of heart failure, John Cleland was instrumental in this thesis. He gave me the freedom to develop my work and ideas while also being there to provide guidance, support, answers, and a critical eye. I cannot thank him enough. I would also like to thank my statistical supervisors, Sarah Barry and Jim Lewsey, for their invaluable guidance, support, and willingness to work through ever more complex assumptions testing methodologies, and to Jim especially for stepping in near the end of a thesis. My thanks also to Maria Wolters for being my computer science supervisor and continually providing suggestions and support ranging from network analysis to visualisation and to LaTeX help. Finally, heartfelt thanks to David McAllister as my epidemiology supervisor. He

provided invaluable guidance in suggesting how to work with this complex cohort. This PhD and thesis would not be what it is without their help, guidance, support, suggestions, encouragement, and domain expertise. They have each forgotten more in their individual fields than I will ever learn, but combined, the results can impact future healthcare and hopefully improve lives.

I am grateful for the contributions of all of my co-authors and my annual review team for providing valuable ideas, constructive input, and domain expertise. In particular, I would like to thank Yola Jones, Pierpaolo Pellicori, Danny Mackay, and Fraser Graham.

And finally, to my friends and family, thank you for everything.

Declaration

I declare that the contents of this thesis are my own work and have not been submitted for any other degree at the University of Glasgow or any other institution. Where the work of others has been used, it has been indicated and appropriately referenced.

Jocelyn Friday

Chapter 1

Introduction

> There is no such uncertainty as a sure thing.
>
> *Robert Burns*
> *Poet*

Heart failure (HF) was first declared an emerging epidemic in 1997 (Braunwald 1997, Roger 2021) based on the increasing numbers of HF hospital admissions. Since then, the burden attributed to HF has increased to the point where it is on par with the four most common cancers combined (Conrad et al. 2018). Yet, these impacts are likely underestimated due to the lack of a practical, universal, and standardised definition of HF.

1.1 Definition of Heart Failure

There is widespread agreement that HF is common and often debilitating. This is all while being associated with poor outcomes, both in terms of recurrent hospitalisations and high mortality rates and consuming substantial healthcare resources at considerable cost (Conrad et al. 2018, McDonagh et al. 2021, Cleland et al. 2011). Given these factors, it is surprising that HF lacks a robust definition, the absence of which confounds the ability to understand HF's epidemiology and prognosis and undermines the estimates needed for health service provisioning for diagnosis and management (Cleland et al. 2021, Davies et al. 2001).

The first European Society of Cardiology (ESC) guidelines for the diagnosis of HF were published in 1995 (The Task Force on Heart Failure of the European Society of Cardiology 1995). The guidelines recognised the complexity of defining HF, but due to the absence of widely accepted non-invasive measures of congestion at the time, HF was defined by symptoms due to cardiac dysfunction (see Figure 1.1). The most important symptom is breathlessness, which initially only occurs on exertion, but as HF progresses, also when lying down (orthopnoea) or

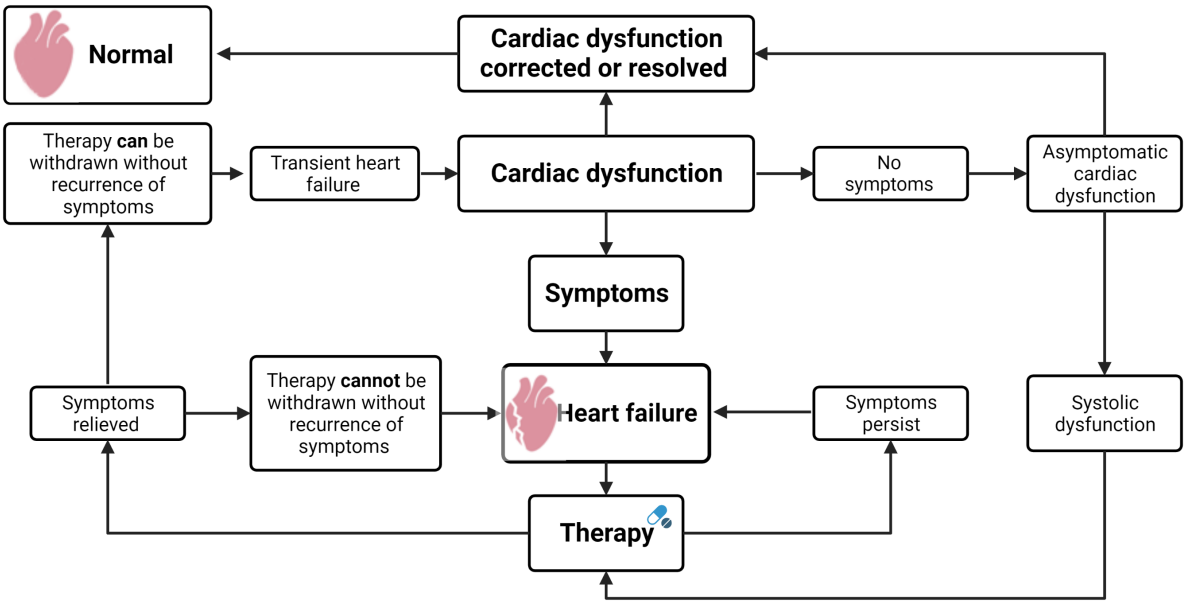


Figure 1.1: An illustration of the relationship between cardiac dysfunction, HF, and resolved HF. The figure was adapted to show the central role therapy played in the diagnosis of HF (Remme & Swedberg 2001).

at rest when sitting up (pulmonary oedema), which is a medical emergency. However, breathlessness is non-specific and can simply be due to an individual doing more than their usual physical activity (e.g., running to catch a bus or walking up hills against inclement weather). Additionally, individuals may be unfit or overweight due to other medical conditions such as smoking-induced chronic bronchitis and asthma. Another common symptom of HF is swelling of the ankles and legs (peripheral oedema), which can also occur due to prolonged periods of sitting or standing, wearing tight-fitting socks, high levels of salt intake, varicose veins, arthritis, or a variety of other medical problems (including certain medications).

These guidelines went on to say that clinical response to the HF treatment was not sufficient in and of itself to make a diagnosis, although the patient should show some response to treatment. Additionally, "treatment may obscure a diagnosis of [HF] by relieving the patient's symptoms" (The Task Force on Heart Failure of the European Society of Cardiology 1995). Of note, a small study of general practitioners (GPs) showed that this test was often done in clinical practice. Yet, the research found it to be an ineffective diagnostic tool[1] (Kelder et al. 2011).

Subsequent ESC Guidelines have not fundamentally changed since this definition. In 2021, a consensus of European, American, and Japanese experts endorsed this definition (Bozkurt et al. 2021). However, other experts disagreed and argued for change (Cleland et al. 2021).

As Cleland et al. point out, the problem with the 1995 and subsequent definitions is that the

[1]The researchers explicitly mention that rationale for such a diagnostic test was obvious, going so far as to point out the inclusion of 4.5 kg weight reduction as a major diagnostic criterion for original Framingham diagnostic criteria (McKee et al. 1971, Kelder et al. 2011).

symptoms and signs of HF are subjective, late manifestations of the condition, and are non-specific. Additionally, as was pointed out in 1995, the symptoms and signs may be relieved without improving the underlying cardiac dysfunction or prognosis (The Task Force on Heart Failure of the European Society of Cardiology 1995). The problem of subjectivity lies both with the patient and the healthcare professional. Patients often attribute symptoms to getting older, being unfit, or to another condition, such as lung disease or their weight (i.e., obesity). Many will find that reducing physical activity avoids the symptoms, and this can continue until the problem becomes so severe that symptoms occur even at rest, which probably accounts for why a first diagnosis of HF is usually made during a hospital admission, a rate of almost 80% of diagnoses in England (Bottle et al. 2018). For many older patients, it may simply be too difficult for them to get an appointment or to travel to the GP surgery, while others do not like to 'bother' their doctor. Busy healthcare professionals may not ask about symptoms of HF during routine check-ups for conditions like high blood pressure (hypertension) or after a heart attack (myocardial infarction [MI]). Alternatively, others may concur with their patient that the reason for their symptoms is that they are just getting older, are not fit, and are overweight. Many patients are already on treatments commonly used for the treatment of HF, such as angiotensin-converting enzyme inhibitor (ACEi), beta-blockers, and thiazides, for problems such as hypertension or prior MI. Yet many patients are prescribed loop diuretics (LD) for reasons that are not always obvious.

The most commonly used method of cardiac imaging for patients with HF is echocardiography. The interpretation of echocardiograms has inter- and intra-observer error (Barnhart et al. 2016, Rusterborgh et al. 1990, O'Dell 2019). The left ventricular ejection fraction (LVEF) is the most used measurement when diagnosing and assessing HF. LVEF is the stroke volume (the volume of blood pumped out of the left ventricle [LV] with each heartbeat) divided by the volume of blood in the left ventricle when it is completely full. Conventionally, HF is classified into phenotypes by LVEF, into reduced (HFrEF) when $\leq$40%, mildly reduced (HFmrEF) when the LVEF is 41-49%, and preserved (HFpEF) when the LVEF $\geq$50%. HFpEF includes LVEF values in the normal range (55-70%). However, an LVEF measurement taken in clinic should be expected to vary about 3%-8% (O'Dell 2019). One of the reasons the HFmrEF classification was introduced was to reduce the diagnostic error between HFrEF and HFpEF, with HFmrEF acting as a zone of diagnostic uncertainty (Savarese, Stolfo, Sinagra & Lund 2022). More recently, and not uniformly accepted yet, a new classification for HF patients with a supra-normal LVEF of >70% (HFsnEF) have been recognised, for which a new in-class treatment may be coming onto the market (cardiac myosin inhibitors) (Olivotto et al. 2020). Based on measurement variability and different aetiologies, HF, especially HFpEF, cannot be diagnosed through LVEF alone (Myhre et al. 2020, Shah et al. 2014, Beale et al. 2018).

As noted above, the identification of HF based on symptoms and signs is highly subjective. A comparative study of four different HF diagnostic criteria (Framingham instrument [McKee

et al. 1971], Boston instrument [Carlson et al. 1985], Gothenburg criteria [Eriksson et al. 1987], and the ESC principles [The Task Force on Heart Failure of the European Society of Cardiology 1995]) found only moderate diagnostic agreement at best between the four diagnostic criteria (Di Bari et al. 2004). (See Roger for a comprehensive table comparing the four different diagnostic criteria for HF used in epidemiological research (Roger 2013). The criteria remain the same even though the table was published in 2013.) Criteria which were based on the presence of symptoms and signs suggested a much higher prevalence of HF within the same community (up to 20.8%) than were captured using the ESC principles (9.0%) [2].

An alternative approach to defining HF avoids the need for symptoms and signs, and instead defines HF as cardiac dysfunction due to congestion (Cleland et al. 2021). This definition uses the same methods of assessing cardiac dysfunction as before, but rather than ascertaining the presence of symptoms and signs, the focus moves towards assessing congestion. Previously, congestion was assessed through clinical assessment for peripheral oedema, heart sounds, and distention of the veins in the neck. Each of these has a subjective component, and low specificity, and all are late manifestations of disease (Martens & Mullens 2018, Girerd et al. 2018). Additionally, the latter two require considerable clinical skill and experience to identify. Over the last 20 years, an objective tool in the form of the measurement of natriuretic peptides (e.g., B-type natriuretic peptide [BNP] and N-terminal pro-B-type natriuretic peptide [NT-proBNP]) have proved their diagnostic and prognostic value, becoming an established tool for measuring congestion levels (Bay et al. 2003, Girerd et al. 2018, Núñez et al. 2022). Additionally, there is growing recognition that the atria (the receiving chambers of the heart [see Figure 1.1]) may be a more sensitive and specific measurement of congestion due to cardiac dysfunction (Cuthbert et al. 2021, Cleland et al. 2021). Using the presence of congestion and cardiac dysfunction as the backbone, an alternative definition to HF may be a raised natriuretic peptide with a dilated atrium and, specifically, a left atrium (LA) when focusing on left-sided heart disease, the most common group of causes of HF (Pazos-López et al. 2011).

Echocardiograms and natriuretic peptide testing are not part of routine care, especially within the National Health Service (NHS). These tests are usually only ordered once a healthcare professional suspects or wishes to exclude the diagnosis of HF[3].

As noted above, the many treatments used for HF are also indicated for the treatment of other conditions such as hypertension or treatment for MI. However, LD is the one treatment that has few other indications for treatment other than for congestion due to cardiac dysfunction, and it is considered essential for the management of congestion (McDonagh et al. 2021, National Institute for Health and Care Excellence 2018*a*). Due to its fundamental nature and limited

[2]The ESC definition is the only criteria to require objective evidence of cardiac dysfunction at rest, which in the context of population-level studies means the evaluation of appropriate tests, which is not always done in practice (Roger 2013)

[3]Or potentially other cardiac issues in the case of echocardiograms.

other indications, LD could be considered a marker of congestion, which could potentially mask the underlying diagnosis (see Figure 1.1). Patients with symptoms and signs of HF will almost always be prescribed a LD in accordance with guidelines. It is less clear how many patients are initiated on LD for its other indications, resistant hypertension or end-stage kidney disease. However, many HF patients will also have these conditions (McDonagh et al. 2021). Regarding the treatment of resistant hypertension, thiazides are one of the first-line treatments for hypertension but are not recommended for HF symptom relief (McDonagh et al. 2021, Williams et al. 2018). When combined, thiazides and LD constitute a powerful, and potentially dangerous, diuretic combination that may cause large imbalances in electrolytes, such as hypokalaemia (see Section 2.5.9), dehydration, renal dysfunction, and hypotension (Trullàs et al. 2022).

1.2 Epidemiology of Heart Failure

Currently, an estimated 1% to 3% of the population has a diagnosis of HF, but the prevalence is higher in older individuals (Savarese, Becher, Lund, Seferovic, Rosano & Coats 2022, Benjamin et al. 2017, Jacobs et al. 2017, Davies et al. 2001). While the age-related prevalence may be falling in wealthier countries, the overall prevalence increases as the proportion of the population aged >60 years increases (Conrad et al. 2018). Where there are adequate records, the age-standardised prevalence varies dramatically, while the incidence is more uniform (Savarese, Becher, Lund, Seferovic, Rosano & Coats 2022). The incidence of HF is less well documented than the prevalence, with some estimates suggesting 2-3 cases per 1000 population within the wealthiest countries (Savarese, Becher, Lund, Seferovic, Rosano & Coats 2022) and others estimating that one in five people will receive a diagnosis of HF during their lifetime (Lloyd-Jones et al. 2002).

However, attempts to describe the epidemiology of HF only serve to highlight how poor and heterogeneous the evidence-base is, with Germany showing a higher estimated prevalence and incidence of HF (3.9% and 6.5 per 1,000 population age-standardised, respectively), compared with Belgium (1.5% and 2.7 per 1,000 person-years) or United Kingdom (UK) (1.6% and 3.5 per 1,00 person-years) (Savarese, Becher, Lund, Seferovic, Rosano & Coats 2022). These discrepancies may reflect a financial incentive to make a diagnosis of HF created by healthcare systems that reinforce subjective biases in different directions or due to the general structure of the healthcare systems in general and their referral mechanisms.

1.3 Causes of Heart Failure

On a functional level, the heart is a muscular pump consisting of two receiving chambers (the atria) and two pumping chambers (the ventricles) (see Figure 1.2). The right side of the heart normally handles low-pressure blood flow, collecting deoxygenated blood from the body and

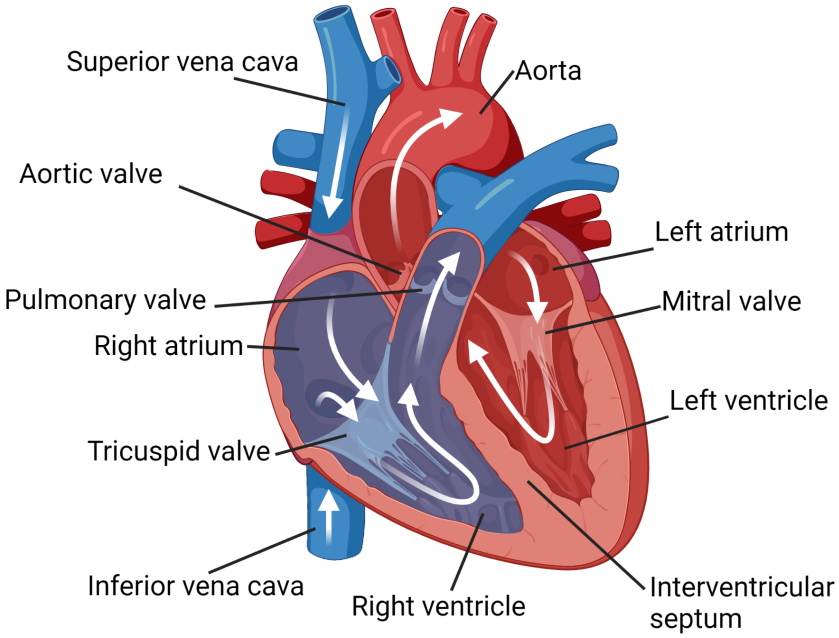


Figure 1.2: An illustration of a heart showing how blood flows into and out of the heart (white arrows). The portions in blue indicate deoxygenated (oxygen-poor) blood as it returns from the body. The portions in red indicate oxygenated (oxygen-rich) blood that will be circulated to the body.

pumping it to the lungs for oxygenation, while the left side of the heart is designed to deal with the higher pressures necessary to pump blood around the body.

Chronic HF is a cardio-renal syndrome defined by the presence of cardiac dysfunction that leads to water and salt retention and the clinical manifestations of congestion (Cleland et al. 2021, Martens et al. 2015). As a syndrome, it has numerous etiologies, and it is often very difficult to determine the primary aetiology in highly comorbid patients (McMurray & Stewart 2000); regardless of aetiology, the result is the same: poor patient prognosis. The most common cause of HF is LV dysfunction, which will often be due to hypertension, coronary artery disease (CAD), or their combined effects (McDonagh et al. 2021). It may further be complicated by valve disease and/or atrial fibrillation (AF). CAD may cause a MI, which in turn will often lead to the damage and the loss of large amounts of cardiac muscle. Other aetiologies include diseases in the primary heart muscle (e.g., dilated cardiomyopathy) and myocardial infiltration (amyloid) (McDonagh et al. 2021).

HF is more often found in patients with chronic kidney disease (CKD). This could either be because HF and CKD have shared associated comorbidities (e.g., hypertension, diabetes mellitus [DM], and ischaemic heart disease [IHD]), or because CKD impairs the kidneys' ability to excrete water and salt (Cleland et al. 2021).

1.4 Detecting Cardiac Dysfunction

There are two broad approaches to the detection of cardiac dysfunction: patients reporting symptoms and signs and diagnostic testing. In the first instance, patients may complain about breathlessness or ankle swelling, which should arouse the suspicion of the healthcare professional. In the second instance, the healthcare professional might take pro-active steps to exclude or make a diagnosis of HF in patients who are in a high-risk group for developing it (e.g., either due to older age, long-standing hypertension, AF, a history of MI, or history of chemotherapy for the treatment of cancer). Occasionally, the healthcare professional may 'stumble' over the diagnosis as an incidental finding (e.g., on a chest X-ray or echocardiogram) during tests done for other reasons.

Both European (McDonagh et al. 2021) and UK (National Institute for Health and Care Excellence 2018*b*) guidelines strongly advise testing natriuretic peptides when identifying dysfunction, with an emphasis on NT-proBNP in the UK. If the tests are normal, this essentially excludes the presence of congestion and, therefore, the presence of significant left or right ventricular dysfunction. However, normal test results do not exclude the possibility that low levels of CAD could lead to a MI in time, with the associated elevated risks and damage.

There are two situations where plasma concentrations of natriuretic peptides can be relatively normal in the presence of cardiac disease. The first is when the patient is morbidly obese (Madamanchi et al. 2014), and the second is when the patient has a thickening of the fibrosis capsule surrounding the heart (constrictive pericarditis) (Leya et al. 2005). However, even in the presence of these conditions, natriuretic peptides are usually still raised if congestion is present (Leya et al. 2005, Madamanchi et al. 2014, Cleland et al. 2021).

In the absence of LV dysfunction, natriuretic peptides may be elevated when patients have severely reduced estimated glomerular filtration rate (eGFR) (Wiley et al. 2010), as severely reduced eGFR increases water and salt retention, or when patients are in AF (Brady et al. 2022). Both severely reduced eGFR and AF provide further stress to the heart and are associated with increased risk of mortality and morbidity. Acknowledging these factors on natriuretic peptide levels, results should be interpreted taking into account the patient's eGFR, heart rhythm, age, and sex, at a minimum, when identifying elevated levels (Welsh et al. 2022).

If natriuretic peptide levels are normal, further investigations for cardiac dysfunction or HF are of little added value. If values are elevated, further information is needed, usually in the form of an echocardiogram. If the echocardiogram shows a dilated LA, this confirms congestion due to cardiac dysfunction. The patient could be considered to have HF, even in the absence of symptoms and signs, which may only be present when the congestion is severe (Cleland et al. 2021).

1.5 Treating Heart Failure

1.5.1 Treatments Designed to Delay Progression and Improve Prognosis

Over the last three decades, treatments for HFrEF have been introduced through a series landmark randomised controlled trials (RCTs), starting with angiotensin-converting enzyme inhibitor (ACEi) in the 1987 (Consensus Trial Study Group 1987), then beta-blockers (CIBIS Investigators and Committees 1994, Australia/New Zealand Heart Failure Research Collaborative Group 1997) and mineralocorticoid receptor antagonists (MRA) (Pitt et al. 1999) in the 1990s, implantable cardioverter defibrillator (ICD) and cardiac resynchronisation therapy (CRT) devices in the 2000s (Moss et al. 2002, 2009), angiotensin receptor-neprilysin inhibitor (ARNi) in 2014 (McMurray et al. 2014), and sodium-glucose co-transporter-2 inhibitor (SGLT2i) in 2019 (McMurray et al. 2019, McDonagh et al. 2021). Beta-blockers and CRT can improve LV function and prognosis.

There is a growing consensus that patients with HFmrEF benefit from the use of ACEi, beta-blockers, MRA, and SGLT2i (Xiang et al. 2022). However, they do not benefit to the same extent as HFrEF patients, though they tend to have a better overall prognosis, and therefore there is less room for improvement (at least in the short-to-medium-term) (Chioncel et al. 2017).

Patients with HFpEF who are not overtly congested have an even better prognosis than HFmrEF (Chioncel et al. 2017, Pellicori et al. 2016). Recent trials suggest that SGLT2i may improve symptoms and reduce hospitalisations (Anker et al. 2021), although whether they reduce mortality to the whole class is controversial (Requena-Ibanez et al. 2022). This group of patients often have hypertension, AF, anaemia, and a high BMI, and treating these problems may improve symptoms and signs and prognosis.

The classification of HFsnEF, as a phenotype has only recently been recognised. This is in part due to the advent of cardiac myosin inhibitors (Ho et al. 2020), which target this population and lend further support for recognising this distinct phenotype.

All of this being said, patients with overt congestion have a poor prognosis regardless of LV phenotype (Cleland et al. 2011).

1.5.2 Treatments Designed to Improve Congestion and Symptoms

Diuretics

Diuretics are the primary treatment for congestion in order to relieve the associated symptoms and signs (Faris et al. 2012, Pellicori et al. 2016, Cleland et al. 2012), and might be considered to be the only available treatment for congestion other than complex and expensive procedures such as haemofiltration. Diuretics work by increasing water and salt excretion via the kidneys,

but they may also increase potassium excretion, reduce blood pressure, impair renal function, and activate the renin-angiotensin-aldosterone system (RAAS), which may have adverse consequences (Wile 2012, Cody et al. 1982, Feigenbaum et al. 2000, Cohn et al. 1984).

Loop Diuretics

Loop diuretics (LDs), also known as high ceiling diuretics, can induce a powerful dieresis. They act on the loop of Henlé, targeting the sodium potassium chloride co-transporter. The primary indication for LD is for the treatment of water and salt overload due to cardiac dysfunction (i.e., congestive HF) (Ellison & Felker 2017). They may also be used to treat resistant hypertension and in end-stage renal disease to manage water and salt retention, although at the risk of accelerating the decline in renal function (National Institute for Health and Care Excellence 2018*a*, Clarke et al. 1995, Khan et al. 2016).

The dieresis caused by LD can cause gout, impaired eGFR, inconvenience due to increased and urgent need for urination, and potentially exacerbate DM (National Institute for Health and Care Excellence 2018*a*, Ellison & Felker 2017). Additionally, LD therapy has been associated with increased calcium excretion, bone loss, and a potential increased risk of bone fractures (Lim et al. 2008).

A study of a regional population in England conducted in 1995 estimated that 2-3% of the entire population, and perhaps 3-4% of adults (higher in older adults) received recurrent prescriptions for LD (Clarke et al. 1995).

Other Diuretics

Thiazides (Trullàs et al. 2022), acetazolamide (Mullens et al. 2022), MRA (Cleland et al. 2020), and SGLT2i (Voors et al. 2022) also have diuretic properties, though modest and are generally insufficient to treat clinical congestion by themselves. For patients with severe congestion, these agents may be combined with LD to enhance a LD induced dieresis. Further research is required to determine if these medications are effective alternatives for preventing the reappearance of congestion. However, due to their diuretic effect, they might be able to manage sub-clinical congestion (i.e., an asymptomatic rise in natriuretic peptides).

1.6 Prognosis of Heart Failure

While mortality rates have improved for younger and middle-aged HF patients, HF is associated with high levels of morbidity and mortality (Conrad et al. 2018, 2019). Comparative analysis of five epidemiological studies suggested that HF patient prognosis follows a bi-phasic pattern, with higher mortality in the first six months (Khand et al. 2000), presumably reflecting a combination of the severity of HF itself and the effect of the final insult that caused the slide into

HF. The prognosis of those that survive the initial onslaught of HF is only marginally better with a 5-year mortality around 50% (Hobbs et al. 2007, Mahmood et al. 2014). Prognosis will vary according to the patient's age, aetiology, severity of cardiac and renal dysfunction, comorbidities, and the treatments the patient receives (Mosterd & Hoes 2007, McDonagh et al. 2021). Commodities such as COPD (Hawkins et al. 2009, Ehteshami-Afshar et al. 2021) and anaemia (Graham et al. 2020, Anand & Gupta 2018) are associated with a worse prognosis, while hypertension (Raphael et al. 2009) and moderate obesity (Clark et al. 2014) are associated with a better prognosis.

1.6.1 Prognosis of Heart Failure Treated with Loop Diuretics

Patients with HF and congestion have a worse prognosis, whether congestion is clinical, or evidenced by raised plasma concentrations of natriuretic peptides, or dilated LA (Faris et al. 2012, Guglin 2012, Cleland et al. 2021). The main treatment for controlling congestion is a LD, and therefore, it is no surprise that their use is associated with a worse outcome (Pellicori et al. 2016, Okumura et al. 2016, Domanski et al. 2003). The treatment with a LD is a sign of more severe congestion and advanced disease (Lim et al. 2008, Pellicori et al. 2016). There is little doubt that diuretics provide immediate benefits with regard to congestion (Eshaghian et al. 2006). Some evidence from RCTs suggests that LDs relieve symptoms in the longer term (Faris et al. 2012).

Results from observational studies suggest that potassium-sparing diuretics are associated with a reduced risk of death from HF (Domanski et al. 2003). And a RCT of spironolactone, a MRA that reduces potassium excretion, showed an improvement in mortality (Cleland et al. 2020). However, it was attributed to other properties (e.g., reduction of blood pressure, reduction of congestion, and possible anti-fibrotic effects).

However, thiazides also cause potassium wasting but were associated with a reduction in any HF, stroke, death from cardiovascular causes, and all-cause mortality (ACM) in a large randomised trial of elderly hypertensive individuals (Beckett et al. 2008).

While the long-term effect of diuretics on HF progression is unclear (Eshaghian et al. 2006), they seem to be the metaphoric 'canary in a coal mine'. The use of a diuretic and the type of diuretic being used indicate the severity of a HF diagnosis (Pellicori et al. 2016). For this reason, a patient's prescription history should be considered a 'low tech' tool for identifying patients who might have HF.

1.7 The Loop Diuretic - Heart Failure Interface

1.7.1 Loop Diuretic and Heart Failure in Primary Care

In a series of studies using primary care data out of Nottinghamshire, one study encompassing six general practices serving a total of 22,000 patients aged >30 years found that 505 were receiving LD prescription (2.3%) and of whom only 56% had a diagnosis of HF (Clarke et al. 1994). Of those who were receiving LD without a diagnosis of HF, less than half had relevant investigations. The same authors also looked at prescribing data for Nottinghamshire (a population of almost one million) and found that the equivalent of just over one million milligrams of furosemide (or 26,214 40 mg tablets [which is the most common dose]) was prescribed each day to an estimated 32,510 patients or about 3.3% of the population, and only about half were thought to have HF (Clarke et al. 1995). In a more recent series of reports from a primary care population of 60,728 patients, 1,301 patients (2.2%) were prescribed LD (66.1% were women), but only 264 (20.3%) had a diagnosis of HF (Sparrow et al. 2003*b*). Additionally, 238 (18.3%) patients reported ankle swelling and 110 (8.5%) reported exertional breathlessness. These patients on a LD were subsequently invited for a full clinical assessment, and only 737 agreed to an assessment, with 621 completing the analysis (Sparrow et al. 2003*a*). Of these 621 patients, 314 (51%) had an LVEF <40% with similar percentages in those with or without a diagnosis of HF. Of those with an LVEF $\geq$40, the median BNP was 42 (1^{st}-3^{rd} quartile: 32-79) ng/L, indicating that most patients had values close to the ESC HF guidelines ($\geq$35 ng/L) (McDonagh et al. 2021), though they were not markedly elevated. In comparison, the median BNP for patients with an LVEF <40% was 58.5 (38-97) ng/L, which are only slightly higher than those with the higher LVEF. These data leave considerable uncertainty with regard to a diagnosis of HF. That being said, no attempt was made to withdraw LD, nor re-assess BNP after initiation (see Figure 1.1).

Cuthbert et al. found that about 2.9% of people within a primary care practice (with a population of about 9,300 patients) were taking LD, but only about a quarter of these patients had a diagnosis of HF, which rose by 24 patients (about 50%) after patients records were reviewed (Cuthbert et al. 2020). During follow-up, the incidence of HF at two years for patients with a LD in the absence of a HF diagnosis at baseline was 13%, and ACM was 23%.

In Barcelona, a cross-sectional study of two primary care centres (a population of 39,000 individuals) identified 595[4] patients (1.5%) with a diagnostic label of HF, of whom 321 (53.9% of those with HF; 0.8% overall) were on a LD (Verdú-Rotellar et al. 2017). However, the diagnosis of HF was confirmed in only 319 (54%) of the total identified HF population. Of those with a confirmed diagnosis of HF, 207 (64.9%) were on a LD. The number of patients on LD without

[4]An additional 21 patients were identified with a diagnostic label of HF, but were excluded from analysis due to censoring.

a diagnosis of HF was not reported.

In summary, these observational studies suggest that LD are often prescribed in the absence of a clear diagnosis of HF. Additionally, in primary care, the diagnosis of HF is often suspect. Additionally, LD may be associated with an adverse prognosis.

1.7.2 Loop Diuretics and Heart Failure in Populations with Atrial Fibrillation

AF commonly precipitates symptoms and signs of congestion, for which patients are treated with LD, but many of these patients will not have a recorded diagnosis of HF.

An analysis of data from two RCT of a new anticoagulant for patients in AF found that 49% of the patients either had documented LV dysfunction or were taking LD (Cleland et al. 2007). Still, only 25% had both. Patients with both LD and LV dysfunction had a worse outcome than those with LV dysfunction who were not on LD.

Using the Clinical Practice Research Datalink (CPRD), which predominately covers England, Zakeri et al. found similar percentages (18%) of patients with AF also were on LD and did not have a diagnosis of HF, and a further 13% had both a diagnosis of HF and were on LD (Zakeri et al. 2021). The prognosis of patients with AF on LD was similar whether or not they had a diagnostic label of HF. Of the patients on LD without a diagnosis of HF, 60% were women.

1.7.3 Loop Diuretics and Heart Failure in Populations with Type-2 Diabetes Mellitus

There is a growing awareness that HF is a common cardiovascular (CV) complication of type-2 diabetes mellitus (T2DM) (McAllister et al. 2018, Dunlay et al. 2019). However, many patients receive a LD without a diagnosis of HF.

The Diabetes Collaborative Registry in the United States of America reported that out of the 1,322,640 patients with T2DM, 225,125 (17%) were on a LD, and of whom 91,969 (41%) had a diagnosis of HF (Arnold et al. 2020). However, 14% of the patients with documented LV function on LD without a diagnosis of HF had an LVEF <50%, suggesting a missed diagnosis. Other reports from this registry found that 26% of the patients had a diagnosis of HFrEF and only 64% were on a LD (Arnold et al. 2019).

A RCT of SGLT2i within patients with T2DM found that the outcomes for people taking LD were worse than for those who were not, with or without a diagnosis of HF (Pellicori et al. 2021).

1.7.4 Loop Diuretics and Heart Failure

In clinical trials for HF, the use of LD is associated with worse outcomes, and it is a stronger marker of prognosis than either a history of of HF hospitalisation or natriuretic peptides (Coiro et al. 2021, Damman et al. 2016). A meta-analysis of observational studies that included 96,959 patients with a diagnosis of HF found that the LD was associated with increased rates of ACM and HF hospitalisations (Kapelios et al. 2022). An analysis of dilated cardiomyopathy patients found that the use of LD was a strong predictor of adverse outcomes (Nuzzi et al. 2023). Pellicori et al. investigated whether LD were associated with an adverse outcome after adjusting for natriuretic peptides and echocardiographic assessment of congestion and concluded that LD use reflected the severity of congestion (Pellicori et al. 2016). In a propensity score-matched study of hospitalised HF patients without prevalent LD, the initiation of LD was associated with increased risk of mortality in the absence of congestion but was protective when congestion was severe (Faselis et al. 2021). Similarly, Testani et al. reported that in patients with HFrEF and LD, the prognosis was better when blood urea nitrogen was normal and decreased when blood urea nitrogen was elevated (Testani et al. 2014). These findings reinforce the need to be careful when prescribing LD.

1.7.5 Withdrawing Loop Diuretics

An anecdotal study of haemodynamic monitoring suggested that diuretic withdrawal often leads to a rise in intra-cardiac pressures and worsening congestion (Braunschweig et al. 2002). However, many patients may be able to stop LD for long periods of time if their HF is well controlled, and their plasma concentration of natriuretic peptides (measuring congestion) is fairly low (Galve et al. 2005, Romano et al. 2017).

Clinical trials have found that the symptoms and signs of congestion worsen when LD are withdrawn from patients with HF (Carter Grinstead et al. 1994, Damman et al. 2016, Dovancescu et al. 2017, Dauw et al. 2021). A RCT of LD withdrawal in patients on LD therapy in primary care without a diagnosis of HF found that the withdrawal often led to worsening congestion, most frequently due to the development of symptoms and signs for HF (Walma et al. 1997).

1.7.6 The Furosemide Test

Kelder et al. conducted a study to test the hypothesis that a trial of furosemide could help GPs distinguish between people with and without HF[5], where there is considerable diagnostic uncertainty (Kelder et al. 2011). They concluded that while the rationale for this test was obvious, the results do not support its use as a diagnostic tool. However, they found that NT-proBNP dropped

[5]This was based on the ancillary diagnostic described in the Framingham Heart Study (McKee et al. 1971), where patients needed to lose $\geq$4.5 kg within 5 days due to diuresis induced by LD therapy.

substantially after initiation of the LD and was still statistically significant after univariate testing. Additionally, most showed symptom improvement regardless of whether or not the primary care physician ultimately gave them a diagnostic label of HF. Ultimately, this study provides further evidence of the inadequacy of subjective medical opinion.

1.8 Thesis Outline

1.8.1 Overall Aim and Objectives

The overall aim of this thesis is to investigate the pharmacoepidemiology of LD dispensing and its relationship to the diagnosis of HF and to prognosis. Specifically,

1. Describe the prevalence of people receiving LD therapy and/or a diagnostic label of HF and their characteristics, comorbidities, concurrent medications, hospitalisations, and mortality compared with patients with a broad range of CV disease (predominately hypertension and CAD).
2. Describe the pattern, nature, and frequency of hospital admissions in the year before initiating LD therapy or receiving a new diagnosis of HF, and the changes in comorbidities and concurrent medications that occur before and after these incident events.
3. Determine the sequence of 'events' with respect to the initiation of LD therapy or a new diagnosis of HF and their relationship to mortality.
4. Investigating the relationship between the presence of LD therapy and a diagnosis of HF with mortality after adjusting for age, sex, socioeconomic status, and concurrent comorbidities in patients known to have IHD, a group of patients who much more likely to have myocardial damage and dysfunction.

1.8.2 Thesis Structure

This thesis is structured to give a general introduction of the problem, and an explanation of the available data, before delving into the analyses and then summarising with a final discussion and concluding remarks. Chapter 1 provided a general introduction to the clinical syndrome of HF, including the challenges and uncertainties regarding its diagnosis, the central role of congestion, and why LD usage is a potential pharmacoepidemiological marker of congestion. Chapter 2 provides an explanation of the available data, including a background into using routinely collected healthcare data for research, what healthcare data were used for this thesis, and how the data were prepared for analysis. Chapter 3 is the first of the four analysis chapters and addresses Objective 1 by describing the prevalent levels of HF, LD usage within a regional population, including patient characteristics and five-year morbidity and mortality rates. Knowing what the

chronic population is, Chapter 4 'moves backwards' in the patient's timeline to address Objective 2 by describing the events and changes leading up to the diagnosis of HF or the initiation of a repeat LD prescription. Chapter 5 addresses Objective 3 and describes the temporal relationship of incident diagnosis of HF and/or the initiation of LD therapy on mortality. The final analytical chapter, Chapter 6, addresses Objective 4 by describing the population with a history IHD according to a diagnosis ofHF or repeat LD, and these patients outcome.

Chapter 2

Data

> I have no data yet. It is a capital mistake to theorise before one has data.
>
> *Sherlock Holmes*
> *Arthur Conan Doyle*

2.1 Introduction

Data are the foundation for research, as they define the questions that can be addressed. Within the medical, statistical, computational, and epidemiological fields, trends are moving toward using larger and more heterogeneous data sources to the point where 'big data' are commonplace and routine within research. According to the Oxford English Dictionary, references to an 'information explosion' have been around since 1941 (Oxford English Dictionary 2008*b*), while the term 'big data' in reference to computing first appeared in 1980 (Oxford English Dictionary 2008*a*) and gained popular attention after Roy Williams recognised in 2003 that 'big data' are a gold mine of information, and not just a pile of tapes [or records] (Williams 2003). 'Big data' refers to the increasing volume of data arriving at ever faster velocities and containing a greater variety of data (Laney 2001), also known as the three 'Vs'. Due to the larger and more complex datasets, especially from a variety of new data sources, the manipulation and management present logistical challenges. Since 2003, the number of sources, types, and complexity of data have increased dramatically. This is particularly evident within the public health community due to the introduction and growth of transactional databases providing detailed longitudinal records of care and outcomes of patients, including electronic patient records (EPR) created by routine functioning of healthcare systems, such as data collected in electronic health records (EHRs) and administrative datasets (NHS Research Scotland 2023, Franklin & Schneeweiss 2017).

2.2 Electronic Patient Records

Both EHRs and administrative data are sources of EPR commonly used in epidemiologic research. In the case of EHRs, the data are generated and recorded during routine, everyday clinical care (Denaxas & Morley 2015). The data can be diverse in structure and source. In terms of composition, records will either be in a structured format, such as diagnostic codes, or an unstructured format, such as clinical notes or image data. Data can be integrated from a plethora of sources, including, but not limited to, data from primary and secondary care, blood tests, other investigations, and procedures. Recently, genomic data and information from wearable technology such as the Apple Watch or Fitbits have started to be integrated into records. In the case of administrative data, the data are generated, unsurprisingly, for administrative purposes, primarily for billing or insurance claims, legal, healthcare utilisation, census, or vital records data (Rothman et al. 2021).

In both cases, the medical ontology[1] regarding the controlled medical vocabulary and concepts and their relationship to the EPR data, including what is recorded, how it's recorded, and how it's transmitted are vital to informing research questions and outcomes. The path from raw EPR to publishable results is far from linear.

2.2.1 Routinely Collected Data

Going from raw EPR to a 'research-ready' dataset starts with understanding the benefits and pitfalls of working with data which were not specifically collected for a given research question. For epidemiologic research purposes, EPR data fit the definition of secondary data, as they mostly constitute transactional information collected for purposes other than research (Rothman et al. 2021). As EPR are a by-product of routine healthcare, a record exists because the patient needed and sought care.

EPR are unlike clinical registries and trials in many aspects. In a registry or trial, individuals who meet a set of inclusion and exclusion criteria are invited to participate by researchers with a particular interest in a condition. The individuals must be known to the researchers directly or indirectly through their health records unless a public campaign seeks volunteers. Participation in a registry or trial usually requires patient consent, and a complete set of data is collected on a specified date, with follow-up visits at regular intervals. The result is a well-curated dataset for analysis, but each step introduces bias.

In contrast, retrospective observational research using routinely generated EPR in a public health system can be more representative of the general population, and generally does not require an individual's consent to be included. There will still be biases, especially for conditions where

[1] In a modern philosophical context, *an ontology* is a theory of what exists (i.e., what data can exist and the rules that govern them) (West 2011).

there is no robust, objective definition (such as HF), where diagnosis depends on healthcare professionals' skills, opinions, and record-keeping. Deciding on a baseline for the start of follow-up is different than for a registry or clinical trial. For research using EPR, the baseline is determined by an event or date, and having a complete set of contemporary data points is very unlikely. Records and tests during follow-up will occur at irregular intervals according to the changing needs of the patient.

The availability and coverage of records depend on the healthcare system. In the United States of America, the presence of an individual's EHRs, or lack thereof, could be attributed to many factors, including good health, personal finances, insurance coverage, or an individual's physical location (e.g., proximity to a specialist, accessibility of resources). Research has been conducted using routinely collected data ranging from a single hospital department (Hsu et al. 2011) to an entire country, such as with Clinical Practice Research Datalink (CPRD) (Denaxas et al. 2012). While data from a single department or hospital will likely have highly detailed, timely, and accurate patient information, the EHRs will not include records pertaining to care received outwith the hospital, which might be required to investigate the antecedents and progression of diseases leading to a referral. The benefits of using routinely collected data from regional or national environments include access to more extensive, unselected sections of the population, albeit with potentially less structured and possibly less granular data (e.g., these datasets rarely contain data on quality of life, or in some cases even details on symptoms or physical examination) (Lash et al. 2014).

2.2.2 Routinely Collected Data in Scotland

NHS Greater Glasgow & Clyde Health Board (NHS GG&C) is well suited for research using large-scale, longitudinal EPR based medical research. Approximately 5 million people live in Scotland and are served by a single, unified health system (National Health Service Scotland [NHS Scotland]) (NHS Research Scotland 2012*b*). About 23% of the population of Scotland resides in NHS GG&C, which has high levels of socioeconomic deprivation (about twice the national average) (National Records of Scotland 2018). NHS Scotland provides most services free of charge, including prescriptions, primary care consultation, investigations, hospital care, medical procedures, and emergency services (NHS Inform 2022). Because there is no payment at the point of service, costs and insurance coverage are not a barrier to patients receiving care, and hence to the completeness of EPR. By removing payment at the point of service, many of the reasons for missing or absent EPR records due to costs or insurance coverage are removed. Additionally, Scotland, and especially Glasgow, has particularly high incidence rates for diseases including heart disease, stroke, and cancer (Walsh et al. 2017).

Scotland has a long history of EPR captured from birth through death using individual Community Health Index (CHI) numbers. CHI numbers allow for the unique identification and tracking

of patients across NHS Scotland's services (NHS Digital 2022*a*). The CHI number is the Scottish equivalent to England and Wales's NHS number. CHI numbers are assigned to each patient upon first registration with the system (NHS National Services Scotland nd*a*). CHI numbers are ten digits long, with the first six digits taken from the date of birth (DOB) in two-digit format (`DDMMYY`), two random digits, a sex-based digit (i.e., even for women and odd for men), and an arithmetical check digit (NHS Digital 2022*a*).

2.2.3 Data Safe Havens

NHS Scotland has a long tradition of linking and using health service data for research to improve patient outcomes, measure long-term outcomes in clinical trials, assess the safety of new medical interventions, and support the understanding of patterns in health and illness across whole populations. Within the context of Scottish NHS EPR, Data Safe Havens form an integral part of Scotland's health informatics capabilities where it is not practicable to obtain individual patient consent for participation (The Scottish Government 2015). A Safe Haven is a secure research environment supported by trained staff and information governance processes where EPR can be linked with other data and made available in a de-identified form for analysis while protecting patient identity (NHS Research Scotland 2012*a*, The Scottish Government 2015, NHS Research Scotland 2023). Safe Havens are structured such that the National Safe Haven contains information for all residents within Scotland, and four regional Safe Havens support it:

- Health Informatics Centre (HIC) - NHS Tayside with the University of Dundee
- DataLock - NHS Lothian with the University of Edinburgh
- West of Scotland Safe Haven - NHS GG&C with the University of Glasgow (recently formed from the Glasgow Safe Haven).
- Grampian Data Safe Haven (DaSH) - NHS Grampian and the University of Aberdeen

The regional Safe Havens work independently in full compliance with relevant codes of practice, legislation, and statutory orders in accordance with current professional practice. They are responsible for providing information about patients residing within their territories (The Scottish Government 2015). Together, these Safe Havens form a federated network to work together to support research across Scotland.

Dataset Classifications

Whether it's the National Safe Haven or a regional Safe Haven, data sets available for linkage are classified into three tiers based on area coverage and data generation source. Tier 1 datasets are the most curated of the three tiers. The data are collated at a national level and contain information from everyday care, such as community-based prescriptions and hospitalisations

(Watson 2020). Following on, Tier 2 datasets are generated locally or regionally to help advance that location's services and to benefit their patients, such as programs to enhance general practice outcomes. Finally, Tier 3 datasets are generated by individual research projects conducted within the Safe Haven and are not currently available for request (Watson 2020).

Data Linkage and Anonymisation

As was alluded to earlier, the universal usage of CHI numbers allows for patient records to be linked from registration through to death or migration out of Scotland. Each Safe Haven provides a data linkage and de-identification process in strict compliance with guidelines set out in the Safe Haven Charter (The Scottish Government 2015). The charter states that linked datasets should be kept only for the minimum time necessary to meet the original purpose of data linkage. Additionally, EPR should have direct identifiers (e.g., name, CHI number, date of birth, address) removed where it is not practicable to obtain consent. This process should occur as soon as reasonably practicable. Data available within the Safe Haven analytic platforms must not contain personal identifiers (The Scottish Government 2015). Only approved researchers are allowed to access de-identified data on the analytic platforms.

2.3 West of Scotland Safe Haven

The Glasgow Safe Haven, a partnership between the Robertson Centre for Biostatistics at the University of Glasgow and the NHS Greater Glasgow & Clyde Health Board (NHS GG&C), was recently expanded to become the West of Scotland (WoS) Safe Haven, one of the four regional Safe Havens within Scotland. WoS includes NHS GG&C, Lanarkshire, Ayrshire & Arran, Dumfries & Galloway, and Forth Valley, but not all regions contribute the same amount of granular data in the same format. Accordingly, the analysis focused solely on NHS GG&C datasets in this thesis.

WoS Safe Haven provides a secure platform for researchers, and other approved users, to access de-identified EPR including NHS EHRs and administrative data relating to residents living within NHS GG&C (Watson 2020). The NHS GG&C Health Board serves the people of East Dunbartonshire, East Renfrewshire, Glasgow City, Inverclyde, Renfrewshire, and West Dunbartonshire (NHS Greater Glasgow & Clyde Health Board 2021), containing about 23% of the entire Scottish population (National Records of Scotland 2018).

The WoS Safe Haven provided the data for NHS GG&C for this project. Ethical approval was sought and obtained to access a cohort of patients treated for cardiovascular disease, ranging from hypertension to HF, to describe the prevalence and incidence of LD dispensing, of HF and the combination (LD and HF), relevant patient demographics, characteristics, and common comorbidities, treatment patterns, and outcomes.

2.3.1 Ethical Approval

Ethical approval and research governance for research carried out on the WoS Safe Haven are managed via a Local Privacy Advisory Committee to protect patient privacy and confidentiality. Ethical approval was granted with the project identifier GSH/18/CA/002.

2.3.2 Cohort Build

The GSH/18/CA/002 cohort consists of all individuals with a record in Scottish Morbidity Records - General/Acute Inpatient and Day Case (SMR01) or Scottish Morbidity Records - Mental Health Inpatient and Day Case (SMR04) for CAD, PAD, or HF; individuals with a record in General Practice Local Enhanced Services (GP LES) for PAD, CAD, LVSD, or HF; individuals in thenNHS GG&C's existing HF database (ATHENA); or individuals with a record for a medication commonly associated with the treatment of these diseases or hypertension (which is very common and a powerful risk factor for HF) in the Prescribing Information System (ACEi, ARB, beta-blocker, MRA or loop diuretic) on or after 31st December 2009 with full coverage ending on 31st March 2018. Codes are listed in Table A.1.

2.3.3 Data Sources and Specifications

The GSH/18/CA/002 dataset consists of all records available from patients identified by the cohort build (see above) from the following sources: Demographics, Deaths, Scottish Care Information Store (SCI Store), General Practice Local Enhanced Services (GP LES), Prescribing Information System (PIS), Scottish Morbidity Records - Outpatient Attendance (SMR00), SMR01, SMR04, Scottish Care Information-Diabetes Collaboration (SCI Diabetes), MUSE ECG, MUSE ECG Matrix, Xcelera, and EchoPAC. With the exceptions of SCI Store and SCI Diabetes, data were acquired using the source's default parameters. SCI Store's default dataset was limited to the biochemistry and haematology disciplines. SCI Diabetes's default dataset was extended to include data in the `dataitemids` '178', diabetes mellitus (DM) type, and '335', date of a DM diagnosis. Additionally, this is the first large-scale project to link NHS GG&C ECG (MUSE ECG and MUSE ECG Matrix) and electrocardiography (Xcelera and EchoPAC) to the default data packages.

2.4 Data Cleaning and Preparation

The cohort comprised >360,000 patients spanning 8 years. For such a large dataset, data-formatting, preparation, and cleaning are essential. These steps are an iterative and time-intensive process of exploration and experimentation (Lohr 2014). In particular, data cleaning is where the raw data are checked for accuracy, consistency, and completeness. This includes careful

scrutiny of the raw data for outright errors and correction of errors where possible (Rothman et al. 2021).

In conjunction with data cleaning, the creation of a research-quality dataset entails refining which data are used and how they are structured in order to facilitate answering the research question(s) (Leek 2019). Particular attention must be paid to the required data format by downstream tools; level of required data granularity; ease of manipulation and use; validity and accuracy underlying the data collection; and documentation of potential biases (Leek 2019). This process occurs in tandem with combining the relevant and required data points needed to answer said question(s).

Three main tools were used to assist with data cleaning and preparation of the GSH/18/CA/002 dataset. These tools were: i) a relational database, ii) statistical software, and iii) a scripting language for network graph preparation.

2.4.1 Relational Databases

A relational database was designed and implemented in order to clean, structure, subset, and format raw EPR into 'research ready' tables for statistical analysis. The EPR made available to researchers come in comma-separated value (CSV) files. Each file contains information from a separate data source (i.e., death records, community-based prescriptions, laboratory records, hospital admissions). Many files contain information on numerous variables, such as blood results for haemoglobin and serum creatinine, each with a different standard unit and 'normal' range. The research database was designed and structured so that each Table contains one set of homogeneous records (e.g., haemoglobin tests) where records can be uniquely identified. Tables were linked without redundant information (Halpin & Morgan 2008). See Section A.2 for an explanation of relational databases.

SQL Server

For the scope of work conducted for this body of work, data were prepared for analysis using Structured Query Language (SQL) Server Management Studio version 17.8.1 running Microsoft SQL Server Management Studio version 14.0.17277.0.

Data Modelling

The SQL database was structured using schemas, which define how data are organised within a relational database. They are commonly used to communicate the architecture of the database, and the process of designing the database schema is known as data modelling (Halpin & Morgan 2008).

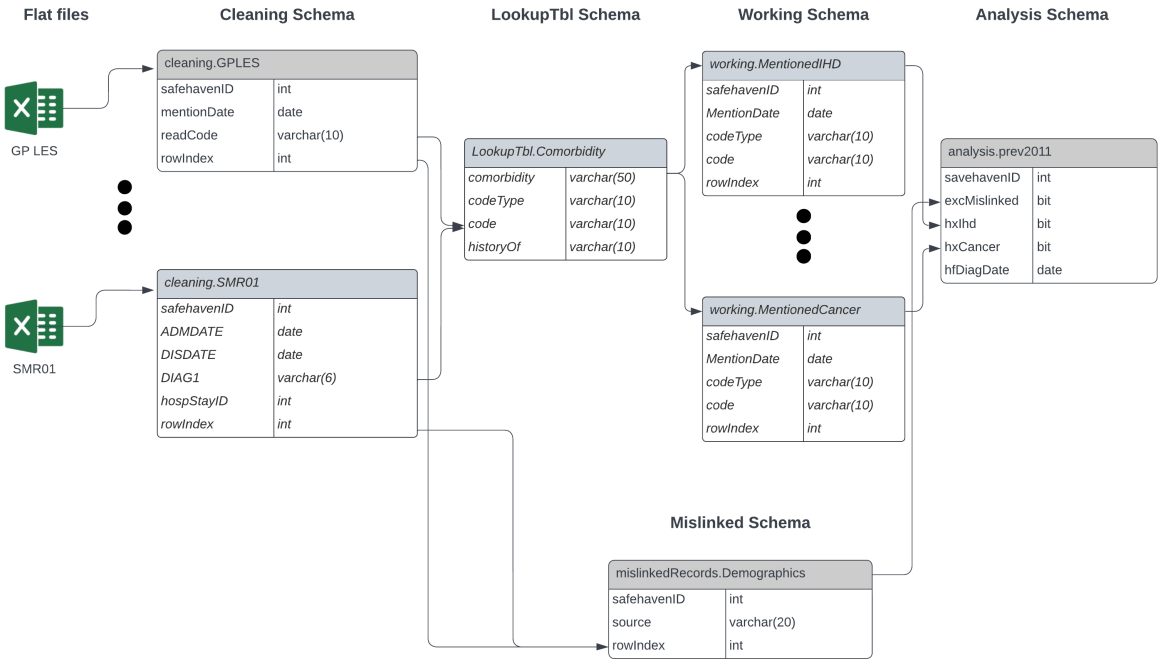


Figure 2.1: An illustration of how the six schemas interlink to create tables ready for input to R or Python.

To that aim, the dataset was organised using six schemas: `staging`, `lookupTbl`, `cleaning`, `working`, `mislinkedRecords`, and `analysis` (see Figure 2.1). As an overview, the tables in `staging` hold the data as it arrives as the flat files provided by Safe Haven. These tables assume that all columns contain `varchars`, a variable length series of numbers, letters, or characters, to limit forced or incorrect data conversion errors. The `lookupTbl` schema holds lookup tables that define various disease definitions, drug classifications, or measurement groupings. The `cleaning` schema refers to the cleaned data presented in the `staging` schema. The tables in `working` are long format tables that hold information about a given test, disease, or medical history. The `mislinkedRecords` schema holds records which are classified as mislinked and are to be ignored. Finally, the `analysis` tables hold data which have been developed for input into R for abstracts, papers, reports, and presentations. Going into detail per schema: in the `cleaning` schema, columns that have been correctly formatted (e.g. numeric values are now numeric and not strings, bit columns are now bits and not varchars, etc.). Unique row indexes (`rowIndex`) have been added to all tables, which remain with the record into `working` tables. Row indexes are unique within each Table but are reused between tables. Finally, columns have been added that apply to all values in each table. In cases where an event date and time occurred are stored in one column, these were split into separate date and time columns. Many of the tables will have an `included` column, which indicates that a record should be excluded from future analysis for various reasons (e.g., impossible dates, null values, or numeric overflow errors).

Tables within the working schema hold all records pertaining to its particular condition or measurement. To create these tables, records from the cleaning schema were parsed into long format tables with the help of the lookup tables in the lookupTbl schema. Tables that begin with 'Mentioned' hold all coded references of that disease. Tables in the working schema that do not begin with 'Mentioned' either hold all lab records for the named test or records for the named echocardiography measurement. For example, the working.MentionedIHD Table holds all coded references of IHD, while working.Haemoglobin holds all haemoglobin test records regardless of anaemia status.

The analysis schema holds tables which are ready to be imported into the statistical software for analysis and visualisation. These tables are used to prepare presentations, abstracts, exploratory analyses, or papers. Unless names are appended with _tdc, tables are formatted to have one row per patient. Where names are appended with _tdc, tables are formatted where there might be multiple rows per patient depending on the number of events of interest (see Section 3.2.1 for an explanation of time-dependent covariates). Data held analysis tables were imported from both the cleaning and working schemas.

2.4.2 Statistical Software

Statistical analyses were performed using R.

R is a free, open-source programming language for data analysis, statistics, and data visualisation (Teetor 2011). R version 4.0.5 (R Core Team 2020) was used to carry out the statistical analysis. In addition, packages used throughout the analysis include the tidyverse ecosystem (Wickham et al. 2019), survival (Therneau & Gramsch 2000), viridis (Garnier et al. 2021*b*), lubridate (Grolemund & Wickham 2011), and RODBC (Ripley & Lapsley 2021).

2.4.3 Scripting

Python is a free, open-source scripting language, and was used as the scripting language of choice. In brief, it was used to manipulate hospital admission data into ordered pairs of admissions, per patient, for input into graphing software. This was done within Spyder version 3.3.6 (Raybaut 2009), an open-source scientific development environment for Python, using the 64-bit Python version 3.7.4.

2.5 GSH/18/CA/002 Dataset

As referenced in sections 2.3.1 and 2.3.3, the project used for this thesis was approved under the reference number GSH/18/CA/002. The following section includes a breakdown of information contained within each linked data source and information pertaining to data cleaning.

2.5.1 Deaths

The deaths file is a Tier 1 dataset containing combined records of death from the General Register Office, sourcing data from National Records Scotland (NRS) deaths, Acute, Cancer, Deaths and Mental Health (ACaDME), and CHI records. With the exception of death records obtained from CHI, each record contains information including date of death (DOD), location of death, the underlying cause of death (COD), and space for up to 10 contributing COD. Since 1st January 2000, CODs are coded in accordance with International Classification of Diseases, 10th Revision (ICD-10) (National Records of Scotland 2017) (see Section 2.5.1 for an explanation of ICD-10). In the case of death records obtained from national CHI records, only the DOD was provided.

Defining the Canonical Date of Death

In addition to the deaths file, deaths were recorded in SMR01, SMR04, and the demographics file. The algorithm used to define a patient's DOD is displayed in Figure 2.2 and is as follows: if only one record of death exists per `safehavenID`, use that record as the definitive date of death. If more than one record of death exists, check to see if the death source is from NRS Deaths. If it is, use this record. If a record was provided in the Deaths dataset, use that record. Finally, if no record was provided by the Deaths dataset, use the most recent record of death from SMR01, SMR04, or the demographics file as the canonical DOD (see Section A.3 for code).

Identifying Mislinked Records

Individuals were classified as mislinked and were excluded from analysis where significant amounts of activity continued long after the canonical date of death as defined above (see Section 2.5.1). Records of clinical activity continuing long after a patient's recorded death occur either because the patient is still alive and the death record is false or because the death record is correct and more than one individual has been linked to the same `safehavenID`. In either case, Safe Haven does not provide enough personally identifiable information needed to untangle which case is correct; therefore, all instances where this happened were excluded.

Significant amounts of activity long after the canonical date of death were defined as having at least five days of activity more than 60 days after death from any data source. The five-day activity threshold was chosen in order to allow for the filing or processing of records after death from multiple data sources. For instance, data are correctly linked to `safehavenID` in cases where investigations or prescriptions were made prior to death, but results were not available, or pharmacies initiated reimbursement procedures after death (see Section 2.5.4 for further information). The 60-day threshold was chosen to allow for two prescription billing cycles, as dispensing days are usually registered at the end of the month (see Figure 2.5 for

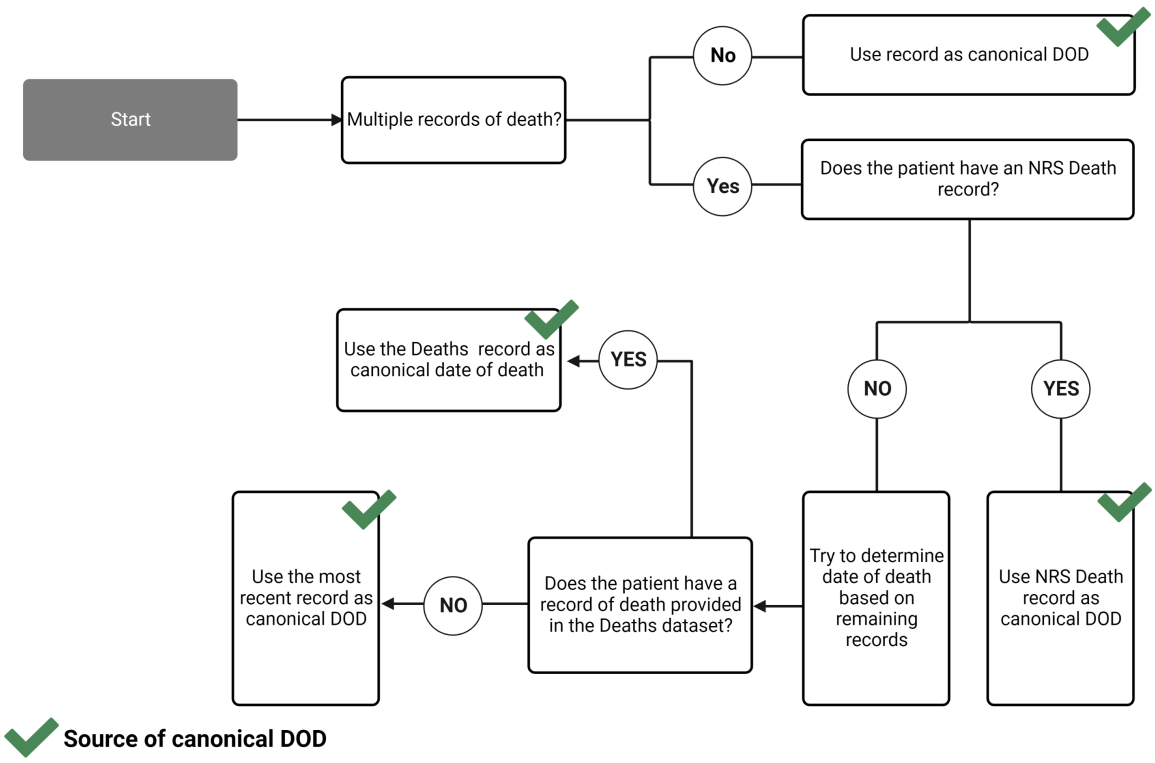


Figure 2.2: Diagram detailing the process of defining a patient's canonical DOD.

an example of dispensing patterns). Using this definition, 1,773 (0̃.5%) `safehavenIDs` were classified as mislinked and were excluded from analysis (code provided in A.4). This compares with the 70,008 validated deaths or 2% of all recorded deaths.

Cause of Death

Cause of death (COD) was available for most individuals with a death record. Within records with available COD, the underlying COD was recorded under `COD`. Within Scotland and the UK, the underlying COD is defined according to the WHO's definition as either the disease or injury which initiated the series of events leading directly to death or the circumstances of the accident or violence which produced the fatal injury (World Health Organization 2022*a*, National Records of Scotland 2019). If the certifying medical personnel cannot choose a single underlying COD, NRS uses the internationally agreed mortality coding rules in ICD-10 to select the underlying cause of death (Calderwood & Slater 2018). Additionally, up to ten contributory CODs may be recorded. These are listed in ascending order based on their location within the series of events leading to death, with the first recorded as `COD0` and the last recorded under `COD9`. Of note, the UK certificate advises against the coding of HF as a COD (Berlin 2009), as well as terms such as cardiac arrest, renal failure, liver failure, and shock.

International Classification of Diseases, 10th Revision

The International Classification of Disease (ICD) was originally a system to classify causes of death but has since expanded its scope to include non-fatal diseases, medical procedures, impairments, disabilities and handicaps (World Health Organization 2016). The 10th revision was adopted by the WHO in May 1990 and went into effect on 1st January 1993 (World Health Organization 2022*b*). More formally, the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) coding standard is a hierarchical standard provided by the WHO to enable systematic health recording and collection of statistics on disease in primary, secondary, tertiary care, and death certificates internationally and over time (World Health Organization 2022*b*). The codes translate potentially complicated medical diagnoses and other health problems into a finite set of alphanumeric codes, permitting easy storage and analysis (World Health Organization 2016).

Internationally, many countries have developed country-specific modifications to the WHO's version of the ICD-10 codes (Jetté et al. 2010). Universally, codes are at least three characters long, and the maximum can vary (World Health Organization 2016, Jetté et al. 2010). Within the UK, ICD-10 codes range between 4 and 6 characters long. The first character is a letter, following international standards, and the second two characters are always numbers, then a period followed by an alphanumeric character (NHS National Services Scotland nd*c*). In the case of a 3-character code, the UK fills in the fourth character with an 'x' (NHS Digital 2022*b*). If present, the sixth character is the dagger 'D' or asterisk 'A' indicator, though these can be present in the fifth position, where there are either modified 3-character or standard 4-character codes (NHS National Services Scotland nd*b*).

2.5.2 Demographics

The demographics data are collated from a collection of sources based on CHI numbers. The data made available within the dataset are acquired largely from NRS and records available to the NHS GG&C Safe Haven team. Demographics data that were used include date of birth (DOB), sex, and Scottish Index of Multiple Deprivation (SIMD) from 2012 by quintile.

Date of Birth

The canonical date of birth (DOB) was obfuscated by the NHS GG&C Safe Haven team. In the `'YYYY-MM-DD'` date format, DOBs were uniformly obfuscated by setting the day part of the date to be the middle of the month while maintaining the month and year values. For example, a birthday of 1922-01-09 would be changed to 1922-01-15.

Sex

The Demographics `sex` field was taken as the authoritative version for an individual's sex based on emails with NHS GG&C's Safe Haven team. Biological sex was used throughout this thesis as reported results refer to biological factors and do not identify, psychological, or cultural factors, which would allow reporting by gender.

Throughout this work, women were used as the reference category when sex was included in models.

Scottish Index of Multiple Deprivation

Scottish Index of Multiple Deprivation (SIMD) is an area-based measurement of socioeconomic deprivation assigned to residents of Scotland based on where they live. Scottish residents' SIMD 2012 status was calculated by the Scottish Government using thirty-one indicators from seven different aspects of deprivation: income, employment, health, education, housing, geographic access, and crime. The indicators are combined using a weighted sum to create a single index, providing a relative ranking for each small geographic area in Scotland. Areas average about 800 individuals (The Scottish Government 2012). It is important to note that SIMD can only measure an area's level of deprivation, not an individual's level. The absence of deprivation should not necessarily be correlated with affluence. The terms most deprived or least deprived were used to refer to the areas and not to the individuals living in those areas (The Scottish Government 2012).

Throughout this work, SIMD 2012 quantiles were used when referencing or adjusting for socioeconomic deprivation. When used in models, SIMD was treated as ordinal data due to its non-continuous nature (e.g., it is impossible to have a SIMD quintile rank of 3.5), and the least deprived quintile, SIMD 5, was used as the reference.

Ethnicity

While the Demographics file did not provide an authoritative record of ethnicity, the algorithm for defining deprivation is included here due to its logical association with demographics.

Ethnicity was recorded in multiple datasets, including the SMR datasets and SCI Diabetes. Each dataset has a different level of granularity (e.g., 'White' versus 'White - Scottish' or 'White - British'). In order to provide a level of standardisation across datasets, a patient's ethnicity was classified as either White, Black, Asian, Other, or Missing. Patient ethnicity was determined using the following steps.

1. Gather all references to patient ethnicity from the database, including the event date (i.e., `ADMDATE` or `DATE`) and `safehavenID`.

Grouped Ethnicity	Grouped Frequency	Singular Frequency
White	241,769 (66%)	White 241,769 (66%)
Missing	112,193 (31%)	Missing 112,193 (31%)
Other	10,823 (3%)	Asian 7,998 (2%) Black 1,037 (<1%) Other 1,788 (<1%)
Data are n(%).		

Table 2.1: Grouped ethnicity classification counts and relative percentages.

2. Group ethnicity into Asian, Black, other, White, and missing (see Table A.2 for list of codes).

3. Assign the most recently recorded ethnicity group 'other' where the ethnicity is known.

4. Assign all patients with no record of ethnicity as 'ethnicity missing'.

Ethnicity classifications were grouped in order to meet clinical governance requirements for the minimum number of patients.

2.5.3 General Practice Local Enhanced Services

Local Enhanced Services (LES) for general practice surgeries (GPs) is a service for which general practice surgeries receive additional payments for demonstrating a high-quality service for specific conditions including coronary heart disease, DM, stroke, chronic obstructive pulmonary disease (COPD), HFrEF (but not HFpEF), learning disabilities, and nationally enhanced services for drug misuse. Surgeries can subscribe to any number of the LES, without needing to cover every service. The General Practice Local Enhanced Services (GP LES) is a dataset which contains information about patients who received care under the LES scheme. Within NHS GG&C, 82.8% of surgeries provided data to GP LES (information based on a query to the West of Scotland Safe Haven Data Manager [Hamilton 2022]).

The General Practice Local Enhanced Services (GP LES) dataset is a Tier 1 dataset with coverage stopping at the end of 2018 (see Figure 2.12). Each GP LES record contains a `safehavenID`, the event date (`EventDate`), a Read code describing the entry (`READCODE`), a user-editable description to complement said code (`Description`), a flag for if the record pertains to a prescription (`IsPrescription`), a flag for if the record pertains to numerical values (`IsValue`), two value fields (`Value1` and `Value2`), the local enhanced service area (e.g., 3 for diabetes and 4 for HF) (`LESAreaID`).

Read Codes

Read Codes are a hierarchical controlled clinical vocabulary for terms and short phrases (Robinson et al. 1997, Pringle 1990, Chisholm 1990). The first widely used version of Read Codes was standardised to 4-byte set codes, which was then extended to a 5-byte unified set. Version 2 added a term code to hold an 'idea' or 'concept', where the preferred term appends '00' and additional synonyms append term codes 11-99 (Booth 1994). For example, if the original 5-byte Read Code was 'G30..' for acute myocardial infarction, the 5-byte version 2 code, with the preferred term code, is 'G30..00' for 'Acute myocardial infarction' and the first synonym, 'Attack - heart' for heart attack, is 'G30..11', followed by 'Coronary thrombosis', 'G30..12'. The NHS GG&C's GP LES dataset uses the 5-byte set of codes without the term code, which means synonyms are mapped onto the same five-digit code. For example, 'G580.00', 'Congestive heart failure', and 'G580.11', 'Congestive cardiac failure' both map onto 'G580.'. Additionally, trailing space holders (.) have been removed due to formatting errors or deliberate elimination. This means 'G580.11' maps to 'G580' and 'G58..00', 'heart failure' maps to 'G58'.

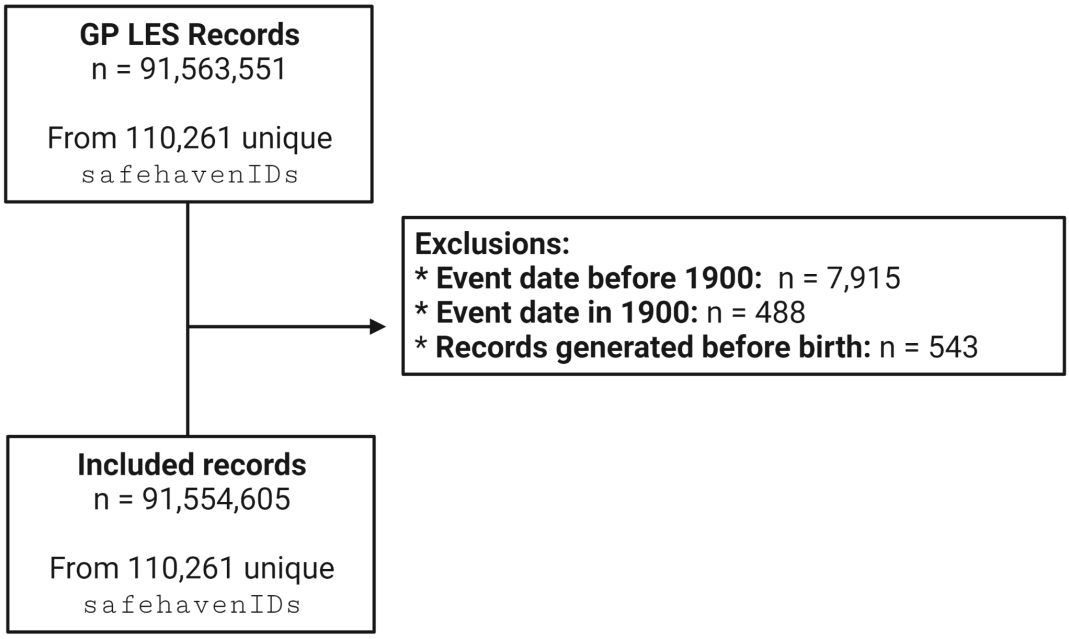


Figure 2.3: Flow diagram of GP LES records exclusion criteria

Cleaning General Practice Local Enhanced Services Data

In cleaning the GP LES data, records were excluded where the `EventDate`, the date attributed to the record, was clearly impossible or where the records were dated as originating before the patient's birth (see Figure 2.3).

2.5.4 Prescribing Information System

The Prescribing Information System (PIS) is a fairly unique resource that enables pharmaco-epidemiological research due to its population coverage and record linkage. PIS covers all NHS medications prescribed, dispensed and reimbursed in the community-setting within Scotland (Alvarez-Madrazo et al. 2016). Prescriptions written in hospitals and dispensed in the community-setting are also included in the dataset (Information Services Division Scotland 2022*a*). Of note, the NHS GG&C version of PIS only holds records of dispensed prescriptions. PIS uses the CHI number to link individuals prescribing and dispensing data to their other health records data since 2009, with a coverage that is almost 100% for prescribed and dispensed items (Alvarez-Madrazo et al. 2016).

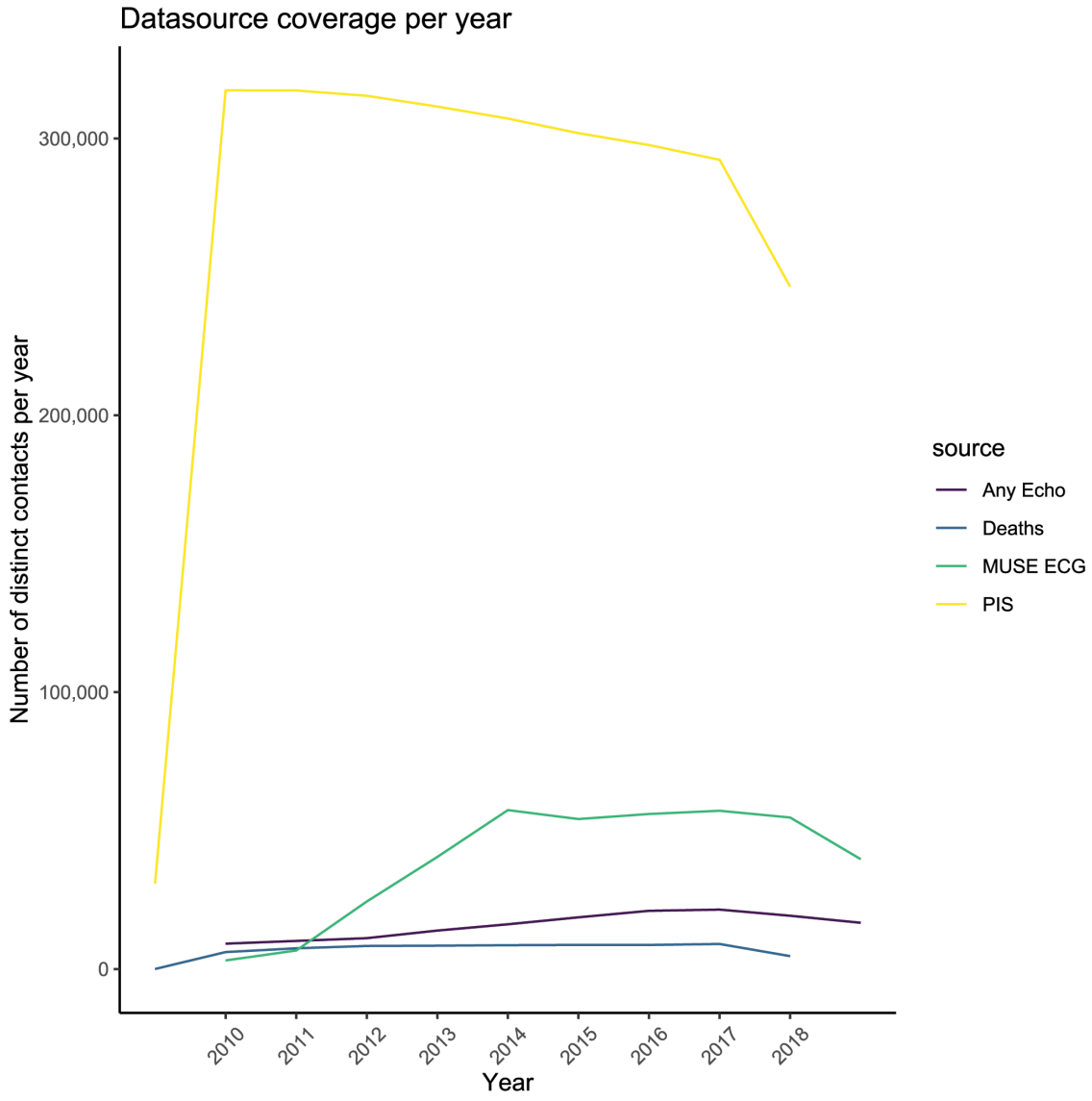


Figure 2.4: The number of distinct contacts per year for any echocardiography (Any Echo; see Section 2.5.7), deaths recorded in the Deaths file (Deaths; see Section 2.5.1), electrocardiograms (MUSE ECGs; see Section 2.5.6), and prescriptions (PIS; see Section 2.5.4).

For each reimbursed prescription, PIS provides the approved name, product name, formulation, and strength using the British National Formulary (BNF) chapter and item codes (see Section 2.5.4). Importantly, PIS does not provide information on how often a medication should be taken, how many pills should be taken at one time, nor at what time of day. Additionally, records do not explicitly record the reasoning or timing of when treatment was started, changed, or terminated (Williams, Brown, Peek & Buchan 2016).

Use of Prescribing and Dispensing Date

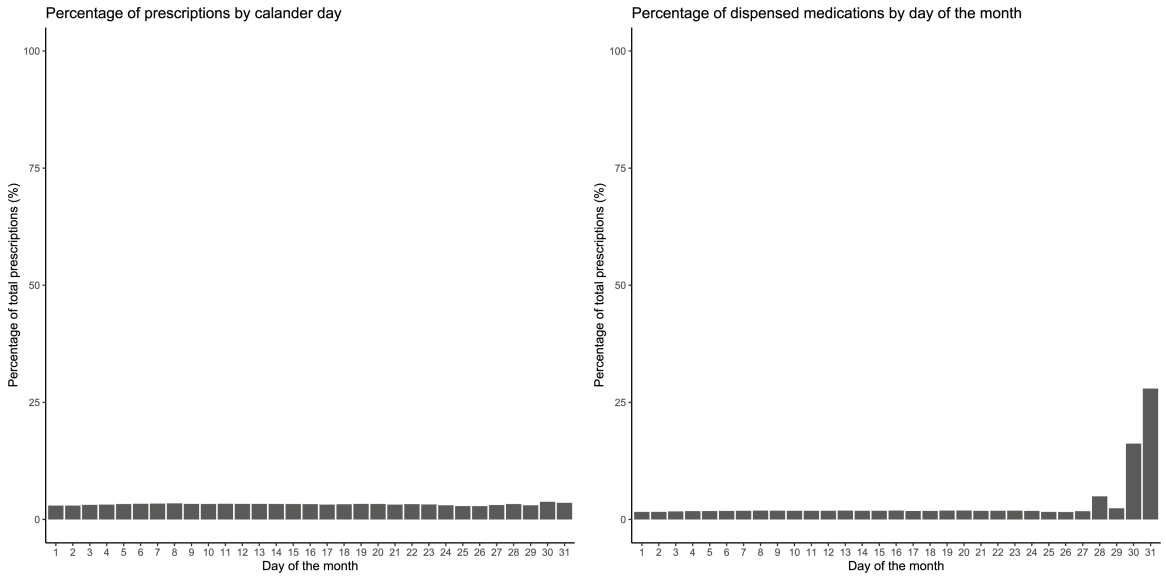
Each prescription record is accompanied by a prescribing date (`PRESC_DATE`), indicating when the medication was prescribed to the patient, and a dispensing date (`DISP_DATE`), when the patient acquired the medication. The PIS data within the GSH/18/CA/002 dataset has two quirks involving the prescription and dispensing dates that need to be considered. Regarding the prescribing date, there were over a thousand prescriptions for individual medications where the patient, medication, and prescribed date were the same, but each row had a different dispensed date. One would assume that these are repeat prescriptions, but the pattern was rare before 2013. When this pattern isn't present, the prescribed date defaulted to the dispensed date for prescriptions after the initial prescription. That is, the prescribed date changed even if the prescription was repeated.

Concerning the dispensing date, recorded dates likely represent when the pharmacy was reimbursed for the prescription (typically the last day of the month) rather than the date when the medication was dispensed to the patient. This record pattern is shown in 2.5, where prescription dates are uniform throughout the month, while dispensing dates tend to fall on the last day of the month. This is likely an artefact due to Scotland's free at-the-point-of-contact prescriptions, where pharmacies are reimbursed monthly rather than on the day when the patient collects the medication.

The following two data handling assumptions were put into place. First, when referring to the first prescription of a type, the `PRESC_DATE` was used, as it would not be affected by any of the inconsistent repeat prescription updates. Second, when defining a repeat prescription, the spacing between prescriptions was calculated using the `DISP_DATE` instead of the `PRESC_DATE`.

British National Formulary

PIS uses British National Formulary (BNF) codes to specify prescriptions. The BNF is a joint publication reference book published by the British Medical Association and Royal Pharmaceutical Society containing the standard list of medications used within the NHS (Joint Formulary Committee 2019). For each medication, the BNF gives summaries of product characteristics, treatment summaries, drug classification, indication and dose information, side effects, impor-



(a) Prescription date (PRESC_DATE) (b) Dispensing date (DISP_DATE)

Figure 2.5: Spread of recorded prescription days (PRESC_DATE) across the month versus spread of recorded dispensing days (DISP_DATE), a clear reimbursement artefact.

tant safety information, evidence grading, and legal categories. The NHS Business Service Authority assigns a unique, hierarchical code to drugs and chemicals using a legacy version of the BNF's hierarchy (French 2017), which is why this code is referred to as the BNF code.

Furosemide Tablets 40mg (generic)

Chapter	Section	Paragraph	Sub paragraph	Chemical substance	Product	Strength and formulation	Generic equivalent
02	**02**	**02**	**0**	**L0**	**AA**	**BD**	*****
Cardio-vascular System	Diuretics	Loop Diuretics	Loop Diuretics	Furosemide	Furosemide	Furosemide _tap 40 mg	

'AA' here always indicates 'generic'

Figure 2.6: A breakdown of the BNF code for a generic 40 mg tablet of furosemide. 'AA' in the 'Product' section always indicates that the medication is a generic version. The asterisk indicates that any code could be entered in this section.

The first nine characters of the BNF code specify the chemical level of the medication. Within these nine characters, the first two characters indicate the chapter of the BNF that the medication is from. For example, drugs in BNF Chapter 2 (Cardiovascular System) will always begin with '02'. The code is then further subdivided into sections (e.g., Diuretics, contained within Chapter 2 Section 2 of the BNF, all begin with '0202'). The remaining six characters provide more detailed information about the medication, including whether the product is branded or generic,

its strength, and formulation (see Figure 2.6 for a breakdown of a 9-character BNF code).

Classifying Prescriptions

There are two primary ways to classify prescriptions. The first, and most straightforward way, is to classify prescriptions using the BNF Chapter, Section, or Paragraph (see Section 2.5.4). The benefit of this classification mechanism is that it easily groups medications without incorporating potential selection bias or classification errors. However, combination medications will often be in a separate paragraph from the constituent chemicals.

The alternative, which was used for this analysis, is to classify medications based on their active chemicals, meaning that loop diuretics were classified as all medications which included a member of the loop diuretic family. See Table A.3 for a complete list of classifications.

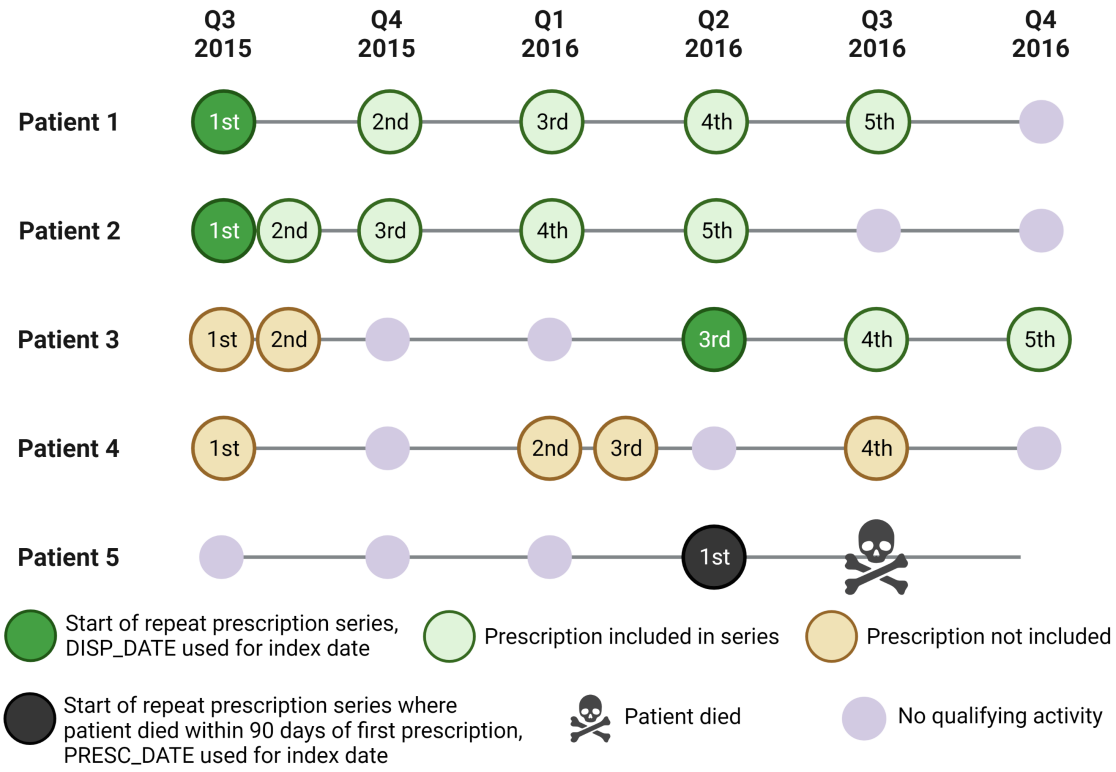


Figure 2.7: Examples of prescription patterns found within PIS. Patients 1, 2, 3, and 5 would be classified as being on a repeat prescription, while patient 4 would be considered to be on an 'as needed' administration direction. For patients 3 and 4, the two prescriptions close together do not account as they occurred within the same quarter and were not followed by dispensing in the next quarter. Note that to be included as a repeat prescription, the prescription must be followed by a serial dispensing or death within 90 days (about 5% of patients who met the definition this way died within 90 days of the first prescription of LD.

Repeat Prescriptions

The decision was made to focus on repeat prescriptions of loop diuretics relevant to the management of chronic disease, instead of people who received a single prescription or a prescription that was renewed infrequently. PIS does not provide a variable to indicate whether or not an individual dispensed prescription is part of a larger repeat prescription. To address this gap, patients were defined to be on a repeat prescription if they either had a prescription dispensed over two consecutive quarters or died within 90 days of their first dispensed prescription. For example, in Figure 2.7, patients 1 through 3 fulfil the first criterion, while patient 5 fulfils the second criterion. By this definition, only patient 4 fails to meet the requirements for being on a repeat prescription. Within NHS GG&C, of those determined to be on a repeat prescription of a loop diuretic, 94.5% of the population met the first criteria, and only 5.5% were determined to be on a repeat prescription due to death within 90 days of the first prescription. Where patients met the definition due to receiving a dispensed medication over two consecutive quarters, the date of initiation is defined by the qualifying medication's dispensing date (`DISP_DATE`). Where the patient met the definition due to death within 90 days, the date of initiation is defined by the qualifying medication's prescription date (`PRESC_DATE`)[2].

The SQL code used to identify repeat prescriptions is provided in Section A.7.

Duplicate Prescriptions

In cases where duplicate records existed (i.e., identical across all fields), only one record was retained for use in the analysis. This decision was made in order to avoid erroneous identification of repeat prescriptions within the dataset.

2.5.5 Scottish Morbidity Records

Scottish Morbidity Records (SMR) are Tier 1 datasets which contain individual-level healthcare data for patients treated within Scotland. The type of record denotes the general type of healthcare received and/or the medical status of the patient. The five SMR data sets routinely used for research include:

- SMR00 - Outpatient Appointments & Attendance
- SMR01 - General/Acute Inpatient & Day Cases
- SMR02 - Maternity Inpatient & Day Cases
- SMR04 - Mental Health Inpatient & Day Cases

[2]As the `DISP_DATE` usually falls at the end of the month (see Section 2.5.4), this date often occurs after the patient's DOD. For this reason, and because there is no risk of running into the `PRESC_DATE` update inconsistency mentioned in the same section (Section 2.5.4), it was deemed safe to use the `PRESC_DATE` as the reference date.

- SMR06 - Scottish Cancer Registry

The GSH/18/CA/002 dataset contains linked information from SMR00, SMR01, and SMR04. Data coverage across Scotland for these data sources is reported to be complete with $\geq 99\%$ coverage (Public Health Scotland 2020*b*). Within the cohort, the number of distinct patient contacts is presented per year in Figure 2.8.

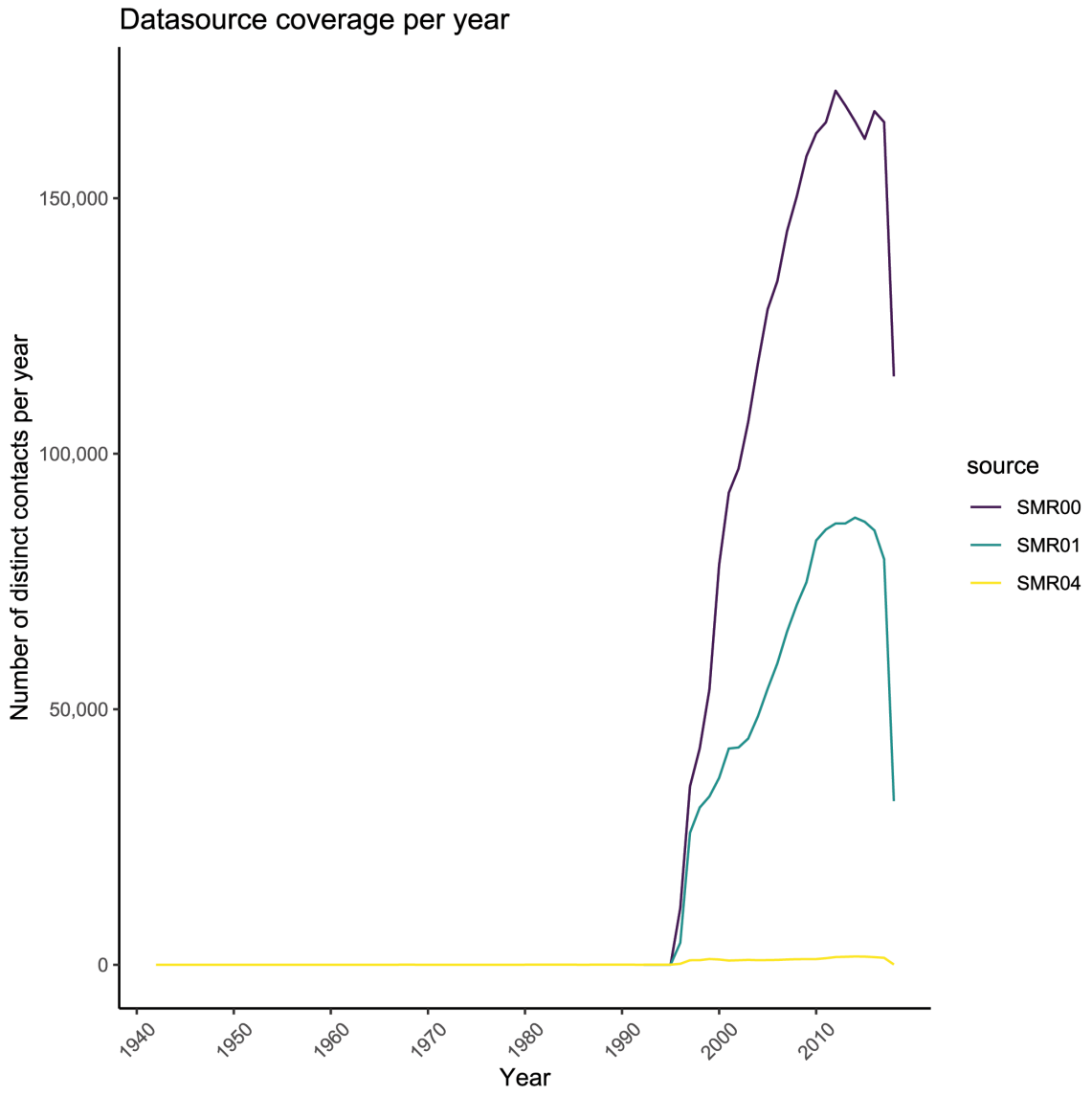


Figure 2.8: The number of distinct contacts per year by Scottish Morbidity Records dataset.

SMR00 - Outpatient Appointments & Attendance

SMR00 contains information on outpatient appointments, attendance, and procedures performed. A record is generated when a patient either has outpatient clinical interaction or where the patient meets with a healthcare provider responsible for care outwith an outpatient clinic session (NHS National Services Scotland nd*d*). The value of SMR00 lies in being able to track

patient contact with a specialist. Unfortunately, this rarely includes information on diagnosis or procedures.

SMR01 - General/Acute Inpatient & Day Case

SMR01 contains information regarding all general and acute inpatient and day cases from all NHS hospitals in Scotland. Each row of data corresponds to an episode of care. Patients receive a new episode of care each time they change specialty, significant facility[3], or consultant for medical reasons (Information Services Division Scotland 2022*b*).

Each episode of care contains some demographic information about the patient, admission, discharge, procedures if performed, and diagnostic factor(s) contributing to the episode. The demographic information contained within each row is limited to ethnicity, age, and SIMD decile and quintile. Admission information covers admission date (`ADMDATE`), admission type (i.e., emergency, urgent, or routine in `ADMTYPE`), where the patient was admitted or transferred from (`ADMTRANS`), what specialty the patient was treated by (`SPEC`), and what hospital the patient was admitted to (`HOSP`). Discharge information covers discharge date (`DISDATE`), discharge type (e.g., regular discharge, death, or transfer in `DISTYPE`), and where the patient was discharged or transferred to (`DISTRANS`). Each record must have the first diagnostic position (`DIAG1`) populated, which defines the primary diagnosis or main problem treated within the episode of care, and may have up to five additional positions populated with diagnosis information classified using ICD-10 codes (see Section 2.5.1). Data quality assurance assessments have suggested coding accuracy levels $\geq$88% using the first 4 digits of the ICD-10 code for `DIAG1`, but accuracy declines for `DIAG2` - `DIAG6`, including under-reporting of common conditions such as HF and AF/AFL (Public Health Scotland 2019, Khand et al. 2005, National Services Scotland Information Services Division 2019). However, coding may be more accurate for some conditions which have a large objective component to diagnosis (e.g., cancer, MI), but much less accurate for those which have a large subjective component (e.g., HF), or where the problem is not considered a primary problem (e.g., AF) (Khand et al. 2005).

Additionally, each record has space for up to four procedures (`OPxA` [where `x` is the procedure number 1 - 4]) with the potential for additional information (e.g., laterality, aborted, or unsuccessful are coded in `OPxB` [where `x` is the procedure number 1 - 4]) codes recorded using Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 (OPCS-4) (see Section 2.5.5). Where applicable, the procedure coded in `OP1A` is considered the primary or main procedure for that episode of care. As with diagnostic codes, duality assurance assessments have shown coding accuracy levels $\geq$94% using the first four digits of the OPCS-4 code, with $\geq$97% of hospitals reporting codes (Public Health Scotland 2019).

SMR04 - Mental Health Inpatient & Day Cases

SMR04 contains information regarding mental health inpatient and day cases. The SMR04 dataset has a similar format to that of SMR01(see Section 2.5.5) with regard to the information provided. Data points are recorded within episodes of care, and contain patient demographic, admission, discharge, and diagnostic information. However, these columns are not populated within the GSH/18/CA/002 dataset as NHS GG&C transports patients to general hospitals to undergo procedures and medical intervention, which would be recorded in SMR01. For this reason, patients are still at risk for an SMR01 admission while they are receiving care under the purview of an SMR04 contributing facility. SMR04 admissions tend to be for longer stays than are found in SMR01admissions.

Defining Hospital Admissions

The length and number of hospital admissions both reflect the severity of the disease and the resources expended. Unique hospital admissions must be identified and grouped from the recorded episodes of care in either SMR01 or SMR04.

According to National Services Scotland, a continuous inpatient stay is defined as an unbroken period of time that a patient spends admitted in hospital (Redpath 2018). Taking into account the example provided in Redpath's report, and the lack of information on the time of day for discharges and re-admissions, single hospital admission was defined as the set of all episodes of care that differed by no more than one calendar day between one episode's discharge and the subsequent episode's admission date as shown by the green arrow in Figure 2.9. This holds true whether or not patients transferred between hospitals or NHS Boards (Anwar et al. 2011). This is particularly important as many cardiac patients are transferred to and from the Golden Jubilee National Hospital for specialist care. If two episodes of care differ by at least one day, these two episodes are considered to belong to different hospital stays, and the event is classified as a new admission (see the blue arrow in Figure 2.9).

Hospital admissions were identified and labelled with unique identification numbers, `hosp - StayID`, within SMR01 and SMR04. This number is unique only within its given dataset, so admission '1' in SMR01 might refer to a different patient and separate admission than admission '1' in SMR04.

The length of hospital admission was defined as the difference between the first admission date and the last discharge date for each `hospStayID`. This value is recorded in the `lenOfStay` field.

Three patients were excluded from analyses because they had hospital admissions lasting more

[3] A division of medicine or density covering a specific area of clinical activity and identified within one of the Royal Colleges or Faculties (Information Services Division Scotland 2022*c*).

Other columns containing information about hospital stay

SafeHavenID	HOSP	ADMTYPE	ADMTRANS	DISTYPE	DISTRANS	ADMREAS	ADMDATE	DISDATE	...	RefAdmit	HospStayID	LenOfStay
1	QEUH	30	11	12	52	10	2021-01-01	2021-01-02	...	1	1	10
1	Jubilee	18	50	10	10	11	2021-01-02	2021-01-10	...	0	1	10
1	GRI	10	69	40	00	NULL	2021-10-11	2021-10-30	...	1	2	19

HOSP, hospital; QEUH, Queen Elizabeth University Hospital; Golden Jubilee National Hospital; GRI, Glasgow Royal Infirmary; ADMTYPE, admission type (i.e., routine [numbers 10-19], urgent [20-22] or emergency [30-39]); ADMTRANS, admission/transfer from (i.e. private residence [numbers 10-19], institution [20-29], temporary place of residence [30-39], transfer within the same Health Board/Health Care provider [40-49, 4A-4H], transfer from another Health Board/Health Care Provider [50-59, 5A-5H], other type of location [60-69]), DISTYPE, discharge type (i.e., regular [10-19], irregular [20-29], or death [40-43]); DISTRANS, discharge/transfer to (i.e., died [00, 01], private residence [10-19], institution [20-29], temporary place of residence [30-39], transfer within the same Health Board/ Health Care Provider [40-49, 4A-4H], transfer to another Health Board/Health Care Provider [50-59, 5A-5H], other type of location [60-69]); ADMREAS, admission reason; [ADMDATE, admission date; DISDATE, discharge date; refAdmit, reference admission indicating the first record of the hospital stay; HospStayID, unique hospital stay number to identify a unique admission; LenOfStay, total length of hospital admission.

Figure 2.9: Example of admission and discharge information contained in the Scottish Morbidity Records General/Acute Inpatient and Day Case (SMR01) episodes of care. Episodes of care are considered part of the same hospital stay if the difference between an admission and discharge dates was at most one day (green arrow), regardless of the hospital; otherwise, the records were considered from two different admissions (blue arrow). This allows for a transfer at 11:30 p.m. with an arrival after midnight to be registered as the same admission.

than 2,000 days (or 5.5 years). These patients were excluded because such a length of stay in an acute hospital is improbable, and if true, they would not have received any community-dispesned prescriptions.

Admission Type

The admission type was defined using the first record of stay's (see Section 2.5.5) 'Admission Type' (`ADMTYPE`) field. For SMR01, the 'Admission Type' is a two-digit number with values 10 through 19 defined as a 'Routine Admission', values 20 through 22 defined as an 'Urgent Admission', and values 30 through 39 defined as an 'Emergency Admission' (NHS National Service Scotland 2019). For analysis purposes, admissions were grouped as emergency or non-emergency to remove any possible ambiguity acquired when trying to describe what classifies an 'Urgent Admission.'

Identifying First Record of Stay

For hospital admissions spanning more than one episode of care (see Section 2.5.5 for the definition of continuous hospital admission), hospital episodes were ordered within an admission in order to identify the admission reason. Hospital episodes of care were ranked to prioritise earliest dates, most urgent admission reason, and transfer to other institutions over discharges home or death in order to find the most probable first episode of care. This was done using the following logic:

1. Prefer the earliest admission date (rank `ADMDATE` in ascending order).

2. Prefer an admission/transfer from a private residence before admission from an institution or a transfer within the same health board/health care provider (rank `ADMTRANS` in ascending order).
3. Prefer emergency admissions over urgent or routine admissions (rank `ADMTYPE` in descending order).
4. Prefer a discharge/transfer to another health board/ health care provider before an institution, private residence, or finally death (rank `DISTRANS` in descending order).
5. Prefer the earliest discharge date (rank `DISDATE` in ascending order).
6. Prefer missing admission reasons, as it is the most common, followed by an acute admission with no additional detail added, then admission for treatment, pre-operative preparation, and so on, finishing with geriatric palliative care (rank `ADMREAS` in ascending order).

Using the above logic, the reference admission flag (`refAdmit`) was set to 1 for the highest-ranked episode of care, while the other episodes were given a 0 flag.

Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4

The Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 (OPCS-4) coding standard is developed, maintained, licensed, and supported by NHS Digital's Terminology and Classifications Delivery Service and governed by Crown Copyright (NHS Digital 2019). OPCS-4 is a hierarchical coding standard used to classify operations, procedures, and interventions conducted within the NHS. OPCS-4 codes are four characters long and have a similar structure to ICD-10 codes. OPCS-4 codes start with a letter followed by three digits. A full stop (.) separates the second and third digit (NHS Digital 2021).

2.5.6 Electrocardiography

The electrocardiogram (ECG) is a record of the heart's electrical activity, including rate, rhythm, and conduction abnormalities. It is essential for the diagnosis of many cardiovascular diseases, including arrhythmias (e.g., AF), problems with electrical conduction (e.g., left or right bundle branch block), left ventricular hypertrophy (e.g., due to hypertension), or myocardial infarction (MI) (both acute and previous) (Hampton & Hampton 2019). An annotated ECG cardiac cycle is presented in Figure A.2.

The MUSE ECG dataset is a Tier 2 dataset covering the 12-lead ECG data acquired from all NHS GG&C electrocardiographs that were connected to General Electric (GE)'s MUSE Cardiology Information System (Watts & Graham 2022). Safe Haven provided the automatically generated

SQL data for electrocardiogram (ECG) variables and diagnostic rhythm information from the recorded ECGs, including but not limited to, ventricular rate, QRS duration, information on ST-T wave abnormalities, information about conduction delays, and the numerical waveform values (e.g., PR, QRS, QT, and ST intervals) from the 12 leads.

Data Cleaning

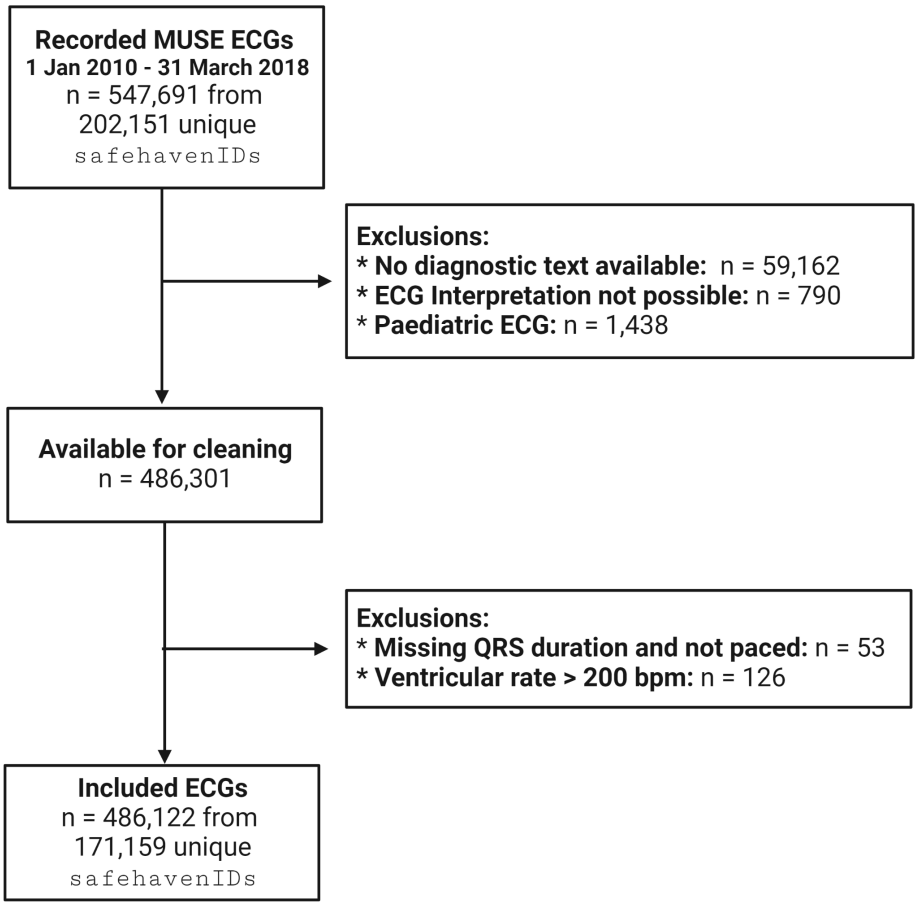


Figure 2.10: Diagram detailing the number of ECGs excluded for given reasons and the number ultimately available for analysis, where n indicates the number of unique ECGs.

ECGs were excluded from analysis for the following reasons (see Figure 2.10):

- Where the diagnostic text was either missing or completely redacted,
- Where the diagnostic text stated that ECG interpretation was not possible,
- Where the diagnostic text indicated a paediatric case, as the interpretation of values varies from that originating from adult values,
- Where the QRS was missing, and the diagnostic text did not indicate the presence of pacing, as the other numerical values were also missing, making further analysis impractical,

- And finally, where the ventricular rate was above 200 beats per minute, a heart rate that is incompatible with life if sustained, could be an artefact and would be followed soon after by a further ECG if a rate and rhythm compatible with survival were restored.

ECGs will often have been recorded during a hospitalisation for acute illness, which should be kept in mind when interpreting data.

Heart Rhythm

The heart rhythms used for analysis were obtained from the diagnostic statement text derived by GE's proprietary ECG algorithm without clinical over-reading or correction of this data. Heart rhythms were identified and classified using regular expression matching and wild card searches. Rhythms were assigned in the order of sinus, AF, atrial flutter (AFL), supraventricular tachycardia (SVT), idioventricular, junctional, other, paced[4], and undetermined (see Table A.4 for search terms).

Rhythms were grouped into four discrete groups based on atrial activity to be sensitive to the lack of clinical review: sinus, AF/AFL/SVT, Other, and undetermined. Table 2.2 contains the classifications, counts, and relative frequencies.

Grouped Rhythm	**Grouped Frequency**	**Singular Frequency**
Sinus	399,584 (82.20)	Sinus 399,584 (82.20)
AF/AFL/ SVT	68,142 (14.02)	Atrial fibrillation 61,319 (12.61) Atrial flutter 5,942 (1.22) Supraventricular tachycardia 881 (0.18)
Other	13,927 (2.86)	Idioventricular 155 (0.03) Junctional 4,522 (0.93) Other 1,531 (0.31) Paced 7,719 (1.59)
Undetermined	4,469 (0.92)	Undetermined 4,469 (0.92)
Data are n(%). AF, atrial fibrillation; AFL, supraventricular tachycardia; SVT, supraventricular tachycardia.		

Table 2.2: Grouped ECG rhythm classification counts and relative percentages.

Heart Rate

Heart rate was defined as the recorded ventricular rate, `VentRate`, provided by the MUSE ECG System. A normal ventricular rate for an adult in sinus rhythm is 60-100 beats per minute (Wyatt 2006).

[4]Paced is only assigned as a rhythm in the absence of an intrinsic rhythm, as patients usually have an intrinsic rhythm in addition to having an active cardiac implantable electronic device.

QRS Duration

The QRS complex represents the depolarisation of the ventricles. The QRS duration measures the amount of time taken for the excitation to travel through the ventricles (Hampton & Hampton 2019) (see Figure A.2 for an annotated ECG). A normal QRS duration is less than 0.12 s (120 ms). Values $\geq$120 ms reflect problems with conduction from the atria (the chambers that receive venous blood from the systemic and pulmonary circulation) to the ventricles (the main 'pumping' chambers of the heart). QRS prolongation is also associated with left ventricular dilation and dysfunction and with a poorer prognosis for patients with HF (Iuliano et al. 2002).

ST-T Abnormalities

ST-T abnormalities refer to changes in either the ST segment or T wave. The ST segment of the cardiac cycle lies between the end of the QRS complex and the start of the T wave or the end of ventricle depolarisation and the start of the ventricular repolarisation (see Figure A.2 for an annotated ECG). In a normal ECG, the ST segment is an electrically neutral area and, as such, should be isoelectric, flat (i.e., neither positive nor negative) and at the baseline of the cardiac cycle. The T wave measures the repolarisation of the ventricles.

Abnormalities in the ST segment and T wave are common and reflect abnormal ventricular repolarisation (Rautaharju et al. 2009). They can have numerous causes, including IHD, pericarditis, digoxin therapy, ventricular hypertrophy, and electrolyte abnormalities (Hampton & Hampton 2019) including hypokalaemia and hyperkalaemia (see Section 2.5.9).

QT Interval Correction

The QT interval is measured from the start of the Q wave to the end of the T wave of the cardiac cycle, measuring the ventricular depolarisation and repolarisation of a heartbeat (see Figure A.2 for an annotated ECG). The length of the QT interval is dependent on and inversely related to heart rate; as the heart rate increases, the QT interval decreases. Numerous formulae have been proposed to take into account the impact of the heart rate on the QT interval (QTc) including, but not limited to, Bazett's original correction published in 1920 (Bazett 1920), Fridericia (Fridericia 1920), Framingham (Sagie et al. 1992), Hodges (Hodges 1997), and Rautaharju (Rautaharju et al. 1992). Numerous studies point to the inferiority of the Bazett correction (Luo et al. 2004, Malik 2001, Sagie et al. 1992, Vandenberk et al. 2016). Based on the combination of expert opinion (Macfarlane 2021), universal use, and its improved one-year mortality risk stratification (Vandenberk et al. 2016), the Fridericia QTc formula was implemented as seen below.

$$QTc = QT(Heart\,rate/60)^{1/3} = QT(RR)^{-1/3} \tag{2.1}$$

Based on the Fridericia correction, QTcs were considered prolonged if they were longer than 450 ms for men and 460 ms for women (Rautaharju et al. 2009). The prolongation can be caused by numerous factors, including by QT-prolonging cardioactive medications (e.g., amiodarone and dronedarone) (Joint Formulary Committee 2019), and hypokalaemia (see Section 2.5.9. However, in HF, the most common reason is an increase in QRS duration. A prolonged QT is associated with an increased risk of ventricular tachycardia and sudden death (Algra et al. 1991, Rautaharju et al. 2009).

2.5.7 Echocardiography

An echocardiogram is a non-invasive form of medical imaging, which uses ultrasound to assess and measure the heart's structure and function. The measurements collected during an echocardiogram are used to diagnose and monitor heart valve and muscle disease and inform treatment decisions. Echocardiograms are usually reported in a standard format but focus on abnormal findings; when a report says nothing about a heart valve or heart function being abnormal, that generally indicates that no problem was observed.

Systems and Data Cleaning

NHS GG&C uses two echocardiography platforms, GE Healthcare's EchoPAC and Philip's Xcelera system. The EchoPAC system is more commonly used, representing about 81% of the recorded echocardiograms, and the remaining 19% were taken from the Xcelera platform. Structured echocardiogram measurements stored within the Image Vault system were available for analysis, although there is some uncertainty about which echocardiograms were migrated here (Watts & Graham 2022).

For the purposes of this thesis, the small number of stress echocardiograms was excluded. In addition, EchoPAC echocardiograms where the `AverageType` was labelled as 'U', indicating undefined values for a measurement, were also excluded (see Figure 2.11).

Averaging Protocol

The echocardiogram data provided by NHS GG&C Safe Haven often have repeated measurements within the same test. The following averaging protocol has been implemented to assign one, potentially composite, measurement to an individual per day. The steps are as follows:

1. Average the values associated with the highest-ranked measurement per test, which fall within the range of believable (e.g., average all believable biplane measurements). This value becomes the measurement assigned to a particular measurement within the echocardiogram.

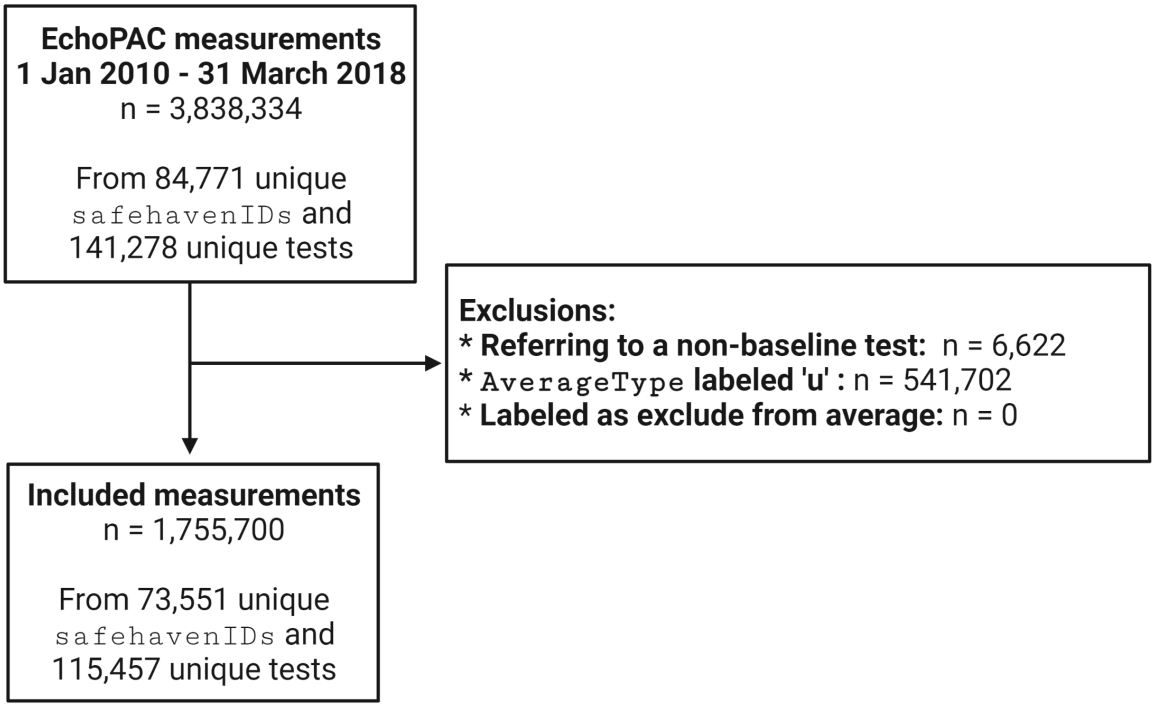


Figure 2.11: Flow chart of EchoPAC exclusion criteria.

2. Take the average of the values found in Step 1 if a patient has more than one baseline echocardiogram recorded in a day. This is the measurement value assigned to an individual on a particular day per system.
3. Combine EchoPAC and Xcelera results from Step 2.
4. If a patient has echocardiograms recorded on more than one system on the same day, prefer values from EchoPAC, the more prevalent dataset.
5. If applicable, classify the resulting value. Otherwise, use the value produced in Step 4.

Averaging multiple values of the same test type is intended to smooth out inconsistencies between operators and measurement errors.

Left Ventricular Ejection Fraction

Left ventricular (LV) ejection fraction (LVEF) is a measurement of the LV's pumping capacity measured as the difference in the volume of blood in the LV at the end of the filling period (diastole) compared with the end of contraction (systole). Simpson's biplane was the preferred method, followed by the method of disks (MOD) (additional measurement types used for LVEF are reported in Table A.7). Only measurement values between 1% and 90% were included for averaging.

LVEF is used to phenotype HF and select the most appropriate treatments (McDonagh et al.

2021, Butler et al. 2014). ESC guidelines suggest classifying patients with an LVEF $\leq$40% as HF with reduced LVEF (HFrEF), those with an LVEF between 41% and 49% as HF with mildly reduced LVEF (HFmrEF), and LVEF $\geq$50% as HF with preserved LVEF (HFpEF) (McDonagh et al. 2021).

Mitral Valve Regurgitation

Mitral valve regurgitation (MR) is common (Iung et al. 2019) and occurs when the mitral valve is unable to form a seal, allowing blood to flow backwards from the LV to LA. The presence of MR was defined by the presence of a measurement listed in Table A.9.

Tricuspid Valve Regurgitation

Tricuspid valve regurgitation (TR) occurs when the tricuspid valve is unable to form a seal, allowing blood to flow backwards from the RV into the LA. The presence of TR was defined according to Section A.12.4. Values below 0 mm Hg or above 170 mm Hg were classified as biologically impossible and were excluded. Valid measurements were then averaged according to 2.5.7

Aortic Stenosis

Aortic stenosis (AS) is another common and serious valve disease (Carabello & Paulus 2009, Iung et al. 2019). It occurs when the aortic valve gets 'stuck', restricting blood flow from the LV to the aorta. AS was measured by the gradient (mm Hg). Values were identified and ranked according to measurement modalities listed in Table A.10. Values less than 0 mm Hg and above 170 mm Hg were classified as biologically impossible and were excluded. Valid measurements were then averaged according to Section 2.5.7.

Aortic velocity (m/s) was calculated using Equation 2.2 for aortic gradient values greater than 0 mm Hg.

$$velocity = \sqrt{\frac{gradient}{4}} \tag{2.2}$$

Thresholds of $\leq$2.5 m/s, (2.5-3.0) for mild, [3.0-4.0) for moderate, and $\geq$4.0 for severe AS (Baumgartner et al. 2017).

Aortic Regurgitation

Aortic regurgitation (AR) occurs when the aortic valve is unable to form a seal, allowing blood to leak backwards from the aorta into the LV. The presence of AR was determined by the presence of a measurement, therefore showing an attempt to quantify its presence (See Section A.12.5).

Reporting Chamber Thresholds

Current ESC Cardiovascular Imaging recommendations (Lang et al. 2015) report volume metrics adjusted for body surface area (BSA) in an attempt to remove bias. Unfortunately, BSA values and BSA indexed values were sparse compared with the corresponding unadjusted variable. For the sake of data coverage, it was decided to report and use the unadjusted variables using the 2006 recommendation's unadjusted variable thresholds for the left atrial area, left atrial volume, left atrial diameter, and right atrial area (Lang et al. 2006).

Left Atrial Diameter

The left atrial (LA) diameter is a simple measurement of LA size that is usually reported (Khankirawatana et al. 2004). LA volume is more accurate but much less often measured and reported (Lester et al. 1999). Even so, LA diameter is a strong predictor of first cardiovascular events (Kizer et al. 2006).

LA diameter measurements were identified and ranked according to Table A.8. Values below 1 cm and above 10 cm were excluded as being improbable. The remaining measurements were averaged according to the protocol to arrive at a single measurement per patient per day. LA diameter was considered increased when $\geq$4.0 cm for men and $\geq$3.8 cm for women (Lang et al. 2015).

Excluded Measurements

Left atrial area (LAA), right atrial volume (RAV), right atrial area (RAA), inferior vena cava (IVC), and tricuspid annular plane systolic excursion (TAPSE) were identified but not included in the subsequent analysis due to incomplete record coverage. Details of how these measurements were identified and cleaned can be found in Appendix SectionA.13.

2.5.8 Scottish Care Information-Diabetes Collaboration

Scottish Care Information-Diabetes Collaboration (SCI Diabetes) is a Tier 1 dataset holding the electronic clinical registry records pertaining to the treatment of people with DM in Scotland (Livingstone et al. 2012). It holds some records dating back to the mid-1920s (see Figure 2.12), but full coverage with automatic capture based on assigned Read Code started in 2000. It has a national estimated capture of $\geq$99% of all people diagnosed with DM (Livingstone et al. 2012).

2.5.9 Scottish Care Information Store

Scottish Care Information Store (SCI Store) is a Tier 2 dataset covering all Scottish NHS Health Boards and contains clinical reports from biochemistry, haematology, pathology, microbiology,

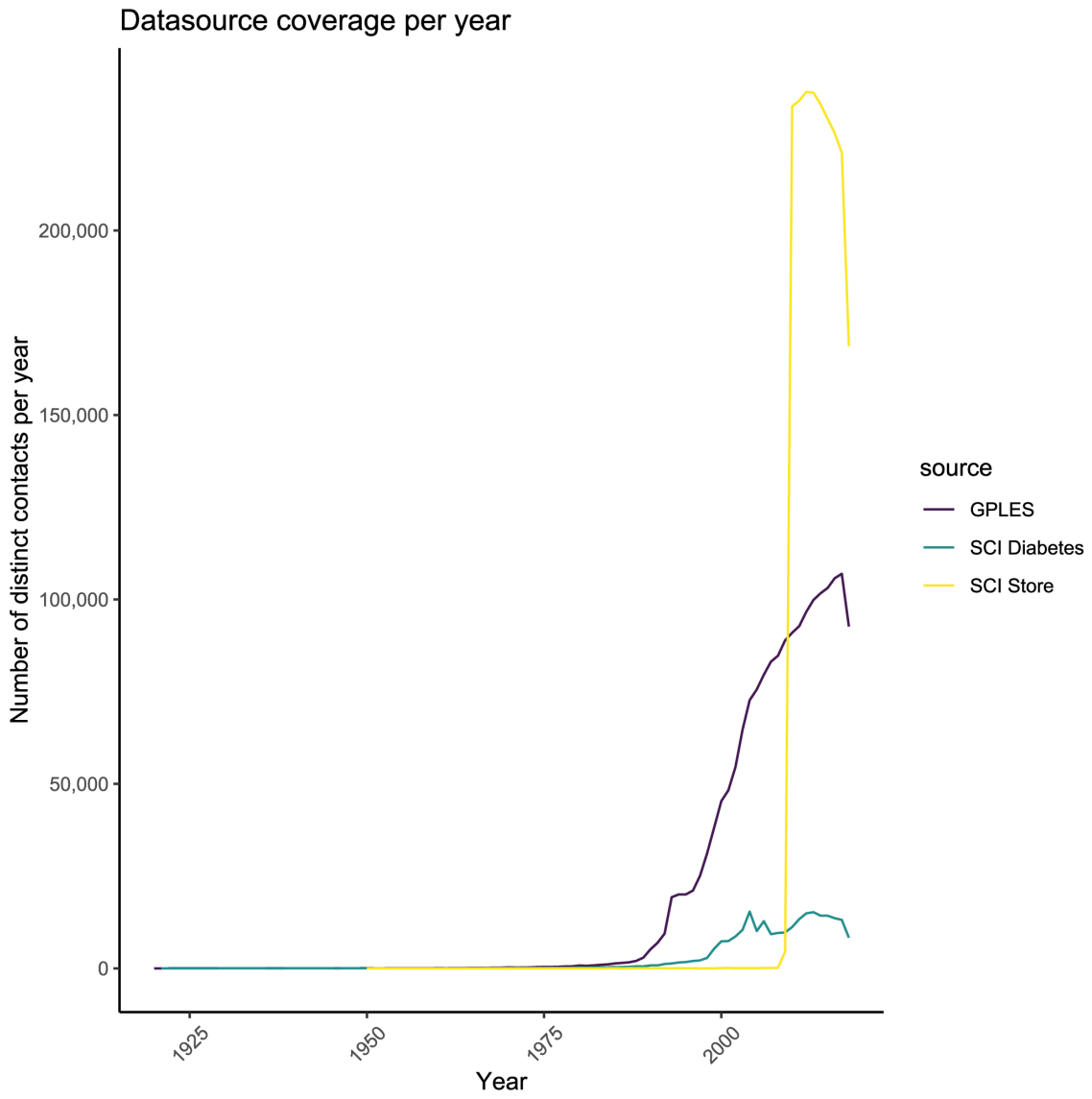


Figure 2.12: The number of distinct contacts per year datasets with coverage starting before 2009. GP LES, General Practice Local Enhanced Services (see Section 2.5.3); SCI Diabetes, Scottish Care Information-Diabetes Collaboration (see Section 2.5.8); SCI Store, Scottish Care Information Store (see Section 2.5.9)

and radiology (NHS National Services Scotland 2015). For the GSH/18/CA/002 dataset, SCI Store was limited to information on haematology and biochemistry to focus on blood tests. The extract has full coverage between 1st January 2010 through 31st March 2018 (see Figure 2.12). Test types were identified using the `CLINICALCODEVALUE` field. The following lab tests were selected due to their clinical importance.

Haemoglobin

Haemoglobin is a protein found in red blood cells that binds to and transports oxygen. It was identified using the `CLINICALCODEVALUE` '423..'. Haemoglobin test records were excluded

where the sample date (SAMPELDATE) or test value was missing. Values below 2 g/dL or above 30 g/dL were classified as biologically improbable and excluded (see Figure 2.13a for per-year box plots of haemoglobin results). Values prior to 2010 were also excluded due to incomplete reporting.

If a patient had multiple haemoglobin results recorded on the same day, the lowest test result was used to avoid including post-transfusion values.

Serum Creatinine and Estimated Glomerular Filtration Rate

Creatinine is a waste product from muscle tissue. Normal serum levels are based predominantly on an individual's age and sex; high levels indicate impaired renal function. Serum Creatinine values were identified using the CLINICALCODEVALUE '44J3.'. During data cleaning, values from before 2010 were excluded due to incomplete records, and extreme values were verified against the patients' other results (i.e., a series of extreme measurements over several days are likely to be true, a single aberrant measurement is likely to be false). See Figure 2.13b for a box plot of serum creatinine results per year with the minimum found value of 8 µmol/L, which occurred within a series of similar values, indicated by the horizontal dotted line.

Renal function was further assessed using the estimated glomerular filtration rate (eGFR). The eGFR was calculated from serum creatinine values using the isotope dilution mass spectrometry (IDMS)-traceable Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation below (Levey et al. 2009).

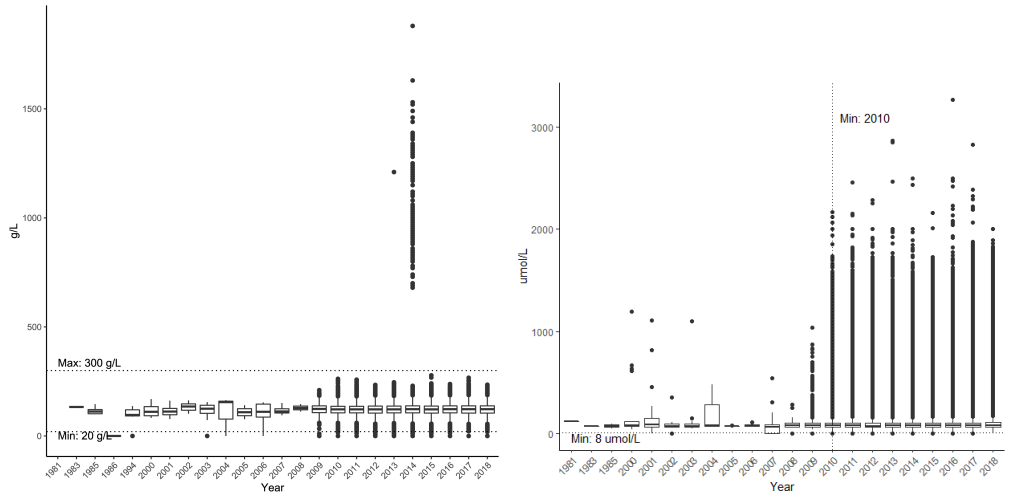
$$eGFR = 141 \times min(SCr/\kappa, 1)^{\alpha} \times max(SCr/\kappa, 1)^{-1.209} \times 0.993^{Age} \times A \times B \qquad (2.3)$$

SCr is serum creatinine (mg/dL), κ is 0.7 for women and 0.9 for men, α is -0.329 for women and -0.411 for men, Age is the individual's age on the date of the test, A is 1.159 if the patient is a woman and 1 otherwise, and B is 1.159 if the patient is African American and 1 otherwise.

In calculating eGFR values, it was assumed that there were no African Americans in the cohort and that all serum creatinine values were IDMS standardised. Moreover, recent research suggests that eGFR should not be adjusted for ethnicity (Diao et al. 2021). Serum creatinine values were converted from µmol/L to mg/dL by a conversion factor of 0.01131, as required by Equation 2.3. Results were rounded to the nearest integer. If a patient had multiple tests recorded on the same day, the eGFR derived from the highest serum creatinine value (i.e., the lowest eGFR) was used.

The CKD-EPI equation was chosen over the Modification of Diet in Renal Disease Study Equation (MDRD) for several reasons. First and foremost, CKD-EPI was developed using a larger database, including 16 additional studies, and has been shown to be more accurate for eGFR

values >60 mL/min/1.73m^2 (Levey et al. 2009). Secondly, CKD-EPI has been shown to be less biased for groups at increased risk for CKD, including elderly patients, diabetics, and obese individuals compared with the MDRD equation (Stevens et al. 2010). Finally, while the MDRD equation has historical precedence, the CKD-EPI equation was developed by the same team three years later to address the shortcomings of the MDRD equation (Levey et al. 2006).



(a) Haemoglobin (Hb) (b) Serum creatinine

Figure 2.13: Box plots of raw haemoglobin and serum creatinine values by year. Dotted lines indicate exclusion limits.

Serum Urea

Serum urea is a waste product from ammonia breakdown and depends on the liver and renal function. Normal ranges are 2.5-6.5 mmol/L (Creed & Hargreaves 2016). Increased urea levels suggest impaired renal dysfunction but can also be caused by dehydration, a catabolic state induced by acute illness or high protein intake (Raine et al. 2018). Low levels of urea are rarely pathological but can be caused by starvation, high water or alcohol intake, or liver failure (Raine et al. 2018).

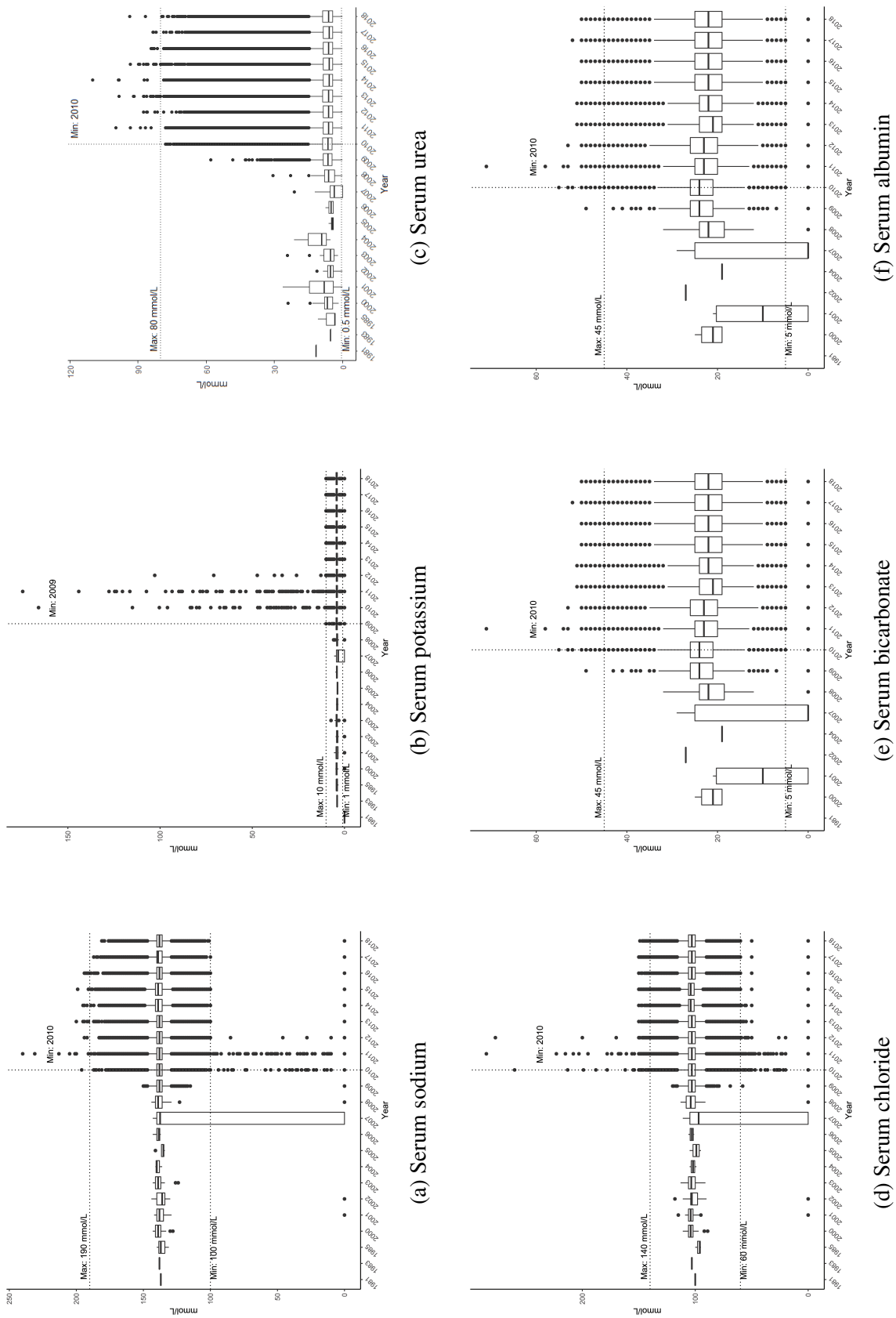
Serum urea test results were identified using the `CLINICALCODEVALUE` '44J9.'. Figure 2.14c displays box plots for all identified test values. Values below 0.5 mmol/L and above 80 mmol/L were classified as biologically improbable and were excluded. Values prior to 2010 were excluded due to incomplete reporting.

Serum Albumin

Serum albumin is a protein mainly produced in the liver (Merlot et al. 2014, Gburek et al. 2021, Bos et al. 1989). Physiologically, it plays a key role in maintaining hydrostatic pressure (i.e., it helps keep water in the vasculature rather than leaking into tissues). Normal values range

between 35-50 g/L, with dehydration causing values to rise and inflammation, liver cirrhosis, renal disease (nephrotic syndrome), and pregnancy causing values to drop (Raine et al. 2018).

Serum albumin tests were identified using the `CLINICALCODEVALUE` '44M4.'. Figure 2.14f displays box plots for all identified values. Test values below 4 g/L or above 80 g/L were classified as biologically improbable and excluded. Values prior to 2012 were also excluded due to incomplete reporting.



(a) Serum sodium (b) Serum potassium (c) Serum urea

(d) Serum chloride (e) Serum bicarbonate (f) Serum albumin

Figure 2.14: Box plots of raw haematology and biochemistry test results by year. Exclusion limits are indicated by dotted lines.

Serum Sodium

Table salt is sodium chloride. The ability to regulate body salt content enabled life to live and emerge from the oceans. The concentrations of sodium (Na+) and chloride (Cl-) ion concentrations are physiologically tightly regulated (Edelman et al. 1958). Increased sodium intake causes water retention to 'dilute' sodium to normal concentrations. However, water retention may lead to increased blood pressure and tissue fluids (oedema), which is why people with high blood pressure or HF are often advised to reduce salt intake. Increased sodium loss will normally be associated with increased urinary excretion of water to maintain a normal sodium concentration. The normal range for serum sodium is 135 - 145 mmol/L (National Institute for Health and Care Excellence 2020, Raine et al. 2018).

Hyponatraemia (a low Na+ concentration) reflects an excess of water compared with sodium, which may be caused by increased sodium loss due to diuretic agents or to water retention due to excess secretion of anti-diuretic hormone, which may be due to lung infections, cancer, and a side effect of anti-depressant medicines (Anderson et al. 1985, Mann 2008, Adrogué & Madias 2000*b*, Ashraf et al. 1981, Fichman et al. 1971). Hyponatraemia becomes increasingly common as the severity of HF progresses and diuretic dose increases, and is associated with a poor prognosis (Anderson 1986).

Hypernatraemia is a less common electrolyte disorder that reflects relative dehydration (Adrogué & Madias 2000*a*). It is also associated with a poor prognosis (Palevsky et al. 1996).

Serum sodium test values were identified using the `CLINICALCODEVALUE` '44I5..'. Box plots for all identified test values are displayed in Figure 2.14a. Values below 100 mmol/L and above 190 mmol/L were classified as biologically improbable and excluded. Values prior to 2010 were also excluded due to incomplete reporting.

Serum Chloride

Chloride and sodium concentrations are highly correlated. Each anion (negative charge; predominantly chloride [Cl-] and bicarbonate [HCO3-]) must be balanced by a cation (positive charge; predominantly sodium and potassium) (Berend et al. 2012, Powers 1999). As for sodium, the kidney is the main organ regulating chloride retention and excretion (Walker et al. 1990). The normal concentration range for serum chloride is between 95 - 105 mmol/L (Singer & Webb 2009). Low concentrations (hypochloraemia) are usually associated with hyponatraemia (Berend et al. 2012). Hypochloraemia and hyponatraemia are associated with diuretic resistance in HF and an adverse prognosis (Hanberg et al. 2016).

High serum chloride concentrations (hyperchloraemia) are usually associated with hypernatraemia and share common causes (Walker et al. 1990). Moderate increases in serum chloride are associated with acute kidney injury (Suetrong et al. 2016).

Serum chloride test results were identified using the CLINICALCODEVALUE '44I6.'. Box plots for all identified test values are displayed in Figure 2.14d. Values below 60 mmol/L and above 140 mmol/L were classified as biologically improbable and excluded. Values prior to 2010 were also excluded due to incomplete reporting.

Serum Potassium

For serum potassium, a slight deviation (less than 1.0 mmol/L) from the normal range (3.5 - 5.5 mmol/L) is associated with increased morbidity (e.g., neuromuscular weakness and paralysis, and arrhythmias) and mortality (Walker et al. 1990). Hypokalaemia, defined as serum potassium below 3.5 mmol/L, may be caused both by thiazide and loop diuretics (Joint Formulary Committee 2019). Unlike haemoglobin and serum creatinine, the interpretation of serum potassium concentrations is independent of sex and age (Walker et al. 1990).

Serum potassium test results were identified using the CLINICALCODEVALUE '44I4.'. Box plots for all identified test values are displayed in Figure 2.14b. Values below 1 mmol/L and above 10 mmol/L were classified as biologically improbable and were excluded. Values prior to 2010 were excluded due to incomplete reporting.

Serum Bicarbonate

Serum bicarbonate (HCO3-) is the other major anion in the blood. Low bicarbonate suggests acidosis, which may be due to renal failure, respiratory failure, circulatory failure, or diabetic keto-acidosis (Navaneethan et al. 2011, Shah et al. 2009). Severe acidosis (a very low HCO3-) indicates a medical emergency. Diuretics may often cause an alkalosis (a high HCO3-). The normal serum range for bicarbonate is 23 - 28 mmol/L (Singer & Webb 2009).

Serum bicarbonate test results were identified using the CLINICALCODEVALUE '44I7.'. Box plots for all identified test values are displayed in Figure 2.14e. Values below 5 mmol/L and above 45 mmol/L were classified as biologically improbable and excluded. Values prior to 2010 were also excluded due to incomplete reporting. Bicarbonate used to be part of the routine biochemistry panel but has since been removed, accounting for the decline in the number of tests performed.

2.6 Derived Variables

2.6.1 Heart Failure

HF was defined as the first record for HF in any diagnostic position using a comprehensive list of codes. Codes were identified through literature (Conrad et al. 2018), expert opinion, and code searches. Codes recorded in secondary care were identified using ICD-10 codes in SMR01 and

SMR04 (see Table A.14). Codes recorded in primary care were identified using Read Codes in GP LES (see tableA.15). In cases where patients were diagnosed during a hospital admission, the diagnosis date is taken as the admission date for the episode of care which first mentioned HF.

Of note, outpatient attendance with or for HF are not recorded in SMR00, nor in SCI Diabetes (McAllister et al. 2018).

Identifying Incident Heart Failure

Incident HF was identified by excluding individuals whose first record indicated pre-existing HF (Conrad et al. 2018). See Table A.16 For a list of clinical codes used to exclude records of HF where the record refers to an existing, non-*de novo* event.

2.6.2 Comorbidities

Comorbidities were defined as a record in any diagnostic position on or before the date of interest (i.e., date of diagnosis, cohort inclusion, or state change). Records were identified using code lists adapted from CALIBER (Kuan et al. 2019) unless otherwise specified below. Patients without a diagnostic code were assumed to be free from a given condition. This is a standard methodology employed with large-scale, observational EPR datasets (Koudstaal et al. 2017, Conrad et al. 2018).

Where definitions differed from CALIBER, the differences are listed below.

Anaemia

Anaemia was defined using the World Health Organisation (WHO)'s sex-adjusted definition of a haemoglobin concentration less than 12.0 g/dL for women, assuming no pregnancies, and 13.0 g/dL for men (World Health Organization 2011). As the median age of the populations of interest was >70 years and patients were required to have a cardiovascular problem, this seemed justifiable. See Section 2.5.9 for a description of how haemoglobin tests were identified and cleaned.

Atrial Fibrillation/ Flutter

The definition of atrial fibrillation/flutter (AF/AFL) was modified from the CALIBER phenotype definition to include only cases with a coded record of a diagnosis. Using the phenotype's diagnostic classifications, the accepted diagnoses were classified either as historical or diagnosed in primary or secondary care. Diagnoses were not inferred from the prescription of anticoagulants or digoxin because there are other reasons for prescribing these agents. However, AF/AFL may account for the majority of such prescriptions, so the use of these medications was reported.

Cancer

The definition of a history of cancer was modified from the CALIBER phenotype definition by excluding non-melanoma skin cancers. The code list includes 'Other myeloproliferative disease', which CALIBER defined based on chronic or unspecified myelofibrosis, including Waldenstrom macroglobulinaemia and other myeloproliferative conditions listed under ICD-10 Chapter II, but are not included in the Charlson index (D'Hoore et al. 1993).

Chronic Kidney Disease

Chronic kidney disease (CKD) was defined as the presence of two eGFR results (see Section 2.5.9 for calculations) of <60 mL/min/1.73m^2 at least 90 days apart (National Institute for Health and Care Excellence 2021). Of note, the standards do not mention how to classify an individual if he/she has a reading $\geq$60 mL/min/1.73m^2 between the two low readings. Higher readings were ignored in these cases, as repeated low readings suggest chronic renal impairment. The eGFR tests were identified in SCI Store (see Section 2.5.9).

Diabetes Mellitus

Diabetes mellitus (DM) was defined using a combination of the CALIBER phenotype code list and the SCI Diabetes registry (see Section 2.5.8). Patients were considered to have a diagnosis of DM on a given date if they had a record in SMR01, SMR04, GP LES, or deaths which met the CALIBER definition or a registry record indicating a diagnosis.

Ischaemic Heart Disease

Ischaemic heart disease (IHD) was not included as a CALIBER phenotype; therefore, IHD was defined using a combination of previously published research, searching CALIBER for definitions of known components, and mapping Read codes to ICD-10 codes. The Read Code list was obtained based on previously published research (Reeves et al. 2014) submitted to the Manchester Code Repository. This list formed the basis of search terms used to look up known components in CALIBER (e.g., myocardial infarction (MI), coronary thrombosis, unstable angina, stable angina, and coronary artery disease). Where applicable, Read codes from Reeves were mapped directly onto ICD-10 codes. Codes were taken from SMR01, SMR04, GP LES, and deaths.

Stroke

Stroke was defined as a record for at least one of the following: intracerebral haemorrhage, ischaemic stroke, stroke not elsewhere specified, or subarachnoid haemorrhage. The presence

of one of the above types was defined using the specific stroke's CALIBER phenotype across SMR01, SMR04, GP LES, and deaths.

Thyroid Disease

Thyroid disease was defined as using the CALIBER Thyroid disorders code lists for hyperthyroidism and hypothyroidism in either primary or secondary care. Codes were taken from SMR01, SMR04, GP LES, and deaths.

2.6.3 End of Follow-up

Patients were considered to be alive and contributing data under follow-up until the earliest of death, last record across all data sources, the last date of full data coverage on the 31st March 2018, or the predetermined end date (e.g., 1 year or 5 years of follow-up).

Right censoring follow-up at the last available record ensures that people who emigrated out of the region and ceased receiving medical care in Scotland do not turn into apparently 'immortal' patients.

2.7 Data Limitations

This analysis was subject to several data limitations, including how the cohort was built, lack of in-hospital medications, and inherent limitations in using secondary administrative data.

2.7.1 Cohort Build

The dataset was not built for the sole purpose of this research project. As such, the cohort was built according to Section 2.3.2 and missed desirable codes, including codes for atherosclerotic heart disease; any code for a history of MI, including acute, subsequent and current complications following an acute event; angina pectoris; loop diuretics prescribed as combination therapies (e.g., Co-amilofruse [Amiloride hydrochloride/frusemide]; BNF code: 0202040B0); and some comorbidities recorded in conjunction with other conditions (e.g., hypertensive heart and renal disease with both [congestive] heart failure and renal failure;ICD-10 code: I13.2). However, almost all of these patients will have been identified for inclusion in the cohort by other codes or treatments.

We used the enriched definitions of CAD, PAD, HF, ACEi, ARB, MRA, beta-blockers, and LD when working with the cohort (see Table A.3 for medication classifications and A.6 for comorbidity codes). Based on efforts from the NHS GG&C Safe Haven team, we know using these expanded definitions potentially expands the cohort by 2% if every patient were included in the analysis. For this reason, the expanded definitions were used when considering cohort

inclusion dates, comorbidity status, or grouping for analysis to limit the impact of the restricted cohort build.

2.7.2 Lacking In-Hospital Medications

The prescription data from PIS is limited to community-prescribed and dispensed medications, and hospital-prescribed medications dispensed in-community (see Section 2.5.4). Prescription coverage does not include medications dispensed in the hospital. The implication is that the start date for a particular medication will be delayed from the original initial prescription if the said prescription was started in hospital. Patients who are discharged will usually receive a community prescription for ongoing treatment within a few weeks. However, LD initiation would be missed if it was dispensed in hospital and the patient either died in hospital or within a few days of discharge.

2.7.3 Secondary Administrative Data

The data used in this thesis are secondary administrative EPR data. As such, the analysis contained within this thesis is subject to the benefits and drawbacks of using secondary administrative EPR data. While the analysis benefits from the population-level study size, with high coverage levels for over eight years, it is subject to issues like incomplete clinical records, diagnostic clinical error (both omission and commission), miscoding of events or procedures, and incomplete or false data linkage (Harpe 2009).

Reliant on Coding Accuracy

The presence, and hence absence, of a condition, was defined using recorded ICD-10, Read, OPCS-4, and registry codes, where applicable. This relies on the accuracy and completeness of the medical diagnosis and subsequent accuracy of clinical coding. For a condition such as HF, the accuracy of a clinical diagnosis (both omission and commission) may be low. Validation studies suggest that hospital diagnostic codes and procedures can be highly specific in identifying some conditions but fail to identify, and hence underestimate, the actual disease burden (Stavrou et al. 2012, Wilchesky et al. 2004, Khand et al. 2005). Public Health Scotland audits of SMR01 found that the coding accuracy of the main diagnostic code increased by 1.8 points from 96.3% in the 2015 report to 98.1% in 2020 (Public Health Scotland 2019). However, it is not clear that these data apply to heart failure. At a similarly high level, main operations were coded with a 93.8% accuracy, which increased by 3.7 points by the 2019 - 2020 audit. These values indicated that while coding accuracy is high and increasing over time, they are not perfect, and results should be viewed in accordance. By combining multiple data sources in disease and comorbidity definitions, reporting and prediction capabilities are expected to improve (Huang et al. 2013, Ng et al. 2016).

Incomplete General Practice Local Enhanced Services Coverage

GP LES is a source of important clinical information, including blood pressure, smoking status, and body mass index (BMI). Unfortunately, while a relatively high percentage of GP surgeries contributed data (83%), data are only captured for patients who meet the criteria for a condition with enhanced services (see Section 2.5.3), accounting for just 30% of the entire cohort. Where possible, attempts have been made to find substitutes for these missing clinical variables by using surrogates. For example, recording the diagnosis of hypertension using ICD-10 and Read Codes gives an indication of the prevalence. Still, these numbers are an underestimate as hypertension is rarely the primary cause of hospitalisation. The recording of concurrent medications also helps bridge this knowledge gap. For example, ACEis, ARBs, CCBs, and diuretics can be used to treat hypertension, though they are not exclusively used for this purpose.

Missing SMR00 Coding

The major limitation for SMR00 is that while the specialties are listed, it is rare that an ICD-10 code is recorded for any visit outside of psychiatric care. The same applies for OPCS-4 codes. The value of SMR00 is that researchers can tell if an individual had contact with a given specialty service, although not the reason why.

Chapter 3

Characteristics and Prognosis of Patients Receiving Loop Diuretic Therapy With or Without a Diagnosis of Heart Failure

> If the statistics are boring, you've got the wrong numbers.
>
> *Edward Tufte*
> *Professor emeritus of political science, statistics, and computer science at Yale University*

3.1 Introduction

This chapter sets out to describe the population with chronic HF and already receiving repeat prescriptions for LD in the NHS GG&C region.

HF is characterised by water and salt retention that leads to systemic and pulmonary circulation congestion, and eventually to symptoms, such as exertional breathlessness, and signs such as peripheral oedema (Cleland et al. 2021, Pellicori et al. 2016). Guidelines on HF strongly recommend LD for managing symptoms and signs of congestion (National Institute for Health and Care Excellence 2018*a*, McDonagh et al. 2021, Cleland et al. 2021), but LD can also be used to treat resistant hypertension or to manage congestion due to end-stage kidney disease (Cleland et al. 2021). For patients with HF, the development of congestion and the need for LD to manage it is associated with an adverse prognosis (Pellicori et al. 2016). However, many patients appear to be given LD, presumably for ankle swelling or breathlessness, without investigations for the underlying cause; relief of symptoms and signs may then mask the diagnosis of HF (Cleland

et al. 2021). Accordingly, an exploration of the relationship and consequences of prescribing LD and diagnosing HF is required.

3.1.1 Aims

This chapter aims to describe the NHS GG&C population classified by the presence or absence of a diagnosis of HF or recurrent LD on the 1st January 2012 and their outcome over the following 5 years.

3.2 Background

Time-to-event analysis considers whether or not the event of interest occurred and the amount of time the patients were under observation. As such, it is implicit when reporting morbidity and mortality rates.

3.2.1 Survival Analysis

Survival analysis is a subset of time-to-event analysis, where the outcome of interest is death[1] (Rothman et al. 2021). Within the context of survival analysis using administrative data, there are at least three reasons why an individual might not have the death recorded during follow-up. The most straightforward is the individual survived to the end of follow-up. Alternatively, the patient emigrated out of the data capture area, or the patient died, and the death was not captured in the system. To avoid making assumptions about an individual's survival status, right-censoring on the date of the last record indicates that there is only partial knowledge that the patient was event-free as of that date (Lagakos 1979). Right-censoring assumes that censoring is non-informative about the event of interest (Harrell et al. 2015), as censoring follow-up due to changes in health status related to the outcome will introduce bias. One further practical issue for survival analysis is that all patients must have some amount of time past time zero.

Kaplan-Meier Survival Estimator

The Kaplan-Meier (KM) survival estimator, also known as the product-limit approach or estimate, was named after Edward L Kaplan and Paul Meier (Kaplan & Meier 1958, Etikan et al. 2017). Since its publication, it has become an essential survival analysis tool for measuring and visualising survival data while allowing comparison between groups of individuals (Jager et al. 2008, Etikan et al. 2017, Harrell et al. 2015). Besides being easily understood, the KM

[1] Colloquially, the term 'survival analysis' has expanded to include events of interest other than mortality, though both boil down to time-to-event analysis. For this thesis, survival analysis will always have mortality as the single or joint endpoint, although most will hold during standard time-to-event analysis.

survival estimator's primary benefit is that it is a non-parametric estimate[2] of the distribution of survival times, and it accounts for censoring (Etikan et al. 2017, Harrell et al. 2015). In other words, without imposing a distribution, it estimates the survival probability, S, at time t during follow-up as:

$$S_t = \frac{Number\ of\ patients\ living\ at\ the\ start - Number\ of\ patients\ that\ died}{Number\ of\ patients\ alive\ at\ the\ start} \tag{3.1}$$

(Kaplan & Meier 1958, Goel et al. 2010)

The assumptions one makes when using the KM estimator are:

1. Censored individuals have the same probability of experiencing the event as those who remain under follow-up (i.e., censoring is uninformative).
2. Survival probabilities are the same for individuals regardless of the patient's recruitment date.

(Kaplan & Meier 1958, Jager et al. 2008, Harrell et al. 2015). The estimator is limited in its ability to adjust for continuous or multiple covariates, which is where other survival analysis methods such as Cox proportional hazards (Cox PH) (see Section 3.2.1) and Weibull regression methods are useful for estimating covariate-adjusted survival.

Cox Proportional Hazards Regression

Cox proportional hazards (Cox PH) regression is probably the most popular regression model for survival analysis. It was first introduced by David R Cox in 1972 as a way to obtain an estimate of the effect of one or multiple covariates upon time-to-event or survival time. The response variable of the Cox PH model is the hazard function $\lambda(t|X)$ conditional on covariates $X_i, i = 1, ...n$, which assesses the probability that the event of interest (e.g., death) occurred before a specified time, t:

$$\lambda(t|X) = \lambda_0(t) * exp\left[\sum_{i=1}^{n} \beta_i X_i\right] \tag{3.2}$$

$\lambda_0(t)$ is the baseline hazard corresponding to $X_1 = ... = X_n = 0$, and $\beta_1 ... \beta_n$ are the associated regression coefficients (Cox 1972, Harrell et al. 2015, Wynant & Abrahamowicz 2014).

The Cox PH model is semi-parametric as it does not assume an underlying survival distribution such as one would see in a fully parametric model (e.g., a Weibull regression model). Instead, the two assumptions that are made are:

1. The effects of predictor variables upon survival are proportional over time, and,

[2]Non-parametric estimates are effectively an empirical survival function, and as such aren't smooth. For example, due to the observed data, there might be sharp changes in the survival probability that may not reflect the actual survival probability.

2. There is a linear relationship between the log hazard or log cumulative hazard of each covariate (referred to as the linear effects relationship).

(Cox 1972, Harrell et al. 2015)

When Cox PH regression analysis was used in this thesis, the Efron method was used for handling tied events[3] due to its improved accuracy over the Breslow approximation (Hertz-Picciotto & Rockhill 1997, Harrell et al. 2015). Additionally, the Efron method has the added benefit of being the default implemented tie method within R's `survival` package (Therneau & Gramsch 2000) based on similar reasoning.

Finally, due to the large population size and high statistical power to detect small departures from proportional hazards (PH) (see Assumption 1 above), proportional effects were checked visually using log-log plots for categorical variables. The linearity of log hazards for continuous variables (see Assumption 2 above) was assessed visually by splitting the continuous variable into deciles and plotting the mean value against its log hazard (Hosmer et al. 2008).

Time-Dependent Covariates

In standard survival methods such as Cox PH (see above in Section 3.2.1) and Weibull regression, covariate effects are fixed at time 0. In the case of models which assume PH, the covariate effects remain constant over time (Wynant & Abrahamowicz 2014) (see the fixed covariate example in Figure 3.1). In studies with longer follow-up periods, or where covariates change during follow-up, the use of fixed hazards taken at baseline fails to consider these changes. Instead, time-dependent, or time-varying, covariates allow for models to incorporate changes in patient status which occur during follow-up (Fisher & Lin 1999, Wynant & Abrahamowicz 2014, Zhang et al. 2018). An illustrated example of this is shown in the bottom section of Figure 3.1, where the follow-up time for each patient is divided into shorter intervals.

When implementing time-dependent covariates in Cox PH, the equation looks very similar to the traditional model with the addition of a vector to account for the updated values (see Equation 3.3). The function $\lambda_0(t)$ remains the baseline hazards as described in Equation 3.2, but adds a vector of time-dependent covariates, $X(t)$, where

$$\lambda(t|X = x) = \lambda_0(t)exp[\beta' X(t)] \tag{3.3}$$

β is the vector of coefficients (Zhang et al. 2018, Therneau & Grambsch 2000). Of note, as long as the intervals within each patient's follow-up do not overlap, there is only one hazard ratio for the time-dependent covariant as the equations only use information from at one most one row of

[3]Tied events are when events share identical recorded times.

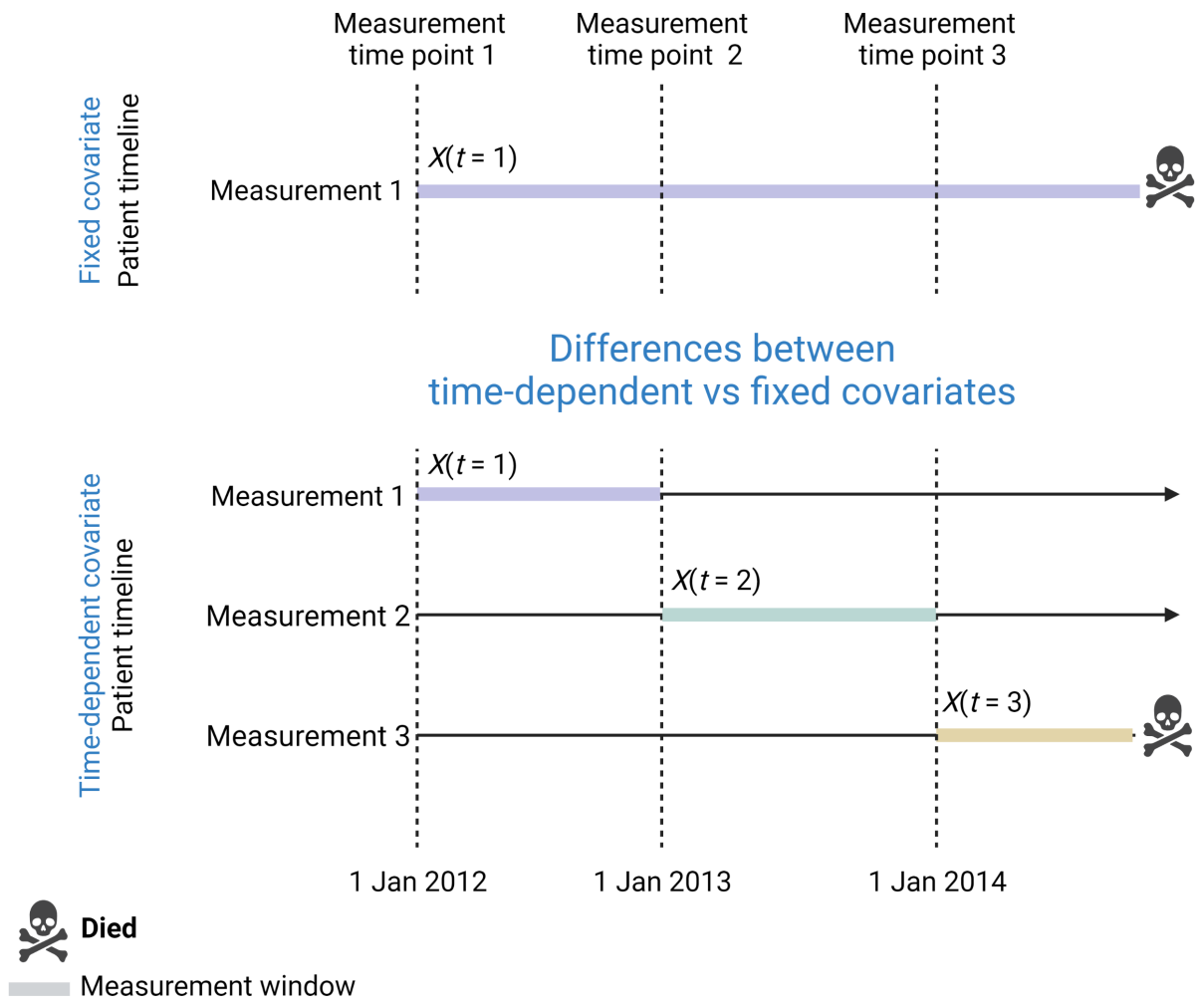


Figure 3.1: Schematic illustration of the difference between using a standard variable set at baseline (top) versus a variable with time-dependent covariates in a Cox PH regression model. If the covariate is fixed and determined at baseline, the effect of the covariate also remains fixed. It must remain proportional for the PH assumption to remain true. With time-dependent covariates, the follow-up time of each subject is divided into shorter intervals, and the covariate value (X) is updated at each time point (t).

the dataset per individual at any time point (see the patient timeline for time-dependent covariate section in Figure 3.1) (Therneau & Grambsch 2000, Zhang et al. 2018).

Competing Risks

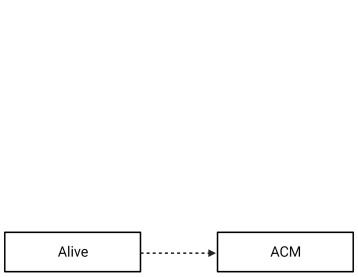
In the standard (non-competing risk) survival analysis, there is either one event of interest, or there is a composite outcome which usually combines a non-fatal event with mortality (see Figure 3.2a). A common example of a composite outcome within HF research is HF hospitalisation or ACM. Where this is the case, standard analysis techniques include KM (see Section 3.2.1) and Cox proportional hazards regression modelling (see Section 3.2.1); however, in settings where a competing event prevents or precludes patients from experiencing the event of interest

or where there are multiple mutually exclusive events, the standard approach to survival analysis no longer holds. For example, if a patient dies before being diagnosed with heart failure, that patient can never develop HF as illustrated in Figure 3.2b. Instead, in this situation, there are multiple paths that a patient can take. Still, the first event of interest (e.g., either HF hospitalisation or ACM) can only occur once per patient (although recurring-events analysis can also be considered).

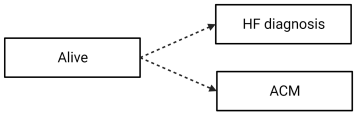
The analysis of competing risks has been used in medical statistics and epidemiology since at least the work of Bernoulli (1760) on the improved life-expectancy after eliminating smallpox. In the modern context, Kalbfleich & Prentice introduced the concept of the cumulative incidence function (CIF) as a non-parametric estimation of the marginal probability that failure or event type j occurs before any specified time t within follow-up

$$CIF_j(t_f) = \sum_{f'=1}^{f} \hat{S}(t_{f'-1}) \times \hat{h}_j(t_{f'}) \tag{3.4}$$

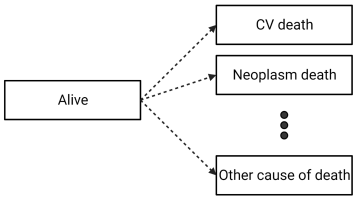
where the probability of experiencing event j at time tf is the product of surviving the prior time points ($\hat{S}(t_{f'-1})$) and cause-specific hazards at time tf ($\hat{h}_j(t_{f'})$) (Kalbfleich & Prentice 2002, Kleinbaum et al. 2012). The CIF allows for estimating the marginal probability for an event of interest in a competing risk setting without requiring one to assume that the competing risks are independent (Kalbfleich & Prentice 2002, Kleinbaum et al. 2012).



(a) Standard survival analysis



(b) Non-fatal event with ACM as a competing risk



(c) Specific causes of death

Figure 3.2: Common scenarios for survival and competing risk analysis

3.2.2 Person-Time at Risk

Within time-to-event analysis, the amount of time a patient is at risk of experiencing an event is referred to as person-time at risk (Porta 2016, Rothman et al. 2021). For example, if a study is investigating the number of HF admissions within a 5-year window for a prevalent HF population, patients accrue time at risk provided they are eligible to be admitted to hospital (i.e., are alive, under follow-up, and are not at currently in hospital). If only the first event is of interest (e.g., first HF readmission after diagnosis), then the patient is no longer at risk after they experience the event.

3.3 Methods

3.3.1 Study Population

Patients were eligible for inclusion if they were alive and aged 18 years or older on 1st January 2012 but were only included if their data were available for at least the preceding 12 months to avoid under-reporting of diagnoses and LD prescribing. Patients could be included if they were dispensed an ACEi, ARB, beta-blocker, MRA, or LD between 31st December 2009 through 31st December 2011 for any reason, including hypertension. The presence of a dispensed medication was identified using BNF codes. see Section 2.5.4 for information on prescribing records and Table A.3 for qualifying codes. Patients could also be included if they had a record of CAD, PAD, or HF in that same time frame, even if not dispensed the above medicines. Diagnoses were identified using either Read or ICD-10 codes in any position (see Table A.6 for codes used to define cohort inclusion).

3.3.2 Patient Identification and Classification

Repeat LD dispensing was defined as the first time a LD was dispensed in two consecutive quarters or if the patient died within 90 days of the first LD prescription (see Section 2.5.4 for further details of identifying a repeat prescription, and Table A.3 for list of qualifying LD medications). The initiation of LD therapy during follow-up was defined as the first time the above definition was met. Prevalent HF was defined according to Section 2.6.1. Patients were excluded if the first record of HF during follow-up indicated a pre-existing diagnosis (see Section 2.6.1 and Table A.16 for qualifying codes), suggesting incomplete data capture (see Figure B.1 for the patient flow diagram). A first diagnosis of HF associated with a fatal hospitalisation (1,093 deaths) was not counted as HF.

Patients were classified at baseline into four discrete groups based on prevalent records of HF and LD therapy prior to the 31st December 2011, as: 'Neither', 'LD Only', 'HF Only', and 'Both: LD + HF'. Where groups are reported without age adjustments, the 'neither' group was split into two subgroups based on baseline age: neither (18-59 yrs) and neither ($\geq$60 yrs) to allow for fairer comparisons against groups with LD and/or HF.

3.3.3 Study Outcomes

Patients were followed from 1st January 2012 until 31st December 2016. Follow-up was right censored at the time of the last available EPR to avoid uncertainty about survival status. The main outcome of interest was 5-year all-cause mortality (ACM). The primary covariate of interest was group membership as defined by LD prescription and diagnosis of HF. The 5-year cumulative incidence of HF, initiation of repeat LD dispensing, cause-specific hospitalisation

rates, and cause-specific mortality were also reported. The cause of hospitalisation (i.e., admission reason) was defined using the primary diagnosis from the first episode of care within an admission (see Section 2.5.5) and mapped onto twelve disease categories. Cause of death was defined using the underlying cause of death (see Section 2.5.1) in a patient's death record, and these were mapped onto five disease categories (see B.1 for mappings).

3.3.4 Patient Characteristics

Baseline characteristics were reported based on patient status as of 1st January 2012, including age, sex (see Section 2.5.2), ethnicity (see Section 2.5.2), quintile of SIMD using the 2012 status (see Section 2.5.2), comorbidities, current medication, most recent blood tests, and, when available, the results from the closest ECG and echocardiogram to 1st January 2012.

Reported comorbidities included: hypertension, DM, thyroid disease, AF/AFL, CAD (including MI), valve disease, PAD, stroke, COPD, cancer, and dementia. The presence of a comorbidity was defined as a recorded diagnosis according to 2.6.2 on or before 1st January 2012. Patients without a diagnosis were assumed to be free from that condition.

In addition to ACEi, ARB, beta-blockers, MRA and LD, other reported medication classes, including calcium channel blockers (CCBs) (including diltiazem/verapamil and dihydropyridines), digoxin, thiazides and thiazide-related diuretics (Thiazides+), low dose aspirin, lipid regulators, bronchodilators, thyroid medications, and hypoglycaemic agents (including insulin). Patients were considered to be on these medications if dispensed in the 180 days prior to, and including, 1st January 2012. Medicines were identified by BNF codes (see Table A.3 for classifications).

The results of the most recent blood tests in the two years prior were reported for haemoglobin (Hb), estimated glomerular filtration rate (eGFR), and serum results for sodium, potassium, urea, bicarbonate, and chloride (see Section 2.5.9 for additional information on the blood tests). Anaemia was defined using the WHO's definition as a haemoglobin <12.0 g/dL for women and <13.0 g/dL for men (see Section 2.6.2). The eGFR was calculated from serum creatinine using the CKD-EPI equation (Levey et al. 2009) without adjusting for ethnicity (see Section 2.5.9).

When available, left atrial (LA) diameter (see Section 2.5.7) and left ventricular ejection fraction (LVEF) (see Section 2.5.7) were obtained from the closest echocardiogram to 1st January 2012 (within the time-window of 31st December 2009 through 31st December 2016) and similarly for the electrocardiogram (ECG), heart rhythm (see Section 2.5.6), heart rate (see Section 2.5.6), QRS duration (see Section 2.5.6), calculated QTc (see Section 2.5.6), and reports of acute myocardial infarction (acute MI).

3.3.5 Statistical Analysis

Patient characteristics were presented as numbers and percentages for categorical data and median (1^{st} - 3^{rd} quartile) values for continuous data. The number and percentage of complete records were reported, and percentages of categorical variables refer to complete cases.

To estimate the prevalence of the cohort within the wider NHS GG&C population, we applied the mid-year 2012 population estimate (National Records of Scotland 2018), using 5-year, sex-stratified, age-bands as the denominator from ages 29 to 90 years plus.

Admission rates were calculated as the number of admissions per patient-year at risk (see Section 3.2.2) where the patient was at risk of being admitted (i.e., alive, not in hospital, and contributing observed EPR). Allowances were made for the competing risk of death when estimating cumulative incidence.

A Cox PH regression model was used to assess between-group differences in all-cause mortality (ACM) with a robust sandwich-type estimator due to the potential lack of statistical independence between chronic comorbidities. Assumptions were checked per Section 3.2.1. The model was adjusted for age by decade centred around 65 years (the median age of the cohort), sex, SIMD, and the presence of comorbidities at baseline. Results are reported as HR with 95% confidence interval (95% CI). The estimated competing risks of death due to cardiovascular disease, infection, neoplasm, or other causes of death were also calculated.

Time-dependent covariates (see Section 3.2.1) were used to assess the impact of disease progression on morbidity and mortality. The dates of HF diagnosis and the start of repeat LD were used to determine group and comorbidity status. (See appendix Section B.5 for a further explanation). Crude 5-year morbidity and mortality rates per patient-year at risk were calculated with time-dependent covariates. Additionally, ACM was modelled using time-dependent covariates where age by decade, sex, and SIMD used 2012 values, while comorbidity status and the groups were updated at each change point.

Statistical analysis was performed using R (see Section 2.4.2) (R Core Team 2021), using the following main packages `survival` (Therneau & Gramsch 2000), `RODBC` (Ripley & Lapsley 2021), `tidyverse` (Wickham et al. 2019), `viridis` (Garnier et al. 2021*b*), `cmprsk` (Gray 2020), `lubridate` (Grolemund & Wickham 2011), `survminer` (Kassambara et al. 2021), `broom` (Robertson et al. 2021), `reshape2` (Wickham 2007), `ggfortify` (Horikoshi & Tang 2018), `gridExtra` (Auguie 2017), `forcats` (Wickham 2021), and `ggpubr` (Garnier et al. 2021*a*).

3.4 Results

The eligibility criteria were met by 198,898 individuals who contributed a median of 5 (1st - 3rd quartiles: 3 - 5) years of follow-up for a total of 898,999 patient-years. Of the estimated NHS GG&C population in 2012, more than 50% of the male population aged >70 years and >75 years for women met the criteria (see Figure 3.3). On January 1st 2012, the cohort included 161,935 (81%) patients with neither a diagnosis of HF nor repeat LD (of whom 89,699 were aged $\geq$ 60 years), 23,963 (12%) who were dispensed repeat LD but had no record of a diagnosis of HF, 5,156 (3%) who had a diagnosis of HF but were not dispensed a LD, and 7,844 (4%) who had both a diagnosis of HF and were dispensed repeat LD. The estimated prevalence of HF for the whole NHS GG&C population (approximately 1 million individuals $\geq$18 years, and 0.2 million <18 years) was 1.3% of the adult population.

Patients with neither a diagnosis of HF nor repeat LD aged $\geq$60 years had a median age of 72 (1st - 3rd quartiles: 66 - 78) years, 48,184 (54%) were women, 16,222 (18%) had DM, 23,638 (26%) had a history of CAD, including 7,916 (9%) who had a history of MI (see Table 3.1). Additionally, 16,603 (20%) had an eGFR <60 mL/min/1.73 m^2 and 16,735 (24%) were anaemic (see Table 3.3). A history of AF/AFL (6,429 [7%]), stroke (7,177 [8%]), cancer (7,308 [8%]), and dementia (1,995 [2%]) were less common. Only 34,038 (38%) had a record of hypertension, although it's likely that this diagnosis was under-recorded in individual datasets available within the GSH/18/CA/002 dataset. For the 10,725 (46% of those with an available echocardiogram) patients with an available measurement, LVEF <50% in 1,138 (11%). For the 18,886 (82% of those with an available echocardiogram) patients with measurements, the LA was dilated in 9,539 (51%) (see Table 3.4).

Patients dispensed LD without a record of a diagnosis of HF had a median age of 75 (1st - 3rd quartiles: 65 - 83) years, were more likely to be women (16,775 [70%]), more likely to have an eGFR <60 mL/min/1.73 m^2 (7,749 [34%]) (see Table 3.3), and anaemia (7,696 [36%]), but less likely to have a diagnosis of CAD (7,390 [31%]) or history of MI (2,485 [10%]) (see Table 3.1). LVEF was <50% for 475 of 3,728 patients (13%) with measurements (see Figure 3.4a), and the LA was dilated in 4,467 of 7,268 patients (61%) according to the criteria that were applied (see Figure 3.4b and Table 3.4).

Patients with a diagnosis of HF who were not dispensed repeat LD were younger with a median age of 69 (1st - 3rd quartiles: 59 - 78) years. They were more likely to be men (3,486 [68%]), to have a diagnosis of CAD (3,860 [75%]), and a history of MI (2,734 [53%]), but less likely to have had an eGFR <60 mL/min/1.73 m^2 (1,007 [20%]), or to have had anaemia (1,220 [28%]). LVEF was <50% for 334 of 1,005 patients (34%) with measurements, and the LA was dilated in 1,060 of 1,787 patients (59%).

Patients with both a diagnosis of HF and treated with repeat LD were older, with a median age of

77 (1^{st} - 3^{rd} quartiles: 68 - 83) years. In addition, they had other adverse prognostic characteristics, including a high prevalence of DM (2,303 [29%]), COPD (2,203 [28%]), a history of CAD (5,266 [67%]), atrial arrhythmias (1,525 [33% of those with an ECG]), and a high proportion with an LVEF <50% (652 of 1,768 patients [37%]) (see Figure 3.4a) and/or dilated left atrium (2,500 of 3,413 patients [73%]) (see Figure 3.4b).

The rates of stroke (10-15%), cancer (8-10%), and dementia (3-6%) were similar for patients with repeat LD, with a diagnosis of HF, or both. These levels were only slightly higher than for patients with neither a diagnosis of HF nor repeat LD aged $\geq$60 years.

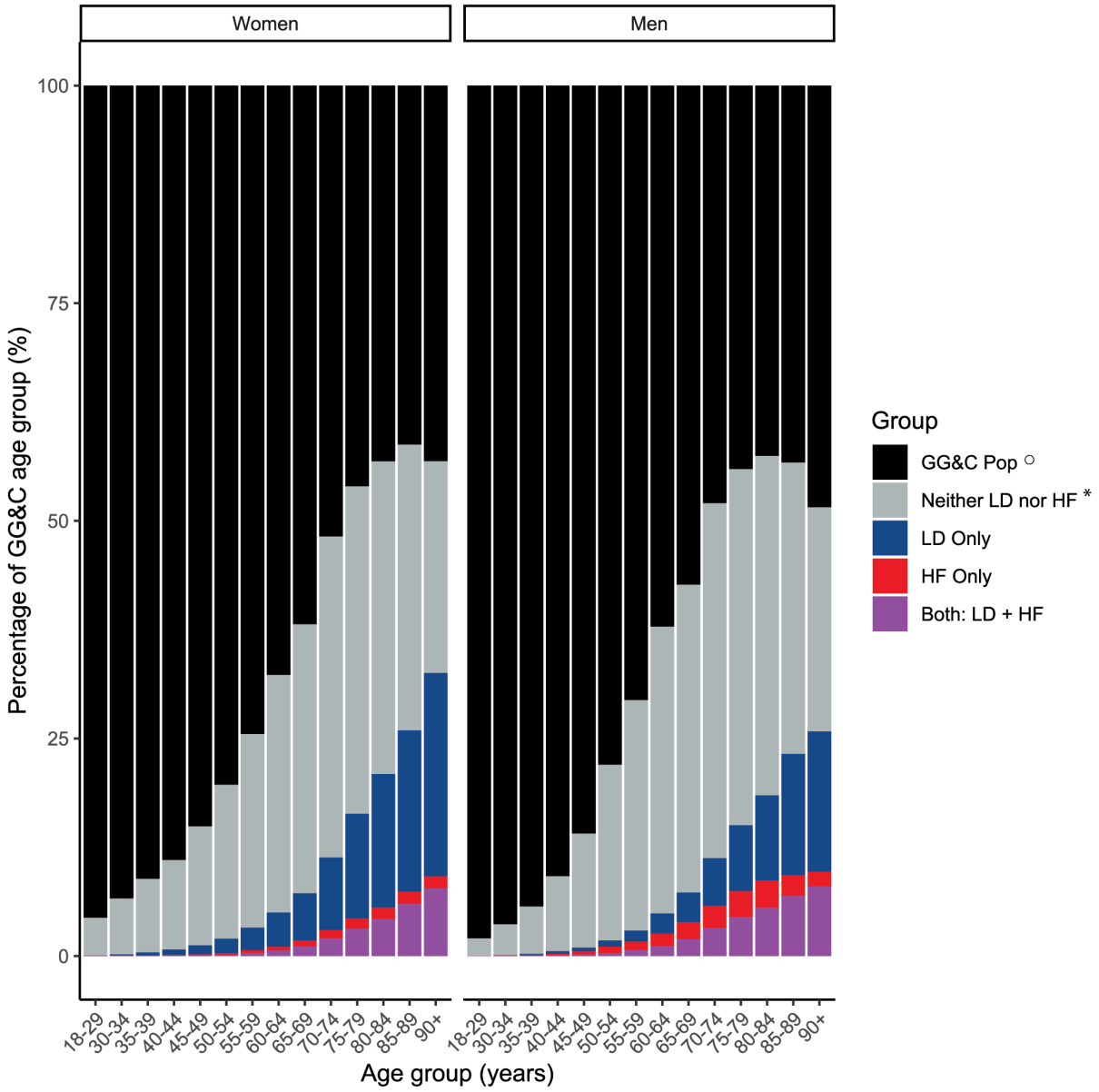


Figure 3.3: Breakdown of NHS GG&C population by sex, age group, and group based on the estimated mid-year population estimate for 2012.

○ Population not taking CV medications (i.e., RAASi, beta-blockers, or LD) or a diagnosis of CAD, PAD, or HF.

∗ Receiving CV medications or diagnosed with CAD or PAD, but neither on repeat LD nor diagnosed with HF.

Table 3.1: Baseline patient demographics and comorbidities as of 1st January 2012.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: LD + HF**
n	72,236	89,699	23,963	5,156	7,844
Age (years)	50 (41 - 55)	72 (66 - 78)	75 (65 - 83)	69 (59 - 78)	77 (68 - 83)
Sex					
Women	38,838 (54%)	48,184 (54%)	16,775 (70%)	1,670 (32%)	3,959 (50%)
Men	33,398 (46%)	41,515 (46%)	7,188 (30%)	3,486 (68%)	3,885 (50%)
Ethnicity					
White	44,164 (61%)	70,461 (79%)	20,660 (86%)	4,425 (86%)	7,151 (91%)
Missing	25,182 (35%)	17,302 (19%)	2,902 (12%)	577 (11%)	523 (7%)
Other	2,890 (4%)	1,936 (2%)	401 (2%)	154 (3%)	170 (2%)
Socioeconomic deprivation (SIMD)					
1 (most deprived)	30,476 (42%)	31,527 (35%)	10,525 (44%)	2,258 (44%)	3,457 (44%)
2	13,046 (18%)	16,252 (18%)	4,524 (19%)	917 (18%)	1,529 (19%)
3	9,857 (14%)	11,914 (13%)	3,236 (14%)	613 (12%)	1,021 (13%)
4	8,170 (11%)	11,904 (13%)	2,636 (11%)	582 (11%)	898 (11%)
5 (least deprived)	10,687 (15%)	18,102 (20%)	3,042 (13%)	786 (15%)	939 (12%)
Comorbidities∗					
H/o hypertension	14,728 (20%)	34,038 (38%)	9,110 (38%)	2,312 (45%)	4,104 (52%)
DM	8,824 (12%)	16,222 (18%)	4,858 (20%)	952 (18%)	2,303 (29%)
Thyroid disease	709 (1%)	2,013 (2%)	1,083 (5%)	194 (4%)	502 (6%)
CAD	7,432 (10%)	23,638 (26%)	7,390 (31%)	3,860 (75%)	5,266 (67%)
Of which is MI	3,420 (5%)	7,916 (9%)	2,485 (10%)	2,734 (53%)	3,157 (40%)
Valve disease	323 (< 1%)	1,338 (1%)	1,139 (5%)	507 (10%)	1,570 (20%)

Continuation of Table 3.1

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	LD Only	HF Only	Both: LD + HF
n	72,236	89,699	23,963	5,156	7,844
AF/AFL	1,022 (1%)	6,429 (7%)	3,893 (16%)	1,232 (24%)	3,547 (45%)
PAD	623 (1%)	2,431 (3%)	923 (4%)	291 (6%)	670 (9%)
Stroke	1,668 (2%)	7,177 (8%)	2,385 (10%)	609 (12%)	1,183 (15%)
COPD	3,321 (5%)	8,636 (10%)	4,616 (19%)	895 (17%)	2,203 (28%)
Cancer	1,383 (2%)	7,308 (8%)	2,280 (10%)	417 (8%)	796 (10%)
Dementia	28 (< 1%)	1,995 (2%)	1,195 (5%)	143 (3%)	456 (6%)

∗ History of a coded record on or before 1 January 2012.

H/o, History of;

CAD, coronary artery disease; MI, myocardial infarction;AF/AFL, atrial fibrillation/flutter;

COPD,chronic obstructive pulmonary disease.

Table 3.2: Baseline medications based on a dispensed prescription within the 180 days before 1st January 2012.

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	LD Only	HF Only	Both: LD + HF
n	72,236	89,699	23,963	5,156	7,844
Age (years)	50 (41 - 55)	72 (66 - 78)	75 (65 - 83)	69 (59 - 78)	77 (68 - 83)
ACEi or ARB	34,436 (48%)	60,624 (68%)	11,638 (49%)	3,992 (77%)	5,769 (74%)
ACEi	27,746 (38%)	45,540 (51%)	8,498 (35%)	3,320 (64%)	4,572 (58%)
ARB	7,459 (10%)	16,406 (18%)	3,687 (15%)	785 (15%)	1,465 (19%)
Beta-blocker	27,941 (39%)	40,555 (45%)	8,512 (36%)	3,542 (69%)	5,054 (64%)
MRA	462 (1%)	600 (1%)	971 (4%)	173 (3%)	1,179 (15%)
CCB	12,901 (18%)	32,213 (36%)	7,516 (31%)	1,220 (24%)	1,737 (22%)
Diltiazem/Verapamil	1,045 (1%)	3,872 (4%)	1,801 (8%)	254 (5%)	351 (4%)

Continuation of Table 3.2

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: LD + HF**
n	72,236	89,699	23,963	5,156	7,844
Dihydropyridine	11,907 (16%)	28,485 (32%)	5,784 (24%)	982 (19%)	1,403 (18%)
Digoxin	115 (< 1%)	1,461 (2%)	1,792 (7%)	366 (7%)	1,772 (23%)
Thiazides+	11,151 (15%)	31,103 (35%)	1,499 (6%)	591 (11%)	307 (4%)
Low dose aspirin	12,091 (17%)	40,474 (45%)	10,727 (45%)	3,383 (66%)	4,409 (56%)
Oral anticoagulants	646 (1%)	4,114 (5%)	2,789 (12%)	722 (14%)	2,253 (29%)
Lipid regulators	21,382 (30%)	57,287 (64%)	13,975 (58%)	3,999 (78%)	5,743 (73%)
Bronchodilators	6,691 (9%)	11,580 (13%)	5,988 (25%)	839 (16%)	2,108 (27%)
Thyroid medications	3,375 (5%)	8,360 (9%)	3,218 (13%)	350 (7%)	949 (12%)
Hypoglycaemic Agents	8,049 (11%)	13,689 (15%)	4,390 (18%)	694 (13%)	1,862 (24%)
Insulin†	2,085 (3%)	2,030 (2%)	1,196 (5%)	146 (3%)	650 (8%)
Other hypoglycaemic agents	6,781 (9%)	12,771 (14%)	3,798 (16%)	626 (12%)	1,515 (19%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

† Either alone or in combination with another agent.

Dilt/Verap, Diltiazem/Verapamil; Thiazides+, thiazides and related.

Table 3.3: Baseline patient blood tests based on the most recent value in the prior two years before 1^{st} January 2012.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: LD + HF**
n	72,236	89,699	23,963	5,156	7,844
Haemoglobin (Hb) (g/dL)					
Hb available	49,507 (69%)	69,406 (77%)	21,125 (88%)	4,427 (86%)	7,295 (93%)
Women: Hb	13.3 (12.5 - 14.0)	12.9 (12.0 - 13.8)	12.6 (11.5 - 13.6)	12.6 (11.6 - 13.6)	12.2 (11.1 - 13.3)
Men: Hb	14.9 (14.0 - 15.7)	14.1 (13.0 - 15.1)	13.3 (11.9 - 14.5)	14.1 (13.0 - 15.1)	13.2 (11.7 - 14.4)

Continuation of Table 3.3

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: LD + HF**
n	72,236	89,699	23,963	5,156	7,844
Anaemic ◇	5,872 (12%)	16,735 (24%)	7,696 (36%)	1,220 (28%)	3,247 (45%)
Estimated glomerular filtration rate (eGFR)∇ (mL/min/1.73m^2)					
eGFR available	57,305 (79%)	83,815 (93%)	22,575 (94%)	4,962 (96%)	7,656 (98%)
eGFR	99 (90 - 106)	78 (64 - 88)	71 (52 - 85)	81 (64 - 92)	61 (43 - 79)
eGFR [20 - 60)	1,797 (3%)	16,620 (19%)	7,243 (32%)	935 (19%)	3,497 (48%)
eGFR <20	132 (<1%)	238 (<1%)	506 (2%)	72 (1%)	214 (3%)
Serum results (mmol/L)					
Urea available	57,362 (79%)	83,839 (93%)	22,582 (94%)	4,963 (96%)	7,656 (98%)
Urea	4.8 (3.9 - 5.8)	6.0 (4.9 - 7.3)	6.4 (5.0 - 8.4)	5.8 (4.7 - 7.3)	7.6 (5.8 - 10.3)
Sodium available	57,308 (79%)	83,823 (93%)	22,574 (94%)	4,963 (96%)	7,656 (98%)
Sodium	139 (137 - 140)	139 (137 - 140)	139 (137 - 141)	139 (137 - 140)	139 (137 - 141)
Chloride available	57,301 (79%)	83,811 (93%)	22,570 (94%)	4,963 (96%)	7,656 (98%)
Chloride	104 (102 - 106)	104 (101 - 106)	103 (101 - 106)	104 (102 - 106)	103 (100 - 105)
Potassium available	56,855 (79%)	83,501 (93%)	22,516 (94%)	4,947 (96%)	7,647 (97%)
Potassium	4.2 (4.0 - 4.5)	4.3 (4.0 - 4.5)	4.2 (3.9 - 4.5)	4.3 (4.1 - 4.6)	4.2 (3.9 - 4.6)
Bicarbonate available	35,246 (49%)	51,731 (58%)	14,836 (62%)	3,059 (59%)	5,011 (64%)
Bicarbonate	23 (21 - 25)	24 (22 - 26)	24 (21 - 26)	23 (21 - 25)	24 (21 - 26)

Data are frequencies(%) for categorical values or median (1st - 3rd quartile) for continues values.

◇ Using the WHO definition of anaemia assuming no pregnancies.

∇ Calculated eGFR assuming no pregnancies and without adjusting for ethnicity (see Section 2.5.9).

Table 3.4: Closest ECG and echocardiogram to 1st January 2012 between 31st December 2009 and censoring.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: LD + HF**
n	72,236	89,699	23,963	5,156	7,844
ECG available	24,516 (34%)	42,515 (47%)	13,302 (56%)	2,908 (56%)	4,654 (59%)
Heart rate (bpm)	75 (65 - 88)	73 (62 - 86)	76 (65 - 90)	70 (60 - 83)	74 (64 - 88)
QRS duration (ms)	88 (80 - 96)	88 (80 - 98)	88 (80 - 100)	96 (86 - 112)	100 (86 - 128)
Heart rhythm					
Sinus	23,786 (97%)	35,890 (84%)	10,070 (76%)	2,243 (77%)	2,622 (56%)
AF/AFL/SVT	583 (2%)	5,448 (13%)	2,629 (20%)	513 (18%)	1,525 (33%)
Other	106 (< 1%)	878 (2%)	449 (3%)	132 (5%)	405 (9%)
Undetermined	41 (< 1%)	299 (1%)	154 (1%)	20 (1%)	102 (2%)
QTc available	23,522 (96%)	40,648 (96%)	12,634 (95%)	2,809 (97%)	4,467 (96%)
QTc (ms)‡	414 (401 - 429)	421 (405 - 438)	426 (408 - 446)	427 (408 - 448)	435 (412 - 462)
Prolonged QTc‡	493 (9%)	2,141 (14%)	1,051 (18%)	327 (22%)	848 (32%)
ST-T abnormality	3,944 (16%)	10,741 (25%)	4,359 (33%)	954 (33%)	1,858 (40%)
Acute MI recorded in hosp	68 (< 1%)	172 (< 1%)	49 (< 1%)	15 (1%)	29 (1%)
Acute MI by ECG	97 (<1%)	213 (1%)	64 (< 1%)	25 (1%)	39 (1%)
Echocardiogram available	8,988 (12%)	23,136 (26%)	8,842 (37%)	2,197 (43%)	4,220 (54%)
LVEF available	4,649 (52%)	10,725 (46%)	3,728 (42%)	1,005 (46%)	1,768 (42%)
LVEF (%)	67 (61 - 73)	66 (59 - 72)	66 (58 - 72)	57 (44 - 67)	57 (42 - 67)
≥ 50%	4,270 (92%)	9,587 (89%)	3,253 (87%)	661 (66%)	1,116 (63%)
(41% - 49%)	213 (5%)	607 (6%)	234 (6%)	159 (16%)	253 (14%)
≤ 40%	166 (4%)	531 (5%)	241 (6%)	185 (18%)	399 (23%)

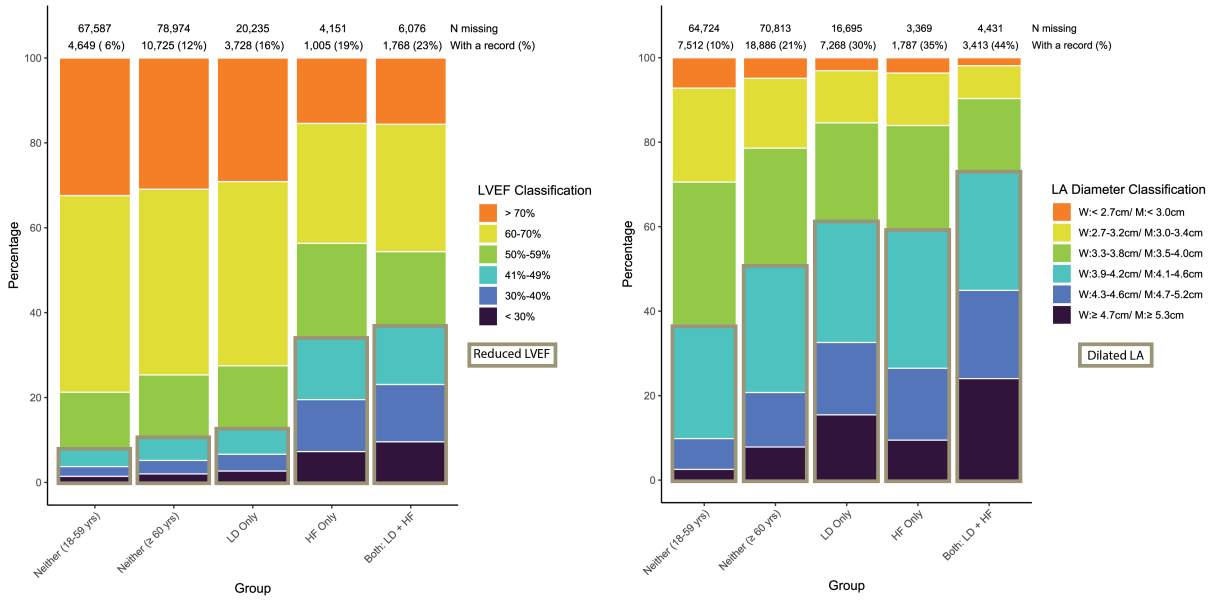
Continuation of Table 3.4

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: LD + HF**
n	72,236	89,699	23,963	5,156	7,844
LA diameter available	7,512 (84%)	18,886 (82%)	7,268 (82%)	1,787 (81%)	3,413 (81%)
Women: LA diameter (cm)	4.1 (3.9 - 4.4)	4.2 (3.9 - 4.4)	4.4 (4.1 - 4.8)	4.3 (4.1 - 4.7)	4.6 (4.2 - 5.0)
Men: LA diameter (cm)	4.4 (4.2 0 4.7)	4.5 (4.3 - 4.9)	4.7 (4.3 - 5.1)	4.6(4.3 - 5.0)	4.8 (4.4 - 5.3)
LA dilated	2,760 (37%)	9,539 (51%)	4,467 (61%)	1,060 (59%)	2,500 (73%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

‡ Corrected QT interval using the Fridericia formula (see Section 2.5.6).

LA, Left Atrial.



(a) Left ventricular ejection fraction (LVEF) (b) Left atrial (LA) diameter

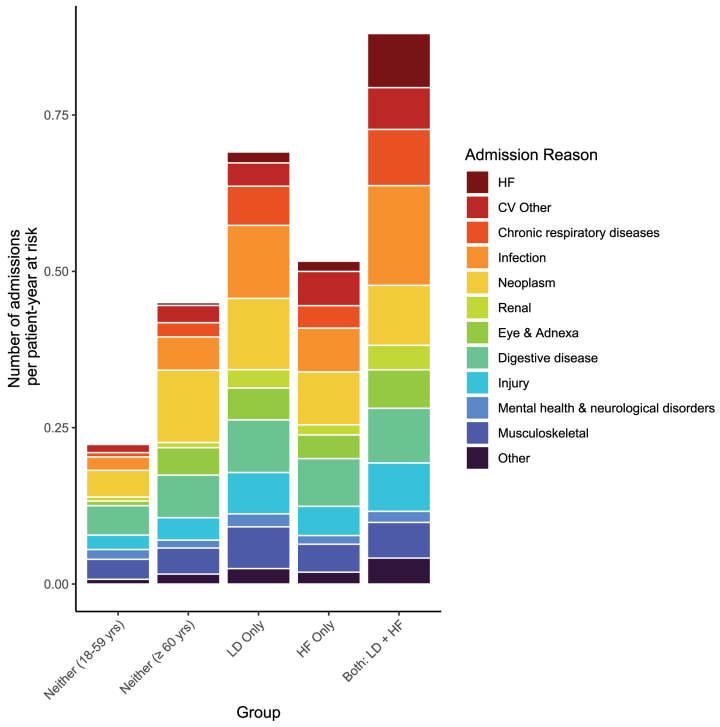
Figure 3.4: Breakdown of available recorded measurements taken at the closest echocardiogram by baseline cohort group. Grey boxes indicate reduced LVEF or dilated LA, respectively.

3.4.1 Changes in Classification Over Time

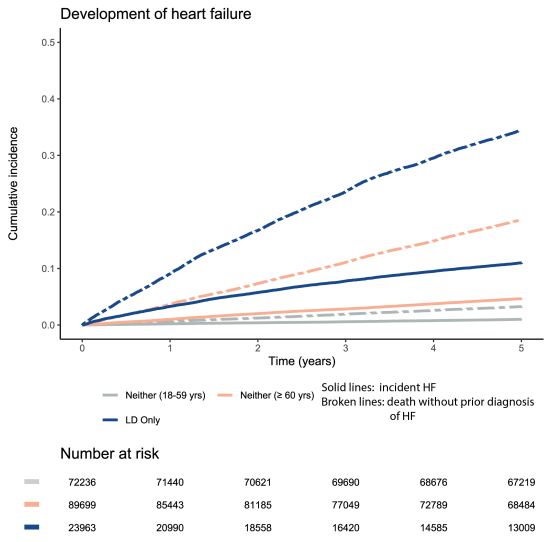
During 5 years of follow-up, 1,119 (22%) patients with a diagnosis of HF were initiated on repeat LD. Conversely, 2,635 (11%) patients with repeat LD subsequently received a diagnosis of HF (see Figures 3.5b and 3.5c). The sequence of events for each classification of patients over the 5 years of follow-up is shown in Figure 3.13.

3.4.2 Hospitalisations

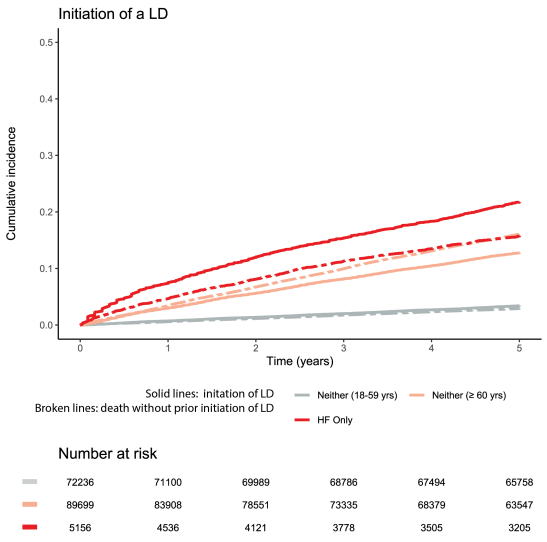
Compared to patients who had neither a diagnosis of HF nor repeat LD, rates of hospital admission per patient-year at risk were higher for those on a repeat LD whether or not they had a diagnosis of HF. Patients with neither a diagnosis of HF nor repeat LD aged 18-59 years were admitted 123,362 times over the 5 years of follow-up (0.36 admissions per patient-year at risk; with 38,667 [54%] having at least one admission). Patients with neither a diagnosis of HF nor repeat LD aged $\geq$60 years were admitted 225,847 times over the 5 years of follow-up (0.61 admissions per patient-year at risk; with 62,921 [70%] having at least one admission). Patients taking a LD who did not have a diagnosis of HF at baseline were admitted 107,730 times over the 5 years of follow-up (0.99 admissions per patient-year at risk with 26,933 [75%] having at least one admission). Patients with a diagnosis of HF who did not receive a repeat LD at baseline were admitted 22,701 times over the 5 years of follow-up (0.93 admissions per patient-year at risk with 6,700 [77%] having at least one admission). Patients with both a diagnosis of HF and repeat LD were admitted 58,245 times over the 5 years of follow-up (1.51 admissions per



(a) Raw hospital rate classified by admission reason



(b) Development of HF



(c) Initiation of a LD

Figure 3.5: 5-year morbidity. (a) Hospital admission rate. (b) Cumulative incidence of a diagnosis of HF during follow-up with ACM acting as the competing risk. Patients are grouped based on the presence or absence of a repeat prescription for LD at baseline and age where LD was absent. (c) Cumulative incidence of the initiation of a repeat LD during follow-up with ACM acting as the competing risk classified by presence or absence of a diagnosis of HF and age where a diagnosis of HF was absent.

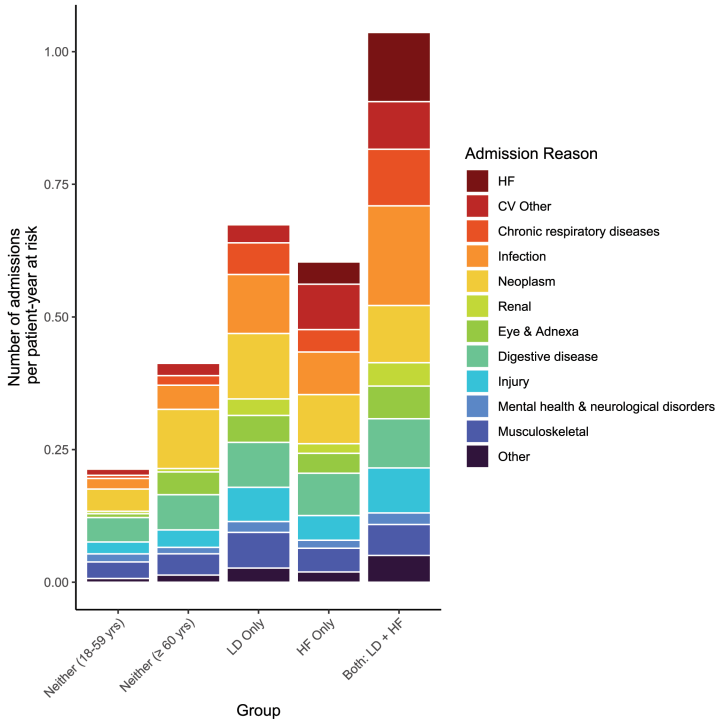
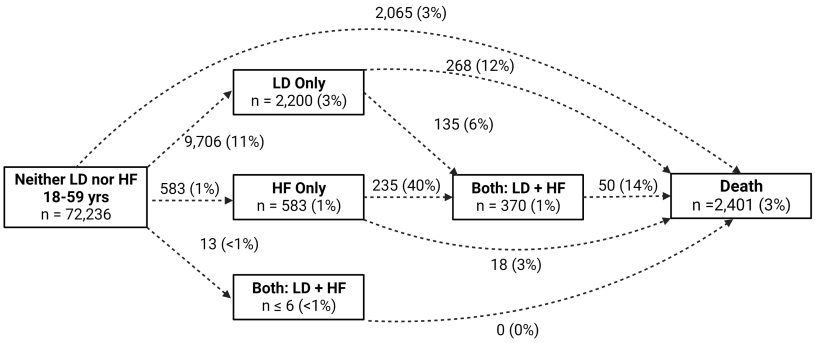
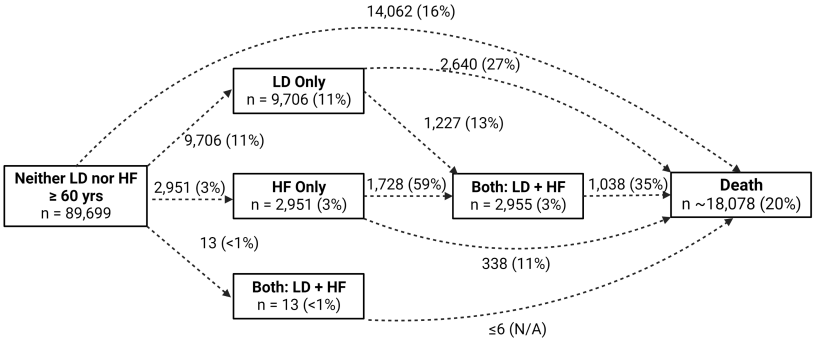


Figure 3.6: Admission rate per patient-year at risk by type of hospital admission and group, where group status is a time-dependent covariate between 1st January 2012 and 31st December 2016.

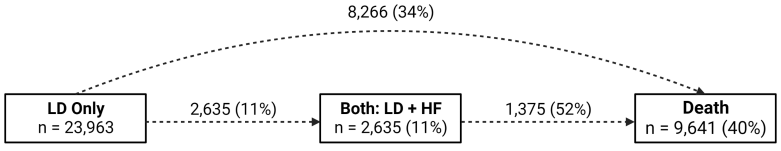
patient-year at risk with 12,923 [87%] having at least one admission) (see Figure 3.5a). Time-dependent analysis shows that rates of admissions increased based on updated HF and LD status (see Figures 3.8 and 3.6), especially for cardiovascular and infection-related admissions. HF or other CV problems accounted for only a small minority of admissions for all patient groups.



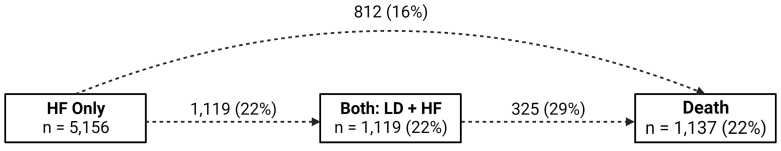
(a) Neither (18-59 yrs)



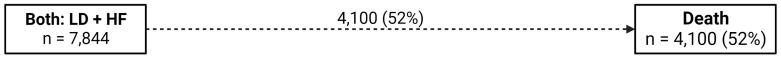
(b) Neither ($\geq$60 yrs)



(c) LD Only

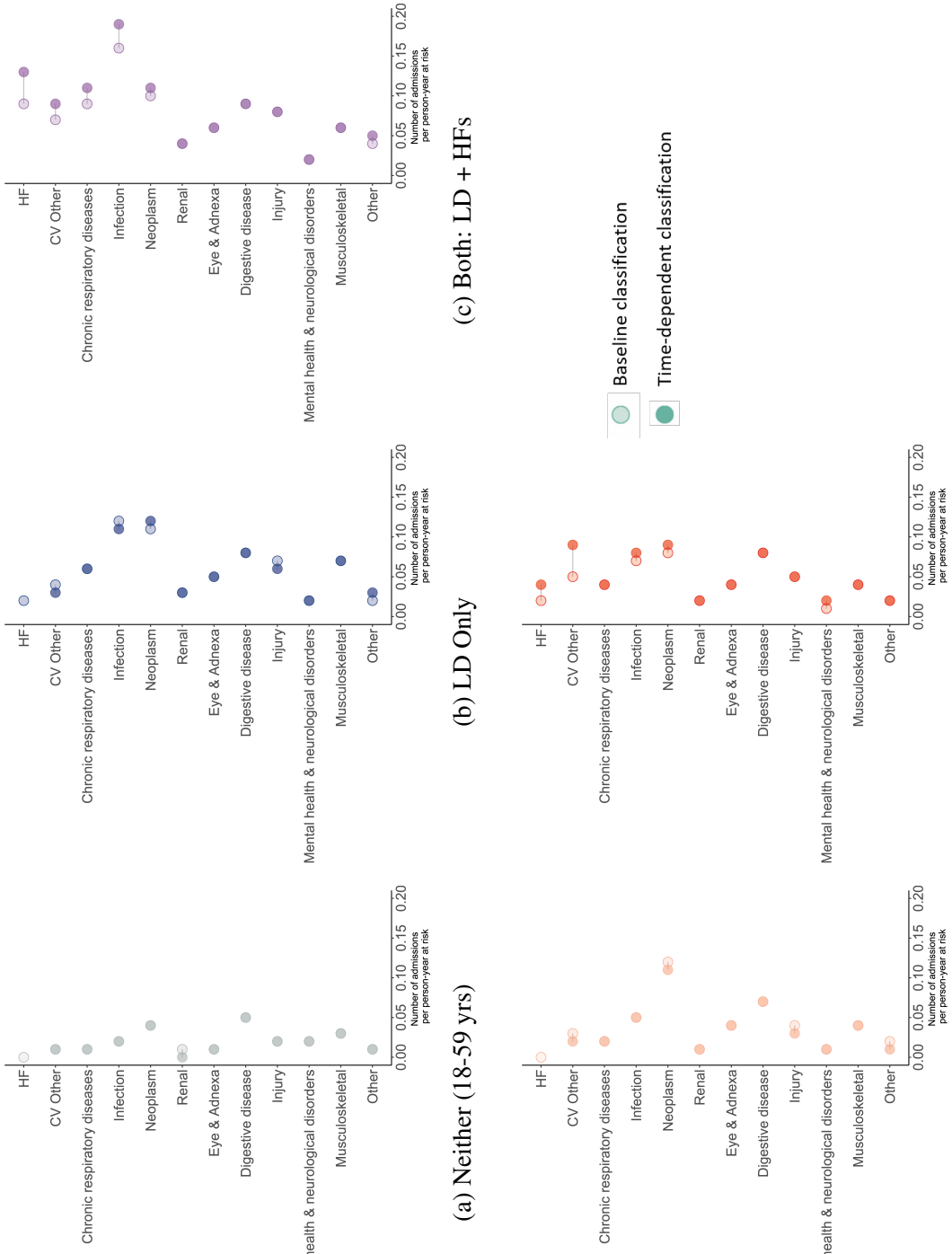


(d) HF Only



(e) Both: LD + HF

Figure 3.7: Transition diagrams illustrating how many patients started in each baseline group (left most boxes), and how many were subsequently diagnosed with HF, initiated on LD therapy, and died between 1st January 2012 through 31st December 2016. The percentages in the boxes are calculated using the baseline group size as the denominator, while transitions are calculated based on those eligible for each transition.



(a) Neither (18-59 yrs)

(b) LD Only

(c) Both: LD + HFs

(d) Neither ($\geq$60 yrs)

(e) HF Only

Figure 3.8: Admission rates using baseline classification or time-dependent changes in classification. If circles completely overlap, only time-dependent results can be seen.

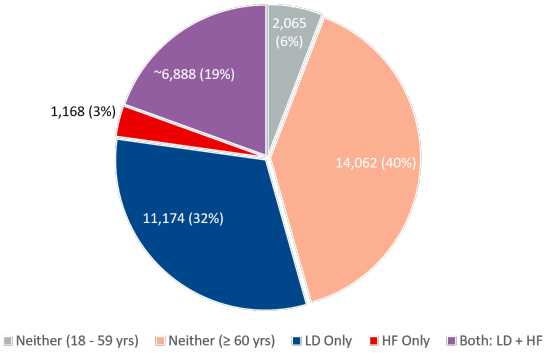
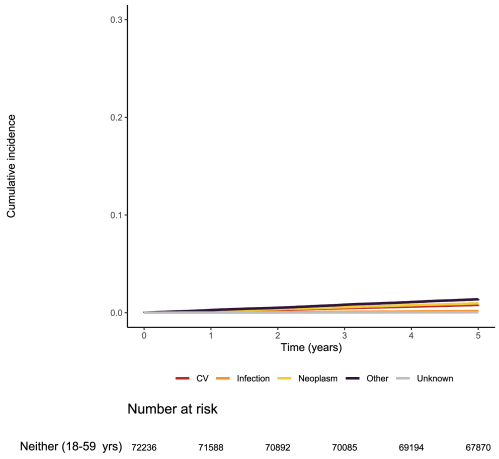


Figure 3.9: A breakdown of the total deaths by group status using time-dependent groups based on LD and HF status among people on cardiovascular therapy (predominantly for hypertension) or with a prior diagnosis of CAD, PAD, or HF.

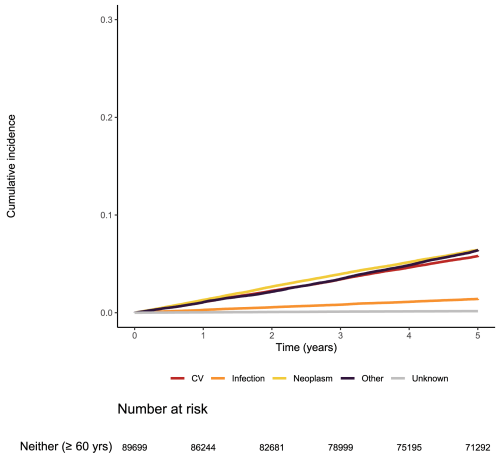
3.4.3 Mortality

For mortality, the PH assumption was not met for the first 14 days, but this applied to only 328 deaths (0.9% of all deaths) (see appendix Section B.4 for further information). Accordingly, the PH assumption was considered a reasonable summary of between-group differences. Using the 'neither' group as a reference, the adjusted HR (which includes age) for 5-year ACM for those dispensed LD without a diagnosis of HF was 1.8 (95% CI: 1.8 -1.9), for those with a diagnosis of HF who were not dispensed LD, the HR was 1.2 (95% CI: 1.1 - 1.3), and for those with both a diagnosis of HF and dispensed LD was 2.3 (95% CI: 2.2 - 2.4) (see Figure 3.11). Applying time-dependent covariates increased HRs, particularly for patients with both HF and dispensed repeat LD (see Figure 3.12).

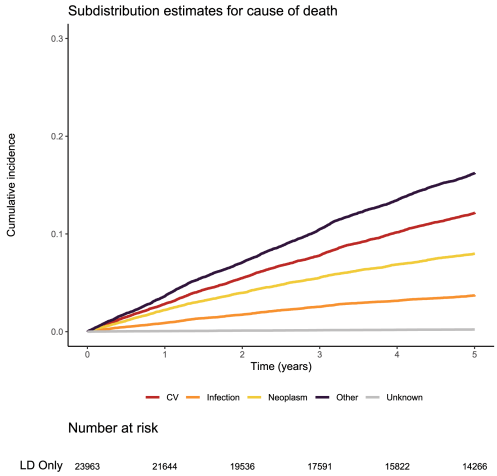
Using baseline classification, cardiovascular mortality at five years for patients who had neither a diagnosis of HF nor were dispensed LD aged below and above 60 years was 1% and 6% respectively (3% and 20% for ACM), for patients with a diagnosis of HF who were not on repeat LD was 8% (22% for ACM), for patients dispensed LD who did not have a diagnosis of HF was 12% (40% ACM). For patients with a diagnosis of HF and repeat LD, it was 23% (52% for ACM) (see Figure 3.10). There were similar rates of deaths from neoplasms for each of these groups (see Figures 3.10 for cumulative incidence and 3.13 for changes with time-dependent covariates).



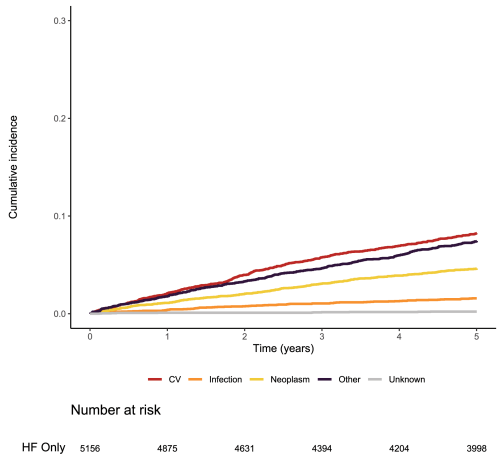
(a) Neither (18-59 yrs)



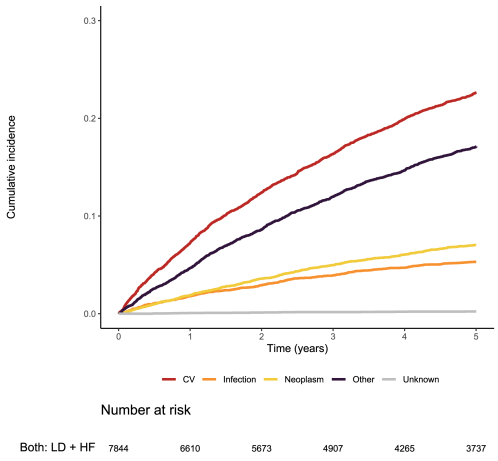
(b) Neither (≥60 yrs)



(c) LD Only

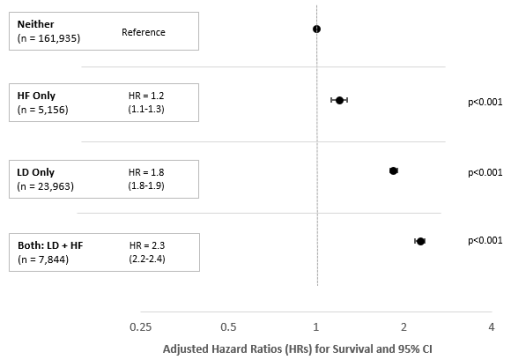


(d) HF Only

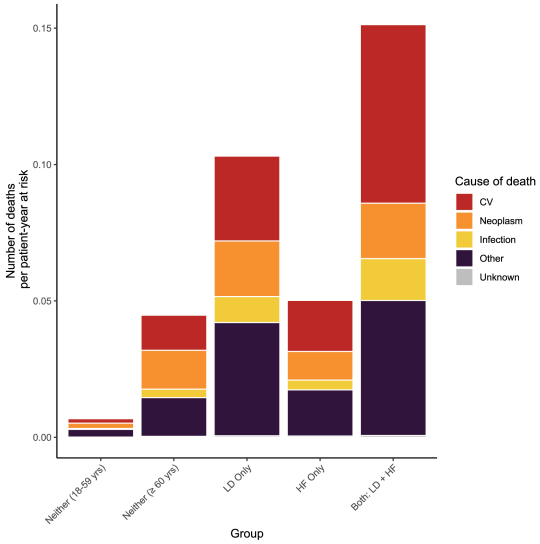


(e) Both: LD + HF

Figure 3.10: 5-year subdistribution of causes of death by group.

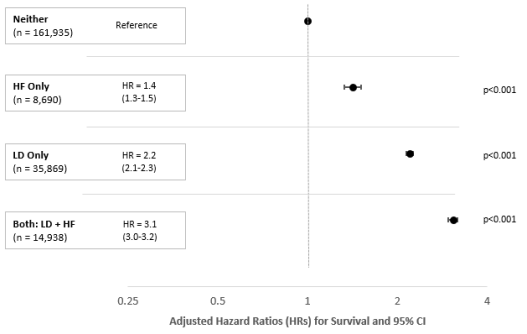


(a) Forest plot of 5-year mortality using baseline group classification

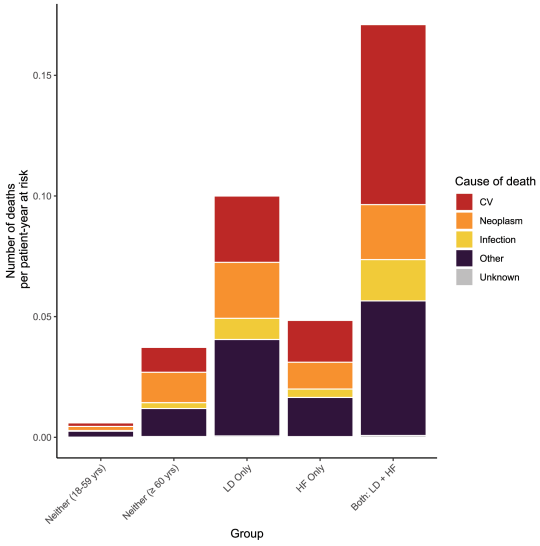


(b) Cause specific mortality rate per patient-year at risk

Figure 3.11: 5-year mortality using baseline classification. (a) Forest plot of HR with 95% CI for ACM by baseline group between 1st January 2012 through 31st December 2016. The model was adjusted for age per decade (centred around the population median of 65 years), sex, SIMD, and comorbidity status. (b) Causes of death per patient-year at risk from 1st through 31st December 2016 classified by the presence of a repeat prescription for LD and a diagnosis of HF.

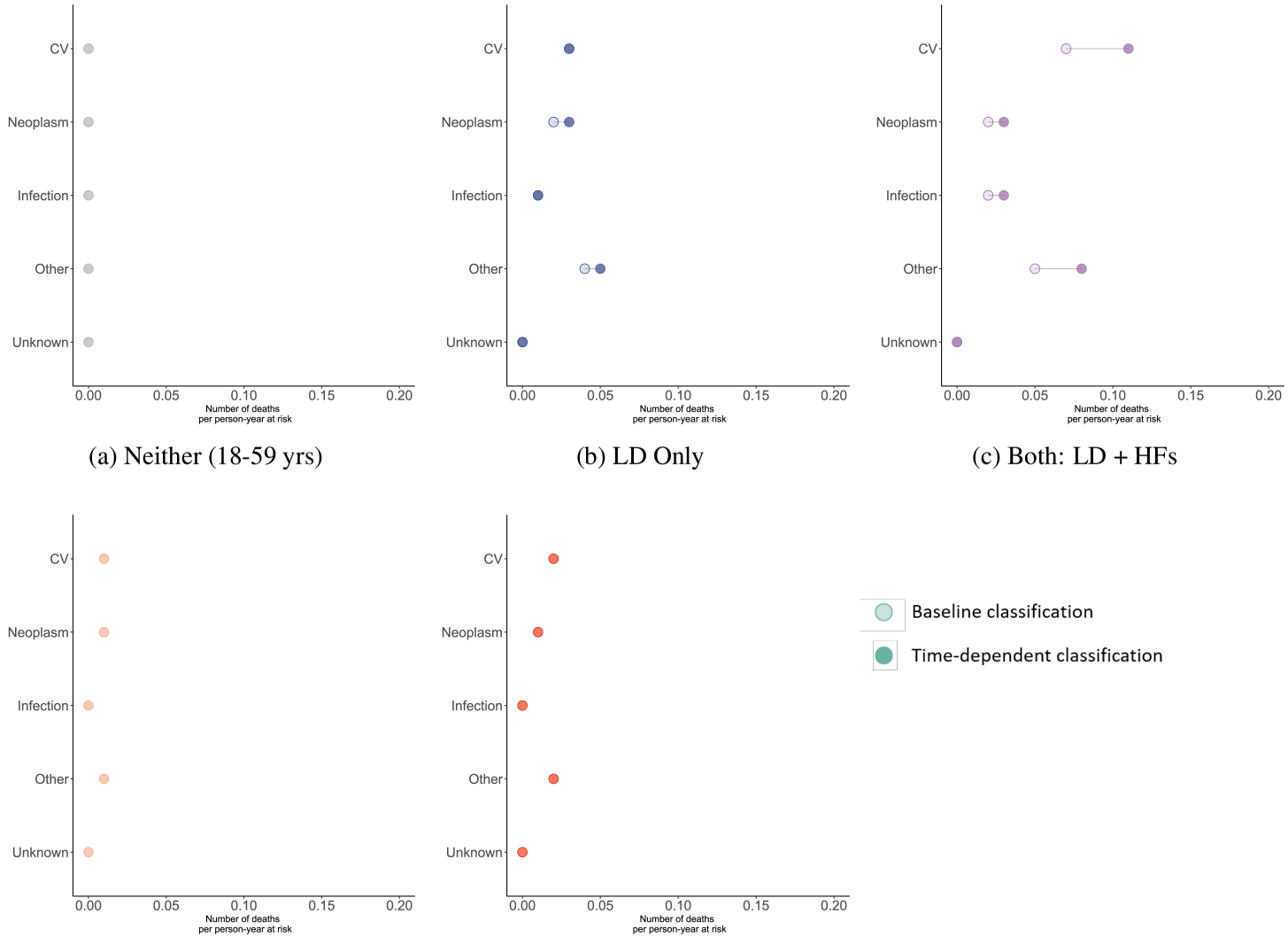


(a) Forest plot of 5-year mortality using time-dependent group classification



(b) Cause specific mortality rate per patient-year at risk

Figure 3.12: 5-year mortality using time-dependent group classification. (a) Forest plot of HR with 95% CI for ACM by time-dependent group status between 1st January 2012 through 31st December 2016. The model was adjusted for age per decade on 1st January 2012 (centred around the population median of 65 years), sex, SIMD status in 2012, and updated comorbidity status. (b) Causes of death per patient-year at risk from 1st through 31st December 2016 classified by the time-varying presence of a repeat prescription for LD and a diagnosis of HF.



(a) Neither (18-59 yrs)

(b) LD Only

(c) Both: LD + HFs

(d) Neither (≥60 yrs)

(e) HF Only

Figure 3.13: Mortality rates using baseline classification or time-dependent changes in classification. If circles completely overlap, only time-dependent results can be seen.

3.5 Discussion

The analysis suggests that for patients with CV disease, mortality is more closely associated with taking LD than with a diagnosis of HF only, even after adjusting for age and other risk factors. The estimated prevalence of HF in Glasgow of 1.3% is broadly consistent with results reported elsewhere in the United Kingdom (Conrad et al. 2018). Still, many more patients (3.2%) were receiving repeated LD prescriptions. Only one in four patients treated with LD had a diagnosis of HF, and only 11% were subsequently diagnosed with HF over the following five years. The prognosis of patients treated with LD, even without a diagnosis of HF, was substantially worse than that of patients with HF who were not taking LD. Those patients had only a slightly worse prognosis than patients in the neither ($\geq$60 yrs) group with other cardiovascular diseases, predominantly hypertension, who were not treated with a repeat LD. Patients with HF taking LD had the worst prognosis. Hospitalisation rates were also higher for patients taking LD with or without a diagnosis of HF. However, regardless of patient classification, the primary reason for admission was usually for conditions other than CV causes. In summary, being treated with repeat LD is a marker of a more extensive and serious health problem than a diagnosis of HF only. This might be explained by a substantial under-diagnosis of HF amongst patients with prescribed LD.

The observation that LDs are associated with an increased risk of an adverse prognostic event in the absence of a diagnosis of HF is not unique to the NHS GG&C population. International trials of AF (Cleland et al. 2007), type-2 DM (Pellicori et al. 2021) show that patients treated with LD often do not have a diagnosis of HF, yet have worse outcomes than those with a diagnosis of HF who are not treated with LD. Those with both a diagnosis of HF and receiving LD had the worst outcome. A study of LD prescribing in patients with AF in England also found that patients treated with LD in the absence of HF had a similar prognosis to patients with HF (Zakeri et al. 2021).

The current criteria used to define HF (McDonagh et al. 2021) are not robust, especially within the context of routinely collected data, as they rely heavily on the identification and recording of symptoms and signs, such as pitting oedema, ankle swelling, and breathlessness. These symptoms and signs lack specificity and may not be obvious until the HF is at an advanced stage (Cleland et al. 2021, Bozkurt et al. 2021). The diagnosis of HF is often not made until symptoms and signs are so severe that the patient needs to be hospitalised (Conrad et al. 2018, Mosterd & Hoes 2007, Shah et al. 2009). In this analysis, diagnostic uncertainty exists as only a few patients initiated on repeat LD had an available echocardiogram (37%). Additionally, natriuretic peptides were not available to make or exclude the diagnosis. A normal LVEF and the absence of severe valve disease might provide false reassurance that the patient does not have HF. Within the LD only group, many patients had a history of hypertension, DM, anaemia, AF/AFL, and impaired kidney function, which are common comorbidities that may cause or

exacerbate HF. The diagnosis of HFpEF, particularly when there are other causes of symptoms and signs such as COPD or kidney disease, is difficult. Many patients treated with LD who did not have a diagnosis of HF were older women with LA dilation, which is consistent with the demographic profile and diagnosis of HFpEF. This diagnostic uncertainty might explain why many patients were treated with LD but were not diagnosed with HF.

Besides treating HF, the other medical reasons for prescribing LD are the treatment of end-stage kidney disease or resistant hypertension. This analysis shows that patients treated with LD without a diagnosis of HF rarely have severe kidney dysfunction. Both the guidelines for hypertension (Williams et al. 2018) and the BNF (National Institute for Health and Care Excellence 2018*a*) recommend the use of LD for resistant hypertension. Unfortunately, the lack of blood pressure measurements is a limitation of the GHS/18/CA/002 dataset. However, based on the low percentages of patients prescribed an MRA, a mainstay treatment for resistant hypertension, one suspects that this is not the defining aspect of the group. In any case, resistant hypertension is a common cause of HF, and is, in of itself, common and deadly, yet recent research (Groenland et al. 2021) reports much lower mortality rates than are seen for the LD only patients within NHS GG&C.

For patients with congestion, LD are probably life-saving and improving their quality of life. But the use of LD in patients who are not overtly congested might activate the neuroendocrine systems and cause renal and electrolyte disturbances, which are linked to an increased risk of morbidity and mortality (Lim et al. 2008, Bruderer et al. 2014, Flamenbaum 1986). The overall prognosis of the LD only group is composed of individual outcomes, which will reflect the diverse reasons patients are taking the LD. For those who receive the LD due to ankle swelling in the absence of CV disease (e.g., due to a medication side effect such as found with NSAIDs or dihydropyrimidines [Joint Formulary Committee 2019]), these patients should have a better outcome than those with severe cardiac dysfunction. For some, the poor prognosis may be driven by problems other than HF, including COPD (Vozoris et al. 2018) and CAD (Schartum-Hansen et al. 2015).

Additionally, patients may be initiated on LD while in hospital for several reasons, including treatment of breathlessness due to an acute illness such as a chest infection, when there is diagnostic uncertainty about HF being a comorbidity, or for patients who are immobile for long periods of time and develop gravitational ankle oedema. In these situations, the initiation of the LD may not be followed up with subsequent investigations for HF. In cases where the diagnosis of HF is incorrect, the LD may be continued if the symptoms and signs improve, even in the absence of a diagnosis of HF, and even when other treatments (e.g., antibiotics in the case of a chest infection) are responsible for the improvement. In these cases, one can imagine that the LD will be added to the list of discharge medications without further thought or investigation. In these cases, the prescription may be continued in the community if the primary care physician

continues the prescriptions automatically or trusts that the secondary care physicians ordered the appropriate investigations. One can draw a line from a hospital discharge to a patient being on a perpetual LD prescription. However, the analysis in the following chapter suggests that when LD is initiated in the absence of a diagnosis of HF, <50% were admitted to hospital in the year prior.

Current guidelines state that symptoms and signs of congestion, for which guidelines strongly recommend LD, are essential diagnostic criteria for HF (McDonagh et al. 2021). However, many patients with a coded diagnosis of HF were not treated with LD. These patients often had a history of MI, a reduced LVEF, and were prescribed RAASi and beta-blockers; although lack of treatment with LD indicates that they had few or no symptoms of congestion. This strongly implies that these patients did not fulfil conventional guideline criteria for a HF diagnosis. Moreover, the prognosis of these patients was only slightly worse than patients in the neither group made up of patients with CV problems who did not have HF, casting further doubt on the validity of a diagnosis of HF for many patients in the HF only group.

3.5.1 Strengths and Limitations

The dataset used in this chapter pulled patients into the cohort from a region encompassing 23% of the Scottish population (National Records of Scotland 2018). Some limitations in using this dataset are highlighted in Section 2.7. Unlike many other large administrative datasets (Denaxas et al. 2012, Public Health Scotland 2020*a*, Jones et al. 2019), this dataset included ECG and echocardiogram results. However, many tests will not have been conducted on MUSE-connected electrocardiographs (see Section 2.5.6) and will therefore not be available for analysis. This accounts for some of the apparent diagnostic shortfalls observed.

Hypertension was probably under-reported given the high prescription rates for ACEi, ARB, CCB and thiazide diuretics. However, the prevalence of DM and percentage of the cohort on hypoglycaemic therapy and prevalence of COPD and percentage on bronchodilators were closely matched, suggesting that these are useful pharmaco-epidemiological markers of disease, just as LD might be for diagnosed and undiagnosed HF.

3.5.2 Future Work

This chapter flags the prescribing and diagnostic discrepancies between repeat LD use and the diagnosis of HF. Prospective analysis into why these patients are receiving LD is necessary to understand whether the LD prescriptions are appropriate or not and, if inappropriate, whether they are harmful.

Additionally, further work is needed to investigate if the LD prescriptions were initiated to treat symptoms and signs of HF, to determine whether or not investigations needed to diagnose HF

were done but not recorded, and whether or not LD can be safely discontinued in patients who appear to have had them inappropriately prescribed (Cleland et al. 2011).

3.6 Conclusion

This analysis highlights that a large proportion of patients with a cardiovascular problem (be it receiving medication for hypertension or a diagnosis of CAD) will start a repeat LD before they die. However, only about one in four patients who receive repeat LD will also be diagnosed with HF. Rates of hospital admissions are similarly elevated for patients treated with either repeat LD or are diagnosed with HF compared with those with neither. Where investigated, the prevalence of LA dilation is similar. Furthermore, mortality is more strongly associated with the repeat LD than with a diagnosis of HF. Over the 5 years of follow-up, even after adjusting for the number of admissions per patient-year at risk, the hospital admission and mortality rates are greater for patients with repeat LD with or without a diagnosis of HF than for patients with HF only. However, some patients are likely prescribed LD inappropriately for relatively benign problems. Consequently, the prognosis of LD patients with a more serious underlying CV disease will likely be similar to those with HF. This is explored in Chapter 6.

Ultimately, HF lacks a robust and practical definition, creating diagnostic uncertainty and limiting the utility of epidemiological estimates of prevalence, incidence, morbidity, and mortality, thus underpinning healthcare planning and provisioning. As others have pointed out, a better definition is required (Cleland et al. 2021).

The next chapter moves backwards in the diagnostic timeline to look at events and common patterns leading up to the diagnosis of HF and the initiation of repeat LD.

Chapter 4

Pattern of Admissions Pre-Dating New-Onset Heart Failure or Initiation of Loop Diuretics

> Sometimes something catastrophic can occur in a split second that changes a person's life forever; other times one minor incident can lead to another and then another and another, eventually setting off just as big a change in a body's life.
>
> *Jeannette Walls*
> *Half Broke Horses 2009*

4.1 Introduction

There are many causes of HF, and most patients will have more than one contributing factor. Common markers for developing HF are age, hypertension, DM, CAD, and chronic renal dysfunction (Kao et al. 2015, Cleland & McGowan 1999, Gho et al. 2018, Roger 2021). These markers may also be considered risk factors because they have adverse effects on heart function. In healthy individuals, the heart has a large reserve capacity; during exercise, cardiac output can increase four-fold, and individuals can cope with the intravenous infusions of several litres of fluid. For HF to occur, this reserve capacity must be overwhelmed. Many patients who develop HF will have a long asymptomatic prodromal phase during which this reserve capacity is eroded. The final 'straw' is an event such as an infection, AF, MI, or anaemia, which will precipitate evidence for congestion, and if the clinician is alert, a diagnosis of HF. The smaller

the reserve capacity, the smaller the insult required. For a patient with good cardiac reserve, even a substantial MI may not be enough to cause HF; for a patient with little residual reserve, a minor episode of myocardial ischaemia may be enough. In summary, HF rarely occurs 'out of the blue.' There is usually a long prior medical history leading up to the diagnosis.

The consequences of developing congestion and receiving a diagnosis of HF are also of interest. Using the initiation of LD as a surrogate for the development of congestion and primary and secondary care records for the diagnosis of HF, which may or may not be correctly diagnosed, a diagnosis of HF that is not associated with the need for diuretic treatment may be considered a suspect diagnosis. An impaired heart function leading to a low LVEF does not mean that the patient has HF if there is no evidence of congestion (Cleland et al. 2021). On the other hand, the natural history of patients newly initiated on LD is unknown. Patients initiated on LD may subsequently develop HF, and patients initially diagnosed with HF may or may not be initiated on LD therapy. Using EPR records to map out the pathways of disease progression and contacts with the healthcare service could be of considerable interest.

Accordingly, an investigation of patterns of events leading up to and following a diagnosis of HF and/or the initiation of LD, including hospital admissions, results of blood tests, comorbid conditions, and medications is required. When available, results of routinely collected echocardiograms and ECGs were also reported.

4.1.1 Aims

This Chapter describes the pattern of hospital admissions in the year leading up to a new diagnosis of HF or the initiation of LD. Changes in prescriptions and comorbid diagnoses, and results of key blood tests are also reported prior to, at the time of, and shortly after a new diagnosis of HF or the initiation of LD.

4.2 Background

The following descriptions are intended as an introduction to concepts that will be implemented in this Chapter.

4.2.1 Look-Back Periods

Identifying the date of a HF diagnosis or the date when LD prescriptions were initiated is a relatively simple task for an individual patient with a complete medical record. Defining these events for a cohort of patients using routinely-collected, linked EHRs is more complex. Diagnostic information is entered manually and often has a substantial subjective component (i.e., it is an opinion rather than a definitive test), especially for a diagnosis such as HF. This leads

to errors of both 'commission' (i.e., recording an incorrect diagnosis) and, more importantly, omission (i.e., failing to make a diagnosis or failing to record it)[1]. Information on prescriptions and dispensing is acquired through electronic transactions and, therefore, is much less prone to error.

Unlike a registry, where all patients have a baseline evaluation date, administrative records provide a dynamic, longitudinal record and highly variable lengths of observed time before the event of interest. Also, for a registry, the information to be collected is determined prior to including patients with the intent of asking one or more questions, whereas, for administrative records, the question must be based on information which has already been collected as part of routine healthcare. The question must be tailored to the data rather than tailoring which data are collected to the question.

Incident cases, defined as new occurrences of an event of interest within a given interval, require the use of a look-back period (Rothman et al. 2021, Kim et al. 2020), which is a set length of time where patients are required to have available EPR prior to the event of interest. This allows the identification of prevalent cases and the exclusion of patients for whom the condition of interest cannot be ruled out. A balance must be struck between shorter look-back periods, which carry the risk of missing prevalent cases, and long look-back periods, which exclude patients with shorter length of coverage and/or reduces the length of available follow-up data for analysis. Choosing a suitable length for the look-back window depends on the event of interest (Kim et al. 2020, Sulo et al. 2015, Czwikla et al. 2017).

4.2.2 Immortal Time Bias

Commonly, observational studies investigate the association of risk factors with the elapsed time until the event of interest occurs. Researchers must be careful to avoid immortal time bias, defined as a period of follow-up, where, due to the definition of exposure, the outcome of interest cannot occur (Yadav & Lewis 2021, Suissa 2007, Rothman et al. 2021). The patient is not truly immortal but must be event-free until the occurrence of the event, which classifies them into a particular group. The first recorded instances of immortal time bias in epidemiology were found in the Stanford Heart Transplant (Clark et al. 1971) and the Texas Heart Institute (Messmer et al. 1969) studies (Suissa 2007). In both instances, Mitchell H. Gail (1972) pointed out that the apparent excellent results were due to the patients having had a heart transplant in order to be included in the analysis, thereby guaranteeing patients survived transplant surgery for that group (Suissa 2007). Re-analysis of the Stanford Heart Transplant data accounting for immortal time using time-dependent analysis (see Section 3.2.1) showed that the purported major survival

[1]These errors can be amplified and expanded in claim datasets and similar administrative datasets as coders have the same biases as found in EHRs, but they are also added issues such as a lack of common terminology between the coder and physician and incomplete or non-standard documentation (Hosseini et al. 2021).

benefit disappeared (HR: 0.93, p = 0.9) (Mantel & Byar 1974). For this reason, and the nature of linked EPR databases (Agarwal et al. 2018), retrospective observational studies must be careful in defining exposure in order to avoid introducing immortal time or selection biases. Commonly, observational studies investigate the association of risk factors with the elapsed time until the event of interest occurs.

4.2.3 Data to Networks

On a superficial level, a patient's EPR represents a complex series of contacts with the healthcare system, where each contact contributes information to the patient's overall EPR. Within a cohort of patients, common, discrete sequences of events, or even a common single event, present potential targets for diagnostic or therapeutic interventions. Compiling the individual sequence of events, or paths, into a network of events turns a convoluted and complicated collection of data into a digestible research question for graph theory and network analysis.

Graph Theory

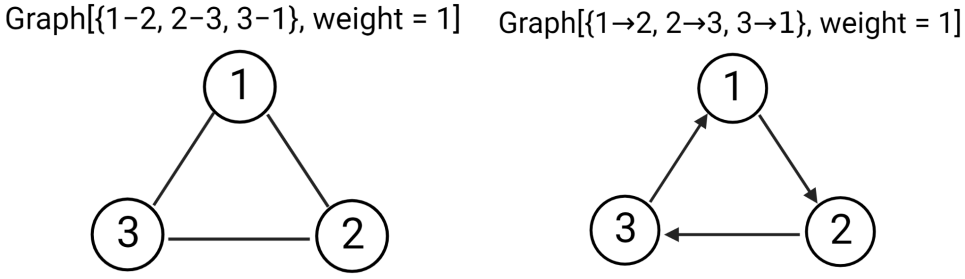
Graph theory, as a discipline, traces back to Leonhard Euler's answer to the Köngsberg Bridge Problem from the 18th century (Euler 1741). The problem asks if the seven bridges in Köngsberg can be traversed in a single trip without doubling back, ending at the same point where the trip begins. If the Köngsberg Bridge Problem was represented as a graph, the landmasses would be the nodes connected by the bridges as edges (Carlson 2010). Two simplified examples of graphs formed by three nodes and three edges are illustrated in Figure 4.1. Continuing with the example of the bridges, in an undirected graph, people are allowed to travel in either direction, while the Köngsberg Bridge Problem specified a directed graph such as seen in Figure 4.1b, meaning that people are only allowed to travel across the bridge in one direction.

Network Analysis

Network science or network analysis is a tool to describe the structure and dynamics of complex, real-world systems, tackling heterogeneous, temporal, and adaptive patterns of interactions (Iñiguez et al. 2020). Network analysis inherited its intrinsic components from graph theory, where nodes represent elements and the edges indicate interactions. From this foundation, network analysis has evolved from rigorous proofs and theory towards ad hoc mathematical concepts to quantify the observations found in real-world data (Iñiguez et al. 2020).

As such, a few common measurements and their definitions in the context of directed graphs are:

- The weight of a degree, node, or edge is a numerical representation of importance or frequency. For example, the weight of an edge might be the number of times that edge is



(a) An undirected graph

(b) A directed graph

Figure 4.1: Basic examples of an undirected and a directed graph. The circles represent the set of nodes within the graph, and the lines connecting the nodes represent the set of edges within the graph. In an undirected graph, the edge can be traversed in either direction, while in a directed graph, the direction of travel for the edge follows the arrow. The text above each graph illustrates how the graph is specified through edges (i.e., pairs of nodes), with 'weight' specifying the edge thickness.

traversed.

- The average path length is the average number of edges along the shortest path for all possible pairs of nodes.
- The node degree is the number of entry and exit edges through the node.
- The weighted degree of a node is the sum of the weights of the entry and exit edges that pass through the node.
- The average degree is the total number of edges divided by the total number of nodes.
- The diameter of a network is the maximum number of steps needed to travel between nodes.

(Khokhar 2015) From its origins in logic, topography, and later social networks, network analysis has evolved into a growing field of interest within medical and epidemiological research. Various aspects of network analysis have been employed in the context of hospital admissions (Bean et al. 2017), diagnostic pathways (Jeong et al. 2017), disease trajectories in large-scale populations (Siggaard et al. 2020), and even progression of HF in T2DM patients (Hossain et al. 2019).

4.3 Methods

4.3.1 Patient Identification

Using the GSH/18/CA/002 dataset (see Section 2.3.2), the incidence of HF and of new LD dispensing were investigated from 1st January 2012 through 31st March 2017. Incident HF cases

were identified based on the first Read or ICD-10 code across primary and secondary care. As described in Section 2.6.1, patients were excluded if their first HF code indicated a pre-existing diagnosis of HF (see Table A.16 for a list of codes). New dispensing of LD was defined as the first instance where a LD was dispensed over two consecutive quarters or where the patient died within 90 days of their first prescription (see Section 2.5.4 for further details and Table A.6 for qualifying medications).

Patients were excluded if they were classified as mislinked; were aged <18 years or were censored before 1st January 2012; or had no record of HF or repeat LD dispensing during the study window. In addition, patients were excluded if they had a record of HF or LD in the look-back period or they had less than 5 years of EPR prior to the earliest event (HF diagnosis or LD initiation); or the first record of HF indicated that the patient had long-standing HF (see Section 2.6.1 and Table A.16 for codes). See Figure C.1 for a flow diagram of patient inclusion and exclusion numbers.

4.3.2 Patient Classification

Patients were grouped into six discrete groups based on the presence and order of a diagnosis of HF or the initiation of LD. Patients were classified as 'HF Only' if the patient did not receive a repeat prescription for LD within the following year and as 'LD Only' if they did not receive a diagnosis of HF in the following year. Patients who experienced both events within the following year were classified as 'Both: Together,' if the events occurred within a 30-day window, with the HF date as the index date. If the window was >30 days, then patients were classified as either 'Both: LD First' or 'Both: HF First' using the first event as the classifier and index date[2] (see Figure 4.2 for an example of a Both: HF First patient timeline.) Patients whose first event was a HF diagnosis during a hospital admission that they did not survive, thereby precluding the detection of the LD prescription (see Section 2.5.4), were classified as 'HF Death as First Record'.

The one-year window for classifying the 'both' groups was chosen to reduce the effect of immortal time bias (see Section 4.2.2). See Figure 4.2 for an illustration of a patient timeline showing how the start date is calculated using the ordering of the HF diagnosis and LD initiation.

4.3.3 Patient Characteristics

Baseline patient demographic information 12 months prior to diagnosis of HF or the initiation of LD were identified (referred to as the start date in Figure 4.2), including age, sex (see Section 2.5.2), ethnicity (see Section 2.5.2), and quintile of socioeconomic deprivation (see Section

[2]The definition of the 'Both' groups differs from those in Chapter 5, as those 'Both' groups do not have a one-year boundary window

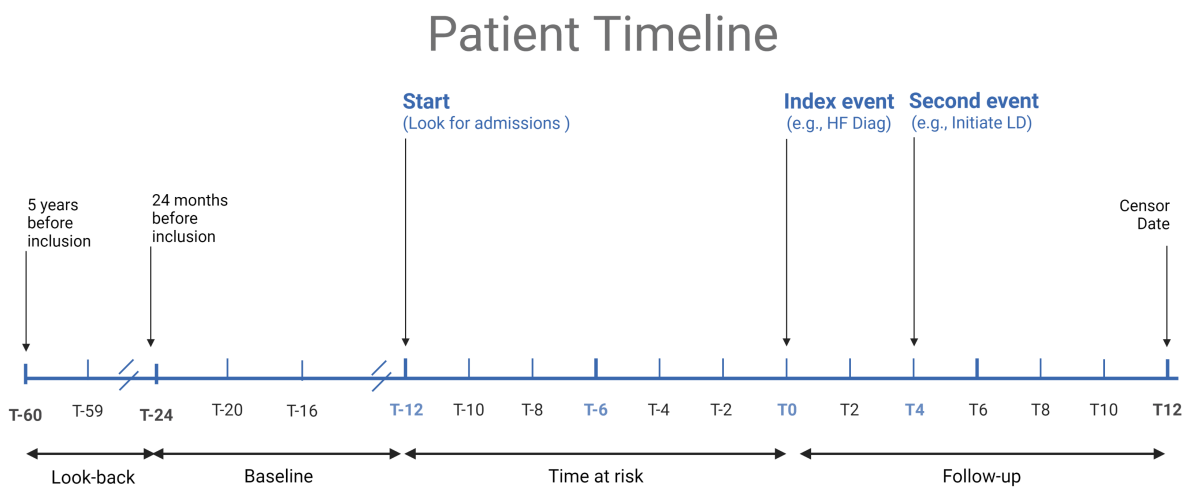


Figure 4.2: Patient timeline illustrating that the first of either HF diagnosis or LD initiation is the index event. The index event date[3] is the inclusion date (T0 on the timeline), the start of follow-up, and the terminal node of the admission network graphs. T-n (e.g., T-6), where n indicates the number of months, shows the number of months before the inclusion date, or Tn (e.g., T4) shows the number of months after inclusion. 'Start' is the initial node of the admission network graphs and the date patients' time at risk for being admitted to hospital starts.

2.5.2). Where applicable, the breakdown of how patients met the definition of starting a repeat LD prescription and the location of HF diagnosis were reported.

To describe patients' haematology and biochemistry test results (see Section 2.5.9) leading up to the diagnosis of HF or the initiation of LD, the most recent haemoglobin (see Section 2.5.9), calculated estimated glomerular filtration rate (eGFR), and serum results for urea, albumin, sodium, chloride, potassium, and bicarbonate, in the baseline period indicated in Figure 4.2 (24 to 12 months prior to the index event) were reported. Anaemia was classified as a haemoglobin below 12.0 g/dL for women, assuming no pregnancies, and 13.0 g/dL for men (see Section 2.6.2). The eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation (Levey et al. 2009) without adjusting for ethnicity (see Section 2.5.9).

Co-morbid conditions and dispensing of medicines of interest were recorded for different intervals (24-12 months prior, 12-0 months prior, and 0-12 months after an index event) to describe changes in diagnoses and medication levels. For co-morbid conditions, the history of hypertension, DM, thyroid disease, IHD, AF/AFL, valve disease, PAD, CKD, COPD, stroke, cancer, and dementia were reported. See Section 2.6.2 for an explanation of how conditions were defined. Patients with a history of three or more conditions and those with an implanted cardiac electrophysiological (EP) device (i.e., pacemaker, ICD, or CRT) were also reported. For medications of interest, dispensing of ACEi, ARB, MRA, beta-blockers, digoxin, CCB, thiazides and thiazide related diuretics (Thiazides+), low-dose aspirin, oral anticoagulants, lipid regulating medications, bronchodilators, thyroid medications, and hypoglycaemic agents including insulin were

reported. Those with three or more medications dispensed[4] were reported. Medications were identified by BNF code and classified by active agents (see Section 2.5.4).

Using the closest echocardiogram to the index date within the three-year window of 24 months before the index event through 12 months after, LA diameter and LVEF were reported when recorded (see Section 2.5.7). Similarly, from the closest ECG, heart rhythm, heart rate, QRS duration and durations $\geq$120 ms, calculated QTc and prolonged QTc, and whether or not the ECG detected an acute MI (see Section 2.5.6).

4.3.4 Hospital Admissions

The pattern and number of hospitalisations were investigated in the year before an incident diagnosis of HF or the initiation of a repeat LD. Admissions were defined as a hospital stay lasting more than one calendar day. Admissions were classified using the ICD-10 code in the first diagnostic position. For those who received a diagnosis of HF in hospital, the admission was excluded from the admissions analyses, although these numbers are reported. Admissions were classified using the overarching ICD-10 Chapter. ICD-10 Chapters XV, pregnancy, childbirth, and the puerperium; XVI, certain conditions originating in the perinatal period; XVII, congenital malformations, deformations and chromosomal abnormalities; XX, external causes of morbidity and mortality; and XXII, codes for special purposes were grouped into 'Other' in order to meet information governance requirements for minimum reporting that do not permit reporting of small groups of patients that might make them identifiable. For both counts and graphs, Chapter IX, Diseases of the Circulatory System, was split into the most frequent disease categories: AF, cardiac electrophysiology other (Card EP Other), acute MI, IHD, cerebrovascular disease (Cerebrovasc Dis), hypertensive disease, pulmonary disease (Pulm Dis), and cardiovascular other (CV Other). A breakdown of the four most common admissions Chapters is shown in Table C.1.

Network construction

The sequences of patient admissions were modelled as weighted directed graphs in which nodes represent the reason for admission and edges represent the order of events. Each group's directed graph was constructed from combining individual admissions paths, starting with a 'Start' node 12 months prior to a diagnosis of HF or the initiation of LD, followed by a node for each admission, and terminating with a node labelled according to the patient's group. Paths with fewer than six patients were grouped into 'Other Paths' as required by data-privacy rules and were removed from further analysis to improve clarity.

Edge weights are proportional to all node pairs within each group. The edge weight (w_i) was

[4] ARNi was not reported separately because it was only licensed for use near the end of the study window.

calculated for each edge (e_i) in the set of observed edges of length (n) per the following:

$$w_i = \left(\frac{e_i \, frequency}{\sum_{i=1}^{n} e_i \, frequency} \right) * 100 \tag{4.1}$$

where $frequency$ is the number of times the edge was present in a group's individual admission paths. Edge weights were multiplied by 100 in order to aid visual interpretation. The edges were proportionally weighted in order to allow for direct comparisons between the different groups' networks that have varying numbers of admitted patients (Bean et al. 2017).

Graphical layout

Graphs were generated in five distinct steps:

1. Calculate proportional edge weights,
2. Fix the virtual 'Start' and terminal node positions,
3. Apply the layout algorithm,
4. Add node labels,
5. And scale node font size.

The 'Start' node represents the first time patients are eligible to be admitted (see Figure 4.2). It was positioned and fixed (i.e., 'settled') at the top of the graphical window. The terminal node (i.e., one of 'LD Only','Both: LD First','Both: Together', 'Both: HF First', 'HF Only', or 'HF Death as First Record') was fixed at the bottom of the graphical window.

The Yifan Hu layout algorithm Hu (2005) was chosen as it provides the best compromise between visual interpretability and computational efficiency, avoiding the local minima found when using Fruchterman and Reigold Fruchterman & Reingold (1991), Gephi Home-brewed ForceAtlas, and ForceAtlas 2 layouts (Jacomy et al. 2014), while ensuring reproducibility. The Gephi Yifan Hu algorithm was applied using the default parameters, with the exception of setting the optimal distance to 1,500 to allow for easier visual interpretation.

Finally, the node font size was scaled between 1 and 2 based on each node's degree rank. This means that font sizes increase based on how well-connected the node is; more connections indicate an admission reason that is less likely to occur in isolation. This adjustment was applied to the base Arial 24 Plain font.

The Graphs were created using Gephi, version 0.9.2 (Bastian et al. 2009), and open-source software for network analysis and visualisation.

4.3.5 Statistical Analysis

Patient characteristics for categorical data are presented as frequencies (%), and continuous data are reported as medians with (first and third quartiles). Where applicable, the number and percentage of patients with missing data are displayed for each variable.

When reporting hospital admissions rates, the number of admissions was split using the primary reason for admission. The numbers per patient-year at risk were calculated in the 12 months leading up to the index event in order to normalise the number of admissions by the number of patients in each group eligible to be admitted (e.g., survivors not already in hospital). For a further explanation of patient-time at risk, see Section 3.2.2.

Statistical analysis was conducted in R (see Section 2.4.2) (R Core Team 2021), using the packages listed in Section 3.3.5.

4.4 Results

From an estimated population of 1,217,020 individuals[5], of whom 248,077 were aged >60 years, 28,244 people developed incident HF or started repeat LD prescriptions between 1st January 2012 through 31st March 2017. Initiation of LD without a subsequent diagnosis of HF within a year accounted for the majority of cases (18,596 patients; 66%; see Table 4.1). Their median age was 73 years, and 62% were women. Analysis of events subsequent to the first year is described in Chapter 5. Compared to the group initiated on LD only, patients initiated on LD who also received a diagnosis of HF had a rather similar median age (approximately 75 years) but a rather smaller proportion of women (approximately 48%), with little variation in age or sex depending on the order of LD/HF events. Patients with a diagnosis of HF who did not receive LD were younger (median age 68 years) and more likely to be men (63%). Those who did not survive an index HF event were older (median age 79 years), and 50% were women. About half of the patients were admitted to hospital in the year before the index event, and >70% had some form of secondary care contact with little difference between the groups. The 'location' of a diagnosis of HF was in primary care in about 25% of cases treated with LD but 45% for those who were not initiated on LD. Most patients had multiple co-morbid diseases prior to an index event (see Table 4.1 for demographics and Figure 4.6 for co-morbid diseases).

[5]Using the 2012 mid-year population estimate

Table 4.1: Patient demographics at time of first key.

Variable	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**	**HF Death as First Record**
n	18,596	3,957	1,251	1,854	1,876	710
Age (years)	73 (62 - 81)	68 (57 - 78)	77 (69 - 84)	75 (66 - 83)	73 (65 - 81)	79 (71 - 86)
Sex						
Women	11,529 (62%)	1,455 (37%)	609 (49%)	898 (48%)	855 (46%)	356 (50%)
Men	7,067 (38%)	2,502 (63%)	642 (51%)	956 (52%)	1,021 (54%)	354 (50%)
Ethnicity						
White	16,283 (88%)	3,531 (89%)	1,154 (92%)	1,731 (93%)	1,698 (91%)	622 (88%)
Missing	1,949 (10%)	317 (8%)	70 (6%)	82 (4%)	116 (6%)	75 (11%)
Other	364 (2%)	109 (3%)	27 (2%)	41 (2%)	62 (3%)	13 (2%)
Socioeconomic deprivation (SIMD)						
1 (most deprived)	7,698 (41%)	1,599 (40%)	486 (39%)	818 (44%)	753 (40%)	321 (45%)
2	3,467 (19%)	719 (18%)	260 (21%)	335 (18%)	334 (18%)	127 (18%)
3	2,590 (14%)	542 (14%)	161 (13%)	250 (13%)	243 (13%)	87 (12%)
4	2,179 (12%)	458 (12%)	140 (11%)	192 (10%)	218 (12%)	72 (10%)
5 (least deprived)	2,662 (14%)	639 (16%)	204 (16%)	259 (14%)	328 (17%)	103 (15%)
Secondary care contact in the year leading up to diagnosis or initiation						
Any contact	13,454 (72%)	3,132 (79%)	1,047 (84%)	1,453 (78%)	1,455 (78%)	548 (77%)
CV clinic	2,072 (11%)	830 (21%)	227 (18%)	251 (14%)	283 (15%)	46 (6%)
Admitted	9,144 (49%)	1,836 (46%)	643 (51%)	912 (49%)	817 (44%)	397 (56%)
Num. admissions	1 (1 - 2)	1 (1 - 2)	1 (1 - 2)	1 (1 - 2)	1 (1 - 2)	1 (1 - 2)
Other paths	1,564 (17%)	356 (19%)	210 (33%)	247 (%)	208 (25%)	150 (26%)

Continuation of Table 4.1

Variable	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together	HF Death as First Record
n	18,596	3,957	1,251	1,854	1,876	710
Not admitted	9,452 (51%)	2,121 (54%)	608 (49%)	942 (51%)	1,059 (56%)	313 (44%)
Admission rate ‡	0.95	0.77	1.00	0.84	0.73	1.14
Background LD and HF information						
Criteria for inclusion in LD groups						
Death <90 days	1,361 (7%)	N/A	49 (4%)	87 (5%)	156 (8%)	N/A
Consec qtrs	17,234 (93%)	N/A	1,202 (96%)	1,767 (95%)	1,720 (92%)	N/A
Location of HF diagnosis						
GP	N/A	1,779 (45%)	338 (27%)	454 (24%)	476 (25%)	N/A
Hospital	N/A	2,178 (55%)	913 (73%)	1,400 (76%)	1,400 (75%)	710 (100%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

‡ Admission rate defined as the number of admissions per patient-year at risk;

Num. admissions; Number of admissions; Death, Death within 90 days of first prescription;

Consec qtrs, two consecutive quarters.

4.4.1 Admissions

Of the 28,244 incident cases, 13,749 (49%) had a hospital admission in the year before the index event. Regardless of the index event, those who were hospitalised had similar numbers of admissions with a median of 1 admission (1^{st} - 3^{rd} quartile: 1 - 2) (see Figure C.2 for age and sex split hospital admission rates per patient-year at risk.). Figure 4.3 shows the six admission network graphs. Where patients in the LD only group were admitted, the most common pathway was a single admission for non-specific symptoms and signs (ICD-10 Chapter XVIII: 912 cases; 10%). The next most common pathway was a single respiratory admission (X: 789 cases; 9%) (see Table C.2). A similar pattern was observed in cases where dispensing of LD preceded a diagnosis of HF. For these groups, the most common pathway was a single respiratory admission (ICD-10 Chapter X: 75 cases; 12%), and a single admission for non-specific symptoms and signs was the second most common pathway (ICD-10 Chapter XVIII: 71 cases, 11%). Where a diagnosis of HF occurred (either alone or with or before dispensing of a LD), the most common path was a single admission for an acute MI (Both: Together: 91 [11%]; Both: HF First: 108 [12%]; HF only: 435 [20%]). The second most common path was either a respiratory admission (Both: Together: 86 cases [8%]) or an admission for non-specific symptoms and signs (Both: HF First: 87 cases [10%]; and HF only: 180 cases [8%]). For those with a first record of HF at the time of death, a single admission for non-specific symptoms and signs (52 cases; 16%), followed by a single respiratory admission (39 cases; 12%) were the most common pathways.

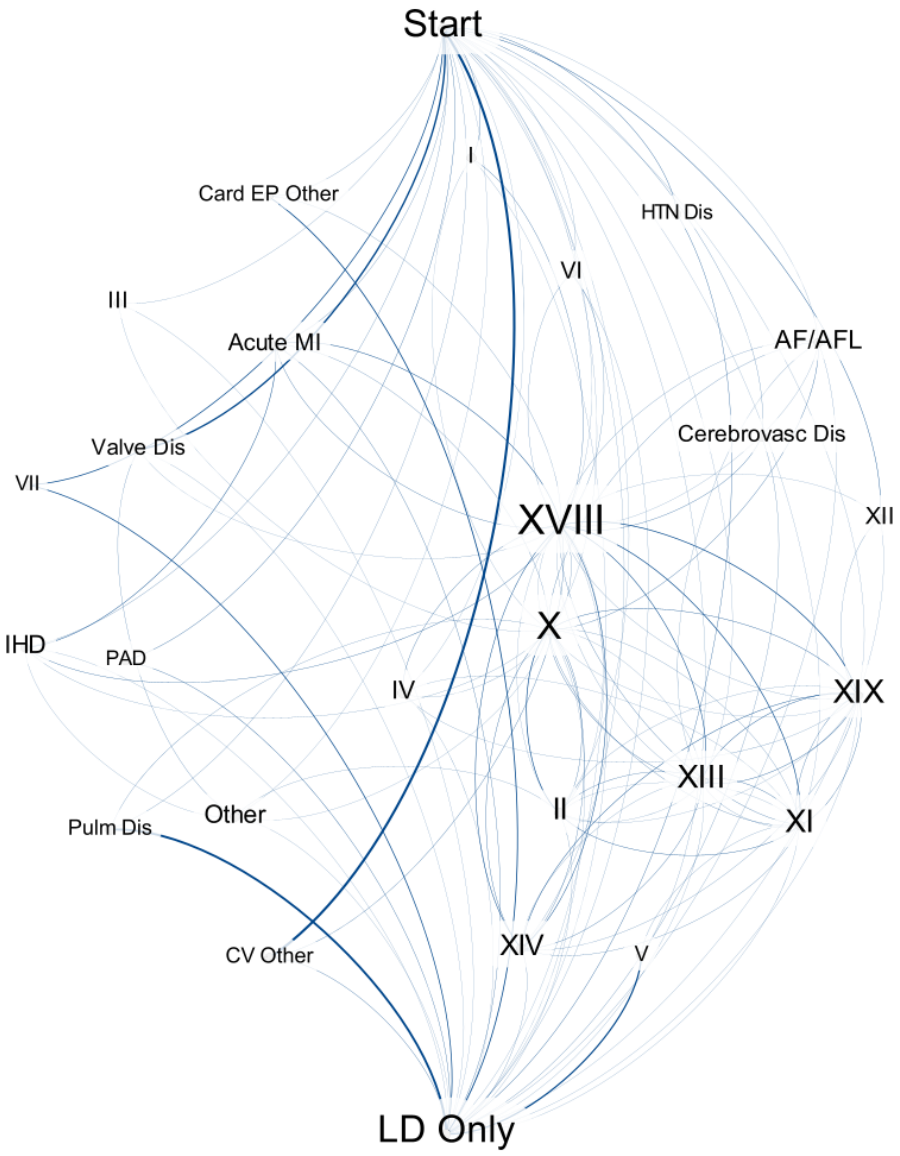
For patients whose index event was dispensing of LD with or without a subsequent HF diagnosis, the top-ranked node by weighted degree [6] was nonspecific symptoms and signs (LD only: weighted degree of 20.9; Both: LD First: weighted degree 16.2). For patients whose index event was the diagnosis of HF with or before the dispensing of LD, the top-ranked node by weighted degree was acute MI (Both: Together: weighted degree of 22.4; Both: HF First: weighted degree of 28.0). For patients whose index event was the diagnosis of HF either alone or at death, the top-ranked node by weighted degree was injury, poisoning, or consequences of external causes (HF only: weighted degree 15.3; HF death as first record: 27.2).

Across the six groups, the most common reasons for admission in the year prior to the event were CV disease (especially acute MI or AF/AFL), respiratory disease, neoplasms, and admissions for non-specific symptoms and signs (see Figures 4.5a and C.2a). The LD only group contributed the majority of admissions in all Chapters and categories except acute MI. Admissions for respiratory infections and exacerbation of COPD predominate for patients in the LD only group. Figures 4.5b and C.2b show the rates adjusted for person-time at risk, which provides a different perspective. Patients in the group have lower per-patient year at risk rates of admissions, and patients with HF (with or without LD) generally have higher rates. However, rates of

[6] The weighted degree of a node is defined as the number of connected edges which have been ponderated by the edge's weight. Higher values indicate a more centrally involved node.

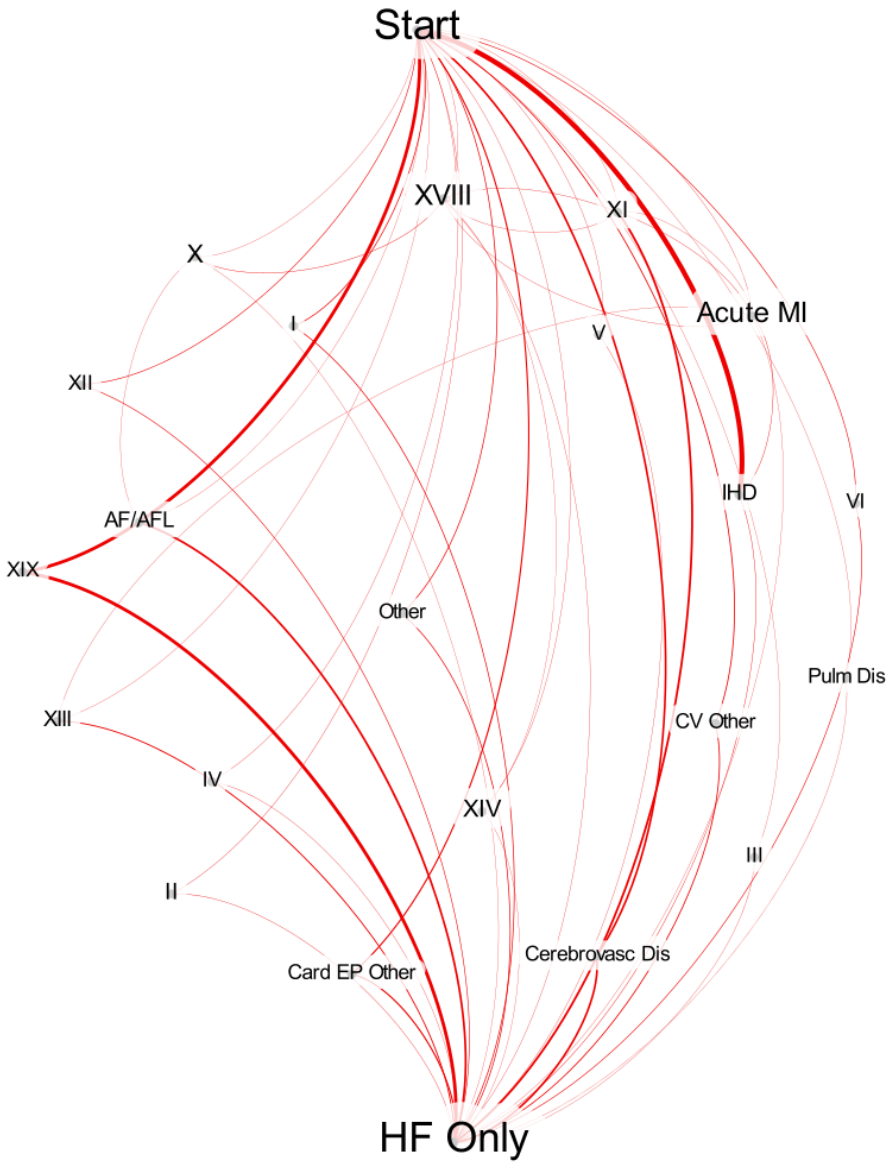
respiratory and non-specific symptoms and signs hospitalisations are similar to or greater in the LD only group than for cardiovascular disease.

In Figure 4.3a, the node with the highest weighted degree (the number of paths in and out ponderated by the number of patients who experienced the event) is admissions for symptoms and signs (ICD-10 Chapter XVIII), with important contributions from respiratory (X), injuries (XIX), and cardiovascular.



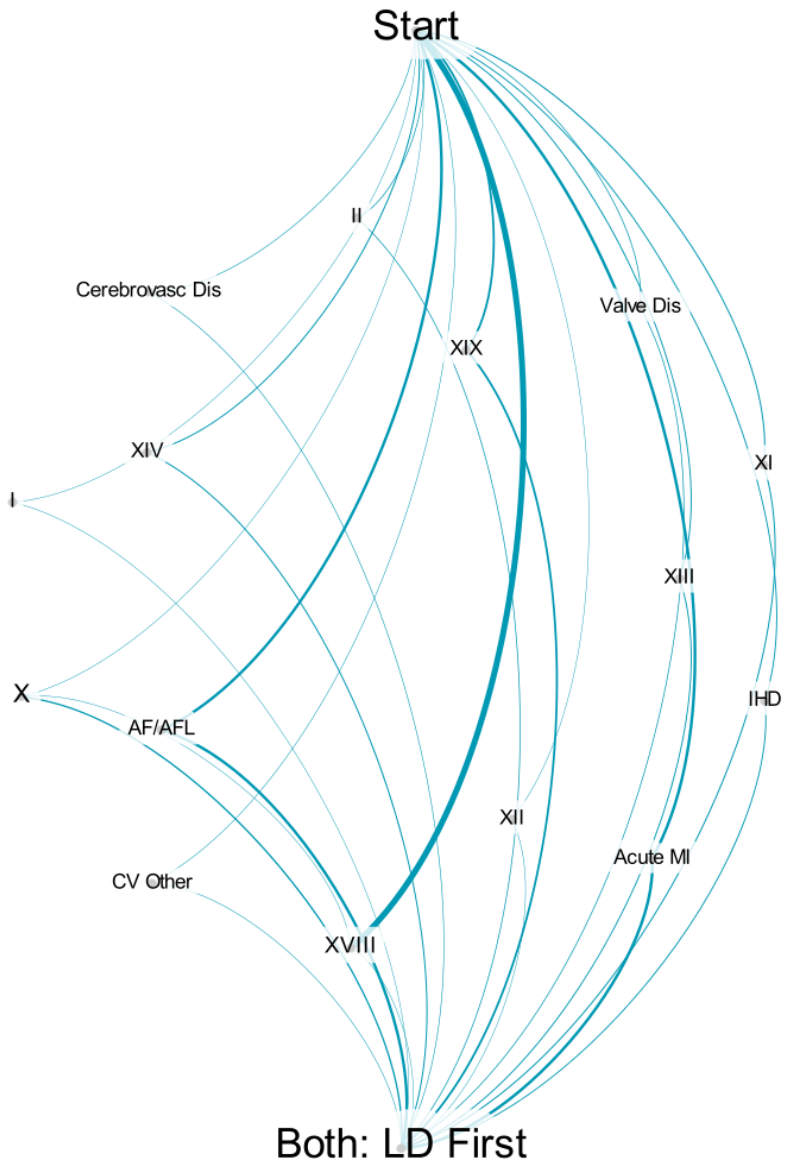
(a) LD Only

Figure 4.3: Representative admissions patterns starting from 12 months before an index event (in this case LD initiation). Edge thickness represents the relative flow of patients, and node size represents the relative degree ranking. The networks are directed, but directionality is not represented for clarity at this scale. ICD-10 Chapters: I, infections & parasitic diseases; II, neoplasms; III, blood and blood-forming organs; IV, endocrine, nutritional & metabolic diseases; V, mental and behavioural; VI, nervous system; VII, eye and adnexa; VIII, ear and mastoid process; IX, circulatory system split up into atrial fibrillation/flutter (AF/AFL), cardiac electrophysiology other (Card EP Other), ischaemic heart disease (IHD), acute myocardial infarction (Acute MI), cerebrovascular disease (Cerebrovasc. Dis), hypertensive disease (HTN Dis), peripheral arterial disease (PAD), pulmonary disease (Pulm Dis), valve disease (Valve Dis.), and other cardiovascular (Other CV); X, respiratory system; XI, digestive system; XII, skin and subcutaneous tissue; XIII, musculoskeletal system; XIV, genitourinary system; XVIII, symptoms and signs; XIX, injury; XXI, factors influencing health status.



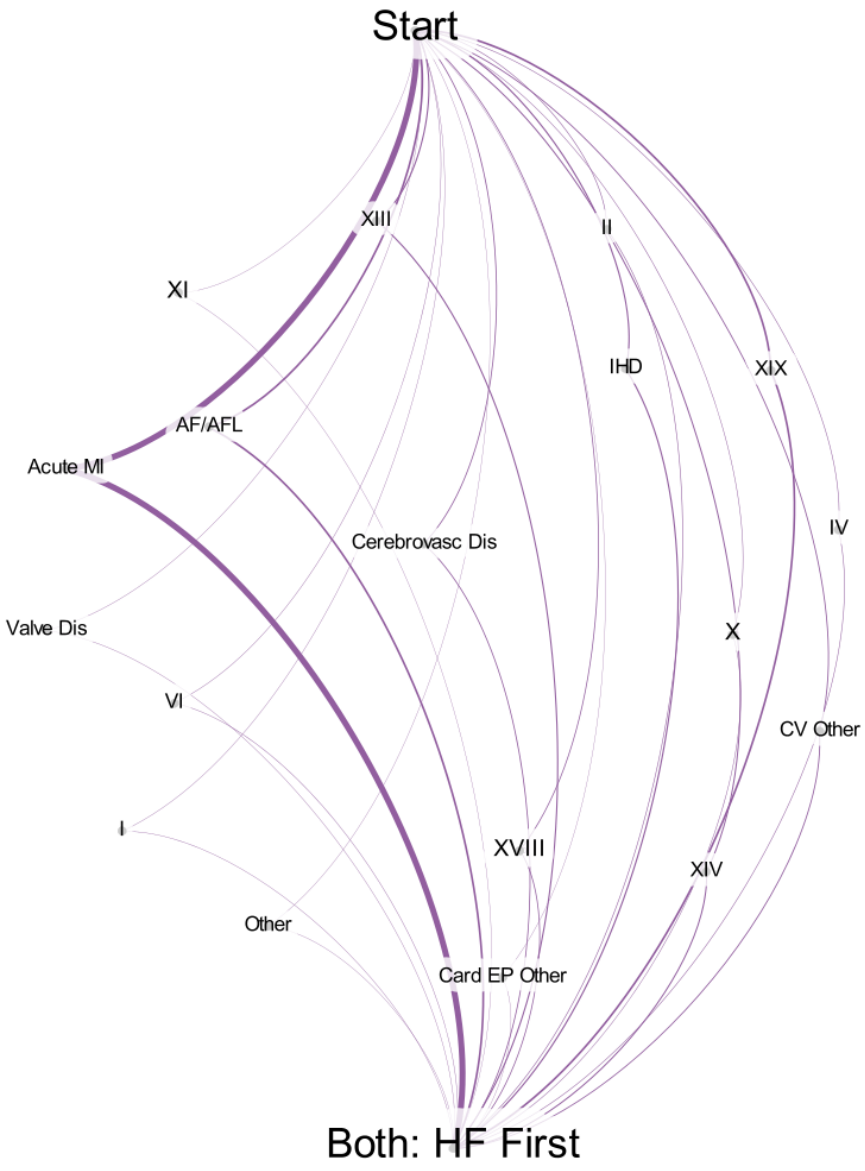
(b) HF Only

Network graphs representing the admissions patterns starting 12 months before an index event (in this group, diagnosis of HF without initiation of LD). See Figure 4.3a for an explanation of abbreviations. The most connected nodes are acute MI and symptoms and signs (ICD-10 Chapter XVIII), with a much less frequented respiratory disease (ICD-10 Chapter X) compared with any of the groups with LD.



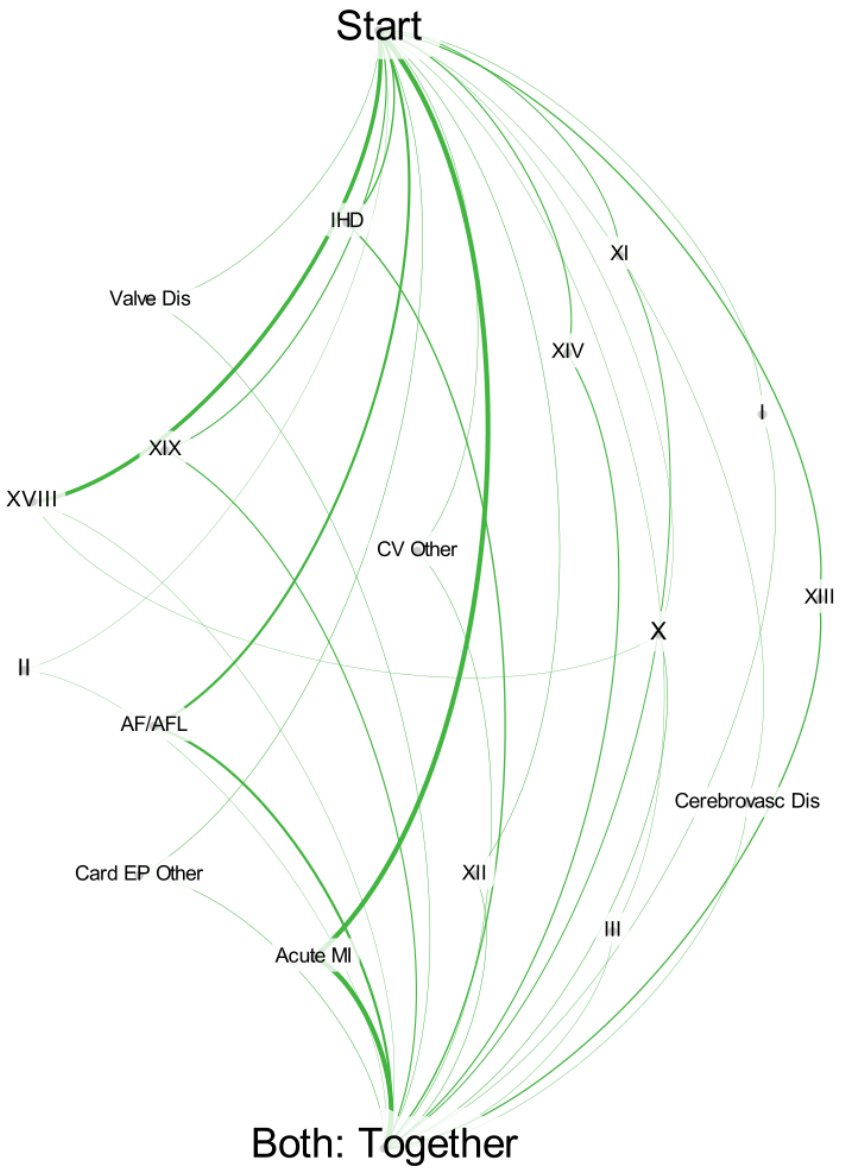
(c) Both: LD First

Network graphs representing the admissions patterns starting 12 months before an index event (in this group, initiation of LD followed by a diagnosis of HF). See Figure 4.3a for an explanation of abbreviations. The pathways in this graph are similar to those seen in the LD only group's graph, with admissions for respiratory disease (ICD-10 Chapter X) and symptoms and signs (ICD-10 Chapter XVIII), but with acute MI and AF/AFL being more common.



(d) Both: HF First

Network graphs representing the admissions patterns starting 12 months before an index event (in this group, diagnosis of HF followed by initiation of LD). See Figure 4.3a for an explanation of abbreviations. This graph is dominated by a single admission for acute MI or AF/AFL, while symptoms and signs (ICD-10 Chapter XVIII) and respiratory (ICD-10 Chapter X) admissions are well-connected to other admissions.



(e) Both: Together

Representative admissions graphs starting from 12 months before an index event (in this group, diagnosis of HF and initiation of LD within 30 days). See Figure 4.3a for an explanation of abbreviations. This graph has similar admissions patterns to those in the previous 'Both: HF First' group.

(f) HF Death as First Record

Network graphs representing the admissions patterns starting 12 months before an index event (in this case, HF diagnosis in hospital where the patient did not survive the index event. See Figure 4.3a for an explanation of abbreviations. This graph is dominated by a single admission for injuries (ICD-10 Chapter XIX), while symptoms and signs (ICD-10 Chapter XVIII) and respiratory (ICD-10 Chapter X) admissions are well-connected.

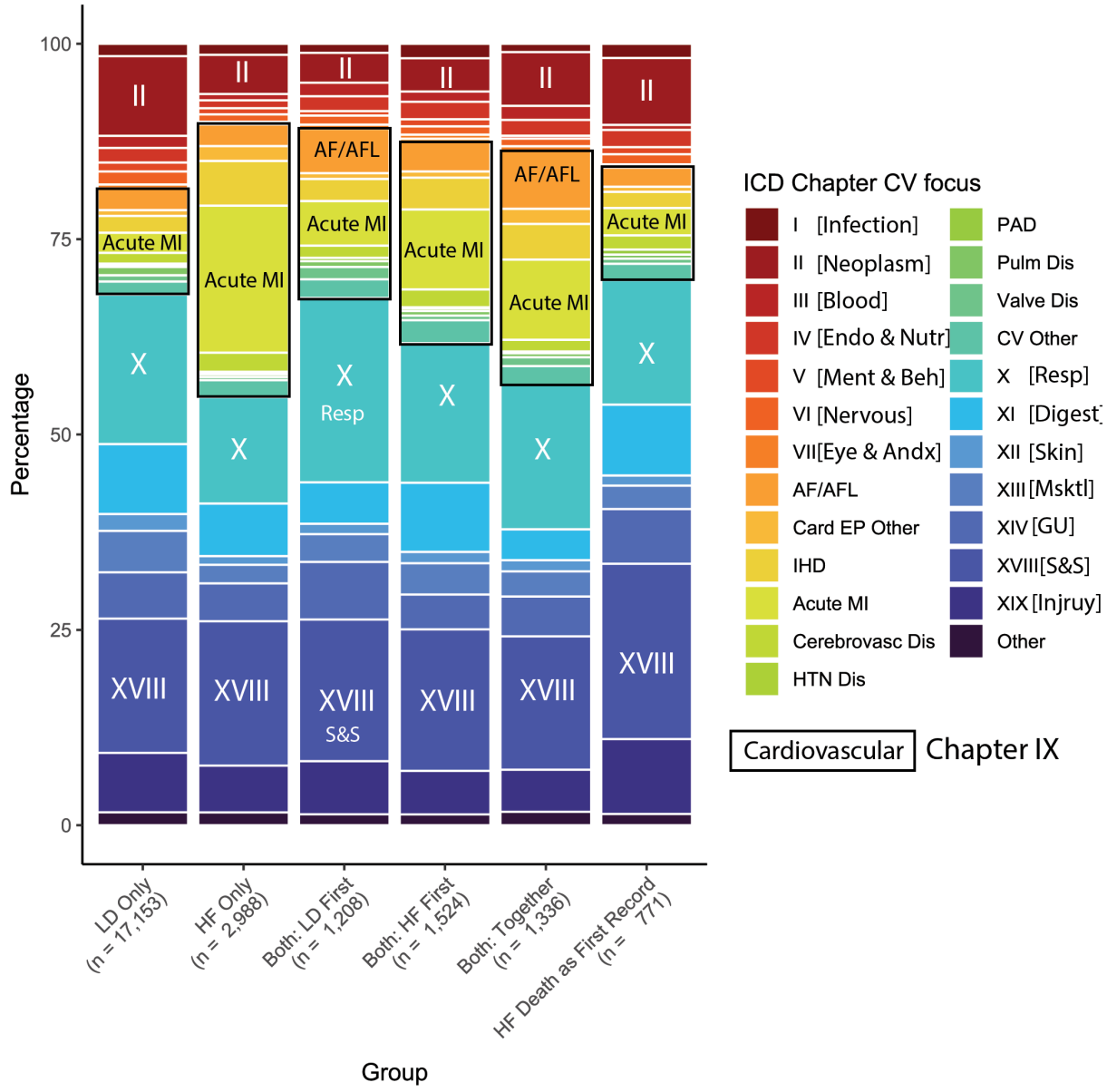
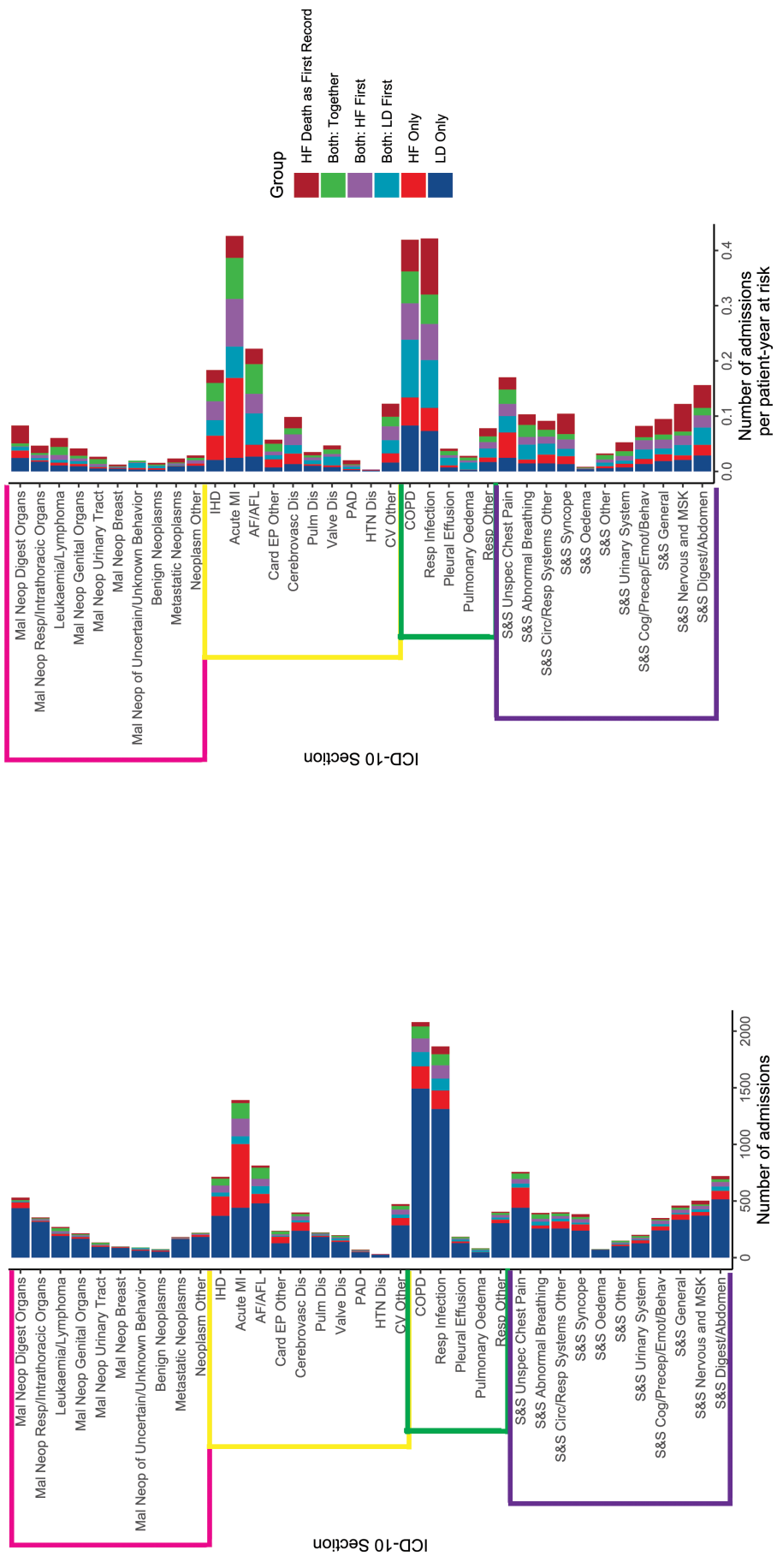


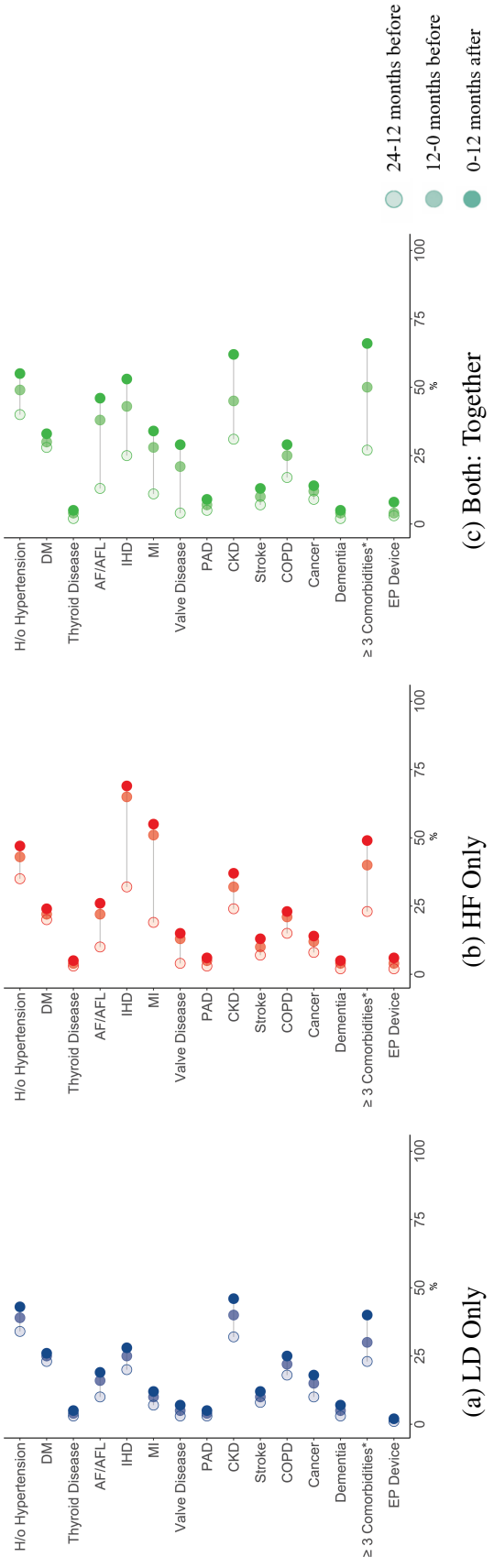
Figure 4.4: Hospital admissions in the year prior to the event per ICD-10 Chapter presented as a proportion of the total number of hospital admissions; circulatory system admissions are shown within the black outline boxes. See Figure 4.3a for an explanation of abbreviations.



(a) Number of admissions

(b) Number of admissions, per patient-year at risk

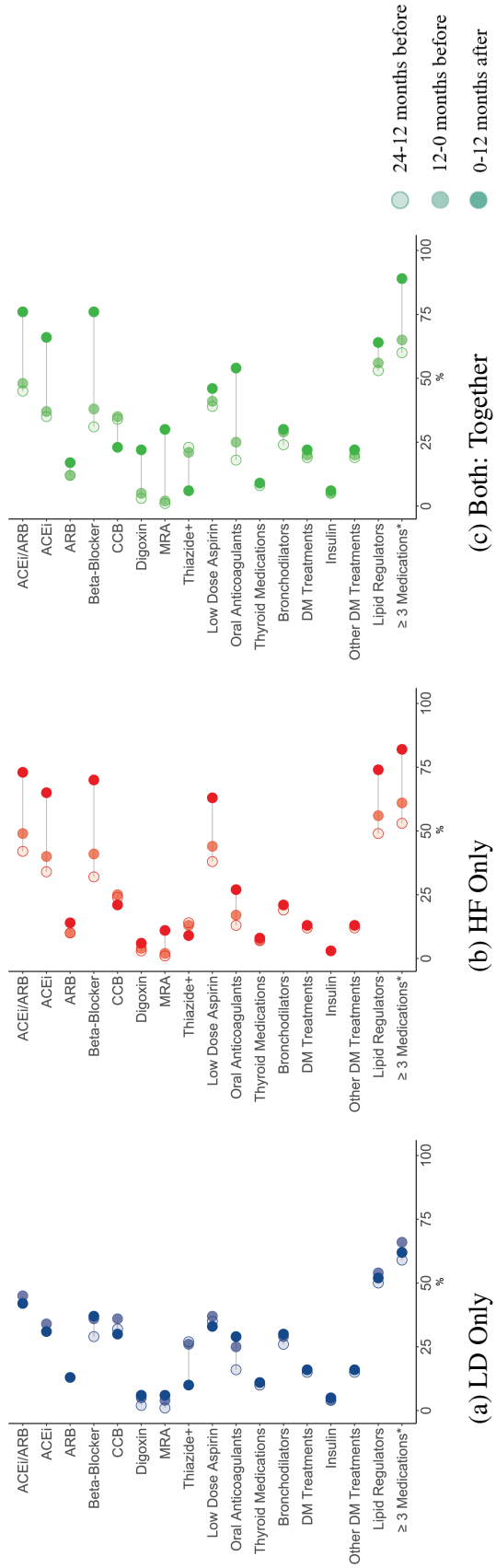
Figure 4.5: Admissions by primary cause in the year prior to the index event for the four most frequent ICD-10 Chapters: **II**, neoplasms (box with pink borders); **IX**, circulatory system (box with yellow borders); **X**, respiratory system (box with green borders); and **XVIII**, symptoms and signs (box with purple borders).



(a) LD Only (b) HF Only (c) Both: Together

(d) Both: LD First (e) Both: HF First (f) HF Death as First Record

Figure 4.6: Changes in the prevalence of comorbidities in the two years leading up to diagnosis and the year following. A patient is considered to have a comorbidity if they have a record during or before each interval. ∗ Does not include MI or EP devices.



(a) LD Only

(b) HF Only

(c) Both: Together

(d) Both: LD First

(e) Both: HF First

(f) HF Death as First Record

Figure 4.7: Changes in medication levels in the two years leading up to diagnosis and the year following. A patient is considered to be adherent to medication if they had a community-based prescription dispensed within an interval. $*$ ACEi and ARB are counted separately.

Of note, the network graphs highlight common pathways and centrally connected reasons for admission. Admissions reasons that occur frequently, but not often in a set order nor well connected to other admissions, will not appear significant within the graphs but will contribute to numbers in Figures 4.5 and C.2.

4.4.2 Changes in Comorbidities and Medications

For those with incident HF, substantial increases in the diagnosis of IHD, MI, AF/AFL, and CKD in the prior 12 months were observed (see Figure 4.6), with medications used to treat these conditions following a similar pattern (see Figure 4.7). In contrast, for patients initiated on LD who did not go on to receive a diagnosis of HF, there was little change in the rate of cardiovascular diagnosis and, apart from a reduction in the prescription of Thiazides+ diuretics, few changes in medications.

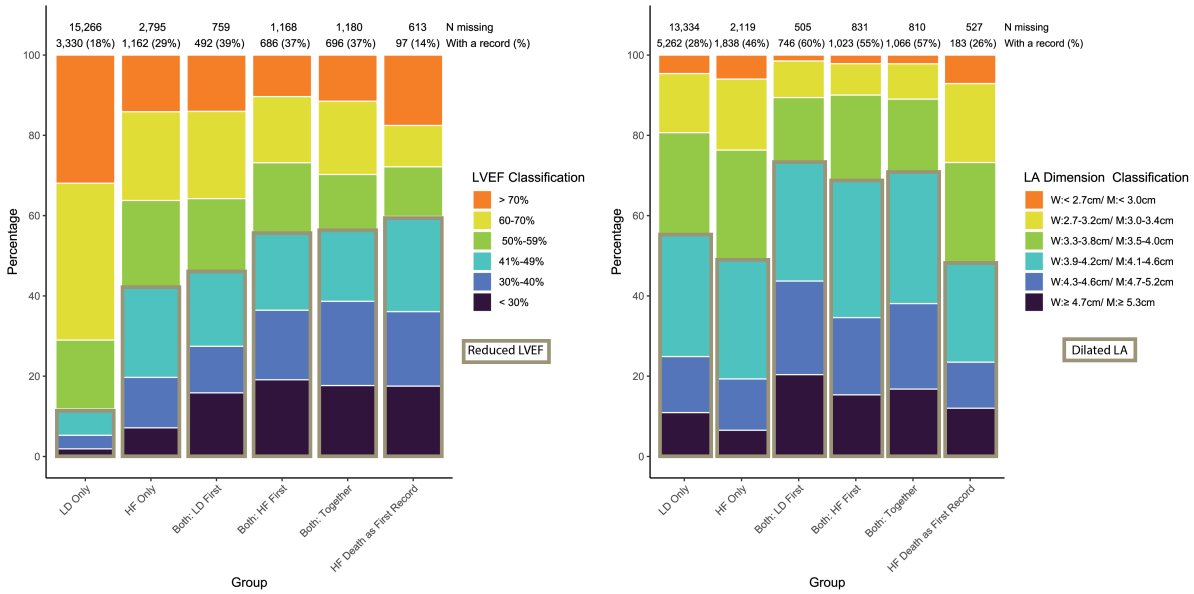
4.4.3 Investigations

Fewer than half of patients in any LD or HF category had echocardiographic data available. Results may not have been migrated to a central system, and normal values may also have been less likely to be entered. When LA diameter was reported, it was increased in about 50% of patients in the LD only or HF only groups and in about 70% of those who fulfilled both criteria (see Figure 4.8b). When measured, LVEF was < 50% in about half of the cases with an index event of HF (see Figure 4.8a). Of patients prescribed a LD alone, only 12% had an LVEF < 50% but 32% had a supra-normal LVEF (> 70%).

ECG data were available for 59% of patients in the cohort overall, with availability increasing to about 72% where HF and LD were both present. ECGs done outwith NHS GG&C or where the machine was not MUSE compatible were not available for analysis. For patients in the LD only or HF only groups, approximately 80% were in sinus rhythm, and approximately 17% had atrial arrhythmias (predominantly AF). For patients with HF initiated on LD, approximately 65% were in sinus rhythm, and approximately 30% had atrial arrhythmias (again, predominately AF) (see Table 4.2 and Figure 4.9). For those with ECG data, the LD only group had the shortest median QRS duration of 88 ms, while patients with both LD and HF had a median duration between 94 and 96 ms and 22% had QRS duration $\geq$120 ms. QTc prolongation was less common for the LD only group at 16% and highest for patients who fulfilled both LD and HF criteria (28 - 31%). The prevalence of ST-T abnormalities was highest when there was a diagnosis of HF, reflecting high levels of prior MI.

Blood tests showed that a substantial proportion of patients had anaemia; 25% for those in the HF only group and >30% in all other groups. Moderately impaired renal function (eGFR <60 mL/min/1.73m^2) was common, affecting about 20% of patients but few patients had an eGFR

<30mL/min/1.73m^2. About 20% of patients have a serum urea value above the normal range (2.5 - 7.8 mmol/L), indicating impaired renal function or dehydration. Serum sodium (normal range 135 - 145 mmol/L) was generally normal, but about 11% of patients had a value <135 mmol/L. Serum chloride (normal range 96 - 106 mmol/L) was increased in about 25% of cases. Serum bicarbonate (normal range 23 - 29 mmol/L) was low in about 50% of patients and substantially reduced in about 25% of cases, suggesting metabolic acidosis, which may reflect renal dysfunction. As the number of anions (negatively charged ions) and cations (positively charged ions) in blood should generally be balanced, serum chloride and bicarbonate are expected to have this sort of reciprocal relationship. Serum potassium (normal range 3.5 - 5.0 mmol/L) was generally normal. Only a few patients had values <3.5 mmol/L or >6.0 mmol/L (see Table 4.3).



(a) Left ventricular ejection fraction (LVEF) (b) Left atrial (LA) diameter

Figure 4.8: Distribution of (a) LVEF, and (b) LA diameter based on echocardiogram recorded closest in time to index event for patients classified by prescription of LD and a diagnosis of HF. The grey boxes indicate a reduced ejection fraction (LVEF < 50%) or a dilated LA.

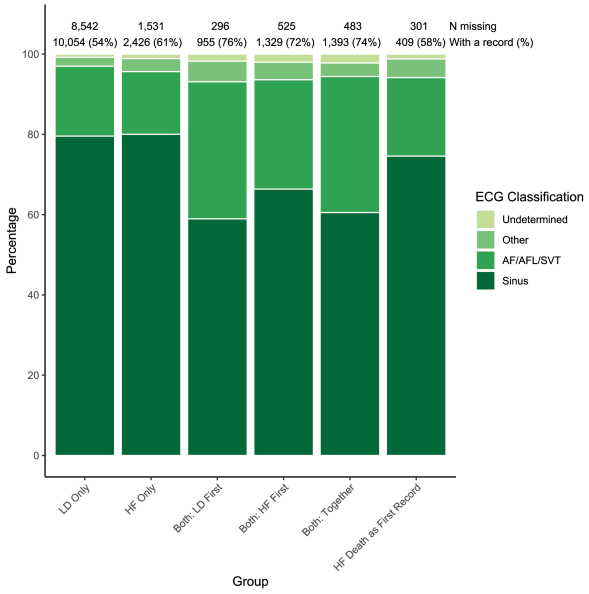


Figure 4.9: Distribution of heart rhythms based on ECG recorded closest in time to index event for patients classified by prescription of LD and a diagnosis of HF.

Table 4.2: Rhythm classification by group including the percentage of patients without a 12-leadECG. Data are taken from the test recorded closest recorded ECG to the index event (e.g., first of HF diagnosis or LD initiation) within the 3-year study window.

Variable	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**	**HF Death as First Record**
n	18,596	3,957	1,251	1,854	1,876	710
Age (years)	73 (62 - 81)	68 (57 - 78)	77 (69 - 84)	75 (66 - 83)	73 (65 - 81)	79 (71 - 86)
ECG available	10,054 (54%)	2,426 (61%)	955 (76%)	1,329 (72%)	1,393 (74%)	409 (58%)
Heart rate (bpm)	79 (67 - 93)	76 (64 - 91)	82 (68 - 98)	84 (70 - 102)	88 (71 - 108)	88 (74 - 105)
QRS duration	88 (80 - 98)	92 (84 - 108)	96 (84 - 116)	94 (84 - 114)	96 (84 - 114)	92 (82 - 110)
QRS $\geq$ 120 ms	938 (9%)	413 (17%)	208 (22%)	294 (22%)	306 (22%)	69 (17%)
Rhythm						
Sinus	8,006 (80%)	1,937 (80%)	566 (59%)	885 (67%)	842 (60%)	300 (73%)
AF/AFL/SVT	1,744 (17%)	381 (16%)	323 (34%)	362 (27%)	474 (34%)	83 (20%)
Other	224 (2%)	82 (3%)	49 (5%)	58 (4%)	47 (3%)	18 (4%)
Undetermined	80 (1%)	26 (1%)	17 (2%)	24 (2%)	30 (2%)	8 (2%)
QTc available ■	9,721 (97%)	2,309 (95%)	926 (97%)	1,274 (96%)	1,348 (97%)	393 (96%)
QTc (ms)	423 (406 - 442)	429 (409 - 452)	437 (414 - 462)	435 (412 - 462)	432 (410 - 458)	429 (408 - 454)
Prolonged QTc	1,422 (15%)	538 (23%)	284 (31%)	398 (31%)	380 (28%)	102 (26%)
ST-T abnormality	3,048 (30%)	986 (41%)	434 (45%)	638 (48%)	671 (48%)	195 (48%)
Acute MI△	52 (2%)	104 (10%)	12 (3%)	35 (5%)	34 (5%)	14 (7%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

■ QTc could not be calculated in patients whose QRS complex or RR interval was suppressed.

△ ECG detected acute MI

Table 4.3: Blood results based on the most recent value in the year before the index event (e.g., first of HF diagnosis or LD initiation).

Variable	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**	**HF Death as First Record**
n	18,596	3,957	1,251	1,854	1,876	710
Age (years)	73 (62 - 81)	77 (69 - 84)	73 (65 - 81)	75 (66 - 83)	68 (57 - 78)	79 (71 - 86)
Haemoglobin (Hb) (g/dL)						
Hb available	13,260 (71%)	2,388 (60%)	894 (71%)	1,208 (65%)	1,207 (64%)	505 (71%)
Women: Hb	12.8 (11.8 - 13.7)	12.9 (11.8 - 13.8)	12.7 (11.7 - 13.8)	12.6 (11.6 - 13.5)	12.7 (11.7 - 13.8)	12.3 (11.1 - 13.2)
Men: Hb	13.7 (12.4 - 14.8)	14.3 (13.1 - 15.3)	13.5 (12.3 - 14.6)	13.7 (12.4 - 14.9)	14.0 (12.7 - 15.1)	13.1 (11.7 - 14.2)
Anaemic ◇	4,105 (31%)	599 (25%)	302 (34%)	423 (35%)	371 (31%)	221 (44%)
Estimated glomerular filtration rate (eGFR)∇ (mL/min/1.73m^2)						
eGFR available	15,002 (81%)	2,781 (70%)	1,044 (83%)	1,419 (77%)	1,415 (75%)	561 (79%)
eGFR	78 (62 - 90)	80 (65 - 92)	73 (56 - 86)	75 (58 - 87)	77 (60 - 89)	70 (53 - 84)
eGFR [30 - 60)	2,865 (19%)	462 (17%)	268 (26%)	350 (25%)	321 (23%)	150 (27%)
eGFR <30	382 (3%)	65 (2%)	33 (3%)	30 (2%)	22 (2%)	32 (6%)
Measured serum results (mmol/L)						
Urea available	15,009 (81%)	2,786 (70%)	1,044 (83%)	1,419 (77%)	1,415 (75%)	562 (79%)
Urea	5.8 (4.5 - 7.4)	5.7 (4.6 - 7.1)	6.3 (4.8 - 8.1)	6.2 (4.8 - 7.8)	6.0 (4.8 - 7.5)	6.2 (4.9 - 8.4)
Albumin available	14,521 (78%)	2,689 (68%)	1,015 (81%)	1,360 (73%)	1,361 (73%)	543 (76%)
Albumin	36 (34 - 38)	37 (35 - 39)	36 (34 - 39)	36 (34 - 38)	37 (34 - 39)	35 (33 - 38)

Continuation of Table 4.3

Variable	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together	HF Death as First Record
n	18,596	3,957	1,251	1,854	1,876	710
Albumin <30 mmol/L	1,063 (7%)	123 (5%)	61 (6%)	78 (6%)	72 (5%)	59 (11%)
Sodium available	14,974 (81%)	2,778 (70%)	1,038 (83%)	1,414 (76%)	1,414 (75%)	560 (79%)
Sodium	139 (137 - 141)	139 (137 - 141)	139 (137 - 140)	139 (137 - 141)	139 (137 - 141)	138 (136 - 140)
Sodium < 135 mmol/L	1,583 (11%)	224 (8%)	104 (10%)	155 (11%)	160 (11%)	80 (14%)
Chloride available	15,001 (81%)	2,782 (70%)	1,044 (83%)	1,419 (77%)	1,415 (75%)	561 (79%)
Chloride	104 (101 - 106)	104 (102 - 106)	104 (101 - 106)	104 (101 - 106)	104 (101 - 106)	103 (101 - 106)
Potassium available	14,864 (80%)	2,749 (69%)	1,037 (83%)	1,406 (76%)	1,404 (75%)	556 (78%)
Potassium	4.3 (4.0 - 4.6)	4.3 (4.1 - 4.6)	4.3 (4.0 - 4.7)	4.3 (4.1 - 4.6)	4.3 (4.0 - 4.6)	4.3 (4.0 - 4.7)
Potassium >6.0 mmol/L	<6 (<1%)	<6 (<1%)	0 (0%)	0 (0%)	<6 (<1%)	0 (0%)
Potassium <3.5 mmol/L	405 (3%)	45 (2%)	26 (3%)	32 (2%)	23 (2%)	8 (1%)
Bicarbonate available	4,743 (26%)	970 (25%)	323 (26%)	466 (25%)	399 (21%)	179 (25%)
Bicarbonate	23 (21 - 26)	23 (21 - 25)	23 (21 - 25)	23 (21 - 25)	23 (21 - 26)	22 (20 - 25)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

◇ Using the WHO definition of anaemia assuming no pregnancies.

∇ Calculated eGFR assuming no pregnancies and without adjusting for ethnicity (see Section 2.5.9).

4.5 Discussion

This Chapter illustrates the similarities and some of the differences in the events leading up to the initiation of LD therapy and/or a diagnosis of HF.

Out of the six groups, by far, the largest group of patients were those who were initiated on LD and did not receive a diagnosis of HF. Across the groups, there were some differences in age and sex, but in the 12 to 24 months prior to the index event, there was little difference in baseline comorbidity and medication levels. In the year prior to the index event, 49% of the entire cohort had a hospital admission, but only a minority of these admissions were for CV disease, with many of the other admissions classified as admission for non-specific symptoms and signs without a definitive diagnosis, or for respiratory disease. However, in the year prior to an index HF event, there were more admissions for IHD (and acute MI in particular), AF/AFL, and valve disease, plus increased numbers of patients with CKD, all of which are well-recognised factors that conspire in the progression of HF. These patterns in admissions and diagnosis were not seen when the initiation of LD was the index event.

Previous studies have also shown that symptoms consistent with HF, such as oedema and breathlessness, are often recorded in the year before diagnosis (Sandhu et al. 2021, Bottle et al. 2018, Koudstaal et al. 2017). An event such as an acute MI, the onset of AF, or a diagnosis of valve disease may not only trigger the development of HF but also increase the likelihood that the diagnosis will be investigated and recorded. Previous reports suggest that the presence of comorbidities such as COPD, CKD, AF, DM, and hypertension increase the probability of receiving a diagnosis of HF in secondary care rather than in the primary care setting, and that patients diagnosed in secondary care have a poorer prognosis (Ezekowitz et al. 2011).

Of those patients diagnosed with HF, 44%[7] did not receive a LD within the subsequent year. Compared with the other groups, these patients were younger (difference of 5 years between the median ages), more often men, often had a prior MI, often had an LVEF <50%, and were more likely to receive their diagnosis of HF in a primary care setting. It is possible that these patients were given the diagnostic label of HF due to their medical history and reduced LVEF, rather than because they had symptoms and signs of HF. The lack of an initiated LD suggests they were unlikely to have clinical evidence of congestion. Additionally, in the year before the diagnosis of HF, these patients had a substantial increase in the number of recorded comorbidities and increasing numbers of medication classes, implying either more complex disease or heightened health surveillance and intervention.

The other 56% of patients diagnosed with HF were also initiated on LD within a year window of the diagnosis. Their median age was approximately 75 years, and about half were women. The

[7]This disregards the HF death as first record group as they are a special case compared with the other groups due to definition and death precluding the initiation of LD.

network analysis of the hospital admissions found a mixed picture of admissions for symptoms and signs without a definitive diagnosis, respiratory disease, and smaller contributions from acute MI and AF/AFL in the year prior to the index event. Most patients were diagnosed in secondary care, and about half of the patients had an LVEF <50%, and most had a dilated LA, plus many had an atrial arrhythmia, all supporting a diagnosis of HF. The pattern and changes in comorbidities and medications appeared similar to those with a diagnosis of HF who were not initiated on LD, though a larger decline in Thiazides+ usage.

For every patient diagnosed with HF, two patients were initiated on LD without receiving a diagnosis of HF in the following year, although the diagnosis might have been prevented in some cases as 7% of the LD only group died within 90 days of their first LD prescription date. This group of patients had a median age of 73 years, were more likely to be women, and at two years prior to the index event (initiation of the LD), had a similar pattern in the prevalence of comorbidities and medications compared with patients with HF. However, there was little increase in the prevalent comorbidities recorded in the year prior to the index event, and levels of medication usage were less likely to change after the index event. The only substantial change in medications (other than the introduction of LD) was a reduction in the use of Thiazides+. This is expected as the combination of thiazide and LD has an extremely powerful diuretic effect that may be dangerous when they are co-administered unintentionally. Additionally, this group of patients are more likely to have an ECG indicating sinus rhythm. They were less likely to have an echocardiogram recorded, and where recorded, a smaller proportion had an LVEF <50%. However, many had a dilated LA, and many had an LVEF >70%; recently a new HF phenotype with a supra-normal LVEF (HFsnEF) has been proposed (Wehner et al. 2019).

It is unclear what proportion of the patients initiated on LD for symptoms and signs of HF, but it may be substantial. The current criteria for the diagnosis of HF are not robust because they have a substantial subjective component which requires the patient to report symptoms and, subsequently, the healthcare professional to 'hear' the patient and act. Based on admissions alone, and not including primary care records, many patients are treated for problems such as breathlessness and ankle swelling without further investigations or diagnosis.

Breathlessness is a common feature of heart and lung diseases, and it may be impossible for patients, and difficult for doctors, to tell what is the main cause of the breathlessness (Pellicori et al. 2020). Many HF patients will also have lung disease, which one expects to contribute to a patient's breathlessness since they share common risk factors (e.g., smoking, air pollution, and older age) (Pellicori et al. 2022). In cases where the treatments for chronic lung disease are not effectively relieving symptoms, the doctor may be tempted to initiate a trial LD to see if it helps. It is unclear how often such patients benefit from LD initiation. The fact that patients in this analysis fill repeated prescriptions suggests that they might receive a benefit. The alternative explanation is that the prescriptions continue to be filled due to prescribing inertia (i.e., the

automatic renewal of prescriptions) and patients dutifully adhering to instructions.

LD may also be used to treat resistant hypertension (National Institute for Health and Care Excellence 2018*a*). However, few patients were receiving three (12%) or more (3%) other anti-hypertensive agents, and very few received an MRA, which is an excellent treatment for resistant hypertension. LD may also be used to treat severe renal disease, but very few patients had (3%) or developed an eGFR <30 mL/min/1.73m^2.

Many reports suggest that women with CV disease are less likely to be investigated and are managed differently compared with men. These results reinforce these concerns; the majority of patients treated with LD were women, and this group of patients was less likely to have an ECG or an echocardiogram recorded, which probably contributed to many missed diagnoses.

When initiation of a LD was the index event, the node representing admissions for non-specific symptoms and signs (ICD-10 Chapter XVIII) admissions was frequent and well-connected to other admissions. Public Health Scotland (PHS) only assigns this code when no specific or precise diagnosis has been made by the end of the period of care that can account for presentation (Christie 2022). Within Chapter XVIII, the most common 4-digit discharge code was R07.4 'Chest pain, unspecified'. In SMR01, a patient admitted with chest pain, raised troponin, an abnormal ECG, and diagnosed with acute anterior MI would be coded as ICD-10 code I21.0 for 'Acute transmural myocardial infarction of anterior wall' with an extra digit[8] (I21.09) to indicate MI without confirmation of ST elevation. As MI was diagnosed, the chest pain symptom should not be recorded. On the other hand, if a patient were admitted with chest pain, normal troponin, and no changes on the ECG, a discharge code of R07.4 'Chest pain, unspecified' would be recorded (Christie 2022).

This analysis found that 710 patients (11% of first admissions for HF) died during the first admission for HF without having received a LD prior to this admission. Consistent with previous reports, they were older compared with the LD/HF groups. However, despite having HF diagnosed during the admission, only 341 (49%) had a CV reason recorded as the underlying cause of death on their death certificate. These patients were clearly different from other HF patients in terms of age, prior admissions patterns, and comorbidity and prescription rates. No treatment has been shown to reduce in-patient mortality for HF. Clearly, death precludes obtaining the longer-term benefits of guideline-recommended therapy for HF. It is possible that some patients might have had undiagnosed HF for months prior to their terminal admission but did not receive treatment for congestion with LD prior to the event. Other patients will have experienced a large 'insult' to the heart, such as a massive MI or valve rupture, which subsequently caused HF, resulting in death in rapid succession. However, it is also possible that the term 'heart failure' was used to describe a patient dying from multiple-organ failure. Further research is required to determine whether the diagnosis of HF was accurate in these patients.

[8]This extra digit is referred to as the Scottish 5st digit.

4.5.1 Strengths and Limitations

Beyond the benefits of the cohort size mentioned in Chapter 3, with regard to this Chapter, strengths include the complete coverage and longitudinal nature of the SMR01 and PIS datasets for all Health Boards, both regional and specialised, throughout Scotland.

The main limitation of this analysis is when a patient experiences both the initiation of LD and a diagnosis of HF within a year, the network and admissions analyses do not include admissions that occur between the first (index) and second key events. Also, the GSH/18/CA002 dataset does not provide access to in-hospital prescribing of LD, so rates of LD use may be under-reported for those who died in hospital or shortly after discharge and the initiation date for LD will have been delayed until patients start receiving their community-based prescriptions.

As was referenced in Section 2.7.3, the lack of out-patient specialist diagnostic information may mean that the diagnosis of HF is either missed or appears at a later date within the available EPR. Information on ECGs was only available from MUSE-compatible electrocardiographs. As such, some ECGs may not have been included in EPR, which is less likely for electrocardiographs located within cardiology departments and wards. Few patients had measurements of natriuretic peptides, biomarkers of congestion that, when normal, rule out a diagnosis of HF.

4.5.2 Future Work

There was substantial variation in the proportion of women among the different LD/HF groups. Stratification of analysis by sex should be considered in future, especially as women are more likely to have HFpEF, a form of HF associated with different pathophysiology, treatment, and outcome (Pellicori & Cleland 2015, Beale et al. 2018, Ho et al. 2012). This analysis could be further expanded through the implementation of sequence analysis to capture the time information around the admissions, including lengths of admissions and time gaps between sequential admissions, followed by cluster analysis, which could identify similar subgroups of patients within these larger groups, similar to an analysis performed in those with alcohol use disorder (Han et al. 2019). Furthermore, the network graphs do not incorporate events between the index and second key events (e.g., diagnosis of HF for the Both: LD First group). Incorporating this information will help to clarify the ordering and reasoning behind the comorbidity and medication changes illustrated in Figures 4.6 and 4.7.

To avoid immortal time biases, further analysis is needed to understand the temporal relationship of an incident HF diagnosis and the initiation of LD therapy on mortality. This will be addressed in Chapter 5.

4.6 Conclusions

This analysis highlights the differences and similarities in the admissions pathways leading to an incident diagnosis of HF or the initiation of LD. The pattern of events preceding a diagnosis of HF is consistent with the known risk factors and precipitants of HF. However, the pattern of events preceding the initiation of LD has not been previously reported and appears very different to that of HF, and is perhaps best described as a combination of 'neglect' and respiratory disease. This is a large group of patients composed of predominantly older women, which accounts for many more hospital admissions than those with a diagnosis of HF. Many are admitted for a respiratory disease which might masquerade as or conceal a diagnosis of HF, and, as shown in Chapter 3, and will be shown again in the coming Chapters 5 and 6, the presence of a repeat LD is associated with poor prognosis. It is likely that many, perhaps most, of these patients have undiagnosed HF. If so, the incidence and prevalence of HF may be up to three times higher than current estimates. For such a common, deadly, and treatable disease, this seems rather important.

Chapter 5

Relationship of incident HF and initiation of LD and the sequence of these events on prognosis

> There are no secrets that time does not reveal.
>
> *Jean Racine*
> *French Dramatist*

5.1 Introduction

Understanding the trajectory of a patient's 'journey' from what they would consider healthy to having a disease, and modelling the progression of that patient's condition either to recovery (unfortunately rare for patients with HF) or death could offer important insights with regard to improving the scientific understanding of the disease. Additionally, this knowledge could be used to alert healthcare professionals to changes in the patient's risk, act as a signal that further investigations and treatment may be appropriate, and/or identify potential new therapeutic targets for research (Cleland 2002, Cleland et al. 2006). For example, if only a few HF patients die before they receive a LD, then attention should be focused on the initiation of the LD. Did the addition of LD reflect disease progression and worsening congestion, with the initiation of LD being a prognostic marker (Pellicori et al. 2016, 2019)? Alternatively, did the initiation of the LD drive disease progression by activating the neuro-endocrine system, causing metabolic disturbances?

Data from RCTs of anticoagulants for AF (Cleland et al. 2007), and hypoglycaemic therapy for DM (Pellicori et al. 2021) found that many patients are treated with LD at baseline but did not

have a diagnosis of HF; plus they had a poor prognosis, similar to those with HF. On the other hand, many patients who had a diagnosis of HF but were not treated with LD at baseline; these patients had a fairly good prognosis. However, these analyses are 'static' in that classifications are based on the trial baseline and do not consider changes in patient status, nor the incident diagnosis of HF nor initiation of LD during follow-up.

Understanding patients' progression from a healthy state may provide further insights into the relative importance of a diagnosis of HF and the initiation of LD, and help with future RCT designs that investigate whether observed relationships are causal. For instance, a RCT of LD withdrawal in the absence of evidence of congestion could be done to determine the effect on symptoms, patient well-being, and prognosis. A RCT conducted in the Netherlands (where 23% of older individuals are reported to receive LD [van Kraaij et al. 1998]) found that attempting to withdraw LD often led to the appearance of symptoms and signs (S&S) of HF (Walma et al. 1997). On the other hand, a RCT of initiating LD in patients with cardiac dysfunction and sub-clinical congestion might also be designed to investigate effects on symptoms, patients' well-being, and prognosis. Surprisingly, there is no substantial RCT on LD initiation in HF patients despite the fact that LD are considered to be a mainstay of treatment for congestion. This might be because LD are so effective at relieving S&S.

The following describes how patients transition from a state of health through the diagnosis of HF and initiation of LD through to death in a population with a broad range of CV conditions to determine which between diagnosis of HF and initiation of LD is the key event associated with prognosis, or are both required.

5.1.1 Aims

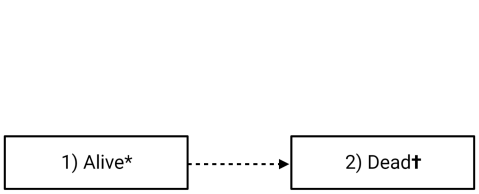
This chapter assesses the prognostic relationships between the initiation of a LD and the recording of a diagnosis of HF and whether the sequence of these events matters.

5.2 Background

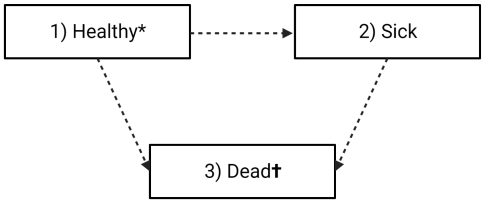
Time-dependent covariates (see Section 3.2.1) are one way to include updated information on patients during follow-up. Another option to investigate more complex relationships and disease pathways is multi-state models.

5.2.1 Multi-State Models

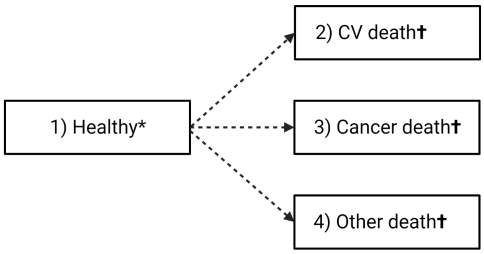
Multi-state models are representations of movement between a finite number of discrete states, where events (e.g., diagnosis, therapy, recovery, hospital admissions, and death[1]) are the transitions between the states (Hougaard 1999, Meira-Machado et al. 2009). The most basic of these models are standard survival models with two states, an initial and terminal states (see Figure 5.1a). Slightly more complex multi-state models include competing risk models (see Figure 5.1c and Section 3.2.1 for further information) where individuals start in an initial state and can progress to one of the competing states (e.g. CV death, death caused by a neoplasm, or death from other causes). Even more complex multi-state models provide a framework for including non-fatal events which occur prior to censoring or entering the final 'absorbing' state (i.e., dead) (see Figure 5.1b) (Putter et al. 2007, Andersen & Keiding 2002).



(a) 2-state survival model



(b) 3-state illness-death model



(c) 4-state competing risks model

Figure 5.1: Three examples of multi-state models: Figure (a) shows a simple mortality model like one would see in a KM plot or Cox PH regression analysis. Boxes represent states, and arrows represent transitions between the states. Figure (c) extends the simple mortality model to include competing risk states by including additional causes of mortality. Figure (b) extends the mortality model to include a potential, transitory disease state, which patients may transition through.
$*$ Inclusion state(s); † the terminal state.

Multi-state modelling uses a patient's data directly to model the time-dependent probabilities of moving between states of interest. It allows for continuous-time analysis using the event and censoring dates, instead of relying on discrete cycles or tunnel states[2]. Additionally, predictions

[1] In many cases, death can represent both the terminal state and the event which transitions individuals from their prior state to the death state.

[2] Separate, temporary states that allow for the incorporation of information from prior states or temporary adjustments that last for more than one cycle or state (Sonnenberg & Beck 1993, van Rosmalen et al. 2013).

from multi-state models enable intermediate states, allowing patients to enter and leave based on their individual event history. This contrasts with competing risk cumulative incidence analysis, where state entry is terminal.

Due to their robustness and flexibility, multi-state models are growing in popularity and are being employed across diverse health research topics (Klein & Shu 2002, Sutradhar et al. 2010, Putter et al. 2006, Gasperoni et al. 2017, Williams, Lewsey, Mackay & Briggs 2016, Foucher et al. 2005).

Markov Multi-State Models

The Markov property states that the future depends only on the present and does not depend on the past (Markov 1961), and was named after the Russian mathematician Andrey Markov. In the context of multi-state models, a multi-state Markov model assumes that the probability of transitioning to a new state depends only on the current values (Meira-Machado et al. 2009, Putter et al. 2007). These models are called 'clock forward' models (see Figure 5.2) as there is a single time scale from inclusion through follow-up. Due to the Markov assumption, transition probabilities can be calculated using the Kolmogorov differential equation (Meira-Machado et al. 2009).

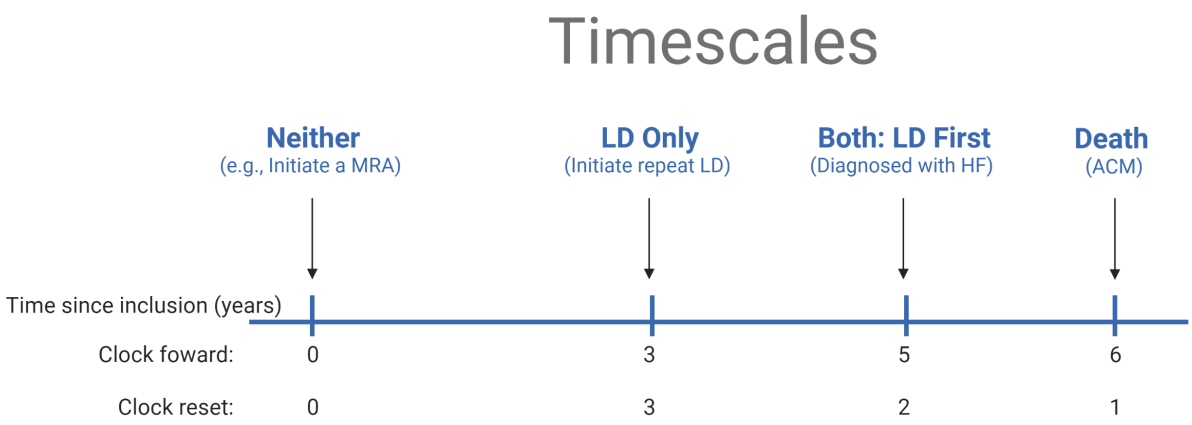


Figure 5.2: Markov (i.e., "clock forward") uses a single time scale based on time since entering the initial state; while semi-Markov (i.e., "clock reset") modes create a sequence of embedded Markov models, each with its own time scale.

Semi-Markov Multi-State Models

When the Markov property is not met, a semi-Markov multi-state model, also called a 'clock reset' model, can be employed. 'Clock reset' models break the Markov assumption because the time scale depends on a patient's history up to and including time since the current state was reached (Putter et al. 2007, de Wreede et al. 2010). Suppose one assumes that time since inclusion depends on the history of the process only through the current state, combined with

time since entering that state, and not on future events. In that case, the resulting multi-state model is formed from a sequence of embedded Markov models, each having a different time scale (Putter et al. 2007). Therefore, the probability of transitioning to another state depends on the current state and the time of entry into said state (Meira-Machado et al. 2009).

Due to the varying time scales found when implementing semi-Markov models, state occupancy and transition probabilities do not have an analytical solution. For this reason, probabilities must be sampled from a simulated approach adapted to survival data and multi-state models in order to estimate transition probabilities and standard errors (de Wreede et al. 2010).

5.3 Methods

5.3.1 Patient Identification

Using the GSH/18/CA/002 dataset (see Section 2.3.2 for specifics), the incidence of HF and newly initiated repeat LD prescription cases were investigated between 1st January 2012 through the 31st March 2017 against a population defined by the first of CAD, PAD, HF, or a dispensed ACEi, ARB, MRA, beta-blocker, or LD prescription between 31st December 2009 through the 31st March 2017.

Patients with a record of CAD, PAD, HF, ACEi, ARB, MRA, beta-blockers, or LD were identified using ICD-10, Read, and BNF codes (see Table A.11 for relevant codes).

An incident HF diagnosis was defined as the first Read or ICD-10 code across primary care and hospital admissions with HF recorded in any diagnostic position (see Section 2.6.1 for further information) unless the patient did not survive that diagnostic HF admission. Such cases of HF were excluded for several reasons. It appears that many frail patients, hospitalised for various reasons, including terminal cancers and infections, will have HF recorded as a diagnostic code. Also, it was not possible to obtain information on in-patient prescribing of LD. Finally, if the interval between diagnosis and death is short, there is little opportunity for diagnostic investigation or effective treatment.

A history of repeat LD was defined as having community-prescribed LD dispensed over two consecutive quarters or where the patient died within 90 days of the first prescription (see Section 2.5.4 and Table A.3 for qualifying medications). The initiation of LD therapy was defined as the first time the above definition was met.

5.3.2 Study Population

The study population was restricted to adults aged 18 years or older on 1st January 2012. The patient's inclusion date was defined as the later of 1st January 2012 or the first date of a qualifying

inclusion reason (see Figures D.1 and D.2) through 31st March 2017, inclusive. This date range allows for a period after prescribing records started to become available (2010) to determine incident versus prevalent HF and repeat LD usage and, where appropriate, to determine the temporal ordering of the two events. The inclusion window ended on 31st March 2017, allowing at least one year of follow-up within the limitations of the dataset for all patients.

Patients were excluded (see Figures D.1 and D.2) if they were classified as mislinked (see Section 2.5.1); if they did not have at least five years of EPR records before inclusion to provide a look-back window (see Section 4.2.1); if they had records of HF or repeat LD usage in the 5-year look-back window; if they were censored before the 1st January 2012; if they did not have follow-up after inclusion, or if there was uncertainty about the HF diagnosis date (e.g., the patient's first record of HF indicated a pre-existing diagnosis [see Table A.16 for a list of codes], or where the diagnosis date was missing). See Figure D.1 for examples of patient timelines resulting in inclusion or exclusion.

5.3.3 Study Outcomes

Patients were followed from cohort inclusion until death or 31st March 2018. Follow-up was right censored at the time of the last medical contact or investigation (see Section 2.6.3). The primary outcome of interest was ACM. The primary covariant of interest was the patient state defined by the initiation of LD or a new diagnosis of HF, and, where applicable, the temporal order of these events. First-, second- and third-year crude cause-specific hospitalisation and mortality rates were also reported from inclusion and subsequent initiation or diagnosis. Cause-specific hospital admissions were classified using the first diagnostic position and admission type of a hospital stay (see Sections 2.5.5 and 2.5.5). The cause of death was categorised by the underlying cause (see Section 2.5.1). Classifications of the cause of death and hospital admissions by disease category are reported in Table D.6.

5.3.4 Patient Classification and Multi-State Model Analysis

In order to investigate the implications of the temporal relationship between the initiation of a repeat LD and diagnosis of HF on ACM, a uni-directional multi-state model was developed where each state represents patients' LD and HF status. Models were prepared according to the flow diagram outlined in Figure 5.3 with more details in the following sections.

Defining States and Transition Matrix

The main multi-state model was designed using seven states to encode the presence or absence of LD and HF and the order of events, where applicable. The seven states and the requirements for being in each state are:

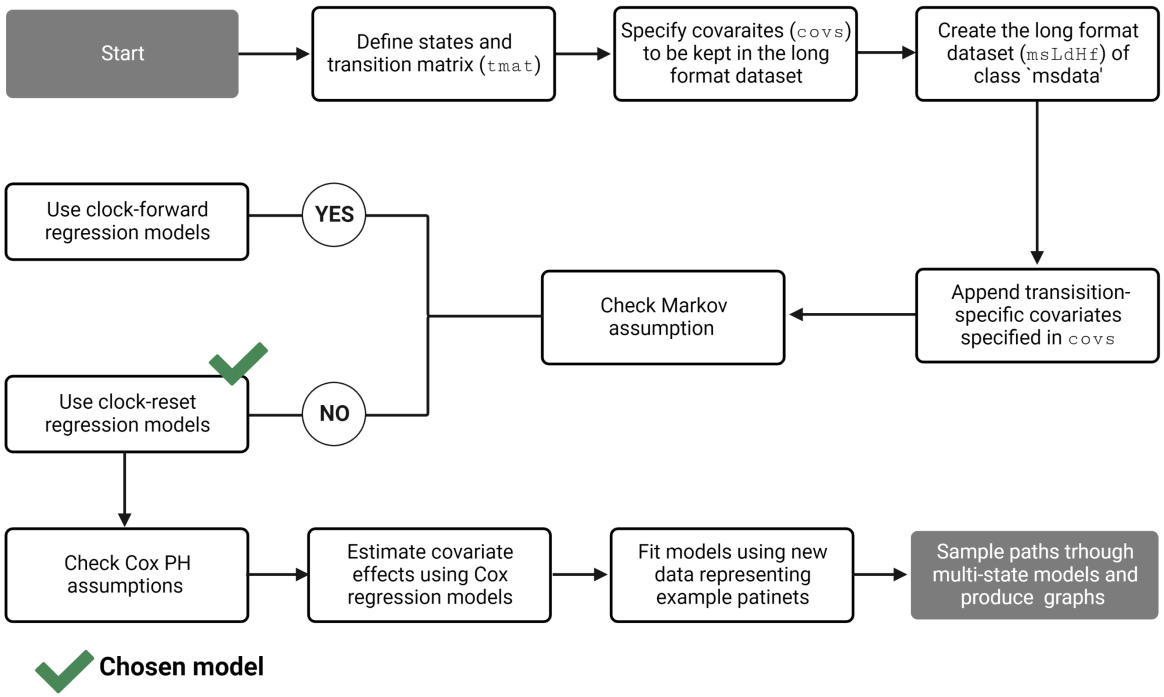


Figure 5.3: Flow diagram illustrating the steps taken to build a multi-model using the `mstate` and `survival` packages in R.

1. Neither - absence of a diagnosis of HF or repeat LD prescription/dispensing.

   Note: These patients were either receiving treatment for CV disease (predominantly hypertension) and/or had a diagnosis of CAD/PAD.

2. LD Only - Recurrent LD dispensing in the absence of a HF diagnosis.

3. HF Only - A diagnosis of HF in the absence of recurrent LD dispensing.

4. Both: LD First - Receiving a HF diagnosis >30 days after starting recurrent LD.

5. Both: HF First - Initiation of recurrent LD dispensing >30 days after a diagnosis of HF.

6. Both: Together - Diagnosis of HF and initiation of LD within a 30-day window[3].

7. Death - ACM.

Patients can enter the model in one of four inclusion states: Neither, LD Only, HF Only, and Both: Togethers (as illustrated in Figure 5.4), and can transition between zero and three times depending on their initial inclusion state, and whether or not they are subsequently diagnosed with HF, initiated on a repeat LD, or die. Events causing transitions are shown in Table D.1.

Of note, in cases where a record for CAD or PAD occurred (thus triggering cohort entry) on

[3]The 30-day window was chosen over a 60- or 90-window to reduce the amount of immortal time bias incurred while taking into account the delay observed in the dispensing dates (`DISP_DATE`) (see Section 2.5.4).

the same day as the diagnosis of HF, the patient was forced through the 'neither' state for 0.5 days before transitioning, under the assumption that either of these conditions predated the HF diagnosis. In cases where the initiation of an ACEi, ARB, MRA, or beta-blocker occurred (thus triggering cohort entry) on the same day as the HF diagnosis, it was assumed that the HF diagnosis predated (and initiated) the medication's prescription. Therefore, the patient would not pass through the 'neither' state based on the medication's presence. Finally, due to the prescription and dispensing date issue highlighted in Section 2.5.4, the date for Both: Together inclusion is based on the HF diagnosis date, as the diagnosis of HF is likely to be the reason for initiating LD. This is the same logic as applied in Chapter 4.

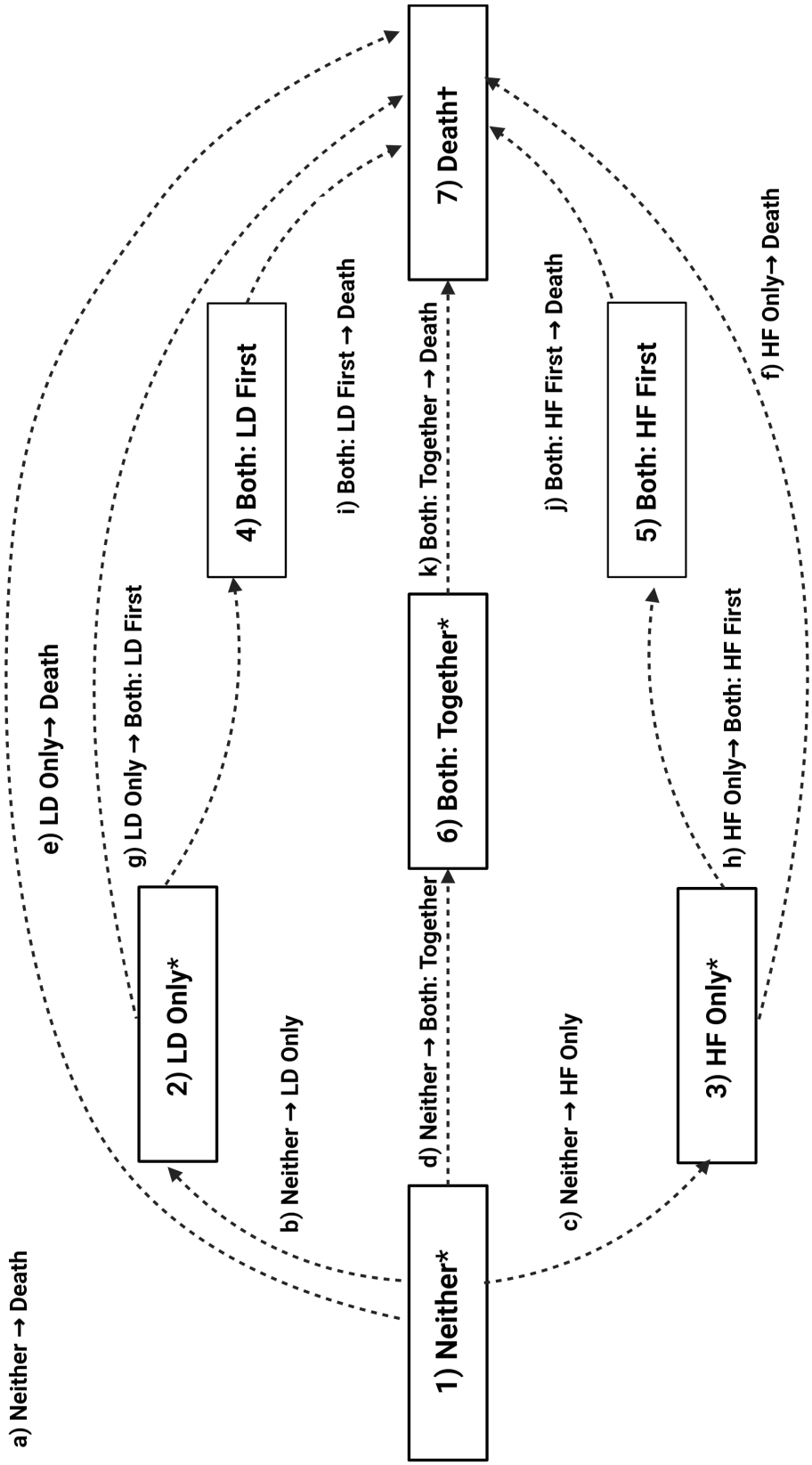


Figure 5.4: Multi-state diagram of the seven states used to track LD initiation and HF diagnosis with inclusion between 1 January 2012 through 31 March 2017 and an additional year of potential follow-up. Alphabetical labels indicate allowed transitions.
∗ Potential inclusion states.
† the 'absorbing' terminal state.

Patients entering directly into HF Only, LD Only, or Both: Together may have a different natural history and are therefore shown as separate multi-state models.

Specifying Covariates and Creating an Analysis-Ready Dataset

This process is split into three distinct steps: i) specifying the covariates, ii) creating the long format dataset, and iii) appending the transition-specific covariates.

First, the covariates used in the per-transition adjustments were: age per decade centred around 60 years (the median age) at cohort inclusion, sex (see Section 2.5.2), SIMD quintile using 2012 status (see Section 2.5.2), and inclusion year. The list of covariates was restricted to limit potential biases incurred from incorporating under-reported comorbidity levels. Age, sex, and SIMD were chosen due to the higher confidence levels with regard to the accuracy, combined with their known relationship to ACM, and the differences seen in age and sex between those with only a HF diagnosis and those with only a LD, as seen in Chapter 4. The inclusion year was included to adjust for unmeasured confounding. For later use, these variable names were stored in the `covs` vector.

Second, the transition matrix (`tmat`) (see Section 5.3.4 for further details) and the covariates specified above (`covs`) were combined to create a long format dataset objects of class `msdata`. Briefly, the long format `msdata` object has one row of data for every transition an individual patient is at risk of experiencing based on their event history (i.e., a patient in the 'HF only' state will have one row representing the potential transition to Both: HF First, and one row for potential transition to death). This pattern is replicated for all patients within the subset. For example, a patient who starts LD before being diagnosed with HF will not have rows corresponding to the probable transitions of HF only to Both: HF First or Both: Together to Death.

Third, and finally, transition-specific covariates specified in `covs` (see the first step) were appended to the `msdata` object, creating one column per transition per covariate (or covariate level for factors). (See Section D.7 for extracts of the implementation code for the three steps.)

Check Markov Assumption

Two methods were used to test the Markov property. The first method was the 'state-arrival extended' multi-state model method (Williams et al. 2017) using a covariate to indicate time spent in the 'neither' state. The second method was the 'Stratified Commenges-Andersen test' (Titman & Putter 2020). This test induces a non-Markov process by assuming unobserved frailty terms exist and is shared across the transitions of the multi-state model. The null hypothesis corresponds to zero variance of survival time in the stratified transitions. (Titman & Putter 2020). The Stratified Commenges-Andersen test was implemented using the `ca_test` function in the R package `frailtyEM`, version 1.0.1 (Balan & Putter 2019). (See Section D.7 for extracts of

	(a) High risk for LD only	(b) High risk for HF only
Age (years) at inclusion	80	65
Sex	Woman	Man
SIMD	1 (most deprived)	1 (most deprived)
Inclusion year	2012	2012

Table 5.1: Risk profiles used for prediction in the multi-state models

the implementation code.)

While the following is a result, rather than strictly methodology, it is included here as it dictated the type of multi-state model implemented. Testing the Markov assumption results in a test statistic of $1.29e^{-8}$ and a p-value of $2.44e^{-12}$, indicating that the null hypothesis should be rejected. This indicates that there is a non-zero variance across the stratified transitions, indicating that the Markov assumption fails; accordingly, a semi-Markov (clock reset) model should be used.

Check Assumptions and Fit Cox PH Models

Clock reset semi-Markov Cox PH models were used to assess the transition hazards. Due to the population sizes, assumptions for inclusion age, SIMD, and sex were assessed according to Section 3.2.1. The inclusion year was assumed to have a linear effect because, regardless of the outcome, the results still represent an average variation over time, just with a larger variation.

Estimate Covariate Effects and Fit Models

Models were adjusted for age by decade (centred around 60), sex, SIMD, and inclusion year while stratifying the transitions. See Section D.7 for extracts of the implementation code. Cox PH results are reported as HR with 95% 95% CI. Predictions were made based on two patient profiles as shown in Table 5.1. Predictions were made using the time of cohort inclusion and were repeated using each of the four models (i.e., those entering into Neither, LD Only, HF Only, or Both: Together).

The multi-state models were prepared using R (see Section 2.4.2) and the `mstate` (de Wreede et al. 2010) and `survival` (Therneau & Gramsch 2000) packages.

5.3.5 Patient Characteristics

For each patient, baseline characteristics were reported and subsequent initiation of LD or diagnosis of HF during follow-up (referred to as the reference date), and the final transition state (referred to as the foregoing state) before death, if they died.

Demographic information included age, sex (see Section 2.5.2), ethnicity (see Section 2.5.2), and SIMD quintile using the 2012 status (see Section 2.5.2).

Contact with secondary care was defined as either a record of clinic attendance in SMR00 or a hospital discharge (including day cases) within 30 days of initiating the repeat LD. The number and percentage of patients who attended a cardiovascular-based clinic (including cardiology, cardiac surgery, and vascular surgery specialities) within the same window were also reported.

Comorbidities included: history of hypertension, DM, thyroid disease, AF/AFL, CAD (including MI), valve disease, PAD, stroke, COPD, cancer, and dementia. The presence of these conditions was defined as a record on or before the date of interest (see Section 2.6.2 for definitions).

The current medication classes reported included: ACEi, ARB, beta-blockers, MRA, CCB including diltiazem and verapamil, digoxin, Thiazides+, low dose aspirin, lipid regulators, bronchodilators, thyroid medications, and hypoglycemic agents, including insulin. Patients were considered to be on these medications if dispensed in the six months prior to the reference date. Prescriptions were identified by BNF codes (see Table A.3 for the classifications).

The most recent blood test results reported in the two years prior to the reference date included: haemoglobin values stratified by sex (see Section 2.5.9) and the percentage with anaemia (see Section 2.6.2); calculated eGFR (see Section 2.5.9) and percentages of patients with values $\geq$30 and <60 mL/min/1.73m^2 and those with values <30 mL/min/1.73m^2; and serum values for potassium (see Section 2.5.9) and percentages of patients with values <3.5 mmol/L or values >6.0 mmol/L; sodium (see Section 2.5.9) and percentages of patients with values >135 mmol/L; urea (see Section 2.5.9); albumin (see Section 2.5.9) and percentages of patients with values $\geq$30 to <35 mmol/L and values <30 mmol/L; chloride (see Section 2.5.9); and bicarbonate (see Section 2.5.9). The eGFR was calculated according to Section 2.5.9.

From the closest 12-lead ECG to the reference date, reported results included heart rate (see Section 2.5.6) and rhythm (see Section 2.5.6), QRS duration (see Section 2.5.6, QTc (see Section 2.5.6) including proportion which was prolonged, and ST-T abnormalities (see Section 2.5.6) including those reported an acute MI.

From the closest echocardiogram to the reference date, LVEF (see Section 2.5.7), LA diameter (see Section 2.5.7), aortic velocity (see Section 2.5.7), tricuspid regurgitation (TR) (see Section 2.5.7), and the presence of aortic regurgitation (AR) (see Section 2.5.7) and mitral regurgitation (MR) (see Section 2.5.7) were reported.

5.3.6 Statistics

Patient characteristics are presented as numbers and percentages for categorical data and median (1st - 3rd quartile) values for continuous data. For categorical variables, percentages refer to

complete cases.

First, second, and third-year hospital admissions rates per state occupied were calculated as the number of admissions within the year divided by the number of patient-years at risk (see Section 3.2.2) where patients were eligible to be admitted (e.g., alive, not in hospital, and under follow-up). Patients were censored upon changing state. Similarly, first, second, and third-year crude mortality rates were calculated by dividing the number of within-year causes of death by the number of days the patient was alive.

Analysis was conducted in R (see Section 2.4.2) (R Core Team 2021), using the packages listed in Section 3.3.5 plus `mstate` (Putter et al. n.d., de Wreede et al. 2010, 2011).

5.4 Results

5.4.1 Inclusion Status

Eligibility criteria were met by 229,820 people, contributing a total of 1,082,434 years of follow-up, with a median time of 6 (1^{st} - 3^{rd} quartile: 3 - 6) years of follow-up. Note, this is a larger cohort than that identified using prevalent cases in Chapter 3 because patients could enter the cohort after 1^{st} January 2012. The cohort included 226,278 (98%) patients who entered the cohort due to a diagnosis of CAD, PAD, or a dispensed prescription of an ACEi, ARB, MRA, beta-blocker, or a single LD prescription, of whom 107,727 were aged $\geq$60 years.

Most patients who received LD met other inclusion criteria, but 2,371 (1%) patients started a repeat LD prescription as their first intimation of risk, which was associated with hospitalisation or hospital clinic visits in 957 (40%) cases. This group of patients were more likely to be women (1,523 [64%]) and often had anaemia (807 [38%]) (see Table 5.4), and a history of cancer (518 [22%]) (see Table 5.2). They were less likely to have a diagnosis of AF/AFL (146 [6%]) or CAD (85 [4%]); or to be on concurrent ACEi or ARB (52 [2%]) (see Table 5.3).

Similarly, most patients who were diagnosed with HF met other inclusion criteria, but 1,171 (1%) patients only met the inclusion criteria by virtue of a new diagnosis of HF (of whom 307 (<1%) were simultaneously initiated on a LD). Patients who were included due to a diagnosis of HF and concurrently started a LD were more likely to be men (172 [56%]) and have a related hospitalisation (281 [92%]). They were also more likely to have a history of AF/AFL (121 [39%]), or valve disease (71 [23%]), but less likely to be anaemic (84 [28%]).

Patients who entered the cohort due to a diagnosis of HF without a recurrent LD prescription were similar to the patients where both events occurred together, although they were less likely to have valve disease (132 [15%]).

Patients in the ‘neither’ group aged <60 years are clearly different from the other groups based

on age and were included predominantly because they were treated with ACEi/ARB or beta-blockers with relatively few patients having a history of CAD (10,394 [9%]), suggesting that treatment was given mainly for hypertension, although a record of this diagnosis was often not available, probably due to the limited primary care records. These younger patients had better renal function (only 1,624 [1%] had an eGFR <60 mL/min/1.73m^2), a higher haemoglobin, and higher serum albumin, but the results of other blood tests were similar amongst the groups.

Patients in the 'neither' group aged $\geq$60 years, had a similar median age to patients who entered the cohort due to initiation of LD or a diagnosis of HF, and often had a history of hypertension (39,647 [37%]), DM (25,781 [24%]), and CAD (28,419 [26%]), including MI (10,507 [10%]), and often received ACEi (45,786 [43%]), ARB (15,057 [14%]), and beta-blockers (40,301 [37%]). Compared to patients with HF, this group of patients was less likely to have a history of AF/AFL (8,557 [8%]), valve disease (1,878 [2%]), or COPD (12,640 [12%]). Compared to patients in the LD only group, this group of patients was less likely to have a history of cancer (9,825 [9%]).

Table 5.2: Baseline patient demographics and comorbidities on patient inclusion.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: Together**
n	118,551	107,727	2,371	864	307
Age (years)	48 (37 - 54)	72 (66 - 79)	70 (57 - 81)	71 (59 - 82)	70 (57 - 80)
Sex					
Women	68,795 (58%)	58,611 (54%)	1,523 (64%)	413 (48%)	135 (44%)
Men	49,756 (42%)	49,116 (46%)	848 (36%)	451 (52%)	172 (56%)
Ethnicity					
White	74,049 (62%)	86,479 (80%)	1,959 (83%)	764 (88%)	266 (87%)
Missing	40,306 (34%)	18,912 (18%)	388 (16%)	78 (<1%)	38 (12%)
Other	4,196 (4%)	2,336 (2%)	24 (1%)	22 (3%)	<6 (<1%)
Socioeconomic deprivation (SIMD)					
1 (most deprived)	51,583 (44%)	37,826 (35%)	992 (42%)	329 (38%)	117 (38%)
2	21,743 (18%)	19,358 (18%)	414 (17%)	158 (18%)	58 (19%)
3	15,928 (13%)	14,387 (13%)	352 (15%)	133 (15%)	39 (13%)
4	12,804 (11%)	14,337 (13%)	284 (12%)	103 (12%)	36 (12%)
5 (least deprived)	16,493 (14%)	21,819 (20%)	329 (14%)	141 (16%)	57 (19%)
Healthcare contact					
HF diagnosed in PC	N/A	N/A	N/A	255 (30%)	51 (17%)
LD started within 30 days post contact					
Any secondary care	N/A	N/A	957 (40%)	N/A	297 (97%)
Hospital discharge ⊠	N/A	N/A	600 (25%)	N/A	281 (92%)
Specialist clinic	N/A	N/A	582 (25%)	N/A	70 (23%)
CV specialist visit ∮	N/A	N/A	39 (2%)	N/A	28 (9%)

Continuation of Table 5.2

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	LD Only	HF Only	Both: Together
n	118,551	107,727	2,371	864	307
Comorbidities *					
H/o hypertension	17,620 (15%)	39,647 (37%)	319 (13%)	194 (22%)	66 (21%)
DM	14,825 (13%)	25,781 (24%)	228 (10%)	88 (10%)	25 (8%)
Thyroid disease	1,136 (1%)	2,701 (3%)	58 (2%)	30 (3%)	17 (6%)
CAD	10,394 (9%)	28,419 (26%)	85 (4%)	31 (4%)	9 (3%)
Of which is MI	5,398 (5%)	10,507 (10%)	19 (1%)	9 (1%)	<6 (1%)
Valve disease	546 (<1%)	1,978 (2%)	48 (2%)	132 (15%)	71 (23%)
AF/AFL	1,486 (1%)	8,557 (8%)	146 (6%)	243 (28%)	121 (39%)
PAD	974 (1%)	3,317 (3%)	18 (1%)	10 (1%)	<6 (1%)
Stroke	2,423 (2%)	9,265 (9%)	108 (5%)	95 (11%)	15 (5%)
COPD	6,178 (5%)	12,640 (12%)	485 (20%)	215 (25%)	82 (27%)
Cancer	2,544 (2%)	9,825 (9%)	518 (22%)	112 (13%)	39 (13%)
Dementia	45 (<1%)	2,649 (2%)	77 (3%)	55 (6%)	7 (2%)
EP Devices	206 (<1%)	1,313 (1%)	20 (1%)	14 (2%)	<6 (2%)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

☒ SMR01 hospital discharge includes day-cases and inpatient stays.

∮ CV specialties include cardiology, cardiac surgery, cardiothoracic surgery, and vascular surgery.

∗ History of a coded record on or before the inclusion date.

PC, primary care; EP device, cardiac electrophysiology device (e.g., pacemaker, ICD, and CRT).

Table 5.3: Concurrent medication levels defined by a dispensed medication within the 180 days leading up to patient inclusion.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: Together**
n	118,551	107,727	2,371	864	307
Age (years)	48 (37 - 54)	72 (66 - 79)	70 (57 - 81)	71 (59 - 82)	70 (57 - 80)
ACEi or ARB	35,328 (30%)	59,625 (55%)	52 (2%)	10 (1%)	<6 (1%)
ACEi	29,526 (25%)	45,786 (43%)	48 (2%)	7 (1%)	<6 (1%)
ARB	6,482 (5%)	15,057 (14%)	<6 (<1%)	<6 (<1%)	0 (0%)
Beta-blocker	44,843 (38%)	40,301 (37%)	65 (3%)	7 (1%)	<6 (<1%)
MRA	558 (<1%)	619 (1%)	40 (2%)	0 (0%)	0 (0%)
CCB	13,585 (11%)	36,064 (33%)	414 (17%)	135 (16%)	40 (13%)
Diltiazem/Verapamil	1,158 (1%)	4,349 (4%)	60 (3%)	22 (3%)	8 (3%)
Dihydropyridine	12,485 (11%)	31,901 (30%)	356 (15%)	115 (13%)	33 (11%)
Digoxin	130 (<1%)	1,678 (2%)	61 (3%)	14 (2%)	<6 (1%)
Thiazides+	10,344 (9%)	31,356 (29%)	350 (15%)	67 (8%)	20 (7%)
Low dose aspirin	13,588 (11%)	43,475 (40%)	361 (15%)	119 (14%)	38 (12%)
Oral anticoagulants	896 (1%)	4,903 (5%)	137 (6%)	39 (5%)	22 (7%)
Lipid regulators	24,322 (21%)	62,759 (58%)	552 (23%)	171 (20%)	54 (18%)
Bronchodilators	11,222 (9%)	15,659 (15%)	699 (29%)	210 (24%)	106 (35%)
Thyroid medications	5,139 (4%)	10,135 (9%)	226 (10%)	52 (6%)	24 (8%)

Continuation of Table 5.3

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: Together**
n	118,551	107,727	2,371	864	307
Hypoglycaemic agents	10,485 (9%)	15,396 (14%)	127 (5%)	31 (4%)	8 (3%)
Insulin†	2,729 (2%)	2,295 (2%)	19 (1%)	9 (1%)	0(0%)
Other hypoglycaemic agents	8,756 (7%)	14,346 (13%)	113 (5%)	23 (3%)	8 (3%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

† Either alone or in combination with another agent.

Dilt/Verap, Diltiazem/Verapamil; Thiazides+, thiazides and related.

Table 5.4: Most recent baseline patient blood results taken in the two years before the inclusion date.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: Together**
n	118,551	107,727	2,371	864	307
Age (years)	48 (37 - 54)	72 (66 - 79)	70 (57 - 81)	71 (59 - 82)	70 (57 - 80)
Haemoglobin (Hb) (mg/dL)					
Reported	84,635 (71%)	86,857 (81%)	2,147 (91%)	801 (93%)	302 (98%)
Women: Hb	13.4 (12.5 - 14.1)	13.0 (12.0 - 13.9)	12.7 (11.5 - 13.8)	12.7 (11.4 - 13.9)	13.0 (11.8 - 14.2)
Men: Hb	15.0 (14.1 - 15.8)	14.2 (13.1 - 15.1)	13.1 (11.5 - 14.5)	14.3 (12.5 - 15.4)	14.4 (13.0 - 15.5)
Anaemia◇	9,958 (12%)	20,213 (23%)	807 (38%)	256 (32%)	84 (28%)

Continuation of Table 5.4

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: Together**
n	118,551	107,727	2,371	864	307
Estimated glomerular filtration rate (eGFR)∇ (mL/min/1.73m^2)					
Reported	90,789 (77%)	100,695 (93%)	2,199 (93%)	814 (94%)	303 (99%)
eGFR	101 (91 - 108)	79 (65 - 88)	85 (71 - 96)	79 (63 - 93)	78 (66 - 90)
eGFR [30 - 60)	1,349 (1%)	17,053 (17%)	228 (10%)	155 (19%)	50 (17%)
eGFR <30	275 (<1%)	1,312 (1%)	26 (1%)	28 (3%)	6 (2%)
Serum values (mmol/L)					
Reported urea	90,882 (77%)	100,723 (93%)	2,199 (93%)	814 (94%)	304 (99%)
Urea	4.6 (3.7 - 5.6)	5.9 (4.8 - 7.2)	5.5 (4.2 - 6.9)	6.0 (4.5 - 8.0)	6.3 (5.1 - 7.9)
Reported albumin	88,010 (74%)	98,364 (91%)	2,166 (91%)	801 (93%)	302 (98%)
Albumin	39 (37 - 41)	37 (35 - 39)	35 (30 - 38)	35 (31 - 38)	35 (32 - 37)
Albumin [30-35)	7,265 (8%)	14,560 (15%)	516 (24%)	220 (27%)	104 (34%)
Albumin <30	2,912 (3%)	4,307 (4%)	537 (25%)	145 (18%)	33 (11%)
Reported sodium	90,843 (77%)	100,707 (93%)	2,201 (93%)	814 (94%)	304 (99%)
Sodium	139 (137 - 140)	139 (137 - 140)	139 (136 - 141)	138 (136 - 140)	138 (136 - 140)
Sodium < 135	4,230 (5%)	9,133 (9%)	288 (13%)	155 (19%)	53 (17%)
Reported chloride	90,797 (77%)	100,691 (93%)	2,199 (93%)	814 (94%)	304 (99%)
Chloride	104 (103 - 106)	104 (102 - 106)	103 (100 - 106)	104 (101 - 106)	104 (101 - 107)
Reported Potassium	89,903 (76%)	100,270 (93%)	2,185 (92%)	806 (93%)	302 (98%)
Potassium	4.2 (3.9 - 4.5)	4.3 (4.0 - 4.6)	4.2 (3.9 - 4.5)	4.3 (4.0 - 4.6)	4.2 (4.0 - 4.5)
Potassium >6.0	26 (<1%)	43 (<1%)	<6 (<1%)	<6 (<1%)	0 (0%)
Potassium <3.5	1,617 (2%)	2,446 (2%)	120 (5%)	53 (7%)	9 (3%)

Continuation of Table 5.4

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**LD Only**	**HF Only**	**Both: Together**
n	118,551	107,727	2,371	864	307
Reported bicarbonate	39,196 (33%)	53,518 (50%)	752 (32%)	287 (33%)	90 (29%)
Bicarbonate	23 (21 - 25)	24 (22 - 26)	23 (21 - 26)	23 (20 - 25)	22 (19 - 25)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

◇ Using the WHO definition of anaemia (see Section 2.6.2).

∇ Calculated eGFR assuming no pregnancies and without adjusting for ethnicity (see Section 2.5.9).

5.4.2 Patient Progression and Multi-State Analysis

Patient Movement

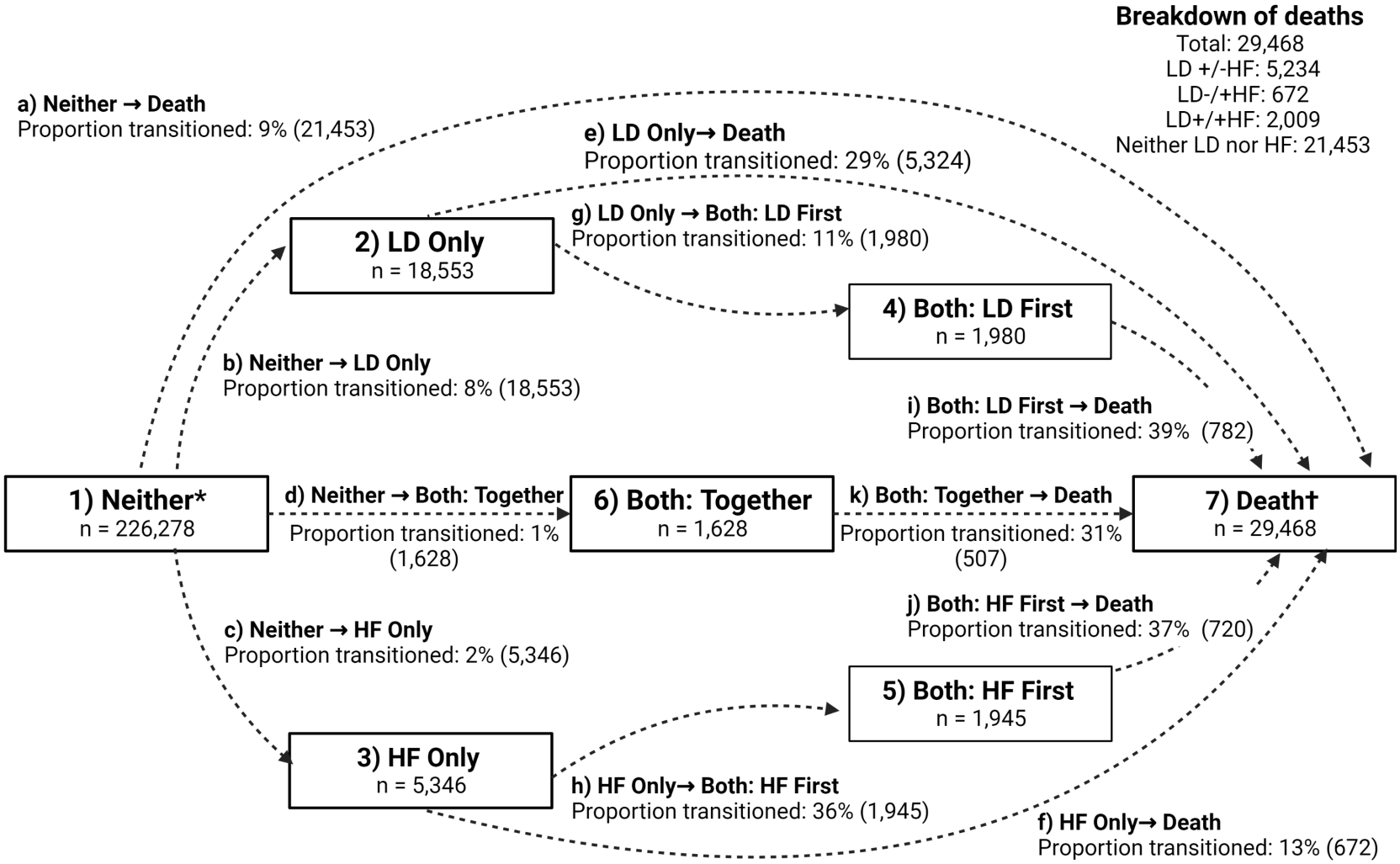
From the baseline inclusion state of neither (226,278 [98% of cohort]), those who entered the cohort in the absence of a diagnosis of HF or a repeat LD prescription were the least likely to experience a subsequent qualifying event, with only 25,527 (11%) receiving a diagnosis of HF or being initiated on LD during follow-up and 21,463 (9%) dying without experiencing either of these events (see Figure 5.5a). Where an event did occur, patients were more likely to be initiated on a repeat LD in the absence of a HF diagnosis than to be diagnosed with HF. From entering the LD Only state, patients were more likely to die without receiving a diagnosis of HF (29%) than to receive a diagnosis of HF (11%). In contrast, patients entering the HF Only state were more likely to be initiated on a repeat LD (36%) than to die prior to receiving a LD (13%). Once patients received both a diagnosis of HF and were initiated on LD, the prognosis was poor, especially when the two events were >30 days apart (39% for Both: LD First, 31% for Both: Together, and 37% for Both: HF First).

In the 'neither' group, the incidence of those entering the LD Only state was 18.1 cases per 1,000 patient-years at risk, the incidence of those entering the HF Only state was 5.1 cases per 1,000 patient-years at risk, and the incidence of those entering the Both: Together state was 1.5 cases per 1,000 patient-years at risk.

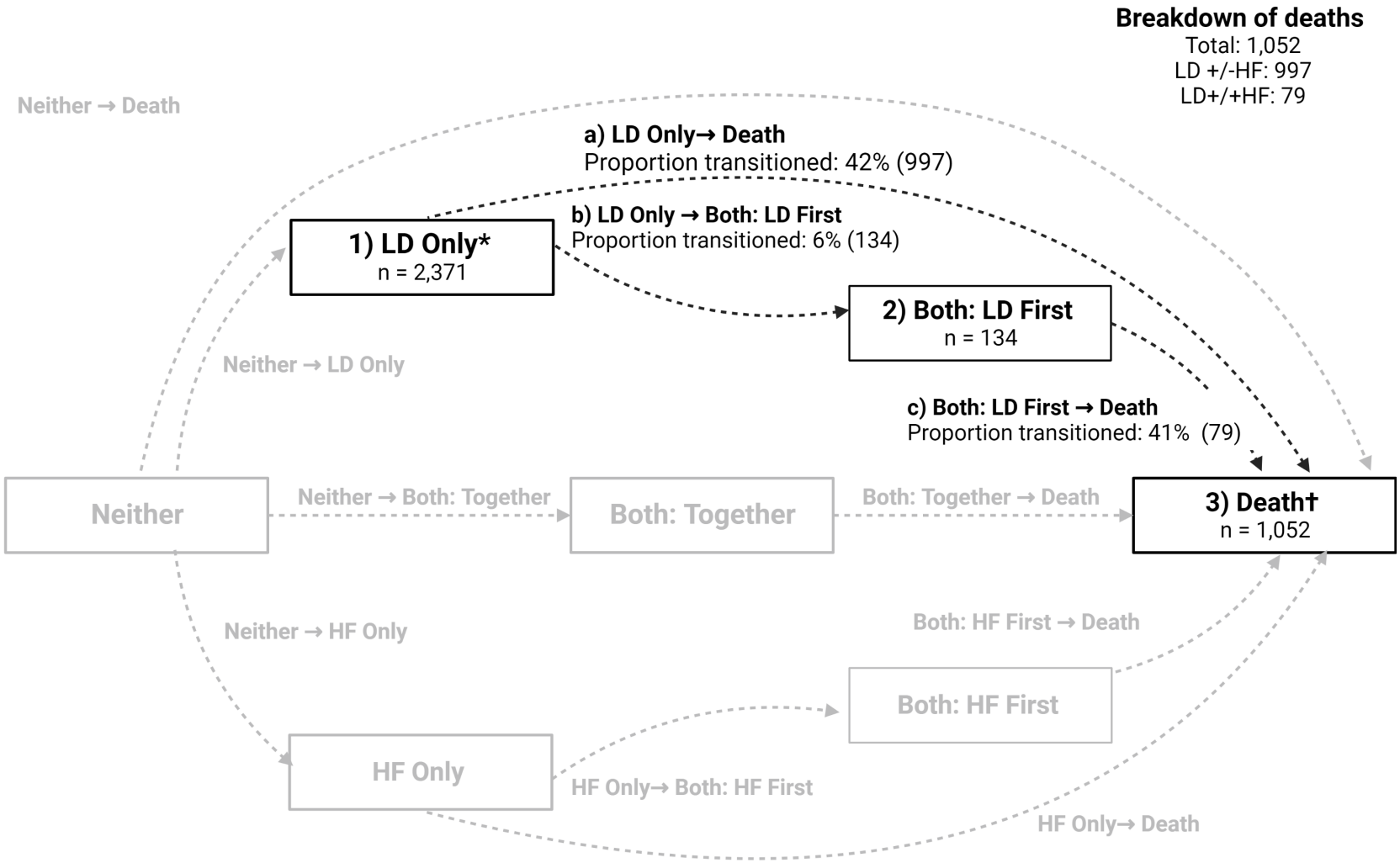
For those who entered the cohort due to LD dispensing in the absence of a HF diagnosis (2,371 [1% of cohort]), only 134 (6%) were subsequently diagnosed with HF (see Figure 5.5b), a lower percentage than for patients who entered the cohort through the Neither state. Patients were more likely to die without receiving a diagnosis of HF (997 [42%]) than to receive a diagnosis of HF. Of those diagnosed with HF, 55 patients (41%) died. Patients who started LD after cohort inclusion were also more likely to die than to be diagnosed with HF (see Figures 5.5a and 5.5b).

For those who entered the cohort due to a diagnosis of HF (864 [<1% of the cohort]) in the absence of repeat LD, 402 patients (47%) were subsequently initiated on LD, while 145 patients (17%) died without being initiated on LD and 162 (40%) of those initiated on LD died (see Figure 5.5c). Patients diagnosed with HF subsequent to inclusion were also more likely to be initiated on repeat LD than to die without first receiving a LD (see Figures 5.5a and 5.5c).

Only 307 (<1% of the cohort) patients entered the cohort with a diagnosis of HF and initiation of LD occurring together. Across all four models, this group of patients had a lower mortality rate (see Figure 5.5d) before adjusting for age, sex, or SIMD.

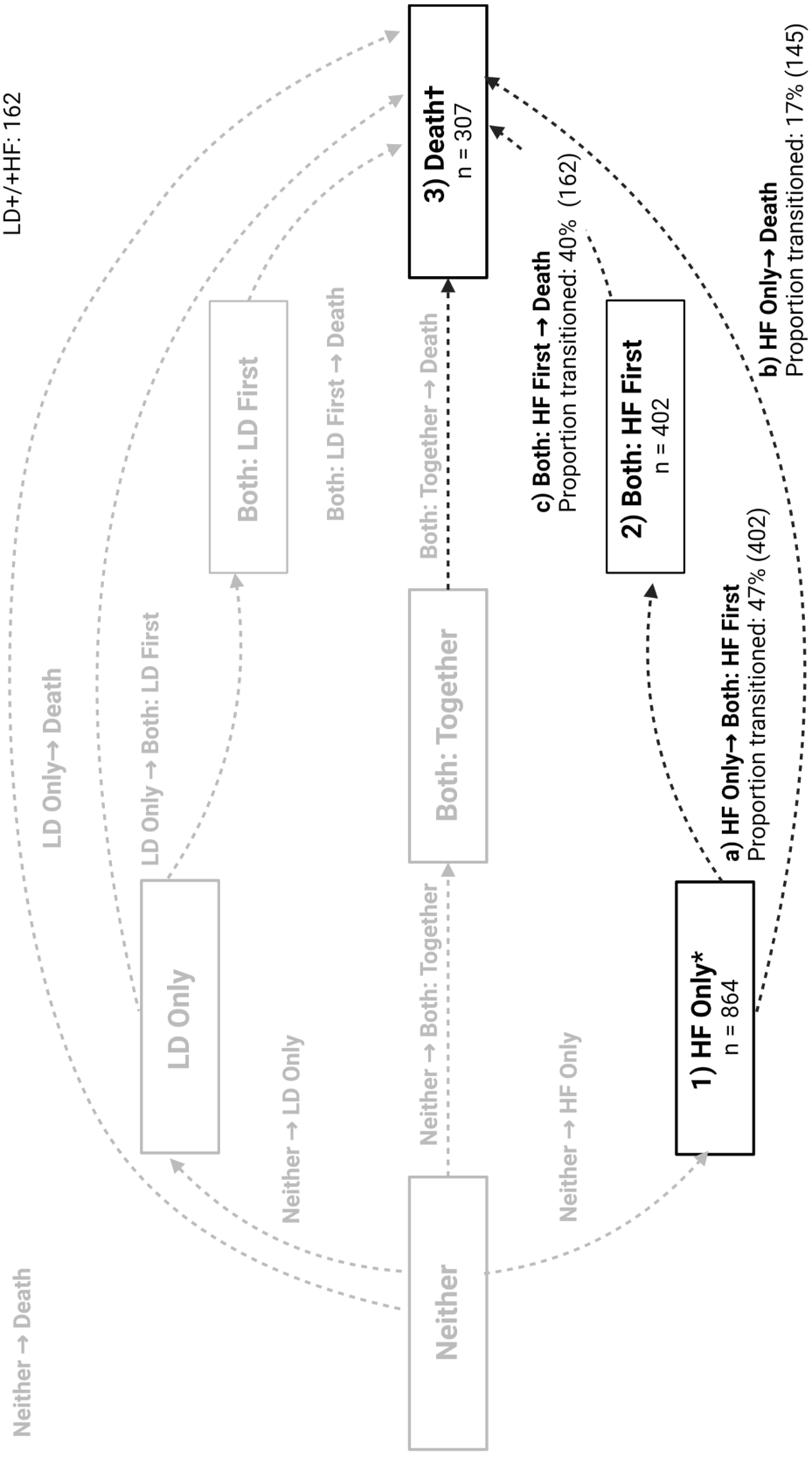


(a) Multi-state diagram showing allowed transitions for patients who enter the cohort as neither having a diagnosis of HF nor being treated with repeat LD. Patients entered the cohort between 1st January 2012 and 31st March 2017, inclusive. Information in black indicates the subset of the main multi-state diagram accessible to those who enter neither. Alphabetical labels indicate allowed transitions.
∗ Inclusion state; † the 'absorbing' terminal state.



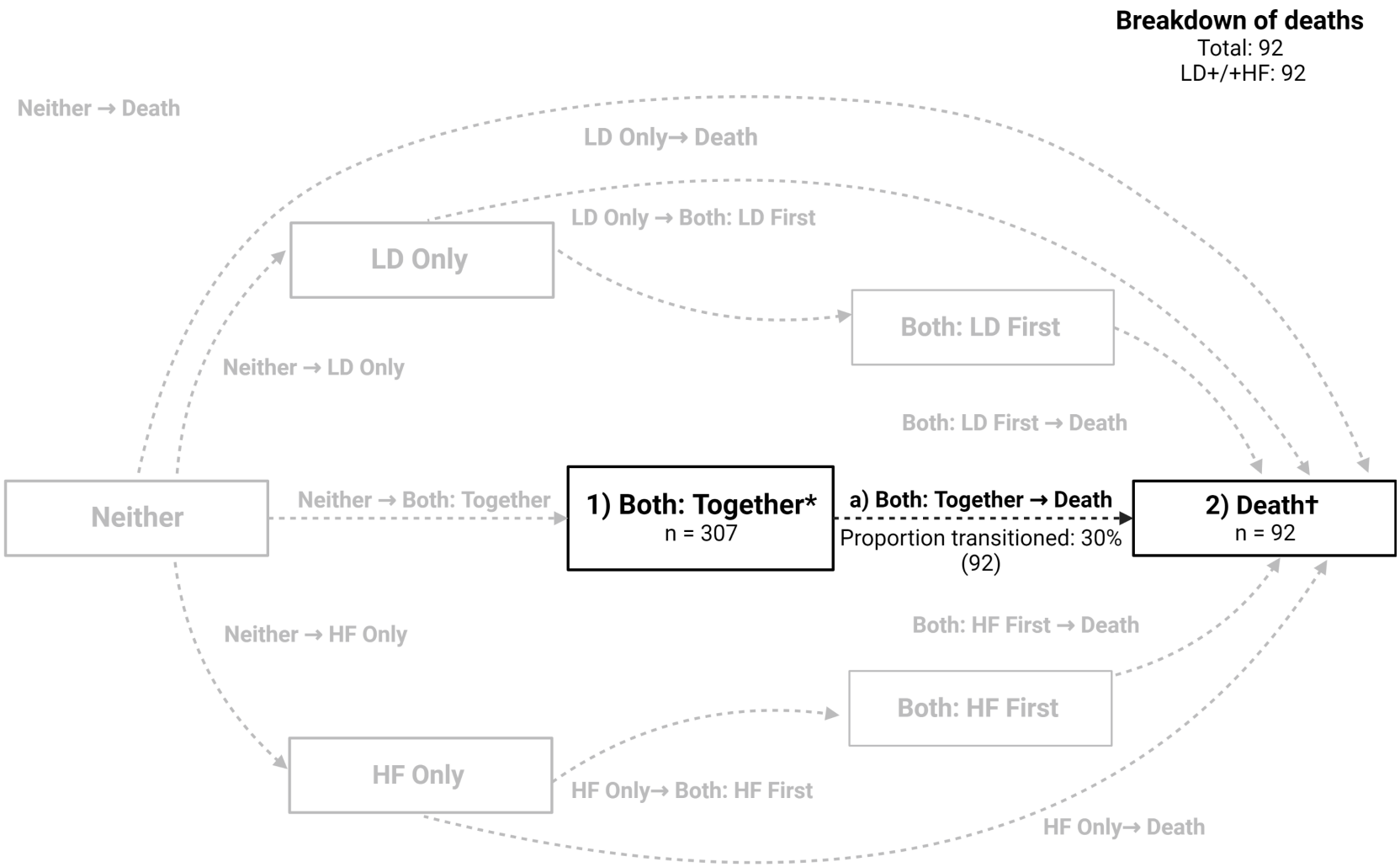
(b) Multi-state diagram showing allowed transitions for patients who entered the cohort as LD only at any time between 1st January 2012 and 31st March 2017, inclusive. Information in black indicates the subset of the main multi-state diagram accessible to those who enter LD only. Alphabetical labels indicate allowed transitions.

∗ Inclusion state; † the 'absorbing' terminal state.



(c) Multi-state diagram showing allowed transitions for patients who entered the cohort as HF only at any time between 1st January 2012 and 31st March 2017, inclusive. Information in black indicates the subset of the main multi-state diagram accessible to those who enter HF only. Alphabetical labels indicate allowed transitions.

∗ Inclusion state; † the 'absorbing' terminal state.



(d) Multi-state diagram showing allowed transitions for patients who entered the cohort as Both: Together at any time between 1st January 2012 and 31st March 2017, inclusive. Information in black indicates the subset of the main multi-state diagram accessible to those who enter Both: Together. Alphabetical labels indicate allowed transitions.

∗ Inclusion state; † the 'absorbing' terminal state.

Patient Characteristics

Regardless of the initial reason for inclusion, the final transition state before the end of follow-up or death (referred to as the foregoing state) was LD only for 18,810 patients, HF only for 2,347 patients, and both HF and LD for 6,396 patients, with similar numbers in the latter group initiating LD first, having a diagnosis of HF first, or both events occurring together.

Patients in the foregoing LD Only state were more likely to be women (11,721 [62%]) (see Table 5.5), and where measurements were available, were likely to have an LVEF $\geq$50% (4,144 [89%]) (see Figure 5.6a). They were less likely to have a history of MI (1,943 [10%]) or valve disease (1,019 [5%]), but no more likely to have a history of COPD or cancer than patients with HF. In the 180 days leading up to entering the foregoing state, patients initiated on LD were somewhat less likely to receive ACEi/ARB (8,076 [43%]) or beta-blockers (6,481 [34%]).

Patients with HF, whether or not they received LD, were more likely to have CAD and a history of MI.

Patients in the foregoing HF Only state were younger with a median age of 69 years (1^{st} - 3^{rd} quartiles: 58 - 79) and more likely to be men (2,424 [63%]). In the 180 days leading up to entering the foregoing state, these patients were more likely to receive beta-blockers (1,543 [40%]) than patients who remained in the 'neither' group, regardless of age.

Patients with both HF and repeat LD in the foregoing state were more more likely to have AF/AFL (Both: LD First: 1,010 [48%]; Both: HF First: 971 [41%]; and Both: Together: 812 [42%]). In the 180 days leading up to entering the foregoing state, patients were somewhat more likely to receive ACEi/ARB (Both: LD First: 1,177 [56%]; Both: HF First: 1,629 [69%]; and Both: Together: 927 [48%]) than those without both events. These patients were more likely to receive a beta-blocker (Both: LD First: 1,106 [52%]; Both: HF First: 1,544 [66%]; and Both: Together: 759 [39%]) in the same time interval. These patients also had the worst renal function and lowest haemoglobin.

All groups of patients dispensed LD in the foregoing state were more likely to have an eGFR <60 mL/min/1.73m^2 and anaemia (see Table 5.7), and COPD.

Rates of cancer, dementia, and stroke were similar across the groups.

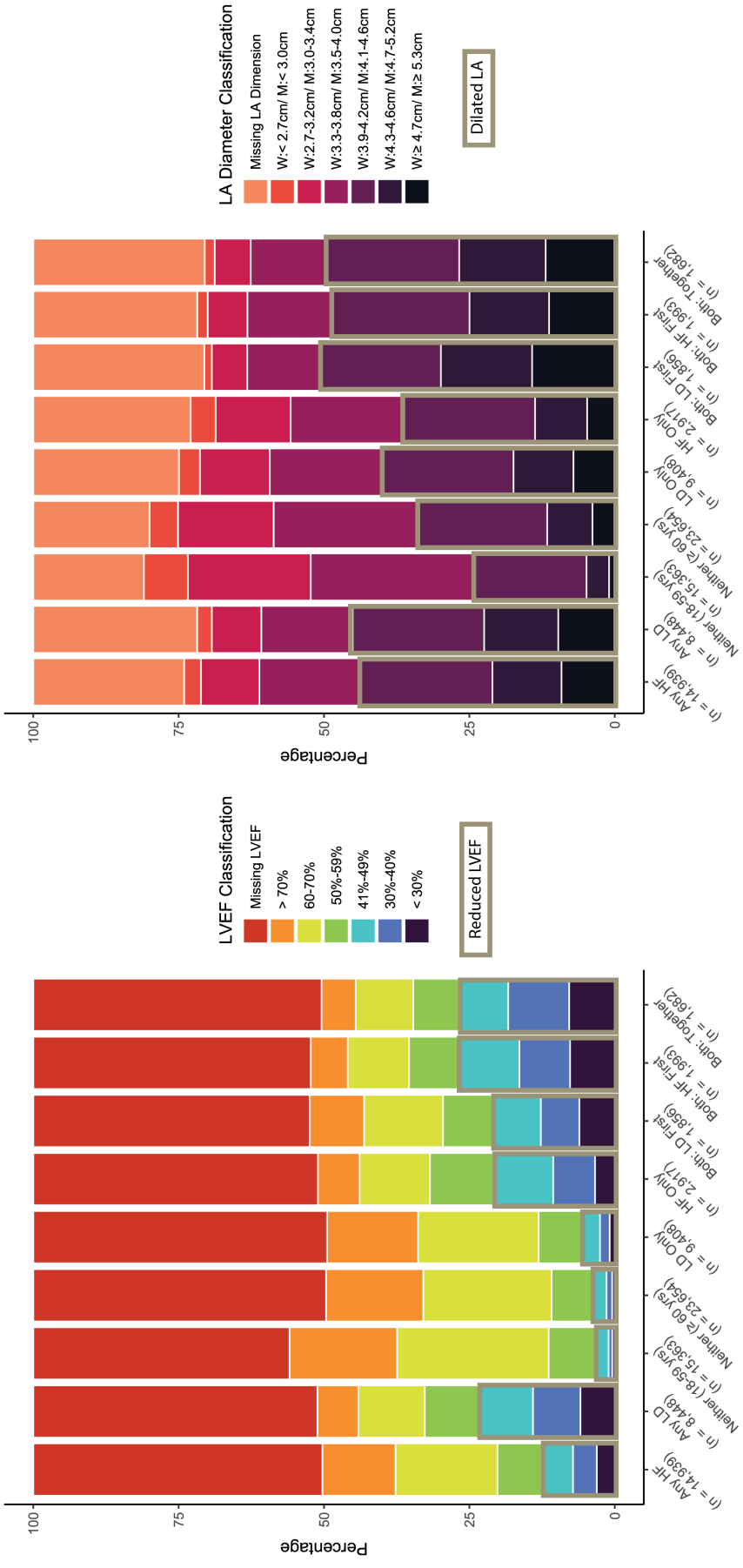
Additionally, where repeat LD was initiated alone or in advance of a HF diagnosis, patients had lower rates of contact with secondary care[4] before initiating repeat LD (LD only: 8,408 [45%]; Both: LD First: 1,023 [48%]) compared with patients who were diagnosed with HF before or with the initiation of repeat LD (Both: Together: 1,681 [87%]; Both: HF First: 1,444 [62%]). Only a small minority of all groups attended a CV[5] specialist in the 30 days prior to initiation of

[4]This includes any admission or clinic visit across SMR00, SMR01, or SMR04.

[5]Included specialties are cardiology, cardiac surgery, cardiothoracic surgery, and vascular surgery

repeat LD or a diagnosis of HF.

Patients diagnosed with HF were more likely to have an echocardiogram, and if so, a reduced LVEF (see Figure 5.6). Additional echocardiographic results are available in Figure D.4.



(a) Left ventricular ejection fraction (LVEF) (b) Let atrial (LA) diameter

Figure 5.6: Grey boxes indicate either reduced LVEF or dilated LA respectively.

Table 5.5: Patient demographics and comorbidities grouped according to entry into the last transition state prior to the end of observation or death (i.e., the foregoing state). Due to space limitations, information for patients in the neither (18-59 yrs) and neither (≥60 yrs) stats are displayed in Table D.2.

Variable	**Neither**	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**
n	87,940	18,810	3,863	2,114	2,347	1,935
Age (years)	71 (65 - 77)	74 (62 - 82)	69 (58 - 79)	79 (71 - 85)	77 (68 - 84)	75 (66 - 82)
Sex						
Women	47,371 (54%)	11,721 (62%)	1,439 (37%)	1,117 (53%)	1,123 (48%)	890 (46%)
Men	40,569 (46%)	7,089 (38%)	2,424 (63%)	997 (47%)	1,224 (52%)	1,045 (54%)
Ethnicity						
White	68,605 (78%)	16,430 (87%)	3,448 (89%)	1,979 (94%)	2,179 (93%)	1,755 (91%)
Missing	17,367 (20%)	2,013 (11%)	307 (8%)	89 (4%)	111 (5%)	115 (6%)
Other	1,968 (2%)	367 (2%)	108 (3%)	46 (2%)	57 (2%)	65 (3%)
Socioeconomic deprivation (SIMD)						
1 (most deprived)	30,292 (34%)	7,836 (42%)	1,562 (40%)	840 (40%)	1,016 (43%)	769 (40%)
2	15,595 (18%)	3,472 (18%)	704 (18%)	465 (22%)	422 (18%)	357 (18%)
3	11,633 (13%)	2,605 (14%)	533 (14%)	265 (13%)	323 (14%)	239 (12%)
4	11,919 (14%)	2,187 (12%)	438 (11%)	231 (11%)	253 (11%)	228 (12%)
5 (least deprived)	18,501 (21%)	2,710 (14%)	626 (16%)	313 (15%)	333 (14%)	342 (18%)
Healthcare contact						
HF diagnosed in PC	N/A	N/A	1,745 (45%)	590 (28%)	678 (29%)	525 (27%)
LD started within 30 days post contact						
Any secondary care	N/A	8,408 (45%)	N/A	1,023 (48%)	1,444 (62%)	1,681 (87%)
Hospital discharge ☒	N/A	4,809 (26%)	N/A	550 (26%)	1,032 (44%)	1,507 (78%)
Specialist clinic	N/A	4,973 (26%)	N/A	613 (29%)	661 (28%)	483 (25%)

Continuation of Table 5.5

Variable	Neither	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together
n	87,940	18,810	3,863	2,114	2,347	1,935
CV specialist visit ∮	N/A	796 (4%)	N/A	168 (8%)	314 (13%)	203 (10%)
Comorbidities *						
H/o hypertension	31,389 (36%)	7,604 (40%)	1,739 (45%)	1,286 (61%)	1,288 (55%)	1,028 (53%)
DM	20,596 (23%)	4,722 (25%)	892 (23%)	750 (35%)	698 (30%)	616 (32%)
Thyroid disease	1,997 (2%)	797 (4%)	151 (4%)	136 (6%)	151 (6%)	83 (4%)
CAD	21,694 (25%)	4,885 (26%)	2,586 (67%)	1,080 (51%)	1,401 (60%)	939 (49%)
Of which is MI	7,766 (9%)	1,943 (10%)	1,983 (51%)	588 (28%)	977 (42%)	592 (31%)
Valve disease	1,203 (1%)	1,019 (5%)	516 (13%)	554 (26%)	570 (24%)	454 (23%)
AF/AFL	5,881 (7%)	2,996 (16%)	881 (23%)	1,010 (48%)	971 (41%)	812 (42%)
PAD	2,557 (3%)	737 (4%)	196 (5%)	167 (8%)	236 (10%)	148 (8%)
Stroke	7,270 (8%)	2,240 (12%)	475 (12%)	409 (19%)	439 (19%)	247 (13%)
COPD	9,068 (10%)	4,240 (23%)	806 (21%)	674 (32%)	713 (30%)	502 (26%)
Cancer	7,787 (9%)	2,907 (15%)	475 (12%)	310 (15%)	295 (13%)	239 (12%)
Dementia	1,953 (2%)	1,008 (5%)	134 (3%)	118 (6%)	168 (7%)	72 (4%)
EP Devices	885 (1%)	398 (2%)	143 (4%)	132 (6%)	173 (7%)	91 (5%)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

☒ SMR01 hospital discharge includes day-cases and inpatient stays.

∮ CV specialties include cardiology, cardiac surgery, cardiothoracic surgery, and vascular surgery.

∗ History of a coded record on or before inclusion date.

PC, primary care; EP device, cardiac electrophysiology device (e.g., pacemaker, ICD, and CRT).

Table 5.6: Medicines dispensed in the 180 days prior to the last transition state before the end of observation or death (i.e., the foregoing state) with patients grouped according to the foregoing state. Due to space limitations, information for patients classified as neither (18-59 yrs) or neither (≥60 yrs) is displayed in Table D.3.

Variable	**Neither**	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**
n	87,940	18,810	3,863	2,114	2,347	1,935
Age (years)	71 (65 - 77)	74 (62 - 82)	69 (58 - 79)	79 (71 - 85)	77 (68 - 84)	75 (66 - 82)
ACEi or ARB	49,520 (56%)	8,076 (43%)	1,809 (47%)	1,177 (56%)	1,629 (69%)	927 (48%)
ACEi	38,189 (43%)	5,937 (32%)	1,473 (38%)	877 (41%)	1,364 (58%)	722 (37%)
ARB	12,233 (14%)	2,400 (13%)	382 (10%)	331 (16%)	336 (14%)	227 (12%)
Beta-blocker	33,436 (38%)	6,481 (34%)	1,543 (40%)	1,106 (52%)	1,544 (66%)	759 (39%)
MRA	452 (1%)	753 (4%)	75 (2%)	155 (7%)	324 (14%)	36 (2%)
CCB	29,000 (33%)	6,304 (34%)	914 (24%)	683 (32%)	767 (33%)	673 (35%)
Diltiazem/Verapamil	3,167 (4%)	1,027 (5%)	149 (4%)	123 (6%)	123 (5%)	103 (5%)
Dihydropyridine	25,965 (30%)	5,361 (29%)	776 (20%)	570 (27%)	653 (28%)	576 (30%)
Digoxin	1,078 (1%)	828 (4%)	125 (3%)	228 (11%)	328 (14%)	83 (4%)
Thiazides+	25,383 (29%)	4,357 (23%)	445 (12%)	216 (10%)	374 (16%)	364 (19%)
Low dose aspirin	34,727 (39%)	6,318 (34%)	1,568 (41%)	818 (39%)	1,235 (53%)	745 (39%)
Oral anticoagulants	3,318 (4%)	2,341 (12%)	396 (10%)	640 (30%)	633 (27%)	314 (16%)
Lipid regulators	51,151 (58%)	9,692 (52%)	2,069 (54%)	1,337 (63%)	1,610 (69%)	1,073 (55%)
Bronchodilators	11,444 (13%)	5,103 (27%)	758 (20%)	697 (33%)	702 (30%)	518 (27%)
Thyroid medications	8,050 (9%)	2,069 (11%)	286 (7%)	235 (11%)	240 (10%)	173 (9%)

Continuation of Table 5.6

Variable	Neither	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together
n	87,940	18,810	3,863	2,114	2,347	1,935
Hypoglycaemic agents	12,233 (14%)	2,958 (16%)	532 (14%)	488 (23%)	449 (19%)	400 (21%)
Insulin†	1,699 (2%)	781 (4%)	120 (3%)	133 (6%)	118 (5%)	102 (5%)
Other hypoglycaemic agents	11,436 (13%)	2,608 (14%)	469 (12%)	423 (20%)	401 (17%)	360 (19%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

† Either alone or in combination with another agent.

Dilt/Verap, Diltiazem/Verapamil; Thiazides+, thiazides and related.

Table 5.7: Most recent blood test results in the two years prior to entering the last transition state before the end of observation or death (i.e., the foregoing state) with patients grouped according to the foregoing state. Due to space limitations, information for patients classified as neither (18-59 yrs) or neither (≥60 yrs) is displayed in Table D.4.

Variable	Neither	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together
n	87,940	18,810	3,863	2,114	2,347	1,935
Age (years)	71 (65 - 77)	74 (62 - 82)	69 (58 - 79)	79 (71 - 85)	77 (68 - 84)	75 (66 - 82)
Haemoglobin (Hb) (mg/dL)						
Reported	69,800 (79%)	17,673 (94%)	3,630 (94%)	2,094 (99%)	2,317 (99%)	1,901 (98%)
Women: Hb	13.1 (12.1 - 13.9)	12.5 (11.3 - 13.6)	12.9 (11.6 - 13.9)	12.1 (10.9 - 13.3)	12.0 (10.9 - 13.3)	12.4 (11.0 - 13.7)
Men: Hb	14.2 (13.2 - 15.2)	13.1 (11.5 - 14.5)	14.3 (12.9 - 15.4)	12.9 (11.3 - 14.3)	13.0 (11.4 - 14.5)	13.5 (12.0 - 14.8)
Anaemia◇	14,821 (21%)	7,288 (41%)	1,001 (28%)	984 (47%)	1,123 (48%)	784 (41%)

Continuation of Table 5.7

Variable	Neither	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together
n	87,940	18,810	3,863	2,114	2,347	1,935
Estimated glomerular filtration rate (eGFR)∇ (mL/min/1.73m^2)						
Reported	81,903 (93%)	18,190 (97%)	3,736 (97%)	2,109 (100%)	2,338 (100%)	1,913 (99%)
eGFR	79 (66 - 89)	77 (60 - 89)	80 (62 - 92)	60 (42 - 77)	68 (50 - 83)	70 (52 - 85)
eGFR [30 - 60)	12,820 (16%)	3,667 (20%)	669 (18%)	778 (37%)	730 (31%)	568 (30%)
eGFR <30	899 (1%)	686 (4%)	193 (5%)	255 (12%)	130 (6%)	95 (5%)
Serum values (mmol/L)						
Reported urea	81,927 (93%)	18,191 (97%)	3,736 (97%)	2,109 (100%)	2,338 (100%)	1,914 (99%)
Urea	5.8 (4.8 - 7.1)	6.0 (4.6 - 7.8)	5.8 (4.6 - 7.6)	7.7 (5.8 - 11.0)	7.1 (5.4 - 9.7)	6.8 (5.2 - 9.0)
Reported Albumin	79,860 (91%)	17,986 (96%)	3,686 (95%)	2,102 (99%)	2,329 (99%)	1,905 (98%)
Albumin	38 (36 - 39)	35 (31 - 38)	36 (33 - 39)	34 (31 - 37)	33 (30 - 36)	35 (32 - 37)
Albumin [30 - 35)	10,526 (13%)	4,904 (27%)	869 (24%)	726 (35%)	856 (37%)	679 (36%)
Albumin <30	2,965 (4%)	3,510 (20%)	423 (11%)	367 (17%)	540 (23%)	235 (12%)
Reported sodium	81,913 (93%)	18,195 (97%)	3,737 (97%)	2,109 (100%)	2,338 (100%)	1,914 (99%)
Sodium	139 (137 - 140)	139 (136 - 141)	138 (136 - 140)	138 (136 - 141)	139 (136 - 140)	138 (135 - 140)
Sodium <135	6,915 (8%)	2,392 (13%)	518 (14%)	368 (17%)	314 (13%)	371 (19%)
Reported chloride	81,900 (93%)	18,191 (97%)	3,735 (97%)	2,109 (100%)	2,338 (100%)	1,914 (99%)
Chloride	104 (102 - 106)	103 (100 - 106)	104 (102 - 106)	102 (99 - 105)	102 (99 - 105)	104 (100 - 106)
Reported Potassium	81,535 (93%)	18,152 (97%)	3,716 (96%)	2,106 (100%)	2,337 (100%)	1,909 (99%)
Potassium	4.3 (4.0 - 4.5)	4.3 (4.0 - 4.6)	4.3 (4.0 - 4.6)	4.2 (3.9 - 4.6)	4.2 (3.9 - 4.6)	4.3 (4.0 - 4.6)
Potassium >6.0	32 (<1%)	25 (<1%)	15 (<1%)	19 (1%)	<6 (<1%)	<6 (<1%)

Continuation of Table 5.7

Variable	Neither	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together
n	87,940	18,810	3,863	2,114	2,347	1,935
Potassium <3.5	1,915 (2%)	711 (4%)	135 (4%)	145 (7%)	110 (5%)	90 (5%)
Reported bicarbonate	43,115 (49%)	6,751 (36%)	1,483 (38%)	890 (42%)	1,072 (46%)	705 (36%)
Bicarbonate	24 (22 - 26)	23 (21 - 25)	22 (20 - 25)	23 (21 - 26)	23 (20 - 25)	22 (20 - 25)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

◇ Using the WHO definition of anaemia (see Section 2.6.2).

∇ Calculated eGFR assuming no pregnancies and without adjusting for ethnicity (see Section 2.5.9).

Table 5.8: Data taken from the closest 12-lead ECG to entering the last transition state before the end of observation or death (i.e., the foregoing state) with patients grouped according to the foregoing state. Due to space limitations, information for patients classified as neither (18-59 yrs) or neither (≥60 yrs) is displayed in Table D.5.

Variable	Neither	LD Only	HF Only	Both: LD First	Both: HF First	Both: Together
n	87,940	18,810	3,863	2,114	2,347	1,935
Age (years)	71 (65 - 77)	74 (62 - 82)	69 (58 - 79)	79 (71 - 85)	77 (68 - 84)	75 (66 - 82)
ECG available	48,795 (55%)	13,873 (74%)	3,124 (81%)	1,938 (92%)	2,070 (88%)	1,719 (89%)
Rhythm						
Sinus	43,147 (88%)	11,138 (80%)	2,508 (80%)	1,071 (55%)	1,348 (65%)	1,032 (60%)
Heart rate (bpm)	72 (62 - 84)	78 (67 - 91)	73 (62 - 86)	79 (67 - 92)	76 (64 - 90)	81 (68 - 97)
AF/AFL/SVT	4,635 (9%)	2,315 (17%)	485 (16%)	710 (37%)	588 (28%)	586 (34%)
Heart rate (bpm)	88 (73 - 109)	87 (71 - 107)	94 (77 - 118)	86 (72 - 102)	87 (73 - 105)	96 (78 - 126)
Other	759 (2%)	310 (2%)	94 (3%)	110 (6%)	104 (5%)	63 (4%)
Heart rate (bpm)	68 (60 - 80)	70 (60 - 83)	66 (60 - 79)	66 (60 - 77)	69 (60 - 79)	69 (60 - 84)

Continuation of Table 5.8

Variable	**Neither**	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**
n	87,940	18,810	3,863	2,114	2,347	1,935
Undetermined	254 (<1%)	110 (<1%)	37 (1%)	47 (2%)	30 (1%)	38 (2%)
Heart rate (bpm)	88 (71 - 108)	88 (72 - 109)	88 (72 - 108)	84 (72 - 99)	100 (71 - 119)	82 (69 - 112)
QRS duration	86 (80 - 96)	86 (80 - 98)	92 (84 - 106)	96 (84 - 116)	98 (84 - 118)	96 (84 - 116)
QRS duration $\geq$120 ms	3,574 (7%)	1,235 (9%)	513 (16%)	447 (23%)	500 (24%)	393 (23%)
QTc available ■	46,680 (53%)	13,386 (71%)	2,984 (77%)	1,887 (89%)	1,974 (84%)	1,660 (86%)
QTc (ms)	419 (404 - 435)	423 (407 - 442)	428 (408 - 451)	437 (415 - 463)	436 (413 - 464)	433 (411 - 459)
Prolonged QTc	4,246 (9%)	1,876 (14%)	678 (23%)	591 (31%)	624 (32%)	473 (28%)
ST-T abnormality	11,134 (23%)	4,060 (29%)	1,198 (38%)	874 (45%)	952 (46%)	826 (48%)
Acute MI△	258 (2%)	67 (2%)	116 (10%)	22 (3%)	36 (4%)	36 (4%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

■ QTc could not be calculated in patients whose QRS complex or RR interval was suppressed.

△ ECG detected acute MI.

Hospital Admissions and Mortality

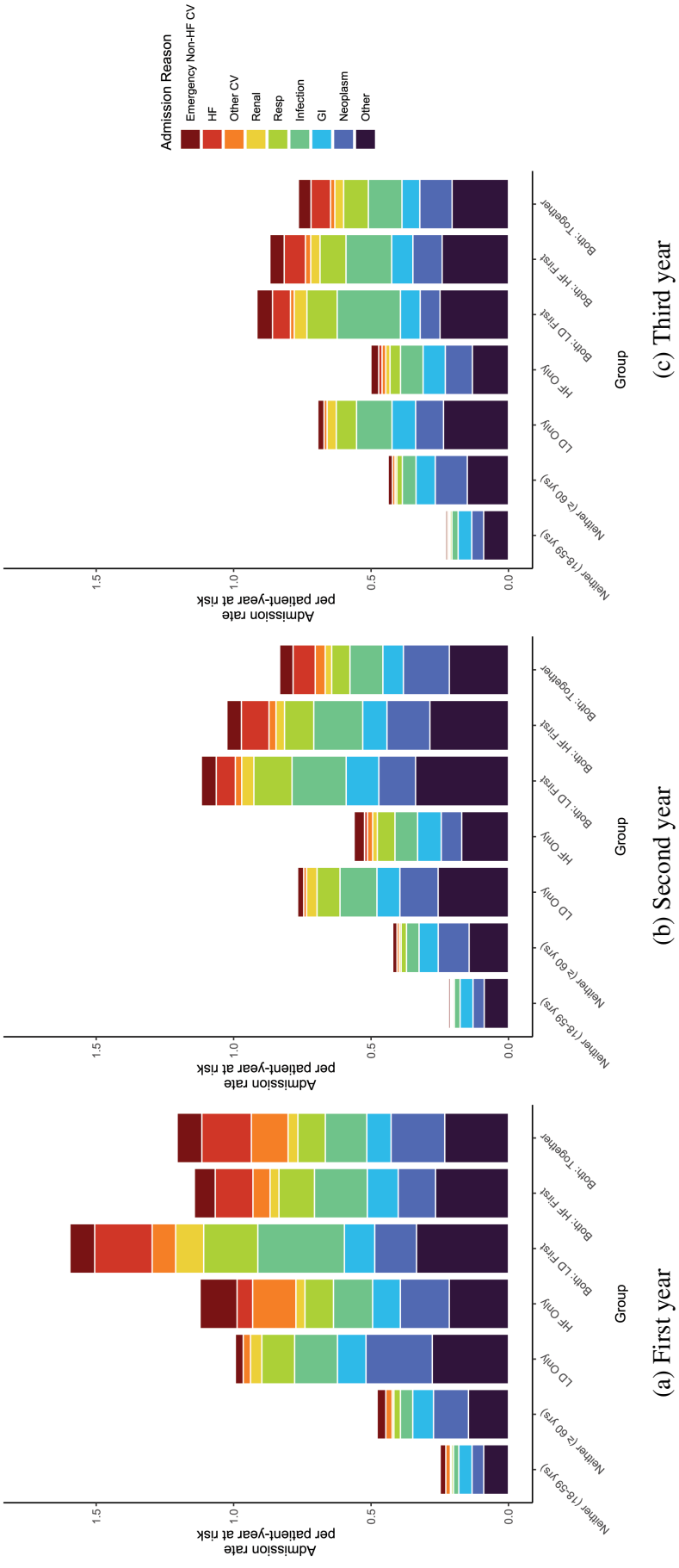
Patients entering the LD Only and HF Only states had similarly increased rates of hospitalisation compared with the neither groups, which were higher still in patients with both HF and repeat LD (especially when LD were initiated before a diagnosis of HF) (see Figure 5.7). Rates of hospitalisation were stable and low across the first, second, and third years of follow-up for patients aged $\geq$60 years who remained with neither a diagnosis of HF nor a repeat LD. For patients with a diagnosis of HF only, rates of hospitalisation were only slightly higher than the neither ($\geq$60 yrs) group in years two and three after the HF diagnosis. For all four groups with repeat LD, rates of hospital admission (adjusted for patient-time at risk of admission allowing for time spent in hospital and mortality, lost to follow-up, or transitioned to another state) fell in years two and three but remained markedly higher compared with the 'neither' groups.

For the 'neither' groups, cancer, gastrointestinal (GI) problems, infection, and other miscellaneous reasons for admission predominated in all three years; few admissions were for CV reasons. For the LD only group in year one, the pattern was similar to the neither ($\geq$60 yrs) group apart from a generally higher rate of admissions and a higher proportion of admissions for respiratory and renal problems. By definition, this group could not be admitted for HF. For all four groups of patients with HF, CV admissions made a large contribution to admissions in the first year, which diminished in years two and three.

For all groups with repeat LD, admissions in years two and three were predominantly for either respiratory problems or infections, disregarding miscellaneous other reasons (see Table D.7 for list of 6 most common codes). In groups with both a diagnosis of HF and repeat HF, only a small proportion of admissions was ascribed to HF and relatively few to CV reasons. Admissions patterns in years two and three for the HF only group were rather similar to the neither ($\geq$60 yrs) group, with few admissions coded as CV.

For patients in states defined by new-onset HF, admissions for HF accounted for most of the higher rates of hospitalisation in the first year, whereas rates from infection and neoplasm remained relatively stable throughout.

One, two, and three-year mortality per patient-year at risk was high for all four groups of patients initiated on a repeat LD, compared with either of the neither groups (see tables 5.9 and D.8), particularly when repeat LD preceded a diagnosis of HF. Mortality was higher in the first year. Deaths attributed to neoplasms were highest in the year after entering LD only. Otherwise, mortality rates attributed to neoplasms and infections remained markedly stable across all three years. Patients with only a diagnosis of HF had a similar one-year mortality rate to those who were diagnosed with HF and initiated on repeat LD together, but the mortality of the HF only group fell to rates similar to that of patients aged $\geq$60 years who neither had HF nor received repeat LD.



(a) First year (b) Second year (c) Third year

Figure 5.7: Hospital admission rate per patient-year at risk by group and cause of admission.

Table 5.9: 1-, 2-, and 3-year crude mortality by cause of death per patient-year at risk based on time since entering each state using dates of HF diagnosis and LD initiation to transition states.

Variable	**Neither**	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**
n	226,278	20,924	6,210	2,114	2,347	1,935
Died in the 1st year	4,609 (2%)	3,247 (16%)	447 (7%)	457 (22%)	412 (18%)	271 (14%)
CV	1,197 (<0.01)	613 (0.01)	168 (0.01)	198 (0.05)	169 (0.03)	132 (0.03)
Neoplasm	1,616 (<0.01)	1,278 (0.03)	98 (0.01)	60 (0.02)	64 (0.01)	43 (0.01)
Infection	291 (<0.01)	191 (<0.01)	46 (<0.01)	51 (0.01)	35 (0.01)	17 (<0.01)
Other	1,463 (<0.01)	1,149 (0.02)	135 (0.01)	147 (0.04)	142 (0.03)	78 (0.02)
Unknown	42 (<0.01)	16 (<0.01)	0 (0.00)	<6 (<0.01)	<6 (<0.01)	<6 (<0.01)
In 2nd year†	209,414 (93%)	14,881 (71%)	3,357 (54%)	1,306 (62%)	1,691 (72%)	1,474 (76%)
Died in the 2nd year	3,735 (<0.01)	1,307 (0.04)	138 (0.02)	200 (0.09)	196 (0.06)	117 (0.04)
CV	1,023 (<0.01)	309 (0.01)	42 (0.01)	73 (0.03)	85 (0.02)	62 (0.02)
Neoplasm	1,316 (<0.01)	316 (0.01)	30 (<0.01)	26 (0.01)	23 (0.01)	20 (0.01)
Infection	245 (<0.01)	101 (<0.01)	11 (<0.01)	20 (0.01)	19 (0.01)	9 (<0.01)
Other	1,123 (<0.01)	577 (0.02)	54 (0.01)	80 (0.04)	68 (0.02)	26 (0.01)
Unknown	28 (<0.01)	<6 (<0.01)	<6 (<0.01)	<6 (<0.01)	<6 (<0.01)	0 (0.00)
In 3rd year†	185,339 (82%)	10,490 (50%)	2,516 (41%)	783 (37%)	1,152 (49%)	1,050 (54%)
Died in the 3rd year	3,485 (0.01)	868 (0.04)	89 (0.02)	108 (0.10)	130 (0.07)	91 (0.05)
CV	913 (<0.01)	203 (0.01)	29 (0.01)	48 (0.04)	57 (0.03)	37 (0.02)
Neoplasm	1,148 (<0.01)	174 (0.01)	23 (<0.01)	15 (0.01)	20 (0.01)	17 (0.01)
Infection	184 (<0.01)	81 (<0.01)	<6 (<0.01)	<6 (<0.01)	13 (0.01)	10 (0.01)
Other	1,204 (<0.01)	407 (0.02)	32 (0.01)	38 (0.03)	40 (0.02)	27 (0.01)

Continuation of Table 5.9						
Variable	**Neither**	**LD Only**	**HF Only**	**Both: LD First**	**Both: HF First**	**Both: Together**
n	226,278	20,924	6,210	2,114	2,347	1,935
Unknown	36 (<0.01)	<6 (<0.01)	<6 (<0.01)	<6 (<0.01)	0 (0.00)	0 (0.00)

Data are rates (patient-year at risk) for mortality values or number (%) for those who entered the state and remained within the said state at a given year after entering.

† Alive and in specified state at the year after entering said state.

Assumptions Testing

The assumptions of linearity and PH were tested for per decade, sex, and SIMD for the four multi-state models, respectively. Linearity for age per decade was broadly met for the models where patients started in either the Neither, LD Only, or HF Only states (see Figure D.3). It was not met for the model where patients started in the Both: Together state (see Figure D.3d). This is the smallest group of patients and thus will be subject to more 'noise'. For parsimony, linearity was assumed to use only one degree of freedom, allowing an easier comparison against the other models. Sex met the PH assumption across the four models, and SIMD also showed PH. Inclusion year was assumed to be linear as it was a stand-in for unmeasured covariates and changes in clinical practice.

Models and Predictions from Cohort Inclusion

Across the four models, the hazard of transitioning increased with age and levels of socioeconomic deprivation, though many of the per-transition adjustments were not statistically significant (see Table D.9). With some exceptions, men were more likely to transition, but few results were statistically significant. Men who started in the HF only group were more likely to die without receiving a LD (HR [95% CI]: 1.1 [0.79 - 1.6]).

Figure 5.8 shows, for the two risk profiles presented in Table 5.1, predictions of transitioning through a particular sequence of states (a path) based on the patient's characteristics and initial state.

When patients started in the Neither state, most patients did not transition further.

The probabilities of patients transitioning based on being a patient at high risk of starting LD therapy (see Table 5.1 for profile) are shown in figures 5.8a, D.5a, and D.5b. Initially, the most likely path was to be initiated on repeat LD. Although, within half a year, the most likely path was ACM without a diagnosis of HF or the initiation of repeat LD. Where patients only experienced one of the two events, patients were more likely to be initiated on repeat LD than diagnosed with HF, and patients initiated on LD were more likely to die than those diagnosed with HF. When both events occurred together, survival was better than when these events occurred sequentially.

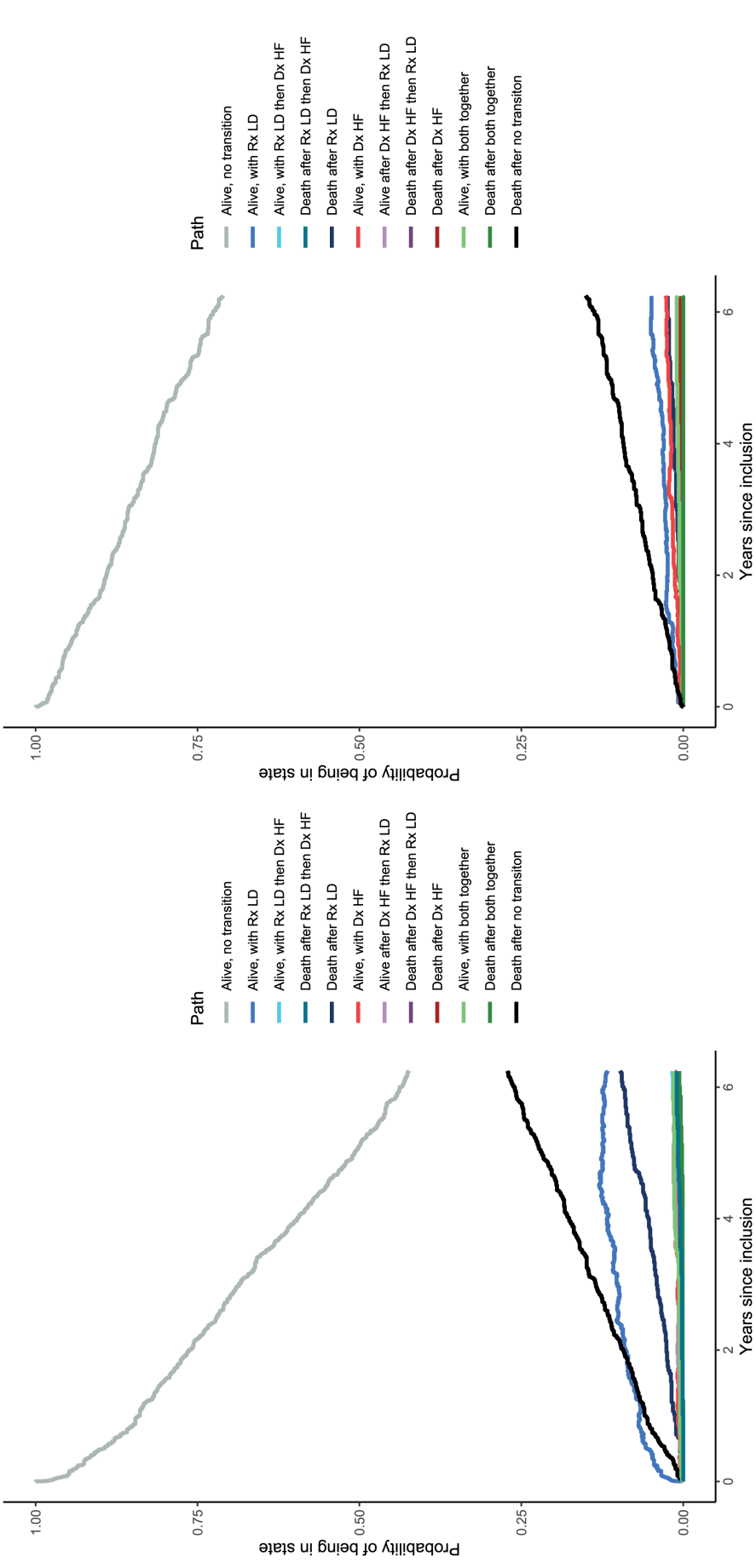
The probabilities of patients transitioning based on being at high risk of being diagnosed with HF (see Table 5.1 for profile) are illustrated in figures 5.8b, D.5e, and D.5f. While the probability of a transition is still low, the most frequent transition is dying without being initiated on repeat LD or being diagnosed with HF. After 6 years from follow-up from cohort inclusion, there is a similar probability of being alive with only a diagnosis of HF and being dead having only been started on repeat LD without a diagnosis of HF.

The probability of receiving both a diagnosis of HF and being initiated on a repeat LD was low. When both events occurred together, survival was better than when these events occurred sequentially.

Using the same two risk profiles in patients who start off in the LD Only state, the probability of transitioning exceeded 50% after 3-4 years follow-up (see Figures 5.8c and 5.8d), with the most likely outcome in both cases being death without receiving a diagnosis of HF. Based on the 80-year-old woman's profile provided (see Figure 5.8c), there is little difference in survival at 6 years whether or not the patient receives a diagnosis of HF.

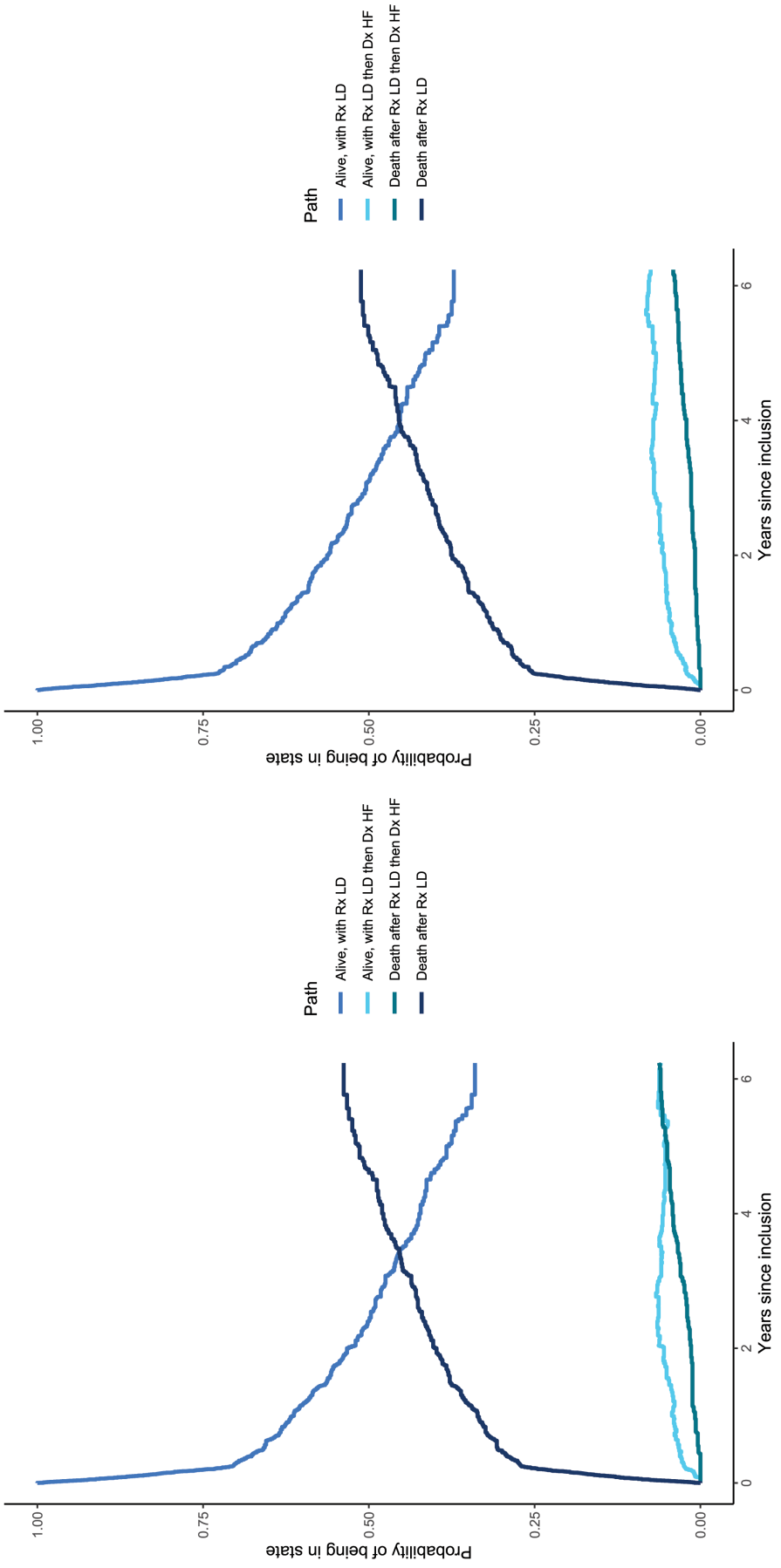
Using the same two risk profiles for patients who start in the HF Only state, the probability of remaining alive without receiving repeat LD drops below 50% by the first year (see Figures 5.8e and 5.8f). In both risk profiles, patients are more likely to start repeat LD than they are to die. Mortality after being initiated on repeat LD is high.

Using the same two risk profiles for patients who start in the Both: Together state, the probability of remaining alive drops below 50% at 4 years post-inclusion based on the 80-year-old woman's profile (see Figure 5.8g), and just before 6 years for the 65-year-old man's profile (see Figure 5.8h). Unfortunately, these paths do not have 95% CI, so it is impossible to tell how much of this result is due to the small sample size, lower event rate, and the non-linear effect of age. The lack of a 95% CI is also a limitation for other models and predictions, but they comprise many more patients, so predictions should be more accurate.



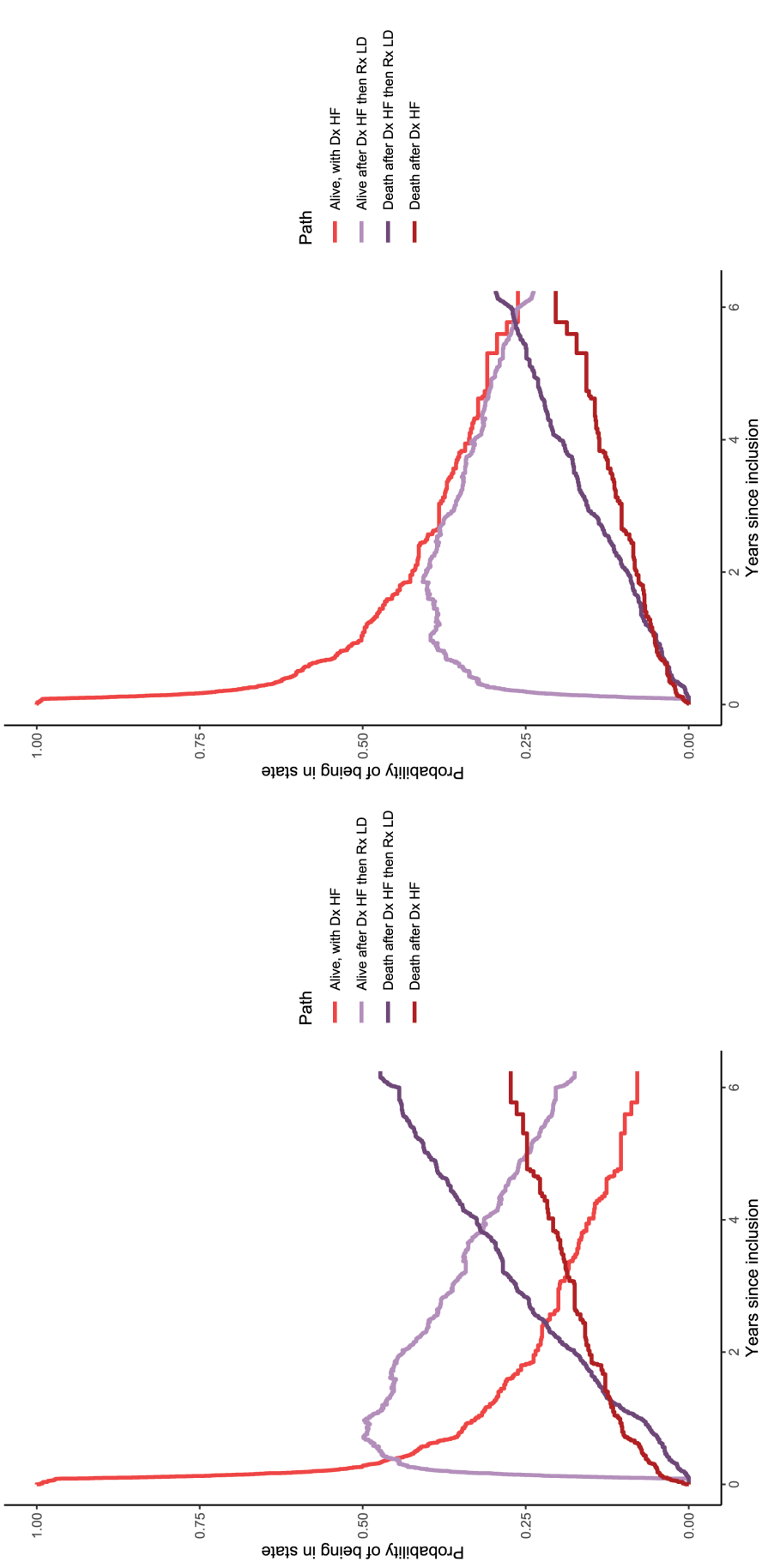
(a) Neither mode: (a) High risk for LD only (woman aged 80 years) (b) Neither mode: (b) High risk for HF only (man aged 65 years)

Figure 5.8: Probabilities of following each path over time after entering the cohort in the Neither state (no heart failure [HF] and no repeat loop diuretic [LD]).



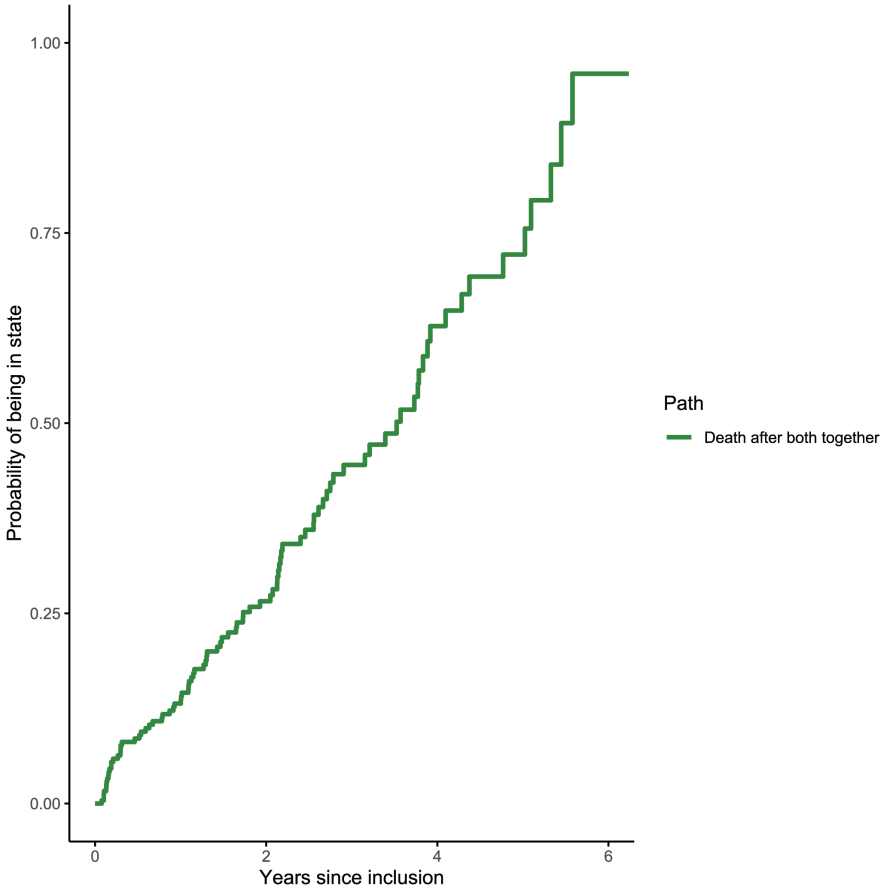
(c) LD Only mode: (a) High risk for LD only (woman aged 80 years) (d) LD Only mode: (b) High risk for HF only (man aged 65 years)

Figure 5.8: Probabilities of following each path over time after entering the cohort in the LD Only state (loop diuretic (LD) but no HF).

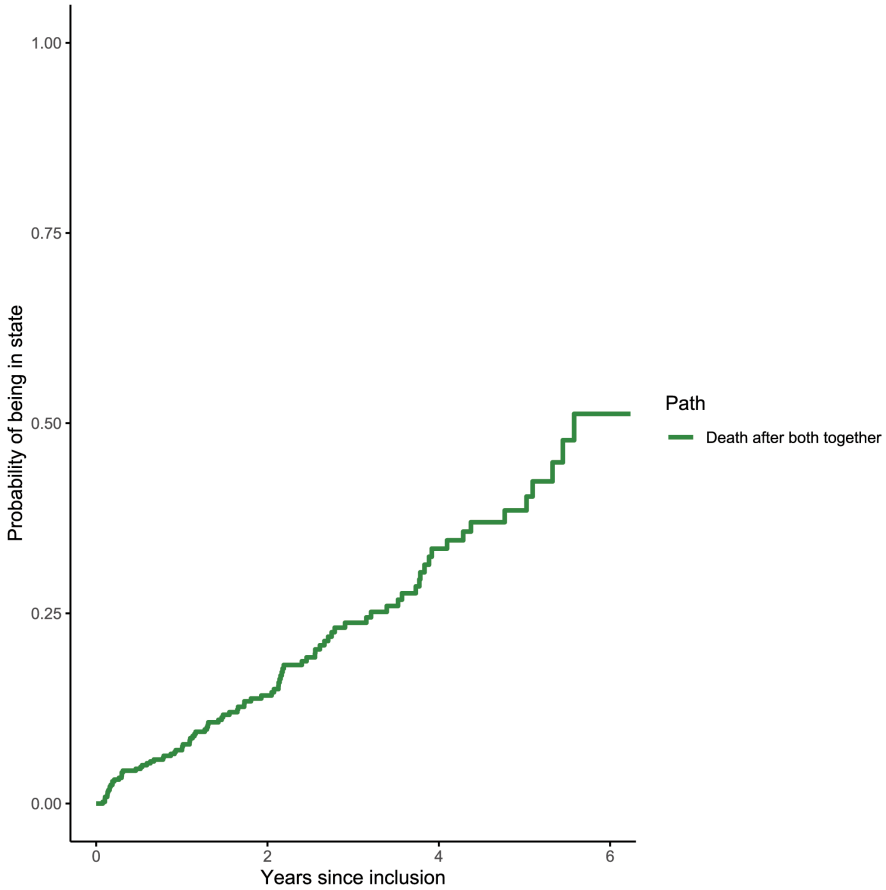


(e) HF Only model: (a) High risk for LD only (woman aged 80 years) (f) HF Only model: (b) High risk for HF only (man aged 65 years)

Figure 5.8: Probabilities of following each path over time after entering the cohort in the HF Only state (heart failure (HF) without repeat LD).



(g) Both: Together mode: (a) High risk for LD only (woman aged 80 years)



(h) Both: Together mode: (b) High risk for HF only (man aged 65 years)

Figure 5.8: Probability of death after entering the cohort in the Both: Together state. This model only has two states: alive or dead, and one possible path

5.5 Discussion

This analysis provides insights into the pattern and sequence of initiating LD and diagnosing HF and the consequences with regard to morbidity and mortality in patients with a broad range of CV disease, predominantly hypertension and CAD.

Most of the patients (78%) who initially entered the cohort with neither repeat LD nor a diagnosis of HF survived to be censored and did not transition out of the Neither state. Of the patients who entered the Neither state, 9% of patients died during follow-up without first transitioning to either LD Only, HF Only, or Both: Together, which accounted for 69% of all deaths. Most of these deaths were ascribed to non-CV causes. The most common first non-fatal transition was due to the initiation of LD in the absence of a HF diagnosis (18.1 cases per 1,000 patient-years at risk), which was three times more common than receiving a diagnosis of HF. Only a few patients (0.7%) were diagnosed with HF and initiated LD within 30 days. Before the end of follow-up, 4.0 % of the cohort received a diagnosis of HF at some point, but only 62% of those patients ever started a repeat LD. Few patients with HF died without starting a repeat LD (2.3% of all deaths from the baseline neither group, or 0.3% of the entire group). Before the end of follow-up, 2.5% of the cohort had been diagnosed with HF and initiated on repeat LD at any point, of whom just over one-third died (6.8% of all deaths from the baseline neither group). Before the end of follow-up, 9.8% of the baseline neither group were initiated on repeat LD at any point. Of those who entered the LD Only state, 29% died without receiving a diagnosis of HF, accounting for 18% of all deaths. Overall, 9.7% of patients were initiated on LD at some point, but only 25% of these patients ever received a diagnosis of HF.

Each index event was associated with a higher rate of hospitalisation in the following year, which subsequently subsided. About one-third of the admissions in the first year were admissions for CV and renal problems if patients had a diagnosis of HF; the proportion of CV related admissions was smaller in those who were only initiated on LD. However, the admissions rate was higher in patients with both a diagnosis of HF and on repeat LD. In years two and three following inclusion or transition, the rate of admissions fell for those with HF only to rates which were similar to those aged $\geq$60 years with neither LD nor HF. In contrast, at the same time windows, all groups who received LD had hospital admission rates well above those with neither LD nor HF. However, again, only a minority of admissions were for CV or renal reasons. Many admissions were coded as relating to cataracts, anaemia, fractured radial heads, respiratory, and infection reasons. However, rates of HF admissions may be under-reported as it is often coded in the secondary diagnostic position, rather than the primary (Cleland et al. 2011, Cleland & McGowan 1999) if it's not 'forgotten' altogether (Khand et al. 2005). Moreover, infections (mainly chest and urinary tract) and anaemia are common causes of worsening HF. Taking this into account, HF may have caused or complicated many more admissions than this analysis suggests.

For the relatively small number of patients that entered the cohort solely with a diagnosis of HF, 47% were initiated on a repeat LD during follow-up, which was associated with a doubling in mortality. For the somewhat large group that entered the cohort due to receiving a repeat LD, only 5.6% were subsequently diagnosed with HF; mortality (around 41%) was similar whether or not these patients were diagnosed with HF. Without adjusting for age, sex, and SIMD, those who entered the cohort with both HF and LD had a similar mortality to those who entered the cohort with LD only. In summary, the multi-state models suggest that being dispensed a LD is associated with a poor outcome compared with the diagnosis of HF alone. It is unclear if this reflects undiagnosed HF or an adverse effect of LD on prognosis, which should be investigated further. However, the majority woman makeup combined with the limited amounts of echocardiographic data suggests that many patients had a dilated LA with a normal or increased LVEF, suggesting heart failure with preserved ejection fraction (HFpEF) or heart failure with supra-normal ejection fraction (HFsnEF) (Wehner et al. 2019, Ho et al. 2012).

Identifying congestion due to cardiac dysfunction is key to determining the true burden of HF (Cleland et al. 2021). So far, guidelines have relied on S&S as measures of congestion. These S&S are clearly important as therapeutic targets in their own right but are often late manifestations of HF and are subjective, both on the part of the patient and on the healthcare professional. Biomarkers and cardiac imaging (especially the atria) are more sensitive and objective methods for assessing congestion, but this assumes that the tests have been ordered in the first place. For the tests to be ordered, someone must suspect HF or wish to exclude the diagnosis. Alternatively, screening programmes for those at risk could be implemented. Repeat prescription of LD constitutes a readily attainable, simple, inexpensive, and objective method for identifying a high-risk population which is not usually investigated for the diagnosis of HF, at least currently.

HF is not the only reason to prescribe LD. Other reasons include resistant hypertension, nephrotic syndrome, and end-stage kidney disease, but many patients may be prescribed them for breathlessness or ankle swelling (even when it might be a side effect of treatments such as amlodipine) (Savage et al. 2020, Reddi 2017). This analysis shows that many patients initiated on LD had low serum albumin, which could exacerbate or even account for peripheral oedema. Presumably, reduced hepatic synthesis of albumin was the cause, due either to hepatic congestion from HF, the inflammatory response associated with many chronic conditions or general frailty. Many studies have shown that serum albumin is a strong predictor of outcome in HF (Cleland et al. 2014). Prescriptions of dihydropyridine CCB, which may cause oedema, were similar across groups. In summary, although LD may be a risk marker, further investigations are required to determine the nature and best management of that risk.

5.5.1 Strengths and Limitations

Beyond the benefits of the cohort size mentioned in Chapter 3 and the complete and longitudinal nature of the dataset mentioned in Chapter 4, one of the major strengths of the cohort is the broad inclusion criteria which did not mandate a diagnosis of CV disease, and used CV therapies as a pharmacological signal of CV disease which was readily obtainable, in contrast to cohorts defined by the presence of hypertension (Savage et al. 2020), CAD (Schartum-Hansen et al. 2015), or AF/AFL (Zakeri et al. 2021). This allowed for a more complete and comprehensive estimate of incident LD dispensing, including 2,571 patients who did not have pre-existing records of CAD, CAD, HF, or CV therapy. However, due to the nature of PIS, it was not possible to determine LD daily dosing, nor was it possible reliably to determine changes in LD prescribing patterns (e.g., upwards or downwards titrations, prescription interruptions, or terminations of therapy entirely). For this reason, the multi-state models were designed to show patients' forward progression from inclusion, incident HF diagnosis and LD dispensing to ACM. The models do not account for pausing or terminating LD therapy or the possibility that patients might recover from HF[6], which is rare (Halliday et al. 2019). Due to these limitations, and to reduce model complexity while improving interoperability, the choice was made to develop uni-directional multi-state models.

Another limitation is the reported echocardiogram, and ECG results might be done several years after the HF diagnosis or LD initiation date. The decision was made to include information from the closest record based on the assumption that any record is preferential to no data, and many of the measurements (e.g. LVEF and LA dilation) will usually not change substantially over a period of a few years.

5.5.2 Future Work

This is the first work that describes the mortality and morbidity based on the temporal relationship of incident LD therapy and HF diagnosis in a regional cohort. The intention is to extend this work to the English population using CPRD data. Additionally, the multi-state models should be extended to investigate the effect of interruptions or terminations of LDs, changes in LD dosage, and accounting for the presence of comorbidities on outcomes.

Due to the nature of the PIS dataset, the choice was made to focus on the initiation of LD dispensing, thus fixing patients' status at a binary level, which fails to take into account the complexity of prescribing and patient adherence, including up or down titration based on perceived diuretic need. Additionally, the presence of other medications may alter diuresis (e.g., MRA, SGLT2i, NSAIDs, and thiazolidinediones) and prognosis (Testani et al. 2014, Okumura et al. 2016), thus affecting PH.

[6]Coded as the Read Code 'Heart failure resolved'.

The multi-state models were designed to analyse the impact a diagnosis of HF and/or the initiation of a repeat LD has on ACM. Currently, the models include adjustments for age, sex, and SIMD status in addition to LD and HF status.

While these models account for easily adjustable factors, and age and sex are known to have a substantial impact on ACM rates, these models do not adjust for the fact that HF rarely occurs in isolation, but rather as part of a cardio-renal pathway, and is often associated with other comorbidities such as AF/AFL, DM, CAD, and MI, all of which adversely impact ACM rates.

5.6 Conclusions

This analysis highlights that the diagnosis of HF or the initiation of LD dispensing is, as expected, usually preceded by other CV conditions. With regard to incidence, LD is initiated almost three times more often than HF is diagnosed, and only a minority of patients initiated on LD are subsequently diagnosed with HF. In contrast, most patients diagnosed with HF are subsequently initiated on LD, and those that don't have a low mortality rate. Hospitalisation rates were higher in the first year after either index event (HF diagnosis/LD initiation) but remained high the subsequent second and third years only for patients on a LD. Most of the admissions (using the primary diagnosis code) were for reasons other than cardiovascular or renal reasons. HF and renal causes may have been coded in subsequent contributory reasons of admission. Mortality was higher for patients taking LD with or without a diagnosis of HF, but especially high for those initiated on the LD >30 days in advance of the diagnosis of HF. This is perhaps due to delays in introducing disease-controlling therapy, or potentially, the diagnosis of HF indicates a failure of the LD to control S&S of congestion.

Chapter 6

Outcome of Patients with Ischaemic Heart Disease Classified by Loop Diuretic Use and Diagnosis of Heart Failure

> Without data, you're just another person with an opinion.
>
> *Edwards Deming*
> *Statistician*

6.1 Introduction

This analysis steps out of the chronological pattern of disease development found in Chapters 3, 4, and 5. Instead, it focuses on the pharmaco-epidemiology of LD use and its relationship to HF and prognosis in a population of patients with diagnosed ischaemic heart disease (IHD). IHD is a term applied to a collection of clinical syndromes characterised by a lack of myocardial blood supply based on demand. IHD is generally caused by atherosclerosis and narrowing or occlusion of one or more coronary arteries, which may lead to myocardial damage and an increased risk of developing HF, particularly heart failure with reduced ejection fraction (HFrEF). Many IHD patients will experience an acute MI, often as the first presentation of IHD. The initial management of an acute MI aims to reduce the loss of healthy myocardium tissue by revascularisation (either mechanically or by thrombolysis) in order to reduce the subsequent severity of LV dysfunction, which is a major determinant for the development of HF. The phrase 'time is muscle' is used as the longer the delay to revascularisation, the greater the damage to the myocardium. Unfortunately, many patients experiencing an acute MI do not receive treatment in a timely manner for various reasons (e.g., delays in transport, travel distance, patient hesitation to seek medical

help, delays in medical assistance). For these patients, the myocardial damage may already be severe when they arrive at hospital. Subsequent treatment for acute MI involves medications to reduce the risk of coronary re-occlusion (anti-platelet and anti-thrombotic agents), progression of atherosclerosis (lipid-regulating and anti-hypertensive agents), ventricular enlargement, also known as remodelling, (ACEi and beta-blockers), and blockade of the effects of aldosterone, thereby reducing blood pressure and therefore reducing the load on the LV. Aldosterone blockade may also prevent hypokalaemia (an adverse side effect of many diuretics including LD see Section 2.5.9) and reduce myocardial fibrosis (MRA and ACEi) (Ibanez et al. 2017).

Epidemiological studies from 20 years ago showed that most people who initially survived an acute MI but subsequently died, developed HF between the two events (Torabi et al. 2008, 2009). It is possible that contemporary epidemiology might differ due to the widespread introduction of PCI for acute management (Torabi et al. 2008), but this avenue of enquiry has not been adequately explored.

However, many patients with IHD and a diagnosis of HF do not have a documented history of acute MI (Sigurdsson et al. 1995, Danielsen et al. 2017), with up to one in three acute MIs going unrecognised (Sigurdsson et al. 1995). This could be the case for several reasons. Many patients with an acute MI do not seek urgent medical attention for varied reasons. An epidemiologic cohort study with the initial inclusion criteria based on participation in a cross-sectional study with follow-up starting in 1967 (Harris et al. 2007) showed that more than half of people with MI scars on imaging did not recall an acute event required for the diagnosis of an acute MI (Schelbert et al. 2012). Additionally, studies like this will likely underestimate the true rate of unrecognised acute MI because many patients who met the original enrolment criteria in 1967 will not have survived long enough to be included in the subsequent cohort study, which started enrolling patients in 2004. Importantly, researchers believe that the unrecognised MI scars are not an imaging artefact as the presence of scars predicts outcomes similarly whether or not they are associated with a prior recognised and diagnosed acute MI (Yang et al. 2020, Schelbert et al. 2012, Kwong et al. 2006). Another explanation for the association between IHD and HF, independent of MI, is that IHD may operate synergistically with hypertension, another important risk factor for HF, to cause chronic myocardial damage to recurrent ischaemia or low-grade inflammation (Libby et al. 2009) leading to myocardial fibrosis. This, compounded by myocardial scars due to MI, may compound this problem. The combination of IHD and hypertension may also conspire to cause AF and, along with DM, chronic renal disease. All of these conditions will increase the risk of developing HF. Finally, because most patients with HF are aged >60 years and have risk factors for IHD, they may also have coronary disease, which is not the cause of myocardial dysfunction.

Previous chapters have described the prevalent and incident use of LD, and have shown that many of these patients have little or no coded evidence of cardiac disease. It is possible that

many of these patients were receiving LD inappropriately in the absence of cardiac or renal disease. The current consensus on the definition of HF stipulates that patients have symptoms of congestion and cardiac dysfunction (Bozkurt et al. 2021), which some experts view as a flawed definition (Cleland et al. 2021). The use of LD may be considered a pharmacological surrogate for reported symptoms and signs of congestion. Additionally, having a diagnosis of IHD indicates that the patient has a documented cardiac problem. A patient treated with a LD who has a diagnosis of IHD might be considered to fulfil the current definition of HF. If the analysis of patients with IHD shows similar results to previous chapters, this lends further support to the view that a large proportion of patients treated with a LD have HF.

Accordingly, this chapter shows the investigation of mortality for patients with IHD classified according to recurrent LD dispensing or a diagnosis of HF.

6.1.1 Aims

This chapter sets out to describe the NHS GG&C population with a history of IHD by the presence or absence of a diagnosis of HF or recurrent dispensing of LD on the 1st January 2013 and their outcome through 31st March 2018.

6.2 Background

Missing Data

For research, and particularly research involving EPR data, the presence of missing data is the rule, not the exception. Identifying what data are missing, how the missingness relates to the outcome(s) of interest and algorithms for handling missing data are preliminary steps for analysing EPR data. The pattern of missingness can be classified in one of three ways:

1. Missing completely at random (MCAR) is the strictest of the three classifications. Missing data are a random subset of all data for the variable in question where the observed and missing data will have similar distributions.

2. Missing at random (MAR) is a slightly more relaxed classification compared with MCAR, in that there might be a systematic difference between the missing and observed data, but the probability of missingness depends only on the observed data.

3. Finally, missing not at random (MNAR) is the classification assigned when the probability of data being missing depends on the value itself within the context of the research question. For example, investigations are less likely to be done on clearly healthy individuals (Bhaskaran & Smeeth 2014, Rothman et al. 2021, Dong & Peng 2013).

For example, within this dataset, the absence of haemoglobin results was associated with lower

ACM. Based on previous analysis using the GSH/18/CA/002 dataset, the absence of a haemoglobin result was associated with lower ACM (Graham et al. 2020, 2023); therefore, missing blood test values were classified as MNAR and handled accordingly.

There are several ways to handle continuous MNAR data, with some methods being less prone to bias than others. What follows is a brief overview and justification for decisions about handling missing blood test results. The simplest method is to restrict the analysis to complete cases, although complete case analysis has the potential to badly bias results (Rothman et al. 2021, Harrell et al. 2015).

A slightly more involved method of handling missingness is to compute replacement values using algorithms such as multiple imputations by chained equations. These imputation methods generally assume that data are MCAR, or MAR at a minimum, to maintain consistency between imputed and existing values. In this case, neither MCAR nor MAR was met, meaning that imputed results will be biased to the point where complete case analysis might be preferable to multiple imputation (Hughes et al. 2019), depending on the situation and data.

An alternative is to understand the dependence of results on missing values by measuring the changes in results when a credible range of values are substituted for the missing values (Lash et al. 2014). This method benefits from its transparency and being upfront about how the missingness impacts results. Additionally, it is straightforward to implement, allowing research to be clear regarding the limitations of the analysis.

6.3 Methods

6.3.1 Study Population

Patients were eligible for inclusion if they were alive and aged 18 years or older on 1st January 2013 but were only included if their EPR were available for at least the preceding 12 months to avoid under-reporting of diagnoses and LD prescribing. Patients could be included if they had a record of IHD in the three years between 31st December 2009 through 1st January 2013. The presence of a diagnosis was identified using either Read or ICD-10 codes in any position (see tables E.1 and E.2 [Reeves et al. 2014] for codes used to define cohort).

6.3.2 Patient Identification and Classification

Repeat LD prescription was defined as the first time a LD was dispensed in two consecutive quarters or if the patient died within 90 days of the first LD prescription date (see Section 2.5.4 for further details of identifying repeat dispensing, and Table A.3 for a list of qualifying LD medications). The initiation of LD therapy during follow-up was defined as the first time the above definition was met. Prevalent HF was defined according to Section 2.6.1. Patients were

excluded if the first record of HF during follow-up indicated a pre-existing diagnosis (see Section 2.6.1 and Table A.16 for qualifying codes), suggesting incomplete data capture (see Figure E.1 for the patient flow diagram). A first diagnosis of HF associated with a fatal hospitalisation was not counted as HF.

Patients were classified at baseline into discrete groups based on prevalent records of HF and LD therapy on or prior to the 1st January 2013, as: 'No LD/No HF', LD Only', 'HF Only', and 'Both: LD + HF.

6.3.3 Study Outcomes

Patients were followed from 1st January 2013 until 31st March 2018, inclusive. Follow-up was right censored at the time of last available EPR to avoid uncertainty about survival status. The primary covariate of interest was group membership as defined by LD dispensing and diagnosis of HF. The cumulative incidence of HF, initiation of repeat LD dispensing, event-free survival, and cause-specific rates of mortality were also estimated. The cause of death was defined using the underlying cause of death in a patient's death record (see Section 2.5.1 for further information), mapped onto five disease categories (see Table B.1).

6.3.4 Patient Characteristics

Baseline characteristics were framed as patient's status as of 1st January 2013, including age, sex, ethnicity, and quintile of SIMD using the 2012 status (see Section 2.5.2), comorbidities, current medications, most severe blood test result within the prior two years, and, when available, the results from the most recent ECG, and most severe echocardiogram results.

Reported comorbidites included: history of hypertension, DM, thyroid disease, AF/AFL, MI, valve disease, PAD, stroke, COPD, cancer, and dementia. With the exception of DM, comorbidities were defined according to Section 2.6.2. As an exception to the above, DM or concurrent hypoglycaemic agents (Diabetes+) was defined as the composite of a coded record of DM or the presence of a dispensed medication for the treatment of DM within and inclusive of the 180 days prior to 1st January 2013 (see Table A.3 for qualifying insulin and hypoglycaemic medications). The presence of a comorbidity was defined as a coded diagnosis on or before 1st January 2013. Patients without a diagnosis were considered to be free from the condition. Additionally, due to the focus on IHD, the history of percutaneous coronary intervention (PCI) and coronary artery bypass surgery (CABG) were reported based on the history of a procedure defined using OPCS-4 codes.

As with Chapters 3 - 5, the medication classes reported included ACEi, ARB, beta-blocker, MRA, CCB including diltiazem or verapamil and dihydropyridine, digoxin, Thiazides+, low dose aspirin, lipid regulators, bronchodilators, thyroid medications, and hypoglycaemic agents

including insulin (see Section 2.5.4 and Table A.3 for further details). Patients were considered to be on these medications if a prescription was dispensed in the 180 days prior to 1st January 2013, inclusive.

In order to determine those at higher risk of morbidity and mortality, the lowest results in the prior two years for haemoglobin and eGFR were reported. Anaemia was defined according to the WHO's definition assuming no pregnancies (see Section 2.6.2 for further details). The eGFR was calculated according to Section 2.5.9.

Echocardiographic measurements were often missing, but information on the most abnormal LVEF, LA diameter, aortic velocity (measuring aortic stenosis AS), tricuspid regurgitation (TR), and presence of mitral regurgitation (MR) was reported where available (see Section 2.5.7 for more information). As 12-lead ECGs were more common, information on heart rhythm, QRS duration, QTc and QTc prolongation, and ST-T abnormalities were reported from the closest report (see 2.5.6).

6.3.5 Statistical Analysis

As with Chapters 3 - 5, patient characteristics were presented as numbers and percentages for categorical data and median (1st-3rd quartile) values for continuous data. For categorical variables, percentages refer to complete cases.

To estimate the prevalence of IHD within the wider NHS GG&C population, the mid-year 2013 population estimate was applied (National Records of Scotland 2018).

A Cox PH regression model with time-dependent covariates (see sections 3.2.1 and 3.2.1 for further information) was used to assess between-group differences in ACM while taking into account disease progression during follow-up. The model was adjusted for age on 1st January 2013, sex, and SIMD using 2013 values, while time-dependent covariates were used for group status, comorbidity levels, and the most severe haemoglobin and serum creatinine results from the past two years were used to assess the between-group differences on ACM. Due to the large sample size and high statistical power to detect small departures from PH, proportional effects were checked using log-log plots.

To account for missing haemoglobin and serum creatinine values, the time-dependent Cox PH regression analysis was repeated five times, where missing haemoglobin and serum creatinine values were, in turn, replaced with the third and ninety-seventh percentile values sampled from a sex and age-matched results from patients with available records. This analysis was done to assess the impact missing values had on the primary study outcome.

Analysis was conducted using R (see Section 2.4.2) (R Core Team 2021), using the packages mentioned in Section 3.3.5.

6.4 Results

The eligibility criteria were met by 24,921 patients with IHD who contributed a total of 109,889 years of follow-up with a median of 5.2 (1st - 3rd quartiles: 4.4 - 5.2) years. The estimated prevalence of IHD for the adult NHS GG&C population was 2.5%. On 1st January 2013, the cohort included 15,200 (61%) patients who had no record of HF nor repeat LD dispensing, 3,806 (15%) who received repeat LD but did not have a diagnosis of HF, 2,384 (10%) who had a diagnosis of HF but had not received repeat LD, and 3,531 (14%) who had both a diagnosis of HF and received repeat LD.

Table 6.1: Baseline patient demographics, comorbidities, and blood tests on 1st January 2013.

Variable	**No LD/No HF**	**LD Only**	**HF Only**	**Both: LD + HF**
n	15,200	3,806	2,384	3,531
Age (years)	68 (59 - 76)	77 (68 - 83)	69 (58 - 77)	76 (69 - 83)
Sex				
Women	6,266 (41%)	2,341 (62%)	723 (30%)	1,632 (46%)
Men	8,934 (59%)	1,465 (38%)	1,661 (70%)	1,899 (54%)
Ethnicity				
White	13,259 (87%)	3,583 (94%)	2,126 (89%)	3,315 (94%)
Missing	1,322 (9%)	137 (4%)	180 (8%)	106 (3%)
Other	619 (4%)	86 (2%)	78 (3%)	110 (3%)
Socioeconomic deprivation (SIMD)				
1 (most deprived)	6,364 (42%)	1,753 (46%)	1,033 (43%)	1,518 (43%)
2	2,791 (18%)	698 (18%)	439 (18%)	696 (20%)
3	1,939 (13%)	515 (14%)	300 (13%)	510 (14%)
4	1,763 (12%)	404 (11%)	280 (12%)	400 (11%)
5 (least deprived)	2,343 (15%)	436 (11%)	332 (14%)	407 (12%)
Contact with secondary care in 2012				
Any secondary care	11,221 (74%)	3,256 (86%)	1,832 (77%)	3,163 (90%)
Cardiology specialist	3,780 (25%)	944 (25%)	772 (32%)	1,458 (41%)
Comorbidities $*$				
H/o hypertension	7,887 (52%)	2,385 (63%)	1,170 (49%)	2,189 (62%)
Diabetes+	3,813 (25%)	1,292 (34%)	580 (24%)	1,381 (39%)
Thyroid disease	600 (4%)	278 (7%)	108 (5%)	251 (7%)
MI	600 (4%)	278 (7%)	108 (5%)	251 (7%)
Valve disease	658 (4%)	433 (11%)	266 (11%)	820 (23%)
AF/AFL	1,638 (11%)	1,020 (27%)	454 (19%)	1,493 (42%)
PAD	916 (6%)	385 (10%)	207 (9%)	545 (15%)
Stroke	1,473 (10%)	646 (17%)	287 (12%)	655 (19%)

Continuation of Table 6.1

Variable	**No LD/No HF**	**LD Only**	**HF Only**	**Both: LD + HF**
n	15,200	3,806	2,384	3,531
COPD	2,829 (19%)	1,231 (32%)	515 (22%)	1,160 (33%)
Cancer	1,230 (8%)	454 (12%)	215 (9%)	396 (11%)
Dementia	477 (3%)	290 (8%)	64 (3%)	217 (6%)
Procedures ∗				
PCI	4,066 (27%)	560 (15%)	938 (39%)	754 (21%)
CABG	1,311 (9%)	406 (11%)	368 (15%)	439 (12%)
Blood Results ∗∗				
Haemoglobin (Hb) (g/dL)				
Recorded	13,799 (91%)	3,686 (97%)	2,212 (93%)	3,457 (98%)
Women: Hb	12.1 (10.7 - 13.1)	11.2 (9.6 - 12.5)	11.6 (9.9 - 12.9)	10.3 (8.9 - 11.7)
Men: Hb	13.6 (12.0 - 14.6)	11.8 (10.0 - 13.4)	13.3 (11.7 - 14.4)	11.4 (9.7 - 13.2)
Anaemia◇	5,759 (42%)	2,417 (66%)	1,056 (48%)	2,617 (76%)
Estimated glomerular filtration rate (eGFR)∇ (mL/min/1.73m^2)				
Recorded	14,810 (97%)	3,779 (99%)	2,339 (98%)	3,505 (99%)
Women: eGFR	71 (54 - 85)	53 (35 - 69)	66 (44 - 80)	37 (25 - 57)
Men: eGFR	79 (62 - 92)	56 (38 - 77)	74 (57 - 90)	45 (29 - 66)
eGFR <60	3,860 (26%)	2,168 (57%)	772 (33%)	2,535 (72%)

Data are frequencies (%) for categorical values or median (1st - 3rd quartile) for continuous values.

∗ Record on or before 1st January 2013.

∗∗ Based on the most extreme value in the two years before 1st January 2013.

◇ Using the WHO definition of anaemia (see Section 2.6.2).

∇ Calculated eGFR assuming no pregnancies and without adjusting for ethnicity (see Section 2.5.9).

H/o; History of; Diabetes+, a coded diagnosis of DM or a hypoglycaemic medication dispensed in the 180 days on or before 1st January 2013.

Table 6.2: Baseline patient medications dispensed between 180 days before through 1st January 2013 by group.

Variable	**No LD/No HF**	**LD Only**	**HF Only**	**Both: LD + HF**
n	15,200	3,806	2,384	3,531
Age (years)	68 (59 - 76)	77 (68 - 83)	69 (58 - 77)	76 (69 - 83)
ACEi or ARB	9,166 (60%)	2,174 (57%)	1,872 (79%)	2,587 (73%)
ACEi	7,498 (49%)	1,638 (43%)	1,572 (66%)	2,065 (58%)
ARB	1,862 (12%)	603 (16%)	332 (14%)	615 (17%)
Beta-blocker	9,921 (65%)	2,114 (56%)	1,805 (76%)	2,487 (70%)
MRA	80 (1%)	111 (3%)	116 (5%)	645 (18%)
Spironolactone	62 (78%)	101 (91%)	66 (57%)	479 (74%)
Eplerenone	18 (22%)	10 (9%)	52 (45%)	175 (27%)
CCB	4,927 (32%)	1,440 (38%)	576 (24%)	792 (22%)
Diltiazem/Verapamil	1,310 (9%)	515 (14%)	113 (5%)	168 (5%)
Dihydropyridine	3,677 (24%)	949 (25%)	469 (20%)	637 (18%)
Digoxin	191 (1%)	257 (7%)	71 (3%)	588 (17%)
Thiazides+	2,155 (14%)	180 (5%)	186 (8%)	127 (4%)
Low dose aspirin	11,602 (76%)	2,600 (68%)	1,842 (77%)	2,305 (65%)
Oral anticoagulants	1,665 (11%)	1,064 (28%)	395 (17%)	1,405 (40%)
Lipid regulators	13,001 (86%)	3,108 (82%)	2,090 (88%)	2,929 (83%)
Bronchodilators	2,746 (18%)	1,137 (30%)	456 (19%)	1,007 (29%)
Thyroid medications	1,286 (8%)	511 (13%)	185 (8%)	423 (12%)
Hypoglycaemic agents	2,250 (15%)	828 (22%)	331 (14%)	905 (26%)
Insulin†	445 (3%)	269 (7%)	83 (3%)	343 (10%)
Other hypoglycaemic agents	2,054 (14%)	708 (19%)	291 (12%)	729 (21%)

Data are frequencies (%) for categorical values or median (1st - 3rd quartile) for continuous values.

† Either alone or in combination with another agent.

Table 6.3: Measurements from the closest 12-lead ECG to 1st January 2013 (within the time-window of 31st December 2009 through 31st March 2018). Baseline status is defined based on patient's LD and HF status on the 1st January 2013.

Variable	**No LD/No HF**	**LD Only**	**HF Only**	**Both: LD + HF**
n	15,200	3,806	2,384	3,531
ECG available	10,338 (68%)	2,669 (76%)	1,643 (69%)	2,876 (76%)
Heart rate (bpm)	69 (59 - 81)	72 (63 - 86)	68 (60 - 81)	73 (62 - 86)
QRS duration (ms)	88 (80 - 98)	100 (88 - 126)	94 (84 - 108)	90 (82 - 102)
Heart rhythm				
Sinus	9,201 (89%)	1,691 (63%)	1,380 (84%)	2,179 (76%)
AF/AFL/SVT	866 (8%)	677 (25%)	188 (11%)	556 (19%)
Other	215 (2%)	258 (10%)	67 (4%)	114 (4%)
Undetermined	56 (1%)	43 (2%)	8 (<1%)	27 (1%)
QTc available	9,730 (94%)	2,530 (95%)	1,563 (95%)	2,700 (94%)
QTc (ms)‡	421 (405 - 438)	436 (415 - 462)	424 (408 - 446)	429 (411 - 449)
Prolonged QTc‡	1,074 (11%)	806 (32%)	286 (18%)	495 (18%)
ST-T abnormality	2,518 (24%)	1,072 (40%)	534 (33%)	1,051 (37%)
Acute MI recorded in hosp	49 (<1%)	24 (1%)	21 (1%)	11 (<1%)
Most abnormal results				
Echo available	6,929 (46%)	2,441 (69%)	1,332 (56%)	2,248 (59%)
LVEF available	3,878 (26%)	1,391 (39%)	776 (33%)	1,165 (31%)
LVEF (%)	64 (55 - 70)	48 (35 - 61)	53 (40 - 64)	62 (52 - 68)
$\geq$ 50%	3,222 (83%)	634 (46%)	431 (56%)	906 (78%)
41% - 49%	391 (10%)	288 (21%)	155 (20%)	151 (13%)
$\leq$ 40%	265 (7%)	469 (34%)	190 (24%)	108 (9%)
LAD available	6,048 (40%)	2,181 (62%)	1,173 (49%)	1,997 (52%)
Women:LAD (cm)	3.8 (3.4 - 4.2)	4.4 (3.8 - 5.0)	3.9 (3.4 - 4.4)	4.1 (3.6 - 4.6)
Men:LAD (cm)	4.1 (3.7 - 4.5)	4.7 (4.2 - 5.2)	4.3 (3.8 - 4.8)	4.5 (4.0 - 5.1)
LAD dilated	3,151 (52%)	1,722 (79%)	711 (61%)	1,375 (69%)

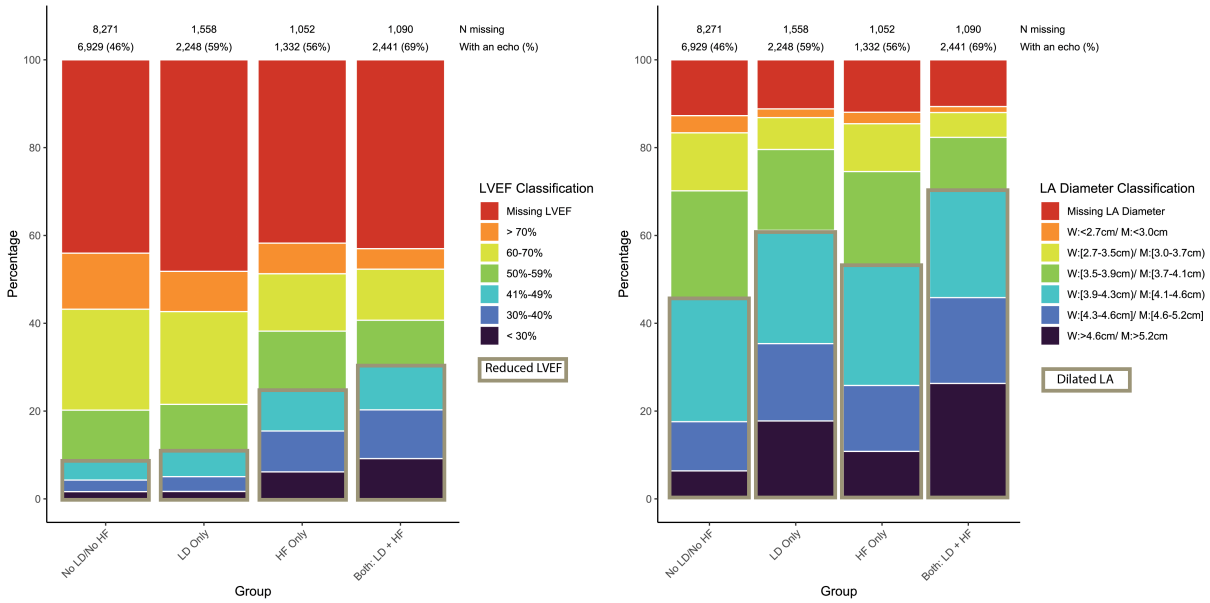
Data are frequencies(%) for categorical values or median (1st - 3rd quartile) for continues values.
‡ Corrected QT interval calculated using the Fridericia formula (see Section 2.5.6).
Echo, Echocardiogram; LAD, Left atrial diameter.

For patients with IHD, those who received repeat LD no record of a diagnosis of HF were older, with a median age of 77 (1st - 3rd quartile: 68 - 83) years, more likely to be women (2,341

[62%]), more likely to have an eGFR <60 mL/min/1.73m^2 (3,779 [57%]), and more likely to be anaemic (2,417 [66%]) (see Table 6.1) compared with the other groups. Even in this subset with IHD, patients taking LD without a diagnosis of HF were less likely to have a history of MI compared with those with a diagnosis of HF. LVEF was <50% for 259 (22%) and >70% in 206 (18%) of the 1,165 patients with an available LVEF measurement (see Table 6.3 and Figure 6.1). The LA was dilated in 1,375 (69%) of the 1,997 patients with an available LA diameter measurement.

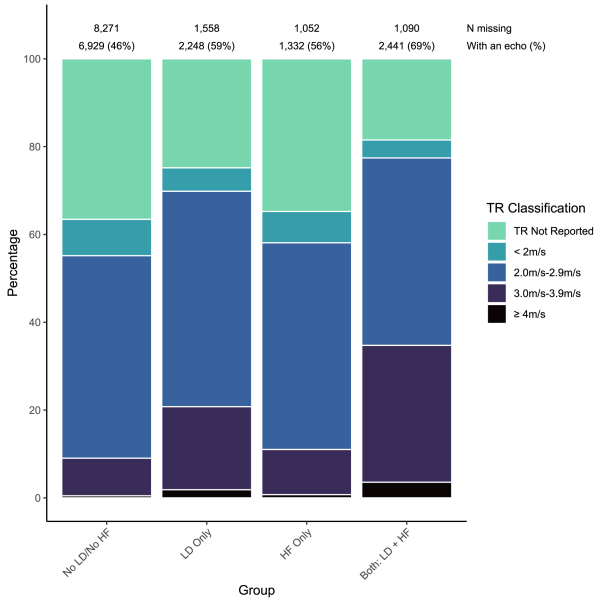
Patients with IHD and HF who had not received repeat LD dispensing at baseline were younger with a median age of 69 (1st - 3rd quartile: 58 - 77) years, more likely to be men (1,661 [70%]), and to have a history of MI (1,905 [80%]), but less likely to have an eGFR <60 mL/min/1.73m^2 (772 [33%]), or to be anaemic (1,056 [48%]). LVEF was <50% for 345 (44%) and >70% for 93 (12%) of the 776 patients with an available LVEF measurement. The LA was dilated in 711 (61%) of the 1,173 patients with an available LA diameter measurement.

Patients with IHD and HF who had received repeat LD dispensing at baseline were of a similar age to those with only a LD prescription with a median age of 76 (1st - 3rd quartile: 69 - 83) years, and more likely to be men (1,899 [54%]). They had high levels of prior MI, though lower than those with HF alone (2,358 [67%]). They were more likely to have AF/AFL (1,493 [42%]) and more likely to have an eGFR <60 mL/min/1.73m^2 (2,535 [72%]) and anaemia (2,617 [76%]) than the other groups. LVEF was <50% for 757 (54%) and >70% for 114 (8%) of the 1,391 patients with an available LVEF measurement. The LA was dilated in 1,722 (79%) of the 2,181 patients with an available LA diameter measurement.

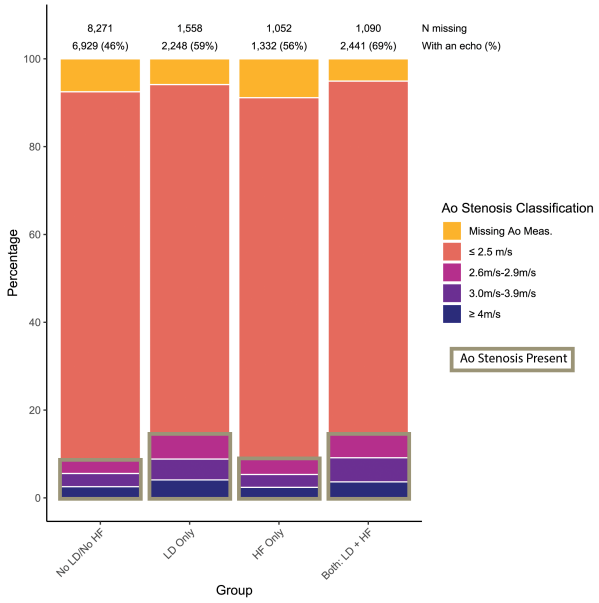


(a) Left ventricular ejection fraction (LVEF)

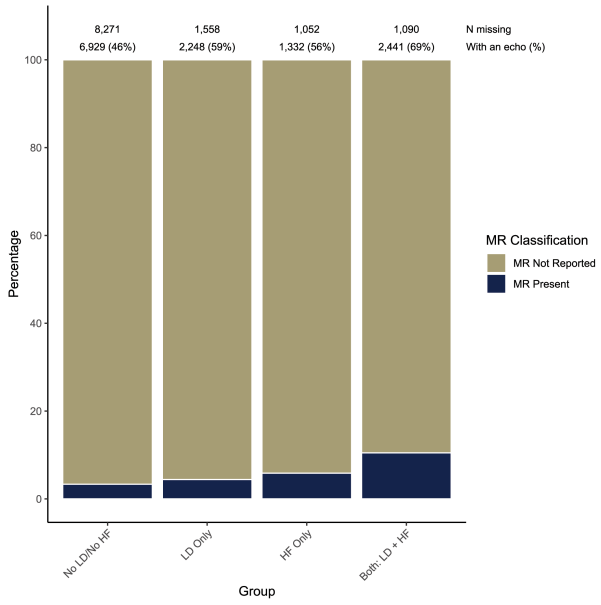
(b) Left atrial (LA) diameter



(c) Tricuspid regurgitation (TR)



(d) Aortic stenosis (AS)

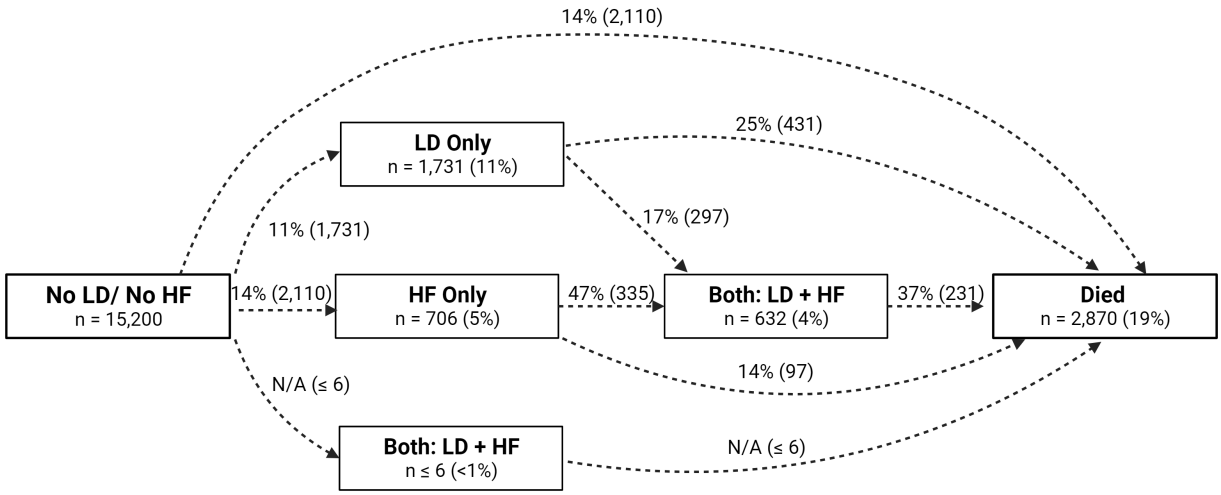


(e) Mitral regurgitation (MR)

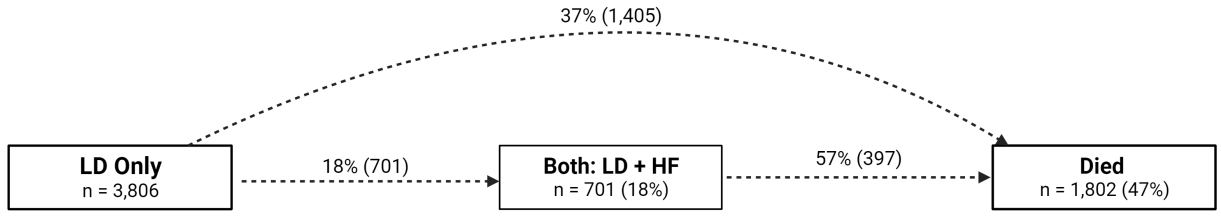
Figure 6.1: Most abnormal echocardiogram result for left ventricular ejection fraction (LVEF) (a), left atrial (LA) diameter (b), tricuspid regurgitation (TR) (c), aortic stenosis (AS) (d), and mitral regurgitation (MR) (e) using patients' baseline status. Patients were classified according to their LD and HF status on the 1st January 2013. Grey boxes indicate reduced LVEF, dilated LA, or AS.

6.4.1 Morbidity and Mortality from Baseline (Figures 6.2 - 6.4)

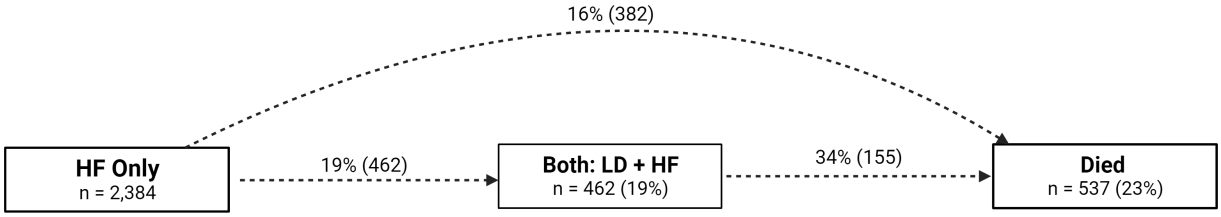
For patients with IHD who were neither taking repeat LD nor had a diagnosis of HF at baseline, over the following 5 years, 1,555 (10%) of the group were initiated on repeat LD only without a



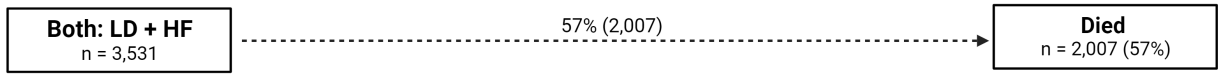
(a) No LD/No HF



(b) LD Only



(c) HF Only



(d) Both: LD + HF

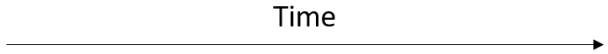
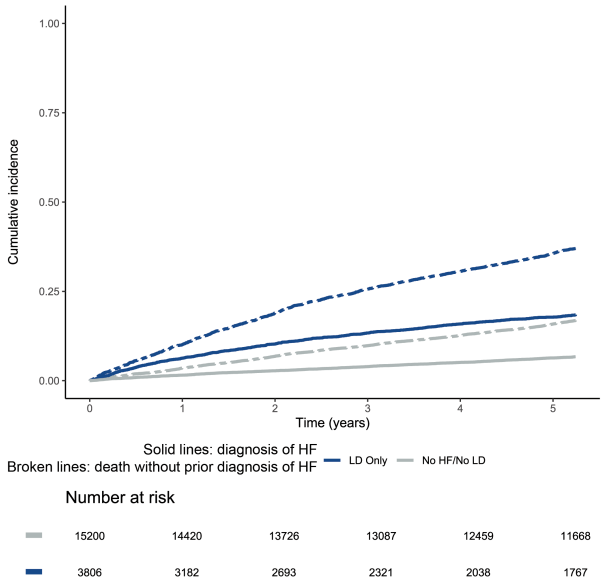
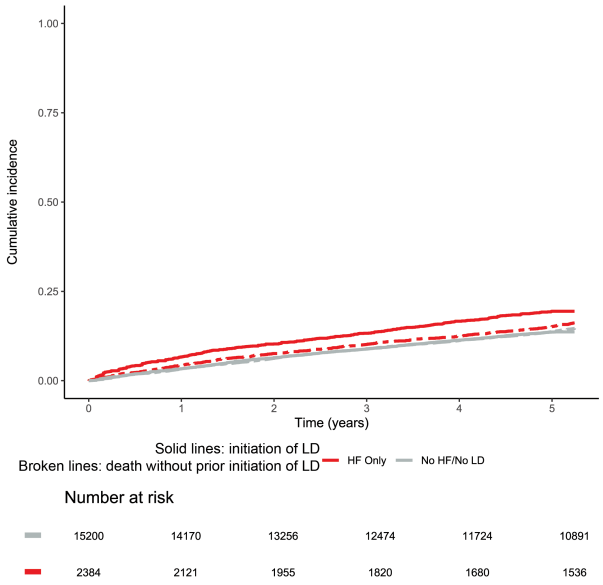


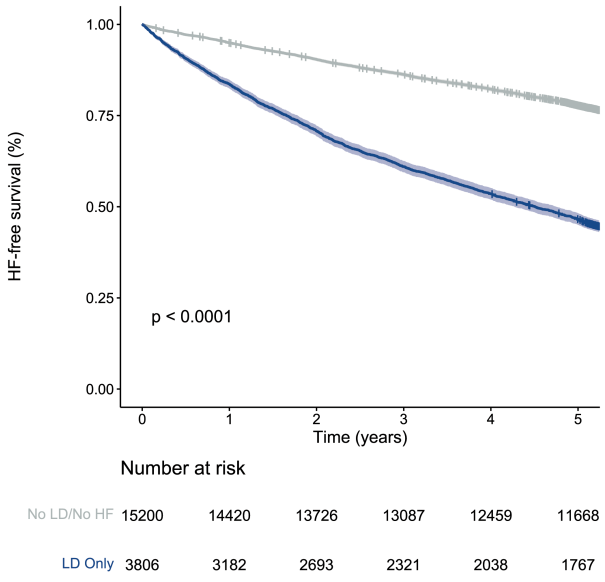
Figure 6.2: Transition diagrams illustrating how many patients started in each baseline group (left most boxes), and how many were subsequently diagnosed with HF, initiated on LD therapy, or died between 1st January 2013 through 31st March 2018. Percentages within the boxes are out of the baseline group, while percentages on transition are out of those eligible for said transition.



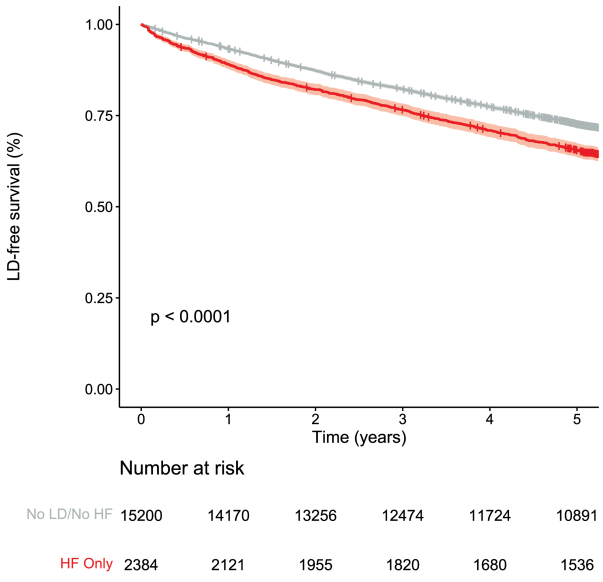
(a) Competing risks by LD status



(b) Competing risks by HF status

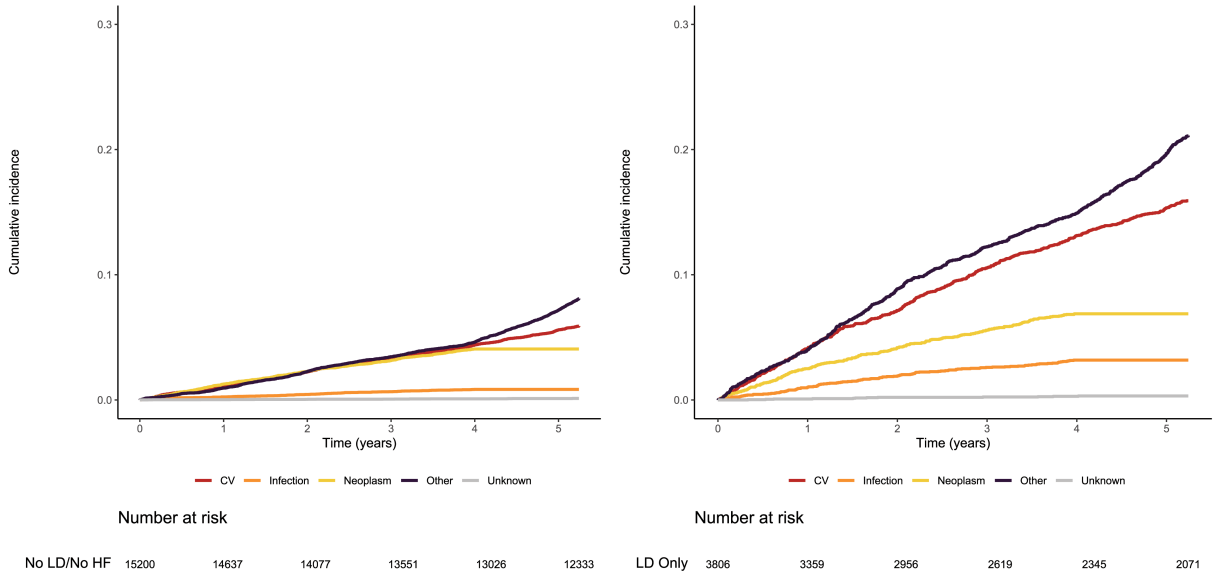


(c) Event free survival by LD status



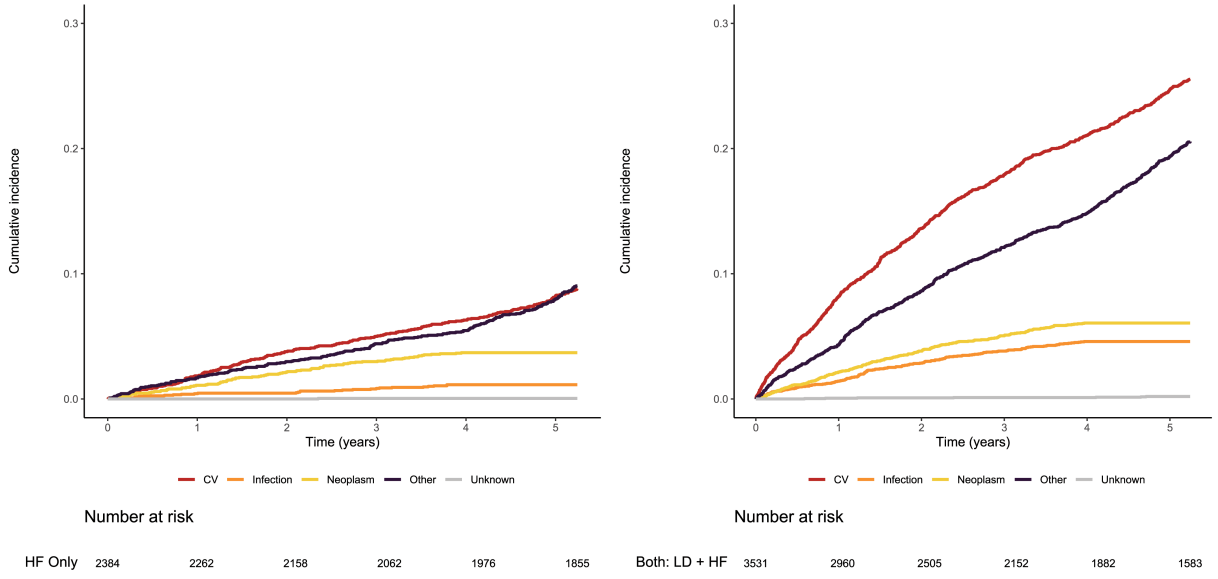
(d) Event free survival by HF status

Figure 6.3: (a & b): Estimation of the cumulative incidence of a diagnosis of HF, or initiation of a repeat LD prescription respectively, while taking into account ACM removing patients from those at risk of being diagnosed with HF or initiating LD therapy, respectively. (c & d): Event-free survival where patients are censored at the first of diagnosis of HF (or the initiation of LD respectively) or ACM. Patients were classified according to their HF and repeat LD status on the 1st January 2013.



(a) No LD/No HF

(b) LD Only



(c) HF Only

(d) Both: LD + HF

Figure 6.4: Cumulative incidence of the cause of death by baseline group classification between 1st January 2013 through 31st March 2018.

subsequent diagnosis of HF, 371 (2%) were diagnosed with HF only without a subsequent repeat LD, and just over 632 (4%) of the group were both initiated on repeat LD and were diagnosed with HF. In all, 2,870 (19%) died from any cause (see Figure 6.2a). The most common cause of death in patients who had IHD but who had neither a diagnosis of HF nor were initiated on repeat LD at baseline was miscellaneous other reasons (an estimated 7% of the No LD/No HF group), followed by CV reasons (an estimated 6%), then neoplasms, and finally infections (see Figure 6.4a).

For patients with IHD and repeat LD only at baseline, over the following 5 years, 18% (701 patients) were diagnosed with HF, and 47% (1,802 patients) died (see Figure 6.2b). The most common cause of death was for miscellaneous other reasons (an estimated 20% of the LD only), followed by CV causes (an estimated 15%), then neoplasm, and infection (see Figure 6.4b).

For patients with IHD and HF only at baseline, over the following 5 years, 19% (462 patients) started a repeat LD during follow-up, and 23% of the group died (537 patients) (see Figure 6.2c). The most common causes of death were CV causes and miscellaneous other reasons (each with an estimated 8% of the HF only), followed by neoplasms (with an estimated 4%), the majority of which was due to cancer (see Figure 6.4c).

For patients with IHD and both a diagnosis of HF and a repeat LD at baseline, 57% (2,007 patients) died over the following 5 years (see Figure 6.2d). Across those years, the most common cause of death was for CV reasons (an estimated 25% of Both: LD + HF), followed by miscellaneous other reasons (an estimated 19%), then neoplasms and infections (see Figure 6.4d).

Overall, patients with IHD but were neither taking repeat LD nor had a diagnosis of HF at baseline had the best 5-year event-free survival of remaining alive and free of a diagnosis of HF and never starting repeat LD. Patients with either repeat LD or a diagnosis of HF at baseline were more likely to experience a subsequent event. In contrast, patients with LD only were more likely to die (see Figures 6.3a and 6.3d). In contrast, patients with a diagnosis of HF only were more likely to be initiated on repeat LD than to die beforehand (see Figures 6.3b and 6.3c).

Finally, 7,216 patients died during the 5 years of follow-up, of whom 71% were either initiated on repeat LD or had a diagnosis of HF (or both) prior to death. Of those that died, 25% were on a repeat LD only, 7% had a diagnosis of HF only, and 39% had a diagnosis of HF and were initiated on repeat LD. Altogether, of those that died, 64% were initiated on a repeat LD, with or without a diagnosis of HF, and 45% received a diagnosis of HF, with or without receiving repeat LD.

6.4.2 Mortality Using Time-Dependent Covariates

Missing Data

There were low levels of missing haemoglobin values with a maximum of only 9% of patients missing values from the group with IHD but without a diagnosis of HF nor repeat LD. Patients were more likely to have serum creatinine measured as the highest level of missingness accounted for only 3% of those with IHD but without a diagnosis of HF not repeat LD. When testing for potential biases, the largest variation in the HR occurred in the haemoglobin results, with a difference in the adjusted HR starting at 2.1 (95% CI: 2.0 - 2.3) for 97^{th} percentile and increasing to 2.3 (95% CI: 2.1 - 2.4) for 3^{rd} percentile (see Figure E.2). In all cases, the significance levels were <0.001. The differences in the HR found when substituting the missing variables

were small. Where differences were found, the 95% CI for these values were still statistically significant; therefore, substituting missing values with the median value from the age and sex-matched test values was considered a reasonable substitution for the MNAR haemoglobin and serum creatinine values.

Assumptions Testing

The PH and linearity assumptions on mortality for group status, age, haemoglobin, and serum creatinine values were tested. The PH assumption for group status was not met for the first 36 days, but this applied to only 217 deaths (3% of all deaths). Accordingly, PH was considered to be a reasonable summary of between-group differences (see Figure E.3a). The linearity assumption was deemed to be met based on plotting the log-hazards of the mean value of age per decile of age (see Figure E.3b). There was a slight deviation from linearity in the haemoglobin values (see Figure E.3c). The addition of a sex interaction term did not add statistical information. For these reasons, the linearity assumption of haemoglobin on mortality was considered to be reasonable. Finally, the linearity assumption was not met for serum creatinine, particularly due to the influence of particularly high values (see Figure E.3d). The addition of two- and three- way age and sex interaction terms did not provide additional statistical information. For these reasons, combined with ease of interpretation, the linearity assumption was used.

Time-Dependent Covariates

Using time-dependent covariates and the 'No LD/No HF' group as a reference, the adjusted (which includes age) HR for mortality for those dispensed LD without a diagnosis of HF was 1.7 (95% CI:1.6 - 1.8; p-value <0.001), for those with HF who were not dispensed LD was 1.1 (95% CI: 0.96 - 1.2; p-value = 0.21), and for those with HF and treated with LD was 2.2 (95% CI: 2.0 - 2.3; p-value <0.001) (see Figure 6.5).

6.5 Discussion

This analysis provides further evidence that the presence of repeat LD is a marker of poor prognosis, but now in a cohort with well-defined heart disease (i.e., IHD) rather than a cohort predominated by those with hypertension and without another well-established cardiac disease. A multi-variable model, adjusted for age, sex, and other prognostic markers, suggests that mortality is associated with the use of repeat LD regardless of HF status, rather than with the diagnosis of HF only. Death was preceded by the presence of repeat LD dispensing in the community setting in about two-thirds of cases. On the other hand, death was rarely preceded by diagnosed HF unless the patient also received a repeat LD. Based on patients' status in 2013, those with repeat LD only experienced substantially higher rates of ACM and CV mortality compared with

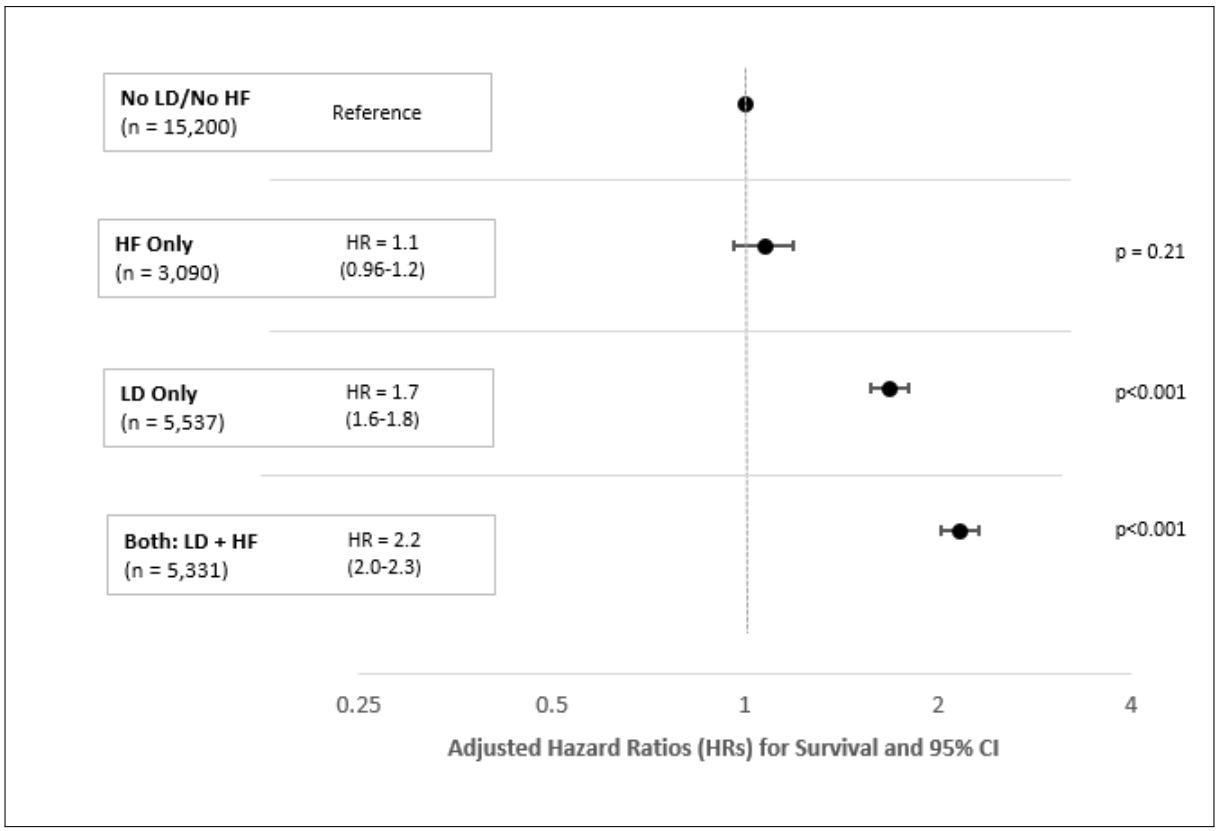


Figure 6.5: Forest plot of HR with 95% CI for ACM by time-dependent group between 1st January 2013 through 31st March 2018. The model was adjusted for age, sex, SIMD, time-dependent comorbidity status, and the lowest haemoglobin and highest serum creatinine results in the preceding two years.

patients with a diagnosis of HF only. Patients with both a diagnosis of HF who were receiving repeat LD had the highest ACM and CV mortality of all groups.

Within the four groups, while never the most common cause of death, deaths due to neoplasms, predominately cancer, made up a substantial subset. The absolute rate of neoplasm-related deaths was somewhat higher in groups who were on repeat LD at baseline compared with those who were not. Previous analyses have suggested that HF is associated with an increased risk of cancer (Roderburg et al. 2021, Hasin et al. 2016), although one should be sceptical about the existence of a causal relationship (Anker et al. 2020). Instead, it's easier to believe that HF and cancer share common causes, such as older age, smoking, hypertension, DM, and clonal hematopoiesis of indeterminate potential (Meijers & de Boer 2019), or that cancer causes inflammation, anaemia, and increases the demands on the heart. Additionally, chemotherapy and radiation therapy may cause myocardial damage (Shakir & Rasul 2009), hypertension (Cohen et al. 2023), and anaemia (Groopman & Itri 1999), which can lead to HF. Finally, increasing awareness of the effects of cancer and cancer treatment on heart function might increase the rate of investigations and, consequently, diagnosis of HF and initiation of LD. However, the possibility that HF increases the risk of cancer by inducing inflammation cannot be completely

discounted. However, if this is the case, within this analysis, the cancer burden appears to be higher in patients with a repeat LD than in those with HF only.

For the groups which had IHD and were treated with repeat LD at baseline, a substantial proportion of the deaths were due to infection, which broke down predominantly to one of respiratory, urinary tract infections (UTI), or septicaemia. This is consistent with the clinical epidemiology of HF (Shen et al. 2021), as in particular, congestion within the lungs may increase the susceptibility of the upper respiratory airways to allow infection to progress into the lung tissue (i.e., developing pneumonia). In addition, UTIs are highly prevalent in patients with prostatic disease or DM (Storme et al. 2019). Additionally, infection as the underlying cause of death might be more common as HF may be associated with impaired immune response and the patient's ability to fight off infections (Strassheim et al. 2019).

Most patients with IHD and at least one of a diagnosis of HF or a repeat LD (or both) had contact with a hospital in the year prior to inclusion in this cohort. The age of patients treated with repeat LD was similar whether or not they had a diagnosis of HF. Still, those who did not have a diagnosis of HF on cohort inclusion were more likely to be women and less likely to have had input from cardiology specialists. The comorbidity profile was also similar for these two groups except for cardiac-related comorbidities (i.e., MI, AF/AFL, valve disease). The lower percentages of contact with a cardiology specialist might reduce the likelihood of receiving a diagnosis for these. Still, if these patients genuinely had fewer such conditions, they might be less likely to be referred to a cardiology service. One would expect that both issues are present. Ultimately, lower percentages of contact with cardiology specialists may account for missed diagnosis of HF.

LD usage is associated with the activation of the renin-angiotensin-aldosterone system (RAAS) (Mentz et al. 2015) and a decline in eGFR (Damman et al. 2016). A low eGFR is associated with iron deficiency and anaemia (Graham et al. 2022). Both are probably mediated through an inflammatory response and increase with hepatic hepcidin production, which blocks iron absorption (Masini et al. 2022). Within the IHD population, at baseline, patients with a diagnosis of HF and repeat LD had decreased renal function with a median eGFR of 37 mL/min/1.73m^2 compared with patients with LD only. They were also more likely to be anaemic, which suggests that the patients with repeat LD only might have less advanced disease.

Within the IHD cohort, compared with patients with repeat LD only, those who also had a diagnosis of HF were more likely to be dispensed treatments that have been shown to reduce mortality for patients with HFrEF, including ACEi or ARB, beta-blocker, and MRA. Additionally, this is while they were less likely to receive treatments that have an adverse effect on prognosis for HF patients such as diltiazem and verapamil (McDonagh et al. 2021, Joint Formulary Committee 2019). However, levels of LA dilation, which may be the best echocardiographic barometer of cardiac function (Cleland et al. 2021), were similar for patients with repeat LD, whether or

not they had a diagnosis of HF. The high levels of LA dilation and higher LVEF values are consistent with a population with high levels of HFpEF/HFsnEF in those treated with LD. The lack of a diagnosis of HF may be due to the greater levels of diagnostic uncertainty concerning these phenotypes, which deviate from the classic reduced LVEF phenotype; consequently, patients are less likely to receive a diagnosis of HF (Huusko et al. 2020).

Patients with both a diagnosis of HF and repeat LD were more likely to receive digoxin, which is consistent with the higher rate of AF/AFL. Almost all of the patients were either on low dose aspirin or an oral anticoagulant, in addition to lipid regulating therapy, which is consistent with a diagnosis of IHD. Patients on a repeat LD, with or without a diagnosis of HF, had similar levels of dispensed bronchodilators (around 30%), which is a pharmacological marker of respiratory disease.

Within the IHD cohort, there were few substantial differences between patients with HF only and those with neither a diagnosis of HF nor a repeat LD. Patients with a diagnosis of HF were more likely to be men, to have a history of of MI, to have a lower LVEF (median of 53% versus 64%), and to receive an ACEi or ARB and beta-blocker; otherwise, the two groups were similar, including LA diameter. Due to these similarities, it is possible that many of the HF only patients received a diagnosis of HF because they had an MI and a reduced LVEF, where measurements were available, rather than because these patients developed evidence of congestion.

6.5.1 Strengths and Limitations

As has been the case throughout this thesis, one of the major limitations is the lack of complete GP record coverage combined with the restricted information held within GP LES (see Section 2.5.3). In particular, this analysis lacks blood pressure, BMI, and smoking status information. These are all known contributors and accelerating factors for the development of IHD and HF. Additionally, while the GSH/18/CA/002 dataset contains information on outpatient attendance, it does not contain cardiac diagnostic information (see Section 2.5.5). Due to the incomplete EPR, it is possible that some diagnoses of HF in the primary care or outpatient settings were missed. Two of the strengths contained within this dataset are the complete set of blood test results (see Section 2.5.9) and prescribing/dispensing records (see Section 2.5.4).

Another limitation is that echocardiogram and ECG data will often not be included in the EPR when the tests were done outwith NHS GG&C. Additionally, tests were not recorded where the electrocardiograph was not connected to General Electric (GE)'s MUSE Cardiology Information System (see Section 2.5.6) or where the echocardiogram data was not uploaded to Image Vault (see Section 2.5.7).

Finally, blood tests for natriuretic peptides, biomarkers which can be used to rule out the presence of cardiac dysfunction, were rarely available.

6.5.2 Future Work

As it stands, there are at least two areas which require further investigation: the first is replicating the analysis for the group of patients within the GHS/18/CA/002 dataset who do not have a diagnosis of IHD, and the second is incorporating parametric survival models into the analysis. Replicating the analysis in patients who met the inclusion criteria set out in Section 3.3.1 by the 1st January 2013 but did not have a history of IHD is important because if patients with a repeat LD only are still at an increased risk of ACM, whether or not they have a diagnosis of IHD, as this will lend credence to the argument that these patients have underlying heart disease.

Incorporating parametric survival models into the analysis is beneficial as the assumptions inherent in the Cox PH regression model used in this chapter were not met for several covariates. In particular, the PH assumption was not met for the first 36 days of follow-up for the patient group covariate, nor was the linearity assumption met for the effects of serum creatinine. While parametric survival models require variables to meet a baseline parametric distribution, included are accelerated failure time (AFT) models, which assume that the failure rate accelerates or decelerates over time with a constant factor. In addition to potentially meeting assumptions, parametric models which use an AFT are an appealing avenue for further analysis as the survival and hazard functions are specified (unlike in a Cox PH regression model). They are more robust to omitted covariates (Keiding et al. 1997), and the AFT value is easier to interpret as it is based on the survival curve rather than the hazard function (Kleinbaum et al. 2012).

6.6 Conclusion

This analysis suggests that patients with IHD treated with repeat LD are rather similar whether or not they receive a diagnostic label of HF. Those with a diagnosis of HF are more likely to have a reduced LVEF, a low eGFR, and anaemia, all of which suggest more advanced disease. These patients are more likely to be referred to a cardiologist. Ultimately, they have the worst prognosis overall. Patients treated with repeat LD who were not diagnosed with HF may have had a higher prevalence of HFpEF/HFsnEF. On the other hand, patients with IHD who were not treated with LD had similar characteristics whether or not they had a diagnosis of HF apart from their history of MI and reduced LVEF, which were more common in patients with a diagnosis of HF. However, the prognosis of patients with IHD who were not on a repeat LD was similar whether or not they had a diagnosis of HF.

In summary, patients with IHD and repeat LD appear similar in many respects whether or not they received a diagnosis of HF. Patients who had IHD and a diagnosis of HF, but were not on a repeat LD, appear to be more alike to patients with neither a diagnosis of HF nor on repeat LD. For patients with IHD, mortality is associated with the use of repeat LD rather than with the diagnosis of HF. This suggests that the presence of a repeat LD, rather than the diagnostic label

of HF, may be a better measure of congestion and prognosis within this cohort. In other words, the repeated use of LD may be a better marker of HF than the presence of ICD-10 or Read codes. If this continues to hold under further analysis and validation studies, the implications for the epidemiology and clinical management of HF are numerous.

Chapter 7

Discussion

> Only the careless leave a possibility
> unattended due to assumptions.
>
> *Laurie R King*
> *Locked Rooms 2006*

The findings in this thesis suggest that the diagnosis of HF is more often missed than made, and when made, often made in error.

The fundamental problem may be in the current definition of HF, which relies on the subjective opinions of both the patients and healthcare professionals. Currently, whether or not the diagnosis of HF is made likely depends on if the patient is seen by someone with expertise in this field (often a cardiologist) and how the patient presents (e.g., with an MI). These factors will impact whether tests for cardiac function are done, combined with the fact that the interpretation of these investigations is subjective (e.g., an LVEF <50% is not sufficient to fulfil the diagnostic criteria for HF on its own[1]). A low LVEF in the absence of evidence of congestion is not associated with a markedly poor prognosis (Yeboah et al. 2012), unlike HF regardless of LVEF (Curtis et al. 2003).

It appears that whatever triggers the prescription and repeat dispensing of LD is associated with a poor outcome. The most obvious explanation for repeat dispensing is for treating congestion that, if associated with cardiac dysfunction, fulfils the definition of HF (McDonagh et al. 2021). However, symptoms and signs of congestion are still subjective, while the repeated dispensing of LD is an objective marker of disease. Alternative objective markers of congestion include plasma concentrations of natriuretic peptides. However, plasma concentrations need to be interpreted in light of the patient's age, sex, eGFR, heart rhythm, and BMI. If plasma concentrations of natriuretic peptides (a blood test that can be done in primary[2] or secondary care) are nor-

[1] Nor is a low LVEF sufficient to rule out a diagnosis

[2] As of 2020, NT-proBNP testing is performed in 12 out of the 13 reporting health boards (Hogg 2020)

mal, then congestion, and therefore, a diagnosis of HF is unlikely. If plasma concentrations are elevated, and an echocardiogram showing atrial dilation confirms the presence of cardiac congestion, this confirms congestive HF (McDonagh et al. 2021). Within this analysis, most patients treated with LD had a dilated LA whether or not they had a diagnosis of HF. Additionally, many of the patients with a diagnosis of HF who were not taking repeat LD (HF only) did not have a dilated LA (although many patients with IHD in all groups [see Chapter 6] exhibited LA dilation indicating some cardiac stress).

When the diagnosis of HF (an opinion) is associated with repeated dispensing of LD (a fact), where the repeated dispensing suggests the presence of symptoms and signs of congestion (opinion), the prognosis is poor. This combination of characteristics provides considerable confidence that the diagnosis of HF is probably correct. However, a diagnosis of HF in the absence of LD treatment is suspect. In contrast, repeated prescription of a LD in the absence of a diagnosis of HF should raise the suspicion that the diagnosis has been missed.

Patients receiving LD without a diagnosis of HF were more likely to be women and less likely to have recently seen a cardiologist and, therefore, were less likely to have had their cardiac function investigated. They were also less likely to receive treatments other than the LD for CV disease. Additionally, they were more likely to have an LVEF $\geq$50% (with many >70%) compared with patients with a diagnosis of HF. This difference might be attributable to diagnostic omission. Patients with both a diagnosis of HF and repeat LD had only a marginally worse prognosis than patients with LD only not labelled as having a diagnosis of HF. Others have also found that women are less likely to be investigated and treated for CV disease in general (Peters et al. 2018, Hyun et al. 2017, Woodward 2019), and HF in particular (Bozkurt & Khalaf 2017).

From a data science perspective, the complexity of modelling the temporal aspect of disease and patient progression were factors which needed to be accounted for and influenced the planning of analyses and how results were displayed. Beyond the patient's baseline characteristics, comorbidities, medications, blood tests, and results of other investigations on cohort or study inclusion, nothing was fixed, and capturing these changes was important for investigating patient prognosis. In a population with a high mortality rate such as HF, developing HF drastically changes a patient's overall survival expectations, including expected increases in hospital admissions and a decrease in quality of life. A static analysis from baseline fails to capture these changes, and updating factors with time can be difficult to report, display, and present such that the underlying message is easily understood while balancing the analytical requirements.

7.1 Loop Diuretic Dispensing

Each analysis in this thesis shows that regardless of how the population was defined, whether considering prevalent or incident cases, and even when restricting to those with IHD, the number

of patients receiving repeat LD was two or three times higher than the number of patients with a diagnosis of HF. Only 25% of the patients taking repeat LD received a diagnosis of HF, and up to 40% of patients with HF were not initiated on repeat LD. However, patients with repeat LD had many similar characteristics whether or not they had a diagnosis of HF (apart from LVEF and cardiology specialist input). Those with both had a slightly worse prognosis than the LD only group. In contrast, patients who were not on repeat LD but had a diagnostic label of HF were more similar in terms of characteristics and outcomes to the rest of the cohort of a similar age with a broad range of CV disease. In other words, the presence of repeat LD seems to be a better method of stratifying patients with CV disease than an administrative code for HF. Of note, patients initiated on repeat LD well in advance of receiving the diagnosis of HF (Both: LD First) had a particularly poor prognosis, despite applying adjustments to mitigate age and sex differences.

End-stage kidney disease and resistant hypertension might account for some LD use, but the prevalence and incidence of these problems appeared low. Further review of case notes (potentially facilitated by natural language processing) might reveal reasons for LD prescribing and the reasons for a diagnosis of HF in the absence of LD prescribing (Cleland et al. 2011). Clinical 'bedside' research is ongoing in order to investigate the reasons for initiating LD and what proportion of patients have evidence of congestion based on biomarkers and ultrasound. If these patients actually have undiagnosed HF, Chapter 4 shows that many patients lost the opportunity to receive guideline-recommended therapy for patients with HFrEF/HFmrEF including RAASi and beta-blockers. This therapeutic deficit may now be increased since the introduction of two new therapeutic classes, ARNi and SGLT2i, which may also improve CV outcomes for patients with HFpEF (Solomon et al. 2019, McDonagh et al. 2021, Anker et al. 2021). Until recently, it could be argued that it was not a serious matter to miss a diagnosis of HFpEF because treatments were purely for symptoms rather than aiming to reduce hospitalisations and mortality. Now that there is substantial evidence that MRA, ARNi, and SGLT2i improve outcomes, diagnosing HFpEF has become more important. Even more recently, a new class of agent[3] directly specifically at hypertrophic cardiomyopathy (one of the causes of HFsnEF) has been introduced (Olivotto et al. 2020, Bello & Pellegrini 2022), providing a rationale for a separate phenotype with an LVEF >70% as a cause of HF. The efficacy of this agent is being explored in a broad range of HFsnEF (Spertus et al. 2021, Schenk & Fields 2023). Now, regardless of LVEF, missing or delaying a HF diagnosis postpones or misses opportunities to improve patient well-being and outcomes (Solomon et al. 2019, McDonagh et al. 2021).

On the other hand, for those with a diagnosis of HF, guidelines strongly recommend LD therapy for the management of symptoms and signs of congestion (McDonagh et al. 2021). Not prescribing a LD suggests that these patients had few or no signs of congestion and, therefore,

[3]The first-in-class molecule is called Mavacamten.

did not fulfil the conventional guideline criteria for a HF diagnosis. Moreover, Chapters 3 and 6 show that the prognosis of those patients was fairly similar to the reference group (those without HF nor LD).

Further research, including prospective research, is needed to determine why treatment with LD has such an adverse impact on prognosis. LD could be a marker of congestion, a powerful driver of disease progression and risk. On a rudimentary level, the syndrome of HF is congestion due to cardiac dysfunction (Cleland et al. 2021). Alternatively, if the LD is inappropriately used, one wonders if it adversely affects patient outcomes by causing RAAS activation, electrolyte disturbances, and renal dysfunction.

7.2 Sex Differences

Differences in the distribution of the sexes were pronounced in groups where HF or repeat LD appeared alone. Those receiving repeat LD without a diagnosis of HF were more likely to be older women. In contrast, those with a diagnosis of HF and not receiving LD were more likely to be younger men. Men are more likely to have an MI at a younger age, which is likely to flag them for cardiology specialist attention, in addition to reducing the LVEF (de Miguel-Yanes et al. 2021). Men, with or without IHD, are also more likely to have HFrEF (Lam et al. 2019). Currently, many cardiology services are designed to focus on this group of patients and to neglect and overlook (until now) patients with HFpEF for the reasons given above. Women are more likely to develop HF due to chronic hypertension, AF, iron deficiency anaemia, and renal dysfunction; additionally, they are more likely to have HFpEF, which is characterised by an LVEF 50%-70% (HFpEF) or >70% (HFsnEF), to have a dilated LA (Lam et al. 2019), which was commonly observed in the LD only group. Patients with both a diagnosis of HF and repeat LD had a broad range of LVEF, with a similar number of men and women in the group.

Across the analyses, many of the LD only patients had hypertension, DM, anaemia, AF/AFL, and impaired kidney function, which are common comorbidities and are known to cause or exacerbate HF. Due to the high percentages of older women with high levels of LA dilation in the LD only groups, these profiles are suggestive of HFpEF.

For those with a diagnosis of HF in the absence of the strongly recommended LD therapy (McDonagh et al. 2021), the missing treatment for congestion (National Institute for Health and Care Excellence 2018*a*) suggests that these patients either had few or no signs of the condition and therefore did not fulfil the conventional guideline criteria for HF. Moreover, Chapters 3 and 6 showed that the prognosis of these patients was either not statistically worse or only slightly worse than the reference patients.

7.3 Complexity of Modelling Temporal Factors

Throughout this thesis, patient classification based on the existence and order of a diagnosis of HF and repeat LD was vital to building an overall picture showing the size of the problem and how patients progressed to and through health states. Each chapter took pains to avoid immortal time bias (see Section 4.2.2) while still describing meaningful relationships and subsequent events in an understandable way. Chapters 3 and 6 used time-dependent covariates to account for changes in risk associated with developing HF and to account for initiating repeat LD. Chapter 5 went a step further by using multi-state models to incorporate the ordering of the diagnosis of HF and initiation of repeat LD into predicting outcomes, which is much easier to fit within the clinical understanding of HF and existing clinical practice.

For Chapters 3, 5, and 6, the decision was made to report the number of patients who passed through a given state as a percentage of the baseline group and to report the number of people who experienced a given transition as the percentage of those eligible to experience said transition. This was done because, outside of the inclusion and terminal states, patients transitioned in and out of the states throughout follow-up. Reporting the number of patients in a given state at a given time would have been a Sisyphean task to calculate, report, or display coherently. For similar reasons, Chapter 4 calculated changes in comorbidities and prescription levels one year after inclusion based on the baseline population size.

7.4 Strengths and Limitations

One strength of this thesis is the data used in Chapters 3-6, which are from a large regional, real-world cohort drawn from a population which accounts for about 23% of the entire Scottish population (National Records of Scotland 2018). This dataset captures numerous variables throughout a patient's interaction with the NHS, including attendance at outpatient clinics, hospital admissions, haematology and biochemistry results, dispensed prescriptions, and, uniquely, measurements from ECG and echocardiograms. Additionally, the cohort was defined to allow for the identification of LD usage, regardless of the patient's other diagnosed comorbidities. Due to this definition, this thesis was able to show a more realistic size disparity between long-term LD usage compared with those with a diagnosis of HF. The similarities of the codes used to define HF to other observational secondary cohorts (Conrad et al. 2018, 2019) provide some external validation to these results. This is backed up by similar patterns of results found in other research (Zakeri et al. 2021, Pellicori et al. 2021). Nevertheless, further validation and prospective analysis are required. In particular, to understand why patients are on a repeat LD prescription in the absence of a diagnosis of HF. Based on the available clinical variables, one strongly suspects a missed diagnosis of HF. However, this cannot be confirmed or ruled out without further testing, such as natriuretic peptide concentrations and echocardiograms. With regard

to validating results shown in Chapters 3, 5, and 6, CPRD data could be used for validation and to generalise the results further.

While the dataset is a strength in terms of its size, coverage, and longitudinal nature, it lacked in-hospital medications, outpatient diagnostic codes, and information on well-known contributing risk factors for HF including blood pressure, smoking status, and BMI.

7.5 Future Work

There are several opportunities for further work looking at the relationship between LD dispensing, the diagnosis of HF, and their relationship on outcomes and the wider implications. Throughout this thesis, the analysis of LD dispensing focused only on reported LD usage once the minimum repeat prescription threshold was met. The impact of this is two-fold. First, as mentioned in Section 5.5.2, this thesis does not investigate the consequences of patients interrupting or terminating LD treatment. However, this may be uncommon once the patient is established on repeat LD dispensing. Stopping repeat LD therapy may lead to the reappearance of congestion, potentially to a HF admission. On the other hand, the improvement of cardiac function due to disease-modifying therapies may allow LD therapy to be stopped safely and should be associated with improved outcomes. Second, due to the minimum dispensing threshold, this thesis under-reports the number of patients with dispensed LD within NHS GG&C. If LD prescription is a marker for disease, then it is possible that even a single prescription could trigger concerns about diagnosis and prognosis.

Third, due to the limitations of PIS and the challenges of determining daily dosage (see Section 2.5.4), the decision was made to focus on patients who met the minimum threshold of being on a repeat LD prescription. From early exploratory analysis, patients who did not meet this threshold account for 70% of all patients receiving a dispensed LD within a community-based pharmacy within NHS GG&C. Further work is required to understand these patients, and if one-off or intermittent LD treatment is related to HF, and if these patients have a poorer prognosis than those who have never had LD dispensed.

Fourth, there are multiple places within this thesis where applying other analysis techniques would be beneficial. For example, as mentioned in Chapter 4, implementing sequence analysis to expand on the admissions patterns work presented could provide further insight into subgroups which share similar trajectories toward the diagnosis of HF and/or the initiation of LD above and beyond displaying common patterns and central nodes. Additionally, rather than simply grouping patients based on the presence or absence of repeat LD dispensing and/or diagnosis of HF because it's intuitive, cluster analysis could be beneficial to find subgroups which share distinct risk profiles and potentially provide further targets for both retrospective and prospective analysis, especially for Chapters 3 and 6. For Chapters 3 and 6, the covariates were chosen based

on the literature and expert clinical opinion, but this could be more formally investigated using causal inference, particularly surrounding the use of LD. Finally, another lens to the analysis conducted within Chapter 6 would be to leave the frequentist realm and implement a Bayesian approach. The use of independent normal priors can readily accommodate covariate measurement error and missing data, including where the actual value is assumed to have a non-linear relationship to the outcome (Bartlett & Keogh 2018), which was seen with the serum creatinine values in relation to mortality.

Finally, this analysis was conducted within a single regional population, where most of the population is white with high levels of socioeconomic deprivation and CV disease. Further work is required to see if similar patterns in prescription and diagnostic levels exist and if those with a repeat LD still have higher morbidity and mortality rates in larger, more diverse populations. To this end, ethical approval and data have been acquired from CPRD to see if these patterns hold across a representative set of the English population. Suppose these patterns hold within England and ideally can be replicated across the UK. In that case, the natural progression is to seek to replicate the analysis in a diversified set of countries and healthcare systems If there are similar disproportional relationships between LD dispensing and the diagnosis of HF and those on a LD have similarly poor outcomes across multiple countries and healthcare systems, one could argue for a change to the definition of HF to include the use of LD rather than the neither specific nor objective symptoms and signs currently required by the ESC. Even without these additional analyses, these analyses should be convincing enough to prompt NHS GG&C, and potentially PHS, to investigate policy changes requiring further medical investigations in those who are initiated on a LD in the absence of a diagnosis of HF, impaired renal function, or resistant hypertension.

Chapter 8

Conclusions

> Begin thus from the first act, and proceed; and, in conclusion, at the ill which thou hast done, be troubled, and rejoice for the good.
>
> *Pythagoras*

8.1 Summary of Chapters

The contents of this thesis have sought to address the overall aim and objectives presented in Section 1.8.1. Chapter 3 described the prevalence of people receiving LD therapy and/or had a diagnostic label of HF, including their characteristics, comorbidities, concurrent medications, and 5-year hospitalisations and mortality compared with patients with a broad range of CV disease (predominantly hypertension and CAD) (Objective 1). This analysis highlighted the discrepancy in the levels of prevalent LD usage comparison to the prevalent levels of HF diagnosis. The majority of patients with repeat LD only were women who had similar characteristics to the 'typical' HFpEF patient. In contrast, the majority of the HF only patients were men, and many had a history of MI. Finally, the risk of mortality was higher for those with LD than for a diagnosis of HF only.

Chapter 3 reported the 'steady-state' of dynamics of HF diagnosis and chronic LD usage. Chapter 4 stepped backwards in the chronology of disease progression to describe the pattern, nature, and frequency of hospital admissions in the year before patients initiated repeat LD therapy or received a diagnosis of HF (Objective 2). It highlights that approximately 50% of new HF and LD patients are admitted to hospital in the year before diagnosis or initiation, and the differences and similarities in the pathways to incident HF or the initiation of LD. In particular, the pattern of events leading to a diagnosis of HF is expected based on existing epidemiology studies. The

results in this chapter are novel in that this is the first instance where the pattern of events leading up to the initiation of LD have been reported. They are very different to those seen in the HF population, with many being admitted for unspecific symptoms and signs, particularly un-specified chest pain.

From events leading up to diagnosis or therapy initiation, Chapter 5 reports on how the ordering of events (initiation of LD therapy and HF diagnosis) have on hospital admission and mortality rates (Objective 3). In particular, patients starting a LD had higher admissions and mortality rates than those with HF only. Even after adjusting for age and sex, mortality was highest in those where the diagnosis of HF occurred after the initiation of LD therapy.

Rather than describing the prevalent and incident HF and LD populations, Chapter 6 focused on the outcomes of the IHD patients in relation to the use of LD and a diagnosis of HF to address Objective 4. In a population at an already increased risk of mortality, the risk of ACM is associated with the LD rather than with the diagnosis of HF only after adjusting for comorbidities and most severe haemoglobin and serum creatinine values.

8.2 Final Conclusions

The analyses contained within this thesis are important. It identifies that many thousands of patients in NHS GG&C are treated repeatedly with LD but do not have a diagnosis of HF. They have a high risk of hospitalisation and death. Prescribing and dispensing of LD can be readily identified in routinely collected NHS data. Based on patient characteristics, many of these patients probably have undiagnosed HF for which treatments exist which improve symptoms and quality of life, reduce hospitalisation rates, and reduce the risk of mortality. Validating these analyses in a wider CPRD dataset acquired from the whole of England might identify up to 1.5 million people taking LD who do not have a diagnosis of HF but are at high risk for morbidity and mortality.

Some of the LD only patients will have other serious conditions, particularly respiratory disease and cancer, that may cause diagnostic confusion with HF, which would count for missed diagnosis. However, these do not appear to account for the majority of cases. It is also possible that LDs are a marker for some unidentified serious medical condition. This possibility needs to be explored.

The final possibility is perhaps even more alarming. Perhaps many of these patients are receiving LD inappropriately, and this is causing the increase in morbidity and mortality rates. Perhaps LD can be safely withdrawn for many patients or substituted with Thiazides+, SGLT2i, MRA, or even with digoxin. All of these have diuretic properties and good safety profiles. Of course, severe renal dysfunction requires cautious use of all CV medications, including these and LD.

Ultimately, clinical trials are required to determine the optimal strategy for the management of congestion.

Appendix A

Supplementary material for Chapter 2

Unless explicitly noted otherwise, three- and four-digit ICD-10 codes include all codes below them. BNF codes include all codes below them.

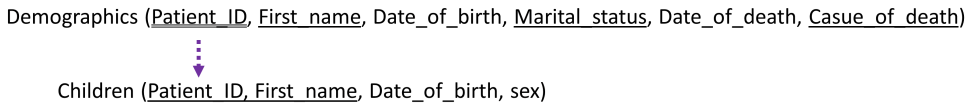
A.1 Codes Used to Define Cohort

Dataset	Code Type	Variable	Code	Description
SMR01/04	ICD-10	PAD	I73	Other peripheral vascular disease
SMR01/04	ICD-10	CAD	I24	Other Acute ischaemic heart disease
SMR01/04	ICD-10	CAD	I254	Coronary artery aneurysm and dissection
SMR01/04	ICD-10	CAD	Z951	Presence of aortocoronary bypass graft
SMR01/04	ICD-10	CAD	Z955	Presence of coronary angioplasty implant and graft
SMR01/04	ICD-10	HF	I50	Heart failure
SMR01/04	ICD-10	HF	I110	Hypertensive heart disease with (congestive) heart failure
SMR01/04	ICD-10	HF	I130	Hypertensive heart and renal disease with (congestive) heart failure
GP LES	Read	PAD	G734	Echocardiogram shows left ventricular systolic dysfunction
GP LES	Read	CAD	G340	Coronary artery disease (CAD)
GP LES	Read	HF	G5yy9	Left ventricular systolic dysfunction (LVSD)
GP LES	Read	HF	585f	Echocardiogram shows left ventricular systolic dysfunction
GP LES	LES Area	HF	HF	Heart failure
PIS	BNF	ACEi	0205051	Angiotensin-converting-enzyme inhibitor (ACEi)
PIS	BNF	ARB	0205025	Angiotensin II receptor blocker (ARB)
PIS	BNF	Beta-blocker	0204	Beta-blocker
PIS	BNF	LD	020202	Loop diuretics (LD)
PIS	BNF	MRA	020203	Mineralocorticoid receptor antagonists (MRA)

SMR01, Scottish Morbidity Records for Acute Inpatient and Day Cases; SMR04, Mental Health Inpatient and Day Cases;
GP LES, General Practice Local Enhanced Services; PIS, Prescribing Information System for community-based prescriptions; LES Area, local enhanced service area; ICD-10, International Classification of Diseases, 10th Revision; BNF, British National Formulary;
PAD, peripheral arterial disease; CAD, coronary artery disease; HF, heart failure; ACEi, angiotensin-converting enzyme inhibitor;
ARB, angiotensin receptor blocker; LD, loop diuretic; MRA, mineralocorticoid receptor antagonists

Table A.1: Database and codes used to determine the size of the cohort using community-based prescriptions, primary care, hospital admission records, and mental health inpatient and day cases care records

A.2 Explanation of Relational Database



Demographics:

Patient_ID	First_name	Date_of_birth	Marital_status	Date_of_death	Cause_of_death
1	Katherine	1485-12-16	Divorced	1536-01-07	Cancer
2	Anne	1501-01-01	Married	1536-05-19	Beheaded
3	Jane	1508-01-01	Married	1537-10-24	Postnatal complications
4	Anne	1515-07-22	Divorced	1557-07-16	Cancer
5	Catherine	1521-01-01	Married	1542-02-13	Beheaded
6	Catherine	1512-01-01	Widow	1548-07-05	Complications of childbirth

Children:

Patient_ID	First_name	Date_of_birth	Sex
1	Henry	1511-01-01	Man
1	Mary	1516-02-8	Woman
2	Elizabeth	1533-09-07	Woman
3	Edward	1537-10-24	Man

Figure A.1: A representation of a two-table relational database with the schematic from above the populated tables.

With all relational databases, the design, structure, and implementation are fundamental to its ability to store and manipulate data in useful ways. In the example presented in Figure A.1, a two-table relational database illustrates how to link a patient's demographic information to their children, if they had any. In this example, patients in the Demographics table may be uniquely identified by their `Patient_ID` or through the combination of `First_name`, `Martial_status`, and `Cause_of_death`. Either case is called a candidate key for the Demographics table and is underlined in the schema. In this case, the `Patient_ID` is the primary way to identify patients throughout the database, indicated by the double underline. The primary key in tables with only one candidate key is identified by a single underline. The `Patient_ID` column in the Children table is called a foreign key as it originates from the `Patient_ID` column in the Demographics table, indicated by the purple arrow.

Using this kind of design strategy, it's easy to see how designing and implementing a relational database fits easily into the structural information held within EPR, where tables exist for hospitalisations, GP records, ECGs, patient demographics, and the like. The unique patient identifier is a natural foreign key to link points of information from disparate data sources.

A.3 Code Used to Define Canonical Date of Death

```
-- Create and populate Death Records from all data
```

```
-- sources with stated deaths
create table working.Deaths60Days5Events (
    safehavenID varchar(3000),
    Location_Code varchar(3000),
    Death_Location varchar(3000),
    Death_Source varchar(3000),
    FlatFileName varchar(3000),
    DOD date,
    );

-- working.DeathRecords contains all records of
-- death across the different datasets.
select
    distinct safehavenID
into #DistinctSHIDs
from working.DeathRecords;

Declare @SH_ID varchar(3000)
Declare @maxDOD date
Declare @countValue int
Declare @deathRecord int
Declare @flatFileName varchar(3000)

while exists(select * from #DistinctSHIDs)
begin
    select
        top 1 @SH_ID=safehavenID
    from #DistinctSHIDs
    select
        *
    into #GroupedShIds
    from working.DeathRecords
        where safehavenID = @SH_ID

    select
        @countValue =  count (*)
    from  #GroupedShIds
```

```
if(@countValue = 1)
begin
    insert into working.Deaths60Days5Events
        (safehavenID, Location_Code, Death_Location,
        Death_Source, FlatFileName, DOD)
    select
        safehavenID, Location_Code, Death_Location,
        Death_Source, FlatFileNameSource, DOD
    from working.DeathRecords
        where safehavenID = @SH_ID;
    print 'there is one safehavenID'
End
else
Begin
/* Prefer NRS Deaths- use NRS record if it exists */
    select
        @countValue = count (*)
    from #GroupedShIds
        where death_source like 'NRS%'

    if (@countValue > 0)
    begin
        insert into working.Deaths60Days5Events
            (safehavenID, Location_Code, Death_Location,
            Death_Source, FlatFileName, DOD)
        select
            top 1 safehavenID, Location_Code,
            Death_Location, Death_Source,
            FlatFileNameSource, DOD
        from working.DeathRecords
            where
                working.DeathRecords.safehavenID = @SH_ID
                and working.DeathRecords.death_source
                = 'NRS_DEATHS'
        print 'there is an NRS record'
    end
    else
    begin
```

```
        select
            @deathRecord = count(*)
        from #GroupedShIds
            where FlatFileNameSource like 'Death%'
        /* Use most recent DOD from Deaths if it exist*/
        if(@deathRecord > 0)
        begin
        select
            @maxDOD = max(DOD)
        from #GroupedShIds
            where FlatFileNameSource like 'Death%'

        select
            @flatFileName = flatFileNameSource
        from #GroupedShIds
            where FlatFileNameSource like 'Death%'
            and DOD = @maxDOD

        insert into working.Deaths60Days5Events
            (safehavenID, Location_Code, Death_Location,
            Death_Source, FlatFileName, DOD)
        select
            top 1 safehavenID, Location_Code,
            Death_Location, Death_Source,
            FlatFileNameSource, DOD
        from working.DeathRecords
            where working.DeathRecords.safehavenID
            = @SH_ID
            and working.DeathRecords.DOD = @maxDOD
            and working.DeathRecords.FlatFileNameSource
            = @flatFileName

        print 'Picking max death record value'
    end
    /*finally, choose the most recent DOD if no death
    record exists*/
    else
    begin
```

```
            select
                @maxDOD = max(DOD)
            from #GroupedShIds

            insert into working.Deaths60Days5Events
                (safehavenID, Location_Code, Death_Location,
                Death_Source, FlatFileName, DOD)
            select
                top 1 safehavenID, Location_Code,
                Death_Location, Death_Source,
                FlatFileNameSource, DOD
            from working.DeathRecords
                where working.DeathRecords.safehavenID = @SH_ID
                and working.DeathRecords.DOD = @maxDOD

            print 'picking max value'
        end
    end
end

drop table #GroupedShIds
delete #DistinctSHIDs where safehavenID = @SH_ID
end
drop table #DistinctSHIDs
```

A.4 Code Used to Identify Mislinked Records

```
-- Gather all records which occurred more than 60 days
-- after death
-- Example of code for extracting SMR01 admissions
insert into working.EventsAfterDeath60Days5Events (
    safehavenID, DOD, Death_Source,
    DeathSourceFlatFileName, Event_After_DOD_Date,
    Even_Description, EventSourceFileName)
Select
    d.safehavenID, DOD, death_source,
    d.flatfilename, smr01.ADMDATE,
    ADMTYPEDesc, 'SMR01'
```

```
from working.Deaths60Days5Events d inner join
    cleaning.SMR01 smr01
    on d.safehavenID = smr01.safehavenID
    where dateadd(day, 60, d.DOD) < smr01.ADMDate;

-- Repeat above for all data sources, adding records to
-- working.EventsAfterDeath60Days5Events

-- Select all safehavenIDs where patients have more than
-- 5 distinct event days after 60 days post canonical DOD

select *
into mislinkedRecords.MiscodedDeaths60Days5Events
from working.EventsAfterDeath60Days5Events
    where exists (
    select
        safehavenID
    from working.EventsAfterDeath60Days5Events
        group by safehavenID
        having count(distinct event_after_DOD_date) > 5);
```

A.5 Classification of Ethnicity

Table A.2: Ethnicity classifications used to group the ethnicity values recorded by the various data sources.

Ethnicity Classification	Recorded Ethnicity
Asian	'Any other Asian background'
	'Arab'
	'Arab, Arab Scottish or Arab British'
	'Bangladeshi'
	'Bangladeshi, Bangladeshi Scottish or Bangladeshi British'
	'Chinese'
	'Chinese, Chinese Scottish Chinese British'
	'Indian'
	'Indian, Indian Scottish or Indian British'
	'Other - Asian, Asian Scottish or Asian British'
	'Other Asian, Asian Scottish or Asian British'
	'Pakistani'

Continuation of Table A.2

Ethnicity Classification	**Recorded Ethnicity**
Asian	'Pakistani, Pakistani Scottish or Pakistani British'
Black	'African'
	'African, African Scottish or African British'
	'Any other Black background'
	'Black African'
	'Black, Black Scottish or Black British'
	'Black Other'
	'Caribbean'
	'Caribbean, Caribbean Scottish or Caribbean British'
	'Other African'
	'Other - African, Caribbean or Black'
	'Other African, Caribbean or Black'
	'Other Caribbean or Black'
White	'Any Other White background'
	'Any other white ethnic group'
	'British'
	'English'
	'Irish'
	'Northern Irish'
	'Welsh'
	'White'
	'White - British'
	'White - English'
	'White - Gypsy/Traveller'
	'White - Irish'
	'White - Northern Irish'
	'White - Other British'
	'White - Other white ethnic group'
	'White - Polish'
	'White - Scottish'
	'White - Welsh'
Other	'Any mixed or multiple ethnic groups'
	'Any other ethnic background'
	'Other - Other ethnic group'
	'Other ethnic group'
	'Other Ethnic group - Other'

Continuation of Table A.2

Ethnicity Classification	Recorded Ethnicity
Missing	NULL
	'Not Known'
	'Refused' item
	'Refused/ Not provided by patient'

A.6 Codes Used to Identify and Group Medication Chemicals

Table of codes used to identify and group medication chemicals. Prescriptions are classified under a given category when the BNF selection code matches the provided number of characters in the target BNF code.

Table A.3: Prescription classifications used to group the medication chemicals using BNF codes.

Category	BNF Selection Code	Description
ACEi	0205051	Angiotensin-converting enzyme inhibitors
Amiodarone/Dronedarone	02030230D	Amiodarone hydrochloride
	0203020X	Dronedarone hydrochloride
ARB	0205052	Angiotensin receptor blockers
	0206020Z0	Valsartan/Amlodipine
ARNi	0205052AE	Sacubitril/Valsartan
Anti-arrhythmic	0203020C	Adenosine
	0203020F	Disopyramide
	0203020G	Disopyramide phosphate
	0203020I	Flecainide acetate
	0203020L	Lidocaine hydrochloride
	0203020P	Mexiletine hydrochloride
	0203020S	Procainamide hydrochloride
	0203020R	Propafenone hydrochloride
	0203020U	Quinidine sulfate
Antiplatelet	0209	Antiplatelet
Low dose aspirin	0202000A0	Aspirin
	0204000AC	Bisoprolol fumarate/aspirin
	0209000V0	Dipyridamole and Aspirin
Beta-blocker	0204	Beta-adrenergic blocking agents
CCB	020602	Calcium channel blockers
	0205052AB	Olmesartan medoxomil/amlodipine

Continuation of Table A.3

Category	BNF Selection Code	Description
CCB	0205052AC	Olmesartan medoxomil/amlodipine/ hydrochlorothiazide
Bronchodilators	0301	Bronchodilators
Hypoglycemic therapy	060101	Insulin
	060102	Anti-diabetic drugs
	0601023AM	SGLT2i
	0601023AP	
	0601023AG	
	0601023AL	
	0601023AV	
	0601023AN	
	0601023AR	
	0601023AX	
Digoxin	0201010F	Digoxin
Lipid-regulating	0212	Lipid-regulating
Loop diuretic	020202	Loop diuretics
	0202040D0	Amiloride HCI with loop diuretics
	0202040B0	Co-amilofruse (Amiloride hydrochloride/ frusemide)
	0202040T0	Spironolactone with loop diuretics
	0202040U0	Triamterene with loop diuretics
	0202080D0	Bumetanide/Amiloride hydrochloride
	0202080C0	Bumetanide/potassium
	0202080K0	Furosemide/potassium
MRA	0202030X0	Eplerenone
	0202030S0	Spironolactone
	0202040G0	Co-flumactone (Hydroflumethiazide/ spironolactone)
	0202040T0	Spironolactone with loop diuretics
	0202040S0	Spironolactone with thiazides
NSAIDs	100101	Non-steroidal anti-inflammatory drugss
Oral anticoagulants	020802	Oral anticoagulants
Thiazides+	020201	Thiazides and related

Continuation of Table A.3

Category	BNF Selection Code	Description
Thiazides+	0202040H	Co-triamterzide (Triamterene/hydrochlorothiazide)
	0202040C0	Co-amilozide (Amiloride hydrochloride/ hydrochlorothiazide)
	0204000Y0	Co-prenozide (Oxprenolol hydrochloride/ cyclopenthiazide)
	020400040	Co-tenidone (Atenolol/chlortalidone)
	0205051AB	Perindopril tosilate/indapamide
	0202040A0	Amiloride hydrochloride with thiazides
	0202040S0	Spironolactone with thiazides
	0202040V0	Triamterene with thiazides
	0202080B0	Bendroflumethiazide/potassium
	0205052Y0	Olmesartan medoxomil/ hydrochlorothiazide
	020400010	Pindolol with diuretic
	020400030	Timolol with diuretic
	0204000F0	Atenolol with diuretic
	0204000Q0	Propranolol hydrochloride with diuretic
	0204000W0	Metoprolol tartrate with diuretic
	0205051H0	Enalapril maleate with diuretic
	0205051K0	Lisinopril with diuretic
	0205051N0	Perindopril erbumine with diuretic
	0205051P0	Quinapril hydrochloride with diuretic
	0205051Z0	Perindopril arginine with diuretic
	0205052A0	Irbesartan with diuretic
	0205052P0	Losartan potassium with diuretic
	0205052R0	Telmisartan with diuretic
	0205052X0	Valsartan with diuretic
	0205052AC	Olmesartan medoxomil/amlodipine/ hydrochlorothiazide
	0205051G0	Co-zidocapt (Hydrochlorothiazide/captopril)

Continuation of Table A.3

Category	BNF Selection Code	Description
Thyroid disorders	060201	Thyroid hormones
	0602020D	Antithyroid drugs
	0602020G	
	0602020N	

ACEi, Angiotensin-converting enzyme inhibitor; ARB, Angiotensin receptor blocker; ARNi, Angiotensin receptor-neprilysin inhibitor; CCB, Calcium channel blocker; SGLT2i, Sodium-glucose co-transporter-2 inhibitor; MRA, Mineralocorticoid receptor antagonists; NSAIDs,Non-steroidal anti-inflammatory drugs; Thiazides+, Thiazides and related.

A.7 Code Used to Define Repeat Prescriptions

SQL Server code used to identify repeat loop diuretic prescriptions and the associated start date.

```
-- safehavenID is the anonymised patient identifier
-- PRESC_DATE is the date of prescription
-- DISP_DATE is the date that the medication was dispensed
-- PI_BNF_Item_Code is the BNF Item code down to dosage amount
-- and formulation (e.g., tab, cream, etc)

--** Step 1: Create a table with all drugs containing any loop
-- diuretic
select distinct
    safehavenID, PRESC_DATE, DISP_DATE, PI_BNF_Item_Code
into #allLD
from [working].[Pharmacy_DrugsOfInterest]
    where ([category] like 'Loop%') and
          DISP_DATE <= '2018-04-01'


 --** Step 2: Add indexes to allow for efficient lookups
 -- and referencing.
 -- safehavenID index
CREATE NONCLUSTERED INDEX [NonClusteredIndex-20190416-1]
    ON #allLD
(
```

```
    [safehavenID] ASC
) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
    SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
GO

-- DISP_DATE index
CREATE NONCLUSTERED INDEX [NonClusteredIndex-20190416-2]
    ON #allLD
(
    [DISP_DATE] ASC
) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
  SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
  ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
GO

--** Step 3: Select individuals who died within 90 days
-- of the first prescription of a loop diuretic

-- select first prescriptions
select *
into #firstLD
from  (
    select safehavenID, PRESC_DATE, ROW_NUMBER()
    over (partition by safehavenID order by PRESC_DATE) rn
from #allLD
) t
    where rn = 1

-- Select individuals who died within 90 days of first
-- prescription
select
    m.safehavenID, dod, PRESC_DATE
into #dodWithin90Days
from #firstLD m inner join
working.Deaths60Days5Events d
    on m.safehavenID = d.safehavenID
    where DATEDIFF(day, PRESC_DATE, dod) <= 90
```

```
-- clean up
drop table #firstLD

--** Step 4: Identify repeat prescriptions of a loop diuretic
-- over two consecutive quarters

-- Order prescriptions based on DISP_DATE, calculate
-- prevDisp and nextDisp, add row number. Results saved
-- in a new table - #series
select
    safehavenID, PRESC_DATE, DISP_DATE,
    lead(DISP_DATE, 1) over (partition by safehavenID order
    by DISP_DATE) nextDisp,
    lag(DISP_DATE, 1) over (partition by safehavenID order
    by DISP_DATE) prevDisp,
    ROW_NUMBER () over (order by safehavenID, DISP_DATE) rn
into #series
from #allLD

-- add series and newSeries indicator
alter table #series add series int
-- indicates if a row starts a new series
alter table #series add newSeries bit -- 0 - no, 1 - yes
go

-- add index to rn
CREATE NONCLUSTERED INDEX [NonClusteredIndex-20181121-080854]
ON #series
(
    rn ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
 SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
 ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)

GO

-- Add a flag to indicate if the prescription is part of a series
```

```
update #series
set series = case when prevDisp is null and nextdisp is null
                       then 1
                  when prevDisp >=  DISP_DATE
                       then 2
                  when prevDisp <= DATEADD(QUARTER, 1,
                    DISP_DATE)
                       then 2
                  else 1
         end
go

-- Calculate if a record represents a new series
-- 0 - no, 1 - yes
update #series
set newSeries = case when series = 1
                       then 1
                     when prevdisp is null
                       then 1
                     when DISP_DATE > DATEADD(QUARTER, 1,
                      prevDisp)
                       then 1
                     when DISP_DATE <= DATEADD(day, 1,
                      prevDisp)
                       then 0
                     else 0
end
select * from #series order by safehavenID, DISP_DATE

 -- Add a column to indicate the year and quarter that a
 -- prescription is dispensed in
alter table #series add holding int
go

update #series
set holding = datepart(YYYY, DISP_DATE)*100 +
              datepart(quarter, DISP_DATE)
```

```
-- update add linked
--* linked = 1 if the current row is 1-quarter different
-- in the same year
--* linked = 97 if current row is Q1 in new year and
-- previous row is Q4 of the previous year
--* linked = 0 if the current row is the same as the
-- previous row
--* linked is NULL if the first prescription drug based on
-- DISP_DATE
--* linked other - not consecutive quarters
select
    *, holding - lag(holding, 1)
    over(partition by safehavenID order by DISP_DATE)
        as linked
into #series2
from #series
    order by safehavenID, DISP_DATE


-- Count the number of consecutive quarters
alter table #series2 add numQuarter int
go

-- Create a table to iterate over
select rn into #rows from #series2
SET ANSI_PADDING ON


GO
-- Add indexes to speed up selection
CREATE NONCLUSTERED INDEX [NonClusteredIndex-20181121-080866]
ON #rows
(
    rn ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
 SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
 ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
```

```
GO

CREATE NONCLUSTERED INDEX [NonClusteredIndex-20181121-080867]
ON #series2
(
    rn ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
 SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
 ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)

GO

-- Go row by row to identify prescriptions which fall
-- within consecutive quarters
Declare @rowNum int
Declare @quartIndex int

set @quartIndex = 1
while exists (select * from #rows)
begin
    select top 1 @rowNum = rn from #rows

    update #series2
    set numQuarter = case when linked is null
                           then 1
                         when linked is null and
                          nextDisp is null
                           then 1
                        -- first dispensed prescription
                         when prevDisp is null
                           then 1
                        -- same quarter
                         when linked = 0
                           then @quartIndex
                        -- next quarter within a year
                         when linked = 1
                           then  @quartIndex +1
                        -- continuing series from last year
```

```
                            when linked = 97
                              then @quartIndex + 1
                            else 1
                        end
            where rn = @rowNum

        select @quartIndex = numQuarter from #series2
            where rn = @rowNum

delete #rows where rn = @rowNum
end

-- clean up
drop table #rows
go

alter table #series2 add pharmID int

select rn into #rows from #series2
SET ANSI_PADDING ON


GO
-- Add index for faster data retrieval
CREATE NONCLUSTERED INDEX [NonClusteredIndex-20181121-080868]
ON #rows
(
    rn ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)

GO

Declare @rowNum int
Declare @seriesIndex int

set @seriesIndex = 0
```

```
while exists (select * from #rows)
    begin
        select top 1 @rowNum = rn from #rows
        update #series2
        set pharmID = case when numQuarter = 1
                            and linked is null
                                then @seriesIndex + 1
                           when numQuarter = 1
                            and linked <> 0
                            and linked <> 1 and linked <> 97
                                then @seriesIndex +1
                           else @seriesIndex
                      end
        where rn = @rowNum
        select @seriesIndex = pharmID
        from #series2
            where rn = @rowNum
        delete #rows where rn = @rowNum
    end

    -- clean up
drop table #rows
go

-- Select all patients who have loop diuretic series
-- spanning two or more consecutive quarters.
with pharmIDs(safehavenID, PRESC_DATE, min_DISP_DATE,
max_DISP_DATE, numQuarter, pharmID) as
(
    select
        min(safehavenID) as safehavenID,
        min(PRESC_DATE) as PRESC_DATE,
        min(DISP_DATE) as min_DISP_DATE,
        max(DISP_DATE) as max_DISP_DATE,
        max(numQuarter) as numQuarter, pharmID
    from #series2
        group by pharmID having max(numQuarter) >= 2
)
```

```
-- Select the first repeat prescription event per patient
-- ranked by series start date.
select *
into #series3
from
(
    select
        *,  ROW_NUMBER() over (partition by safehavenID
        order by min_DISP_DATE asc) as rn
    from pharmIDs
) t
    where rn = 1

--** Step 5: Create table
-- Create a holding table
create table #holding(
    safehavenID int,
    eventDate date,
    eventType varchar(200)
);
go

--** Step 6: Populate holding table with dead and
-- repeat dates
insert into #holding (safehavenID, eventDate, eventType)
select
    [safehavenID], [PRESC_DATE], 'dead'
from #dodWithin90Days

insert into #holding (safehavenID, eventDate, eventType)
select
    [safehavenID], min_DISP_DATE, 'repeat'
from #series3

--** Step 7: Create a table by Selecting the first qualifying
-- event per safehavenID
select
    h.[safehavenID], h.[eventdate], h.[eventType]
```

```
into working.LDDeadOr2ConsecQrt
from
(
    select
        [safehavenID],[eventdate], [eventType],
        ROW_NUMBER () over (partition by safehavenID
        order by [eventdate] asc) rn
    from #holding
) h
    where h.rn = 1
        order by [safehavenID]

--** clean up
drop table #holding
drop table #series
drop table #series2
drop table #series3
drop table #allLD
drop table #dodWithin90Days
```

A.8 Annotated Electrical Cardiac Cycles

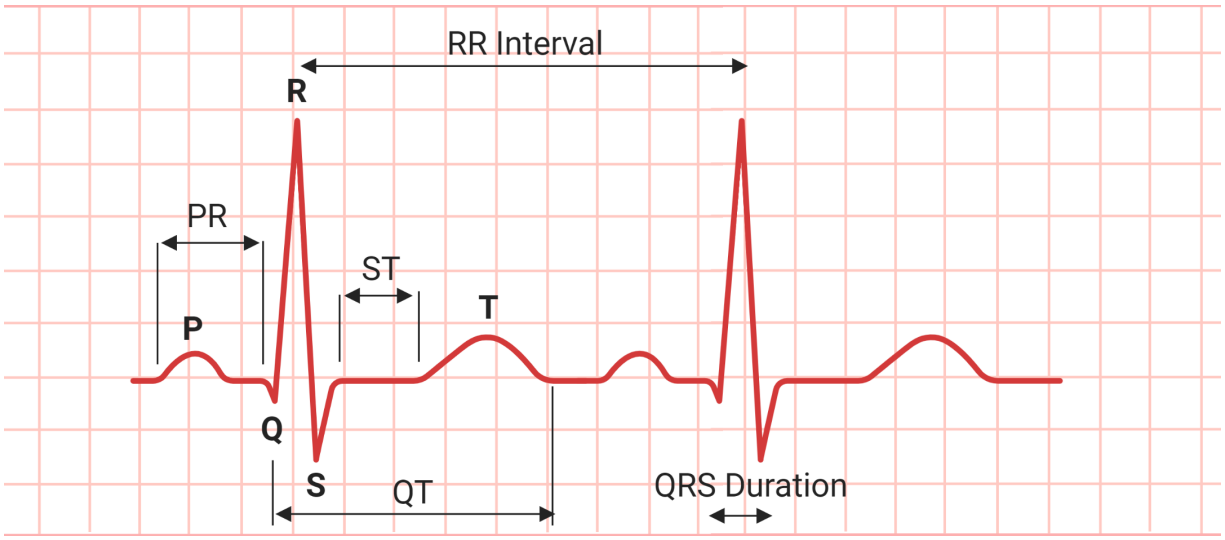


Figure A.2: The basic shape of two normal cardiac cycles as shown on the ECG.

As seen in Figure A.2, the P, Q, R, S, and T deflections are called waves; the Q, R, and S waves together make up the QRS complex. The interval between the S wave and the T wave is the ST segment. The duration of the QRS complex indicates the amount of time needed for

the electrical signal to travel through the ventricles. A normal QRS duration is 0.12s (120 ms) Hampton & Hampton (2019). Each square across represents 0.2 s (200 ms), and when properly calibrated, each vertical square represents half of a millivolt (mV). The relationship between the distance, or the number of squares, (amount of time) between the two R waves (the RR interval) gives the heart rate.

A.9 Regular Expressions Used to Identify Rhythms

Table A.4: Search terms were used to classify the heart rhythm using the automatically generated ECG diagnostic text. Classifications are assigned in the order of sinus, AF/AFL, other, paced, and undetermined.

General Classification	Rhythm	Search Term
Sinus	Sinus	Sinus rhythm
		Sinus rhythm in %
		Sinus rhythm and %
		Sinus rhythm with %
		Possible ectopic atrial rhythm
		Possible ectopic atrial rhythm with %
		Possible ectopic atrial bradycardia
		Possible ectopic atrial bradycardia with%
		Possible ectopic atrial tachy%
		Normal sinus rhythm
		Sinus bradycardia
		Marked sinus bradycardia
		Sinus tachycardia
AF/AFL/ SVT	AF	Atrial fibrillation
	AF/AFL	Atrial flutter
		Atrial flutter with %
	SVT	Supraventricular tachycardia
		Supraventricular tachycardia with %
		Narrow QRS tachycardia
Other	Idioventricular	Idioventricular rhythm
		Idioventricular rhythm with %
	Junctional	Junctional rhythm
		Junctional rhythm with %
		Probable junctional with %
		Probable junctional rhythm with %
		Accelerated junctional rhythm

Continuation of Table A.4

General Classification	Rhythm	Search Term
Other		Probable junctional rhythm
		Probable junctional tachy%
		% probable junctional tachy
		Junctional bradycardia
		Junctional bradycardia with %
	Other	Irregular Low right Atrial rhythm
		Low right atrial
		Wide QRS
		Wide QRS rhythm
		Wide QRS rhythm with %
		Wide QRS tachy%
		Narrow QRS
Paced	Paced	Electronic ventricular pacemaker
		AV dual-paced rhythm
		AV sequential or dual chamber electronic pacemaker
		Suspected unspecified pacemaker failure
		Ventricular-paced rhythm
		Electronic atrial pacemaker
		Atrial-paced rhythm
		Demand pacemaker
Undetermined	Undetermined	Undetermined rhythm

AV, atrial and ventricular; AV, atrial and ventricular; %, the SQL wildcard search character; AF, atrial fibrillation; AFL, atrial flutter; SVT, supraventricular tachycardia.

A.10 Regular Expressions Used to Identify the Presence of ST-T Abnormalities

Table A.5: Search terms used to identify ST-T wave abnormalities in the automatically generated ECG diagnostic text. Classifications are assigned in order of appearance in the table, with the absence of ST-T abnormalities determined first.

No ST-T Abnormalities	ST-T Abnormalities Present
Normal sinus rhythm; Normal ECG	T wave abnormality, %
Normal sinus rhythm*	T wave amplitude has increased %

Continuation of Table A.5

No ST-T Abnormalities	**ST-T Abnormalities Present**
Nonspecific T wave abnormality no longer evident %	T wave inversion more evident
% ST no longer depressed	ST & T wave abnormality
% ST no longer elevated	Marked ST wave abnormality, %
T wave inversion no longer evident	Marked T wave abnormality
Minimal voltage criteria for LVH, may be normal variant	Marked ST wave abnormality %
	Marked ST abnormality
	ST & Marked T wave abnormality
	Nonspecific T wave abnormality %
	Nonspecific T wave abnormality
	Nonspecific ST wave abnormality
	Nonspecific ST abnormality
	Nonspecific ST and T wave abnormality %
	Non-specific change in ST segment in %
	ST elevation
	ST elevation in %
	ST elevation now present
	ST elevation has replaced %
	ST elevation now present
	ST less elevated in Anterior leads
	ST depression
	Early repolarization
	Junctional ST depression
	T wave inversion now evident
	T wave inversion less evident
	T wave inversion in Inferior leads
	T wave inversion in %
	T wave amplitude has decreased in %
	Right ventricular hypertrophy with repolarization abnormality
	Left ventricular hypertrophy with repolarization abnormality
	% with QRS widening and repolarization abnormality
	% with repolarization abnormality

Continuation of Table A.5

No ST-T Abnormalities	ST-T Abnormalities Present
	Abnormal QRS-T angle, consider primary T wave abnormality

* Diagnostic text does not provide further information

%, the SQL wildcard search character; LVH, left ventricular hypertrophy.

A.11 Codes Used to Define Inclusion Conditions

Table A.6: Codes used to identify inclusion conditions. Unless otherwise specified, all three and four character ICD-10 and OPCS-4 codes contain nested codes.

Inclusion Reason	Code Type	Code	Description
HF	ICD-10	See Table A.14	Heart failure
	Read	See Table A.15	
ARB	BNF	See Table A.3	Angiotensin II receptor blockers
ACEi	BNF	See Table A.3	Angiotensin-converting-enzyme inhibitors
LD	BNF	See Table A.3	Loop diuretics
MRA	BNF	See Table A.3	Mineralocorticoid-receptor antagonists
Beta-blocker	BNF	See Table A.3	Beta blockers
PAD	Read	14F7	H/O: arterial lower limb ulcer
	Read	14NB	H/O: Peripheral vascular disease procedure
	Read	2G63	Ischaemic toe
	Read	7A100	Emerg aortic bypass by anastomosis axillary to femoral art
	Read	7A101	Bypass aorta by anastomosis axillary to femoral artery NEC
	Read	7A102	Axillo-bifemoral bypass graft
	Read	7A103	Axillo-unifemoral PTFE bypass graft
	Read	7A12	Other bypass of bifurcation of aorta
	Read	7A120	Emerg bypass bifurc aorta by anast aorta to femoral artery
	Read	7A121	Bypass bifurc aorta by anastom aorta to femoral artery NEC
	Read	7A12111	Aorto bifemoral graft
	Read	7A12112	Dacron aortofemoral Y graft
	Read	7A123	Bypass bifurcation aorta by anastom aorta to iliac artery

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	7A12311	Aorto biiliac graft
	Read	7A12312	Dacron aortoiliac Y graft
	Read	7A12y	Other specified other bypass of bifurcation of aorta
	Read	7A12z	Other bypass of bifurcation of aorta NOS
	Read	7A192	Open embolectomy of bifurcation of aorta
	Read	7A41	Other bypass of iliac artery
	Read	7A410	Emerg bypass iliac art by iliac/femoral art anastomosis NEC
	Read	7A411	Bypass iliac artery by iliac/femoral artery anastomosis NEC
	Read	7A412	Emerg bypass iliac artery by femoral/femoral art anast NEC
	Read	7A41211	Emergency femoro-femoral prosthetic cross over graft
	Read	7A413	Bypass iliac artery by femoral/femoral art anastomosis NEC
PAD	Read	7A41311	Femoro-femoral prosthetic cross over graft
	Read	7A414	Emerg bypass comm iliac art by aorta/com iliac art anast NEC
	Read	7A416	Emerg bypass leg artery by aorta/com fem art anastomosis NEC
	Read	7A419	Bypass common iliac artery by aorta/com iliac art anast NEC
	Read	7A41B	Bypass leg artery by aorta/com femoral art anastomosis NEC
	Read	7A41C	Bypass leg artery by aorta/deep femoral art anastomosis NEC
	Read	7A41F	Ilio-femoral prosthetic cross over graft
	Read	7A41y	Other specified other bypass of iliac artery
	Read	7A41z	Other bypass of iliac artery NOS
	Read	7A42	Reconstruction of iliac artery
	Read	7A420	Endarterectomy and patch repair of iliac artery
	Read	7A42011	Endarterectomy and patch repair of common iliac artery

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	7A42012	Iliac endarterectomy and patch
	Read	7A421	Endarterectomy of iliac artery NEC
	Read	7A42111	Endarterectomy of common iliac artery NEC
	Read	7A42y	Other specified reconstruction of iliac artery
	Read	7A42z	Reconstruction of iliac artery NOS
	Read	7A43	Other open operations on iliac artery
	Read	7A430	Repair of iliac artery NEC
	Read	7A43011	Repair of common iliac artery NEC
	Read	7A431	Open embolectomy of iliac artery
	Read	7A43111	Open embolectomy of common iliac artery
	Read	7A433	Open insertion of iliac artery stent
PAD	Read	7A440	Percutaneous transluminal angioplasty of iliac artery
	Read	7A441	Percutaneous transluminal embolectomy of iliac artery
	Read	7A443	Insertion of iliac artery stent
	Read	7A444	Percutaneous transluminal insertion of iliac artery stent
	Read	7A44y	Other specified transluminal operation on iliac artery
	Read	7A44z	Transluminal operation on iliac artery NOS
	Read	7A47	Other emergency bypass of femoral artery or popliteal artery
	Read	7A470	Emerg bypass femoral art by fem/pop art anast c prosth NEC
	Read	7A471	Emerg bypass popliteal art by pop/pop art anast c prosth NEC
	Read	7A472	Emerg bypass femoral art by fem/pop a anast c vein graft NEC
	Read	7A473	Emerg bypass pop art by pop/pop art anast c vein graft NEC
	Read	7A474	Emerg bypass femoral art by fem/tib art anast c prosth NEC

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	7A476	Emerg bypass femoral art by fem/tib a anast c vein graft NEC
	Read	7A477	Emerg bypass pop art by pop/tib art anast c vein graft NEC
	Read	7A47B	Emerg bypass pop art by pop/peron art anast c vein graft NEC
	Read	7A47C	Emerg bypass femoral artery by fem/fem art anastomosis NEC
	Read	7A47D	Emerg bypass popliteal artery by pop/fem art anastomosis NEC
	Read	7A47y	Other emergency bypass of femoral or popliteal artery NOS
	Read	7A47z	Other emergency bypass of femoral or popliteal artery NOS
	Read	7A48	Other bypass of femoral artery or popliteal artery
	Read	7A480	Bypass femoral artery by fem/pop art anast c prosthesis NEC
	Read	7A481	Bypass popliteal artery by pop/pop a anast c prosthesis NEC
	Read	7A482	Bypass femoral artery by fem/pop art anast c vein graft NEC
PAD	Read	7A483	Bypass popliteal artery by pop/pop a anast c vein graft NEC
	Read	7A484	Bypass femoral artery by fem/tib art anast c prosthesis NEC
	Read	7A485	Bypass popliteal artery by pop/tib a anast c prosthesis NEC
	Read	7A486	Bypass femoral artery by fem/tib art anast c vein graft NEC
	Read	7A487	Bypass popliteal artery by pop/tib a anast c vein graft NEC
	Read	7A488	Bypass femoral artery by fem/peron a anast c prosthesis NEC
	Read	7A48A	Bypass femoral artery by fem/peron a anast c vein graft NEC
	Read	7A48B	Bypass popliteal art by pop/peron art anast c vein graft NEC
	Read	7A48C	Bypass femoral artery by femoral/femoral art anastomosis NEC
	Read	7A48D	Bypass popliteal artery by pop/fem artery anastomosis NEC
	Read	7A48E	Femoro-femoral prosthetic cross over graft
	Read	7A48y	Other bypass of femoral artery or popliteal artery NOS

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	7A48z	Other bypass of femoral artery or popliteal artery NOS
	Read	7A49	Reconstruction of femoral artery or popliteal artery
	Read	7A490	Endarterectomy and patch repair of femoral artery
	Read	7A491	Endarterectomy and patch repair of popliteal artery
	Read	7A492	Endarterectomy of femoral artery NEC
	Read	7A493	Endarterectomy of popliteal artery NEC
	Read	7A494	Profundoplasty femoral artery & patch repair deep fem artery
	Read	7A495	Profundoplasty and patch repair of popliteal artery
	Read	7A496	Profundoplasty of femoral artery NEC
	Read	7A497	Profundoplasty of popliteal artery NEC
	Read	7A498	Reconstruction of femoral artery with vein graft
PAD	Read	7A499	Reconstruction of popliteal artery with vein graft
	Read	7A49y	Reconstruction of femoral or popliteal artery OS
	Read	7A49z	Reconstruction of femoral or popliteal artery NOS
	Read	7A4A	Other open operations on femoral artery or popliteal artery
	Read	7A4A0	Repair of femoral artery NEC
	Read	7A4A1	Repair of popliteal artery NEC
	Read	7A4A2	Open embolectomy of femoral artery
	Read	7A4A211	Open thrombectomy of femoral artery
	Read	7A4A212	Open femoral embolectomy
	Read	7A4A3	Open embolectomy popliteal artery
	Read	7A4A311	Open thrombectomy of popliteal artery
	Read	7A4A7	Repair of femoral artery with temporary silastic shunt

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	7A4A8	Repair of popliteal artery with temporary silastic shunt
	Read	7A4Ay	Other open operation on femoral or popliteal artery OS
	Read	7A4B0	Percutaneous transluminal angioplasty of femoral artery
	Read	7A4B1	Percutaneous transluminal angioplasty of popliteal artery
	Read	7A4B2	Percutaneous transluminal embolectomy of femoral artery
	Read	7A4B3	Percutaneous transluminal embolectomy of popliteal artery
	Read	7A4B4	Percutaneous transluminal embolisation of femoral artery
	Read	7A4B5	Percutaneous transluminal embolisation of popliteal artery
	Read	7A4B8	Percut translum thrombolysis femoral graft streptokinase
	Read	7A4B9	Percutaneous transluminal insertion of stent femoral artery
	Read	7A50	Revision of reconstruction of artery
PAD	Read	7A500	Revision of reconstruction involving aorta
	Read	7A501	Revision of reconstruction involving iliac artery
	Read	7A502	Revision of reconstruction involving femoral artery
	Read	7A503	Revision of reconstruction of popliteal artery
	Read	A3A0F	Gas gangrene-foot
	Read	C107	Diabetes mellitus with peripheral circulatory disorder
	Read	C1070	Diabetes mellitus, juvenile +peripheral circulatory disorder
	Read	C1071	Diabetes mellitus, adult, + peripheral circulatory disorder
	Read	C1073	IDDM with peripheral circulatory disorder
	Read	C1074	NIDDM with peripheral circulatory disorder
	Read	C107z	Diabetes mellitus NOS with peripheral circulatory disorder
	Read	C108G	Insulin dependent diab mell with peripheral angiopathy

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	C109F	Non-insulin-dependent d m with peripheral angiopath
	Read	C109F11	Type II diabetes mellitus with peripheral angiopathy
	Read	C109F12	Type 2 diabetes mellitus with peripheral angiopathy
	Read	C10EG	Type 1 diabetes mellitus with peripheral angiopathy
	Read	C10FF	Type 2 diabetes mellitus with peripheral angiopathy
	Read	G700	Aorto-iliac disease
	Read	G702	Extremity artery atheroma
	Read	G702z	Extremity artery atheroma NOS
	Read	G73	Other peripheral vascular disease
	Read	G731	Thromboangiitis obliterans
	Read	G7310	Buerger's disease
PAD	Read	G731z	Thromboangiitis obliterans NOS
	Read	G732	Peripheral gangrene
	Read	G7320	Gangrene of toe
	Read	G7321	Gangrene of foot
	Read	G733	Ischaemic foot
	Read	G73y	Other specified peripheral vascular disease
	Read	G73y0	Diabetic peripheral angiopathy
	Read	G73y1	Peripheral angiopathic disease EC NOS
	Read	G73y2	Acrocyanosis
	Read	G73y4	Acroparaesthesia - Schultze's type
	Read	G73y5	Acroparaesthesia - Nothnagel's type
	Read	G73y511	Nothnagel's vasomotor acroparaesthesia

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	G73y6	Acroparaesthesia - unspecified
	Read	G73y	Erythrocyanosis
	Read	G73y8	Erythromelalgia
	Read	G73yz	Other specified peripheral vascular disease NOS
	Read	G73z	Peripheral vascular disease NOS
	Read	G73z0	Intermittent claudication
	Read	G73z011	Claudication
	Read	G73zz	Peripheral vascular disease NOS
	Read	G740	Aortoiliac obstruction
	Read	G7424	Embolism and thrombosis of the femoral artery
	Read	G7425	Embolism and thrombosis of the popliteal artery
PAD	Read	G7426	Embolism and thrombosis of the anterior tibial artery
	Read	G7427	Embolism and thrombosis of the dorsalis pedis artery
	Read	G7429	Embolism and thrombosis of a leg artery NOS
	Read	G742z	Peripheral arterial embolism and thrombosis NOS
	Read	G74y0	Embolism and/or thrombosis of the common iliac artery
	Read	G74y1	Embolism and/or thrombosis of the internal iliac artery
	Read	G74y2	Embolism and/or thrombosis of the external iliac artery
	Read	G74y3	Embolism and thrombosis of the iliac artery unspecified
	Read	Gyu74	[X]Other specified peripheral vascular diseases
	Read	M271	Ischaemic leg ulcer
	Read	M2710	Ischaemic ulcer diabetic foot
	Read	M2713	Arterial leg ulcer

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	M2714	Mixed venous and arterial leg ulcer
	Read	R0542	[D]Gangrene of toe in diabetic
	Read	R0543	[D]Widespread diabetic foot gangrene
	Read	R0550	[D]Failure of peripheral circulation
	Read	R055011	[D]Peripheral circulatory failure
	ICD-10	I731	Thromboangiitis obliterans [Buerger]
	ICD-10	I738	Other specified peripheral vascular diseases
	ICD-10	I739	Peripheral vascular disease, unspecified
	ICD-10	I743	Embolism and thrombosis of arteries of lower extremities
	ICD-10	I744	Embolism and thrombosis of arteries of extremities, unspecified
	ICD-10	I745	Embolism and thrombosis of iliac artery
PAD	OPCS-4	L50	Other emergency bypass of iliac artery
	OPCS-4	L51	Other bypass of iliac artery
	OPCS-4	L52	Reconstruction of iliac artery
	OPCS-4	L530	Other open operations on iliac artery
	OPCS-4	L531	Repair of iliac artery NEC
	OPCS-4	L532	Open embolectomy of iliac artery
	OPCS-4	L541	Percutaneous transluminal angioplasty of iliac artery
	OPCS-4	L542	Percutaneous transluminal embolectomy of iliac artery
	OPCS-4	L544	Percutaneous transluminal insertion of stent into iliac artery
	OPCS-4	L548	Other specified transluminal operations on iliac artery
	OPCS-4	L549	Unspecified transluminal operations on iliac artery
	OPCS-4	L58	Other emergency bypass of femoral artery

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
PAD	OPCS-4	L59	Other bypass of femoral artery
	OPCS-4	L60	Reconstruction of femoral artery
	OPCS-4	L620	Other open operations on femoral artery
	OPCS-4	L621	Repair of femoral artery NEC
	OPCS-4	L622	Open embolectomy of femoral artery
	OPCS-4	L628	Other specified other open operations on femoral artery
	OPCS-4	L629	Unspecified other open operations on femoral artery
	OPCS-4	L631	Percutaneous transluminal angioplasty of femoral artery
	OPCS-4	L632	Percutaneous transluminal embolectomy of femoral artery
	OPCS-4	L633	Percutaneous transluminal embolisation of femoral artery
	OPCS-4	L635	Percutaneous transluminal insertion of stent into femoral artery
	OPCS-4	L650	Revision of reconstruction of artery
	OPCS-4	L651	Revision of reconstruction involving aorta
	OPCS-4	L652	Revision of reconstruction involving iliac artery
	OPCS-4	L653	Revision of reconstruction involving femoral artery
CAD	ICD-10	I20	Angina pectoris
	ICD-10	I21	Acute myocardial infarction
	ICD-10	I22	Subsequent myocardial infarction
	ICD-10	I23	Certain current complications following acute myocardial infarction
	ICD-10	I24	Other acute ischaemic heart diseases
	ICD-10	I25	Chronic ischaemic heart disease
	Read	14AL	H/O: Treatment for ischaemic heart disease
	Read	G310	Dressler's syndrome

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
	Read	G33z5	Post infarct angina
	Read	889A	Diab mellit insulin-glucose infus acute myocardial infarct
	Read	6A2	Coronary heart disease annual review
	Read	6A4	Coronary heart disease review
	Read	8B3k	Coronary heart disease medication review
	Read	8H2V	Admit ischaemic heart disease emergency
	Read	G3	Ischaemic heart disease
	Read	G31	Other acute and subacute ischaemic heart disease
	Read	G3110	Myocardial infarction aborted
	Read	G311011	MI - myocardial infarction aborted
	Read	G31y	Other acute and subacute ischaemic heart disease
CAD	Read	G31y2	Subendocardial ischaemia
	Read	G31y3	Transient myocardial ischaemia
	Read	G31yz	Other acute and subacute ischaemic heart disease NOS
	Read	G34	Other chronic ischaemic heart disease
	Read	G340	Coronary atherosclerosis
	Read	G344	Silent myocardial ischaemia
	Read	G34y	Other specified chronic ischaemic heart disease
	Read	G34y1	Chronic myocardial ischaemia
	Read	G34yz	Other specified chronic ischaemic heart disease NOS
	Read	G34z	Other chronic ischaemic heart disease NOS
	Read	G34z0	Asymptomatic coronary heart disease
	Read	G3y	Other specified ischaemic heart disease

Continuation of Table A.6

Inclusion Reason	Code Type	Code	Description
CAD	Read	G3z	Ischaemic heart disease NOS
	Read	Gyu3	[X]Ischaemic heart diseases
	Read	Gyu32	[X]Other forms of acute ischaemic heart disease
	Read	Gyu33	[X]Other forms of chronic ischaemic heart disease

A.12 Regular Expressions Used to Identify Echocardiography Measurements

A.12.1 Left Ventricular Ejection Fraction

Regular expression text used to identify left ventricular ejection fraction (LVEF) in the EchoPAC and Xcelera data sets. Measurements were ranked based on accuracy. The modified Simpson method is based on tracing LV cavity areas from 2D echocardiograms. The Teicholz formula was calculated using LV dimensions measured from M-mode echocardiograms.

Rank	EchoPAC	Xcelera
1	EF(Biplane)_03	EF Biplane%
2	EF(Biplane)	EF(Bi-plane)(a2DQ)
3	EF(MOD 4AC)	EF(Bi-plane)(aCMQ)
4	EF(MOD A2C)	EF(MOD-bp)
5	2D/EF(MOD)	EF(MOD-sp4)
6	EF(A-L A4C)	EF(MOD-sp2)
7	EF(A-L A2C)	EF(Cubed)
8	2D/EF(A-L)	EF(Teich)
9	2D/(Cube)	EF(sp4-el)
10	2D/(Teich)	EF(sp2-el)
11	MM/EF(Cube)	10_EF(APR)(a2DQ)
12	MM/EF(Teich)	10_EF(AP2)(A2DQ)
12		10_EF(AP4)(aCMQ)
12		10_EF(AP2)(aCMQ)

Table A.7: Ranked echocardiogram measurements used to identify LVEF measurements. Scanning modalities are ordered first by accuracy and second by prevalence (e.g., Simpson's biplane is preferred over the method of disks [MOD]).

A.12.2 Left Atrial Diameter

Regular expression text used to identify left atrial (LA) diameter measurements in the EchoPAC and Xcelera data sets.

Rank	EchoPAC	Xcelera
1	2D/LA	LA dimension
2	MM/LA	

Table A.8: Ranked echocardiogram measurements used to identify LA diameter measurements. Scanning modalities are ordered first by accuracy and second by prevalence.

A.12.3 Mitral Regurgitation

Mitral regurgitation (MR) was defined as present based on a recorded attempt to measure the regurgitation. Regular expression text in Table A.9 was used to identify a measurement in the EchoPAC and Xcelera data sets.

Rank	EchoPAC	Xcelera
1	PISA/MR/Radius	MR PISA
2	PISA/MR/Velocity	MR ERO
3	PISA/MR/Flow	
4	PISA/MR/Vmax	
5	PISA/MR/VTI	
6	PISA/MR/ERO	
7	PISA/MR/RV	

Table A.9: Ranked echocardiogram measurements were used to identify the presence of MR.

A.12.4 Tricuspid Regurgitation

Tricuspid regurgitation (TR) variables were identified identified in EchoPAC using the `ParameterId` 'TR maxPG' or `NAME` 'TR max PG' in Xcelera. Values below 0 mm Hg or above 170 mm Hg were classified as biologically impossible and were excluded.

A.12.5 Aortic Stenosis

Aortic stenosis (AS) variables were identified and ranked according to Table A.10.

Values below 0 mm Hg and above 170 mm Hg were classified as biologically impossible and were excluded.

Rank	EchoPAC	Xcelera
1	AV maxPG	Ao max PG
2		Ao max PG (full)
3		AS max PG

Table A.10: Regular expression text used to identify AS measurements in the EchoPAC and Xcelera data sets. Ranked echocardiogram measurements were used to identify the presence of AS. Scanning modalities are ordered first by accuracy and second by prevalence.

A.12.6 Aortic Regurgitation

The presence of aortic regurgitation (AR) was defined as the presence of measurement in EchoPAC identified using the `ParameterId` 'AR PHT' or `NAME` 'AI P1/2t' in Xcelera.

A.13 Excluded Echocardiography Measurements

The following measurements were not used in further analysis due to disparate machine coverage.

A.13.1 Left Atrial Area

Left atrial area (LAA) variables used to identify and rank values are listed in appendix Table A.11. Values below 3 cm^2 and above 80 cm^2 were classified as biologically impossible and were excluded. Ultimately, LAA values were not included in the analysis due to incomplete coverage before 2015 (see Figure A.3b). Regular expression text used to identify LAA measurements in the EchoPAC and Xcelera data sets.

Rank	EchoPAC	Xcelera
1	LAAd(A4C)	LA 4cs area
2	LAAd(A2C)	LA 2cs area
3	2D/LA Area	

Table A.11: Ranked echocardiogram measurements used to identify and measure LAA. Scanning modalities are ordered first by accuracy and second by prevalence.

A.13.2 Left Atrial Volume

Left atrial volume (LAV) variables used to identify and rank values are listed in appendix Table A.12. Values below 5 mL and above 800 mL were classified as biologically impossible and were

excluded. Ultimately, LAV values were not included in the analysis due to incomplete coverage before 2015 (see Figure A.3c).

Rank	EchoPAC	Xcelera
1	LAEDV(MOD A4C)	LA 4cs Vol
2	LAEDV(A-L 4AC)	LA 2cs Vol
3	LADisksD(A4C)	
4	LAEDV(MOD A2C)	
5	LAEDV(A-L 2AC)	
6	2D/LAEDV(A-L)	
7	LADisksD(A2C)	
8	LAEDV(MOD BP)	

Table A.12: Ranked echocardiogram variables used to identify and measure LAV. Scanning modalities are ordered first by accuracy and second by prevalence.

A.13.3 Right Atrial Area

Right atrial area (RAA) variables were identified and ranked according to appendix Table A.13. Values below 3 cm^2 and above 80 cm^2 were classified as biologically impossible and were excluded. Ultimately, RAA values were not included in the analysis due to incomplete coverage before 2015 (see Figure A.3a).

Rank	EchoPAC	Xcelera
1	2D/RA Area	Ra Area
2	RAAd(A4C)	

Table A.13: Ranked echocardiogram measurements used to identify and measure RAA. Scanning modalities are ordered first by accuracy and second by prevalence.

A.13.4 Right Atrial Volume

Right atrial volume was not present in the Xcelera system and was therefore not included in the analysis, as patients scanned on the Xcelera system were missing values.

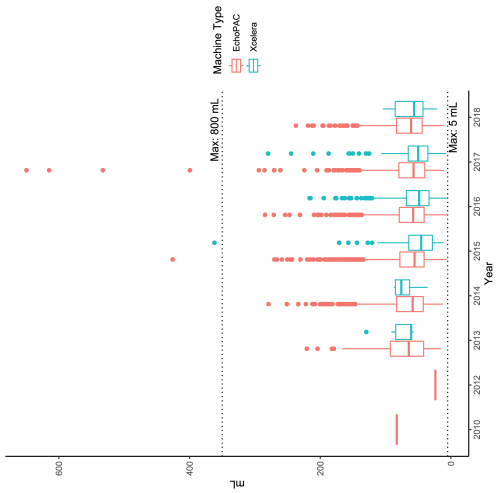
A.13.5 Inferior Vena Cava

Inferior vena cava (IVC) measurements was identified in EchoPAC using the `ParameterId` '2D/IVC'. In Xcelera, values were identified using the `NAME` 'IVC exp'. Values below 3 mm and above 55 mm were classified as biologically impossible and were excluded. Ultimately, IVC

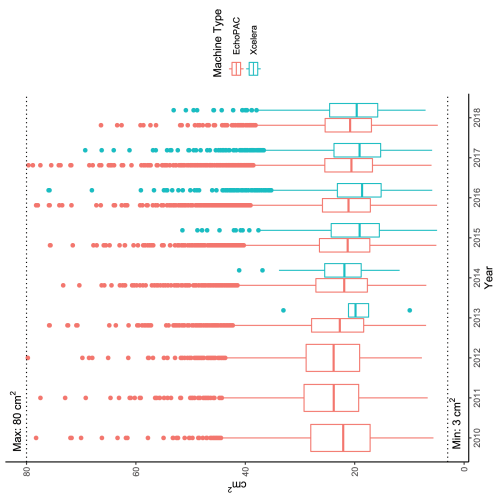
values were not included in the analysis due to incomplete coverage before 2016 (see Figure A.3e).

A.13.6 Tricuspid Annular Plane Systolic Excursion

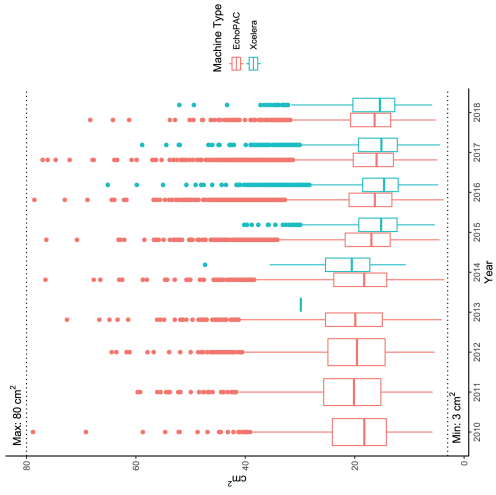
Tricuspid annular plane systolic excursion (TAPSE) measurements were identified in EchoPAC using the `ParameterId` 'MM/TAPSE'. In Xcelera, values were identified using the `NAME` 'TAPSE'. Values below 0.5 cm and above 4.25 cm were classified as biologically impossible and were excluded. Ultimately, TAPSE values were not included in the analysis due to incomplete coverage before 2015 (see Figure A.3d).



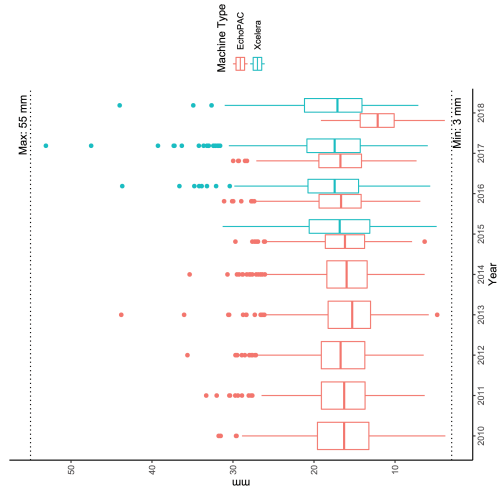
(c) Left atrial volume (LAV)



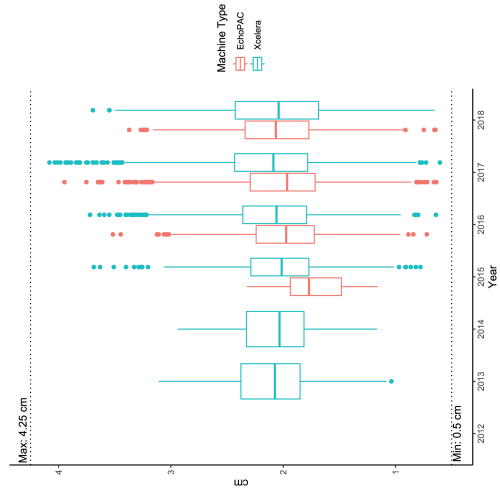
(b) Left atrial area (LAA)



(a) Right atrial area (RAA)



(e) Inferior vena cava (IVC)



(d) TAPSE

Figure A.3: Box plots of raw echocardiogram values by measurement, machine type, and year. Dotted lines indicate exclusion limits.

A.14 Codes Used to Identify Heart Failure

Table A.14: ICD-10 codes used to identify patients with HF in hospital in any diagnostic position based on previously published research (Conrad et al. 2018).

Code Type	Code	Description
ICD-10	I50	Heart failure
ICD-10	I50.0	Congestive heart failure
ICD-10	I50.1	Left ventricular failure
ICD-10	I50.9	Heart failure, unspecified
ICD-10	I42.0	Dilated cardiomyopathy (Congestive cardiomyopathy)
ICD-10	I42.9	Cardiomyopathy, unspecified (Cardiomyopathy (primary)(secondary) NOS)
ICD-10	I11.0	Hypertensive heart disease with (congestive) heart failure
ICD-10	I25.5	Ischaemic cardiomyopathy
ICD-10	I13.0	Hypertensive heart and renal disease with (congestive) heart failure
ICD-10	I13.2	Hypertensive heart and renal disease with both (congestive) heart failure and renal failure

Table A.15: Clinical Read codes used to identify patients with HF in GP LES. Codes were adapted from previously published research (Conrad et al. 2018) and from searching for the term 'heart failure' in the NHS coding dictionary.

Code Type	Read Code	Safe Haven Search Term	Description	In Conrad et al.
Read	G581.00	G581	Left ventricular failure	Y
Read	G58..00	G58	Heart failure	Y
Read	G580.11	G580	Congestive cardiac failure	Y
Read	G580.00	G580	Congestive heart failure	Y
Read	G5yy900	G5yy9	Left Ventricular Systolic Dysfunction	Y

Continuation of Table A.15

Code Type	Read Code	Safe Haven Search Term	Description	In Conrad et al.
Read	G55..00	G55	Cardiomyopathy	Y
Read	585f.00	585f	Echocardiogram shows left ventricular systolic dysfunction	Y
Read	G58..11	G58	Cardiac failure	Y
Read	G581.13	G581	Impaired left ventricular function	Y
Read	33BA.00	33BA	Impaired Left Ventricular Function	Y
Read	1O1..00	1O1	Heart failure confirmed	Y
Read	G5yyA00	G5yyA	Left Ventricular Diastolic dysfunction	Y
Read	G551.00	G551	Hypertrophic obstructive cardiomyopathy	Y
Read	G41z.11	G41z	Chronic cor pulmonale	Y
Read	G58z.00	G58z	Heart failure NOS	Y
Read	585g.00	585g	Echo shows LVDD	Y
Read	G554400	G5544	Primary dilated cardiomyopathy	Y
Read	662g.00	662g	New York Heart Association classification - class II	Y
Read	G581000	G5810	Acute left ventricular failure	Y
Read	G580.12	G580	Right heart failure	Y
Read	G554300	G5543	Hypertrophic non-obstructive cardiomyopathy	Y
Read	G580.14	G580	Biventricular failure	Y
Read	G343.00	G343	Ischaemic cardiomyopathy	Y
Read	G58z.12	G58z	Cardiac failure NOS	Y
Read	662h.00	662h	New York Heart Association classification - class III	Y
Read	G582.00	G582	Acute heart failure	Y
Read	G55z.00	G55z	Cardiomyopathy NOS	Y
Read	G5yyD00	G5yyD	Left ventricular cardiac dysfunction	Y

Continuation of Table A.15

Code Type	Read Code	Safe Haven Search Term	Description	In Conrad et al.
Read	662f.00	662f	NYHA class f - i	Y
Read	G583.00	G583	Heart failure with normal ejection fraction	Y
Read	G580100	G5801	Chronic congestive heart failure	Y
Read	G555.00	G555	Alcoholic cardiomyopathy	Y
Read	G580.13	G580	Right ventricular failure	Y
Read	G55y.11	G55y	Secondary dilated cardiomyopathy	Y
Read	G580200	G5802	Decompensated cardiac failure	Y
Read	8H2S.00	8H2S	Admit heart failure emergency	Y
Read	G554200	G5542	Familial cardiomyopathy	Y
Read	G583.12	G583	Heart failure with preserved ejection fraction	Y
Read	G584.00	G584	Right ventricular failure	Y
Read	G554000	G5540	Congestive cardiomyopathy	Y
Read	G580300	G5803	Compensated cardiac failure	Y
Read	G559.00	G559	Arrhythmogenic right ventricular cardiomyopathy	Y
Read	G583.11	G583	HFNEF - heart failure with normal ejection fraction	Y
Read	8CMW800	8CMW8	Heart failure clinical pathway	Y
Read	G580400	G5804	Congestive heart failure due to valvular disease	Y
Read	G558100	G5581	Cardiomyopathy in myotonic dystrophy	Y
Read	Gyu5M00	Gyu5M	[X]Other hypertrophic cardiomyopathy	Y
Read	G1yz100	G1yz1	Rheumatic left ventricular failure	Y
Read	662i.00	662i	New York Heart Association classification - class IV	Y
Read	661M500	661M5	Heart failure self-management plan agreed	Y
Read	8CMK.00	8CMK	Has heart failure management plan	Y

Continuation of Table A.15

Code Type	Read Code	Safe Haven Search Term	Description	In Conrad et al.
Read	G21z100	G21z1	Hypertensive heart disease NOS with CCF	Y
Read	G211100	G2111	Benign hypertensive heart disease with CCF	Y
Read	G232.00	G232	Hypertensive heart & renal dis wth (congestive) heart failure	Y
Read	8CeC.00	8CeC	Preferred place of care for next exacerbation heart failure	Y
Read	G580000	G5800	Acute congestive heart failure	N
Read	G234.00	G234	Hyperten heart & renal dis + both (congestv) heart and renal fail	N

Table A.16: Clinical codes used to exclude first HF diagnosis not referring to a *de novo* event.

Code Type	Read Code	Safe Haven Search Term	Description
Read	9N4s.00	9N4s	Did not attend practice nurse heart failure clinic
Read	8B29.00	8B29	Cardiac failure therapy
Read	679X.00	679X	Heart failure education
Read	8HHb.00	8HHb	Referral to heart failure nurse
Read	23E1.00	23E1	O/E - pulmonary oedema
Read	9hH..00	9hH	Exception reporting: heart failure quality indicators
Read	9hH0.00	9hH0	Excepted heart failure quality indicators: Patient unsuitabl
Read	9h1..00	9h1	Exception reporting: LVD quality indicators
Read	9h11.00	9h11	Excepted from LVD quality indicators: Patient unsuitable
Read	9h12.00	9h12	Excepted from LVD quality indicators: Informed dissent
Read	14A6.00	14A6	H/O: heart failure
Read	14AM.00	14AM	H/O: Heart failure in last year
Read	662p.00	662p	Heart failure 6 month review

Continuation of Table A.16

Code Type	Read Code	Safe Haven Search Term	Description
Read	662T.00	662T	Congestive heart failure monitoring
Read	662W.00	662W	Heart failure annual review
Read	8CL3.00	8CL3	Heart failure care plan discussed with patient
Read	8HBE.00	8HBE	Heart failure follow-up
Read	8HHz.00	8HHz	Referral to heart failure exercise programme
Read	8Hk0.00	8Hk0	Referred to heart failure education group
Read	9hH1.00	9hH1	Excepted heart failure quality indicators: Informed dissent
Read	9Or..00	9Or	Heart failure monitoring administration
Read	9Or0.00	9Or0	Heart failure review completed
Read	9Or1.00	9Or1	Heart failure monitoring telephone invite
Read	9Or2.00	9Or2	Heart failure monitoring verbal invite
Read	9Or3.00	9Or3	Heart failure monitoring first letter
Read	9Or4.00	9Or4	Heart failure monitoring second letter
Read	9Or5.00	9Or5	Heart failure monitoring third letter
Read	9On..00	9On	Left ventricular dysfunction monitoring administration
Read	9On0.00	9On0	Left ventricular dysfunction monitoring first letter
Read	9On1.00	9On1	Left ventricular dysfunction monitoring second letter
Read	9On2.00	9On2	Left ventricular dysfunction monitoring third letter
Read	9On3.00	9On3	Left ventricular dysfunction monitoring verbal invite
Read	9On4.00	9On4	Left ventricular dysfunction monitoring telephone invite
Read	2126400	21264	Heart Failure Resolved
Read	8HgD.00	8HgD	Discharge from heart failure nurse service
Read	8HTL000	8HTL0	Referral to rapid access heart failure clinic

Continuation of Table A.16

Code Type	Read Code	Safe Haven Search Term	Description
Read	8IB8.00	8IB8	Referral to heart failure exercise programme not indicated
Read	8IE1.00	8IE1	Referral to heart failure exercise programme declined
Read	8IE0.00	8IE0	Referral to heart failure education group declined
Read	12CR.00	12CR	FH: Hypertrophic obstructive cardiomyopathy

Appendix B

Supplementary Material for Chapter 3

B.1 Definitions of Disease Categories

Table B.1: Definitions of disease categories used for classifying causes of death and hospitalisation.

<table>
<tr><th>Causes of Death</th><th>Causes of Hospitalisation</th></tr>
<tr><td rowspan="2">CV ICD-10 Chapter IX 'Diseases of the circulatory system' (code range I00-I99), except for those classified as an infection.</td><td>Heart Failure: ICD-10 codes: I50 'Heart Failure' (includes I50.0, I50.1, I50.9), I42.0 'Dilated cardiomyopathy', I42.9 'Cardiomyopathy, unspecified', I11.0 'Hypertensive heart disease with (congestive) heart failure', I25.5 'Ischaemic cardiomyopathy', I13.0 'Hypertensive heart and renal disease with (congestive) heart failure', I13.2 'Hypertensive heart and renal disease with both (congestive) heart failure and renal failure.</td></tr>
<tr><td>Other Cardiovascular disorders: ICD-10 Chapter IX 'Disorders of the circulatory system' (I00 - I99), excluding codes related to heart failure or infections.</td></tr>
<tr><td colspan="2">Neoplasms: ICD-10 Chapter II 'Neoplasms' (C00 - D48)</td></tr>
<tr><td colspan="2">Infections: Individual codes listed in Table B.4</td></tr>
<tr><td>Other</td><td>Chronic respiratory disease: individual codes listed in Table B.3</td></tr>
</table>

Continuation of Table B.1

Causes of Death	Causes of Hospitalisation
Other: Any coding not falling into any of the above categories.	**Digestive diseases:** ICD-10 Chapter XI 'Diseases of the digestive system' (K00-K93), excluding codes categorised as infections.
	Eye & adnexa: ICD-10 Chapter VII 'Disease of the eye and adnexa' (H00-H59), excluding codes categorised as infections.
	Injury: ICD-10 Chapter XIX 'Injury, poisoning and certain other consequences of external causes' (S00-T98), and ICD-10 Chapter XX 'External Causes of morbidity and mortality' (V01-Y09).
	Mental health & neurological disorders: ICD-10 Chapter IV 'Diseases of the nervous system' (G00-G99), excepting selected codes categorised as infections and ICD-10 Chapter V 'Mental and behavioural disorders' (F00-F99).
	Musculoskeletal disorders: ICD-10 Chapter XIII 'Diseases of the musculoskeletal system and connective tissue'
	Renal: ICD-10 sub-chapters 'Renal failure' (N17-N19), 'Glomerular disease (N00-N08), 'Renal tubule-interstitial disease' (N10-N16), 'Other disorders of kidney and ureter' (N25-N29) where not classified as infections. (See Table B.5)
	Other: ICD-10 Chapter XVIII 'Symptoms, signs, and abnormal clinical laboratory findings, not elsewhere classified (R00-R99) as well as any code not falling into the above categories.
Unknown: No cause of death recorded |	

Table B.2: Codes used to define digestive disease admissions in SMR01. Unless explicitly noted otherwise, 3 and 4-digit ICD-10 lookup codes contain all codes below them.

ICD-10	Code Description
K00	Disorders of tooth development and eruption
K01	Embedded and impacted teeth
K02	Dental caries
K03	Other diseases of hard tissues of teeth
K04	Diseases of pulp and periapical tissues
K05	Gingivitis and periodontal diseases
K06	Other disorders of gingiva and edentulous alveolar ridge
K07	Dentofacial anomalies [including malocclusion]
K08	Other disorders of teeth and supporting structures
K09	Cysts of oral region, not elsewhere classified
K10	Other diseases of jaws
K11	Diseases of salivary glands
K12	Stomatitis and related lesions
K13	Other diseases of lip and oral mucosa
K14	Diseases of tongue
K20	Oesophagitis
K21	Gastro-oesophageal reflux disease
K22	Other diseases of oesophagus
K23	Disorders of oesophagus in diseases classified elsewhere
K25	Gastric ulcer
K26	Duodenal ulcer
K27	Peptic ulcer, site unspecified
K28	Gastrojejunal ulcer
K29	Gastritis and duodenitis
K30	Functional dyspepsia
K31	Other diseases of stomach and duodenum
K35	Acute appendicitis
K36	Other appendicitis
K37	Unspecified appendicitis
K38	Other diseases of appendix
K40	Inguinal hernia
K41	Femoral hernia
K42	Umbilical hernia
K43	Ventral hernia
K44	Diaphragmatic hernia

Continuation of Table B.2

ICD-10	Code Description
K45	Other abdominal hernia
K46	Unspecified abdominal hernia
K50	Crohn's disease [regional enteritis]
K51	Ulcerative colitis
K52	Other noninfective gastroenteritis and colitis
K55	Vascular disorders of intestine
K56	Paralytic ileus and intestinal obstruction without hernia
K57	Diverticular disease of intestine
K58	Irritable bowel syndrome
K59	Other functional intestinal disorders
K60	Fissure and fistula of anal and rectal regions
K61	Abscess of anal and rectal regions
K62	Other diseases of anus and rectum
K63	Other diseases of intestine
K64	Haemorrhoids and perianal venous thrombosis
K65	Peritonitis
K66	Other disorders of peritoneum

Table B.3: Codes used to define chronic respiratory disease admissions in SMR01. Unless explicitly noted otherwise, 3-digit ICD-10 lookup codes contain all codes below them.

ICD-10	Code Description
D86.0	Sarcoidosis of lung
D86.1	Sarcoidosis of lymph nodes
D86.2	Sarcoidosis of lung with sarcoidosis of lymph nodes
D86.8	Sarcoidosis of other and combined sites
D86.9	Sarcoidosis, unspecified
J38.0	Paralysis of vocal cords and larynx
J38.6	Stenosis of larynx
J39.0	Retropharyngeal and parapharyngeal abscess
J39.2	Other diseases of pharynx
J39.8	Other specified diseases of upper respiratory tract
J41	Simple and mucopurulent chronic bronchitis
J42	Unspecified chronic bronchitis
J43	Emphysema
J44	Other chronic obstructive pulmonary disease
J45	Asthma

Continuation of Table B.3

ICD-10	Code Description
J46	Status asthmaticus
J47	Bronchiectasis
J60	Coalworker pneumoconiosis
J61	Pneumoconiosis due to asbestos and other mineral fibres
J63	Pneumoconiosis due to other inorganic dusts
J64	Unspecified pneumoconiosis
J67.0	Farmer lung
J67.9	Hypersensitivity pneumonitis due to unspecified organic dust
J69.0	Pneumonitis due to food and vomit
J81	Pulmonary oedema
J82	Pulmonary eosinophilia, not elsewhere classified
J84	Other interstitial pulmonary diseases
J90	Pleural effusion, not elsewhere classified
J92.9	Pleural plaque with presence of asbestos
J93.1	Other spontaneous pneumothorax
J93.9	Pneumothorax, unspecified
J94.1	Fibrothorax
J94.8	Other specified pleural conditions
J96.1	Chronic respiratory failure
J96.9	Respiratory failure, unspecified
J98.1	Pulmonary collapse
J98.4	Other disorders of lung
J98.8	Other specified respiratory disorders
J98.9	Respiratory disorder, unspecified

Table B.4: Codes used to define infection admissions in SMR01. Unless explicitly noted otherwise, 3 and 4-digit ICD-10 lookup codes contain all codes below them.

Infection Classification	ICD-10	Code Description
Infectious diseases	A00-B99	Certain infectious and parasitic disease
Other respiratory infections	H65	Nonsuppurative otitis media
	H65.0	Acute serous otitis media
	H65.1	Other acute nonsuppurative otitis media
	H65.2	Chronic serous otitis media
	H65.3	Chronic mucoid otitis media
	H65.4	Other chronic nonsuppurative otitis media
	H66.0	Acute suppurative otitis media

Continuation of Table B.4

Infection Classification	ICD-10	Code Description
Other respiratory infections	H66.1	Chronic tubotympanic suppurative otitis media
	H66.2	Chronic atticoantral suppurative otitis media
	H72	Perforation of tympanic membrane
	H73	Other disorders of tympanic membrane
	H80	Otosclerosis
	H83	Other diseases of inner ear
	J01	Acute sinusitis
	J02	Acute pharyngitis
	J03	Acute tonsillitis
	J04	Acute laryngitis and tracheitis
	J05	cute obstructive laryngitis [croup] and epiglottitis
	J06	Acute upper respiratory infections of multiple and unspecified sites
	J09	Influenza due to identified zoonotic or pandemic influenza virus
	J10	Influenza due to identified seasonal influenza virus
	J11	Influenza, virus not identified
	J12	Viral pneumonia, not elsewhere classified
	J13	Pneumonia due to Streptococcus pneumoniae
	J14	Pneumonia due to Haemophilus influenzae
	J15	Bacterial pneumonia, not elsewhere classified
	J16	Pneumonia due to other infectious organisms, not elsewhere classified
	J17	Pneumonia in diseases classified elsewhere
	J18	Pneumonia, organism unspecified
	J20	Acute bronchitis
	J21	Acute bronchiolitis
	J22	Unspecified acute lower respiratory infection
	J32.9	Chronic sinusitis, unspecified
	J40	Bronchitis, not specified as acute or chronic
	J85.1	Abscess of lung with pneumonia
	J86.9	Pyothorax without fistula
	K67.3	Tuberculous peritonitis
	K93.0	Tuberculous disorders of intestines, peritoneum and mesenteric glands

Continuation of Table B.4

Infection Classification	ICD-10	Code Description
Other respiratory infections	N74.1	Female tuberculous pelvic inflammatory disease
UTI	N11.0	Non-obstructive reflux-associated chronic pyelonephritis
	N11.8	Other chronic tubulo-interstitial nephritis
	N15.0	Balkan nephropathy
	N15.1	Renal and perinephric abscess
	N30	Cystitis
	N34	Irradiation cystitis
	N39.0	Cystitis, unspecified
Other infectious diseases	G00	Bacterial meningitis, not elsewhere classified
	G03	Meningitis due to other and unspecified causes
	G04	Encephalitis, myelitis and encephalomyelitis
	H70.1	Chronic mastoiditis
	I00	Rheumatic fever without mention of heart involvement
	I01	Rheumatic fever with heart involvement
	I02	Rheumatic chorea
	I30	Acute pericarditis
	I33	Acute and subacute endocarditis
	I40	Acute myocarditis
	L03	Cellulitis

Table B.5: Codes used to define renal admissions in SMR01. Unless explicitly noted otherwise, 3 and 4-digit ICD-10 lookup codes contain all codes below them.

ICD-10	Code Description
N00	Acute nephritic syndrome
N01	Rapidly progressive nephritic syndrome
N02	Recurrent and persistent haematuria
N03	Chronic nephritic syndrome
N04	Nephrotic syndrome
N05	Unspecified nephritic syndrome
N06	Isolated proteinuria with specified morphological lesion
N07	Hereditary nephropathy, not elsewhere classified
N08	Glomerular disorders in diseases classified elsewhere
N10	Acute tubulo-interstitial nephritis
N11.1	Chronic obstructive pyelonephritis

Continuation of Table B.5

ICD-10	Code Description
N11.9	Chronic tubulo-interstitial nephritis, unspecified
N12	Tubulo-interstitial nephritis, not specified as acute or chronic
N13	Obstructive and reflux uropathy
N14	Drug- and heavy-metal-induced tubulo-interstitial and tubular conditions
N16	Renal tubulo-interstitial disorders in diseases classified elsewhere
N17	Acute renal failure
N18	Chronic kidney disease
N19	Unspecified kidney failure
N25	Disorders resulting from impaired renal tubular function
N26	Unspecified contracted kidney
N27	Small kidney of unknown cause
N28	Other disorders of kidney and ureter, not elsewhere classified
N29	Other disorders of kidney and ureter in diseases classified elsewhere

B.2 Patient Flow Diagram

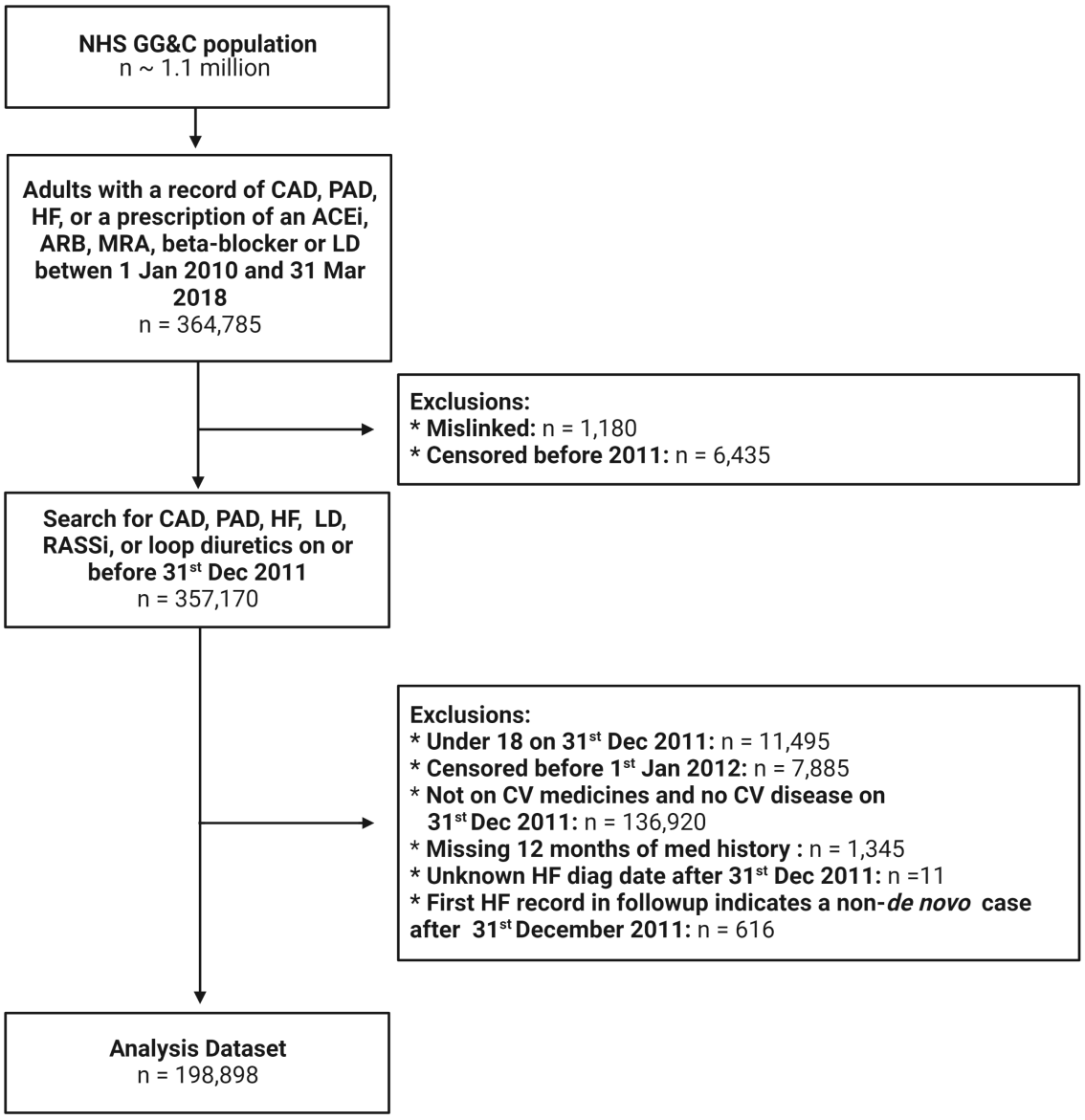
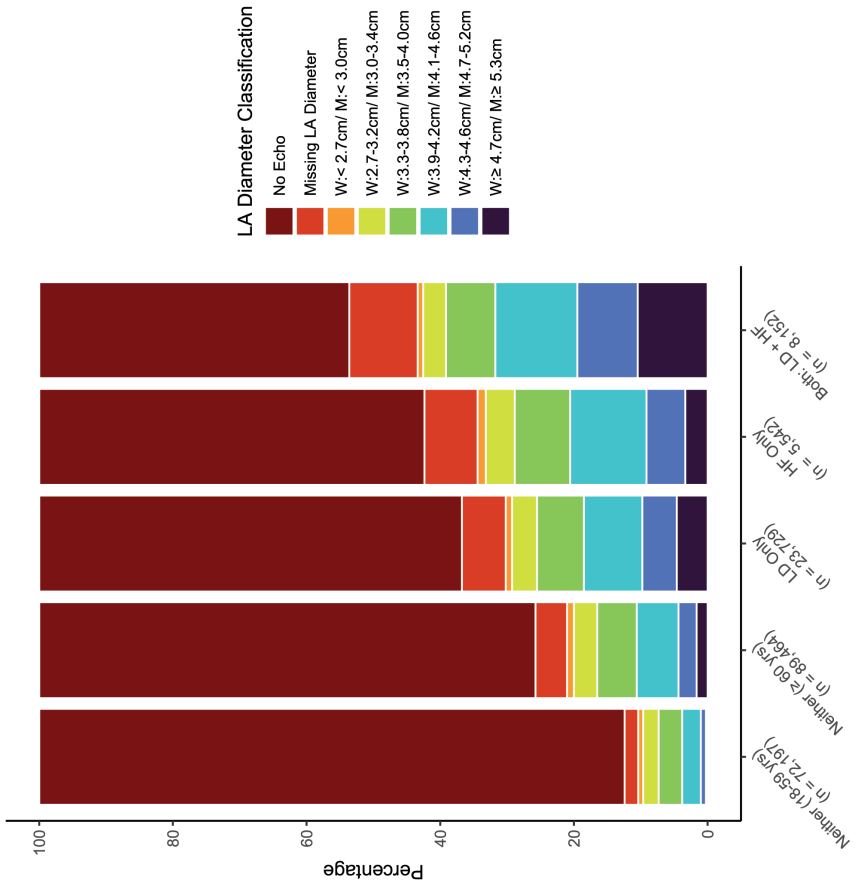
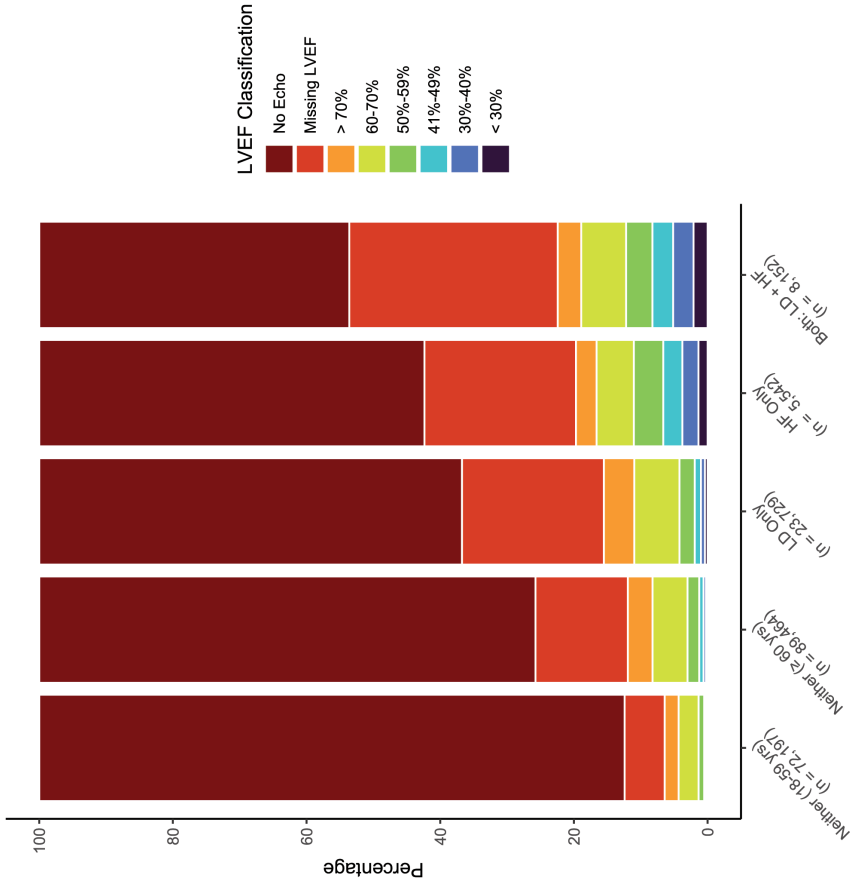


Figure B.1: Flow diagram of inclusion and exclusion criteria for NHS GG&C patients.

B.3 Additional Echocardiogram Measurements



(b) Left atrium/atrial (LA) diameter



(a) Left ventricular ejection fraction (LVEF)

Figure B.2: Breakdown of recorded measurements taken at the closest echocardiogram to 1st December 2012 by baseline group.

B.4 Assumptions Testing

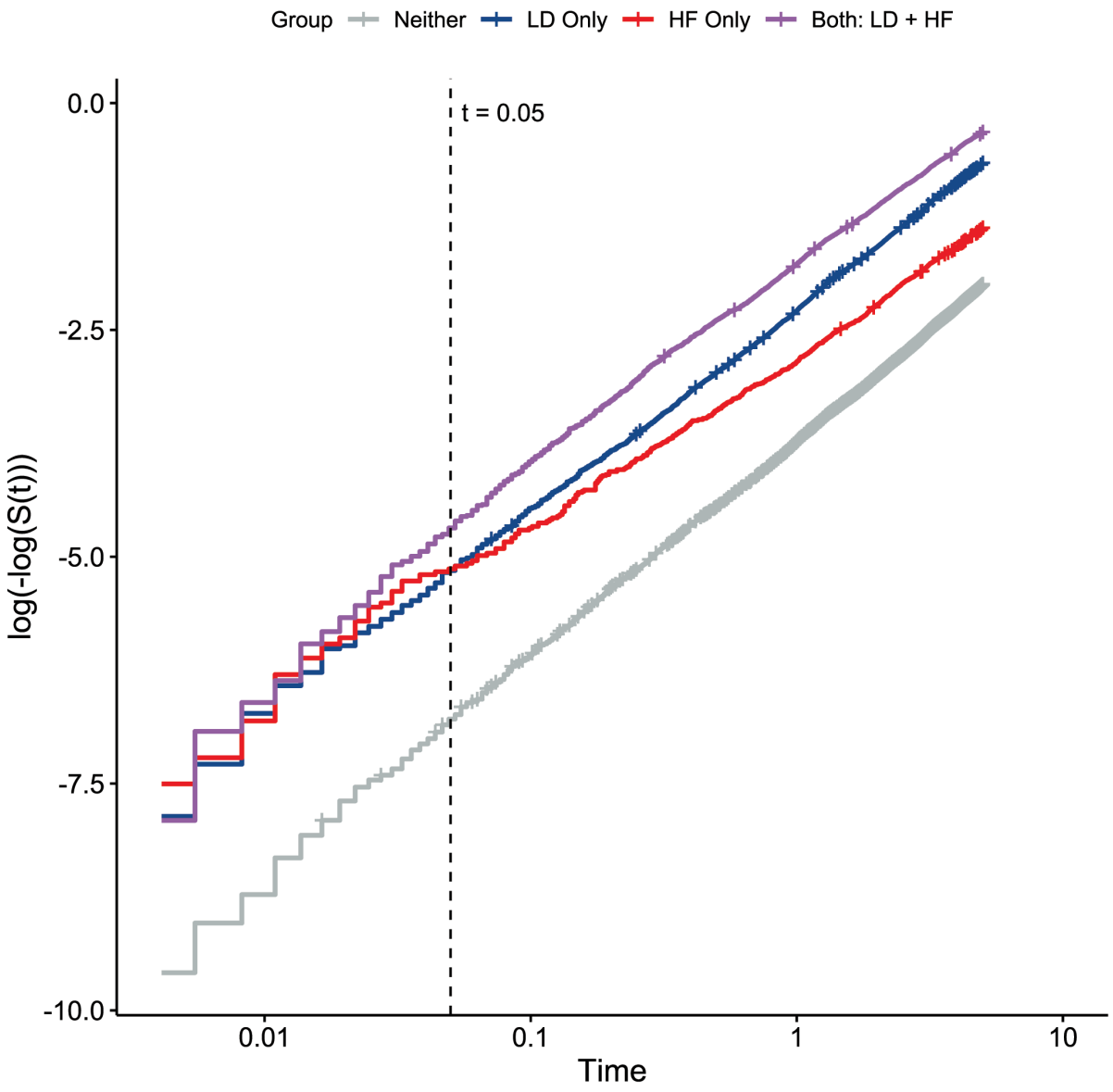


Figure B.3: Log-log plots testing proportional hazards over time (years). The proportional hazards assumption is violated at t = 0.05 years (or 18.25 days).

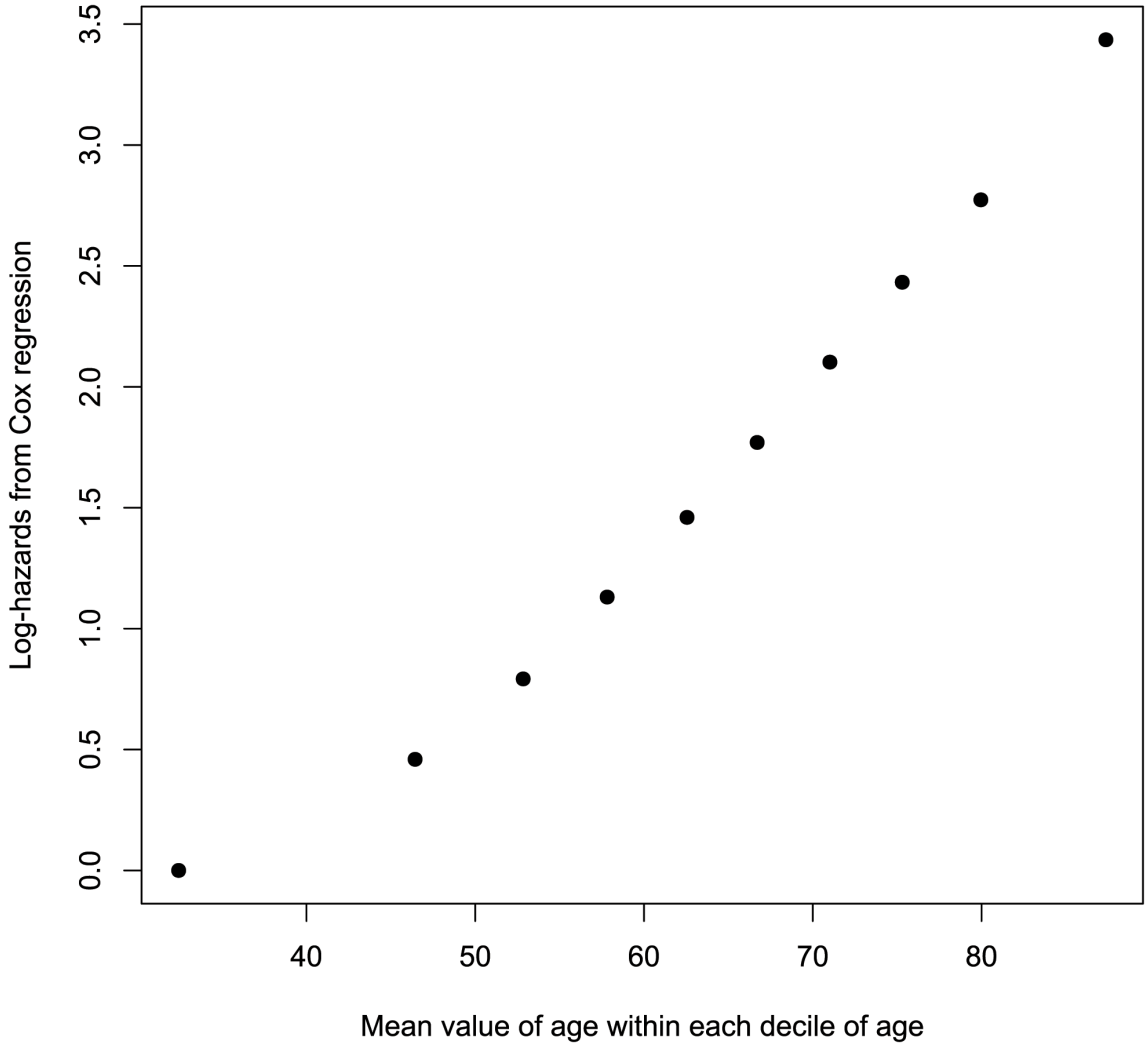


Figure B.4: Log-hazard plotted against mean age within age decile to test the linearity of hazard age.

B.5 Time-Dependent Covariant Analysis

Time-dependent analysis was conducted where the diagnosis of HF and the initiation of a LD were triggers for updating group and comorbidity status.

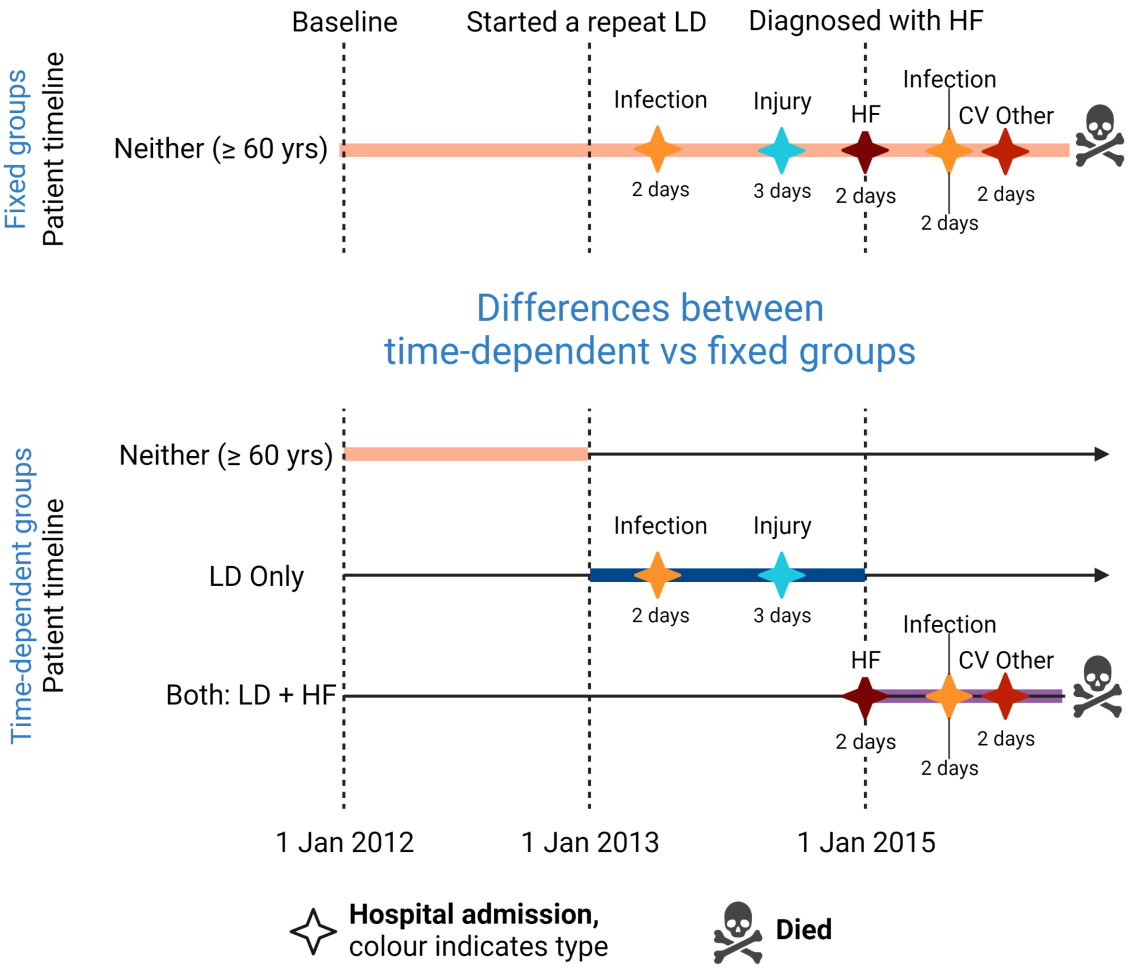


Figure B.5: Schematic illustration of the difference between using fixed group set at baseline (top) versus a time-varying group based on a diagnosis of HF and the initiation of a LD. Here, the patient died on 31st December 2016. Subtracting the 11 days where the patient was admitted to hospital, said patient contributed two infection admissions, one injury admission, one heart failure admission, and one other CV admission, plus 1,816 days or 4.97 patient-years of follow-up to the neither (≥60 yrs) group admissions calculations. With regard to mortality, the patient contributed 1,827 days to the neither (≥60 yrs) group and the death is also attributed to the neither (≥60 yrs).

Using the example presented in Figure B.5, the top schematic illustrates patient time at risk calculated using baseline group classification. In contrast, the bottom schematic illustrates patient-time at risk calculated using time-dependent groups based on the presence or absence of LD and HF. If calculations are conducted using the fixed baseline group, the patient's follow-up started on 1st January 2012 and ended when the patient died on 31st December 2016. Subtracting the 11 days the patient was admitted to hospital, the patient contributed two infection admissions, one injury admission, one heart failure admission, and one other CV admission, plus 1,816 days or 4.97 patient-years of follow-up to the neither (≥60 yrs) group admissions calculations. With regard to mortality, the patient contributed 1,827 days to the neither (≥60

yrs) group, and the death is also attributed to the neither (≥60 yrs).

If the calculations are conducted using the time-dependent groups, that same patient contributed no admissions and 367 days or 1 patient-year of admission-free follow-up to the neither (≥60 yrs) group. They then contributed one infection admission, one injury admission, and 726 admission-free days (or 1.99 admission-free patient-years) of follow-up to the LD only group. Finally, the patient changed groups for the second time based on the HF diagnosis in hospital. From this point, the patient contributed one HF admission, one infection admission, and one other CV admission plus 725 admission-free days (or 1.98 admission-free patient-years) of follow-up to the Both: LD + HF before dying on 31st December 2016. The patient contributed 367 days alive to neither (≥60 yrs) and 731 days to both LD only and Both: LD + HF, but the death record is attributed to the Both: LD + HF group.

Appendix C

Supplementary material for Chapter 4

C.1 Definitions Causes of Hospitalisations and Death

Note: all 3- and 4-digit ICD-10 codes contain codes specified below unless otherwise noted.

Table C.1: Classification of hospital admissions by ICD-10 Chapter.

ICD-10 Chapter	Description
I	**Certain infectious and parasitic diseases:** (ICD-10 code range: A00-B99)
II	**Neoplasms:** (C00-D48)
	Leukaemia/Lymphoma: C900 'Multiple myeloma', C901 'Plasma cell leukaemia', C910 'Acute lymphoblastic leukaemia [ALL]', C911 'Chronic lymphocytic leukaemia of B-cell type', C913 'Prolymphocytic leukaemia of B-cell type', C916 'Prolymphocytic leukaemia of T-cell type', C917 'Other lymphoid leukaemia', C918 'Mature B-cell leukaemia Burkitte-type', C919 'Lymphoid leukaemia, unspecified', C92 'Myeloid leukaemia', C93 'Monocytic leukaemia', C940 'Acute erythroid leukaemia', C942 'Acute megakaryoblastic leukaemia', C943 'Mast cell leukaemia', C947 'Other specified leukaemias', C95 'Leukaemia of unspecified cell type', C81 'Hodgkin lymphoma', C82 'Follicular lymphoma', C83 'Non-follicular lymphoma', C84 'Mature T/NK-cell lymphomas', C85 'Other and unspecified types of non-Hodgkin lymphoma', C883 'Immunoproliferative small intestinal disease', C887 'Other malignant immunoproliferative diseases', C889 'Malignant

Continuation of Table C.1

ICD-10 Chapter	Description
II	immunoproliferative disease, unspecified', C914 'Hairy cell leukaemia', C915 'Adult T-cell lymphoma/leukaemia [HTLV-1-associated]', C96 'Other and unspecified malignant neoplasms of lymphoid, haematopoietic and related tissue'.
	Malignant neoplasms of genital organs (**Mal Neop Genital Organs**)**:** C51-C58 'Malignant neoplasms of female genital organs' and C60-63 'Malignant neoplasms of male genital organs'
	Malignant neoplasms of urinary tract (**Mal Neop Urinary Tract**)**:** (C64-68)
	Malignant neoplasm of breast (**Mal Neop Breast**)**:** C50
	Malignant neoplasms of uncertain /unknown behaviour (Mal Neop of uncertain/unknown behaviour): (D37 – D48)
	Benign neoplasms: (D10-D36)
	Metastatic neoplasms: C77 'Secondary and unspecified malignant neoplasm of lymph nodes', C78 'Secondary malignant neoplasm of respiratory and digestive organs', and C79 'Secondary neoplasm of other and unspecified sites'.
	Neoplasm other: ICD-10 Chapter II 'Neoplasms' (code range: C00-D48), excluding codes relating to malignant neoplasms of digestive organs, malignant neoplasms of respiratory and intrathoracic organs, leukaemia or lymphoma, malignant neoplasms of genital organs, malignant neoplasms of urinary tract, malignant neoplasm of breast, malignant neoplasms of uncertain or unknown behaviour, benign neoplasms, and metastatic neoplasms.
III	**Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism:** (D50 – D89)
IV	**Endocrine, nutritional and metabolic disease:** (E00 – E90)
V	**Mental and behavioural disorders:** (F00 – F99)
VI	**Diseases of the nervous system:** (G00-G99)
VII	**Diseases of the eye and adnexa:** (H00-H59)
IX*	**AF/AFL:** I48 'Atrial fibrillation and flutter'
	Cardiac electrophysiology other (Card EP Other): I44 'Atrio-ventricular and left bundle-branch block', I45 'Other conduction disorders', I46 'Cardiac arrest', I47 'Paroxysmal tachycardia', and I49 'Other cardiac arrhythmias'.
	Acute MI: I21 'Acute myocardial infarction'

Continuation of Table C.1

ICD-10 Chapter	Description
	IHD: I20 'Angina pectoris', I22 'Subsequent myocardial infarction', I23 'Certain current complications following acute myocardial infarction', I25.0 Atherosclerotic cardiovascular disease, so described', I25.1 'Atherosclerotic heart disease', I25.2 'Old myocardial infarction', I25.3 'Aneurysm of heart', I25.4 'Coronary artery aneurysm and dissection', I25.6 'Silent myocardial ischaemia', I25.8 'Other forms of chronic ischaemic heart disease', and I25.9 'Chronic ischaemic heart disease, unspecified'.
	Cerebrovascular disease (Cerebrovasc Dis): I60 'Subarachnoid haemorrhage', I61 'Intracerebral haemorrhage', I62 'Other non-traumatic intracranial haemorrhage', I63 'Cerebral infarction', I64 'Stroke, not specified as haemorrhage or infarction', I65 'Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction', I66 'Occlusion and stenosis of pre-cerebral arteries, not resulting in cerebral infarction', I67 'Other cerebrovascular disease', I68 Cerebrovascular disorders in disease classified elsewhere', and I69 'Sequelae of cerebrovascular disease'.
IX*	**Hypertensive disease (HTN Dis):** I10 'Essential (primary) hypertension', I11.9 'Hypertensive heart without (congestive) heart failure, I12'Hypertensive renal disease', I13.1 'Hypertensive heart and renal disease with renal failure', I13.9 'Hypertensive heart and renal disease, unspecified', and I15 'Secondary hypertension'.
	PAD: I73.1 'Thromboangiitis obliterans [Buerger]', I73.8 'Other specified peripheral vascular disease', I73.9 'Peripheral vascular disease, unspecified', I74.3 'Embolism and thrombosis of arteries in lower extremities', I74.4 'Embolism and thrombosis of arteries of extremities, unspecified', and I74.5 'Embolism and thrombosis of iliac artery'.
	Pulm Dis: I26, 'Pulmonary embolism', I27 'Other pulmonary heart disease', and I28 'Other diseases of pulmonary vessels'.

Continuation of Table C.1

ICD-10 Chapter	Description
IX*	**Valve disease (Valve Dis):** I05 'Rheumatic mitral valve disease', I06 'Rheumatic aortic valve disease' , I07 'Rheumatic tricuspid valve disease' , I08 'Multiple valve disease', I34 'Non-rheumatic mitral valve disorders', I35 'Nonrheumatic aortic valve disorders', I36 'Non-rheumatic tricuspid valve disorders', and I37 'Pulmonary valve disorders'.
	HF: 150 'Heart failure', I42.0 'Dilated cardiomyopathy', I42.9 'Cardiomyopathy, unspecified', I11.0 'Hypertensive heart disease with (congestive) heart failure, I25.5 'Ischaemic cardiomyopathy', I13.0 'Hypertensive heart and renal disease with (congestive) heart failure', I13.2 'Hypertensive heart and renal disease with both (congestive) heart failure and renal failure.
	Other cardiovascular disorders (CV Other): ICD-10 Chapter IX 'Diseases of the circulatory system' (code range: I00-I99), excluding codes relating to AF/AFL, other cardiac electrophysiology, acute MI, IHD, cerebrovascular disease, Pulm Dis, hypertensive disease, PAD, valve disease, and HF.
X	**Disease of the respiratory system:**(J00 – J99)
	Resp infection: J00 'Acute nasopharyngitis [common cold]', J01 'Acute sinusitis', J02 'Acute pharyngitis', J03 'Acute tonsillitis', J04 'Acute laryngitis and tracheitis', J05 'Acute obstructive laryngitis [croup] and epiglottitis', J06 'Acute upper respiratory infections of multiple and unspecified sites', J09 'Influenza due to identified zoonotic or pandemic influenza virus', J10 'Influenza due to identified seasonal influenza virus', J11 'Influenza, virus not identified', J12 'Viral pneumonia, not elsewhere classified', J13 'Pneumonia due to Streptococcus pneumoniae', J14 'Pneumonia due to Haemophilus influenzae', J15 'Bacterial pneumonia, not elsewhere classified', J16 'Pneumonia due to other infectious organisms, not elsewhere classified', J17 'Pneumonia in diseases classified elsewhere', J18 'Pneumonia, organism unspecified', J20 'Acute bronchitis', J21 'Acute bronchiolitis', J22 'Unspecified acute lower respiratory infection'
	COPD: J43 'Emphysema' and J44 'Other chronic obstructive pulmonary disease'
	Pulmonary oedema: J81 'Pulmonary oedema'

Continuation of Table C.1

ICD-10 Chapter	Description
	Pleural effusions: J90 'Pleural effusion, not elsewhere classified' and J91 'Pleural effusion in conditions classified elsewhere'
	Resp other: ICD-10 Chapter X codes not included in resp infection, COPD, and pulmonary oedema.
XI	**Diseases of the digestive system:** (K00 – K14)
XII	**Disease of the skin and subcutaneous tissue:** (L00 – L99)
XIII	**Diseases of the musculoskeletal system and connective tissue:** (M00 – M99)
XIV	**Diseases of the genitourinary system:** (N00 – N99)
XVIII	**Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified:** (R00 – R99)
	S&S Abnormal breathing: R06 'Abnormalities of breathing'
	S&S Unpsec chest pain: R07.4 'Chest pain, unspecified'
	S&S Circ/Resp systems other: R00 'Abnormalities of heart beat', R01 'Cardiac murmurs and other cardiac sounds', R02 'Gangrene, not elsewhere classified', R03 'Abnormal blood-pressure reading, without diagnosis', R04 'Haemorrhage from respiratory passages', R05 'Cough', R07.0 'Pain in throat', R07.1 'Chest pain on breathing', R07.2 'Precordial pain', R07.3 'Other chest pain', R09 'Other symptoms and signs involving the circulatory and respiratory systems'
	S&S Syncope: R55 'Syncope and collapse'
	S&S Nervous and MSK: R25 'Abnormal involuntary movements', R26 'Abnormalities of gait and mobility', R27 'Other lack of coordination', R29 'Other symptoms and signs involving the nervous and musculoskeletal systems'
	S&S Digest/Abdomen: R10 'Abdominal and pelvic pain', R11 'Nausea and vomiting', R12 'Heartburn', R13 'Dysphagia', R14 'Flatulence and related conditions', R15 'Faecal incontinence', R16 'Hepatomegaly and splenomegaly, not elsewhere classified', R17 'Unspecified jaundice', R18 'Ascites', R19 'Other symptoms and signs involving the digestive system and abdomen'
	S&S Urinary system: R30 'Pain associated with micturition', R31 'Unspecified haematuria', R32 'Unspecified urinary incontinence', R33 'Retention of urine', R34 'Anuria and oliguria', R35 'Polyuria', R36 'Urethral discharge', R39 'Other symptoms and signs involving the urinary system'

Continuation of Table C.1

ICD-10 Chapter	Description
XVIII	**S&S Cog/Precep/Emot/Behav:** R40 'Somnolence, stupor and coma', R41 'Other symptoms and signs involving cognitive functions and awareness', R42 'Dizziness and giddiness', R43 'Disturbances of smell and taste', R44 'Other symptoms and signs involving general sensations and perceptions', R45 'Symptoms and signs involving emotional state', R46 'Symptoms and signs involving appearance and behavior'
	S&S Oedema: R60 'Oedema, not elsewhere classified'
	S&S General: R50 'Fever of other and unknown origin', R5 'Headache', R52 'Pain, not elsewhere classified', R53 'Malaise and fatigue', R54 'Senility', R56 'Convulsions, not elsewhere classified', R57 'Shock, not elsewhere classified', R58 'Haemorrhage, not elsewhere classified', R59 'Enlarged lymph nodes', R61 'Hyperhidrosis', R62 'Lack of expected normal physiological development', R63 'Symptoms and signs concerning food and fluid intake', R64 'Cachexia', R65 'Systematic Inflammatory Response Syndrome [SIRS]', R68 'Other general symptoms and signs', R69 'Unknown and unspecified causes of morbidity'.
	S&S Other: ICD-10 Chapter XVIII codes not in S&S Abnormal breathing, S&S Unspec Chest Pain, S&S Circ/Resp Systems Other, S&S Syncope, S&S Nervous and MSK, S&S Digest/Abdomen, S&S Urinary system, S&S Cog/Precep/Emot/Behav, S&S Oedema, and S&S General.
XIX	**Injury, poisoning, and certain other consequences of external causes:** (S00 – S99, T00 – T98)
Other	**Other ICD-10 Chapters grouped due to reporting requirements:** Chapter XV 'Pregnancy, childbirth, and the puerperium' (O00 – O99), Chapter XVI 'Certain conditions originating in the perinatal period' (P00 – P96), Chapter XVII 'Congenital malformations, deformations, and chromosomal abnormalities' (Q00 – Q99), Chapter XX 'External causes of morbidity and mortality', and Chapter XXII 'Codes for special purposes'

Mal neop, malignant neoplasms; Pulm Dis, Pulmonary heart disease and diseases of pulmonary circulation; Resp infection, respiratory infection; COPD, chronic obstructive pulmonary disease; S&S,symptoms and signs; Unspec Chest Pain, unspecified chest pain.
* Chapter split into constituent causes of hospitalisation for admission network graphs.

C.2 Patient Flow Diagram

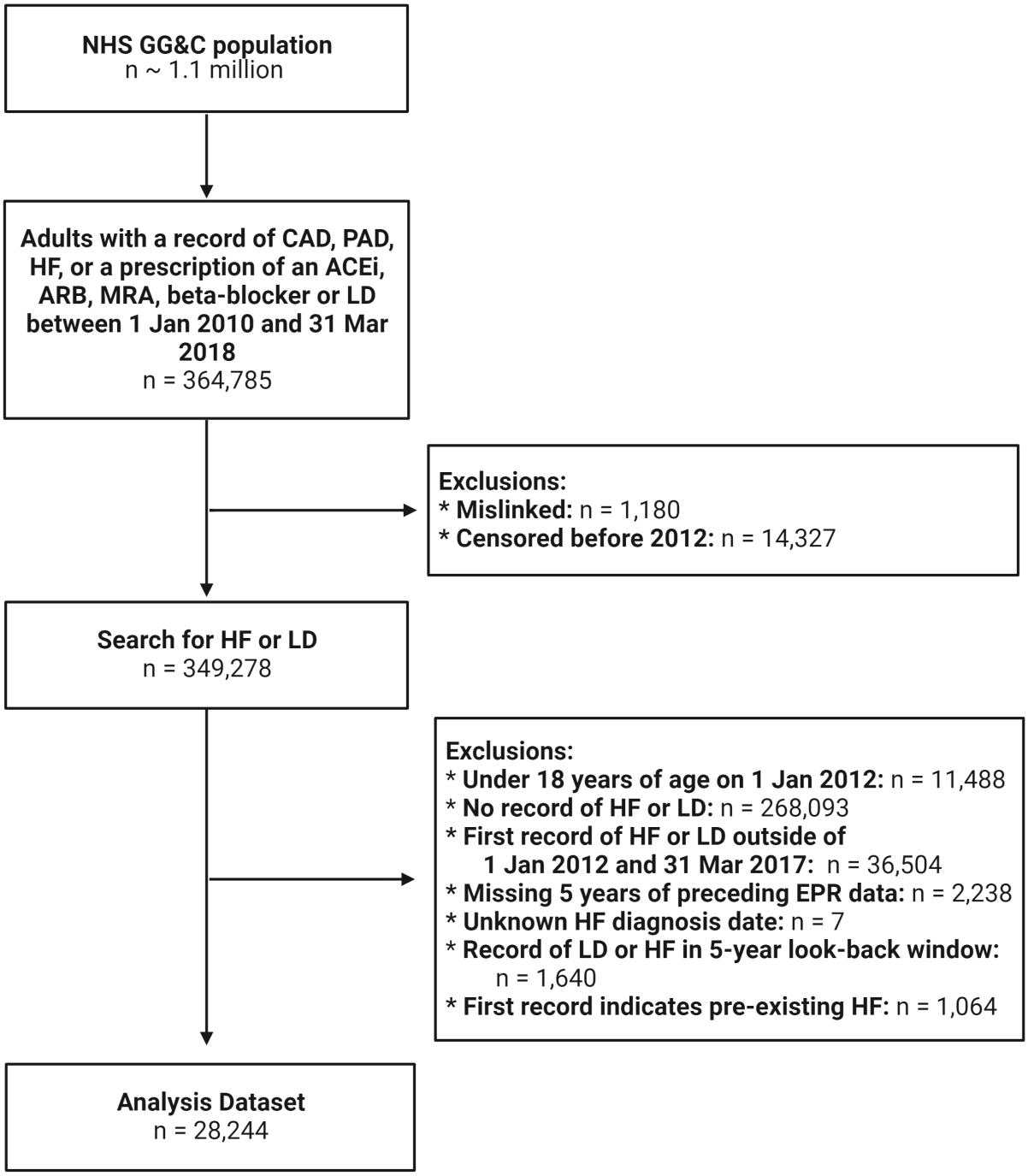


Figure C.1: Flow diagram of inclusion and exclusion criteria for NHS GG&C patients

C.3 Hospital Admissions

Note: A path is the order and number of admissions experienced by a patient, 'Start' node 12 months prior to a diagnosis of HF or the initiation of LD, followed by a node for each admission, and terminating with a node labelled according to the patient's group. Nodes are hospital admissions, and edges are ordered pairs of admissions. The node degree is the number of entry and exit edges through a node. Weights are a scaling metric (see equation 4.1) normalised to represent how frequently an edge is traversed. The weighted degree of a node is the sum of the weights on the entry and exit edges through the node. The average path length is the average number of edges along the shortest path for all possible pairs of nodes. The average degree is defined by the total number of edges divided by the total number of nodes.

Table C.2: Statistics on the network graphs representing the admissions patterns starting 12 months before an index event (e.g., first of diagnosis of HF or initiation of LD).

Variable	**LD Only**	**Both: LD First**	**Both: Together**	**Both: HF First**	**HF Only**	**HF Death as First Record**
n	18,596	1,251	1,876	1,854	3,957	710
Admitted	9,144 (49%)	643 (51%)	817 (44%)	912 (49%)	1,836 (46%)	397 (56%)
Num. admit	17,153	1,208	1,336	1,524	2,988	771
Path Statistics						
Avg. path length	1.8	1.0	1.0	1.0	1.1	1.0
Top 3 paths based on occurrence (%)*	XVIII: 912 (10); X: 789 (9); & XIX: 482 (5)	X: 75 (12); XVIII: 71 (11); & Acute MI: 34 (5)	Acute MI: 91 (11); X: 86 (8); &XVIII: 76 (9)	Acute MI: 108 (12); XVIII: 87 (10);& X: 87 (10)	Acute MI: 435 (20); XVIII: 180 (8); &IHD: 107 (5)	XVIII: 52 (16%), X: 39 (12%), & XIX: 32 (10%)
Node Statistics						
Avg. degree	5.0	1.2	2.0	2.0	2.4	1.8
Degree	6 (3.5-16)	2 (2-2)	2 (2-2.5)	2 (2-2.5)	2 (2-4.5)	2 (2-4)

Continuation of Table C.2

Variable	**LD Only**	**Both: LD First**	**Both: Together**	**Both: HF First**	**HF Only**	**HF Death as First Record**
n	18,596	1,251	1,876	1,854	3,957	710
Top 3 nodes by degree (degree)*	XVIII (30), X (25), & XIX (18)	X (5), XVIII (3),& Acute MI (2)	X (5), II (4), & XVIII (3)	XVIII (4), X (4), & XI (4)	Acute MI (9), XVIII (9), & X (6)	XVIII (4), X (4), & XIX (2)
Avg. weighted degree	3.7	5.9	5.3	5.0	4.3	8.3
Weighted degree	4.9 (3.2 - 9.8)	5.5 (3.9 - 12.9)	6.2 (2.6 - 9.2)	4.9 (2.7 - 8.5)	4.2 (3.3 - 7.9)	9.4 (8.5 - 15.7)
Top 3 nodes by weighted degree (weight)*	XVIII (20.9), X (14.0), & XIII (12.7)	XVIII (16.2), Acute MI (13.3), & AF/AFL (12.9)	Acute MI (22.4), AF/AFL (12.1), & XVIII (11.1)	Acute MI (28.0), XIX (9.1), & AF/AFL (9.1)	XIX (15.3) IHD (13.5), & XIII (10.3)	XIX (27.2), II (11.9), & X (11.9)
Edge Statistics						
Diameter	3	2	2	2	3	2
Top 3 edges by weight (weight)	Start→ CV Other (5.7), Pulm Dis→ LD Only (5.1), & Start→ Valve Dis (3.8)	Start→ XVIII (13.9), Start→ Acute MI (6.7), Acute MI→ Both: LD First (6.7)	Start→ Acute MI (11.2) Acute MI→ Both: Together (11.2), & Start →AF/AFL (9.4)	Start→ Acute MI (14.0), Acute MI→ Both: HF First (14.0), & Start→ AF/AFL (4.5)	Start→ IHD (7.6), Start→ XIX (7.6), & XIX→ HF Only (7.6)	Start→ XIX (13.6), XIX→ HF Death as First Record (13.60), & Start→ HF Death as First Record (6.0)

Continuation of Table C.2

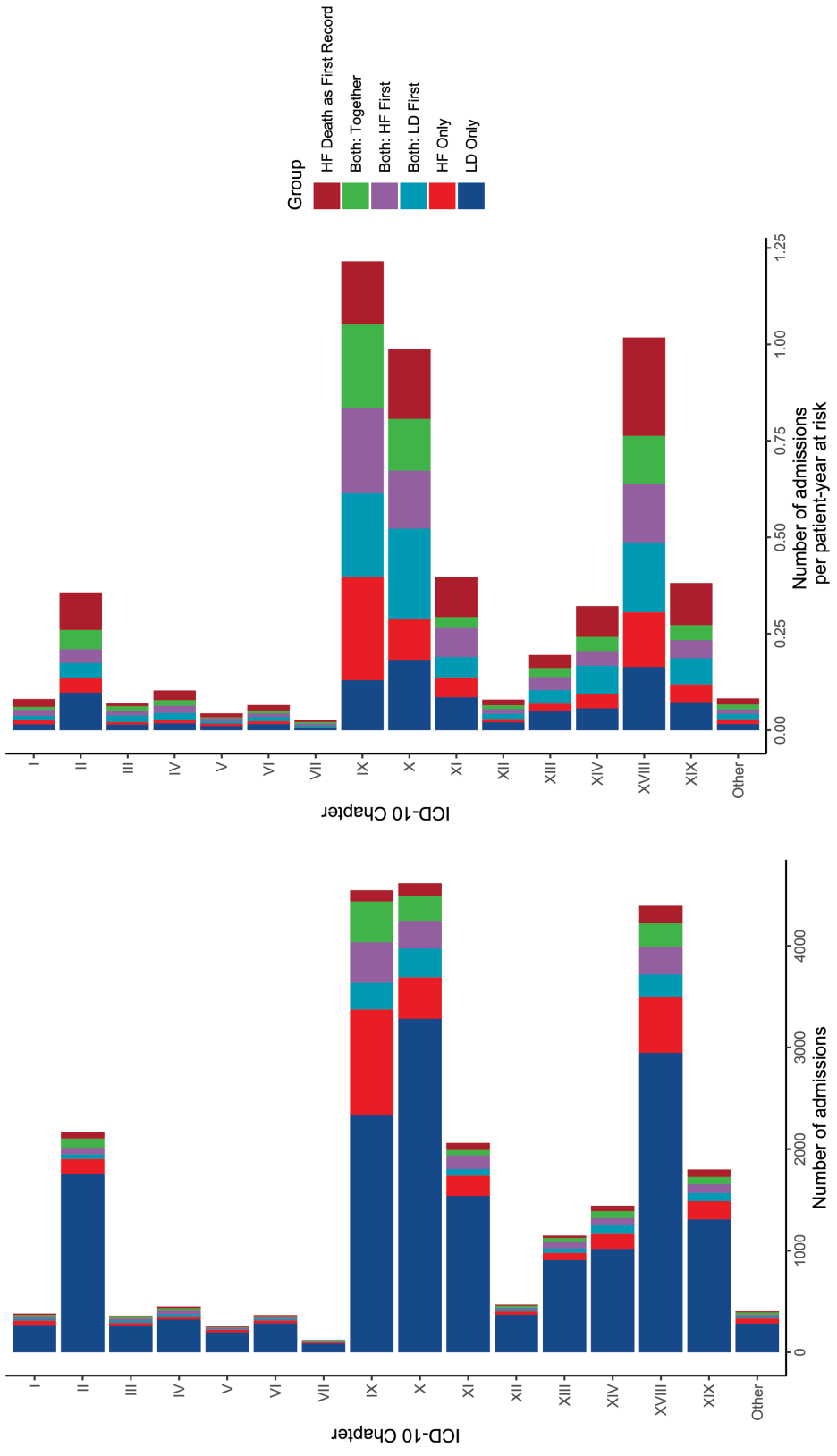
Variable	LD Only	Both: LD First	Both: Together	Both: HF First	HF Only	HF Death as First Record
n	18,596	1,251	1,876	1,854	3,957	710

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values unless otherwise specified.

* Start and terminal nodes excluded from reporting.

Avg, average; Num. admit, Total number of admissions; Valve Dis, valve disease;

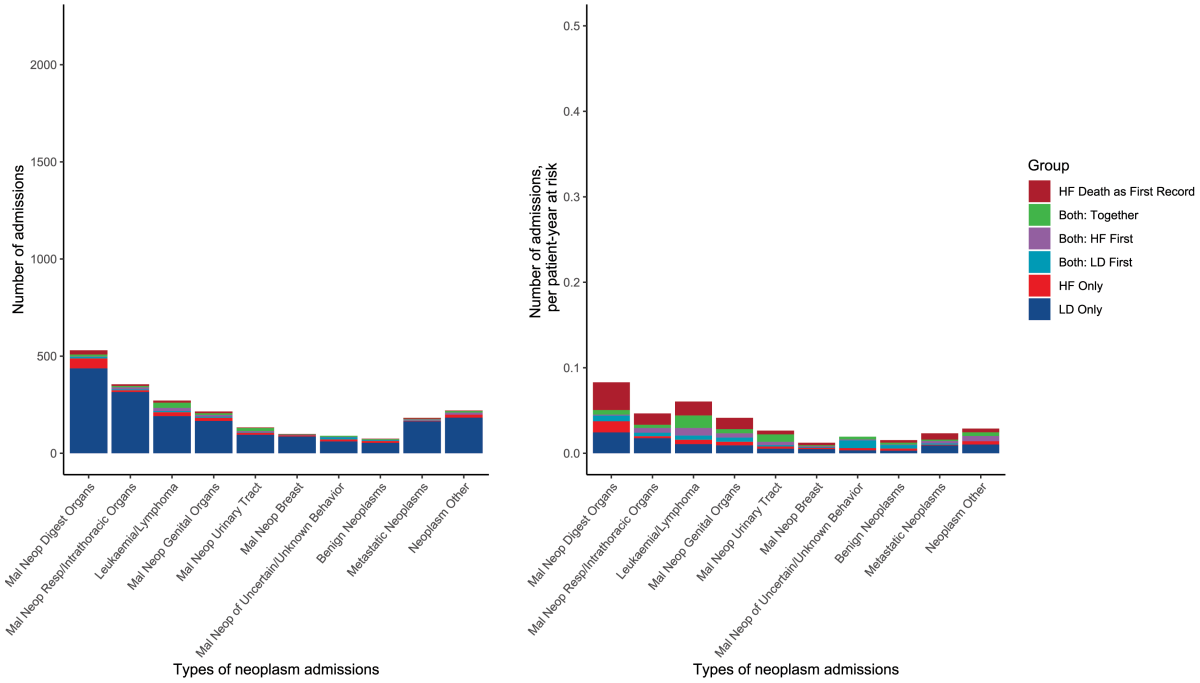
ICD-10 Chapters: I, infections & parasitic diseases; II, neoplasms; III, blood and blood -forming organs; IV, endocrine, nutritional & metabolic; V, mental and behavioural; VI, nervous system; VII, eye and adnexa; VIII, ear and mastoid process; IX, circulatory system split up into AF/AFL, Card EP Other, IHD, acute MI, cerebrovascular disease (Cerebrovasc. Dis), hypertensive disease (HTN Dis), peripheral arterial disease (PAD), pulmonary disease (Pulm Dis), and Other CV; X, respiratory system; XI digestive system; XII, Skin and subcutaneous tissue; XIII, musculoskeletal system; XIV, genitourinary system; XVIII, non-specific S&S; XIX injury; XXI, factors influencing health status.



(a) Number of admissions

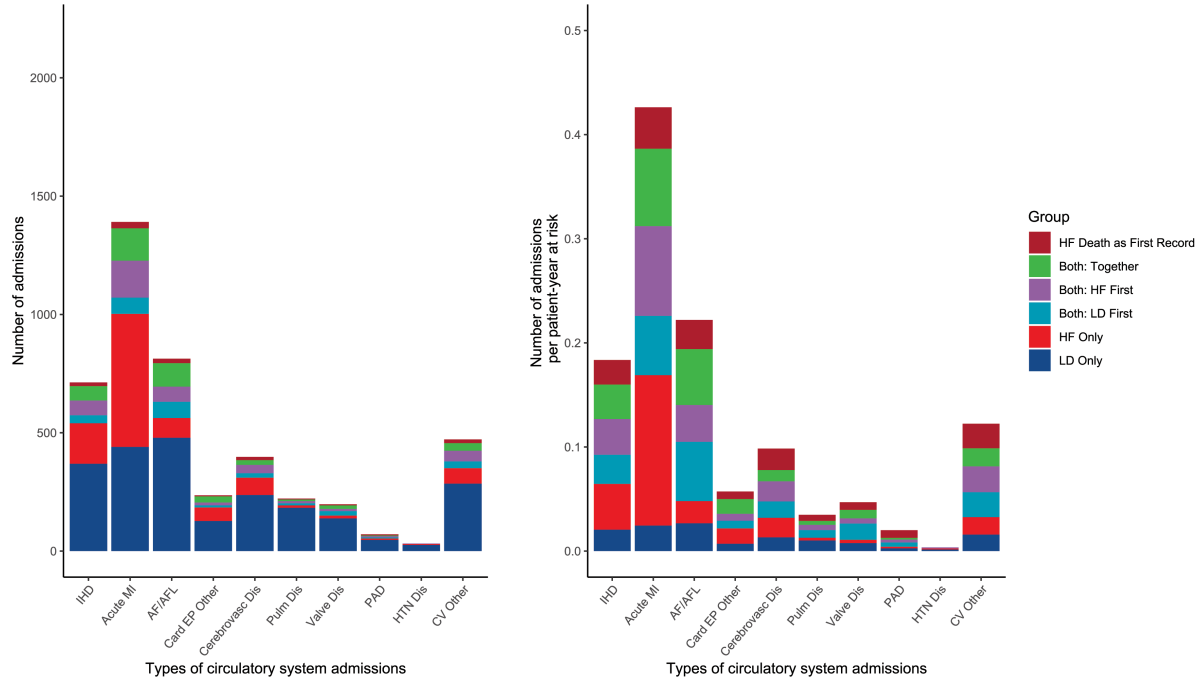
(b) Scaled per patient-year at risk

Figure C.2: Admissions by primary reason in the first year prior to an event by ICD-10 Chapter. ICD-10 Chapters: I, infections & parasitic diseases; II, neoplasms; III, blood and blood-forming organs; IV, endocrine, nutritional & metabolic; V, mental and behavioural; VI, nervous system; VII, eye and adnexa; VIII, ear and mastoid process; IX, circulatory system; X, respiratory system; XI digestive system; XII, Skin and subcutaneous tissue; XIII, musculoskeletal system; XIV, genitourinary system; XVIII, symptoms and signs; XIX injury; XXI, factors influencing health status.



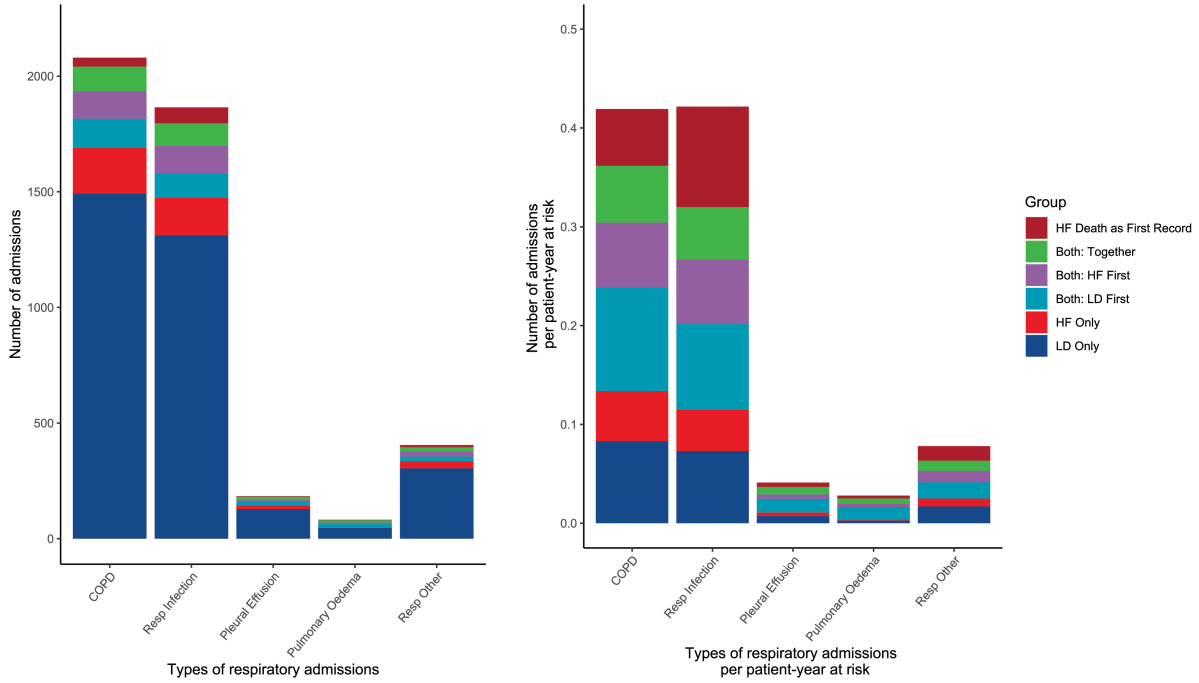
(a) Number of admissions (b) Scaled per patient-year at risk

Figure C.3: Neoplasm admissions in 12 months before inclusion



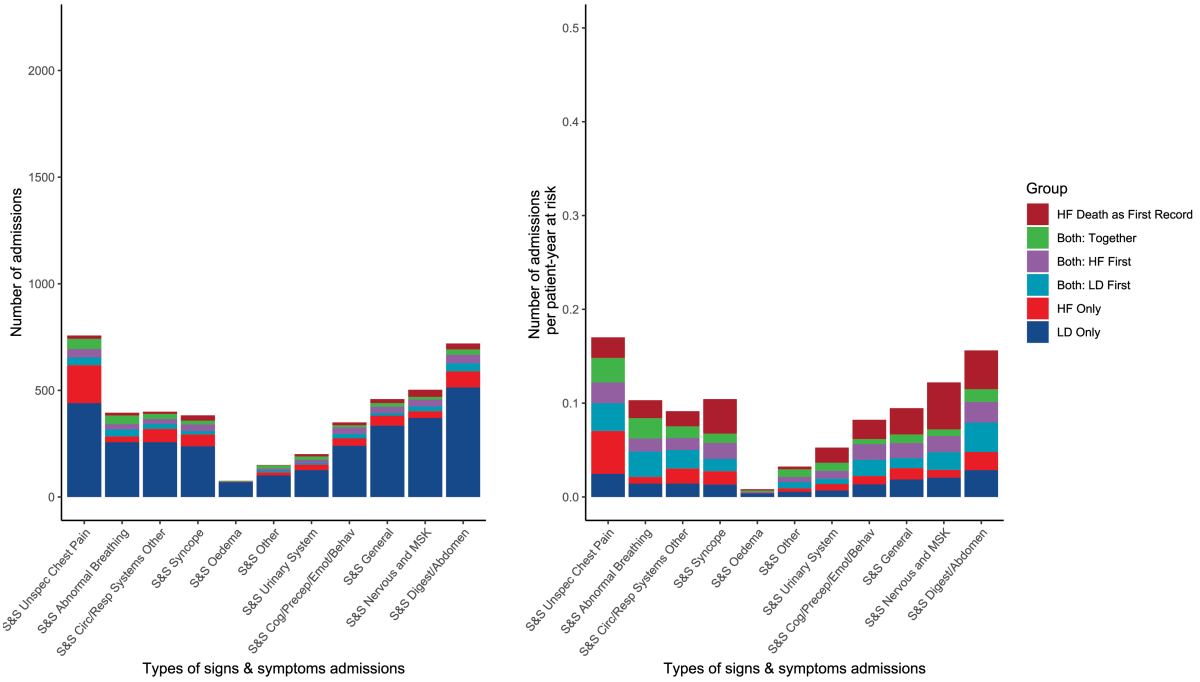
(a) Number of admissions (b) Scaled per patient-year at risk

Figure C.4: Circulatory system admissions in 12 months before inclusion



(a) Number of admissions (b) Scaled per patient-year at risk

Figure C.5: Respiratory system admissions in 12 months before inclusion



(a) Number of admissions (b) Scaled per patient-year at risk

Figure C.6: Symptoms and signs (S&S) admissions in 12 months before inclusion

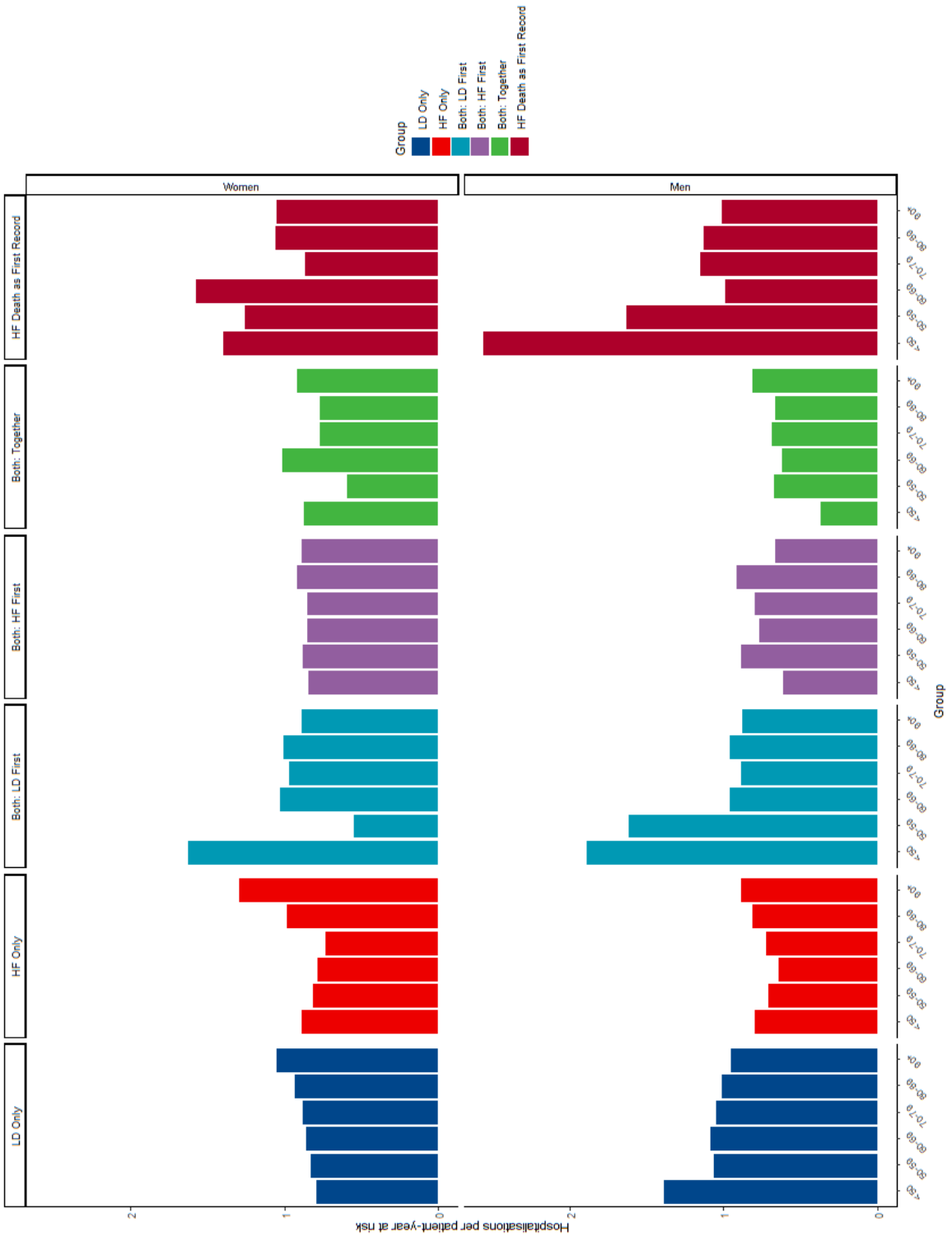
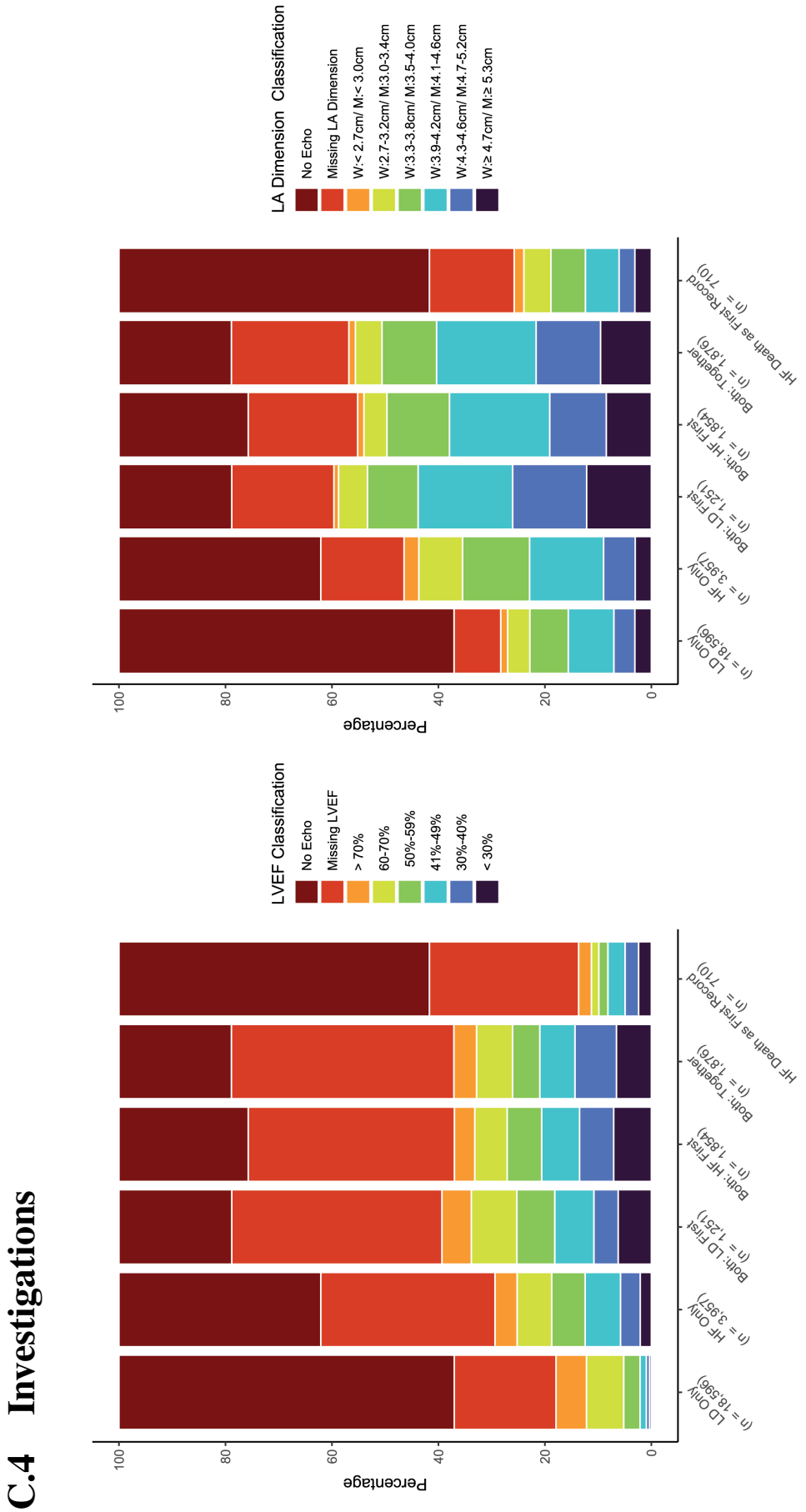


Figure C.7: Hospital admissions per patient-year at risk, stratified by age, sex, and group.



(a) Left ventricular ejection fraction (LVEF)

(b) Left atrial (LA) diameter

Figure C.8: Results from the closest echocardiogram to the index event within the observation window (two years before and one year following the HF diagnosis or LD initiation).

C.4 Investigations

Appendix D

Supplementary Material for Chapter 5

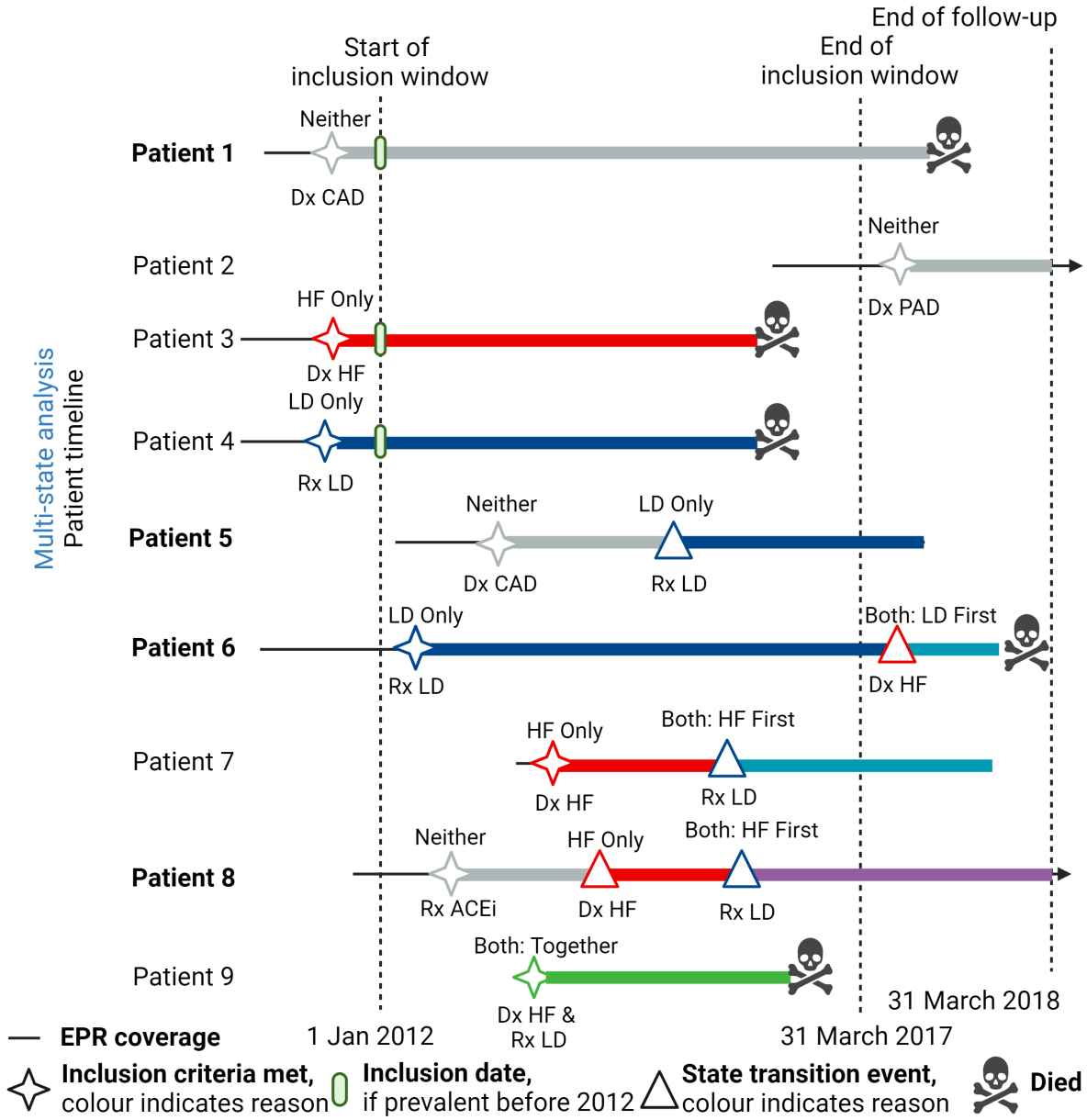


Figure D.1: Schematic showing examples for patient inclusion and exclusion, including the events that cause state changes. Patients in **bold** are included in the analysis. Patients 5, 6, and 8 experienced at least one state change due to receiving a diagnosis of HF or starting a repeat LD prescription before they were censored. Patient 2 was excluded as his or her inclusion date after the 31st March 2017. Patient 3 was excluded due to a pre-existing diagnosis of HF on the 1st January 2012, and the analogous reasoning applies to patient 4 where a repeat LD was started before 1st January 2012. Finally, patients 7 and 9 were excluded due to lacking at least five years of EPR data prior to inclusion. Dx, Diagnosis of; Rx, Prescription of.

D.1 Patient Flow Diagram

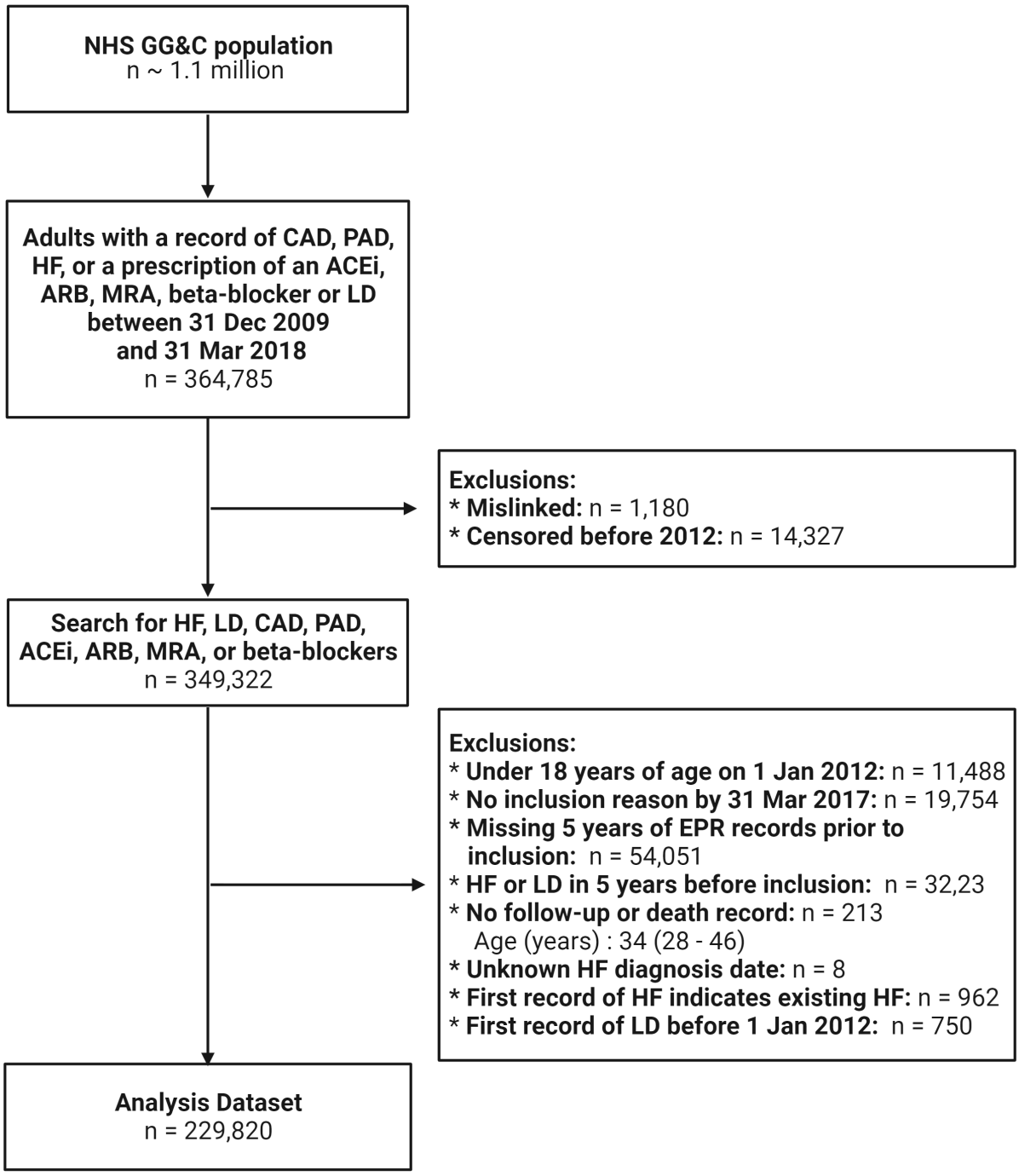


Figure D.2: Patient flow diagram of inclusion and exclusion criteria for NHS GG&C patients. The number of prevalent LD or HF cases is larger than the number reported in Chapter 3 as case history includes records back to 2000, not just mentions between 2009 through 2011.

D.2 Reasons for State Transitions

Table D.1: List of allowed patient transitions between states based on receiving a diagnosis of HF, initiating a repeat LD, and/or dying.

From	To	Cause
Before cohort entry	Neither	Diagnosis of CAD, PAD, or a prescription of an ACEi, ARB, MRA, beta-blocker, or a one-off LD prescription.
	LD Only	Starting a repeat LD prescription in the absence of a record for CAD, PAD, or a prescription of an ACEi, ARB, MRA, beta-blocker before 1 January 2012.
	HF Only	Diagnosis of HF in the absence of a record of CAD, PAD, or a prescription of an ACEi, ARB, MRA, beta-blocker, or a repeat LD prescription before 1 January 2012.
	Both: Together	Diagnosis of HF and initiation of a repeat LD prescription within 30-days of each other in the absence of a record of CAD, PAD, or a prescription of an ACEi, ARB, MRA, beta-blocker, or a repeat LD prescription before 1 January 2012.
Neither	LD Only	Starting a repeat LD prescription in the absence of a heart failure diagnosis
	HF Only	Diagnosis of HF in the absence of a repeat LD prescription
	Both: Together	Diagnosis of HF and initiation of a repeat LD prescription within 30 days.
	Death	Dying before a diagnosis of HF or initiating a repeat LD prescription
LD Only	Both: LD First	Diagnosis of HF more than 30-days after the initiation of a repeat LD prescription
	Death	Dying before receiving a diagnosis of HF
HF Only	Both: HF First	Starting a repeat LD prescription more than 30-days after a HF diagnosis.
	Death	Dying without starting a repeat LD prescription
Both: LD First	Death	Dying after starting a repeat LD prescription and being diagnosed with HF.
Both: HF First	Death	Dying after receiving a HF diagnosis and starting a repeat LD prescription.
Both: Together	Death	Dying after starting a repeat LD prescription and being diagnosed with HF.

D.3 Tables Showing Patient Progression

Table D.2: Patient demographics and comorbidities based on the date of cohort inclusion for patients who were not subsequently diagnosed with HF nor were initiated on repeat LD, the date of receiving a diagnosis of HF, and the date of initiating a repeat LD prescription.

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**Any LD**	**Any HF**
n	112,811	87,940	25,206	10,259
Age (years)	47 (37 - 54)	71 (65 - 77)	74 (64 - 82)	74 (64 - 82)
Sex				
Women	65,816 (58%)	47,371 (54%)	14,851 (59%)	4,569 (45%)
Men	46,995 (42%)	40,569 (46%)	10,355 (41%)	5,690 (55%)
Ethnicity				
White	69,121 (61%)	68,605 (78%)	22,343 (89%)	9,361 (91%)
Missing	39,720 (35%)	17,367 (20%)	2,328 (9%)	622 (6%)
Other	3,970 (4%)	1,968 (2%)	535 (2%)	276 (3%)
Socioeconomic deprivation (SIMD)				
1 (most deprived)	48,532 (43%)	30,292 (34%)	10,461 (42%)	4,187 (41%)
2	20,716 (18%)	15,595 (18%)	4,716 (19%)	1,948 (19%)
3	15,241 (14%)	11,633 (13%)	3,432 (14%)	1,360 (13%)
4	12,308 (11%)	11,919 (14%)	2,899 (12%)	1,150 (11%)
5 (least deprived)	16,014 (14%)	18,501 (21%)	3,698 (15%)	1,614 (16%)
Healthcare contact				
HF diagnosed in PC	N/A	N/A	1,793 (7%)	3,538 (34%)
LD started within 30 days post-secondary care contact				
Any	N/A	N/A	12,556 (50%)	4,148 (40%)
Hospital ☒	N/A	N/A	7,898 (31%)	3,089 (30%)
Clinic	N/A	N/A	6,730 (27%)	1,757 (17%)
CV specialist ∮	N/A	N/A	1,481 (6%)	685 (7%)
Comorbidities *				
H/o hypertension	15,977 (14%)	31,389 (36%)	11,019 (44%)	5,260 (51%)
DM	13,537 (12%)	20,596 (23%)	6,738 (27%)	2,937 (29%)
Thyroid disease	1,027 (1%)	1,997 (2%)	1,117 (4%)	492 (5%)
CAD	8,894 (8%)	21,694 (25%)	8,091 (32%)	5,894 (57%)
Of which is MI	4,464 (4%)	7,766 (9%)	3,908 (16%)	4,034 (39%)
Valve disease	427 (<1%)	1,203 (1%)	2,238 (9%)	1,981 (19%)
AF/AFL	5,386 (21%)	3,529 (34%)	1,285 (1%)	5,882 (7%)

Continuation of Table D.2

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	112,811	87,940	25,206	10,259
PAD	845 (1%)	2,557 (3%)	1,241 (5%)	705 (7%)
Stroke	2,194 (2%)	7,270 (8%)	3,227 (13%)	1,509 (15%)
COPD	5,362 (5%)	9,068 (10%)	5,941 (24%)	2,633 (26%)
Cancer	2,281 (2%)	7,787 (9%)	3,675 (15%)	1,288 (13%)
Dementia	39 (<1%)	1,953 (2%)	1,300 (5%)	456 (4%)
EP devices	168 (<1%)	885 (1%)	770 (3%)	461 (4%)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

☒ SMR01 hospital discharge includes day-cases and inpatient stays.

∮ CV specialties include cardiology, cardiac surgery, cardiothoracic surgery, and vascular surgery.

∗ History of a coded record on or before the inclusion date.

PC, primary care; EP devices, cardiac electrophysiology devices (e.g., pacemaker, ICD, or CRT).

Table D.3: Concurrent medication levels for patients defined by a dispensed medication in the 180 days leading up to cohort inclusion for patients who were not subsequently diagnosed with HF nor were initiated on repeat LD by age, the date of receiving a diagnosis of HF, and the date of initiating a repeat LD prescription.

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	112,811	87,940	25,206	10,259
Age (years)	47 (37 - 54)	71 (65 - 77)	74 (64 - 82)	74 (64 - 82)
ACEi or ARB	33,186 (29%)	49,520 (56%)	12,263 (49%)	4,997 (49%)
ACEi	27,764 (25%)	38,189 (43%)	9,350 (37%)	3,908 (38%)
ARB	5,987 (5%)	12,233 (14%)	3,344 (13%)	1,224 (12%)
Beta-blocker	43,284 (38%)	33,436 (38%)	10,228 (41%)	4,264 (42%)
MRA	438 (<1%)	452 (1%)	1,368 (5%)	279 (3%)
CCB	12,431 (11%)	29,000 (33%)	8,595 (34%)	3,065 (30%)
Diltiazem/Verapamil	990 (1%)	3,167 (4%)	1,400 (6%)	508 (5%)
Dihydropyridine	11,488 (10%)	25,965 (30%)	7,308 (29%)	2,589 (25%)
Digoxin	100 (<1%)	1,078 (1%)	1,579 (6%)	537 (5%)
Thiazides+	9,508 (8%)	25,383 (29%)	5,659 (22%)	1,467 (14%)
Low dose aspirin	12,285 (11%)	34,727 (39%)	9,425 (37%)	4,066 (40%)

Continuation of Table D.3

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	112,811	87,940	25,206	10,259
Oral anticoagulants	763 (1%)	3,318 (4%)	3,968 (16%)	1,621 (16%)
Lipid regulators	22,360 (20%)	51,151 (58%)	13,806 (55%)	5,794 (56%)
Bronchodilators	10,053 (9%)	11,444 (13%)	6,954 (28%)	2,580 (25%)
Thyroid medications	4,787 (4%)	8,050 (9%)	2,708 (11%)	920 (9%)
Hypoglycaemic agents	9,528 (8%)	12,233 (14%)	4,280 (17%)	1,840 (18%)
Insulin†	2,387 (2%)	1,699 (2%)	1,123 (4%)	459 (4%)
Other hypo-glycaemic agents	7,984 (7%)	11,436 (13%)	3,798 (15%)	1,631 (16%)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

† Either alone or in combination with another agent.

Table D.4: Most recent blood results in the prior two years prior to cohort inclusion for patients who were not subsequently diagnosed with HF nor were initiated on repeat LD by age, the date of receiving a diagnosis of HF, and starting a repeat LD prescription.

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	112,811	87,940	25,206	10,259
Age (years)	47 (37 - 54)	71 (65 - 77)	74 (64 - 82)	74 (64 - 82)
Haemoglobin (Hb) (mg/dL)				
Reported	79,776 (71%)	69,800 (79%)	23,911 (95%)	9,881 (96%)
Women: Hb	13.4 (12.6 - 14.1)	13.1 (12.1 - 13.9)	12.4 (11.2 - 13.6)	12.5 (11.1 - 13.7)
Men: Hb	15.0 (14.2 - 15.8)	14.2 (13.2 - 15.2)	13.1 (11.6 - 14.5)	13.8 (12.1 - 15.0)
Anaemia◇	8,954 (11%)	14,821 (21%)	10,083 (42%)	3,699 (37%)
Estimated glomerular filtration rate (eGFR)∇ (mL/min/1.73m^2)				
Reported	85,626 (76%)	81,903 (93%)	24,521 (97%)	10,052 (98%)
eGFR	101 (92 - 108)	79 (66 - 89)	75 (58 - 88)	72 (52 - 86)
eGFR [30 - 60)	1,105 (1%)	12,820 (16%)	5,596 (23%)	2,656 (26%)
eGFR <30	191 (<1%)	899 (1%)	1,037 (4%)	695 (7%)
Serum values (mmol/L)				
Reported urea	85,717 (76%)	81,927 (93%)	24,522 (97%)	10,053 (98%)
Urea	4.6 (3.7 - 5.6)	5.8 (4.8 - 7.1)	6.2 (4.8 - 8.2)	6.5 (5.0 - 8.9)

Continuation of Table D.4

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	112,811	87,940	25,206	10,259
Reported albumin	82,939 (74%)	79,860 (91%)	24,284 (96%)	9,974 (97%)
Albumin	39 (37 - 41)	38 (36 - 39)	35 (31 - 37)	35 (32 - 38)
Albumin [30-35)	6,401 (8%)	10,526 (13%)	7,147 (29%)	2,990 (30%)
Albumin <30	2,482 (3%)	2,965 (4%)	4,662 (19%)	1,376 (14%)
Reported sodium	85,679 (76%)	81,913 (93%)	24,526 (97%)	10,055 (98%)
Sodium	139 (137 - 140)	139 (137 - 140)	139 (136 - 141)	138 (136 - 140)
Sodium <135	485 (1%)	998 (1%)	3,251 (13%)	1,741 (17%)
Reported chloride	85,634 (76%)	81,900 (93%)	24,522 (97%)	10,052 (98%)
Chloride	104 (103 - 106)	104 (102 - 106)	103 (100 - 106)	104 (101 - 106)
Reported Potassium	84,773 (75%)	81,535 (93%)	24,479 (97%)	10,017 (98%)
Potassium	4.2 (4.0 - 4.5)	4.3 (4.0 - 4.5)	4.3 (4.0 - 4.6)	4.3 (4.0 - 4.6)
Potassium >6.0	23 (<1%)	32 (<1%)	28 (<1%)	473 (5%)
Potassium <3.5	1,438 (2%)	1,915 (2%)	974 (4%)	51 (1%)
Reported bicarb.	36,554 (32%)	43,115 (49%)	9,460 (38%)	4,097 (40%)
Bicarbonate	23 (21 - 25)	24 (22 - 26)	23 (21 - 26)	22 (20 - 25)

Data are frequencies (%) for categorical values or median (1^{st} - 3^{rd} quartile) for continuous values.

◇ Using the WHO definition of anaemia (see Section 2.6.2).

∇ Calculated eGFR assuming no pregnancies and without adjusting for ethnicity (see Section 2.5.9).

Table D.5: ECG results from the closest test to cohort inclusion for patients who were not diagnosed with HF nor were initiated on repeat LD by age, the date of receiving a diagnosis of HF, and the date of initiating a repeat LD.

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	112,811	87,940	25,206	10,259
Age (years)	47 (37 - 54)	71 (65 - 77)	74 (64 - 82)	74 (64 - 82)
ECG available	47,578 (42%)	48,795 (55%)	19,600 (78%)	8,851 (86%)
Rhythm				
Sinus	46,577 (98%)	43,147 (88%)	14,697 (75%)	5,989 (68%)
Heart rate (bpm)	75 (65 - 87)	72 (62 - 84)	78 (67 - 91)	76 (65 - 91)

Continuation of Table D.5

Variable	**Neither (18-59 yrs)**	**Neither (≥60 yrs)**	**Any LD**	**Any HF**
n	112,811	87,940	25,206	10,259
AF/AFL/SVT	775 (2%)	4,635 (9%)	4,120 (21%)	2,338 (26%)
Heart rate (bpm)	100 (81 - 126)	88 (73 - 109)	87 (72 - 107)	92 (75 - 115)
Other	162 (<1%)	759 (2%)	568 (3%)	363 (4%)
Heart rate (bpm)	78 (67 - 98)	68 (60 - 80)	69 (60 - 80)	68 (60 - 80)
Undetermined	64 (<1%)	254 (<1%)	215 (1%)	161 (2%)
Heart rate (bpm)	86 (73 - 101)	88 (71 - 108)	85 (70 - 108)	85 (71 - 110)
QRS duration (ms)	86 (80 - 94)	86 (80 - 96)	88 (80 - 102)	94 (84 - 112)
QRS ≥120 ms	905 (2%)	3,574 (7%)	2,540 (13%)	1,853 (21%)
QTc available ■	45,629 (40%)	46,680 (53%)	2,540 (13%)	1,853 (21%)
QTc (ms)	413 (400 - 427)	419 (404 - 435)	426 (408 - 447)	432 (411 - 457)
Prolonged QTc	1,856 (4%)	4,246 (9%)	3,491 (18%)	2,330 (27%)
ST-T abnormality	7,190 (15%)	11,134 (23%)	6,701 (34%)	3,842 (43%)
Acute MI△	198 (3%)	258 (2%)	160 (2%)	221 (6%)

Data are frequencies(%) for categorical values or median (1^{st} - 3^{rd} quartile) for continues values.

■ QTc could not be calculated in patients whose QRS complex or RR interval were suppressed.

△ ECG detected acute MI.

D.4 Classification of Causes of Death and Hospitalisation

Table D.6: Definitions of disease categories used for classifying causes of death and hospitalisation.

<table>
<tr><th>Cause of Death</th><th>Cause of Hospitalisation</th></tr>
<tr><td rowspan="3">CV ICD-10 Chapter IX 'Diseases of the circulatory system' (code range I00-I99)</td><td>Heart Failure:ICD-10 codes: I50 'Heart Failure' (includes I50.0, I50.1, I50.9), I42.0 'Dilated cardiomyopathy', I42.9 'Cardiomyopathy, unspecified', I11.0 'Hypertensive heart disease with (congestive) heart failure', I25.5 'Ischaemic cardiomyopathy', I13.0 'Hypertensive heart and renal disease with (congestive) heart failure', I13.2 'Hypertensive heart and renal disease with both (congestive) heart failure and renal failure.</td></tr>
<tr><td>Emergency Non-HF CV:ICD-10 Chapter IX 'Diseases of the circulatory system' (I00 - I99) admission not classified as HF with an emergency admission (<code>ADMTYPE</code> 30-39) (see Section 2.5.5 for an explanation of classifying admissions</td></tr>
<tr><td>Other Cardiovascular disorders: ICD-10 Chapter IX 'Disorders of the circulatory system' (I00 - I99), excluding codes related to HF, emergency non-HF CV admissions, or infections.</td></tr>
<tr><td colspan="2">Neoplasms: ICD-10 Chapter II 'Neoplasms' (C00 - D48)</td></tr>
<tr><td rowspan="3">Other: Any coding not falling into either of the above categories.</td><td>Gastrointestinal (GI) Disease: ICD-10 Chapter XI 'Diseases of the digestive system' (K00-K93), excluding codes categorised as infections. Individual codes listed in Table B.2</td></tr>
<tr><td>Chronic respiratory disease: individual codes listed in Table B.3</td></tr>
<tr><td>Infections: Individual codes listed in Table B.4</td></tr>
</table>

Continuation of Table D.6

Cause of Death	Cause of Hospitalisation
Other: Any coding not falling into either of the above categories.	**Renal:** ICD-10 subchapters 'Renal failure' (N17-N19), 'Glomerular disease (N00-N08), 'Renal tubule-interstitial disease' (N10-N16), 'Other disorders of kidney and ureter' (N25-N29) where not classified as infections. Individual codes are listed in Table B.5.
	Other: ICD-10 Chapter XVIII 'Symptoms, signs, and abnormal clinical laboratory findings, not elsewhere classified (R00-R99) as well as any code not falling into the above categories.
Not applicable	**Emergency non-CV*:** Any code not falling into the ICD-10 Chapter IX 'Diseases of the circulatory system' (I00 - I99) with an emergency admission (`ADMTYPE` 30-39).
Unknown: No cause of death recorded	**Not applicable**

* Includes all emergency admissions of a given type, regardless if that admission is also classified as an infection or heart failure admission.
CV, cardiovascular

D.5 Other Causes of Hospital Admissions

Table D.7: Within hospital admissions classified as 'other', the top six reasons for hospital admission in the 1st year of group status, where group status is a time-dependent covariate updated based on the diagnosis of HF and initiation of LD.

Group	ICD-10 code	Description	Num. of Admissions
Neither (18-59 yrs)	M06.9	Rheumatoid arthritis, unspecified	399
	F10.3	Mental and behavioural disorders to use of alcohol Withdrawal state	396
	H26.9	Cataract, unspecified	336
	T39.1	Poisoning by nonopioid analgesics, antipyretics, and antirheumatics - 4-aminophenol derivatives	301
	M54.5	Lower back pain	240
	D64.9	Anaemia, unspecified	194

Continuation of Table D.7

Group	ICD-10 code	Description	Num. of Admissions
Neither (≥60 yrs)	H26.9	Cataract, unspecified	3159
	M06.9	Rheumatoid arthritis, unspecified	576
	D64.9	Anaemia, unspecified	554
	M17.1	Other primary gonarthrosis	467
	H35.3	Degeneration of macula and posterior pole	456
	S72.0	Fracture of neck of femur	400
LD Only	H26.9	Cataract, unspecified	634
	D64.9	Anaemia, unspecified	338
	M79.8	Other specified soft tissue disorders	174
	S72.0	Fracture of neck of femur	136
	D50.9	Iron deficiency anaemia	126
	M069	Rheumatoid arthritis, unspecified	116
HF Only	H26.9	Cataract, unspecified	121
	D64.9	Anaemia, unspecified	64
	D50.9	Iron deficiency anaemia, unspecified	26
	S72.0	Fracture of neck of femur	22
	T814	Infection following a procedure, not elsewhere classified	21
	M798	Other specified soft tissue disorders	21
Both: LD First	D64.9	Anaemia, unspecified	67
	H26.9	Cataract, unspecified	61
	S72.0	Fracture of neck of femur	17
	D50.9	Iron deficiency anaemia, unspecified	16
	M79.8	Other specified soft tissue disorders	12
	M05.1	Rheumatoid lung disease	10

Continuation of Table D.7

Group	ICD-10 code	Description	Num. of Admissions
Both: HF First	H26.9	Cataract, unspecified	73
	D64.9	Anaemia, unspecified	73
	S72.0	Fracture of neck of femur	24
	D50.9	Iron deficiency anaemia, unspecified	18
	M06.9	Rheumatoid arthritis, unspecified	15
	T39.1	Poisoning by nonopioid analgesics, antipyretics, and antirheumatics - 4-aminophenol derivatives	14
Both: Together	H26.9	Cataract, unspecified	76
	D64.9	Anaemia, unspecified	42
	D50.9	Iron deficiency anaemia, unspecified	12
	D50.8	Other iron deficiency, anaemia	12
	S72.0	Fracture of neck of femur	10
	M79.8	Other specified soft tissue disorders	9

D.6 Raw Causes of Death

Table D.8: 1-, 2-, and 3-year crude mortality by cause of death per patient-year at risk based on time since entering the cohort as 'neither' split by age, and those who receive a diagnosis of HF or initiate a repeat LD.

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	118,551	107,727	25,206	10,259
Died in the 1st year	718 (<0.01)	3,891 (0.01)	4,072 (0.19)	1,443 (0.16)
CV	134 (<0.01)	1,063 (<0.01)	979 (0.04)	618 (0.07)
Neoplasm	200 (<0.01)	1,416 (<0.01)	1,398 (0.06)	618 (0.07)
Infection	32 (<0.01)	259 (<0.01)	261 (0.01)	135 (0.02)
Other	344 (<0.01)	1,119 (<0.01)	1,415 (0.06)	447 (0.05)
Unknown	8 (<0.01)	34 (<0.01)	19 (<0.01)	<6 (<0.01)
In 2nd year†	113,047 (95%)	96,367 (89%)	18,965 (75%)	7,678 (75%)
Died in the 2nd year	591 (<0.01)	3,144 (0.01)	1,841 (0.11)	676 (0.10)
CV	130 (<0.01)	893 (<0.01)	547 (0.03)	274 (0.04)
Neoplasm	176 (<0.01)	1,140 (<0.01)	390 (0.02)	107 (0.02)
Infection	28 (<0.01)	217 (<0.01)	152 (0.03)	59 (0.01)
Other	250 (<0.01)	873 (<0.01)	745 (0.05)	232 (0.04)

Continuation of Table D.8

Variable	Neither (18-59 yrs)	Neither (≥60 yrs)	Any LD	Any HF
n	118,551	107,727	25,206	10,259
Unknown	7 (<0.01)	21 (<0.01)	7 (<0.01)	<6 (<0.01)
In 3rd year†	99,384 (84%)	85,955 (80%)	13,615 (54%)	5,622 (55%)
Died in the 3rd year	572 (<0.01)	2,913 (0.01)	1,271 (0.11)	373 (0.08)
CV	125 (<0.01)	788 (<0.01)	366 (0.03)	152 (0.03)
Neoplasm	165 (<0.01)	983 (<0.01)	237 (0.02)	68 (0.02)
Infection	21 (<0.01)	163 (<0.01)	120 (0.01)	24 (0.01)
Other	252 (<0.01)	952 (<0.01)	544 (0.05)	126 (0.03)
Unknown	9 (<0.01)	27 (<0.01)	<6 (<0.01)	<6 (<0.01)

Data are rates (patient-year at risk) for mortality values or number (% of those available) for those who entered the state and remained within the said state at a given year after entering.
† Alive and in specified state at the year after entering said state.

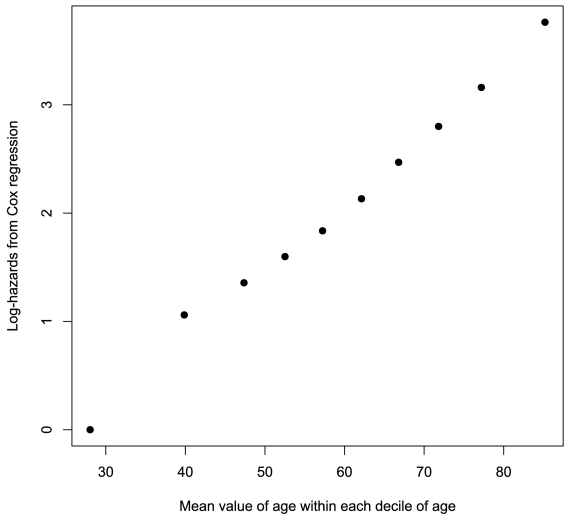
D.7 Example R Code for the Multi-State Models

```
library(mstate)
Create the multi-state transition matrix
tmat <- transMat(
          x = list(c(2, 3, 6, 7), #Neither
                   c(4, 7), #LD Only
                   c(5, 7), #HF Only
                   c(7), #Both: LD First
                   c(7), #Both: HF First
                   c(7), #Both: Together
                   c()), #Death
            names = c("Neither", "LD  Only", "HF Only",
                      "Both: LD First", "Both: HF First",
                      "Both: Together", "Death")
)

# Pick out covariates to keep
covs <- c("age", inclusionAge", "inclusionYear", "sex", "SIMD",
          "neitherTime", "ldOnlyTime", "hfOnlyTime",
          "bothLdFirstTime", "bothHfFirstTime",
```

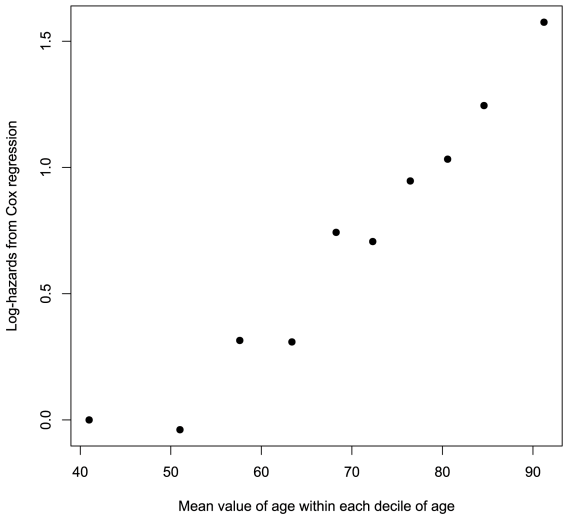
```
          "bothTogetherTime", "timeInNeither")

# Create a long format of data for multi-state analysis
# Flags are 1 if state reached, 0 otherwise
# Time is time to entry if state entered, or time until
# censoring if the state is not entered
 msLdHf <- msprep(
            time = c("neitherTime", "ldOnlyTime",
                     "hfOnlyTime", "bothLdFirstTime",
                     "bothHfFirstTime", "bothTogetherTime",
                     "timeInNeither"),
            status = c("neitherFlag", "ldOnlyFlag",
                       "hfOnlyFlag", "bothLdFirstFlag",
                       "bothHfFirstFlag",
                       "bothTogetherFlag", "died"),
            data = included,
            trans = tmat,
            id = c("SafeHavenID"),
            # All patients start at personal time zero,
            # from their initial state
            start = list(state = as.numeric(as.character(
                                 included$startState)),
                         time = as.numeric(as.character(
                                 included$startTime))),
            keep = covs
)

# Append transition specific variables
 msLdHf <- expand.covs( msLdHf, covs, append = TRUE,
                     longnames = TRUE)

#check Markov assumption
library(frailtyEM)
fit <- coxph(Surv(Tstart, Tstop, status) ~ strata(trans),
             control = coxph.control(timefix = TRUE),
             data = data, x=TRUE, model = TRUE)
testMarkov <- ca_test(fit, id = data$trans)
```

```
#fit clock-reset (cr) cox model
crCox <- coxph(Surv(time, status) ~ strata(trans) +
               transition-specific adjustments for SIMD,
               sex, and age,
               data = msLdHf)
```

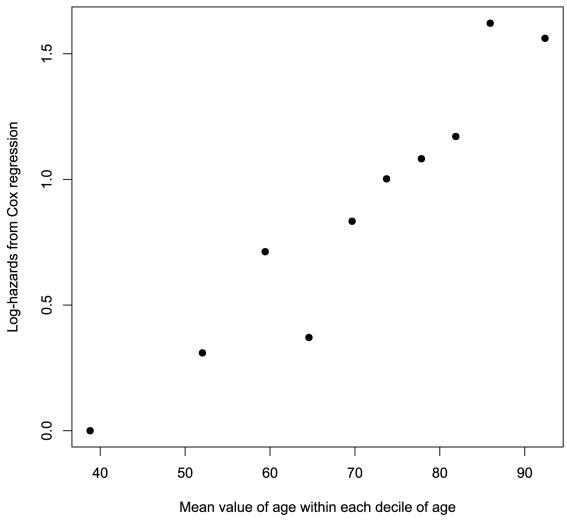
D.8 Testing Linear Age Assumption



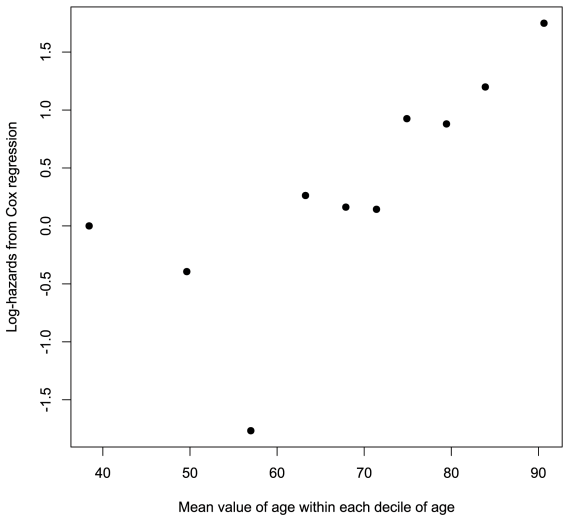
(a) Neither model



(b) LD Only model

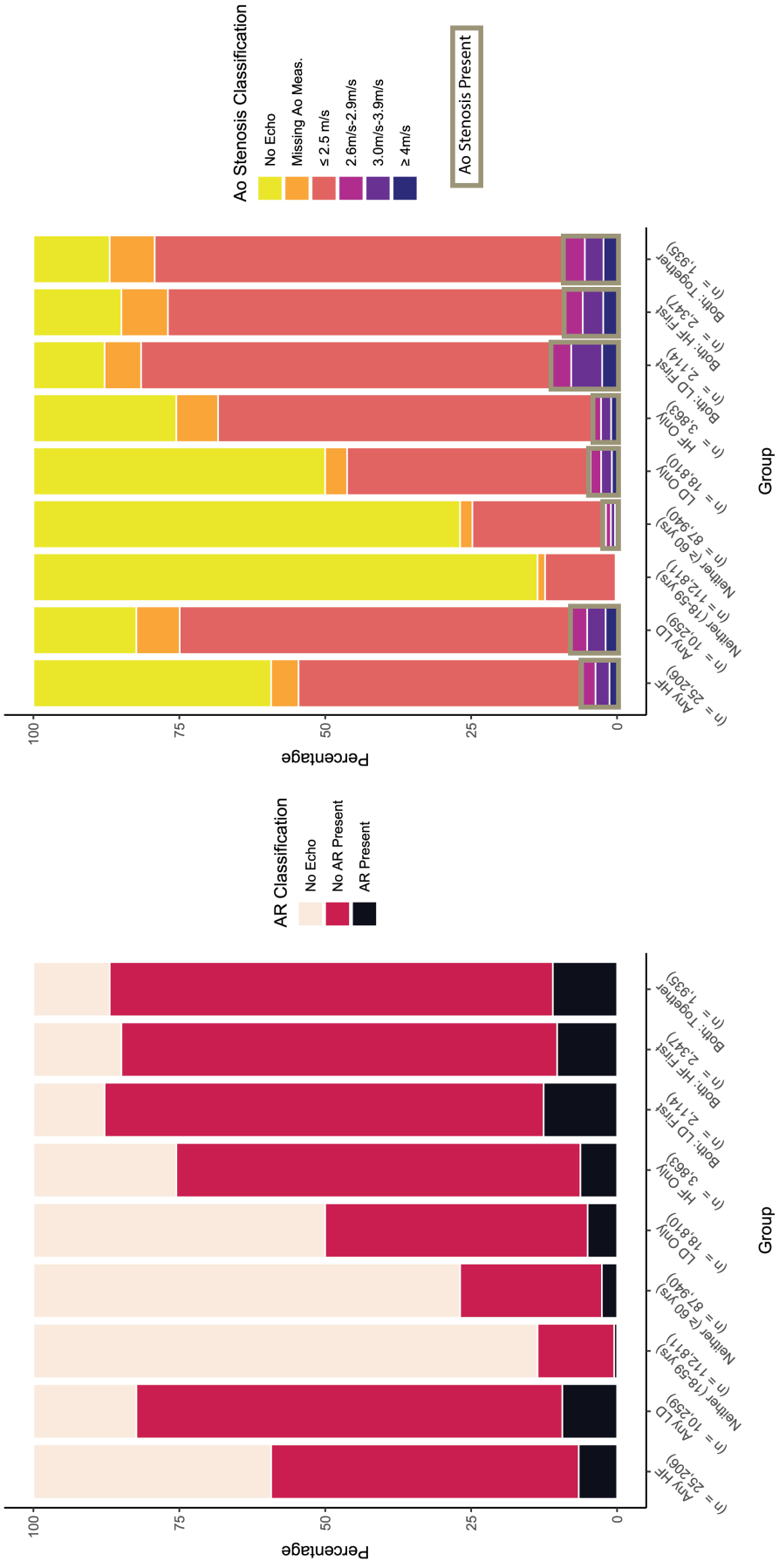


(c) HF Only model



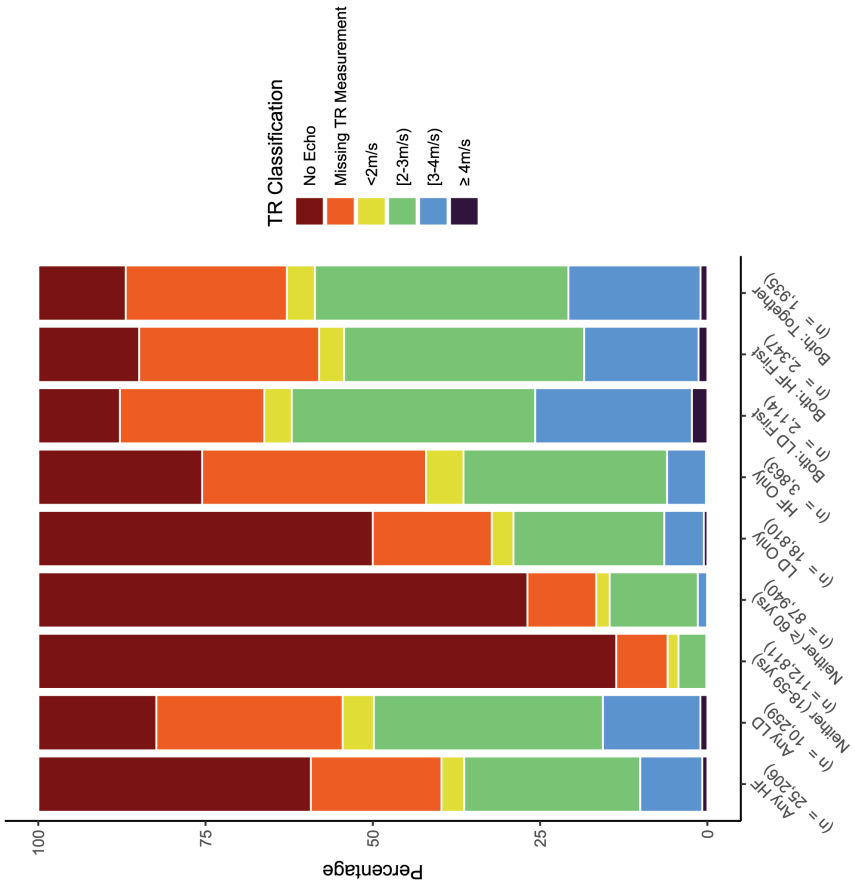
(d) Both: Together model

Figure D.3: Assumption testing of the linear effect of age by plotting log-hazards against mean age within each decile. This assumption testing was replicated for the four multi-state models. Upon entering the cohort, patients were assigned to a model based on their HF and LD status.



(a) Aortic regurgitation (AR) (b) Aortic stenosis (AS)

Figure D.4: Echocardiogram measurements from the closest test to initiating LD regardless of HF status, receiving a diagnosis of HF regardless of LD status, and entry into the state prior to the end of observation or death with patients grouped by the presence and order of a diagnosis of HF and initiation of LD. Grey boxes indicate the presence of AS.



(c) Mitral regurgitation (MR)

(d) Tricuspid regurgitation (TR)

Echocardiogram measurements from the closest test to initiating LD regardless of HF status, receiving a diagnosis of HF regardless of LD status, and entry into the state prior to the end of observation or death with patients grouped by the presence and order of a diagnosis of HF and initiation of LD.

D.9 Cox Proportional Hazards Results for Multi-State Models

Table D.9: Cox PH results from multi-state diagram based on cohort inclusion status.

Variable	**Neither**	**LD Only**	**HF Only**	**Both: Together**
n	226,278	2,371	864	307
Inc age (years)	72 (66 - 79)	70 (57 - 81)	71 (59 - 82)	70 (57 - 80)
	HR (95% CI); p	HR (95% CI); p	HR (95% CI); p	HR (95% CI); p
		Neither to LD Only		
Age (per decade)*	1.8 (1.8 - 1.9); p < 0.001	N/A	N/A	N/A
SIMD				
5 (least deprived)	Reference	N/A	N/A	N/A
4	1.2 (1.2 - 1.3); p < 0.001	N/A	N/A	N/A
3	1.4 (1.3 - 1.5); p < 0.001	N/A	N/A	N/A
2	1.5 (1.4 - 1.6); p < 0.001	N/A	N/A	N/A
1 (most deprived)	1.7 (1.6 - 1.7); p < 0.001	N/A	N/A	N/A
Sex				
Women	Reference	N/A	N/A	N/A
Men	0.86 (0.83 - 0.88) ; p < 0.001	N/A	N/A	N/A
Inc year	1.2 (1.2 - 1.3); p < 0.001	N/A	N/A	N/A
		Neither to HF Only		
Age (per decade)*	1.7 (1.6 - 1.7); p < 0.001	N/A	N/A	N/A
SIMD				
5 (least deprived)	Reference	N/A	N/A	N/A
4	1.1 (0.99 - 1.2); p = 0.057	N/A	N/A	N/A
3	1.3 (1.2 - 1.5); p < 0.001	N/A	N/A	N/A

Continuation of Table D.9

Variable	Neither	LD Only	HF Only	Both: Together
n	226,278	2,371	864	307
2	1.4 (1.2 - 1.5); p < 0.001	N/A	N/A	N/A
1 (most deprived)	1.6 (1.5 - 1.7); p < 0.001	N/A	N/A	N/A
Sex				
Women	Reference	N/A	N/A	N/A
Men	2.0 (1.8 - 2.1) ; p < 0.001	N/A	N/A	N/A
Inc year	1.2 (1.2 - 1.2); p < 0.001	N/A	N/A	N/A
Neither to Both: Together				
Age (per decade)*	2.0 (1.9 - 2.1); p < 0.001	N/A	N/A	N/A
SIMD				
5 (least deprived)	Reference	N/A	N/A	N/A
4	1.1 (0.88 - 1.3); p = 0.550	N/A	N/A	N/A
3	1.1 (0.92 - 1.3); p = 0.306	N/A	N/A	N/A
2	1.2 (1.1 - 1.5); p = 0.009	N/A	N/A	N/A
1 (most deprived)	1.4 (1.2 - 1.6); p < 0.001	N/A	N/A	N/A
Sex				
Women	Reference	N/A	N/A	N/A
Men	1.6 (1.4 - 1.7) ; p < 0.001	N/A	N/A	N/A
Inc year	1.3 (1.2 - 1.3); p < 0.001	N/A	N/A	N/A
Neither to Death				
Age (per decade)*	2.3 (2.3 - 2.4); p < 0.001	N/A	N/A	N/A

Continuation of Table D.9

Variable	**Neither**	**LD Only**	**HF Only**	**Both: Together**
n	226,278	2,371	864	307
SIMD				
5 (least deprived)	Reference	N/A	N/A	N/A
	1.3 (1.2 - 1.3); p = 0.057	N/A	N/A	N/A
3	1.5 (1.4 - 1.6); p < 0.001	N/A	N/A	N/A
2	1.6 (1.5 - 1.7); p < 0.001	N/A	N/A	N/A
1 (most deprived)	2.0 (1.9 - 2.1); p < 0.001	N/A	N/A	N/A
Sex				
Women	Reference	N/A	N/A	N/A
Men	1.4 (1.4 - 1.5) ; p < 0.001	N/A	N/A	N/A
Inc year	1.1 (1.1 - 1.1); p < 0.001	N/A	N/A	N/A
		LD Only to Both: LD First		
Age (per decade)*	1.4 (1.3 - 1.4); p < 0.001	1.5 (1.3 - 1.7); p < 0.001	N/A	N/A
SIMD				
5 (least deprived)	Reference	Reference	N/A	N/A
4	0.95 (0.80 - 1.1); p = 0.544	1.4 (0.63 - 2.9); p = 0.443	N/A	N/A
3	0.94 (0.79 - 1.1); p = 0.459	1.8 (0.88 - 3.6); p = 0.112	N/A	N/A
2	1.3 (1.1 - 1.5); p = 0.002	2.1 (1.1 - 4.1); p = 0.029	N/A	N/A
1 (most deprived)	1.1 (0.96 - 1.3); p = 0.189	2.2 (1.2 - 4.0); p = 0.011	N/A	N/A
Sex				
Women	Reference	Reference	N/A	N/A
Men	1.6 (1.5 - 1.8) ; p < 0.001	1.9 (1.4 - 2.7); p < 0.001	N/A	N/A
Inc year	0.89 (0.85 - 0.94); p < 0.001	0.95 (0.84 - 1.1); p = 0.444	N/A	N/A

Continuation of Table D.9

Variable	Neither	LD Only	HF Only	Both: Together
n	226,278	2,371	864	307
LD Only to Death				
Age (per decade)*	1.6 (1.5 - 1.6); p < 0.001	1.4 (1.3 - 1.5); p < 0.001	N/A	N/A
SIMD				
5 (least deprived)	Reference	Reference	N/A	N/A
4	1.1 (1.0 - 1.3); p = 0.022	1.1 (0.91 - 1.4); p = 0.255	N/A	N/A
3	1.2 (1.1 - 1.4); p < 0.001	1.0 (0.80 - 1.3); p = 0.895	N/A	N/A
2	1.2 (1.1 - 1.3); p = 0.002	1.2 (0.99 - 1.5); p = 0.063	N/A	N/A
1 (most deprived)	1.3 (1.2 - 1.4); p < 0.001	1.1 (0.95 - 1.4); p = 0.160	N/A	N/A
Sex				
Women	Reference	Reference	N/A	N/A
Men	1.5 (1.4 - 1.6) ; p < 0.001	1.5 (1.4 - 1.8); p < 0.001	N/A	N/A
Inc year	1.0 (1.00 - 1.05); p = 0.078	1.0 (0.96 - 1.0); p = 0.893	N/A	N/A
HF Only to Both: HF First				
Age (per decade)*	1.4 (1.4 - 1.5); p < 0.001	N/A	1.2 (1.2 - 1.3); p < 0.001	N/A
SIMD				
5 (least deprived)	Reference	N/A	Reference	N/A
4	1.2 (0.96 - 1.4); p = 0.118	N/A	0.94 (0.64 - 1.4); p = 0.772	N/A
3	1.3 (1.1 - 1.5); p = 0.002	N/A	0.88 (0.61 - 1.3); p = 0.503	N/A
2	1.3 (1.1 - 1.6); p < 0.001	N/A	0.96 (0.69 - 1.4); p = 0.829	N/A
1 (most deprived)	1.4 (1.2 - 1.6); p < 0.001	N/A	1.4 (1.0 - 1.8); p = 0.026	N/A
Sex				
Women	Reference	N/A	Reference	N/A

Continuation of Table D.9

Variable	Neither	LD Only	HF Only	Both: Together
n	226,278	2,371	864	307
Men	0.78 (0.72 - 0.86) ; p < 0.001	N/A	0.88 (0.72 - 1.1); p = 0.209	N/A
Inc year	0.89 (0.85 - 0.93); p < 0.001	N/A	0.97 (0.90 - 1.0); p = 0.307	N/A
HF Only to Death				
Age (per decade)*	2.0 (1.9 - 2.2); p < 0.001	N/A	1.9 (1.6 - 2.1); p < 0.001	N/A
SIMD				
5 (least deprived)	Reference	N/A	Reference	N/A
4	1.2 (0.87 - 1.6); p = 0.295	N/A	0.99 (0.51 - 1.9); p = 0.971	N/A
3	1.5 (1.1 - 2.0); p = 0.004	N/A	1.2 (0.68 - 2.1); p = 0.529	N/A
2	1.5 (1.2 - 2.0); p = 0.002	N/A	1.1 (0.61 - 1.8); p = 0.841	N/A
1 (most deprived)	1.7 (1.3 - 2.1); p < 0.001	N/A	1.4 (0.83 - 2.2); p = 0.225	N/A
Sex				
Women	Reference	N/A	Reference	N/A
Men	0.82 (0.70 - 0.96) ; p = 0.013	N/A	1.1 (0.79 - 1.6); p = 0.550	N/A
Inc year	0.89 (0.81 - 0.97); p = 0.009	N/A	1.0 (0.91 - 1.2); p = 0.583	N/A
Both: LD First to Death				
Age (per decade)*	1.6 (1.5 - 1.7); p < 0.001	1.5 (1.2 - 2.0); p < 0.001	N/A	N/A
SIMD				
5 (least deprived)	Reference	Reference	N/A	N/A
4	1.2 (0.92 - 1.6); p = 0.183	0.52 (0.16 - 1.6); p = 0.266	N/A	N/A
3	1.2 (0.93 - 1.6); p = 0.149	0.67 (0.24 - 1.9); p = 0.459	N/A	N/A
2	1.2 (0.98 - 1.6); p = 0.076	0.73 (0.29 - 1.8); p = 0.500	N/A	N/A
1 (most deprived)	1.4 (1.1 - 1.8); p = 0.001	0.45 (0.18 - 1.1); p = 0.077	N/A	N/A

Continuation of Table D.9

Variable	**Neither**	**LD Only**	**HF Only**	**Both: Together**
n	226,278	2,371	864	307
Sex				
Women	Reference	Reference	N/A	N/A
Men	1.1 (0.98 - 1.3) ; p = 0.098	1.2 (0.70 - 2.1); p = 0.497	N/A	N/A
Inc year	1.0 (0.91 - 1.1); p = 0.970	1.0 (0.82 - 1.2); p = 0.976	N/A	N/A
Both: HF First to Death				
Age (per decade)*	1.6 (1.5 - 1.8); p < 0.001	N/A	1.5 (1.3 - 1.7); p < 0.001	N/A
SIMD				
5 (least deprived)	Reference	N/A	Reference	N/A
4	0.90 (0.67 - 1.2); p = 0.484	N/A	0.76 (0.41 - 1.4); p = 0.385	N/A
3	1.0 (0.78 - 1.4); p = 0.834	N/A	0.82 (0.47 - 1.5); p = 0.502	N/A
2	1.0 (0.79 - 1.3); p = 0.479	N/A	0.56 (0.30 - 1.0); p = 0.058	N/A
1 (most deprived)	1.2 (0.99 - 1.5); p = 0.067	N/A	1.1 (0.76 - 1.7); p = 0.508	N/A
Sex				
Women	Reference	N/A	Reference	N/A
Men	1.1 (0.98 - 1.3) ; p = 0.098	N/A	1.2 (0.86 - 1.6); p = 0.330	N/A
Inc year	0.96 (0.87 - 1.1); p = 0.372	N/A	1.0 (0.93 - 1.2); p = 0.461	N/A
Both: Together to Death				
Age (per decade)*	1.7 (1.5 - 1.8); p < 0.001	N/A	N/A	1.6 (1.3 - 1.9); p < 0.001
SIMD				
5 (least deprived)	Reference	N/A	N/A	Reference
4	1.2 (0.88 - 1.7); p = 0.230	N/A	N/A	0.67 (0.26 - 1.7); p = 0.260
3	1.1 (0.77 - 1.5); p = 0.673	N/A	N/A	1.2 (0.53 - 2.7); p = 0.671

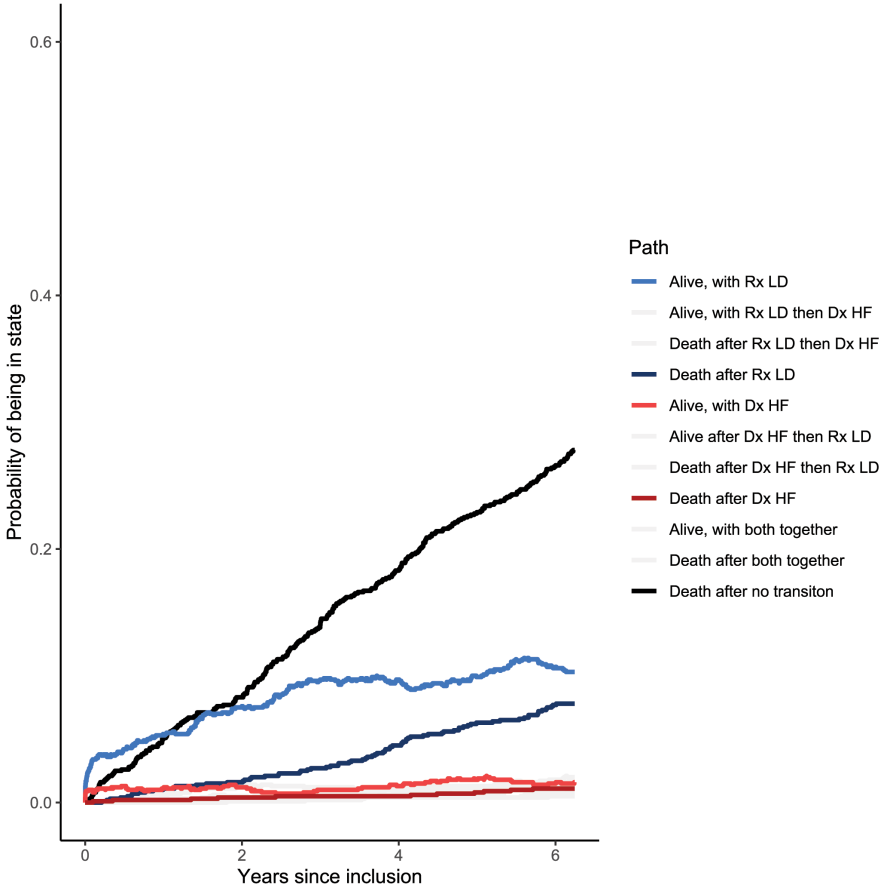
Continuation of Table D.9

Variable	Neither	LD Only	HF Only	Both: Together
n	226,278	2,371	864	307
2	1.1 (0.83 - 1.5); p = 0.479	N/A	N/A	1.9 (1.0 - 3.7); p = 0.050
1 (most deprived)	1.3 (0.99 - 1.7); p = 0.058	N/A	N/A	1.7 (0.94 - 3.2); p = 0.078
Sex				
Women	Reference	N/A	N/A	Reference
Men	1.0 (0.86 - 1.2) ; p = 0.732	N/A	N/A	1.1 (0.71 - 1.6); p = 0.725
Inc year	0.95 (0.86 - 1.0); p = 0.324	N/A	N/A	1.1 (0.98 - 1.3); p = 0.101

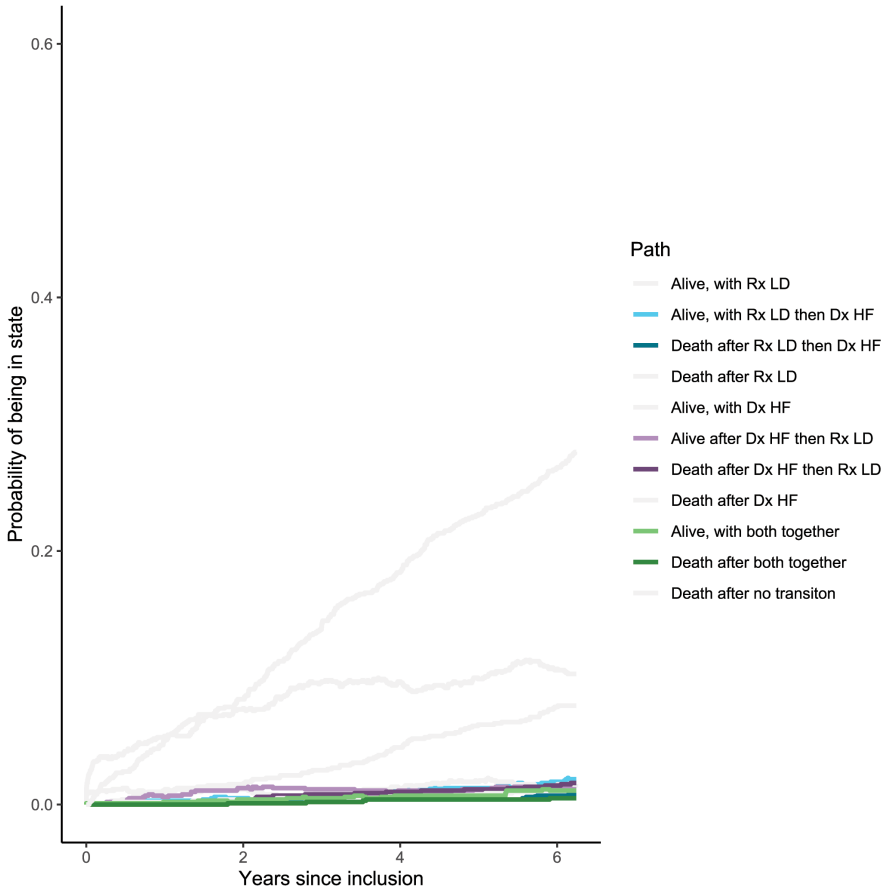
Inc age, age at cohort inclusion.

*Age (per decade), age per decade at cohort inclusion year centred around 60 years (the median age).

D.10 Additional Multi-State Figures



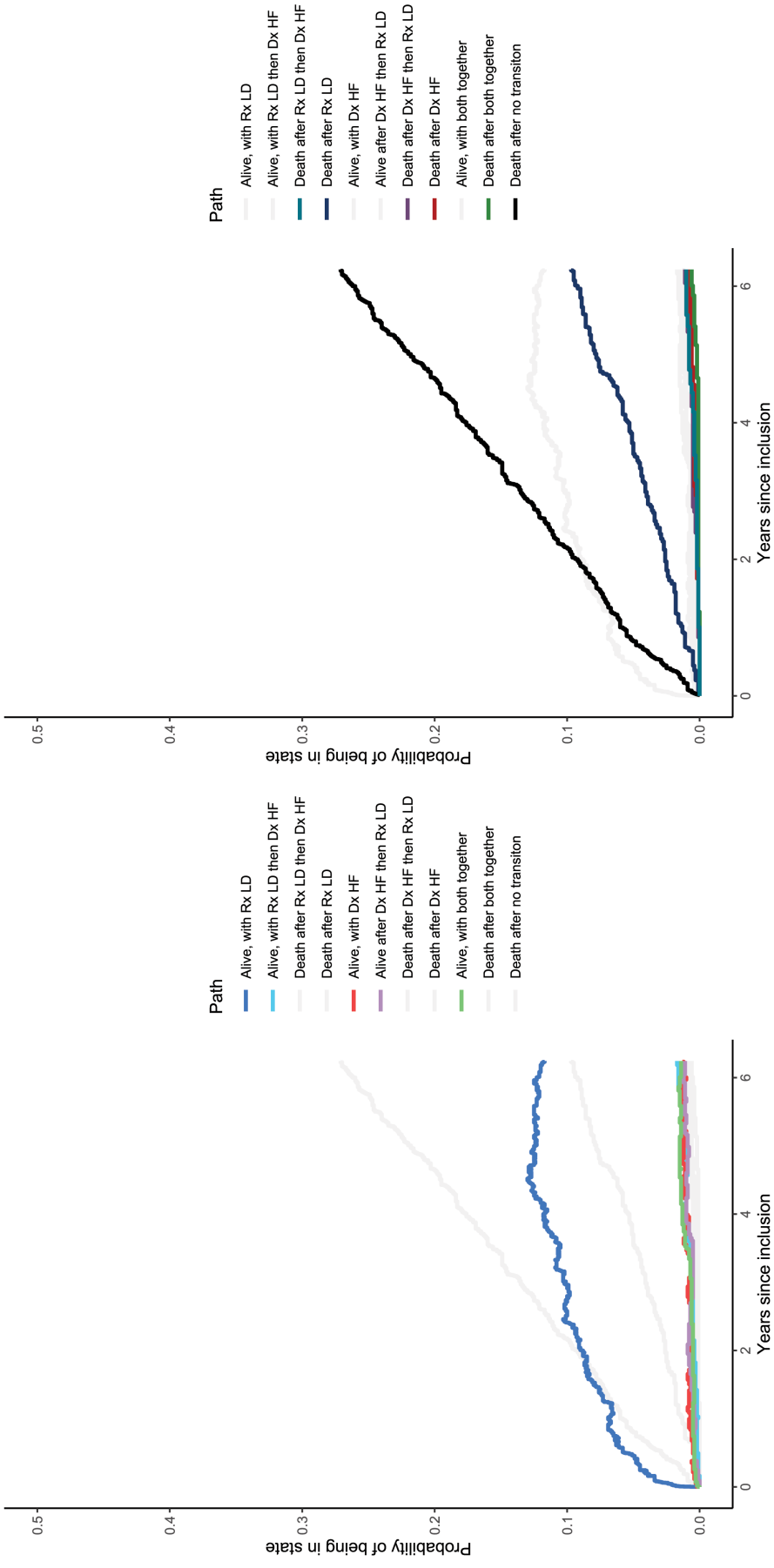
(a) Emphasising paths where patients don't transition or only experience one of HF diagnosis or starting a LD



(b) Emphasising paths where patients are diagnosed with HF and start LD before censoring or death.

.

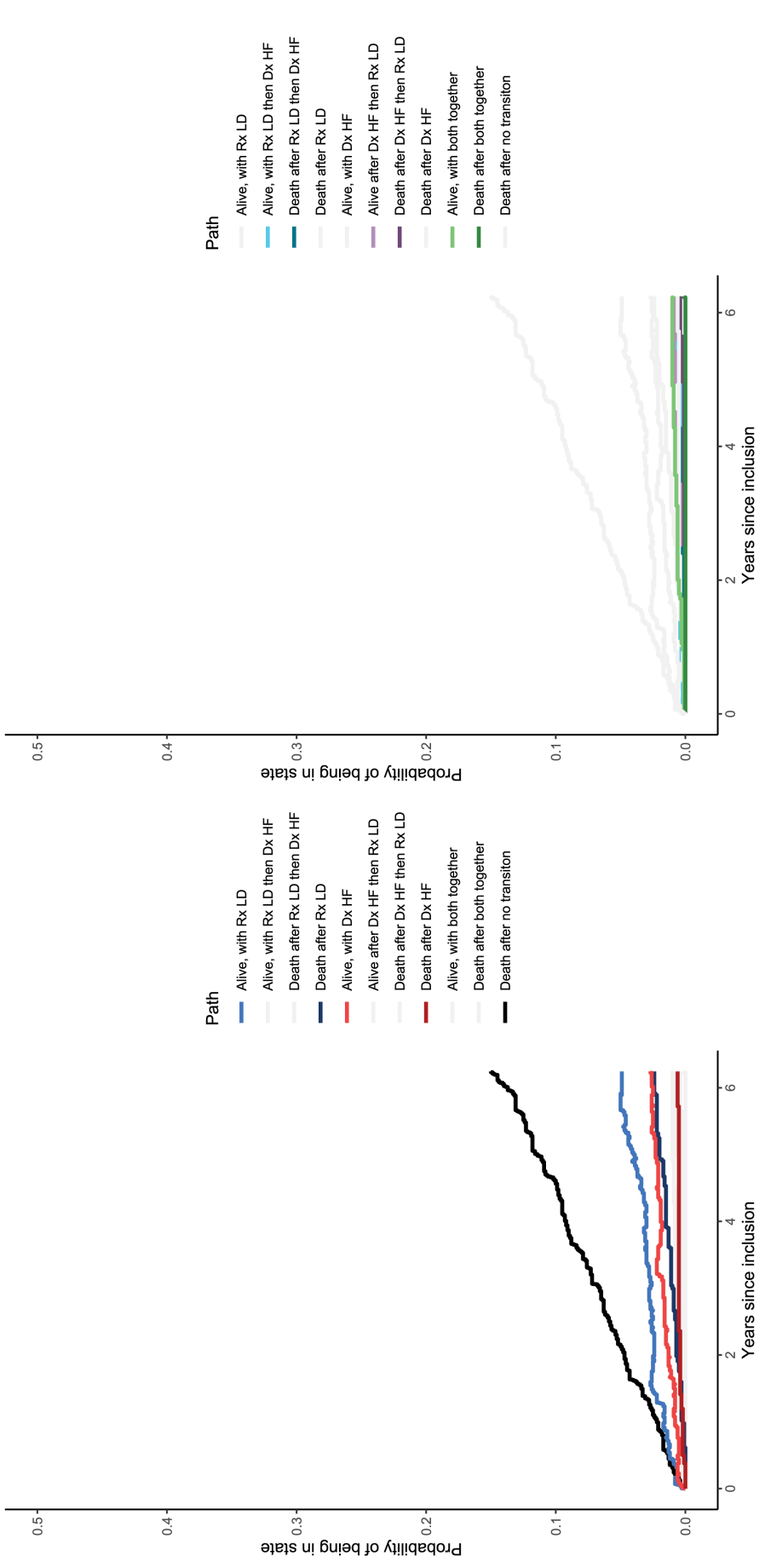
Figure D.5: Predictions, at the time of cohort inclusion, based on entering the cohort in the Neither and having a profile at high risk for entering the LD Only state (woman aged 80 years), against the probability of being in each state over time.



(c) Emphasising paths where patients survive to be censored.

(d) Emphasising paths where patients died during follow-up.

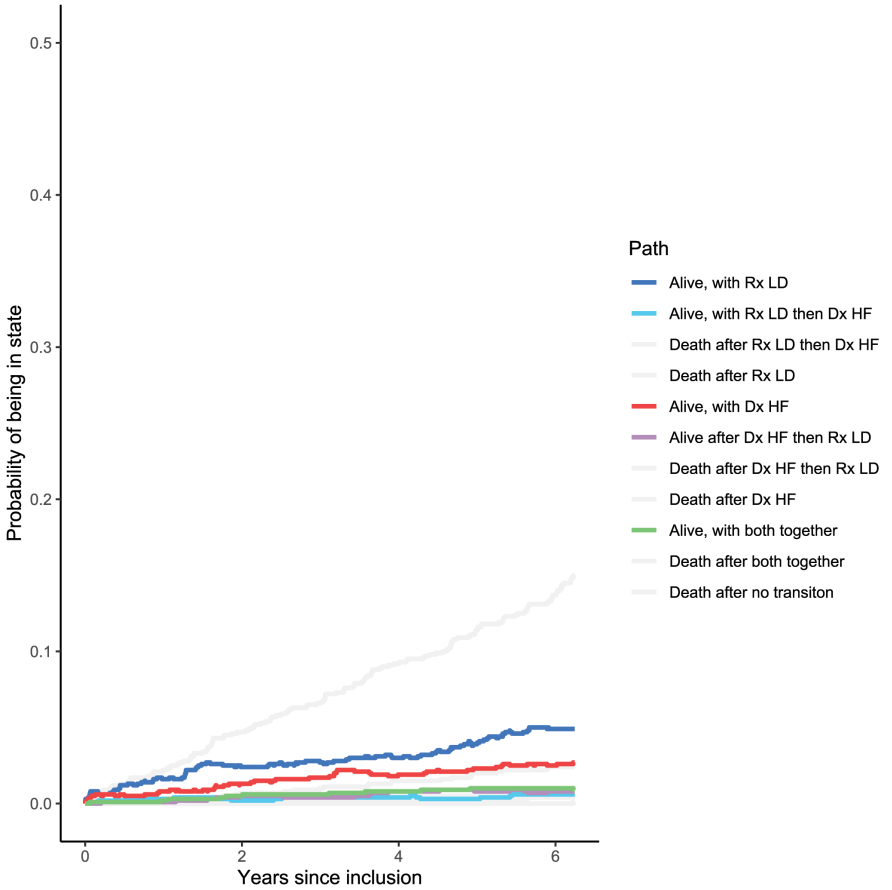
Figure D.5: Predictions, at the time of cohort inclusion based on entering the cohort in the Neither and having a profile at high risk for entering the LD Only state (woman aged 80 years), against the probability of being in each state over time.



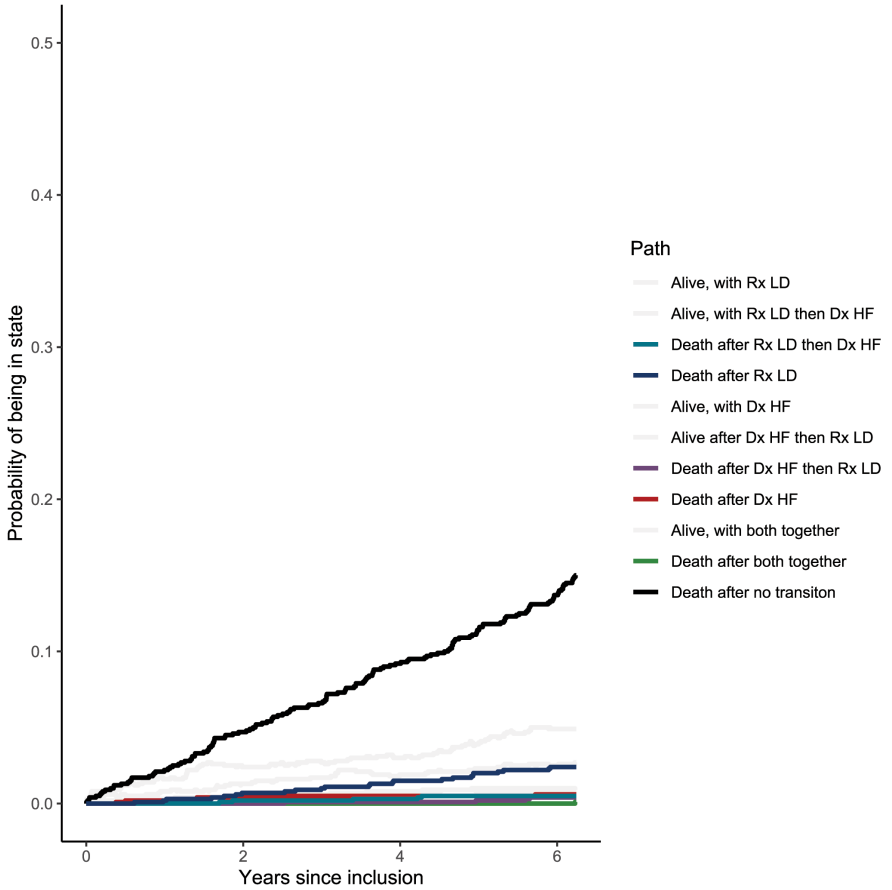
(e) Emphasising paths where patients don't transition or only experience one of HF diagnosis or starting a LD.

(f) Emphasising paths where patients are diagnosed with HF and start LD before censoring or death.

Figure D.5: Predictions, at the time of cohort inclusion based on entering the cohort in the Neither and having a profile at high risk for entering the HF Only state (man aged 65 years), against the probability of being in each state over time.



(g) Emphasising paths where patients survive to be censored.



(h) Emphasising paths where patients died during follow-up.

Figure D.5: Predictions, at the time of cohort inclusion based on entering the cohort in the Neither and having a profile at high risk for entering the HF Only state (man aged 65 years), against the probability of being in each state over time.

Appendix E

Supplementary Material for Chapter 6

E.1 Codes Used to Identify Ischaemic Heart Disease

Table E.1: Clinical ICD-10 codes used to identify patients with IHD. Note: all 3 and 4-digit ICD-10 codes contain the codes nested below them.

Code Type	Code	Description
ICD-10	I20	Angina pectoris
ICD-10	I21	Acute myocardial infarction
ICD-10	I22	Subsequent myocardial infarction
ICD-10	I23	Certain current complications following acute myocardial infarction
ICD-10	I24	Other acute ischaemic heart diseases
ICD-10	I25	Chronic ischaemic heart disease

Table E.2: Clinical Read codes used to identify patients with IHD. Codes are from the list published by Reeves et al. (2014).

Read Code	Safe Haven Search Term	Description
G3...00	G3	Ischaemic heart disease
G30..00	G30	Acute myocardial infarction
G300.00	G300	Acute anterolateral infarction
G301.00	G301	Other specified anterior myocardial infarction
G301000	G301000	Acute anteroapical infarction
G30..11	G30	Attack - heart
G301100	G3011	Acute anteroseptal infarction
G30..12	G30	Coronary thrombosis

Continuation of Table E.2

Read Code	Safe Haven Search Term	Description
G30..13	G30	Cardiac rupture following myocardial infarction (MI)
G30..14	G30	Heart attack
G30..15	G30	MI - acute myocardial infarction
G30..16	G30	Thrombosis - coronary
G30..17	G30	Silent myocardial infarction
G301z00	G301z	Anterior myocardial infarction NOS
G302.00	G302	Acute inferolateral infarction
G303.00	G303	Acute inferoposterior infarction
G304.00	G304	Posterior myocardial infarction NOS
G305.00	G305	Lateral myocardial infarction NOS
G306.00	G306	True posterior myocardial infarction
G307.00	G307	Acute subendocardial infarction
G307000	G3070	Acute non-Q wave infarction
G307100	G3071	Acute non-ST segment elevation myocardial infarction
G308.00	G308	Inferior myocardial infarction NOS
G309.00	G309	Acute Q-wave infarct
G30A.00	G30A	Mural thrombosis
G30B.00	G30B	Acute posterolateral myocardial infarction
G30X.00	G30X	Acute transmural myocardial infarction of unspecif site
G30X000	G30X0	Acute ST segment elevation myocardial infarction
G30y.00	G30y	Other acute myocardial infarction
G30y000	G30y0	Acute atrial infarction
G30y100	G30y1	Acute papillary muscle infarction
G30y200	G30y2	Acute septal infarction
G30yz00	G30yz	Other acute myocardial infarction NOS
G30z.00	G30z	Acute myocardial infarction NOS
G31..00	G31	Other acute and subacute ischaemic heart disease
G310.00	G310	Postmyocardial infarction syndrome
G310.11	G310	Dressler's syndrome
G3...11	G3	Arteriosclerotic heart disease
G311.00	G311	Preinfarction syndrome
G311000	G3110	Myocardial infarction aborted
G311011	G311011	MI - myocardial infarction aborted
G311100	G3111	Unstable angina
G311.11	G311	Crescendo angina

Continuation of Table E.2

Read Code	Safe Haven Search Term	Description
G311.12	G311	Impending infarction
G311.13	G311	Unstable angina
G311.14	G311	Angina at rest
G311200	G3112	Angina at rest
G311300	G3113	Refractory angina
G311400	G3114	Worsening angina
G311500	G3115	Acute coronary syndrome
G311z00	G311z	Preinfarction syndrome NOS
G3...12	G3	Atherosclerotic heart disease
G312.00	G312	Coronary thrombosis not resulting in myocardial infarction
G3...13	G3	IHD - Ischaemic heart disease
G31y.00	G31y	Other acute and subacute ischaemic heart disease
G31y000	G31y0	Acute coronary insufficiency
G31y100	G31y1	Microinfarction of heart
G31y200	G31y2	Subendocardial ischaemia
G31y300	G31y3	Transient myocardial ischaemia
G31yz00	G31yz	Other acute and subacute ischaemic heart disease NOS
G32..00	G32	Old myocardial infarction
G32..11	G32	Healed myocardial infarction
G32..12	G32	Personal history of myocardial infarction
G33..00	G33	Angina pectoris
G330.00	G330	Angina decubitus
G330000	G3300	Nocturnal angina
G330z00	G330z	Angina decubitus NOS
G331.00	G331	Prinzmetal's angina
G331.11	G331	Variant angina pectoris
G332.00	G332	Coronary artery spasm
G33z.00	G33z	Angina pectoris NOS
G33z000	G33z0	Status anginosus
G33z100	G33z1	Stenocardia
G33z200	G33z2	Syncope anginosa
G33z300	G33z3	Angina on effort
G33z400	G33z4	Ischaemic chest pain
G33z500	G33z5	Post infarct angina
G33z600	G33z6	New onset angina

Continuation of Table E.2

Read Code	Safe Haven Search Term	Description
G33z700	G33z7	Stable angina
G33zz00	G33zz	Angina pectoris NOS
G34..00	G34	Other chronic ischaemic heart disease
G340.00	G340	Coronary atherosclerosis
G340000	G3400	Single coronary vessel disease
G340100	G3401	Double coronary vessel disease
G340.11	G340	Triple vessel disease of the heart
G340.12	G340	Coronary artery disease
G342.00	G342	Atherosclerotic cardiovascular disease
G343.00	G343	Ischaemic cardiomyopathy
G344.00	G344	Silent myocardial ischaemia
G34y.00	G34y	Other specified chronic ischaemic heart disease
G34y000	G34y0	Chronic coronary insufficiency
G34y100	G34y1	Chronic myocardial ischaemia
G34yz00	G34yz	Other specified chronic ischaemic heart disease NOS
G34z.00	G34z	Other chronic ischaemic heart disease NOS
G34z000	G34z0	Asymptomatic coronary heart disease
G35..00	G35	Subsequent myocardial infarction
G350.00	G350	Subsequent myocardial infarction of anterior wall
G351.00	G351	Subsequent myocardial infarction of inferior wall
G353.00	G353	Subsequent myocardial infarction of other sites
G35X.00	G35X	Subsequent myocardial infarction of unspecified site
G36..00	G36	Certain current complication follow acute myocardial infarct
G360.00	G360	Haemopericardium/current comp folow acut myocard infarct
G361.00	G361	Atrial septal defect/curr comp folow acut myocardal infarct
G362.00	G362	Ventric septal defect/curr comp fol acut myocardal infarctn
G363.00	G363	Ruptur cardiac wall w'out haemopericard/cur comp fol ac MI
G364.00	G364	Ruptur chordae tendinae/curr comp fol acute myocard infarct

Continuation of Table E.2

Read Code	Safe Haven Search Term	Description
G365.00	G365	Rupture papillary muscle/curr comp fol acute myocard infarct
G366.00	G366	Thrombosis atrium; auric append&vent/curr comp foll acute MI
G38..00	G38	Postoperative myocardial infarction
G380.00	G380	Postoperative transmural myocardial infarction anterior wall
G381.00	G381	Postoperative transmural myocardial infarction inferior wall
G384.00	G384	Postoperative subendocardial myocardial infarction
G38z.00	G38z	Postoperative myocardial infarction unspecified
G3y..00	G3y	Other specified ischaemic heart disease
G3z..00	G3z	Ischaemic heart disease NOS

E.2 Inclusion Reasons

Table E.3: Breakdown of the number of included patients by group, code type, IHD code, and code description. Only the first code between 31st December 2009 through 1st January 2013 is included below.

Group	Code Type	Code	Count
No LD/No HF	ICD-10	I200	293
No LD/No HF	ICD-10	I2000	37
No LD/No HF	ICD-10	I2001	45
No LD/No HF	ICD-10	I2002	22
No LD/No HF	ICD-10	I2009	70
No LD/No HF	ICD-10	I201	11
No LD/No HF	ICD-10	I208	20
No LD/No HF	ICD-10	I209	1,300
No LD/No HF	ICD-10	I210	33
No LD/No HF	ICD-10	I2100	56
No LD/No HF	ICD-10	I2101	22
No LD/No HF	ICD-10	I2109	9
No LD/No HF	ICD-10	I211	56
No LD/No HF	ICD-10	I2110	55
No LD/No HF	ICD-10	I2111	45
No LD/No HF	ICD-10	I2119	13
No LD/No HF	ICD-10	I212	10

Continuation of Table E.3

Group	Code Type	Code	Count
No LD/No HF	ICD-10	I2120	7
No LD/No HF	ICD-10	I2121	<6
No LD/No HF	ICD-10	I2129	<6
No LD/No HF	ICD-10	I213	15
No LD/No HF	ICD-10	I2130	103
No LD/No HF	ICD-10	I2131	<6
No LD/No HF	ICD-10	I2139	14
No LD/No HF	ICD-10	I214	119
No LD/No HF	ICD-10	I2140	<6
No LD/No HF	ICD-10	I2141	<6
No LD/No HF	ICD-10	I2149	<6
No LD/No HF	ICD-10	I219	144
No LD/No HF	ICD-10	I2190	414
No LD/No HF	ICD-10	I2191	32
No LD/No HF	ICD-10	I2199	91
No LD/No HF	ICD-10	I220	<6
No LD/No HF	ICD-10	I2200	<6
No LD/No HF	ICD-10	I221	<6
No LD/No HF	ICD-10	I2211	<6
No LD/No HF	ICD-10	I228	13
No LD/No HF	ICD-10	I2280	<6
No LD/No HF	ICD-10	I2289	<6
No LD/No HF	ICD-10	I229	6
No LD/No HF	ICD-10	I2290	12
No LD/No HF	ICD-10	I240	16
No LD/No HF	ICD-10	I248	135
No LD/No HF	ICD-10	I249	8
No LD/No HF	ICD-10	I250	<6
No LD/No HF	ICD-10	I251	2,551
No LD/No HF	ICD-10	I252	1,121
No LD/No HF	ICD-10	I253	15
No LD/No HF	ICD-10	I254	35
No LD/No HF	ICD-10	I256	<6
No LD/No HF	ICD-10	I258	126
No LD/No HF	ICD-10	I259	2,359
No LD/No HF	Read	G3	467

Continuation of Table E.3

Group	Code Type	Code	Count
No LD/No HF	Read	G3	1,095
No LD/No HF	Read	G3	<6
No LD/No HF	Read	G3	<6
No LD/No HF	Read	G30	297
No LD/No HF	Read	G30	132
No LD/No HF	Read	G30	<6
No LD/No HF	Read	G30	<6
No LD/No HF	Read	G300	<6
No LD/No HF	Read	G301	<6
No LD/No HF	Read	G3011	<6
No LD/No HF	Read	G301z	33
No LD/No HF	Read	G302	<6
No LD/No HF	Read	G303	<6
No LD/No HF	Read	G304	<6
No LD/No HF	Read	G305	<6
No LD/No HF	Read	G307	<6
No LD/No HF	Read	G3070	<6
No LD/No HF	Read	G3071	115
No LD/No HF	Read	G308	44
No LD/No HF	Read	G30A	<6
No LD/No HF	Read	G30X0	36
No LD/No HF	Read	G30y0	<6
No LD/No HF	Read	G30yz	<6
No LD/No HF	Read	G30z	<6
No LD/No HF	Read	G30z	183
No LD/No HF	Read	G310	<6
No LD/No HF	Read	G311	<6
No LD/No HF	Read	G311	6
No LD/No HF	Read	G311	44
No LD/No HF	Read	G3110	<6
No LD/No HF	Read	G3111	14
No LD/No HF	Read	G3112	<6
No LD/No HF	Read	G3114	6
No LD/No HF	Read	G3115	112
No LD/No HF	Read	G3115	<6
No LD/No HF	Read	G31y0	<6

Continuation of Table E.3

Group	Code Type	Code	Count
No LD/No HF	Read	G31y3	<6
No LD/No HF	Read	G32	13
No LD/No HF	Read	G33	<6
No LD/No HF	Read	G33	1,793
No LD/No HF	Read	G33	<6
No LD/No HF	Read	G33z	105
No LD/No HF	Read	G33z	<6
No LD/No HF	Read	G33z	<6
No LD/No HF	Read	G33z3	14
No LD/No HF	Read	G33z4	<6
No LD/No HF	Read	G33z5	<6
No LD/No HF	Read	G33z6	31
No LD/No HF	Read	G33z7	22
No LD/No HF	Read	G33zz	163
No LD/No HF	Read	G340	95
No LD/No HF	Read	G340	32
No LD/No HF	Read	G340	18
No LD/No HF	Read	G3400	13
No LD/No HF	Read	G3401	8
No LD/No HF	Read	G344	<6
No LD/No HF	Read	G34y1	<6
No LD/No HF	Read	G34z0	<6
No LD/No HF	Read	G3y	<6
No LD/No HF	Read	G3z	780
LD Only	ICD-10	I200	92
LD Only	ICD-10	I2000	10
LD Only	ICD-10	I2001	15
LD Only	ICD-10	I2002	7
LD Only	ICD-10	I2009	19
LD Only	ICD-10	I201	<6
LD Only	ICD-10	I208	8
LD Only	ICD-10	I209	457
LD Only	ICD-10	I210	6
LD Only	ICD-10	I2100	<6
LD Only	ICD-10	I2101	<6
LD Only	ICD-10	I2109	<6

Continuation of Table E.3

Group	Code Type	Code	Count
LD Only	ICD-10	I211	8
LD Only	ICD-10	I2110	7
LD Only	ICD-10	I2111	7
LD Only	ICD-10	I2120	<6
LD Only	ICD-10	I2121	<6
LD Only	ICD-10	I213	<6
LD Only	ICD-10	I2130	20
LD Only	ICD-10	I2139	6
LD Only	ICD-10	I214	21
LD Only	ICD-10	I2149	<6
LD Only	ICD-10	I219	38
LD Only	ICD-10	I2190	100
LD Only	ICD-10	I2191	9
LD Only	ICD-10	I2199	28
LD Only	ICD-10	I220	<6
LD Only	ICD-10	I221	<6
LD Only	ICD-10	I228	<6
LD Only	ICD-10	I2280	<6
LD Only	ICD-10	I229	<6
LD Only	ICD-10	I2290	8
LD Only	ICD-10	I232	<6
LD Only	ICD-10	I240	<6
LD Only	ICD-10	I248	27
LD Only	ICD-10	I249	<6
LD Only	ICD-10	I250	<6
LD Only	ICD-10	I251	602
LD Only	ICD-10	I252	349
LD Only	ICD-10	I253	<6
LD Only	ICD-10	I254	<6
LD Only	ICD-10	I258	51
LD Only	ICD-10	I259	1,070
LD Only	Read	G3	58
LD Only	Read	G3	171
LD Only	Read	G30	29
LD Only	Read	G30	12
LD Only	Read	G301z	<6

Continuation of Table E.3

Group	Code Type	Code	Count
LD Only	Read	G3071	15
LD Only	Read	G308	7
LD Only	Read	G30X0	<6
LD Only	Read	G30z	15
LD Only	Read	G311	<6
LD Only	Read	G311	<6
LD Only	Read	G311	15
LD Only	Read	G3111	<6
LD Only	Read	G3112	<6
LD Only	Read	G3115	15
LD Only	Read	G3115	<6
LD Only	Read	G32	<6
LD Only	Read	G33	266
LD Only	Read	G33z	12
LD Only	Read	G33z3	<6
LD Only	Read	G33z4	<6
LD Only	Read	G33zz	20
LD Only	Read	G34	<6
LD Only	Read	G340	<6
LD Only	Read	G340	<6
LD Only	Read	G340	<6
LD Only	Read	G3400	<6
LD Only	Read	G3401	<6
LD Only	Read	G34z	<6
LD Only	Read	G3z	130
HF Only	ICD-10	I200	28
HF Only	ICD-10	I2000	<6
HF Only	ICD-10	I2001	8
HF Only	ICD-10	I2002	<6
HF Only	ICD-10	I2009	<6
HF Only	ICD-10	I208	<6
HF Only	ICD-10	I209	81
HF Only	ICD-10	I210	21
HF Only	ICD-10	I2100	22
HF Only	ICD-10	I2101	17
HF Only	ICD-10	I2109	<6

Continuation of Table E.3

Group	Code Type	Code	Count
HF Only	ICD-10	I211	11
HF Only	ICD-10	I2110	23
HF Only	ICD-10	I2111	13
HF Only	ICD-10	I2119	<6
HF Only	ICD-10	I212	<6
HF Only	ICD-10	I2120	<6
HF Only	ICD-10	I2121	<6
HF Only	ICD-10	I2129	<6
HF Only	ICD-10	I213	<6
HF Only	ICD-10	I2130	20
HF Only	ICD-10	I2131	<6
HF Only	ICD-10	I2139	<6
HF Only	ICD-10	I214	14
HF Only	ICD-10	I2149	6
HF Only	ICD-10	I219	31
HF Only	ICD-10	I2190	89
HF Only	ICD-10	I2191	11
HF Only	ICD-10	I2199	31
HF Only	ICD-10	I220	<6
HF Only	ICD-10	I2200	<6
HF Only	ICD-10	I2210	<6
HF Only	ICD-10	I228	7
HF Only	ICD-10	I2280	<6
HF Only	ICD-10	I2281	<6
HF Only	ICD-10	I229	<6
HF Only	ICD-10	I2290	7
HF Only	ICD-10	I2299	<6
HF Only	ICD-10	I240	<6
HF Only	ICD-10	I248	6
HF Only	ICD-10	I249	<6
HF Only	ICD-10	I251	576
HF Only	ICD-10	I252	339
HF Only	ICD-10	I253	<6
HF Only	ICD-10	I254	12
HF Only	ICD-10	I255	<6
HF Only	ICD-10	I258	37

Continuation of Table E.3

Group	Code Type	Code	Count
HF Only	ICD-10	I259	327
HF Only	Read	G3	31
HF Only	Read	G3	134
HF Only	Read	G30	69
HF Only	Read	G30	34
HF Only	Read	G30	<6
HF Only	Read	G30	<6
HF Only	Read	G300	<6
HF Only	Read	G301	<6
HF Only	Read	G301z	11
HF Only	Read	G304	<6
HF Only	Read	G305	<6
HF Only	Read	G3071	13
HF Only	Read	G308	<6
HF Only	Read	G30X0	8
HF Only	Read	G30yz	<6
HF Only	Read	G30z	39
HF Only	Read	G311	<6
HF Only	Read	G311	<6
HF Only	Read	G311	<6
HF Only	Read	G3115	11
HF Only	Read	G31y0	<6
HF Only	Read	G32	<6
HF Only	Read	G33	87
HF Only	Read	G33	<6
HF Only	Read	G33z	<6
HF Only	Read	G33z3	<6
HF Only	Read	G33z6	<6
HF Only	Read	G33z7	<6
HF Only	Read	G33zz	<6
HF Only	Read	G340	10
HF Only	Read	G340	<6
HF Only	Read	G340	<6
HF Only	Read	G3400	<6
HF Only	Read	G3401	<6
HF Only	Read	G343	<6

Continuation of Table E.3

Group	Code Type	Code	Count
HF Only	Read	G34z0	<6
HF Only	Read	G3z	91
Both: LD + HF	ICD-10	I200	71
Both: LD + HF	ICD-10	I2000	8
Both: LD + HF	ICD-10	I2001	13
Both: LD + HF	ICD-10	I2002	<6
Both: LD + HF	ICD-10	I2009	13
Both: LD + HF	ICD-10	I201	<6
Both: LD + HF	ICD-10	I208	<6
Both: LD + HF	ICD-10	I209	206
Both: LD + HF	ICD-10	I210	17
Both: LD + HF	ICD-10	I2100	22
Both: LD + HF	ICD-10	I2101	15
Both: LD + HF	ICD-10	I2109	<6
Both: LD + HF	ICD-10	I211	<6
Both: LD + HF	ICD-10	I2110	<6
Both: LD + HF	ICD-10	I2111	8
Both: LD + HF	ICD-10	I2119	<6
Both: LD + HF	ICD-10	I212	<6
Both: LD + HF	ICD-10	I2120	<6
Both: LD + HF	ICD-10	I2129	<6
Both: LD + HF	ICD-10	I213	6
Both: LD + HF	ICD-10	I2130	34
Both: LD + HF	ICD-10	I2131	<6
Both: LD + HF	ICD-10	I2139	6
Both: LD + HF	ICD-10	I214	48
Both: LD + HF	ICD-10	I2140	<6
Both: LD + HF	ICD-10	I2149	<6
Both: LD + HF	ICD-10	I219	47
Both: LD + HF	ICD-10	I2190	113
Both: LD + HF	ICD-10	I2191	15
Both: LD + HF	ICD-10	I2199	32
Both: LD + HF	ICD-10	I2209	<6
Both: LD + HF	ICD-10	I221	<6
Both: LD + HF	ICD-10	I2210	<6
Both: LD + HF	ICD-10	I228	10

Continuation of Table E.3

Group	Code Type	Code	Count
Both: LD + HF	ICD-10	I2280	<6
Both: LD + HF	ICD-10	I229	<6
Both: LD + HF	ICD-10	I2290	9
Both: LD + HF	ICD-10	I2299	<6
Both: LD + HF	ICD-10	I240	12
Both: LD + HF	ICD-10	I248	29
Both: LD + HF	ICD-10	I251	745
Both: LD + HF	ICD-10	I252	542
Both: LD + HF	ICD-10	I253	16
Both: LD + HF	ICD-10	I254	9
Both: LD + HF	ICD-10	I255	11
Both: LD + HF	ICD-10	I256	<6
Both: LD + HF	ICD-10	I258	69
Both: LD + HF	ICD-10	I259	888
Both: LD + HF	Read	G3	33
Both: LD + HF	Read	G3	105
Both: LD + HF	Read	G3	<6
Both: LD + HF	Read	G30	37
Both: LD + HF	Read	G30	<6
Both: LD + HF	Read	G30	25
Both: LD + HF	Read	G301	<6
Both: LD + HF	Read	G301z	<6
Both: LD + HF	Read	G3071	17
Both: LD + HF	Read	G308	<6
Both: LD + HF	Read	G30A	<6
Both: LD + HF	Read	G30X	<6
Both: LD + HF	Read	G30X0	<6
Both: LD + HF	Read	G30y2	<6
Both: LD + HF	Read	G30z	21
Both: LD + HF	Read	G311	<6
Both: LD + HF	Read	G311	7
Both: LD + HF	Read	G3111	<6
Both: LD + HF	Read	G3114	<6
Both: LD + HF	Read	G3115	11
Both: LD + HF	Read	G32	<6
Both: LD + HF	Read	G33	92

Continuation of Table E.3

Group	Code Type	Code	Count
Both: LD + HF	Read	G33z	<6
Both: LD + HF	Read	G33z6	<6
Both: LD + HF	Read	G33z7	<6
Both: LD + HF	Read	G33zz	<6
Both: LD + HF	Read	G340	8
Both: LD + HF	Read	G340	<6
Both: LD + HF	Read	G340	<6
Both: LD + HF	Read	G3401	<6
Both: LD + HF	Read	G343	<6
Both: LD + HF	Read	G3z	76

E.3 Patient Flow Diagram

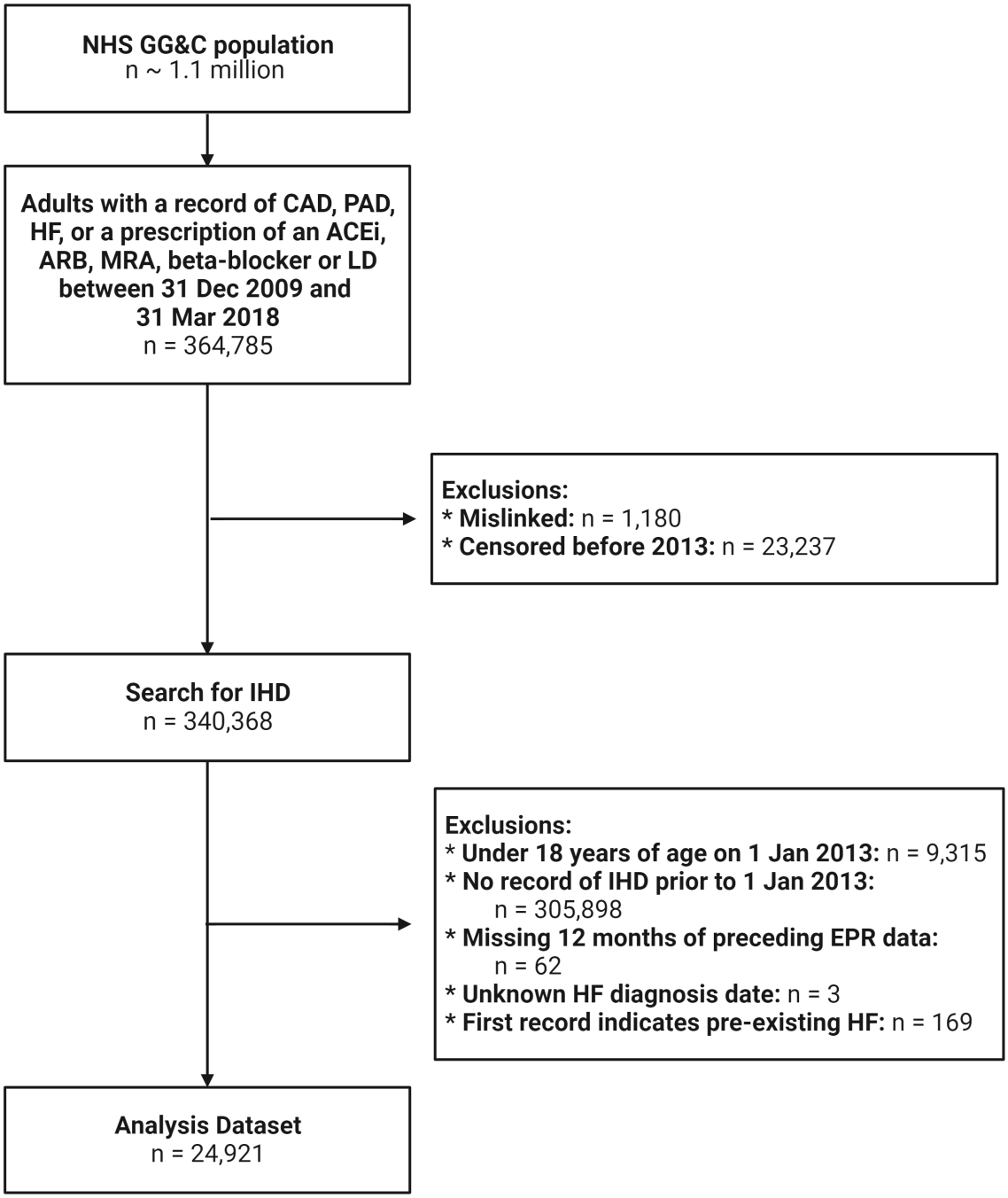


Figure E.1: Flow diagram of inclusion and exclusion criteria for NHS GG&C patients.

E.4 Sensitivity of Extreme Haemoglobin and Serum Creatinine Results

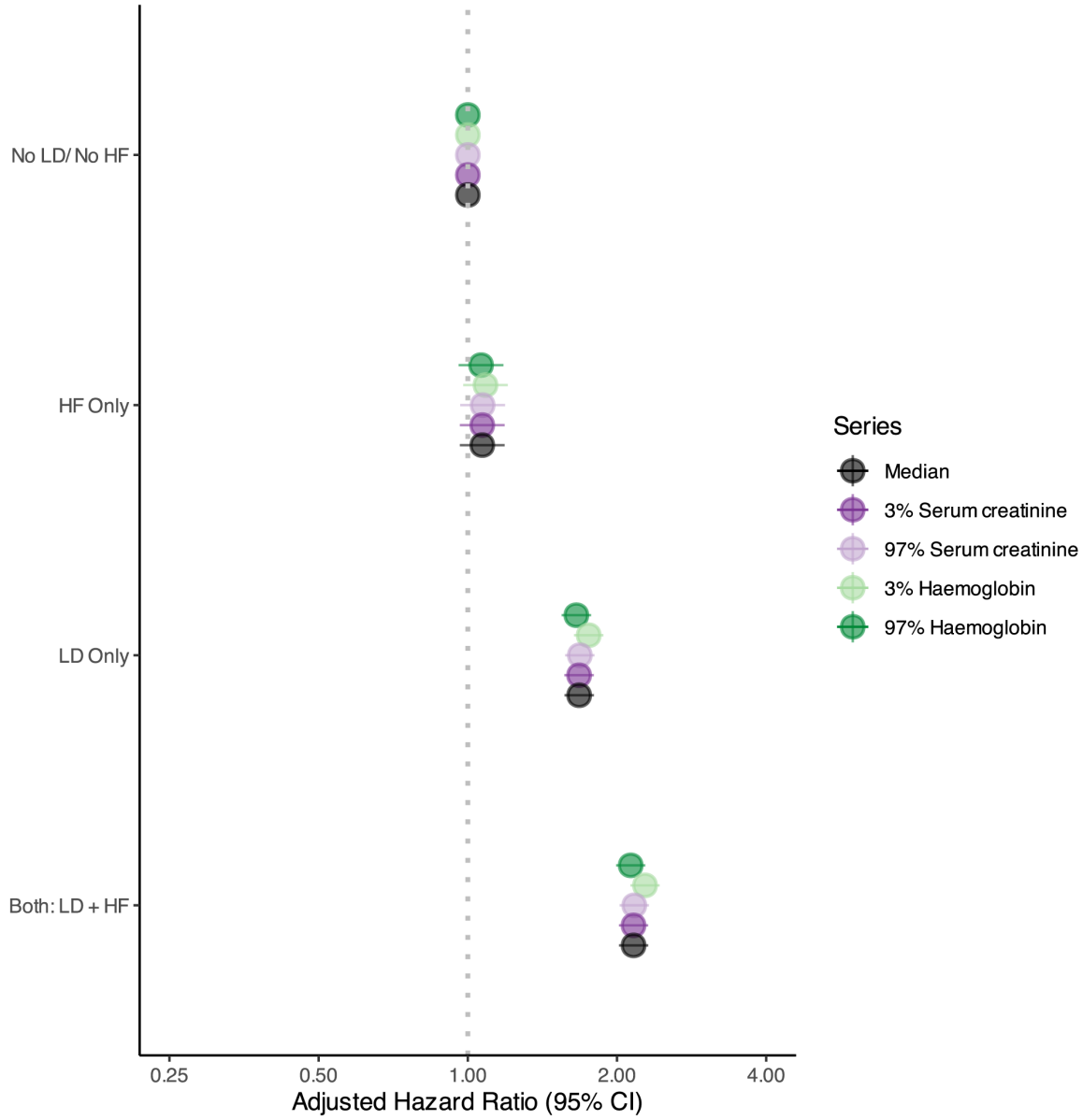
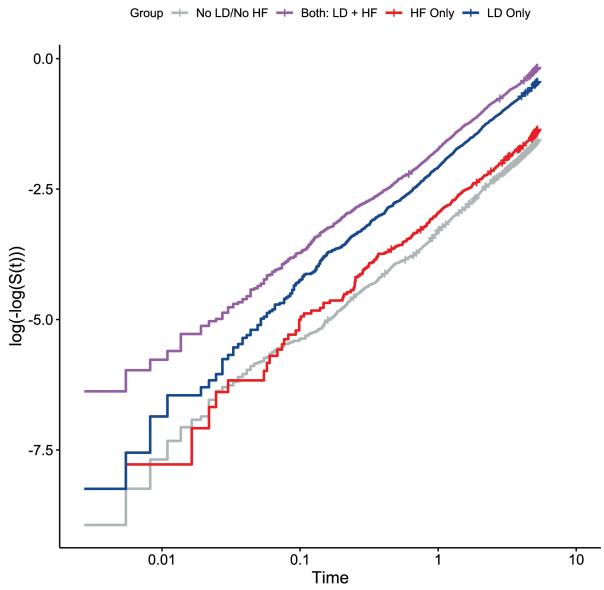
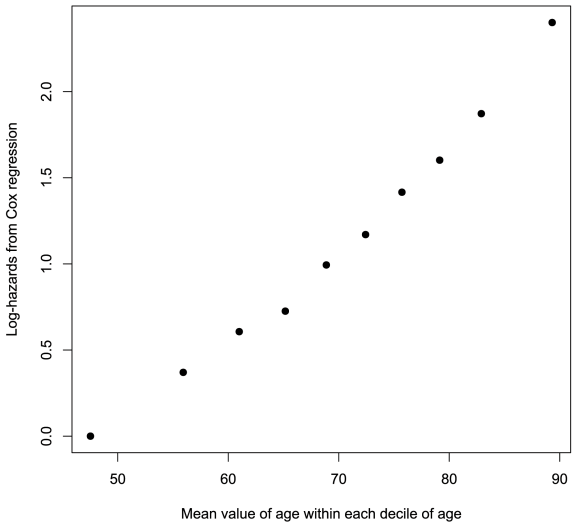


Figure E.2: Changes in the adjusted hazard ratio (HR) and 95% confidence interval (95% CI) for the main covariate of interest based on replacing the missing haemoglobin and serum creatinine with the third, fiftieth, and ninety-seventh percentile from sex and age-matched members of the cohort with measured results. The Cox PH regression modes were built using time-dependent covariates based on the presence of a HF diagnosis and repeat LD prescription. Models were adjusted for age in 2013, sex, SIMD quintile, and the history of the following: AF/AFL, Diabetes+ MI, PAD, hypertension, thyroid disease, cancer, COPD, stroke, and dementia.

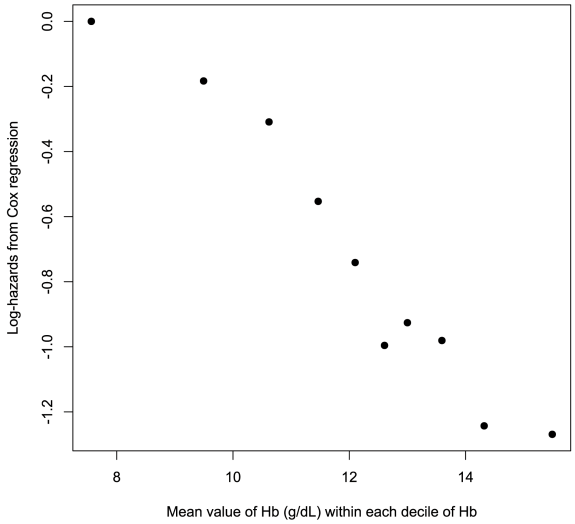
E.5 Assumptions Testing



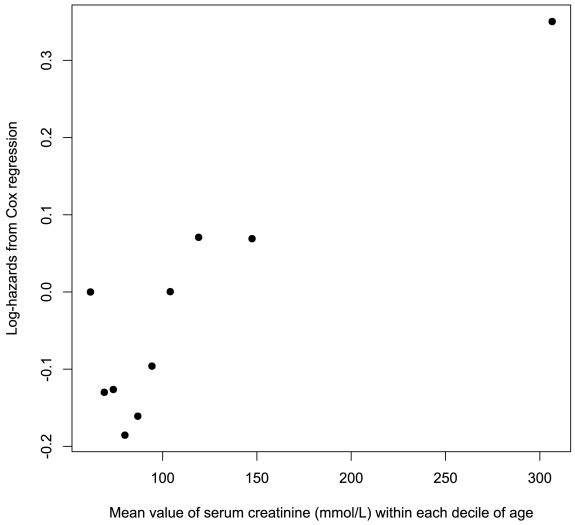
(a) Time-dependent group status



(b) Age in 2013



(c) Time-dependent haemoglobin (Hb) values



(d) Time-dependent serum creatinine

Figure E.3: Testing Cox PH assumptions of PH and linearity assumptions. (a) the PH assumption is not met for the first 36 days, although only this accounted for a small proportion of the deaths. (b) The linearity assumption for patient age in 2013 was met. (c) The linearity assumption for the most severe haemoglobin shows a slight deviation from linearity, but the assumption is close enough to provide a meaningful result. (d) The linearity assumption for serum creatinine was unmet due to the outlier in the highest decile of serum creatinine values.

Bibliography

Adrogué, H. J. & Madias, N. E. (2000*a*), 'Hypernatremia', *New England Journal of Medicine* **342**(20), 1493–1499. PMID: 10816188.
URL: *https://doi.org/10.1056/NEJM200005183422006*

Adrogué, H. J. & Madias, N. E. (2000*b*), 'Hyponatremia', *New England Journal of Medicine* **342**(21), 1581–1589. PMID: 10824078.
URL: *https://doi.org/10.1056/NEJM200005253422107*

Agarwal, P., Moshier, E., Ru, M., Ohri, N., Ennis, R., Rosenzweig, K. & Mazumdar, M. (2018), 'Immortal time bias in observational studies of time-to-event outcomes: Assessing effects of postmastectomy radiation therapy using the national cancer database', *Cancer control : journal of the Moffitt Cancer Center* **25**(1), 1073274818789355–1073274818789355.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/30021466*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6053873/

Algra, A., Tijssen, J. G., Roelandt, J. R., Pool, J. & Lubsen, J. (1991), 'QTc prolongation measured by standard 12-lead electrocardiography is an independent risk factor for sudden death due to cardiac arrest.', *Circulation* **83**(6), 1888–1894.
URL: *https://www.ahajournals.org/doi/abs/10.1161/01.CIR.83.6.1888*

Alvarez-Madrazo, S., McTaggart, S., Nangle, C., Nicholson, E. & Bennie, M. (2016), 'Data resource profile: The Scottish National Prescribing Information System (PIS)', *International Journal of Epidemiology* **45**(3), 714–715f.
URL: *https://doi.org/10.1093/ije/dyw060*

Anand, I. S. & Gupta, P. (2018), 'Anemia and iron deficiency in heart failure', *Circulation* **138**(1), 80–98.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.118.030099*

Andersen, P. K. & Keiding, N. (2002), 'Multi-state models for event history analysis', *Statistical Methods in Medical Research* **11**(2), 91–115. PMID: 12040698.
URL: *https://doi.org/10.1191/0962280202SM276ra*

Anderson, R. J. (1986), 'Hospital-associated hyponatremia', *Kidney International*

29(6), 1237–1247.
URL: *https://www.sciencedirect.com/science/article/pii/S0085253815337170*

Anderson, R. J., chung, H.-m., Kluge, R. & Scherier, R. W. (1985), 'Hyponatremia: A prospective analysis of its epidemiology and the pathogenetic role of vasopressin', *Annals of Internal Medicine* **102**(2), 164–168. PMID: 3966753.
URL: *https://www.acpjournals.org/doi/abs/10.7326/0003-4819-102-2-164*

Anker, M. S., Hülsmann, M. & Cleland, J. G. (2020), 'What do patients with heart failure die from? A single assassin or a conspiracy?', *European Journal of Heart Failure* **22**(1), 26–28.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.1689*

Anker, S. D., Butler, J., Filippatos, G., Ferreira, J. a. P., Bocchi, E., Böhm, M., Brunner–La Rocca, H.-P., Choi, D.-J., Chopra, V., Chuquiure-Valenzuela, E., Giannetti, N., Gomez-Mesa, J. E., Janssens, S., Januzzi, J. L., Gonzalez-Juanatey, J. R., Merkely, B., Nicholls, S. J., Perrone, S. V., Piña, I. L., Ponikowski, P., Senni, M., Sim, D., Spinar, J., Squire, I., Taddei, S., Tsutsui, H., Verma, S., Vinereanu, D., Zhang, J., Carson, P., Lam, C. S. P., Marx, N., Zeller, C., Sattar, N., Jamal, W., Schnaidt, S., Schnee, J. M., Brueckmann, M., Pocock, S. J., Zannad, F. & Packer, M. (2021), 'Empagliflozin in heart failure with a preserved ejection fraction', *New England Journal of Medicine* **385**(16), 1451–1461. PMID: 34449189.
URL: *https://doi.org/10.1056/NEJMoa2107038*

Anwar, H., Fischbacher, C. M., Leese, G. P., Lindsay, R. S., McKnight, J. A., Wild, S. H. & on behalf of the Scottish Diabetes Research Network Epidemiology Group (2011), 'Assessment of the under-reporting of diabetes in hospital admission data: a study from the Scottish Diabetes Research Network Epidemiology Group', *Diabetic Medicine* **28**(12), 1514–1519.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1464-5491.2011.03432.x*

Arnold, S. V., Jones, P. G., Beasley, M., Cordova, J., Goyal, A., Fonarow, G. C. & Seman, L. (2020), 'Heart failure documentation in outpatients with diabetes and volume overload: an observational cohort study from the Diabetes Collaborative Registry', *Cardiovascular Diabetology* **19**(1), 212.
URL: *https://doi.org/10.1186/s12933-020-01190-6*

Arnold, S. V., Yap, J., Lam, C. S., Tang, F., Tay, W. T., Teng, T. H., McGuire, D. K., Januzzi, J. L., Fonarow, G. C., Masoudi, F. A. & Kosiborod, M. (2019), 'Management of patients with diabetes and heart failure with reduced ejection fraction: An international comparison', *Diabetes, Obesity and Metabolism* **21**(2), 261–266.
URL: *https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.13511*

Ashraf, N., Locksley, R. & Arieff, A. I. (1981), 'Thiazide-induced hyponatremia associated with death or neurologic damage in outpatients', *The American Journal of Medicine* **70**(6), 1163–1168.
URL: *https://www.sciencedirect.com/science/article/pii/0002934381908226*

Auguie, B. (2017), *gridExtra: Miscellaneous Functions for "Grid" graphics*. R package version 2.3.
URL: *https://CRAN.R-project.org/package=gridExtra*

Australia/New Zealand Heart Failure Research Collaborative Group (1997), 'Randomised, placebo-controlled trial of carvedilol in patients with congestive heart failure due to ischaemic heart disease', *The Lancet* **349**(9049), 375–380.
URL: *https://www.sciencedirect.com/science/article/pii/S0140673697800086*

Balan, T. A. & Putter, H. (2019), 'frailtyem: An R package for estimating semiparametric shared frailty models', *Journal of Statistical Software* **90**(7), 1–29.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v090i07*

Barnhart, H. X., Yow, E., Crowley, A. L., Daubert, M. A., Rabineau, D., Bigelow, R., Pencina, M. & Douglas, P. S. (2016), 'Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting', *Statistical Methods in Medical Research* **25**(6), 2939–2958. PMID: 24831133.
URL: *https://doi.org/10.1177/0962280214534651*

Bartlett, J. W. & Keogh, R. H. (2018), 'Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration', *Statistical Methods in Medical Research* **27**(6), 1695–1708. PMID: 27647812.
URL: *https://doi.org/10.1177/0962280216667764*

Bastian, M., Heymann, S. & Jacomy, M. (2009), 'Gephi: An open source software for exploring and manipulating networks'.
URL: *http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154*

Baumgartner, H., Hung, J., Bermejo, J., Chambers, J. B., Edvardsen, T., Goldstein, S., Lancellotti, P., LeFevre, M., Miller Jr, F. & Otto, C. M. (2017), 'Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the european association of cardiovascular imaging and the american society of echocardiography', *European Heart Journal-Cardiovascular Imaging* **18**(3), 254–275.

Bay, M., Kirk, V., Parner, J., Hassager, C., Nielsen, H., Krogsgaard, K., Trawinski, J., Boesgaard, S. & Aldershvile, J. (2003), 'NT-proBNP: a new diagnostic screening tool to differentiate between patients with normal and reduced left ventricular systolic function',

Heart **89**(2), 150–154.
URL: *https://heart.bmj.com/content/89/2/150*

Bazett, H. C. (1920), 'An analysis of the time relations of electrocardiograms', *Heart* **7**, 353–370.

Beale, A. L., Meyer, P., Marwick, T. H., Lam, C. S. & Kaye, D. M. (2018), 'Sex differences in cardiovascular pathophysiology', *Circulation* **138**(2), 198–205.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.118.034271*

Bean, D. M., Stringer, C., Beeknoo, N., Teo, J. & Dobson, R. J. B. (2017), 'Network analysis of patient flow in two UK acute care hospitals identifies key sub-networks for a&e performance', *PLOS ONE* **12**(10), 1–16.
URL: *https://doi.org/10.1371/journal.pone.0185912*

Beckett, N. S., Peters, R., Fletcher, A. E., Staessen, J. A., Liu, L., Dumitrascu, D., Stoyanovsky, V., Antikainen, R. L., Nikitin, Y., Anderson, C., Belhani, A., Forette, F., Rajkumar, C., Thijs, L., Banya, W. & Bulpitt, C. J. (2008), 'Treatment of hypertension in patients 80 years of age or older', *New England Journal of Medicine* **358**(18), 1887–1898. PMID: 18378519.
URL: *https://doi.org/10.1056/NEJMoa0801369*

Bello, J. & Pellegrini, M. V. (2022), Mavacamten, *in* 'StatPearls [Internet]', StatPearls Publishing.

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., de Ferranti, S. D., Floyd, J., Fornage, M., Gillespie, C., Isasi, C. R., Jimenez, M. C., Jordan, L. C., Judd, S. E., Lackland, D., Lichtman, J. H., Lisabeth, L., Liu, S., Longenecker, C. T., Mackey, R. H., Matsushita, K., Mozaffarian, D., Mussolino, M. E., Nasir, K., Neumar, R. W., Palaniappan, L., Pandey, D. K., Thiagarajan, R. R., Reeves, M. J., Ritchey, M., Rodriguez, C. J., Roth, G. A., Rosamond, W. D., Sasson, C., Towfighi, A., Tsao, C. W., Turner, M. B., Virani, S. S., Voeks, J. H., Willey, J. Z., Wilkins, J. T., Wu, J. H., Alger, H. M., Wong, S. S., Muntner, P. & American Heart Association Statistics, Committee and Stroke Statistics, Subcommittee (2017), 'Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association', *Circulation* **135**(10), e146–e603.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/28122885*

Berend, K., van Hulsteijn, L. H. & Gans, R. O. (2012), 'Chloride: The queen of electrolytes?', *European Journal of Internal Medicine* **23**(3), 203–211.
URL: *https://www.sciencedirect.com/science/article/pii/S0953620511002779*

Berlin, A. (2009), 'Death certification: topical tips for GPs', *London Journal of Primary Care* **2**(2), 130–137.

Bernoulli, D. (1760), 'Essai d'une nouvelle analyse de la mortalité causée par la petite vérole,

et des avantages de l'inoculation pour la prévenir', *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem* pp. 1–45.

Bhaskaran, K. & Smeeth, L. (2014), 'What is the difference between missing completely at random and missing at random?', *International journal of epidemiology* **43**(4), 1336–1339.

Booth, N. (1994), 'What are the Read Codes?', *Health Libraries Review* **11**(3), 177–182.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2532.1994.1130177.x*

Bos, O. J., Fischer, M. J., Wilting, J. & Janssen, L. H. (1989), 'Mechanism by which warfarin binds to human serum albumin: Stopped-flow kinetic experiments with two large fragments of albumin', *Biochemical Pharmacology* **38**(12), 1979–1984.
URL: *https://www.sciencedirect.com/science/article/pii/0006295289904978*

Bottle, A., Kim, D., Aylin, P., Cowie, M. R., Majeed, A. & Hayhoe, B. (2018), 'Routes to diagnosis of heart failure: observational study using linked data in England', **104**(7), 600–605.
URL: *https://heart.bmj.com/content/heartjnl/104/7/600.full.pdf*

Bozkurt, B., Coats, A. J., Tsutsui, H., Abdelhamid, M., Adamopoulos, S., Albert, N., Anker, S. D., Atherton, J., Böhm, M., Butler, J., Drazner, M. H., Felker, G. M., Filippatos, G., Fonarow, G. C., Fiuzat, M., Gomez-Mesa, J., Heidenreich, P., Imamura, T., Januzzi, J., Jankowska, E. A., Khazanie, P., Kinugawa, K., Lam, C. S., Matsue, Y., Metra, M., Ohtani, T., Francesco Piepoli, M., Ponikowski, P., Rosano, G. M., Sakata, Y., SeferoviĆ, P., Starling, R. C., Teerlink, J. R., Vardeny, O., Yamamoto, K., Yancy, C., Zhang, J. & Zieroth, S. (2021), 'Universal Definition and Classification of Heart Failure: A Report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure', *Journal of Cardiac Failure* **27**(4), 387–413.
URL: *https://www.sciencedirect.com/science/article/pii/S1071916421000506*

Bozkurt, B. & Khalaf, S. (2017), 'Heart failure in women', *Methodist DeBakey Cardiovascular Journal* .

Brady, P. F., Chua, W., Nehaj, F., Connolly, D. L., Khashaba, A., Purmah, Y. J. V., Ul-Qamar, M. J., Thomas, M. R., Varma, C., Schnabel, R. B., Zeller, T., Fabritz, L. & Kirchhof, P. F. (2022), 'Interactions between atrial fibrillation and natriuretic peptide in predicting heart failure hospitalization or cardiovascular death', *Journal of the American Heart Association* **11**(4), e022833.
URL: *https://www.ahajournals.org/doi/abs/10.1161/JAHA.121.022833*

Braunschweig, F., Linde, C., Eriksson, M., Hofman-Bang, C. & Rydén, L. (2002), 'Continous haemodynamic monitoring during withdrawal of diuretics in patients with congestive heart

failure', *European Heart Journal* **23**(1), 59–69.
URL: *https://doi.org/10.1053/euhj.2001.2690*

Braunwald, E. (1997), 'Cardiovascular medicine at the turn of the millennium: Triumphs, concerns, and opportunities', *New England Journal of Medicine* **337**(19), 1360–1369. PMID: 9358131.
URL: *https://doi.org/10.1056/NEJM199711063371906*

Bruderer, S., Bodmer, M., Jick, S. S. & Meier, C. R. (2014), 'Use of diuretics and risk of incident gout: A population-based case–control study', *Arthritis & Rheumatology* **66**(1), 185–196.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/art.38203*

Butler, J., Fonarow, G. C., Zile, M. R., Lam, C. S., Roessig, L., Schelbert, E. B., Shah, S. J., Ahmed, A., Bonow, R. O., Cleland, J. G. F., Cody, R. J., Chioncel, O., Collins, S. P., Dunnmon, P., Filippatos, G., Lefkowitz, M. P., Marti, C. N., McMurray, J. J., Misselwitz, F., Nodari, S., O'Connor, C., Pfeffer, M. A., Pieske, B., Pitt, B., Rosano, G., Sabbah, H. N., Senni, M., Solomon, S. D., Stockbridge, N., Teerlink, J. R., Georgiopoulou, V. V. & Gheorghiade, M. (2014), 'Developing therapies for heart failure with preserved ejection fraction: Current state and future directions', *JACC: Heart Failure* **2**(2), 97–112.
URL: *https://www.sciencedirect.com/science/article/pii/S2213177914000213*

Calderwood, C. & Slater, A. (2018), 'Guidene for doctors completing medical certificates of the cause of death (MCCD) and its quality assurance'.
URL: *https://www.gov.scot/publications/medical-certificates-of-cause-of-death-guidance-on-completion/*

Carabello, B. A. & Paulus, W. J. (2009), 'Aortic stenosis', *The Lancet* **373**(9667), 956–966.
URL: *https://www.sciencedirect.com/science/article/pii/S0140673609602117*

Carlson, K. J., Lee, D. C.-S., Goroll, A. H., Leahy, M. & Johnson, R. A. (1985), 'An analysis of physicians' reasons for prescribing long-term digitalis therapy in outpatients', *Journal of Chronic Diseases* **38**(9), 733–739.
URL: *https://www.sciencedirect.com/science/article/pii/0021968185901158*

Carlson, S. C. (2010), 'Königsberg bridge problem'.
URL: *https://www.britannica.com/science/Konigsberg-bridge-problem*

Carter Grinstead, W., Francis, M. J., Marks, G. F., Tawa, C. B., Zoghbi, W. A. & Young, J. B. (1994), 'Discontinuation of chronic diuretic therapy in stable congestive heart failure secondary to coronary artery disease or to idiopathic dilated cardiomyopathy', *The American Journal of Cardiology* **73**(12), 881–886.
URL: *https://www.sciencedirect.com/science/article/pii/000291499490815X*

Chioncel, O., Lainscak, M., Seferovic, P. M., Anker, S. D., Crespo-Leiro, M. G., Harjola, V.-P., Parissis, J., Laroche, C., Piepoli, M. F., Fonseca, C., Mebazaa, A., Lund, L., Ambrosio, G. A., Coats, A. J., Ferrari, R., Ruschitzka, F., Maggioni, A. P. & Filippatos, G. (2017), 'Epidemiology and one-year outcomes in patients with chronic heart failure and preserved, mid-range and reduced ejection fraction: an analysis of the ESC Heart Failure Long-Term Registry', *European Journal of Heart Failure* **19**(12), 1574–1585.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.813*

Chisholm, J. (1990), 'The Read clinical classification', *BMJ (Clinical research ed.)* **300**(6732), 1092–1092.
URL: *https://www.bmj.com/content/300/6732/1092*

Christie, M. (2022), 'Defining comorbidities'.

CIBIS Investigators and Committees (1994), 'A randomized trial of beta-blockade in heart failure. The Cardiac Insufficiency Bisoprolol Study (CIBIS)', *Circulation* **90**(4), 1765–1773.
URL: *https://www.ahajournals.org/doi/abs/10.1161/01.CIR.90.4.1765*

Clark, A. L., Fonarow, G. C. & Horwich, T. B. (2014), 'Obesity and the obesity paradox in heart failure', *Progress in Cardiovascular Diseases* **56**(4), 409–414.
URL: *https://www.sciencedirect.com/science/article/pii/S0033062013001771*

Clark, D. A., Stinson, E. B., Griepp, R. B., Schroeder, J. S., Shumway, N. E. & Harrison, D. C. (1971), 'Cardiac transplantation in man: Vi. prognosis of patients selected for cardiac transplantation', *Annals of Internal Medicine* **75**(1), 15–21.

Clarke, K., Gray, D. & Hampton, J. (1994), 'Evidence of inadequate investigation and treatment of patients with heart failure', *British heart journal* **71**(6), 584—587.
URL: *https://europepmc.org/articles/PMC1025459*

Clarke, K. W., Gray, D. & Hampton, J. R. (1995), 'How common is heart failure? Evidence from PACT (Prescribing Analysis and Cost) data in Nottingham', *Journal of Public Health* **17**(4), 459–464.
URL: *https://doi.org/10.1093/oxfordjournals.pubmed.a043163*

Cleland, J. G. (2002), 'How to assess new treatments for the management of heart failure: composite scoring systems to assess the patients' clinical journey', *European Journal of Heart Failure* **4**(3), 243–247.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1016/S1388-9842%2802%2900039-9*

Cleland, J. G., Carubelli, V., Castiello, T., Yassin, A., Pellicori, P. & Antony, R. (2012), 'Renal dysfunction in acute and chronic heart failure: prevalence, incidence and prognosis', *Heart Fail Rev* **17**(2), 133–49.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/22426973*

Cleland, J. G., Chiswell, K., Teerlink, J. R., Stevens, S., Fiuzat, M., Givertz, M. M., Davison, B. A., Mansoor, G. A., Ponikowski, P., Voors, A. A., Cotter, G., Metra, M., Massie, B. M. & O'Connor, C. M. (2014), 'Predictors of postdischarge outcomes from information acquired shortly after admission for acute heart failure', *Circulation: Heart Failure* **7**(1), 76–87.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.113.000284*

Cleland, J. G. F., Charlesworth, A., Lubsen, J., Swedberg, K., Remme, W. J., Erhardt, L., Di Lenarda, A., Komajda, M., Metra, M., Torp-Pedersen, C. & Poole-Wilson, P. A. (2006), 'A comparison of the effects of carvedilol and metoprolol on well-being, morbidity, and mortality (the "patient journey") in patients with heart failure: A report from the Carvedilol Or Metoprolol European Trial (COMET)', *Journal of the American College of Cardiology* **47**(8), 1603–1611.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109706002907*

Cleland, J. G. F., Ferreira, J. P., Mariottoni, B., Pellicori, P., Cuthbert, J., Verdonschot, J. A. J., Petutschnigg, J., Ahmed, F. Z., Cosmi, F., Brunner La Rocca, H.-P., Mamas, M. A., Clark, A. L., Edelmann, F., Pieske, B., Khan, J., McDonald, K., Rouet, P., Staessen, J. A., Mujaj, B., González, A., Diez, J., Hazebroek, M., Heymans, S., Latini, R., Grojean, S., Pizard, A., Girerd, N., Rossignol, P., Collier, T. J., Zannad, F. & the HOMAGE Trial Committees and Investigators (2020), 'The effect of spironolactone on cardiovascular function and markers of fibrosis in people at increased risk of developing heart failure: the heart 'OMics' in AGEing (HOMAGE) randomized clinical trial', *European Heart Journal* **42**(6), 684–696.
URL: *https://doi.org/10.1093/eurheartj/ehaa758*

Cleland, J. G. F., Pfeffer, M. A., Clark, A. L., Januzzi, J. L., McMurray, J. J., Mueller, C., Pellicori, P., Richards, M., Teerlink, J. R., Zannad, F. & Bauersachs, J. (2021), 'The struggle towards a Universal Definition of Heart Failure—how to proceed?', *European Heart Journal* **42**(24), 2331–2343.
URL: *https://doi.org/10.1093/eurheartj/ehab082*

Cleland, J. G. F., Shelton, R., Nikitin, N., Ford, S., Frison, L. & Grind, M. (2007), 'Prevalence of markers of heart failure in patients with atrial fibrillation and the effects of ximelagatran compared to warfarin on the incidence of morbid and fatal events: A report from the SPORTIF III and V trials', *European Journal of Heart Failure* **9**(6-7), 730–739.
URL: *https://doi.org/10.1016/j.ejheart.2007.01.013*

Cleland, J. G., McDonagh, T., Rigby, A. S., Yassin, A., Whittaker, T., Dargie, H. J. & National Heart Failure Audit Team for England and Wales (2011), 'The national heart failure audit for England and Wales 2008-2009', *Heart* **97**(11), 876–86.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/21173198*

Cleland, J. G. & McGowan, J. (1999), 'Heart failure due to ischaemic heart disease:

Epidemiology, pathophysiology and progression', *Journal of Cardiovascular Pharmacology* **33**, S17–S29.
URL: *https://journals.lww.com/cardiovascularpharm/Fulltext/1999/00003/Heart_Failure_-due_to_Ischaemic_Heart_Disease_.3.aspx*

Cody, R. J., Franklin, K. W. & Laragh, J. H. (1982), 'Postural hypotension during tilt with chronic captopril and diuretic therapy of severe congestive heart failure', *American Heart Journal* **103**(4, Part 1), 480–484.
URL: *http://www.sciencedirect.com/science/article/pii/0002870382903337*

Cohen, J. B., Brown, N. J., Brown, S.-A., Dent, S., van Dorst, D. C., Herrmann, S. M., Lang, N. N., Oudit, G. Y., Touyz, R. M. & null null (2023), 'Cancer Therapy–Related Hypertension: A Scientific Statement From the American Heart Association', *Hypertension* **80**(3), e46–e57.
URL: *https://www.ahajournals.org/doi/abs/10.1161/HYP.0000000000000224*

Cohn, J. N., Levine, T. B., Olivari, M. T., Garberg, V., Lura, D., Francis, G. S., Simon, A. B. & Rector, T. (1984), 'Plasma norepinephrine as a guide to prognosis in patients with chronic congestive heart failure', **311**(13), 819–823.
URL: *https://www.nejm.org/doi/full/10.1056/NEJM198409273111303*

Coiro, S., Girerd, N., McMurray, J. J., Pitt, B., Swedberg, K., van Veldhuisen, D. J., Lamiral, Z., Rossignol, P. & Zannad, F. (2021), 'Diuretic therapy as prognostic enrichment factor for clinical trials in patients with heart failure with reduced ejection fraction', *Clinical Research in Cardiology* **110**(8), 1308–1320.
URL: *https://doi.org/10.1007/s00392-021-01851-w*

Conrad, N., Judge, A., Canoy, D., Tran, J., Pinho-Gomes, A. C., Millett, E. R. C., Salimi-Khorshidi, G., Cleland, J. G., McMurray, J. J. V. & Rahimi, K. (2019), 'Temporal trends and patterns in mortality after incident heart failure: A longitudinal analysis of 86,000 individuals', *JAMA Cardiol* **4**(11), 1102–1111.

Conrad, N., Judge, A., Tran, J., Mohseni, H., Hedgecott, D., Crespillo, A. P., Allison, M., Hemingway, H., Cleland, J. G., McMurray, J. J. V. & Rahimi, K. (2018), 'Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals', *Lancet* **391**(10120), 572–580.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/29174292*

Consensus Trial Study Group (1987), 'Effects of enalapril on mortality in severe congestive heart failure', *New England Journal of Medicine* **316**(23), 1429–1435.

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society.*

Series B (Methodological) **34**(2), 187–220.
URL: *http://www.jstor.org/stable/2985181*

Creed, F. & Hargreaves, J. (2016), *Oxford handbook of critical care nursing*, Oxford University Press.

Curtis, J. P., Sokol, S. I., Wang, Y., Rathore, S. S., Ko, D. T., Jadbabaie, F., Portnay, E. L., Marshalko, S. J., Radford, M. J. & Krumholz, H. M. (2003), 'The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure', *Journal of the American College of Cardiology* **42**(4), 736–742.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109703007897*

Cuthbert, J. J., Gopal, J., Crundall-Goode, A. & Clark, A. L. (2020), 'Are there patients missing from community heart failure registers? An audit of clinical practice', *European Journal of Preventive Cardiology* **26**(3), 291–298.
URL: *https://doi.org/10.1177/2047487318810839*

Cuthbert, J. J., Pellicori, P., Flockton, R., Kallvikbacka-Bennett, A., Khan, J., Rigby, A. S., Girerd, N., Zannad, F., Cleland, J. G. & Clark, A. L. (2021), 'The prevalence and clinical associations of ultrasound measures of congestion in patients at risk of developing heart failure', *European Journal of Heart Failure* **23**(11), 1831–1840.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.2353*

Czwikla, J., Jobski, K. & Schink, T. (2017), 'The impact of the lookback period and definition of confirmatory events on the identification of incident cancer cases in administrative data', *BMC Medical Research Methodology* **17**(1), 122.
URL: *https://doi.org/10.1186/s12874-017-0407-4*

Damman, K., Kjekshus, J., Wikstrand, J., Cleland, J. G., Komajda, M., Wedel, H., Waagstein, F. & McMurray, J. J. (2016), 'Loop diuretics, renal function and clinical outcome in patients with heart failure and reduced ejection fraction', *Eur J Heart Fail* **18**(3), 328–36.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/26693947*

Danielsen, R., Thorgeirsson, G., Einarsson, H., Örn Ólafsson, Aspelund, T., Harris, T. B., Launer, L. & Gudnason, V. (2017), 'Prevalence of heart failure in the elderly and future projections: the AGES-Reykjavík study', *Scandinavian Cardiovascular Journal* **51**(4), 183–189. PMID: 28366010.
URL: *https://doi.org/10.1080/14017431.2017.1311023*

Dauw, J., Martens, P., Tersalvi, G., Schouteden, J., Deferm, S., Gruwez, H., De Moor, B., Nijst, P., Dupont, M. & Mullens, W. (2021), 'Diuretic response and effects of diuretic omission in ambulatory heart failure patients on chronic low-dose loop diuretic therapy', *European*

Journal of Heart Failure **23**(7), 1110–1119.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.2145*

Davies, M., Hobbs, F., Davis, R., Kenkre, J., Roalfe, A., Hare, R., Wosornu, D. & Lancashire, R. (2001), 'Prevalence of left-ventricular systolic dysfunction and heart failure in the echocardiographic heart of england screening study: a population based study', *The Lancet* **358**(9280), 439–444.
URL: *https://www.sciencedirect.com/science/article/pii/S0140673601056203*

de Miguel-Yanes, J. M., Jiménez-García, R., Hernandez-Barrera, V., de Miguel-Díez, J., Muñoz-Rivas, N., Méndez-Bailón, M., Pérez-Farinós, N., López-Herranz, M. & Lopez-de Andres, A. (2021), 'Sex differences in the incidence and outcomes of acute myocardial infarction in spain, 2016–2018: A matched-pair analysis', *Journal of Clinical Medicine* **10**(8).
URL: *https://www.mdpi.com/2077-0383/10/8/1795*

de Wreede, L. C., Fiocco, M. & Putter, H. (2010), 'The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models', *Computer Methods and Programs in Biomedicine* **99**, 261–274.

de Wreede, L. C., Fiocco, M. & Putter, H. (2011), 'mstate: An R package for the analysis of competing risks and multi-state models', *Journal of Statistical Software* **38**(7), 1–30.
URL: *https://www.jstatsoft.org/v38/i07/*

Denaxas, S. C., George, J., Herrett, E., Shah, A. D., Kalra, D., Hingorani, A. D., Kivimaki, M., Timmis, A. D., Smeeth, L. & Hemingway, H. (2012), 'Data resource profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER)', *International Journal of Epidemiology* **41**(6), 1625–1638.
URL: *https://doi.org/10.1093/ije/dys188*

Denaxas, S. C. & Morley, K. I. (2015), 'Big biomedical data and cardiovascular disease research: opportunities and challenges', *European Heart Journal - Quality of Care and Clinical Outcomes* **1**(1), 9–16.
URL: *https://doi.org/10.1093/ehjqcco/qcv005*

Di Bari, M., Pozzi, C., Cavallini, M. C., Innocenti, F., Baldereschi, G., De Alfieri, W., Antonini, E., Pini, R., Masotti, G. & Marchionni, N. (2004), 'The diagnosis of heart failure in the community. comparative validation of four sets of criteria in unselected older adults: the ICARe Dicomano Study', *J Am Coll Cardiol* **44**(8), 1601–8.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/15489092*

Diao, J. A., Inker, L. A., Levey, A. S., Tighiouart, H., Powe, N. R. & Manrai, A. K. (2021), 'In search of a better equation — performance and equity in estimates of kidney function', *New*

England Journal of Medicine **384**(5), 396–399. PMID: 33406354.
URL: *https://doi.org/10.1056/NEJMp2028243*

Domanski, M., Norman, J., Pitt, B., Haigney, M., Hanlon, S. & Peyster, E. (2003), 'Diuretic use, progressive heart failure, and death in patients in the studies of left ventricular dysfunction (SOLVD)', *Journal of the American College of Cardiology* **42**(4), 705–708.

Dong, Y. & Peng, C.-Y. J. (2013), 'Principled missing data methods for researchers', *SpringerPlus* **2**(1), 222.
URL: *https://doi.org/10.1186/2193-1801-2-222*

Dovancescu, S., Pellicori, P., Mabote, T., Torabi, A., Clark, A. L. & Cleland, J. G. (2017), 'The effects of short-term omission of daily medication on the pathophysiology of heart failure', *European Journal of Heart Failure* **19**(5), 643–649.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.748*

Dunlay, S. M., Givertz, M. M., Aguilar, D., Allen, L. A., Chan, M., Desai, A. S., Deswal, A., Dickson, V. V., Kosiborod, M. N., Lekavich, C. L., McCoy, R. G., Mentz, R. J., Piña, I. L. & On behalf of the American Heart Association Heart Failure and Transplantation Committee of the Council on Clinical Cardiology; Council on Cardiovascular and Stroke Nursing; and the Heart Failure Society of America (2019), 'Type 2 diabetes mellitus and heart failure: A scientific statement from the American Heart Association and the Heart Failure Society of America: This statement does not represent an update of the 2017 ACC/AHA/HFSA heart failure guideline update', *Circulation* **140**(7), e294–e324.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000691*

D'Hoore, W., Sicotte, C. & Tilquin, C. (1993), 'Risk adjustment in outcome assessment: the Charlson comorbidity index', *Methods of information in medicine* **32**(05), 382–387.

Edelman, I. S., Leibman, J., O'Meara, M. P. & Birkenfeld, L. W. (1958), 'Interrelations between serum sodium concentration, serum osmolarity and total exchangeable sodium, total exchangeable potassium and total body water', *J Clin Invest* **37**(9), 1236–56.

Ehteshami-Afshar, S., Mooney, L., Dewan, P., Desai, A. S., Lang, N. N., Lefkowitz, M. P., Petrie, M. C., Rizkala, A. R., Rouleau, J. L., Solomon, S. D., Swedberg, K., Shi, V. C., Zile, M. R., Packer, M., McMurray, J. J., Jhund, P. S. & Hawkins, N. M. (2021), 'Clinical characteristics and outcomes of patients with heart failure with reduced ejection fraction and chronic obstructive pulmonary disease: Insights from PARADIGM-HF', *Journal of the American Heart Association* **10**(4), e019238.
URL: *https://www.ahajournals.org/doi/abs/10.1161/JAHA.120.019238*

Ellison, D. H. & Felker, G. M. (2017), 'Diuretic treatment in heart failure', *N Engl J Med*

377(20), 1964–1975.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/29141174*

Eriksson, H., Caidaul, K., Larsson, B., Ohlson, L.-O., Welin, L., Wilhelmsen, L. & Svärdsudd, K. (1987), 'Cardiac and pulmonary causes of dyspnoea—validation of a scoring test for clinical-epidemiological use: The Study of Men Born in 1913', *European Heart Journal* **8**(9), 1007–1014.
URL: *https://doi.org/10.1093/oxfordjournals.eurheartj.a062365*

Eshaghian, S., Horwich, T. B. & Fonarow, G. C. (2006), 'Relation of loop diuretic dose to mortality in advanced heart failure', *Am J Cardiol* **97**(12), 1759–64.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/16765130*

Etikan, I., Abubakar, S. & Alkassim, R. (2017), 'The Kaplan-Meier estimate in survival analysis', *Biom Biostatistics Int J* **5**(2), 00128.

Euler, L. (1741), 'Solutio problematis ad geometriam situs pertinentis', *Commentarii academiae scientiarum Petropolitanae* pp. 128–140.

Ezekowitz, J. A., Kaul, P., Bakal, J. A., Quan, H. & McAlister, F. A. (2011), 'Trends in heart failure care: has the incident diagnosis of heart failure shifted from the hospital to the emergency department and outpatient clinics?', *European Journal of Heart Failure* **13**(2), 142–147.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hfq185*

Faris, R., Flather, M., Purcell, H., Poole-Wilson, P., & Coats, A. (2012), 'Diuretics for heart failure', *Cochrane Database of Systematic Reviews* (2).
URL: *https://doi.org//10.1002/14651858.CD003838.pub3*

Faselis, C., Lam, P. H., Patel, S., Arundel, C., Filippatos, G., Deedwania, P., Zile, M. R., Wopperer, S., Nguyen, T., Allman, R. M., Fonarow, G. C. & Ahmed, A. (2021), 'Loop diuretic prescription and long-term outcomes in heart failure: Association modification by congestion', *The American Journal of Medicine* **134**(6), 797–804.
URL: *https://www.sciencedirect.com/science/article/pii/S0002934320311207*

Feigenbaum, M. S., Welsch, M. A., Mitchell, M., Vincent, K., Braith, R. W. & Pepine, C. J. (2000), 'Contracted plasma and blood volume in chronic heart failure', *Journal of the American College of Cardiology* **35**(1), 51–55.

Fichman, M. P., Voherr, H., Kleeman, C. R. & Telfer, N. (1971), 'Diuretic-induced hyponatremia', *Annals of Internal Medicine* **75**(6), 853–863. PMID: 4944156.
URL: *https://www.acpjournals.org/doi/abs/10.7326/0003-4819-75-6-853*

Fisher, L. D. & Lin, D. Y. (1999), 'Time-dependent covariates in the cox proportional-hazards

regression model', *Annual Review of Public Health* **20**(1), 145–157.
URL: *https://doi.org/10.1146/annurev.publhealth.20.1.145*

Flamenbaum, W. (1986), 'Diuretic use in the elderly: Potential for diuretic-induced hypokalemia', *The American Journal of Cardiology* **57**(2), A38–A43.
URL: *http://www.sciencedirect.com/science/article/pii/0002914986910052*

Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand, J.-F. & Daurès, J.-P. (2005), 'A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease', *Biometrical Journal* **47**(6), 825–833.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.200410170*

Franklin, J. M. & Schneeweiss, S. (2017), 'When and how can real world data analyses substitute for randomized controlled trials?', *Clinical Pharmacology & Therapeutics* **102**(6), 924–933.
URL: *https://ascpt.onlinelibrary.wiley.com/doi/abs/10.1002/cpt.857*

French, L. (2017), 'Prescribing data: BNF codes'.
URL: *https://www.bennett.ox.ac.uk/blog/2017/04/prescribing-data-bnf-codes/#: :text=The%20BNF%20codes%20from%20this,BNF%20a%20drug%20is%20from.*

Fridericia, L. (1920), 'Die systolendauer im elektrokardiogramm bei normalen menschen und bei herzkranken i: Beziehung zwischen der pulsfrequenz und der dauer des ventrikelelektrokardiogramms bei normalen menschen in der ruhe', *Acta Medica Scandinavica* **53**, 469–486.

Fruchterman, T. M. J. & Reingold, E. M. (1991), 'Graph drawing by force-directed placement', *Software: Practice and Experience* **21**(11), 1129–1164.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.4380211102*

Gail, M. H. (1972), 'Does cardiac transplantation prolong life? A reassessment', *Annals of Internal Medicine* **76**(5), 815–817.

Galve, E., Mallol, A., Catalan, R., Palet, J., Méndez, S., Nieto, E., Diaz, A. & Soler-Soler, J. (2005), 'Clinical and neurohumoral consequences of diuretic withdrawal in patients with chronic, stabilized heart failure and systolic dysfunction', *European Journal of Heart Failure* **7**(5), 892–898.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1016/j.ejheart.2004.09.006*

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro, A., Sciaini, Marco, Scherer & Cédric (2021*a*), *ggpubr: 'ggplot2' Based Publication Read Plots*. R package version 0.6.2.
URL: *https://sjmgarnier.github.io/viridis/*

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro, A., Sciaini, Marco, Scherer &

Cédric (2021*b*), *viridis - Colorblind-Friendly Color Maps for R*. R package version 0.6.2.
URL: *https://sjmgarnier.github.io/viridis/*

Gasperoni, F., Ieva, F., Barbati, G., Scagnetto, A., Iorio, A., Sinagra, G. & Di Lenarda, A. (2017), 'Multi-state modelling of heart failure care path: A population-based investigation from Italy', *PLOS ONE* **12**(6), e0179176.
URL: *https://doi.org/10.1371/journal.pone.0179176*

Gburek, J., Konopska, B. & Gołąb, K. (2021), 'Renal handling of albumin—from early findings to current concepts', *International Journal of Molecular Sciences* **22**(11).
URL: *https://www.mdpi.com/1422-0067/22/11/5809*

Gho, J. M. I. H., Schmidt, A. F., Pasea, L., Koudstaal, S., Pujades-Rodriguez, M., Denaxas, S., Shah, A. D., Patel, R. S., Gale, C. P., Hoes, A. W., Cleland, J. G., Hemingway, H. & Asselbergs, F. W. (2018), 'An electronic health records cohort study on heart failure following myocardial infarction in England: incidence and predictors', **8**(3), e018331.
URL: *https://bmjopen.bmj.com/content/bmjopen/8/3/e018331.full.pdf*

Girerd, N., Seronde, M.-F., Coiro, S., Chouihed, T., Bilbault, P., Braun, F., Kenizou, D., Maillier, B., Nazeyrollas, P., Roul, G., Fillieux, L., Abraham, W. T., Januzzi, J., Sebbag, L., Zannad, F., Mebazaa, A. & Rossignol, P. (2018), 'Integrative assessment of congestion in heart failure throughout the patient journey', *JACC: Heart Failure* **6**(4), 273–285.
URL: *https://www.sciencedirect.com/science/article/pii/S2213177917306881*

Goel, M. K., Khanna, P. & Kishore, J. (2010), 'Understanding survival analysis: Kaplan-Meier estimate', *Int J Ayurveda Res* **1**(4), 274–8.

Graham, F., Friday, J., Pellicori, P. & Cleland, J. (2020), 'Haemoglobin and serum markers of iron deficiency in people with or at increased risk of heart failure', *Heart* **106**(Suppl 2), A60–A61.
URL: *https://heart.bmj.com/content/106/Suppl_2/A60*

Graham, F. J., Friday, J. M., Pellicori, P., Greenlaw, N. & Cleland, J. G. (2023), 'Assessment of haemoglobin and serum markers of iron deficiency in people with cardiovascular disease', *Heart* **109**(17), 1294–1301.
URL: *https://heart.bmj.com/content/109/17/1294*

Graham, F. J., Masini, G., Pellicori, P., Cleland, J. G., Greenlaw, N., Friday, J., Kazmi, S. & Clark, A. L. (2022), 'Natural history and prognostic significance of iron deficiency and anaemia in ambulatory patients with chronic heart failure', *European Journal of Heart Failure* **24**(5), 807–817.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.2251*

Gray, B. (2020), *cmprsk: Subdistribution analysis of Competing Risks*. R package version

2.2-10.
URL: *https://CRAN.R-project.org/package=cmprsk*

Groenland, E. H., Bots, M. L., Asselbergs, F. W., de Borst, G. J., Kappelle, L. J., Visseren, F. L. & Spiering, W. (2021), 'Apparent treatment resistant hypertension and the risk of recurrent cardiovascular events and mortality in patients with established vascular disease', *International Journal of Cardiology* **334**, 135–141.
URL: *https://www.sciencedirect.com/science/article/pii/S0167527321007361*

Grolemund, G. & Wickham, H. (2011), 'Dates and times made easy with lubridate', *Journal of Statistical Software* **40**(3), 1–25. R package version 1.8.0.
URL: *https://www.jstatsoft.org/v40/i03/*

Groopman, J. E. & Itri, L. M. (1999), 'Chemotherapy-Induced Anemia in Adults: Incidence and Treatment', *JNCI: Journal of the National Cancer Institute* **91**(19), 1616–1634.
URL: *https://doi.org/10.1093/jnci/91.19.1616*

Guglin, M. (2012), 'Congestion is the driving force behind heart failure', *Curr Heart Fail Rep* **9**(3), 219–27.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/22699923*

Halliday, B. P., Wassall, R., Lota, A. S., Khalique, Z., Gregson, J., Newsome, S., Jackson, R., Rahneva, T., Wage, R., Smith, G., Venneri, L., Tayal, U., Auger, D., Midwinter, W., Whiffin, N., Rajani, R., Dungu, J. N., Pantazis, A., Cook, S. A., Ware, J. S., Baksi, A. J., Pennell, D. J., Rosen, S. D., Cowie, M. R., Cleland, J. G. F. & Prasad, S. K. (2019), 'Withdrawal of pharmacological treatment for heart failure in patients with recovered dilated cardiomyopathy (TRED-HF): an open-label, pilot, randomised trial', *The Lancet* **393**(10166), 61–73.
URL: *https://www.sciencedirect.com/science/article/pii/S014067361832484X*

Halpin, T. A. & Morgan, T. (2008), *Information modeling and relational databases*, 2nd edn, Morgan Kaufmann Publishers, Burlington, MA.

Hamilton, A. (2022), 'GP LES coverage'.

Hampton, J. & Hampton, J. (2019), *The ECG made easy e-book*, Elsevier Health Sciences.

Han, X., Jiang, F., Zhou, H., Needleman, J., Guo, M., Chen, Y., Liu, Y. & Tang, Y. (2019), 'Hospitalization Pattern, Inpatient Service Utilization and Quality of Care in Patients With Alcohol Use Disorder: A Sequence Analysis of Discharge Medical Records', *Alcohol and Alcoholism* **55**(2), 179–186.
URL: *https://doi.org/10.1093/alcalc/agz081*

Hanberg, J. S., Rao, V., ter Maaten, J. M., Laur, O., Brisco, M. A., Wilson, F. P., Grodin, J. L.,

Assefa, M., Broughton, J. S., Planavsky, N. J., Ahmad, T., Bellumkonda, L., Tang, W. H. W., Parikh, C. R. & Testani, J. M. (2016), 'Hypochloremia and diuretic resistance in heart failure', *Circulation: Heart Failure* **9**(8), e003180.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.116.003180*

Harpe, S. E. (2009), 'Using secondary data sources for pharmacoepidemiology and outcomes research', *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* **29**(2), 138–153.
URL: *https://doi.org/10.1592/phco.29.2.138*

Harrell, F. E. et al. (2015), *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, Vol. 3, Springer.

Harris, T. B., Launer, L. J., Eiriksdottir, G., Kjartansson, O., Jonsson, P. V., Sigurdsson, G., Thorgeirsson, G., Aspelund, T., Garcia, M. E., Cotch, M. F., Hoffman, H. J., Gudnason, V. & for the Age, G. S. S. I. (2007), 'Age, Gene/Environment Susceptibility–Reykjavik Study: Multidisciplinary Applied Phenomics', *American Journal of Epidemiology* **165**(9), 1076–1087.
URL: *https://doi.org/10.1093/aje/kwk115*

Hasin, T., Gerber, Y., Weston, S. A., Jiang, R., Killian, J. M., Manemann, S. M., Cerhan, J. R. & Roger, V. L. (2016), 'Heart failure after myocardial infarction is associated with increased risk of cancer', *Journal of the American College of Cardiology* **68**(3), 265–271.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109716331308*

Hawkins, N. M., Petrie, M. C., Jhund, P. S., Chalmers, G. W., Dunn, F. G. & McMurray, J. J. (2009), 'Heart failure and chronic obstructive pulmonary disease: diagnostic pitfalls and epidemiology', *European Journal of Heart Failure* **11**(2), 130–139.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hfn013*

Hertz-Picciotto, I. & Rockhill, B. (1997), 'Validity and efficiency of approximation methods for tied survival times in cox regression', *Biometrics* **53**(3), 1151–1156.
URL: *http://www.jstor.org/stable/2533573*

Ho, C. Y., Mealiffe, M. E., Bach, R. G., Bhattacharya, M., Choudhury, L., Edelberg, J. M., Hegde, S. M., Jacoby, D., Lakdawala, N. K., Lester, S. J., Ma, Y., Marian, A. J., Nagueh, S. F., Owens, A., Rader, F., Saberi, S., Sehnert, A. J., Sherrid, M. V., Solomon, S. D., Wang, A., Wever-Pinzon, O., Wong, T. C. & Heitner, S. B. (2020), 'Evaluation of mavacamten in symptomatic patients with nonobstructive hypertrophic cardiomyopathy', *Journal of the American College of Cardiology* **75**(21), 2649–2660.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109720348567*

Ho, J. E., Gona, P., Pencina, M. J., Tu, J. V., Austin, P. C., Vasan, R. S., Kannel, W. B., D'Agostino, R. B., Lee, D. S. & Levy, D. (2012), 'Discriminating clinical features of heart failure with preserved vs. reduced ejection fraction in the community', *European Heart Journal* **33**(14), 1734–1741.
URL: *https://doi.org/10.1093/eurheartj/ehs070*

Hobbs, F. D., Roalfe, A. K., Davis, R. C., Davies, M. K., Hare, R. & Midlands Research Practices, C. (2007), 'Prognosis of all-cause heart failure and borderline left ventricular systolic dysfunction: 5 year mortality follow-up of the echocardiographic heart of england screening study (echoes)', *Eur Heart J* **28**(9), 1128–34.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/17459902*

Hodges, M. (1997), 'Rate correction of the QT interval', *Cardiac Electrophysiology Review* **1**(3), 360–363.

Hogg, J. (2020), Nt probnp testing in nhs scotland 2020, Report, Scottish Clinical Biochemistry Network.
URL: *https://www.clinicalbiochemistry.scot.nhs.uk/wp-content/uploads/2020/11/Report-NTproBNP-Query-2020.pdf*

Horikoshi, M. & Tang, Y. (2018), *ggfortify: Data Visualization Tools for Statistical Analysis Results*. R package version 0.4.13.
URL: *https://CRAN.R-project.org/package=ggfortify*

Hosmer, D. W. J., Lemeshow, S. & May, S. (2008), *Model Development*, John Wiley & Sons, Ltd, chapter 5, pp. 132–168.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470258019.ch5*

Hossain, M. E., Khan, A. & Uddin, S. (2019), Understanding the progression of congestive heart failure of type 2 diabetes patient using disease network and hospital claim data, *in* 'International conference on complex networks and their applications', Springer, pp. 774–788.

Hosseini, N., Kimiafar, K., Mostafavi, S. M., Kiani, B., Zendehdel, K., Zareiyan, A. & Eslami, S. (2021), 'Factors affecting the quality of diagnosis coding data with a triangulation view: A qualitative study', *The International Journal of Health Planning and Management* **36**(5), 1666–1684.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/hpm.3254*

Hougaard, P. (1999), 'Multi-state models: A review', *Lifetime Data Analysis* **5**(3), 239–264.
URL: *https://doi.org/10.1023/A:1009672031531*

Hsu, J. C., Friday, J., Lee, B. K., Azadani, P. N., Lee, R. J., Badhwar, N., Tseng, Z. H., Olgin, J. E. & Marcus, G. M. (2011), 'Predictors of axillary vein location for vascular access during

pacemaker and defibrillator lead implantation', *Pacing and clinical electrophysiology* **34**(12), 1585–1592.

Hu, Y. (2005), 'Efficient, high-quality force-directed graph drawing', *Mathematica journal* **10**(1), 37–71.
URL:
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.353.4547&rep=rep1&type=pdf

Huang, L., Midthune, D., Krapcho, M., Zou, Z., Horner, M.-J. & Feuer, E. J. (2013), 'Adjusting for reporting delay in cancer incidence when combining different sets of cancer registries', *Biom J.* **55**(5), 755–770.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201100191*

Hughes, R. A., Heron, J., Sterne, J. A. C. & Tilling, K. (2019), 'Accounting for missing data in statistical analyses: multiple imputation is not always the answer', *International Journal of Epidemiology* **48**(4), 1294–1304.
URL: *https://doi.org/10.1093/ije/dyz032*

Huusko, J., Purmonen, T., Toppila, I., Lassenius, M. & Ukkonen, H. (2020), 'Real-world clinical diagnostics of heart failure patients with reduced or preserved ejection fraction', *ESC Heart Failure* **7**(3), 1039–1048.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ehf2.12665*

Hyun, K. K., Redfern, J., Patel, A., Peiris, D., Brieger, D., Sullivan, D., Harris, M., Usherwood, T., MacMahon, S., Lyford, M. & Woodward, M. (2017), 'Gender inequalities in cardiovascular risk factor assessment and management in primary healthcare', *Heart* **103**(7), 492–498.
URL: *https://heart.bmj.com/content/103/7/492*

Iñiguez, G., Battiston, F. & Karsai, M. (2020), 'Bridging the gap between graphs and networks', *Communications Physics* **3**(1), 88.
URL: *https://doi.org/10.1038/s42005-020-0359-6*

Ibanez, B., James, S., Agewall, S., Antunes, M. J., Bucciarelli-Ducci, C., Bueno, H., Caforio, A. L. P., Crea, F., Goudevenos, J. A., Halvorsen, S., Hindricks, G., Kastrati, A., Lenzen, M. J., Prescott, E., Roffi, M., Valgimigli, M., Varenhorst, C., Vranckx, P., Widimský, P. & Group, E. S. D. (2017), '2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC)', *European Heart Journal* **39**(2), 119–177.
URL: *https://doi.org/10.1093/eurheartj/ehx393*

Information Services Division Scotland (2022*a*), 'Prescribing Information System (PIS)'.
URL: *https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=9*

Information Services Division Scotland (2022*b*), 'SMR01 - summary of rules'.
URL: *https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=S&ID=997&Title=SMR01%20-%20Summary%20of%20Rules*

Information Services Division Scotland (2022*c*), 'Specialty/discipline'.
URL: *https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=S&ID=997&Title=SMR01%20-%20Summary%20of%20Rules*

Iuliano, S., Fisher, S. G., Karasik, P. E., Fletcher, R. D. & Singh, S. N. (2002), 'QRS duration and mortality in patients with congestive heart failure', *American Heart Journal* **143**(6), 1085–1091.
URL: *https://www.sciencedirect.com/science/article/pii/S0002870302000261*

Iung, B., Delgado, V., Rosenhek, R., Price, S., Prendergast, B., Wendler, O., De Bonis, M., Tribouilloy, C., Evangelista, A., Bogachev-Prokophiev, A. et al. (2019), 'Contemporary presentation and management of valvular heart disease: the eurobservational research programme valvular heart disease ii survey', *Circulation* **140**(14), 1156–1169.

Jacobs, L., Efremov, L., Ferreira, J. P., Thijs, L., Yang, W. Y., Zhang, Z. Y., Latini, R., Masson, S., Agabiti, N., Sever, P., Delles, C., Sattar, N., Butler, J., Cleland, J. G. F., Kuznetsova, T., Staessen, J. A., Zannad, F. & Heart, O. i. A. i. (2017), 'Risk for incident heart failure: A subject-level meta-analysis from the heart "OMics" in AGEing (HOMAGE) Study', *J Am Heart Assoc* **6**(5).
URL: *https://www.ncbi.nlm.nih.gov/pubmed/28465299*

Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. (2014), 'Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software', *PLOS ONE* **9**(6), 1–12.
URL: *https://doi.org/10.1371/journal.pone.0098679*

Jager, K. J., van Dijk, P. C., Zoccali, C. & Dekker, F. W. (2008), 'The analysis of survival data: the kaplan–meier method', *Kidney International* **74**(5), 560–565.
URL: *https://www.sciencedirect.com/science/article/pii/S0085253815533681*

Jeong, E., Ko, K., Oh, S. & Han, H. W. (2017), 'Network-based analysis of diagnosis progression patterns using claims data', *Scientific Reports* **7**(1), 15561.
URL: *https://doi.org/10.1038/s41598-017-15647-4*

Jetté, N., Quan, H., Hemmelgarn, B., Drosler, S., Maass, C., Oec, D.-G., Moskal, L., Paoin, W.,

Sundararajan, V., Gao, S., Jakob, R., Üstün, B., Ghali, W. A. & the IMECCHI Investigators (2010), 'The development, evolution, and modifications of ICD-10: Challenges to the international comparability of morbidity data', *Medical Care* **48**(12), 1105–1110.
URL: *http://www.jstor.org/stable/25767019*

Joint Formulary Committee (2019), *British National Formulary*, 78 edn, BMJ Group and Pharmaceutical Press, London.

Jones, K. H., Ford, D. V., Thompson, S. & Lyons, R. (2019), 'A profile of the SAIL Databank on the UK secure research platform', *International Journal of Population Data Science* **4**(2).

Kalbfleich, J. D. & Prentice, R. L. (2002), *Competing Risks and Multistate Models*, John Wiley & Sons, Ltd, chapter 8, pp. 247–277.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118032985.ch8*

Kao, D. P., Lewsey, J. D., Anand, I. S., Massie, B. M., Zile, M. R., Carson, P. E., McKelvie, R. S., Komajda, M., McMurray, J. J. & Lindenfeld, J. (2015), 'Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response', *Eur J Heart Fail* **17**(9), 925–35.

Kapelios, C. J., Bonou, M., Malliaras, K., Athanasiadi, E., Vakrou, S., Skouloudi, M., Masoura, C. & Barbetseas, J. (2022), 'Association of loop diuretics use and dose with outcomes in outpatients with heart failure: a systematic review and metaanalysis of observational studies involving 96,959 patients', *Heart Failure Reviews* **27**(1), 147–161.
URL: *https://doi.org/10.1007/s10741-020-09995-z*

Kaplan, E. L. & Meier, P. (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* **53**(282), 457–481.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452*

Kassambara, A., Kosinski, M. & Biecek, P. (2021), *Survminer: Drawing survival curves using 'ggplot2'*. R package version 0.4.9.
URL: *https://CRAN.R-project.org/package=survminer*

Keiding, N., Anddersen, P. K. & Klein, J. P. (1997), 'The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates', *Statistics in Medicine* **16**(2), 215–224.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819970130%2916%3A2%3C215%3A%3AAID-SIM481%3E3.0.CO%3B2-J*

Kelder, J. C., Cramer, M. J., Rutten, F. H., Plokker, H., Grobbee, D. E. & Hoes, A. W. (2011), 'The furosemide diagnostic test in suspected slow-onset heart failure: popular but not useful', *European Journal of Heart Failure* **13**(5), 513–517.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hfr010*

Khan, Y. H., Sarriff, A., Adnan, A. S., Khan, A. H. & Mallhi, T. H. (2016), 'Chronic kidney disease, fluid overload and diuretics: A complicated triangle', *PLOS ONE* **11**(7), 1–13.
URL: *https://doi.org/10.1371/journal.pone.0159335*

Khand, A., Gemmel, I., Clark, A. L. & Cleland, J. G. (2000), 'Is the prognosis of heart failure improving?', *Journal of the American College of Cardiology* **36**(7), 2284–2286.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109700009955*

Khand, A. U., Shaw, M., Gemmel, I. & Cleland, J. G. (2005), 'Do discharge codes underestimate hospitalisation due to heart failure? Validation study of hospital discharge coding for heart failure', *Eur J Heart Fail* **7**(5), 792–7.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/16054867*

Khankirawatana, B., Khankirawatana, S. & Porter, T. (2004), 'How should left atrial size be reported? Comparative assessment with use of multiple echocardiographic methods', *American Heart Journal* **147**(2), 369–374.
URL: *https://www.sciencedirect.com/science/article/pii/S0002870303005982*

Khokhar, D. (2015), *Gephi cookbook*, Packt Publishing Ltd.

Kim, M., Chae, K.-H., Chung, Y.-J., Hwang, H., Lee, M., Kim, H.-K., Cho, H.-H., Kim, M.-R., Jung, C.-Y. & Kim, S. (2020), 'The effect of the look-back period for estimating incidence using administrative data', *BMC Health Services Research* **20**(1), 166.
URL: *https://doi.org/10.1186/s12913-020-5016-y*

King, L. R. (2006), *Locked Rooms: A Mary Russell Novel*, Vol. 8, Bantam.

Kizer, J. R., Bella, J. N., Palmieri, V., Liu, J. E., Best, L. G., Lee, E. T., Roman, M. J. & Devereux, R. B. (2006), 'Left atrial diameter as an independent predictor of first clinical cardiovascular events in middle-aged and elderly adults: The Strong Heart Study (SHS)', *American Heart Journal* **151**(2), 412–418.
URL: *https://www.sciencedirect.com/science/article/pii/S0002870305004631*

Klein, J. P. & Shu, Y. (2002), 'Multi-state models for bone marrow transplantation studies', *Statistical Methods in Medical Research* **11**(2), 117–139. PMID: 12040693.
URL: *https://doi.org/10.1191/0962280202sm277ra*

Kleinbaum, D. G., Klein, M. et al. (2012), *Survival analysis: a self-learning text*, Vol. 3, Springer.

Koudstaal, S., Pujades-Rodriguez, M., Denaxas, S., Gho, J., Shah, A. D., Yu, N., Patel, R. S., Gale, C. P., Hoes, A. W., Cleland, J. G., Asselbergs, F. W. & Hemingway, H. (2017), 'Prognostic burden of heart failure recorded in primary care, acute hospital admissions, or both: a population-based linked electronic health record cohort study in 2.1 million people',

Eur J Heart Fail **19**(9), 1119–1127.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/28008698*

Kuan, V., Denaxas, S., Gonzalez-Izquierdo, A., Direk, K., Bhatti, O., Husain, S., Sutaria, S., Hingorani, M., Nitsch, D., Parisinos, C. A., Lumbers, R. T., Mathur, R., Sofat, R., Casas, J. P., Wong, I. C. K., Hemingway, H. & Hingorani, A. D. (2019), 'A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service', *The Lancet Digital Health* **1**(2), e63–e77.
URL: *https://doi.org/10.1016/S2589-7500(19)30012-3*

Kwong, R. Y., Chan, A. K., Brown, K. A., Chan, C. W., Reynolds, H. G., Tsang, S. & Davis, R. B. (2006), 'Impact of unrecognized myocardial scar detected by cardiac magnetic resonance imaging on event-free survival in patients presenting with signs or symptoms of coronary artery disease', *Circulation* **113**(23), 2733–2743.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.105.570648*

Lagakos, S. W. (1979), 'General right censoring and its impact on the analysis of survival data', *Biometrics* **35**(1), 139–156.
URL: *http://www.jstor.org/stable/2529941*

Lam, C. S. P., Arnott, C., Beale, A. L., Chandramouli, C., Hilfiker-Kleiner, D., Kaye, D. M., Ky, B., Santema, B. T., Sliwa, K. & Voors, A. A. (2019), 'Sex differences in heart failure', *European Heart Journal* **40**(47), 3859–3868c.
URL: *https://doi.org/10.1093/eurheartj/ehz835*

Laney, D. (2001), '3d data management: Controlling data volume, velocity and variety', *META group research note* **6**(70), 1.

Lang, R. M., Badano, L. P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F. A., Foster, E., Goldstein, S. A., Kuznetsova, T., Lancellotti, P., Muraru, D., Picard, M. H., Rietzschel, E. R., Rudski, L., Spencer, K. T., Tsang, W. & Voigt, J.-U. (2015), 'Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging', *European Heart Journal - Cardiovascular Imaging* **16**(3), 233–271.
URL: *https://doi.org/10.1093/ehjci/jev014*

Lang, R. M., Bierig, M., Devereux, R. B., Flachskampf, F. A., Foster, E., Pellikka, P. A., Picard, M. H., Roman, M. J., Seward, J., Shanewise, J., Solomon, S., Spencer, K. T., St. John Sutton, M. & Stewart, W. (2006), 'Recommendations for chamber quantification', *European Journal of Echocardiography* **7**(2), 79–108.
URL: *https://doi.org/10.1016/j.euje.2005.12.014*

Lash, T. L., Fox, M. P., MacLehose, R. F., Maldonado, G., McCandless, L. C. & Greenland, S. (2014), 'Good practices for quantitative bias analysis', *International Journal of Epidemiology* **43**(6), 1969–1985.
URL: *https://doi.org/10.1093/ije/dyu149*

Leek, J. (2019), 'Research quality data and research quality databases'.
URL: *https://simplystatistics.org/posts/2019-05-29-research-quality-data-and-research-quality-databases/*

Lester, S. J., Ryan, E. W., Schiller, N. B. & Foster, E. (1999), 'Best method in clinical practice and in research studies to determine left atrial size', *The American Journal of Cardiology* **84**(7), 829–832.
URL: *https://www.sciencedirect.com/science/article/pii/S0002914999004464*

Levey, A. S., Coresh, J., Greene, T., Stevens, L. A., Zhang, Y., Hendriksen, S., Kusek, J. W., Van Lente, F. & Collaboration*, f. t. C. K. D. E. (2006), 'Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate', *Annals of Internal Medicine* **145**(4), 247–254.
URL: *https://doi.org/10.7326/0003-4819-145-4-200608150-00004*

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F. & Greene, T. J. A. o. i. m. (2009), 'A new equation to estimate glomerular filtration rate', *Annals of internal medicine* **150**(9), 604–612.

Leya, F. S., Arab, D., Joyal, D., Shioura, K. M., Lewis, B. E., Steen, L. H. & Cho, L. (2005), 'The efficacy of brain natriuretic peptide levels in differentiating constrictive pericarditis from restrictive cardiomyopathy', *Journal of the American College of Cardiology* **45**(11), 1900–1902.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109705007801*

Libby, P., Ridker, P. M. & Hansson, G. K. (2009), 'Inflammation in atherosclerosis: From pathophysiology to practice', *Journal of the American College of Cardiology* **54**(23), 2129–2138.
URL: *https://www.sciencedirect.com/science/article/pii/S073510970903006X*

Lim, L. S., Fink, H. A., Kuskowski, M. A., Taylor, B. C., Schousboe, J. T., Ensrud, K. E. & Group, O. F. i. M. S. (2008), 'Loop diuretic use and increased rates of hip bone loss in older men: The osteoporotic fractures in men study', *Archives of Internal Medicine* **168**(7), 735–740.
URL: *https://dx.doi.org/10.1001/archinte.168.7.735*

Livingstone, S. J., Looker, H. C., Hothersall, E. J., Wild, S. H., Lindsay, R. S., Chalmers, J., Cleland, S., Leese, G. P., McKnight, J., Morris, A. D., Pearson, D. W. M., Peden, N. R.,

Petrie, J. R., Philip, S., Sattar, N., Sullivan, F. & Colhoun, H. M. (2012), 'Risk of cardiovascular disease and total mortality in adults with type 1 diabetes: Scottish registry linkage study', *PLOS Medicine* **9**(10), 1–11.
URL: *https://doi.org/10.1371/journal.pmed.1001321*

Lloyd-Jones, D. M., Larson, M. G., Leip, E. P., Beiser, A., D'Agostino, R. B., Kannel, W. B., Murabito, J. M., Vasan, R. S., Benjamin, E. J. & Levy, D. (2002), 'Lifetime risk for developing congestive heart failure', *Circulation* **106**(24), 3068–3072.

Lohr, S. (2014), 'For big-data scientists,'janitor work'is key hurdle to insights'.
URL: *https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html*

Luo, S., Michler, K., Johnston, P. & Macfarlane, P. (2004), 'A comparison of commonly used QT correction formulae: The effect of heart rate on the QTc of normal ECGs', *Journal of Electrocardiology* **37**, 81–90.
URL: *https://www.sciencedirect.com/science/article/pii/S0022073604001001*

Macfarlane, P. (2021), '[favourite] QTc and bounds for believable refractory period?'. Email communication regarding which QTc correction to use and the maximum believable heart rate.

Madamanchi, C., Alhosaini, H., Sumida, A. & Runge, M. S. (2014), 'Obesity and natriuretic peptides, bnp and nt-probnp: Mechanisms and diagnostic implications for heart failure', *International Journal of Cardiology* **176**(3), 611–617.
URL: *https://www.sciencedirect.com/science/article/pii/S0167527314015137*

Mahmood, S. S., Levy, D., Vasan, R. S. & Wang, T. J. (2014), 'The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective', *The lancet* **383**(9921), 999–1008.

Malik, M. (2001), 'Problems of heart rate correction in assessment of drug-induced QT interval prolongation', *Journal of Cardiovascular Electrophysiology* **12**(4), 411–420.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1540-8167.2001.00411.x*

Mann, S. J. (2008), 'The silent epidemic of thiazide-induced hyponatremia', *The Journal of Clinical Hypertension* **10**(6), 477–484.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-7176.2008.08126.x*

Mantel, N. & Byar, D. P. (1974), 'Evaluation of response-time data involving transient states: An illustration using heart-transplant data', *Journal of the American Statistical Association* **69**(345), 81–86.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10480131*

Markov, A. A. (1961), 'Theory of algorithms', *Journal of Symbolic Logic* **27**(2), 444. English Translation by Jacques J. Schorr-kon, and Program for Scientific Translations staff. Published for the National Science Foundation, Washington, D.C., and the Department of Commerce by the Israel Program for Scientific Translations, Jerusalem.

Martens, P. & Mullens, W. (2018), 'How to tackle congestion in acute heart failure', *Korean J Intern Med* **33**(3), 462–473.

Martens, P., Nijst, P. & Mullens, W. (2015), 'Current approach to decongestive therapy in acute heart failure', *Current Heart Failure Reports* **12**(6), 367–378.
URL: *https://doi.org/10.1007/s11897-015-0273-5*

Masini, G., Graham, F. J., Pellicori, P., Cleland, J. G. F., Cuthbert, J. J., Kazmi, S., Inciardi, R. M. & Clark, A. L. (2022), 'Criteria for iron deficiency in patients with heart failure', *Journal of the American College of Cardiology* **79**(4), 341–351.
URL: *https://www.jacc.org/doi/abs/10.1016/j.jacc.2021.11.039*

McAllister, D. A., Read, S. H., Kerssens, J., Livingstone, S., McGurnaghan, S., Jhund, P., Petrie, J., Sattar, N., Fischbacher, C., Kristensen, S. L., McMurray, J., Colhoun, H. M. & Wild, S. H. (2018), 'Incidence of hospitalization for heart failure and case-fatality among 3.25 million people with and without diabetes mellitus', *Circulation* **138**(24), 2774–2786.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.118.034986*

McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak, M., Lam, C. S. P., Lyon, A. R., McMurray, J. J., Mebazaa, A., Mindham, R., Muneretto, C., Francesco Piepoli, M., Price, S., Rosano, G. M. C., Ruschitzka, F., Kathrine Skibelund, A. & Group, E. S. D. (2021), '2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed 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', *European Heart Journal* **42**(36), 3599–3726.
URL: *https://doi.org/10.1093/eurheartj/ehab368*

McKee, P. A., Castelli, W. P., McNamara, P. M. & Kannel, W. B. (1971), 'The natural history of congestive heart failure: The Framingham Study', **285**(26), 1441–1446.
URL: *https://www.nejm.org/doi/full/10.1056/NEJM197112232852601*

McMurray, J. J., Packer, M., Desai, A. S., Gong, J., Lefkowitz, M. P., Rizkala, A. R., Rouleau, J. L., Shi, V. C., Solomon, S. D., Swedberg, K. & Zile, M. R. (2014), 'Angiotensin–neprilysin inhibition versus enalapril in heart failure', *New England Journal of*

Medicine **371**(11), 993–1004. PMID: 25176015.
URL: *https://doi.org/10.1056/NEJMoa1409077*

McMurray, J. J., Solomon, S. D., Inzucchi, S. E., Køber, L., Kosiborod, M. N., Martinez, F. A., Ponikowski, P., Sabatine, M. S., Anand, I. S., Bělohlávek, J., Böhm, M., Chiang, C.-E., Chopra, V. K., de Boer, R. A., Desai, A. S., Diez, M., Drozdz, J., Dukát, A., Ge, J., Howlett, J. G., Katova, T., Kitakaze, M., Ljungman, C. E., Merkely, B., Nicolau, J. C., O'Meara, E., Petrie, M. C., Vinh, P. N., Schou, M., Tereshchenko, S., Verma, S., Held, C., DeMets, D. L., Docherty, K. F., Jhund, P. S., Bengtsson, O., Sjöstrand, M. & Langkilde, A.-M. (2019), 'Dapagliflozin in patients with heart failure and reduced ejection fraction', *New England Journal of Medicine* **381**(21), 1995–2008. PMID: 31535829.
URL: *https://doi.org/10.1056/NEJMoa1911303*

McMurray, J. J. & Stewart, S. (2000), 'Epidemiology, aetiology, and prognosis of heart failure', *Heart* **83**(5), 596–602.

Meijers, W. C. & de Boer, R. A. (2019), 'Common risk factors for heart failure and cancer', *Cardiovascular Research* **115**(5), 844–853.
URL: *https://doi.org/10.1093/cvr/cvz035*

Meira-Machado, L., de Uña-Alvarez, J., Cadarso-Suárez, C. & Andersen, P. K. (2009), 'Multi-state models for the analysis of time-to-event data', *Statistical methods in medical research* **18**(2), 195–222.
URL: *https://pubmed.ncbi.nlm.nih.gov/18562394*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2692556/

Mentz, R. J., Stevens, S. R., DeVore, A. D., Lala, A., Vader, J. M., AbouEzzeddine, O. F., Khazanie, P., Redfield, M. M., Stevenson, L. W., O'Connor, C. M., Goldsmith, S. R., Bart, B. A., Anstrom, K. J., Hernandez, A. F., Braunwald, E. & Felker, G. M. (2015), 'Decongestion strategies and renin-angiotensin-aldosterone system activation in acute heart failure', *JACC Heart Fail* **3**(2), 97–107.

Merlot, A. M., Kalinowski, D. S. & Richardson, D. R. (2014), 'Unraveling the mysteries of serum albumin—more than just a serum protein', *Frontiers in physiology* **5**, 299.

Messmer, B., Leachman, R., Nora, J. & Cooley, D. (1969), 'Survival-times after cardiac allografts', *The Lancet* **293**(7602), 954–956. Originally published as Volume 1, Issue 7602.
URL: *https://www.sciencedirect.com/science/article/pii/S0140673669918571*

Moss, A. J., Hall, W. J., Cannom, D. S., Klein, H., Brown, M. W., Daubert, J. P., Estes, N. M., Foster, E., Greenberg, H., Higgins, S. L., Pfeffer, M. A., Solomon, S. D., Wilber, D. & Zareba, W. (2009), 'Cardiac-resynchronization therapy for the prevention of heart-failure

events', *New England Journal of Medicine* **361**(14), 1329–1338. PMID: 19723701.
URL: *https://doi.org/10.1056/NEJMoa0906431*

Moss, A. J., Zareba, W., Hall, W. J., Klein, H., Wilber, D. J., Cannom, D. S., Daubert, J. P., Higgins, S. L., Brown, M. W. & Andrews, M. L. (2002), 'Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction', *New England Journal of Medicine* **346**(12), 877–883. PMID: 11907286.
URL: *https://doi.org/10.1056/NEJMoa013474*

Mosterd, A. & Hoes, A. W. (2007), 'Clinical epidemiology of heart failure', *Heart* **93**(9), 1137–46.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/17699180*

Mullens, W., Dauw, J., Martens, P., Verbrugge, F. H., Nijst, P., Meekers, E., Tartaglia, K., Chenot, F., Moubayed, S., Dierckx, R., Blouard, P., Troisfontaines, P., Derthoo, D., Smolders, W., Bruckers, L., Droogne, W., Ter Maaten, J. M., Damman, K., Lassus, J., Mebazaa, A., Filippatos, G., Ruschitzka, F. & Dupont, M. (2022), 'Acetazolamide in acute decompensated heart failure with volume overload', *New England Journal of Medicine* **387**(13), 1185–1195.
URL: *https://doi.org/10.1056/NEJMoa2203094*

Myhre, P. L., Vaduganathan, M. & Greene, S. J. (2020), 'Diagnosing heart failure with preserved ejection fraction in 2019: the search for a gold standard', *European Journal of Heart Failure* **22**(3), 422–424.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.1638*

National Institute for Health and Care Excellence (2018*a*), 'British National Formula - BNF diuretics treatment summary'.
URL: *https://bnf.nice.org.uk/treatment-summary/diuretics.html*

National Institute for Health and Care Excellence (2018*b*), 'Chronic heart failure in adults: diagnosis and management'.
URL: *https://www.nice.org.uk/guidance/ng106*

National Institute for Health and Care Excellence (2020), 'Hyponatraemia'.
URL: *https://cks.nice.org.uk/topics/hyponatraemia/*

National Institute for Health and Care Excellence (2021), 'Chronic kidney disease: assessment and management'.
URL: *https://www.nice.org.uk/guidance/ng203*

National Records of Scotland (2017), 'Vital events - deaths - background information'.
URL: *https://www.nrscotland.gov.uk/files/statistics/vital-events/coding-causes-of-death.pdf*

National Records of Scotland (2018), Mid-year population estimates by pre-april 2014 NHS Board areas by single year of age and sex: 1981-2013, Report, 'National Records of Scotland'.
URL: *https://www.nrscotland.gov.uk/files//statistics/time-series/population/hbe8113-pre-14-nhs-board-areas-revised.xlsx*

National Records of Scotland (2019), 'The medical certificate of the cause of death'.
URL: *https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths/deaths-background-information/death-certificates-and-coding-the-causes-of-death/the-medical-certificate-of-the-cause-of-death*

National Services Scotland Information Services Division (2019), 'Assessment of SMR01 data Scotland 2014-2015'.
URL: *https://www.isdscotland.org/Products-and-Services/Data-Quality/docs/Assessment-of-SMR01-Data-2014-15-report-181019.pdf*

Navaneethan, S. D., Schold, J. D., Arrigain, S., Jolly, S. E., Wehbe, E., Raina, R., Simon, J. F., Srinivas, T. R., Jain, A., Schreiber, M. J., J. & Nally, J. V., J. (2011), 'Serum bicarbonate and mortality in stage 3 and stage 4 chronic kidney disease', *Clin J Am Soc Nephrol* **6**(10), 2395–402.

Ng, K., Steinhubl, S. R., deFilippi, C., Dey, S. & Stewart, W. F. (2016), 'Early detection of heart failure using electronic health records: Practical implications for time before diagnosis, data diversity, data quantity, and data density', *Circ Cardiovasc Qual Outcomes* **9**(6), 649–658.

NHS Digital (2019), 'DCB0084: OPCS Classification of Interventions and Procedures'.
URL:
https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0084-opcs-classification-of-interventions-and-procedures

NHS Digital (2021), 'National clinical coding standards OPCS-4'.
URL: *https://classbrowser.nhs.uk/ref_books/OPCS-4.9_NCCS-2021.pdf*

NHS Digital (2022*a*), 'Community health index number'.
URL: *https://www.datadictionary.nhs.uk/attributes/community_health_index_number.html*

NHS Digital (2022*b*), 'ICD-10 code'.
URL: *https://www.datadictionary.nhs.uk/data_elements/icd-10_code.html*

NHS Greater Glasgow & Clyde Health Board (2021), 'Who we are'.
URL: *https://www.nhsggc.scot/about-us/who-we-are/*

NHS Inform (2022), 'Accessing and using the NHS in Scotland'.

URL: *https://www.nhsinform.scot/care-support-and-rights/health-rights/access/accessing-and-using-the-nhs-in-scotland*

NHS National Service Scotland (2019), 'SMR01 crib sheet'.
URL: *https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Crib-Sheets/SMR01_CRIB_220219.pdf*

NHS National Services Scotland (2015), 'SCI Store product overview'.
URL: *https://www.sci.scot.nhs.uk/products/store/General/SCI%20Store%20-%20Product%20Description.pdf*

NHS National Services Scotland (nd*a*), 'CHI number'.
URL: *https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=C&ID=128&Title=CHI%20Number*

NHS National Services Scotland (nd*b*), 'Code formats for ICD-10'.
URL: *https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/General-Clinical-Information/Diagnostic-Section/Code-Formats-for-ICD10.asp*

NHS National Services Scotland (nd*c*), 'General information on the ISCD & related health problems (ICD-10)'.
URL: *https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/General-Clinical-Information/Diagnostic-Section/General%20Information%20on%20the%20ISCD.asp*

NHS National Services Scotland (nd*d*), 'SMR00 - summary of rules'.
URL: *https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=S&ID=996&Title=SMR00%20-%20Summary%20of%20Rules*

NHS Research Scotland (2012*a*), 'Data Safe Haven'.
URL: *https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens*

NHS Research Scotland (2012*b*), 'Our population'.
URL: *https://www.nhsresearchscotland.org.uk/research-in-scotland/data/our-population*

NHS Research Scotland (2023), 'Data Safe Haven'.
URL: *https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens*

Nuzzi, V., Cannatà, A., Pellicori, P., Manca, P., Stolfo, D., Gregorio, C., Barbati, G., Bromage, D. I., McDonagh, T., Cleland, J. G. F., Merlo, M. & Sinagra, G. (2023), 'Diuretic dose trajectories in dilated cardiomyopathy: prognostic implications', *Clinical Research in Cardiology* **112**(3), 419–430.
URL: *https://doi.org/10.1007/s00392-022-02126-8*

Núñez, J., de la Espriella, R., Rossignol, P., Voors, A. A., Mullens, W., Metra, M., Chioncel,

O., Januzzi, J. L., Mueller, C., Richards, A. M., de Boer, R. A., Thum, T., Arfsten, H., González, A., Abdelhamid, M., Adamopoulos, S., Anker, S. D., Gal, T. B., Biegus, J., Cohen-Solal, A., Böhm, M., Emdin, M., Jankowska, E. A., Gustafsson, F., Hill, L., Jaarsma, T., Jhund, P. S., Lopatin, Y., Lund, L. H., Milicic, D., Moura, B., Piepoli, M. F., Ponikowski, P., Rakisheva, A., Ristic, A., Savarese, G., Tocchetti, C. G., Van Linthout, S., Volterrani, M., Seferovic, P., Rosano, G., Coats, A. J. & Bayes-Genis, A. (2022), 'Congestion in heart failure: a circulating biomarker-based perspective. a review from the biomarkers working group of the heart failure association, european society of cardiology', *European Journal of Heart Failure* **24**(10), 1751–1766.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.2664*

O'Dell, W. G. (2019), 'Accuracy of left ventricular cavity volume and ejection fraction for conventional estimation methods and 3d surface fitting', *Journal of the American Heart Association* **8**(6), e009124.
URL: *https://www.ahajournals.org/doi/abs/10.1161/JAHA.118.009124*

Okumura, N., Jhund, P. S., Gong, J., Lefkowitz, M. P., Rizkala, A. R., Rouleau, J. L., Shi, V. C., Swedberg, K., Zile, M. R., Solomon, S. D., Packer, M. & McMurray, J. J. (2016), 'Effects of sacubitril/valsartan in the PARADIGM-HF Trial (prospective comparison of ARNI with ACEI to determine impact on global mortality and morbidity in heart failure) according to background therapy', *Circ Heart Fail* **9**(9).
URL: *https://www.ncbi.nlm.nih.gov/pubmed/27618854*

Olivotto, I., Oreziak, A., Barriales-Villa, R., Abraham, T. P., Masri, A., Garcia-Pavia, P., Saberi, S., Lakdawala, N. K., Wheeler, M. T., Owens, A., Kubanek, M., Wojakowski, W., Jensen, M. K., Gimeno-Blanes, J., Afshar, K., Myers, J., Hegde, S. M., Solomon, S. D., Sehnert, A. J., Zhang, D., Li, W., Bhattacharya, M., Edelberg, J. M., Waldman, C. B., Lester, S. J., Wang, A., Ho, C. Y., Jacoby, D., Bartunek, J., Bondue, A., Van Craenenbroeck, E., Kubanek, M., Zemanek, D., Jensen, M., Mogensen, J., Thune, J. J., Charron, P., Hagege, A., Lairez, O., Trochu, J.-N., Axthelm, C., Duengen, H.-D., Frey, N., Mitrovic, V., Preusch, M., Schulz-Menger, J., Seidler, T., Arad, M., Halabi, M., Katz, A., Monakier, D., Paz, O., Viskin, S., Zwas, D., Olivotto, I., Brunner-La Rocca, H. P., Michels, M., Dudek, D., Oko-Sarnowska, Z., Oreziak, A., Wojakowski, W., Cardim, N., Pereira, H., Barriales-Villa, R., García Pavia, P., Gimeno Blanes, J., Hidalgo Urbano, R., Rincón Diaz, L. M., Elliott, P., Yousef, Z., Abraham, T., Afshar, K., Alvarez, P., Bach, R., Becker, R., Choudhury, L., Fermin, D., Jacoby, D., Jefferies, J., Kramer, C., Lakdawala, N., Lester, S., Marian, A., Masri, A., Maurer, M., Nagueh, S., Owens, A., Owens, D., Rader, F., Saberi, S., Sherrid, M., Shirani, J., Symanski, J., Turer, A., Wang, A., Wever-Pinzon, O., Wheeler, M., Wong, T. & Yamani, M. (2020), 'Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled,

phase 3 trial', *The Lancet* **396**(10253), 759–769.
URL: *https://doi.org/10.1016/S0140-6736(20)31792-X*

Oxford English Dictionary (2008*a*), *"big, adj. and adv."*, Oxford University Press.
URL: *https://www.oed.com/view/Entry/18833*

Oxford English Dictionary (2008*b*), *"information, n."*, Oxford University Press.
URL: *https://www.oed.com/view/Entry/95568?redirectedFrom=information*

Palevsky, P. M., Bhagrath, R. & Greenberg, A. (1996), 'Hypernatremia in hospitalized patients', *Annals of Internal Medicine* **124**(2), 197–203. PMID: 8533994.
URL: *https://www.acpjournals.org/doi/abs/10.7326/0003-4819-124-2-199601150-00002*

Pazos-López, P., Peteiro-Vázquez, J., Carcía-Campos, A., García-Bueno, L., de Torres, J. P. A. & Castro-Beiras, A. (2011), 'The causes, consequences, and treatment of left or right heart failure', *Vascular Health and Risk Management* **7**, 237–254. PMID: 21603593.
URL: *https://www.tandfonline.com/doi/abs/10.2147/VHRM.S10669*

Pellicori, P. & Cleland, J. G. (2015), 'Update on management of heart failure with preserved ejection fraction', *Current Opinion in Cardiology* **30**(2), 173–178.
URL: *https://journals.lww.com/co-cardiology/Fulltext/2015/03000/Update_on_-management_of_heart_failure_with.9.aspx*

Pellicori, P., Cleland, J. G. F. & Clark, A. L. (2020), 'Chronic obstructive pulmonary disease and heart failure: A breathless conspiracy', *Heart Failure Clinics* **16**(1), 33–44.
URL: *https://www.sciencedirect.com/science/article/pii/S155171361930087X*

Pellicori, P., Cleland, J. G., Zhang, J., Kallvikbacka-Bennett, A., Urbinati, A., Shah, P., Kazmi, S. & Clark, A. L. (2016), 'Cardiac dysfunction, congestion and loop diuretics: their relationship to prognosis in heart failure', *Cardiovasc Drugs Ther* **30**(6), 599–609.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/27819111*

Pellicori, P., Fitchett, D., Kosiborod, M. N., Ofstad, A. P., Seman, L., Zinman, B., Zwiener, I., Wanner, C., George, J., Inzucchi, S. E., Testani, J. M. & Cleland, J. G. (2021), 'Use of diuretics and outcomes in patients with type 2 diabetes: findings from the EMPA-REG OUTCOME trial', *European Journal of Heart Failure* **23**(7), 1085–1093.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.2220*

Pellicori, P., McConnachie, A., Carlin, C., Wales, A. & Cleland, J. G. F. (2022), 'Predicting mortality after hospitalisation for copd using electronic health records', *Pharmacological Research* **179**, 106199.
URL: *https://www.sciencedirect.com/science/article/pii/S104366182200144X*

Pellicori, P., Shah, P., Cuthbert, J., Urbinati, A., Zhang, J., Kallvikbacka-Bennett, A., Clark,

A. L. & Cleland, J. G. (2019), 'Prevalence, pattern and clinical relevance of ultrasound indices of congestion in outpatients with heart failure', *European Journal of Heart Failure* **21**(7), 904–916.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ejhf.1383*

Peters, S. A., Colantonio, L. D., Zhao, H., Bittner, V., Dai, Y., Farkouh, M. E., Monda, K. L., Safford, M. M., Muntner, P. & Woodward, M. (2018), 'Sex differences in high-intensity statin use following myocardial infarction in the united states', *Journal of the American College of Cardiology* **71**(16), 1729–1737.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109718333102*

Pitt, B., Zannad, F., Remme, W. J., Cody, R., Castaigne, A., Perez, A., Palensky, J. & Wittes, J. (1999), 'The effect of spironolactone on morbidity and mortality in patients with severe heart failure', *New England Journal of Medicine* **341**(10), 709–717. PMID: 10471456.
URL: *https://doi.org/10.1056/NEJM199909023411001*

Porta, M. (2016), *A Dictionary of Epidemiology*, Oxford University Press.
URL:
https://www.oxfordreference.com/view/10.1093/acref/9780199976720.001.0001/acref-9780199976720

Powers, F. (1999), 'The role of chloride in acid-base balance', *Journal of Infusion Nursing* **22**(5), 286.
URL: *https://journals.lww.com/journalofinfusionnursing/Fulltext/1999/09000/The_Role_-of_Chloride_in_Acid_Base_Balance.7.aspx*

Pringle, M. (1990), 'The new agenda for general practice computing', *BMJ (Clinical research ed.)* **301**(6756), 827–828.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1663967/*

Public Health Scotland (2019), Assessment of SMR01 (acute inpatient and day case) data Scotland 2019-2020, Report, Public Health Scotland,.
URL: *https://beta.isdscotland.org/media/7465/assessment-of-smr01-data-scotland-report-2019-v1.pdf*

Public Health Scotland (2020*a*), 'ISD datamarts'.
URL: *https://www.isdscotland.org/Products-and-Services/Datamarts/ISD-Datamarts/*

Public Health Scotland (2020*b*), 'SMR completeness'.
URL: *https://www.isdscotland.org/Products-and-Services/Data-Support-and-Monitoring/SMR-Completeness/*

Putter, H., Fiocco, M. & Geskus, R. B. (2007), 'Tutorial in biostatistics: competing risks and

multi-state models', *Statistics in Medicine* **26**(11), 2389–2430.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2712*

Putter, H., Geskus, R. B. & Fiocco, M. (n.d.), 'Tutorial in biostatistics: Competing risks and multi-state models', *Statistics in Medicine* (26), 2389–2430. R package version 0.3.1.

Putter, H., van der Hage, J., de Bock, G. H., Elgalta, R. & van de Velde, C. J. (2006), 'Estimation and prediction in a multi-state model for breast cancer', *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **48**(3), 366–380.

R Core Team (2020), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

R Core Team (2021), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Raine, T., Collins, G., Hall, C. & Hjelde, N. (2018), *Oxford handbook for the foundation programme*, Oxford University Press, USA.

Raphael, C. E., Whinnett, Z. I., Davies, J. E., Fontana, M., Ferenczi, E. A., Manisty, C. H., Mayet, J. & Francis, D. P. (2009), 'Quantifying the paradoxical effect of higher systolic blood pressure on mortality in chronic heart failure', *Heart* **95**(1), 56–62.
URL: *https://heart.bmj.com/content/95/1/56*

Rautaharju, P. M., Surawicz, B. & Gettes, L. S. (2009), 'AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: Part IV: The ST segment, T and U waves, and the QT interval a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology', *Journal of the American College of Cardiology* **53**(11), 982–991.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109708041363*

Rautaharju, P., Zhou, S., Wong, S., Calhoun, H., Berenson, G., Prineas, R. & Davignon, A. (1992), 'Sex differences in the evolution of the electrocardiographic QT interval with age', *The Canadian journal of cardiology* **8**(7), 690—695.
URL: *http://europepmc.org/abstract/MED/1422988*

Raybaut, P. (2009), 'Spyder-documentation', *Available online at: pythonhosted.org* .

Reddi, A. (2017), *Fluid, Electrolyte and Acid-Base Disorders: Clinical Evaluation and*

Management, Springer International Publishing.
URL: *https://books.google.co.uk/books?id=7vnlswEACAAJ*

Redpath, D. (2018), 'Methodologies of counting patient continuous inpatient stays (CIS) using SMR01 inpatient data'.
URL: *https://www.isdscotland.org/About-ISD/Methodologies/_docs/CIS_COUNTING_-paper_V0%202.pdf*

Reeves, D., Springate, D. A., Ashcroft, D. M., Ryan, R., Doran, T., Morris, R., Olier, I. & Kontopantelis, E. (2014), 'Can analyses of electronic patient records be independently and externally validated? the effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis', *BMJ Open* **4**(4).
URL: *https://bmjopen.bmj.com/content/4/4/e004952*

Remme, W. J. & Swedberg, K., T. (2001), 'Guidelines for the diagnosis and treatment of chronic heart failure', *European Heart Journal* **22**(17), 1527–1560.
URL: *https://doi.org/10.1053/euhj.2001.2783*

Requena-Ibanez, J. A., Santos-Gallego, C. G., Zafar, M. U. & Badimon, J. J. (2022), 'SGLT2-inhibitors on HFpEF patients. role of ejection fraction', *Cardiovascular Drugs and Therapy* .
URL: *https://doi.org/10.1007/s10557-022-07371-7*

Ripley, B. & Lapsley, M. (2021), *RODBC: ODBC Database Access*. R package version 1.3-19.
URL: *https://CRAN.R-project.org/package=RODBC*

Robertson, D., Hayes, A. & Couch, S. (2021), *broom: Convert statistical Objects to Tidy Tibbles*. R package version 0.7.10.
URL: *https://CRAN.R-project.org/package=broom*

Robinson, D., Schulz, E., Brown, P. & Price, C. (1997), 'Updating the Read Codes: user-interactive maintenance of a dynamic clinical vocabulary', *Journal of the American Medical Informatics Association: JAMIA* **4**(6), 465–472.
URL: *https://academic.oup.com/jamia/article/4/6/465/786188*

Roderburg, C., Loosen, S. H., Jahn, J. K., Gänsbacher, J., Luedde, T., Kostev, K. & Luedde, M. (2021), 'Heart failure is associated with an increased incidence of cancer diagnoses', *ESC Heart Failure* **8**(5), 3628–3633.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ehf2.13421*

Roger, V. L. (2013), 'Epidemiology of heart failure', *Circulation Research* **113**(6), 646–659.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.113.300268*

Roger, V. L. (2021), 'Epidemiology of heart failure', *Circulation Research*

128(10), 1421–1434.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.121.318172*

Romano, G., Vitale, G., Bellavia, D., Agnese, V. & Clemenza, F. (2017), 'Is diuretic withdrawal safe in patients with heart failure and reduced ejection fraction? a retrospective analysis of our outpatient cohort', *European Journal of Internal Medicine* **42**, e11–e13.
URL: *https://doi.org/10.1016/j.ejim.2017.03.025*

Rothman, K. J. et al. (2021), *Modern Epidemiology*, 4 edn, Wolters Kluwer.

Rusterborgh, H., Mayala, A., Forster, T., Vletter, W., van der Borden, B., Sutherland, G. R. & Roelandt, J. (1990), 'The reproducibility of continuous wave Doppler measurements in the assessment of mitral stenosis or mitral prosthetic function: The relative contributions of heart rate, respiration, observer variability and their clinical relevance', *European Heart Journal* **11**(7), 592–600.
URL: *https://doi.org/10.1093/oxfordjournals.eurheartj.a059764*

Sagie, A., Larson, M. G., Goldberg, R. J., Bengtson, J. R. & Levy, D. (1992), 'An improved method for adjusting the QT interval for heart rate (the Framingham Heart Study)', *Am J Cardiol* **70**(7), 797–801.

Sandhu, A. T., Tisdale, R. L., Rodriguez, F., Stafford, R. S., Maron, D. J., Hernandez-Boussard, T., Lewis, E. & Heidenreich, P. A. (2021), 'Disparity in the setting of incident heart failure diagnosis', *Circulation: Heart Failure* **14**(8), e008538.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.121.008538*

Savage, R. D., Visentin, J. D., Bronskill, S. E., Wang, X., Gruneir, A., Giannakeas, V., Guan, J., Lam, K., Luke, M. J., Read, S. H., Stall, N. M., Wu, W., Zhu, L., Rochon, P. A. & McCarthy, L. M. (2020), 'Evaluation of a common prescribing cascade of calcium channel blockers and diuretics in older adults with hypertension', *JAMA Internal Medicine* .
URL: *https://doi.org/10.1001/jamainternmed.2019.7087*

Savarese, G., Becher, P. M., Lund, L. H., Seferovic, P., Rosano, G. M. C. & Coats, A. J. S. (2022), 'Global burden of heart failure: a comprehensive and updated review of epidemiology', *Cardiovascular Research* **118**(17), 3272–3287.
URL: *https://doi.org/10.1093/cvr/cvac013*

Savarese, G., Stolfo, D., Sinagra, G. & Lund, L. H. (2022), 'Heart failure with mid-range or mildly reduced ejection fraction', *Nature Reviews Cardiology* **19**(2), 100–116.
URL: *https://doi.org/10.1038/s41569-021-00605-5*

Schartum-Hansen, H., Løland, K. H., Svingen, G. F. T., Seifert, R., Pedersen, E. R., Nordrehaug, J. E., Bleie, Ø., Ebbing, M., Berge, C., Nilsen, D. W. T. & Nygård, O. (2015), 'Use of loop diuretics is associated with increased mortality in patients with suspected

coronary artery disease, but without systolic heart failure or renal impairment: An observational study using propensity score matching', *PloS one* **10**(6), e0124611–e0124611.
URL: *https://pubmed.ncbi.nlm.nih.gov/26030195*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4452510/

Schelbert, E. B., Cao, J. J., Sigurdsson, S., Aspelund, T., Kellman, P., Aletras, A. H., Dyke, C. K., Thorgeirsson, G., Eiriksdottir, G., Launer, L. J., Gudnason, V., Harris, T. B. & Arai, A. E. (2012), 'Prevalence and Prognosis of Unrecognized Myocardial Infarction Determined by Cardiac Magnetic Resonance in Older Adults', *JAMA* **308**(9), 890–896.
URL: *https://doi.org/10.1001/2012.jama.11089*

Schenk, A. & Fields, N. (2023), 'Mavacamten – a targeted therapy for hypertrophic cardiomyopathy', *Journal of Cardiovascular Pharmacology* **81**(5), 317–326.
URL: *https://journals.lww.com/cardiovascularpharm/Fulltext/9900/Mavacamten___a_-targeted_therapy_for_hypertrophic.159.aspx*

Shah, S. J., Katz, D. H. & Deo, R. C. (2014), 'Phenotypic spectrum of heart failure with preserved ejection fraction', *Heart failure clinics* **10**(3), 407–418.
URL: *https://pubmed.ncbi.nlm.nih.gov/24975905*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4076705/

Shah, S. N., Abramowitz, M., Hostetter, T. H. & Melamed, M. L. (2009), 'Serum bicarbonate levels and the progression of kidney disease: A cohort study', *American Journal of Kidney Diseases* **54**(2), 270–277.
URL: *https://www.sciencedirect.com/science/article/pii/S0272638609005113*

Shakir, D. K. & Rasul, K. I. (2009), 'Chemotherapy induced cardiomyopathy: pathogenesis, monitoring and management', *J Clin Med Res* **1**(1), 8–12.

Shen, L., Jhund, P. S., Anand, I. S., Bhatt, A. S., Desai, A. S., Maggioni, A. P., Martinez, F. A., Pfeffer, M. A., Rizkala, A. R., Rouleau, J. L., Swedberg, K., Vaduganathan, M., Vardeny, O., van Veldhuisen, D. J., Zannad, F., Zile, M. R., Packer, M., Solomon, S. D. & McMurray, J. J. (2021), 'Incidence and outcomes of pneumonia in patients with heart failure', *Journal of the American College of Cardiology* **77**(16), 1961–1973.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109721005775*

Siggaard, T., Reguant, R., Jørgensen, I. F., Haue, A. D., Lademann, M., Aguayo-Orozco, A., Hjaltelin, J. X., Jensen, A. B., Banasik, K. & Brunak, S. (2020), 'Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million danish patients', *Nature Communications* **11**(1), 4952.
URL: *https://doi.org/10.1038/s41467-020-18682-4*

Sigurdsson, E., Thorgeirsson, G., Sigvaldason, H. & Sigfusson, N. (1995), 'Unrecognized

myocardial infarction: Epidemiology, clinical characteristics, and the prognostic role of angina pectoris: The Reykjavik study', *Annals of Internal Medicine* **122**(2), 96–102. PMID: 7993002.
URL: *https://www.acpjournals.org/doi/abs/10.7326/0003-4819-122-2-199501150-00003*

Singer, M. & Webb, A. (2009), *Oxford Handbook of Critical Care*, Oxford University Press.
URL: *https://doi.org/10.1093/med/9780199235339.001.0001*

Solomon, S. D., McMurray, J. J., Anand, I. S., Ge, J., Lam, C. S., Maggioni, A. P., Martinez, F., Packer, M., Pfeffer, M. A., Pieske, B., Redfield, M. M., Rouleau, J. L., van Veldhuisen, D. J., Zannad, F., Zile, M. R., Desai, A. S., Claggett, B., Jhund, P. S., Boytsov, S. A., Comin-Colet, J., Cleland, J., Düngen, H.-D., Goncalvesova, E., Katova, T., Kerr Saraiva, J. F., Lelonek, M., Merkely, B., Senni, M., Shah, S. J., Zhou, J., Rizkala, A. R., Gong, J., Shi, V. C. & Lefkowitz, M. P. (2019), 'Angiotensin–neprilysin inhibition in heart failure with preserved ejection fraction', *New England Journal of Medicine* **381**(17), 1609–1620. PMID: 31475794.
URL: *https://doi.org/10.1056/NEJMoa1908655*

Sonnenberg, F. A. & Beck, J. R. (1993), 'Markov models in medical decision making: A practical guide', *Medical Decision Making* **13**(4), 322–338. PMID: 8246705.
URL: *https://doi.org/10.1177/0272989X9301300409*

Sparrow, N., Adlam, D., Cowley, A. & Hampton, J. R. (2003*a*), 'The diagnosis of heart failure in general practice: implications for the UK National Service Framework', *European Journal of Heart Failure* **5**(3), 349–354.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1016/S1388-9842%2803%2900046-1*

Sparrow, N., Adlam, D., Cowley, A. & Hampton, J. R. (2003*b*), 'Difficulties of introducing the national service framework for heart failure into general practice in the UK', *European Journal of Heart Failure* **5**(3), 355–361.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1016/S1388-9842%2803%2900047-3*

Spertus, J. A., Fine, J. T., Elliott, P., Ho, C. Y., Olivotto, I., Saberi, S., Li, W., Dolan, C., Reaney, M., Sehnert, A. J. & Jacoby, D. (2021), 'Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): health status analysis of a randomised, double-blind, placebo-controlled, phase 3 trial', *The Lancet* **397**(10293), 2467–2475.
URL: *https://doi.org/10.1016/S0140-6736(21)00763-7*

Stavrou, E., Pesa, N. & Pearson, S.-A. (2012), 'Hospital discharge diagnostic and procedure codes for upper gastro-intestinal cancer: how accurate are they?', *BMC Health Services Research* **12**(1), 331.
URL: *https://doi.org/10.1186/1472-6963-12-331*

Stevens, L. A., Schmid, C. H., Greene, T., Zhang, Y. L., Beck, G. J., Froissart, M., Hamm, L. L., Lewis, J. B., Mauer, M., Navis, G. J., Steffes, M. W., Eggers, P. W., Coresh, J. & Levey, A. S. (2010), 'Comparative performance of the CKD Epidemiology Collaboration (CKD-EPI) and the Modification of Diet in Renal Disease (MDRD) Study Equations for estimating GFR levels above 60 ml/min/1.73 m^2', *American Journal of Kidney Diseases* **56**(3), 486–495.
URL: *https://www.sciencedirect.com/science/article/pii/S0272638610007985*

Storme, O., Tirán Saucedo, J., Garcia-Mora, A., Dehesa-Dávila, M. & Naber, K. G. (2019), 'Risk factors and predisposing conditions for urinary tract infection', *Ther Adv Urol* **11**, 1756287218814382.

Strassheim, D., Dempsey, E. C., Gerasimovskaya, E., Stenmark, K. & Karoor, V. (2019), 'Role of inflammatory cell subtypes in heart failure', *Journal of immunology research* **2019**, 2164017–2164017.
URL: *https://pubmed.ncbi.nlm.nih.gov/31565659*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6745095/

Suetrong, B., Pisitsak, C., Boyd, J. H., Russell, J. A. & Walley, K. R. (2016), 'Hyperchloremia and moderate increase in serum chloride are associated with acute kidney injury in severe sepsis and septic shock patients', *Critical Care* **20**(1), 315.
URL: *https://doi.org/10.1186/s13054-016-1499-7*

Suissa, S. (2007), 'Immortal time bias in pharmacoepidemiology', *American Journal of Epidemiology* **167**(4), 492–499.
URL: *https://doi.org/10.1093/aje/kwm324*

Sulo, G., Igland, J., Vollset, S. E., Nygård, O., Egeland, G. M., Ebbing, M., Sulo, E. & Tell, G. S. (2015), 'Effect of the lookback period's length used to identify incident acute myocardial infarction on the observed trends on incidence rates and survival', *Circulation: Cardiovascular Quality and Outcomes* **8**(4), 376–382.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCOUTCOMES.114.001703*

Sutradhar, R., Barbera, L., Seow, H., Howell, D., Husain, A. & Dudgeon, D. (2010), 'Multistate analysis of interval-censored longitudinal data: Application to a cohort study on performance status among patients diagnosed with cancer', *American Journal of Epidemiology* **173**(4), 468–475.
URL: *https://doi.org/10.1093/aje/kwq384*

Teetor, P. (2011), *R cookbook: Proven recipes for data analysis, statistics, and graphics*, " O'Reilly Media, Inc.".

Testani, J. M., Brisco, M. A., Turner, J. M., Spatz, E. S., Bellumkonda, L., Parikh, C. R. &

Tang, W. W. (2014), 'Loop diuretic efficiency', *Circulation: Heart Failure* **7**(2), 261–270.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.113.000895*

The Scottish Government (2012), 'The Scottish Index of Multiple Deprivation 2012'.
URL: *https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2012/12/scottish-index-multiple-deprivation-2012-executive-summary/documents/scottish-index-multiple-deprivation-2012-executive-summary/scottish-index-multiple-deprivation-2012-executive-summary/govscot%3Adocument/00410895.pdf*

The Scottish Government (2015), 'Charter for Safe Havens in Scotland: Handling unconsented data from National Health Service patient records to support research and statistics.'.
URL:
https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/documents/

The Task Force on Heart Failure of the European Society of Cardiology (1995), 'Guidelines for the diagnosis of heart failure', *European Heart Journal* **16**(6), 741–751.
URL: *https://doi.org/10.1093/oxfordjournals.eurheartj.a060992*

Therneau, T. M. & Grambsch, P. M. (2000), *The Cox Model*, Springer New York, New York, NY, pp. 39–77.
URL: *https://doi.org/10.1007/978-1-4757-3294-8_3*

Therneau, T. M. & Gramsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, Springer, New York, New York. Version 3.3.-1.
URL: *https://CRAN.R-project.org/package=survival*

Titman, A. C. & Putter, H. (2020), 'General tests of the Markov property in multi-state models', *Biostatistics* .
URL: *https://doi.org/10.1093/biostatistics/kxaa030*

Torabi, A., Cleland, J. G., Khan, N. K., Loh, P. H., Clark, A. L., Alamgir, F., Caplin, J. L., Rigby, A. S. & Goode, K. (2008), 'The timing of development and subsequent clinical course of heart failure after a myocardial infarction', *European Heart Journal* **29**(7), 859–870.
URL: *https://doi.org/10.1093/eurheartj/ehn096*

Torabi, A., Rigby, A. S. & Cleland, J. G. F. (2009), 'Declining in-hospital mortality and increasing heart failure incidence in elderly patients with first myocardial infarction', *Journal of the American College of Cardiology* **55**(1), 79–81.
URL: *https://www.sciencedirect.com/science/article/pii/S0735109709033233*

Trullàs, J. C., Morales-Rull, J. L., Casado, J., Carrera-Izquierdo, M., Sánchez-Marteles, M., Conde-Martel, A., Dávila-Ramos, M. F., Llácer, P., Salamanca-Bautista, P., Pérez-Silvestre, J., Plasín, M. Á., Cerqueiro, J. M., Gil, P., Formiga, F., Manzano, L. & trial investigators, C.

(2022), 'Combining loop with thiazide diuretics for decompensated heart failure: the CLOROTIC trial', *European Heart Journal* **44**(5), 411–421.
URL: *https://doi.org/10.1093/eurheartj/ehac689*

van Kraaij, D. J., Jansen, R. W., de Gier, J. J., Gribnau, F. W. & Hoefnagels, W. H. (1998), 'Prescription patterns of diuretics in Dutch community-dwelling elderly patients', *Br J Clin Pharmacol* **46**(4), 403–7.

van Rosmalen, J., Toy, M. & O'Mahony, J. F. (2013), 'A mathematical approach for evaluating Markov models in continuous time without discrete-event simulation', *Medical Decision Making* **33**(6), 767–779. PMID: 23715464.
URL: *https://doi.org/10.1177/0272989X13487947*

Vandenberk, B., Vandael, E., Robyns, T., Vandenberghe, J., Garweg, C., Foulon, V., Ector, J. & Willems, R. (2016), 'Which QT correction formulae to use for QT monitoring?', *Journal of the American Heart Association* **5**(6), e003264.
URL: *https://www.ahajournals.org/doi/abs/10.1161/JAHA.116.003264*

Verdú-Rotellar, J. M., Frigola-Capell, E., Alvarez-Pérez, R., da Silva, D., Enjuanes, C., Domingo, M., Mena, A. & Muñoz, M.-A. (2017), 'Validation of heart failure diagnosis registered in primary care records in two primary care centres in Barcelona (Spain) and factors related. a cross-sectional study', *European Journal of General Practice* **23**(1), 107–113. PMID: 28376668.
URL: *https://doi.org/10.1080/13814788.2017.1305104*

Voors, A. A., Angermann, C. E., Teerlink, J. R., Collins, S. P., Kosiborod, M., Biegus, J., Ferreira, J. P., Nassif, M. E., Psotka, M. A., Tromp, J., Borleffs, C. J. W., Ma, C., Comin-Colet, J., Fu, M., Janssens, S. P., Kiss, R. G., Mentz, R. J., Sakata, Y., Schirmer, H., Schou, M., Schulze, P. C., Spinarova, L., Volterrani, M., Wranicz, J. K., Zeymer, U., Zieroth, S., Brueckmann, M., Blatchford, J. P., Salsali, A. & Ponikowski, P. (2022), 'The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial', *Nature Medicine* **28**(3), 568–574.
URL: *https://doi.org/10.1038/s41591-021-01659-1*

Vozoris, N. T., Wang, X., Austin, P. C., O'Donnell, D. E., Aaron, S. D., To, T. M. & Gershon, A. S. (2018), 'Incident diuretic drug use and adverse respiratory events among older adults with chronic obstructive pulmonary disease', *British Journal of Clinical Pharmacology* **84**(3), 579–589.
URL: *https://bpspubs.onlinelibrary.wiley.com/doi/abs/10.1111/bcp.13465*

Walker, H. K., Hall, W. D. & Hurst, J. W. (1990), *Clinical methods: the history, physical, and laboratory examinations*, Butterworths.

Walls, J. (2009), *Half Broke Horses: A True-life Novel*, Simon and Schuster.

Walma, E., van Dooren, C., Prins, A., van der Does, E. & Hoes, A. (1997), 'Withdrawal of long term diuretic medication in elderly patients: a double blind randomised trial', *BMJ* **315**(7106), 464–468.
URL: *https://www.bmj.com/content/315/7106/464*

Walsh, D., McCartney, G., Collins, C., Taulbut, M. & Batty, G. D. (2017), 'History, politics and vulnerability: explaining excess mortality in Scotland and Glasgow', *Public health* **151**, 1–12.

Watson, S. (2020), 'Safe Haven'.
URL: *https://www.quest.scot.nhs.uk/hc/en-gb/articles/115005312709-Safe-Haven*

Watts, M. & Graham, K. (2022), 'Questions regarding source and number of ECGs and Echos'. Email communication regarding coverage and completeness of MUSE ECGs and Xcelera and EchoPAC echocardiograms.

Wehner, G. J., Jing, L., Haggerty, C. M., Suever, J. D., Leader, J. B., Hartzel, D. N., Kirchner, H. L., Manus, J. N. A., James, N., Ayar, Z., Gladding, P., Good, C. W., Cleland, J. G. F. & Fornwalt, B. K. (2019), 'Routinely reported ejection fraction and mortality in clinical practice: where does the nadir of risk lie?', *European Heart Journal* **41**(12), 1249–1257.
URL: *https://doi.org/10.1093/eurheartj/ehz550*

Welsh, P., Campbell, R. T., Mooney, L., Kimenai, D. M., Hayward, C., Campbell, A., Porteous, D., Mills, N. L., Lang, N. N., Petrie, M. C., Januzzi, J. L., McMurray, J. J. & Sattar, N. (2022), 'Reference ranges for NT-proBNP (N-terminal pro-B-type natriuretic peptide) and risk factors for higher NT-proBNP concentrations in a large general population cohort', *Circulation: Heart Failure* **15**(10), e009427.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.121.009427*

West, M. (2011), Some general principles for conceptual, integration, and enterprise data models, *in* 'Developing High Quality Data Models', Morgan Kaufmann, Boston, chapter 6, pp. 63–78.
URL: *https://www.sciencedirect.com/science/article/pii/B9780123751065000063*

Wickham, H. (2007), 'Reshaping data with the reshape package', *Journal of Statistical Software* **21**(12), 1–20. R package version 1.4.4.
URL: *http://www.jstatsoft.org/v21/i12/paper*

Wickham, H. (2021), *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 0.5.1.
URL: *https://CRAN.R-project.org/package=forcats*

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Muller, K., Ooms, J., Robinson, D., Siedel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019), 'Welcome to the Tidyverse', *Journal of open source software* **4**(43), 1686. R package version 1.3.1.

Wilchesky, M., Tamblyn, R. M. & Huang, A. (2004), 'Validation of diagnostic codes within medical services claims', *Journal of Clinical Epidemiology* **57**(2), 131–141.
URL: *https://www.sciencedirect.com/science/article/pii/S0895435603002464*

Wile, D. (2012), 'Diuretics: a review', *Ann Clin Biochem* **49**(Pt 5), 419–31.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/22783025*

Wiley, C. L., Switzer, S. P., Berg, R. L., Glurich, I. & Dart, R. A. (2010), 'Association of b-type natriuretic peptide levels with estimated glomerular filtration rate and congestive heart failure', *Clinical medicine & research* **8**(1), 7–12.

Williams, B., Mancia, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., Clement, D. L., Coca, A., de Simone, G., Dominiczak, A., Kahan, T., Mahfoud, F., Redon, J., Ruilope, L., Zanchetti, A., Kerins, M., Kjeldsen, S. E., Kreutz, R., Laurent, S., Lip, G. Y. H., McManus, R., Narkiewicz, K., Ruschitzka, F., Schmieder, R. E., Shlyakhto, E., Tsioufis, C., Aboyans, V., Desormais, I. & ESC Scientific Document Group (2018), '2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH)', *European Heart Journal* **39**(33), 3021–3104.
URL: *https://doi.org/10.1093/eurheartj/ehy339*

Williams, C., Lewsey, J. D., Briggs, A. H. & Mackay, D. F. (2017), 'Cost-effectiveness analysis in R using a multi-state modeling survival analysis framework: A tutorial', *Medical Decision Making* **37**(4), 340–352. PMID: 27281337.
URL: *https://doi.org/10.1177/0272989X16651869*

Williams, C., Lewsey, J., Mackay, D. & Briggs, A. (2016), 'Estimation of survival probabilities for use in cost-effectiveness analyses', *Medical Decision Making* **37**.

Williams, R. (2003), *Grids and the Virtual Observatory*, John Wiley & Sons, Ltd, chapter 38, pp. 837–858.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/0470867167.ch38*

Williams, R., Brown, B., Peek, N. & Buchan, I. (2016), 'Making medication data meaningful: illustrated with hypertension', *Studies in health technology and informatics* **228**, 247–251.

Woodward, M. (2019), 'Cardiovascular disease and the female disadvantage', *International*

Journal of Environmental Research and Public Health **16**(7).
URL: *https://www.mdpi.com/1660-4601/16/7/1165*

World Health Organization (2011), 'Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity'.
URL: *https://apps.who.int/iris/handle/10665/85839*

World Health Organization (2016), *International statistical classification of diseases and related health problems - 10th revision*, Vol. 3, 5th edn, World Health Organization, France.
URL: *https://apps.who.int/iris/bitstream/10665/246208/1/9789241549165-V1-eng.pdf*

World Health Organization (2022*a*), 'Cause of death'.
URL:
https://www.who.int/standards/classifications/classification-of-diseases/cause-of-death

World Health Organization (2022*b*), 'Classification of disease'.
URL: *https://www.who.int/standards/classifications/classification-of-diseases*

Wyatt, J. P. (2006), *Oxford Handbook of Emergency Medicine*, Oxford University Press.

Wynant, W. & Abrahamowicz, M. (2014), 'Impact of the model-building strategy on inference about nonlinear and time-dependent covariate effects in survival analysis', *Statistics in Medicine* **33**(19), 3318–3337.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6178*

Xiang, B., Zhang, R., Wu, X. & Zhou, X. (2022), 'Optimal Pharmacologic Treatment of Heart Failure With Preserved and Mildly Reduced Ejection Fraction: A Meta-analysis', *JAMA Network Open* **5**(9), e2231963–e2231963.
URL: *https://doi.org/10.1001/jamanetworkopen.2022.31963*

Yadav, K. & Lewis, R. J. (2021), 'Immortal Time Bias in Observational Studies', *JAMA* **325**(7), 686–687.
URL: *https://doi.org/10.1001/jama.2020.9151*

Yang, Y., Li, W., Zhu, H., Pan, X.-F., Hu, Y., Arnott, C., Mai, W., Cai, X. & Huang, Y. (2020), 'Prognosis of unrecognised myocardial infarction determined by electrocardiography or cardiac magnetic resonance imaging: systematic review and meta-analysis', *BMJ* **369**.
URL: *https://www.bmj.com/content/369/bmj.m1184*

Yeboah, J., Rodriguez, C. J., Stacey, B., Lima, J. A., Liu, S., Carr, J. J., Hundley, W. G. & Herrington, D. M. (2012), 'Prognosis of individuals with asymptomatic left ventricular systolic dysfunction in the Multi-Ethnic Study of Atherosclerosis (MESA)', *Circulation* **126**(23), 2713–2719.
URL: *https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.112.112201*

Zakeri, R., Morgan, A. D., Sundaram, V., Bloom, C., Cleland, J. G. F. & Quint, J. K. (2021), 'Under-recognition of heart failure in patients with atrial fibrillation and the impact of gender: a UK population-based cohort study', *BMC Medicine* **19**(1), 179.
URL: *https://doi.org/10.1186/s12916-021-02048-8*

Zhang, Z., Reinikainen, J., Adeleke, K. A., Pieterse, M. E. & Groothuis-Oudshoorn, C. G. M. (2018), 'Time-varying covariates and coefficients in cox regression models', *Annals of translational medicine* **6**(7), 121–121.
URL: *https://pubmed.ncbi.nlm.nih.gov/29955581*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6015946/