THE UNIVERSITY
of EDINBURGH

Improving diagnosis in acute cardiac care using statistical machine learning

THE UNIVERSITY *of* EDINBURGH

Dimitrios Doudesis

DEGREE OF DOCTOR OF PHILOSOPHY

The University of Edinburgh

2022

To Iliana

Declaration

This thesis represents research undertaken in the Centre for Cardiovascular Sciences (University of Edinburgh) and the Usher Institute (University of Edinburgh) between September 2018 and February 2022. I declare that I have composed this thesis and that the work included is either my own or represents collaborative work in which I have made a substantial contribution.

This thesis reports findings from patients participating in the High-STEACS clinical trial, designed and led by my supervisor, Professor Nicholas L. Mills. Participants contributing to the individual patient-level data-meta analysis were recruited from thirteen different countries by Dr Kuan Ken Lee, Dr Mohamed Anwar and Professor Nicholas L. Mills, all of whom obtained the necessary approvals to share patient-level data.

Chapters 4 has been accepted for publication in peer-reviewed journals. The thesis has not been accepted in any previous applications for a degree and all sources of information have been acknowledged.

This work was supported by a Precision Medicine Doctoral Training Programme Scholarship from Medical Research Council (MR/N013166/1).

Dimitrios Doudesis 28th February 2022

Abstract

Cardiovascular disease affects more than half of people in the United Kingdom and remains the most common cause of death. Each year more than 25 million persons attend an Emergency Department, with chest pain or breathlessness being the most common presentations. These patients are often admitted to hospital because of concerns that they may have a life-threatening condition, such as acute myocardial infarction or decompensated heart failure. Despite the availability of specific and sensitive cardiac biomarkers, the diagnosis is not straightforward, resulting in unnecessary hospital admission or misdiagnosis. The aim of this thesis is to use cardiac biomarkers and statistical machine learning to develop clinical decision support tools that improve the diagnosis of patients presenting to the Emergency Department with possible acute cardiac conditions.

In 20,761 consecutive patients from the High Sensitivity Troponin in the Evaluation of Acute Coronary Syndrome (High-STEACS) trial, we validated a previously developed machine learning algorithm to assess its diagnostic performance for myocardial infarction in routine clinical practice. The myocardial-ischemic-injury-index (MI^3) algorithm, which incorporates age, sex, and two troponin measurements of a patient, had excellent discrimination for the index diagnosis of myocardial infarction, and moreover, it predicted subsequent events too. However, the analysis showed that MI^3 performance was not well calibrated in patients with intermediate probability, and there was considerable heterogeneity across important subgroups such

as age, sex, presenting symptom of chest pain, cerebrovascular disease and renal function.

It is well known that cardiac troponin concentrations are influenced not only by the age and sex of the patient, but also by the time since symptom onset and comorbidities. Hence, we used patients from the High-STEACS trial and three external cohorts to develop and validate CoDE-ACS (Collaboration for the Diagnosis and Evaluation of Acute Coronary Syndrome); a decision support tool that uses machine learning to incorporate clinical variables along with serial troponin measures and other laboratory tests. CoDE-ACS accurately predicts the likelihood of myocardial infarction and provides a more individualised diagnostic assessment. When CoDE-ACS was compared with the guideline-recommended clinical pathways, the performance was more consistent across important patient subgroups with better negative and positive predicted value.

We have subsequently developed and validated a decision support tool for patients with suspected acute heart failure; a condition where symptoms mimic many other conditions making the diagnosis challenging. To address this, we developed and externally validated CoDE-HF (Collaboration for the Diagnosis and Evaluation of Heart Failure) in 10,369 patients from 13 countries to improve the diagnosis and evaluation of acute heart failure. CoDE-HF combines blood natriuretic peptide concentrations as a continuous measure and simple objective clinical variables known to be associated with

acute heart failure. First, we used the N-terminal pro-B-type natriuretic peptide (NT-proBNP), as it is the most common test used in clinical practice. Then, we retrained CoDE-HF to support the use of two other natriuretic peptides (B-type natriuretic peptide [BNP] and mid-regional pro atrial natriuretic peptide [MR-proANP]). Last, we compared my solution with the guideline-recommended approach showing that my new decision support tool could achieve a better overall performance, including in complex patients with comorbidities.

My findings suggest that a precision medicine approach combining machine learning algorithms with clinical variables and cardiac biomarkers could improve the diagnostic information provided to clinicians when assessing patients with suspected myocardial infarction and heart failure in the Emergency Department.

Lay Summary

The British Heart Foundation estimates that more than half of the population will develop a heart or circulatory disease during their lifetime. These are some of the most common causes of death in the United Kingdom. Two of the most common conditions are heart attack and heart failure. More than 25 million admissions to hospitals each year are related to symptoms of heart attacks or heart failure. Despite the availability of specific and sensitive blood tests, the diagnosis can be challenging. Artificial intelligence (AI) uses data to help humans to solve problems quicker and more accurately than was previously possible. My aim in this thesis is to develop and test AI-guided tools to help doctors and nurses diagnose these conditions earlier and more accurately in the Emergency Department.

In 20,761 patients from a large Scottish dataset with suspected heart attacks, we tested a previously developed AI tool to assess its performance in routine clinical practice data, including in the prediction of future heart attacks. This AI tool, which uses age, sex, and a blood test that measures heart damage (cardiac troponin), had a good overall performance, and moreover, it could predict the chance of future events too. However, the analysis showed that its performance was not as good in patients at intermediate heart attack risk and there was less consistency in higher risk patients.

It is well known that cardiac troponin is influenced by patient factors, such as older age, sex and pre-existing medical conditions. We used a large study of

patients who presented to hospitals across Scotland with symptoms of a heart attack to develop and test CoDE-ACS (Collaboration for the Diagnosis and Evaluation of Acute Coronary Syndrome). This is an AI tool that combines these variables along with other blood tests, and can accurately predict the probability of heart attack. When we compared CoDE-ACS with the current approach used in hospitals, the performance was more consistent and accurate.

For the second half of the thesis, we focused on the development and testing of an AI tool for heart failure. This condition can be hard to diagnose because symptoms such as shortness of breath, coughing and leg swelling are similar to other illnesses. To address this, we developed and tested the CoDE-HF (Collaboration for the Diagnosis and Evaluation of Heart Failure) tool in 10,369 patients from 13 countries to help improve the diagnosis and identification of heart failure. CoDE-HF combines blood test results with simple clinical variables known to be associated with acute heart failure. First, we used a blood test that measures a small protein released in heart failure (NT-proBNP) as it is the most common test used in hospitals. Then, we retrained CoDE-HF, so it can also work with two other proteins (BNP and MR-proANP) that can be also used in the diagnosis of acute heart failure. Last, we compared my solution with the guideline-recommended approach showing that my new AI tool could achieve a better overall performance, including those with previous medical conditions where the diagnosis of heart failure is more challenging.

My findings suggest that by combining AI with patient information and blood tests, we can achieve a more personalised approach when it comes to the diagnosis of heart attack and heart failure in the Emergency Department.

Acknowledgements

The work presented in this thesis was only possible with the support and expert supervision of Professor Nicholas Mills (Professor of Cardiology and Consultant Cardiologist), Dr Athanasios Tsanas (Senior Lecturer) and Dr Atul Anand (Clinical Senior Lecturer and Consultant Geriatrician). I am incredibly grateful to have had the opportunity to work for and learn from them and look forward to continuing to do so in the years to come.

I would like to thank the Medical Research Council for their support through the Precision Medicine Doctoral Training Programme (MR/N013166/1). Without their generosity, this thesis would not have been possible.

In addition, I wish to acknowledge the support of all members of the 'High-STEACS' research team. In particular, Dr Kuan Ken Lee, with whom I have spent and enjoyed the last couple of years collaborating on many different projects.

I would like to acknowledge the support of my family and friends.

Finally, my thanks to Iliana. For everything. She has been with me during both my MSc and PhD studies and has been a constant source of support.

Abbreviations

ANP – Atrial Natriuretic Peptide

AUC – Area Under the receiver-operating-characteristic Curve

AUC-PR – Area Under the Precision-Recall Curve

BNP – B-type Natriuretic Peptide

CABG – Coronary Artery Bypass Grafting

CAD – Coronary Artery Disease

CART – Classification and Regression Trees

CHD – Coronary Heart Disease

CI – Confidence Interval

CNN – Convolutional Neural Networks

CNP – C-type Natriuretic Peptides

CoDE-ACS – Collaboration for the Diagnosis and Evaluation of Acute Coronary Syndrome

CoDE-HF – Collaboration for the Diagnosis and Evaluation of Heart Failure

CVD – Cardiovascular Disease

ESC – European Society of Cardiology

ECG – Electrocardiogram

FN – False Negative

FP – False Positive

HFpEF – Heart Failure with preserved Ejection Fraction

HFrEF – Heart Failure with reduced Ejection Fraction

High-STEACS – High-Sensitivity Troponin in the Evaluation of patients with suspected Acute Coronary Syndrome

IPD – Individual Patient Data

MI^3 – Myocardial-Ischemic-Injury-Index

MICE – Multivariate imputation by chained equations

MR-proANP – Mid-Regional proANP

NHS – National Health Service

NPV – Negative Predictive Value

NSTEMI – Non-ST-segment Elevation Myocardial Infarction

NT-proBNP – N-terminal piece of proBNP

STEMI – ST-segment Elevation Myocardial Infarction

PCI – Percutaneous Coronary Intervention

PPV – Positive Predictive Value

ProVa CoDE-HF – Prospective Validation of the CoDE-HF

RNN – Recurrent Neural Networks

ROC – Receiver Operating Characteristic

TP – True Positive

TN – True Negative

URL – Upper Reference Limit

XGBoost – Extreme Gradient Boosting

Contents

Figures and tables

CHAPTER 1

Introduction

Chapter 1 Introduction

1.1 Overview of cardiovascular conditions

Cardiovascular diseases claim the lives of 1.8 million European citizens each year and affects the lives of 60 million people living with cardiovascular disease across the continent.[1] Despite the severity of the coronavirus pandemic, cardiovascular disease continues to be Europe's leading cause of death. Furthermore, the prevalence of cardiovascular disease is expected to rise as our population ages.

Chest pain and shortness of breath are the most typical symptoms of acute cardiovascular disease. Together the assessment of people with these symptoms is responsible for 1 in 4 presentations to the Emergency Department and 1 in 10 unplanned hospital admissions. Patients and clinicians do not want to overlook a potentially life-threatening cardiovascular condition such as acute myocardial infarction or heart failure. Prompt and accurate diagnosis is critical and significantly impacts patient outcomes. However, assessing life-threatening cardiovascular disease can be difficult because it may be indistinguishable from other prevalent conditions with similar symptoms.

1.2 Progression of disease

Despite advancements in cardiovascular disease detection and treatment, as well as significant reductions in myocardial infarction mortality over the last decade, over 170,000 patients die in the United Kingdom each year.[2] Our population is ageing, in part due to advances in acute cardiac care, resulting in more individuals living with the repercussions of chronic cardiac problems than ever before.

The coronavirus pandemic caused significant disruptions in healthcare throughout Europe. Concerns that acute medical care would be overburdened led to imposition of limitations on citizen movement. As a result, emergency presentations for acute cardiovascular diseases declined in Europe, whereas cardiovascular deaths rose.[3] This emphasises the insufficiency of the existing patient assessment approach and the necessity to build effective care pathways.

1.3 Machine learning and precision medicine

Artificial intelligence has the potential to transform clinical practice. Machine learning, in particular, has been promoted as an objective, repeatable method for combining many quantitative data to improve risk prediction. Machine learning algorithms can combine cardiac biomarkers with key clinical data and other routine tests to quickly inform the diagnostic probability of various acute conditions and guide clinical decisions for patients who present to the Emergency Department with suspected cardiovascular disease.

Emergency Departments are demanding and often high-pressure environments in which clinicians must aggregate, examine, and synthesise vast amounts of new or previously available data. Decision support tools that use machine learning algorithms could improve clinical judgments and surpass current diagnostic pathways based on single tests with binary thresholds, resulting in more consistent, effective and personalised care.

1.4 Scope and objectives

There are two core aims to this thesis. Firstly, it will seek to develop and validate machine learning algorithms that incorporate blood biomarkers and other objective clinical variables to improve the diagnosis of acute cardiac conditions. These algorithms will be deployed as decision support tools that can be applied at the patient's bedside to aid in diagnosis and risk stratification in the Emergency Department. Secondly, it will compare the diagnostic performance of the proposed machine learning algorithms with the current national and international clinical practice guidelines using fixed threshold approaches.

The following hypotheses will be addressed:

i) The myocardial-ischemic-injury-index (MI^3) machine learning algorithm can predict the index diagnosis of myocardial infarction and risk of subsequent myocardial infarction or cardiovascular death at one year in a consecutive patients with suspected acute coronary syndrome population (Chapter 4).

ii) A machine learning model that incorporates cardiac troponin as a continuous variable and other objective clinical variables can predict the index diagnosis of myocardial infarction in a consecutive patient population with suspected acute coronary syndrome (Chapter 5).

iii) A machine learning model that incorporates natriuretic peptides as a continuous variable and other objective clinical variables can predict the index diagnosis of acute heart failure in patients with breathlessness (Chapter 6 and Chapter 7).

1.5 Structure of the thesis

The thesis begins in Chapter 2 with background information on the physiology of the heart, acute cardiac conditions and their diagnostic cardiac biomarkers. Chapter 3 provides a detailed description of the datasets used in this thesis, the methodology which is applicable in the following chapters and explains some of the evaluation metrics widely used to assess the diagnostic performance of the proposed algorithms. Chapter 4 evaluates the performance of a previously developed machine learning algorithm (MI^3) in a consecutive and heterogeneous patient population. Chapter 5 evaluates the diagnostic performance of guideline-recommended cardiac troponin thresholds and describes the development and external validation of a new decision-support tool that uses machine learning to calculate the probability of myocardial infarction for each patient. Chapter 6 evaluates the diagnostic performance of guideline-recommended N-terminal pro-B-type natriuretic peptide (NT-proBNP) thresholds and describes the development and external validation of a decision-support tool that uses machine learning to calculate the probability of acute heart failure for each patient. Chapter 7 evaluates the diagnostic performance of guideline-recommended B-type natriuretic peptide (BNP) and mid-regional proANP (MR-proANP) thresholds and describes the development and external validation of a decision-support tool that uses machine learning to calculate the probability of acute heart failure for each patient. Chapter 8 draws conclusions and describes the next steps.

CHAPTER 2

Background

Chapter 2 Background

2.1 Overview

The symptoms of acute heart failure and acute myocardial infarction are among the most common reasons for presenting to a hospital worldwide. Together the assessment of people with these symptoms is responsible for 1 in 4 presentations to the Emergency Department and 1 in 10 unplanned hospital admissions. However, the diagnosis of life-threatening cardiovascular disease can be challenging as the signs and symptoms can mimic those of other differential conditions.

Recently, the national and international clinical guidelines recommend the use of specific and sensitive cardiac biomarkers, such as cardiac troponin and natriuretic peptides, to aid with the diagnoses and risk-stratification of these patients. However, current practice relies on fixed thresholds to rule-in or rule-out these conditions based on the average person and does not recognise that cardiac biomarkers are influenced by age, sex, and comorbidities. Hence, whether these cardiac biomarkers are used effectively with optimal diagnostic performance is unclear.

2.2 The cardiovascular system

The heart is a muscle that works as a pump to provide blood supply throughout the body. Together with the arteries, veins and capillaries, this forms the cardiovascular system.

2.2.1 Physiology of the heart

The heart, which has the size of an adult's fist, consists of three layers. The pericardium, which is a thin sac that surrounds and protects the heart. The myocardium, is the middle layer which is also the muscle of the heart and pumps the blood, and the endocardium, which lines the chambers and the valves of the heart.

The inside of the heart is divided into four chambers – two on top and two on the bottom. The upper chambers are the left atrium and the right atrium (or the atria), while the lower chambers are the left ventricle and the right ventricle. The atria receive blood entering the heart, while the ventricles pump the blood out of the heart. The chambers are connected by cardiac valves that control the flow of blood.

The four cardiac valves that ensure the blood will flow only in one direction are:

- The aortic valve, located between the left ventricle and the aorta which carries blood to the whole body.

- The mitral valve, which separates the left atrium from the left ventricle.
- The pulmonary valve, located between the right ventricle and the pulmonary artery which carries blood to the lungs.
- The tricuspid valve, which separates the right atrium from the right ventricle.

The left side of the heart is larger and more muscular compared to the right side. That is because the left side pumps blood to the whole body, while the right side pumps blood only to the lungs.

2.2.2 Circulatory system

The circulatory system consists of the systemic circulation, which circulates the blood from the heart to the whole body and back again, and the pulmonary circulation, which carries the blood from the heart to the lungs and back again. The circulatory system is made up of blood vessels. There are three main types of blood vessels:

- Arteries, which take reoxygenated blood that comes from the lungs to the heart back to the whole body.
- Veins, which take the deoxygenated blood that has circulated to the whole body back to the heart.
- Capillaries, which connect the arteries and the veins via small, thin blood vessels.

2.2.3 Coronary arteries

The heart is an organ, and like every other organ, it requires blood to survive. Blood that needs to be oxygen-rich and full of nutrients. Although its chambers are full of that blood, the heart receives its blood via its own vascular system, called coronary circulation. The arteries that form the coronary circulation branch off the aorta and are called coronary arteries. There are two main coronary arteries. The left main coronary artery, which supplies blood to the left side of the heart, and the right main coronary artery, which supplies blood mainly to the right side of the heart. These coronary arteries branch off into smaller arteries, which help supply blood to the heart. When the blood supply to the heart is reduced or stopped due to narrowing or blockage of the coronary arteries, this is called coronary artery disease (CAD) or coronary heart disease (CHD) and has severe consequences.

2.3 Acute cardiac diseases

Cardiovascular disease (CVD) is a class of diseases that involve the heart or blood vessels. The CVD umbrella term includes diseases such as CAD, angina, myocardial infarction, cardiac disease, hypertension, stroke and more.

Two of the most common cardiac diseases, also known as heart diseases, are myocardial infarction and heart failure. The main difference between these two diseases is that myocardial infarction is a disease arising from the coronary blood vessels, while heart failure is caused by dysfunction in the muscle and function of the heart. Most typically, a myocardial infarction occurs suddenly, while heart failure happens gradually.

2.3.1 Myocardial injury

Myocardial injury or myocardial necrosis refers to the death of the tissue in the heart, and recently it was acknowledged as a separate entity.[4] It can arise from both ischaemic and non-ischaemic conditions, as well as cardiac and non-cardiac conditions. It is defined using circulating blood markers such as cardiac troponin (see section 1.4 diagnostic biomarkers). If the cardiac troponin concentration is above the 99th percentile upper reference limit (URL), this is defined as myocardial injury.[4,5] Myocardial injury can be classified either as acute or chronic. Acute if there is a rise and/or fall of the cardiac troponin concentration, and chronic if the change between two cardiac troponin measurements is less or equal than 20%.[5,6]

2.3.1.1 Aetiology of myocardial injury

Atherosclerosis is a pathologic process that causes CAD[7,8] and has spread worldwide.[9,10] It is caused by plaque (fatty deposits) building up in the inner lining of the blood vessels. These fatty deposits can be made up of cholesterol, fatty and other substances found in the blood. The plaque can cause the arteries to narrow or even block, reducing the blood flow and oxygen supply to the heart and other organs such as the brain. Atherosclerosis is a slow, progressive disease that may not cause symptoms for decades, even though it can start as early as childhood. The most typical risk factors for atherosclerosis are smoking, high cholesterol levels, diabetes, hypertension, obesity and lack of exercise. Atherosclerosis can happen to any artery in the body. Depending on which artery is affected, atherosclerosis can lead to the injury of the heart or a having stroke. If atherosclerosis is in the coronary arteries, symptoms such as chest pain or pressure (angina) can be experienced for moderate to severe cases of atherosclerosis.

2.3.1.2 Symptoms

Chest pain or angina is one of the most common reasons for presentation to hospitals worldwide,[11] with more than one million attendances each year in the United Kingdom alone.[2] When a part of the heart does not get enough blood, a patient will start feeling pain, discomfort or pressure in the chest. However, this discomfort can travel to the arm, shoulder, neck and jaw. Women and older people more commonly experience these symptoms.[12,13] There are two types of angina. Stable angina, which can be controlled with

medication or usually rest after an exercise and is predictable, and unstable angina, which cannot be controlled and can happen at any time, even when at rest or sleeping.

2.3.1.3 Myocardial ischaemia

Ischaemia is the restriction in blood supply to any organ of the body. When the ischaemia affects the heart muscle, it is called cardiac ischemia or myocardial ischemia.

Although the most common symptoms of myocardial ischaemia are acute chest pain, shortness of breath, or pain in the shoulder, arm, neck and jaw, myocardial ischaemia can have no symptoms at all. The latter is called asymptomatic or silent myocardial ischaemia and can be detected through electrocardiographic monitoring.[14] The most common cause of myocardial ischaemia is atherosclerotic plaque.[15] As the restriction of blood supply to organs results in the death of the organ (or part of it), myocardial ischaemia will ultimately lead to myocardial infarction.

2.3.1.4 Myocardial infarction

Myocardial infarction, commonly known as heart attack, is a condition that affects the myocardium. Infarction refers to the necrosis (death) of the myocardial tissue due to the lack of oxygen-rich blood. Plaque rupture or erosion resulting in the creation of a blockage in the coronary arteries is the most common cause of myocardial infarction. Myocardial infarction is a form of myocardial injury but requires clinical evidence of acute myocardial ischaemia.[4,16]

Myocardial infarction can be differentiated between ST-segment elevation myocardial infarction (STEMI) and non-ST-segment elevation myocardial infarction (NSTEMI). Of the two, the former is the most acute form, as it usually occurs from a complete blockage occurs in a coronary artery. The diagnosis is made based on the electrocardiogram (ECG) trace.

2.3.1.5 Classification of myocardial infarction

From the first universal definition of myocardial infarction in 2007 by the Global Task Force[17] the discovery of novel biomarkers that can identify acute myocardial injury more accurately has led to the fourth universal definition in 2018.[4] The fourth universal definition of myocardial infarction is based on a classification system with five subcategories as follows:

- Type 1 myocardial infarction, which is caused by atherothrombotic coronary artery disease which is usually precipitated by plaque disruption (Figure 2.1).
- Type 2 myocardial infarction, due to a mismatch between oxygen supply and demand that is unrelated to acute coronary atherothrombosis (Figure 2.2).
- Type 3 myocardial infarction, which is defined as cardiac death with symptoms suggestive of myocardial ischaemia that occurred before an elevation of the cardiac biomarker values.
- Type 4 myocardial infarction, associated with percutaneous coronary intervention (PCI). This is further subcategorised into:

    - Type 4a myocardial infarction, which is related to PCI procedure that occurred within 48 hours after the index procedure.
    - Type 4b myocardial infarction, due to stent/scaffold thrombosis.
    - Type 4c myocardial infarction, due to in-stent restenosis or restenosis following balloon angioplasty in the infarct territory.
- Type 5 myocardial infarction, associated with coronary artery bypass grafting (CABG), i.e. cardiac surgery, occurring within 48 hours of the procedure.

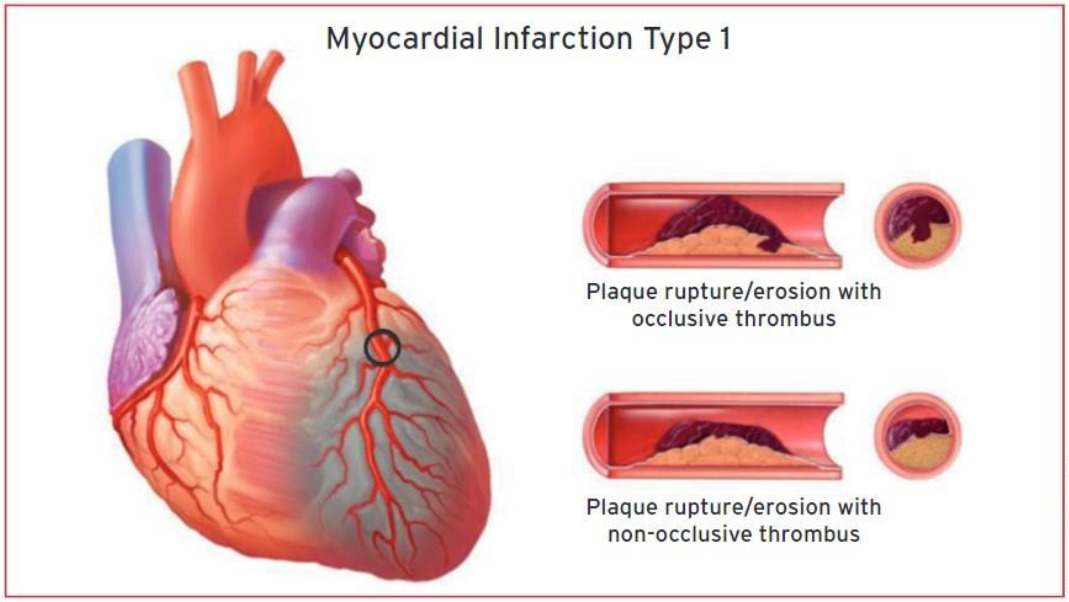


Figure 2.1 Myocardial infarction type 1.
Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). European Heart Journal 2018;40:237-69 by permission of Oxford University Press.

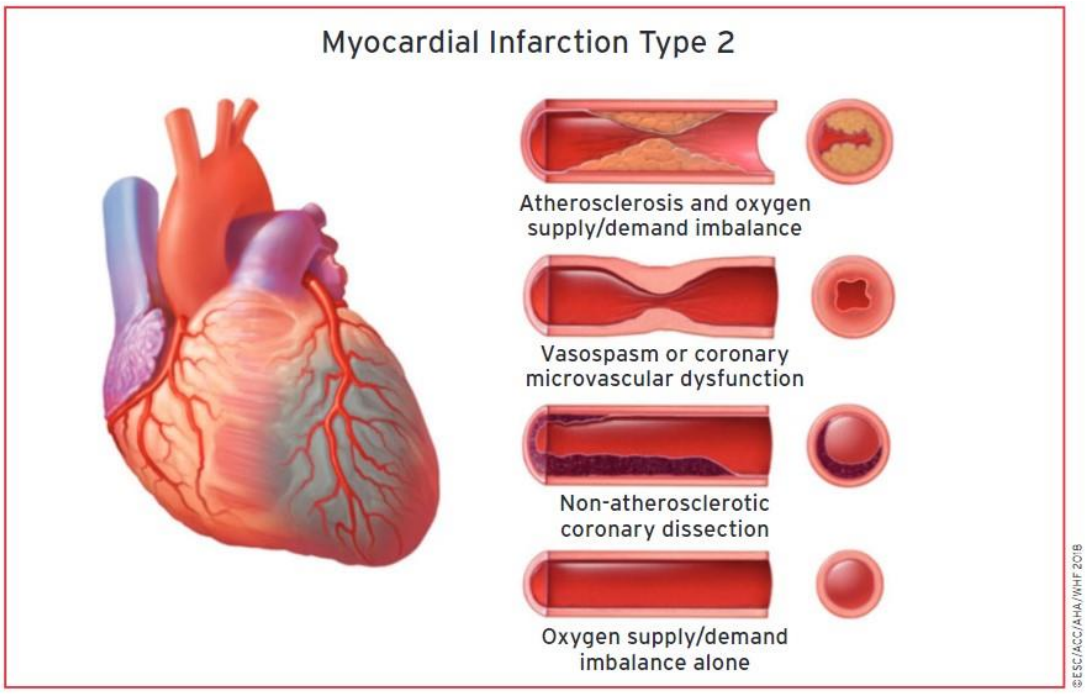


Figure 2.2 Myocardial infarction type 2.
Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). European Heart Journal 2018;40:237-69 by permission of Oxford University Press.

2.3.2 Heart failure

Heart failure is a progressive illness that occurs when the heart can no longer provide adequate blood to the meet body's requirements.[18] It can be classified as either heart failure with reduced ejection fraction (HFrEF) or heart failure with preserved ejection fraction (HFpEF). The ejection fraction refers to the proportion of blood pumped from the ventricles per contraction, and it is defined as ≤ 40% for HFrEF (Figure 2.3). When it comes to CVD, heart failure is a condition that can be challenging to diagnose although common. Heart failure has been associated with substantial morbidity and mortality risk.[19]

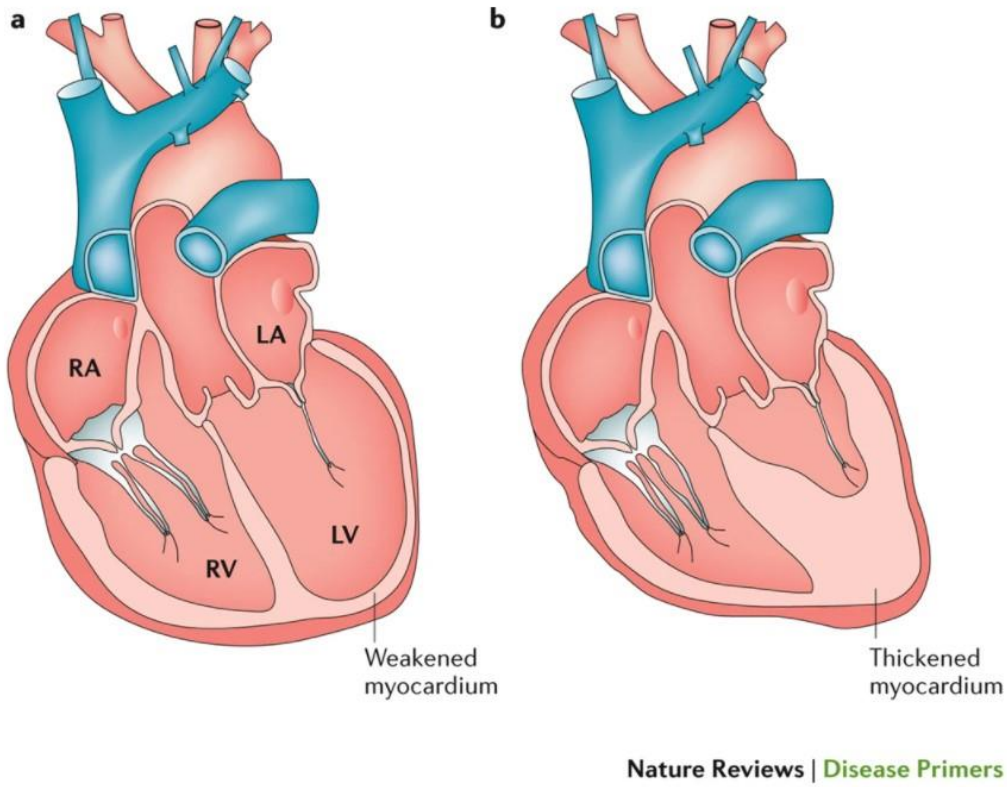


Figure 2.3 Different mechanisms as underlying causes of these conditions. a) In patients with HFrEF, the ejection fraction is defined as ≤40%. The weakened ventricular myocardium can no longer pump the blood adequately. b) In patients with HFpEF, the ventricular myocardium fails to relax during diastole because the tissue has thickened or stiffened. As a result, the chamber cannot adequately fill with blood. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle.
Bloom MW, Greenberg B, Jaarsma T, et al. Heart failure with reduced ejection fraction. Nature Reviews Disease Primers 2017;3:17058 by permission of Springer Nature.

2.3.2.1 Aetiology of heart failure

Acute myocardial infarction is the most common cause of heart failure worldwide.[20-22] Besides myocardial infarction, other CAD, diabetes mellitus, and hypertension are also considered precursors of heart failure.[23-25] Although, there are differences between HFrEF and HFpEF patients when it comes to risk factors, as older people and women tend to develop HFpEF,[26] while those with an injury that has affected the heart tend to develop HFrEF.[27] Patients with chronic and stable HFrEF, can be affected by changes in their diet, excess of fluids or salt,[19] or factors such as ischaemia and infection in the heart.[28]

2.3.2.2 Symptoms

The symptoms of heart failure can be similar to other medical conditions, making diagnosis difficult.[29-32] These symptoms include shortness of breath (dyspnoea), fatigue and limitations in exercise tolerance and fluid accumulation. However, the most common symptom is breathlessness, since the lungs are struggling to take in additional required oxygen as the heart does not function properly enough to supply blood, hence oxygen, to the body. Depending on if the left or the right ventricle loses its ability to pump blood, different symptoms can be experienced. For the left ventricle, the most common symptom is the fluid accumulation in the lungs (pulmonary oedema) that causes the breathlessness and inability to lie flat/sleep flat (orthopnoea), with less common symptoms being cough and fingernails or lips that turn

blue. In contrast, for the right ventricle, the most common symptom is fluid accumulation in the legs (pedal oedema), which can become very extensive, while less common symptoms are nausea, loss of appetite and frequent urination.

2.3.2.3 Acute vs chronic heart failure

When it comes to hospital presentations, heart failure can be divided into acute heart failure or chronic heart failure.[33] The term “acute” is used to describe a rapid onset or worsening of symptoms that require urgent evaluation and treatment, typically leading to urgent hospital admission. Acute heart failure is a leading cause of hospital admissions in older patients. The mortality can range from 5% to 10% for in-hospital[34-37] and from 25% to 45% for 1-year post-discharge.[38-40] If the diagnosis of heart failure has been established, but there is no worsening of symptoms, this is called chronic heart failure. However, if chronic heart failure rapidly deteriorates, this is called decompensated heart failure, which often leads to an emergency presentation or unscheduled hospital admission.[19,33,41,42]

2.4 Diagnostic biomarkers

As the signs and symptoms of acute myocardial infarction and acute heart failure can be similar to other medical conditions, the use of clinical biomarkers offers the potential of assisting with their difficult and challenging diagnosis of them. New blood tests are able to measure these blood biomarkers and identify even small areas of heart muscle death. Hence, the national and international guidelines recommend the use of such clinical biomarkers with an aim for a more objective patient assessment and risk stratification. The recommended clinical biomarker for the diagnosis of myocardial injury is cardiac troponin, while for the diagnosis of acute heart failure, natriuretic peptides are recommended.[4,42,43]

2.4.1 Blood biomarkers of myocardial injury

2.4.1.1 Cardiac troponin

Cardiac troponin consists of three subunits, troponin C, troponin T and troponin I. In the first universal definition of myocardial infarction in 2007, the Global Task also recommended the use of cardiac troponin I and T as the preferred biomarker for myocardial injury.[17] Previous cardiac biomarkers, such as creatinine kinase and lactate dehydrogenase, were no longer recommended as they offered lower specificity than cardiac troponin.[44-46] In the setting of acute myocardial injury, the release of cardiac troponin in the blood is regulated in a time-dependent fashion.[47-49] This has led to the widespread adoption of cardiac troponin, as more than 95% of Europe's

laboratories use cardiac troponin as the preferred biomarker for the diagnosis of myocardial injury.[50] Nowadays, technological advances allow us to measure cardiac troponin even more accurately with the use of high-sensitivity cardiac troponin assays.

2.4.1.2 High-sensitivity cardiac troponin assays

Cardiac troponin assays were used to measure cardiac troponin until around 10 years ago.[51] However, the advance of technology has led to the development of high-sensitivity cardiac troponin assays, which allows for the quantification of extremely low levels of cardiac troponin with excellent precision.[52] Although, there are still analytical differences between different assays from different manufacturers, the improvement in sensitivity and specificity has allowed cardiac troponin testing to be performed in a majority of healthy individuals to define the 99th percentile URL. Although cardiac troponin will rise to indicate a myocardial injury, the underlying cause cannot always be clear. Some of the most common causes of elevated cardiac troponin values are sepsis, pulmonary embolism, chronic kidney disease and chronic inflammation.[45,53-57] All these conditions can cause small cardiac troponin elevations, which now can be detected by the high-sensitivity cardiac troponin assays. Moreover, it has been observed that older patients presenting to the emergency department have elevated concentrations of cardiac troponin.[58-60] However, guidelines do not recommend age-specific thresholds for the diagnosis of myocardial infarction, while they recommended sex-specific thresholds to avoid to under diagnosing

myocardial infarction in women.[61] These have led to diagnostic uncertainty as the specificity of the test for type 1 myocardial infarction has been reduced.[62,63]

2.4.1.3 Cardiac troponin for the diagnosis of acute myocardial infarction

The 99th percentile URL is used as the cut-off point for the diagnosis of myocardial injury and can be different for each different assay. However, the significant differences between men and women have led to the recommendation to use sex-specific 99th percentile URL for hs-cTn assays.[64-67] Many clinical pathways have been proposed over the recent years.[61,68-73] The development of the high-sensitivity cardiac troponin assays has also allowed the time required between the presentation and the second cardiac troponin measurement to be shortened. Previously the clinical pathways for the diagnosis of myocardial infarction required two cardiac troponin measurements at least 6 hours apart, while the European Society of Cardiology (ESC) rule-in and rule-out algorithm (decision chart) requires 3 hours.[43,68] Recent studies have shown that very low cardiac troponin concentrations can be used to rapidly identify and discharge patients with low risk of myocardial infarction.[61,68,69,74,75]Hence, the latest guidelines now recommend to use new rapid rule-in and rule-out algorithms (decision charts), the 0 h/1 h algorithm or the 0 h/2 h algorithm (Figure 2.4).[43] Although this has enabled myocardial infarction to be ruled out earlier, resulting in shorter stays in the emergency department and lower costs[69,73-79], it has led

to the identification of more patients with myocardial injury and diagnostic uncertainty, which could result in unnecessary treatment and harm. Hence, the optimal approach is uncertain. In Chapter 5, we are going to evaluate the guideline recommend clinical pathway performance and address its limitation with the use of machine learning algorithms.

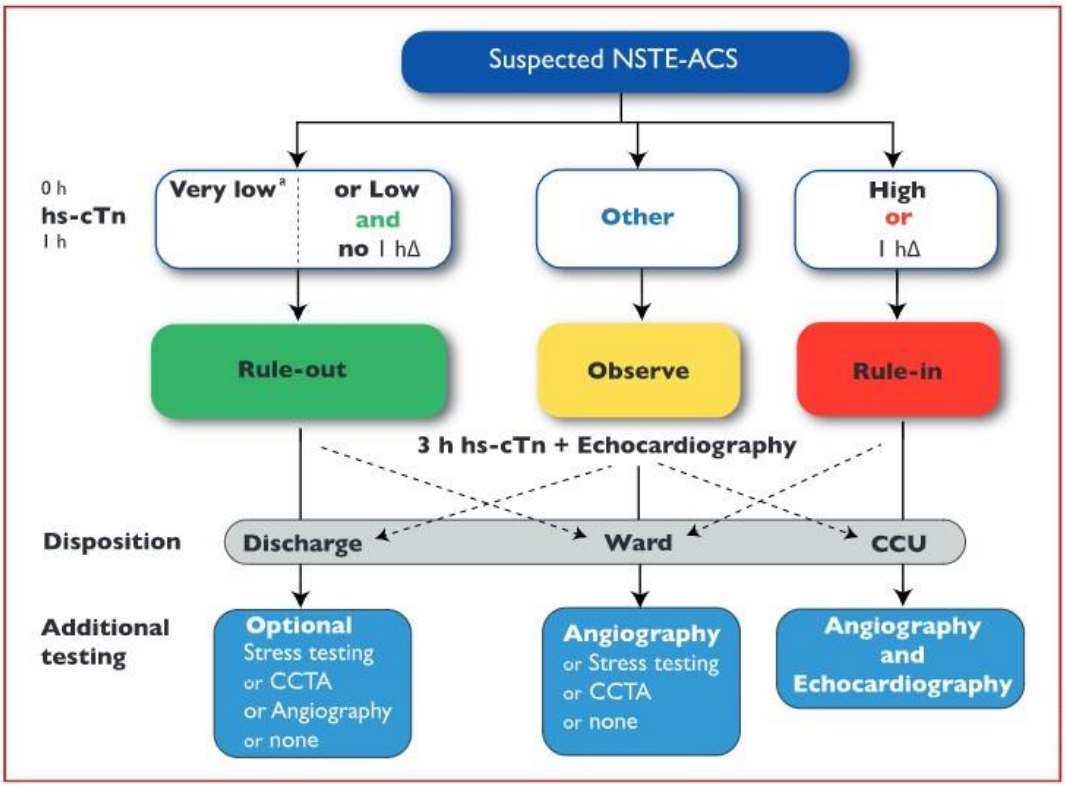


Figure 2.4 0 h/1 h rule-out and rule-in algorithm, or decision chart, using high-sensitivity cardiac troponin assays in patients presenting with suspected non-ST-segment elevation acute coronary syndrome to the emergency department. Cut-offs are assay specific. CCU = coronary care unit; CCTA = coronary computed tomography angiography; hs-cTn = high-sensitivity cardiac troponin; NSTE-ACS = non-ST-segment elevation acute coronary syndrome; [a]Only applicable if chest pain onset >3 h.
Collet JP, Thiele H, Barbato E, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. European Heart Journal 2021;42:1289-367 by permission of Oxford University Press.

2.4.2 Blood biomarkers of heart failure

Accurate and timely diagnosis of acute heart failure can be challenging, especially in patients presenting for the first time with breathlessness. Therefore both national and international guidelines recommend natriuretic peptides as the gold standard biomarkers to aid with the diagnosis.[42,80] The natriuretic peptides are the most extensively studied and used biomarkers in heart failure.[81]

2.4.2.1 Natriuretic peptides

Natriuretic peptides are peptide hormones synthesised by the heart, brain and other organs. The release of these peptides by the heart is usually in response to heart failure. The main physiological action of natriuretic peptides is to reduce arterial pressure by decreasing blood volume and systemic vascular resistance.

There are three main types of natriuretic peptides available; atrial natriuretic peptide (ANP), B-type natriuretic peptide (BNP) and C-type natriuretic peptides (CNP). BNP is one of the natriuretic peptides synthesised largely by the ventricles (as well as in the brain where it was first identified). Proteolysis is a process in which a protein is broken down partially into peptides or completely into amino acids. Proteolysis of BNP results in the formation of proBNP after removal of a 26-amino acid signal peptide.[82] A subsequent proteolysis of proBNP results in an N-terminal piece of proBNP (NT-proBNP).

Both BNP and NT-pro-BNP are sensitive diagnostic markers for heart failure.[33,83] However, BNP has a half-life of approximately 20 minutes, which results in a lower clinical utility compared with NT-proBNP, which has a half-life of approximately 1 to 2 hours.[84] Both ANP and CNP are not used as diagnostic biomarkers, as the former is less consistent and the latter is primarily produced in the in the central nervous system rather than in the heart.[85] However, the recent developments in natriuretic peptides assays has allowed the measurement of the precursor hormone of ANP, called mid-regional proANP (MR-proANP), which is a more stable and reliable diagnostic biomarker.

Natriuretic peptides are known to be influenced by multiple patient factors, which make the already difficult diagnosis of heart failure even more challenging. This is crucial, as most patients with heart failure are older with one or more comorbidities.[86-88] Two of the most common factors that influence natriuretic peptides are renal disease and atrial arrhythmias, both known to increase natriuretic peptide concentrations.[87,89-91] Obesity, on the other hand, affects natriuretic peptides in the opposite way, as it lowers the natriuretic peptide concentrations.[92-94]

2.4.2.2 Natriuretic peptide assays

The ESC guidelines recommend measuring BNP or NT-proBNP to aid in the diagnosis of heart failure. Although both BNP and NT-proBNP are well correlated and performance can be similar in most cases, there are some

important differences when it comes to their assays. Even though the BNP assay is highly sensitive and fairly specific for diagnosing heart failure, the fact that all BNP assays use different antibodies and materials without any standardisation makes it difficult to generalise their performance.[95] On the other hand, NT-proBNP assays are generally considered harmonised because they use the same calibrators and antibodies for peptide detection as Roche Diagnostics holds the patent for the antibodies used by the other manufacturers' assays. However, there are still analytical differences between different NT-proBNP assays from different manufacturers, which were measured to be <10% variation in measured values.[96] These differences between the assays of these biomarkers play a significant role in their application in clinical practice.

2.4.2.3 Natriuretic peptides for the diagnosis of acute heart failure

The current national and international guidelines recommend the use of fixed thresholds to aid in the diagnosis of acute heart failure.[33,42,97] The cut-offs to rule-out heart failure with BNP, NT-proBNP and MR-proANP are 100 pg/mL, 300 pg/mL and 120 pg/mL, respectively. Concentrations below these thresholds make the diagnosis of acute heart failure unlikely. The diagnostic performance of the BNP and NT-proBNP thresholds has been validated in multiple cohorts[98-100], while fewer data are available for the MR-proANP performance.[101] Age is maybe the most influential factor of natriuretic peptides concentrations as baseline BNP and NT-proBNP levels are known to increase with age.[102] To address this in the latest guidelines, age-specific

rule-in cut-offs have been recommended for NT-proBNP (450, 900, and 1,800 pg/mL for those <50 years, 50-75 years, and >75 years, respectively) to improve the diagnosis of acute heart failure.[103] However, the diagnostic performance of NT-proBNP have mainly been performed in relatively small, selected patient cohorts, limiting the generalisability of study findings across clinically important subgroups. In Chapters 6 and 7, we are going to address these limitations with the use of machine learning algorithms.

In the next chapter, we will describe the methodology that was used throughout this thesis and introduce and explain important evaluation metrics to assess the models developed in chapters 4 to 7.

CHAPTER 3

Methodology

Chapter 3 Methodology

3.1 Overview

This thesis uses data from cohort studies from Scotland, Australia and New Zealand and data derived from a systematic review and individual patient-level meta-analysis from 13 countries around the world.

The statistical machine learning models, methodology, performance evaluation metrics and a description of the R Shiny decision support tools are described in this chapter.

Continuous variables were compared using parametric and non-parametric tests as appropriate. Statistical significance was taken as a two-sided $P<0.05$. All analyses were performed using R (versions 3.6.3 to 4.1.2).

3.2 Data

Chapter 4 uses data from the High-STEACS trial.[73] High-Sensitivity Troponin in the Evaluation of patients with suspected Acute Coronary Syndrome (High-STEACS) was a stepped-wedge cluster randomised controlled trial evaluating the effect of implementation of a high-sensitivity cardiac troponin I assay in consecutive patients presenting with suspected acute coronary syndrome across 10 secondary and tertiary hospitals in Scotland (ClinicalTrials.gov number, NCT01852123). Between June 10, 2013, and March 3, 2016, all patients with suspected acute coronary syndrome who met the inclusion criteria were included in this study with demographic and clinical data collected prospectively through a dynamic linkage process and housed within the national safe haven. High-STEACS enrolled 48,282 consecutive patients, of which almost half were women (47%), and 10,360 (21%) had myocardial injury detected by the high-sensitivity cardiac troponin assay. The primary outcome was type 1 myocardial infarction or cardiovascular death at one year. All index and primary outcome events were adjudicated according to the Universal Definition of Myocardial Infarction by a panel of ten clinicians blind to the study phase and cardiac troponin concentration during the index presentation.[4,68]

Chapter 5 includes participants of the High-STEACS trial for the development and internal validation of the decision support tool. The IMPACT (Improved Assessment of Chest pain Trial), ADAPT (2-Hour Accelerated Diagnostic Protocol to Assess Patients With Chest Pain Symptoms Using Contemporary

Troponins as the Only Biomarker) and SPACE (Signal Peptide in Acute Coronary Events) cohorts from Australia and New Zealand were used for the external validation.[104-106]

Chapter 6 and chapter 7 use data from a systematic review and individual patient-level meta-analysis of studies that evaluated the diagnostic performance of natriuretic peptides in patients with suspected acute heart failure. The clinical team systematically searched Embase, Medline and the Cochrane Central Register of Controlled Trials to update a previous review by Roberts et al.[107] with studies published up to 7 February 2019 using the following keywords: ‘heart failure’ and ‘natriuretic peptide’. Studies were eligible if they met the following prespecified inclusion criteria: (1) enrolled patients ≥18 years with suspected acute heart failure in an acute care setting, (2) measured natriuretic peptides, and (3) adjudicated diagnosis of acute heart failure using an acceptable reference standard. Two investigators independently screened all studies identified in the systematic literature search, and conflicts were adjudicated by a third using a prespecified protocol (PROSPERO register: CRD42019159407). The corresponding authors of all eligible cohorts were contacted to request anonymised individual patient-level data. For these analyses, fourteen studies provided individual patient-level data in 10,369 patients with suspected acute heart failure across 13 countries.

3.3 Statistical analysis

The following sections aim to provide a general overview of the statistical models used throughout this thesis and subsequently describe the methodological steps followed to develop the decision support tools.

3.3.1 Statistical models

3.3.1.1 Logistic regression

Despite having the word ‘regression’ in its name, logistic regression is probably the most commonly used statistical model to use when it comes to binary classification problems.[108-111] Logistic regression uses the maximum likelihood estimation, and it is fully parametric. It estimates probabilities using the underlying logit function and its inverse, which is the sigmoid function. The sigmoid function maps probabilities to the range [0, 1].

The formula for the logit function:

$$logit(p) = a + bX. \tag{3.1}$$

The formula for the sigmoid function:

$$p = \frac{e^{(a+bX)}}{1 + e^{(a+bX)}} = \frac{1}{1 + e^{-(a+bX)}}, \tag{3.2}$$

where, p is the probability, a is the intercept, b is the vector with the coefficients and X is a dependent variables vector ($X = [x_1, x_2, \ldots, x_n]$, $i = 1, 2, \ldots, N$).

However, there are no model hyperparameters to be tuned. Models' coefficients are adjusted to allow for dependence between the variables. Logistic regression is useful for inference and interpretation as it can be written down as an equation.

The logistic regression model is part of a broad class of models known as generalised linear models (GLM).[112-114] GLMs include linear regression, Poisson regression, ANOVA, and many more. However, one of the assumptions for these models is that the observations are uncorrelated. So, when there is a correlation between observations, e.g. when there are clusters in the dataset, generalized linear mixed models (GLMM) are used as an extension to the GLMs, to allow a random effect to be included as a linear predictor.[115] For the second half of the thesis (chapter 6 and chapter 7), the study identifier was included as a random effect variable whilst all other variables were fitted as fixed effects variables. Furthermore, non-linear relationships between continuous variables and the diagnosis were evaluated using multivariable fractional polynomial methods to maximise the performance of the model.[116] For example, due to the positive skew in the biomarker concentrations, a logarithmic transformation was introduced in the model.

3.3.1.2 Naïve Bayes

Naive Bayes is a supervised machine learning model based on Bayes' Theorem with a 'naïve' assumption of independence among features.[117,118]

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (3.3)$$

where A and B are independent events, $P(A|B)$ and $P(B|A)$ are conditional probabilities, $P(A)$ and $P(B)$ are the prior probabilities of observing each event, respectively and $P(B) \neq 0$.

For classification problems the equation 3.3 can be rewritten as follows:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}, \quad (3.4)$$

where, y is class variable and X is a dependent variables vector ($X = [x_1, x_2, \ldots, x_n]$, $i = 1, 2, \ldots, N$). There are no model hyperparameters to be tuned; however, a kernel density estimation function was used to achieve higher accuracy levels when developing the model. Naïve Bayes was a model comparator in chapters 5, 6 and 7.

3.3.1.3 Random forest

Random forest is a well-known powerful non-parametric supervised machine learning model.[108,119-121] It is an ensemble technique that combines a large number of decision trees using a bagging (bootstrap aggregation) approach to improve the overall performance. The bagging approach grows multiple classification trees in parallel where each tree gives a classification which are called votes. For a new sample, the final decision concerning the classification is accomplished by averaging the estimated class probability

over the ensemble of trees to provide a more accurate and stable prediction. A problem with this process is that it can overfit the data. The solution to this problem is to tune the hyper-parameters. The random forest hyper-parameters were tuned during the development of this model through a grid search strategy using 10-fold cross validation. The hyper-parameters tuned were the number of trees in the forest, the number of variables randomly sampled as candidates at each split, the tree's maximum depth, and the minimum number of samples required to split an internal node. Random forest was a model comparator in chapters 5, 6 and 7.

3.3.1.4 Gradient boosting

Gradient boosting is a supervised machine learning model similar to random forest.[108,122] Gradient boosting employs an ensemble of weak learners, usually decision trees, which are combined in an iterative process. The difference is that in the gradient boosting the decision trees are built sequentially instead of in parallel. The sequential approach aims to improve the model’s accuracy, where each tree attempts to correct the errors of the preceding stage. The final model is a weighted contribution of each decision tree, while a Bernoulli distribution is used as a loss function towards the computation of the probabilities. As gradient boosting is a non-parametric model, it has hyper-parameters that can be tuned to improve the model’s performance. These are the learning rate (shrinkage parameter applied to each tree in the expansion), the interaction depth (maximum depth of each tree, expresses the highest level of variable interactions allowed), the

minimum number of observations in the terminal nodes, and the fraction of the training set observations randomly selected for each subsequent tree. Gradient boosting was the model evaluated in chapter 4. However, as the scope of chapter 4 was to validate a previously developed algorithm trained on a different dataset, these hyper-parameters were set.

3.3.1.5 Extreme gradient boosting

Extreme gradient boosting (XGBoost) is a supervised machine learning model initially proposed by Chen and Guestrin.[123] As mentioned above, gradient boosting employs an ensemble technique to iteratively improve model accuracy for regression and classification problems. This ensemble-based model is achieved by creating sequential models, using decision trees as learners where subsequent models attempt to correct errors of the preceding models.[122]. In the boosting method, individuals that the previous model misclassified are assigned a higher weight to increase their chance of being selected in subsequent models. Each model is subsequently fitted in a step-wise fashion to minimise loss functions such as absolute or squared errors (the number of predicted values differ from the true values). The word ‘extreme’ refers to the re-engineering of gradient boosting to significantly improve the speed of the model by pushing the limits of computational resources. The XGBoost model's output is a probability computed by performing an inverse-logit transformation of the sum of the weights of the terminal nodes of the trained model.

The mathematical formula for the gradient boosting model can be described as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k\,(X), f_k \in F, \tag{3.5}$$

where f is an function that map each variable vector X ($X = [x_i, x_2, \ldots, x_n]$, $i = 1, 2, \ldots, N$) to the outcome y_i, K is the number of Classification and Regression Trees (CART) and F is the space of function containing all CART.[124]

XGBoost optimises an objective function of the form:

$$Obj = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{3.6}$$

where the first term is a loss function, l, which evaluates how well the model fits the data by measuring the difference between the prediction $\hat{y}_i$ and the outcome y_i. The second term, the regularization term, is used by XGBoost to avoid overfitting by penalizing the complexity of the model. Furthermore, to improve and fully leverage the advantages of XGBoost, the hyper-parameters of the algorithm defined below were tuned through a grid search strategy using 10-fold cross validation.

The hyper-parameter values for the model that were tuned are the number of iterations (trees), the learning rate (shrinkage parameter applied to each tree in the expansion), the interaction depth (maximum depth of each tree,

expresses the highest level of variable interactions allowed), the minimum number of observations in the terminal nodes, the fraction of the training set observations randomly selected for each subsequent tree and the fraction of variables randomly sampled for each tree. XGBoost was the selected model for the decision support tools developed in chapters 5, 6 and 7.

3.3.2 Feature selection

Feature selection, or variable selection, is called the process of minimising the number of input variables into the model.[125,126] The number of input variables should be reduced to lower the computational cost of modelling and to increase the model's performance by removing noise. Features can be classified as relevant when they influence the predicted outcome, irrelevant when they do not have an influence or redundant when other features can take their place.

The feature importance can be interpreted as the mean decrease in accuracy that occurs when that feature is added or subtracted. Using out-of-bag data, each time a decision tree is created. Then the feature importance is calculated for each feature in the dataset. This shows how useful or important this feature is during the construction of the multiple decision trees. The higher the relative relevance of a feature is, the more it is used to do splits in the decision trees. For each model, the feature importance across all of the decision trees is then aggregated, and then the features are ranked so that they can be compared to each other (Figure *3.1*).

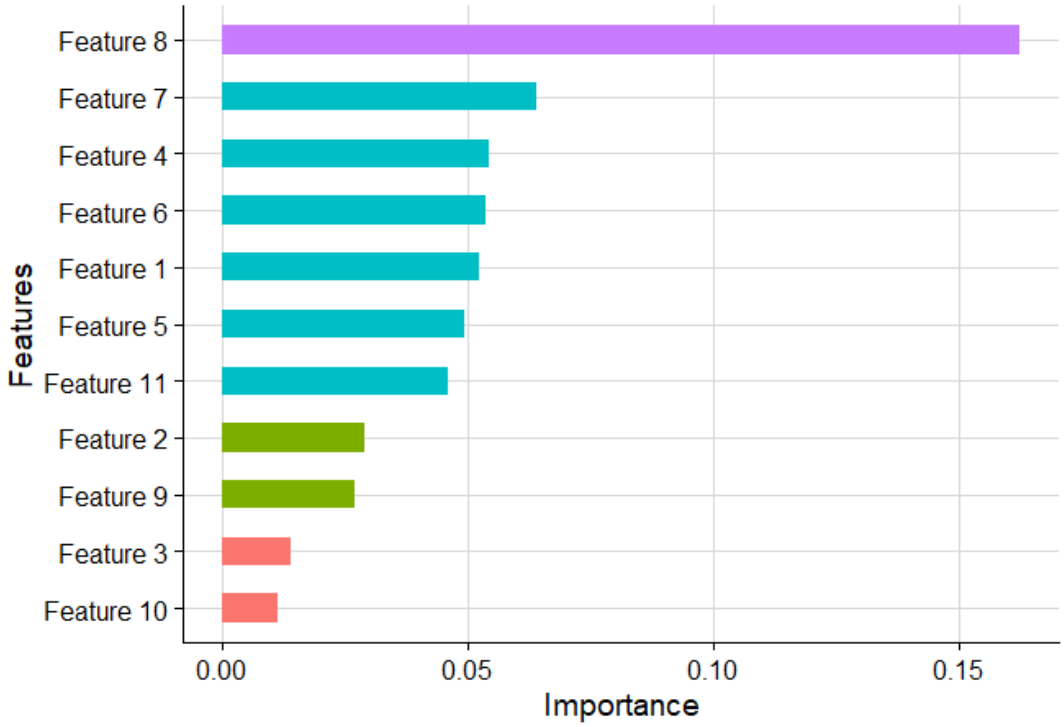


Figure 3.1 In this example, the 11 features available in the example dataset are ranked based on their feature importance. The different colours are for aesthetic purposes only as they help to group them visually. Here, feature 8 is the most important feature, which in these kinds of problems is usually the biomarker. It is up to the researcher to decide how many and which features are going to be used for the model as it always depends on the problem.

For this thesis, and as the goal was to develop models that can be used at the Emergency Department, we worked together with the clinical team to select the features for each model. Our goal was to select the best subset of features that contained the features with the highest feature importance and select features that would be readily available or easy to collect in a busy Emergency Department. This could maximise the clinical utility of the models. For example, subjective features were excluded from the analysis. Features that required a time-consuming process to be collected or are difficult to order tests were excluded. In this way, the developed models in chapter 5, chapter 6 and chapter 7, consist of simple, objective, clinical features that are known to be associated with the outcome, which were found to have the highest relative importance in our model training phase.

3.3.3 Cross validation

Validating the stability of the developed model is always necessary to make sure random factors do not influence the findings, findings are reproducible, and to expect the model to perform correctly with unseen data. Moreover, validation brings assurance that the developed models have correctly identified the majority of the patterns in the data and are not sensitive to noise.[127]

Although cross validation can reduce overfitting, it can only minimise it and not completely protect the model from overfitting. This can be noticed when the model does not perform well in unseen data where one of the reasons can be the use of not representative data in the model development phase. However, even if the performance between the test and train set is similar, this does not mean that there is no overfitting. Cross validation can obtain excellent results when dealing with medical datasets, which are typically relatively small datasets with many characteristics.[128]

There are many different techniques when it comes to the validation of the developed model. Some of them are the holdout method, where a proportion of usually 10% to 30% is left out from the training data and used as a validation set. The leave-n-out cross-validation, where this approach leaves out n data points from the training set and then validates the model on the n data points that were left out. When n = 1, this is known as leave one out cross validation. The k-fold cross validation, where the original dataset is split

into k folds, where the k-1 folds create the training dataset used to train the model and the left out fold is used as a test dataset to evaluate its performance. This process is repeated k times, leaving out a different fold each time. In general, a k=5 or k=10 is preferred.

When dealing with medical datasets, an important factor is the prevalence of the disease. This could affect the model performance as a random split of the dataset could create imbalanced folds. For that reason, a slight variation of the k-fold cross validation was used in this thesis (chapter 5, chapter 6 and chapter 7), called stratified k-fold cross validation, such that each fold has the same prevalence of the disease. In order to make the developed models more general and unbiased, we chose to perform a stratified 10-fold cross validation repeated ten times to improve the statistical confidence. Last, when there were missing values in the dataset, we chose to create ten different imputed datasets. We repeated ten iterations of the 10-fold stratified cross validation process for each one. Then we aggregate the results.

3.3.4 Imputation

Imputation is the process of replacing missing data with substitute values while retaining the majority of the dataset's data/information. These strategies are utilised because eliminating data from a dataset every time is impractical and might result in a significant reduction in the dataset's size, which raises issues about biassing the dataset and leads to inaccurate analysis.[129-131]

There are many options for performing an analysis with missing data. The first and most simple thing is to perform a complete case (listwise deletion) analysis where all rows of data with at least one missing value are deleted. However, this will reduce the dataset's size, hence the power of the analysis, which is not preferable, especially when dealing with medical datasets where their size is relatively small already. Moreover, if the missing values are not missing at random, this technique can introduce bias.

Another approach is to perform a single dataset imputation, filling a single value for a missing data element without specifying an explicit model for the partially missing data. The most popular strategy is to choose the mean/median as a replacement for numerical variables, in which missing values are substituted with the variable mean/median value. The mode, i.e. the value with the highest frequency, is selected in this approach for categorical variables.

However, in this thesis, we used a multiple imputation approach as this is a relatively flexible and general purpose approach when dealing with missing data. In multiple imputation three steps need to be followed. First, the dataset is imputed, but multiple times, similar to single imputation. Second, an analysis is performed for each imputed dataset, resulting in the same number of analyses as imputed datasets. Finally, the results of all the analyses are being pooled together in an aggregating way. Similar to single imputation, there are different approaches to performing multiple imputation. Two different approaches were selected as the nature of the two datasets that were used in this thesis was different.

3.3.4.1 Multivariate imputation by chained equations

Multivariate imputation by chained equations (MICE) is a multiple imputation method that was used to impute the missing values for the analysis in chapter 5.[132] Chapter 5 uses data from the High-STEACS registry. Certain assumptions need to be fulfilled for this method, such as that the data are missing at random or completely at random. Using linear regression for continuous variables and logistic regression for categorical variables, MICE imputes one variable each time for all the multiple datasets, using the rest of the variables to make predictions about the missing values of the variable's that is being imputed. Then repeats the process for the rest of the variables.

3.3.4.2 Joint modelling multiple imputation

Chapter 6 and chapter 7 use individual patient data (IPD) meta-analysis data for the analyses; hence a MICE method would not be appropriate for this kind of data. A joint modelling multiple imputation method has been selected to account for missing data across studies in the training cohort for the models. Multiple datasets were imputed using joint-modelling multiple imputation with random study specific covariance matrices fitted with a Markov chain Monte Carlo algorithm.[133]

3.4 Evaluation of diagnostic performance

3.4.1 Confusion matrix

The confusion matrix is a 2x2 table that contains four outputs provided by the binary classification model. The diagnostic performance of each model can be evaluated using this 2x2 table (Table 3.1), illustrating the agreement between the model and the outcome.

Table 3.1 Confusion matrix.

		Predicted values	
		Negative -	Positive +
Actual values	Negative -	True Negatives	False Positives
	Positive +	False Negatives	True Positives

Many diagnostic metrics can be derived for this table; however, four measures will be derived from the confusion matrix to evaluate and compare models and pathways. These are the sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) and are used throughout this thesis (chapter 4 to 7).

Sensitivity, also known as true positive rate or recall, is the proportion of actual true positive predictions over all actual positive conditions.

$$Sensitivity = \frac{true\ positive}{true\ positive + false\ negative} = \frac{TP}{TP + FN}. \quad (3.7)$$

Specificity, also known as true negative rate or selectivity, is the proportion of actual true negative predictions over all actual negative conditions.

$$Specificity = \frac{true\ negative}{true\ negative + false\ positive} = \frac{TN}{TN + FP}. \quad (3.8)$$

NPV is the proportion of actual true negative predictions over all negative predictions.

$$NPV = \frac{true\ negative}{true\ negative + false\ negative} = \frac{TN}{TN + FN}. \quad (3.9)$$

PPV, also known as precision, is the proportion of actual true positive predictions over all positive predictions.

$$PPV = \frac{true\ positive}{true\ positive + false\ positive} = \frac{TP}{TP + FP}. \quad (3.10)$$

When it comes to rule out patients, i.e., those with a score lower than a predefined threshold, sensitivity and NPV will be assessed. For ruling in patients, i.e., those with a score greater than or equal to a high predefined threshold, specificity and PPV will be assessed. (Figure 3.2)

For the selection of each threshold, the criterion is to maximise the percentage of patients being ruled out and ruled in, leaving as few patients as possible in the intermediate group, where further testing and assessment is required. This approach was followed throughout every chapter of this thesis.

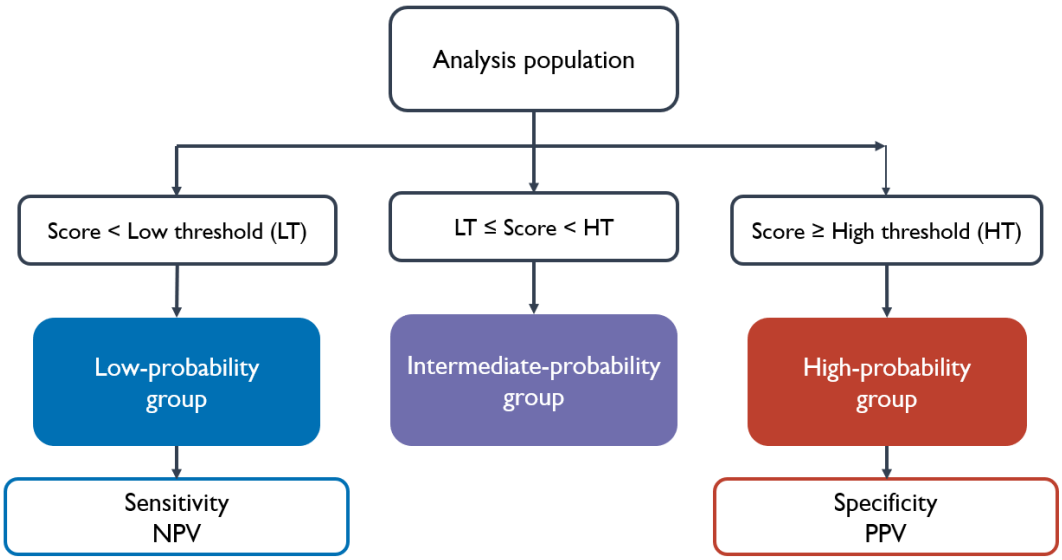


Figure 3.2 Example of how the classification pathway was developed and assessed.

3.4.2 ROC curve

A receiver operating characteristic (ROC) curve (Figure 3.3) shows the trade-off between sensitivity and specificity of a classification model at various thresholds. Models with curves that are closer to the top-left corner perform better. The diagonal line is used as a baseline and shows how a random model would perform. One of the important characteristics of the ROC curve is that it does not depend on the prevalence since it is based on sensitivity and specificity. Both sensitivity and specificity are probabilities conditioned on the true class label. Therefore, they will be the same regardless of the prevalence. As a result, this makes it useful for evaluating classification models used in medicine which are predicting diseases.

Area Under the ROC Curve (AUC) measures the area underneath the entire ROC curve. The AUC provides an overall measure of performance across all possible classification thresholds, and it is used to compare the predictive accuracy of different models. However, it doesn’t take into account the different misclassification costs arising from false-negative and false-positive diagnoses.

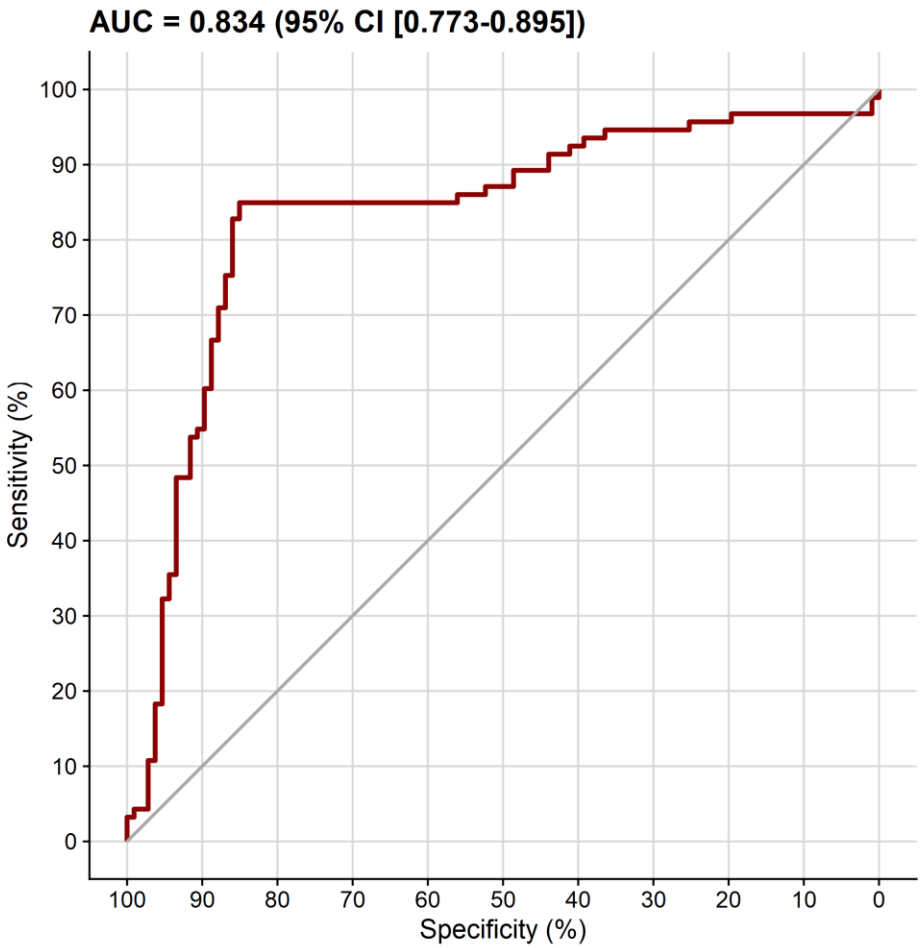


Figure 3.3 Example of a ROC curve with its AUC.

When the dataset that is used is imbalanced, i.e., has a low prevalence of the disease, it is preferred to use a precision-recall curve instead of a ROC curve, as it shows the trade-off between PPV (precision) and sensitivity (recall) of a classification model at various thresholds (Figure 3.4)]. Similarly to the AUC, we can summarise the information in a precision-recall curve with the area under the precision-recall curve (AUC-PR). Both AUC and AUC-PR can take values between zero and one, with one to indicate the best performance. However, a model with high discrimination is not necessarily well calibrated.

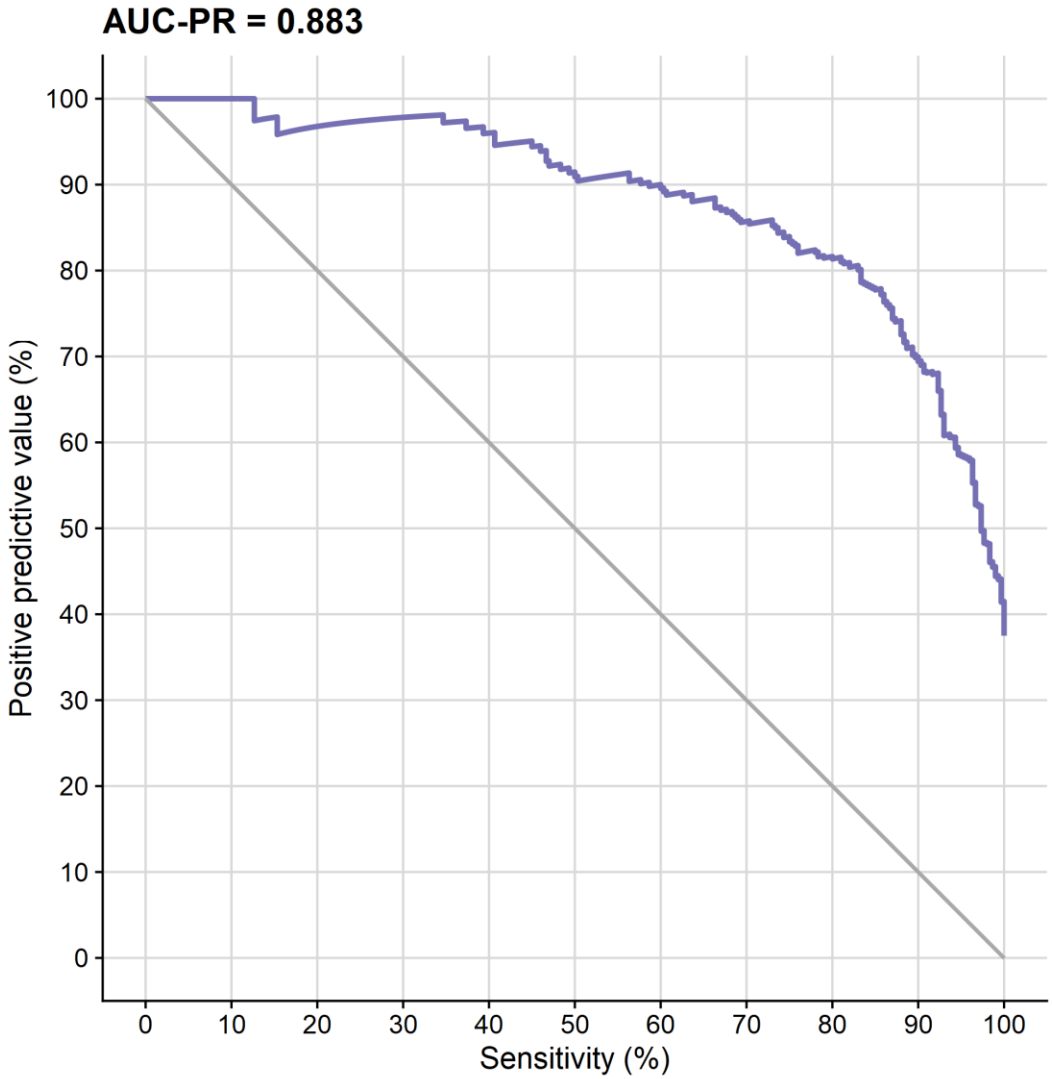


Figure 3.4 Example of a precision-recall curve with its AUC-PR.

3.4.3 Calibration

Throughout this thesis, instead of directly predicting if a patient has the disease or not, the likelihood of a patient having the disease is predicted. In this way, probabilities can be interpreted and provide more sophisticated ways to evaluate the model's skill and reflect on the risk continuum.

A calibration plot (Figure 3.5) has a diagonal line representing perfect calibration. The y-axis has the observed relative frequency of the outcome and on the x-axis is the model's predicted probabilities. The position of the points relative to the diagonal can help to interpret the probabilities; for example:

- Below the diagonal line: The model is over-diagnosing the disease.
- Above the diagonal line: The model is under-diagnosing the disease.

Brier score is an evaluation metric for assessing the quality of a model's calibration overall. The lower the Brier score is for a set of predictions, the better the predictions are calibrated. Note that the Brier score, in its most common formulation, takes on a value between zero and one as it is very similar to the mean squared error, except that it only applies to prediction probability scores with values ranging from zero to one.

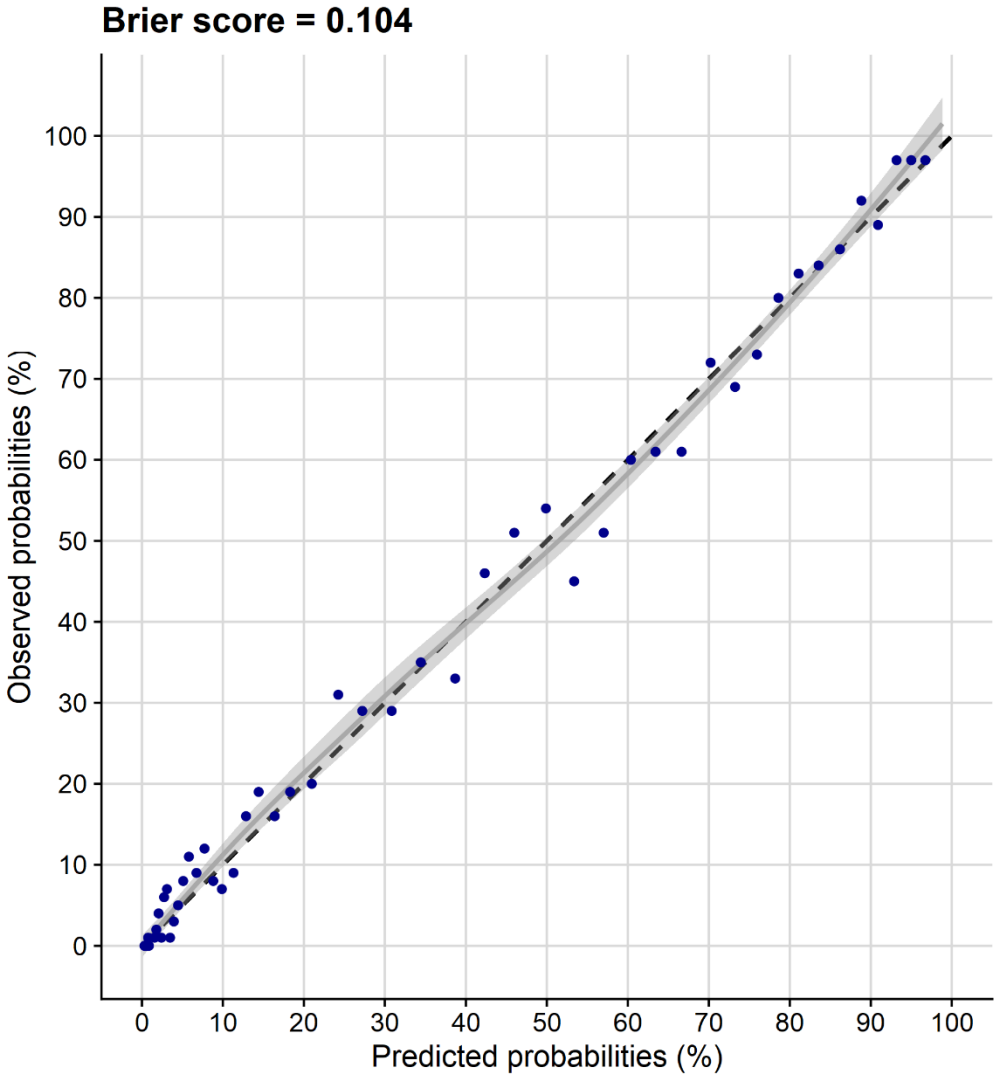


Figure 3.5 Example of calibration plot with a loess line and the Brier score.

3.4.4 Meta-regression

Chapter 6 and chapter 7 use individual patient data (IPD) meta-analysis data for the analyses. To evaluate the heterogeneity between the different studies selected, we are going to use an extension of meta-analysis, called meta-regression. In concept, meta-regression is very similar to a subgroup analysis with continuous variables where these variables reveal if there is a linear association with the student effect size similar in essence to simple regressions, in which an outcome variable is predicted according to the values of one or more explanatory variables. In this thesis, we are going to evaluate if the prevalence is associated with the guideline recommended thresholds performance to rule-out and rule-in the diagnosis of acute heart failure (see Chapter 6).

3.4.5 Algorithm stochasticity

Many machine learning models are referred to as “stochastic” models based on their behaviour and their performance. Stochastic refers to a changeable process where the result contains some element of uncertainty and randomness. It is a mathematical phrase that is connected to "randomness," "probabilistic," and can be used to contrast with "deterministic" thinking.

To adjust for this randomness in the model behaviour and to increase the reproducibility of the analysis code when shared with other researchers, we have set the random seed in R to be the same before any stochastic process. These stochastic processes do not necessary have to do the model development, but also any other part of the methodology when randomness is introduced, for example, in the k-fold cross validation.

Setting a seed in R means to initialize a pseudorandom number generator. This helps with having the exact same results every time the code runs, while it other circumstances the results could be slightly different (Table *3.2*).

Table 3.2 Example of the random seed in model's performance for the different diagnostic metrics.

Random seed	Rule-out		Rule-in	
	Sensitivity	**NPV**	**Specificity**	**PPV**
123	95.5% (91.8-98.5%)	99.6% (99.8-99.9%)	80.7% (79.1-82.3%)	83.5% (82.0-84.8%)
1234	94.7% (90.7-98.3%)	99.5% (99.1-99.8%)	80.2% (78.5-81.8%)	83.1% (81.6-84.5%)
789	95.5% (92.0-98.5%)	99.6% (99.8-99.9%)	80.5% (79.0-82.3%)	83.2% (81.8-84.7%)
1	94.7% (90.7-98.2%)	99.5% (99.1-99.9%)	80.1% (78.6-81.7%)	83.0% (81.6-84.4%)

3.5 Decision support tools

Shiny is an R package that can be used to create interactive web applications straight from R. Users can host and control Shiny applications online thanks to Shiny Server. A reactive programming style is used by the Shiny package to streamline the creation of web apps that use R in their core. R processes running separate Shiny applications on various URLs and ports can be managed by the Shiny Server. The advantages of using Shiny Server over just running Shiny in R are numerous. Some of the advantages are the following:

- Host numerous web applications at once, each with its own URL.
- Support web browsers that don't support WebSocket, including Internet Explorer 8 & 9.
- Allow users of the system to create and maintain their own Shiny applications.
- Make sure that any R processes that crash or are terminated restart for the user who requests the application the following time.

Shiny Server is an open-source back-end program that makes a significant difference. The most important one, especially when it comes to applications that work with patient sensitive information, is that both the Shiny app and any data it requires can be deployed in way that they will not ever leave the University of Edinburgh controlled settings environment.

To go beyond what Shiny offers and extend the capabilities and appearance of the app, we incorporated HTML and CSS code to customise the user interface of the apps.

Below are screenshots of the interface from the app that we created for the CoDE-HF algorithm. Each app can be accessed both from a smartphone (Figure 3.6) and a laptop/desktop (Figure 3.7). CoDE-HF can be accessed via https://decision-support.shinyapps.io/code-hf/ while CoDE-ACS can be accessed via https://decision-support.shinyapps.io/code-acs/.

Each app has two extra tabs: Settings and About. The ‘Settings’ tab allows different healthcare facilities to set their own clinical settings regarding ruling out and ruling in patients. This more flexible approach could be more conservative (higher NPV) or tailored to limit the observe zone or abolish it completely. The ‘About’ tab contains useful information about each app itself. These are current guidance for use, the evidence section, which refers to the publication, terms and conditions if you want to use the app, the funders of this project (British Heart Foundation and Medical Research Council) and contact information where users could report bugs and issues.

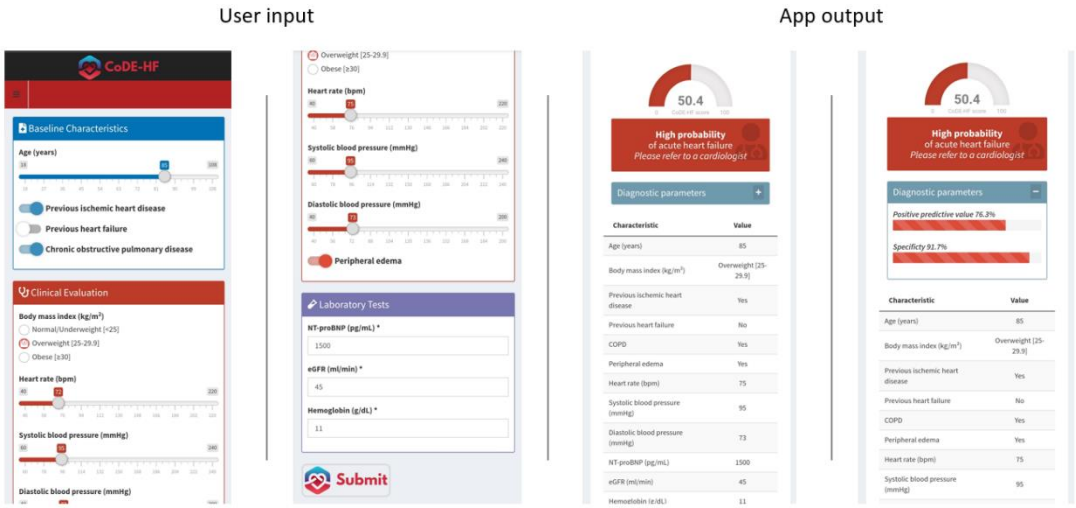


Figure 3.6 The user interface, with an example, of the CoDE-HF app accessed from a phone.

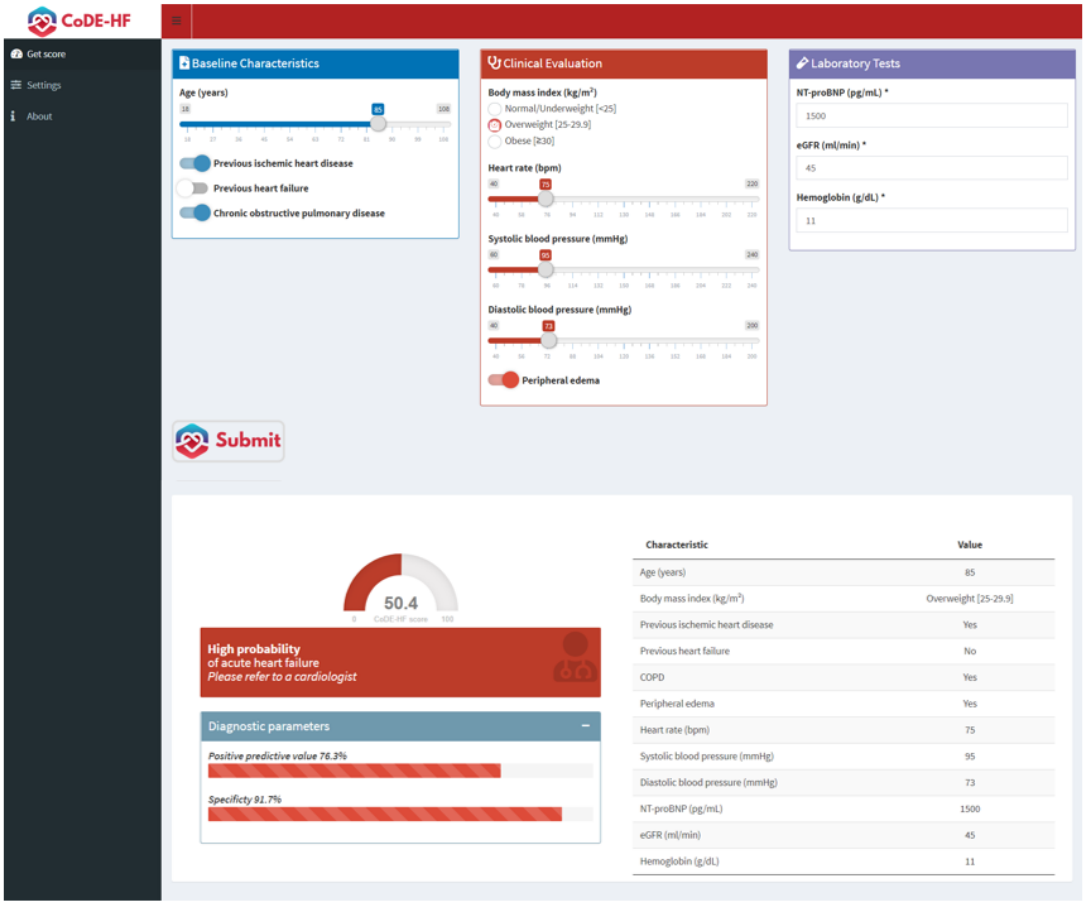


Figure 3.7 The user interface, with an example, of the CoDE-HF app accessed from a laptop/desktop.

Having in mind the usability and user experience, we applied a user-centered approach and an adaptive design. We conducted qualitative and quantitative studies to evaluate how clinicians use existing clinical guidance in the assessment of patients with suspected acute heart failure and myocardial infarction. This has given us insights that aid the initial prototype development.

To further improve in those areas, we will work with potential clinical users of the CoDE-HF and CoDE-AC apps to design the interface using dummy data as case studies. Performance testing will then move to a studio setting where clinicians will use CoDE-HF and CoDE-AC to assess patient vignettes. During these tasks clinicians will be asked to ‘think aloud’ while interacting with the apps and the patient vignette. An observer will record notes real time during this task analysis event revealing how the tool fits the task, highlighting why clinicians may perform certain actions, and revealing any knowledge gaps in use of the tool or barriers to its use in clinical assessment.

Last, to minimise human input errors and improve the utility of the developed decision support tools, we have secured funding from a BHF Translational Award to develop a product that integrates into the NHS clinical workflow and collects patient and user feedback.

The next chapter is the first chapter that was published during the PhD period. It is about a previous developed algorithm, and how it performs in data that represent clinical practice. Previous evaluation of this algorithm was in selected small datasets which could not answer the key question about the generalisability of the algorithm.

CHAPTER 4

Validation of the myocardial-ischemic-injury-index (MI^3) machine learning algorithm to guide the diagnosis of myocardial infarction in a heterogeneous population

Doudesis D, Lee KK*, Yang J, et al. Validation of the myocardial-ischaemic-injury-index machine learning algorithm to guide the diagnosis of myocardial infarction in a heterogenous population: a prespecified exploratory analysis. The Lancet Digital Health 2022;4(5):e300-e308.*

DOI: 10.1016/s2589-7500(22)00025-5.

* Contributed equally

Chapter 4 Validation of the myocardial-ischemic-injury-index (MI^3) machine learning algorithm to guide the diagnosis of myocardial infarction in a heterogeneous population

4.1 Overview

We recently introduced the myocardial-ischemic-injury-index (MI^3), a machine learning algorithm that predicts the likelihood of myocardial infarction in patients with suspected acute coronary syndrome. Whether this algorithm performs well in routine clinical practice or predicts subsequent events is unknown.

MI^3 was validated in a prespecified exploratory analysis from a multi-centre randomised trial that included consecutive patients with suspected acute coronary syndrome undergoing serial high-sensitivity cardiac troponin I measurement. Patients with ST-segment elevation myocardial infarction were excluded. MI^3 incorporates age, sex, and two troponin measurements to compute a value (0-100) reflecting an individual's likelihood of myocardial infarction during the index visit and estimates diagnostic performance metrics at the computed score. Model performance for an index diagnosis of myocardial infarction, and for subsequent myocardial infarction or

cardiovascular death at one year was determined using previously defined low- and high-probability MI[3] thresholds (1.6 and 49.7, respectively).

In total, 20,761 patients (64±16 years, 46% women) were included of whom 3,272 (15.8%) had myocardial infarction. MI[3] had an area under the receiver-operating-characteristic curve of 0.949 (95% confidence interval 0.946-0.952) identifying 12,983 (62.5%) patients as low-probability (sensitivity 99.3% [99.0-99.6%], negative predictive value 99.8% [99.8-99.9%]), and 2,961 (14.3%) as high-probability (specificity 95.0% [94.6-95.3%], positive predictive value 70.4% [68.7-72.0%]). At one year, subsequent myocardial infarction or cardiovascular death occurred more often in high-probability compared to low-probability patients (17.6% [520/2,961] *versus* 1.5% [197/12,983], P<0.001).

In consecutive patients undergoing serial cardiac troponin measurement for suspected acute coronary syndrome, the MI[3] algorithm accurately estimates the likelihood of myocardial infarction and predicts subsequent adverse cardiovascular events. In consecutive patients undergoing serial cardiac troponin measurement for suspected acute coronary syndrome, the MI[3] algorithm accurately estimates the likelihood of myocardial infarction and predicts subsequent adverse cardiovascular events.

4.2 Introduction

Myocardial infarction is a condition characterised by myocardial necrosis secondary to acute myocardial ischaemia, and is the most common cause of death worldwide.[134] In recognition of this, guidelines emphasise the importance of early diagnosis and treatment to reduce mortality, and clinicians have a low threshold for referring patients for further investigation.[135] Although, patients with suspected myocardial infarction account for 1 in 20 attendances in the Emergency Department[11], the diagnosis is ultimately ruled out in 80 to 90% of patients.[69,136]

Accelerated diagnostic pathways aim to promote earlier discharge in those considered low-risk and improve the targeting of treatment to high-risk patients.[137-140] However, these pathways have some limitations. First, they use fixed cardiac troponin thresholds for all patients, which do not account for age or comorbidities that are known to influence troponin concentrations.[140,141] Second, they are based on fixed time-points for serial testing, which can be challenging in a busy Emergency Department, and such pathways may not be generalisable to all health care systems. Third, up to 1 in 3 patients are neither ruled-out, nor ruled-in, using these pathways and questions often remain for these individuals. For example, how likely were my symptoms due to a heart attack and would I benefit from further testing?

The myocardial-ischemic-injury-index (MI^3) is an algorithm developed using the machine learning technique, gradient boosting, to compute an

individualised probability of myocardial infarction on a scale of 0-100 for patients with suspected acute coronary syndrome.[142] The MI^3 score is computed using age, sex, cardiac troponin concentration, and the rate of change in troponin concentration when re-measured at a second flexible time point. Whilst the algorithm performed well when validated in data pooled from seven diagnostic cohort studies, it has not been evaluated in a more heterogeneous patient population, where a greater burden of comorbid conditions may impact on performance. Furthermore, it is not clear whether it provides information about cardiovascular risk beyond the initial diagnosis.

In consecutive patients with suspected acute coronary syndrome, we evaluate whether MI^3 can predict the index diagnosis of myocardial infarction and risk of subsequent myocardial infarction or cardiovascular death at one year.

4.3 Methods

4.3.1 Study population and approvals

High-Sensitivity Troponin in the Evaluation of patients with suspected Acute Coronary Syndrome (*High-STEACS*) is a stepped-wedge cluster randomised controlled trial that evaluated the implementation of a high-sensitivity cardiac troponin I assay in consecutive patients with suspected acute coronary syndrome, across ten secondary and tertiary care hospitals in Scotland.[143] All patients with suspected acute coronary syndrome attending the Emergency Department were identified by the attending clinician at the time troponin was requested, using an electronic form integrated into the clinical care pathway. For this prespecified exploratory analysis of the trial, patients were eligible for inclusion if they presented with suspected acute coronary syndrome and had at least two serial cardiac troponin measurements. Patients were excluded if they presented with ST-segment elevation myocardial infarction, since they were not included in the original development of the algorithm, or if there was insufficient clinical information to adjudicate the diagnosis.

The High-STEACS trial was approved by the Scotland A Research Ethics Committee, the Public Benefit and Privacy Panel for Health and Social Care, and by each National Health Service (NHS) Health Board. As randomisation was at the hospital level, consent was not sought from individual patients. All data were collected prospectively from the electronic patient record, deidentified and linked in a data repository within a secure NHS Safe Haven

(DataLoch, Edinburgh, United Kingdom). This exploratory analysis was pre-specified in the trial protocol, however, due to its observational nature the statistical analysis plan was not reviewed by the trial steering committee.

4.3.2 Cardiac troponin

Cardiac troponin testing was performed at presentation and was repeated 6 or 12 hours after the onset of symptoms at the discretion of the attending physician and in accordance with national guidelines.[144] All patients had troponin measured using the investigational high-sensitivity assay (ARCHITECTSTAT high-sensitive troponin I assay; Abbott Laboratories, Abbott Park, IL, USA) throughout the trial, but this was only used to guide clinical decisions in the implementation phase. This assay has an inter-assay coefficient of variation of less than 10% at 4.7 ng/L[145], and a 99th centile upper reference limit of 34 ng/L in men and 16 ng/L in women.[145]

4.3.3 Adjudication of myocardial infarction and outcomes

All diagnoses in patients with high-sensitivity cardiac troponin I concentrations above the 99th centile were adjudicated and classified according to the Third Universal Definition of Myocardial Infarction as previously described.[143] In brief, two physicians independently reviewed all clinical information with discordant diagnoses resolved by a third reviewer. Type 1 myocardial infarction was defined as myocardial necrosis (any high-sensitivity cardiac troponin I concentration above the sex-specific 99th centile

with a rise and/or fall in hs-cTnI concentration where serial testing was performed) in the context of a presentation with suspected acute coronary syndrome with symptoms or signs of myocardial ischemia on the electrocardiogram. Patients with myocardial necrosis, symptoms or signs of myocardial ischemia, and evidence of increased myocardial oxygen demand or decreased supply secondary to an alternative condition without evidence of acute atherothrombosis were defined as type 2 myocardial infarction. Type 4b myocardial infarction was defined where myocardial ischaemia and myocardial necrosis were associated with stent thrombosis documented at angiography. We used regional and national registries to ensure complete follow-up for the trial population.[143]

The primary and key secondary outcomes were myocardial infarction (type 1 or type 4b) during the index visit, subsequent myocardial infarction (type 1 or type 4b) or cardiovascular death at one year, and all-cause death at one year, respectively.

4.3.4 MI^3 algorithm

MI^3 is an algorithm derived using the machine learning technique, gradient boosting. It computes a value of 0 to 100 for each patient using their age, sex, serial cardiac troponin concentrations and the time interval between sampling which corresponds to an individualised estimate of the likelihood of a diagnosis of type 1 myocardial infarction.[142]

4.3.5 Statistical analysis

Model discrimination and calibration was assessed by calculating the area under the receiver-operating-characteristic curve (AUC), and by visual inspection of the precision-recall and calibration curve, respectively. Diagnostic performance was evaluated using previously defined low-probability and high-probability thresholds (MI^3 scores of 1.6 and 49.7, respectively).[142] These thresholds were defined in the cohort used to train the algorithm based on pre-specified performance criteria (sensitivity ≥99.0% and negative predictive value (NPV) ≥99.5% for low-probability specificity ≥90% and positive predictive value (PPV) ≥75% for high-probability). We report the sensitivity, specificity, NPV, and PPV for these thresholds, along with 95% confidence intervals calculated using 1,000 bootstrapped samples. Subgroup analysis was performed by age, sex, presenting symptom of chest pain, prior ischaemic heart disease, myocardial infarction, diabetes mellitus and cerebrovascular disease and stratified by renal function and the time from symptom onset to presentation. MI^3 performance was also validated based on the time interval between blood sampling (<3 hours, 3 to 6 hours, and >6 hours). In an exploratory analysis, we evaluated diagnostic performance for a composite endpoint of type 1 or type 2 myocardial infarction during the index hospital admission. Survival free from subsequent myocardial infarction or cardiovascular death at one year, or death from any cause at one year was determined in patients grouped according to their MI^3 score (low-probability <1.6; intermediate-probability 1.6 to 49.6; high-probability ≥49.7). All

analyses were conducted using R (version 3.6.3; The R Foundation for Statistical Computing).

4.4 Results

Of 48,282 patients in the High-STEACS trial, 20,761 patients (64±16 years, 46% women) were included in this analysis (Figure 4.1). There were no differences in sex distribution, symptoms, or laboratory markers, including cardiac troponin concentrations, but the analysis population was on average 3 years older than the trial population, and patients were more likely to have a prior history of ischemic heart disease and to be established on preventative medication (Table *4.1*). In the analysis population, cardiac troponin concentrations were above the 99th centile in 27.9% (5,788/20,761) of patients at presentation or on serial testing. The adjudicated diagnosis was type 1 myocardial infarction in 15.8% (3,272/20,761) and type 2 myocardial infarction in 4.4% (916/20,761).

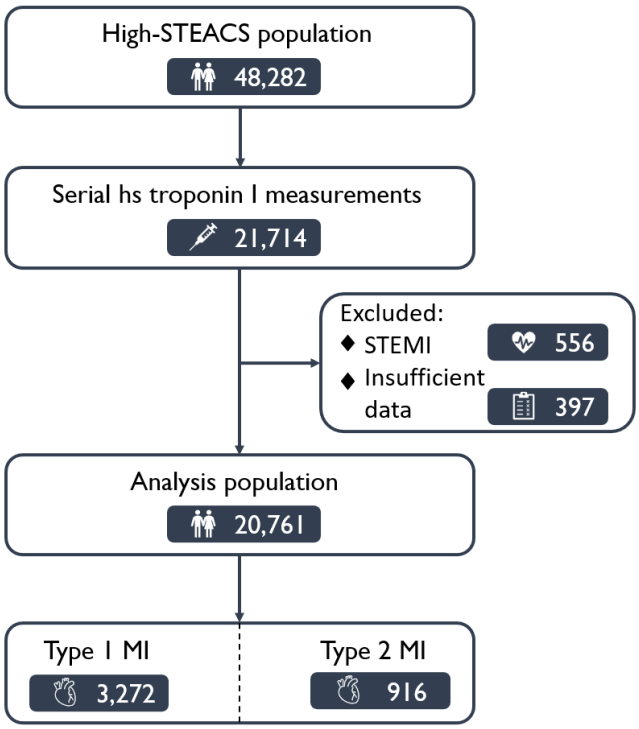


Figure 4.1 Flow diagram of the analysis population.

Table 4.1 Baseline characteristics of the analysis population stratified by MI^3 probability score.

	High-STEACS		MI^3 group		
	All participants	**Analysis population**	**Low probability**	**Intermediate probability**	**High probability**
Number of participants	48,282	20,761	12,983 (63%)	4,817 (23%)	2,961 (14%)
Age, years	61 (±17)	64 (±16)	59 (±15)	72 (±14)	69 (±14)
Sex					
Women	22,562 (47%)	9,597 (46%)	6,241 (48%)	2,225 (46%)	1,131 (38%)
Men	25,720 (53%)	11,164 (54%)	6,742 (52%)	2,592 (54%)	1,830 (62%)
Presenting complaint*					
Chest pain	34,540 (81%)	15,878 (85%)	10,430 (91%)	3,291 (76%)	2,157 (79%)

Dyspnoea	2,175 (5%)	709 (4%)	171 (2%)	326 (8%)	212 (8%)
Palpitation	1,269 (3%)	336 (2%)	164 (1%)	131 (3%)	41 (2%)
Syncope	2,495 (6%)	868 (5%)	393 (3%)	335 (8%)	140 (5%)
Other	2,188 (5%)	706 (4%)	309 (3%)	227 (5%)	170 (6%)
Previous medical conditions					
Myocardial infarction	4,214 (9%)	2,504 (12%)	1,317 (10%)	777 (16%)	413 (14%)
Ischaemic heart disease	11,912 (25%)	6,746 (32%)	3,666 (28%)	2,126 (44%)	954 (32%)
Cerebrovascular disease	2,949 (6%)	1,414 (7%)	624 (5%)	564 (12%)	226 (8%)
Diabetes mellitus	3,518 (7%)	1,960 (9%)	781 (6%)	687 (14%)	492 (17%)
Previous revascularisation					

Percutaneous coronary intervention	3,682 (8%)	2,229 (11%)	1,330 (10%)	597 (12%)	302 (10%)
Coronary artery bypass grafting	782 (2%)	446 (2%)	216 (2%)	163 (3%)	67 (2%)
Medications at presentation					
Aspirin	13,163 (27%)	7,021 (34%)	3,934 (30%)	1,993 (41%)	1,094 (37%)
Dual anti-platelet therapy†	1,605 (3%)	965 (5%)	515 (4%)	298 (6%)	152 (5%)
Statin	19,366 (40%)	9,957 (48%)	5,609 (43%)	2,819 (59%)	1,529 (52%)
ACE inhibitor or ARB	15,618 (32%)	7,948 (38%)	4,390 (34%)	2,292 (48%)	1,266 (43%)
Beta-blocker	13,173 (27%)	6,804 (33%)	3,844 (30%)	1,898 (39%)	1,062 (36%)
Oral anticoagulant‡	3,253 (7%)	1,529 (7%)	663 (5%)	650 (13%)	216 (7%)

Haematology and clinical chemistry measurements					
Haemoglobin, g/L	136 (±22)	135 (±21)	137 (±20)	130 (±24)	134 (±24)
Estimated glomerular filtration mL/min	54 (±13)	54 (±12)	57 (±9)	50 (±14)	49 (±15)
Peak high-sensitivity cardiac troponin I, ng/L	4 [4]	5 [16]	2 [3]	19 [29]	133 [534]

Values are mean ± standard deviation; median [interquartile range]; n (%).
*A presenting symptom was missing in 2,264 (11%). †Two medications from aspirin, clopidogrel, prasugrel, or ticagrelor. ‡Includes warfarin or novel oral anticoagulants.
ACE, angiotensin converting enzyme; ARB, angiotensin receptor blockers.

4.4.1 MI^3 calibration and discrimination

The MI^3 algorithm performed well overall with an AUC of 0.949 (95% confidence interval [CI] 0.946 to 0.952) with excellent discrimination between those with and without type 1 myocardial infarction. Discrimination was similar in patients evaluated during the validation and implementation phases of the trial (AUC 0.949 [95% CI 0.944 to 0.954] versus 0.948 [95% CI 0.945 to 0.952]). However, calibration was not good in patients with intermediate MI^3 scores. MI^3 scores between 10 and 40 underestimated the observed risk whilst scores between 65 and 86 overestimated the observed risk (Figure 4.2). MI^3 identified 62.5% (12,983/20,761) of patients as low-probability for type 1 myocardial infarction at the pre-specified threshold (MI^3 score <1.6), with a sensitivity and NPV of 99.3% (99.0 to 99.6%) and 99.8% (99.8 to 99.9%), respectively (Figure 4.3, Table *4.2*). MI^3 identified 14.3% (2,961/20,761) of patients as high-probability at the pre-specified threshold (MI^3 score ≥49.7), with a specificity and PPV of 95.0% (94.6 to 95.3%) and 70.4% (68.7 to 72.0%), respectively (Figure 4.3 and Table *4.2*).

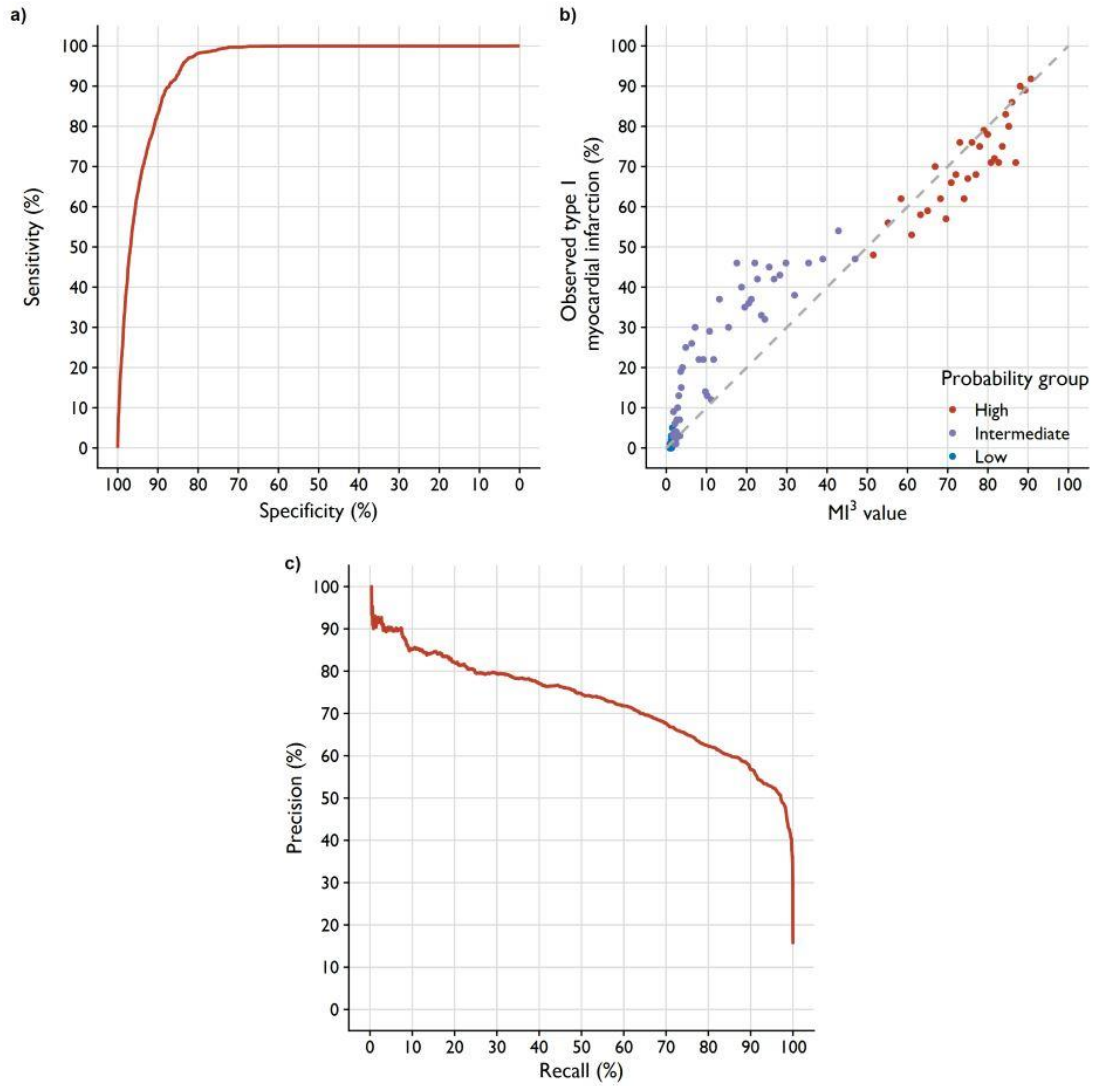


Figure 4.2 (a) Receiver-operating-characteristic (ROC) curve illustrating discrimination of the MI^3 algorithm for type 1 myocardial infarction. (b) Calibration of the MI^3 algorithm with the observed proportion of patients with type 1 myocardial infarction. The dashed line represents perfect calibration. Each point represents 100 patients. (c) Precision recall curve illustrating discrimination of the MI^3 algorithm for type 1 myocardial infarction.

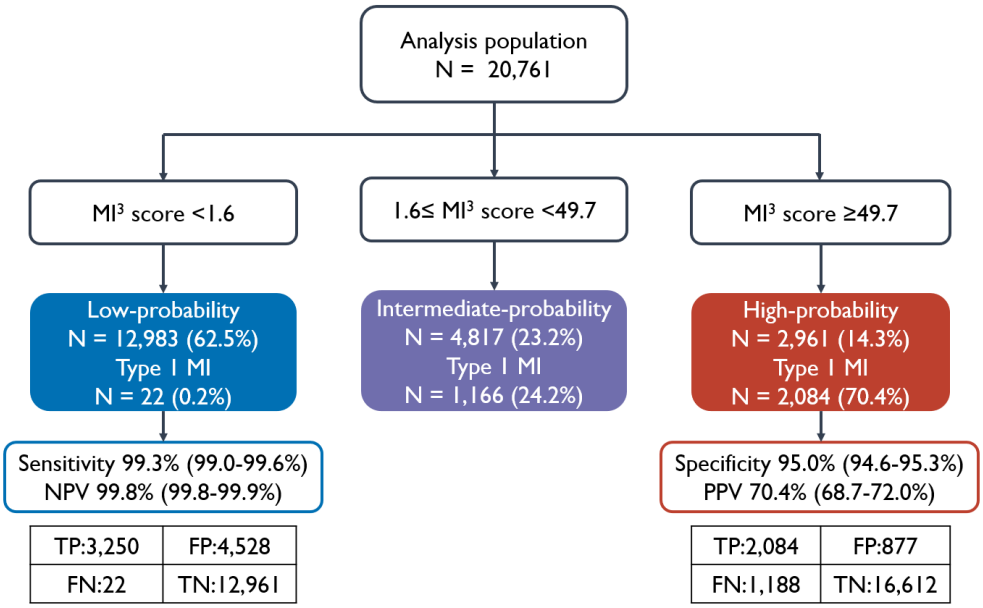


Figure 4.3 Performance of MI³ at example thresholds.
TP = true positive, FP = false positive, FN = false negative, TN = true negative, NPV = negative predictive value, PPV = positive predictive value.

Table 4.2 Performance of MI^3 at example thresholds. Primary outcome: type 1 myocardial infarction.
TP = true positive, FP = false positive, FN = false negative, TN = true negative, NPV = negative predictive value, PPV = positive predictive value.

	Example MI^3 threshold	**TN**	**FN**	**TP**	**FP**	**Sensitivity**	**NPV**	**Specificity**	**PPV**
Low probability	1.6	12,961	22	3,250	4,528	99.3% (99.0-99.6%)	99.8% (99.8-99.9%)	74.1% (73.5-74.8%)	41.8% (41.2-42.4%)
High Probability	49.7	16,612	1,188	2,084	877	63.7% (62.2-65.3%)	93.3% (93.1-93.6%)	95.0% (94.6-95.3%)	70.4% (68.7-72.0%)

The AUC differed when stratifying patients by age, sex, presenting symptom of chest pain, renal function, prior ischaemic heart disease, myocardial infarction, diabetes mellitus and cerebrovascular disease, while there was no difference when stratifying by time from symptom onset to presentation (Figure 4.4). There was no heterogeneity in sensitivity or NPV for the low-probability threshold (Figure 4.5, Figure 4.6), while, in some groups, there was significant heterogeneity in the specificity and PPV for the high-probability threshold (Figure 4.7, Figure 4.8). In particular, the PPV for the high-probability threshold was significantly higher in patients with a primary presenting symptom of chest pain compared to those with other presenting symptoms (80.2% [78.7 to 81.8%] versus 34.7% [32.1 to 37.5%]).

MI^3 performed similarly in patients who had serial measurements within 3 hours (2,439 [11.8%], AUC 0.972 [95% CI 0.966 – 0.979]) or between 3 and 6 hours (4,098 [19.7%], AUC 0.968 [95% CI 0.963 to 0.973]). The MI^3 algorithm also performed well in patients with serial measurements >6 hours apart, but discrimination was lower in this group (14,224 [68.5%], AUC 0.939 [95% CI 0.935 to 0.942]) (Figure 4.9).

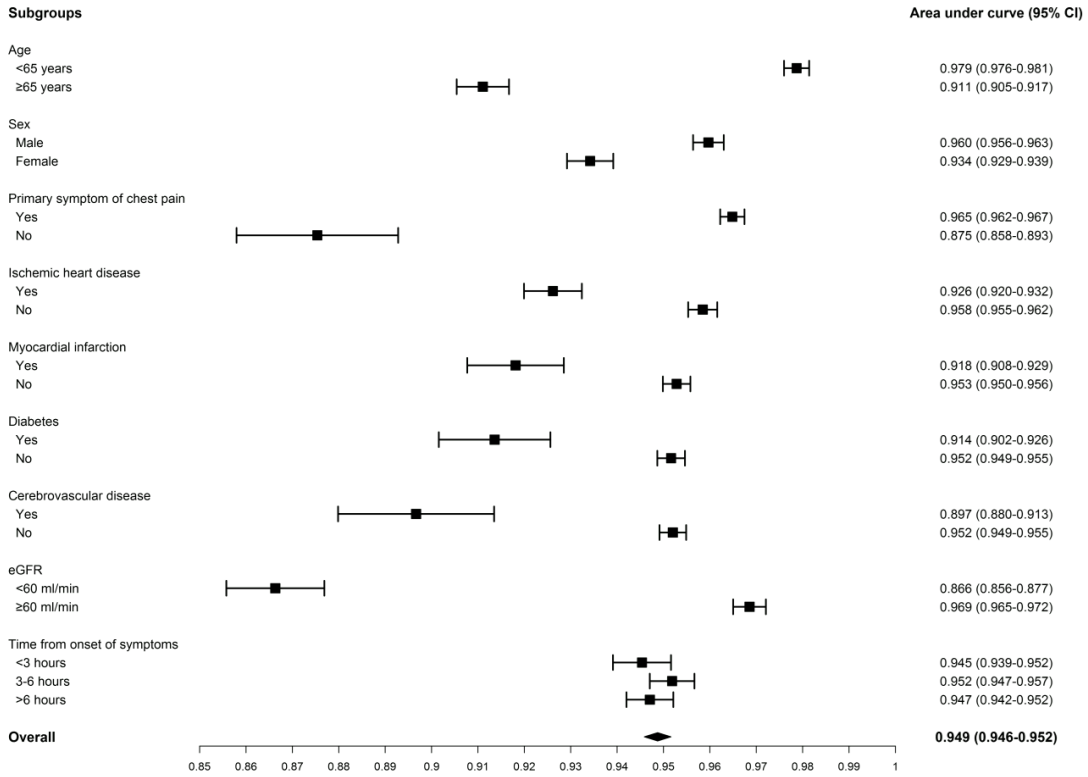


Figure 4.4 Forest plot of the discrimination (AUC) across patient subgroups. AUC = area under curve.

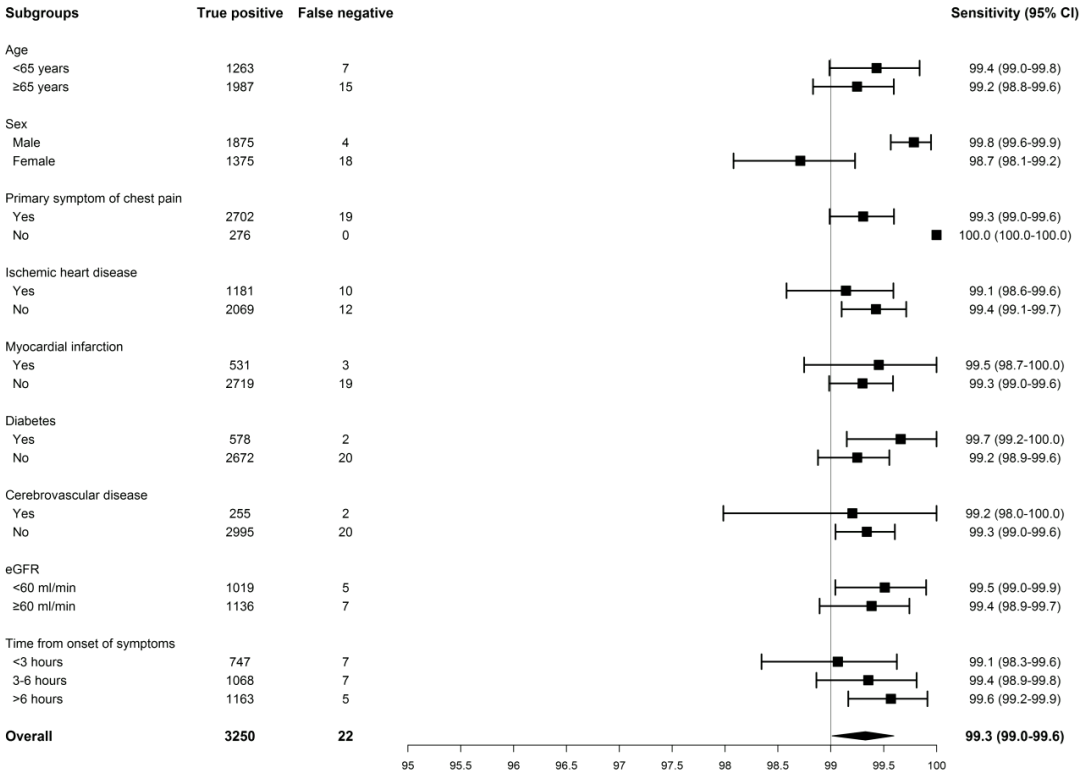


Figure 4.5 Forest plot of the sensitivity across patient subgroups.
TP = true positive, FN = false negative. The vertical line represents the performance target.

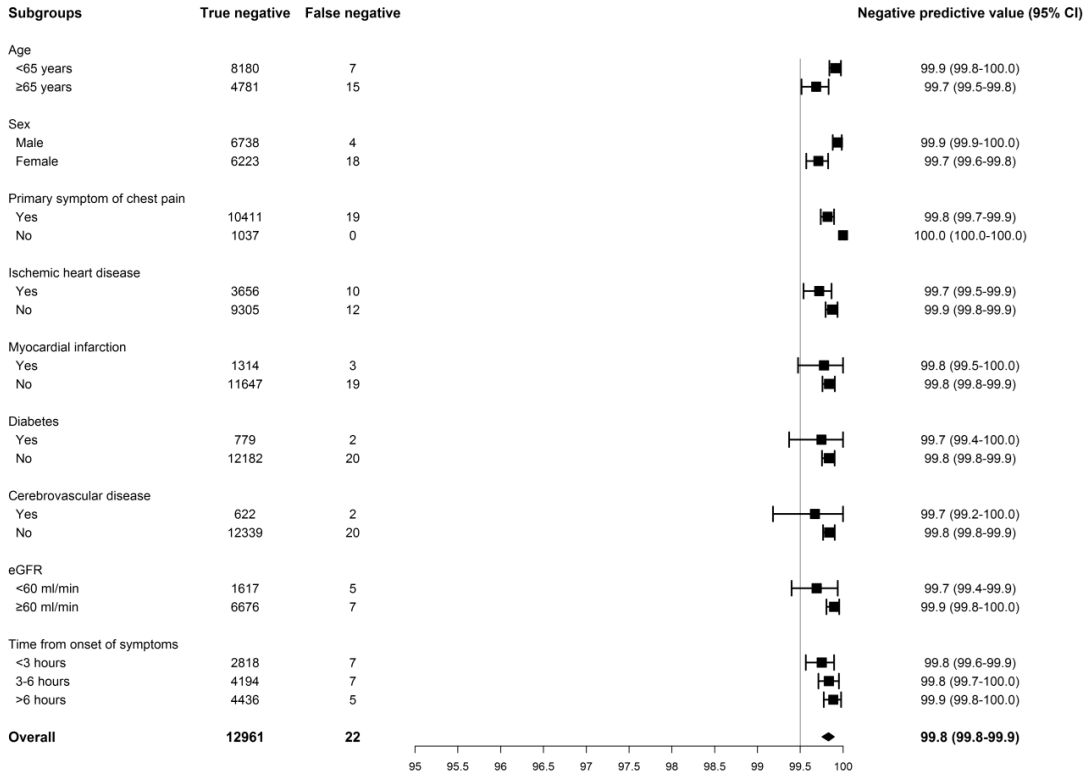


Figure 4.6 Forest plot of the negative predictive value across patient subgroups.
NPV = negative predictive value, TN = true negative, FN = false negative. The vertical line represents the performance target.

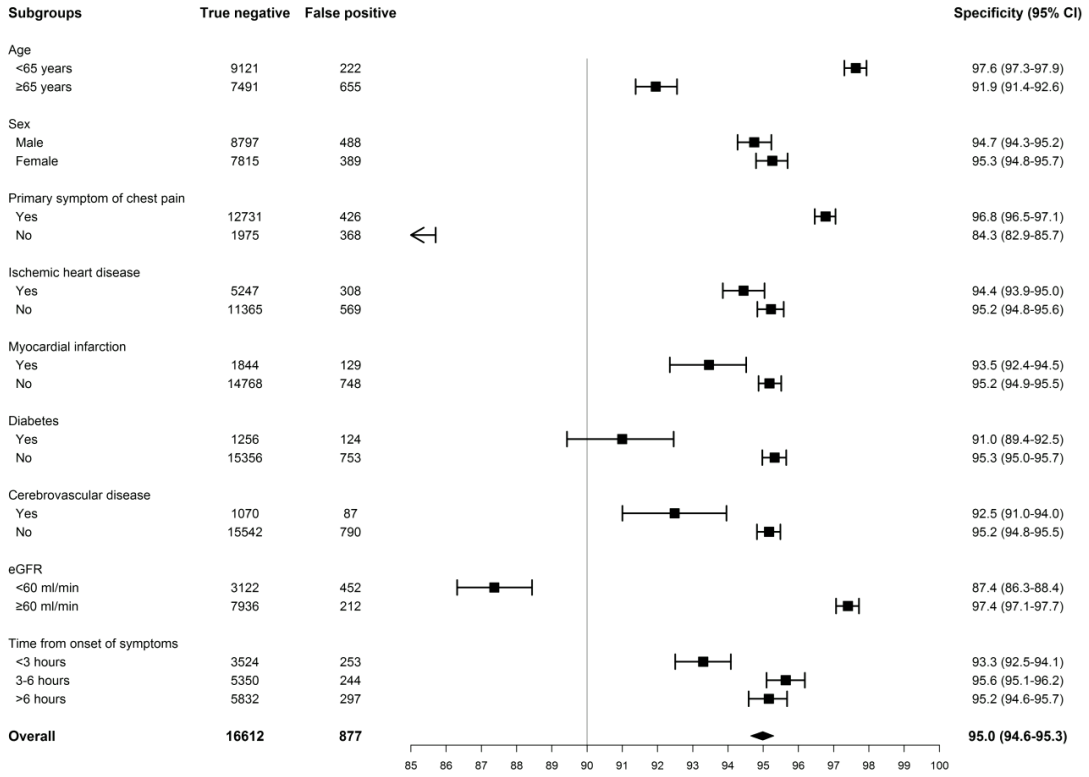


Figure 4.7 Forest plot of the specificity across patient subgroups.
TN = true negative, FP = false positive. The vertical line represents the performance target.

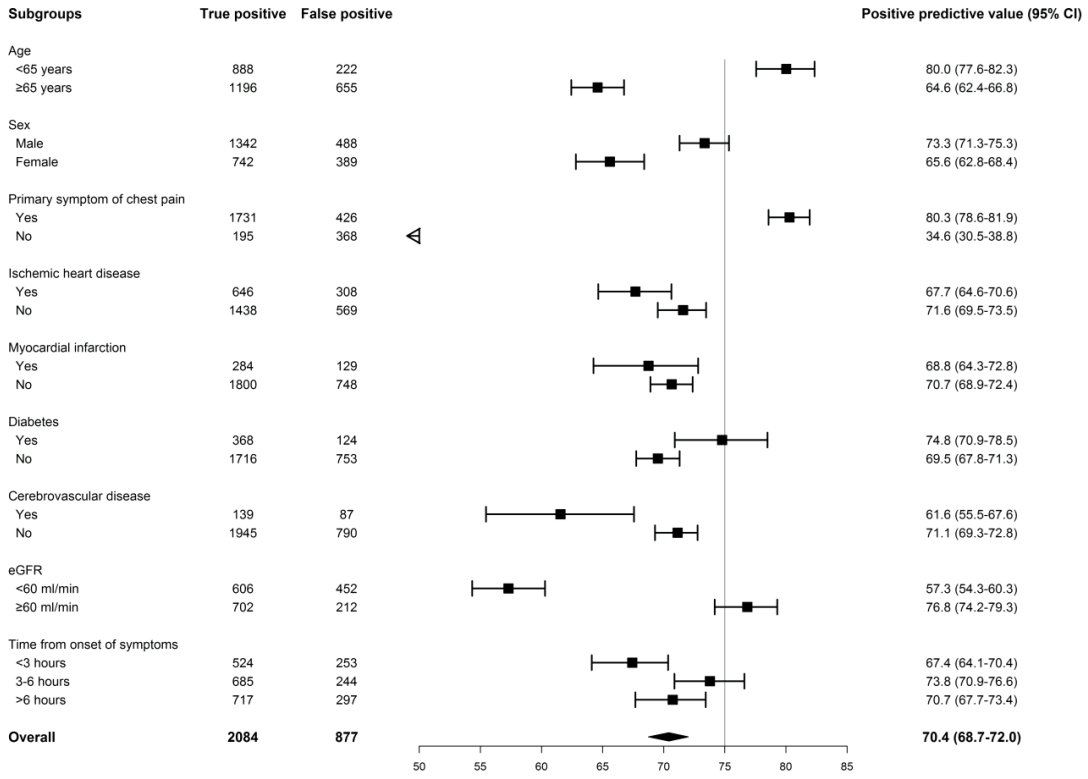


Figure 4.8 Forest plot of the positive predictive value across patient subgroups. PPV = positive predictive value, TP = true positive, FP = false positive. The vertical line represents the performance target.

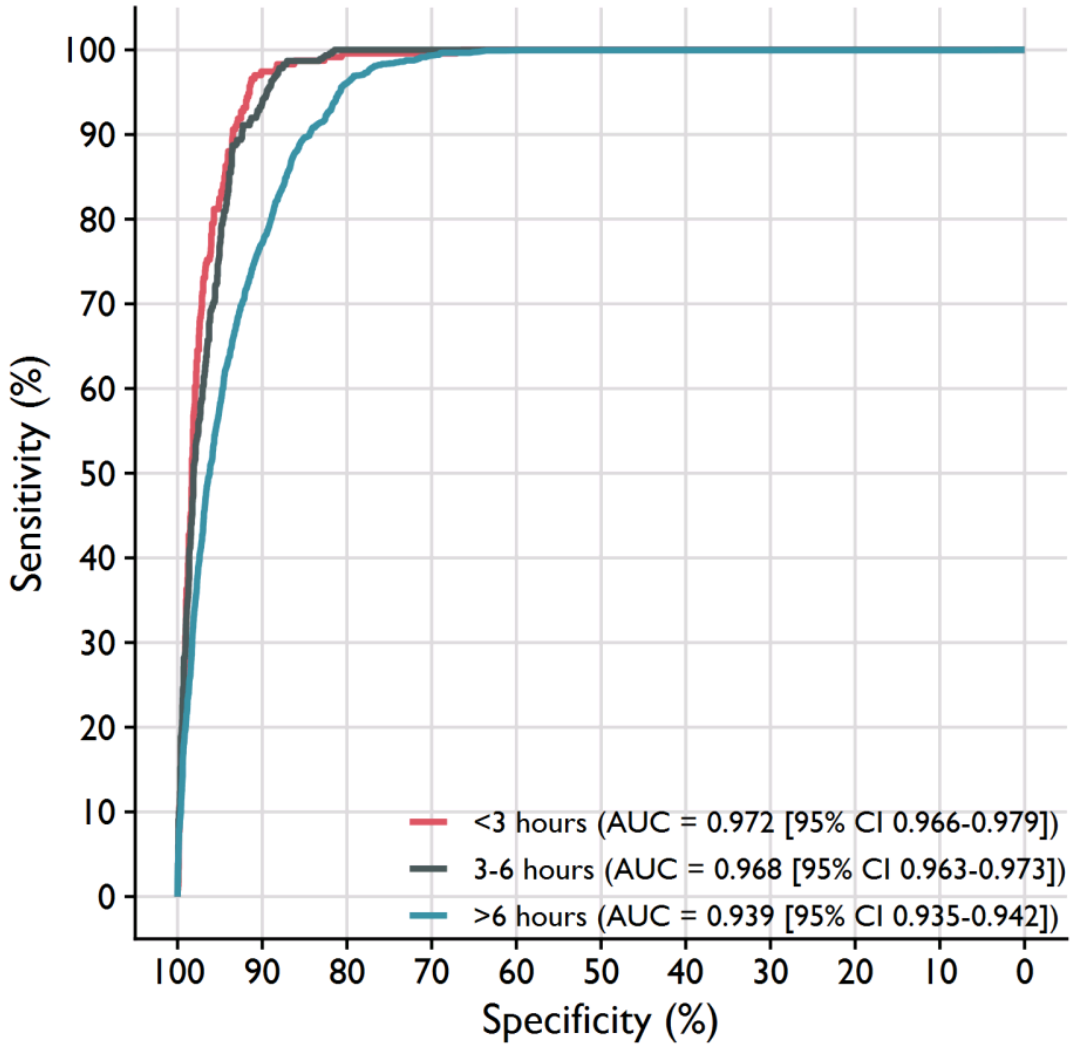


Figure 4.9 Receiver-operating-characteristic (ROC) curve illustrating discrimination of the MI^3 algorithm for type 1 myocardial infarction, stratified by the time difference between blood samples.

4.4.2 Exploratory analysis for type 1 or type 2 myocardial infarction

In the analysis population, 4,194 patients (20.2%) had an adjudicated diagnosis of type 1 or type 2 myocardial infarction. Discrimination was improved for the composite outcome of type 1 or type 2 myocardial infarction (AUC 0.963 [95% CI 0.960 to 0.965], Figure 4.10) compared to type 1 myocardial infarction alone (AUC 0.949 [95% CI 0.946 to 0.952]), however, calibration was not so good. The performance of the high-probability threshold was improved (specificity 97.1% [96.8 to 97.3%], PPV 83.7% [82.3 to 85.0%]) identifying 14.3% (2,961/20,761), while the low-probability threshold identified 62.5% (12,983/20,761) with no difference in performance (sensitivity 99.3% [99.0 to 99.5%], NPV 99.8% [99.7 to 99.9%]) (Figure 4.11).

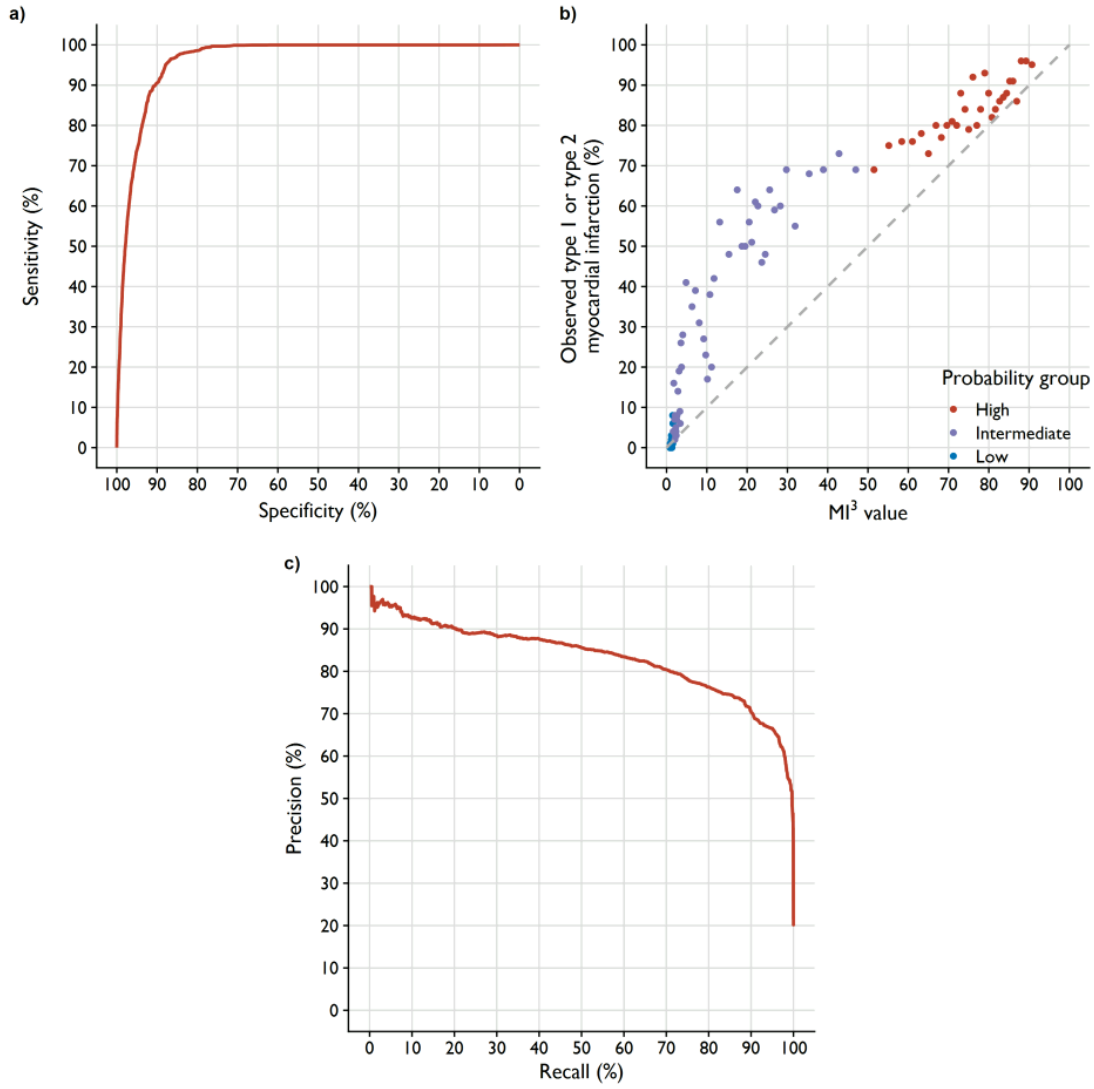


Figure 4.10 (a) Receiver-operating-characteristic (ROC) curve illustrating discrimination of the MI^3 algorithm for type 1 or type 2 myocardial infarction. (b) Calibration of the MI^3 algorithm with the observed proportion of patients with type 1 or type 2 myocardial infarction. The dashed line represents perfect calibration. Each point represents 100 patients. (c) Precision recall curve illustrating discrimination of the MI^3 algorithm for type 1 or type 2 myocardial infarction.

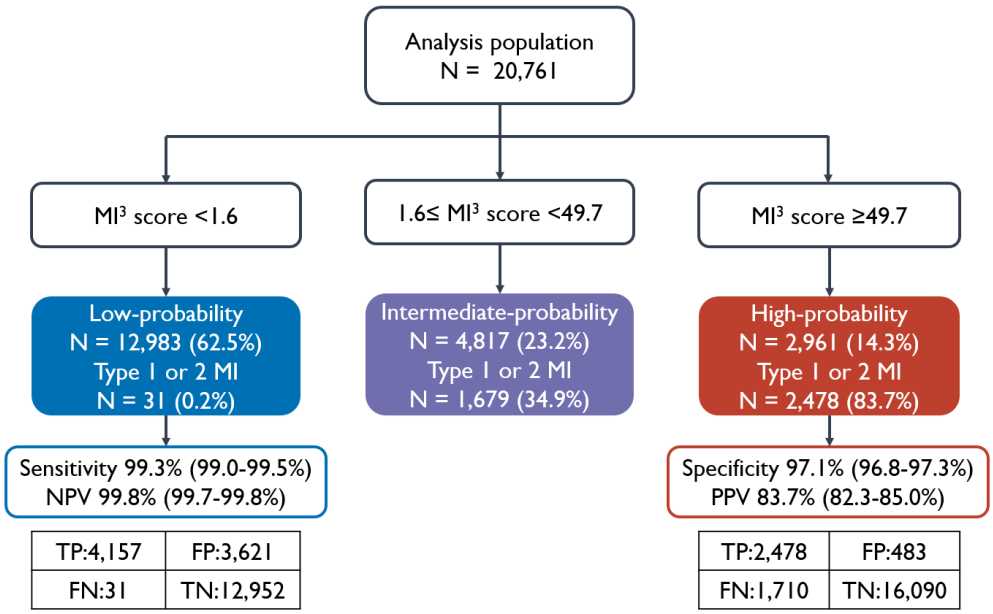


Figure 4.11 Performance of MI³ at example thresholds. Secondary outcome: type 1 or type 2 myocardial infarction.
TP = true positive, FP = false positive, FN = false negative, TN = true negative, NPV = negative predictive value, PPV = positive predictive value.

4.4.3 MI^3 and outcomes at one year

In the analysis population, 1,300 (6.3%) patients experienced either a subsequent myocardial infarction or cardiovascular death at one year. Patients identified by MI^3 as high-probability of index myocardial infarction (2,961/20,761, 14.3%) were more likely to experience a subsequent myocardial infarction or cardiovascular death compared to those identified as low-probability (17.6% [520/2,961] versus 1.5% [197/12,983], P<0.001; Figure 4.12). Death from any cause occurred in 1,671 (8%) patients, and MI^3 was also a good predictor of all-cause mortality.

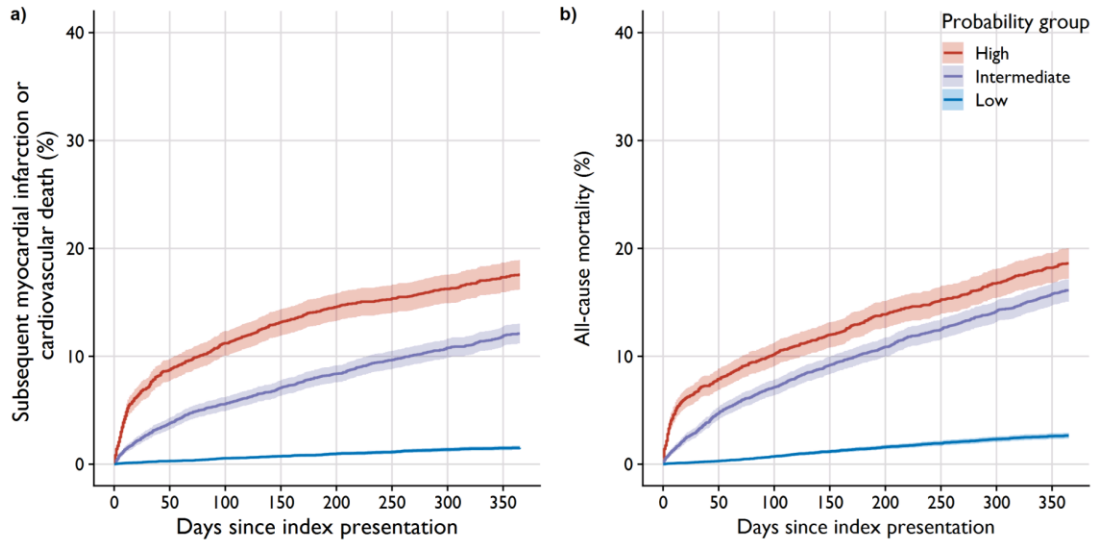


Figure 4.12 (a) Cumulative incidence of myocardial infarction or cardiovascular death over one year and (b) death from any cause stratified by MI^3 score: low-probability, MI^3 score <1.6; intermediate-probability, 1.6 ≤ MI^3 score < 49.7; high-probability, MI^3 score ≥ 49.7. Log-rank between groups for both endpoints, P<0.001.

4.5 Discussion

We validated the MI^3 machine learning algorithm for the diagnosis of myocardial infarction in a large cohort of consecutive patients undergoing serial cardiac troponin measurement for suspected acute coronary syndrome in a multi-centre randomised trial. We make several observations that could inform its application in clinical practice. First, MI^3 had excellent discrimination for type 1 myocardial infarction in a patient population that reflects routine clinical practice. However, calibration was not good in patients with intermediate MI^3 scores. Second, at the pre-specified thresholds, sensitivity and NPV were consistent across patient subgroups, however, specificity and PPV varied significantly. Third, MI^3 provided insights beyond the index presentation, identifying patients at risk of future adverse cardiovascular events. Patients identified as high-probability by the algorithm had a ten-times higher rate of subsequent myocardial infarction or cardiovascular death at one year compared to those who were classified as low-probability.

4.5.1 MI^3 performance

Our study population has several characteristics that enable a robust evaluation of the MI^3 algorithm. Compared to the pooled cohort used in the initial validation of MI^3,[142] our external validation population is three times larger and consists of consecutive patients undergoing serial troponin measurements, improving the generalisability of our findings. The mean age

in our study population is over 5 years older, with a more balanced sex distribution and a higher prevalence of comorbidities. These features are likely to have resulted in a patient population where there is more diagnostic complexity. Although the rule-in performance was more heterogeneous across patient subgroups, MI^3 had an excellent rule-out performance across the study population. This is perhaps unsurprising given that high-sensitivity cardiac troponin, a key variable in this algorithm, is integral to the diagnosis of myocardial infarction and is known to be influenced by both age and sex.[71,146]

Given that the pre-test probability of type 1 myocardial infarction and cardiac troponin concentrations can differ substantially in different patient subgroups[147] it is perhaps intuitive that a diagnostic algorithm that combines cardiac troponin and clinical parameters has a good diagnostic performance. Indeed, there have been numerous statistical models developed to aid in the diagnosis and prognostication of acute cardiovascular conditions, including type 1 myocardial infarction.[148-151] However, very few have been successfully implemented into clinical practice due to barriers such as the number and complexity of the variables that are required in the models and the lack of adequate validation to have sufficient confidence in the diagnostic performance.[152-154] We have not compared gradient boosting to other models or forms of statistical modelling, nor have we evaluated whether discrimination or calibration can be improved by including additional parameters. The objectivity and simplicity of the variables used by MI^3 are

perhaps the algorithm's most important strength. The three variables used in this algorithm: age, sex and troponin, are objective and consistently obtained, with high reproducibility and accuracy, in a busy clinical setting. Furthermore, the initial validation of this algorithm was performed in an international multi-centre patient population, and its diagnostic performance has now been validated in a large consecutive patient population that reflects clinical practice.

4.5.2 MI^3 and high-sensitivity cardiac troponin

Our data further supports the potential clinical application of decision support tools that incorporate key patient factors in the interpretation of cardiac biomarkers. High-sensitivity cardiac troponin is well known to vary significantly according to various patient factors such as age and sex, however it is difficult to account for the complex relationships between these variables using a threshold-based approach. In fact, many institutions worldwide have not yet implemented the sex-specific 99th centile thresholds recommended by the Universal Definition of Myocardial Infarction.[155,156] MI^3 enables more accurate and individualised clinical decisions by taking into account the patient's age and sex in a manner that can be easily interpreted. Furthermore, the ability to include serial troponin concentrations at flexible time-points reduces the potential of misinterpretation compared to an approach that recommends the use of fixed absolute changes in cardiac troponin at specific time-points.[157,158] In our cohort, MI^3 was able to exclude myocardial infarction in the majority of patients with a high negative predictive

value irrespective of when testing was performed, whilst identifying one in seven patients with a high-probability of myocardial infarction.

The application of this algorithm in practice would represent a substantial change in the approach to the assessment of patients with suspected acute coronary syndrome. Our current practice is based exclusively on the use of single or multiple cardiac troponin thresholds with serial testing performed at fixed timepoints. By using cardiac troponin as a continuous measure and incorporating rate of change rather than an absolute change in troponin concentration, MI^3 may be more flexible and easier to implement in busy Emergency Departments. To our knowledge no similar algorithms are available and none report the likelihood of myocardial infarction for individual patients or associated diagnostic metrics to guide clinical decision making. Whilst we have validated the performance of the algorithm in triaging patients as low, intermediate, or high risk, in practice we would anticipate that clinical decisions are guided by individualised estimates of the diagnostic likelihood. Further studies are required to evaluate whether care guided by these estimates, and the provision of diagnostic metrics, changes clinical decision making or the use of subsequent cardiac testing in practice.

4.5.3 Clinical practice and outcomes

Although the training and testing of this algorithm has been published previously,[142] this is the first time that MI^3 has been validated in a consecutive patient population that reflects the way it would be applied in

clinical practice. This is an essential step in understanding how the algorithm will perform in practice where troponin testing is guided by clinical need rather than by a research protocol. A lack of external validation and evaluations of algorithm performance in routine care is one of the main reasons that few machine learning algorithms are used in practice today. Furthermore, in addition to validating the diagnostic performance of MI^3, we provide, for the first-time, data on outcomes following discharge from hospital. The association with adverse cardiovascular outcomes at one year is reassuring and suggests that the algorithm is appropriately risk stratifying patients who are likely to benefit most from further diagnostic testing and treatment beyond the index visit.

4.5.4 Future directions

Although MI^3 had a good overall diagnostic performance in our cohort, there are several aspects that can potentially be improved. First, we observed that MI^3 was not well calibrated in patients with intermediate scores. This group of patients are the most challenging to diagnose in clinical practice. One of the advantages of using a learning algorithm over other pathways is that further training is possible, which may be required to improve calibration for this group in different healthcare settings. Alternatively, the use of additional features to refine the estimates of probability in this group could be explored. Second, although performance of the low-probability threshold was consistent across important patient subgroups, we observed heterogeneity in the positive predictive value of the high-probability threshold, particularly

when stratified according to the primary presenting symptom. This is consistent with our prior research[159,160] and likely reflects the greater prevalence of non-ischaemic myocardial injury and type 2 myocardial infarction in our consecutive patient population as compared to the cohorts used to train the algorithm where some patient selection was inevitable. It is possible that an algorithm that incorporates other clinical features may perform more consistently across these subgroups when identifying patients at high-probability of type 1 myocardial infarction. Third, we used serial cardiac troponin measurements for both the rule-in and rule-out of myocardial infarction here. Algorithms that can risk stratify patients using only cardiac troponin concentrations at presentation are in development and may further improve efficiency. Finally, although MI^3 had good performance for the prediction of type 1 or type 2 myocardial infarction, it was not developed to distinguish between the two. Future algorithms to diagnose and differentiate between type 1 and type 2 myocardial infarction would be useful given the diagnostic challenge in clinical practice and that the treatment for these conditions differs.

4.5.5 Limitations

We acknowledge several limitations in our study design. In the majority of patients in our cohort, serial cardiac troponin measurements were performed 6 hours apart. This is significantly longer than recommended by current international guidelines.[157] However, in our subgroup analysis stratified by time of serial sampling, the diagnostic performance of MI^3 remained good

regardless of the time interval between serial troponin measurements. Although MI^3 includes sex as a parameter in the model discrimination was not as good in women compared to men. This likely reflects differences in the use of sex-specific and uniform thresholds to diagnose myocardial infarction between the data sets used to train and to validate the algorithm.[142] In the High-STEACS trial, sex-specific thresholds were used in practice and to adjudicate the diagnosis of myocardial infarction.[71] Performance could be improved with additional training of the algorithm in healthcare settings that use sex-specific diagnostic thresholds in practice. Furthermore, although our analysis demonstrated that the outcomes at one year in patients classified as low-probability by MI^3 was reassuring, future studies evaluating outcomes after MI^3 is implemented are needed to confirm the safety of this algorithm in clinical practice.

4.6 Conclusions

In consecutive patients undergoing serial cardiac troponin measurement for suspected acute coronary syndrome, the MI^3 machine learning algorithm can accurately estimate the likelihood of myocardial infarction and predict subsequent adverse cardiovascular events. The model could improve the diagnostic pathways for myocardial infarction by accurately identifying high-risk patients to be targeted for prompt individualised treatment, and by allowing early discharge in low-risk patients.

In the next chapter, we will develop a new and improved machine learning algorithm to assist with the diagnosis of myocardial infarction having in mind the limitations of the machine learning algorithm described in this chapter and try to improve them.

CHAPTER 5

Machine learning to optimise use of cardiac troponin in the diagnosis of acute myocardial infarction

Chapter 5 Machine learning to optimise use of cardiac troponin in the diagnosis of acute myocardial infarction

5.1 Overview

Guidelines recommend fixed cardiac troponin thresholds for the assessment of patients with suspected acute coronary syndrome, however, performance varies in important patient groups as concentrations are influenced by age, sex and comorbidities. This limitation can be addressed using machine learning algorithms.

Machine learning algorithms were developed that integrate cardiac troponin concentrations at presentation or on serial testing with age, sex and clinical features in 10,038 consecutive patients with suspected acute coronary syndrome. The primary outcome was an adjudicated diagnosis of type 1, type 4b or type 4c myocardial infarction. The best performing algorithm was selected for the CoDE-ACS (Collaboration for the Diagnosis and Evaluation of Acute Coronary Syndrome) decision-support tool, and performance was externally validated in 3,035 patients pooled from three prospective studies and in key subgroups.

CoDE-ACS had excellent discrimination and calibration using cardiac troponin at presentation (area under curve 0.959, 95% confidence interval 0.948-0.971) and on serial testing (0.971, 0.962-0.980) in the validation cohort. Using cardiac troponin at presentation, the rule-out score identified 71% (1,885/2,656) of patients without myocardial injury as low-probability of myocardial infarction with a 99.5% (99.1-

99.7%) negative predictive value. The rule-in score identified 65.4% (248/379) of patients with myocardial injury as high-probability of myocardial infarction with an 85.1% (81.1-88.3%) positive predictive value. Performance of the rule-out and rule-in scores was consistent across all patient subgroups.

We developed and externally validated the CoDE-ACS decision-support tool using machine learning to aid the diagnosis of myocardial infarction. CoDE-ACS had excellent overall diagnostic performance, and performed consistently across patient subgroups.

5.2 Introduction

High-sensitivity cardiac troponin assays have led to the development and widespread adoption of accelerated diagnostic pathways to expedite the assessment of patients with suspected acute coronary syndrome.[137,138,140,161] These pathways are now recommended by national and international clinical practice guidelines, but have some important limitations.[135] First, they use fixed cardiac troponin thresholds for all patients, which do not account for age, sex or comorbidities known to influence troponin concentrations.[68,140] Second, they are based on specific timepoints for serial testing, which can be challenging in a busy Emergency Department, and such pathways may not be generalisable to all health care systems. Third, they broadly categorise patients as either low-, intermediate- or high-risk based on troponin thresholds alone, and do not consider other important information, such as the time of symptom onset or findings on the electrocardiogram. Machine learning approaches that integrate cardiac troponin as a continuous measure and these clinical features may provide a more individualised risk assessment and improve the diagnosis of myocardial infarction.

We evaluated the diagnostic performance of guideline-recommended cardiac troponin thresholds across important patient subgroups and developed a decision-support tool that uses machine learning to calculate the probability of myocardial infarction for each patient. We subsequently externally validated the performance of our decision support tool.

5.3 Methods

5.3.1 Study population

The High-Sensitivity Troponin in the Evaluation of Patients With Suspected Acute Coronary Syndrome (High-STEACS; NCT01852123) trial population was used as the derivation cohort to develop machine learning models.[73] High-STEACS was a stepped-wedged cluster-randomised controlled trial that evaluated the implementation of a high-sensitivity cardiac troponin I assay in consecutive patients with suspected acute coronary syndrome presenting to 10 secondary and tertiary hospitals in Scotland between June 10, 2013, and March 3, 2016. The trial design has been described previously[73] and the study was approved by the Scotland A Research Ethics Committee, the Public Benefit and Privacy Panel for Health and Social Care, and by each National Health Service (NHS) Health Board.

Patients were included in this prespecified secondary analysis based on the following criteria: (1) age ≥18 years old, (2) presentation with suspected acute coronary syndrome, (3) cardiac troponin measured using the ARCHITECT$_{STAT}$ high-sensitivity cardiac troponin I assay (Abbott Laboratories), (4) availability of electrocardiographic and physiological data for diagnostic adjudication. Patients with a diagnosis of ST-segment elevation myocardial infarction were excluded given they undergo coronary revascularisation directly without troponin testing in the Emergency Department (Figure 5.1).

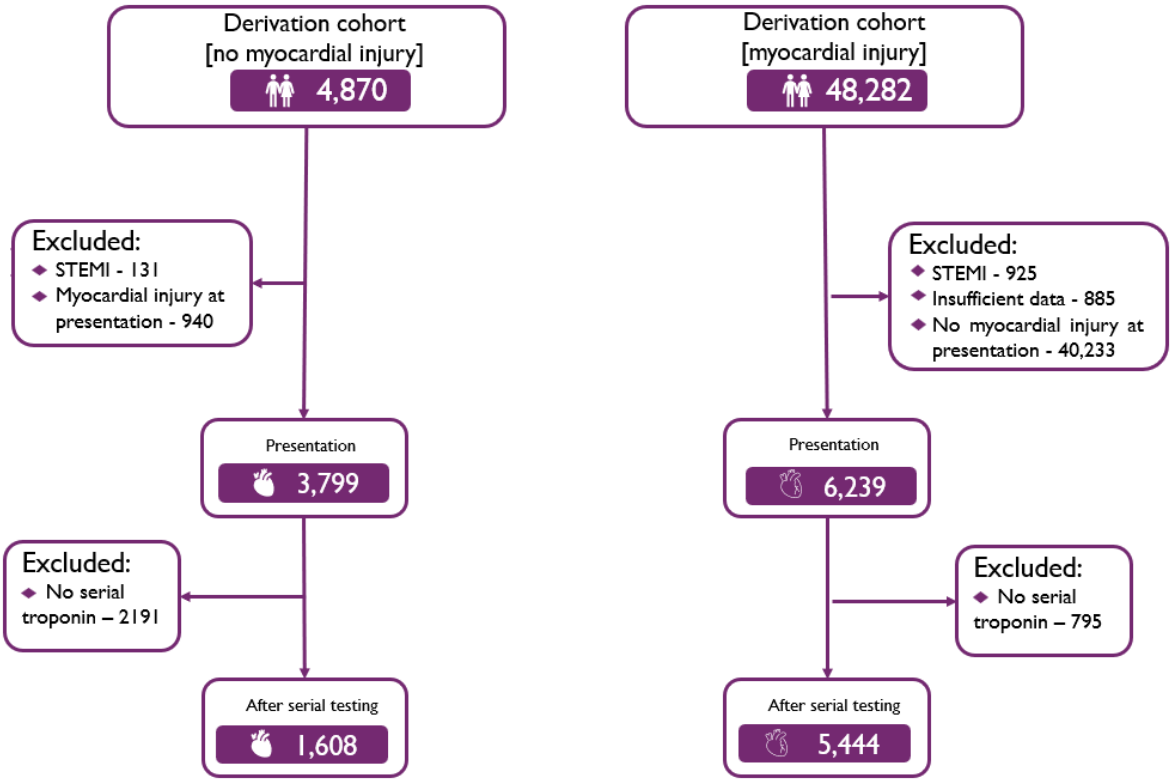


Figure 5.1 Flow diagram of the analysis population.

5.3.2 Outcomes

The machine learning model was developed to predict an adjudicated diagnosis of type 1, type 4b or type 4c myocardial infarction during the index hospital admission. As previously described, the diagnosis of myocardial infarction was adjudicated according to the Fourth Universal Definition of Myocardial Infarction by two clinicians independently, with a third reviewer providing consensus if there was disagreement.[4,73,162]

5.3.3 Performance of guideline recommended cardiac troponin thresholds

We evaluated the diagnostic performance and proportion of patients identified by guideline recommended cardiac troponin thresholds to rule-out (2 ng/L and 5 ng/L) and rule-in (99th centile of 16 ng/L [women] and 34 ng/L [men], and 64 ng/L)[71] myocardial infarction. These were evaluated in the overall population and in key pre-specified subgroups by age, sex, time from symptom onset to troponin measurement, renal impairment, prior ischaemic heart disease, diabetes mellitus, and cerebrovascular disease.

5.3.4 Feature selection and processing

We used high-sensitivity cardiac troponin I concentrations as a continuous measure, and selected 12 objective clinical variables known to be associated with cardiac troponin or myocardial infarction, which were found to have the highest relative importance in our model training phase. These were age, sex, the number of hours

from symptom onset to cardiac troponin measurement, chest pain, known ischaemic heart disease, hyperlipidemia, heart rate, systolic blood pressure, Killip class, evidence of myocardial ischaemia on the electrocardiogram, renal function [estimated glomerular filtration rate calculated using the Chronic Kidney Disease Epidemiology Collaboration formula], and haemoglobin. To maximise the clinical utility of our models, we firstly developed models using cardiac troponin concentration at presentation alone. We subsequently developed models to include a second cardiac troponin concentration, measured at an early and flexible timepoint.

5.3.5 Model development and validation

We evaluated four statistical models in the development of our decision-support tool – logistic regression, naïve bayes, random forest and extreme gradient boosting (XGBoost).[122,123] Given the features that inform diagnosis are likely to differ for the rule in and rule out of myocardial infarction, we developed separate models for those with and without myocardial injury (cardiac troponin I concentration above the sex-specific 99th centile threshold; 16 ng/L in women and 34 ng/L in men).[71] For both models, in the derivation cohort we multiply imputed ten datasets to account for missing data and performed ten iterations of 10-fold cross-validation to compute a score (0-100) that corresponds to an individual patient's probability of having myocardial infarction. We subsequently identified the scores that would classify the highest proportion of patients as high- or low-probability using prespecified criteria to optimise performance to rule-in (80% PPV and 80% specificity) myocardial infarction in those with myocardial injury and to rule-out (99.5% NPV and 95% sensitivity)

myocardial infarction in those without myocardial injury. The models with the best diagnostic performance in the overall population with and without myocardial injury and in pre-specified patient subgroups were selected and integrated into our CoDE-ACS (Collaboration for the Diagnosis and Evaluation of Acute Coronary Syndrome, [https://decision-support.shinyapps.io/code-acs/) decision-support tool.

We externally validated CoDE-ACS in the IMPACT (Improved Assessment of Chest pain Trial), ADAPT (2-Hour Accelerated Diagnostic Protocol to Assess Patients With Chest Pain Symptoms Using Contemporary Troponins as the Only Biomarker) and SPACE (Signal Peptide in Acute Coronary Events) cohorts from Australia and New Zealand.[104-106] All analyses were performed in R version 4.0.3.

5.4 Results

The derivation cohort consists of 10,038 patients (median age 70 years, 48% women) (Table 5.1)**.** In this cohort, during the index hospitalisation, the adjudicated diagnosis was type 1, 4b or 4c myocardial infarction in 49% (3,062/6,239) and 3% (132/3,799) in patients with and without evidence of myocardial injury at presentation.

Table 5.1 Baseline characteristics of the derivation and external validation cohort.

	Derivation cohort		**External validation cohort**	
	No myocardial injury at presentation	**Myocardial injury at presentation**	**No myocardial injury at presentation**	**Myocardial injury at presentation**
Number of patients	3,799	6,239	2,656	379
Age, years	62 (50, 74)	74 (62, 83)	55 (46, 66)	70 (59,79)
Sex				
Female	1,580 (42%)	3,199 (51%)	1,039 (39%)	154 (41%)
Male	2,219 (58%)	3,040 (49%)	1,617 (61%)	225 (59%)
Chest pain at presentation	3,251 (86%)	4,030 (70%)	2,550 (96%)	366 (97%)
Early presenter (≤2 hours from symptom onset)	606 (16%)	1,209 (21%)	588 (22%)	37 (10%)
Previous medical conditions				
Myocardial infarction	606 (18%)	837 (13%)	493 (19%)	127 (24%)
Ischemic heart disease	1,133 (34%)	2, 136 (34%)	618 (23%)	158 (42%)
Cerebrovascular disease	236 (7%)	626 (10%)	142 (5%)	36 (9%)
Diabetes mellitus	513 (16%)	919 (15%)	340 (13%)	86 (23%)
Previous revascularisation				
PCI	360 (11%)	560 (9%)	399 (15%)	88 (23%)
CABG	178 (5%)	161 (3%)	164 (6%)	45 (12%)

Medications at presentation				
Aspirin	720 (30%)	2,267 (36%)	861 (32%)	187 (49%)
Dual anti-platelet therapy†	115 (5%)	309 (5%)	22 (1%)	3 (1%)
ACE or ARB	745 (31%)	2,681 (43%)	335 (13%)	37 (10%)
Beta-blocker	584 (25%)	2,156 (35%)	236 (9%)	36 (10%)
Electrocardiogram result§				
Normal	3,279 (87%)	3,308 (64%)	1,197 (45%)	33 (10%)
Myocardial ischaemia	458 (14%)	1,351 (26%)	137 (5%)	110 (29%)
ST segment elevation	93 (3%)	196 (4%)	18 (1%)	10 (3%)
Physiological parameters§				
Heart rate, beats per minute	77 (65, 90)	81 (68,99)	73 (64, 84)	75 (65, 88)
Systolic blood pressure, mmHg	137 (121, 153)	138 (120, 157)	138 (124, 155)	144 (128, 162)
Haematology and clinical chemistry measurements				
Haemoglobin, g/L	NA	133 (118, 146)	143 (133, 153)	138 (124, 151)
eGFR, ml/min	86 (69, 99)	66 (44, 85)	83 (69, 90)	67 (49, 83)
Presentation high-sensitivity cardiac troponin I	3 (2, 7)	85 (41, 320)	3 (2, 5)	130 (46, 660)
Serial high-sensitivity cardiac troponin I, % ¶	5 (3, 11)	209 (53, 1,786)	3 (2, 5)	192 (56, 1,114)
Peak high-sensitivity cardiac troponin I, ng/L	4 (2, 8)	170 (51, 1,422)	3 (2, 5)	198 (58, 1,127)

Adjudicated diagnosis				
Type 1, 4b or 4c MI	132 (3%)	3,062 (49%)	63 (2%)	271 (72%)
Type 2 MI	33 (1%)	802 (13%)	39 (1%)	34 (9%)
Non-ischemic myocardial injury	21 (1%)	2,343 (38%)	24 (1%)	74 (20%)

Values are mean ± standard deviation; median [interquartile range]; n (%).
*A presenting symptom was missing in 2,264 (11%). †Two medications from aspirin, clopidogrel, prasugrel, or ticagrelor. ‡Includes warfarin or novel oral anticoagulants.
ACE, angiotensin converting enzyme; ARB, angiotensin receptor blockers.

5.4.1 Diagnostic performance of cardiac troponin thresholds in subgroups

In the derivation cohort, in those without myocardial injury at presentation the negative predictive value of cardiac troponin rule-out thresholds of less than 2 ng/L and less than 5 ng/L were 99.5% (95% confidence interval 98.9-100.0%) and 99.6% (99.3-99.8%), respectively (Table 5.2). The corresponding sensitivities were 97.7% (94.7-100.0%) and 93.2% (88.8-97.2%). Overall, 16.0% (609/3,799) and 60.4% (2,295/3,799) of patients without myocardial injury at presentation had troponin concentrations below 2 ng/L and 5 ng/L, respectively. Negative predictive values for both thresholds were lower in patients presenting within 2 hours of symptom onset and in those with evidence of myocardial ischaemia on the electrocardiogram (Figure 5.2, Figure 5.3). Among patients with myocardial injury, the positive predictive value of the sex-specific 99th centile and rule-in threshold of 64 ng/L was 49.4% (48.2-50.7%) and 58.8% (57.2-60.3%), respectively. For both thresholds, there was significant heterogeneity in all subgroups with a lower positive predictive value in those greater than 65 years old, in women, and in those with known ischemic heart disease and impaired renal function (Figure 5.4, Figure 5.5).

Table 5.2 Diagnostic performance of guideline-recommended rule-in and rule-out thresholds for myocardial infarction and CoDE-ACS in the derivation cohort.

A) Rule-out thresholds in patients without myocardial injury at presentation.

	True negative	False negative	True positive	False positive	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out
Cardiac troponin thresholds							
<2 ng/L	606	3	129	3061	99.5 (98.9-100.0)	97.7 (94.7-100.0)	16.0%
<5 ng/L	2286	9	123	1381	99.6 (99.3-99.8)	93.2 (88.8-97.2)	60.4%
CoDE-ACS score							
Presentation (1.2)	2510	5	127	1157	99.8 (99.6-100.0)	96.2 (92.3-99.2)	66.2%
Serial testing (0.6)	1118	2	130	341	99.8 (99.5-100.0)	98.5 (96.1-100.0)	70.4%

NPV = Negative predicted value; CI = confidence interval

B) Rule-in thresholds in patients with myocardial injury at presentation.

	True negative	False negative	True positive	False positive	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in
Cardiac troponin thresholds							
>99th centile	0	0	3085	3154	49.4 (48.2-50.7)	0 (0-0)	100.0%
≥64 ng/L	1657	952	2133	1497	58.8 (57.2-60.3)	52.5 (50.8-54.1)	58.2%
CoDE-ACS score							
Presentation (59.9)	2620	945	2140	534	80.0 (78.4-81.5)	83.1 (81.7-84.4)	42.9%
Serial testing (62.9)	1851	854	2278	461	83.2 (81.7-84.4)	80.1 (78.4-81.6)	50.3%

PPV = Positive predicted value; CI = confidence interval

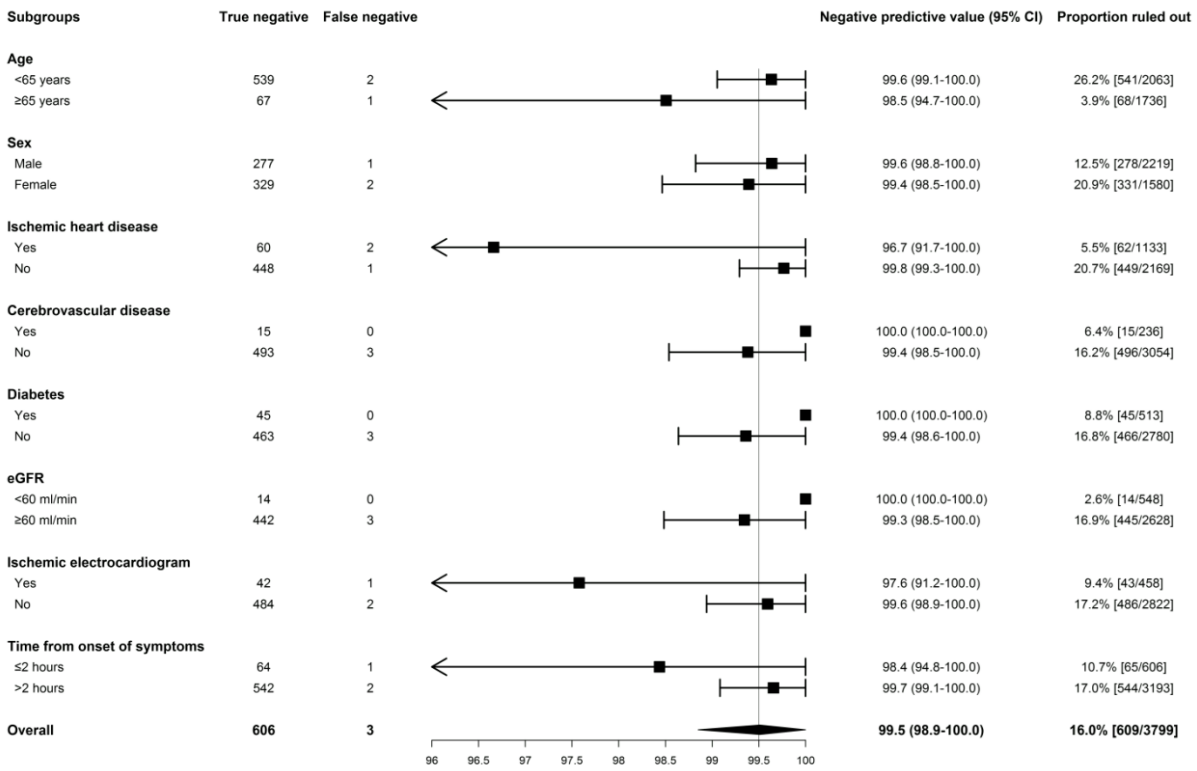


Figure 5.2 Negative predictive value of the 2 ng/L cardiac troponin threshold in the derivation cohort across patient subgroups.

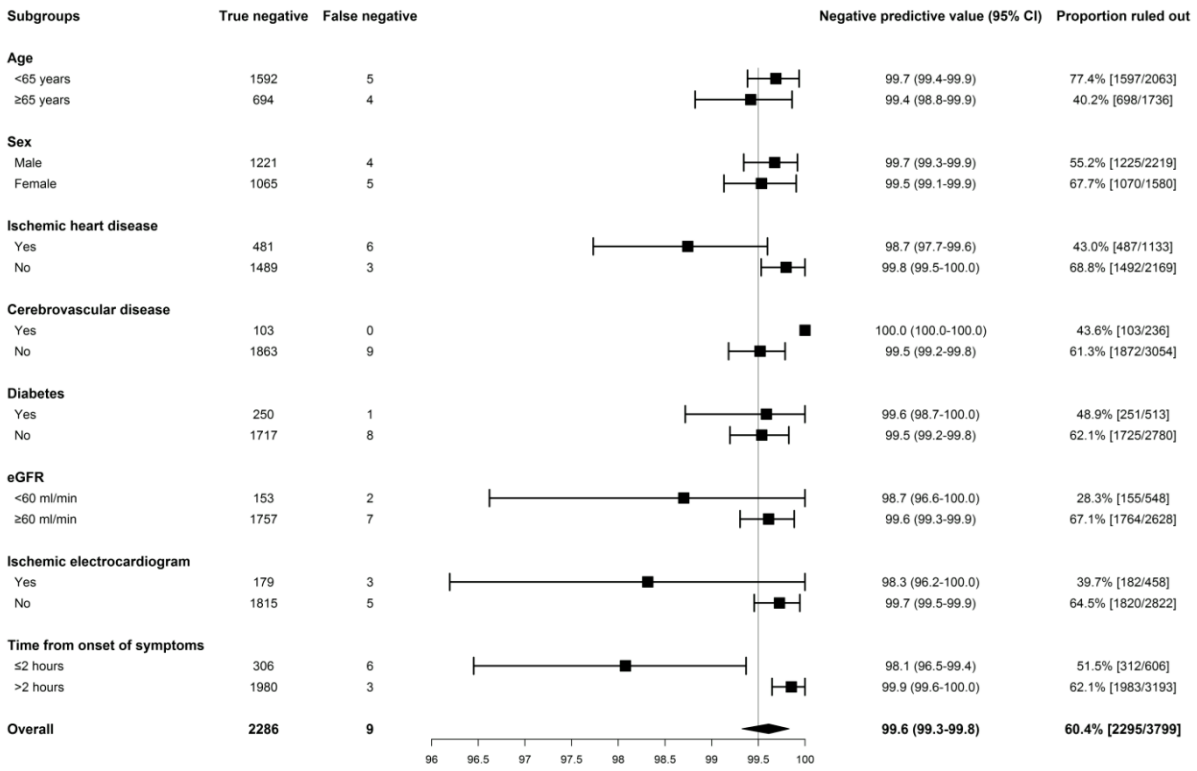


Figure 5.3 Negative predictive value of the 5 ng/L cardiac troponin threshold in the derivation cohort across patient subgroups.

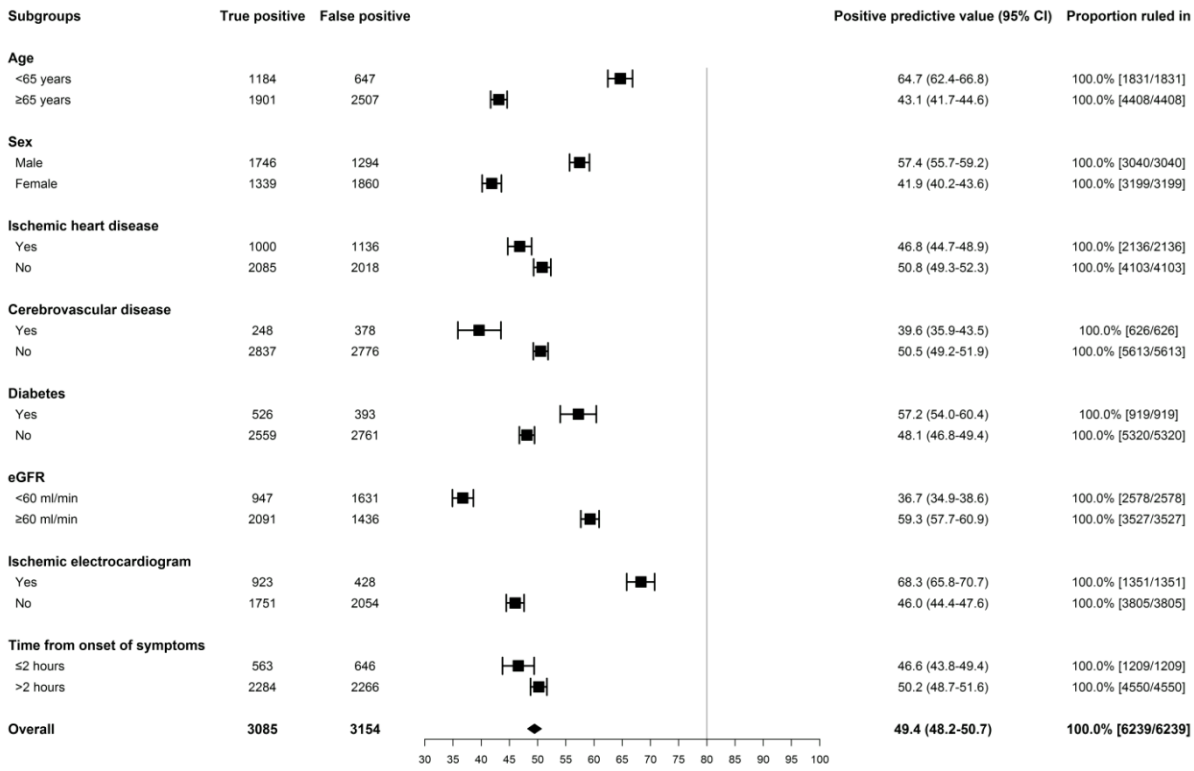


Figure 5.4 Positive predictive value of the sex-specific 99th centile cardiac troponin threshold in the derivation cohort across patient subgroups.

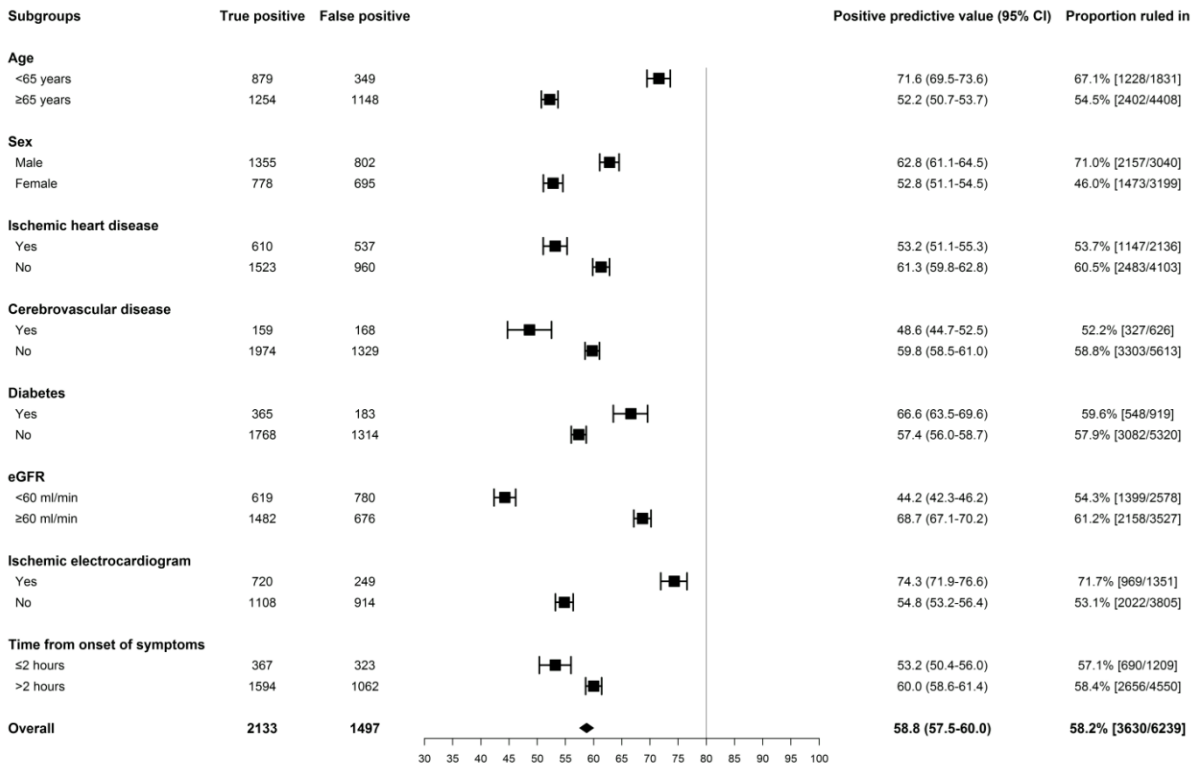


Figure 5.5 Positive predictive value of the 64 ng/L cardiac troponin threshold in the derivation cohort across patient subgroups.

5.4.2 Internal validation of CoDE-ACS

The XGBoost model was the best performing model in the derivation cohort and therefore this model was selected for CoDE-ACS (Table 5.3, Figure 5.6 and Figure 5.7). Using high-sensitivity cardiac troponin concentrations at presentation, a CoDE-ACS rule-out score of 1.2 achieved our prespecified diagnostic performance metrics with a negative predictive value of 99.8% (99.6-100.0%) and sensitivity of 96.2% (92.3-99.2%) in those without myocardial injury at presentation. A CoDE-ACS rule-in score of 59.9 achieved a positive predictive value of 80.0% (78.4-81.5) and specificity of 83.1% (81.7-84.4%) in those with evidence of myocardial injury at presentation. Rule-out and rule-in scores had a consistent performance across all important patient subgroups (Figure 5.8, Figure 5.9). If these scores are used to triage patients with suspected acute coronary syndrome, 66.2% (2,515/3,799) of those without myocardial injury and 42.9% (2,674/6,239) of those with myocardial injury will have myocardial infarction ruled-out and ruled-in respectively at presentation.

Table 5.3 Diagnostic performance of statistical models in the derivation cohort.

A) Rule-out diagnostic performance in patients without myocardial injury.

	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out	AUC	Brier score
Presentation					
Logistic regression	99.8 (99.6-100.0)	97.7 (95.0-100.0)	43.7%	0.898 (0.874-0.923)	0.028
Naïve Bayes	99.9 (99.7-100.0)	99.2 (97.4-100.0)	38.0%	0.875 (0.850-0.900)	0.054
Random Forest	99.8 (99.6-100.0)	97.0 (93.8-99.3)	60.4%	0.904 (0.881-0.927)	0.028
XGBoost	99.8 (99.6-100.0)	96.2 (92.3-99.2)	66.2%	0.914 (0.894-0.934)	0.028
Serial					
Logistic regression	100.0 (100.0-100.0)	100.0 (100.0-100.0)	12.0%	0.890 (0.860-0.921)	0.052
Naïve Bayes	98.0 (97.1-98.8)	83.5 (76.6-89.4)	68.8%	0.870 (0.834-0.906)	0.062
Random Forest	99.8 (99.4-100.0)	98.5 (96.0-100.0)	54.3%	0.964 (0.947-0.981)	0.047
XGBoost	99.8 (99.5-100.0)	98.5 (96.1-100.0)	70.4%	0.973 (0.961-0.985)	0.031

NPV = Negative predicted value; CI = confidence interval.

B) Rule-in diagnostic performance in patients with myocardial injury.

	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in	AUC	Brier score
Presentation					
Logistic regression	80.0 (78.3-81.6)	84.2 (82.9-85.5)	40.0%	0.845 (0.836-0.855)	0.160
Naïve Bayes	80.1 (78.4-81.7)	84.9 (83.7-86.2)	38.3%	0.831 (0.821-0.841)	0.174
Random Forest	80.1 (78.6-81.6)	81.8 (80.3-83.2)	46.4%	0.869 (0.861-0.878)	0.148
XGBoost	80.0 (78.4-81.5)	83.1 (81.7-84.4)	42.9%	0.852 (0.842-0.861)	0.157
Serial					
Logistic regression	82.8 (81.2-84.2)	80.1 (78.4-81.7)	49.2%	0.836 (0.825-0.847)	0.162
Naïve Bayes	82.2 (80.7-83.7)	80.1 (78.4-81.7)	47.5%	0.824 (0.813-0.836)	0.194
Random Forest	83.4 (82.0-84.8)	80.2 (78.5-81.8)	50.6%	0.855 (0.845-0.865)	0.163
XGBoost	83.2 (81.7-84.4)	80.1 (78.4-81.6)	50.3%	0.850 (0.840-0.860)	0.155

PPV = Positive predicted value; CI = confidence interval.

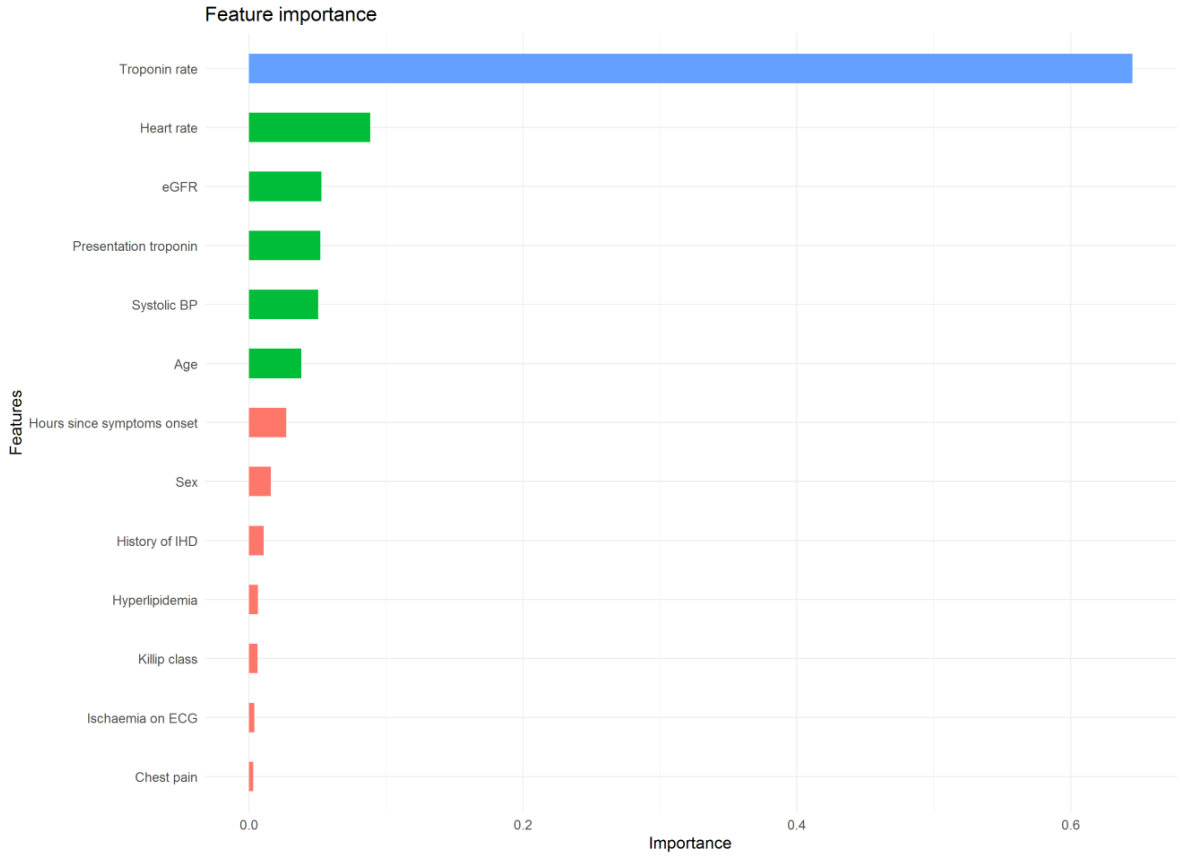


Figure 5.6 Relative feature importance plot for the model developed for patients without myocardial injury.

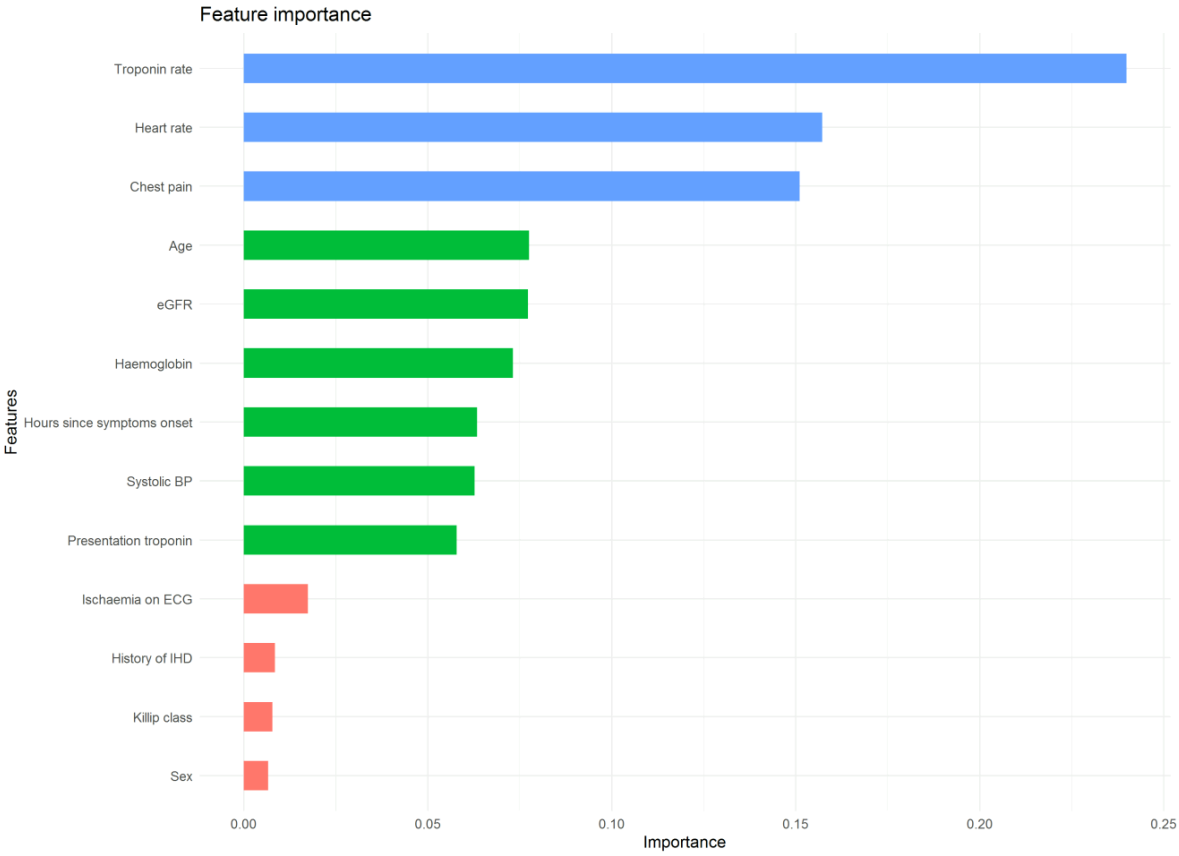


Figure 5.7 Relative feature importance plot for the model developed for patients with myocardial injury.

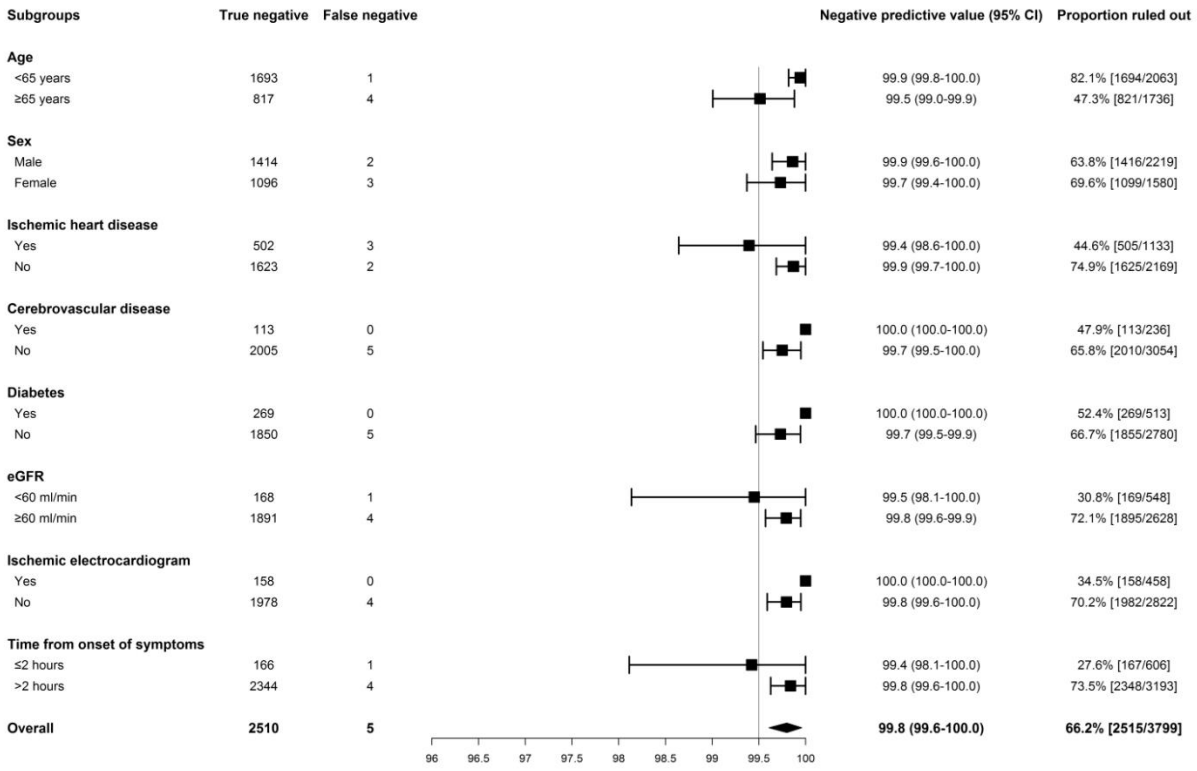


Figure 5.8 Diagnostic performance of the CoDE-ACS score using presentation cardiac troponin concentration in the derivation cohort across patient subgroups (rule-out score).

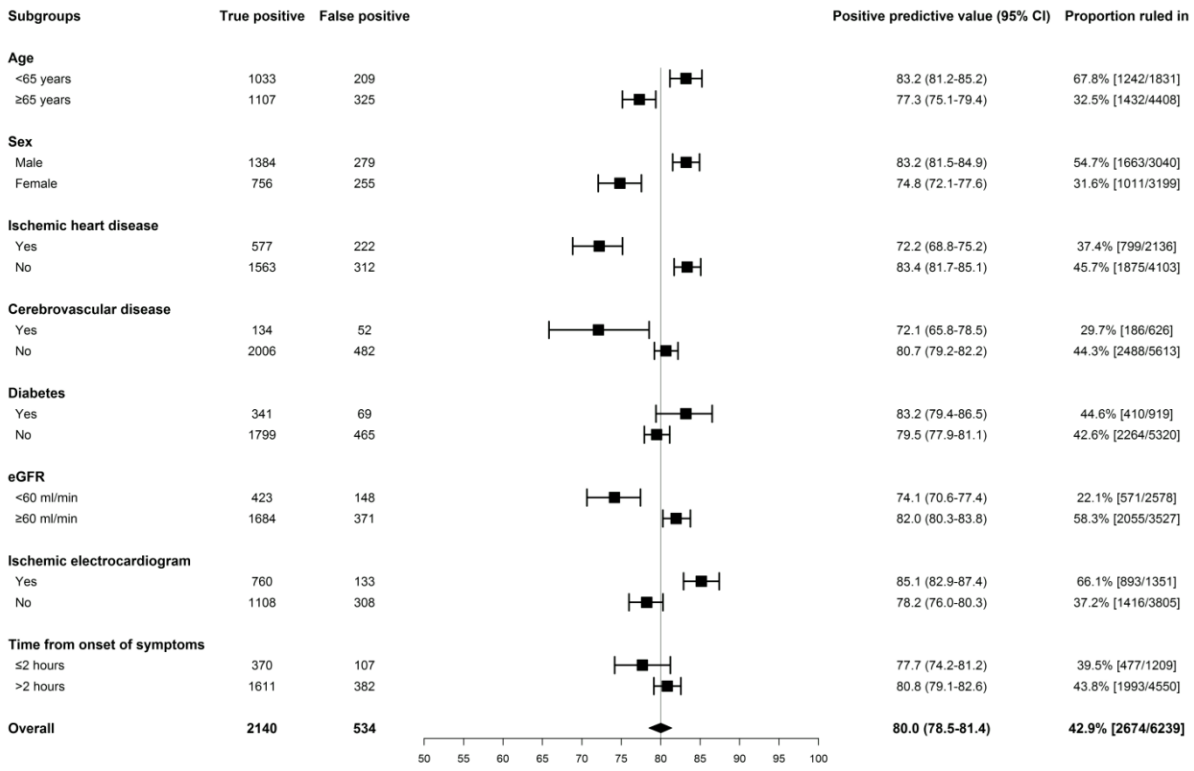


Figure 5.9 Diagnostic performance of the CoDE-ACS score using presentation cardiac troponin concentration in the derivation cohort across patient subgroups (rule-in score).

When the first and second cardiac troponin concentrations were incorporated within the CoDE-ACS model, a CoDE-ACS rule-out score of 0.6 achieved a negative predictive value of 99.8% (99.5-100.0) and a sensitivity of 98.5% (96.1-100.0) (Figure 5.10). The rule-in performance for the of CoDE-ACS, rule-in score of 62.9, improved with serial cardiac troponin concentrations with a positive predictive of 83.2% (81.7-84.4) and a specificity of 80.1 (78.4-81.6). The positive predictive value of CoDE-ACS was improved across all patient subgroups (Figure 5.11).

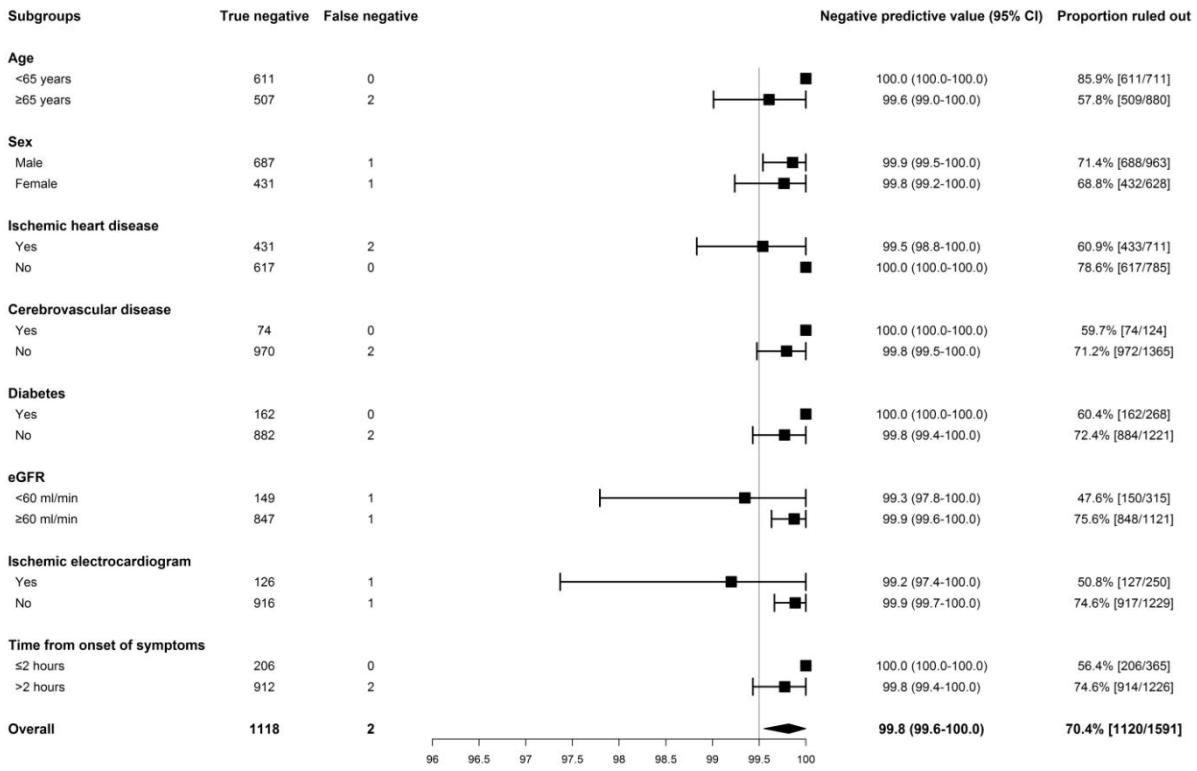


Figure 5.10 Diagnostic performance of the CoDE-ACS score using serial cardiac troponin concentrations across patient subgroups in the derivation cohort (rule-out score).

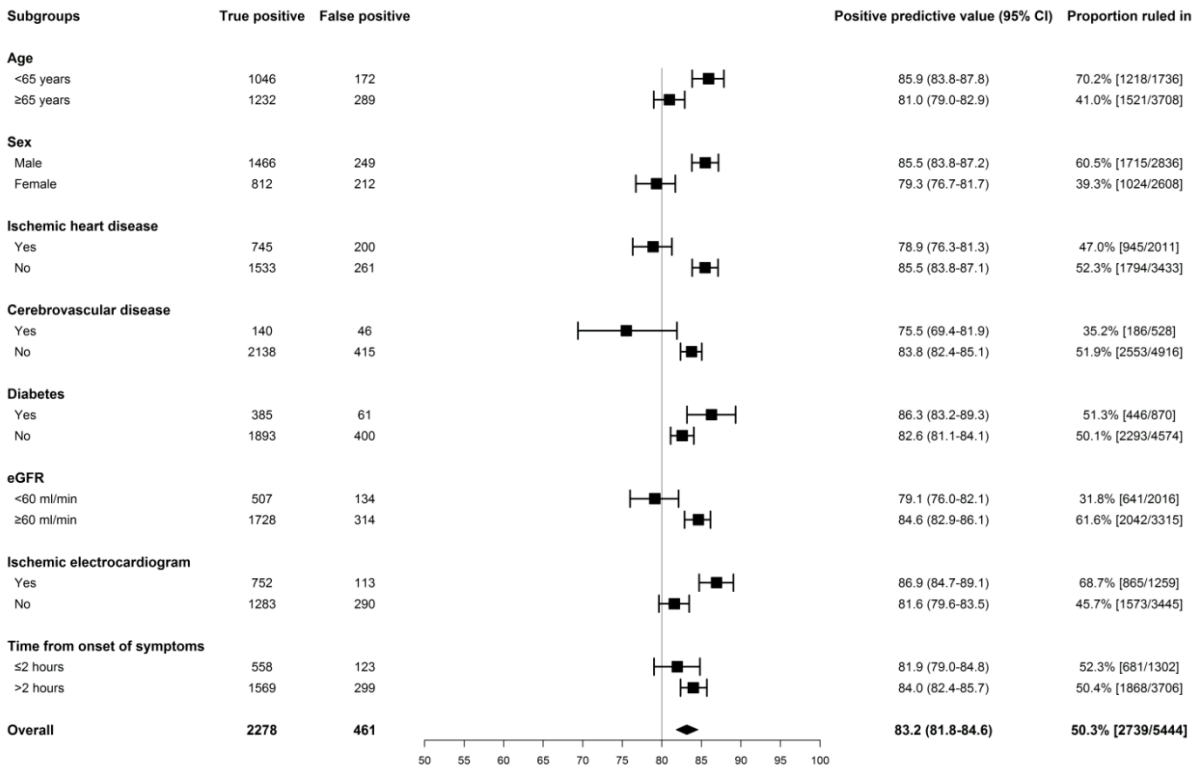


Figure 5.11 Diagnostic performance of the CoDE-ACS score using serial cardiac troponin concentrations across patient subgroups in the derivation cohort (rule-in score).

5.4.3 External validation of CoDE-ACS

The pooled external validation cohort consists of 3,035 patients (median age 57 years, 61% male) (Table 5.1). During the index hospitalisation, myocardial infarction (type 1, 4b or 4c) occurred in 49% (3,062/6,239) and 3% (132/3,799) in patients with and without evidence of myocardial injury at presentation.

In the pooled external validation cohort, the area under curve for CoDE-ACS was 0.959 (0.948-0.971) using the presentation troponin concentration alone and 0.971 (0.962-0.980) using serial troponin results with good calibration (Brier score of 0.040 and 0.039, respectively) (Figure 5.12, Figure 5.13).

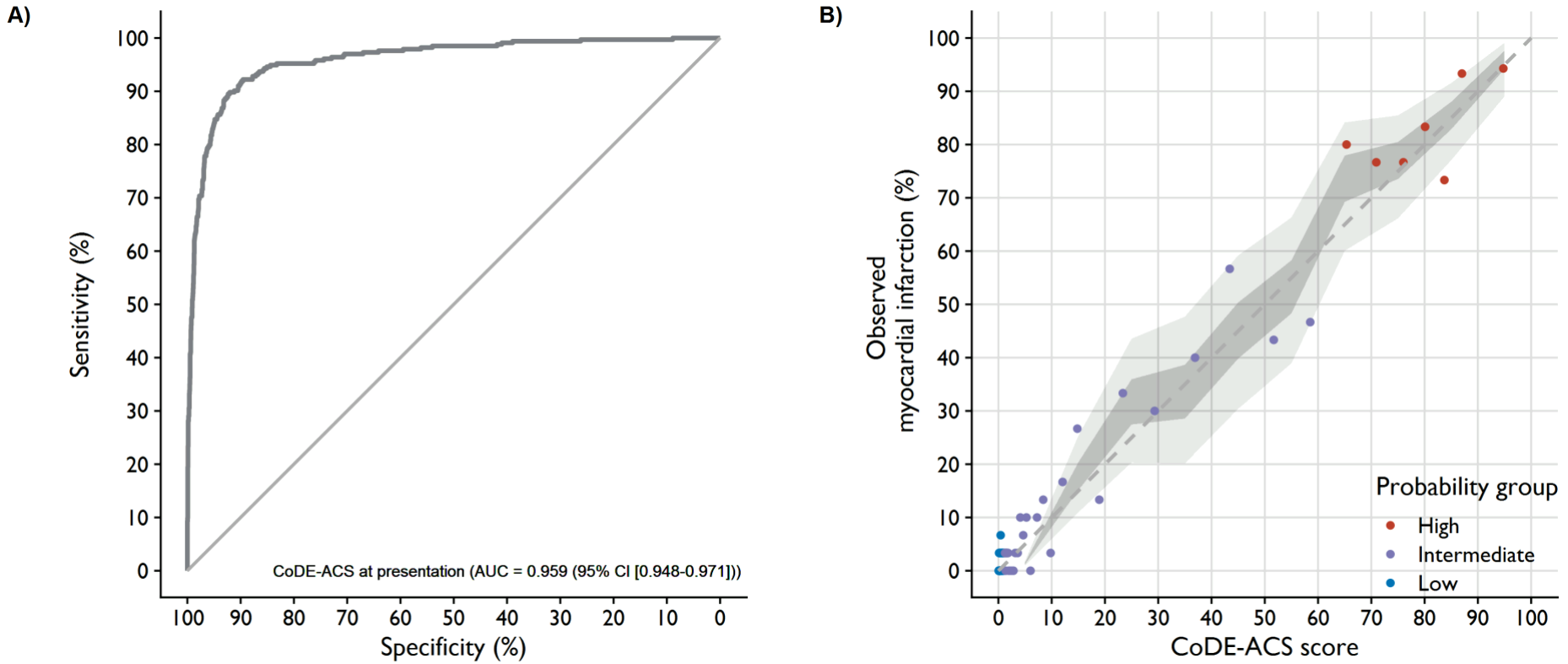


Figure 5.12 Diagnostic performance of the CoDE-ACS score in the external validation cohort using the presentation troponin concentration alone.
A) Receiver-operating-characteristic (ROC) curve illustrating discrimination of the CoDE-ACS for myocardial infarction.
B) Calibration of the CoDE-ACS score with the observed proportion of patients with myocardial infarction. The dashed line represents perfect calibration. Each point represents 100 patients.

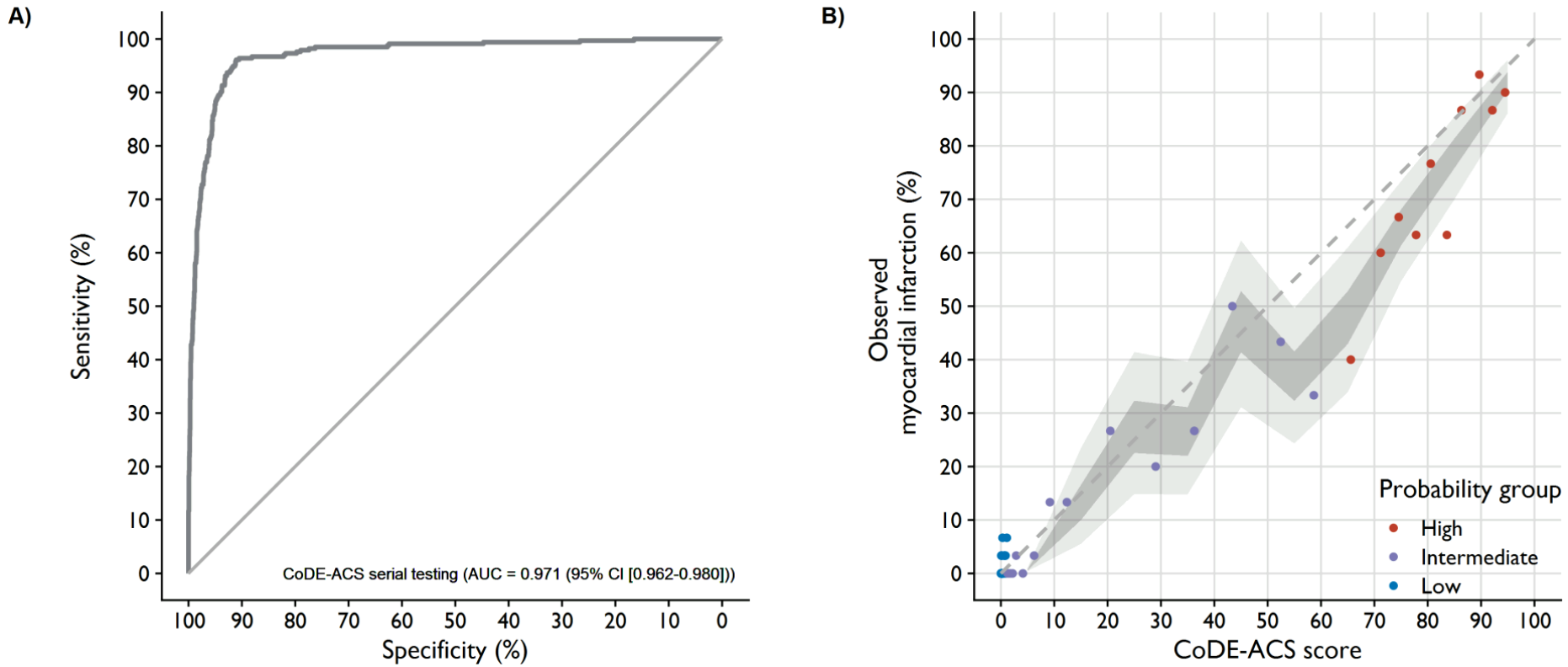


Figure 5.13 Diagnostic performance of the CoDE-ACS in the external validation cohort score using serial troponin results.
A) Receiver-operating-characteristic (ROC) curve illustrating discrimination of the CoDE-ACS for myocardial infarction.
B) Calibration of the CoDE-ACS score with the observed proportion of patients with myocardial infarction. The dashed line represents perfect calibration. Each point represents 100 patients.

In patients without myocardial injury at presentation, a CoDE-ACS score of 1.2 achieved an NPV of 99.5% (99.1-99.7%) and sensitivity of 84.1% (82.7-85.5%) (Table 5.4), whilst in patients with myocardial injury at presentation, a score of 59.9 achieved a PPV of 85.1% (81.1-88.3%) and a specificity of 65.7% (60.8-70.3%) with the presentation troponin result. These rule-in and rule-out scores had similar diagnostic performance across all subgroups (Figure 5.14, Figure 5.15). If these scores were applied in patients with suspected acute coronary syndrome, CoDE-ACS would identify 71% (1,885/2,656) of patients without myocardial injury at presentation at low-probability (<1.2) and 65.4% (248/379) of patients with myocardial injury at presentation at high-probability (≥59.9) of myocardial infarction. With serial troponin concentrations, CoDE-ACS had improved performance to rule-out (NPV of 99.8% [99.5-99.9%] and sensitivity of 84.8% [83.4-86.2%]) and rule-in (PPV of 78.1% [74.0-81.7%] and specificity of 51.4% [46.8-56.0%]) myocardial infarction (Figure 5.16, Figure 5.17, Figure 5.18).

Table 5.4 Diagnostic performance of CoDE-ACS in the external validation cohort.

A) Rule-out thresholds in patients without myocardial injury.

	True negative	False negative	True positive	False positive	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out
Rule-out CoDE-ACS threshold							
Presentation (1.2)	1875	10	53	718	99.5 (99.1-99.7)	84.1 (82.7-85.5)	71.0%
Serial testing (0.6)	2030	5	28	527	99.8 (99.5-99.9)	84.8 (83.4-86.2)	78.6%

NPV = Negative predicted value; CI = confidence interval

B) Rule-in thresholds in patients with myocardial injury.

	True negative	False negative	True positive	False positive	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in
Rule-in CoDE-ACS threshold							
Presentation (59.9)	71	60	211	37	85.1 (81.1-88.3)	65.7 (60.8-70.3)	65.4%
Serial testing (62.9)	74	52	249	70	78.1 (74.0-81.7)	51.4 (46.8-56.0)	71.7%

PPV = Positive predicted value; CI = confidence interval

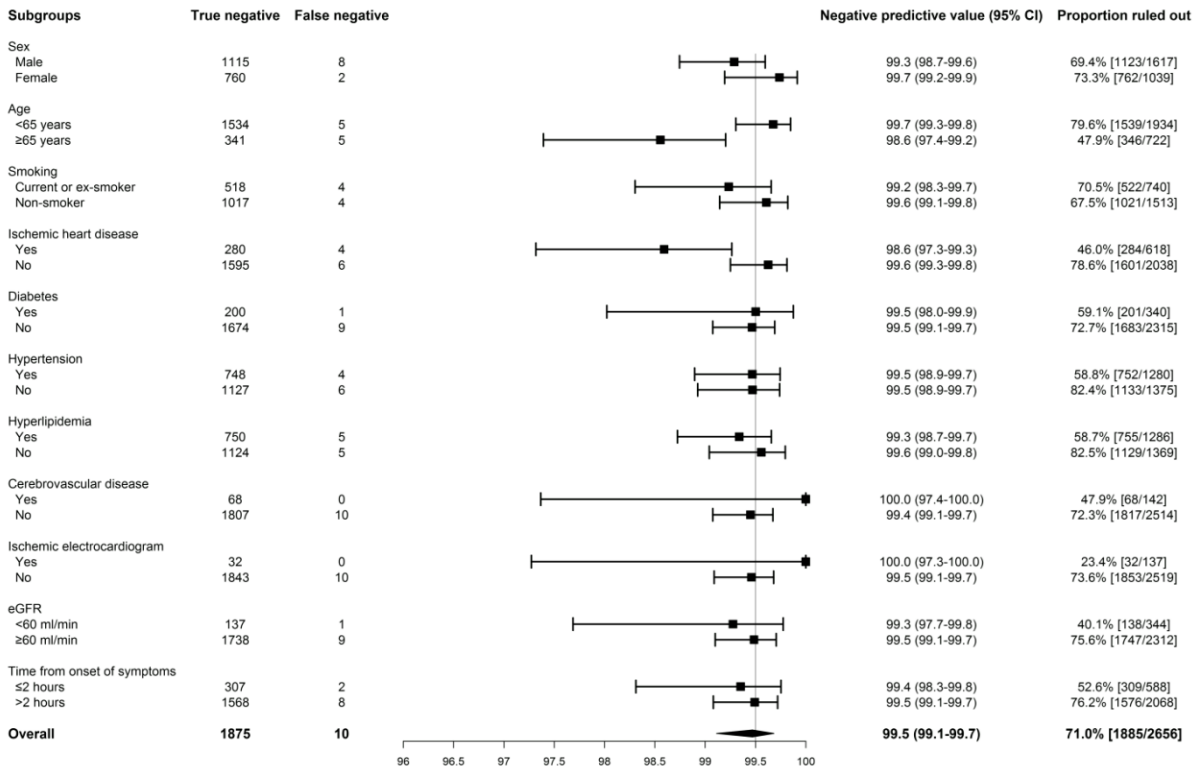


Figure 5.14 Diagnostic performance of the CoDE-ACS score across patient subgroups in the external validation cohort. Negative predictive value of CoDE-ACS using the presentation troponin concentration alone across patient subgroups.

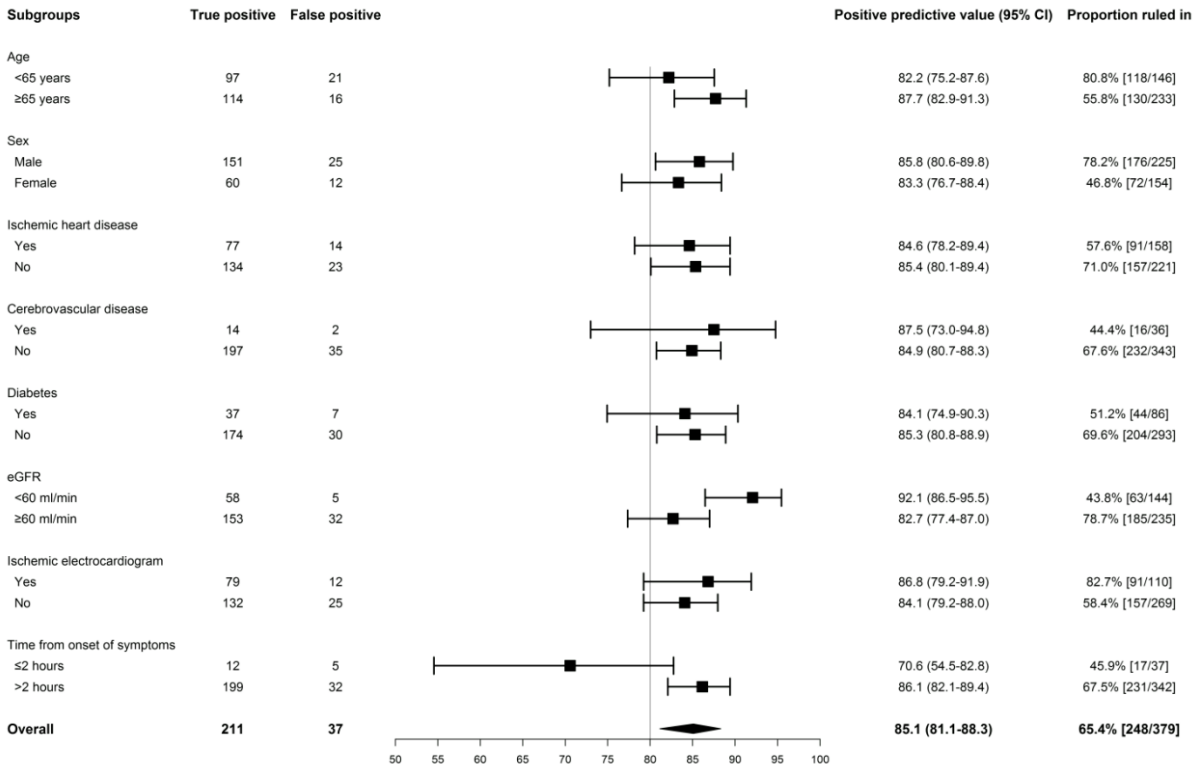


Figure 5.15 Diagnostic performance of the CoDE-ACS score across patient subgroups in the external validation cohort. Positive predictive value of CoDE-ACS using the presentation troponin concentration alone across patient subgroups.

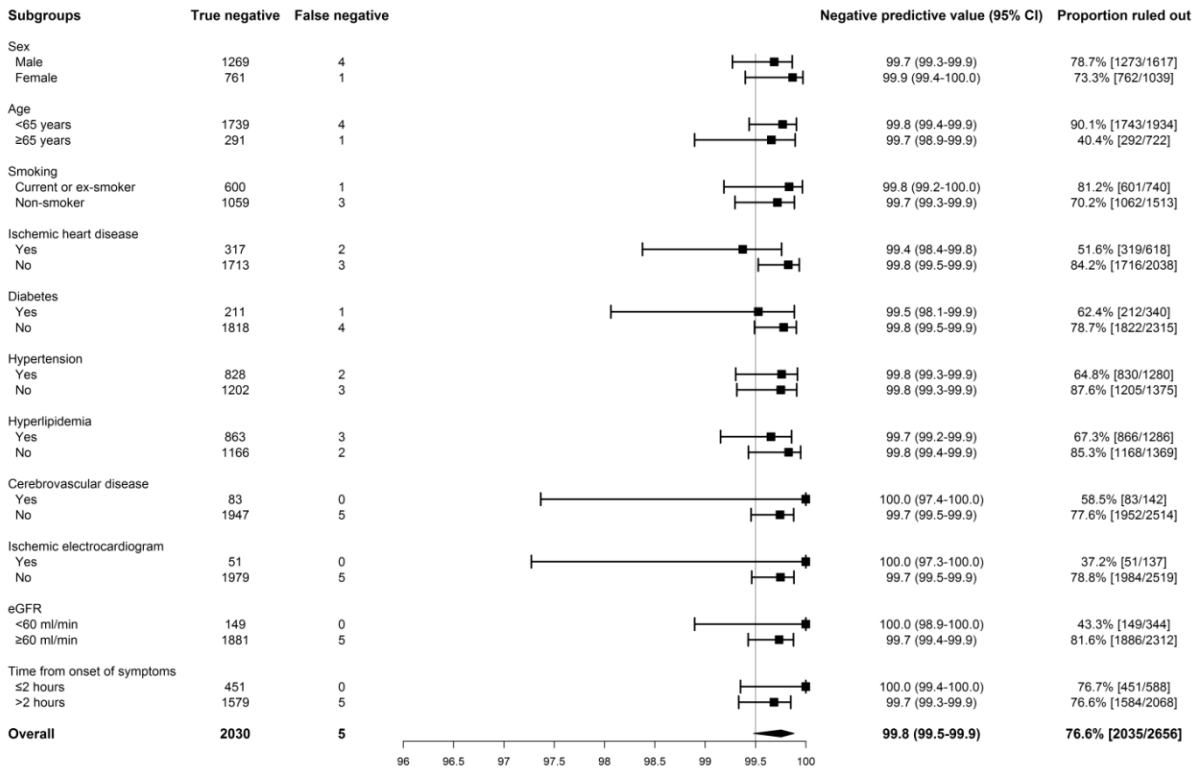


Figure 5.16 Diagnostic performance of the CoDE-ACS score across patient subgroups in the external validation cohort. Negative predictive value of CoDE-ACS using serial cardiac troponin concentrations across patient subgroups.

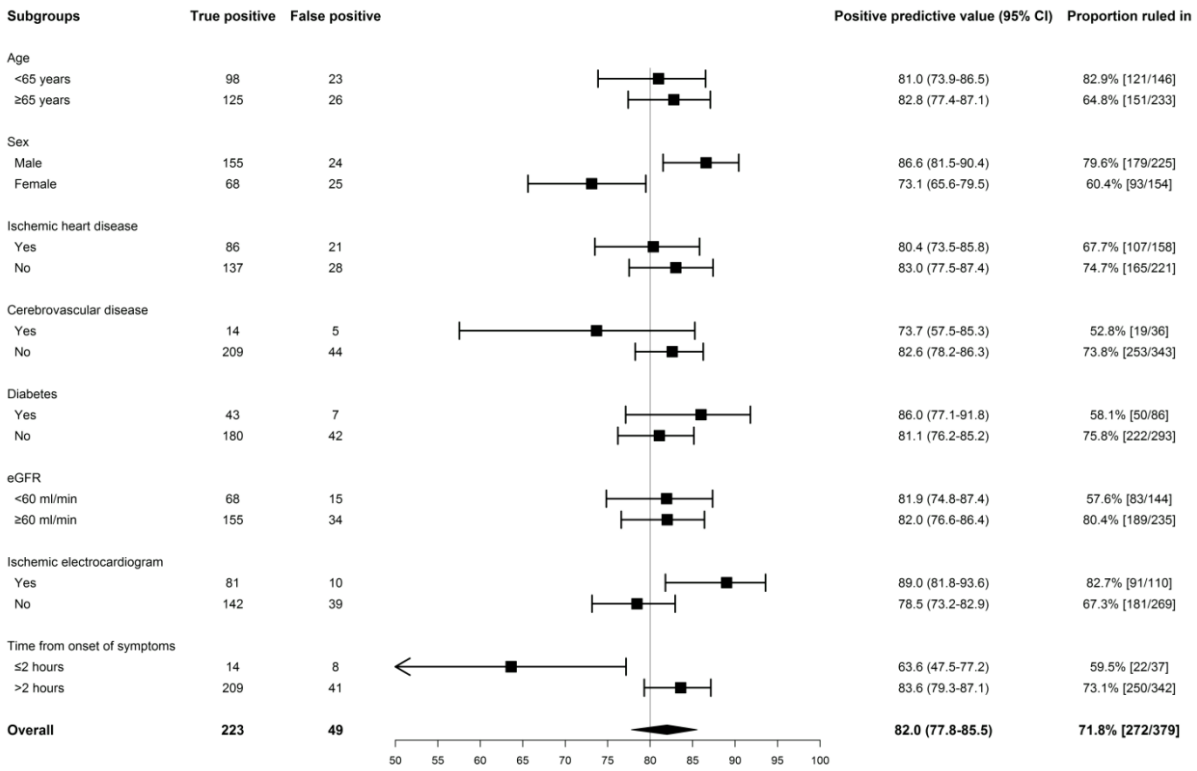


Figure 5.17 Diagnostic performance of the CoDE-ACS score across patient subgroups in the external validation cohort. Positive predictive value of CoDE-ACS using serial cardiac troponin concentrations across patient subgroups.

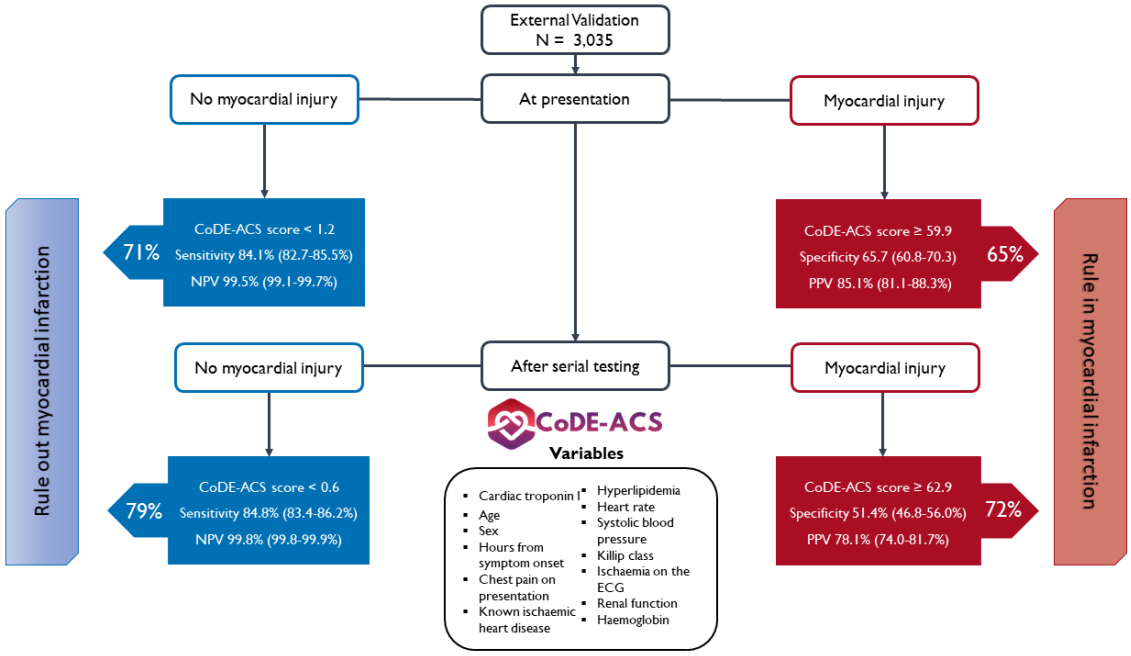


Figure 5.18 Performance of CoDE-ACS as a pathway in the external validation cohort.

Patients who were identified as low-probability for myocardial infarction at presentation had a lower rate of all-cause and cardiac mortality compared to those with an intermediate- and high-probability at 30-days (all-cause mortality: 0.0% *versus* 0.7% and 2.0% respectively; cardiac mortality: 0% *versus* 0.1% and 1.8%) and 1 year (all-cause mortality: 0.6% v*ersus* 5.1% and 8.3% respectively; cardiac mortality: 0.1% *versus* 2.2% and 3.9%, respectively) (Figure 5.19). We observed similar associations in the model incorporating serial troponin measurements (Figure 5.20).

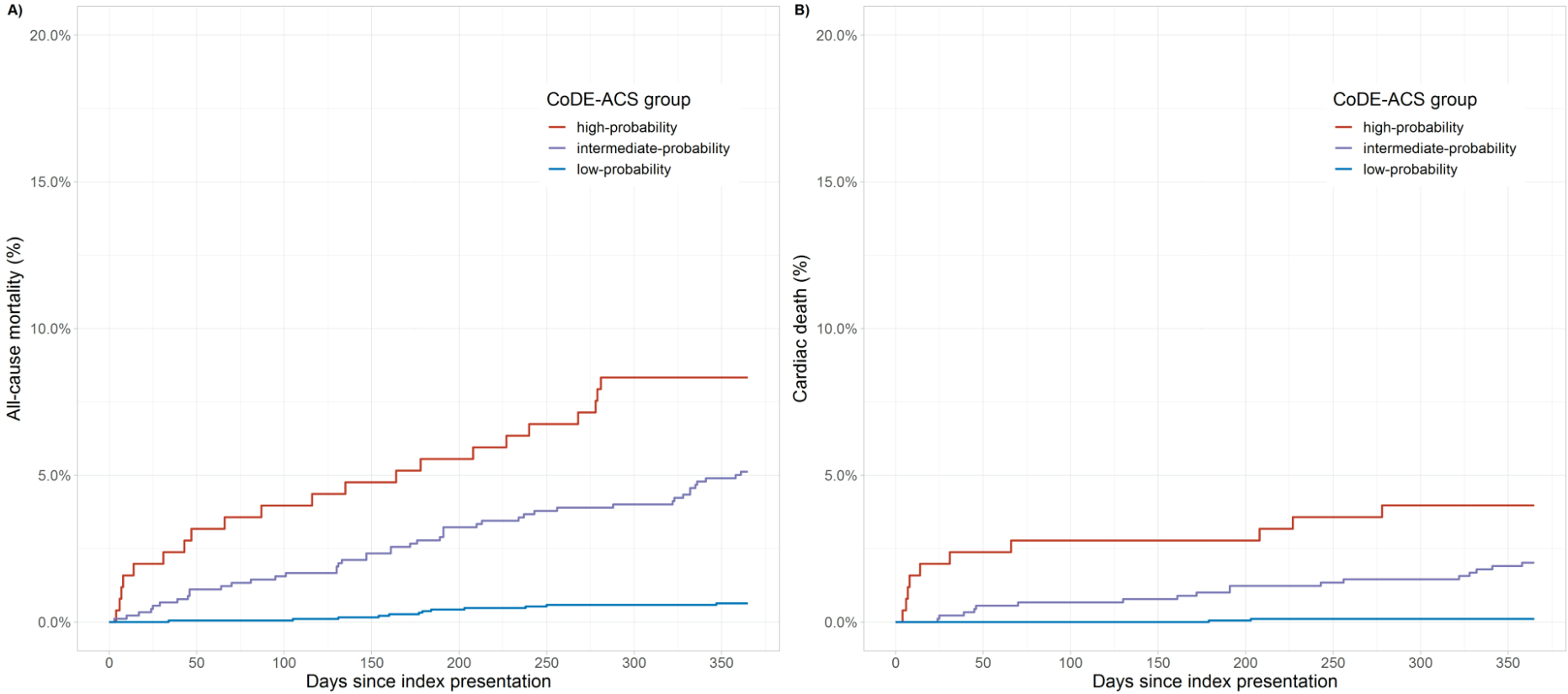


Figure 5.19 Cumulative incidence of mortality in the external validation cohort stratified by CoDE-ACS probability group.
A) All-cause mortality. B) Cardiac mortality.

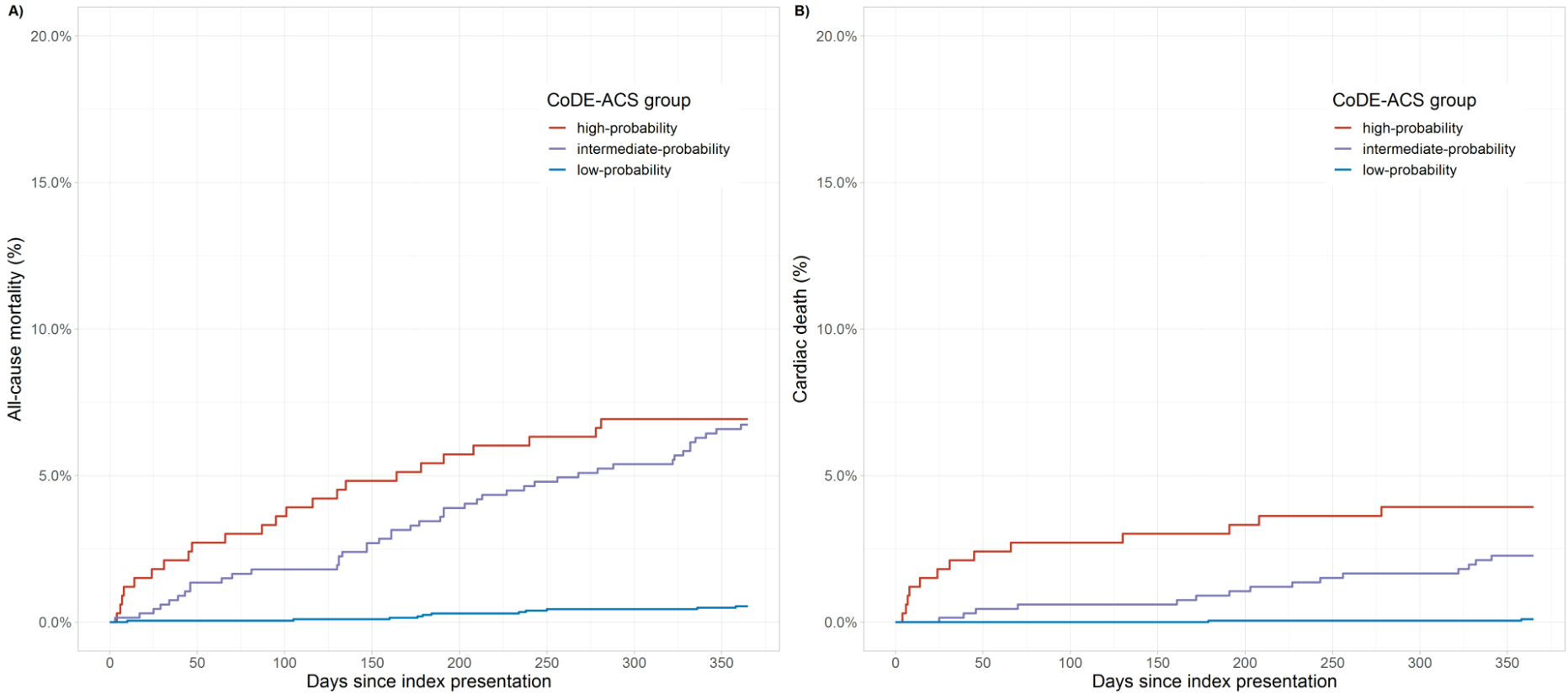


Figure 5.20 Cumulative incidence plot stratified by CoDE-ACS probability groups after serial measurements in the external validation cohort. A) All-cause mortality B) Cardiac mortality.

5.5 Discussion

In a large cohort of consecutive patients presenting with suspected acute coronary syndrome, we developed a decision-support tool called CoDE-ACS using machine learning to aid in the diagnosis of myocardial infarction. Compared to guideline-recommended cardiac troponin thresholds to rule-in and rule-out myocardial infarction, CoDE-ACS had more consistent diagnostic performance across important patient subgroups, such as those with reduced renal function and those who present early following the onset of symptoms. External validation of CoDE-ACS demonstrated excellent diagnostic performance in patients presenting to a different healthcare system and across important subgroups.

5.5.1 CoDE-ACS performance

The use of statistical modelling has several important advantages over the use of fixed troponin thresholds alone. Cardiac troponin is known to be influenced by various factors such as sex, age, renal function and comorbidity burden,[71,146,163] which explains the heterogeneity in diagnostic performance in patient subgroups observed in this study. Furthermore, patients with different patterns of comorbidity have different pre-test probability of having myocardial infarction. We observed that CoDE-ACS provided consistent diagnostic performance across these patient subgroups by incorporating patient factors using machine learning. In contrast to current clinical pathways, where patients presenting within 3 hours of symptom onset or with myocardial ischemia on the electrocardiogram are excluded, CoDE-ACS can be applied in all patients presenting to the Emergency Department irrespective of their symptom onset or risk profile. Furthermore, machine learning enables serial

troponin measurements to be performed at a second flexible timepoint. At present, all guideline recommended pathways require serial cardiac troponin measurement at fixed timepoints, which is challenging to implement precisely in clinical practice. Whilst previous studies have showed that adherence to these pathways was high, one in five to one in three patients did not undergo troponin testing in accordance with the pathway recommendations which may significantly impact on the accuracy and safety of patient assessment.[164,165] Finally, pathways that rely on fixed thresholds may perform variably across healthcare systems due to differences in the way cardiac troponin testing is performed and differences in the prevalence of myocardial infarction. The advantage of using a decision-support tool, such as CoDE-ACS, that generates probabilities of myocardial infarction for individual patients and estimates the diagnostic performance associated with these probabilities is that healthcare systems can apply the algorithm more flexibly. For example, in a healthcare setting that is more conservative, higher CoDE-ACS values could be used to maximise the negative predictive value or in those healthcare settings with less capacity for assessment in the Emergency Department CoDE-ACS values could be applied that reduce the number of patients who are neither ruled out or ruled in but triaged to a period of observation.

5.5.2 Limitations of the MI^3

Recently the first attempt to use a machine learning algorithm for the diagnosis of myocardial infarction was reported. The myocardial-ischemic-injury-index (MI^3) score was developed using gradient boosting, to compute an individualised probability of myocardial infarction for patients with suspected acute coronary syndrome.[142] Whilst

this algorithm overcame several issues with fixed cardiac troponin thresholds, there are important limitations that may limit its implementation. First, MI^3 requires serial cardiac troponin measurements for both the rule-in and rule-out of myocardial infarction which precludes the use of this algorithm during the initial patient assessment. This would significantly limit the efficiency of MI^3 since assessment pathways for patients with suspected acute coronary syndrome currently recommend the use of a single cardiac troponin measurement at presentation to risk stratify patients; an approach that has been shown to be safe and effective shortening duration of stay.[164,165] Second, the MI^3 score was calculated using only age, sex and cardiac troponin concentration. Although the use of these limited and widely available variables may facilitate its implementation due to simplicity, this has also limited its diagnostic performance by not including other important patient factors that influence cardiac troponin. Moreover, specificity and positive predicted value was significantly lower in important subgroups such as older patients, women and those with significant comorbidities such as chronic kidney disease. Finally, MI^3 was developed in a relatively small cohort of selected patients. We recently performed an external validation of the MI^3 algorithm and observed it had poor calibration when applied to a cohort of consecutive patient with suspected acute coronary syndrome. CoDE-ACS overcomes these limitations by including other patient factors that influence cardiac troponin concentration, allowing the use of a single measure of cardiac troponin at presentation and by training the model in a large, unselected patient population.

5.5.3 Using CoDE-ACS to identify errors in ground truth

Moreover, the use of a statistical model, such as CoDE-ACS, for the diagnosis of acute myocardial infarction, could also help with the in-depth review of the clinical cases where the adjudicated diagnosis is different than the clinical diagnosis made by the attending clinician. This is a common problem when it comes in the development of different machine learning algorithm in the field of medicine where the restricted samples size, in combination with the not so accurate labels some times, results in not so precise predictions. When the ground truth is a subjective measurement, such as a difficult diagnosis and not a hard as death, there can be errors in it. We should use the statistical model in combination with clinical expertise and readjudicate those cases when the diagnosis is different between the attending clinician, the adjudicator, and the statistical model. This could result in more accurate datasets, which will result in more accurate diagnostic and prediction models.

5.5.4 Study limitations

We acknowledge several limitations of this analysis. CoDE-ACS was developed and validated using a high-sensitivity cardiac troponin I assay from a single manufacturer. Given cardiac troponin assays are not standardised across different manufacturers, CoDE-ACS will need to be trained and validated across each different assays seperately. Furthermore, there were significant differences in the demographic and patient characteristics between our derivation and external validation cohorts. This is likely to be due to variation in patient selection for troponin testing in different healthcare systems and differences in the way participants were recruited into the different studies. Nevertheless, we demonstrated that CoDE-ACS

had excellent performance in the external validation cohorts despite these significant differences. Finally, the diagnostic performance of CoDE-ACS was evaluated retrospectively and therefore was not used to guide clinical management in this cohort. Prospective validation studies and an evaluation of diagnostic performance following implementation into practice is warranted. Nonetheless, our analysis of longer-term outcomes suggest that CoDE-ACS can appropriately risk-stratifying patients at presentation who are likely to benefit from further specialist investigation and treatment.

5.6 Conclusions

In conclusion, we have developed and externally validated a decision support tool using machine learning to aid in the diagnosis of acute myocardial infarction. The CoDE-ACS algorithm was well calibrated and had excellent discrimination in the internal and external validation cohorts. Furthermore, it has more consistent performance across patient subgroups compared to fixed guideline-recommended thresholds of cardiac troponin and has important advantages compared to previous diagnostic algorithms. Prospective validation is needed to evaluate whether our decision-support tool improves diagnostic accuracy compared to current clinical pathways for the assessment of patients with suspected acute coronary syndrome.

This chapter marks the end of the first half of this thesis. A previous developed algorithm was evaluated for the first time in a heterogeneous population and a newly and improved algorithm was also developed in this chapter. In the second half of this thesis, we will develop a similar machine learning algorithm, but this time to assist with the diagnosis of acute heart failure.

CHAPTER 6

Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study

Doudesis D, Lee KK*, Anwar M*, et al. Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study.*

BMJ 2022;377:e068424.

DOI: 10.1136/bmj-2021-068424.

***** Contributed equally

Chapter 6 Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study

6.1 Overview

The scope of this chapter is to evaluate the diagnostic performance of N-terminal pro-B-type natriuretic peptide (NT-proBNP) thresholds for acute heart failure, and to develop and validate a decision-support tool that combines NT-proBNP concentrations with clinical characteristics. The data that used was from an individual patient-level meta-analysis from 14 studies from 13 countries, including randomised controlled trials and prospective observational studies.

Researchers collaborated to pool individual patient level data for 10,369 patients with suspected acute heart failure were pooled for the meta-analysis to evaluate NT-proBNP thresholds. A decision support tool (Collaboration for the Diagnosis and Evaluation of Heart Failure (CoDE-HF)) that combines NT-proBNP with clinical variables to report the probability of acute heart failure for an individual patient was developed and validated. The main outcome measure was the adjudicated diagnosis of acute heart failure.

Overall, 43.9% (4549/10 369) of patients had an adjudicated diagnosis of acute heart failure (73.3% (2286/3119) and 29.0% (1802/6208) in those with and without previous heart failure, respectively). The negative predictive value of the guideline recommended rule-out threshold of 300 pg/mL was 94.6% (95% confidence interval 91.9% to 96.4%); despite use of age specific rule-in thresholds, the positive predictive value varied at 61.0% (55.3% to 66.4%), 73.5% (62.3% to 82.3%), and 80.2% (70.9% to 87.1%), in patients aged <50 years, 50-75 years, and >75 years, respectively. Performance varied in most subgroups, particularly patients with obesity, renal impairment, or previous heart failure. CoDE-HF was well calibrated, with excellent discrimination in patients with and without previous heart failure (area under the receiver operator curve 0.846 (0.830 to 0.862) and 0.925 (0.919 to 0.932) and Brier scores of 0.130 and 0.099, respectively). In patients without previous heart failure, the diagnostic performance was consistent across all subgroups, with 40.3% (2502/6208) identified at low probability (negative predictive value of 98.6%, 97.8% to 99.1%) and 28.0% (1737/6208) at high probability (positive predictive value of 75.0%, 65.7% to 82.5%) of having acute heart failure.

In an international, collaborative evaluation of the diagnostic performance of NT-proBNP, guideline recommended thresholds to diagnose acute heart failure varied substantially in important patient subgroups. The CoDE-HF decision support tool incorporating NT-proBNP as a continuous measure and

other clinical variables provides a more consistent, accurate, and individualized approach.

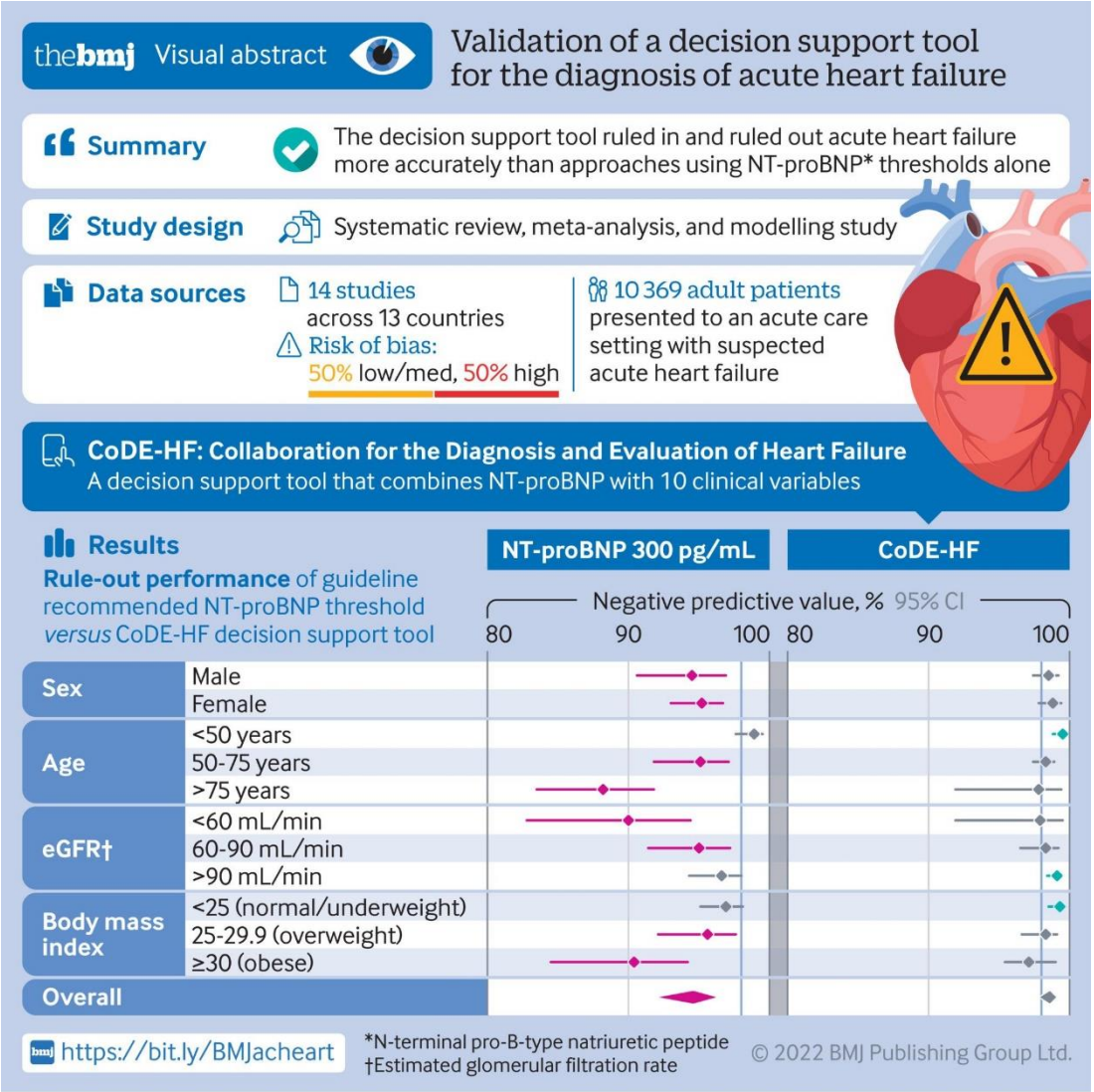


Figure 6.1 Visual abstract © 2022 BMJ Publishing Group Ltd.

6.2 Introduction

Nearly 1 million people are currently living with heart failure in the United Kingdom and the prevalence is projected to rise by approximately 50% over the next 25 years due to our ageing population.[166] Currently, decompensated acute heart failure account for 5% of all unplanned hospital admissions.[167] The accurate and timely diagnosis of acute heart failure can be challenging, and both national and international guidelines recommend natriuretic peptide testing to aid in the diagnosis.[33,80,103,107,168] Despite these recommendations, NT-proBNP testing has not been universally implemented, in part, due to concerns about clinical utility in a real-world setting. Studies investigating the diagnostic performance of NT-proBNP have mainly been performed in relatively small, selected patient cohorts, limiting the generalisability of study findings across clinically important subgroups, such as older patients and those with renal disease or obesity; characteristics that are becoming increasingly prevalent in patients with heart failure.[169-171] Statistical modelling approaches that incorporate patients' characteristics to provide a more individualised assessment may have more consistent diagnostic performance across patient subgroups.[142]

Although many models have been developed to predict prognosis in patients with heart failure, very few have been developed to aid in the diagnosis of acute heart failure. [150,172-177] Previous attempts have many strengths, but have incorporated subjective variables, such as the clinicians' estimation of pre-test probability or the patient's description of symptoms. Furthermore,

they have incorporated NT-proBNP as a binary variable, which does not take into account the dynamic and nonlinear interaction between NT-proBNP and other clinical variables. Previous attempts at developing and validating diagnostic scores have also included a limited number of patients from a single healthcare setting, which has precluded the assessment of performance within subgroups and limited external generalizability.

In this collaborative international analysis, we evaluated the diagnostic performance of guideline recommended NT-proBNP thresholds for acute heart failure across patient subgroups. We subsequently developed and validated a decision support tool for patients with suspected acute heart failure that uses statistical modelling to combine NT-proBNP concentrations with clinical characteristics.

6.3 Methods

6.3.1 Study population

We performed a systematic review to identify studies that evaluated the diagnostic performance of NT-proBNP in patients with suspected acute heart failure. We searched Embase, Medline and the Cochrane Central Register of Controlled Trials to update a previous review by Roberts et al[1] with studies published up to 18th August 2021 using the following keywords: 'heart failure' and 'natriuretic peptide'. Studies were eligible if they met the following prespecified inclusion criteria: (1) enrolled patients ≥18 years with suspected acute heart failure in an acute care setting, (2) measured NT-proBNP on blood samples obtained during the patients' initial assessment on the day of the hospital attendance, and (3) adjudicated diagnosis of acute heart failure using an acceptable reference standard. All studies identified in the systematic literature search were independently screened by two investigators and conflicts adjudicated by a third using a pre-specified protocol (PROSPERO register: CRD42019159407).

The corresponding authors of all eligible cohorts were contacted to request anonymized individual patient-level data on NT-proBNP concentrations, adjudicated diagnosis of acute heart failure, demographics (age, sex, ethnicity), past medical history (heart failure, ischaemic heart disease, diabetes, hypertension, hyperlipidemia, smoking, asthma, chronic obstructive pulmonary disease, chronic kidney disease), physiological variables (heart

rate and blood pressure), and clinical haematology and biochemistry profiles. Accuracy and completeness were checked with all corresponding authors prior to harmonization. All studies were conducted in accordance with the Declaration of Helsinki and with ethical approval to permit sharing of individual patient-level data to conduct this meta-analysis. Risk of bias for each study was assessed using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) tool[178] by two investigators independently, with conflicts resolved by a third.

6.3.2 NT-proBNP threshold analysis

Meta-estimates (95% confidence intervals) of the sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) of the guideline-recommended NT-proBNP rule-out threshold[33,97] (300 pg/mL) and age-specific rule-in thresholds[168] (450, 900, and 1,800 pg/mL for those <50 years, 50-75 years, and >75 years, respectively) for acute heart failure were derived using a two-stage approach, with estimates calculated separately within each study, then pooled across studies by random effects meta-analysis. We further evaluated the performance of these thresholds in prespecified subgroups stratified by age, sex, ethnicity, body mass index, renal function, anaemia and the presence of comorbidities (prior heart failure, hypertension, hyperlipidemia, diabetes mellitus, atrial fibrillation, chronic obstructive pulmonary disease). Using the same approach, we subsequently evaluated the diagnostic performance of NT-proBNP concentrations across a range of concentrations to determine a rule-out threshold that would identify

the highest proportion of patients as low-probability for an NPV ≥98% and a rule-in threshold that would identify the highest proportion of patients as high-probability for a PPV ≥75%.

6.3.3 Model development and validation

We developed and externally validated a decision-support tool (Collaboration for the Diagnosis and Evaluation of Heart Failure [CoDE-HF], https://decision-support.shinyapps.io/code-hf/) using statistical modelling to compute a value (0-100) that corresponds to an individual patient's probability of acute heart failure.

Due to significant differences in comorbidities and the prevalence of acute heart failure, models were developed and validated for patients with and without prior heart failure separately. We used NT-proBNP concentrations as a continuous measure and selected simple objective clinical variables that are known to be associated with acute heart failure, which were found to have the highest relative importance in our model training phase (age, estimated glomerular filtration rate, haemoglobin, body mass index, heart rate, blood pressure, peripheral oedema, prior history of heart failure, chronic obstructive pulmonary disease and ischaemic heart disease).

We evaluated four different statistical models in the development of CoDE-HF – generalised linear mixed-model, random forest, naïve Bayes and extreme gradient boosting (XGBoost).[122,123,179] To account for missing data

across studies in the training cohort, we multiply imputed ten datasets using joint-modelling multiple imputation with random study specific covariance matrices fitted with a Markov chain Monte Carlo algorithm.[133] We performed ten iterations of 10-fold cross-validation for each model and the median score across the iterations and imputed datasets was used as the CoDE-HF score for each patient. We subsequently identified the score in the training cohort that would classify the highest proportion of patients as high- or low-probability of acute heart failure with optimal performance to rule-in (75% PPV and 90% specificity) and rule-out (98% NPV and 90% sensitivity). The performance of each model was assessed across a range of diagnostic metrics (area under the curve [AUC], Brier score, proportion of patients achieving the optimal high- and low-probability criteria, and PPV and NPV across patient subgroups). Brier score is a measure of both discrimination and calibration, and is calculated by taking the mean squared difference between predicted probabilities and the observed outcome. The best performing model was selected for the CoDE-HF decision-support tool. We performed a decision curve analysis and internal-external cross-validation to evaluate the performance of CoDE-HF. In brief, this approach iteratively leaves one study out at a time for external validation and uses the remaining studies for model development. We did not perform imputation in the external validation dataset and therefore external validation was not performed in studies where the majority of variables are completely missing. All analyses were performed in R version 4.1.2.

6.4 Results

6.4.1 Study population

Investigators from 30 eligible studies were contacted, of which, 19 responded. Fourteen studies (12 prospective cohort studies and 2 randomised controlled trials) provided individual patient-level data in 10,369 patients with suspected acute heart failure (mean age 69.3 years, 53.3% male) from 13 countries (Table 6.1, Figure *6.2*, Table 6.2 and Table 6.3).[98,100,173,180-190] All studies were performed in the Emergency Department except one which included patients admitted to the Cardiology and Pulmonology Department (median number of patients in each study was 488 [inter-quartile range 322-1,053]). Overall, 43.9% (4,549/10,369) of patients had an adjudicated diagnosis of acute heart failure. Patients with a prior history of heart failure had a higher prevalence of acute heart failure than those without (73.3% *versus* 29.0% respectively) (Table 6.4 and Table 6.5).

Table 6.1 Baseline characteristics of patients stratified by diagnosis of acute heart failure.

	Overall	Patients with acute heart failure	Patients without acute heart failure
Number of participants	10369	4549	5820
Men	5531 (53.3)	2568 (56.5)	2963 (50.9)
Age, years	69.3 (16.3)	75.0 (12.6)	64.9 (17.5)
<50	1377 (13.3)	222 (4.9)	1155 (19.8)
50-75	4370 (42.1)	1674 (36.8)	2696 (46.3)
>75	4622 (44.6)	2653 (58.3)	1969 (33.8)
Ethnicity			
Black	845 (14.8)	316 (12.5)	529 (16.7)
Caucasian	4112 (72.1)	2028 (80.1)	2084 (65.8)
Other	743 (13.0)	188 (7.4)	555 (17.5)
Past medical history			
Prior heart failure	3119 (33.4)	2286 (55.9)	833 (15.9)
Ischaemic heart disease	2953 (32.3)	1871 (46.8)	1082 (21.0)
Diabetes mellitus	2398 (26.7)	1382 (34.8)	1016 (20.3)
Hypertension	5071 (59.3)	2603 (71.0)	2468 (50.5)
Hyperlipidemia	2269 (41.2)	1160 (50.8)	1109 (34.5)
Current or ex-smoker	2458 (41.3)	918 (37.8)	1540 (43.8)
Asthma	770 (18.5)	98 (6.9)	672 (24.6)
COPD	2117 (29.2)	670 (22.9)	1447 (33.5)
Atrial fibrillation	1701 (20.9)	1243 (33.3)	458 (10.4)
Chronic kidney disease	1215 (18.9)	877 (33.6)	338 (8.8)
Body mass index, kg/m^2	27.7 (7.2)	27.7 (6.8)	27.7 (7.6)
<25	3062 (39.0)	1349 (38.2)	1713 (39.6)
25-29	2473 (31.5)	1172 (33.2)	1301 (30.1)
≥30	2317 (29.5)	1007 (28.5)	1310 (30.3)
Physiological parameters			
Heart rate, beats per minute	91.7 (23.7)	91.5 (25.9)	91.9 (21.9)
Systolic blood pressure, mmHg	140.0 (27.9)	140.2 (30.0)	139.9 (26.4)
Diastolic blood pressure, mmHg	79.7 (17.0)	80.3 (18.3)	79.3 (15.9)

Clinical haematology and biochemistry			
Haemoglobin, g/dL	13.1 (2.1)	12.7 (2.1)	13.4 (2.0)
eGFR, mL/min/1.73 m^2	68.2 (31.3)	56.8 (27.1)	77.2 (31.6)
NT-proBNP, pg/mL	1182 [191 to 4737]	4362 [1883 to 9883]	279 [70 to 1054]

Presented as No. (%), mean (SD) or median [inter-quartile range].
Abbreviations: COPD= chronic obstructive pulmonary disease; eGFR= estimated glomerular filtration rate.

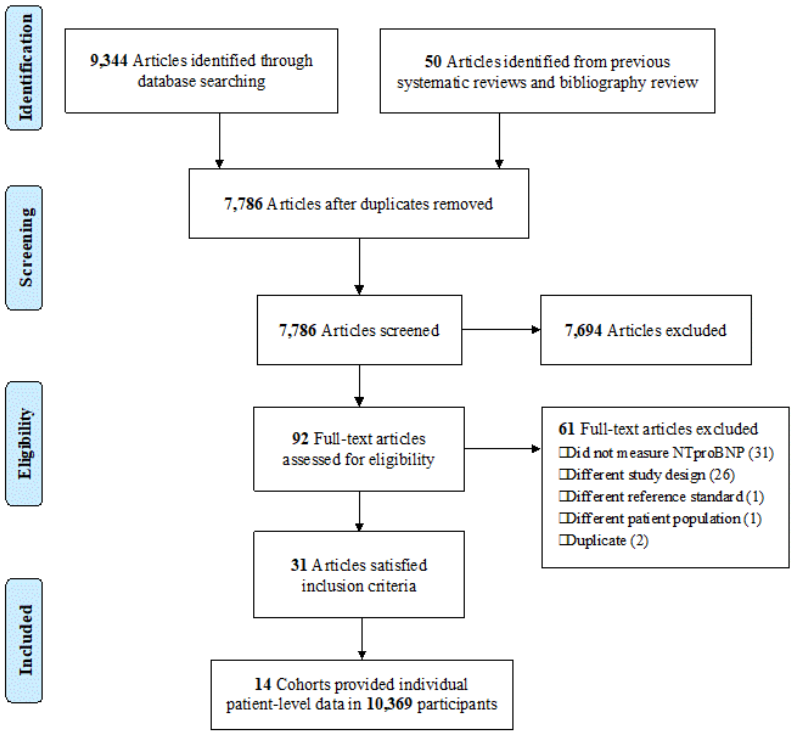


Figure 6.2 Flow diagram of study participants.

Table 6.2 Characteristics of each included study.

Author, year	NT-proBNP assay	Study design	Cohort size	Setting
Bahrmann et al, 2015	Roche	Prospective cohort	303	Emergency department
Behnes et al, 2009	Dade Behring	Prospective cohort	401	Emergency department
Bombelli et al, 2015	Roche	Prospective cohort	895	Emergency department
Chenevier-Gobeaux et al, 2005	Roche	Prospective cohort	380	Emergency department
deFilippi et al, 2007	Roche	Prospective cohort	831	Emergency department
Gargani et al, 2008	Roche	Prospective cohort	149	Cardiology and pulmonary hospital admissions
Ibrahim et al, 2017	Roche	Prospective cohort	1106	Emergency department
Januzzi et al, 2006	Roche	Prospective cohort	1256	Emergency department
Maisel et al, 2010	Roche	Prospective cohort	1623	Emergency department

Moe et al, 2007	Roche	Randomized controlled trial	500	Emergency department
Mueller et al, 2005	Roche	Prospective cohort	251	Emergency department
Nazerian et al, 2010	Roche	Prospective cohort	145	Emergency department
Rutten et al, 2008	Roche	Randomized controlled trial	476	Emergency department
Wussler et al, 2019[19]	Roche	Prospective cohort	2053	Emergency department

(continues)

Author, year	Diagnostic adjudication for acute heart failure	Risk of bias (QUADAS-2)
Bahrmann et al, 2015	Independent adjudication by two cardiologists based on the definition of the ESC guideline. They reviewed all available medical records of the index hospital stay, including the clinical history findings from the physical examination, results of laboratory	Patient selection: high; index test: low; reference standard: low; flow and timing: low; Overall: high.

	tests (excluding NT-proBNP), radiographic studies, ECG, and echocardiography.	
Behnes et al, 2009	Retrospective review by a study physician who had unrestricted access to the records of the patients but was blinded to the results of NT-proBNP measurements.	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.
Bombelli et al, 2015	Adjudication by clinician based on Framingham criteria.	Patient selection: high; index test: low; reference standard: low; flow and timing: low; Overall: high.
Chenevier-Gobeaux et al, 2005	Independent adjudication by two ED clinicians on the basis of clinical examination, medical history, ECG, chest X-ray and blood analysis (including plasma creatinine).	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.
deFilippi et al, 2007	Adjudication by cardiologist who reviewed case report forms blinded to natriuretic peptide results.	Patient selection: high; index test: low; reference standard: low; flow and timing: low; Overall: high.

	Subset of 50 random cases adjudicated by second cardiologist, demonstrating good agreement.	
Gargani et al, 2008	Independent adjudication by two cardiologists who were blinded to the NT-proBNP values, who reviewed all of the medical records pertaining to the patient.	Patient selection: high; index test: low; reference standard: low; flow and timing: high; Overall: high.
Ibrahim et al, 2017	Independent adjudication by an ED specialist and a cardiologist. They were blinded to NT-proBNP measurements but could access medical records, case report forms, and other test results including cardiac imaging as available.	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.
Januzzi et al, 2006	Adjudicated by a panel of physicians or cardiologists utilizing all available clinical data, blinded to NT-proBNP results.	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.

Maisel et al, 2010	Independent adjudication by two cardiologists who reviewed all medical records.	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.
Moe et al, 2007	Independent adjudication by two cardiologists. They were provided with hospital records, including the discharge summary, results of laboratory and radiographic testing, echocardiograms if performed, clinical notes from the time of ED presentation to the 60-day follow-up, and outcome of a telephone interview. Using all available data, the cardiologists assigned a diagnosis without knowledge of the NT-proBNP results.	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.
Mueller et al, 2005	Retrospective review of all medical records by a clinician based on Framingham criterion	Patient selection: low; index test: low; reference standard: low; flow and timing: low; Overall: low.

Nazerian et al, 2010	Independent adjudication by two cardiologists and one respiratory physician, blinded to echocardiogram and NT-proBNP results. The reviewers had access to ED records, clinical notes, components and summary of the Framingham Heart Study Criteria and any additional information that became available during hospital stay.	Patient selection: high; index test: low; reference standard: low; flow and timing: low; Overall: high.
Rutten et al, 2008[18]	Consensus between two clinicians in internal medicine, pulmonology or cardiology.	Patient selection: low; index test: low; reference standard: high; flow and timing: low; Overall: high.
Wussler et al, 2019[19]	Adjudicated by 2 independent cardiologist-internists who had access to all patients' medical records, including clinical history, physical examination, 12-lead electrocardiograms, laboratory findings, chest radiographs, echocardiograms, lung function test results,	Patient selection: high; index test: low; reference standard: high; flow and timing: low; Overall: high.

	computed tomography scans, and response to therapy, as well as autopsy data for patients who died in the hospital.

Table 6.3 Baseline characteristics of patients within each study.

	Overall	Bahrmann et al	Behnes et al	Bombelli et al	Chenevier-Gobeaux et al	De Filippi et al
Number of participants	10369	303	401	895	380	831
Men	5531 (53.3)	148 (48.8)	205 (51.1)	368 (41.1)	189 (49.7)	380 (45.7)
Age, years	69.3 (16.3)	80.4 (5.9)	67.4 (15.6)	85.7 (4.2)	78.5 (12.2)	66.3 (14.9)
< 50	1377 (13.3)	-	53 (13.2)	-	5 (1.3)	125 (15.0)
50-75	4370 (42.1)	79 (26.1)	211 (52.6)	-	121 (31.8)	443 (53.3)
>75	4622 (44.6)	224 (73.9)	137 (34.2)	895 (100.0)	254 (66.8)	263 (31.6)
Ethnicity						
Black	845 (14.8)	0 (0.0)	NR	NR	NR	318 (38.3)
Caucasian	4112 (72.1)	303 (100.0)	NR	NR	NR	499 (60.0)
Other	743 (13.0)	0 (0.0)	NR	NR	NR	14 (1.7)
Past medical history						
Prior Heart failure	3119 (33.4)	224 (73.9)	194 (48.4)	NR	128 (33.8)	287 (36.1)
Ischemic heart disease	2953 (32.3)	138 (45.5)	157 (39.2)	NR	124 (32.6)	263 (33.1)
Diabetes Mellitus	2398 (26.7)	117 (38.6)	120 (30.1)	NR	NR	305 (38.2)
Hypertension	5071 (59.3)	255 (84.4)	268 (67.0)	NR	153 (40.3)	NR

Hyperlipidemia	2269 (41.2)	206 (68.0)	122 (30.7)	NR	NR	NR
Current or ex-smoker	2458 (41.3)	149 (49.2)	206 (53.9)	NR	NR	NR
Asthma	770 (18.5)	NR	21 (5.2)	NR	NR	NR
COPD	2117 (29.2)	87 (28.7)	94 (23.4)	NR	127 (33.4)	NR
Atrial fibrillation	1701 (20.9)	89 (29.4)	70 (17.5)	NR	NR	175 (22.0)
Chronic Kidney Disease	1215 (18.9)	18 (5.9)	71 (17.7)	NR	NR	NR
Body mass index, kg/m^2	27.7 (7.2)	27.1 (5.0)	27.9 (6.2)	NR	NR	30.3 (9.7)
<25	3062 (39.0)	104 (34.3)	118 (32.6)	NR	NR	195 (31.5)
25-30	2473 (31.5)	121 (39.9)	146 (40.3)	NR	NR	172 (27.8)
≥30	2317 (29.5)	78 (25.7)	98 (27.1)	NR	NR	252 (40.7)
Physiological parameters						
Heart rate, beats per minute	91.7 (23.7)	85.2 (23.2)	91.7 (24.0)	NR	NR	NR
Systolic blood pressure, mmHg	140.0 (27.9)	146.5 (27.3)	139.2 (28.6)	NR	NR	NR
Diastolic blood pressure, mmHg	79.7 (17.0)	74.9 (15.7)	79.0 (13.7)	NR	NR	NR
Clinical hematology and biochemistry						
Hemoglobin, g/dL	13.1 (2.1)	12.4 (2.3)	13.2 (1.9)	NR	NR	NR
eGFR, mL/min/1.73m^2	68.2 (31.3)	48.0 (14.8)	66.8 (24.9)	52.9 (25.4)	54.7 (19.9)	63.3 (34.0)

NT-proBNP, pg/mL	1182.2 [191.0, 4737.0]	1594.0 [433.5, 4976.5]	766.0 [150.6, 3155.1]	2937.0 [1014.0, 7598.5]	1678.0 [409.8, 5875.0]	1731.0 [472.0, 6032.5]
Adjudicated diagnosis of heart failure	4549 (43.9)	168 (55.4)	122 (30.4)	405 (45.3)	115 (30.3)	437 (52.6)

(continues)

	Gargani et al	**Ibrahim et al**	**Januzzi et al**	**Maisel et al**	**Moe et al**	**Mueller et al**
Number of participants	149	1106	1256	1623	500	251
Men	98 (65.8)	683 (61.8)	643 (51.2)	848 (52.2)	258 (51.6)	234 (93.2)
Age, years	70.8 (11.0)	62.1 (16.2)	68.4 (15.9)	63.8 (16.9)	70.7 (14.3)	70.2 (14.0)
< 50	5 (3.4)	232 (21.0)	183 (14.6)	332 (20.5)	46 (9.2)	24 (9.6)
50-75	91 (61.1)	601 (54.3)	554 (44.1)	800 (49.3)	231 (46.2)	115 (45.8)
>75	53 (35.6)	273 (24.7)	519 (41.3)	491 (30.3)	223 (44.6)	112 (44.6)
Ethnicity						
Black	0 (0.0)	0 (0.0)	46 (3.7)	471 (29.3)	10 (2.0)	0 (0.0)
Caucasian	149 (100.0)	461 (41.7)	1210 (96.3)	1078 (67.0)	464 (93.0)	251 (100.0)
Other	0 (0.0)	645 (58.3)	0 (0.0)	59 (3.7)	25 (5.0)	0 (0.0)
Past medical history						
Prior Heart failure	46 (30.9)	235 (21.3)	429 (34.2)	563 (35.6)	171 (38.2)	75 (29.9)

Ischemic heart disease	60 (40.3)	110 (11.8)	526 (42.0)	498 (31.7)	135 (31.1)	117 (46.6)
Diabetes Mellitus	57 (38.3)	293 (26.5)	309 (24.6)	457 (28.5)	113 (26.2)	58 (23.1)
Hypertension	79 (53.0)	601 (54.3)	661 (52.6)	1069 (66.9)	262 (58.9)	141 (56.2)
Hyperlipidemia	57 (38.3)	472 (43.1)	NR	NR	NR	NR
Current or ex-smoker	13 (9.6)	245 (22.2)	721 (57.5)	463 (29.4)	311 (65.8)	46 (100.0)
Asthma	2 (1.3)	282 (25.6)	NR	313 (19.9)	88 (19.6)	NR
COPD	33 (22.1)	254 (23.0)	NR	465 (29.5)	126 (30.3)	72 (28.7)
Atrial fibrillation	30 (27.0)	164 (16.1)	280 (22.3)	240 (16.8)	NR	83 (33.1)
Chronic Kidney Disease	43 (29.1)	121 (11.0)	NR	244 (15.6)	NR	74 (29.5)
Body mass index, kg/m^2	27.7 (5.1)	27.4 (7.5)	27.4 (6.3)	29.2 (8.8)	28.5 (7.4)	26.8 (5.0)
<25	42 (30.7)	308 (43.0)	412 (37.4)	490 (35.4)	157 (33.0)	93 (37.1)
25-30	54 (39.4)	195 (27.2)	375 (34.0)	384 (27.7)	162 (34.0)	100 (39.8)
>=30	41 (29.9)	213 (29.7)	316 (28.6)	511 (36.9)	157 (33.0)	58 (23.1)
Physiological parameters						
Heart rate, beats per minute	78.1 (16.6)	93.2 (23.0)	91.5 (24.9)	91.4 (22.8)	86.5 (21.7)	93.8 (26.1)
Systolic blood pressure, mmHg	131.7 (23.4)	139.9 (26.9)	NR	140.9 (28.7)	136.1 (25.4)	140.1 (31.9)
Diastolic blood pressure, mmHg	73.1 (10.8)	79.9 (16.1)	NR	80.8 (17.4)	76.8 (17.0)	81.8 (16.3)
Clinical hematology and biochemistry						
Hemoglobin, g/dL	13.4 (0.4)	13.4 (2.0)	13.0 (2.0)	12.9 (2.2)	13.0 (1.8)	13.7 (2.0)

eGFR, mL/min/1.73m^2	59.9 (23.1)	72.9 (27.7)	72.5 (31.9)	70.4 (30.4)	76.4 (30.1)	94.0 (44.6)
NT-proBNP, pg/mL	2303.0 [430.0, 5034.0]	531.8 [87.7, 3149.1]	1403.5 [150.1, 6284.5]	832.9 [111.5, 4114.5]	1293.5 [212.8, 4199.5]	1222.0 [269.4, 3981.5]
Adjudicated diagnosis of heart failure	122 (81.9)	327 (29.6)	720 (57.3)	562 (34.6)	230 (46.0)	137 (54.6)

(continues)

	Nazerian et al	**Rutten et al**	**Wussler et al**
Number of participants	145	476	2053
Men	74 (51.0)	257 (54.0)	1146 (55.8)
Age, years	77.8 (11.2)	58.6 (17.8)	70.5 (15.5)
< 50	4 (2.8)	144 (30.3)	224 (10.9)
50-75	43 (29.7)	239 (50.2)	842 (41.0)
>75	98 (67.6)	93 (19.5)	987 (48.1)
Ethnicity			
Black	NR	NR	NR
Caucasian	NR	NR	NR
Other	NR	NR	NR
Past medical history			
Prior Heart failure	30 (20.7)	85 (17.9)	652 (32.0)

Ischemic heart disease	47 (32.4)	100 (21.0)	678 (33.2)
Diabetes Mellitus	25 (17.2)	77 (16.2)	467 (22.8)
Hypertension	84 (57.9)	120 (25.2)	1378 (67.5)
Hyperlipidemia	NR	NR	846 (41.9)
Current or ex-smoker	NR	304 (64.5)	NR
Asthma	NR	64 (13.4)	NR
COPD	49 (33.8)	126 (26.5)	684 (33.4)
Atrial fibrillation	50 (34.5)	70 (19.2)	450 (21.9)
Chronic Kidney Disease	15 (10.3)	41 (8.6)	588 (28.7)
Body mass index, kg/m^2	NR	25.6 (5.7)	26.6 (6.1)
<25	NR	247 (52.2)	896 (44.2)
25-30	NR	138 (29.2)	626 (30.9)
>=30	NR	88 (18.6)	505 (24.9)
Physiological parameters			
Heart rate, beats per minute	98.3 (21.2)	97.4 (23.9)	92.6 (23.9)
Systolic blood pressure, mmHg	137.6 (28.2)	147.9 (33.2)	138.5 (26.1)
Diastolic blood pressure, mmHg	76.0 (15.8)	81.5 (19.5)	80.2 (17.4)
Clinical hematology and biochemistry			
Hemoglobin, g/dL	13.5 (2.3)	13.6 (2.3)	13.1 (2.1)

eGFR, mL/min/1.73m^2	52.6 (22.1)	88.7 (40.2)	67.4 (29.2)
NT-proBNP, pg/mL	2355.0 [764.0, 7806.0]	48.8 [8.3, 314.2]	1333.0 [230.0, 5178.0]
Adjudicated diagnosis of heart failure	64 (44.1)	97 (20.4)	1043 (50.8)

Table 6.4 Baseline characteristics of study patients stratified by prior history of heart failure.

	Overall	No prior history of heart failure	Prior history of heart failure
Number of participants	10369	6208	3119
Men	5531 (53.3)	3271 (52.7)	1818 (58.3)
Age, years	69.3 (16.3)	64.7 (16.9)	73.7 (12.7)
< 50	1377 (13.3)	1191 (19.2)	168 (5.4)
50-75	4370 (42.1)	3012 (48.5)	1298 (41.6)
>75	4622 (44.6)	2005 (32.3)	1653 (53.0)
Ethnicity			
Black	845 (14.8)	522 (13.9)	303 (16.8)
Caucasian	4112 (72.1)	2634 (70.0)	1373 (76.2)
Other	743 (13.0)	609 (16.2)	127 (7.0)
Past medical history			
Ischemic heart disease	2953 (32.3)	1228 (20.4)	1687 (55.6)
Diabetes Mellitus	2398 (26.7)	1270 (21.5)	1096 (37.1)
Hypertension	5071 (59.3)	2950 (52.2)	2057 (73.4)
Hyperlipidemia	2269 (41.2)	1156 (32.1)	1092 (58.5)
Current or ex-smoker	2458 (41.3)	1645 (42.3)	767 (38.9)
Asthma	770 (18.5)	621 (22.1)	131 (10.3)
COPD	2117 (29.2)	1361 (28.3)	718 (30.4)
Atrial fibrillation	1701 (20.9)	822 (15.5)	861 (31.1)
Chronic Kidney Disease	1215 (18.9)	493 (11.4)	712 (34.2)
Body mass index, kg/m^2	27.7 (7.2)	27.5 (7.2)	28.0 (7.2)
<25	3062 (39.0)	2053 (40.6)	963 (36.1)
25-30	2473 (31.5)	1532 (30.3)	900 (33.8)
>=30	2317 (29.5)	1477 (29.2)	801 (30.1)
Physiological parameters			
Heart rate, beats per minute	91.7 (23.7)	93.2 (23.5)	88.8 (23.9)
Systolic blood pressure, mmHg	140.0 (27.9)	141.9 (27.1)	136.2 (29.2)
Diastolic blood pressure, mmHg	79.7 (17.0)	81.0 (16.7)	77.1 (17.2)
Clinical hematology and biochemistry			
Hemoglobin, g/dL	13.1 (2.1)	13.4 (2.1)	12.6 (2.1)
eGFR, mL/min/1.73m^2	68.2 (31.3)	77.0 (31.5)	55.1 (25.8)

NT-proBNP, pg/mL	1182.2 [191.0, 4737.0]	421.5 [83.8, 2270.2]	3484.0 [1162.5, 8905.5]
Adjudicated diagnosis of heart failure	4549 (43.9)	1802 (29.0)	2286 (73.3)

Table 6.5 Baseline characteristics of the training and external validation cohorts.

	Training cohort*	External validation cohort†
Number of participants	6315	3159
Men	3334 (52.8)	1829 (57.9)
Age, years	67.8 (16.2)	67.6 (16.3)
< 50	921 (14.6)	456 (14.4)
50-75	2927 (46.3)	1443 (45.7)
>75	2467 (39.1)	1260 (39.9)
Ethnicity		
Black	845 (18.4)	0 (0.0)
Caucasian	3651 (79.5)	461 (41.7)
Other	98 (2.1)	645 (58.3)
Past medical history		
Prior Heart failure	2232 (36.1)	887 (28.2)
Ischemic heart disease	2165 (35.2)	788 (26.5)
Diabetes Mellitus	1638 (28.2)	760 (24.1)
Hypertension	3092 (57.2)	1979 (62.9)
Hyperlipidemia	951 (39.9)	1318 (42.3)
Current or ex-smoker	2213 (45.7)	245 (22.2)
Asthma	488 (16.0)	282 (25.6)
COPD	1179 (28.8)	938 (29.8)
Atrial fibrillation	1087 (21.5)	614 (20.0)
Chronic Kidney Disease	506 (15.4)	709 (22.5)
Body mass index, kg/m^2	28.2 (7.5)	26.8 (6.5)
<25	1858 (36.4)	1204 (43.9)
25-30	1652 (32.3)	821 (29.9)
>=30	1599 (31.3)	718 (26.2)

Physiological parameters		
Heart rate, beats per minute	91.1 (23.7)	92.8 (23.6)
Systolic blood pressure, mmHg	140.9 (29.1)	139.0 (26.4)
Diastolic blood pressure, mmHg	79.4 (17.0)	80.1 (16.9)
Clinical hematology and biochemistry		
Hemoglobin, g/dL	13.0 (2.1)	13.2 (2.1)
eGFR, mL/min/1.73m^2	69.8 (32.7)	69.3 (28.8)
NT-proBNP, pg/mL	1072.7 [164.8, 4435.5]	979.0 [149.9, 4503.2]
Adjudicated diagnosis of heart failure	2774 (43.9)	1370 (43.4)

Presented as No. (%), mean (SD) or median [inter-quartile range]. Abbreviations: COPD= chronic obstructive pulmonary disease;
eGFR= estimated glomerular filtration rate; NT-proBNP= N-terminal pro-B-type natriuretic peptide; CVD= cardiovascular disease.
* Training cohort consists of all studies except Bombelli et al, Wussler et al and Ibrahim et al.
† External validation cohort consists of Wussler et al and Ibrahim et al.

6.4.2 Guideline-recommended and age-specific NT-proBNP thresholds

Pooled meta-estimates of NPV, sensitivity, PPV and specificity of NT-proBNP for the overall population at the guideline recommended rule-out threshold of 300 pg/mL were 94.6% (95% confidence interval, 91.9-96.4%), 96.8% (94.6-98.1%), 62.9% (51.3-73.3%), and 49.3% (35.4-63.4%) respectively (Figure 6.3 and Table 6.6). Overall, 30.4% of patients had NT-proBNP concentrations below 300 pg/mL. However, there was marked heterogeneity across patient subgroups and studies (Figure 6.4, Figure 6.5, Figure 6.6 and Figure 6.6). NPV was lower in patients ≥75 years (88.2% [83.5-91.8%]), and in those with prior heart failure (79.4% [68.4-87.3%]) and obesity (90.4% [84.5-94.2%]).

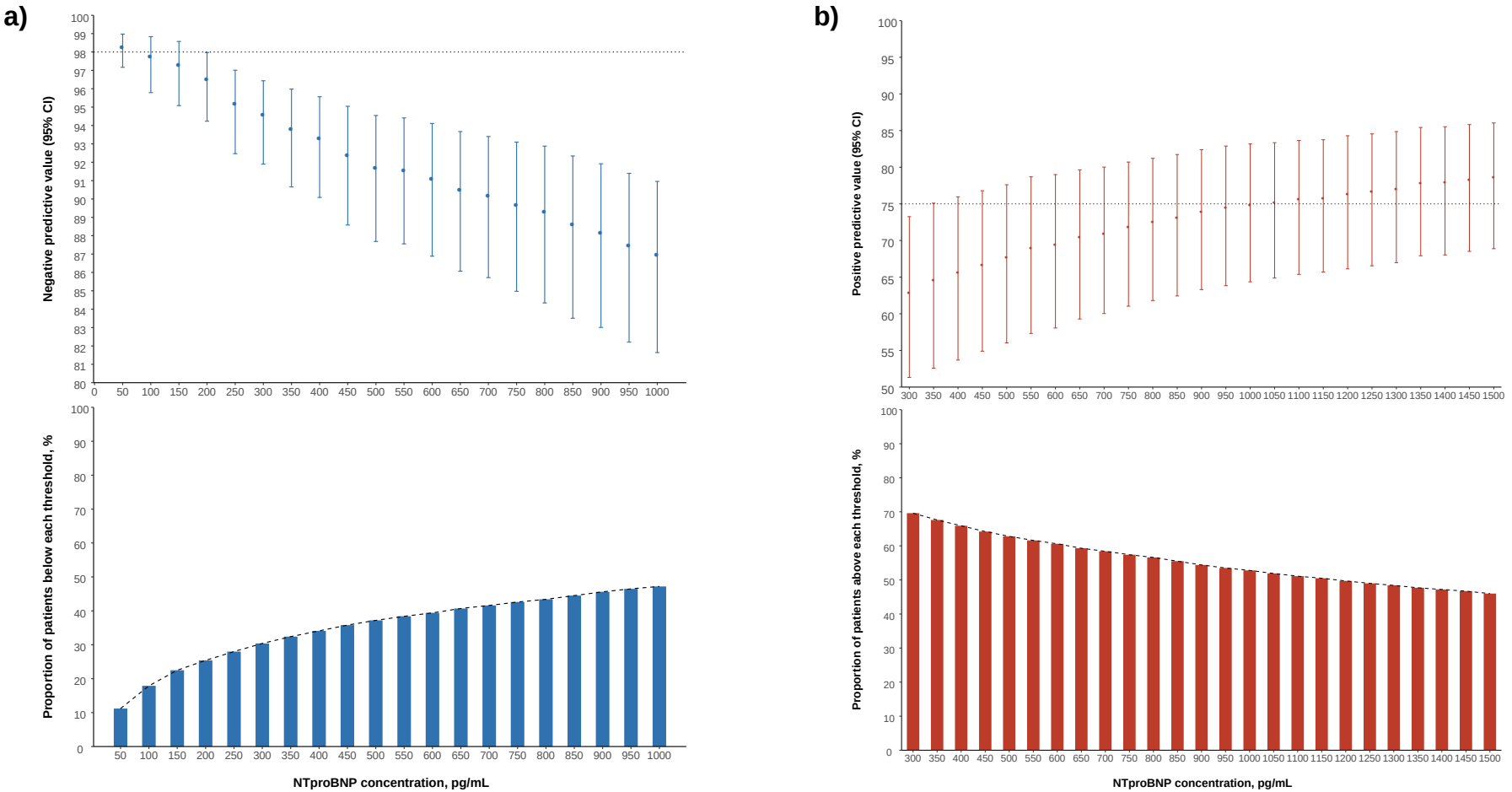


Figure 6.3 NT-proBNP thresholds for acute heart failure.
(a) (top) Negative predictive values of NT-proBNP concentrations to rule-out a diagnosis of acute heart failure. (bottom) Cumulative proportion of patients presenting with suspected acute heart failure with NT-proBNP concentrations below each threshold.
(b) (top) Positive predictive values of NT-proBNP concentrations to rule-in a diagnosis of acute heart failure. (bottom) Cumulative proportion of patients presenting with suspected acute heart failure with NT-proBNP concentrations above each threshold.

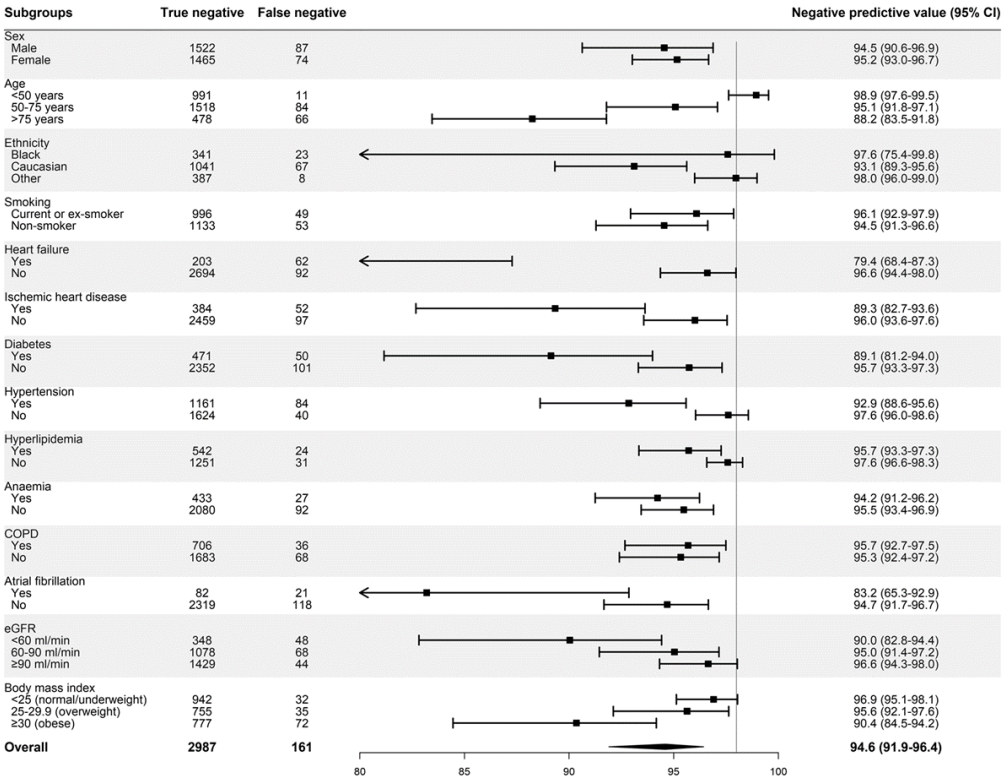
Table 6.6 Diagnostic performance of NT-proBNP for acute heart failure.

NT-proBNP threshold (pg/mL)	True positive	False positive	True negative	False negative	Negative predictive value (NPV)	Positive predictive value (PPV)	Sensitivity	Specificity
50	4529	4676	1144	20	98.3 (97.2-99.0)	49.8 (41.9-57.7)	99.7 (99.2-99.9)	9.1 (3.6-21.3)
100	4507	4011	1809	42	97.8 (95.8-98.8)	53.8 (44.8-62.6)	99.3 (98.5-99.7)	22.8 (13.2-36.4)
150	4481	3557	2263	68	97.3 (95.1-98.6)	56.6 (47.3-65.5)	98.9 (97.8-99.5)	32.3 (21.3-45.7)
200	4451	3280	2540	98	96.6 (94.2-98.0)	58.5 (49.0-67.3)	98.4 (96.9-99.2)	38.1 (26.5-51.3)
250	4417	3045	2775	132	95.2 (92.5-97.0)	60.6 (50.3-70.1)	97.4 (95.5-98.5)	43.9 (31.1-57.6)
300	4388	2833	2987	161	94.6 (91.9-96.4)	62.9 (51.3-73.3)	96.8 (94.6-98.1)	49.3 (35.4-63.4)
350	4359	2648	3172	190	93.8 (90.7-96.0)	64.6 (52.6-75.1)	96.0 (93.4-97.6)	53.8 (39.8-67.3)
400	4331	2500	3320	218	93.3 (90.1-95.6)	65.7 (53.7-76.0)	95.3 (92.4-97.2)	56.5 (42.6-69.4)
450	4292	2364	3456	257	92.4 (88.6-95.0)	66.7 (54.9-76.8)	94.3 (90.7-96.6)	59.6 (46.2-71.7)
500	4262	2247	3573	287	91.7 (87.7-94.5)	67.8 (56.0-77.6)	93.5 (89.8-96.0)	61.8 (48.4-73.6)
550	4243	2143	3677	306	91.6 (87.6-94.4)	69.0 (57.3-78.7)	93.1 (89.1-95.7)	64.3 (51.1-75.6)
600	4219	2066	3754	330	91.1 (86.9-94.1)	69.5 (58.1-79.0)	92.5 (88.2-95.3)	65.7 (53.1-76.5)

650	4177	1967	3853	372	90.5 (86.1-93.7)	70.5 (59.3-79.6)	91.6 (87.1-94.7)	67.7 (55.4-78.0)
700	4155	1897	3923	394	90.2 (85.7-93.4)	71.0 (60.0-80.0)	91.1 (86.5-94.3)	68.9 (57.1-78.7)
750	4126	1829	3991	423	89.7 (85.0-93.1)	71.9 (61.0-80.7)	90.4 (85.7-93.7)	70.5 (58.7-80.1)
800	4096	1769	4051	453	89.3 (84.3-92.9)	72.6 (61.8-81.2)	89.8 (84.7-93.3)	71.8 (60.3-81.0)
850	4061	1695	4125	488	88.7 (83.5-92.3)	73.2 (62.4-81.7)	88.9 (83.7-92.6)	72.9 (61.7-81.8)
900	4025	1613	4207	524	88.2 (83.0-91.9)	74.0 (63.3-82.4)	88.1 (82.7-92.0)	74.3 (63.4-82.9)
950	3993	1553	4267	556	87.5 (82.2-91.4)	74.5 (63.8-82.9)	87.1 (81.4-91.3)	75.4 (64.5-83.7)
1000	3959	1511	4309	590	87.0 (81.6-91.0)	74.9 (64.4-83.2)	86.3 (80.6-90.6)	76.1 (65.6-84.2)
1050	3921	1459	4361	628	86.3 (80.8-90.5)	75.3 (64.9-83.4)	85.4 (79.5-89.8)	76.9 (66.8-84.6)
1100	3883	1412	4408	666	85.9 (80.3-90.1)	75.7 (65.4-83.6)	84.7 (78.7-89.2)	77.6 (67.7-85.2)
1150	3855	1382	4438	694	85.5 (79.7-89.8)	75.9 (65.7-83.8)	83.9 (77.7-88.7)	78.1 (68.5-85.5)
1200	3814	1336	4484	735	84.9 (79.1-89.3)	76.4 (66.1-84.3)	83.0 (76.5-88.0)	79.1 (69.5-86.2)
1250	3780	1300	4520	769	84.4 (78.5-88.9)	76.7 (66.6-84.5)	82.2 (75.6-87.2)	79.7 (70.4-86.6)
1300	3750	1273	4547	799	84.1 (78.1-88.6)	77.1 (67.0-84.9)	81.5 (74.8-86.7)	80.3 (71.2-87.1)
1350	3720	1231	4589	829	83.9 (77.9-88.5)	77.9 (67.9-85.4)	81.0 (74.2-86.3)	81.3 (72.4-87.9)
1400	3685	1211	4609	864	83.4 (77.3-88.1)	78.0 (68.0-85.5)	80.1 (73.1-85.7)	81.7 (72.9-88.1)
1450	3662	1178	4642	887	83.2 (77.0-88.0)	78.4 (68.5-85.8)	79.6 (72.5-85.3)	82.3 (73.7-88.5)

1500	3625	1148	4672	924	82.6 (76.3-87.6)	78.7 (68.9-86.0)	78.7 (71.6-84.4)	82.8 (74.4-88.8)
1550	3608	1125	4695	941	82.5 (76.1-87.4)	79.0 (69.3-86.2)	78.4 (71.1-84.2)	83.2 (74.9-89.1)
1600	3574	1104	4716	975	82.0 (75.7-86.9)	79.2 (69.5-86.4)	77.5 (70.4-83.3)	83.5 (75.4-89.3)
1650	3547	1085	4735	1002	81.6 (75.2-86.6)	79.2 (69.6-86.4)	76.8 (69.5-82.8)	83.8 (75.9-89.4)
1700	3516	1068	4752	1033	81.3 (74.8-86.4)	79.4 (69.8-86.5)	76.2 (69.0-82.2)	84.1 (76.3-89.6)
1750	3484	1039	4781	1065	80.8 (74.4-85.9)	79.7 (70.2-86.7)	75.3 (68.2-81.3)	84.5 (77.0-89.9)
1800	3458	1020	4800	1091	80.5 (74.1-85.6)	79.8 (70.4-86.8)	74.7 (67.6-80.7)	84.8 (77.5-90.0)
1850	3430	995	4825	1119	80.2 (73.8-85.3)	80.4 (70.9-87.4)	74.1 (66.9-80.2)	85.5 (78.1-90.6)
1900	3400	980	4840	1149	79.7 (73.3-84.8)	80.5 (70.9-87.6)	73.2 (66.0-79.3)	85.8 (78.5-90.9)
1950	3360	962	4858	1189	79.2 (72.9-84.4)	80.9 (71.3-87.8)	72.3 (65.2-78.4)	86.2 (79.1-91.2)
2000	3329	942	4878	1220	78.8 (72.4-84.1)	81.3 (71.7-88.1)	71.4 (64.4-77.6)	86.7 (79.8-91.5)

a)


Subgroups
True negative
False negative
Negative predictive value (95% CI)
Sex
Male 1522 87 94.5 (90.6-96.9)
Female 1465 74 95.2 (93.0-96.7)
Age
<50 years 991 11 98.9 (97.6-99.5)
50-75 years 1518 84 95.1 (91.8-97.1)
>75 years 478 66 88.2 (83.5-91.8)
Ethnicity
Black 341 23 97.6 (75.4-99.8)
Caucasian 1041 67 93.1 (89.3-95.6)
Other 387 8 98.0 (96.0-99.0)
Smoking
Current or ex-smoker 996 49 96.1 (92.9-97.9)
Non-smoker 1133 53 94.5 (91.3-96.6)
Heart failure
Yes 203 62 79.4 (68.4-87.3)
No 2694 92 96.6 (94.4-98.0)
Ischemic heart disease
Yes 384 52 89.3 (82.7-93.6)
No 2459 97 96.0 (93.6-97.6)
Diabetes
Yes 471 50 89.1 (81.2-94.0)
No 2352 101 95.7 (93.3-97.3)
Hypertension
Yes 1161 84 92.9 (88.6-95.6)
No 1624 40 97.6 (96.0-98.6)
Hyperlipidemia
Yes 542 24 95.7 (93.3-97.3)
No 1251 31 97.6 (96.6-98.3)
Anaemia
Yes 433 27 94.2 (91.2-96.2)
No 2080 92 95.5 (93.4-96.9)
COPD
Yes 706 36 95.7 (92.7-97.5)
No 1683 68 95.3 (92.4-97.2)
Atrial fibrillation
Yes 82 21 83.2 (65.3-92.9)
No 2319 118 94.7 (91.7-96.7)
eGFR
<60 ml/min 348 48 90.0 (82.8-94.4)
60-90 ml/min 1078 68 95.0 (91.4-97.2)
≥90 ml/min 1429 44 96.6 (94.3-98.0)
Body mass index
<25 (normal/underweight) 942 32 96.9 (95.1-98.1)
25-29.9 (overweight) 755 35 95.6 (92.1-97.6)
≥30 (obese) 777 72 90.4 (84.5-94.2)
Overall 2987 161 94.6 (91.9-96.4)
80
85
90
95
100

b)

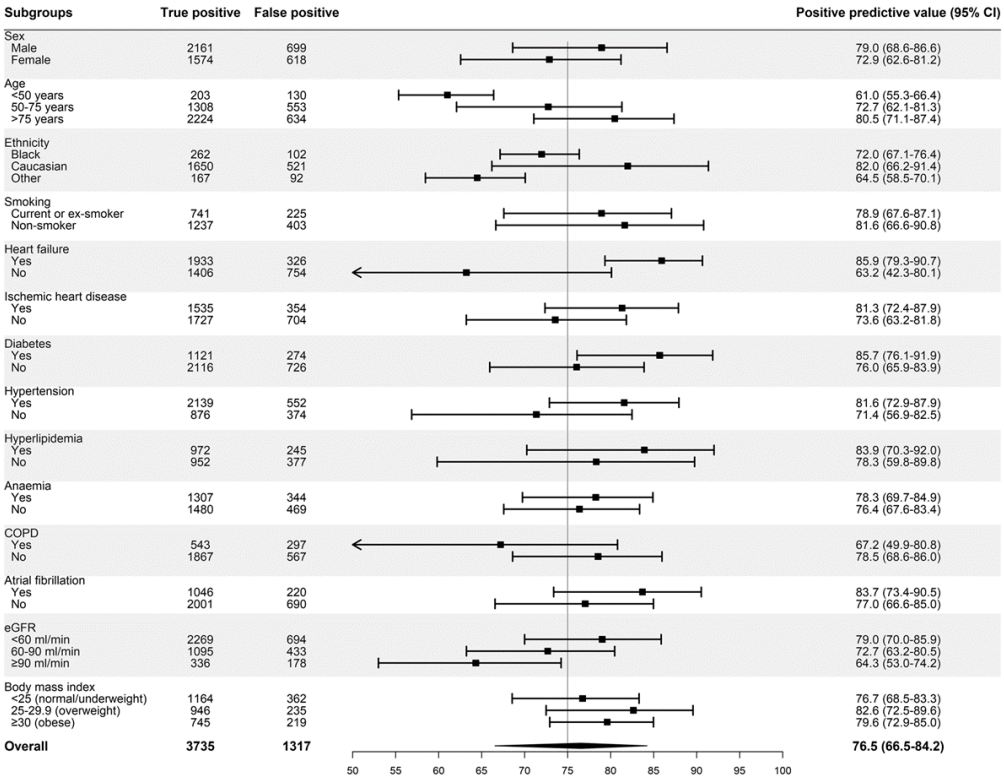


Figure 6.4 Diagnostic performance of guideline-recommended NT-proBNP thresholds across patient subgroups.
(a) Negative predictive value of the NT-proBNP threshold of 300 pg/mL across patient subgroups.
(b) Positive predictive value of age-specific NT-proBNP thresholds across patient subgroups (450, 900, and 1,800 pg/mL for those <50 years, 50-75 years, and >75 years, respectively).
Abbreviations: COPD= chronic obstructive pulmonary disease; eGFR= estimated glomerular filtration rate.

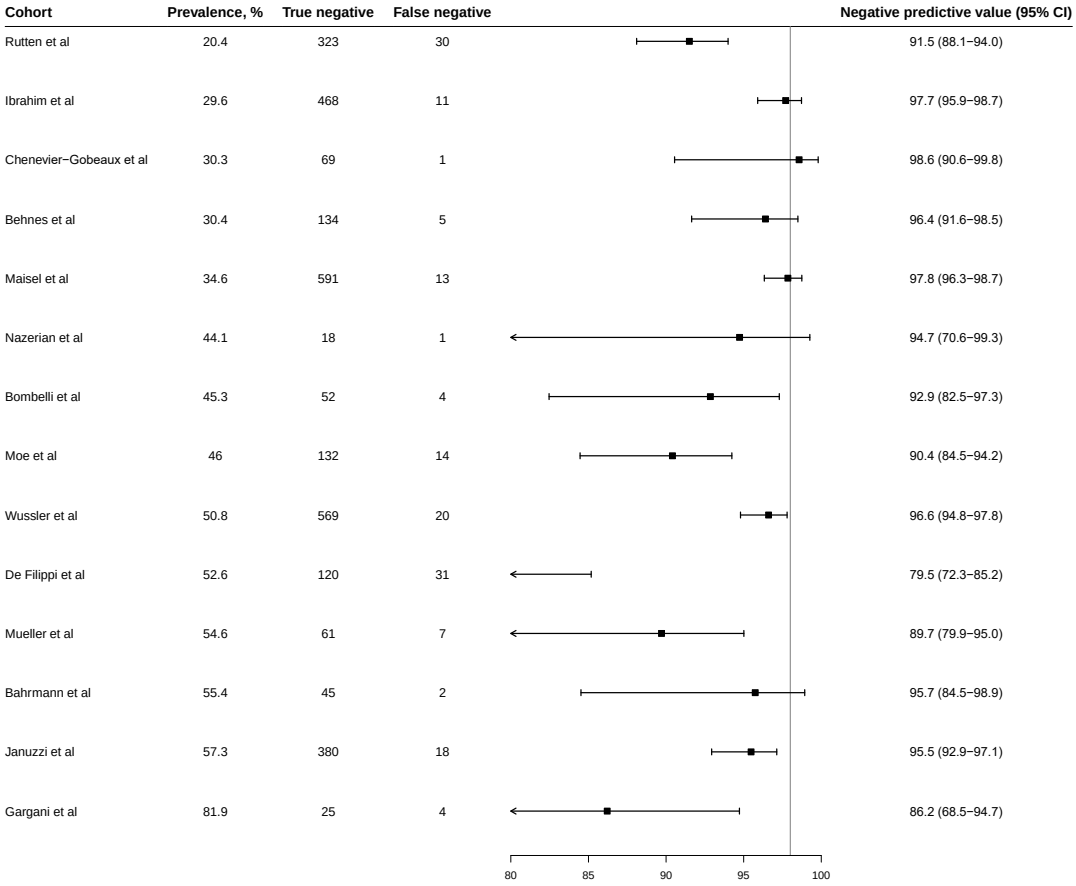


Figure 6.5 Negative predictive value of NT-proBNP at the 300 pg/mL threshold across cohorts.

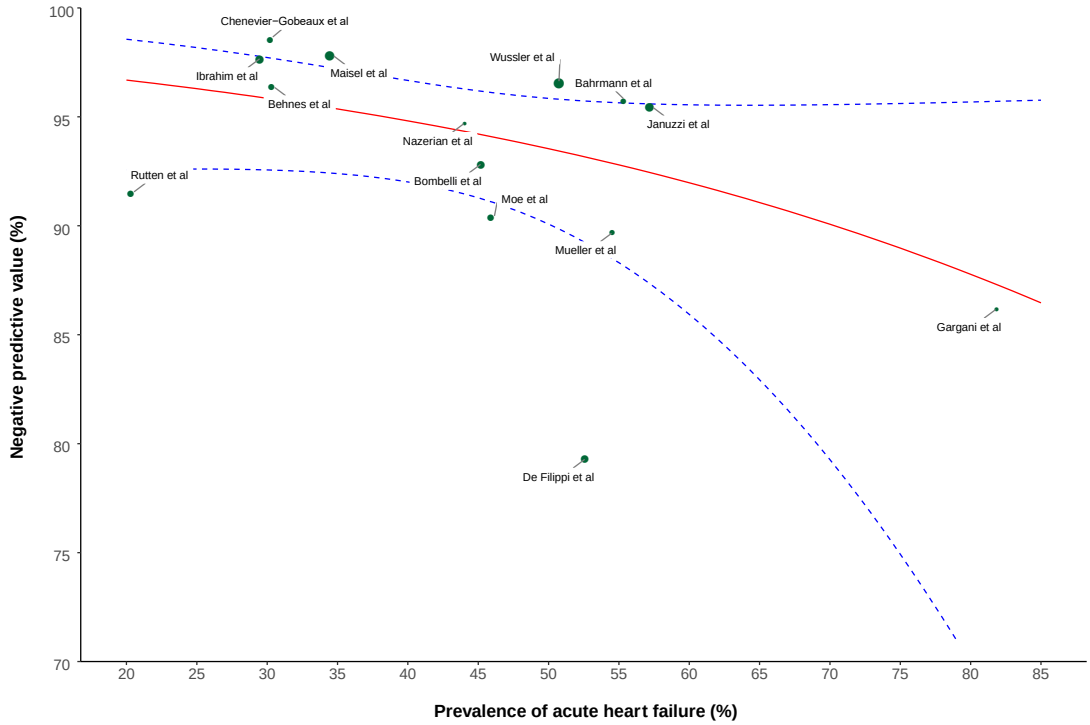


Figure 6.6 Meta-regression of the negative predictive value of NT-proBNP at the threshold of 300 pg/mL by prevalence of acute heart failure.

Pooled meta-estimates of the PPV for age-specific NT-proBNP rule-in thresholds of 450, 900 and 1800 pg/mL were 61.0% (55.3-66.4%), 73.5% (62.3-82.3%) and 80.2% (70.9-87.1%), respectively (Table 6.7). Corresponding specificities were 87.8% (79.5-93.0%), 81.1% (72.6-87.5%) and 73.1% (65.2-79.8%). Overall, 48.9% of patients with suspected acute heart failure had NT-proBNP above these age-specific thresholds. PPVs of the age-specific rule-in thresholds were higher than the uniform 300 pg/mL threshold in subgroups although there was heterogeneity by age groups, renal function and prevalence of acute heart failure (Figure 6.8, Figure 6.9, Figure 6.10, Figure 6.11 and Figure 6.11).

Table 6.7 Diagnostic performance of age-specific thresholds of NT-proBNP for acute heart failure.

Age groups	NT-proBNP threshold (pg/mL)	True positive	False positive	True negative	False negative	Negative predictive value (NPV)	Positive predictive value (PPV)	Sensitivity	Specificity
<50 years	450	203	130	1025	19	98.4 (96.2-99.3)	61.0 (55.3-66.4)	91.4 (87.0-94.5)	87.8 (79.5-93.0)
50-75 years	900	1407	575	2121	267	88.3 (82.9-92.2)	73.5 (62.3-82.3)	83.2 (76.0-88.6)	81.1 (72.6-87.5)
>75 years	1800	2135	621	1348	518	72.2 (63.4-79.7)	80.2 (70.9-87.1)	79.3 (74.2-83.5)	73.1 (65.2-79.8)
All	300	4388	2833	2987	161	94.6 (91.9-96.4)	62.9 (51.3-73.3)	96.8 (94.6-98.1)	49.3 (35.4-63.4)

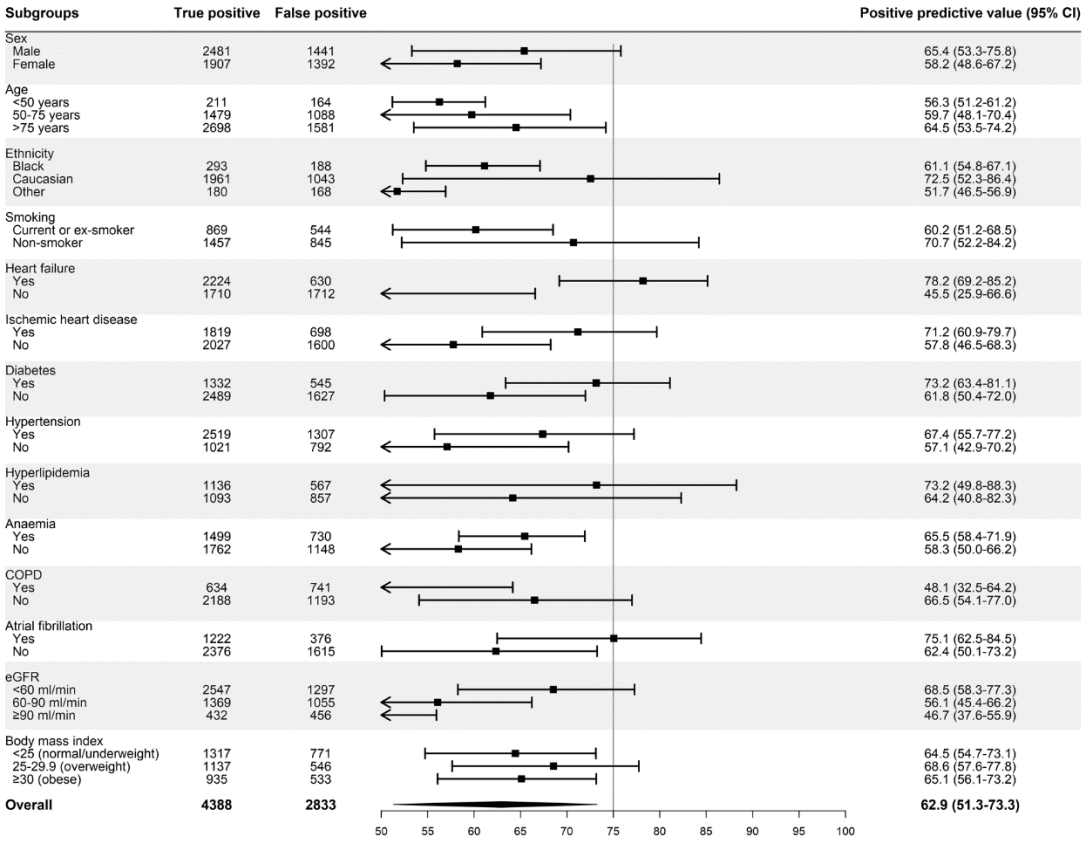


Figure 6.7 Positive predictive value of the 300 pg/mL NT-proBNP threshold across patient subgroups.

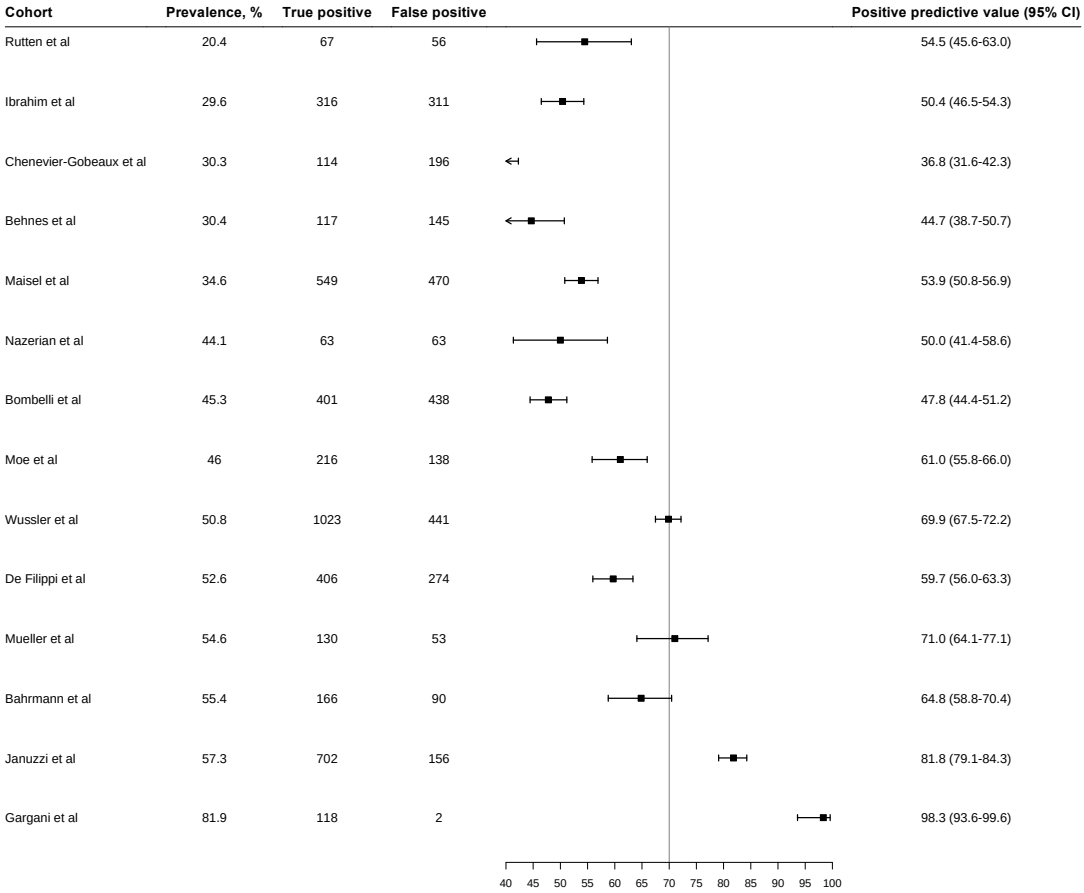


Figure 6.8 Positive predictive value of the NT-proBNP threshold of 300 pg/mL across cohorts.

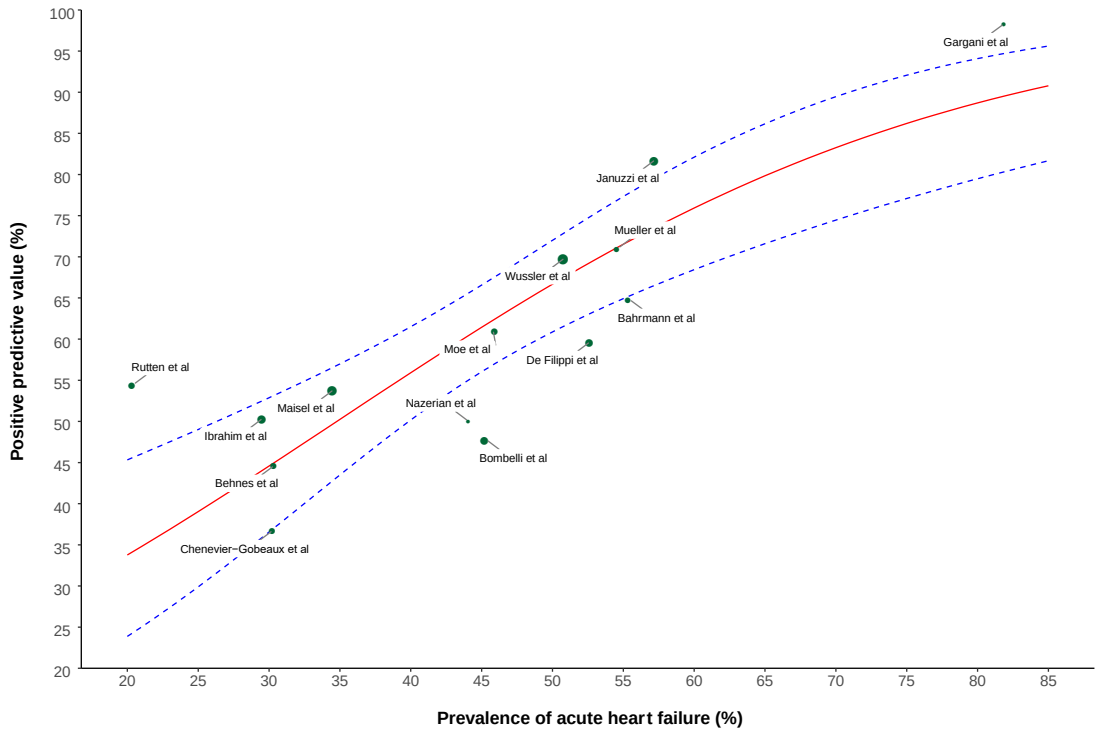


Figure 6.9 Meta-regression of positive predictive value of the 300 pg/mL NT-proBNP threshold by prevalence of acute heart failure.

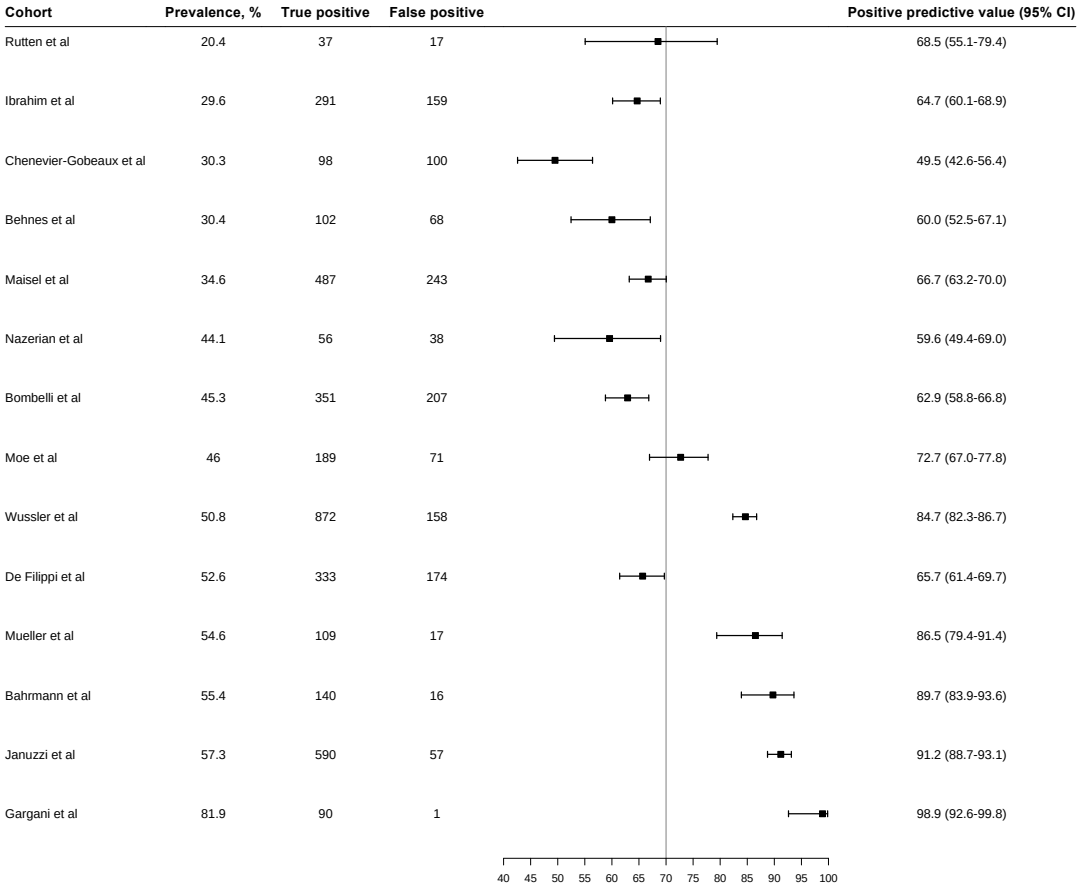


Figure 6.10 Positive predictive value of age-specific thresholds of NT-proBNP across cohorts.

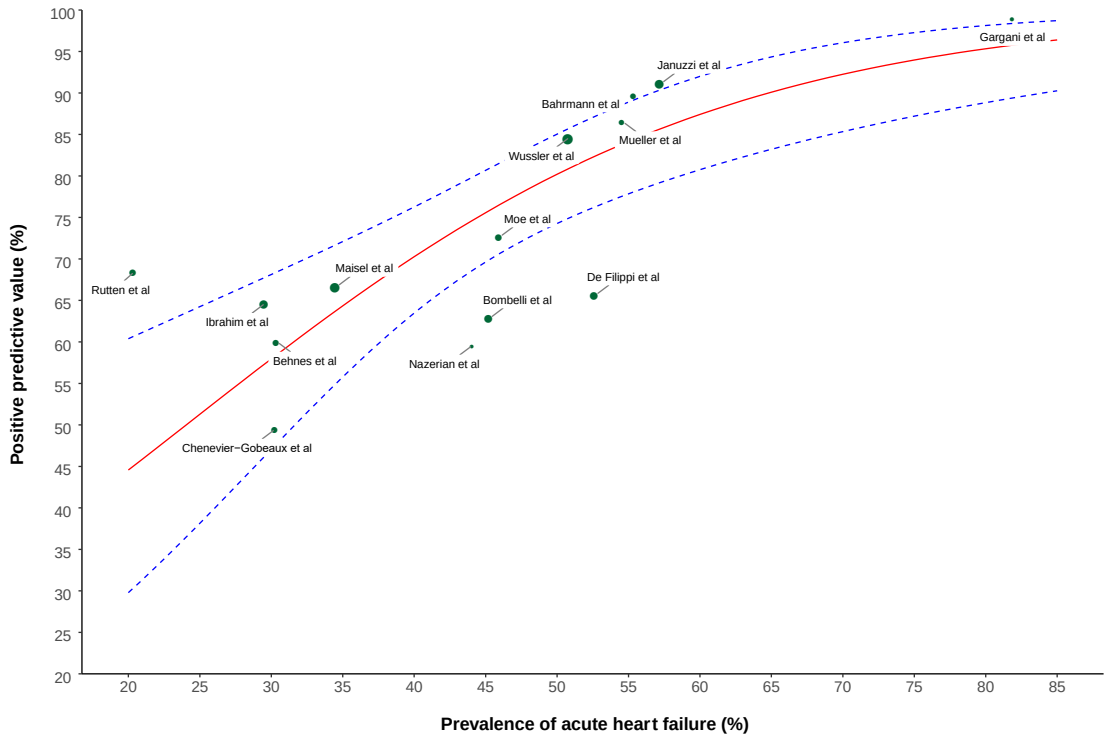


Figure 6.11 Meta-regression of positive predictive value of age-specific thresholds of NT-proBNP by prevalence of acute heart failure.

Overall, seven studies were identified as having a high risk of bias (Table *6.2*). In sensitivity analyses restricted to studies where the adjudication of acute heart failure was blinded to NT-proBNP concentrations and those with a low-risk of bias, diagnostic performance of the guideline-recommended and age-specific NT-proBNP thresholds remained unchanged (Table 6.8 and Table 6.9).

Table 6.8 Sensitivity analysis of the diagnostic performance of guideline-recommended and age-specific thresholds of NT-proBNP for acute heart failure in studies where the reference standard was blinded to NT-proBNP concentration.

Age groups	NT-proBNP threshold (pg/mL)	True positive	False positive	True negative	False negative	Negative predictive value (NPV)	Positive predictive value (PPV)	Sensitivity	Specificity
<50 years	450	170	107	719	12	98.7 (95.2-99.7)	61.7 (54.1-68.7)	93.4 (88.8-96.2)	85.4 (74.2-92.3)
50-75 years	900	1095	487	1514	193	88.4 (81.7-92.8)	73.6 (59.7-84.0)	85.2 (81.8-88.0)	77.7 (69.3-84.3)
>75 years	1800	1571	557	1046	368	72.7 (62.0-81.3)	79.2 (68.3-87.1)	80.9 (77.2-84.1)	69.8 (62.1-76.5)
All	300	3298	2336	2095	111	94.6 (91.4-96.7)	63.2 (49.4-75.1)	97.1 (95.8-98.0)	44.8 (31.1-59.4)

Table 6.9 Sensitivity analysis of the diagnostic performance of guideline-recommended and age-specific thresholds of NT-proBNP for acute heart failure in studies with low risk of bias in patient selection.

Age groups	NT-proBNP threshold (pg/mL)	True positive	False positive	True negative	False negative	Negative predictive value (NPV)	Positive predictive value (PPV)	Sensitivity	Specificity
<50 years	450	127	75	807	10	98.8 (97.7-99.3)	63.1 (54.3-71.1)	94.2 (83.3-98.2)	91.6 (89.1-93.6)
50-75 years	900	834	379	1503	156	90.5 (86.8-93.3)	68.8 (57.6-78.2)	82.6 (71.1-90.2)	81.7 (73.3-87.9)
>75 years	1800	942	278	641	241	75.5 (66.4-82.8)	78.0 (64.7-87.3)	78.2 (69.9-84.7)	75.0 (64.6-83.2)
All	300	2211	1525	2158	99	95.6 (93.0-97.2)	57.6 (47.1-67.5)	95.7 (91.5-97.9)	57.0 (44.1-69.1)

6.4.3 Optimised NT-proBNP thresholds

A NT-proBNP threshold of 100 pg/mL achieved our optimal rule-out criteria with a pooled NPV of 97.8% (95.8-98.8%) and sensitivity of 99.3% (98.5-99.7%) (Table 6.6). However, only 17.9% of patients had NT-proBNP concentrations below 100 pg/mL and the NPV remained lower in older patients and those with prior history of heart failure, ischaemic heart disease and impaired renal function (Table 6.6 and Figure 6.12). Similarly, a NT-proBNP threshold of 1000 pg/mL achieved our optimal rule-in criteria with a PPV of 74.9% (64.4-83.2%) and specificity of 76.1% (65.6-84.2%), although performance was also lower within patient subgroups, particularly those without prior heart failure (Table 6.6 and Figure 6.13).

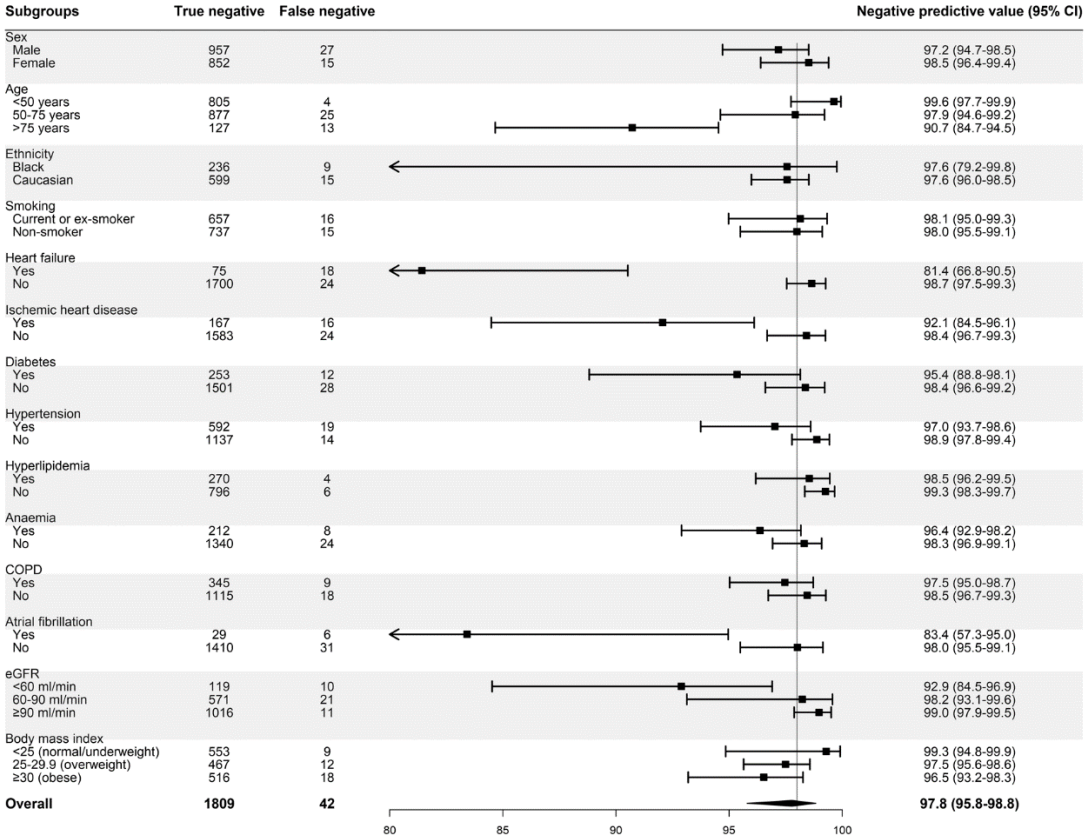


Figure 6.12 Negative predictive value of the NT-proBNP threshold of 100 pg/mL across patient subgroups.

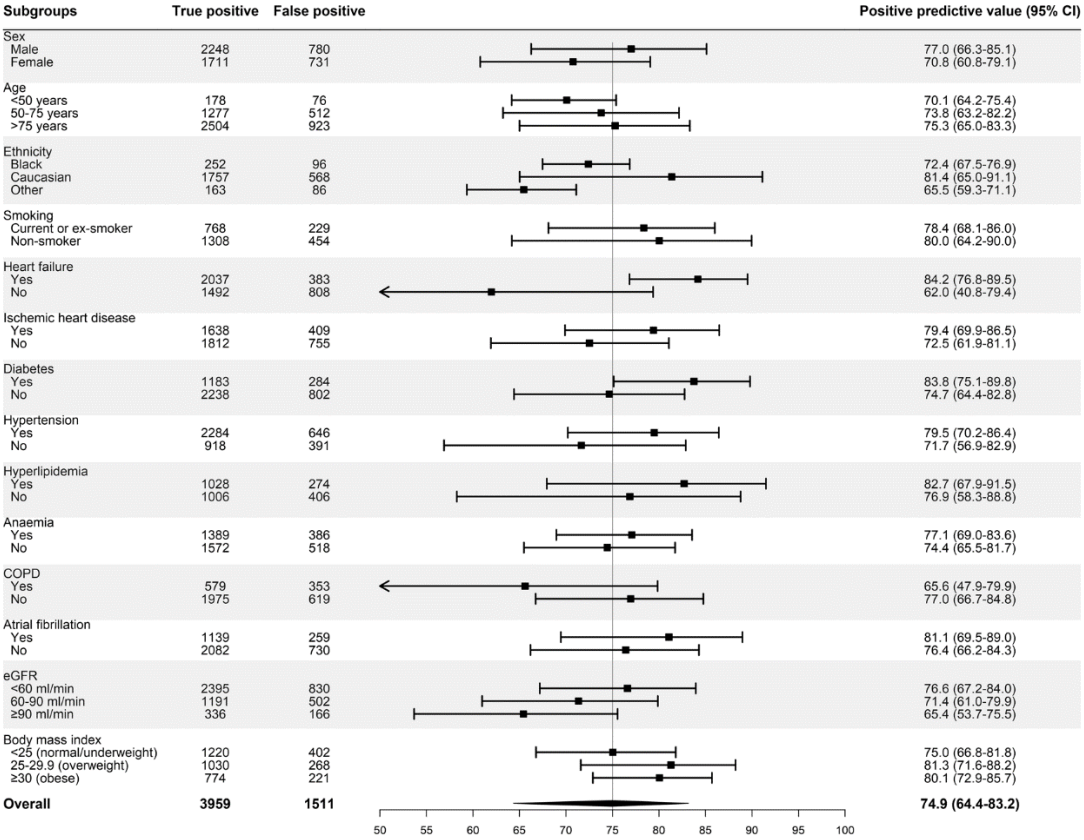


Figure 6.13 Positive predictive value of the NT-proBNP threshold of 1000 pg/mL across patient subgroups.

6.4.4 The CoDE-HF score

The extreme gradient boosting (XGBoost) and generalised linear mixed-model (GLMM) were the best performing models (area under curve in the overall training cohort of 0.925 [0.919 to 0.932] versus 0.931 [0.925 to 0.937] respectively) (Table 6.10, Figure 6.14, Figure 6.15, Figure 6.16 and Figure 6.17).

Table 6.10 Diagnostic performance of all models.

A) Rule-out diagnostic performance.

	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out	AUC	Brier score
Patients without prior heart failure					
XGBoost (NT-proBNP only)	98.5 (97.5-99.1)	99.0 (97.8-99.5)	22.4 %	0.882 (0.873-0.891)	0.123
GLMM	98.6 (97.4-99.2)	98.1 (96.0-99.1)	41.1 %	0.931 (0.925-0.937)	0.094
NB	98.6 (97.2-99.3)	98.5 (96.9-99.3)	33.8 %	0.892 (0.884-0.900)	0.129
RF	98.6 (97.0-99.3)	98.2 (96.6-99.0)	38.4 %	0.914 (0.906-0.921)	0.108
XGBoost (all variables)	98.6 (97.8-99.1)	98.1 (96.9-98.9)	40.3 %	0.925 (0.919-0.932)	0.099

B) Rule-in diagnostic performance.

	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in	AUC	Brier score
Patients without prior heart failure					
XGBoost (NT-proBNP only)	75.8 (53.9-89.3)	94.2 (88.9-97.0)	21.5 %	0.882 (0.873-0.891)	0.123
GLMM	75.1 (67.2-81.7)	91.0 (86.2-94.2)	30.1 %	0.931 (0.925-0.937)	0.094
NB	75.5 (54.6-88.7)	94.4 (90.5-96.8)	19.0 %	0.892 (0.884-0.900)	0.129
RF	75.2 (53.3-88.9)	92.4 (87.0-95.7)	26.7 %	0.914 (0.906-0.921)	0.108
XGBoost (all variables)	75.0 (65.7-82.5)	92.2 (87.5-95.2)	28.0 %	0.925 (0.919-0.932)	0.099
Patients with prior heart failure					
XGBoost (NT-proBNP only)	89.9 (82.7-94.3)	90.2 (86.0-93.3)	27.8 %	0.779 (0.759-0.798)	0.150
GLMM	93.2 (89.9-95.4)	90.1 (82.4-94.6)	47.5 %	0.860 (0.845-0.875)	0.123
NB	91.9 (86.7-95.2)	90.3 (87.2-92.7)	33.6 %	0.812 (0.794-0.829)	0.144
RF	93.5 (88.9-96.2)	90.4 (86.2-93.4)	41.3 %	0.830 (0.813-0.847)	0.137
XGBoost (all variables)	92.7 (89.1-95.2)	90.2 (84.0-94.1)	45.5 %	0.846 (0.830-0.862)	0.130

Abbreviations: GLMM=generalised linear mixed-model; NB=naïve bayes; RF=random forest; XGBoost=extreme gradient boosting.

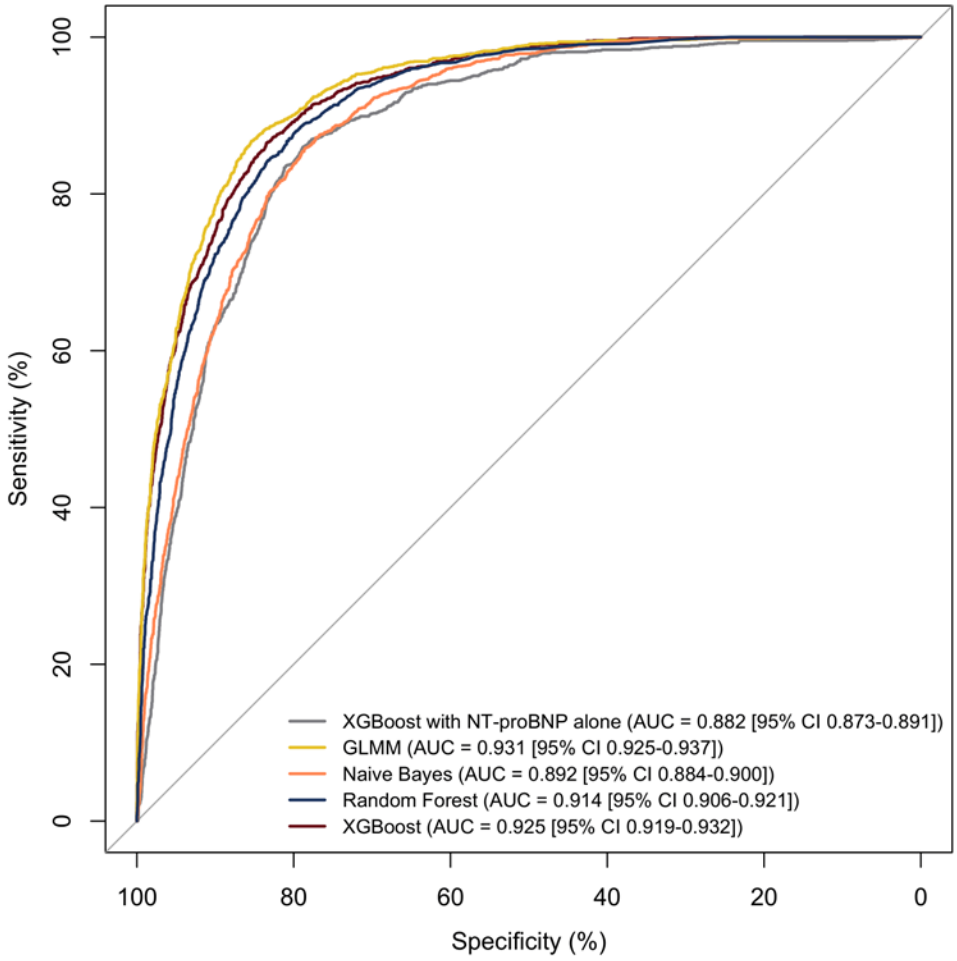


Figure 6.14 Receiver operator curve for all statistical models in the patients without prior heart failure.

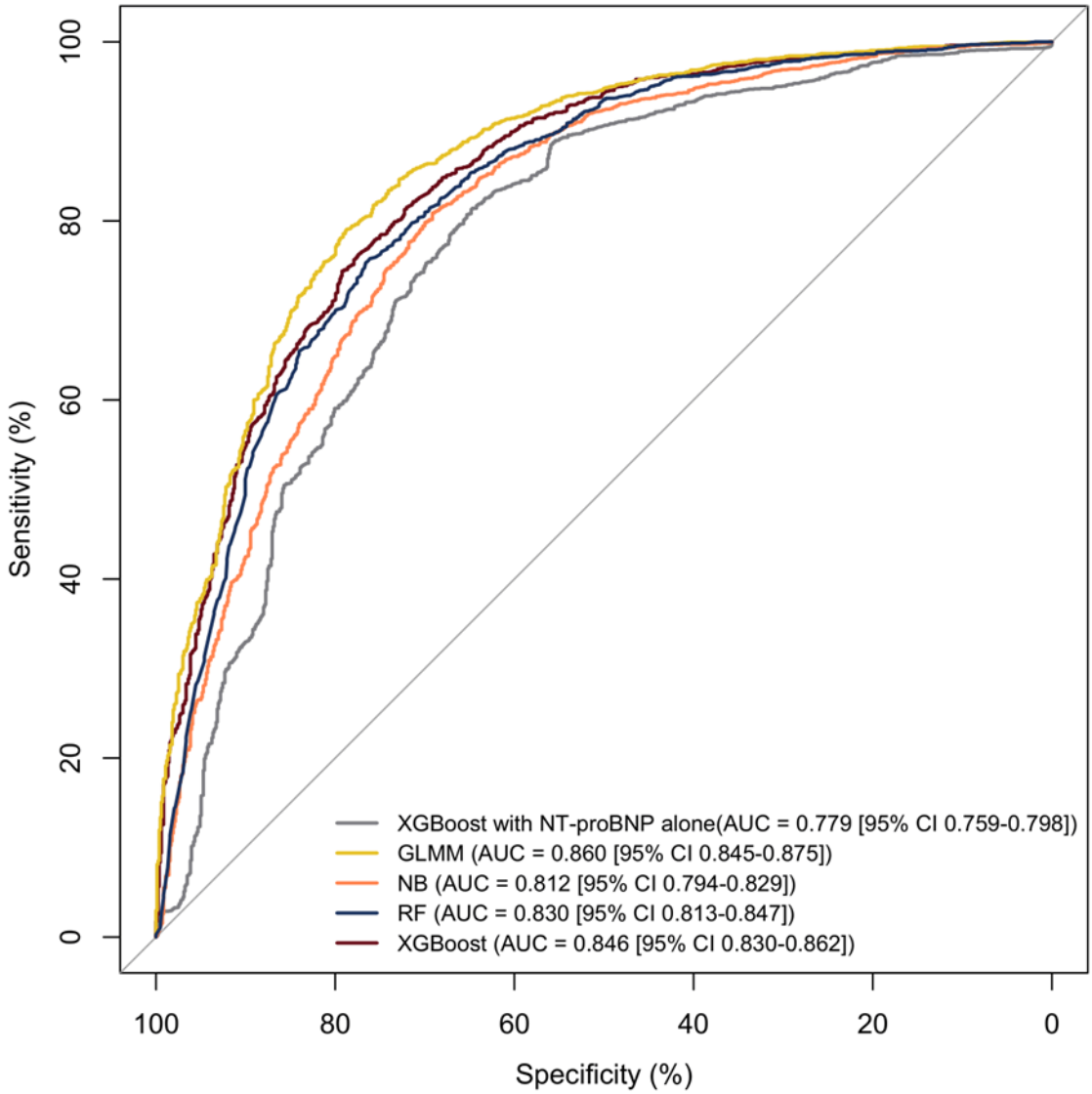


Figure 6.15 Receiver operator curve for all statistical models in the patients with prior heart failure.

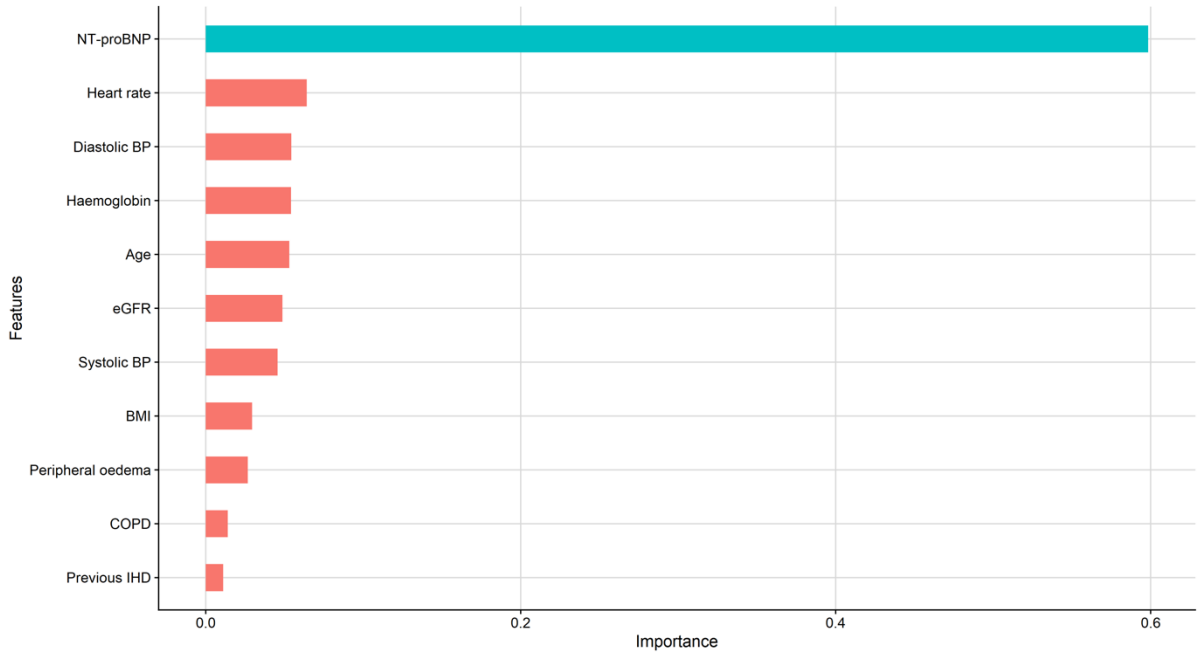


Figure 6.16 Relative feature importance plot for the model developed for patients without prior heart failure.

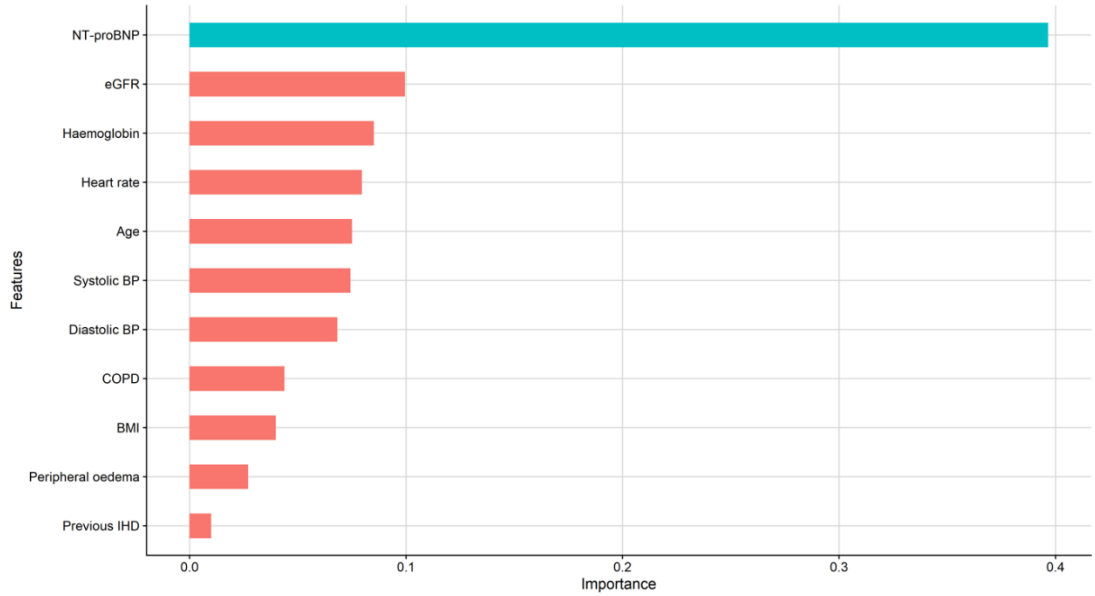


Figure 6.17 Relative feature importance plot for the model developed for patients with prior heart failure.

Whilst the performance of XGBoost was similar to GLMM, a key advantage of XGBoost is its ability to compute a score despite missing values. This is an important functionality we wish to build into the CoDE-HF decision-support tool to facilitate its implementation in clinical practice therefore the XGBoost model was selected as the final model for CoDE-HF.

CoDE-HF was well calibrated with excellent discrimination in those with and without prior heart failure (area under the receiver operator curve [AUC] 0.846 [0.830 to 0.852] and 0.925 [0.919 to 0.932], and Brier scores of 0.130 and 0.099, respectively) (Figure 6.18 and Figure 6.19). A CoDE-HF score of 4.7 achieved an NPV of 98.6% (97.8 to 99.1%) and sensitivity of 98.1% (96.9 to 98.9%) (Table *6.11*), whilst a score of 51.2 achieved a PPV of 75.0% (65.7 to 82.5%) and a specificity of 92.2% (87.5 to 95.2%) in those without prior heart failure. These rule-in and rule-out scores had similar diagnostic performance across all subgroups (Figure 6.20). If these scores were applied in patients with suspected acute heart failure, CoDE-HF would identify 40.3% (2,502/6,208) at low-probability (<4.7) and 28.0% (1,737/6,208) at high-probability (≥51.2) of acute heart failure. In patients with prior heart failure, there was no score which achieved our target rule-out criteria in the training cohort. A CoDE-HF score of 84.5 achieved a PPV of 92.7% (89.1 to 95.2%) and specificity of 90.2% (84.0 to 94.1%). This score would identify 45.5% (1,420/3.119) of patients as high-probability for acute heart failure (Figure 6.21).

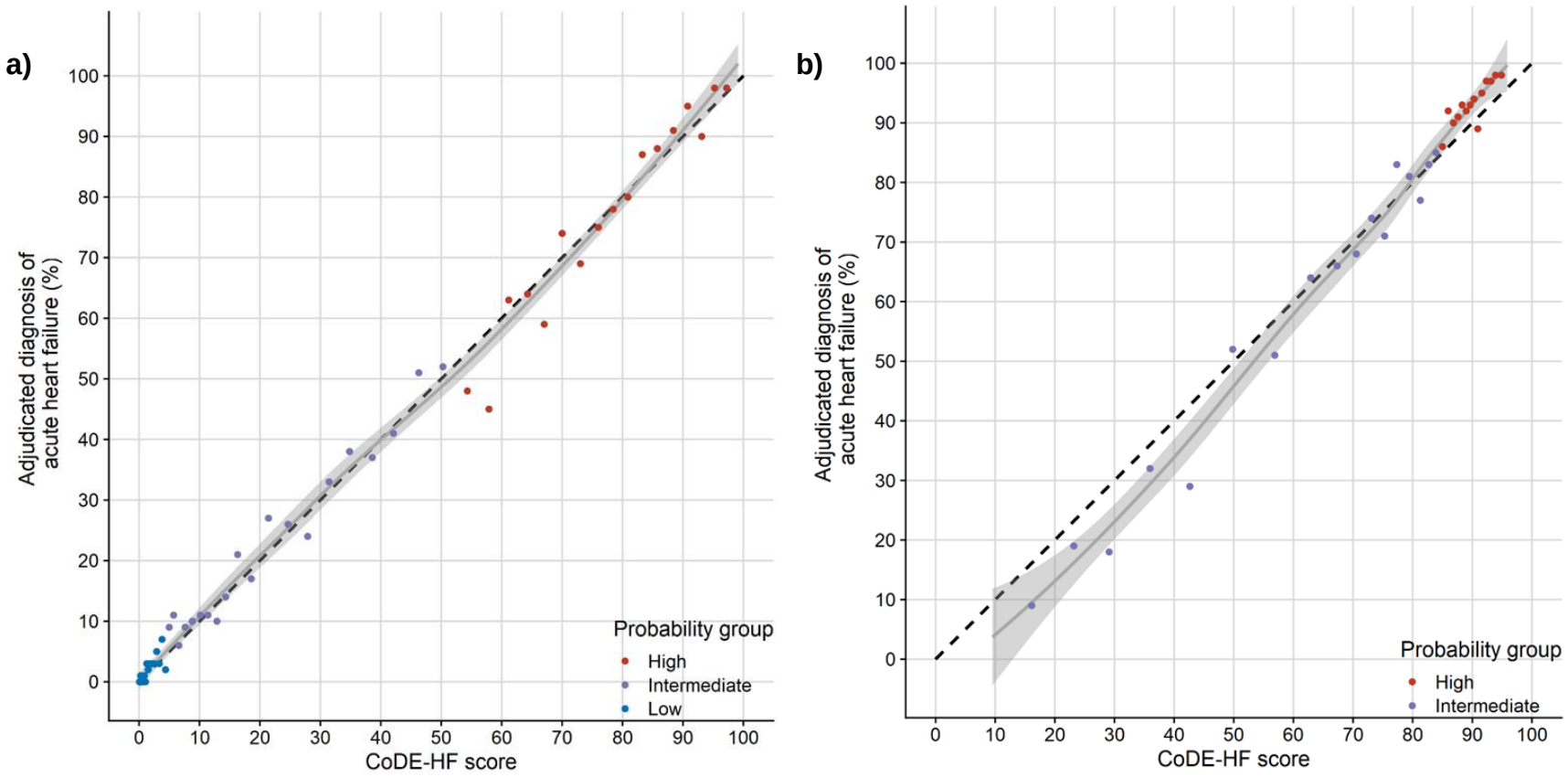


Figure 6.18 Calibration of the CoDE-HF score with the observed proportion of patients with acute heart failure.
(a) Receiver operator curve in patients without prior heart failure (area under curve of 0.851 [0.841-0.860], 0.914 [0.904-0.923] and 0.929 [0.918-0.939] for guideline-recommended NT-proBNP thresholds and the CoDE-HF score in the training cohort and external validation cohort respectively)
(b) Receiver operator curve in patients with prior heart failure (area under curve of 0.738 [0.719-0.756], 0.833 [0.814-0.852] and 0.787 [0.744-0.830] for guideline-recommended NT-proBNP thresholds and the CoDE-HF score in the training cohort and external validation cohort respectively).

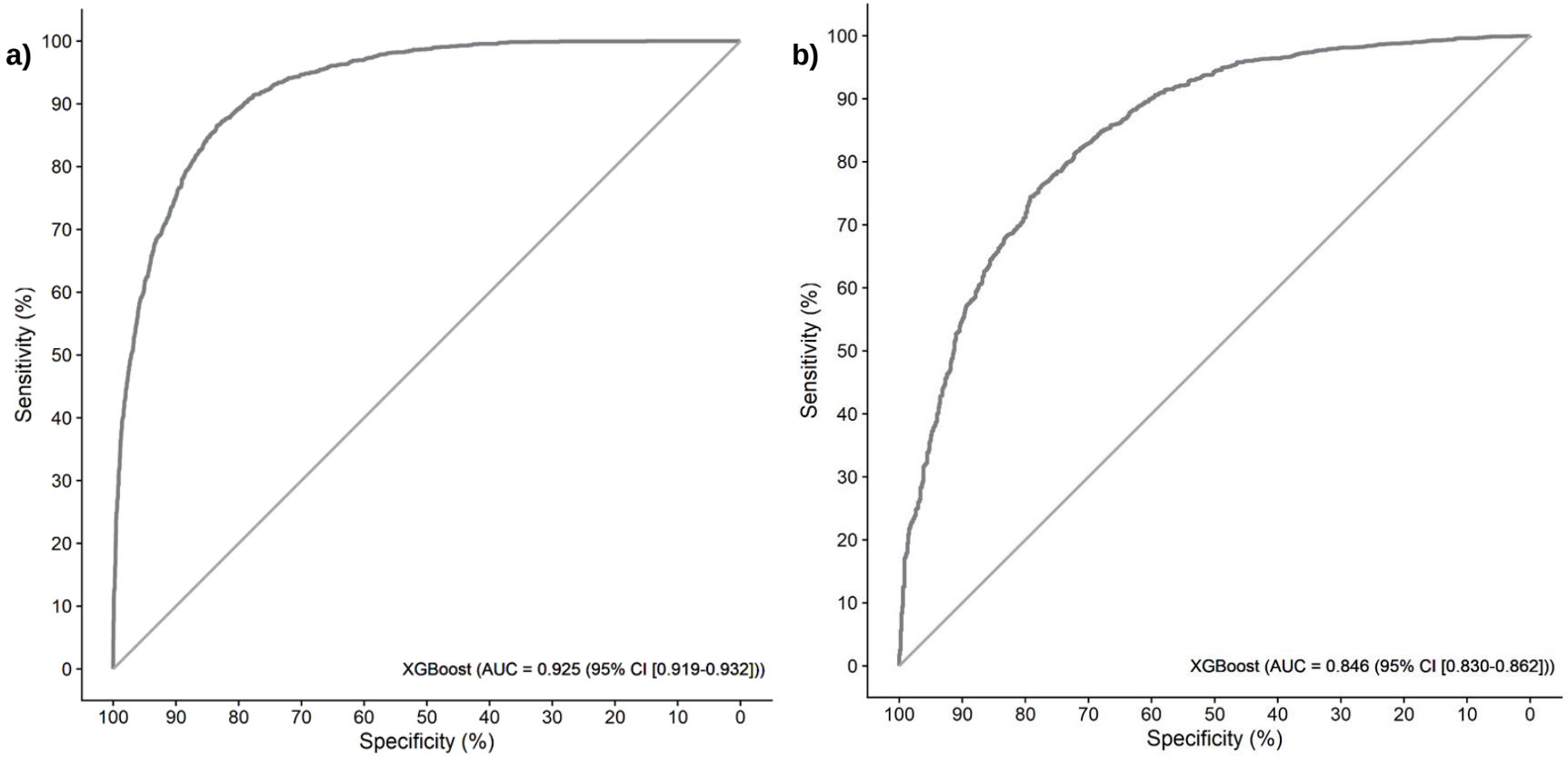


Figure 6.19 Discrimination of the CoDE-HF score.
(a) Receiver operator curve in patients without prior heart failure for the CoDE-HF score (area under curve 0.925 [0.919-0.932]).
(b) Receiver operator curve in patients with prior heart failure for the CoDE-HF score (area under curve 0.946 [0.830-0.862]).

Table 6.11 Diagnostic performance of rule-in and rule-out thresholds and CoDE-HF scores for acute heart failure.

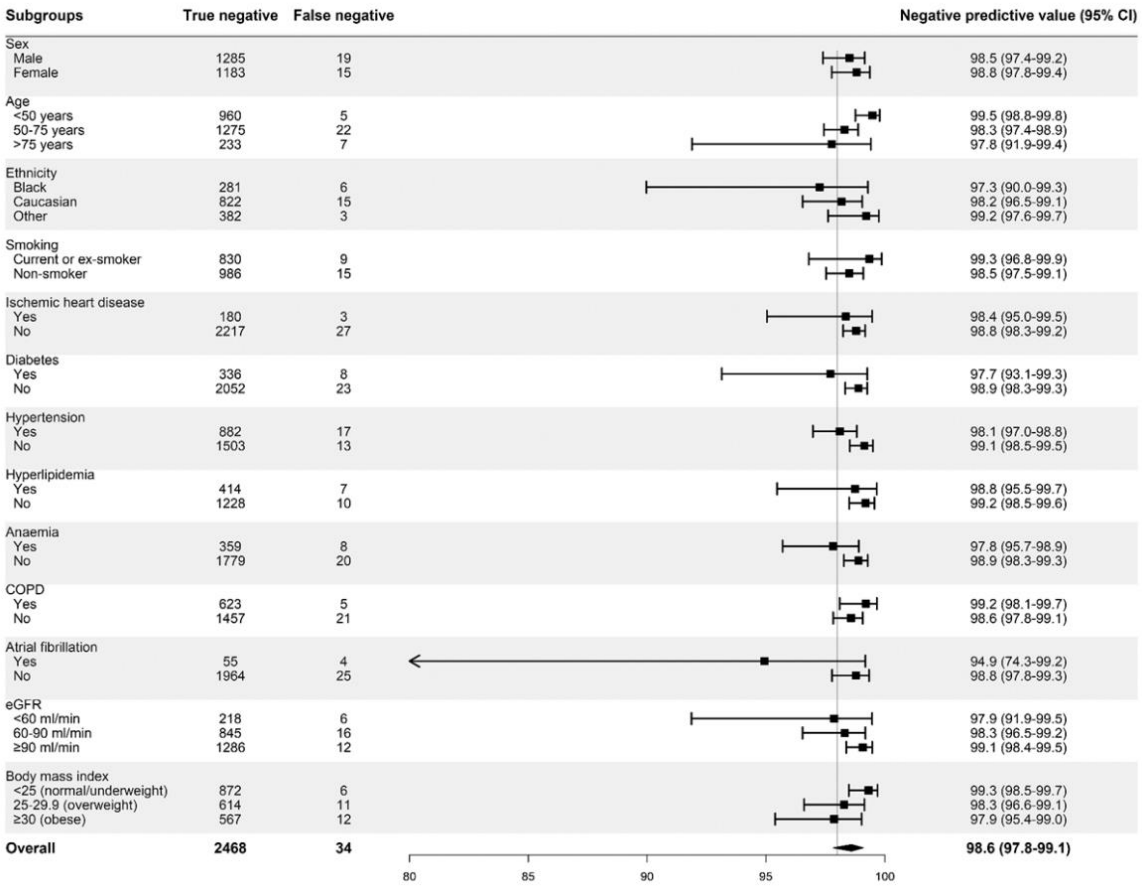
A) Rule-out thresholds and CoDE-HF scores.

	Threshold or score	True positive	False positive	True negative	False negative	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out	AUC	Brier score
All patients										
NT-proBNP	100 pg/mL	4507	4011	1809	42	97.8 (95.8-98.8)	99.3 (98.5-99.7)	17.9%	0.872 (0.865-0.879)	
Patients without prior heart failure										
CoDE-HF	4.7	1768	1938	2468	34	98.6 (97.8-99.1)	98.1 (96.9-98.9)	40.3%	0.925 (0.919-0.932)	0.099

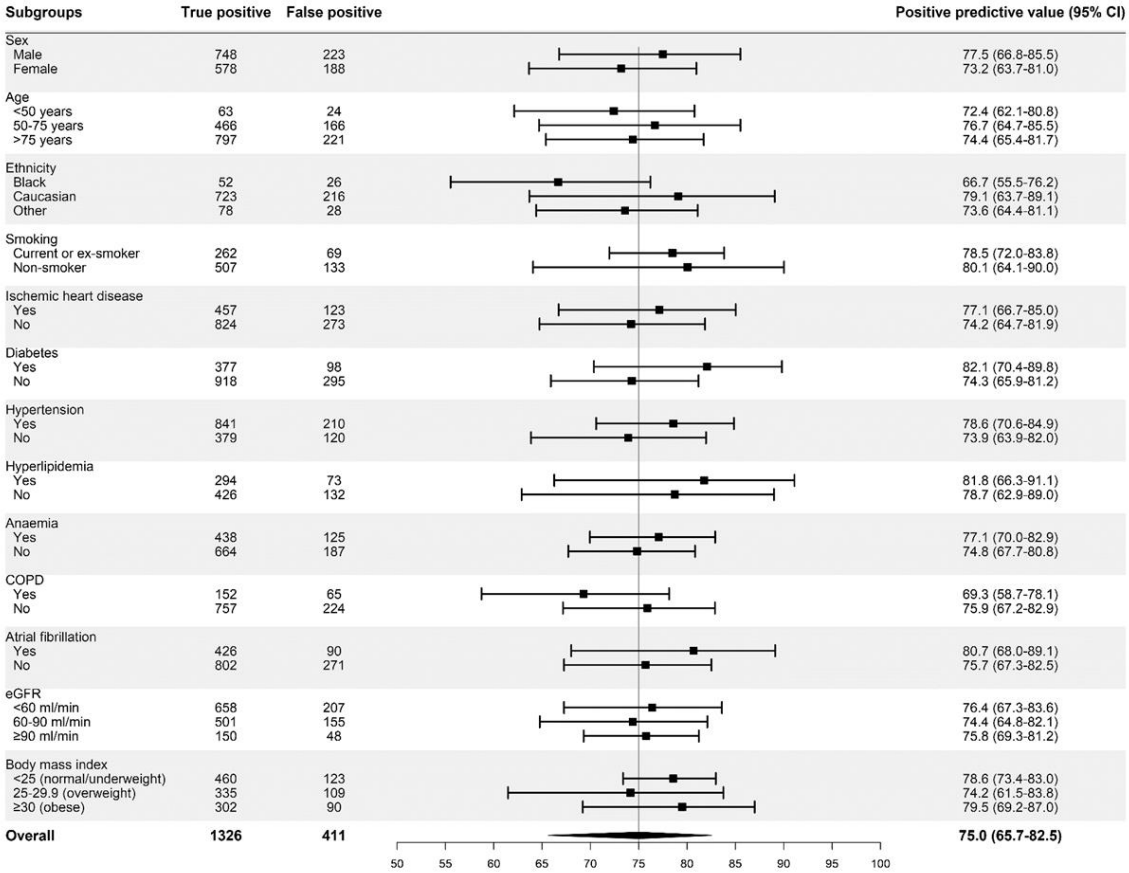
B) Rule-in thresholds and CoDE-HF scores.

	Threshold or score	True positive	False positive	True negative	False negative	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in	AUC	Brier score
All patients										
NT-proBNP	1000 pg/mL	3959	1511	4309	590	74.9 (64.4-83.2)	76.1 (65.6-84.2)	52.8%	0.872 (0.865-0.879)	
Patients without prior heart failure										
CoDE-HF	51.2	1326	411	3995	476	75.0 (65.7-82.5)	92.2 (87.5-95.2)	28.0%	0.925 (0.919-0.932)	0.099
Patients with prior heart failure										
CoDE-HF	84.5	1325	95	738	961	92.7 (89.1-95.2)	90.2 (84.0-94.1)	45.5%	0.846 (0.830-0.862)	0.130

a)


Subgroups
True negative
False negative
Negative predictive value (95% CI)
Sex
Male 1285 19 98.5 (97.4-99.2)
Female 1183 15 98.8 (97.8-99.4)
Age
<50 years 960 5 99.5 (98.8-99.8)
50-75 years 1275 22 98.3 (97.4-98.9)
>75 years 233 7 97.8 (91.9-99.4)
Ethnicity
Black 281 6 97.3 (90.0-99.3)
Caucasian 822 15 98.2 (96.5-99.1)
Other 382 3 99.2 (97.6-99.7)
Smoking
Current or ex-smoker 830 9 99.3 (96.8-99.9)
Non-smoker 986 15 98.5 (97.5-99.1)
Ischemic heart disease
Yes 180 3 98.4 (95.0-99.5)
No 2217 27 98.8 (98.3-99.2)
Diabetes
Yes 336 8 97.7 (93.1-99.3)
No 2052 23 98.9 (98.3-99.3)
Hypertension
Yes 882 17 98.1 (97.0-98.8)
No 1503 13 99.1 (98.5-99.5)
Hyperlipidemia
Yes 414 7 98.8 (95.5-99.7)
No 1228 10 99.2 (98.5-99.6)
Anaemia
Yes 359 8 97.8 (95.7-98.9)
No 1779 20 98.9 (98.3-99.3)
COPD
Yes 623 5 99.2 (98.1-99.7)
No 1457 21 98.6 (97.8-99.1)
Atrial fibrillation
Yes 55 4 94.9 (74.3-99.2)
No 1964 25 98.8 (97.8-99.3)
eGFR
<60 ml/min 218 6 97.9 (91.9-99.5)
60-90 ml/min 845 16 98.3 (96.5-99.2)
≥90 ml/min 1286 12 99.1 (98.4-99.5)
Body mass index
<25 (normal/underweight) 872 6 99.3 (98.5-99.7)
25-29.9 (overweight) 614 11 98.3 (96.6-99.1)
≥30 (obese) 567 12 97.9 (95.4-99.0)
Overall 2468 34 98.6 (97.8-99.1)
80
85
90
95
100

b)



Subgroups	True positive	False positive	Positive predictive value (95% CI)
Sex			
Male	748	223	77.5 (66.8-85.5)
Female	578	188	73.2 (63.7-81.0)
Age			
<50 years	63	24	72.4 (62.1-80.8)
50-75 years	466	166	76.7 (64.7-85.5)
>75 years	797	221	74.4 (65.4-81.7)
Ethnicity			
Black	52	26	66.7 (55.5-76.2)
Caucasian	723	216	79.1 (63.7-89.1)
Other	78	28	73.6 (64.4-81.1)
Smoking			
Current or ex-smoker	262	69	78.5 (72.0-83.8)
Non-smoker	507	133	80.1 (64.1-90.0)
Ischemic heart disease			
Yes	457	123	77.1 (66.7-85.0)
No	824	273	74.2 (64.7-81.9)
Diabetes			
Yes	377	98	82.1 (70.4-89.8)
No	918	295	74.3 (65.9-81.2)
Hypertension			
Yes	841	210	78.6 (70.6-84.9)
No	379	120	73.9 (63.9-82.0)
Hyperlipidemia			
Yes	294	73	81.8 (66.3-91.1)
No	426	132	78.7 (62.9-89.0)
Anaemia			
Yes	438	125	77.1 (70.0-82.9)
No	664	187	74.8 (67.7-80.8)
COPD			
Yes	152	65	69.3 (58.7-78.1)
No	757	224	75.9 (67.2-82.9)
Atrial fibrillation			
Yes	426	90	80.7 (68.0-89.1)
No	802	271	75.7 (67.3-82.5)
eGFR			
<60 ml/min	658	207	76.4 (67.3-83.6)
60-90 ml/min	501	155	74.4 (64.8-82.1)
≥90 ml/min	150	48	75.8 (69.3-81.2)
Body mass index			
<25 (normal/underweight)	460	123	78.6 (73.4-83.0)
25-29.9 (overweight)	335	109	74.2 (61.5-83.8)
≥30 (obese)	302	90	79.5 (69.2-87.0)
Overall	**1326**	**411**	**75.0 (65.7-82.5)**

c)

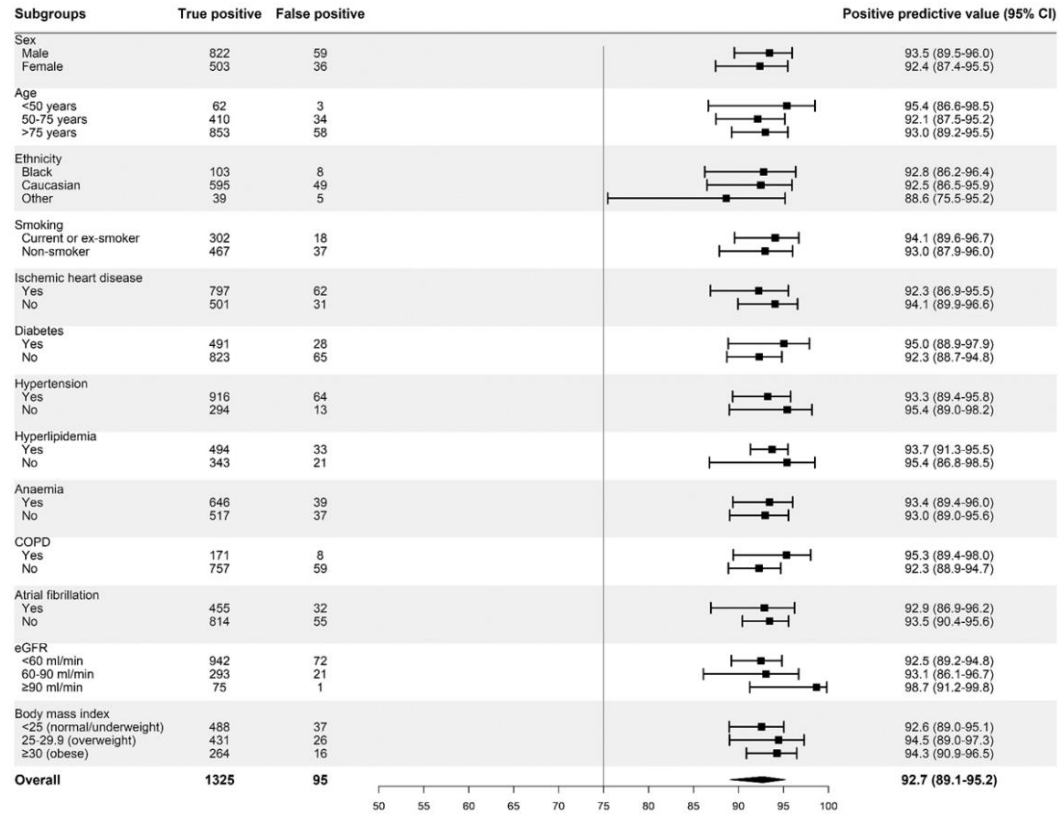


Figure 6.20 Diagnostic performance of the CoDE-HF score across patient subgroups. CoDE-HF incorporates NT-proBNP concentrations as a continuous measure and predefined simple objective clinical variables (age, estimated glomerular filtration rate, haemoglobin, body mass index, heart rate, blood pressure, peripheral oedema, prior history of heart failure, chronic obstructive pulmonary disease and ischaemic heart disease) to provide an individualized assessment of the likelihood of the diagnosis of acute heart failure.
(a) Negative predictive value of the CoDE-HF rule-out score of 4.2 in patients without prior heart failure across patient subgroups.
(b) Positive predictive value of the CoDE-HF rule-in score of 53.4 in patients without prior heart failure across patient subgroups.
(c) Positive predictive value of the CoDE-HF rule-in score of 82.4 in patients with prior heart failure across patient subgroups.

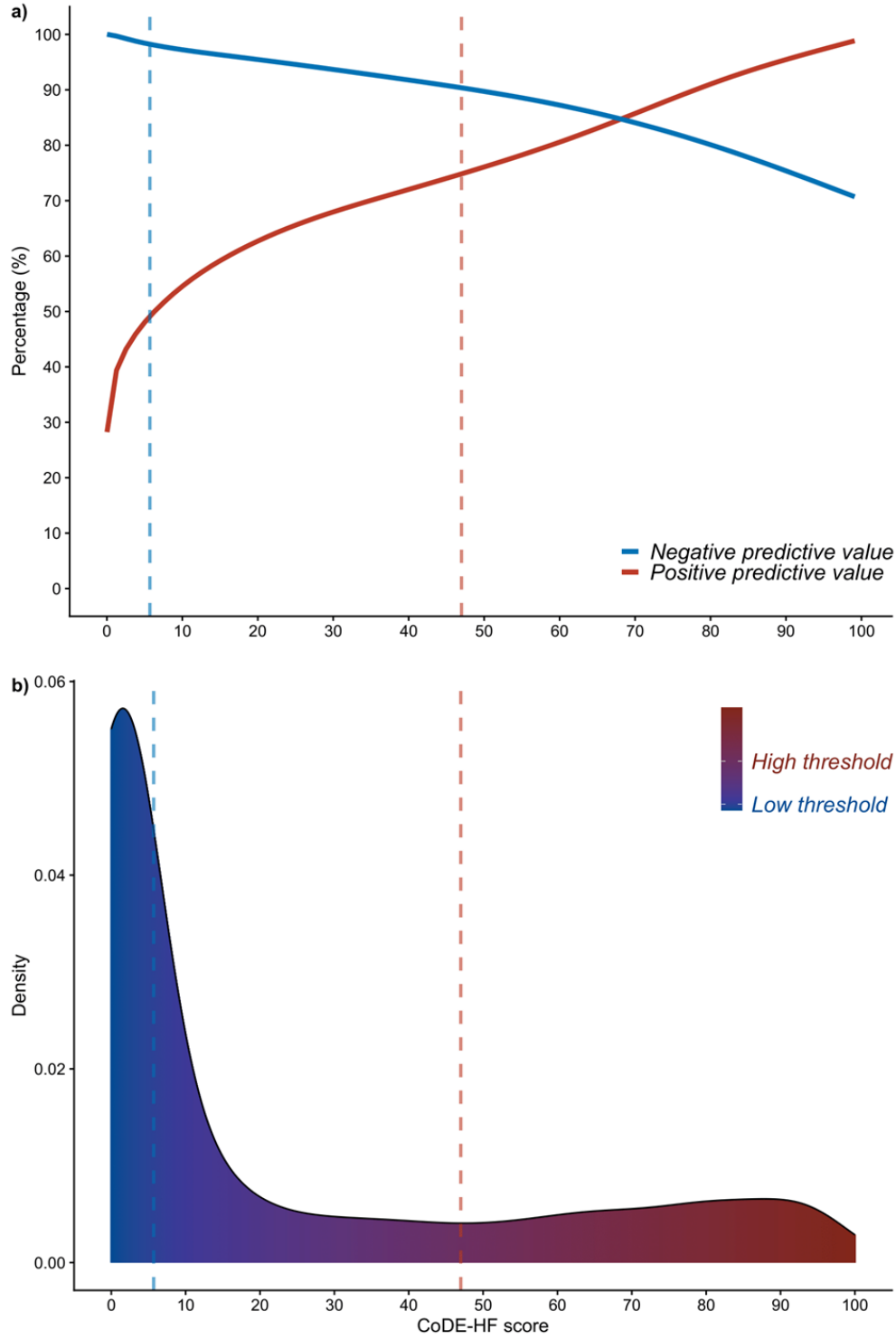


Figure 6.21 Diagnostic performance of the CoDE-HF score in patients without prior heart failure.
(a) Negative and positive predictive values of CoDE-HF scores. Blue vertical dashed line = target rule-out score of 4.7. Red vertical dashed line = target rule-in score of 51.2.
(b) Density plot of CoDE-HF score in patients without prior heart failure. The target rule-out and rule-in scores identify 40.3% of patients as low-probability and 28.0% as high-probability respectively.

In a decision curve analysis, CoDE-HF had superior net benefit compared to the NT-proBNP alone across all threshold probabilities (Figure 6.22). The performance of the CoDE-HF was marginally attenuated when trained without past medical history (AUC of 0.922 [0.916 to 0.929] and 0.841 [0.825 to 0.857] in those without and with prior heart failure). Internal-external cross-validation demonstrated good performance across cohorts for both models (Figure 6.23).

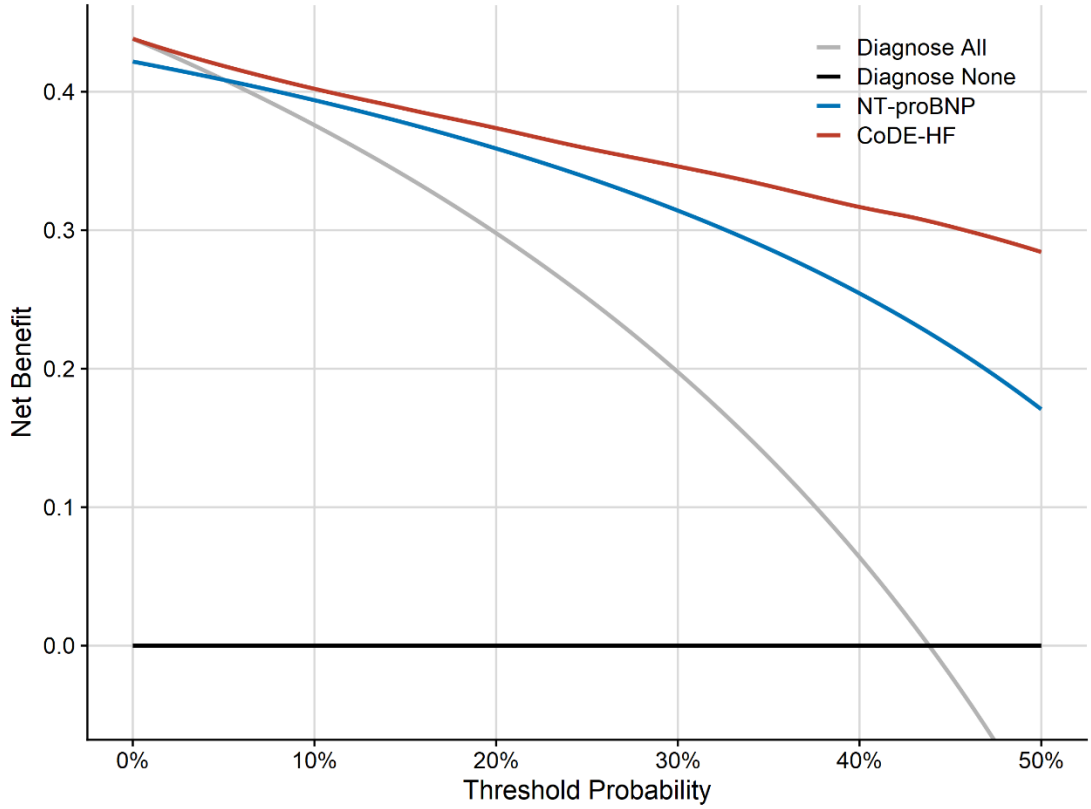
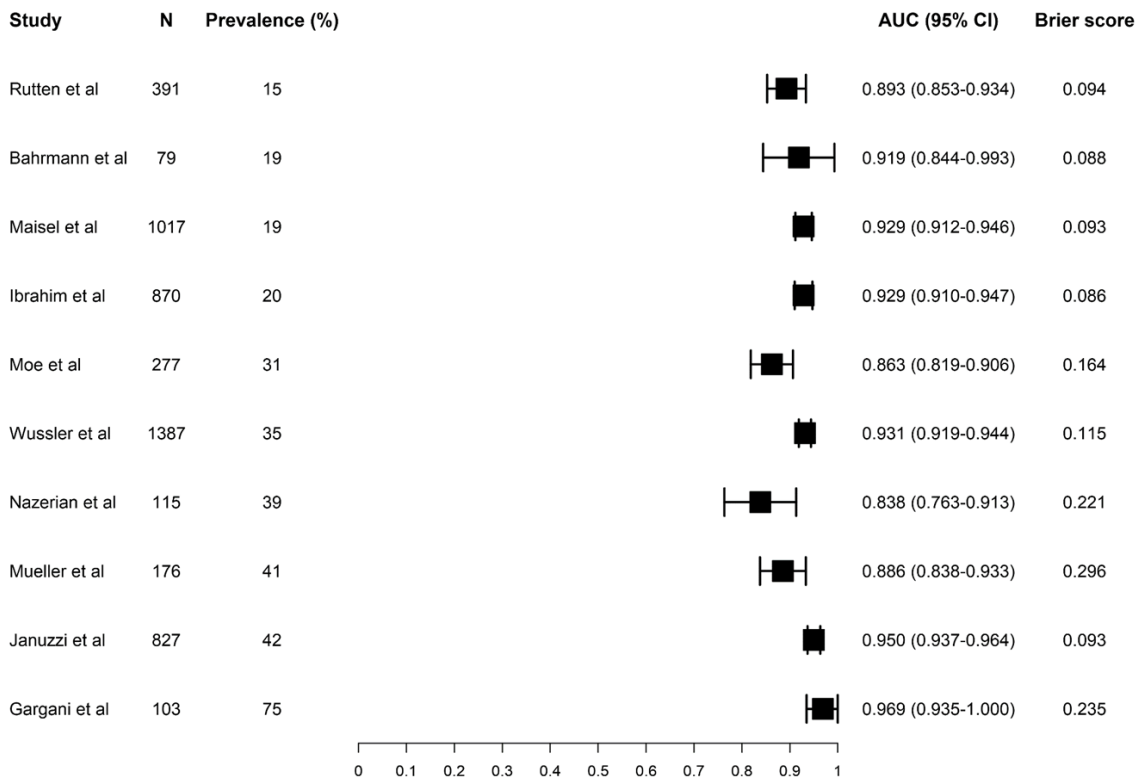


Figure 6.22 Decision curve analysis for CoDE-HF versus NT-proBNP alone.

a)


Study
N
Prevalence (%)
AUC (95% CI)
Brier score
Rutten et al
391
15
0.893 (0.853-0.934)
0.094
Bahrmann et al
79
19
0.919 (0.844-0.993)
0.088
Maisel et al
1017
19
0.929 (0.912-0.946)
0.093
Ibrahim et al
870
20
0.929 (0.910-0.947)
0.086
Moe et al
277
31
0.863 (0.819-0.906)
0.164
Wussler et al
1387
35
0.931 (0.919-0.944)
0.115
Nazerian et al
115
39
0.838 (0.763-0.913)
0.221
Mueller et al
176
41
0.886 (0.838-0.933)
0.296
Januzzi et al
827
42
0.950 (0.937-0.964)
0.093
Gargani et al
103
75
0.969 (0.935-1.000)
0.235
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

b)

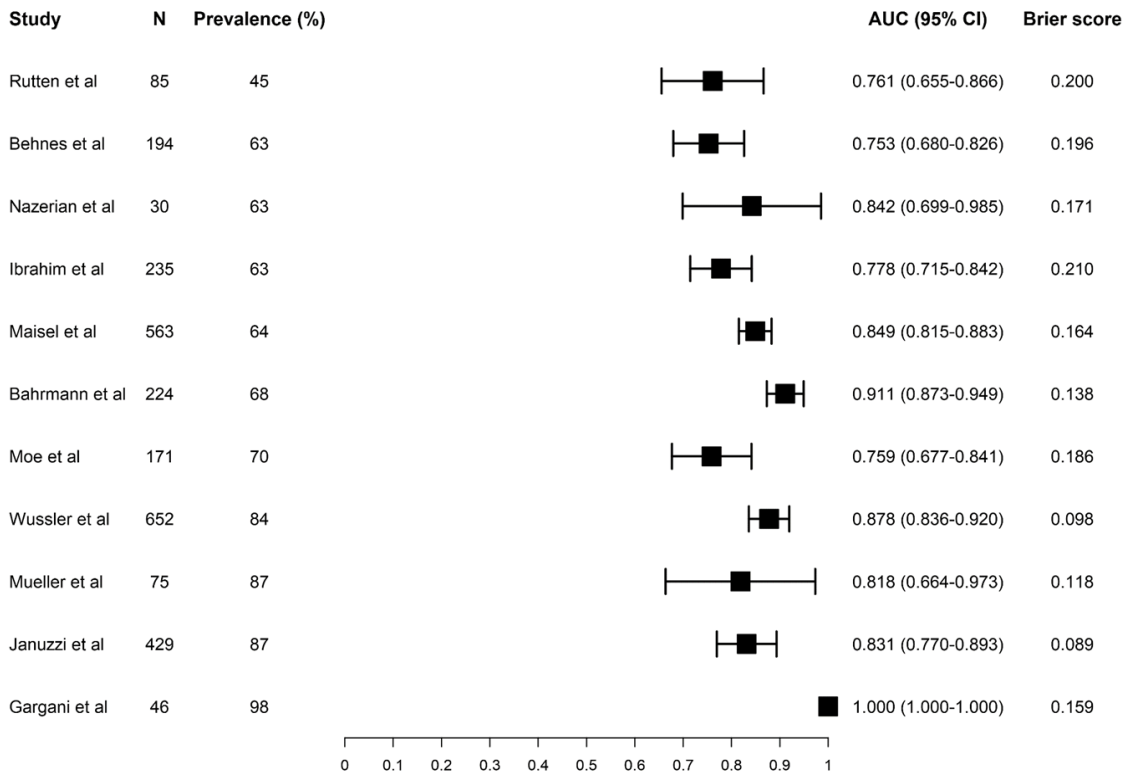


Figure 6.23 Internal-external cross-validation of CoDE-HF.
a) Patients without a prior history of heart failure across studies.
b) Patients with a prior history of heart failure across studies.

Patients who were identified as low-probability by CoDE-HF had a substantially lower rate of all-cause and cardiovascular mortality at 30-days and 1-year compared to those who were identified as intermediate and high-probability (30 day all-cause mortality: 1.0% *versus* 4.0% and 10.4%; 1 year all-cause mortality: 5.9% v*ersus* 17.8% and 33.4% respectively, and 30 day cardiovascular mortality: 0.2% *versus* 0.8% and 4.1%; 1 year cardiovascular mortality: 1.4% *versus* 3.4% and 16.3%, respectively) (Figure 6.24). In those with NT-proBNP concentrations <300 pg/mL compared to those ≥300 pg/mL, the all-cause mortality rates were 0.8% *versus* 7.6% at 30-days and 5.9% *versus* 26.6% at 1-year, respectively, whilst the cardiovascular mortality rates were 0.1% *versus* 2.6% at 30-days, and 1.3% *versus* 10.2% at 1-year, respectively (Figure 6.25).

Table 6.12 Mortality in the pool cohort stratified by CoDE-HF probability group and NT-proBNP threshold of 300 pg/mL.

	CoDE-HF score			**NT-proBNP concentration**	
	low probability	intermediate-probability	high-probability	<300 pg/mL	≥300 pg/mL
Cardiovascular mortality	n = 1239	n = 1721	n = 1593	n = 1448	n = 3105
30 days	2 (0.2%)	14 (0.8%)	66 (4.1%)	2 (0.1%)	80 (2.6%)
1 year	17 (1.4%)	58 (3.4%)	260 (16.3%)	19 (1.3%)	316 (10.2%)
All-cause mortality	n = 1251	n =1858	n =1769	n = 1460	n = 3418
30 days	12 (1.0%)	75 (4.0%)	184 (10.4%)	12 (0.8%)	259 (7.6%)
1 year	74 (5.9%)	330 (17.8%)	590 (33.4%)	86 (5.9%)	908 (26.6%)

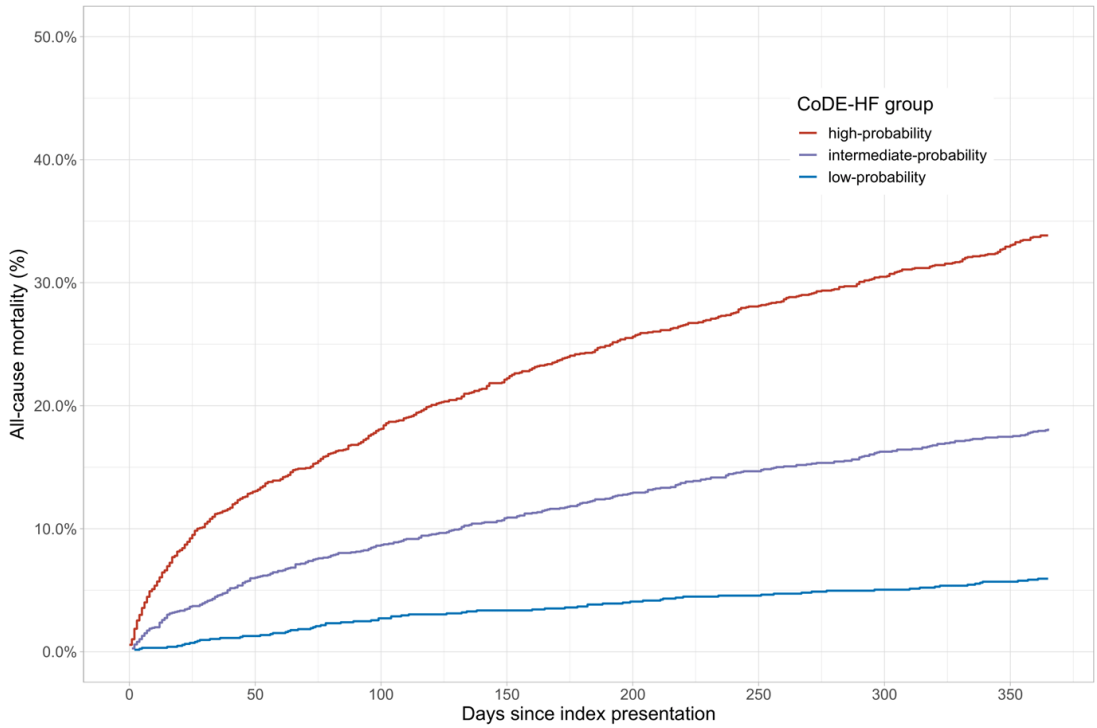


Figure 6.24 Cumulative incidence of all-cause mortality stratified by CoDE-HF probability group.

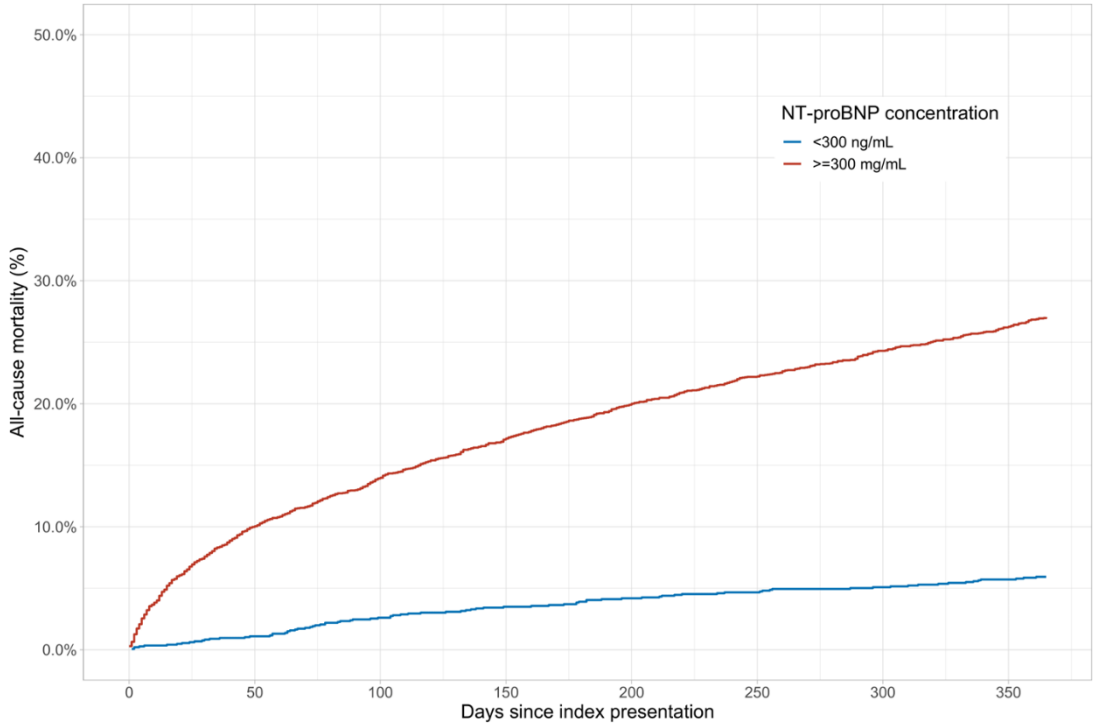


Figure 6.25 Cumulative incidence of all-cause mortality stratified by NT-proBNP concentration.

6.5 Discussion

We performed an individual patient-level data meta-analysis to evaluate the diagnostic performance of NT-proBNP thresholds in over 10,000 patients with suspected acute heart failure recruited in 14 prospective studies from 13 countries, and developed and externally validated a decision-support tool that uses NT-proBNP as a continuous measure. We report several important findings. First, the guideline-recommended threshold to rule-out acute heart failure was heterogenous across important patient subgroups.[107] Whilst performance was good in the overall population and in several subgroups such as younger patients and in women, NPV was substantially lower in older patients, and those with obesity or prior heart failure, where the false negative rates were between one in ten and one in five. Second, age-stratified thresholds performed well to rule-in the diagnosis of acute heart failure. However, the PPV was lower in younger patients. Third, although our optimized NT-proBNP thresholds of 100 pg/mL to rule-out and 1000 pg/mL to rule-in acute heart failure had excellent NPV and PPV in the overall population, performance was lower in older patients, and in those with prior heart failure and obesity. Finally, we developed and validated a decision-support tool, the CoDE-HF score, which had excellent diagnostic performance across all patient subgroups. This decision-support tool ruled-in and ruled-out acute heart failure more accurately than any approach using NT-proBNP thresholds alone.

6.5.1 Strengths of this study

To our knowledge, this is the largest study evaluating the diagnostic performance of NT-proBNP for acute heart failure to date. All studies were prospective and the final diagnosis was adjudicated by a panel of clinicians using all available information. Importantly, the availability of individual patient-level data across a large study population allowed a robust evaluation of diagnostic performance across a full range of possible NT-proBNP thresholds within patient subgroups and the development and validation of a novel diagnostic score.

The majority of national and international guidelines recommend the use of an NT-proBNP threshold of 300 pg/mL to rule-out acute heart failure,[33,97] based on evidence from multiple previous studies[103,107,191,192] which reported a NPV of 98% at this threshold. However, these studies were unable to evaluate the diagnostic performance within important patient subgroups. Our study, which included three-times as many patients compared to a previous study-level meta-analysis,[107] demonstrated a lower overall NPV at the 300 pg/mL threshold, with a pooled meta-estimate of 94.6%. More importantly, the NPV was markedly lower in key subgroups, such as older patients and those with prior heart failure, ischaemic heart disease and obesity. Moreover, nearly 70% of all patients had an NT-proBNP concentration above the 300 pg/mL threshold, highlighting the limitation of using a single threshold in practice. Although a lower threshold of 100 pg/mL achieved an overall NPV of 98%, performance was poorer within important patient subgroups.

Furthermore, the age-specific and optimized thresholds to rule-in acute heart failure had heterogenous performance within patient subgroups, particularly in those without prior heart failure. This heterogeneity in diagnostic performance is particularly concerning as our patient populations are aging and living with more comorbidities. This raises the question as to whether clinical guidelines should continue to recommend use of uniform thresholds when NT-proBNP is influenced by many risk factors and comorbidities.

6.5.2 CoDE-HF performance

To improve the clinical utility of NT-proBNP, we developed and externally validated a clinical decision-support tool, the CoDE-HF score. This score incorporates NT-proBNP as a continuous measure together with simple, objective clinical variables to provide an individualized assessment of the likelihood of the diagnosis of acute heart failure. We demonstrated that the diagnostic performance of the CoDE-HF score was robust across patient subgroups. CoDE-HF was able to rule-out and rule-in the diagnosis of acute heart failure in a larger proportion of patients compared to using optimized NT-proBNP thresholds alone. We believe this finding is intuitive given that NT-proBNP is a continuous marker of risk[193] and concentrations are influenced by other patient factors, such as body-mass index, age and renal function.[194,195] Whilst these proportions were calculated based on pre-specified performance criteria, we acknowledge that these targets may not be universally supported, and different healthcare settings may have different tolerances for risk. The advantage of using a decision-support tool, such as

CoDE-HF, is that clinicians or institutions have the option to select the diagnostic performance criteria used to guide local decisions depending on their priorities and the availability of echocardiography or heart failure specialists.

We anticipate that CoDE-HF, our novel machine learning-based decision-support tool has the potential to improve the triage of patients with suspected acute heart failure presenting to multiple medical specialties and to transform their care by facilitating more accurate diagnosis. Previous studies have showed that prompt and accurate delivery of evidence-based therapies for patients with acute heart failure can lead to a substantial reduction in mortality and length of hospital stay while delays are associated with worse outcomes.[196] Furthermore, CoDE-HF uses routinely collected variables and therefore can be embedded within the clinical workflow as part of the triage pathway to facilitate more efficient assessment. Currently, the vast majority of patients with suspected acute heart failure undergo echocardiography during the course of their hospital admission to guide their care, but only a proportion of these patients ultimately have the diagnosis.[167] This is a relatively time-consuming and resource-intensive specialist investigation. We anticipate that use of CoDE-HF to guide more accurate and judicious use of specialist services such as echocardiography could lead to significant cost and efficiency savings for healthcare systems. Additionally, cost savings could also be made by triaging low-risk patients to outpatient care.

6.5.3 Limitations

We acknowledge several limitations. First, we were able to obtain individual patient-level data in 14 out of 31 studies that met our eligibility criteria. The eligible studies that were not included had a similar prevalence of acute heart failure, publication dates and geographical coverage and therefore we do not think this has introduced selection bias. Second, in combining information from multiple studies, data were missing for some variables in several studies. To maximize the use of information, we have used a hierarchical multiple imputation method. Third, we did not have consistently recorded data from the electrocardiogram and chest X-ray to enable inclusion of these data in our model. The interpretation of NT-proBNP in patients with suspected acute heart failure should be made in conjunction with these investigations,[197] and future studies are needed to determine if approaches that integrate these investigations can improve the performance of CoDE-HF. Fourth, not all studies have adjudicated the diagnosis blinded to the results of NT-proBNP testing. In our sensitivity analysis, diagnostic performance was unchanged when we excluded the two studies where adjudication was unblinded. Fifth, the adjudicated diagnosis of acute heart failure did not differentiate between heart failure with reduced ejection fraction (HFrEF) and preserved ejection fraction (HFpEF).[198] The increasing prevalence of HFpEF in older patients may explain some of the heterogeneity observed across age groups, but current guidelines recommend the same NT-ProBNP threshold for both HFrEF and HFpEF.[33,97] Sixth, although the majority of studies included consecutive patients with acute dyspnoea, the prevalence of acute

heart failure was high and there may have been some selection bias. However, the performance of guideline-recommended and age-specific NT-proBNP thresholds were unchanged in a sensitivity analysis excluding studies at high-risk of bias. Finally, acute heart failure is a clinical syndrome and therefore the diagnostic adjudication itself has inherent uncertainty and variability across studies. This uncertainty may be greater among older adults which may, in part, explain some of the observed heterogeneity in diagnostic performance.

6.6 Conclusions

We demonstrated that the diagnostic performance of guideline-recommended NT-proBNP thresholds for acute heart failure varies across important patient subgroups. We developed and validated the CoDE-HF score, which combines NT-pro-BNP as a continuous measure with clinical variables using statistical modelling to determine the probability of acute heart failure for individual patients. This decision-support tool accurately ruled-in and ruled-out acute heart failure and performed consistently across all subgroups. Prospective studies are now required to evaluate the impact of implementing this decision-support tool on healthcare resource utilisation and patient outcomes.

In this next chapter, we will evaluated if the same algorithm could also work for different biomarkers used for the diagnosis of acute heart failure and if the performance is better than the current approach used in clinical practice.

CHAPTER 7

Machine learning to optimise use of B-type natriuretic peptide and mid-regional pro-atrial natriuretic peptide in the diagnosis of acute heart failure

Chapter 7 Machine learning to optimise use of B-type natriuretic peptide and mid-regional pro-atrial natriuretic peptide in the diagnosis of acute heart failure

7.1 Overview

B-type natriuretic peptide (BNP) and mid-regional pro-atrial natriuretic peptide (MR-proANP) testing are recommended to aid in the diagnosis of acute heart failure. However, the application of these biomarkers for optimal diagnostic performance is uncertain.

We performed a systematic review and harmonised individual patient-level data to evaluate the diagnostic performance of BNP and MR-proANP for the diagnosis of acute heart failure using random-effects meta-analysis. We subsequently developed and externally validated a decision-support tool called CoDE-HF for both BNP and MR-proANP that combines the natriuretic peptide concentrations with clinical variables using machine learning to report the probability of acute heart failure for an individual patient.

Fourteen studies from 12 countries provided individual patient-level data in 8,493 patients for BNP and 3,847 patients for MR-proANP, in whom, 48.3%

(4,105/8,493) and 41.3% (1,611/3,899) had an adjudicated diagnosis of acute heart failure, respectively. The negative predictive values of guideline-recommended thresholds of for BNP (100 pg/mL) and MR-proANP (120 pg/mL) were 93.6% (95% confidence interval 88.4-96.6%) and 95.6% (92.2-97.6%), respectively whilst the positive predictive value of these thresholds were 68.8% (62.9-74.2%) and 64.8% (56.3-72.5%), respectively. We observed significant heterogeneity in the diagnostic performance of these thresholds across important patient subgroups. In the external validation cohort, CoDE-HF was well calibrated with excellent discrimination in those without prior acute heart failure for both BNP and MR-proANP (area under the curve of 0.946 [0.933-0.958] and 0.943 [0.921-0.964], and Brier scores of 0.105 and 0.073, respectively). CoDE-HF performed consistently across all subgroups for both BNP and MR-proANP, and identified 30% and 65.7% at low-probability (negative predictive value of 99.1% [98.8%-99.3%] and 99.1% [98.8%-99.4%]), and 30% and 17.3% at high-probability (positive predictive value of 91.3% [90.7%-91.9%] and 70.0% [68.5%-71.4%]) in those without prior heart failure, respectively.

In an international collaborative analysis, we observed that guideline-recommended thresholds for BNP and MR-proANP to diagnose acute heart failure varied significantly across patient subgroups. A decision-support tool using machine learning to combine natriuretic peptides as a continuous measure and other clinical variables provides a more accurate and individualised approach.

7.2 Introduction

Decompensated acute heart failure is one of the commonest reasons for unplanned hospital attendances.[167] However, an accurate and timely diagnosis is challenging because many other conditions can present with similar signs and symptoms. National and international guidelines therefore recommend the use of fixed natriuretic peptides threshold concentrations to aid in the diagnosis of acute heart failure.[80,103,107,168,199] Nevertheless, natriuretic peptides concentrations are known to be influenced by various patient factors such as body mass index, renal function and age which have important impact on its diagnostic performance.[169-171] This has, in part, limited the application of natriuretic peptides in routine clinical practice.

There are currently three natriuretic peptides recommended for the diagnosis of acute heart failure – N-terminal pro-B-type natriuretic peptide (NT-proBNP), B-type natriuretic peptide (BNP) and mid-regional pro-atrial natriuretic peptide (MR-proANP).[33] We previously demonstrated that guideline-recommended thresholds of NT-proBNP have relatively poor performance in older patients and those with obesity and prior heart failure (Chapter 6). We subsequently developed and validated a decision-support tool called CoDE-HF (Collaboration for the Diagnosis and Evaluation of Heart Failure) that uses machine learning to incorporate NT-ProBNP concentrations as a continuous variable and other simple and objective physiological and patient factors that are routinely collected during the patient's initial clinical assessment to calculate an individualised probability of

acute heart failure for each patient. We demonstrated that CoDE-HF ruled-in and ruled-out acute heart failure more accurately than any approach using NT-ProBNP thresholds alone.

In this collaborative international analysis, we evaluated the diagnostic performance of current guideline-recommended BNP and MR-proANP thresholds for acute heart failure across patient subgroups, and developed and externally validated the CoDE-HF decision-support tool for BNP and MR-proANP.

7.3 Methods

7.3.1 Study population

We performed a systematic review to identify studies that evaluated BNP and MR-proANP in the diagnosis of acute heart failure. We updated a previous review by Roberts et al[107] by searching Embase, Medline and the Cochrane Central Register of Controlled Trials for studies published up to 18th August 2021 using the following keywords: ‘heart failure’ and ‘natriuretic peptide’. Studies were included if they satisfied the following inclusion criteria: (1) enrolled patients ≥18 years with suspected acute heart failure in an acute care setting, (2) measured BNP or MR-proANP on blood samples obtained during the patients’ initial assessment on the day of the hospital attendance, and (3) adjudicated diagnosis of acute heart failure using an acceptable reference standard. Two investigators independently screened all studies identified in the systematic literature search and conflicts adjudicated by a third using a pre-specified protocol (PROSPERO register: CRD42019159407).

We contacted the corresponding authors of all eligible cohorts to request anonymized individual patient-level data on BNP and MR-proANP concentrations, adjudicated diagnosis of acute heart failure, demographics (age, sex, ethnicity), past medical history (heart failure, ischaemic heart disease, diabetes, hypertension, hyperlipidemia, smoking, asthma, chronic obstructive pulmonary disease, chronic kidney disease), physiological

variables (heart rate and blood pressure), and clinical haematology and biochemistry profiles. We checked the accuracy and completeness of the individual patient-level data with all corresponding authors prior to harmonization. All studies were conducted in accordance with the Declaration of Helsinki and with ethical approval to permit sharing of individual patient-level data to conduct this meta-analysis. Two investigators independently evaluated the risk of bias for each study using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) tool[178] with conflicts resolved by a third investigator.

7.3.2 BNP and MR-proANP threshold analysis

We used a two stage approach to calculate meta-estimates (95% confidence intervals) of the sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) of the guideline-recommended BNP and MR-proANP thresholds for acute heart failure (100 pg/mL and 120 pg/mL, respectively).[33,97] We calculated these metrics separately within each study, then pooled them across studies in a binomial-normal random effects model using the DerSimonian and Laird method.[200] We further evaluated the performance of these thresholds in the overall population and subsequently in prespecified subgroups that are known to influence biomarker levels and the diagnosis of acute heart failure (age, sex, ethnicity, body mass index, renal function, anaemia and the presence of comorbidities [prior heart failure, hypertension, hyperlipidemia, diabetes mellitus, atrial fibrillation, chronic obstructive pulmonary disease]).

7.3.3 Model development and validation

We developed and externally validated a decision-support tool (Collaboration for the Diagnosis and Evaluation of Heart Failure [CoDE-HF], https://decision-support.shinyapps.io/code-hf/) using extreme gradient boosting (XGBoost)[123] to compute a value (0-100) that corresponds to an individual patient's probability of acute heart failure. CoDE-HF was developed and validated for both BNP and MR-proANP separately. We ordered the studies chronologically and used the earliest studies that corresponded to 70% of the pooled study population as the training cohort whilst the most recent studies that corresponded to 30% of the pooled study population was used as the external validation cohort. We used this temporal validation approach to test the performance of CoDE-HF in the most contemporary cohorts as patient demographics, the availability of cardiac imaging and prevalence of different subtypes of heart failure may have evolved over time.

We developed and validated our model for patients with and without prior heart failure separately due to significant differences in the demography, comorbidities and the prevalence of acute heart failure in these two groups of patients. We used BNP and MR-proANP concentrations as a continuous measure and selected simple objective clinical variables that are known to be associated with acute heart failure, which were found to have the highest relative importance in our model training phase (age, estimated glomerular filtration rate, haemoglobin, body mass index, heart rate, blood pressure,

peripheral oedema, prior history of heart failure, chronic obstructive pulmonary disease and ischaemic heart disease).

We multiply imputed ten datasets using joint-modelling multiple imputation with random study specific covariance matrices fitted with a Markov chain Monte Carlo algorithm to account for missing data in the training cohort.[133] We performed ten iterations of 10-fold cross-validation for each model. The median score across the iterations and imputed datasets was used as the CoDE-HF score for each patient. We prespecified the high- and low-probability thresholds as the score that would classify the highest proportion of patients with a rule-in performance of 75% PPV and 90% specificity, and a rule-out performance of 98% NPV and 90% sensitivity, respectively. We subsequently evaluated the performance of each model in the external validation cohort across a range of diagnostic metrics (area under the receiver operator curve [AUC], Brier score, proportion of patients achieving the optimal high- and low-probability criteria, and PPV and NPV across patient subgroups). We also evaluated the incidence of all-cause and cardiovascular mortality stratified by CoDE-HF probability groups. All analyses were performed in R version 4.0.3.

7.4 Results

7.4.1 Study population

Fourteen studies from 12 countries provided individual patient-level data in 8,493 patients for BNP (mean age 69 [±16] years, 54% male), and 3,847 patients for MR-proANP (mean age 66 [±17] years, 58% male), in whom, 48.3% (4,105/8,493) and 41.3% (1,611/3,899) had an adjudicated diagnosis of acute heart failure, respectively. (Table 7.1, Figure 7.1, Table 7.2 and Table 7.3).[98,100,173,180-190]. Patients with a prior history of heart failure had a higher prevalence of acute heart failure than those without (75% *versus* 33% and 74% *versus* 27% for BNP and MR-proANP, respectively) (Table 7.4).

Table 7.1 Baseline characteristics of patients stratified by diagnosis of acute heart failure.

	BNP			MR-proANP		
	Overall	Patients with acute heart failure	Patients without acute heart failure	Overall	Patients with acute heart failure	Patients without acute heart failure
Number of participants	8493	4105	4388	3899	1611	2288
Men	4559 (53.7)	2287 (55.7)	2272 (51.8)	2258 (57.9)	1005 (62.4)	1253 (54.8)
Age, years						
<50	1126 (13.3)	271 (6.6)	855 (19.6)	680 (17.4)	94 (5.8)	586 (25.6)
50-75	3639 (43.1)	1569 (38.4)	2070 (47.4)	1826 (46.8)	639 (39.7)	1187 (51.9)
>75	3687 (43.6)	2244 (54.9)	1443 (33.0)	1393 (35.7)	878 (54.5)	515 (22.5)
Ethnicity						
Black	964 (27.7)	389 (24.7)	575 (30.1)	473 (19.1)	118 (13.9)	355 (21.8)
Caucasian	2282 (65.5)	1088 (69.2)	1194 (62.5)	1338 (54.0)	566 (66.9)	772 (47.3)
Other	237 (6.8)	96 (6.1)	141 (7.4)	667 (26.9)	162 (19.1)	505 (30.9)
Past medical history						
Prior heart failure	2943 (36.3)	2219 (56.3)	724 (17.3)	1199 (31.2)	884 (55.3)	315 (14.0)

Ischaemic heart disease	2632 (36.4)	1687 (49.7)	945 (24.7)	1150 (30.0)	746 (47.0)	404 (18.0)
Diabetes mellitus	1756 (26.5)	1029 (32.5)	727 (21.1)	1047 (27.0)	558 (34.7)	489 (21.5)
Hypertension	4167 (62.7)	2242 (72.4)	1925 (54.2)	2529 (65.4)	1241 (77.6)	1288 (56.8)
Hyperlipidemia	1350 (37.5)	751 (44.4)	599 (31.4)	1421 (40.2)	705 (49.5)	716 (33.9)
Current or ex-smoker	864 (31.7)	306 (26.0)	558 (35.9)	672 (27.5)	187 (22.5)	485 (30.1)
Asthma	372 (19.0)	47 (6.4)	325 (26.5)	488 (22.2)	44 (6.3)	444 (29.6)
COPD	2077 (34.4)	725 (26.1)	1352 (41.5)	1060 (27.5)	353 (22.2)	707 (31.3)
Atrial fibrillation	1380 (21.5)	1033 (32.6)	347 (10.6)	722 (19.8)	556 (34.8)	166 (8.1)
Chronic kidney disease	931 (22.2)	696 (34.9)	235 (10.7)	793 (20.7)	572 (36.1)	221 (9.8)
Body mass index, kg/m^2						
<25	2503 (40.6)	1242 (41.3)	1261 (40.0)	1311 (40.1)	572 (39.6)	739 (40.5)
25-29	1727 (28.0)	892 (29.7)	835 (26.5)	973 (29.8)	474 (32.8)	499 (27.3)
≥30	1928 (31.3)	873 (29.0)	1055 (33.5)	986 (30.2)	398 (27.6)	588 (32.2)
Physiological parameters						
Heart rate, beats per minute	92.1 (23.4)	91.6 (25.4)	92.6 (21.3)	92.3 (23.3)	92.0 (26.0)	92.5 (21.1)

Systolic blood pressure, mmHg	139.8 (28.9)	140.7 (31.3)	139.0 (26.5)	139.6 (27.8)	140.4 (30.4)	139.0 (25.8)
Diastolic blood pressure, mmHg	79.2 (18.0)	79.9 (19.5)	78.5 (16.5)	80.7 (17.2)	81.6 (18.9)	80.1 (15.8)
Clinical haematology and biochemistry						
Haemoglobin, g/dL	13.1 (4.9)	12.7 (4.5)	13.4 (5.2)	13.1 (2.1)	12.8 (2.1)	13.4 (2.1)
eGFR, mL/min/1.73 m^2	65.9 (30.5)	56.8 (27.6)	74.5 (30.6)	72.0 (32.0)	58.8 (29.4)	81.8 (30.3)
BNP or MR-proANP, pg/mL	255.1 [60.0, 801.0]	729.0 [353.0, 1265.0]	70.4 [23.1, 189.0]	191.0 [71.3, 385.0]	390.7 [266.8, 598.5]	87.5 [47.5, 175.6]

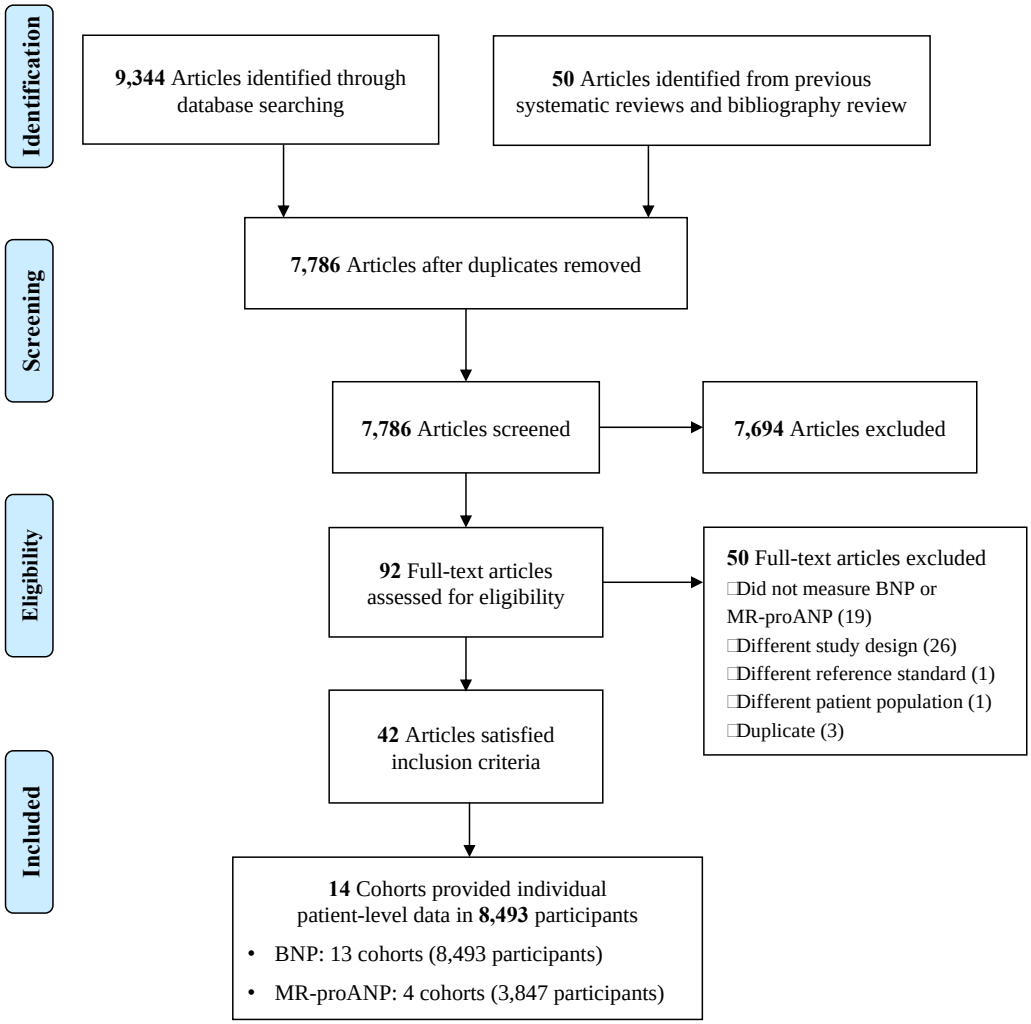


Figure 7.1 Flow diagram of study participants.

Table 7.2 Baseline characteristics of patients with BNP measurements within each study.

	Overall	Chenevier-Gobeaux et al	Chung et al	Collins et al	Coste et al	De Filippi et al	Lokuge et al
Number of participants	8493	380	380	143	330	699	831
Men	4559 (53.7)	189 (49.7)	189 (49.7)	63 (44.1)	154 (46.7)	428 (61.2)	380 (45.7)
Age, years							
< 50	1126 (13.3)	5 (1.3)	5 (1.3)	3 (2.1)	60 (20.8)	52 (7.4)	125 (15.0)
50-75	3639 (43.1)	121 (31.8)	121 (31.8)	41 (28.7)	144 (49.8)	278 (39.8)	443 (53.3)
>75	3687 (43.6)	254 (66.8)	254 (66.8)	99 (69.2)	85 (29.4)	369 (52.8)	263 (31.6)
Ethnicity							
Black	964 (27.7)	NR	NR	NR	167 (50.6)	NR	318 (38.3)
Caucasian	2282 (65.5)	NR	NR	NR	0 (0.0)	NR	499 (60.0)
Other	237 (6.8)	NR	NR	NR	163 (49.4)	NR	14 (1.7)
Past medical history							
Prior Heart failure	2943 (36.3)	128 (33.8)	128 (33.8)	80 (55.9)	198 (63.9)	174 (24.9)	287 (36.1)
Ischemic heart disease	2632 (36.4)	124 (32.6)	124 (32.6)	68 (47.6)	132 (49.6)	NR	263 (33.1)
Diabetes Mellitus	1756 (26.5)	NR	NR	38 (26.6)	NR	NR	305 (38.2)
Hypertension	4167 (62.7)	153 (40.3)	153 (40.3)	77 (53.8)	234 (73.1)	NR	NR
Hyperlipidemia	1350 (37.5)	NR	NR	NR	NR	NR	NR

Current or ex-smoker	864 (31.7)	NR	NR	55 (38.5)	NR	NR	NR
Asthma	372 (19.0)	NR	NR	NR	NR	NR	NR
COPD	2077 (34.4)	127 (33.4)	127 (33.4)	NR	NR	NR	NR
Atrial fibrillation	1380 (21.5)	NR	NR	41 (28.7)	NR	NR	175 (22.0)
Chronic Kidney Disease	931 (22.2)	NR	NR	NR	NR	NR	NR
Body mass index, kg/m^2							
<25	2503 (40.6)	NR	NR	NR	100 (31.7)	NR	195 (31.5)
25-30	1727 (28.0)	NR	NR	NR	61 (19.4)	NR	172 (27.8)
≥30	1928 (31.3)	NR	NR	NR	154 (48.9)	NR	252 (40.7)
Physiological parameters							
Heart rate, beats per minute	91.7 (23.7)	NR	92.1 (23.4)	NR	NR	90.2 (19.7)	NR
Systolic blood pressure, mmHg	140.0 (27.9)	NR	139.8 (28.9)	NR	NR	146.7 (32.1)	NR
Diastolic blood pressure, mmHg	79.7 (17.0)	NR	79.2 (18.0)	NR	NR	83.1 (20.5)	NR
Clinical hematology and biochemistry							
Hemoglobin, g/dL	13.1 (4.9)	NR	NR	NR	NR	NR	NR
eGFR, mL/min/1.73m^2	65.9 (30.5)	54.7 (19.9)	54.7 (19.9)	56.2 (25.3)	65.9 (29.9)	NR	63.3 (34.0)
BNP, pg/mL	255.1 [60.0, 801.0]	286.0 [85.0, 860.5]	286.0 [85.0, 860.5]	488.0 [118.0, 1140.0]	278.0 [58.0, 786.8]	450.0 [130.5, 1020.5]	337.0 [88.5, 811.0]

Adjudicated diagnosis of heart failure	4105 (48.3)	115 (30.3)	115 (30.3)	72 (50.3)	165 (50.0)	417 (59.7)	437 (52.6)

(continued)

	Maisel et al (1)	Maisel et al (2)	Mueller et al	Nakata et al	Plichart et al	Villacorta et al	Wussler et al
Number of participants	1586	1638	251	269	378	70	1619
Men	883 (55.7)	857 (52.3)	234 (93.2)	174 (64.7)	128 (33.9)	33 (47.1)	875 (54.0)
Age, years							
< 50	349 (22.0)	336 (20.5)	24 (9.6)	16 (5.9)	0 (0.0)	10 (14.3)	134 (8.3)
50-75	771 (48.6)	810 (49.5)	115 (45.8)	136 (50.6)	0 (0.0)	23 (32.9)	621 (38.4)
>75	466 (29.4)	492 (30.0)	112 (44.6)	117 (43.5)	378 (100.0)	37 (52.9)	864 (53.4)
Ethnicity							
Black	NR	476 (29.3)	0 (0.0)	NR	0 (0.0)	3 (4.3)	NR
Caucasian	NR	1087 (67.0)	251 (100.0)	NR	378 (100.0)	67 (95.7)	NR
Other	NR	60 (3.7)	0 (0.0)	NR	0 (0.0)	0 (0.0)	NR
Past medical history							
Prior Heart failure	527 (33.2)	568 (35.6)	75 (29.9)	NR	216 (57.1)	26 (37.1)	541 (33.5)
Ischemic heart disease	567 (39.3)	501 (31.6)	117 (46.6)	NR	126 (33.3)	30 (42.9)	575 (35.6)

Diabetes Mellitus	367 (25.4)	461 (28.5)	58 (23.1)	NR	52 (13.8)	22 (31.4)	391 (24.2)
Hypertension	879 (55.4)	1078 (66.9)	141 (56.2)	NR	259 (68.5)	36 (51.4)	1143 (70.9)
Hyperlipidemia	NR	569 (36.8)	NR	NR	68 (18.0)	18 (25.7)	695 (43.3)
Current or ex-smoker	NR	468 (29.4)	46 (18.3)	NR	77 (20.4)	36 (53.7)	NR
Asthma	NR	318 (20.0)	NR	NR	NR	2 (2.9)	NR
COPD	600 (41.4)	470 (29.5)	72 (28.7)	NR	88 (23.3)	31 (44.3)	555 (34.4)
Atrial fibrillation	256 (17.9)	242 (16.8)	83 (33.1)	NR	134 (35.4)	10 (14.3)	395 (24.4)
Chronic Kidney Disease	NR	244 (15.4)	74 (29.5)	NR	94 (24.9)	6 (8.7)	483 (29.9)
Body mass index, kg/m^2							
<25	526 (38.4)	490 (35.1)	93 (37.1)	204 (79.4)	168 (60.0)	20 (28.6)	707 (44.2)
25-30	366 (26.7)	389 (27.9)	100 (39.8)	40 (15.6)	69 (24.6)	36 (51.4)	494 (30.9)
≥30	477 (34.8)	517 (37.0)	58 (23.1)	13 (5.1)	43 (15.4)	14 (20.0)	400 (25.0)
Physiological parameters							
Heart rate, beats per minute	92.1 (23.0)	91.4 (22.8)	93.8 (26.1)	97.6 (26.9)	86.4 (19.2)	89.2 (18.0)	93.0 (24.6)
Systolic blood pressure, mmHg	140.9 (30.1)	140.8 (28.7)	140.1 (31.9)	138.2 (29.4)	129.6 (24.8)	134.7 (29.5)	138.6 (26.7)
Diastolic blood pressure, mmHg	78.8 (18.3)	80.8 (17.3)	81.8 (16.3)	79.4 (20.2)	70.5 (13.4)	80.9 (17.5)	79.7 (17.7)
Clinical hematology and biochemistry							

Hemoglobin, g/dL	12.8 (2.2)	13.7 (2.0)	12.7 (2.4)	11.7 (1.8)	26.4 (37.4)	13.2 (2.1)	12.9 (2.2)
eGFR, mL/min/1.73m^2	69.6 (27.5)	94.0 (44.6)	65.7 (23.9)	39.5 (20.4)	56.4 (21.6)	65.2 (28.6)	70.5 (30.4)
BNP, pg/mL	156.2 [29.8, 680.9]	295.7 [69.2, 911.2]	233.8 [60.6, 623.9]	291.0 [120.0, 670.8]	320.0 [75.2, 1052.0]	338.0 [77.0, 989.0]	164.0 [36.0, 573.8]
Adjudicated diagnosis of heart failure	744 (46.9)	137 (54.6)	131 (48.7)	235 (62.2)	36 (51.4)	903 (55.8)	566 (34.6)

Table 7.3 Baseline characteristics of patients with MR-proANP measurements within each study.

	Overall	Ibrahim et al, 2017	Maisel et al, 2010	Mueller et al, 2005	Wussler et al, 2019
Number of participants	3899	607	1635	251	1406
Men	2258 (57.9)	397 (65.4)	856 (52.4)	234 (93.2)	771 (54.8)
Age, years					
< 50	680 (17.4)	196 (32.3)	335 (20.5)	24 (9.6)	125 (8.9)
50-75	1826 (46.8)	350 (57.7)	807 (49.4)	115 (45.8)	554 (39.4)
>75	1393 (35.7)	61 (10.0)	493 (30.2)	112 (44.6)	727 (51.7)
Ethnicity					
Black	473 (19.1)	0 (0.0)	473 (29.2)	0 (0.0)	NR
Caucasian	1338 (54.0)	0 (0.0)	1087 (67.1)	251 (100.0)	NR
Other	667 (26.9)	607 (100.0)	60 (3.7)	0 (0.0)	NR
Past medical history					
Prior Heart failure	1199 (31.2)	88 (14.5)	565 (35.5)	75 (29.9)	471 (33.7)
Ischemic heart disease	1150 (30.0)	52 (8.6)	499 (31.5)	117 (46.6)	482 (34.4)
Diabetes Mellitus	1047 (27.0)	203 (33.4)	460 (28.5)	58 (23.1)	326 (23.2)
Hypertension	2529 (65.4)	329 (54.2)	1077 (66.9)	141 (56.2)	982 (70.1)
Hyperlipidemia	1421 (40.2)	274 (45.1)	566 (36.7)	NR	581 (42.1)

Current or ex-smoker	672 (27.5)	160 (26.5)	466 (29.4)	46 (18.3)	NR
Asthma	488 (22.2)	171 (28.2)	317 (20.0)	NR	NR
COPD	1060 (27.5)	46 (7.6)	469 (29.5)	72 (28.7)	473 (33.7)
Atrial fibrillation	722 (19.8)	54 (9.9)	241 (16.7)	83 (33.1)	344 (24.5)
Chronic Kidney Disease	793 (20.7)	54 (8.9)	243 (15.4)	74 (29.5)	422 (30.1)
Body mass index, kg/m^2					
<25	1311 (40.1)	111 (46.6)	490 (35.2)	93 (37.1)	617 (44.5)
25-30	973 (29.8)	59 (24.8)	388 (27.9)	100 (39.8)	426 (30.7)
≥30	986 (30.2)	68 (28.6)	515 (37.0)	58 (23.1)	345 (24.9)
Physiological parameters					
Heart rate, beats per minute	92.3 (23.3)	91.8 (21.4)	91.4 (22.8)	93.8 (26.1)	93.2 (24.0)
Systolic blood pressure, mmHg	139.6 (27.8)	138.2 (26.5)	140.8 (28.6)	140.1 (31.9)	138.7 (26.5)
Diastolic blood pressure, mmHg	80.7 (17.2)	80.7 (16.3)	80.8 (17.3)	81.8 (16.3)	80.5 (17.5)
Clinical hematology and biochemistry					
Hemoglobin, g/dL	13.1 (2.1)	13.4 (2.0)	12.9 (2.2)	13.7 (2.0)	13.1 (2.1)
eGFR, mL/min/1.73m^2	72.0 (32.0)	82.4 (29.3)	70.5 (30.3)	94.0 (44.6)	65.3 (29.2)
MR-proANP, pg/mL	191.0 [71.3, 385.0]	84.9 [41.8, 257.1]	174.5 [66.5, 369.4]	212.0 [101.0, 412.0]	243.5 [109.2, 439.8]
Adjudicated diagnosis of heart failure	1611 (41.3)	148 (24.4)	564 (34.5)	137 (54.6)	762 (54.2)

Table 7.4 Baseline characteristics of study patients stratified by prior history of heart failure.

	BNP			MR-proANP		
	Overall	No prior history of acute heart failure	Prior history of acute heart failure	Overall	No prior history of acute heart failure	Prior history of acute heart failure
Number of participants	8493	5175	2943	3899	2648	1199
Men	4559 (53.7)	2657 (51.3)	1675 (56.9)	2258 (57.9)	1474 (55.7)	755 (63.0)
Age, years						
<50	1126 (13.3)	859 (16.7)	231 (7.9)	680 (17.4)	592 (22.4)	80 (6.7)
50-75	3639 (43.1)	2299 (44.6)	1157 (39.6)	1826 (46.8)	1298 (49.0)	507 (42.3)
>75	3687 (43.6)	1998 (38.8)	1534 (52.5)	1393 (35.7)	758 (28.6)	612 (51.0)
Ethnicity						
Black	964 (27.7)	545 (27.0)	389 (28.5)	473 (19.1)	300 (17.5)	167 (23.0)
Caucasian	2282 (65.5)	1367 (67.8)	860 (62.9)	1338 (54.0)	852 (49.8)	451 (62.2)
Other	237 (6.8)	105 (5.2)	118 (8.6)	667 (26.9)	558 (32.6)	107 (14.8)

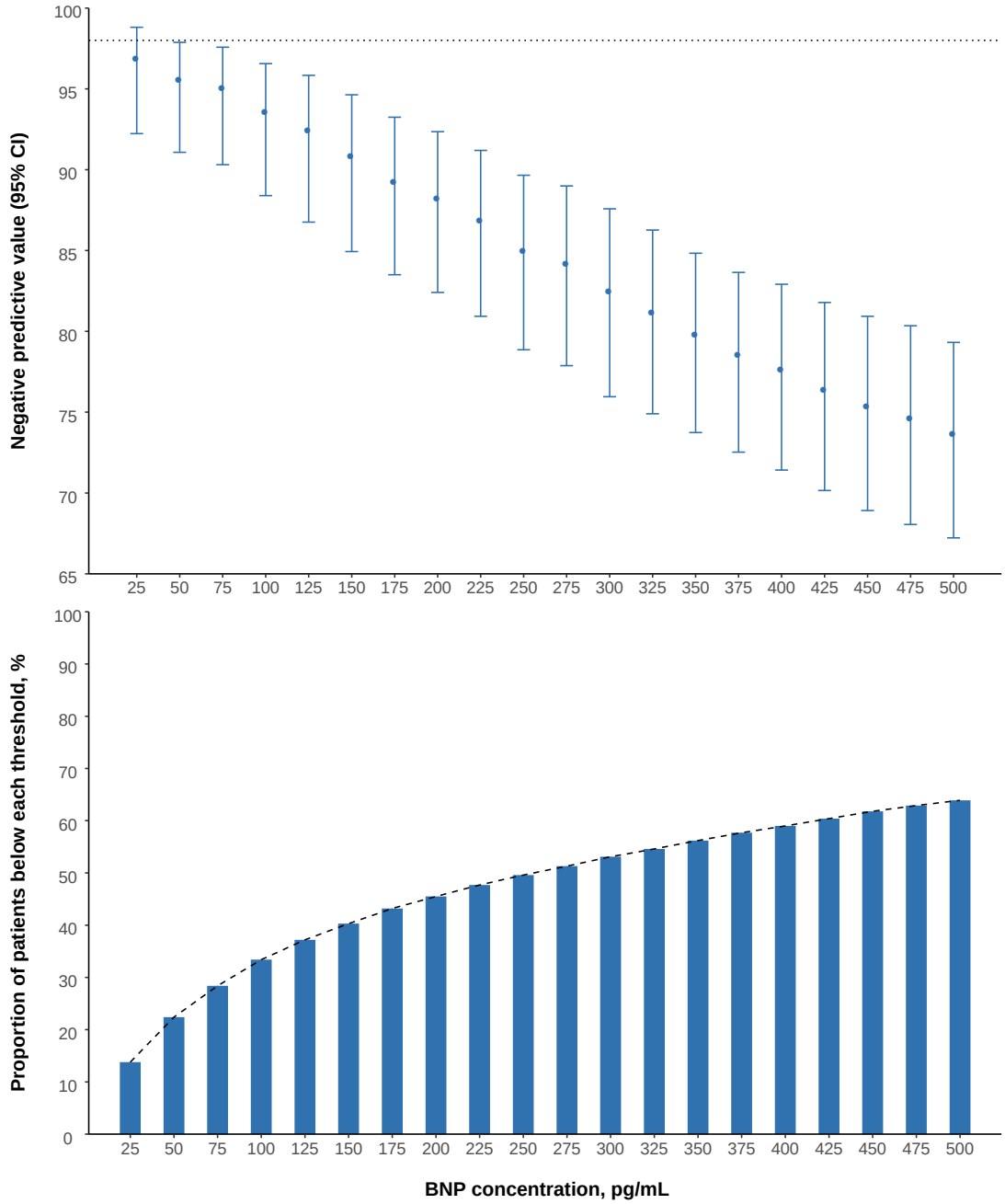
Past medical history						
Ischaemic heart disease	2632 (36.4)	1115 (24.7)	1484 (56.0)	1150 (30.0)	445 (16.9)	687 (58.6)
Diabetes mellitus	1756 (26.5)	912 (21.8)	822 (34.4)	1047 (27.0)	592 (22.4)	444 (37.2)
Hypertension	4167 (62.7)	2283 (55.3)	1831 (74.2)	2529 (65.4)	1525 (57.9)	965 (81.0)
Hyperlipidemia	1350 (37.5)	662 (29.3)	671 (51.2)	1421 (40.2)	780 (32.2)	621 (57.5)
Current or ex-smoker	864 (31.7)	559 (34.5)	289 (27.0)	672 (27.5)	501 (29.5)	155 (21.8)
Asthma	372 (19.0)	277 (22.5)	87 (12.5)	488 (22.2)	397 (26.0)	83 (13.1)
COPD	2077 (34.4)	1343 (34.8)	716 (33.4)	1060 (27.5)	696 (26.4)	345 (29.3)
Atrial fibrillation	1380 (21.5)	645 (16.1)	718 (30.4)	722 (19.8)	345 (14.3)	368 (31.4)
Chronic kidney disease	931 (22.2)	363 (13.7)	560 (36.7)	793 (20.7)	338 (12.8)	447 (38.0)
Body mass index, kg/m^2						
<25	2503 (40.6)	1453 (39.9)	813 (37.4)	1311 (40.1)	917 (42.6)	377 (35.2)
25-29	1727 (28.0)	1016 (27.9)	645 (29.7)	973 (29.8)	597 (27.7)	362 (33.8)
≥30	1928 (31.3)	1172 (32.2)	713 (32.8)	986 (30.2)	640 (29.7)	331 (30.9)
Physiological parameters						

Heart rate, beats per minute	92.1 (23.4)	93.3 (22.9)	89.6 (23.4)	92.3 (23.3)	93.5 (22.7)	89.5 (24.2)
Systolic blood pressure, mmHg	139.8 (28.9)	141.5 (27.8)	137.0 (30.3)	139.6 (27.8)	141.4 (26.7)	135.2 (29.5)
Diastolic blood pressure, mmHg	79.2 (18.0)	80.2 (17.4)	77.3 (18.5)	80.7 (17.2)	81.8 (16.8)	78.2 (17.8)
Clinical haematology and biochemistry						
Haemoglobin, g/dL	13.1 (4.9)	13.3 (5.0)	12.7 (5.1)	13.1 (2.1)	13.3 (2.1)	12.6 (2.1)
eGFR, mL/min/1.73 m^2	65.9 (30.5)	72.3 (31.1)	55.4 (26.9)	72.0 (32.0)	79.3 (31.4)	56.7 (27.5)
BNP or MR-proANP, pg/mL	255.1 [60.0, 801.0]	131.0 [34.0, 513.0]	615.0 [222.0, 1208.5]	191.0 [71.3, 385.0]	114.0 [53.1, 265.2]	386.6 [231.7, 604.7]
Adjudicated diagnosis of heart failure	4105 (48.3)	1724 (33.3)	2219 (75.4)	1611 (41.3)	714 (27.0)	884 (73.7)

7.4.2 Guideline-recommended BNP threshold

Pooled meta-estimates of NPV, sensitivity, PPV and specificity of the guideline-recommended BNP threshold of 100 pg/mL were 93.6% (95% confidence interval, 88.4-96.6%), 96.0% (93.2-97.6%), 68.8% (62.9-74.2%), and 56.5% (48.4-64.3%), respectively (Figure 7.2 and Table 7.5). Overall, 33.4% of patients had BNP concentrations below 100 pg/mL. However, there was marked heterogeneity in the performance of these thresholds across patient subgroups (Figure 7.3). NPV was lower in those with prior heart failure (76.7% [56.2-89.4%]), atrial fibrillation (71.5% [50.4-86.2%]) and obesity (86.4% [77.4-92.7%]). There was no BNP threshold that achieved our pre-specified optimal rule-out criteria (NPV of 98% and sensitivity of 90%). PPVs of the BNP threshold of 100 pg/mL were also heterogenous with lower performance in those without prior heart failure (56.0% [48.0-63.8%]), COPD (53.7% [38.2%-68.5%]) and normal renal function (60.3% [66.0-81.0%]). (Figure 7.4).

a)


Negative predictive value (95% CI)
100
95
90
85
80
75
70
65
25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Proportion of patients below each threshold, %
100
90
80
70
60
50
40
30
20
10
0
25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
BNP concentration, pg/mL

b)

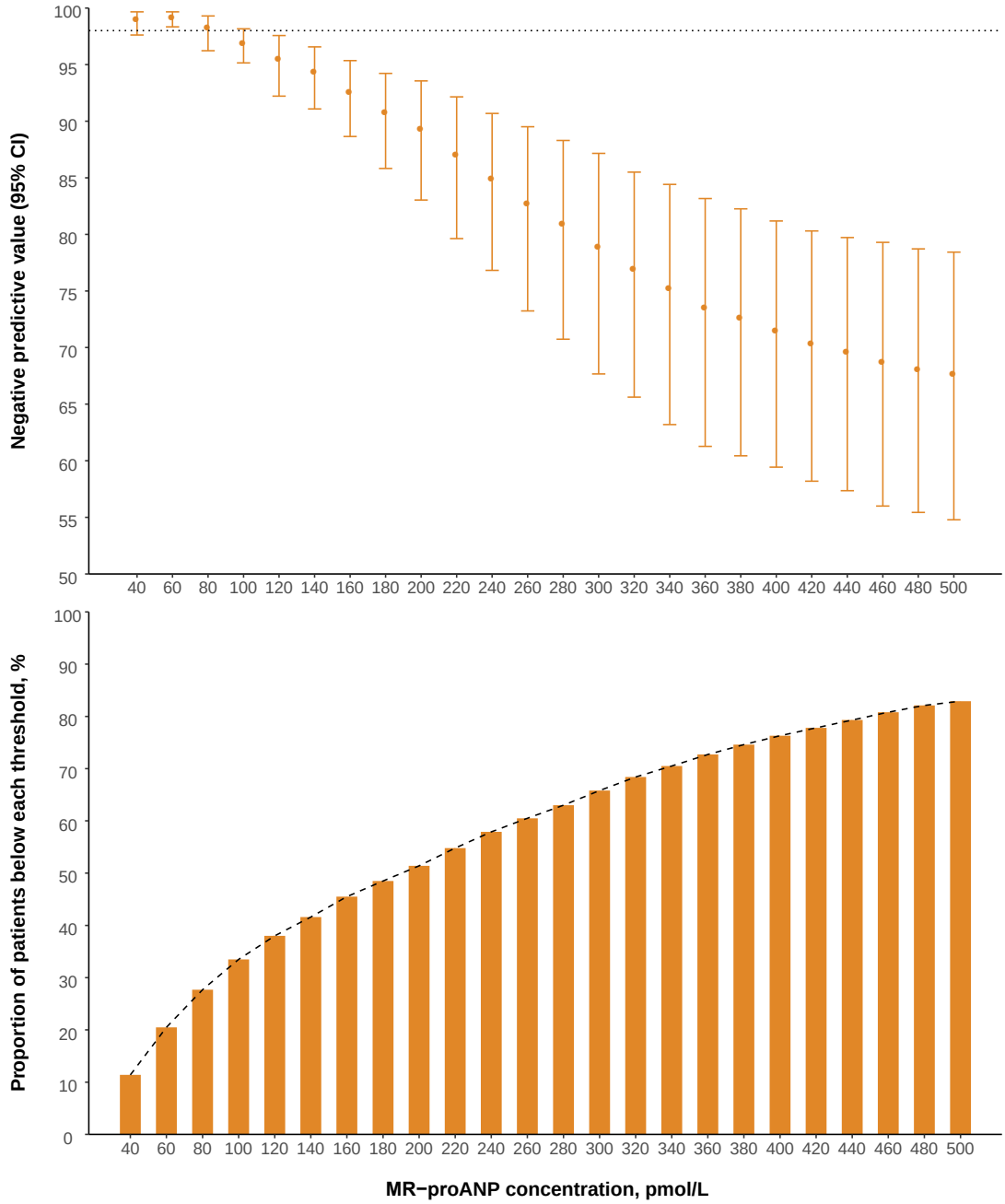


Figure 7.2 BNP and MR-proANP thresholds for acute heart failure.
(a) (top) Negative predictive values of BNP concentrations to rule-out a diagnosis of acute heart failure. (bottom) Cumulative proportion of patients presenting with suspected acute heart failure with BNP concentrations below each threshold.
(b) (top) Negative predictive values of MR-proANP concentrations to rule-out a diagnosis of acute heart failure. (bottom) Cumulative proportion of patients presenting with suspected acute heart failure with MR-proANP concentrations below each threshold.

Table 7.5 Diagnostic performance of BNP and CoDE-HF thresholds for acute heart failure.

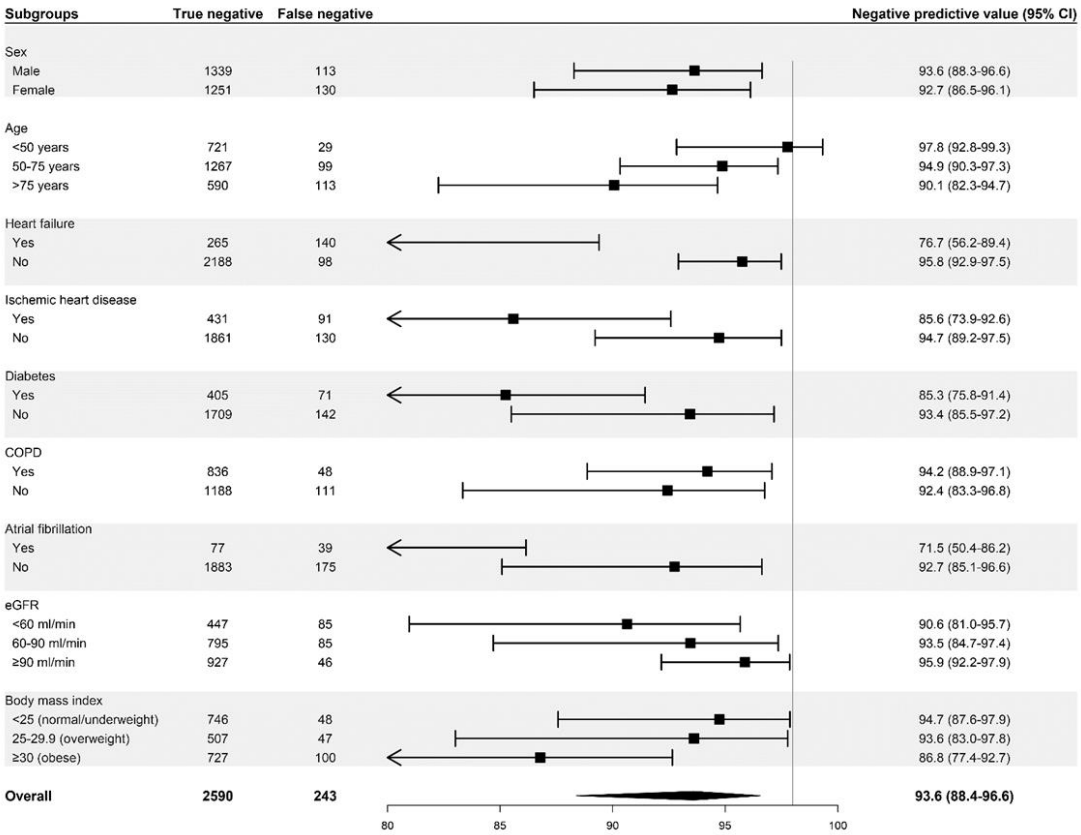
A) Rule-out thresholds and CoDE-HF scores.

	Threshold or score	True positive	False positive	True negative	False negative	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out
All patients								
BNP	100 pg/mL	3862	1798	2590	243	93.6 (88.4-96.6)	96.0 (93.2-97.6)	33.4%
Patients without prior heart failure								
CoDE-HF (internal validation)	3.6	669	882	1015	14	98.6 (97.7-99.2)	98.0 (96.6-98.8)	39.9%
CoDE-HF (external validation)	3.6	429	321	319	3	99.1 (98.8-99.3)	99.3 (99.1-99.5)	30.0%

B) Rule-in thresholds and CoDE-HF scores.

	Threshold or score	True positive	False positive	True negative	False negative	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in
All patients								
BNP	100 pg/mL	3862	1798	2590	243	68.8 (62.9-74.2)	56.5 (48.4-64.3)	66.6%
Patients without prior heart failure								
CoDE-HF (internal validation)	53.2	483	160	1737	200	75.0 (71.0-78.6)	90.9 (86.1-94.2)	24.9%
CoDE-HF (external validation)	53.2	294	28	612	138	91.3 (90.7-91.9)	95.6 (95.2-96.0)	30.0%
Patients with prior heart failure								
CoDE-HF (internal validation)	92.0	646	45	413	554	93.5 (91.4-95.1)	90.1 (85.6-93.4)	41.7%
CoDE-HF (external validation)	92.0	251	3	73	214	98.8 (98.6-99.0)	96.1 (95.6-96.5)	47.0%

a)


Subgroups
True negative
False negative
Negative predictive value (95% CI)
Sex
Male 1339 113 93.6 (88.3-96.6)
Female 1251 130 92.7 (86.5-96.1)
Age
<50 years 721 29 97.8 (92.8-99.3)
50-75 years 1267 99 94.9 (90.3-97.3)
>75 years 590 113 90.1 (82.3-94.7)
Heart failure
Yes 265 140 76.7 (56.2-89.4)
No 2188 98 95.8 (92.9-97.5)
Ischemic heart disease
Yes 431 91 85.6 (73.9-92.6)
No 1861 130 94.7 (89.2-97.5)
Diabetes
Yes 405 71 85.3 (75.8-91.4)
No 1709 142 93.4 (85.5-97.2)
COPD
Yes 836 48 94.2 (88.9-97.1)
No 1188 111 92.4 (83.3-96.8)
Atrial fibrillation
Yes 77 39 71.5 (50.4-86.2)
No 1883 175 92.7 (85.1-96.6)
eGFR
<60 ml/min 447 85 90.6 (81.0-95.7)
60-90 ml/min 795 85 93.5 (84.7-97.4)
≥90 ml/min 927 46 95.9 (92.2-97.9)
Body mass index
<25 (normal/underweight) 746 48 94.7 (87.6-97.9)
25-29.9 (overweight) 507 47 93.6 (83.0-97.8)
≥30 (obese) 727 100 86.8 (77.4-92.7)
Overall 2590 243 93.6 (88.4-96.6)
80
85
90
95
100

b)

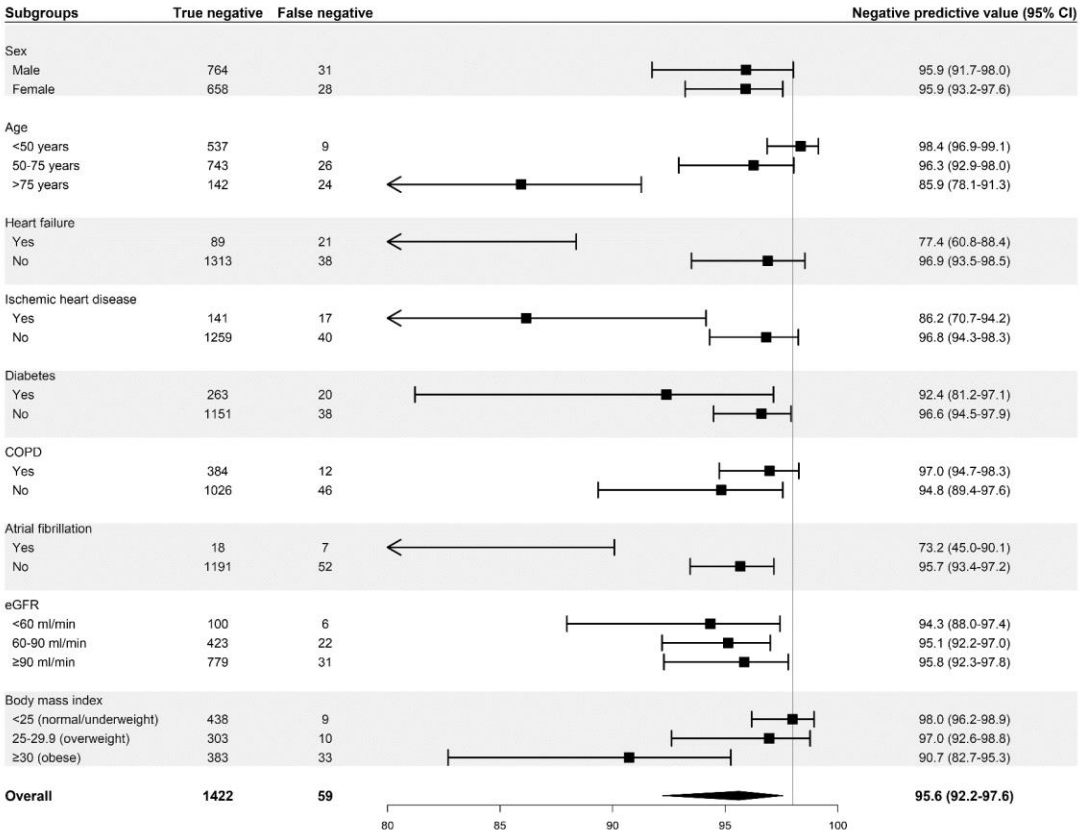
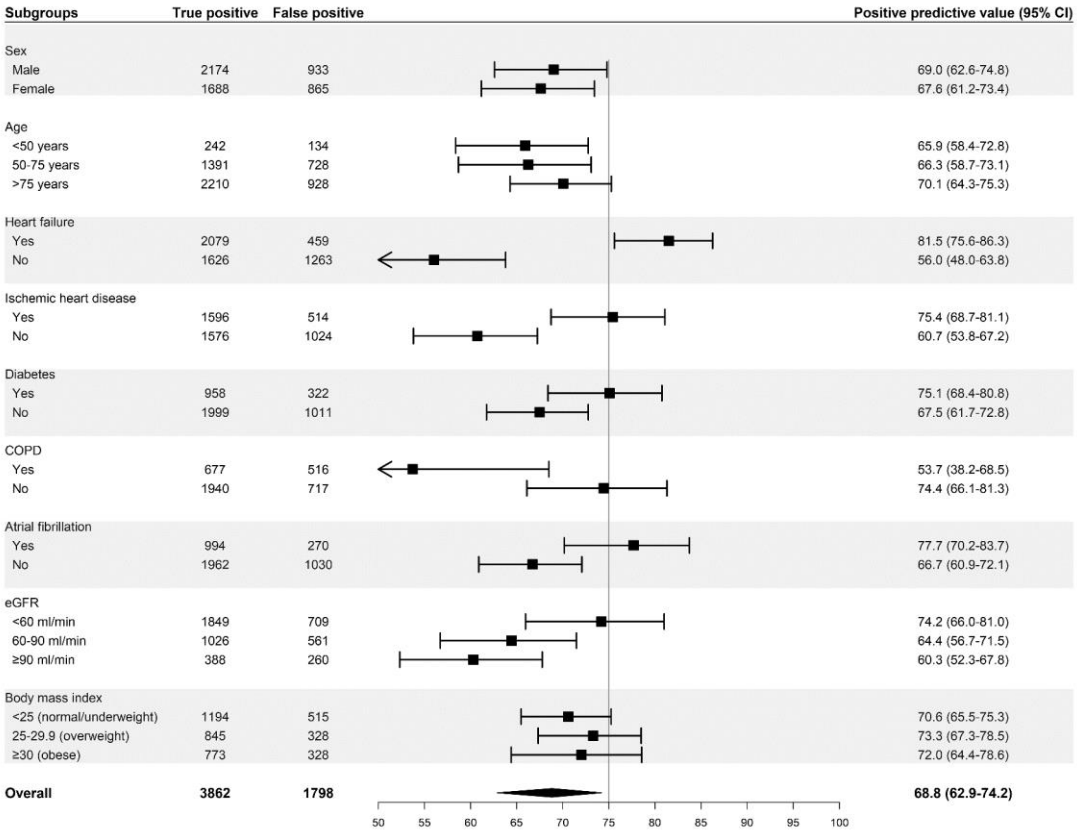


Figure 7.3 Negative predictive value of guideline-recommended BNP and MR-proANP thresholds across patient subgroups.
(a) Negative predictive value of the BNP threshold of 100 pg/mL across patient subgroups.
(b) Negative predictive value of the MR-proANP threshold of 120 pg/mL across patient subgroups.
Abbreviations: COPD= chronic obstructive pulmonary disease; eGFR= estimated glomerular filtration rate.

a)


Subgroups
True positive
False positive
Positive predictive value (95% CI)
Sex
Male 2174 933 69.0 (62.6-74.8)
Female 1688 865 67.6 (61.2-73.4)
Age
<50 years 242 134 65.9 (58.4-72.8)
50-75 years 1391 728 66.3 (58.7-73.1)
>75 years 2210 928 70.1 (64.3-75.3)
Heart failure
Yes 2079 459 81.5 (75.6-86.3)
No 1626 1263 56.0 (48.0-63.8)
Ischemic heart disease
Yes 1596 514 75.4 (68.7-81.1)
No 1576 1024 60.7 (53.8-67.2)
Diabetes
Yes 958 322 75.1 (68.4-80.8)
No 1999 1011 67.5 (61.7-72.8)
COPD
Yes 677 516 53.7 (38.2-68.5)
No 1940 717 74.4 (66.1-81.3)
Atrial fibrillation
Yes 994 270 77.7 (70.2-83.7)
No 1962 1030 66.7 (60.9-72.1)
eGFR
<60 ml/min 1849 709 74.2 (66.0-81.0)
60-90 ml/min 1026 561 64.4 (56.7-71.5)
≥90 ml/min 388 260 60.3 (52.3-67.8)
Body mass index
<25 (normal/underweight) 1194 515 70.6 (65.5-75.3)
25-29.9 (overweight) 845 328 73.3 (67.3-78.5)
≥30 (obese) 773 328 72.0 (64.4-78.6)
Overall 3862 1798 68.8 (62.9-74.2)
50 55 60 65 70 75 80 85 90 95 100

b)

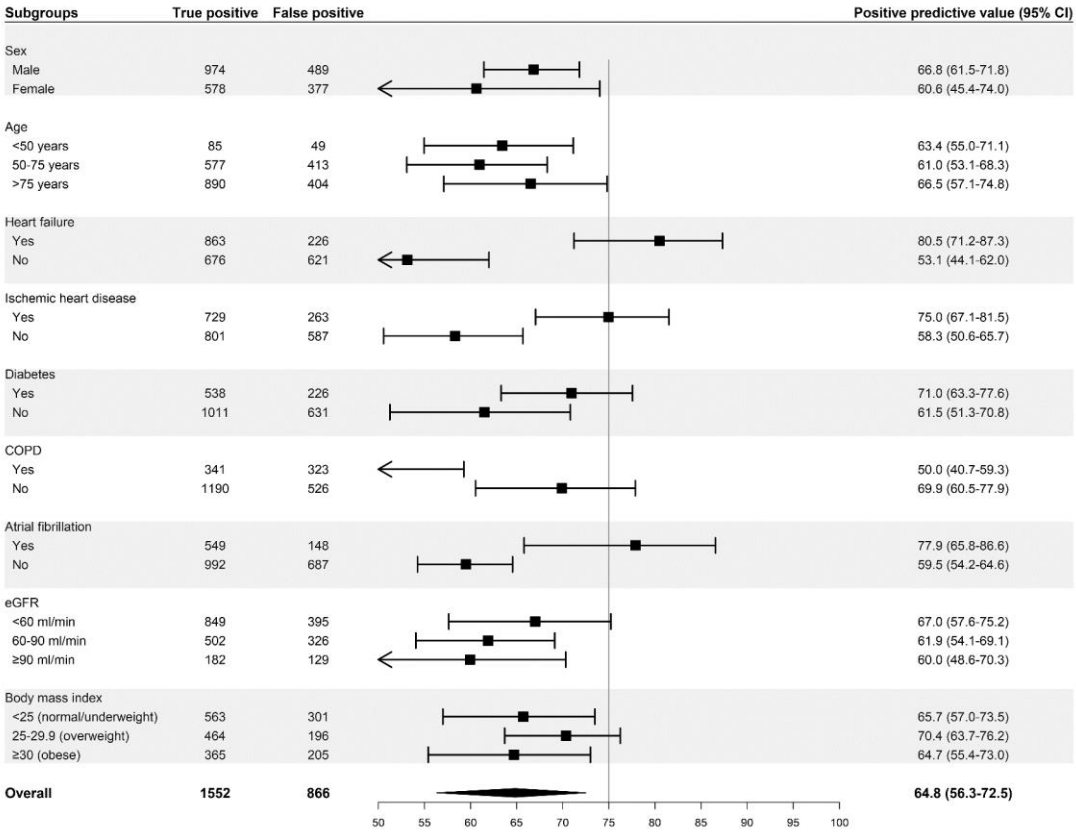


Figure 7.4 Positive predictive value of guideline-recommended BNP and MR-proANP thresholds across patient subgroups.
(a) Positive predictive value of the BNP threshold of 100 pg/mL across patient subgroups.
(b) Positive e predictive value of the MR-proANP threshold of 120 pg/mL across patient subgroups.
Abbreviations: COPD= chronic obstructive pulmonary disease; eGFR= estimated glomerular filtration rate.

7.4.3 Guideline-recommended MR-proANP threshold

Pooled meta-estimates of NPV, sensitivity, PPV and specificity of the guideline-recommended MR-proANP threshold of 120 pg/mL were 95.6% (95% confidence interval, 92.2-97.6), 96.3% (95.3-97.2%), 64.8% (56.3-72.5%), and 63.5% (54.4-71.7%), respectively (Figure 7.2 and Table 7.6). Overall, 38% of patients had MR-proANP concentrations below 120 pg/mL. Similar to BNP, there was marked heterogeneity in the performance of these thresholds across patient subgroups (Figure 7.3). NPV was lower in those with prior heart failure (77.4% [60.8-88.4%]), atrial fibrillation (73.2% [45.0-90.1%]) and ischemic heart disease (86.4% [70.7-88.4%]). MR-proANP threshold of 80 pg/mL achieved our pre-specified optimal rule-out criteria (NPV of 98% and sensitivity of 90%). However, the performance of this threshold remains heterogenous across patient subgroups (Figure 7.5). PPVs of the MR-proANP threshold of 120 pg/mL were also heterogenous with lower performance in those without prior heart failure (53.1% [44.1-62.0%]), COPD (50.0% [40.7-59.3%]) and without atrial fibrillation (59.5% [54.2-64.6%]). (Figure 7.4).

7.5 **Table 7.6** Diagnostic performance of MR-proANP and CoDE-HF thresholds for acute heart failure.

A) Rule-out thresholds and CoDE-HF scores.

	Threshold or score	True positive	False positive	True negative	False negative	NPV (95% CI)	Sensitivity (95% CI)	Proportion ruled out
All patients								
MR-proANP	120 pg/mL	1552	866	1422	59	96.3 (95.3-97.2)	95.6 (92.2-97.6)	38.0%
Patients without prior heart failure								
CoDE-HF (internal validation)	3.5	612	619	885	13	98.6 (97.5-99.2)	97.9 (96.5-98.8)	42.2%
CoDE-HF (external validation)	3.5	86	92	338	3	99.1 (98.8-99.4)	96.6 (96.0-97.2)	65.7%

B) Rule-in thresholds and CoDE-HF scores.

	Threshold or score	True positive	False positive	True negative	False negative	PPV (95% CI)	Specificity (95% CI)	Proportion ruled in
All patients								
MR-proANP	120 pg/mL	1552	866	1422	59	64.8 (56.3-72.5)	63.5 (54.4-71.7)	62.0%
Patients without prior heart failure								
CoDE-HF (internal validation)	52.7	463	131	1373	162	77.5 (72.6-81.7)	90.0 (84.1-93.9)	27.9%
CoDE-HF (external validation)	52.7	63	27	403	26	70.0 (68.5-71.4)	93.7 (92.9-94.4)	17.3%
Patients with prior heart failure								
CoDE-HF (internal validation)	88.5	420	19	267	405	95.7 (93.3-97.2)	90.4 (73.6-96.9)	39.5%
CoDE-HF (external validation)	88.5	38	11	18	21	77.6 (76.2-78.8)	62.1 (60.5-63.6)	55.7%

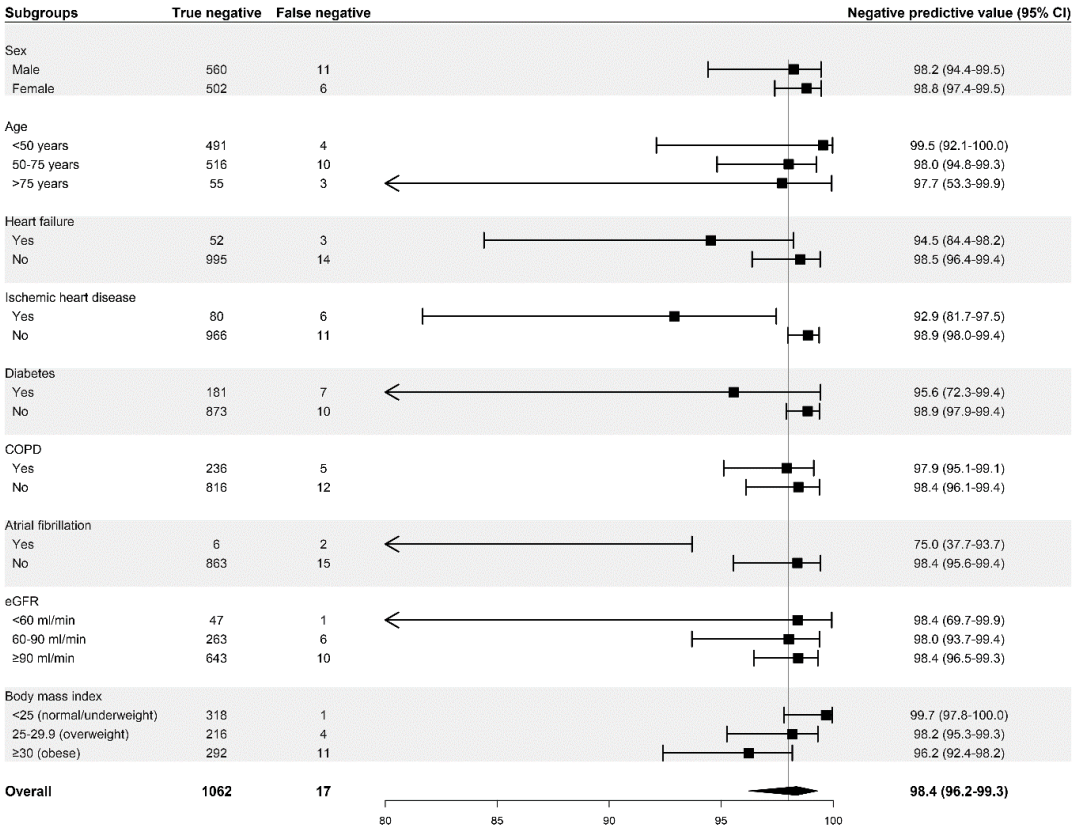


Figure 7.5 Negative predictive value of optimised MR-proANP threshold of 80 pg/mL across patient subgroups.

7.5.1 The CoDE-HF score

In the training cohort, CoDE-HF with BNP had an area under curve [AUC] of 0.914 [0.902-0.926] and Brier score of 0.101 in patients without prior heart failure and an AUC of 0.867 [0.842-0.893] and Brier score of 0.142 in those with prior heart failure (Figure 7.6). CoDE-HF with MR-proANP achieved an AUC 0.927 [0.916-0.938] and Brier score of 0.099 in patients without prior heart failure, and AUC 0.867 [0.842-0.893] and Brier score of 0.117 in patients with prior heart failure (Figure 7.7).

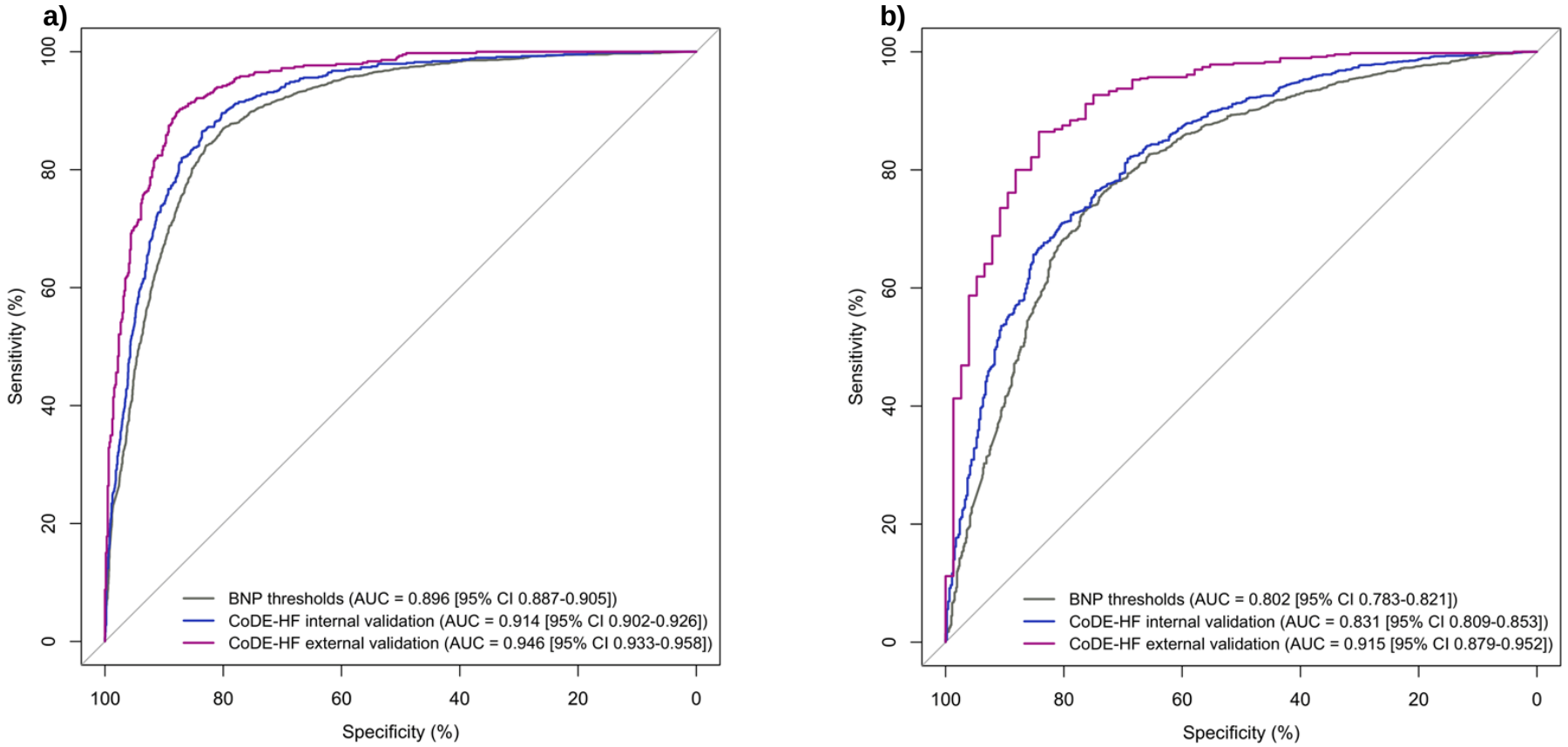


Figure 7.6 Discrimination of the guideline-recommended BNP thresholds and CoDE-HF score.
a) Receiver operator curve in patients without prior heart failure (area under curve of 0.896 [0.887-0.905], 0.914 [0.902-0.926] and 0.946 [0.933-0.958] for guideline-recommended BNP thresholds and the CoDE-HF score in the training cohort and external validation cohort respectively).
b) Receiver operator curve in patients with prior heart failure (area under curve of 0.802 [0.783-0.821], 0.831 [0.809-0.853] and 0.915 [0.879-0.952] for guideline-recommended BNP thresholds and the CoDE-HF score in the training cohort and external validation cohort respectively).

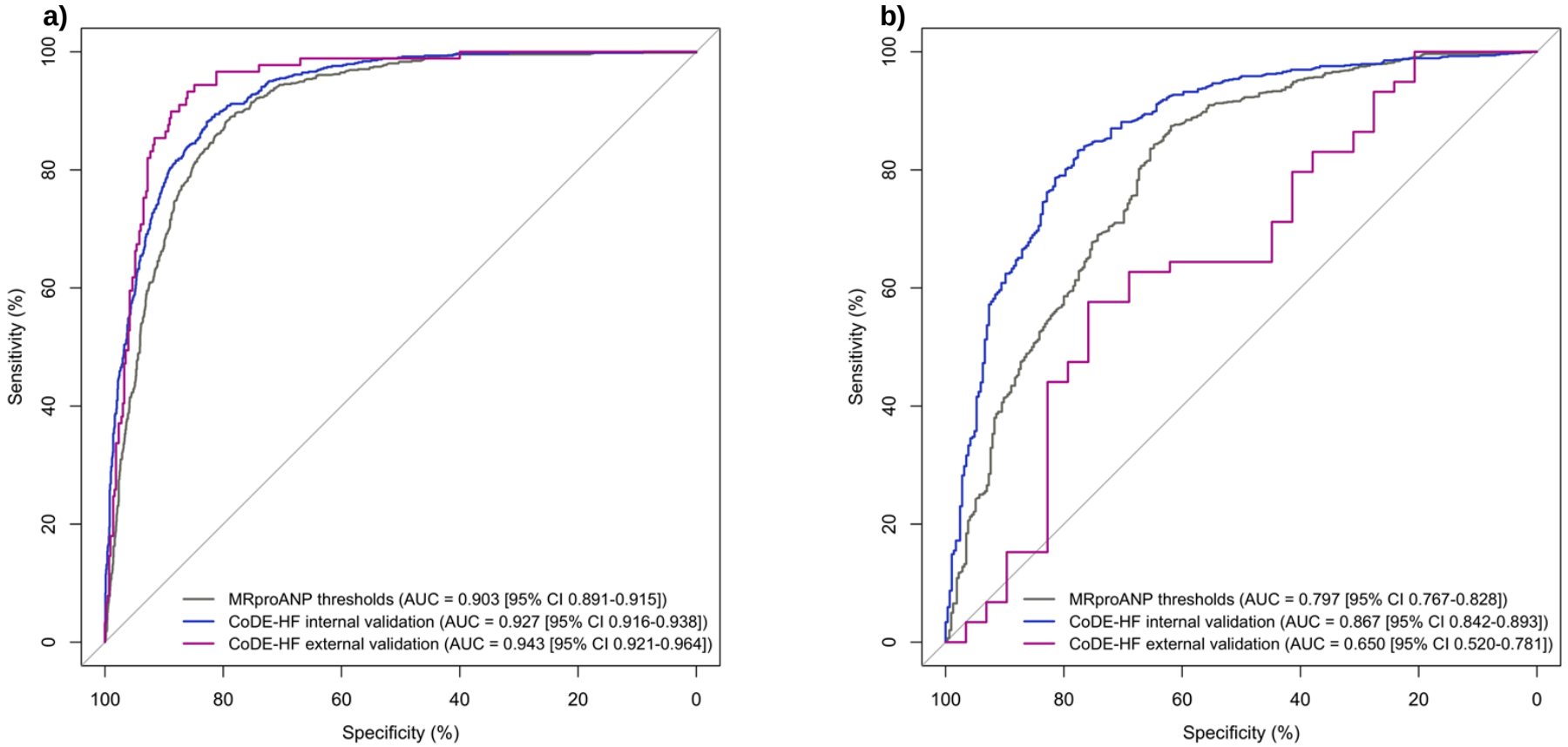


Figure 7.7 Discrimination of the guideline-recommended MR-proANP thresholds and CoDE-HF score.
a) Receiver operator curve in patients without prior heart failure (area under curve of 0.903 [0.891-0.915], 0.927 [0.916-0.938] and 0.943 [0.921-0.964] for guideline-recommended MR-proANP thresholds and the CoDE-HF score in the training cohort and external validation cohort respectively).
b) Receiver operator curve in patients with prior heart failure (area under curve of 0.797 [0.767-0.828], 0.867 [0.842-0.893] and 0.650 [0.520-0.781] for guideline-recommended MR-proANP thresholds and the CoDE-HF score in the training cohort and external validation cohort respectively).

In the external validation cohort, CoDE-HF with BNP had excellent discrimination and calibration in those without (AUC 0.946 [0.933-0.958] and Brier score 0.105) and with prior heart failure (AUC 0.915 [0.879-0.952] and Brier Score 0.082) (Figure 7.6 and Figure 7.8). For BNP, a CoDE-HF score of 3.6 achieved an NPV of 99.1% (98.8-99.3%) and sensitivity of 99.3% (99.1-99.5%) (Figure 7.9 and Table 7.5), whilst a score of 53.2 achieved a PPV of 91.3% (90.7-91.9%) and a specificity of 95.6% (95.2-96.0%) in those without prior heart failure (Figure 7.10 and Table 7.5). These rule-in and rule-out scores had a more consistent diagnostic performance across all subgroups compared to BNP thresholds alone. If these scores were applied in patients with suspected acute heart failure without prior heart failure, CoDE-HF with BNP would identify 30.0% at low-probability and 30.0% at high-probability of acute heart failure, respectively. In patients with prior heart failure, there was no score which achieved our target rule-out criteria in the training cohort. A CoDE-HF score of 92.0 achieved a PPV of 98.8% (98.6-99.0%) and specificity of 96.1% (95.6-96.5%) (Figure 7.11).

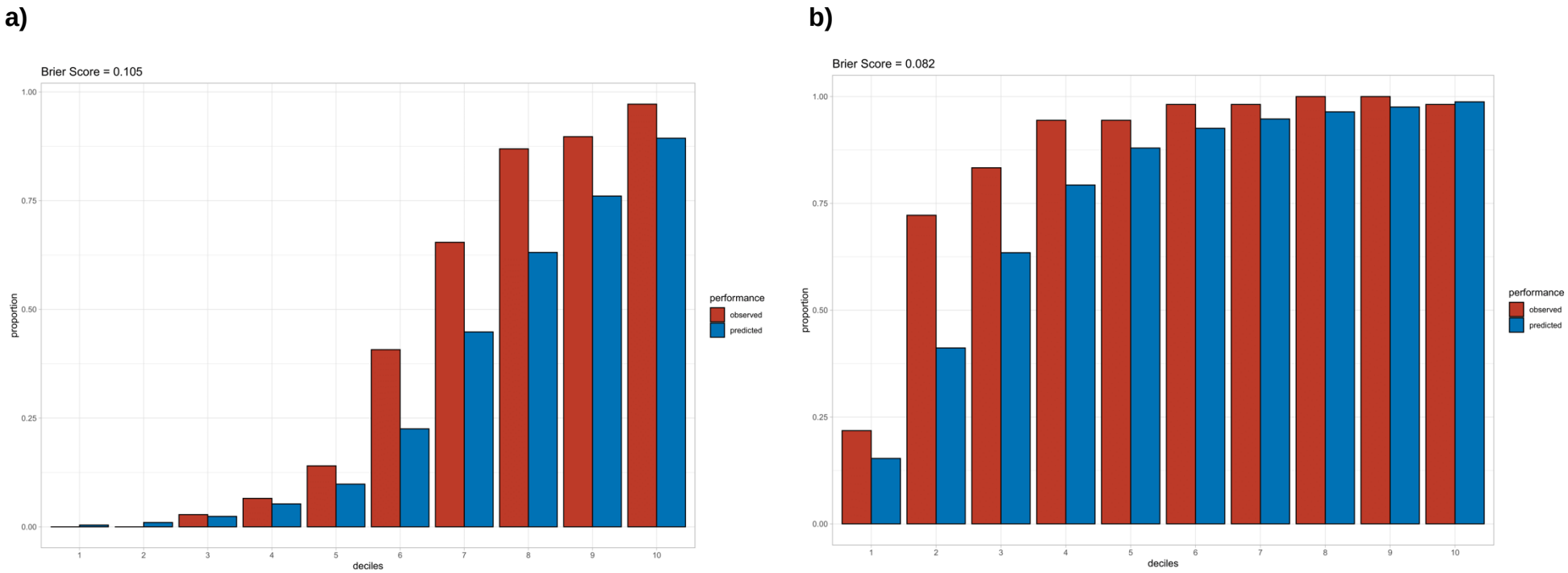
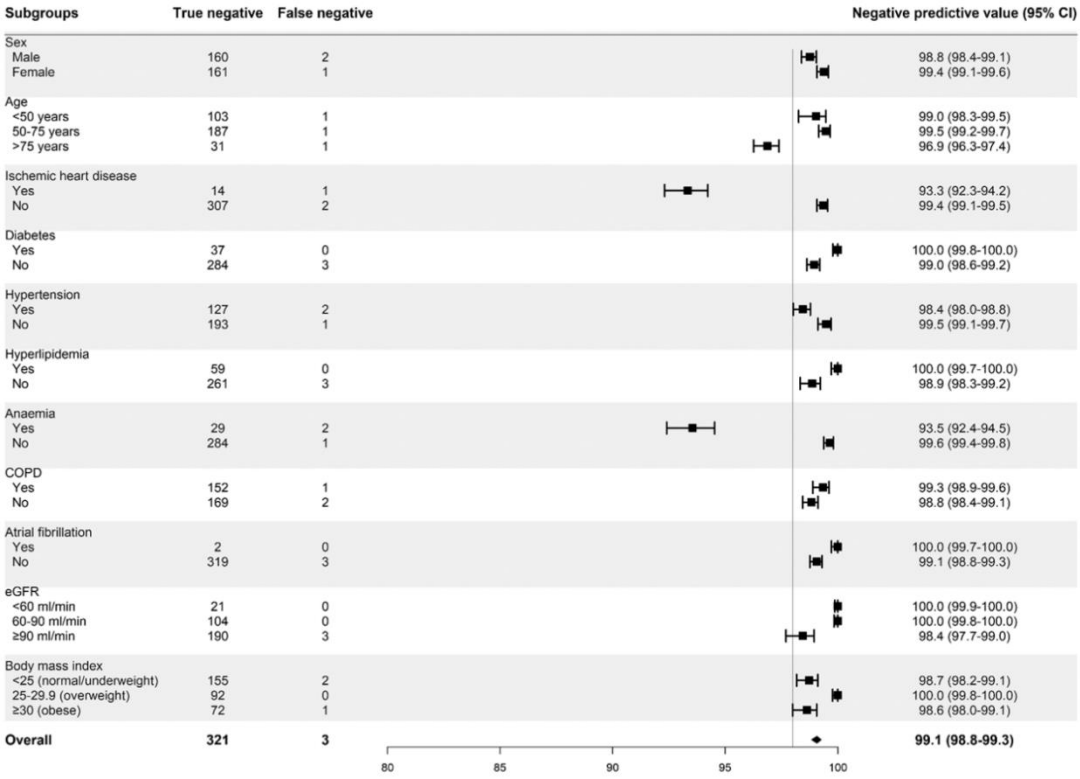


Figure 7.8 Calibration plot of CoDE-HF with BNP in patients with (a) no previous heart failure and (b) previous heart failure in the external validation cohort.

a)



Subgroups	True negative	False negative	Negative predictive value (95% CI)
Sex			
Male	160	2	98.8 (98.4-99.1)
Female	161	1	99.4 (99.1-99.6)
Age			
<50 years	103	1	99.0 (98.3-99.5)
50-75 years	187	1	99.5 (99.2-99.7)
>75 years	31	1	96.9 (96.3-97.4)
Ischemic heart disease			
Yes	14	1	93.3 (92.3-94.2)
No	307	2	99.4 (99.1-99.5)
Diabetes			
Yes	37	0	100.0 (99.8-100.0)
No	284	3	99.0 (98.6-99.2)
Hypertension			
Yes	127	2	98.4 (98.0-98.8)
No	193	1	99.5 (99.1-99.7)
Hyperlipidemia			
Yes	59	0	100.0 (99.7-100.0)
No	261	3	98.9 (98.3-99.2)
Anaemia			
Yes	29	2	93.5 (92.4-94.5)
No	284	1	99.6 (99.4-99.8)
COPD			
Yes	152	1	99.3 (98.9-99.6)
No	169	2	98.8 (98.4-99.1)
Atrial fibrillation			
Yes	2	0	100.0 (99.7-100.0)
No	319	3	99.1 (98.8-99.3)
eGFR			
<60 ml/min	21	0	100.0 (99.9-100.0)
60-90 ml/min	104	0	100.0 (99.8-100.0)
≥90 ml/min	190	3	98.4 (97.7-99.0)
Body mass index			
<25 (normal/underweight)	155	2	98.7 (98.2-99.1)
25-29.9 (overweight)	92	0	100.0 (99.8-100.0)
≥30 (obese)	72	1	98.6 (98.0-99.1)
Overall	**321**	**3**	**99.1 (98.8-99.3)**

b)

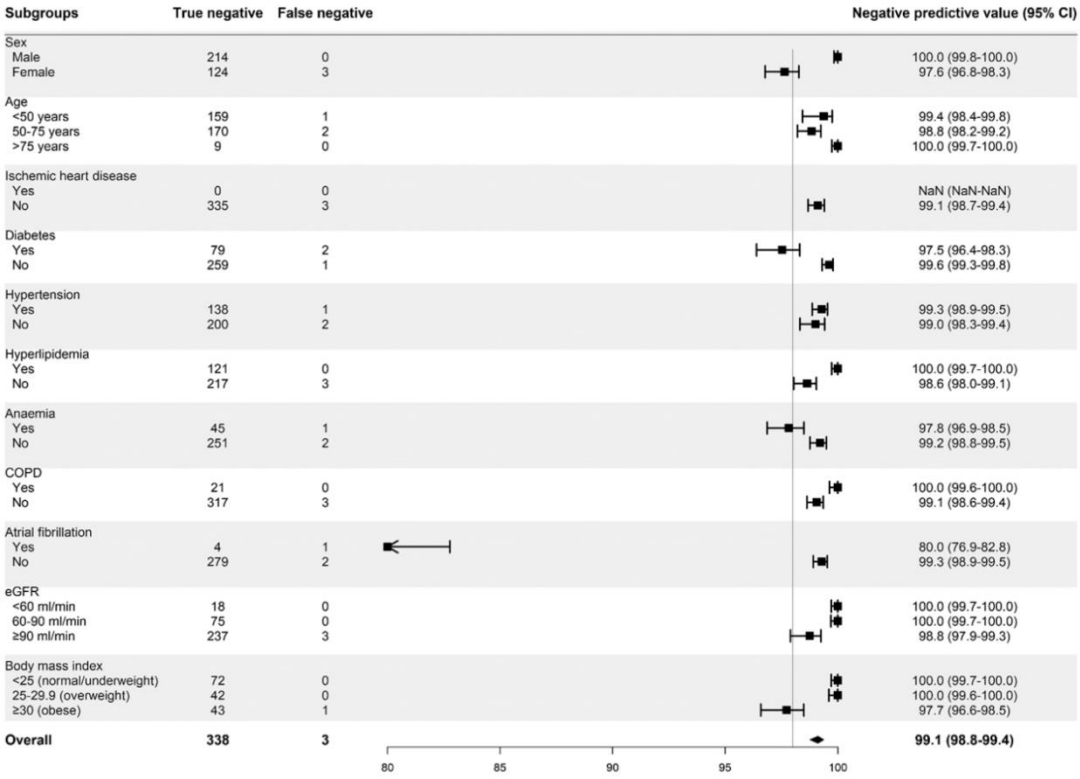


Figure 7.9 Diagnostic performance of the CoDE-HF score across patient subgroups in the external validation cohort.
(a) Negative predictive value of the CoDE-HF with BNP rule-out score of 3.6 in patients without prior heart failure across patient subgroups.
(b) Negative predictive value of the CoDE-HF with MR-proANP rule-out score of 3.5 in patients without prior heart failure across patient subgroups.

a)

Subgroups	True positive	False positive	Positive predictive value (95% CI)
Sex			
Male	131	17	88.5 (87.5-89.4)
Female	142	12	92.2 (91.3-93.0)
Age			
<50 years	5	0	100.0 (99.6-100.0)
50-75 years	67	5	93.1 (92.1-93.9)
>75 years	201	24	89.3 (88.3-90.3)
Ischemic heart disease			
Yes	102	8	92.7 (91.7-93.7)
No	170	21	89.0 (88.1-89.9)
Diabetes			
Yes	63	6	91.3 (89.9-92.5)
No	210	23	90.1 (89.3-90.9)
Hypertension			
Yes	212	21	91.0 (90.1-91.8)
No	60	8	88.2 (86.9-89.4)
Hyperlipidemia			
Yes	107	13	89.2 (87.4-90.7)
No	164	16	91.1 (89.9-92.2)
Anaemia			
Yes	102	11	90.3 (88.9-91.5)
No	167	18	90.3 (89.2-91.3)
COPD			
Yes	47	6	88.7 (87.2-90.0)
No	225	23	90.7 (89.8-91.6)
Atrial fibrillation			
Yes	102	6	94.4 (93.1-95.5)
No	171	23	88.1 (87.2-89.0)
eGFR			
<60 ml/min	139	17	89.1 (87.9-90.2)
60-90 ml/min	108	10	91.5 (90.3-92.6)
≥90 ml/min	23	2	92.0 (90.5-93.2)
Body mass index			
<25 (normal/underweight)	132	12	91.7 (90.5-92.7)
25-29.9 (overweight)	79	6	92.9 (91.6-94.1)
≥30 (obese)	57	10	85.1 (83.4-86.6)
Overall	**273**	**29**	**90.4 (89.7-91.0)**

b)

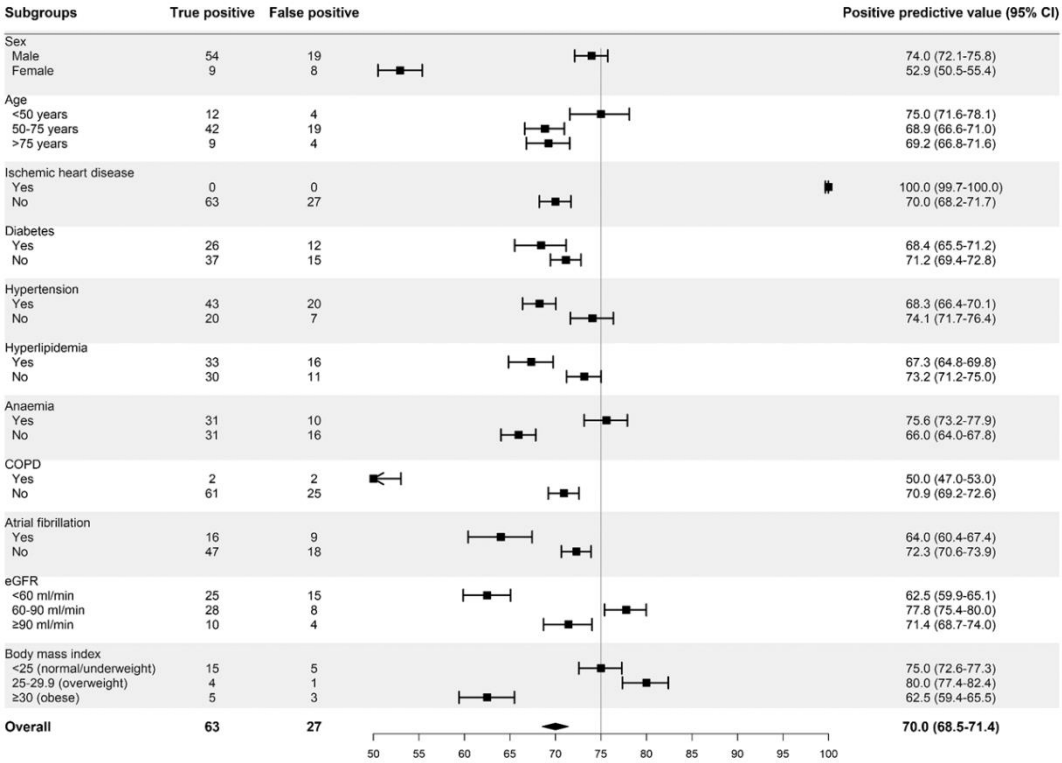
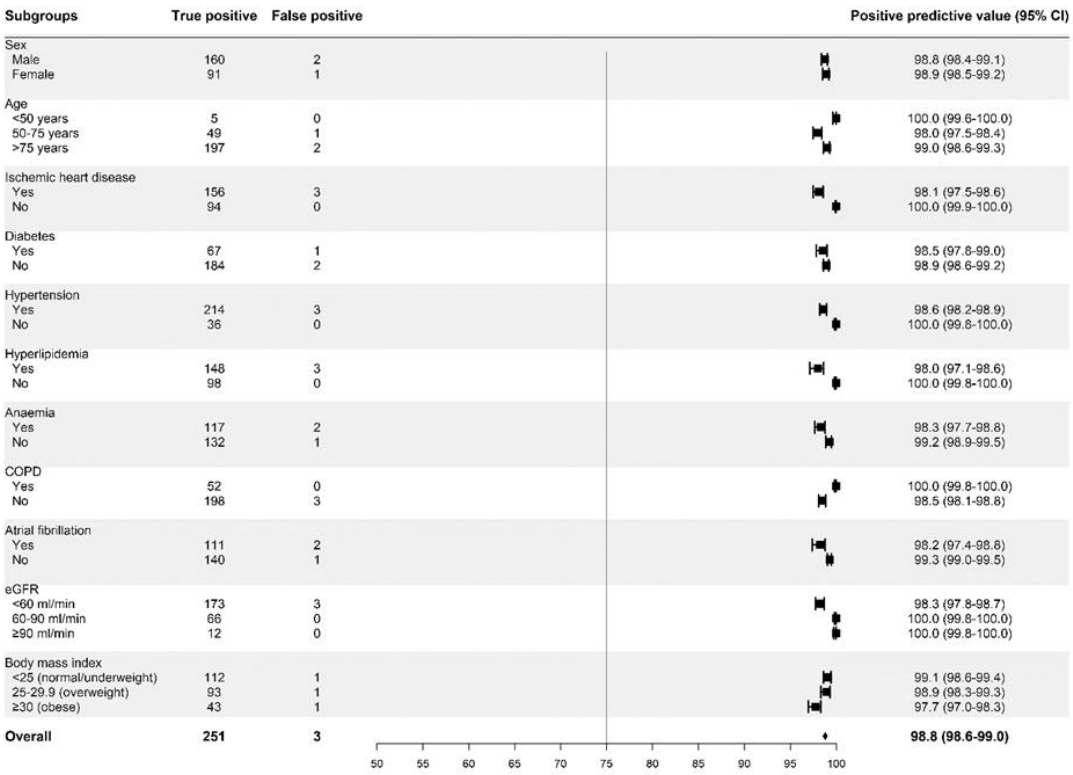


Figure 7.10 Diagnostic performance of the CoDE-HF score across patient subgroups in the external validation cohort.
a) Positive predictive value of the CoDE-HF with BNP rule-in score of 53.2 in patients without prior heart failure across patient subgroups.
b) Positive predictive value of the CoDE-HF with MR-proANP rule-in score of 52.7 in patients without prior heart failure across patient subgroups.

a)



Subgroups	True positive	False positive	Positive predictive value (95% CI)
Sex			
Male	160	2	98.8 (98.4-99.1)
Female	91	1	98.9 (98.5-99.2)
Age			
<50 years	5	0	100.0 (99.6-100.0)
50-75 years	49	1	98.0 (97.5-98.4)
>75 years	197	2	99.0 (98.6-99.3)
Ischemic heart disease			
Yes	156	3	98.1 (97.5-98.6)
No	94	0	100.0 (99.9-100.0)
Diabetes			
Yes	67	1	98.5 (97.8-99.0)
No	184	2	98.9 (98.6-99.2)
Hypertension			
Yes	214	3	98.6 (98.2-98.9)
No	36	0	100.0 (99.8-100.0)
Hyperlipidemia			
Yes	148	3	98.0 (97.1-98.6)
No	98	0	100.0 (99.8-100.0)
Anaemia			
Yes	117	2	98.3 (97.7-98.8)
No	132	1	99.2 (98.9-99.5)
COPD			
Yes	52	0	100.0 (99.8-100.0)
No	198	3	98.5 (98.1-98.8)
Atrial fibrillation			
Yes	111	2	98.2 (97.4-98.8)
No	140	1	99.3 (99.0-99.5)
eGFR			
<60 ml/min	173	3	98.3 (97.8-98.7)
60-90 ml/min	66	0	100.0 (99.8-100.0)
≥90 ml/min	12	0	100.0 (99.8-100.0)
Body mass index			
<25 (normal/underweight)	112	1	99.1 (98.6-99.4)
25-29.9 (overweight)	93	1	98.9 (98.3-99.3)
≥30 (obese)	43	1	97.7 (97.0-98.3)
Overall	**251**	**3**	**98.8 (98.6-99.0)**

b)

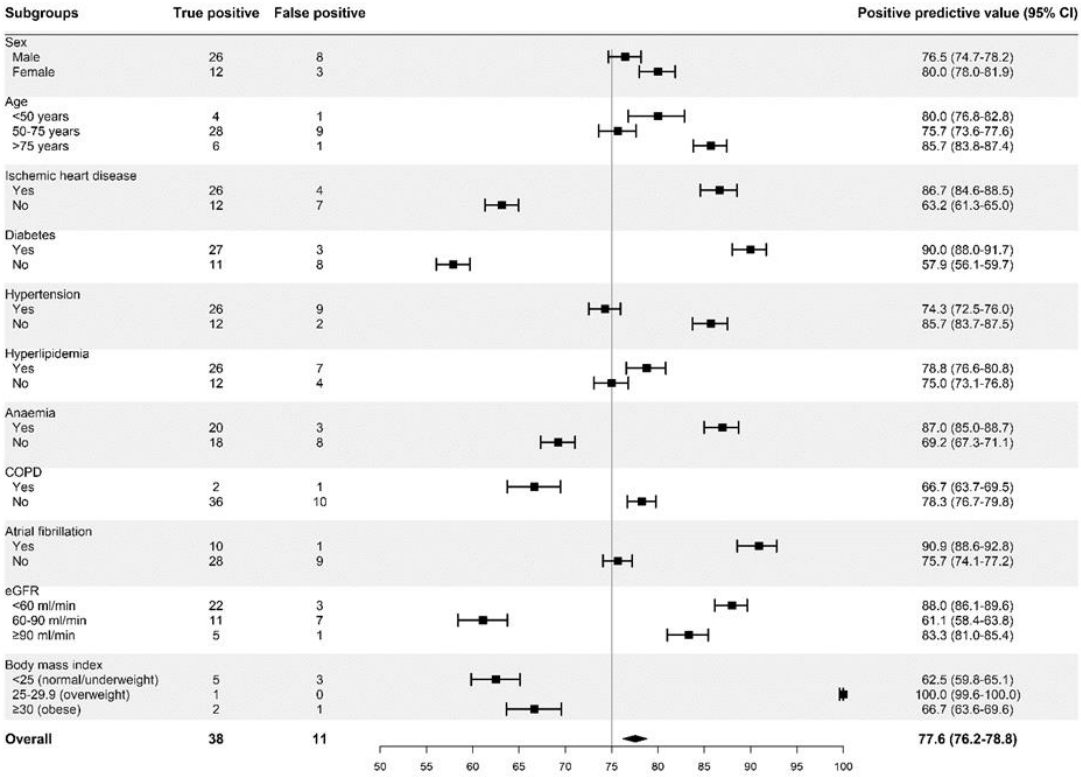


Figure 7.11 Diagnostic performance of the CoDE-HF score across patient subgroups in the external validation cohort.
a) Positive predictive value of the CoDE-HF with BNP rule-in score of 92.0 in patients with prior heart failure across patient subgroups.
b) Positive predictive value of the CoDE-HF with MR-proANP rule-in score of 88.5 in patients without prior heart failure across patient subgroups.

With MR-proANP, CoDE-HF also had good discrimination and calibration in the external validation cohort (AUC of 0.943 [0.921-0.964] and Brier score of 0.073 in patients without prior heart failure and an AUC of 0.650 [0.520-0.781] and Brier score of 0.232 in those with prior heart failure) (Figure 7.7 and Figure 7.12). For MR-proANP, a CoDE-HF score of 3.5 achieved an NPV of 99.1% (98.8-99.4%) and sensitivity of 96.6% (96.0-98.8%) (Figure 7.9 and Table 7.6), whilst a score of 52.7 achieved a PPV of 70.0% (68.5-71.4%) and a specificity of 93.7% (92.9-94.4%) in those without prior heart failure (Figure 7.10 and Table 7.6). Similarly, these rule-in and rule-out scores had more consistent performance across patient subgroups than the biomarker thresholds alone. If these scores were applied in patients with suspected acute heart failure without prior heart failure, CoDE-HF with MR-proANP would identify 65.7% at low-probability and 17.3% at high-probability of acute heart failure. In patients with prior heart failure, a CoDE-HF score of 88.5 achieved a PPV of 77.6 (76.2-78.8) and specificity of 62.1 (60.5-63.6) (Figure 7.11).

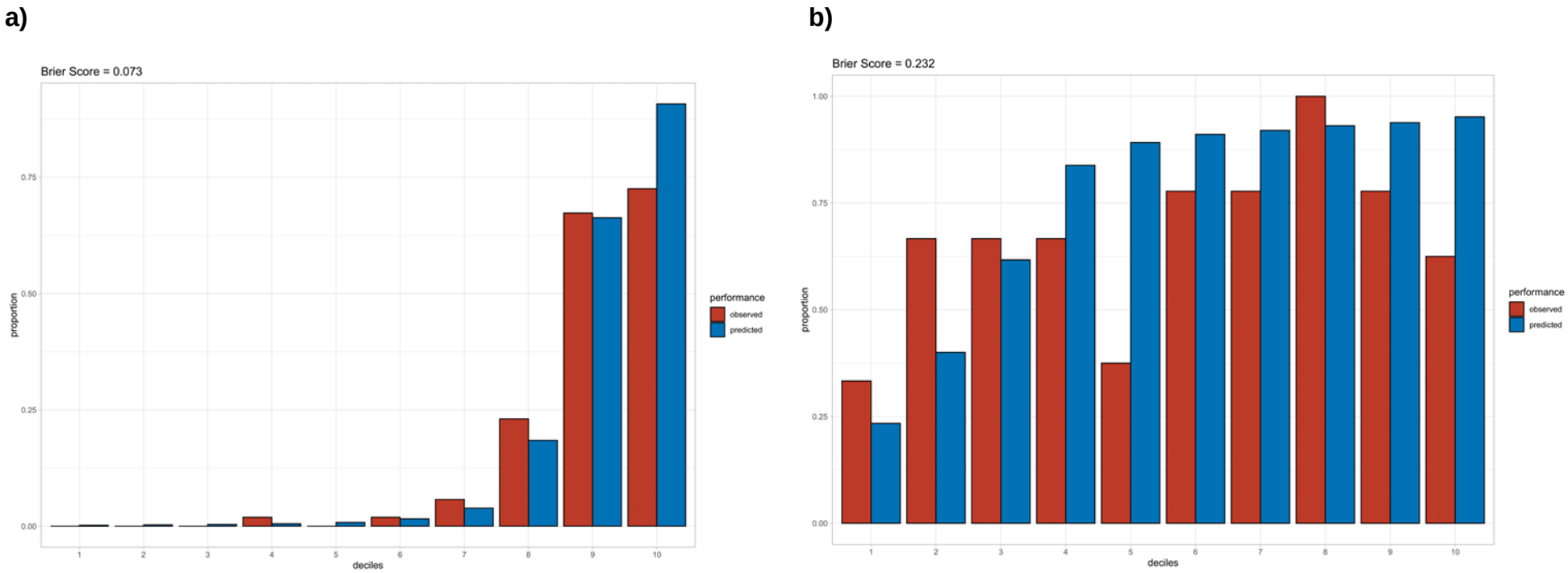


Figure 7.12 Calibration plot of CoDE-HF with MR-proANP in patients with (a) no previous heart failure and (b) previous heart failure in the external validation cohort.

Patients in the external validation cohort who were identified as low-probability by CoDE-HF had a substantially lower rate of all-cause at 30-days and 1-year compared to those who were identified as intermediate and high-probability for both BNP (30 day all-cause mortality: 1.0% *versus* 7.0% and 12.9%; 1 year all-cause mortality: 8.4% v*ersus* 27.7% and 43.4% respectively) and MR-proANP (30 day all-cause mortality: 0.3% *versus* 3.1% and 3.6%; 1 year all-cause mortality: 1.5% v*ersus* 9.4% and 13.7% respectively) (Figure 7.13).

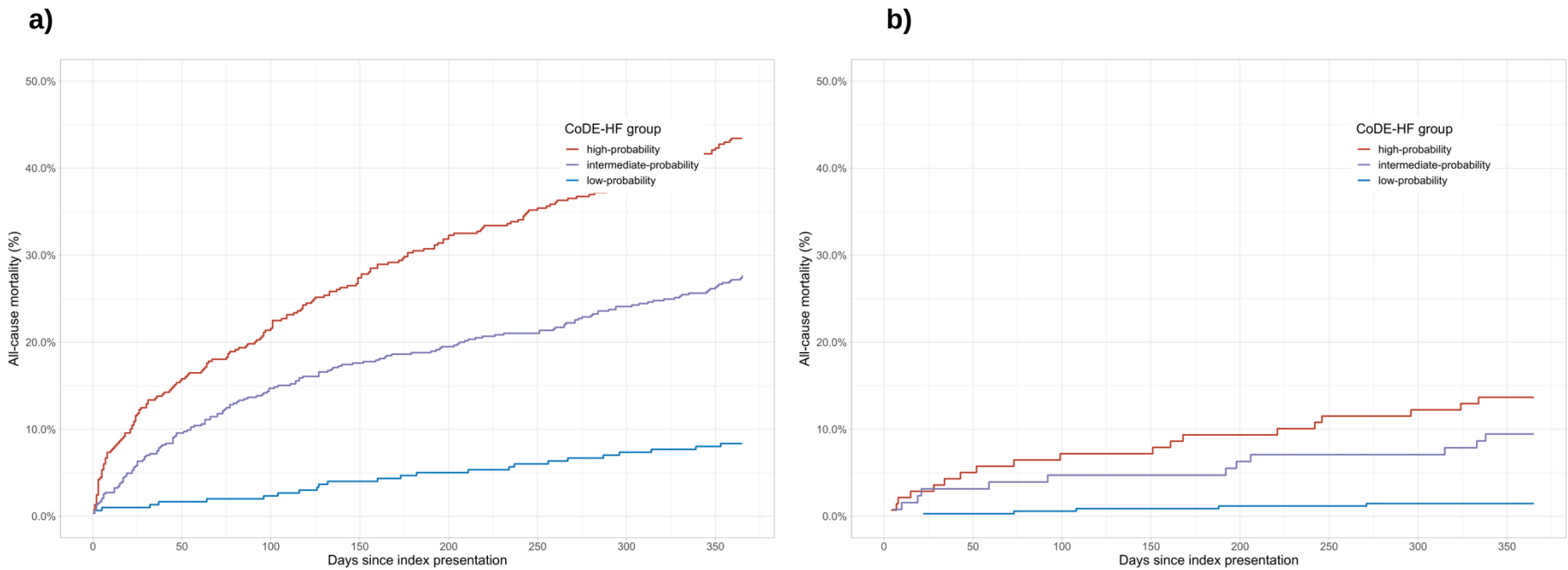


Figure 7.13 Cumulative incidence of all-cause mortality in the external validation cohort stratified by CoDE-HF probability group.
(a) Using the BNP.
(b) Using the MR-proANP.

7.6 Discussion

We performed an individual patient-level meta-analysis to evaluate the diagnostic performance of guideline-recommended thresholds of BNP and MR-proANP in over 8,493 and 3,847 number of patients, respectively, across 14 studies, and subsequently developed and validated a decision-support tool that uses these natriuretic peptides as a continuous variable with patient factors for the diagnosis of acute heart failure. We report several important findings. First, the guideline-recommended thresholds of BNP and MR-proANP to rule out acute heart failure had heterogenous performance across important patient subgroups. NPV was substantially lower in those with prior heart failure, atrial fibrillation and ischemic heart disease where false negative rates were as high as one in five. Second, there was no threshold at which BNP achieved an NPV of 98%. For MR-proANP, an optimised threshold of 80 pg/mL achieved an NPV of 98% however performance remained heterogenous across patient subgroups. Finally, we have developed and validated the CoDE-HF decision-support tool for BNP and MR-proANP using machine learning to combine these natriuretic peptides with simple and objective patient factors to calculate an individualised probability of acute heart failure. We found that CoDE-HF had a more consistent performance across patient subgroups compared to BNP or MR-proANP thresholds alone.

7.6.1 Strengths of this study

To our knowledge, this is the largest study evaluating the diagnostic performance of BNP and MR-proANP for acute heart failure to date. All studies included in this analysis prospectively evaluated the performance of these natriuretic peptides and

performed diagnostic adjudication of the diagnosis of acute heart failure. The availability of individual patient-level data allowed us to evaluate the performance of guideline-recommended thresholds across important patient subgroups. Furthermore, this enabled the evaluation of these natriuretic peptides across a range of alternative thresholds and the development of a decision support tool using machine learning.

7.6.2 CoDE-HF performance

We have previously developed the CoDE-HF decision-support tool using NT-proBNP (Chapter 6). We have now further developed CoDE-HF for BNP and MR-proANP and demonstrated that the use of machine learning substantially improves the diagnostic performance of all three natriuretic peptides. This is intuitive given that all natriuretic peptides share a similar mechanism of release form the myocardium in response to myocardial pressure and volume overload, and are similarly influenced by patient factors such as age, renal function and obesity.[92-94,102,201-204] This is particularly important given the increasing prevalence of heart failure in ageing populations with increasing number of comorbidities.[169-171] The availability of a simple decision-support tool that incorporates routinely collected clinical variables to aid in the interpretation of these biomarkers could significantly improve the efficiency and accuracy of the assessment of patients in busy Emergency Departments. This has major potential to improve patient outcomes by expediating treatment, specialist cardiology referrals and investigations such as echocardiography in patients at high probability for the diagnosis whilst those at low probability could potentially be

discharged from hospital or investigated for other differential diagnoses more promptly.

7.6.3 Limitations of other scores

We are aware of numerous validated prognostic risk scores for patients with an established diagnosis of heart failure.[150,172,173] However there are only a few that have been developed to aid in the diagnosis of acute heart failure.[175,176] Whilst these diagnostic scores have many strengths, they incorporates relatively more subjective variables such as clinicians' estimation of the pre-test probability, patient's description of symptoms and natriuretic peptides as a binary variable which does not take into account the dynamic and nonlinear interaction between this natriuretic peptides and other variables. These previous attempts at developing and validating diagnostic scores have also included a limited number of patients from a single healthcare setting which precluded the assessment of diagnostic performance within important patient subgroups and limits external generalizability.

7.6.4 Limitations of this study

We acknowledge several potential limitations in this study. First, acute heart failure is ultimately a clinical diagnosis and therefore it is likely that there is some inherent heterogeneity in the adjudication of this diagnosis across different studies. Second, the adjudicated diagnosis of acute heart failure did not differentiate between the different underlying aetiologies of heart failure or between heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved ejection fraction (HFpEF). Nevertheless, our decision-support tool was designed to aid in the initial triage of all

patients with suspected acute heart failure regardless of aetiology to identify those who are likely to benefit from early specialist investigation and treatment. Third, prevalence of acute heart failure varies significantly across studies and may have influenced the diagnostic performance of BNP, MR-proANP and CoDE-HF. Whilst diagnostic performance was similar the proportions identified as low probability differed for the evaluation of CoDE-HF with BNP and MR-proANP largely because of differences in the risk profile between cohorts measuring these biomarkers. This heterogeneity reflects the diverse range of settings and populations in which natriuretic peptides and the decision-support tool will be applied in clinical practice and strengthens the generalizability of our study findings. Finally, there is significant missingness in some of the studies included in this analysis. Where possible, we have performed multiple imputation to maximise the use of data in the development of our machine learning model.

7.7 Conclusions

In conclusion, we demonstrated that guideline-recommended thresholds of BNP and MR-proANP have heterogenous performance across important patient subgroups. We developed and externally validated the CoDE-HF decision-support tool for BNP and MR-proANP and demonstrated that CoDE-HF ruled-in and ruled-out acute heart failure more accurately than natriuretic peptide thresholds alone.

This chapter marks the end of the second half of this thesis. The developed algorithm to assist with the diagnosis of acute heart failure works for all the three different biomarkers. In the next and final chapter, we will draw the conclusions derived for this thesis and also discuss the next steps forward.

CHAPTER 8

Conclusions and future directions

Chapter 8 Conclusions and future directions

8.1 Summary of thesis findings

Every year, 1.8 million people die from cardiovascular disease in Europe, and it affects the lives of 60 million people across the continent.[1] Cardiovascular disease remains the most common cause of death in Europe, despite the recent impact of the coronavirus pandemic. Furthermore, as our population ages, the prevalence of cardiovascular disease is predicted to climb.

The most common signs of acute cardiovascular disease are chest discomfort and shortness of breath. People with these symptoms account for one out of every four visits to the emergency room and one out of every ten unexpected hospital hospitalizations. This reflects the high burden of cardiovascular disease in our community and concerns from patients and clinicians alike who do not want to miss a potentially life-threatening cardiovascular condition, such as acute myocardial infarction or heart failure. Timely and accurate diagnosis has a substantial impact on patient outcomes. On the other hand, identifying those with life-threatening cardiovascular disease can be challenging because they can initially appear indistinguishable from other common conditions that present with similar symptoms.

The core aims of this thesis were first to develop and validate machine learning models that incorporate cardiac biomarkers and other objective

clinical variables to improve the diagnosis of acute cardiac conditions. These models were deployed as decision support tools that can be applied at the patient's bedside to aid in diagnosis and risk stratification in the Emergency Department. Secondly, it compared the diagnostic performance of the proposed machine learning models with the current national and international clinical practice guidelines fix threshold approach.

8.1.1 *Validation of the myocardial-ischemic-injury-index (MI^3) machine learning algorithm to guide the diagnosis of myocardial infarction in a heterogeneous population*

The myocardial-ischemic-injury-index (MI^3) is a machine learning algorithm that predicts the likelihood of myocardial infarction in patients with suspected acute coronary syndrome. This was the largest study evaluating the diagnostic performance of a machine learning algorithm for the diagnosis of myocardial infarction in 20,761 patients, and the first to be performed in a consecutive patient population that reflects clinical practice. The MI^3 algorithm had excellent overall discrimination. Our subgroup analysis observed no heterogeneity for the low-probability threshold, while the performance was heterogeneous across subgroups for the high-probability threshold. Moreover, and for the first time, it was reported that patients identified as high-probability by the algorithm of myocardial infarction on the index visit also had a ten-times higher rate of subsequent myocardial infarction or cardiovascular death at one year compared to those who were

classified as low-probability. These findings have potentially important implications for the use and interpretation of this algorithm in clinical practice. MI^3 could improve the diagnostic pathways for myocardial infarction by accurately identifying high-risk patients to be targeted for prompt individualised treatment, and by allowing early discharge in low-risk patients. However, calibration was poor in those at intermediate risk and opportunities to optimise the clinical application of this approach were identified.

8.1.2 Machine learning to optimise use of cardiac troponin in the diagnosis of acute myocardial infarction

The diagnosis of myocardial infarction can be challenging because many patients have elevated cardiac troponin concentrations due to other conditions. CoDE-ACS is the first machine learning algorithm for the diagnosis of myocardial infarction that allows the use of cardiac troponin concentrations at presentation or on serial troponin testing and incorporates important patient factors that influence cardiac troponin. CoDE-ACS had excellent discrimination and good calibration in the derivation and external validation cohorts and unlike MI^3 it performed consistently across important subgroups, such as older patients, females, early presenters and those with comorbidities. Patients identified as low-probability of myocardial infarction on the index visit had a very low-probability of dying from cardiac diseases or other causes at one year. These findings have potentially important implications for the use and interpretation of cardiac troponin in clinical

practice. Prospective validation of CoDE-ACS is needed to evaluate whether this decision-support tool improves diagnostic accuracy compared to current clinical pathways for the assessment of patients with suspected acute coronary syndrome.

8.1.3 Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study

The diagnosis of acute heart failure can be challenging because patients often present with non-specific symptoms. N-terminal pro-B-type natriuretic peptide (NT-proBNP) testing is recommended by most national and international guidelines to aid in the diagnosis of acute heart failure. However, despite guideline recommendations, NT-proBNP testing has not been universally implemented due to concerns about diagnostic performance in clinically important patient subgroups, such as older patients and those with renal disease or obesity. Using individual patient-level data meta-analysis from 14 studies from 13 countries, including randomised controlled trials and prospective observational studies, the guideline-recommended NT-proBNP thresholds for acute heart failure had relatively poor diagnostic performance in important patient subgroups. A newly developed and validated decision-support tool that uses machine learning to combine NT-pro-BNP as a continuous measure with clinical variables ruled-in and ruled-out acute heart failure more accurately than any approach using NT-proBNP

thresholds alone, while it performed consistently across all subgroups. Prospective validation of CoDE-HF is needed to evaluate whether this decision-support tool improves diagnostic accuracy compared to current clinical pathways for the assessment of patients with suspected acute heart failure.

8.1.4 Machine learning to optimise use of B-type natriuretic peptide and mid-regional pro-atrial natriuretic peptide in the diagnosis of acute heart failure

There are currently three natriuretic peptides recommended for the diagnosis of acute heart failure – NT-proBNP, B-type natriuretic peptide (BNP) and mid-regional pro-atrial natriuretic peptide (MR-proANP).[33] We previously demonstrated that guideline-recommended thresholds of NT-proBNP have relatively poor performance in older patients and those with obesity and prior heart failure (Chapter 6). Using data from 13 studies from 12 countries and similar methodology, the CoDE-HF algorithm was further developed to incorporate both the BNP and MR-proANP biomarkers, while its performance was compared with the guideline-recommended thresholds. BNP and MR-proANP concentrations were available in 8,493 and 3,847 patients. The diagnostic performance of guideline-recommended BNP and MR-proANP thresholds varied across patient subgroups, while CoDE-HF had more accurate and consistent performance overall and across subgroups.

8.2 Future directions

8.2.1 Development and validation of the CoDE-ACS decision support tool for high-sensitivity cardiac troponin T

The CoDE-ACS decision support tool was created in Chapter 5 of this thesis to improve the diagnosis of type 1 myocardial infarction using high-sensitivity troponin I. As a next step, we intend to develop further and validate CoDE-ACS to incorporate high-sensitivity cardiac troponin T to broaden the application of this decision-support tool. In August 2021, the Royal Infirmary of Edinburgh hospital changed from cardiac troponin I to cardiac troponin T. This will help me conduct a prospective cohort study of all consecutive patients presenting to the hospital with suspected acute coronary syndrome. The CoDE-ACS for high-sensitivity cardiac troponin T will be developed and internally validated using this dataset. After that, we will work with other researchers to have CoDE-ACS externally validated. We anticipated that the CoDE-ACS will have similar performance and will improve the diagnostic performance of clinical pathways that rely on cardiac troponin T thresholds only.

8.2.2 Prospective validation of the CoDE-HF decision support tool for the diagnosis of acute heart failure

The CoDE-HF decision-support tool was developed using data from multinational collaborative analysis. However, the data were from studies that, on average, were conducted over a decade ago while also using different natriuretic peptide assays. Moreover, the patient population and the disease's prevalence may have changed over these years. Hence, it is unknown what the diagnostic performance of CoDE-HF would be in contemporary practice. For that reason, with Professor Mills and Dr Lee, we secured funding from a Confidence in Concept scheme, funded by the Medical Research Council award, to prospectively evaluate CoDE-HF's performance in a contemporary cohort of patients with suspected acute heart failure. The prospective observational cohort study is called Prospective Validation of the CoDE-HF algorithm for the diagnosis of acute heart failure (ProVa CoDE-HF).

We are currently in the recruitment phase with the aim of enrolling 2,000 consecutive patients presenting to the Emergency Department with suspected acute heart failure. The electronic patient record will be used to assess patient eligibility and capture information about the patient's presenting symptoms, clinical examination findings, previous medical history, appropriate investigations, and diagnostic adjudication. Blood samples submitted by the clinical care team during the initial hospital visit will be used to acquire plasma samples that are surplus to clinical requirements. These

samples will be used to measure the NT-proBNP. The primary outcome will be an adjudicated diagnosis of acute heart failure.

8.2.3 Deep learning for CoDE-ACS and CoDE-HF

The electrocardiogram (ECG) is the most commonly used first-line, non-invasive cardiac test for patients with suspected acute cardiac conditions. However, the diagnostic value of the ECG as used in current practice is limited outwith the identification of patients with overt ST-segment elevation myocardial infarction or bundle branch block.

The primary features of anomalies such as the ST-segment, and T and Q waves associated with ischemia are utilised. For example, during the development of CoDE-ACS, binary variables of ST segment (yes/no), T and Q waves (yes/no) were available in the dataset. However, these binary features were found not to be important. This is primarily because the interpretation is based on predefined rules and subjective manual pattern or feature recognition algorithms that do not always capture the complexity and nuances of an ECG. Only ~50% of the patients ultimately diagnosed with acute myocardial infarction have what are considered to be diagnostic ST-segment changes on their initial ECG.[205,206]

We hypothesize that by incorporating raw data from the ECG into the CoDE-ACS and CoDE-HF decision support tools it will be possible to improve diagnostic and prognostic performance. By performing deep learning on ECGs, new features will be extracted from the raw ECG signal without the

need for laborious hand-crafted feature extraction or the subjective interpretation of the ECG by clinical readers of variable experience. We are going to use multiple types of neural networks that are used for the classification and prediction of outcomes based on raw ECG input. Examples of network architectures include, but will not be limited to convolutional neural networks (CNN) and recurrent neural networks (RNN).

8.2.4 Qualitative evaluation of the decision support tools user experience

For both the CoDE-ACS and CoDE-HF decision support tools and in parallel with their further development and their prospective validation, we want to conduct a qualitative assessment of clinician experience using these tools. This will help me understand the best way to implement these decision support tools in clinical practice. Moreover, this will help me understand better the aspects that influence diagnostic certainty and the utilisation of cardiac imaging or medicines. Despite research showing that some AI-enabled clinical decision support systems may match the diagnostic accuracy of professional doctors, machine learning in healthcare is difficult to implement and maintain. Humans play an important role in the actual application of clinical decision support systems, but a lack of focus on the interaction of social and technical parts of the healthcare system is a major reason for new technology failure.

8.2.5 Implementation of the decision support tools into clinical practice

My aim is to develop CoDE-ACS and CoDE-HF into UKCA/CE marked in vitro diagnostic devices and integrate them into the clinical workflow in the NHS. This will enable me to demonstrate that their use improves the diagnosis of acute myocardial infarction and acute heart failure, and results in better targeting of cardiac imaging and therapies.

Ultimately this will require the development and deployment of a clinical dashboard, which will integrate the data capture for the decision support tools and illustrate their output in the electronic patient record. This will enable me to deliver with my clinical colleagues prospective, randomised trials to assess the impact of implementing CoDE-ACS and CoDE-HF clinical decision support tools on the diagnosis and management of patients with suspected acute cardiac conditions in the Emergency Department.

Published papers during PhD period

Doudesis, D.*, Lee, K. K.*, Anwar M.*, Astengo, F., Chenevier-Gobeaux, C., Claessens, YE., . . . Mills N. L. Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study. *BMJ* 2022; 377 doi:10.1136/bmj-2021-068424

Doudesis, D.*, Lee, K. K.*, Yang, J., Wereski, R., Shah, A. SV., Tsanas, A., . . . Mills N. L. Validation of the myocardial-ischemic-injury-index (MI^3) machine learning algorithm to guide the diagnosis of myocardial infarction in a heterogeneous population. *Lancet Digital Health* doi:10.1016/S2589-7500(22)00025-5

Doudesis, D., & Manataki, A. (2022). Data science in undergraduate medicine: Course overview and student perspectives. *Int J Med Inform*, 159, 104668. doi:10.1016/j.ijmedinf.2021.104668

Lee, K. K., **Doudesis, D.**, Ross, D. A., Bularga, A., MacKintosh, C. L., Koch, O., . . . Mills, N. L. (2021). Diagnostic performance of the combined nasal and throat swab in patients admitted to hospital with suspected COVID-19. *BMC Infect Dis*, 21(1), 318. doi:10.1186/s12879-021-05976-1

Bularga, A., Meah, M. N., **Doudesis, D.,** Shah, A. S. V., Mills, N. L., Newby, D. E., & Lee, K. K. (2021). Duration of dual antiplatelet therapy and stability

of coronary heart disease: a 60 000-patient meta-analysis of randomised controlled trials. *Open Heart*, 8(2). doi:10.1136/openhrt-2021-001707

Lee, K. K., Bularga, A., O'Brien, R., Ferry, A. V., **Doudesis, D.**, Fujisawa, T., . . . Mills, N. L. (2021). Troponin-Guided Coronary Computed Tomographic Angiography After Exclusion of Myocardial Infarction. *J Am Coll Cardiol*, 78(14), 1407-1417. doi:10.1016/j.jacc.2021.07.055

Wereski, R., Kimenai, D. M., Taggart, C., **Doudesis, D.**, Lee, K. K., Lowry, M. T. H., . . . Mills, N. L. (2021). Cardiac Troponin Thresholds and Kinetics to Differentiate Myocardial Injury and Myocardial Infarction. *Circulation*, 144(7), 528-538. doi:10.1161/CIRCULATIONAHA.121.054302

Tibble, H., Chan, A., Mitchell, E. A., Horne, E., **Doudesis, D.**, Horne, R., . . . Tsanas, A. (2020). A data-driven typology of asthma medication adherence using cluster analysis. *Sci Rep*, 10(1), 14999. doi:10.1038/s41598-020-72060-0

Lee, K. K., Bing, R., Kiang, J., Bashir, S., Spath, N., Stelzle, D., **Doudesis, D.,** . . . Shah, A. S. V. (2020). Adverse health effects associated with household air pollution: a systematic review, meta-analysis, and burden estimation study. *Lancet Glob Health*, 8(11), e1427-e1434. doi:10.1016/S2214-109X(20)30343-0

*denotes equal contribution of authorship

References

1. Fighting cardiovascular disease– a blueprint for EU action. https://ehnheart.org/component/attachments/?task=download&id=3185:05748-CVD-plan_digital-edition.

2. British Heart Foundation Heart and Circulatory Diseases Statistics, British Heart Foundation, 2021. Accessed June 7th 2021 (www.bhf.org.uk).

3. Mafham MM, Spata E, Goldacre R, et al. COVID-19 pandemic and admission rates for and management of acute coronary syndromes in England. Lancet 2020;396:381-9.

4. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth Universal Definition of Myocardial Infarction (2018). J Am Coll Cardiol 2018;72:2231-64.

5. Wu AHB, Christenson RH, Greene DN, et al. Clinical Laboratory Practice Recommendations for the Use of Cardiac Troponin in Acute Coronary Syndrome: Expert Opinion from the Academy of the American Association for Clinical Chemistry and the Task Force on Clinical Applications of Cardiac Bio-Markers of the International Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem 2018;64:645-55.

6. Twerenbold R, Jaffe A, Reichlin T, Reiter M, Mueller C. High-sensitive troponin T measurements: what do we gain and what are the challenges? Eur Heart J 2012;33:579-86.

7. Faxon DP, Fuster V, Libby P, et al. Atherosclerotic Vascular Disease Conference: Writing Group III: pathophysiology. Circulation 2004;109:2617-25.

8. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. Nature 2011;473:317-25.

9. Sahin B, Ilgun G. Risk factors of deaths related to cardiovascular diseases in World Health Organization (WHO) member countries. Health Soc Care Community 2022;30:73-80.

10. Virani SS, Alonso A, Aparicio HJ, et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. Circulation 2021;143:e254-e743.

11. Goodacre S, Cross E, Arnold J, Angelini K, Capewell S, Nicholl J. The health care burden of acute chest pain. Heart 2005;91:229-30.

12. Wenger NK. Women and coronary heart disease: a century after Herrick: understudied, underdiagnosed, and undertreated. Circulation 2012;126:604-11.

13. Mehta LS, Beckie TM, DeVon HA, et al. Acute Myocardial Infarction in Women: A Scientific Statement From the American Heart Association. Circulation 2016;133:916-47.

14. Deedwania PC, Carbajal EV. Silent myocardial ischemia. A clinical perspective. Arch Intern Med 1991;151:2373-82.

15. Pepine CJ. Multiple Causes for Ischemia Without Obstructive Coronary Artery Disease. Circulation 2015;131:1044-6.

16. Anderson JL, Morrow DA. Acute Myocardial Infarction. N Engl J Med 2017;376:2053-64.

17. Thygesen K, Alpert JS, White HD. Universal Definition of Myocardial Infarction. Circulation 2007;116:2634-53.

18. Mann, D. L. in Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine (eds Bonow, R. O., Mann, D. L., Zipes, D. P. & Libby, P.) 487–504 (Elsevier Saunders, 2012).

19. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation 2013;128:1810-52.

20. Velagaleti RS, Pencina MJ, Murabito JM, et al. Long-term trends in the incidence of heart failure after myocardial infarction. Circulation 2008;118:2057-62.

21. Gerber Y, Weston SA, Berardi C, et al. Contemporary trends in heart failure with reduced and preserved ejection fraction after myocardial infarction: a community study. Am J Epidemiol 2013;178:1272-80.

22. Ezekowitz JA, Kaul P, Bakal JA, Armstrong PW, Welsh RC, McAlister FA. Declining in-hospital mortality and increasing heart failure incidence in elderly patients with first myocardial infarction. J Am Coll Cardiol 2009;53:13-20.

23. Gheorghiade M, Bonow RO. Chronic heart failure in the United States: a manifestation of coronary artery disease. Circulation 1998;97:282-9.

24. Chen J, Hsieh AF, Dharmarajan K, Masoudi FA, Krumholz HM. National trends in heart failure hospitalization after acute myocardial infarction for Medicare beneficiaries: 1998-2010. Circulation 2013;128:2577-84.

25. Borlaug BA, Paulus WJ. Heart failure with preserved ejection fraction: pathophysiology, diagnosis, and treatment. Eur Heart J 2011;32:670-9.

26. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol 2011;8:30-41.

27. McMurray JJ, Adamopoulos S, Anker SD, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J 2012;33:1787-847.

28. Mazurek JA, Jessup M. Understanding Heart Failure. Heart Fail Clin 2017;13:1-19.

29. Mant J, Doust J, Roalfe A, et al. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. Health Technol Assess 2009;13:1-207, iii.

30. Oudejans I, Mosterd A, Bloemen JA, et al. Clinical evaluation of geriatric outpatients with suspected heart failure: value of symptoms, signs, and additional tests. Eur J Heart Fail 2011;13:518-27.

31. Fonseca C. Diagnosis of heart failure in primary care. Heart Fail Rev 2006;11:95-107.

32. Kelder JC, Cramer MJ, van Wijngaarden J, et al. The diagnostic value of physical examination and additional testing in primary care patients with suspected heart failure. Circulation 2011;124:2865-73.

33. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. European Heart Journal 2021;42:3599-726.

34. Miro O, Garcia Sarasola A, Fuenzalida C, et al. Departments involved during the first episode of acute heart failure and subsequent emergency department revisits and rehospitalisations: an outlook through the NOVICA cohort. Eur J Heart Fail 2019;21:1231-44.

35. Chioncel O, Mebazaa A, Maggioni AP, et al. Acute heart failure congestion and perfusion status - impact of the clinical classification on in-hospital and long-term outcomes; insights from the ESC-EORP-HFA Heart Failure Long-Term Registry. Eur J Heart Fail 2019;21:1338-52.

36. Chioncel O, Mebazaa A, Harjola VP, et al. Clinical phenotypes and outcome of patients hospitalized for acute heart failure: the ESC Heart Failure Long-Term Registry. Eur J Heart Fail 2017;19:1242-54.

37. Nieminen MS, Brutsaert D, Dickstein K, et al. EuroHeart Failure Survey II (EHFS II): a survey on hospitalized acute heart failure patients: description of population. Eur Heart J 2006;27:2725-36.

38. O'Meara E, Clayton T, McEntegart MB, et al. Sex differences in clinical characteristics and prognosis in a broad spectrum of patients with heart failure: results of the Candesartan in Heart failure: Assessment of Reduction in Mortality and morbidity (CHARM) program. Circulation 2007;115:3111-20.

39. Crespo-Leiro MG, Anker SD, Maggioni AP, et al. European Society of Cardiology Heart Failure Long-Term Registry (ESC-HF-LT): 1-year follow-up outcomes and differences across regions. Eur J Heart Fail 2016;18:613-25.

40. Tomasoni D, Lombardi CM, Sbolli M, Cotter G, Metra M. Acute heart failure: More questions than answers. Prog Cardiovasc Dis 2020;63:599-606.

41. Bocchi EA, Vilas-Boas F, Perrone S, et al. I Latin American Guidelines for the Assessment and Management of Decompensated Heart Failure. Arq Bras Cardiol 2005;85 Suppl 3:49-94; 1-48.

42. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37:2129-200.

43. Collet JP, Thiele H, Barbato E, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. Eur Heart J 2021;42:1289-367.

44. Jaffe AS, Ravkilde J, Roberts R, et al. It's time for a change to a troponin standard. Circulation 2000;102:1216-20.

45. Adams JE, 3rd, Abendschein DR, Jaffe AS. Biochemical markers of myocardial injury. Is MB creatine kinase the choice for the 1990s? Circulation 1993;88:750-63.

46. Saenger AK, Jaffe AS. Requiem for a heavyweight: the demise of creatine kinase-MB. Circulation 2008;118:2200-6.

47. Cummins B, Auckland ML, Cummins P. Cardiac-specific troponin-I radioimmunoassay in the diagnosis of acute myocardial infarction. Am Heart J 1987;113:1333-44.

48. Fishbein MC, Wang T, Matijasevic M, Hong L, Apple FS. Myocardial tissue troponins T and I. An immunohistochemical study in experimental models of myocardial ischemia. Cardiovasc Pathol 2003;12:65-71.

49. Katus HA, Remppis A, Looser S, Hallermeier K, Scheffold T, Kubler W. Enzyme linked immuno assay of cardiac troponin T for the detection of acute myocardial infarction in patients. J Mol Cell Cardiol 1989;21:1349-53.

50. Collinson P. High-sensitivity troponin measurements: challenges and opportunities for the laboratory and the clinician. Ann Clin Biochem 2016;53:191-5.

51. Melanson SE, Tanasijevic MJ, Jarolim P. Cardiac troponin assays: a view from the clinical chemistry laboratory. Circulation 2007;116:e501-4.

52. Apple FS, Collinson PO, Biomarkers ITFoCAoC. Analytical characteristics of high-sensitivity cardiac troponin assays. Clin Chem 2012;58:54-61.

53. Gray A, Ward K, Lees F, et al. The epidemiology of adults with severe sepsis and septic shock in Scottish emergency departments. Emerg Med J 2013;30:397-401.

54. Konstantinides S, Geibel A, Olschewski M, et al. Importance of cardiac troponins I and T in risk stratification of patients with acute pulmonary embolism. Circulation 2002;106:1263-8.

55. Roongsritong C, Warraich I, Bradley C. Common causes of troponin elevations in the absence of acute myocardial infarction: incidence and clinical significance. Chest 2004;125:1877-84.

56. Roos A, Bandstein N, Lundback M, Hammarsten O, Ljung R, Holzmann MJ. Stable High-Sensitivity Cardiac Troponin T Levels and Outcomes in Patients With Chest Pain. J Am Coll Cardiol 2017;70:2226-36.

57. Wong CK, Szeto CC, Chan MH, Leung CB, Li PK, Lam CW. Elevation of pro-inflammatory cytokines, C-reactive protein and cardiac troponin T in chronic renal failure patients on dialysis. Immunol Invest 2007;36:47-57.

58. Eggers KM, Lind L, Venge P, Lindahl B. Factors influencing the 99th percentile of cardiac troponin I evaluated in community-dwelling individuals at 70 and 75 years of age. Clin Chem 2013;59:1068-73.

59. Gore MO, Seliger SL, Defilippi CR, et al. Age- and sex-dependent upper reference limits for the high-sensitivity cardiac troponin T assay. J Am Coll Cardiol 2014;63:1441-8.

60. Olivieri F, Galeazzi R, Giavarina D, et al. Aged-related increase of high sensitive Troponin T and its implication in acute myocardial infarction diagnosis of elderly patients. Mech Ageing Dev 2012;133:300-5.

61. Shah AS, Griffiths M, Lee KK, et al. High sensitivity cardiac troponin and the under-diagnosis of myocardial infarction in women: prospective cohort study. BMJ 2015;350:g7873.

62. Becattini C, Vedovati MC, Agnelli G. Prognostic value of troponins in acute pulmonary embolism: a meta-analysis. Circulation 2007;116:427-33.

63. Havekes B, van Manen JG, Krediet RT, et al. Serum troponin T concentration as a predictor of mortality in hemodialysis and peritoneal dialysis patients. Am J Kidney Dis 2006;47:823-9.

64. Apple FS, Sandoval Y, Jaffe AS, Ordonez-Llanos J, Bio-Markers ITFoCAoC. Cardiac Troponin Assays: Guide to Understanding Analytical Characteristics and Their Impact on Clinical Care. Clin Chem 2017;63:73-81.

65. Giannitsis E, Kurz K, Hallermayer K, Jarausch J, Jaffe AS, Katus HA. Analytical validation of a high-sensitivity cardiac troponin T assay. Clin Chem 2010;56:254-61.

66. Goodman SG, Steg PG, Eagle KA, et al. The diagnostic and prognostic impact of the redefinition of acute myocardial infarction: lessons from the Global Registry of Acute Coronary Events (GRACE). Am Heart J 2006;151:654-60.

67. Sandoval Y, Apple FS. The global need to define normality: the 99th percentile value of cardiac troponin. Clin Chem 2014;60:455-62.

68. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. Circulation 2012;126:2020-35.

69. Body R, Carley S, McDowell G, et al. Rapid exclusion of acute myocardial infarction in patients with undetectable troponin using a high-sensitivity assay. J Am Coll Cardiol 2011;58:1332-9.

70. Body R, Mueller C, Giannitsis E, et al. The Use of Very Low Concentrations of High-sensitivity Troponin T to Rule Out Acute Myocardial Infarction Using a Single Blood Test. Academic Emergency Medicine 2016;23:1004-13.

71. Lee KK, Ferry AV, Anand A, et al. Sex-Specific Thresholds of High-Sensitivity Troponin in Patients With Suspected Acute Coronary Syndrome. J Am Coll Cardiol 2019;74:2032-43.

72. Pickering JW, Than MP, Cullen L, et al. Rapid Rule-out of Acute Myocardial Infarction With a Single High-Sensitivity Cardiac Troponin T Measurement Below the Limit of Detection: A Collaborative Meta-analysis. Ann Intern Med 2017;166:715-24.

73. Shah ASV, Anand A, Strachan FE, et al. High-sensitivity troponin in the evaluation of patients with suspected acute coronary syndrome: a stepped-wedge, cluster-randomised controlled trial. Lancet 2018;392:919-28.

74. Mills NL, Churchhouse AM, Lee KK, et al. Implementation of a sensitive troponin I assay and risk of recurrent myocardial infarction and death in patients with suspected acute coronary syndrome. JAMA 2011;305:1210-6.

75. Neumann JT, Sorensen NA, Schwemer T, et al. Diagnosis of Myocardial Infarction Using a High-Sensitivity Troponin I 1-Hour Algorithm. JAMA Cardiol 2016;1:397-404.

76. Ambavane A, Lindahl B, Giannitsis E, et al. Economic evaluation of the one-hour rule-out and rule-in algorithm for acute myocardial infarction using the high-sensitivity cardiac troponin T assay in the emergency department. PLoS One 2017;12:e0187662.

77. Boeddinghaus J, Nestelberger T, Twerenbold R, et al. Direct Comparison of 4 Very Early Rule-Out Strategies for Acute Myocardial

Infarction Using High-Sensitivity Cardiac Troponin I. Circulation 2017;135:1597-611.

78. Ljung L, Lindahl B, Eggers KM, et al. A Rule-Out Strategy Based on High-Sensitivity Troponin and HEART Score Reduces Hospital Admissions. Ann Emerg Med 2019;73:491-9.

79. Wildi K, Boeddinghaus J, Nestelberger T, et al. Comparison of fourteen rule-out strategies for acute myocardial infarction. Int J Cardiol 2019;283:41-7.

80. Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2017;136:e137-e61.

81. Lin DC, Diamandis EP, Januzzi JL, Jr., Maisel A, Jaffe AS, Clerico A. Natriuretic peptides in heart failure. Clin Chem 2014;60:1040-6.

82. Lei W, Zhang Y, Yu G, et al. Cloning and sequence analysis of an Ophiophagus hannah cDNA encoding a precursor of two natriuretic peptide domains. Toxicon 2011;57:811-6.

83. Cao Z, Jia Y, Zhu B. BNP and NT-proBNP as Diagnostic Biomarkers for Cardiac Dysfunction in Both Clinical and Forensic Medicine. Int J Mol Sci 2019;20.

84. Pemberton CJ, Johnson ML, Yandle TG, Espiner EA. Deconvolution analysis of cardiac natriuretic peptides during acute volume overload. Hypertension 2000;36:355-9.

85. Stingo AJ, Clavell AL, Heublein DM, Wei CM, Pittelkow MR, Burnett JC, Jr. Presence of C-type natriuretic peptide in cultured human endothelial cells and plasma. Am J Physiol 1992;263:H1318-21.

86. Gottdiener JS, Arnold AM, Aurigemma GP, et al. Predictors of congestive heart failure in the elderly: the Cardiovascular Health Study. J Am Coll Cardiol 2000;35:1628-37.

87. Heywood JT, Fonarow GC, Costanzo MR, et al. High prevalence of renal dysfunction and its impact on outcome in 118,465 patients hospitalized with acute decompensated heart failure: a report from the ADHERE database. J Card Fail 2007;13:422-30.

88. Yamagata K, Ishida K, Sairenchi T, et al. Risk factors for chronic kidney disease in a community-based population: a 10-year follow-up study. Kidney Int 2007;71:159-66.

89. Hillege HL, Nitsch D, Pfeffer MA, et al. Renal function as a predictor of outcome in a broad spectrum of patients with heart failure. Circulation 2006;113:671-8.

90. Morello A, Lloyd-Jones DM, Chae CU, et al. Association of atrial fibrillation and amino-terminal pro-brain natriuretic peptide concentrations in dyspneic subjects with and without acute heart failure: results from the ProBNP Investigation of Dyspnea in the Emergency Department (PRIDE) study. Am Heart J 2007;153:90-7.

91. Richards M, Di Somma S, Mueller C, et al. Atrial fibrillation impairs the diagnostic performance of cardiac natriuretic peptides in dyspneic patients:

results from the BACH Study (Biomarkers in ACute Heart Failure). JACC Heart Fail 2013;1:192-9.

92. Das SR, Drazner MH, Dries DL, et al. Impact of body mass and body composition on circulating levels of natriuretic peptides: results from the Dallas Heart Study. Circulation 2005;112:2163-8.

93. Mehra MR, Uber PA, Park MH, et al. Obesity and suppressed B-type natriuretic peptide levels in heart failure. J Am Coll Cardiol 2004;43:1590-5.

94. Wang TJ, Larson MG, Levy D, et al. Impact of obesity on plasma natriuretic peptide levels. Circulation 2004;109:594-600.

95. Vasile VC, Jaffe AS. Natriuretic Peptides and Analytical Barriers. Clin Chem 2017;63:50-8.

96. Prontera C, Zaninotto M, Giovannini S, et al. Proficiency testing project for brain natriuretic peptide (BNP) and the N-terminal part of the propeptide of BNP (NT-proBNP) immunoassays: the CardioOrmocheck study. Clin Chem Lab Med 2009;47:762-8.

97. National Institute for Health and Care Excellence, NICE Clinical Guideline 187 [CG187]. Acute Heart Failure. 2014.

98. Januzzi JL, van Kimmenade R, Lainchbury J, et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study. Eur Heart J 2006;27:330-7.

99. Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med 2002;347:161-7.

100. Mueller T, Gegenhuber A, Poelz W, Haltmayer M. Diagnostic accuracy of B type natriuretic peptide and amino terminal proBNP in the emergency diagnosis of heart failure. Heart 2005;91:606-12.

101. Gohar A, Rutten FH, den Ruijter H, et al. Mid-regional pro-atrial natriuretic peptide for the early detection of non-acute heart failure. Eur J Heart Fail 2019;21:1219-27.

102. Wang TJ, Larson MG, Levy D, et al. Impact of age and sex on plasma natriuretic peptide levels in healthy adults. Am J Cardiol 2002;90:254-8.

103. Januzzi JL, Jr., Chen-Tournoux AA, Christenson RH, et al. N-Terminal Pro-B-Type Natriuretic Peptide in the Emergency Department: The ICON-RELOADED Study. J Am Coll Cardiol 2018;71:1191-200.

104. Cullen L, Greenslade JH, Hawkins T, et al. Improved Assessment of Chest pain Trial (IMPACT): assessing patients with possible acute coronary syndromes. Medical Journal of Australia 2017;207:195-200.

105. Than M, Cullen L, Aldous S, et al. 2-Hour Accelerated Diagnostic Protocol to Assess Patients With Chest Pain Symptoms Using Contemporary Troponins as the Only Biomarker. Journal of the American College of Cardiology 2012;59:2091-8.

106. Pickering JW, Young JM, George PM, et al. Early kinetic profiles of troponin I and T measured by high-sensitivity assays in patients with myocardial infarction. Clin Chim Acta 2020;505:15-25.

107. Roberts E, Ludman AJ, Dworzynski K, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ 2015;350:h910.

108. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer New York Inc.; 2001.

109. Menard S. Applied logistic regression analysis: Sage; 2002.

110. Steyerberg EW, Eijkemans MJ, Harrell Jr FE, Habbema JDF. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Medical Decision Making 2001;21:45-56.

111. Van Houwelingen J, Le Cessie S. Logistic regression, a review. Statistica Neerlandica 1988;42:215-32.

112. Nelder JA, Wedderburn RW. Generalized linear models. Journal of the Royal Statistical Society: Series A (General) 1972;135:370-84.

113. Hedeker D. Generalized linear mixed models. Encyclopedia of statistics in behavioral science 2005.

114. Stroup WW. Generalized linear mixed models: modern concepts, methods and applications: CRC press; 2012.

115. Clayton DG. Generalized linear mixed models. Markov chain Monte Carlo in practice 1996;1:275-302.

116. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol 1999;28:964-74.

117. George, Langley P. Estimating Continuous Distributions in Bayesian Classifiers. arXiv pre-print server 2013.

118. Murphy KP. Naive bayes classifiers. University of British Columbia 2006;18:1-8.

119. Breiman L. Machine Learning 2001;45:5-32.

120. Breiman L. Random forests. Machine learning 2001;45:5-32.

121. Ho TK. Random decision forests. Proceedings of 3rd international conference on document analysis and recognition; 1995: IEEE. p. 278-82.

122. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics 2001:1189-232.

123. Chen T, Guestrin C. XGBoost. 2016: ACM.

124. Breiman L. Classification and regression trees. Routledge; 2017.

125. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research 2003;3:1157-82.

126. Kira K, Rendell LA. A practical approach to feature selection. Machine learning proceedings 1992: Elsevier; 1992:249-56.

127. Refaeilzadeh P, Tang L, Liu H. Cross-validation. Encyclopedia of database systems 2009;5:532-8.

128. Browne MW. Cross-validation methods. Journal of mathematical psychology 2000;44:108-32.

129. Donders ART, Van Der Heijden GJ, Stijnen T, Moons KG. A gentle introduction to imputation of missing values. Journal of clinical epidemiology 2006;59:1087-91.

130. Royston P. Multiple imputation of missing values. The Stata Journal 2004;4:227-41.

131. Schafer JL. Multiple imputation: a primer. Statistical methods in medical research 1999;8:3-15.

132. Royston P, White IR. Multiple imputation by chained equations (MICE): implementation in Stata. Journal of statistical software 2011;45:1-20.

133. Quartagno M, Carpenter JR. Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. Stat Med 2016;35:2938-54.

134. Sanchis-Gomar F, Perez-Quilis C, Leischik R, Lucia A. Epidemiology of coronary heart disease and acute coronary syndrome. Ann Transl Med 2016;4:256.

135. Collet JP, Thiele H, Barbato E, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. Eur Heart J 2020.

136. Babuin L, Jaffe AS. Troponin: the biomarker of choice for the detection of cardiac injury. CMAJ 2005;173:1191-202.

137. Body R, Carlton E, Sperrin M, et al. Troponin-only Manchester Acute Coronary Syndromes (T-MACS) decision aid: single biomarker re-derivation and external validation in three cohorts. Emerg Med J 2017;34:349-56.

138. Reichlin T, Schindler C, Drexler B, et al. One-hour rule-out and rule-in of acute myocardial infarction using high-sensitivity cardiac troponin T. Arch Intern Med 2012;172:1211-8.

139. Than M, Flaws D, Sanders S, et al. Development and validation of the Emergency Department Assessment of Chest pain Score and 2 h accelerated diagnostic protocol. Emerg Med Australas 2014;26:34-44.

140. Chapman AR, Anand A, Boeddinghaus J, et al. Comparison of the Efficacy and Safety of Early Rule-Out Pathways for Acute Myocardial Infarction. Circulation 2017;135:1586-96.

141. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. Circulation 2012;126:2020-35.

142. Than MP, Pickering JW, Sandoval Y, et al. Machine learning to predict the likelihood of acute myocardial infarction. Circulation 2019;140:899-909.

143. Shah ASV, Anand A, Strachan FE, et al. High-sensitivity troponin in the evaluation of patients with suspected acute coronary syndrome: a stepped-wedge, cluster-randomised controlled trial. Lancet 2018;392:919-28.

144. Scottish Intercollegiate Guidelines Network (SIGN). Acute coronary syndromes. Edinburgh: SIGN; 2013. (SIGN publication no. 93). [February 2013]. Available from URL: http://www.sign.ac.uk.

145. Shah AS, Anand A, Sandoval Y, et al. High-sensitivity cardiac troponin I at presentation in patients with suspected acute coronary syndrome: a cohort study. Lancet 2015;386:2481-8.

146. Mueller-Hennessen M, Lindahl B, Giannitsis E, et al. Diagnostic and prognostic implications using age- and gender-specific cut-offs for high-sensitivity cardiac troponin T - Sub-analysis from the TRAPID-AMI study. Int J Cardiol 2016;209:26-33.

147. Shah ASV, Sandoval Y, Noaman A, et al. Patient selection for high sensitivity cardiac troponin testing and diagnosis of myocardial infarction: prospective cohort study. BMJ 2017;359:j4788.

148. Eggers KM, Ellenius J, Dellborg M, et al. Artificial neural network algorithms for early diagnosis of acute myocardial infarction and prediction of infarct size in chest pain patients. Int J Cardiol 2007;114:366-74.

149. Green M, Bjork J, Forberg J, Ekelund U, Edenbrandt L, Ohlsson M. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artif Intell Med 2006;38:305-18.

150. Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure: a systematic review and analysis. JACC Heart Fail 2014;2:440-6.

151. Wessler BS, Lai Yh L, Kramer W, et al. Clinical Prediction Models for Cardiovascular Disease: Tufts Predictive Analytics and Comparative Effectiveness Clinical Prediction Model Database. Circ Cardiovasc Qual Outcomes 2015;8:368-75.

152. Califf RM, Pencina MJ. Predictive models in heart failure: who cares? Circ Heart Fail 2013;6:877-8.

153. Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416.

154. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25-34.

155. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth Universal Definition of Myocardial Infarction (2018). Circulation 2018;138:e618-e51.

156. Anand A, Shah ASV, Beshiri A, Jaffe AS, Mills NL. Global Adoption of High-Sensitivity Cardiac Troponins and the Universal Definition of Myocardial Infarction. Clin Chem 2019;65:484-9.

157. Roffi M, Patrono C, Collet JP, et al. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). Eur Heart J 2016;37:267-315.

158. Neumann JT, Twerenbold R, Ojeda F, et al. Application of High-Sensitivity Troponin in Suspected Myocardial Infarction. N Engl J Med 2019;380:2529-40.

159. Lee KK, Noaman A, Vaswani A, et al. Prevalence, Determinants, and Clinical Associations of High-Sensitivity Cardiac Troponin in Patients Attending Emergency Departments. Am J Med 2019;132:110 e8- e21.

160. Wereski R, Kimenai DM, Taggart C, et al. Cardiac Troponin Thresholds and Kinetics to Differentiate Myocardial Injury and Myocardial Infarction. Circulation 2021;144:528-38.

161. Than M, Aldous S, Lord SJ, et al. A 2-Hour Diagnostic Protocol for Possible Cardiac Chest Pain in the Emergency Department. JAMA Internal Medicine 2014;174:51.

162. Chapman AR, Adamson PD, Shah ASV, et al. High-Sensitivity Cardiac Troponin and the Universal Definition of Myocardial Infarction. Circulation 2020;141:161-71.

163. Miller-Hodges E, Anand A, Shah ASV, et al. High-Sensitivity Cardiac Troponin and the Risk Stratification of Patients With Renal Impairment

Presenting With Suspected Acute Coronary Syndrome. Circulation 2018;137:425-35.

164. Anand A, Lee KK, Chapman AR, et al. High-Sensitivity Cardiac Troponin on Presentation to Rule Out Myocardial Infarction: A Stepped-Wedge Cluster Randomized Controlled Trial. Circulation 2021;143:2214-24.

165. Carlton EW, Ingram J, Taylor H, et al. Limit of detection of troponin discharge strategy versus usual care: randomised controlled trial. Heart 2020;106:1586-94.

166. Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. The Lancet 2017.

167. National Heart Failure Audit 2019 Report. University College London: NICOR (National Institute for Cardiovascular Outcomes Research). 2019.

168. Mueller C, McDonald K, de Boer RA, et al. Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. Eur J Heart Fail 2019;21:715-31.

169. Christiansen MN, Kober L, Weeke P, et al. Age-Specific Trends in Incidence, Mortality, and Comorbidities of Heart Failure in Denmark, 1995 to 2012. Circulation 2017;135:1214–23.

170. Baggish AL, van Kimmenade RR, Januzzi JL, Jr. The differential diagnosis of an elevated amino-terminal pro-B-type natriuretic peptide level. Am J Cardiol 2008;101:43–8.

171. Kozhuharov N, Sabti Z, Wussler D, et al. Prospective validation of N-terminal pro B-type natriuretic peptide cut-off concentrations for the diagnosis of acute heart failure. Eur J Heart Fail 2019;21:813-5.

172. Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J 2013;34:1404-13.

173. Wussler D, Kozhuharov N, Sabti Z, et al. External Validation of the MEESSI Acute Heart Failure Risk Score: A Cohort Study. Ann Intern Med 2019;170:248-56.

174. Simpson J, Jhund PS, Lund LH, et al. Prognostic Models Derived in PARADIGM-HF and Validated in ATMOSPHERE and the Swedish Heart Failure Registry to Predict Mortality and Morbidity in Chronic Heart Failure. JAMA Cardiol 2020.

175. Baggish AL, Siebert U, Lainchbury JG, et al. A validated clinical and biochemical score for the diagnosis of acute heart failure: the ProBNP Investigation of Dyspnea in the Emergency Department (PRIDE) Acute Heart Failure Score. Am Heart J 2006;151:48-54.

176. Steinhart B, Thorpe KE, Bayoumi AM, Moe G, Januzzi JL, Jr., Mazer CD. Improving the diagnosis of acute heart failure using a validated prediction model. J Am Coll Cardiol 2009;54:1515-21.

177. Pivetta E, Goffi A, Lupia E, et al. Lung Ultrasound-Implemented Diagnosis of Acute Decompensated Heart Failure in the ED: A SIMEU Multicenter Study. Chest 2015;148:202-10.

178. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.

179. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Statist 2000;28:337-407.

180. Bahrmann P, Bahrmann A, Hofner B, et al. Multiple biomarker strategy for improved diagnosis of acute heart failure in older patients presenting to the emergency department. Eur Heart J Acute Cardiovasc Care 2015;4:137-47.

181. Behnes M, Brueckmann M, Ahmad-Nejad P, et al. Diagnostic performance and cost effectiveness of measurements of plasma N-terminal pro brain natriuretic peptide in patients presenting with acute dyspnea or peripheral edema. Int J Cardiol 2009;135:165–74.

182. Bombelli M, Maloberti A, Rossi S, et al. Clinical value of NT-proBNP assay in the emergency department for the diagnosis of heart failure (HF) in very elderly people. Arch Gerontol Geriatr 2015;61:296-300.

183. Chenevier-Gobeaux C, Claessens YE, Voyer S, Desmoulins D, Ekindjian OG. Influence of renal function on N-terminal pro-brain natriuretic peptide (NT-proBNP) in patients admitted for dyspnoea in the Emergency Department: comparison with brain natriuretic peptide (BNP). Clin Chim Acta 2005;361:167–75.

184. deFilippi CR, Seliger SL, Maynard S, Christenson RH. Impact of renal disease on natriuretic peptide testing for diagnosing decompensated heart failure and predicting mortality. Clin Chem 2007;53:1511-9.

185. Gargani L, Frassi F, Soldati G, Tesorio P, Gheorghiade M, Picano E. Ultrasound lung comets for the differential diagnosis of acute cardiogenic dyspnoea: a comparison with natriuretic peptides. Eur J Heart Fail 2008;10:70–7.

186. Ibrahim I, Kuan WS, Frampton C, et al. Superior performance of N-terminal pro brain natriuretic peptide for diagnosis of acute decompensated heart failure in an Asian compared with a Western setting. Eur J Heart Fail 2017;19:209-17.

187. Maisel A, Mueller C, Nowak R, et al. Mid-region pro-hormone markers for diagnosis and prognosis in acute dyspnea: results from the BACH (Biomarkers in Acute Heart Failure) trial. J Am Coll Cardiol 2010;55:2062–76.

188. Moe GW, Howlett J, Januzzi JL, Zowall H, Canadian Multicenter Improved Management of Patients With Congestive Heart Failure Study I. N-terminal pro-B-type natriuretic peptide testing improves the management of patients with suspected acute heart failure: primary results of the Canadian prospective randomized multicenter IMPROVE-CHF study. Circulation 2007;115:3103-10.

189. Nazerian P, Vanni S, Zanobetti M, et al. Diagnostic accuracy of emergency Doppler echocardiography for identification of acute left ventricular heart failure in patients with acute dyspnea: comparison with Boston criteria and N-terminal prohormone brain natriuretic peptide. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine 2010;17:18–26.

190. Rutten JH, Steyerberg EW, Boomsma F, et al. N-terminal pro-brain natriuretic peptide testing in the emergency department: beneficial effects on hospitalization, costs, and outcome. Am Heart J 2008;156:71-7.

191. Sanz MP, Borque L, Rus A, Vicente B, Ramirez Y, Lasa L. Comparison of BNP and NT-proBNP assays in the approach to the emergency diagnosis of acute dyspnea. Journal of clinical laboratory analysis 2006;20:227–32.

192. Klemen P, Golub M, Grmec S. Combination of quantitative capnometry, N-terminal pro-brain natriuretic peptide, and clinical assessment in differentiating acute heart failure from pulmonary disease as cause of acute dyspnea in pre-hospital emergency setting: study of diagnostic accuracy. Croatian medical journal 2009;50:133–42.

193. Santaguida PL, Don-Wauchope AC, Ali U, et al. Incremental value of natriuretic peptide measurement in acute decompensated heart failure (ADHF): a systematic review. Heart Fail Rev 2014;19:507-19.

194. Maisel A, Mueller C, Adams K, Jr., et al. State of the art: using natriuretic peptide levels in clinical practice. Eur J Heart Fail 2008;10:824-39.

195. Madamanchi C, Alhosaini H, Sumida A, Runge MS. Obesity and natriuretic peptides, BNP and NT-proBNP: mechanisms and diagnostic implications for heart failure. Int J Cardiol 2014;176:611-7.

196. Maisel AS, Peacock WF, McMullin N, et al. Timing of immunoreactive B-type natriuretic peptide levels and treatment delay in acute decompensated heart failure: an ADHERE (Acute Decompensated Heart Failure National Registry) analysis. J Am Coll Cardiol 2008;52:534-40.

197. Galasko GIW, Barnes SC, Collinson P, Lahiri A, Senior R. What is the most cost-effective strategy to screen for left ventricular systolic dysfunction: natriuretic peptides, the electrocardiogram, hand-held echocardiography, traditional echocardiography, or their combination? European Heart Journal 2005;27:193-200.

198. Tschope C, Kasner M, Westermann D, Gaub R, Poller WC, Schultheiss HP. The role of NT-proBNP in the diagnostics of isolated diastolic dysfunction: correlation with echocardiographic and invasive measurements. Eur Heart J 2005;26:2277-84.

199. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021;42:3599-726.

200. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.

201. Levin ER, Gardner DG, Samson WK. Natriuretic peptides. New England Journal of Medicine 1998;339:321-8.

202. Tsutamoto T, Wada A, Sakai H, et al. Relationship between renal function and plasma brain natriuretic peptide in patients with heart failure. Journal of the American College of Cardiology 2006;47:582-6.

203. McCullough PA, Duc P, Omland T, et al. B-type natriuretic peptide and renal function in the diagnosis of heart failure: an analysis from the Breathing Not Properly Multinational Study. Am J Kidney Dis 2003;41:571-9.

204. Takami Y, Horio T, Iwashima Y, et al. Diagnostic and prognostic value of plasma brain natriuretic peptide in non-dialysis-dependent CRF. Am J Kidney Dis 2004;44:420-8.

205. Kumar A, Cannon CP. Acute coronary syndromes: Diagnosis and management, part II. Mayo Clin Proc 2009;84:1021-36.

206. Kumar A, Cannon CP. Acute coronary syndromes: diagnosis and management, part I. Mayo Clin Proc 2009;84:917-38.

Appendix

Copyright licences for figures used in Chapter 2

Appendix

Figure 1 and 2 on Chapter 2

Value added tax	0.00 GBP
Total	0.00 GBP
Title	Improving diagnosis in acute cardiac care using statistical machine learning
Institution name	University of Edinburgh
Expected presentation date	Mar 2022
Portions	Figure 3 and Figure 4 of the clinical classification of myocardial infarction section
Requestor Location	Mr. Dimitrios Doudesis Flat 7, 34 Peffer Bank Edinburgh, EH16 4FE United Kingdom Attn: Mr Dimitrios Doudesis
Publisher Tax ID	GB125506730
Total	0.00 GBP

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Figure 3 on Chapter 2

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 17, 2022

This Agreement between Mr. Dimitrios Doudesis ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5227650668245
License date	Jan 14, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Disease Primers
Licensed Content Title	Heart failure with reduced ejection fraction
Licensed Content Author	Michelle W. Bloom et al
Licensed Content Date	Aug 24, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Improving diagnosis in acute cardiac care using statistical machine learning
Institution name	University of Edinburgh
Expected presentation date	Mar 2022
Portions	Figure 1
Requestor Location	Mr. Dimitrios Doudesis Flat 7, 34 Peffer Bank Edinburgh, EH16 4FE United Kingdom Attn: Mr Dimitrios Doudesis
Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also

seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that

affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Figure 4 on Chapter 4

Title of your work	Improving diagnosis in acute cardiac care using statistical machine learning
Publisher of your work	University of Edinburgh
Expected publication date	Mar 2022
Permissions cost	0.00 GBP
Value added tax	0.00 GBP
Total	0.00 GBP
Title	Improving diagnosis in acute cardiac care using statistical machine learning
Institution name	University of Edinburgh
Expected presentation date	Mar 2022
Portions	Figure 3
Requestor Location	Mr. Dimitrios Doudesis Flat 7, 34 Peffer Bank Edinburgh, EH16 4FE United Kingdom Attn: Mr Dimitrios Doudesis
Publisher Tax ID	GB125506730
Billing Type	Invoice

Billing Address	Mr Dimitrios Doudesis Flat 7, 34 Peffer Bank Edinburgh, United Kingdom EH16 4FE Attn: Mr Dimitrios Doudesis
Total	0.00 GBP

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment

terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions: -Permission is granted for the verbatim reuse of the Material in your academic thesis/ dissertation provided it is not published commercially. -Inclusion under a Creative Commons license or any other Open Access license allowing onward reuse is prohibited. -The credit line and any copyright/ disclaimer notices provided by OUP shall be included on the same page as the Material: Collet JP, Thiele H, Barbato E, et al, 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC), European Heart Journal 2020; 42 (14): 1289-1367 doi:10.1093/eurheartj/ehaa575. Reprinted by permission of Oxford University Press on behalf of the European Society of Cardiology.

v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.