Imperial College London

Big data analytics in cardiovascular sciences

Dr Amit Kaura

PhD

Imperial College London
National Heart and Lung Institute

Statement of originality

I, Amit Kaura, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright declaration

Acknowledgements

Financial support

I am in receipt of a British Heart Foundation Clinical Research Training Fellowship (FS/20/18/34972; awarded 2020) and a National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC) grant (P74354; awarded 2018). The NIHR Health Informatics Collaborative (HIC) is led and funded by NIHR Imperial BRC, with support from NIHR Oxford BRC, NIHR University College London Hospitals BRC, NIHR Guy's & St Thomas' BRC, NIHR Cambridge BRC, NIHR Bristol BRC, NIHR Birmingham BRC, NIHR Leeds BRC, NIHR Leicester BRC, NIHR Manchester BRC, NIHR Royal Marsden BRC, and NIHR Southampton BRC.

Contribution of other researchers

The NIHR HIC data flow (Chapter 2) and database design (Chapter 3) was developed in conjunction with the Imperial College Healthcare NHS trust data science team, led by Mr Benjamin Glampson and Mr Abdulrahim Mulla. The data science team led the development procedures for local data collection, and for developing the secure research database to receive data from primary clinical systems within each NHS Trust contributing data. The contribution of researchers for Chapters 4 to 7 are summarised in the table underneath by their initials.

	Chapter 4	**Chapter 5**	**Chapter 6**	**Chapter 7**
Study	**CRP-RISK**	**SENIOR-NSTEMI**	**TROP-RISK**	**TROP-AF**
Conceived hypothesis	AK, AH, RK, JM	AK, JM	AK, VP, DPF, JM	AK, AA
Study protocol	AK, AH, RK	AK, JACS, AT, JM	AK, VP, DPF, JM	AK, AA
Programming to curate data	AK, BG, JD, AM, JO			
Coding for data analyses	AK	AK, AT	AK	AK
Verification of coding	ADS	JACS, SA	VP, DPF	

The key for researcher initials, as well as their roles is summarised in the following table:

AA	Ahran Arnold	Cardiology Registrar	Imperial College London
ADS	Dr Anoop D Shah	Clinical Lecturer in Clinical Pharmacology	University College London
AH	Dr Adam Hartley	Clinical Research Fellow	Imperial College London
AK	Dr Amit Kaura	Clinical Research Fellow	Imperial College London
AM	Mr Abdulrahim Mulla	Data engineer	Imperial College London
AT	Dr Adam Trickey	Medical statistician	University of Bristol
BG	Mr Benjamin Glampson	Research informatics programme manager	Imperial College London
DPF	Professor Darrel P Francis	Professor of Cardiology	Imperial College London
JACS	Professor Jonathan Sterne	Professor of Epidemiology and Statistics	University of Bristol
JD	Professor Jim Davies	Professor of software engineering	University of Oxford
JM	Professor Jamil Mayet	Professor of Cardiology	Imperial College London
JO	Mr Joe Omigie	Data analysts	King's College London
RK	Dr Ramzi Khamis	Clinical Senior Lecturer in Cardiology	Imperial College London
SA	Dr Sam Abbott	Medical statistician	University of Bristol
VP	Dr Vasileios Panoulas	Cardiology Consultant	Imperial College London

Other support

I would like to thank my PhD supervisors Professor Jamil Mayet, Professor Jonathan Sterne and Professor Darrel Francis for their advice and support in preparing my fellowship application and during my PhD. I would also like to thank my wife, Rina ♥, daughter, Ariya ♥, and my family and friends for supporting me throughout my PhD.

Contents

List of abbreviations

A&E	accident and emergency
AAA	abdominal aortic aneurysm
ACE-I	angiotensin converting enzyme inhibitor
ACS	acute coronary syndrome
ARB	angiotensin receptor blocker
ARNI	angiotensin receptor neprilysin inhibitor
AUROC	area under the receiver operating characteristic curve
AF	atrial fibrillation
APC	admission patient care
BB	beta-blocker
BRC	biomedical research centre
BNP	brain natriuretic peptide
CABG	coronary artery bypass grafting
CAD	coronary artery disease
CAMI	CRP apheresis in Acute Myocardial Infarction
CANTOS	Canakinumab Anti-inflammatory Thrombosis Outcome Study
CATO	Clinical Academic Training Office
CI	confidence interval
CIRT	Cardiovascular Inflammation Reduction Trial
COLCOT	Colchicine Cardiovascular Outcomes Trial
CPRD	Clinical Practice Research Datalink
CRP	C-reactive protein
CRT	cardiac synchronisation therapy
cTn	cardiac troponin
CV	cardiovascular
CVD	cardiovascular disease
dm+d	Dictionary of Medicines and Devices

ECDS	Emergency Care Data Set
ECG	electrocardiogram
EF	ejection fraction
EHR	electronic health record
ETL	extract, transform, load
GP	general practitioner
GSTT	Guy's and St Thomas' NHS Foundation Trust
HES	hospital episode statistics
HF	heart failure
HFmrEF	heart failure with mid-range ejection fraction
HFpEF	heart failure with preserved ejection fraction
HFrEF	heart failure with reduced ejection fraction
HR	heart rate
hsCRP	high-sensitivity C-reactive protein
HIC	Health Informatics Collaborative
HR	hazard ratio
ICD	International Statistical Classification of Diseases and Related Health Problems
ICHNT	Imperial College Healthcare NHS Trust
IDI	integrated discrimination improvement
IG	information governance
IHD	ischaemic heart disease
IL	interleukin
IPD	individual patient data
IPT	inverse probability of treatment
IPTW	inverse probability of treatment weighted
IQR	interquartile range
LoDoCo2	Low Dose Colchicine for secondary prevention of cardiovascular disease

LV	left ventricular
LVEF	left ventricular ejection fraction
LVEDd	left ventricular end diastolic diameter
LVEDs	left ventricular end systolic diameter
LVFS	left ventricular fractional shortening
MACE	major adverse cardiovascular events
MINAP	Myocardial Ischaemia National Audit Project
MR	mineralocorticoid receptor
NCAP	National Cardiac Audit Programme
NHS	National Health Service
NSTEMI	non-ST elevation myocardial infarction
NICOR	National Institute for Cardiovascular Outcomes Research
NIHR	National Institute for Health Research
NRI	net reclassification index
NWL	North West London
OMT	optimal medical therapy
ONS	Office of National Statistics
OP	outpatient
OPCS-4	Office of Population Censuses and Survey's version 4
OUH	Oxford University Hospitals NHS Foundation Trust
PCI	percutaneous coronary intervention
PICO	population, intervention, control, and outcomes
PS	propensity score
RCT	randomised controlled trials
REC	Research Ethics Committee
SFTP	SSH File Transfer Protocol
SNOMED-CT	Systematised Nomenclature of Medicine - Clinical Terms
STEMI	ST elevation myocardial infarction

SUS	Secondary Uses Service
Tn	troponin
Tn +	troponin positive
Tn -	troponin negative
UCL	University College London Hospitals NHS Foundation Trust
ULN	upper limit of normal
VF	ventricular fibrillation
VT	ventricular tachycardia
WCC	white cell count
WHO	World Health Organisation
XML	Extensible Markup Language

List of figures

List of tables

Abstract

Introduction

It has been challenging for researchers to access granular electronic health record (EHR) data at scale in England. The National Institute for Health Research (NIHR) Health Informatics Collaborative (HIC) enables the sharing of routine EHR data across NHS hospitals for research. One emerging prospect is to use big data to traverse the translational spectrum.

As an example of an early discovery phase study, I assessed the effect of invasive versus non-invasive management on the survival of patients aged 80 years or older with non-ST elevation myocardial infarction (NSTEMI) (SENIOR-NSTEMI Study).

As an example of a later implementation phase study, I determined the relationship between the full spectrum of troponin level and mortality in patients in whom troponin testing was performed for clinical purposes (TROP-RISK Study).

Methods

Five NHS Trusts contributed data: Imperial, University College London, Oxford, King's and Guy's and St Thomas'. Microsoft SQL was used to develop a dataset of 257,948 consecutive patients who had a troponin measured between 2010 and 2017. Phenotypically detailed data were extracted, including patient demographics, blood tests, procedural data, and survival status. All studies conducted were retrospective cohort studies.

For the SENIOR-NSTEMI Study, eligible patients were 80 years or older who were diagnosed with NSTEMI. Mortality hazard ratios were estimated comparing invasive with non-invasive management.

For the TROP-RISK Study, the relation between peak troponin level and all-cause mortality was modelled using multivariable adjusted restricted cubic spline Cox regression analyses.

Results

For the SENIOR-NSTEMI Study, 2672 patients with NSTEMI were included who had a median age of 85 (interquartile range (IQR) 82-89) years of whom 59.8% received non-invasive management. During a median follow-up of 2.7 (IQR 1.0-4.5) years, the adjusted cumulative five-year mortality was 40% in the invasive and 63% in the non-invasive group (hazard ratio 0.52, 95% confidence interval 0.43-0.62).

For the TROP-RISK Study, during a median follow-up of 1198 days (IQR 514-1866 days), 55,850 (21.7%) deaths occurred. There was an unexpected inverted U-shaped relation between troponin level and mortality in acute coronary syndrome (ACS) patients (n=120,049). The paradoxical decline in mortality at very high troponin levels may be driven in part by the changing case mix as troponin levels increase; a higher proportion of patients with very high

troponin levels received invasive management.

Conclusion

Routinely collected EHR data can be aggregated across multiple sites to create highly granular datasets for research which can be used to answer research questions that cross the translational spectrum.

The SENIOR-NSTEMI Study demonstrates a survival advantage of invasive compared with non-invasive management of NSTEMI patients aged 80 years or older, who were underrepresented in previous trials.

In the TROP-RISK Study, the inverted U-shaped relationship between troponin level and mortality in ACS patients demonstrates that assembling sufficiently large datasets can cast light on patterns of disease that are impossible to adequately define in single centre studies.

1. Introduction

1.1 Outline of thesis

The National Health Service (NHS) generates vast amounts of data, providing significant opportunities for research innovation by harnessing the value of big data. This thesis will explore the methodology and infrastructure developed to generate big data using hospital EHR system data across five NHS Trusts. I will subsequently be using the datasets generated to answer research questions that cover the translational spectrum from an early phase of discovery to a later implementation phase.

The first chapter of this thesis introduces the use of big data in healthcare research and reviews the strengths and limitations of existing data sources. Chapter 2 describes the NIHR HIC, a programme of infrastructure development across NHS Trusts to share hospital EHR data for research. The NIHR HIC dataset was the data source used for the studies that comprise this thesis. Chapter 3 describes the validation work and methods involved with creating a research ready dataset for statistical analysis.

In Chapter 4, I demonstrate how big data can contribute to early scientific advancement in cardiovascular medicine through discovery of underlying disease mechanisms. I evaluate whether a mildly elevated high-sensitivity C-reactive protein (hsCRP) level was associated with mortality risk, beyond troponin level, in patients with suspected acute coronary syndrome.

In Chapter 5, I further explore the role of big data for early discovery phase translational research by assessing the effect of an intervention in population subgroups that were underrepresented in previous clinical trials. I estimate the effect of invasive versus non-invasive management on the survival of patients aged 80 years or older with non-ST elevation myocardial infarction (NSTEMI).

In Chapter 6, I demonstrate how big data can be used for late translational research through assessment of the epidemiology of cardiovascular disease across unselected populations. I evaluate the relationship between the full spectrum of troponin level and mortality in all patients in whom troponin testing has been performed for clinical purposes.

In Chapter 7, I further demonstrate how similar epidemiological studies can be performed in clinically relevant subgroups. I investigate the relationship between troponin level, coronary angiography, and all-cause mortality in patients presenting to hospital with atrial fibrillation.

The final discussion (Chapter 8) summarises the findings of this thesis with recommendations for clinical practice and future research involving EHR data.

1.2 Using big data for research

Big data is a term that refers to large-volume data sets that are rapidly generated and subsequently transmitted from a variety of different sources. Interactions between patients and the healthcare system generates a wealth of electronic health record (EHR) data that are stored in distinct healthcare clinical information systems. EHR data is held in a number of different settings: primary care-based, secondary care-based, and registries, which may be disease- or procedure-based. Beyond their primary purpose for direct patient care, EHR data are increasingly being used for translational research.[1] In healthcare research, there is potential to use big data to answer research questions that cover the full translational spectrum from an early phase of discovery to a later implementation phase.[1]

Big data can contribute to 'early translational research' in cardiovascular medicine through discovery of underlying disease mechanisms and by assessing the effect of an intervention in population subgroups that were underrepresented in previous clinical trials. For example, in a cohort of over 1 million adults aged 30 years of age or older and initially free from cardiovascular disease, there was a strong association between baseline diastolic blood pressure and future risk of abdominal aortic aneurysm (AAA), compared with the lack of association with systolic blood pressure.[2] This finding into the aetiology of AAA can be used to inform screening and prevention measures in future trials.

In 'late translational research', big data can be used to provide useful information on the epidemiology of cardiovascular disease across the whole population and in clinically relevant sub-groups of disease. Big data can be used to provide population-based measures of disease burden and provide insight on the most effective management strategies. The statistical size and phenotypic resolution of big data in healthcare can be used to investigate relevant subgroups of disease which were difficult to capture with conventional consented cohorts. For example, big data has previously been used to investigate population trends in the incidence and outcomes of acute myocardial infarction.[3]

1.3 Electronic health records for research

To date, it has been challenging, in England, for researchers to access secondary care-based EHR data across multiple centres at scale.[1] Secondary care-based EHR has the potential to provide detailed phenotypic data on patients and their disease profile, including demographics, blood laboratory values, imaging data, procedural data, and clinical notes.[4-6] EHRs hold both 'structured' and 'unstructured' electronic data which are generated during routine clinical care for each patient. Structured EHR data are data stored in defined fields within a patient's record. These include data on patient demographics, blood test results and patient diagnosis using statistical classification systems [such as International Statistical Classification of Diseases and Related Health Problems (ICD-10)[7]]. Unstructured clinical data such as patient clinical notes, discharge summaries, and procedural or imaging reports are recorded in a patient's EHR as unformatted text.[1]

There is a need, in England, for data from secondary care-based EHR systems from multiple sites to be linked together using integration techniques to create a granular-level longitudinal dataset of an individual's health status over time.

1.4 Improving safe access to data for research

The Secretary of State for Health and Social Care's 'Data saves lives' initiative[8] and the Goldacre Report[9], *'Better, broader, safer: using health data for research and analysis',* are both initiatives in the NHS with an aim to invest more in data science and health data. It is apparent that the UK has one of the most advanced data ecosystems worldwide, with some datasets able to link primary care, secondary care and mortality data. However, as highlighted in the Goldacre Report, these ecosystems are becoming increasingly complex with the addition of new data sources and apps. There is a concern regarding how well they will be integrated in current data ecosystems. Additional concerns surround public mistrust in technology, data quality concerns and data access issues.

There is an emphasis throughout the report on the role of Trusted Research Environments (TREs) as a means of providing remote access to health data to approved people for approved purposes.[9] The aspiration is that the TREs will promote transparent data access with a drive for innovation with the facilitation of remote working for efficient working practices. The TRE owners will be the data controllers to prevent duplication of effort for research and overcrowding in terms of data access.

Data quality and understanding the context of data recording are other issues that need to be addressed. This will allow researchers to correctly interpret the data. Furthermore, improvements in clinical coding are required to fully benefit from the use of routine clinical data for research. There is value in clinician coding, especially of the primary reason for hospital admission, as there will be a clinical perspective to the underlying diagnosis. As it stands, clinical coding is usually performed by non-medically trained administrative staff. Additional strategies to improve the quality of clinical coding include the use of specific disease codes as part of a pay- for- performance scheme, however, there is difficulty in making such incentives universal. The future may involve the use of natural language processing to extract relevant information for coding, although there would need to be assurances that patient privacy would be preserved.

1.5 Data protection

All Trusts who contribute EHR data for research have the responsibility to ensure the data is protected as the process of using EHR data for research will involve working with personal and/or confidential information.

The UK General Data Protection Regulation (GDPR)[10] was a law that was introduced after the UK left the European Union. All organisations, including the NHS, have a lawful requirement to comply with this regulation. The UK GDPR and Data Protection Act (DPA)[11] 2018 are usually consulted in parallel as the DPA 2018 covers data processing that does not fall within the law in the EU and provisions for national security. The DPA 2018 outlines six principles that need to be followed when data is being processed. Data must be:

1) accurate, and where relevant, kept up-to-date
2) kept for as long as necessary for the intended purpose
3) lawfully processed in a transparent manner
4) collected for a specific purpose and is processed in accordance with this purpose
5) adequate and not excessive in relation to the intended purpose of collection
6) processed in accordance with appropriate security measures

The NHS and data initiatives for research within the NHS should have a named Data Protection Offer who ensures the DPA principles are satisfied. Additionally, there needs to be processes in place for assessing potential data breaches and acting upon incidents.

1.6 Lawful basis for processing data

Under UK GDPR law[10], as sanctioned by the DPA 2018[11], there is a requirement for a lawful basis for processing personal data. There are six lawful bases which hold equal weight in terms of importance. If a new data process if being implemented, a Data Protection Impact Assessment (DPIA) needs to be completed. At least one of the following must apply before processing personal data:

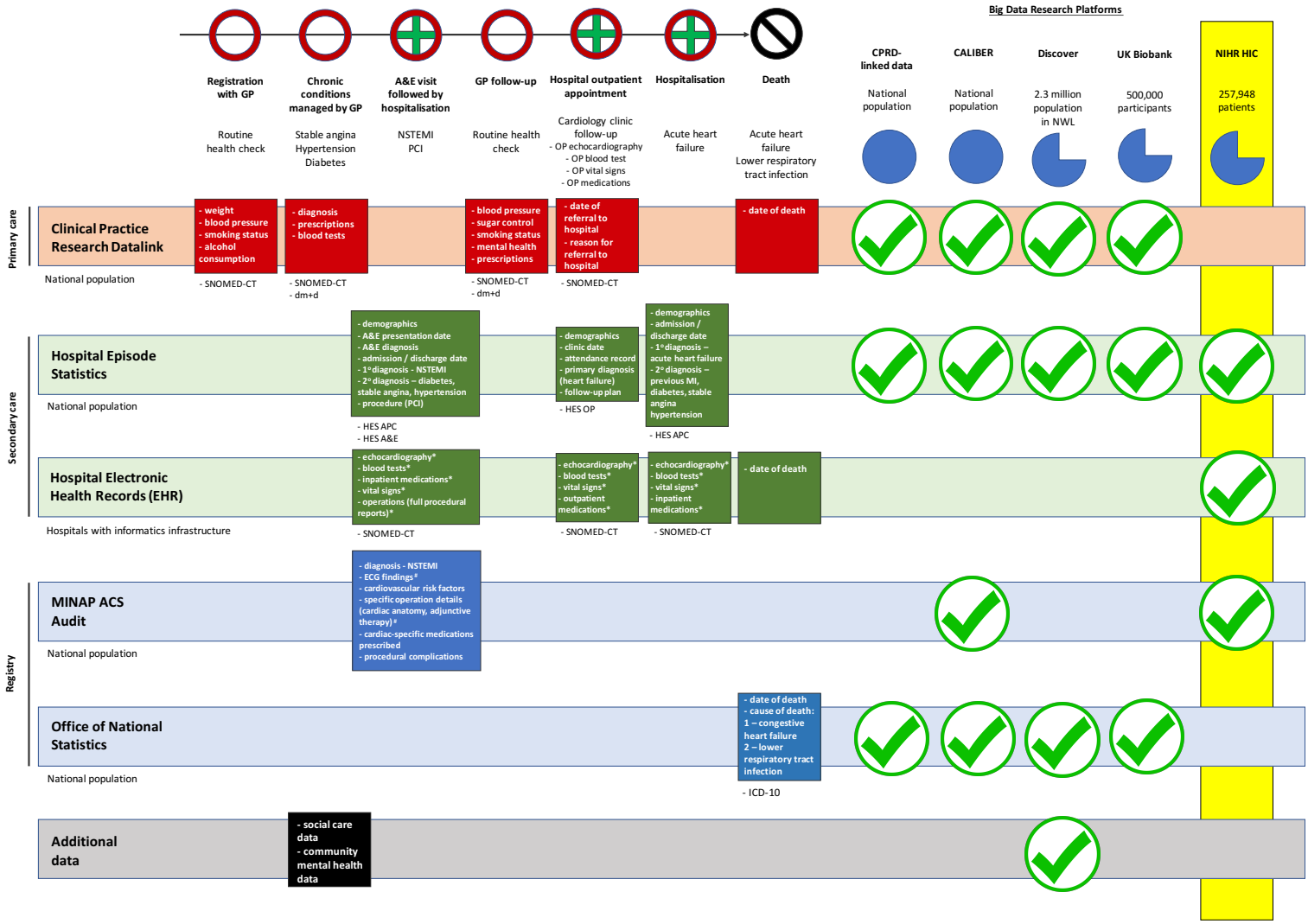
1. Consent – data processing is performed after an individual has provided 'informed' consent for their personal data to be used for a particular purpose.
2. Contract – data processing is necessary for an agreed contract.
3. Legal obligation – data processing is necessary in compliance with law, such as for a Court Order or to assist a regulatory body such as the General Medical Council.
4. Vital interests – data processing is crucial to protect life, although an attempt should be made to firstly ask for consent
5. Public task – data processing is necessary to perform a particular task which is in the public interest or for the NHS Trust's official function. Healthcare services that are provided by the Trust fall under this lawful basis.
6. Legitimate interests – data processing is necessary for the interests of either the Trust or a third party, unless there is a legitimate reason to protect the personal interests of an individual.

Big data research involves accessing confidential patient information without consent. This would involve an application to the Confidentiality Advisory Group (CAG). The CAG independently provides advice on the use of confidential patient data.

1.7 Existing electronic health record data sources

Primary care, secondary care and registry EHR data sources individually allow researchers to capture information for patients during different types of clinical encounters, including routine health check-ups with the general practitioner (GP), management of chronic conditions by the GP, accident and emergency (A&E) visits, hospital admissions, hospital outpatient clinic visits and recording mortality (**Figure 1.1**). The following sections will summarise the data sources currently available in England for research and the data they hold for the different types of clinical encounters, as outlined in **Figure 1.1**.

Figure 1.1 - Electronic health record data sources in England for capturing the patient journey from GP registration to death



A&E, accident and emergency; ACS, acute coronary syndrome; APC, admission patient care; CPRD, Clinical Practice Research Datalink; ECG, electrocardiogram; EHR, electronic health record; GP, general practice; HES, hospital episode statistics; ICD-10, International Statistical Classification of Diseases and Related Health Problems Version 10; NIHR HIC, National Institute for Health Research Health Informatics Collaborative; OP, outpatient; MINAP, Myocardial Ischaemia National Audit Project; NSTEMI, non-ST elevation myocardial infarction; PCI, percutaneous coronary intervention.

1.7.1 Clinical Practice Research Datalink

The Clinical Practice Research Datalink (CPRD) collects anonymised patient data from a UK-wide network of over 2,000 primary care GP practices.[12] These cover 60 million patients of which approximately 16 million are registered active patients as of late 2021. The EHR data is captured from GP practices using the EMIS or Vision software systems.[12] CPRD GOLD contains data contributed by GP practices using Vision EHR software[13], while CPRD Aurum contains similar data from practices using EMIS software.[14] Data are recorded prospectively as part of routine clinical care (**Table 1.1**). In CPRD, non-prescription data, including diagnoses are coded using a combination of SNOMED-CT, Read version 2, or local software-specific (EMIS or Vision) codes. Data on prescriptions are coded using the Dictionary of Medicines and Devices (dm+d) codes.[13]

Although secondary care data, including new patient diagnoses (coded using ICD-10 classification[7]), can be recorded manually by GPs, this information is often not recorded in primary care records. Furthermore, additional information from secondary care may only be available as unstructured free text. For example, blood test results from secondary care following hospitalization may be recorded on discharge letters received by the GP. In **Section 1.8**, I will discuss how primary and secondary care data sources can be linked to provide additional information on patient pathways.

Due to underlying differences in software structure, the data housed in CPRD GOLD and CPRD Aurum are not always comparable.[15] While there is preliminary guidance available for researchers on the differences between the two databases, a standardised dataset is not readily available.

Table 1.1 - Data files supplied by the CPRD database

File type	What it holds	Example of contents
Patient	Demographic and registration status of patients	Patient identifier, month/year of birth, registration status, death date, transfer out date
Practice	Practice administrative data	Practice identifier, geographical region, date practice became 'Up to standard' (when the practice achieves high enough standard of data completeness), last data collection date

Staff	Information about the staff members entering data	Staff identifier, gender, role
Consultation	Administrative information about the consultation	Date of clinical event, date of data entry, type of consultation (telephone, home visit, practice visit), staff identifier and duration of consultation
Clinical	Clinical data regarding medical history	Symptoms, signs and diagnoses.
Additional Clinical Details (ACD)	Specific data about a clinical event	Type of information held, called an 'entity', which provide specific clinical details relating to that entity, e.g. alcohol intake and cigarette consumption
Referral	Details on referrals to secondary care or specialists	Method of referral, referral specialty, urgency of referral
Immunisation	Data associated with immunisations	Reason for immunisation, type, stage, status and the compound used
Test	Test results	Type of test, result, normal range of result, unit of measure
Therapy	Information about therapies including medications and appliances	British National Formulary code, quantity of product, dose, pack size, number of days prescribed

1.7.2 Hospital Episode Statistics

Hospital Episode Statistics (HES) is a database which contains details on A&E attendances (HES A&E), inpatient hospital admissions (HES Admitted Patient Care, APC), and outpatient appointments (HES OP) at NHS Hospitals in England.[16,17] Healthcare providers collect the clinical data locally and is submitted to the Secondary Uses Service (SUS). Data is submitted by the majority of the NHS hospital Trusts on a monthly basis. The submissions to SUS from each site are consolidated and compiled as HES data. It is subsequently validated before making the information available in the HES database. Data quality reports are generated at various stages in the processing cycle.[17]

Different coding systems are used through the different HES datasets. In the HES APC and HES OP datasets, diagnoses are coded using the World Health Organisation International Statistical Classification of Diseases and Related Health Problems (ICD) clinical classification system. The ICD-9 version was used to code entries until 1995, which has since been superseded by the ICD-10 version.[7] In both these datasets, operations and procedures are coded using the Office of Population Censuses and Survey's version 4 clinical classification (OPCS-4).[18]

In the HES A&E dataset, bespoke coding classification systems are used within the data field recording the reason for A&E presentation (AEPATGROUP) and the diagnosis made in A&E (DIAG2_NN).[17] The NHS is currently moving towards using Systematised Nomenclature of Medicine - Clinical Terms (SNOMED) nomenclature which may lead to changes in HES coding in the future.

1.7.2.1 Accident and emergency visits

The HES A&E database provides details relating to the attendance, diagnosis (not ICD-10 coded), investigations (not OPCS coded) and treatment (drug prescribing not recorded) undertaken in the A&E setting.[17,19] A summary of the key fields beyond patient demographics, is shown in **Table 1.2**.

The data captured in the HES A&E dataset create a relatively high-level picture of the patient's journey in A&E. For example, the INVEST2_nn field captures that cardiac enzymes were measured in A&E, however, there is no field which captures the result of the blood test. Similarly, whilst the TREAT2_N field captures that medications were administered, the exact medications prescribed are not recorded.

Since the 2019-2020 financial year, a new national dataset for emergency care, the Emergency Care Data Set (ECDS), has replaced the HES A&E dataset across England. The ECDS was introduced to enable capture the care provided in the hospital emergency setting with more granularity.[17]

While the overall coverage of HES A&E data has increased over the past few years, data completeness for some of the key variables, such as duration to treatment, remain an issue.[19] Furthermore, provisional HES A&E data are publications of HES data on a monthly basis. The data may be incomplete, with counts from this provisional data being lower than those generated in the final dataset from the same time period. Clinical data may also not be complete at the time of submission of data to the SUS.

Table 1.2 - Data fields supplied by the HES A&E database

Field	Field name	Description	Values	Format
Attendance				
AEARRIVALMODE	Arrival Mode	The mode by which a patient arrived at an A&E department	01 = Brought in by ambulance	String
AEATTENDCAT	Attendance category	Initial or follow-up attendance within a particular A&E department	01 = First A&E attendance	String
AEPATGROUP	Patient Group	A coded classification to identify the reason for an Accident and Emergency Episode.	80 = Other than above	String
AEATTENDDISP	Attendance Disposal	The way in which an A&E attendance might end	01 = Admitted to hospital bed	String
Diagnosis				
DIAG2_NN	Diagnosis Code - 2 Character	The A&E diagnosis description at 2-character level covering the diagnosis condition.	20 = Cardiac conditions	String
DIAG3_NN	Diagnosis Code - 3 Character	The A&E diagnosis description at 3-character level, covering the diagnosis condition and the sub-	201 = Cardiac conditions - myocardial ischaemia & infarction	String

		analysis.		
Investigation				
INVEST2_*nn*	A&E Investigation - 2 Character	The A&E investigation description	02 = Electrocardiogram 16 = Cardiac enzymes	String
Treatment				
TREAT2_N	A&E Treatment: 2 character	The A&E treatment description at 2-character level covering the treatment.	51 = Medication administered 40 = Supplemental oxygen 30 = Recording vital signs 12 = Intravenous cannula	String
TREAT3_N	A&E Treatment: 3 character	The A&E treatment description at 3-character level, covering the treatment and the sub-analysis.	511 = Medication administered - oral	String
Patient Pathway				
RTTPERSTAT	RTT period status	The status of an activity, or anticipated activity, for the referral to treatment period.	12 = Consultant referral	String

1.7.2.2 Hospital admission patient care data

HES APC includes data on episodes of treatment that require inpatient admission to NHS hospitals in England.[20,21] The dataset includes high-level data on the complete care pathway during hospitalisation. A summary of the key fields beyond patient demographics, is shown in **Table 1.3**.

Hospitalisations (alternatively known as spells) refers to the duration of hospital stay from admission to discharge. Episodes in the HES APC dataset are defined as time periods within a hospitalisation. Each episode corresponds

to a continuous period of care provided by one consultant under one health care provider.[21] If a patient is transferred between two consultants during a single hospitalisation, this translates to two episodes of care. For each episode of care, up to 20 separate diagnoses (using ICD-10[7]) and 24 procedures or operations (using OPCS-4[18]) may be recorded.

There are limitations with the scope of data recorded for patients admitted to hospital. Data on drugs prescribed and vital signs (such as heart rate, blood pressure and oxygen saturations), are not available in HES APC. In addition, there are no variables which record the blood tests performed during hospitalisation and their results.

Similar to the HES A&E data, the main issues with HES APC data relates to the data being incomplete for a small proportion of patients, with counts from this provisional data being lower than those generated in the final dataset from the same time period.

Table 1.3 - Data fields supplied by the HES APC database

Field	Field name	Description	Values	Format
Administrative				
ADMIDATE	Admission Date (Hospital Provider Spell)	The date the patient was admitted to hospital at the start of a hospital spell. ADMIDATE is recorded on all episodes within the spell.	2022-01-19	Date(YYYY-MM-DD)
ADMIMETH	Method of Admission	A code to identify how the patient was admitted to hospital.	21 = Accident and emergency or dental casualty department of the Health Care Provider	String
MAINSPEF	Main Specialty	The specialty under which the consultant is contracted.	320 = Cardiology	String
DISDATE	Date of Discharge	The date on which the	2022-01-24	Date(YYYY-MM-DD)

		patient was discharged from hospital. It is only populated on the final episode of the spell.		
Clinical information				
DIAG_4_CONCAT	Concatenated Diagnosis Codes - 4 Character	A concatenated string of all 4 character diagnoses (DIAG_4_nn) from the record, separated by commas with no spaces.	I21.4, E11, E10 Acute subendocardial myocardial infarction, Non-insulin-dependent diabetes mellitus, Essential (primary) hypertension	String
OPDATE_nn	Date of Procedure	The date of the procedure recorded in the respectively numbered Procedure Code (OPERTN_nn) field.	2022-01-20	Date(YYYY-MM-DD)
OPERTN_4_nn	Procedure Code - 4 Character	The first four characters of the procedure code (OPERTN_nn).	K49.1 Percutaneous transluminal balloon angioplasty of one coronary artery	String

1.7.2.3 Hospital outpatient data

Information on outpatient appointments in England are recorded in the HES OP dataset.[22] A summary of the key fields beyond patient demographics, are shown in **Table 1.4**.

Similar to the HES datasets discussed above, the granularity of data recorded in the HES OP dataset is also limited. Vital signs, blood tests and drug prescriptions data are all relevant for outpatient care, however, are not recorded in the HES OP dataset.

With regards to data quality, there is a range in the level of completeness of the variables recorded, with fields such as ‘attendance type’ and ‘main specialty’ having high rates of completeness (>98%), and ‘primary diagnosis’ and ‘main procedure’ having low levels of completeness, with only 5% and 26%, respectively.[22]

Table 1.4 - Data fields supplied by the HES OP database

Field	Field name	Description	Values	Format
Administrative				
APPTDATE	Appointment Date	The date of the appointment	2022-05-14	Date(YYYY-MM-DD)
MAINSPEF	Main Specialty	The specialty under which the consultant is contracted.	320 = Cardiology	String
ATENTYPE	Attendance Type	A code to identify if the attendance occurred and whether it was a first or subsequent appointment.	1 = Attended first appointment	String
Clinical information				
DIAG_4_*nn*	Diagnosis Code - 4 Character	The first four characters of the diagnosis code (DIAG_nn)	I50.1 = Left ventricular failure	String
OPERTN_4_*nn*	Procedure Code - 4 Character	The first four characters of the procedure code (OPERTN_*nn*).	- = No operation performed	String

1.7.3 Disease specific audit registry data

There are six separate domains of cardiovascular care that comprise the National Cardiac Audit Programme (NCAP) within the National Institute for Cardiovascular Outcomes Research (NICOR).[23] It involves a partnership of clinicians, statisticians, data scientists and researchers who work together to examine and improve the service delivered to patients admitted to hospital with cardiovascular disease and their outcomes. The Myocardial Ischaemia National Audit Project (MINAP)[24] will be discussed for the patient scenario outlined in **Figure 1.1**.

1.7.3.1 MINAP acute coronary syndrome audit

MINAP is a national disease registry capturing all admissions to hospital with acute coronary syndrome (ACS) across England.[24] At present, MINAP has recorded data for over 1.5 million patients. Data are collected within 130 unique fields that cover the patient journey from the time of symptom onset to the point of patient discharge from hospital.[25] The data fields are categorised in to patient demographics, medical history, clinical assessment, investigations and management during inpatient admission, with a focus on diagnosing and treating ACS. Drug therapy is also recorded pre-admission, during inpatient stay and at discharge from hospital. **Table 1.5** summarises the information coded for the patient scenario in **Figure 1.1**. Whilst not all the 130 data fields are covered in **Table 1.5**, some of the fields that are commonly used for research are displayed. They also provide an example of the level of granularity of the information recorded in the MINAP database.

Although the fields provide information of the patient journey, they are focused on patients with confirmed ACS. They do not capture patients with suspected ACS, characterised by the request of a troponin. Although the level of detail for ACS patients is granular, there are important variables that are missing when carrying out research on patients with ACS using the MINAP dataset. For example, when recording information on troponin, the preferred blood test for assessing the degree of heart muscle damage, the MINAP dataset records the peak troponin level and the troponin assay used to measure the troponin level. The troponin assay variable captures data on whether the hospital measures troponin I or troponin T with either a contemporary or high-sensitivity assay. Each assay yields results in different measurable ranges, with unique cut-off points for the 99th percentile of the upper limit of normal (ULN). The ULN provided by the manufacturer for each troponin assay provides an analytical cut-off between normal and increased (positive) cardiac troponin levels. A positive troponin is defined as a result above the ULN for each troponin assay. As a result, when using troponin level in analyses, it is important

to standardise the different troponin assays between the academic centres by scaling the results using the ratio of the observed troponin value divided by the ULN for each troponin assay.

As a result, there are limitations with using MINAP data for research. Even if a research question being addressed may not primarily focus on troponin, due to the correlation between troponin level and patient outcomes, such as mortality, it is appropriate to adjust analyses using multicentre cohorts by 'standardised' troponin levels. It is important to appreciate, however, that MINAP was not designed for research, but to audit the provision of guideline indicated care.

In terms of data quality, there is evidence that reporting of ACS patients may not capture all patients eligible for entry into MINAP.[26] Case ascertainment is calculated as the ratio of the number of cases that were coded as an ACS using HES data in England to the number of cases recorded in the MINAP dataset. In the most recent MINAP annual report in 2021, the case ascertainment rate for England was approximately 90%.

Table 1.5 - Data fields supplied by the MINAP ACS registry[27]

Field	Field name	Description	Values	Format
Admission				
2.39	AdmissionMethod	The mode by which a patient was admitted to hospital	1. Direct admission via emergency service	String
3.17	AdmissionWard	The unit to which the patient is admitted	1. Cardiac care unit	String
2.23	PlaceFirstLeadECGPerformed	Place 1st 12 lead ECG performed	2. In hospital	String
2.01	AdmissionDiagnosis	The working diagnosis	3. Acute coronary syndrome	String
History				
2.05	PreviousMI	Previous acute myocardial infarction	0. No	String
2.07	Hypertension	Hypertension	1. Yes	String
2.18	PreviousIntervention	Previous PCI	9. Unknown	String

Examination				
2.41	KillipClass	Killip class	1. No evidence of heart failure	String
2.21	HeartRate	Heart rate on admission	64 bpm	Numeric
Drugs on admission				
2.04	Aspirin	Where was Aspirin/other anti-platelet given?	1. Already on aspirin / antiplatelet drug	String
2.24	BetaBlockerUse	Any beta blocker in regular use prior to this admission.	0. No	String
Drugs in hospital				
3.20	UnfractionatedHeparin	Unfractionated heparin	1. Yes	String
3.22	ThienopyridineInhibitor	Thienopyridine Platelet Inhibitor	1. Yes	String
3.32	Angiotensin	Angiotensin converting enzyme inhibitor or angiotensin receptor blocker	1. Yes	String
3.43	OralBetaBlocker	Oral Beta Blocker	1. Yes	String
Drugs on discharge				
4.05	Betablocker	Beta blocker	1. Yes	String
4.07	Statin	Statin	1. Yes	String
4.08	AspirinSecondary	Aspirin	1. Yes	String
4.31	DischargedOnTicagrelor	Ticagrelor	1. Yes	String
4.06	ACEInhibitor	ACE or ARB	1. Yes	String
Investigations/Interventions				
4.13	CoronaryAngiography	Coronary angiography	1. Protocol driven investigation performed in this hospital	String
4.14	CoronaryIntervention	Coronary Intervention	1. Percutaneous coronary intervention	
Tests				

3.19	PeakTroponin	Peak Troponin	441 ng/ml	Numeric
3.37	TroponinAssay	Troponin Assay	3. High sensitivity Troponin T	String
2.31	LeftVentricularEjectionFraction	LV Ejection Fraction	2. Moderate	Numeric
Treatment				
2.03	ECGDeterminingTreatment	ECG Determining Treatment	4. T wave changes only	String
2.36	InfarctionSite	Site of Infarction	2. Inferior	String
Complications				
3.14	CardiacArrest	Was there a Cardiac Arrest? If so, where?	1. No arrest	String
4.03	BleedingComplications	Bleeding Complications	5. Any bleed with Hb fall <30 g	String

1.7.4 Office of National Statistics death registry data

Coding of cause of death is recorded in the Office of National Statistics (ONS) database.[28] In clinical practice, the cause(s) of death are recorded using free text. Based on this, specialist software is used to automatically convert the free text terms to an ICD-10 code, with unrecognised causes of death being manually coded by trained coders. **Table 1.6** summarises the information coded for the patient scenario in **Figure 1.1**. The main issue with the ONS death registration data is that the cause of death is not always complete.

Table 1.6 - Data fields supplied by the ONS death registry

Field	Field name	Description	Values	Format
AGE_AT_DEATH	Age at death	The patient's age, at time of death	88	Number
CAUSE_OF_DEATH	Original underlying cause of death	A valid ICD-9 or ICD-10 diagnosis code. There are fifteen diagnosis fields that can be populated per record.	A – Left ventricular failure B – Lower respiratory tract infection C – Diabetes D – Hypertension	String
CAUSE_OF_DEATH_ROW_POS_nn	Original underlying cause of death row position	There are fifteen row positions that relate to the CAUSE_OF_DEATH field.	A – 1 B – 2 C – 3 D – 4	Number
PLACE_OF_DEATH	Patient's place of death	This is a free text format of the place of death of the patient.	Hospital	String

1.8 Linkage of data sources

Linking existing EHR data sources via big data research platforms allows researchers to create a longitudinal data record of the patient's health journey from GP registration to death (**Figure 1.1**). The following sections will summarise the big data research platforms currently available for research.

1.8.1 CPRD linked data

For a fee, data from patients in the CPRD dataset can be linked to a range of other data sources, including to the ONS and HES datasets.[29] Access to linked data is dependent on research protocol approval by a governance committee. Whilst this linkage enables CPRD to provide a fuller picture of the patient care record, it remains limited by the issues inherent with each of the individual datasets as discussed in **Section 1.7**.

1.8.2 CALIBER

Building on the CPRD linked dataset described in **Section 1.7.1**, the

CALIBER platform additionally links the MINAP ACS registry to the CPRD, HES and ONS national EHR data sources.[30] As a result, there is further scope to generate evidence to inform health care and public health policy for cardiovascular disease, especially in patients with ACS.[31]

1.8.3 Discover

The Discover data is a dataset of linked primary care, secondary care, community mental health and social care EHRs for over 2.3 million patients who are registered with a GP in North West London (NWL).[32] It extends the primary and secondary care elements of the CPRD-linked dataset from residents in NWL to cover some additional variables. In particular, the Discover dataset also captures social care and community mental health data which allows for research in these additional areas of heath care.

1.8.4 Other sources of primary care data

There are a number of primary care data sources beyond CPRD, including The Health Improvement Network (THIN)[33] and QResearch[34].

THIN uses data extracted from the Vision primary care computer software, which dates back to 1994. It currently contains the EHRs of 19.7 million patients across 850 GP practices, of which 2.9 million are active patients.[33,35] The database contains information on symptoms, drug prescriptions, diagnoses, test results, health indicators and measures of social deprivation. The data is catalogued using a combination of SNOMED CT codes, Read codes, ICD-10 codes and ATC codes for medications.

The QResearch database was more recently established in 2003, supported by EMIS Health and the University of Oxford.[34] Data are available from approximately 1,500 GP practices across the UK using the EMIS primary care computer system. At present, the database holds EHRs of over 35 million patients.

1.8.5 UK Biobank

The UK Biobank study is a prospective cohort study which was established to investigate lifestyle and genetic determinants of different diseases.[36] Between 2006 and 2010, 500,000 participants living in the UK,

aged between 40 and 69 years old were recruited. This open-access resource involved collecting extensive questionnaire data, physical measurements and biological samples at recruitment. In large subsets of the cohort there is ongoing data collection, including repeat baseline and physical examination measurements, physical genotyping, online questionnaires and multi-modal imaging. All patients are also followed up for new diagnoses through linkage to national EHR datasets, including CPRD, HES and ONS.

While the Biobank study provides extensive health information from half a million individuals, the sampling population is volunteer-based and is therefore not representative of the UK population.[37] As a result, estimates of both incidence and prevalence should be interpreted with caution.

1.9 Hospital electronic health records

The NIHR HIC[38] was established in 2014, in response to a challenge set by Dame Sally Davies, the UK's Chief Medical Officer, to make routinely available clinical data available from multiple sites for translational research. The NIHR HIC programme involves developing the infrastructure across a network of 25 NHS Trusts, supported by their NIHR BRC university partners. This initiative was originally an arrangement of five NHS Hospital Trusts which hosted comprehensive NIHR BRCs; Imperial College Healthcare NHS Trust, Oxford University Hospitals NHS Foundation Trust, University College London Hospitals NHS Foundation Trust, Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust.

The overarching aim of the NIHR HIC is to improve the availability of high quality, granular level, routinely collected clinical data across multiple sites, in order to make it available for collaborative translational research. The NIHR HIC provides an opportunity to leverage the millions of data points generated in the NHS each year for patient care. Once processes for data flow between collaborating sites have been developed, the NIHR HIC infrastructure allows automated data collection from electronic health record (EHR) systems resulting in a dramatic reduction in the time and cost associated with data collection for research. The availability of highly granular clinical data across a network of leading hospitals also creates new opportunities for engagement and collaboration with both academic and industry partners.

EHR data extraction allows capture of data fields from secondary care that not collected using national HES data, including blood test results, medications prescriptions, vital signs and the results of investigations.

There are however a number of challenges which need to be addressed

before linked data from multiple sites can be used for research:

- Individual NHS Trusts may use different EHR systems with customised variations in what data items they store. In addition to the primary EHR system, each Trust may also have a number of hospital or department specific systems.
- As all NHS Trusts have individual responsibility to protect their own patient data, a governance framework is required to enable data to be shared across organisations.
- The various data items collected when a patient presents to hospital may not be recorded electronically at all Trusts.
- Protocols for the clinical management of patients presenting with a particular condition may differ between NHS Trusts, as well between individual doctors. As a result, there may be variation in the data items collected for patients presenting to hospital with similar conditions.
- Data item definitions may not be standardised between NHS Trusts.
- Data for a particular data item may be collected differently between each NHS Trust. For example, each laboratory may use a different assay or platform for measuring a troponin level (marker of cardiac muscle damage) in a blood test.
- Useful clinical data may be stored in unstructured free text which may more difficult to extract and use for statistical analyses compared to extracting and using discrete, structured data.
- Important clinical data may not be collected electronically at all NHS Trusts.

My thesis will explore how the NIHR HIC infrastructure was developed to overcome these challenges.

1.10 Capturing patients presenting to hospital with cardiovascular disease

1.10.1 Patients presenting to hospital requiring a troponin blood test

Despite great strides in reducing the worldwide burden of cardiovascular disease and mortality over the past few decades, cardiovascular disease continues to cause significant morbidity and mortality. Approximately 7 million people in England are living with cardiovascular disease, which contributes to around one in four of all deaths in England each year.[39] Ischaemic heart disease (IHD), which is the most common type of cardiovascular disease in the UK, involves the reduction of blood flow to the myocardium due to atherosclerosis in the coronary arteries. It was the third most common cause of death in the UK in 2020, after COVID-19 and dementia, and is the most

common cause of pre-mature death.[40]

Acute coronary syndrome (ACS) is a subcategory of IHD whereby the atherosclerotic plaque ruptures, leading to acute coronary obstruction, reduced myocardial perfusion and troponin elevation.[41] Patients with suspected ACS therefore present acutely to hospital where they undergo troponin testing as per national and international guidance.

As the NIHR HIC dataset involves extracting data from secondary care EHRs, my thesis will focus on identifying patients presenting to hospital with suspected ACS, characterised by the request of a troponin blood test.

1.10.2 Additional blood tests in patients presenting to hospital with a troponin blood test

Patients with suspected ACS also undergo a number of additional blood tests on presentation to hospital. Inflammation has been extensively studied and postulated as mechanistic in atherothrombosis.[42] Much debate has been on whether the immune system plays a protective or pathogenic role in atherogenesis and subsequent cardiovascular events. C-reactive protein (CRP) is the most widely evaluated of these biomarkers, given its relatively long circulating half-life and commonplace use in clinical practice.

For my thesis, I will endeavour to capture the results of additional blood tests beyond troponin for patients admitted to hospital with a troponin measurement to answer research questions on the mechanisms in atherothromobosis.

1.10.3 Patient diagnosis following troponin measurement

A patient with a troponin measurement may be diagnosed with ACS or a non-ACS condition. ACS encompasses unstable angina, non-ST elevation myocardial infarction (NSTEMI) and ST elevation myocardial infarction (STEMI).[41] A diagnosis of ACS is based on three clinical factors, which include clinical presentation, changes on the electrocardiogram (ECG) and measurement of troponin. In general, a STEMI is caused by a *complete* blockage of the coronary artery resulting in myocardial necrosis (with a raised troponin) with elevation of the ST-segment on the ECG. In patients with an NSTEMI or unstable angina, there is a *partial* blockage of the coronary artery, which results in myocardial necrosis (with a raised troponin) in NSTEMI but not in unstable angina. The ECG for both NSTEMI or unstable angina may show ST-segment depression, T-wave inversion, or may be normal.[41]

Sometimes the troponin test is positive in isolation, with no other indication to suggest myocardial ischaemia but with clear evidence of another non-ACS diagnosis, such as atrial fibrillation or pneumonia.[43,44] Troponin elevation in non-ACS patients may be a marker of diminished organ-level reserve due to global comorbidity rather than a specific pointer to unstable coronary disease.

For my thesis, I will endeavour to capture the discharge diagnosis for patients admitted to hospital with a troponin measurement to answer research questions in particular subgroups of patients.

1.10.4 Patient management following troponin measurement

Invasive management, in the form of a diagnostic coronary angiogram, is considered the gold standard approach to managing patients with ACS.[41] The alternative to invasive management is medical (non-invasive) management of patients with ACS. In addition to Aspirin, most patients with ACS are usually administered a second antiplatelet agent (Clopidogrel, Ticagrelor or Prasugrel). Anti-thrombin therapy with Fondaparinux Sodium is usually also offered.

For patients who have a non-ACS diagnosis with a positive troponin, there is currently no consensus on the best approach on how to investigate non-ACS patients with some patients undergoing invasive procedures such as coronary angiography to evaluate for the presence of coronary artery disease.[45]

For this thesis, it is therefore crucial to capture whether both ACS and non-ACS patients underwent any invasive procedures to definitively assess for underlying coronary artery disease.

1.11 Aims and objectives

In **Section 1.2**, I introduced the concept of using big data to answer research questions that cover the translational spectrum from an early phase of discovery to a later implementation phase. In **Section 1.3**, I highlighted the rationale behind creating a NIHR HIC dataset of patients presenting to hospital with a troponin test. The Section also highlighted the importance of classifying these patients as ACS or non-ACS, whether additional blood tests were performed, and the management strategy to investigate for underlying coronary artery disease. I will bring together both of these sections to generate the following aims and objectives of my thesis.

1.11.1 Aims

To develop methods to assist in using EHR data from multiple sites for research, and apply them to an exemplar dataset of patients with suspected ACS to answer research questions that cover the translational spectrum from an early phase of discovery to a later implementation phase.

1.11.2 Objectives

The background evidence to support the research study objectives are discussed in their respective chapters.

1.11.2.1 Method-specific objectives

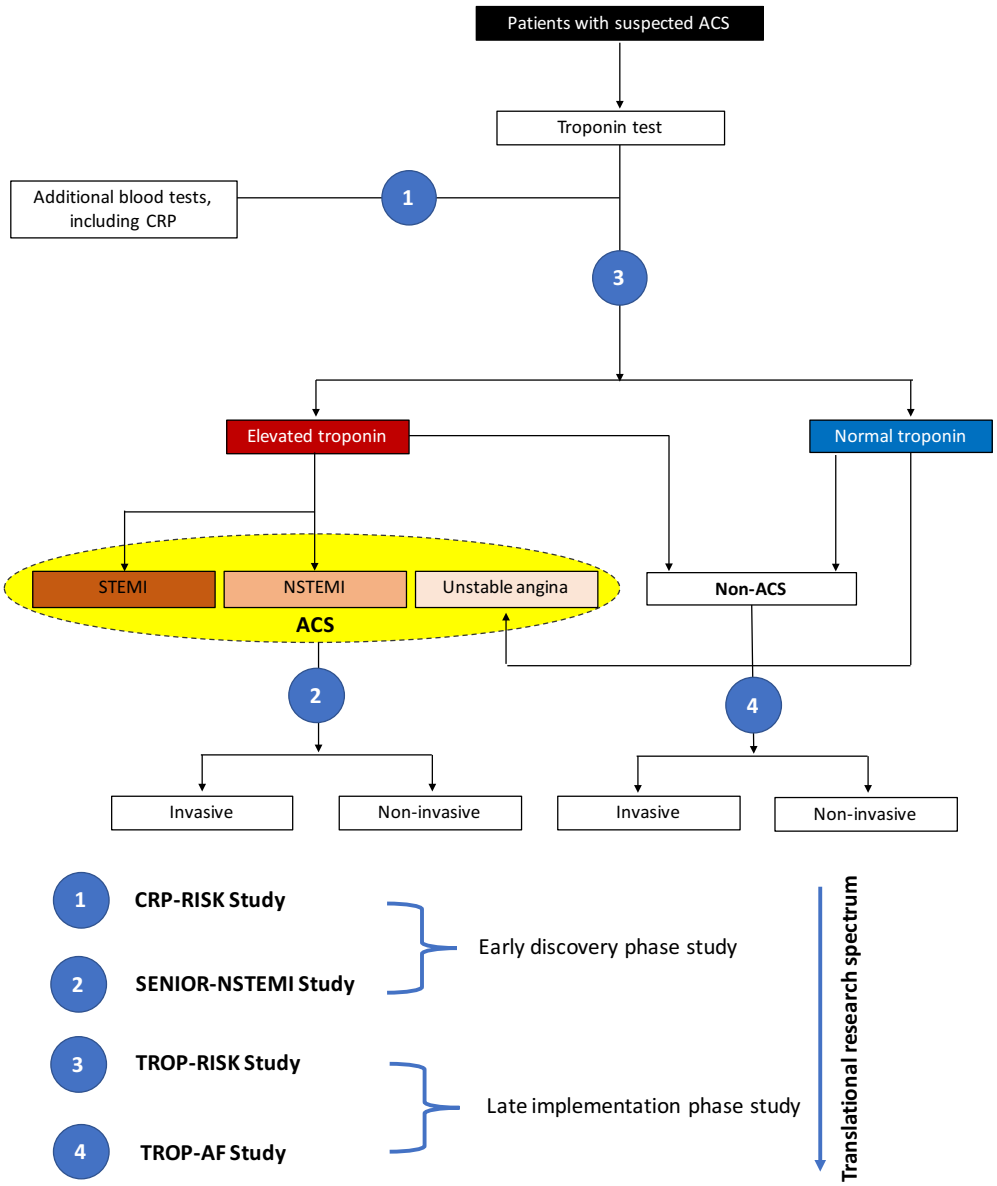
- **a.** Develop the tools to extract and standardise hospital EHR data from multiple sites (Chapter 2).
- **b.** Curation a research-ready data view of all patients presenting to hospital with a troponin measurement (Chapter 3).

1.11.2.2 Research studies

- **a.** Demonstrate how big data can contribute to early scientific advancement in cardiovascular medicine through discovery of underlying disease mechanisms by evaluating whether a mildly elevated high-sensitivity CRP (hsCRP) level is associated with mortality risk, beyond troponin level, in patients with suspected ACS (Chapter 4; CRP-RISK Study).
- **b.** Demonstrate how big data can contribute to early discovery phase translational research by assessing the effect of an intervention in population subgroups that were underrepresented in previous clinical trials by estimating the effect of invasive versus non-invasive management on the survival of patients aged 80 years or older with NSTEMI (Chapter 5; SENIOR-NSTEMI Study).
- **c.** Demonstrate how big data can be used for late translational research through assessment of the epidemiology of cardiovascular disease by
  - i. evaluating the relationship between the full spectrum of troponin level and mortality in all patients in whom troponin testing has been performed for clinical purposes (Chapter 6; TROP-RISK Study).
  - ii. evaluating the relationship between troponin level, coronary angiography, and all-cause mortality in patients presenting to hospital with atrial fibrillation (Chapter 7; TROP-AF Study).

The pathway of patients after presenting to hospital with suspected ACS is illustrated in **Figure 1.2**. It also highlights where on this pathway the research studies will focus.

Figure 1.2 - Pathway of patients presenting to hospital with suspected ACS



ACS, acute coronary syndrome; AF, atrial fibrillation; CRP, C-reactive protein; NSTEMI, non-ST elevation myocardial infarction; STEMI, ST-elevation myocardial infarction.

2. The NIHR Health Informatics Collaborative

2.1 Chapter outline

The NIHR HIC programme aims to overcome the challenges of collecting routine clinical data from hospital EHR systems across multiple sites. This chapter outlines the specifications and procedures for the NIHR HIC and defines the processes for the collection of data for the NIHR HIC Cardiovascular research dataset which was used for all the work in this thesis.

2.2 Procedures for local data collection into local stores

The NIHR HIC Cardiovascular database was designed to support multiple research questions from any interested parties and was therefore not focused purely on data suitable for a specific study/question. However, four exemplar studies were funded to utilise the resource and demonstrate usefulness and functionality. These studies are outlined in Chapters 4-7.

Data are collected automatically from primary clinical systems within each Trust. NHS staff enter data into clinical systems during routine clinical care of patients, or data are generated as a result of clinical tests. Data in the primary clinical systems are processed in accordance with each Trusts clinical guidelines and subject to local quality and governance procedures. Primary clinical systems in each of the Trust differ and processes for extraction and transformation will vary. Data are extracted automatically and validated extract, transform, load (ETL) processes have been designed to import the data into the local secure data stores at Imperial College Healthcare NHS Trust (ICHNT).

The procedure I used for data integration is known as “data consolidation”. Data consolidation involves physically bringing data together from various different systems. This results in a “consolidated” dataset in one local data store at ICHNT. The main goal of data consolidation is to reduce the number of locations where data are stored. The ETL technology supports the data consolidation process by pulling data from the different sources across all Trusts, transforming it into a standard data format, and then transferring it to the local data warehouse.

One benefit of adopting this data consolidation approach is that it allows for data to be pre-processed in advance of any statistical analyses. As a result, the most computationally intensive processes will be performed prior to using the data for research purposes. A pitfall with this approach is that the statistical analyses will be performed on data that has been previously consolidated in the data warehouse, and there it does not provide real-time visibility.

2.2.1 Access to local data stores

Each Trust will have a local store of NIHR HIC NIHR HIC Cardiology data; this collects their information in an identifiable form from clinical systems, the data are then de-identified within the Local NHS Trust. These de-identified data are then passed to the central research database. The research database contains only de-identified information. This database combines the data from each site (including ICHNT) and will be used for research purposes.

Access to identified data is limited to those justified and approved by the local information governance teams and are always NHS staff in accordance with the duty of confidentiality required by law. Local data stores are the only areas that hold any patient identifiers. De-identification is completed at this stage prior to passing data into any research database.

2.2.2 De-identification

The data are pseudonymised locally within each Trust; anonymisation processes are automated and were set up by NHS staff in accordance with advice from local information governance procedures and the NIHR HIC Standard Operating Procedures for data sharing and anonymisation and the Clinical data transfer policy for de-identification and anonymisation. Anonymisation is approved locally by information governance teams before data are sent externally to ICHNT. De-identified data are then shared in accordance with the overarching data sharing agreement.

NHS numbers and hospital numbers are pseudonymised using locally approved procedures. Names are removed from the dataset and date of birth is transformed to year of birth. Date of death is shared, however, is converted in to relevant survival rates on provision of data to researchers. Researchers will never see the full date of death or be able to calculate it from other information. The provision of date of death has been agreed by each of the information governance offices at each of the sites in the de-identification and anonymisation protocol. The exact date of death is required to evaluate mortality after diagnosis. Patients presenting with ACS often have cardiac events and may die within a few months of presentation so an accurate measure of death is required to fully evaluate mortality risk in this cohort.

To redact the date of death to year only would misrepresent the survival of these patients, particularly for those who survive for less than one year. To show the full death patient for patients when they have died will balance the

risk of identifying live patient from combining their data fields with the need for accurate survival data. The clinical leads at our BRCs underlined the importance of having the date in full for any service improvement evaluation or cardiac research.

Each site will hold two versions of the database, one identifiable and one with de-identified, pseudonymised data. The de-identified version is for use in research and shared with the central research database at ICHNT. The identifiable database is held so that if necessary patients can be re-identified if it is of importance to re-contact the patient via their care team.

2.2.3 Data validity and quality

Due to the variation in primary systems, data are transformed to meet the standardised NIHR HIC NIHR HIC Cardiology data model (covered in Chapter 3). Prior to pseudonymisation within the clinical systems, NHS number, date of birth and patient names are automatically checked to remove duplication of patients.

Samples of data are clinically validated to ensure that the transformation process is correct and data are attributed to the correct patients prior to pseudonymisation. After clinical validation is complete the data are transformed to standardised NIHR HIC Cardiology Extensible Markup Language (XML).

2.2.4 Data sharing between Trusts

Data are shared in accordance with the Standard Operating Procedures for data sharing. Data are encrypted in transit via SSH File Transfer Protocol (SFTP) on the N3 network. The N3 is a national broadband network for the NHS, connecting all NHS locations across England. The sFTP is set up and hosted by ICHNT.

The NIHR HIC partners involved in the Cardiovascular theme will extract, create, validate, and share XML documents of anonymised dataset derived from patients who have received a troponin test since 2010 (or as far back in time as possible on each sites EHR system) in scope with an agreed dataset, which is covered in Chapter 3.

2.2.4.1 Scope and staff

Clinical and technical staff within NIHR HIC participating organisations

are expected to access the systems and data as part of the research phase of the project.

Access to the system/data will be limited on the basis of necessity and role. Only staff involved in receiving and validating the XML files, maintaining the database, and clinical staff will be able to view the data or access the ICT system to mitigate the risk of unauthorised access to the data.

2.2.4.2 Data

The dataset used for the research is anonymised data from each centre's source systems in scope with the Cardiovascular theme. The dataset will comprise data from all patients receiving a troponin test since 2010 or the furthest available in time for which the information listed in the NIHR HIC Cardiovascular dataset will be used.

The data items in **Table 2.1** will be anonymised according to local process.

Table 2.1 - Anonymised data items

Data item	Data table	Description	Notes
NHIC_ACS_1	DEMOGRAPHICS	Local Identifier	Provided if NHS number is missing
NHIC_ACS_2	DEMOGRAPHICS	NHS NUMBER	To be pseudonymised (key to be retained at source)
NHIC_ACS_3	DEMOGRAPHICS	PERSON FAMILY NAME	To be removed – LOCAL use
NHIC_ACS_4	DEMOGRAPHICS	PERSON GIVEN NAME	To be removed – LOCAL use
NHIC_ACS_7	DEMOGRAPHICS	YoB	YYYY
NHIC_ACS_8	DEMOGRAPHICS	YoD	DD-MM-YYYY

When there are more non-troponin tests than troponin, biochemistry nodes without troponin tests should still contain the troponin data elements, however these should be set to default values. The appropriate default values are based on the data type (**Table 2.2**).

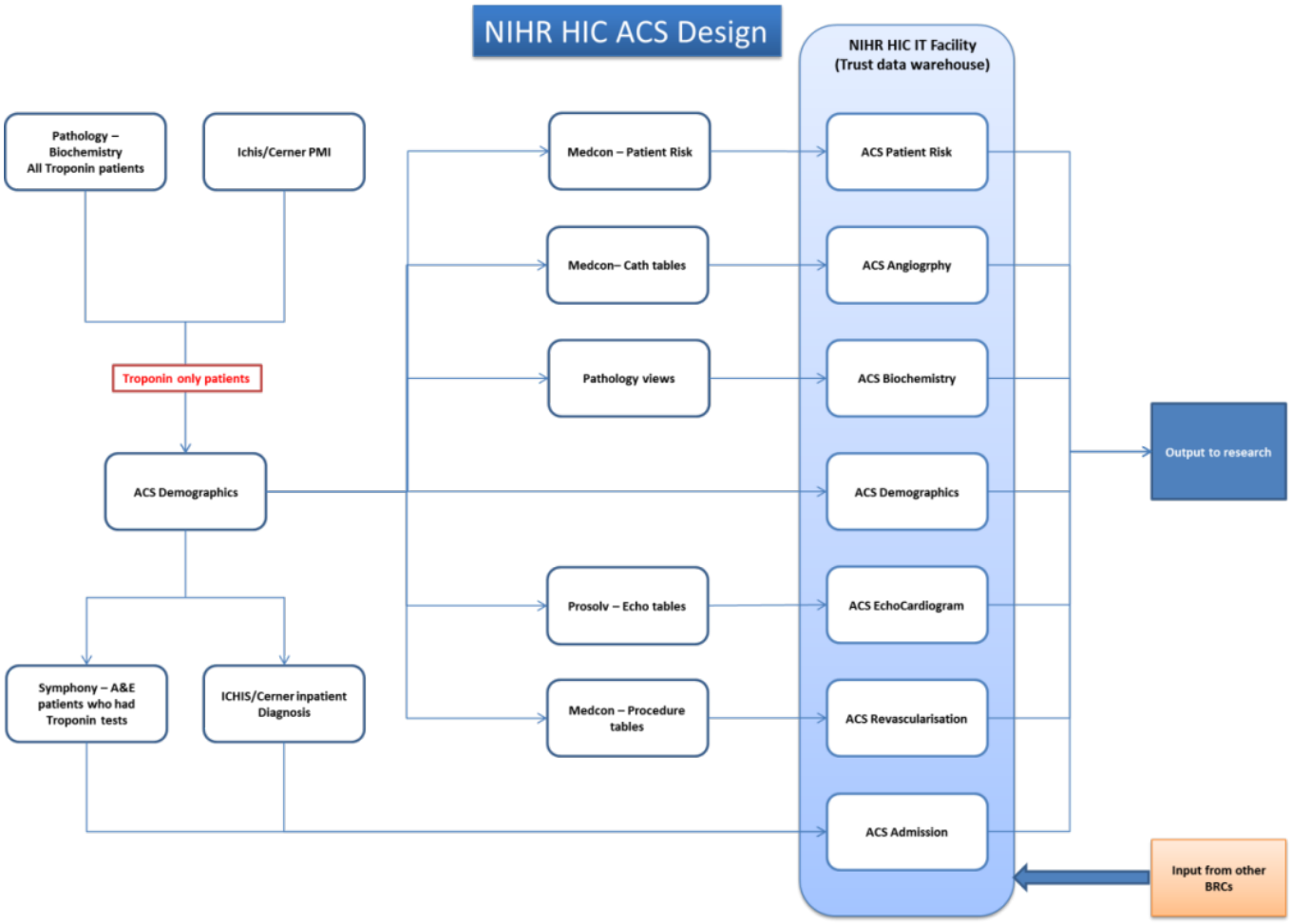
Table 2.2 - Default values for missing data

Data Type	Default Value
Date	1900-01-01
Time	00:00:00
String	Unknown
Numeric	999

2.4.2.3 Systems

Each centre's local solution is out of scope for this document. At Imperial, the NIHR HIC data architecture is defined as shown in **Figure 2.1**. Data from other BRCs will be fed into the NIHR HIC information technology facility.

Figure 2.1 - NIHR HIC data architecture



A&E, accident and emergency; ACS, acute coronary syndrome

2.3 Validation process

Below is a list of high level tests that will be used by our system to validate data that are transferred and shared in a correct way among the collaborators of the Cardiovascular theme.

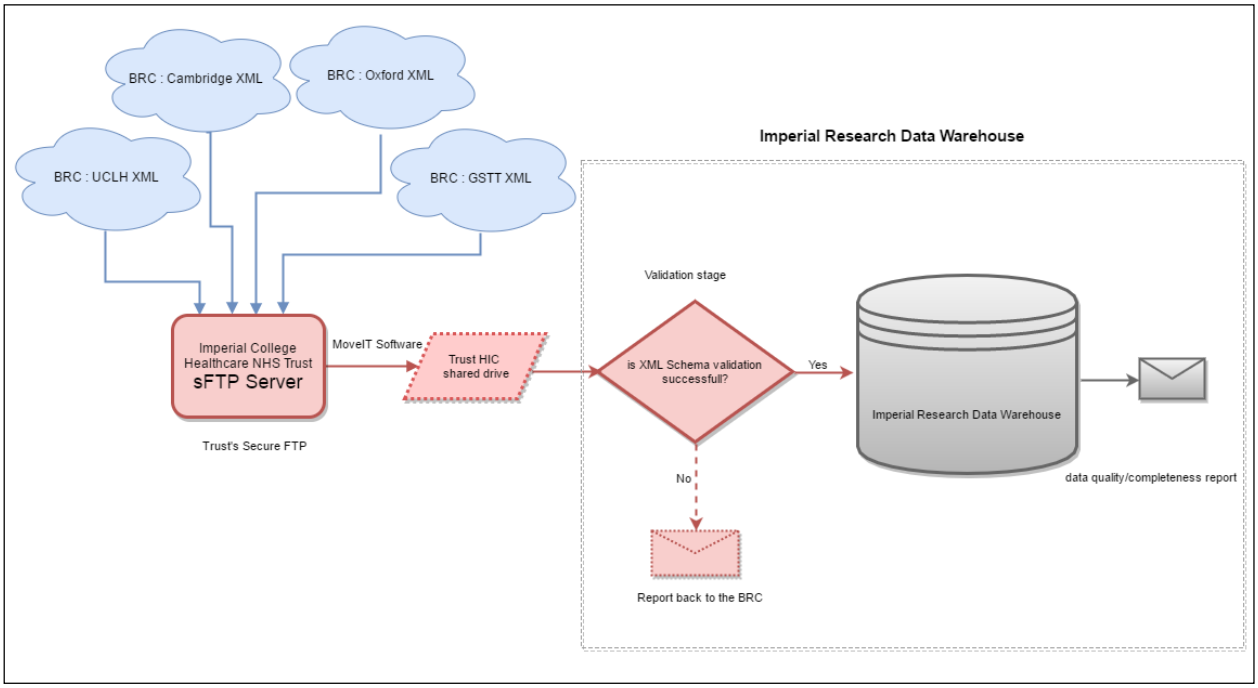
a. Validate XML document
b. Load collaboration data into the ICHNT NIHR HIC Database
c. Generate extract files for use by analytical/statistical software package
d. Execute a sample set of exploratory analyses on the collaboration dataset
e. Validate mappings of reference data sets

ICHNT Validate the data against the NIHR HIC Cardiology XML schema definition to ensure that the structure, data items, units and data types are in accordance with the standardised data model. Data are rejected if validation fails. On rejection, the files are archived and data manager contacts the data provider to review the submission and resend once corrected.

Once imported into the secure research database, data are subject to clinical validation by clinicians; these validations are completed by the clinical researchers using de-identified data. Data are reviewed for data completeness, spread and actual data point values. If data appears invalid it is rejected.

Data quality reports are automatically generated and provided to the research team and local data provider after each submission. These are reviewed after each submission to ensure all areas are populated. The XML import process flow diagram is summarised in **Figure 2.2**.

Figure 2.2 - NIHR HIC XML import process (receiving environment)



BRC, Biomedical Research Centre; GSTT, Guy's and St Thomas' NHS Foundation Trust; sFTP, SSH File Transfer Protocol; UCLH, University College London Hospitals NHS Foundation Trust; XML, Extensible Markup Language

2.4 Research database software

The database is built using Microsoft SQL server 2014 (which is Microsoft's principle database management system software). The installation of the software was carried out by certified technical consultants and tested by the trust ICT team and data warehouse team in accordance with Trust ICT procedures and policy.

2.5 Database management

The Database is fully backed up on a daily basis. Backups are standardised for all trust databases within the trust data warehouse. Backups can be restored at any point, by warehouse staff. Data are secondary copies from clinical systems; at any point the participating trusts can re-extract the data from primary sources.

Once entered on the system data are not changed. All access is read only except via exception, approved by research Informatics Programme Manager and Clinical Leads group.

The database is managed by the data manager and developer. Database changes are controlled by the research informatics Programme manager, and are sanctioned by the NIHR HIC Cardiovascular scientific steering committee. All staff are substantive NHS employees at ICHNT.

Data extracts taken for research are logged at stored within the data warehouse and retained for the period specified in the data request. All information pertinent to the request is retained and tracked by the data manager.

2.6 Security and access

The database is held on a dedicated research server within ICHNT secure data warehouse. The Trust is information governance (IG) toolkit compliant and compliance is externally reviewed annually. Two versions of the database are maintained, one in the test environment and one as the live database. Both versions are subject to the same controls and audits.

Access to the database is via Trust access control policy. Only authorised, users within the data warehouse team can access the research server. Within the data warehouse team, access to the NIHR HIC Cardiovascular research databases is limited to the database developer and the database manager. Access to the data for research is provided as an export from the database (to an approved location by IG team) or via controlled views of specific de-identified layers (managed by access control restrictions). Access will be granted to researchers approved by the NIHR HIC Cardiovascular theme leads.

All researcher users granted direct access to the database will need to have contacts with ICHNT, as the access control policy demands. All users granted access to the research database are NHS Trust employees and are subject to IG training, monitored centrally. Data protection arrangements are built into staff contracts with breach of IG rules being a disciplinary offence that may result in dismissal.

Data access is read only by default; there are no changes made to the individual data points and no manual data entry to the system in the research database; all data are collected from primary clinical systems. Any changes made to systems (data structure/data items collected) must be approved by the Research Informatics Programme Manager and sanctioned by the NIHR HIC Cardiovascular theme leads. Changes are tested in the test environment and database backup before changes are made live.

2.7 Data access for research

Access for researchers will be by extract or controlled access to research analysis system. All researchers must apply to the NIHR HIC Cardiovascular theme leads for review of research proposals prior to being granted access. Access is time limited and permissions are revoked or data must be destroyed on expiry.

All research will be completed on fully de-identified data. This includes further de-identification to remove dates. Dates are converted to delta dates, with date zero being the date of the first troponin test. All further dates are provided as number of days from date zero. Age will be provided in years, at the time of the first troponin test. The source of data (i.e. Trust provider) will only be provided if this is of relevance to the research. Date of death is converted to the number of days since date zero.

2.8 Direct access to research views

The research analysis solution will only hold de-identified data. This is currently a set of de-identified views made available on the research server. Access to these views is strictly controlled via the Trust access control system and researchers will not be able to access any other systems or data.

Designated statisticians will be able to freeze copies of the database, so a static dataset can be used for analysis. This will be retained separately from the live database to allow reproducible analyses.

2.9 Outline of the duration of data retention

The NIHR HIC Cardiovascular dataset is designed for use in long term studies and for reuse in multiple studies. Therefore, there is no specified data retention period. However, as part of the NIHR HIC programme there will be an annual review of the NIHR HIC data collection. This review will examine if the data can be retained for the forthcoming period. It will be specified in the researcher data sharing agreements that responsibility for data retention for academic purposes will rest with the research provided with the data.

2.10 Ethics and approvals

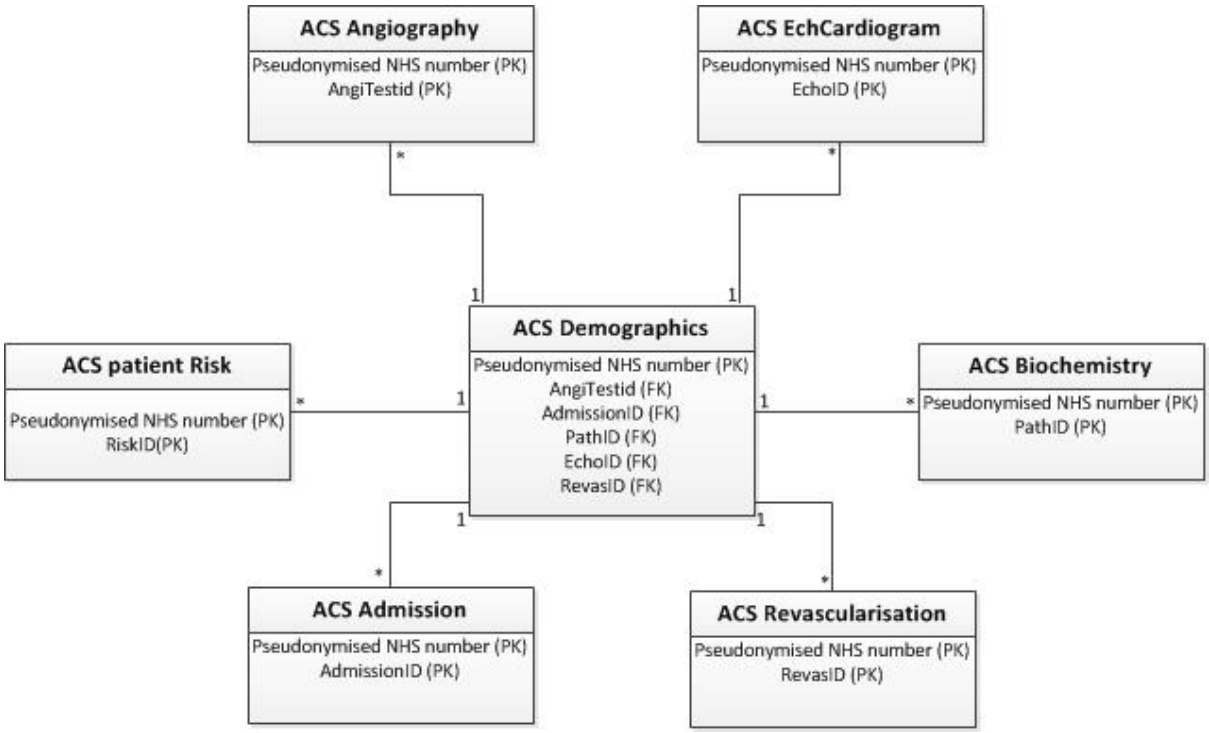
The NIHR HIC dataset was approved by the London-South East Research Ethics Committee (REC reference: 16/HRA/3327).

3. Database design

3.1 Data specification

The database design is shown in **Figure 3.1**.

Figure 3.1 - NIHR HIC database design



Data are coded based on the NIHR HIC Cardiovascular data model, which is standardised for all Trusts submitting data. All data are validated using the NIHR HIC Cardiovascular XML Schema Definition. This ensures that all fields conform to correct data values and types. The full dataset is described in **Table 3.1**.

Table 3.1 - Data specification for NIIHR HIC cardiovascular dataset

Data item No.	Table	MINAP	Data Item Name	Data Item Description	National Code
NHIC_ACS_1	DEMOGRAPHICS		LOCAL PATIENT IDENTIFIER	This is a number used to identify a PATIENT uniquely within a Health Care Provider. It may be different from the PATIENT's case note number and may be assigned automatically by the computer system. This data item is mandatory if missing should be set to UNKNOWN	LOPATID
NHIC_ACS_2	DEMOGRAPHICS	1.03	NHS NUMBER / Subject ID (if **pseudonymised**)	The NHS NUMBER, the primary identifier of a PERSON, is a unique identifier for a PATIENT within the NHS in England and Wales. This will not vary by any ORGANISATION of which a PERSON is a PATIENT. If patient NHS number is missing then data item should be set to UNKNOWN and NHIC_ACS_1 must be populated with data item other than UNKNOWN	NEWNHSNO
NHIC_ACS_3	DEMOGRAPHICS	1.04	PERSON FAMILY NAME	That part of a PERSON's name which is used to describe family, clan, tribal group, or marital association. This is redacted for research purposes	PSUR
NHIC_ACS_4	DEMOGRAPHICS	1.05	PERSON GIVEN NAME	The forename or given name of a PERSON. This is redacted for research purposes	
NHIC_ACS_5	DEMOGRAPHICS	1.07	PERSON GENDER	A PERSON's gender	0 : Not Known 1 : Male 2 : Female 9 : Not Specified
NHIC_ACS_6	DEMOGRAPHICS	1.13	ETHNIC CATEGORY	ETHNIC CATEGORY is the same as attribute ETHNIC CATEGORY CODE. The 16+1 ethnic data categories defined in the 2001 census is the national	

				mandatory standard for the collection and analysis of ethnicity.	
NHIC_ACS_7	DEMOGRAPHICS	1.06	YoB	Year of Birth (local provider may keep this data non anonymised)	
NHIC_ACS_8	DEMOGRAPHICS		Date of Death	Full Date of Death - provided as survival rates for research purposes - delta dates are used to provide survival rates between key events	
NHIC_ACS_9	ADMISSION	3.06	Presentation date	DATE OF PRESENTATION TO HEALTHCARE PROVIDER A&E only	
NHIC_ACS_10	ADMISSION	3.06	Presentation time	TIME OF PRESENTATION TO HEALTHCARE PROVIDER A&E only	
NHIC_ACS_11	ADMISSION		Admission date	DATE OF ADMISSION TO HEALTHCARE PROVIDER	
NHIC_ACS_12	ADMISSION		Admission time	TIME OF ADMISSION TO HEALTHCARE PROVIDER	
NHIC_ACS_13	ADMISSION		ADMISSION TYPE	Elective or non-elective	
NHIC_ACS_14	ADMISSION	2.39	ATTENDANCE MODE	How arrived at hospital described as Admission Route or Arrival Mode e.g. Self refer, Ambulance, inter-hospital	
NHIC_ACS_15	ADMISSION	1.01	HOSPITAL IDENTIFIER	The identifier allocated to the hospital by CCAD.	
NHIC_ACS_16	ADMISSION	2.01	Chest pain	Single episode of chest pain thought to be cardiac in nature where admission was thought appropriate to exclude an ischaemic event. This covers all other admissions where no clear initial diagnosis has been made, but where there is an index of suspicion that the symptoms may be ischaemic in nature.	0 : No 1 : Yes 9: Unknown
NHIC_ACS_17	ADMISSION	2.13	Heart failure	A previously validated diagnosis of heart failure on any therapeutic regime.	0 : No 1 : Yes 9 : Unknown
NHIC_ACS_18	ADMISSION	4.01	DISCHARGE DATE	Indicate the DATE OF DISCHARGE	

NHIC_ACS_19	ADMISSION	4.16	DISCHARGE STATUS	Indicate if the PERSON was alive or dead as of DATE OF DISCHARGE	0 : Alive 1 : Dead 9 : Unknown
NHIC_ACS_20	ADMISSION		Diagnosis CODE1	Primary Diagnosis	
NHIC_ACS_21	ADMISSION		Diagnosis CODE2	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_22	ADMISSION		Diagnosis CODE3	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_23	ADMISSION		Diagnosis CODE4	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_24	ADMISSION		Diagnosis CODE5	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_25	ADMISSION		Diagnosis CODE6	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_146	ADMISSION		Diagnosis CODE7	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_147	ADMISSION		Diagnosis CODE8	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_148	ADMISSION		Diagnosis CODE9	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_149	ADMISSION		Diagnosis CODE10	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	

NHIC_ACS_150	ADMISSION	Diagnosis CODE11	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_151	ADMISSION	Diagnosis CODE12	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_152	ADMISSION	Diagnosis CODE13	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_153	ADMISSION	Diagnosis CODE14	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_154	ADMISSION	Diagnosis CODE15	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_155	ADMISSION	Diagnosis CODE16	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_156	ADMISSION	Diagnosis CODE17	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_157	ADMISSION	Diagnosis CODE18	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_158	ADMISSION	Diagnosis CODE19	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful
NHIC_ACS_159	ADMISSION	Diagnosis CODE20	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful

NHIC_ACS_160	ADMISSION		Diagnosis CODE21	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_161	ADMISSION		Diagnosis CODE22	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_162	ADMISSION		Diagnosis CODE23	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_163	ADMISSION		Diagnosis CODE24	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_164	ADMISSION		Diagnosis CODE25	Secondary Diagnosis (patients may have multiple diagnosis per episode in addition to the primary, the order of secondary diagnosis is not meaningful	
NHIC_ACS_26	RISK FACTORS	2.16	SMOKING STATUS	The SMOKING STATUS of the PERSON at the time the this is recorded. This is a record closest to the time of the admission	0: Never smoked 1 : Ex-smoker 2 : Current smoker 9 : Unknown
NHIC_ACS_27	RISK FACTORS	2.07	PRIOR HISTORY OF HYPERTENSION	History of hypertension of the PERSON at the time the this is recorded. This is a record closest to each admission	0 : No 1 : Yes 9 : Unknown
NHIC_ACS_28	RISK FACTORS	2.17	DIABETES	History of diabetes of the PERSON at the time the this is recorded. This is a record closest to each admission	0 : No 1 : Yes 9 : Unknown
NHIC_ACS_29	RISK FACTORS	2.32	FAMILY HISTORY OF CORONARY HEART DISEASE	Family history of coronary heart disease of the PERSON at the time the this is recorded. This is a record closest to each admission	0 : No 1 : Yes 9 : Unknown

NHIC_ACS_30	RISK FACTORS	2.05	PREVIOUS MI	Any previously validated episode of acute myocardial infarction. This is a record closest to each admission	0 : No 1 : Yes
NHIC_ACS_120	RISK FACTORS		RISK RECORDED DATE AND TIME	Describes when the risk was recorded. This is a record closest to each admission	
NHIC_ACS_31	Biochemistry	3.37	Troponin Assay Type	Troponin Assay : Indicate the type of Troponin	1 : Troponin I 2 : Troponin T
NHIC_ACS_32	Biochemistry	3.37	Troponin Assay Sensitivity	Troponin Assay : Indicate if highly sensitive or standard assay	
NHIC_ACS_34	Biochemistry		Troponin Assay Result	Troponin Assay : Results - This is not the peak troponin result but all troponin measurements	
NHIC_ACS_35	Biochemistry		Troponin Assay Unit	Troponin Assay : Unit	
NHIC_ACS_36	Biochemistry		Troponin Assay lower	Troponin Assay : The LOWER normal value for the testkit use for the Troponin assay	
NHIC_ACS_33	Biochemistry		Troponin Assay higher	Troponin Assay : The HIGHER normal value for the testkit use for the Troponin assay	
NHIC_ACS_37	Biochemistry		Troponin Assay date	Troponin Assay : Collected Date for each Troponin Measurement (ACS_34)	
NHIC_ACS_38	Biochemistry		Troponin Assay time	Troponin Assay : Collected Time for each Troponin Measurement (ACS_34)	
NHIC_ACS_39	Biochemistry	2.34	Creatinine Result	Creatinine Result	
NHIC_ACS_40	Biochemistry		Creatinine Unit	Creatinine Unit	
NHIC_ACS_41	Biochemistry		Creatinine Collected date	Creatinine Collected date	
NHIC_ACS_42	Biochemistry		Creatinine Collected time	Creatinine Collected time	
NHIC_ACS_122	Biochemistry		Creatinine lower	The LOWER normal value for the testkit use for the CREATININE assay	lower
NHIC_ACS_123	Biochemistry		Creatinine higher	The HIGHER normal value for the testkit use for the CREATININE assay	higher
NHIC_ACS_43	Biochemistry		Egfr Result	eGFR Result	
NHIC_ACS_44	Biochemistry		Egfr Unit	eGFR Unit	
NHIC_ACS_45	Biochemistry		Egfr Collected date	eGFR Collected date	
NHIC_ACS_46	Biochemistry		Egfr Collected time	eGFR Collected time	

NHIC_ACS_124	Biochemistry		Egfr lower	The LOWER normal value for the testkit use for the eGFR assay
NHIC_ACS_125	Biochemistry		Egfr higher	The HIGHER normal value for the testkit use for the eGFR assay
NHIC_ACS_47	Biochemistry		Urea Result	Urea Result
NHIC_ACS_48	Biochemistry		Urea Unit	Urea Unit
NHIC_ACS_49	Biochemistry		Urea Collected date	Urea Collected date
NHIC_ACS_50	Biochemistry		Urea Collected time	Urea Collected time
NHIC_ACS_126	Biochemistry		Urea lower	The LOWER normal value for the testkit use for the Troponin assay
NHIC_ACS_127	Biochemistry		Urea higher	The HIGHER normal value for the testkit use for the Troponin assay
NHIC_ACS_51	Biochemistry		Sodium Result	Sodium Result
NHIC_ACS_52	Biochemistry		Sodium Unit	Sodium Unit
NHIC_ACS_53	Biochemistry		Sodium Collected date	Sodium Collected date
NHIC_ACS_54	Biochemistry		Sodium Collected time	Sodium Collected time
NHIC_ACS_128	Biochemistry		Sodium lower	The LOWER normal value for the testkit use for the Sodium assay
NHIC_ACS_129	Biochemistry		Sodium higher	The HIGHER normal value for the testkit use for the Sodium assay
NHIC_ACS_55	Biochemistry		Potassium Result	Potassium Result
NHIC_ACS_56	Biochemistry		Potassium Unit	Potassium Unit
NHIC_ACS_57	Biochemistry		Potassium Collected date	Potassium Collected date
NHIC_ACS_58	Biochemistry		Potassium Collected time	Potassium Collected time
NHIC_ACS_130	Biochemistry		Potassium lower	The LOWER normal value for the testkit use for the Potassium assay
NHIC_ACS_131	Biochemistry		Potassium higher	The HIGHER normal value for the testkit use for the Potassium assay
NHIC_ACS_59	Biochemistry	2.28	Glucose Result	Glucose Result
NHIC_ACS_60	Biochemistry		Glucose Unit	Glucose Unit
NHIC_ACS_61	Biochemistry		Glucose Collected date	Glucose Collected date

NHIC_ACS_62	Biochemistry	Glucose Collected time	Glucose Collected time
NHIC_ACS_132	Biochemistry	Glucose lower	The LOWER normal value for the testkit use for the Glucose assay
NHIC_ACS_133	Biochemistry	Glucose higher	The HIGHER normal value for the testkit use for the Glucose assay
NHIC_ACS_63	Biochemistry	Hba1C Result	Hba1C Result
NHIC_ACS_64	Biochemistry	Hba1C Unit	Hba1C Unit
NHIC_ACS_65	Biochemistry	Hba1C Collected date	Hba1C Collected date
NHIC_ACS_66	Biochemistry	Hba1C Collected time	Hba1C Collected time
NHIC_ACS_134	Biochemistry	Hba1C lower	The LOWER normal value for the testkit use for the HbA1c assay
NHIC_ACS_135	Biochemistry	Hba1C higher	The HIGHER normal value for the testkit use for the HbA1c assay
NHIC_ACS_67	Biochemistry	Cholesterol (Total) Result	Cholesterol (Total) Result
NHIC_ACS_68	Biochemistry	Cholesterol (Total) Unit	Cholesterol (Total) Unit
NHIC_ACS_69	Biochemistry	Cholesterol (Total) Collected date	Cholesterol (Total) Collected date
NHIC_ACS_70	Biochemistry	Cholesterol (Total) Collected time	Cholesterol (Total) Collected time
NHIC_ACS_136	Biochemistry	Cholesterol (Total) lower	The LOWER normal value for the testkit use for the Cholesterol (Total) assay
NHIC_ACS_137	Biochemistry	Cholesterol (Total) higher	The HIGHER normal value for the testkit use for the Cholesterol (Total) assay
NHIC_ACS_71	Biochemistry	Hdl Cholesterol Result	HDL Cholesterol Result
NHIC_ACS_72	Biochemistry	Hdl Cholesterol Unit	HDL Cholesterol Unit
NHIC_ACS_73	Biochemistry	Hdl Cholesterol Collected date	HDL Cholesterol Collected date
NHIC_ACS_74	Biochemistry	Hdl Cholesterol Collected time	HDL Cholesterol Collected time
NHIC_ACS_138	Biochemistry	Hdl Cholesterol lower	The LOWER normal value for the testkit use for the HDL cholesterol assay

NHIC_ACS_139	Biochemistry		Hdl Cholesterol higher	The HIGHER normal value for the testkit use for the HDL cholesterol assay
NHIC_ACS_75	Biochemistry		Triglycerides Result	Triglycerides Result
NHIC_ACS_76	Biochemistry		Triglycerides Unit	Triglycerides Unit
NHIC_ACS_77	Biochemistry		Triglycerides Collected date	Triglycerides Collected date
NHIC_ACS_78	Biochemistry		Triglycerides Collected time	Triglycerides Collected time
NHIC_ACS_140	Biochemistry		Triglycerides lower	The LOWER normal value for the testkit use for the Triglycerides assay
NHIC_ACS_141	Biochemistry		Triglycerides higher	The HIGHER normal value for the testkit use for the Triglycerides assay
NHIC_ACS_79	Biochemistry	2.35	Haemoglobin Result	Haemoglobin Result
NHIC_ACS_80	Biochemistry		Haemoglobin Unit	Haemoglobin Unit
NHIC_ACS_81	Biochemistry		Haemoglobin Collected date	Haemoglobin Collected date
NHIC_ACS_82	Biochemistry		Haemoglobin Collected time	Haemoglobin Collected time
NHIC_ACS_142	Biochemistry		Haemoglobin lower	The LOWER normal value for the testkit use for the HAEMOGLOBIN assay
NHIC_ACS_143	Biochemistry		Haemoglobin higher	The HIGHER normal value for the testkit use for the HAEMOGLOBIN assay
NHIC_ACS_83	Biochemistry		White Cell Count Result	White Cell Count Result
NHIC_ACS_84	Biochemistry		White Cell Count Unit	White Cell Count Unit
NHIC_ACS_85	Biochemistry		White Cell Count Collected date	White Cell Count Collected date
NHIC_ACS_86	Biochemistry		White Cell Count Collected time	White Cell Count Collected time
NHIC_ACS_87	Biochemistry		Platelet Count Result	Platelet Count Result
NHIC_ACS_88	Biochemistry		Platelet Count Unit	Platelet Count Unit

NHIC_ACS_89	Biochemistry		Platelet Count Collected date	Platelet Count Collected date
NHIC_ACS_90	Biochemistry		Platelet Count Collected time	Platelet Count Collected time
NHIC_ACS_91	Biochemistry		Crp Result	CRP Result
NHIC_ACS_92	Biochemistry		Crp Unit	CRP Unit
NHIC_ACS_93	Biochemistry		Crp Collected date	CRP Collected date
NHIC_ACS_94	Biochemistry		Crp Collected time	CRP Collected time
NHIC_ACS_144	Biochemistry		Crp lower	The LOWER normal value for the testkit use for the CRP assay
NHIC_ACS_145	Biochemistry		Crp higher	The HIGHER normal value for the testkit use for the CRP assay
NHIC_ACS_95	Echocardiogram	4.11+	DATE ECHOCARDIOGRAM	Indicate the APPOINTMENT DATE of the ECHOCARDIOGRAM
NHIC_ACS_96	Echocardiogram		TIME ECHOCARDIOGRAM	Indicate the APPOINTMENT Time of the ECHOCARDIOGRAM
NHIC_ACS_97	Echocardiogram	2.31	LV EJECTION FRACTION	State the EJECTION FRACTION of the LEFT VENTRICLE as a percentage
NHIC_ACS_98	Echocardiogram		LVEDD	State the LEFT VENTRICULAR DIASTOLIC DIAMETER measured during the ECHOCARDIOGRAM in 2D mode
NHIC_ACS_99	Echocardiogram		LVESD	State the LEFT VENTRICULAR SYSTOLIC DIAMETER measured during the ECHOCARDIOGRAM in 2D mode
NHIC_ACS_100	Echocardiogram		LVFS	State the LEFT VENTRICULAR FRACTIONAL SHORTENING measured during the ECHOCARDIOGRAM in 2D mode
NHIC_ACS_101	Echocardiogram		LV function	Description of LEFT VENTRICULAR function; Good, Moderate, Poor

NHIC_ACS_102	Angiography	4.13+	DATE OF CATHETERISATION OF HEART	Indicate the DATE of the CATHETERISATION OF HEART	
NHIC_ACS_103	Angiography		TIME OF CATHETERISATION OF HEART	Indicate the TIME of the CATHETERISATION OF HEART	
NHIC_ACS_104	Angiography		SEVERE CAD	Severe coronary artery disease (CAD) if >50% lesion in any vessel between two admissions - This is derived from NVD below	0 : No 1 : Yes
NHIC_ACS_105	Angiography		MAXIMAL STENOSIS	Indicate the maximal stenosis of epicardial vessel - This is derived from NVD below	
NHIC_ACS_106	Angiography		NUMBER OF VESSEL DISEASE (NVD)	Number of coronary vessels with stenosis >50% - If max steatosis in a segment is more than 50% then value is 1. NVD = sum of values for each three segments	0-3
NHIC_ACS_107	Revascularisation	2.18+	PCI	Indicates if a percutaneous coronary intervention (PCI) procedure was conducted	0 : No 1 : Yes
NHIC_ACS_108	Revascularisation		DATE OF PCI	Date of the PCI procedure	
NHIC_ACS_109	Revascularisation		TIME OF PCI	Time of the PCI procedure	
NHIC_ACS_110	Revascularisation	2.19+	CABG	Indicates if a coronary artery bypass grafting (CABG) operation was conducted	0 : No 1 : Yes
NHIC_ACS_111	Revascularisation		DATE OF CABG	Date of the CABG procedure	
NHIC_ACS_112	Revascularisation		TIME OF CABG	Time of the CABG procedure	

3.2 Troponin blood test

I standardised the different troponin assays between the academic centres by scaling the results using the ratio of the observed troponin value divided by the ULN for each troponin assay. Each centre measured troponin I or troponin T using either contemporary or high-sensitivity assays (**Table 3.2**). These tests yielded results in different measurable ranges, with unique cut-off points for the ULN. There is current lack of standardization of the various cardiac troponin assays as a physiological threshold for myocardial injury.[46,47] The ULN provided by the manufacturer for each troponin assay provides an analytical cut-off between normal and increased cardiac troponin levels. A positive troponin was defined as a result above the ULN for each troponin assay.

Table 3.2 - Troponin assays at participating cardiac centres

Assay Manufacturer / Platform	Gender-specific	99th percentile of the ULN (ng/L)	Troponin	Number of patients
Roche Elecsys cTnT	No	10	Tn T	347
Roche Cobas	No	14 (results reported with ULN of 10)	Tn T	17671
Roche Elecsys	No	14 (results reported with ULN of 13)	Tn T	43112
Abbott i-STAT	Female	15	Tn I	10981
Siemens Centaur XP	No	16	Tn I	22354
Roche Elecsys cTnT	No	30	Tn T	9420
Abbott cTnI	No	32	Tn I	50718
Abbott i-STAT	Male	34	Tn I	12853
Abbott Architect	No	40	Tn I	78037
Roche Cardiac Reader	No	50	Tn T	947

cTn, cardiac troponin; Tn, troponin

3.3 Overall summary of the NIHR HIC dataset

The NIHR HIC is an exciting resource for big data research projects spanning secondary care. It harnesses the granular data available in hospital EHRs from five centres across England. In chapters 2 and 3, I presented the work undertaken on data management and dataset curation for the first NIHR HIC Cardiovascular dataset.

The dataset developed captures patients with suspected ACS,

characterised by the request of a troponin test. I will use the PICO (population, intervention, control, and outcomes) criteria to frame each research question at the start of each research study chapter (Chapters 4-7). The PICO criteria[41] was used to ensure that each research question was well formulated to utilise the strengths of the dataset, both in terms of the scope of the patient population captured and the granularity of the data variables extracted.

3.4 Data processing

A number steps were taken to process the data prior to its use for research. The first view of the curated dataset from the raw data tables will be troponin-centric and capture information surrounding the first troponin measurement for each patient. Each patient therefore takes one row in the dataset with no duplication. Time '0' was defined as the date of the first troponin measurement with subsequent dates calculated in 'days' relative to this date.

3.4.1 Hospital admissions

- There were two separate data tables for A&E visits and hospital admissions. In creating one table with both, the A&E visit would need to be linked to an admission if the patient was admitted to hospital from A&E. These tables were joined by linking using the rule that a maximum of a 24 hour gap was assumed to connect an A&E visit t a hospital admission.

- If a patient had their first troponin measurement prior to their first admission, the dataset view will include only capture data surrounding the event related to the time they had their first troponin measured and not data captured within their known admission. For these patients, data surrounding their first troponin admission can be accessed in a seprate data view.

- ICD-10 codes were converted to descriptions, which one code per column, rather than separated by commas or pipes

- To match and an admission to a troponin value, it must fall within one day before admission or one day after discharge.

3.4.2 Echocardiography

- Left ventricular function was converted to the following:
    - 'Normal LV' into 'GOOD'
    - 'MODERATE' into 'Moderate LV Impairment'
    - 'POOR' into 'Severe LV Impairment'
    - 'UNKNOWN' into NULL

- LVFS was calculated when not provided and possible using [LVEDD]-[LVESD])/LVEDD)*100)

- Remove values for LVEF and LVFS where outside rage 0 – 100 N = 40

3.4.3 Coronary artery bypass grafting

- Set coronary artery bypass grafting (CABG) to ‘no’ when not ‘yes’ (i.e. NULL = No)

3.4.4 Angiography

- Data corrections - Incorrectly entered data values of ‘750%’ and ‘950%’ for percentage stenosis – these were set to ‘75%’ and ‘95%’ (These changes were from Imperial data and have been clinically verified changes)

3.4.5 All procedural investigations

- Count of total number of echocardiograms, percutaneous coronary interventions (PCIs), angiographies, and CABG operations across total patient record were added.

- In the exemplar dataset, the procedure takes the closest echocardiogram, angiography, CABG or PCI to the first troponin after -3 days before the troponin measurement to up to 4000 days after the troponin.

3.4.6 Demographics

- If multiple genders or ethnicities are provided then:
    - If exact duplicate – additional duplicates removed
    - If duplicated and one is valid value and remainder are unknown, valid value is kept and remainder are removed.
    - If multiple valid values exist then all are set to unknown.

- Minus ages were set to NULL

- Participants aged under 16 were removed from the dataset.

- Alive or dead status and time to death or last known alive.

- Patient last know alive – max event date for a patient (any data item except date of death)

3.4.7 Biochemistry

- In the exemplar biochemistry tests, the closest to first troponin was included and calculated as closest between -3 days before troponin to a maximum of 4000 days.

TROPONIN

- Troponin unit conversion
  - "ng/ml" and "microg/l" are the same as "μg/l" and are converted to "μg/l"
  - All units are converted to "ng/l" using "ng/l" = "μg/l" * 1000

- For string data, a flag was added if '<' or '>' was reported – the value was split to record only the numeric value.

- Text string results conversions - Any value with "<" or ">" or similar in the results string is converted as follows:

Reported value	Converted Value
<0.001	0.0001
<0.003	0.0001
<0.01	0.0001
<0.02	0.0001
<0.032	0.0001
<0.04	0.0001
<0.05	0.0001
<0.1	0.0001
<0.10	0.0001
<0.2	0.0001
<0.20	0.0001
<1	0.0001
<2	0.0001
<3	0.0001
<30	0.0001
<32	0.0001
>100000	101000
>50	51
>50.00	51
>50.000	51
>50000	51000
>50000.0	51000
>50000.000	51000
>500000.0	510000
Greater than 40.00	41
Greater than 40000	41000

- Any none numeric values (except for text string conversions) are set to NULL.

- Any value over 1000000 μg/l are excluded (n=1)

eGFR (estimated glomerular filtration rate)

- eGFR – unit conversions - eGRF unit “ml/min/1.73^2” = “ml/min” / 1.72
  - If 'mL/min' then 'ml/min/1.73m2' = 'mL/min'/1.73

- Text string results conversions:

Reported value	Converted Value
>90	91
>90	91
Greater than 60	61.00

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.
  - delete '-'
  - delete > 1000

PLATELET COUNT

- Platelet count - Unit conversions - All results provided as “10*9/L” (10*9/L, x10*9/L, x10^9, x10^9/l) were considered the same

- Exclusions
  - Any non-numeric values (except for text string conversions) are set to NULL.

- The following values were set to NULL.

'.50100..'
'42...'
'91..'
'79..'
'49...'
'..99'
'..3024..'
'346..'
'17..'
'6..'

WHITE CELL COUNT

- Unit conversions - All results provided as “10*9/L” (10*9/L, x10*9/L, x10^9, x10^9/l) were considered the same

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

- The following values were set to NULL

'7.1.179.1.17'
'6.12.1612.01.17'
'428930169321012017.1449.'
'428904682005012017.22873.'
'4287267917.18112016.'
'4.1.175.1.17'
'31.12.163.1.17'
'28.11.1629.11.16...'
'25797165618.12.16.','2.12.165.12.16'
'2..11096.','17.11.16.27.11.16.1678.'
'12.12.1613.12.16'
'060117.090117100117..'
'.72.'
'.72.'
'.6800.'
'.64.5.2640121457.11.16'
'.4.0.1157144217117','.2646601.11'
'.26466.'
'.2646319.22'
'.26382125427.11.16.'
'.26017.'
'.25797.'
'.25797.'
'.24111.'
'.23.56'
'.2288008.26'
'.2171326211.'
'.2125610.30.'
'.21178.'
'.21.5620.07.2015'
'.1912.'
'.12.46'
'.112319.46'
'.1016161015.01.17.'
'.07768752641.'
'.07.22','..5.'
'..4577730'
'..4017.17.30.'
'..264660411'
'..2640118.11'

'..26401.'
'..21051.'
'..1857258.'
'...1143.'
'428438952228082016.'
'428739841728112016.'
'428827886520122016.'
'428714447810112016.'
'428739718727112016.'
'427234650625082015.'
'428833275623122016.'
'.211032638026466'
'280916051016.'
'101116111116.'
'010916070916.'
'101016121016.'
'171116191116.'
'071016311016.'
'230117240117.'
'230117240117.'
'150816160816.'
'011116111116.'
'011016021016.'
'021116031116.'
'140916150916.'
'270716280716.'
'121216151216'
'021116041116'
'171116181116'
'090916081116'
'22121615117.'
'8.217192300'
'.1650291116'
'.264011635'
'.264532145'
'264630610.'
'1411715117'
'2213312.02'
'.264662036'
'.25814'
'.25797'
'22491.'
'.26385'
'26211.'
'.25802'
'.26463'
'26118.'
'31454.'
'.21936'
'.26466'
'.25017'

'.26401'
'.21178'
'.26466'
'.24111'
'21178.'
'.22440'
'.21815'
'.1449'
'21414'
'71016'
'26401'
'26401'
'26211'
'26116'
'22881'
'22874'
'21861'
'1449'
'1413'
'1037.'

HbA1c

- HbA1c - Unit conversions - mmol/mol = [% – 2.15] x 10.929 (also applied to upper and lower limits)

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

- If old and new test unit results were received on the same day then old results were excluded.

UREA

- Urea - Unit conversions - “μmol/l” = “mmol/l” * 1000

- Text string results conversions:

Reported value	Converted Value
<0.5	0.5
<1.0	1
>44.6	44.6

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

- Set values less than 0.1 to NULL

CREATININE

- Creatinine - Unit conversions
  - “micromol/L” is same as “μmol/l”
  - “μmol/l” = “mmol/l” * 1000

- Text string results conversions:

Reported value	Converted Value
<0.04	0.04
<10	10
<18	18

- Exclusions - All creatinine with units 'mL/min', 'mmol/24', 'mmol/24Hr, “IU/L”

- Any non-numeric values (except for text string conversions) are set to NULL.

- Any results under 5 μmol/l or over 3000 μmol/l were set to NULL as out of expected range

SODIUM

- Sodium - Unit conversions - All provided as mmol/L

- Text string results conversions:

Reported value	Converted Value
<100	100
<100.0	100
>200	200

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

- Value of '25797' was set to NULL

C-reactive protein (CRP)

- CRP – Corrections - '156T' Changed to 156
- Unit conversions - All provided as mg/L
- Text string results conversions:

Reported value	Converted Value
<0.1	0.1
<0.2	0.2
<0.20	0.2
<0.5	0.5
<0.6	0.6
<1	1
<1.0	1
<2	2
<2.0	2
<5	5
<6	6
<8	8
>156	156
>156.0	156
>160	160
>160.0	160
>320.0	320
>479.99	479.99
>480.0	480
>480.00	480
Greater than 250	250

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

TRIGLYCERIDES

- Triglycerides - Unit conversions - All provided as mmol/L
- Text string results conversions:

Reported value	Converted Value
<0.08	0.08

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

HDL CHOLESTEROL

- HDL Cholesterol - Unit conversions - All provided as mmol/L
- Text string results conversions:

Reported value	Converted Value
<0.1	0.1
<0.13	0.13

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.
- Set results with value less that 0.1 to NULL

GLUCOSE

- Glucose - Unit conversions - All provided as mmol/L
- Text string results conversions:

Reported value	Converted Value
<0.5	0.5
<1.0	1.0
<1.4	1.4
>44.4	44.4

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

- Values greater than 200 were set to null

POTASSIUM

- Potassium - Unit conversions - All provided as mmol/L

- Text string results conversions:

Reported value	Converted Value
>10.0	10
>10.00	10

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

HAEMOGLOBIN

- Haemoglobin - Unit conversions - 'g/dl' = 'g/L' / 10

- Text string results conversions:

Reported value	Converted Value	
>10.0	10	
>10.00	10	
11137	111	
7626459	76	*
7315025	73	*
696700.	69	*
64951510.43'	64	*
4736691	47	*
4426401	44	*
6107	61	*
154370	154	*
11236690	112	*

*Results clinically validated

- Exclusions - Any none numeric values (except for text string conversions) are set to NULL.

3.5 Quality of discharge coding

Patients who were admitted to hospital had International Statistical Classification of Diseases and Related Health Problems (ICD) discharge codes[7]. Payment-by-results is an initiative which uses the coding data to direct health-care funding. Furthermore, routinely collected coding data are increasingly being used for clinical research. The accuracy of routinely collected has previously been assessed against various standards, including a comparison with independent review of case notes or with clinical registry data. This requires reliable data from these data sources, which also depend on the those inputting the data. A systematic review and meta-analysis of discharge coding accuracy has shown that NHS administrative data accuracy has improved in recent years, although it still remains considerably low at 80.3% for primary diagnosis coding.[48]

An independent report assessing the accuracy of clinical coding found differences in the standards of clinical coding across NHS Trusts. While the error rates in coding were as low as 1.1% in some Trusts, others had error rates of up to 45.8%.[49] The performance was worse in the samples that focused on coding patient comorbidities, with consistent underreporting. However, the issues were mainly around those comorbidities considered less relevant with national guidance stating that “non-mandatory” comorbidities should only be coded if they are identified as being clinically relevant. The difficulty lies with clinicians not making the source documentation transparent of whether or not a comorbidity is clinically relevant.

3.6 Using discharge coding to define variables

ICD-10 codes can be used to classify exposure and outcome variables for research. While it is difficult to quantify the level of impact of inaccurate coding on research output, it Is assumed there are no systematic inaccuracies in clinical coding. A research study assessing the impact of an exposure variable on an outcome, which relies on clinical coding to define variables, assumes that the level of inaccuracy will be the same across each variable. The ICD-10 codes are listed in order of diagnostic columns; primary diagnosis, secondary diagnosis and comorbidities, in sequential order.

For the studies performed in this thesis, I used the ICD-10 discharge coding to define variables as follows:

- Patient cohorts
    - ICD-10 code(s) used: first diagnostic column, which defines the primary diagnosis and second diagnostic column, which defines

the secondary diagnosis, where relevant

- Comorbidities
    - ICD-10 codes used: third diagnostic column, onwards
    - Due to concerns regarding the accuracy of using ICD-10 codes for recording “less relevant” co-morbidities, only common comorbidities will be used, such as cardiovascular risk factors (hypertension, smoking, etc.) and common chronic conditions (chronic kidney disease, heart failure, etc.). The prevalence of these comorbidities in the dataset will be compared to the literature to ensure feasibility and accuracy.

- Outcomes
    - ICD-10 codes used: first diagnostic column
    - The first diagnostic column will be used to define the primary reason for readmission to capture relevant outcomes, including:
        - Heart failure
        - Myocardial infarction
        - Bleeding
        - Stroke

3.7 External validation

A number of methodological considerations were considered to improve the external validity of the results produced. External validity refers to the generalisability of the results, not only nationally, but also internationally. Without high external validity, the research results cannot be applied outside of the context of the studies. A number of steps were taken to counter the threat to external validity.

- Representative sample
    - The HIC dataset includes data from five NHS Trusts, which provide care to distinct geographical regions across England. The population residing in each region differ in terms of age, gender and ethnicity to provide a broad demographic profile for research studies.

- Probability sampling
    - To avoid sampling bias, probability sampling will be used, where everyone in the population captured in the HIC dataset will have an equal chance of being selected for a study sample.

- Replicability

- To allow the studies conducted using the HIC dataset to be replicated using other datasets, both nationally and internationally, the methods section will be transparent to report the sample population and statistical methods used for each study.

- Field experiments
  - Factors such as the setting and location may limit the generalisability of the results. The results will be stratified by the contributing centre. Furthermore, sensitivity analyses will be performed, where relevant.

Chapter 4. Early translational research – Discovery of underlying disease mechanism

4.1 CRP-RISK study

In this Chapter, I will explore the role of big data for early discovery phase translational research by demonstrating how big data can contribute to early scientific advancement in cardiovascular medicine through discovery of underlying disease mechanisms. I will evaluate whether a mildly elevated high-sensitivity C-reactive protein (hsCRP) level was associated with mortality risk, beyond troponin level, in patients with suspected acute coronary syndrome.

The rationale for answering this particular research question is summarised in **Table 4.1**.

Table 4.1 - Rationale for CRP-RISK study

PICO Element	Description	Rationale
Population/ Patient/ Problem	Patients in whom troponin testing has been performed for clinical purposes (suspected ACS)	The NIHR HIC dataset is troponin centric, as all patients were only included in the dataset if they had a troponin measured. In accordance with clinical guidelines, all patients with suspected ACS will undergo a troponin test. As a result, there would be no selection bias as all patients presenting to hospital with suspected ACS will be captured.
Intervention	hsCRP level	Levels of hsCRP may reflect background atherosclerotic disease activity. hsCRP levels may prove useful in prognostication beyond routinely studied parameters. Data on hsCRP level was extracted if a hsCRP blood test was measured in patients who underwent troponin testing. However, selection bias may still remain an issue as not all patients who had a troponin measured will undergo a hsCRP blood test.
Comparison	Troponin level	All patients in the dataset will have a troponin blood test measured.
Outcome	All-cause mortality	All-cause mortality is the outcome of interest to most patients. Furthermore, all-cause mortality is the only outcome where the diagnosis is free from bias.

4.2 Abstract

Background

The use of high-sensitivity CRP (hsCRP) as a biomarker to risk stratify patients for advanced cardiovascular therapies has not been well investigated in the modern era. The significance of a mildly elevated hsCRP, beyond troponin, in patients presenting with suspected acute coronary syndromes (ACS), who had a troponin measured, is unknown. I evaluated whether there was an association between mildly elevated hsCRP and mortality risk, beyond troponin, in patients with suspected ACS.

Methods

I conducted a retrospective cohort study based in 257,948 patients who had a troponin measured across 5 centres in the UK between 2010 and 2017. Based on their hsCRP level, patient groups were created (<2, 2-4.9, 5-9.9 and 10-15 mg/L). 3-year mortality was the primary outcome. Multivariable Cox regression was used to assess the association between hsCRP levels and mortality.

Results

A total of 102,337 patients were included (38,390 (hsCRP <2 mg/L), 27,397 (2-4.9 mg/L), 26,957 (5-9.9 mg/L) and 9,593 (10-15 mg/L)). There was a positive, graded relationship between hsCRP level and all-cause mortality at three-years ((HR 1.32 (95% CI, 1.18-1.48) for hsCRP 2.0-4.9mg/L, 1.40 (1.26-1.57) for hsCRP 5-9.9 mg/L, and 2.00 (1.75-2.28) for hsCRP 10-15 mg/L, compared to hsCRP <2 mgL)). This relationship was independent of the troponin result, and was also seen in those with an ACS diagnosis.

Conclusion

In a large cohort of patients with suspected ACS, mildly elevated hsCRP was a clinically relevant prognostic marker in addition to troponin, which points to its potential utility in identifying patients who may benefit from novel treatments targeting inflammation.

4.3 Introduction

Inflammation has been shown to be involved in the mechanism for atherothrombosis.[42] It is debated as to whether the immune system has a protective or pathogenic role in atherothrombosis and subsequent myocardial infarction and other cardiovascular (CV) events. Elements of the immune system can be targeted for novel therapies, or can be measured as potential biomarkers to help risk stratify patients for developing CV disease (CVD).[48,49] The most widely evaluated biomarker is C-reactive protein (CRP) given its common use in clinical practice.

In general and acute coronary syndrome (ACS) populations, observational studies have shown that patients with high levels of CRP have a poorer prognosis, which may be related to underlying atherosclerosis.[48,50-52] Further studies in healthy individuals[53,54] and patients with CVD[55-58] show that baseline CRP levels are independent predictors of CV events and mortality. Rather than a therapeutic target, CRP is likely to be an end-protein product of a pathway involving IL-6.[59,60]

Using high-sensitivity CRP (hsCRP) as a biomarker for selecting patients for more advanced CV therapies targeting the immune system, was demonstrated in CANTOS (Canakinumab Anti-inflammatory Thrombosis Outcome Study). This was a trial randomising patients to either canakinumab (a human monoclonal antibody targeted against an upstream target of IL-6, called IL-1β), or placebo, with the primary outcome being CV events.[61] Over a median of 3.7 years follow-up, patients randomised to the canakinumab arm had a 15% reduction in CVD events, which was associated with a reduction in hsCRP levels.[61] In addition, patients that failed to achieve hsCRP levels <2mg/L whilst on treatment with canakinumab did not have a reduction in major adverse cardiac events.[62]

Results from CIRT (Cardiovascular Inflammation Reduction Trial) raised doubt on the effectiveness of using rheumatological agents to modulate inflammation in atherosclerosis, after the trial showed that low-dose methotrexate had no effect on the incidence of CVD compared to placebo.[63] However, the eligibility criteria of the trial was not restricted to those with elevated levels of hsCRP at baseline. There is recent evidence supportive of a strategy of targeting inflammation in CVD in patients within 30-days of ACS or with chronic coronary disease, with the COLCOT (Colchicine Cardiovascular Outcomes Trial)[64] study and the LoDoCo2 (Low Dose Colchicine for secondary prevention of cardiovascular disease trial)[65] study reporting a reduction in CV endpoints using colchicine, which works on several immune pathways, including targeting IL-1β. However, measurements of inflammation (including CRP) at baseline was not available from these studies.

hsCRP measurments in patients presenting to hospital with ACS may either reflect background atherosclerosis, or the sequelae following the ACS itself. Nevertheless, levels of hsCRP on admission may be useful for prognostic prediction beyond using routinely studied biomarkers, including troponin.

In this study, I evaluated the prognostic significance of mildly raised hsCRP (up to 15 mg/L) in addition to troponin in a large cohort of patients presenting to hospital with suspected ACS. My hypothesis was that a mild elevation in hsCRP in thee patients would be associated with an increased mortality risk, beyond troponin.

4.4 Methods

4.4.1 Study design and participants

This was a retrospective cohort study using the NIHR HIC data of 257,948 patients who had a troponin for a clinical reason, across five centres in the UK between 2010 and 2017.

For patients who had multiple admissions where troponin was measured, the first episode of care was used. There was an intentional focus to those patients with a normal white cell count (WCC) in order to exclude those who were overtly septic as they may have an effect on hsCRP levels. Furthermore, all patients with a hsCRP level of greater than 15mg/L were also excluded for similar reasons. All patients were followed up until death or censoring in April 2017.

All analyses on hsCRP were performed using the first measured result during the index admission. The peak troponin level was used for all troponin analyses. If a patient only had one troponin measured, this was considered the peak troponin level. The different troponin assays were standardised across the cardiac centres using the ratio of the troponin value divided by the ULN for each respective troponin assay. Each centre measured troponin T or troponin I using either high-sensitivity or standard (contemporary) assays (**Table 3.2**). There is are currently no physiological thresholds to define myocardial injury.[46,47]

All patients admitted to hospital had International Statistical Classification of Diseases (ICD) discharge codes. The ICD-10 codes were used to classify patients as having an ACS event during their admission (**Table 4.2**).[7]

Table 4.2 - International Statistical Classification of Diseases (ICD)-10 codes used to indicate an ACS diagnosis

ICD-10 Code	Category
I20.0	Unstable angina
I21.0	Acute transmural myocardial infarction of anterior wall
I21.1	Acute transmural myocardial infarction of inferior wall
I21.2	Acute transmural myocardial infarction of other sites
I21.3	Acute transmural myocardial infarction of unspecified site
I21.4	Acute subendocardial myocardial infarction
I21.9	Acute myocardial infarction, unspecified
I22.0	Subsequent myocardial infarction of anterior wall
I22.1	Subsequent myocardial infarction of inferior wall
I22.8	Subsequent myocardial infarction of other sites
I22.9	Subsequent myocardial infarction of unspecified site
I24.8	Other forms of acute ischaemic heart disease
I24.9	Acute ischaemic heart disease, unspecified

Four categories of hsCRP were defined; <2mg/L (a normal hsCRP), 2-4.9 mg/L, 5-9.9 mg/L and 10-15 mg/L. These four hsCRP groups were also subdivided based on whether the accompanying troponin result was positive (>ULN) and whether patients were diagnosed with an ACS.

4.4.2 Statistical analysis

Comparisons of baseline characteristics between hsCRP groups was analysed using the Kruskal-Wallis test and $\chi 2$ test for trend for continuous and categorical variables, respectively. Spearman's correlation coefficient was used to assess the relationship between hsCRP and troponin levels on a continuous scale. Kaplan-Meier plots were used to display cumulative mortality, with group comparisons made using the log-rank statistic.

Multivariable Cox regression analysis was used to investigate whether the hsCRP independently predicted all-cause mortality after adjusting for demographic and clinical factors. Time-varying coefficients were used with follow-up divided into time intervals. To model non-linear relationships, restricted cubic splines were used in the Cox regression models.

Three models were established to assess the predictive role of hsCRP on mortality beyond standard risk factors and troponin. Model 1 was adjusted for age, gender, creatinine, and haemoglobin; model 2 was also adjusted for troponin; model 3 was also adjusted for hsCRP.

The area under receiver operating characteristic curves (AUROC) was

calculated for each model and compared using the DeLong approach. The continuous net reclassification index (NRI) and integrated discrimination improvement (IDI) was also calculated with the survival data.

The negative predictive values of a negative hsCRP, negative troponin or both being negative was plotted against hypothetical mortality rates in order to evaluate their performance in predicting mortality.

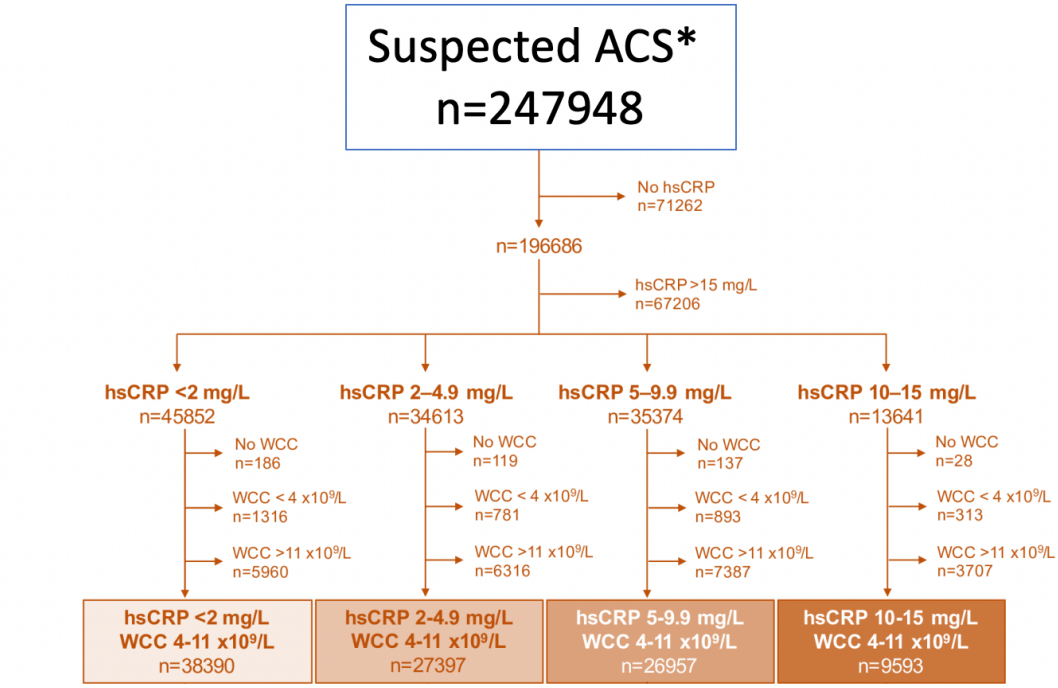
Statistical analyses were performed using R 3.5.0 or MedCalc version 15.8.

4.5 Results

4.5.1 Baseline population characteristics

Of the 257,948 patients presenting with suspected ACS across the five cardiac centres, 71,262 did not have a hsCRP test performed and were excluded, and 67,206 were excluded based on their hsCRP leve beingl >15mg/L. The remaining 129,480 patients were divided into the four hsCRP groups. After excluding those with abnormal WCCs (or those without a WCC blood test), 102,337 patients remained in the analysis. The median duration of follow-up was 3.3 years (interquartile range 1.4-5.1). The study cohort was divided into four hsCRP categories (**Figure 4.1**). The majority of patients were admitted to hospital (53.3% (n=54,534)) and 10.8% (n=5,910) of these diagnosed with an ACS.

Figure 4.1 - Flow of patients through CRP-RISK study



*ACS, acute coronary syndrome; hsCRP, high-sensitivity C-reactive protein; WCC, white cell count. *suspected ACS characterised by the request of a troponin.*

Table 4.3 summarises the limits of detection and the ULN for the hsCRP assays used across the cardiac centres.

Table 4.3 - hsCRP assays at participating cardiac centres

Assay Manufacturer - Platform	99th percentile of the ULN (mg/L)	Limit of detection (mg/L)	Number of patients (Cardiac Centre)	Assay Manufacturer - Platform
Roche - Cobas	4	0.1 (reported <1 mg/L)	19076 (GSTT)	Roche - Cobas
Abbott - Architect	5	0.1	24295 (ICHNT) 33623 (OUH) 14524 (UCL)	Abbott - Architect
Abbott - Architect	10	0.2	10819 (ICHNT)	Abbott - Architect

GSTT; Guy's and St Thomas' NHS Foundation Trust; ICHNT, Imperial College Healthcare NHS Trust; OUH, Oxford University Hospitals NHS Foundation Trust; UCL, University College London Hospitals NHS Foundation Trust; ULN, 99th percentile of the upper limit of normal.

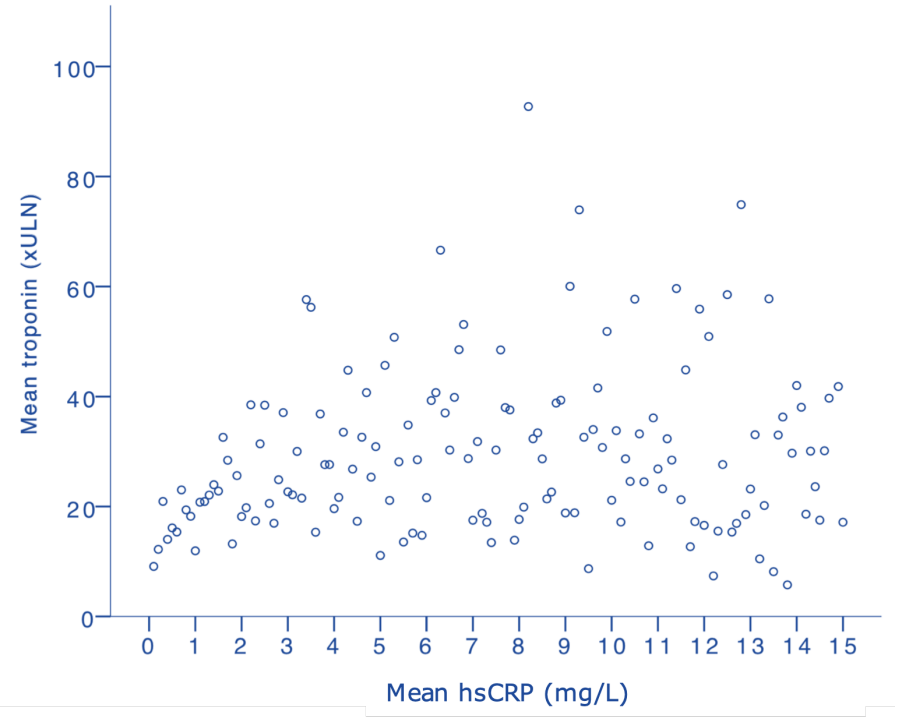
The baseline characteristics for the patients included in the study are shown in **Table 4.4**. There was a weak correlation between hsCRP and troponin level (R=0.15 (p<0.001)) (**Figure 4.2**). The median number of days between the admission hsCRP and peak troponin was 3 days (IQR 0-5), which indicates the admission hsCRP represents a presentation value instead of a value in response to the index event.

Table 4.4 - Baseline clinical characteristics of CRP-RISK study

	hsCRP <2 mg/L (n=38390)		**hsCRP 2-4.9 mg/L (n=27397)**		**hsCRP 5-9.9 mg/L (n=26957)**		**hsCRP 10-15mg/L (n=9593)**		**P-value**
	n (median (IQR) or number (% of total patients))	Total patients (% of all patients)	n (median (IQR) or number (% of total patients))	Total patients (% of all patients)	n (median (IQR) or number (% of total patients))	Total patients (% of all patients)	n (median (IQR) or number (% of total patients))	Total patients (% of all patients)	
General demographics									
Age (years)	59 (45 - 75)	38374 (99.96)	65 (51 - 79)	27381 (99.94)	64 (50 - 79)	26952 (99.98)	70 (54 - 82)	9589 (99.96)	<0.001
Male	21566 (56.20)	38374 (99.96)	14377 (52.51)	27378 (99.93)	14292 (53.03)	26951 (99.98)	4774 (49.79)	9589 (99.96)	<0.001
Ethnicity									
White	22824 (71.04)	32130 (83.69)	17157 (73.69)	23284 (84.99)	16698 (70.54)	23671 (87.81)	6095 (74.37)	8195 (85.43)	<0.001
South Asian	1445 (4.50)		993 (4.26)		879 (3.71)		304 (3.71)		
Black	3018 (9.39)		2197 (9.44)		3367 (14.22)		960 (11.71)		
Other	4843 (15.07)		2937 (12.61)		2727 (11.52)		836 (10.20)		
Haematology / Biochemistry									
Haemoglobin (g/dL)	13.8 (12.7 - 14.9)	38389 (100)	13.6 (12.4 - 14.7)	27395 (99.99)	13.4 (12.1 - 14.5)	26954 (99.99)	13.0 (11.7 - 14.2)	9592 (99.99)	<0.001
White cell count ($x10^9$/L)	7.1 (5.9 - 8.5)	38390 (100)	7.5 (6.2 - 8.9)	27397 (100)	7.6 (6.3 - 9.0)	26957 (100)	7.9 (6.5 - 9.3)	9593 (100)	<0.001
Platelet count ($x10^9$/L)	225 (190 - 264)	38373 (99.96)	230 (192 - 272)	27391 (99.98)	226 (185 - 270)	26941 (99.94)	230 (186 - 281)	9589 (99.96)	<0.001
Creatinine (µmol/L)	74 (64 - 88)	38279 (99.71)	76 (65 - 92)	27338 (99.78)	76 (64 - 94)	26882 (99.72)	78 (65 - 100)	9584 (99.91)	<0.001
Positive troponin	6099 (15.9)	38390 (100)	5937 (21.7)	27397 (100)	7402 (27.5)	26957 (100)	3126 (32.6)	9593 (100)	<0.001
Troponin level (xULN)	0.003 (0.003 - 0.4)	38390 (100)	0.003 (0.003 - 0.7)	27397 (100)	0.23 (0.003 - 1.25)	26957 (100)	0.06 (0.003 - 1.7)	9593 (100)	<0.001

hsCRP, high-sensitivity C-reactive protein; ULN, 99th percentile of the upper limit of normal. *P-values were calculated using Kruskal-Wallis one-way analysis of variance and $\chi 2$ test for trend for continuous and categorical variables, respectively*

Figure 4.2 - Correlation between hsCRP and troponin level



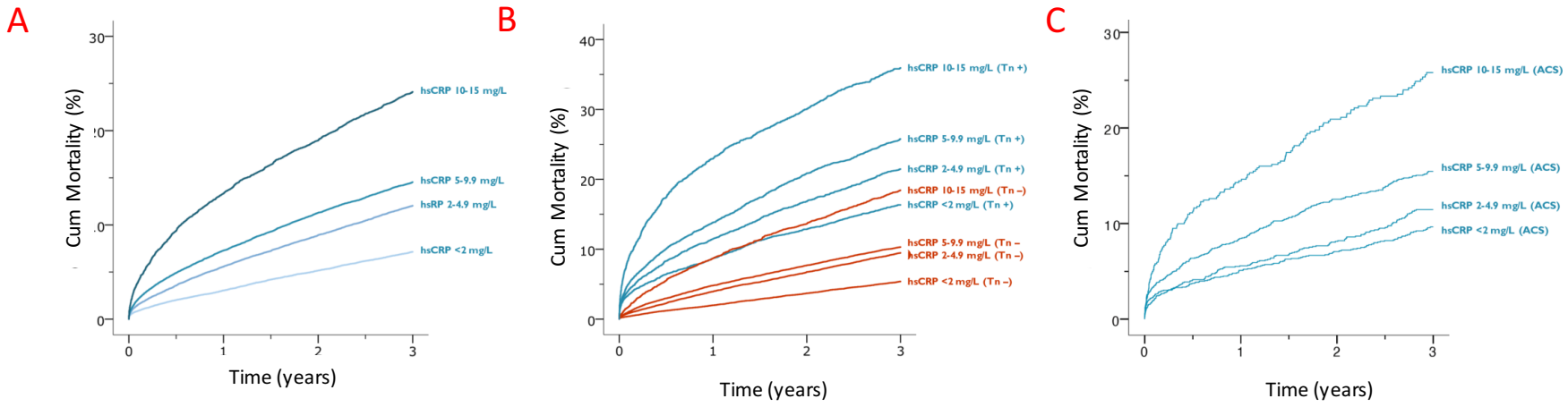
Pearson correlation R=0.019, p<0.0001. hsCRP, high-sensitivity C-reactive protein; ULN, 99th percentile of the upper limit of normal.

4.5.2 Higher hsCRP predicts significantly increased mortality risk

There was an increment in mortality risk with higher consecutive hsCRP groups (**Figure 4.3A**). **Figure 4.3B** shows the highest mortality for patients in the highest hsCRP group (10-15 mg/L) who also had a positive troponin (36% 3 year mortality). The additive effect of hsCRP on top of troponin on mortality risk was seen in all hsCRP groups. This graded relationship was also seen across the hsCRP troponin-negative groups. **Figure 4.3C** shows this relationship persisted when restricting the cohort to those with ACS (**Figure 4.4**).

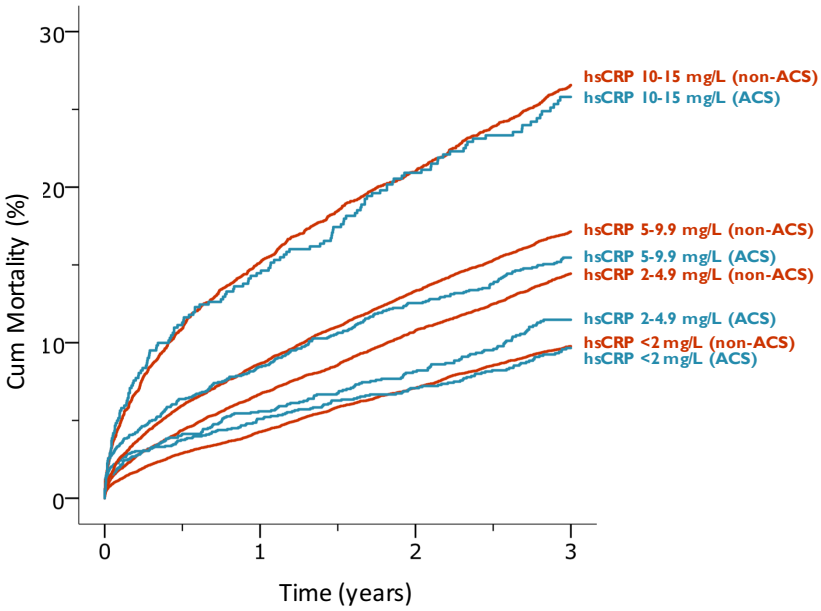
After multivariable adjustment, hsCRP was an independent predictor of mortality over time (**Figure 4.5**). The graded relationship persisted up to 3 years in the study. The hazard ratios for 3-year mortality risk in the hsCRP groups stratified by troponin level are shown in **Table 4.5**.

Figure 4.3 - Unadjusted Kaplan-Meier mortality curves by (A) hsCRP level, (B) hsCRP level and troponin positivity, and (C) hsCRP level in ACS patients



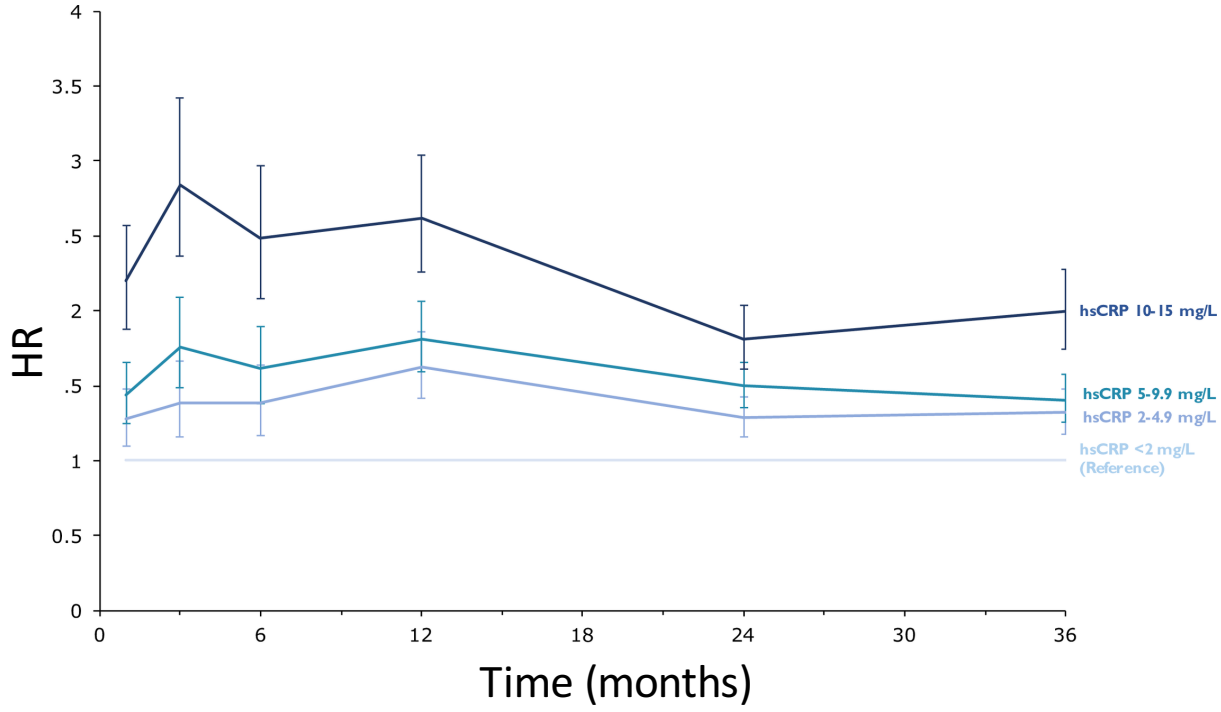
ACS, acute coronary syndrome; hsCRP, high-sensitivity C-reactive protein; Tn +, troponin positive, Tn –, troponin negative.

Figure 4.4 - Unadjusted Kaplan-Meier mortality curves by ACS diagnosis



ACS, acute coronary syndrome; hsCRP, high-sensitivity C-reactive protein.

Figure 4.5 - Multivariable Cox regression analysis with time-varying coefficients



hsCRP, high-sensitivity C-reactive protein. Hazard ratios are adjusted for age, sex, haemoglobin, WCC, platelet count, creatinine and troponin level. The adjusted hazard ratios are displayed at the end of each time period. The error bars indicate 95% confidence intervals.

Table 4.5 - Adjusted hazard ratios for 3-year mortality according to troponin and hsCRP stratified groups

hsCRP (mg/dL)	Troponin level	Number of patients	Hazard ratio (95% confidence interval)	P-value
<2	Negative	32291	Reference	-
2–4.9	Negative	21460	1.48 (1.38 – 1.58)	<0.001
5–9.9	Negative	19555	1.72 (1.60 – 1.84)	<0.001
10–15	Negative	6467	2.51 (2.32 – 2.72)	<0.001
<2	Positive	6099	1.75 (1.61 – 1.90)	<0.001
2–4.9	Positive	5937	2.18 (2.02 – 2.36)	<0.001
5–9.9	Positive	7402	2.45 (2.28 – 2.63)	<0.001
10–15	Positive	3126	3.47 (3.20 – 3.77)	<0.001

Multivariable Cox regression analysis was adjusted for age, gender, haemoglobin, white cell count, platelet count, creatinine.

4.5.3 hsCRP significantly improves mortality risk prediction

Three models were established to assess the predictive role of hsCRP on mortality beyond standard risk factors and troponin. Model 1 was adjusted for age, gender, creatinine, and haemoglobin; model 2 was also adjusted for troponin; model 3 was also adjusted for hsCRP. Each successive model was better able to discriminate mortality risk than its precursor ($p<0.001$) (**Table 4.6**). Inclusion of both hsCRP and troponin resulted in the most robust risk discrimination. The addition of model 3 over model 2 led to an IDI of 0.3% and 0.9% at 30-days and 3-years, respectively (all $p<0{\cdot}001$) and the highest overall NRI (all $p<0{\cdot}001$).

Table 4.6 - hsCRP risk model discrimination, calibration and reclassification

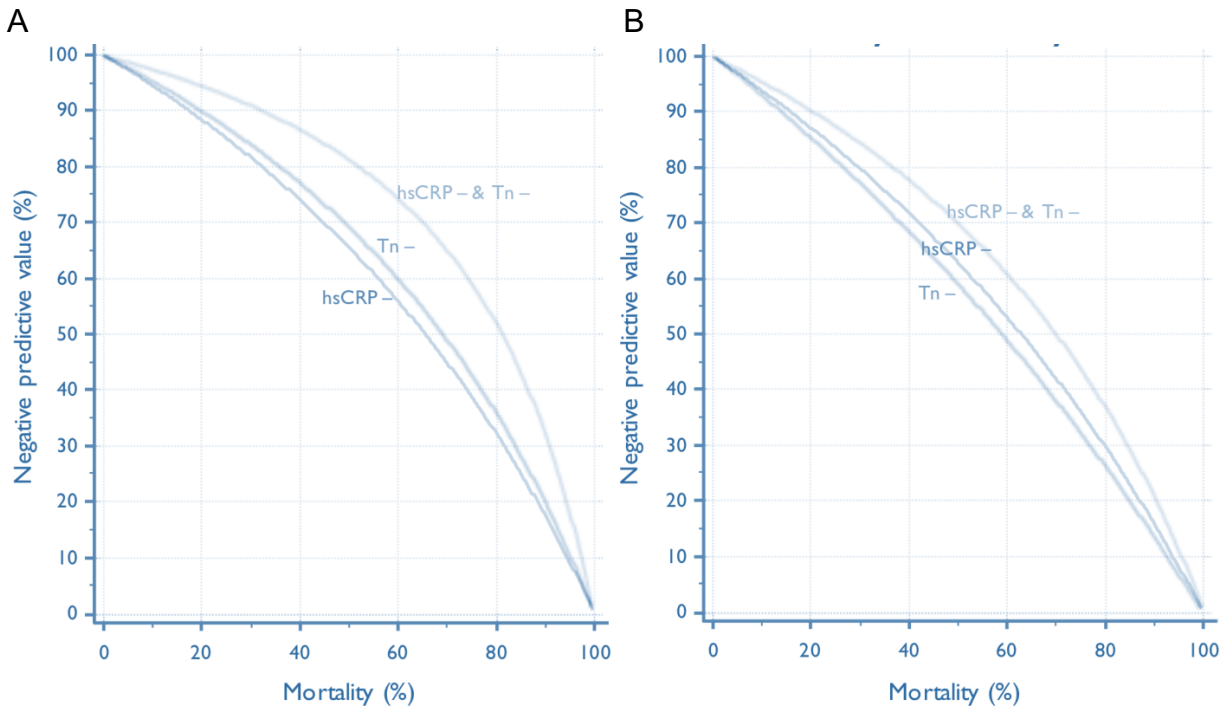
	30-day mortality			*3-year mortality*		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Basic Demographic / Biochemistry	Age, sex, haemoglobin, creatinine	Age, sex, haemoglobin, creatinine	Age, sex, haemoglobin, creatinine	Age, sex, haemoglobin, creatinine	Age, sex, haemoglobin, creatinine	Age, sex, haemoglobin , creatinine
+ Troponin		+ Troponin (positive vs negative)	+ Troponin (positive vs negative)		+ Troponin (positive vs negative)	+ Troponin (positive vs negative)
+ hsCRP			+ hsCRP groups (<2, 2-4.9, 5-9.9, 10-15)			+ hsCRP groups (<2, 2-4.9, 5-9.9, 10-15)
						Discrimination
AUROC	0.747	0.806	0.812	0.790	0.797	0.803
95% CI	0.745 – 0.750	0.803 – 0.808	0.810 – 0.815	0.787 – 0.793	0.794 – 0.800	0.800 – 0.806
P-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
P-value (vs Model 1)	-	<0.001	<0.001	-	<0.001	<0.001
P-value (vs Model 2)	-	-	<0.001	-	-	<0.001
						Reclassification
		vs Model 1	vs Model 2		vs Model 1	vs Model 2
IDI (%)	-	1.1	0.3	-	1.0	0.9
95% CI	-	0.9 – 1.3	0.2 – 0.4	-	0.8 – 1.2	0.7 – 1.1
P-value	-	<0.001	<0.001	-	<0.001	<0.001
NRI (%)	-	42.6	18.4	-	20.9	13.9
95% CI	-	39.7 – 45.3	14.6 – 21.0	-	19.2 – 22.0	12.7 – 14.9
P-value	-	<0.001	<0.001	-	<0.001	<0.001

AUROC, area under the receiver operating characteristic curve; hsCRP, high-sensitivity C-reactive protein; IDI, integrated discrimination improvement; NRI, net reclassification index.

Figure 4.6 shows the relationship between the negative predictive value of hsCRP and troponin with hypothetical mortality.

.

Figure 4.6 - Relationship between negative predictive value of hsCRP and troponin testing with (A) 30-day and (B) 3-year mortality



hsCRP –, negative high-sensitivity C-reactive protein; Tn –, negative troponin

4.6 Discussion

4.6.1 Summary of main findings

There was a positive, graded relationship between mildly elevated hsCRP and all-cause mortality over 3 years in this retrospective cohort study. The mortality risk associated with hsCRP level was independent of age, gender, haemoglobin, WCC, platelet, creatinine and troponin positivity, indicating its potential value as a useful biomarker in patients with suspected ACS. Compared to a normal hsCRP level of <2 mg/L, a mildly elevated hsCRP of between 10-15 mg/L increased mortality risk by 2.2-fold at 30-days, which persisted at 3 years. Surprisingly, patients with a negative troponin but a hsCRP level between 10-15 mg/L had a similar 3-year mortality risk to patients who had a positive troponin and a normal hsCRP level of <2 mg/dL. Patients were also better re-classified into at-risk mortality groups with the inclusion of hsCRP in addition to troponin, indicating its potential utility to risk stratify patients in this clinical setting.

Strict exclusion criteria were used to reduce the proportion of patients with either infective or inflammatory disorders. One of the key findings of this study was that patients with suspected ACS with relatively low levels of hsCRP had an increased mortality risk in both the short- and long-term. The highest hsCRP group (10-15 mg/L) had a hazard ratio of 2.20 at baseline, which only marginally attenuated at 3-years. This long-term risk is clinically significant as it may provide an opportunity to improve the selection of patients for investigation and treatment.[50] It is common for patients with relatively low hsCRP levels to be discharged from hospital with either a diagnosis of non-specific inflammation or the presence of a mild viral infection, with no consideration for mortality risk stratification.

4.6.2 Previous trial evidence

There has been recent interest in investigating the effect of immunomodulatory drugs in CVD following publication of the CANTOS[61], COLCOT[64] and LoDoCo2 trials.[65] However, the results of CIRT[63] highlight that not all anti-inflammatory drugs may show promise in reducing the risk of CVD. In this study, Methotrexate had a minimal effect on hsCRP levels and did not reduce CV events, warranting for the study to be terminated early. Unfortunately, no CRP data was reported in the COLCOT or LoDoCo2 studies. A recent phase 2 trial (RESCUE) assessed the efficacy of ziltivekimab in patients with a hsCRP leveo >2mg/L and chronic kidney disease. The study showed there was a dose-related reduction in hsCRP and a large trial powered on cardiovascular outcomes is planned.[66]

It is important to bear in mind that it is sometimes difficult in clinical practice to determine the exact time point at which an ACS event has occured. The time point at which a diagnosis is made may not coincide with the time of onset of the actual ACS event. In an attempt to determine the effect of baseline hsCRP levels on presentation, rather than the effect of a rise in hsCRP in response to an ACS event, I carried out analyses using the admission hsCRP level and compared it to the ‘peak’ troponin level. For patients who had a raised troponin, the median number of days between the admission hsCRP and peak troponin was 3 days (IQR 0-5). This infers that the admission hsCRP likely represents the presentation value instead of the value in response to the index event.

Individuals who had CRP levels >15 mg/L or an abnormal WCC were excluded in an attempt to limit the population to those without infections or systemic inflammatory conditions. It is unlikely that many patients with ACS were excluded due to the low CRP levels that are measured in ACS patients. In a study on the association of hsCRP levels with major adverse cardiovascular events (MACE) and mortality followinig ACS,[67] the

median baseline hsCRP level was only 10.5 mg/L (IQR, 4.2-30.3 mg/L).

4.6.3 CRP as a mediator for cardiovascular disease

In the multi-centre matched-control study on CRP apheresis in Acute Myocardial Infarction (CAMI-1), investigators evaluated the safety and feasibility of CRP apheresis to reduce the size of the infarct following a STEMI.[68] While the primary endpoint of reduction of infarct size, based on CMR imaging, was not significant in this pilot study, a large randomized trial of CRP apheresis therapy in patients with STEMI is currently under development at the Medical University of Innsbruck. The results of this trial should provide additional insight as to whether CRP is a mediator for CVD, infarct size and subsequent mortality risk.[69] However, the results of trials and observational studies targeting CRP should be reviewed in the context of large mendelian randomization studies that suggest that CRP[59] (unlike the IL-6 receptor[70]) is not causally-related to CVD.

4.6.4 Relationship between CRP and troponin level

The discrimination and net reclassification with the addition of hsCRP to our risk models is an important new finding in patients with suspected ACS. Several previous studies have shown the value of hsCRP in risk scores for predicting mortality risk in those with chronic CVD or primary prevention populations, with the development of scores such as the Reynolds risk score or Framingham risk score.[71]

4.6.5 Limitations

The study had access to all-cause mortality, rather than cause-specific mortality. This limits our understanding of the causes of death in those patients who died with the higher hsCRP levels. By excluding patients with abnormal hsCRP (>15 mg/l) or WCC counts, it is unlikely that sepsis was the underlying cause of death for the excess mortality. Nevertheless, all-cause mortality remains a clinically relevant endpoint. In addition, as with all observational studies, it is impossible to adjust our analyses by all confounding factors.

4.7 Conclusion

In a large cohort of patients with suspected ACS, mildly elevated hsCRP was a clinically relevant prognostic marker in addition to troponin, which points to its potential utility in identifying patients who may benefit from novel treatments targeting inflammation. Clinical guidelines should consider including hsCRP as a biomarker to help risk stratify patients presenting with suspected ACS.

Chapter 5. Early translational cardiovascular research - effect of an intervention in population subgroups that were underrepresented in previous clinical trials

5.1 SENIOR-NSTEMI Study

In this Chapter, I will explore the role of big data for early discovery phase translational research by assessing the effect of an intervention in population subgroups that were underrepresented in previous clinical trials. I will estimate the effect of invasive versus non-invasive management on the survival of patients aged 80 years or older with non-ST elevation myocardial infarction (NSTEMI).

The rationale for answering this particular research question is summarised in **Table 5.1**.

Table 5.1 - Rationale for SENIOR-NSTEMI study

PICO Element	Description	Rationale
Population/ Patient/ Problem	Patients aged 80 years or older admitted to hospital with an NSTEMI	The NIHR HIC dataset is troponin centric, as all patients were only included in the dataset if they had a troponin measured. The diagnosis of NSTEMI includes checking the troponin level to confirm that it is elevated. As a result, there would be no selection bias as all patients admitted to hospital with an NSTEMI were captured. Patients aged 80 years or older were investigated as these were the cohort who were underrepresented in previous trials investigating the best approach to management of patients with NSTEMI.
Intervention	Invasive management (diagnostic coronary angiogram)	Invasive management in the form of a diagnostic coronary angiogram is considered the gold standard approach to managing patients with NSTEMI.
Comparison	Non-invasive management	Non-invasive management was chosen as the comparator. In clinical practice, the alternative to invasive management is medical management of patients with NSTEMI. In addition to Aspirin, most patients with NSTEMI are usually administered a

		second antiplatelet agent (Clopidogrel, Ticagrelor or Prasugrel). Anti-thrombin therapy with Fondaparinux Sodium is usually also offered. For the NIHR HIC troponin dataset, as there were no data available on medications administration, non-invasive management was the definition used for the comparator group. Although most patients in this non-invasive management group would have had medical management, this cannot be confirmed and will be considered as a limitation of the study.
Outcome	All-cause mortality	All-cause mortality is the outcome of interest to most patients. Furthermore, all-cause mortality is the only outcome where the diagnosis is free from bias.

5.2 Abstract

Background

Results from previous trials have shown that there is a lower long-term mortality after invasive compared to non-invasive management among patients with non-ST elevation myocardial infarction (NSTEMI), but these excluded very elderly patients. The aim of this study was to emulate a 'target trial' estimating the effect of invasive compared with non-invasive management on survival in patients with NSTEMI aged ≥80 years admitted during 2010–2017, using routine clinical data from five tertiary centres.

Methods

Propensity scores based on pre-treatment demographic and clinical variables were derived using logistic regression: patients with high probabilities of either invasive or non-invasive management were excluded. Cox regression was used to estimate mortality hazard ratio comparing patients who did and did not have invasive management within 3 days of their peak troponin, and negative binomial regression was used to compare rates of hospital admission for heart failure.

Results

Of 2672 patients with NSTEMI, 157 died within 3 days of their peak troponin, whilst 879 were excluded because of extreme propensity scores. The 1,636 patients (59.8% non-invasive management) included in analyses had median age 85 (interquartile range (IQR) 82-89) years. During median follow-up of 2.7 (IQR 1-4.5) years, there were 723 (44.2%) deaths. Using inverse probability weighting, adjusted cumulative 5-year mortality was 40% and 63% in the invasive and non-invasive management groups, respectively. The mortality hazard ratio comparing invasive with non-invasive management was 0.52 (95% CI 0.43-0.62) after multivariable adjustment for clinical characteristics and propensity score. Invasive management was associated with a lower incidence of hospital admissions for heart failure, compared with non-invasive management (adjusted rate ratio 0.67, 95% CI 0.50–0.88).

Conclusion

Among eligible patients, the survival advantage from invasive compared with non-invasive management may extend to patients age ≥80 years with NSTEMI.

5.3 Introduction

The majority of patients experiencing a non-ST elevation myocardial infarction (NSTEMI) are aged ≥70 years.[72] With improving life expectancy, the proportion of people aged ≥80 years will almost double over the next 20 years.[73] Older patients with an acute coronary syndrome (ACS) are at higher risk of adverse outcomes compared with younger people.[74,75] The rate of coronary angiography lowers with age.[76,77] Only 14% of NSTEMI patients aged ≥85 years receive a coronary angiogram, compared with 83% of those aged <65 years.[77]

Large randomised controlled trials (RCTs) demonstrated a long-term survival advantage for invasive compared with non-invasive management of NSTEMI, but participants' mean age was 66 years. The proportion of elderly patients enrolled into these studies was low, therefore it is unknown if the survival benefit translates to the elderly.[78] Because of the limited evidence to support invasive treatment, many physicians treat elderly patients with NSTEMI symptomatically, recommending consideration of invasive management only when there is ongoing chest pain.

Registry studies have shown that in routine care, frail patients with many comorbidities are more likely to be treated without invasive management while the fittest patients are more likely to have invasive management.[79-81] Although these registry studies tried to control for confounding, their results may also have been affected by immortal time bias.[82,83] Immortal time bias is an issue when patients who died early in their admission, before the chance for invasive therapy to be considered or arranged, were assigned to the non-invasive management group.

Comparative effectiveness research questions should ideally be answered using RCTs.[84] The SENIOR-RITA trial, currently in progress, aims to assess whether an invasive compared with a non-invasive strategy reduces mortality or non-fatal myocardial infarction in NSTEMI patients aged ≥75 years.[85] However, its results are not expected until 2029. Recruitment has been challenging, with many centres having held recruitment or withdrawn.[85] This may reduce the power of the study or lead to delay in its completion.

Causal inference from observational databases can be viewed as an attempt to emulate a 'target' RCT.[86] If the emulation is successful, the analysis will yield the same effect estimates (except for random variability) as the target trial. We estimated the effect of invasive management compared with non-invasive management on the survival of NSTEMI patients with NSTEMI aged ≥80 years using real-world, multi-centre clinical data (SENIOR-NSTEMI study), based on explicit description and emulation of a target trial.

5.4 Methods

5.4.1 Study design and participants

Data were obtained from five collaborating hospitals, which were all tertiary centres with emergency departments. Eligible patients were aged ≥80 years when they were diagnosed with NSTEMI between 2010 and 2017. Classification of NSTEMI was based on the assigned International Statistical Classification of Diseases and Related Health Problems (ICD) discharge codes (I21.4; acute subendocardial myocardial infarction/acute nontransmural myocardial infarction).[7]

We compared patients who did and did not have invasive management, defined as coronary angiography with or without subsequent revascularisation. All included patients were followed up from the time of their peak troponin until death or censoring in April 2017. The outcomes considered were all-cause mortality and the number of hospital admissions for heart failure during follow-up. Vital status was ascertained using routinely collected data on the NHS Spine Application, which was linked to the Office of National Statistics, and thereby to the national registry of deaths. We estimated the intention-to-treat effect: in particular, patients who had invasive management more than three days after their peak troponin were included in the comparison group.

5.4.2 Statistical analysis

To limit the effect of immortal time bias, patients who died within 3 days of admission were excluded from the primary analysis. Propensity scores (PS: patients' probability of receiving invasive management) were derived using a logistic regression model based on pre-treatment variables: patient demographics, blood test results, cardiovascular risk factors, history of cardiovascular disease and other comorbidities. Any non-linear relationships were modelled using smoothing splines.

Patients whose probability of receiving either invasive or non-invasive management was close to zero were removed from the analysis, to ensure that analyses were restricted to patients eligible to receive either treatment strategy (as would be the case in a randomized trial). In eligible patients, inverse probability of treatment weights were defined as 1/PS in patients managed invasively, and 1/(1-PS) in patients managed non-invasively.

Kaplan-Meier plots were used to display cumulative mortality and probability of hospital admission for heart failure over time in each treatment group. We used Cox models to estimate mortality hazard ratios (HR) comparing invasive with non-invasive

management. Three methods were used to control confounding; multivariable adjustment, multivariable adjustment additionally including a continuous propensity score modelled using restricted cubic splines, and inverse-probability-of-treatment (IPT) weighting. Changes in the hazard ratio with age were investigated using a cubic spline. We used negative binominal regression, estimate rate ratios for hospital admissions, comparing invasive with non-invasive management, adjusted for baseline characteristics.

In sensitivity analyses, we explored the robustness of the results following inclusion of those patients who died within 3 days of their peak troponin level. Similar to the primary analysis, we removed those whose probability of receiving either invasive or non-invasive management was near zero. For the remaining patients who died within 3 days of their peak troponin level, we investigated the effect of assuming that they had all been in one or the other treatment group. In a further sensitivity analysis, we randomly assigned these patients to invasive or non-invasive management based on their propensity scores. We generated 20 datasets using this approach, estimated the hazard ratio comparing invasive with non-invasive management in each dataset, then pooled the log hazard ratios and their standard errors using Rubin's rules to estimate the overall hazard ratio with 95% confidence intervals (CI). A further sensitivity analysis was performed by broadening the inclusion criteria of patients based on propensity score.

The analysis of the number of hospital admissions relating to heart failure was examined using negative binomial regression with a log link, including follow-up time as the exposure variable. The model adjusted for invasive management and additionally adjusted for the same variables as in the survival analyses with invasive management as the outcome. Kaplan-Meier plots were used to assess cumulative heart failure admissions over time.

Statistical analyses were performed using R 3.5.0 statistical software (R Core Team, Austria) and Stata version 16.0 (Stata Corp, USA).

5.5 Results

Of 61342 patients aged ≥80 years who had troponin measured during the study period, 2672 had a diagnosis of NSTEMI, of whom 918 (34.4%) had invasive management during their admission. **Figure 5.1** shows a histogram of the numbers of patients undergoing invasive management at different time points relating to the peak troponin level at presentation. Of 1359 deaths (50.9%) among the patients with NSTEMI, 157 were within the first three days of admission and were excluded from the primary analyses (**Figure 5.2**). The patient characteristics of patients who died within the first three days are shown in **Table 5.2**.

Figure 5.1 - Histogram of the numbers of patients undergoing invasive management at different time points relative to the date of the peak troponin level

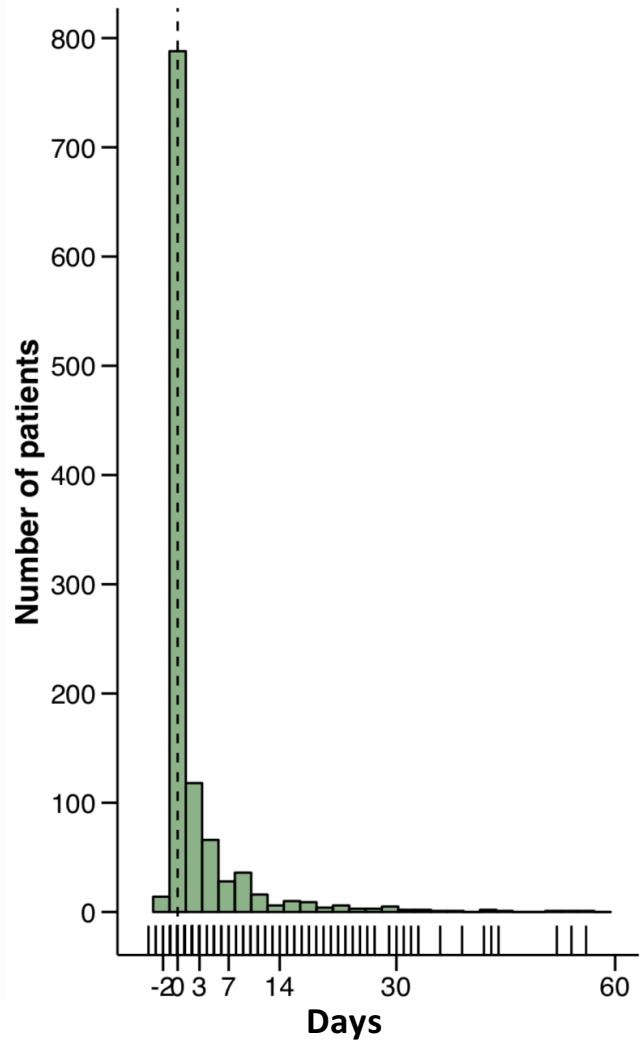
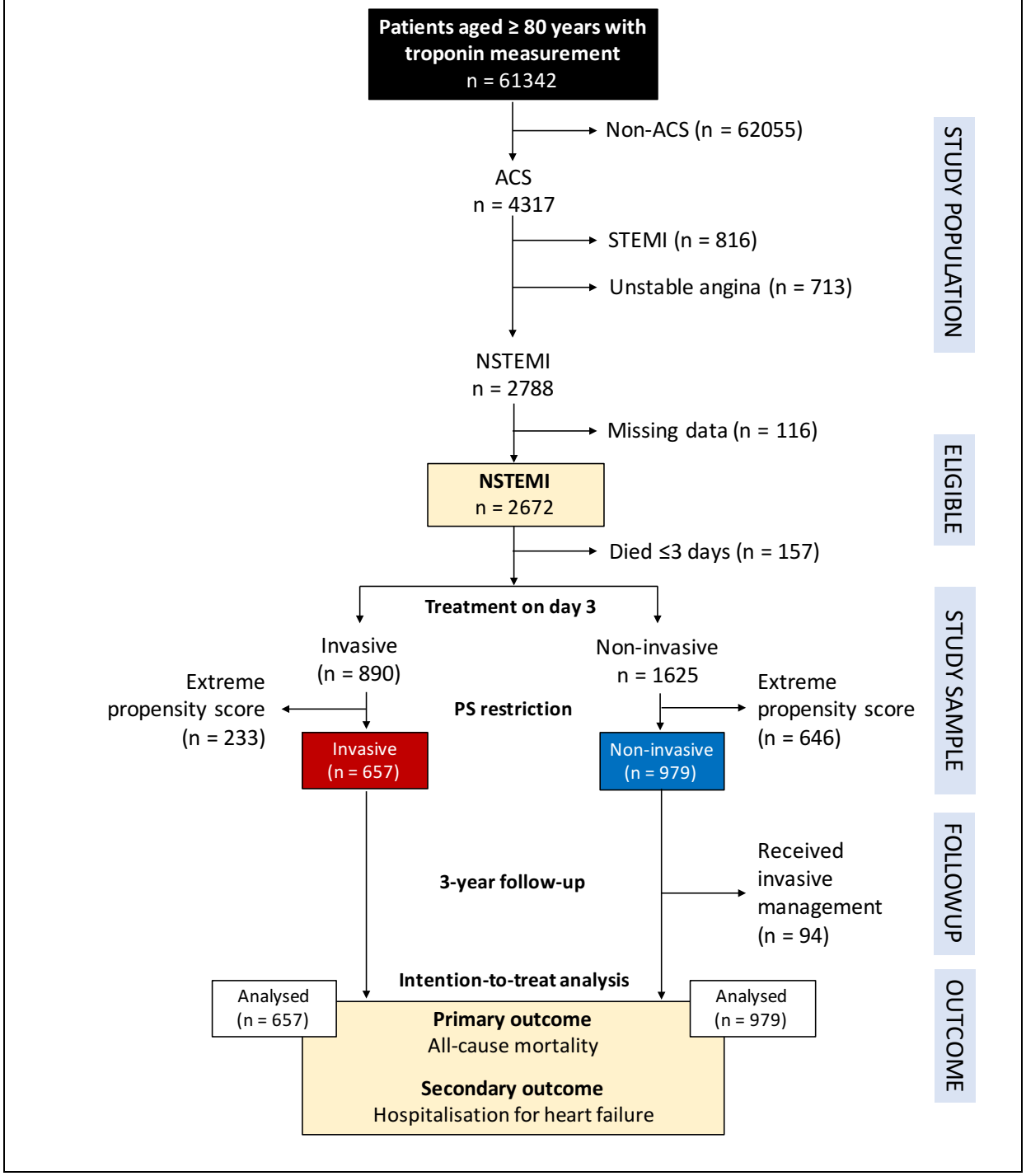


Figure 5.2 - Flow of patients through the study



Invasive management defined as an invasive procedure within 3 days of admission.

Table 5.2 - Characteristics of the 1636 patients undergoing NSTEMI stratified by study subgroup

	Study population (n=1636)	Died at <3 days (n=157)	<25th percentile excluded (n=628)	>90th percentile excluded (n=251)
Demographic characteristics				
Age (year)	85 (82 – 89)	86 (83 – 89)	90 (85 – 94)	83 (81 – 86)
Male sex	907 (55.4)	80 (51.0)	214 (34.1)	196 (78.1)
Haematology and biochemistry results				
C-reactive protein (mg/L)	9.4 (3.3 – 36.4)	44.0 (16.3 – 90.9)	53.6 (13.2 – 124.4)	5.0 (2.0 – 10.7)
Creatinine (μmol/L)	96 (75 – 130)	127 (94 – 180)	101 (76 – 138)	93 (78 – 118)
Haemoglobin (g/dL)	12.2 (10.9 – 13.6)	11.8 (10.1 – 13.3)	11.5 (9.9 – 13.0)	13.1 (12.0 – 14.0)
Platelet count ($x10^9$/L)	223 (179 – 274)	225 (169 – 289)	233 (174 – 302)	214 (178 – 257)
Troponin (xULN)	48.6 (9.3 – 197.5)	120.0 (20.3 – 386.3)	17.6 (6.3 – 74.4)	216.7 (47.5 – 731.7)
White cell count ($x10^9$/L)	9.1 (7.2 – 11.8)	12.0 (9.5 – 15.8)	10.8 (7.9 – 15.2)	9.1 (7.5 – 10.9)
Sodium (mmol/L)	138 (135 – 140)	138 (135 – 142)	138 (134 – 141)	138 (135 – 140)
Potassium (mmol/L)	4.3 (3.9 – 4.6)	4.5 (4.1 – 5.2)	4.3 (3.9 – 4.8)	4.2 (3.9 – 4.5)
Cardiovascular risk factors				
Tobacco use	383 (23.4)	12 (7.6)	23 (3.7)	220 (87.6)
Diabetes mellitus	411 (25.1)	31 (19.7)	129 (20.5)	62 (24.7)
Family history of IHD	141 (8.6)	2 (1.3)	2 (0.3)	115 (45.8)
Hypercholesterolemia	555 (33.9)	26 (16.6)	93 (14.8)	120 (47.8)
Hypertension	946 (57.8)	63 (40.1)	216 (34.4)	167 (66.5)
Cardiovascular disease				
Aortic stenosis	103 (6.3)	8 (5.1)	57 (9.1)	4 (1.6)
Atrial fibrillation	303 (18.5)	29 (18.5)	161 (25.6)	25 (10.0)
Heart failure	353 (21.6)	66 (42.0)	262 (41.7)	16 (6.4)
Previous myocardial infarction	1092 (66.7)	83 (52.9)	104 (16.6)	239 (95.2)
Other comorbidities				
Chronic kidney disease*	158 (9.7)	25 (15.9)	74 (11.8)	13 (5.2)
Malignancy	140 (8.6)	17 (10.8)	102 (16.2)	13 (5.2)
Obstructive lung disease	123 (7.5)	16 (10.2)	103 (16.4)	8 (3.2)

*Median (interquartile range) or number (%). *chronic kidney disease > stage 2; IHD: ischaemic heart disease*

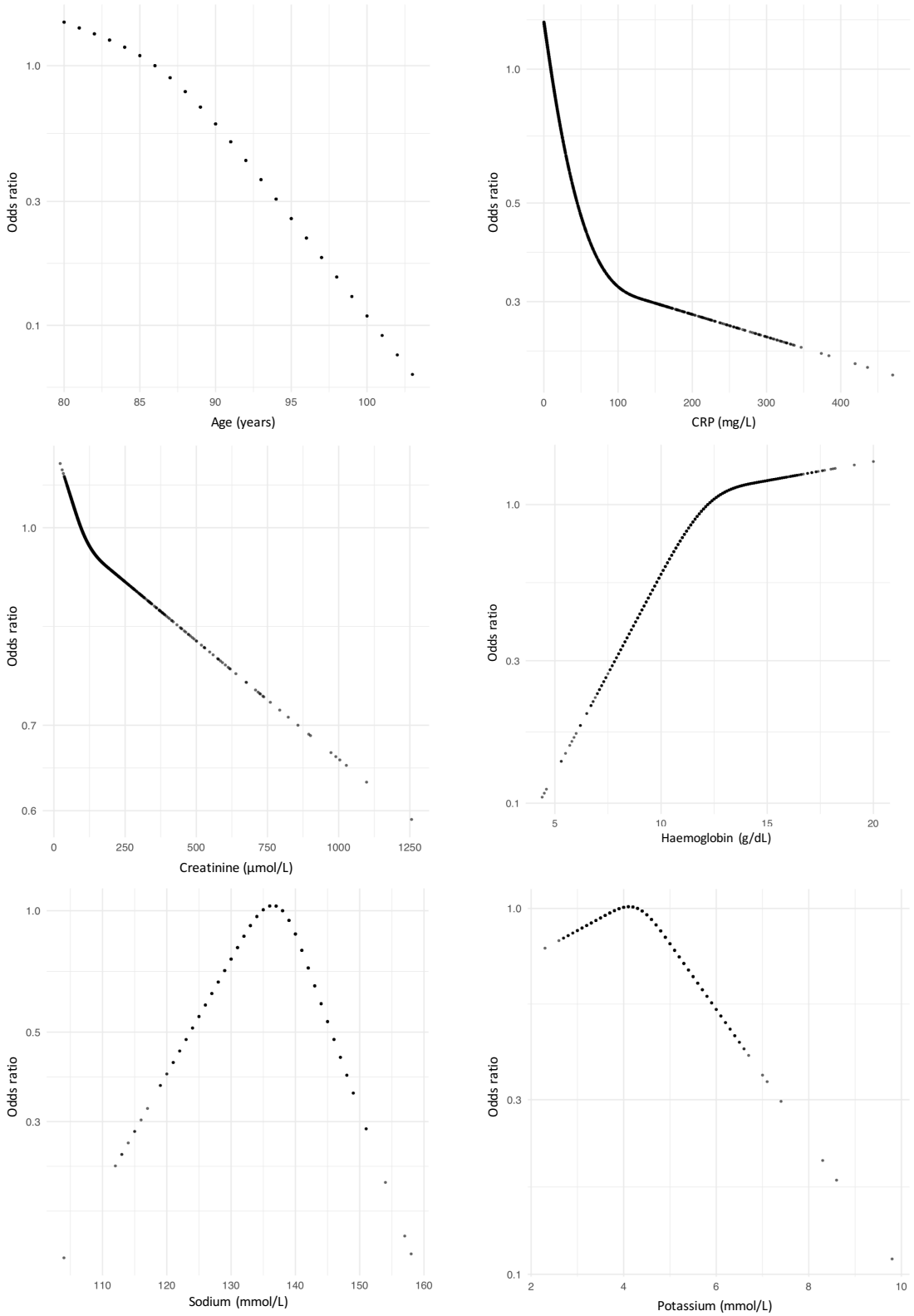
The patient characteristics associated with receipt of invasive management are shown in **Table 5.3** and **Figure 5.3**.

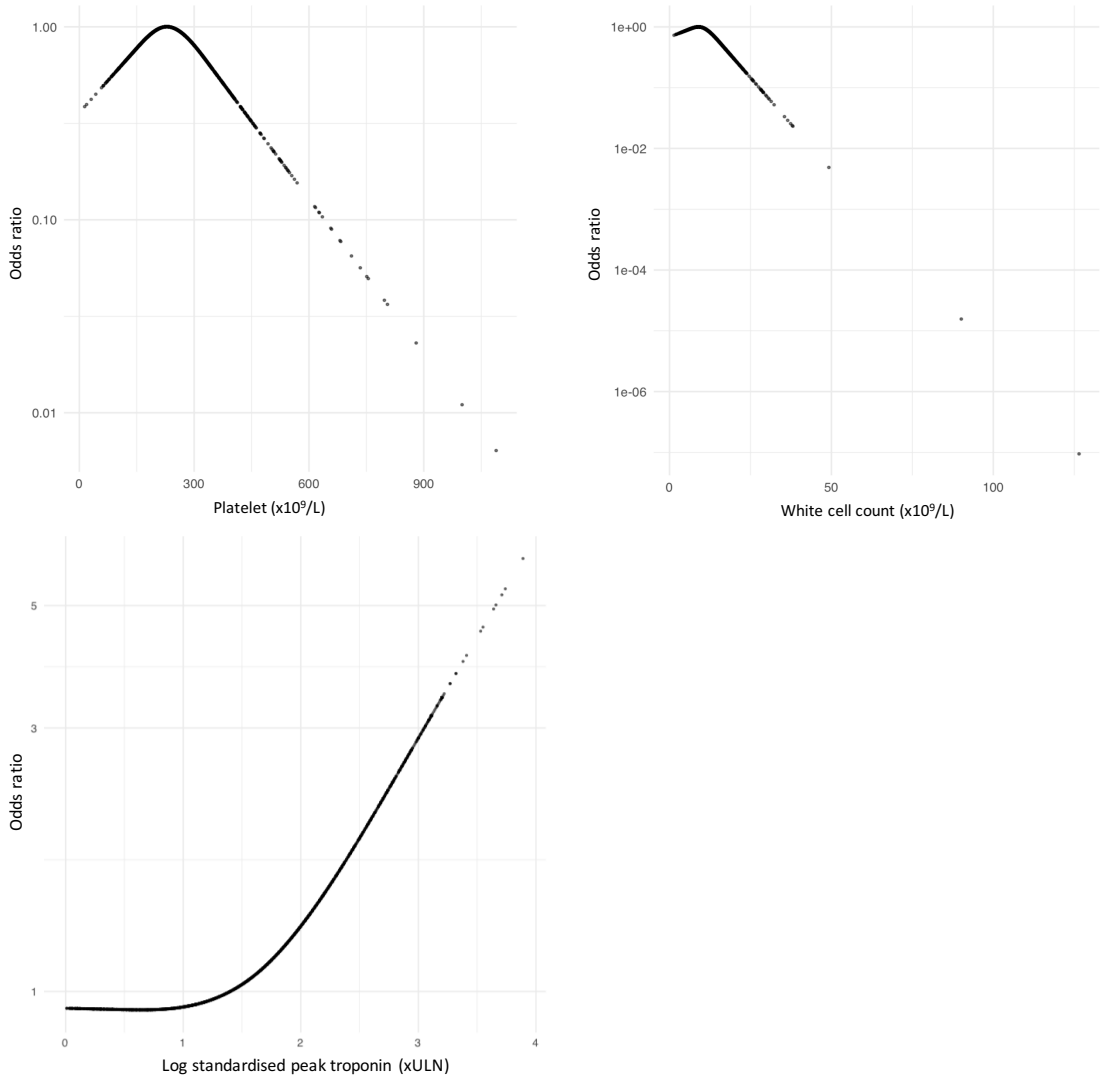
Table 5.3 - Odds ratio of undergoing invasive management

Disease/characteristic	Odds ratio (95% CI)	P-value
Male (vs female)	1.96 (1.66 – 2.32)	<0.0001
Previous MI	4.81 (3.97 – 5.81)	<0.0001
Obstructive lung disease	0.49 (0.36 – 0.68)	<0.0001
Hypercholesterolaemia	1.96 (1.65 – 2.34)	<0.0001
Atrial fibrillation	0.66 (0.53 – 0.82)	<0.0001
Smoker (vs not)	5.20 (4.29 – 6.31)	<0.0001
Aortic stenosis	0.61 (0.43 – 0.88)	0.007
Family history of IHD	5.58 (4.19 – 7.42)	<0.0001
Malignancy	0.62 (0.46 – 0.83)	0.001
Chronic kidney disease	0.71 (0.53 – 0.95)	0.02
Hypertension	1.81 (1.53 – 2.14)	<0.0001
Diabetes	1.12 (0.93 – 1.35)	0.24
Heart failure	0.44 (0.36 – 0.54)	<0.0001

Estimates compared to not having the disease, unless otherwise stated.
MI: myocardial infarction; IHD: ischaemic heart disease

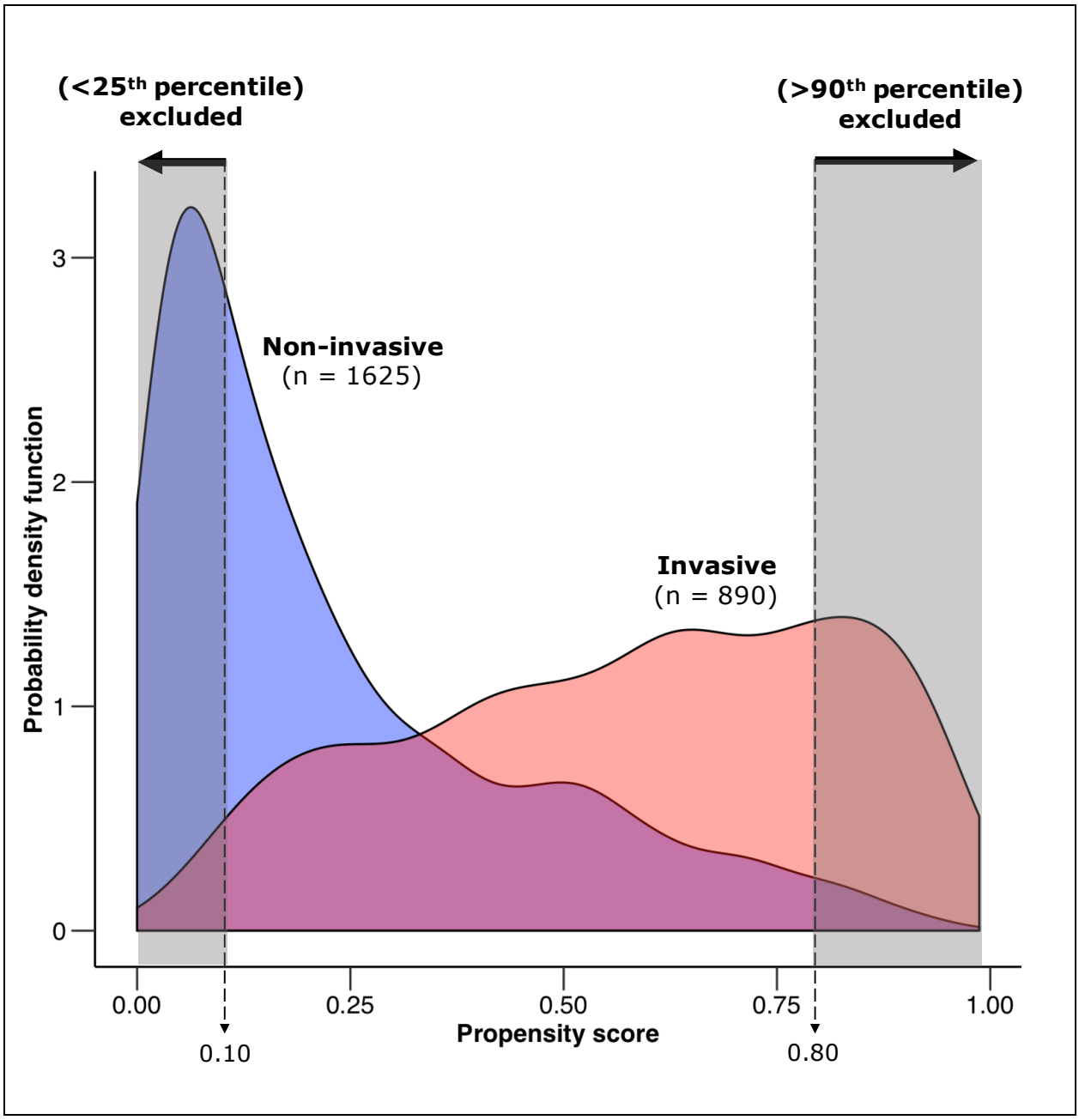
Figure 5.3 - Odds ratios of undergoing invasive management (versus not)





The distributions of the propensity score for patients treated invasively and non-invasively are shown in **Figure 5.4**. The substantial differences between these distributions indicate strong influences of patient characteristics on whether patients were invasively managed. The proportions of patients who died during follow-up in the invasive and non-invasive management groups, according to quantiles of the propensity score, are shown in **Table 5.4**. In both invasively and non-invasively managed patients, the proportion who died was higher in patients with lower propensity scores (those most likely to be non-invasively managed). Below the 25th percentile (PS=0.876), only 19 patients were invasively managed while above the 90th percentile (PS=0.101) only 37 patients were non-invasively managed. Therefore, analyses were restricted to patients whose propensity scores were between the 25th and 90th percentiles of the distribution (**Figure 5.2**). The patient characteristics of patients in the <25th percentile and >90th percentile of the propensity score are shown in **Table 5.2**.

Figure 5.4 - Probability density function of the propensity score for 2515 patients with NSTEMI aged 80 or over according to invasive or non-invasive management strategy



Invasive management defined as invasive procedure within 3 days of peak troponin level. Patients with propensity scores <25% and >90% (grey shaded regions) had a high probability of receiving non-invasive treatment and invasive treatment, respectively, and were excluded from analyses.

Table 5.4 - Proportion of deaths among 2515 patients with NSTEMI aged 80 or over who were treated with invasive or non-invasive management, according to percentiles of the propensity score for the study population who survived up to 3 days from admission

Percentile	Propensity score upper limit	Invasive (n = 890)			Non-invasive (n = 1625)			Hazard ratio (95% CI)
		No.	Deaths		No.	Deaths		
			No.	%		No.	%	
99 to 100	1.000	25	3	12.0	1	1	100	0.08 (0.01 – 0.92)
95 to <99	0.9454	91	13	14.3	9	2	22.2	0.52 (0.12 – 2.33)
90 to <95	0.8762	98	22	22.5	27	13	48.2	0.36 (0.18 – 0.72)
75 to <90	0.7961	262	65	24.8	115	46	40.0	0.50 (0.34 – 0.73)
50 to <75	0.5810	268	96	35.8	361	157	43.5	0.65 (0.51 – 0.84)
25 to <50	0.2723	127	49	38.6	503	310	61.6	0.44 (0.33 – 0.60)
10 to <25	0.1007	16	8	50.0	361	242	67.0	0.62 (0.31 – 1.26)
5 to <10	0.0401	3	1	33.3	123	84	68.3	0.26 (0.04 – 1.89)
1 to <5	0.0222	0			100	73	73.0	
0 to <1	0.0089	0			25	17	68.0	
Overall		**890**	**257**	**28.9**	**1625**	**945**	**58.2**	**0.33 (0.29 – 0.38)**

Invasive management defined as invasive procedure within 3 days of peak troponin level

The 1636 patients (59.8% non-invasive management) included in our analyses had a median age of 85 (IQR 82-89) years (**Table 5.2**). Revascularisation was performed in 61% (n=399) of patients who underwent invasive management. There were significant differences in demographic and clinical factors between the invasive and non-invasive management groups (**Table 5.5**). Those undergoing non-invasive management were older, were more likely to be female, had higher C-reactive protein counts (mg/L), and had lower haemoglobin (g/dL) and troponin (xULN). Additionally, they were less likely to be smokers, to have hypercholesterolemia, hypertension, a family history of IHD, or a previous MI.

Table 5.5 - Characteristics of 1636 patients with NSTEMI aged 80 or over according to invasive or non-invasive management strategy

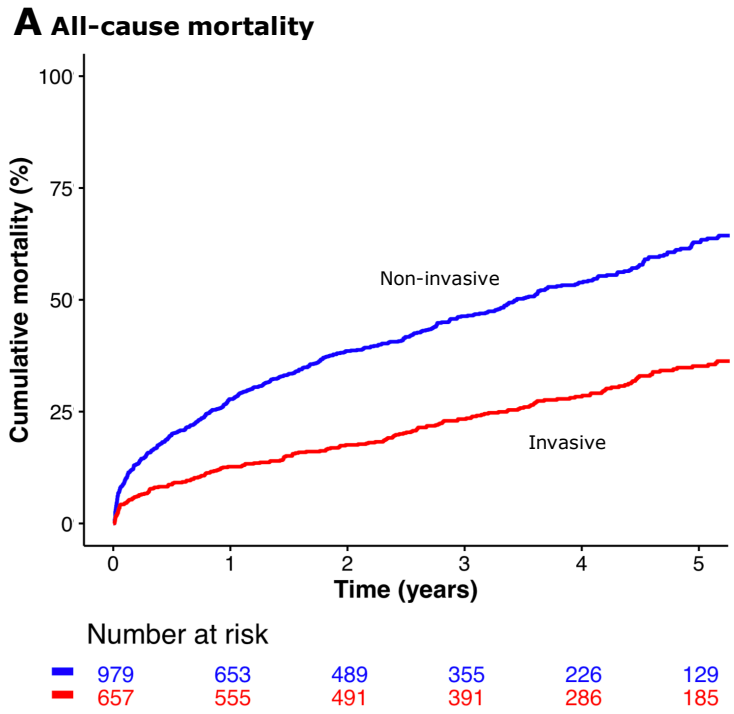
	Invasive management (n = 657)	**Non-invasive management (n = 979)**	**P- Value**
Demographic characteristics			
Age (year)	85 (82 – 88)	86 (82 – 89)	0.001
Male sex	387 (58.9)	520 (53.1)	0.02
Haematology and biochemistry results			
C-reactive protein (mg/L)	7.8 (2.9 – 28.7)	11.7 (3.9 – 41.6)	<0.0001
Creatinine (µmol/L)	96 (74 – 123)	95 (75 – 134)	0.25
Haemoglobin (g/dL)	12.4 (11.1 – 13.7)	12.1 (10.8 – 13.5)	0.002
Platelet count (x10^9/L)	220 (179 – 267)	224 (180 – 279)	0.24
Troponin (xULN)	54.9 (11.0 – 224.1)	42.7 (8.0 – 175.2)	0.02
White cell count (x10^9/L)	8.8 (7.2 – 11.3)	9.3 (7.2 – 12.0)	0.06
Sodium (mmol/L)	138 (135 – 140)	138 (134 – 140)	0.86
Potassium (mmol/L)	4.3 (3.9 – 4.6)	4.3 (3.9 – 4.6)	0.77
Cardiovascular risk factors			
Tobacco use	213 (32.4)	170 (17.4)	<0.0001
Diabetes mellitus	171 (26.0)	240 (24.5)	0.52
Family history of IHD	86 (13.1)	55 (5.6)	<0.0001
Hypercholesterolemia	257 (39.1)	298 (30.4)	<0.0001
Hypertension	415 (63.2)	531 (54.2)	<0.0001
Cardiovascular disease			
Aortic stenosis	36 (5.5)	67 (6.8)	0.30
Atrial fibrillation	114 (17.4)	189 (19.3)	0.33
Heart failure	227 (23.2)	126 (19.2)	0.06
Previous myocardial infarction	500 (76.1)	592 (60.5)	<0.0001
Other comorbidities			
Chronic kidney disease*	59 (9.0)	99 (10.1)	0.50

Malignancy	53 (8.1)	87 (8.9)	0.59
Obstructive lung disease	44 (6.7)	79 (8.1)	0.34

*Median (interquartile range) or number (%). ‡Comparison between troponin negative and troponin positive groups using Mann-Whitney U test for continuous variables and Chi square test for categorical variables. *chronic kidney disease > stage 2 IHD, ischaemic heart disease*

During a median follow-up of 2.7 (IQR 1–4.5) years, there were 723 (44.2%) deaths. At 5-years, cumulative mortality from three days after peak troponin was 33% and 68% in the invasive and non-invasive management groups, respectively (Kaplan-Meier plot, **Figure 5.5A**). An inverse probability of treatment weighted Kaplan-Meier mortality plot demonstrated the mortality rates were higher in the invasive group and slightly lower in the non-invasive group (**Figure 5.6A**), with 5-year cumulative mortality of 40% and 63% in the two groups, respectively.

Figure 5.5 - Unadjusted Kaplan-Meier mortality curves displaying cumulative all-cause mortality and probability of admission for heart failure according to invasive and non-invasive management



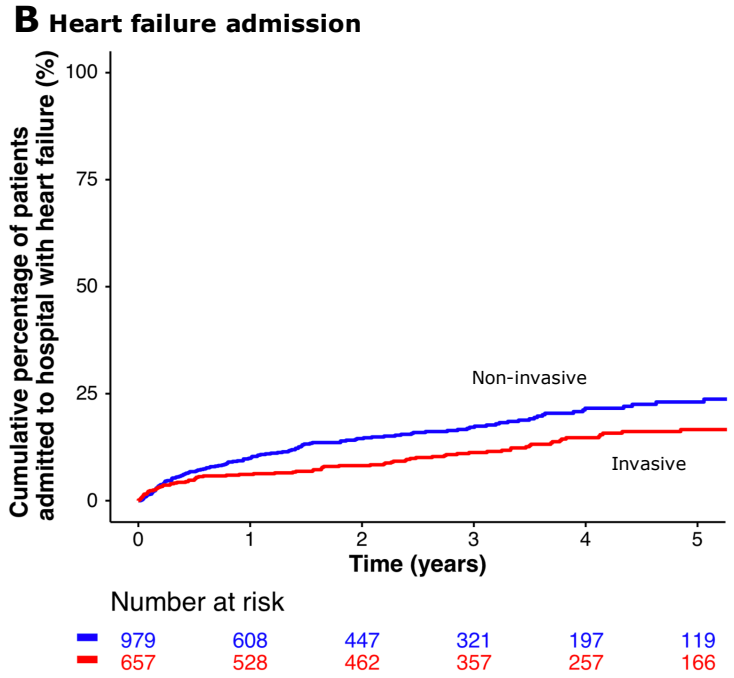
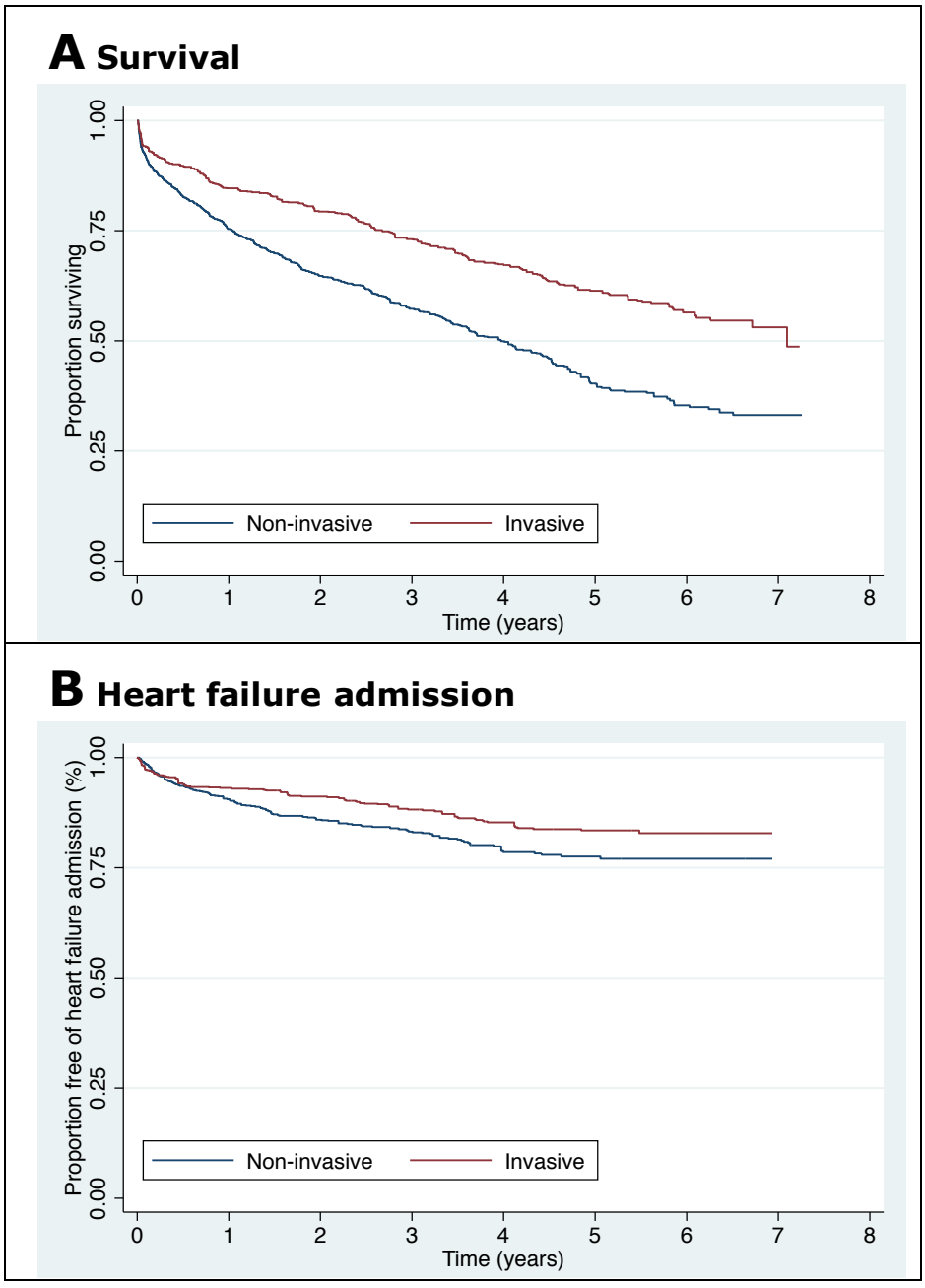


Figure 5.6 - Kaplan-Meier curves displaying cumulative all-cause mortality and probability of admission for heart failure according to invasive and non-invasive management



Plots are weighted according to the inverse probability of treatment received, and so compare outcomes if all eligible patients were invasively or non-invasively managed.

The crude mortality HR comparing invasive with non-invasive management during follow-up was 0.45 (95% CI 0.38-0.52), which was attenuated to 0.52 (95% CI 0.43-0.62) after multivariable adjustment for clinical characteristics and propensity score. The mortality HR remained similar at 0.52 (95% CI 0.43 0.64) following additional inverse probability weighting (**Table 5.6**). The mortality benefit associated with invasive management was significant at both short- and long-term follow-up (**Table 5.7**) and was seen throughout the age range from 80 to 100 years (**Figure 5.7**).

Table 5.6 - Comparison of the estimated treatment effect of invasive management on death using different methods to control for confounding for patients with NSTEMI aged 80 and over

	Invasive (n)	Non-invasive (n)	Crude Hazard Ratio (95% CI)	P-value	Adjusted Hazard Ratio (95% C)	P-value
Main analysis						
Crude	657	979	0.45 (0.38 – 0.52)	<0.0001		
Multivariable* plus PS adjustment	657	979			0.52 (0.43 – 0.62)	<0.0001
IPTW	657	979			0.56 (0.47 – 0.68)	<0.0001
Multivariable* plus PS and IPTW adjustment	657	979			0.52 (0.43 – 0.64)	<0.0001
Sensitivity analysis‡						
Assigning deaths within 3 days of peak troponin to the non-invasive group	657	1070	0.39 (0.33 – 0.46)	<0.0001	0.45 (0.38 – 0.53)	<0.008
Assigning deaths within 3 days of peak troponin to the invasive group	748	979	0.65 (0.57 – 0.76)	<0.0001	0.81 (0.70 – 0.95)	0.01
Assigning deaths within 3 days to a treatment group based on predicted PS	681	1046	0.45 (0.39 – 0.53)	<0.0001	0.53 (0.45 – 0.63)	<0.0001
Restriction of patients to the 10th and 95th centiles	771	1367	0.39 (0.34 – 0.45)	<0.0001	0.50 (0.43 – 0.60)	<0.0001

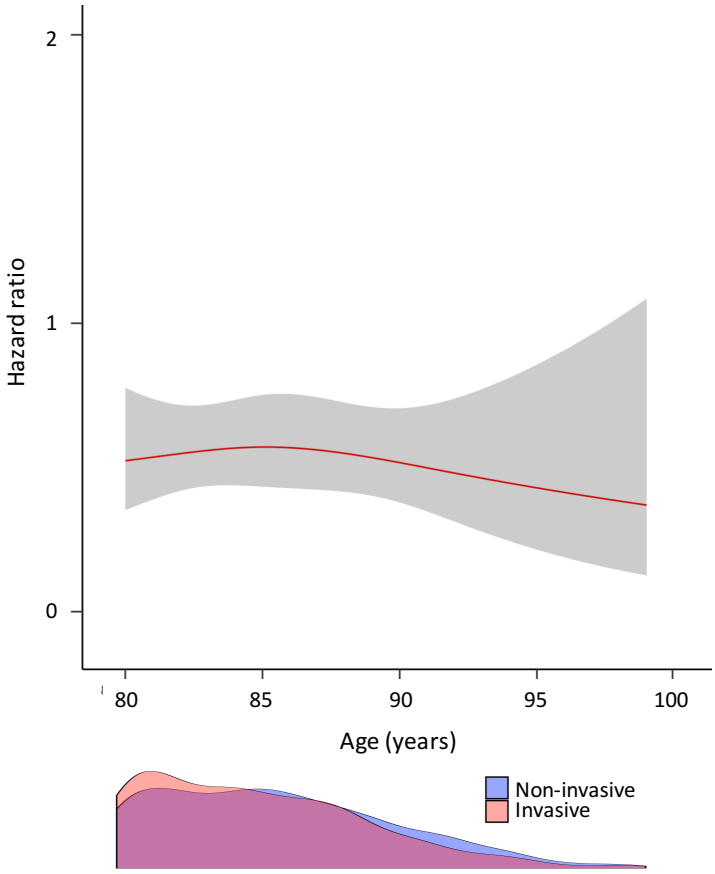
** Adjusted for all patient demographic and clinical variables listed in Table 5.4. ‡All sensitivity analyses used the multivariable plus propensity score adjustment approach to calculate the adjusted hazard ratio. IPTW, inverse probability of treatment weighted; PS, propensity score*

Table 5.7 - Comparison of the estimated treatment effect of invasive management on death at short- and long-term follow-up using different methods to control for confounding for patients with NSTEMI aged 80 and over

Follow-up	Crude Hazard Ratio (95% CI)	P-value	Adjusted* Hazard Ratio (95% CI)	P-value
3 months	0.42 (0.30 – 0.60)	<0.0001	0.61 (0.42 – 0.88)	<0.008
1 year	0.42 (0.33 – 0.54)	<0.0001	0.57 (0.44 – 0.74)	<0.0001
3 years	0.42 (0.35 – 0.51)	<0.0001	0.53 (0.43 – 0.65)	<0.0001
Full follow-up‡	0.45 (0.38 – 0.52)	<0.0001	0.52 (0.43 – 0.62)	<0.0001

**Adjusted for all patient demographic and clinical variables listed in Table 5.4 as well as the propensity score variable. ‡ median follow-up: 3-years*

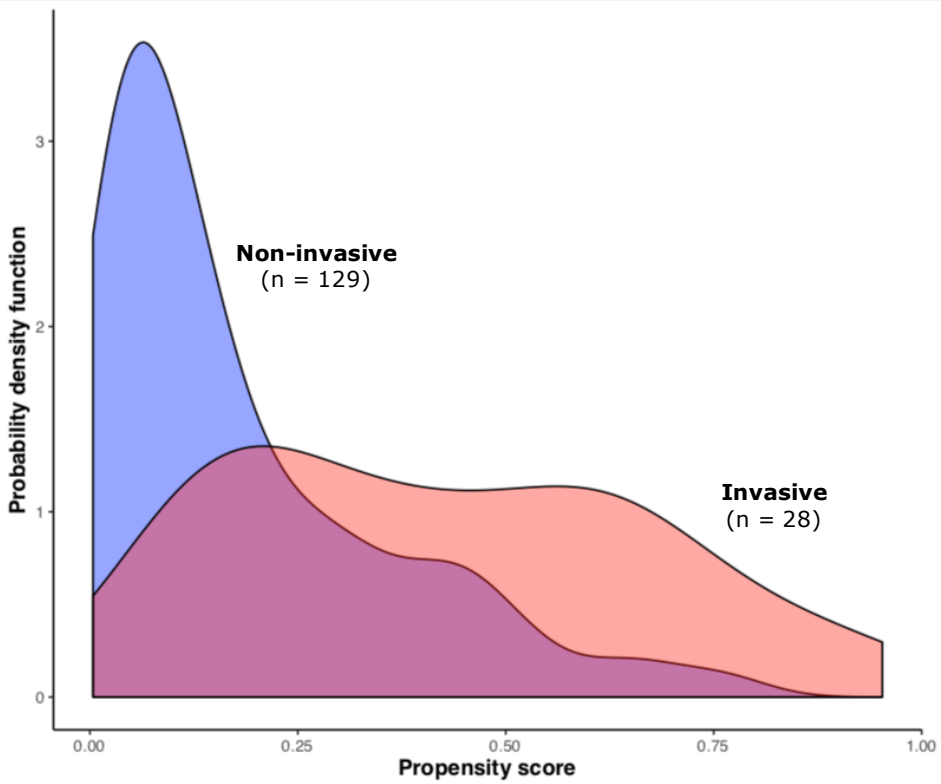
Figure 5.7 - Mortality hazard ratio of invasive management versus non-invasive management across the age spectrum



Restricted cubic spline model adjusted for propensity score and all patient demographic and clinical variables listed in Table 5.4. The lower plot describes the probability distribution of invasive (red shade) and non-invasive management (blue shade) across the age spectrum.

The distributions of the propensity score for the 157 patients treated invasively or non-invasively who died within 3 days of their peak troponin are shown in **Figure 5.8**. The majority (82%) of these patients underwent non-invasive management. Ninety-one of the 157 patients who died within 3 days of their peak troponin were included in the sensitivity analysis after exclusion of those patients whose probability of receiving either invasive or non-invasive management was close to zero. The adjusted hazard ratio was 0.45 (0.38 – 0.53) and 0.81 (0.70 – 0.95) after assigning the 91 patients who died within 3 days of their peak troponin level in either the non-invasive or invasive group, respectively (**Table 5.6**). When patients were assigned to treatment groups based on their predicted propensity scores using multiple imputation, the mortality hazard ratio remained similar to the primary analysis. On further analysis, including patients whose propensity scores were between the 10th and 95th percentiles of the distribution (**Figure 5.4**), led to an adjusted hazard ratio of 0.50 (0.43 – 0.60) (**Table 5.6**).

Figure 5.8 - The probability density functions of the propensity score for the 157 patients treated invasively or non-invasively who died within 3 days of their peak troponin



The numbers of patients subsequently admitted to hospital with heart failure during follow up were 82 (mean 0.21 admissions, range 0 to 7) in the invasive management group and 139 (mean 0.23 admissions, range 0 to 6) in the non-invasive management group. At 5-years, the cumulative percentage of patients admitted to hospital with heart failure was 14% and 26% in the invasive and non-invasive management groups respectively (Kaplan-Meier plot, **Figure 5.5B**). An inverse probability of treatment weighted Kaplan-Meier plot demonstrated similar estimates (Figure 3A), with 5-year cumulative admission rates of 14% and 24% in the invasive and non-invasive groups, respectively (**Figure 5.6B**).

Invasive management was associated with a lower incidence of hospital admissions for heart failure compared to those who underwent non-invasive management (adjusted incidence rate ratio 0.67, 95% CI 0.50–0.88, p=0.004).

5.6 Discussion

5.6.1 Summary of main findings

This is the first study to estimate the effect of invasive compared with non-invasive management on survival in patients with NSTEMI aged 80 years or older using real-world, multi-centre clinical data. We used a framework for comparative effectiveness research using observational data[86] from the NIHR Health Informatics Collaborative that revolved around the explicit description and emulation of a target trial. This study provides evidence that the survival advantage from invasive management may extend to very elderly patients with NSTEMI with an adjusted hazard ratio of 0.52. Additionally, invasive management was associated with a significant reduction in the incidence of heart failure hospitalisations during follow-up.

5.6.2 Treatment decisions in elderly patients with NSTEMI

Elderly patients comprise a growing segment of the population presenting with NSTEMI. The strongest independent negative predictor of invasive management has been demonstrated to be age ≥75 years.[76] In the GRACE registry, coronary angiography was performed in 67% of patients <70 years of age, compared to 33% of patients aged 80 years or older.[81] Similarly, in our study, only one-third of patients underwent invasive management in our cohort. Treatment decisions are often made in the context of careful evaluation of potential risks and benefits, estimated life expectancy and comorbidities. While patients managed non-invasively were more likely to have a lower

haemoglobin and higher creatinine level, patients managed invasively were more likely to have cardiovascular risk factors including smoking, family history of ischaemic heart disease, hypertension and hypercholesterolaemia.

5.6.3 Previous trial evidence

The European Society of Cardiology guidelines suggest that elderly patients should be considered for invasive management and revascularisation.[87] This is a class IIa recommendation, which implies that the expert panel considered that there is conflicting evidence or divergence of opinion but agreed on the side of invasive management. The American Heart Association have a similar guideline recommendation.[88] Only two small RCTs (After Eighty[89], invasive (n=229) and non-invasive (n=228); Italian Elderly ACS Trial[90], invasive (n=154) and non-invasive (n=159)) and two small post-hoc analyses of RCTs (TACTICS-TIMI 18[91] and FIR[92]) have evaluated invasive versus non-invasive management for NSTEMI in the elderly. There was no difference in all-cause mortality between invasive and non-invasive management in all for trials. A recent meta-analysis pooling together data from these four RCTs did not show a reduction in mortality over long-term follow-up with invasive therapy, though there was a trend in that direction (odds ratio 0.84, 95% CI 0.66-1.06, p=0.15, n=1887).[93]

In addition to the relatively small pooled sample size (n=1887), there were a number of additional limitations associated with these RCTs. Whilst both the TACTICS-TIMI 18[91] and FIR[92] subanalyses included only patients aged ≥75 years, they did not minimise the effect of confounders by adjusting the mortality outcome by baseline clinical characteristics of the elderly patients to account for differences between treatment groups. Contamination bias was an issue in both trials with 49% of patients in the non-invasive group undergoing coronary catheterisation of which 32.4% required revascularisation in the TACTICS-TIMI 18 sub-analysis and nearly half of all patients in the non-invasive group requiring revascularisation during follow-up in the FIR trial. Contamination was also present in the Italian Elderly ACS Study with 29% of those randomised to non-invasive management undergoing catheterisation.[90] The intervention effect of invasive therapy would therefore be biased towards the null in these studies. Less than 10% of patients in the non-invasive group underwent invasive management during follow-up in the present study.

There were also inconsistencies between the four RCTs with regards to the timing of angiography post-randomisation, ranging from 4 hours to 7 days, and the duration of patient following randomisation, ranging from 6 months to 5 years.[89-92] The primary outcome for all trials, to which sample size calculations were based, was a composite endpoint, which included additional variables to

all-cause mortality, including myocardial infarction, stroke, the number of rehospitalisations for an acute coronary syndrome and need for urgent revascularisation. All four RCTs were therefore underpowered for all-cause mortality endpoint comparisons between the invasive and non-invasive groups. All-cause mortality is unequivocal and avoids uncertainties in correctly ascribing the cause of death.

The SENIOR-RITA trial is a randomised controlled trial (RCT) currently in progress, with an aim to randomise 2300 patients with NSTEMI aged 75 years or older to either invasive or non-invasive management.[85] With 5-year follow-up planned for all enrolled patients, the primary outcome is a composite of cardiovascular death or non-fatal myocardial infarction, with an estimated study completion date in 2029. Whilst causal inferences can be made from well-conducted RCTs, there is a long time-lag between patient recruitment and study completion, and RCTs, in general, are costly and require heavy time and resource investment. Due to slow recruitment and issues with retaining centres for patient recruitment,[85] the original sample size of 2300 may need to be adjusted at the expense of reduced study power, or the study recruitment period extended leading to a delay in study completion.

5.6.4 Previous observational research

In the absence of RCT evidence, a small number of registry studies[79-81] have compared the mortality of very elderly patients presenting with NSTEMI who received invasive compared with non-invasive therapy. Although they suggested there may be a benefit for invasive therapy the value of these studies is limited by immortal time bias and the inclusion of a group of patients who were very frail and will almost certainly have been managed non-invasively, and also a group of very robust patients who will very likely have been mainly managed invasively. These biases may have exaggerated the effect of invasive therapy on mortality in all three previous observational studies.[79-81] Immortal time refers to a time span during the follow-up period during which the outcome under study, could not have occurred.[82,83] In other words, patients would need to survive long enough to be offered invasive management. Failure to account for the time-varying feature of invasive management can result in biased estimation of the effect of invasive management in favour of the invasive group.

5.6.5 Evaluation of outcomes

In the present study, we undertook a number of steps in the methodology to make the study as similar as possible to a randomised prospective clinical trial:

Firstly, we addressed the problem of immortal time bias by only including patients who survived up to 3 days after the peak troponin. We considered patients whose starting point of invasive management is before 3 days as part of the invasive group, and otherwise managed non-invasively. Patients who died or who were censored before the landmark time were excluded from the analysis. Three days was chosen as the optimal landmark time to capture the large majority of patients who underwent invasive management in our cohort and the median time used in previous RCTs.[89,90]

Secondly, we used propensity scoring to identify patients who were either very well or very unwell, which may skew the data, leaving patients in whom there is a high likelihood of there being a reasonable choice of both methodologies.

Thirdly, as well as a crude comparison model, we used different statistical models to control for confounding factors. This gives a level of assurance that the findings aren't merely a consequence of a particular statistical model.

Using these methodological steps in the elderly patients with an NSTEMI, we have shown that invasive management was associated with an approximately 50% reduction in 3-year mortality risk compared to patients managed non-invasively. The different methods to control for confounding yielded similar treatment effect estimates, with a hazard ratio of 0.52. Datasets in which a large untreated population has a markedly different risk factor distribution than a small treated population are common in observational studies of treatment effects. In such studies, including the present, the low-propensity strata are composed of those patients for whom most clinicians regard treatment as inappropriate, and therefore would not meet the inclusion criteria of RCTs. On the contrary, the high-propensity score strata are composed of patients who are not only more likely to meet the inclusion criteria for trials, but also considered by clinicians as patients who should be undergoing treatment and therefore less likely to be recruited to clinical trials. Our findings reflect a focus in these analyses on patients who would have been eligible to participate in clinical trials on the basis of their characteristics. Restriction to persons whose propensity scores were between 0.25 and 0.90 allowed us to focus even more sharply on this target population, to limit the impact of effect modification across propensity score strata.

5.6.6 Limitations

There are some limitations to this study. The observational design

precludes causal inference, and unmeasured confounding may have persisted despite adjusting our model for all major demographic and cardiovascular risk factors.

We did not have information on whether there was differential receipt of evidence-based cardiac care in the non-invasive management group, including medications prescription. Additionally, we did not have data on post-admission prognostic factors that may have affected choice of invasive or non-invasive management.

Our study reported all-cause mortality because this is incontrovertible and uniformly available. This study could not specify whether the cause of death was related to cardiac pathology. We were also unable to explore other outcomes of particular relevance to the elderly, such as quality of life, independence and quality of life.

5.7 Conclusion

The SENIOR-NSTEMI study is the first study designed to replicate a clinical trial using real-world, multicentre data to investigate the effect of invasive compared with non-invasive management on survival in patients with NSTEMI aged 80 years or older. This study provides evidence that the survival advantage bestowed from invasive management may extend to very elderly patients with NSTEMI.

6. Late translational cardiovascular research - epidemiology of cardiovascular disease across the whole population

6.1 TROP-RISK study

In this Chapter, I demonstrate how big data can be used for late translational research through assessment of the epidemiology of cardiovascular disease across unselected populations. I will evaluate the relationship between the full spectrum of troponin level and mortality in all patients in whom troponin testing has been performed for clinical purposes.

The rationale for answering this particular research question is summarised in **Table 6.1**.

Table 6.1 - Rationale for TROP-RISK study

PICO Element	Description	Rationale
Population/ Patient/ Problem	Patients in whom troponin testing has been performed for clinical purposes (suspected ACS)	The NIHR HIC dataset is troponin centric, as all patients were only included in the dataset if they had a troponin measured. In accordance with clinical guidelines, all patients with suspected ACS will undergo a troponin test. As a result, there would be no selection bias as all patients presenting to hospital with suspected ACS will be captured.
Intervention	Angiography	Invasive coronary angiography is the gold standard investigation for diagnosing coronary artery disease. This may be followed by coronary revascularisation in the form of PCI or CABG.
Comparison	No angiography	Data on whether or not a patient underwent coronary angiography with or without revascularisation was available for all patients.
Outcome	All-cause mortality	All-cause mortality is the outcome of interest to most patients. Furthermore, all-cause mortality is the only outcome where the diagnosis is free from bias.

6.2 Abstract

Introduction

Current evidence suggests a direct relationship between the magnitude of troponin elevation and mortality, albeit over a limited range of troponin levels, and clinicians generally work under the impression that higher troponins signify higher mortality in all age groups. The objective of our study was to use big data to determine the relationship between the full spectrum of troponin level and mortality in patients in whom troponin testing has been performed for clinical purposes.

Methods

As part of the NIHR HIC project, all troponin values measured during the study period (2010 to 2017) were assembled from five cardiovascular centres. Troponin concentrations were standardised as a multiple of each laboratory's 99th-percentile of the upper limit of normal (ULN). All patients were followed up until death or censoring on 1st April 2017. To model the relation between peak troponin level and all-cause mortality, restricted cubic spline Cox regression analyses were used. Splines were adjusted for patient age, gender, haemoglobin, creatinine, white cell count and C-reactive protein.

Results

A total of 257,948 patients had a troponin measured during the study. The peak troponin level was used for all analyses, which was the highest troponin value recorded during the inpatient admission. Troponin levels were standardised across each laboratory's upper limit of normal (ULN). During a median follow-up of 1198 days, there were a total of 55,850 (21.7%) deaths. A positive troponin result was associated with a 3.2 higher mortality hazard (95% CI 3.1-3.2) over a follow-up of 3-years. Age was an effect modifier, with a hazard ratio of 10.6 (8.5-13.3) in 18-29 year olds and 1.5 (1.4 -1.6) in those >90 years. There was a 15 percentage points higher absolute 3-year mortality with a positive troponin (compared to a negative troponin) across all age groups . The excess mortality was concentrated in the first few weeks.

There was a direct positive relationship between troponin level and mortality in patients who did not have an acute coronary syndrome diagnosis (n=120 049), whereas there was an inverted-U shaped relationship in those who had an acute coronary syndrome (n=14 468), with a paradoxical decline in mortality at peak troponin levels greater than 70 xULN. After multivariable adjustment, and stratifying by invasive management, the inverted-U shaped relation reverted to a direct positive relationship in ACS patients.

Conclusion

Regardless of age, a positive troponin result was associated with an increased mortality, even if the troponin level was mildly elevated The excess mortality associated with a raised troponin was mainly concentrated in the first few weeks.

6.3 Introduction

There is a well-established association between plaque rupture, acute coronary obstruction, reduced myocardial perfusion and troponin elevation.[94] The troponin assay is therefore recognised by both European and American guidelines as being the preferred biomarker for diagnosing acute myocardial infarction.[87,88,95,96]

Assays for measuring troponin have improved over the past two decades, allowing for a rapid detection of low levels of troponin with increased precision.[97,98] This has led to a substantial increase in the patient population with a positive troponin of uncertain significance.

The troponin test has been used for as a prognostic biomarker for all-cause mortality in patients suspected of having had a myocardial infarction.[99,100] There is some evidence to suggest a direct relationship between the level of troponin elevation and mortality, although this has been shown over a limited range of troponin levels[100,101] and clinicians generally work under the impression that higher troponins indicate a higher mortality. Additionally, it is often difficult to know how to manage elderly patients with elevated troponin levels.

Management decisions for patients who have small troponin rises, particularly if they have a non-cardiac presentation can be difficult and it would be helpful to understand the prognostic implications of these small troponin rises compared with large rises when making treatment choices.

The first aim was to describe how the prognostic impact of a raised troponin varies across the age spectrum, specifically including the very elderly. The second aim was to investigate the prognostic significance of very small troponin levels compared with larger levels.

6.4 Methods

6.4.1 Study design and participants

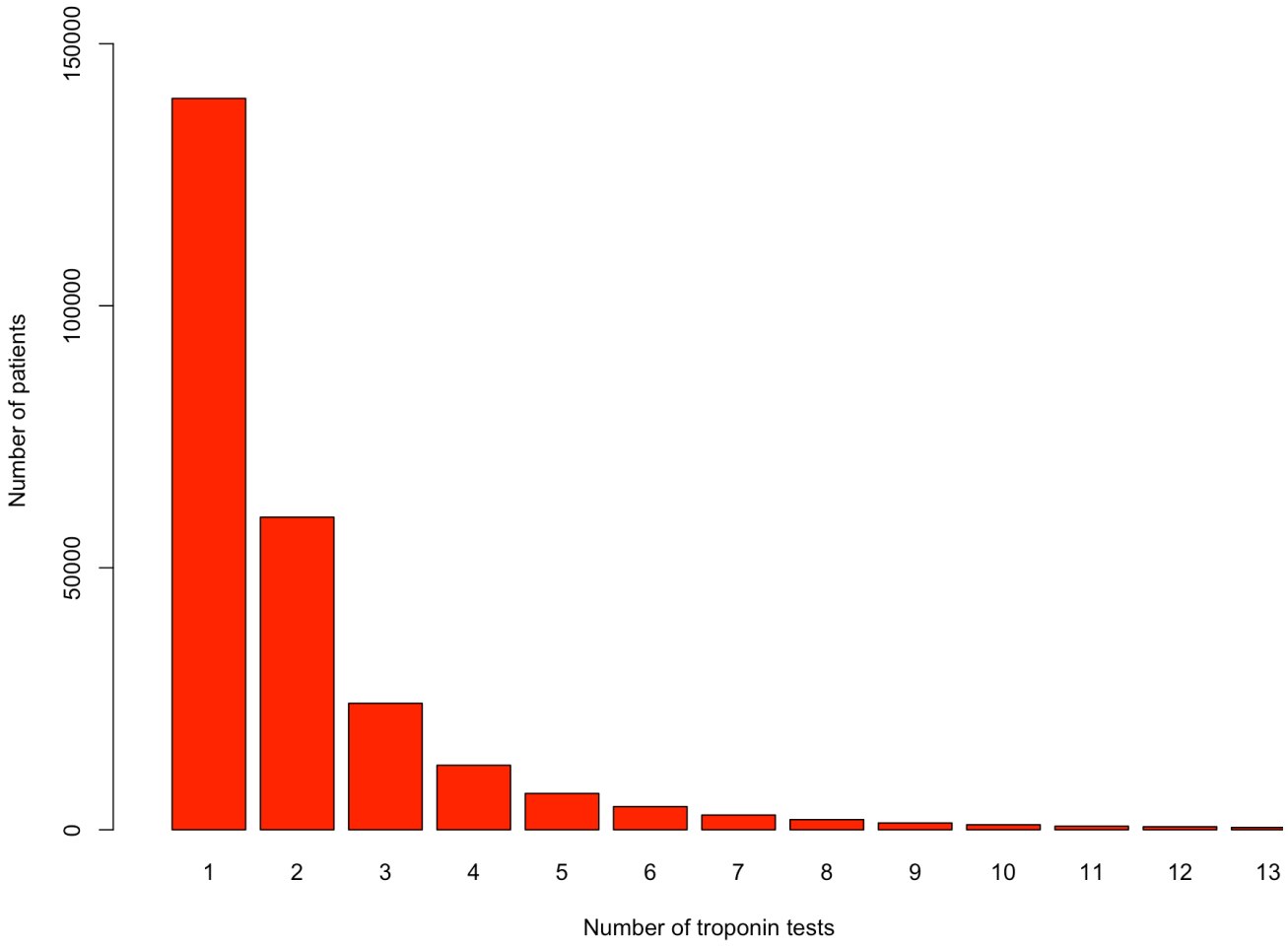
Data was obtained from five collaborating hospitals, which were all tertiary cardiac centres with emergency departments. All patients who underwent troponin measurements (n=257948) between 2010 and 2017 were included. For any patient with more than one hospital episode with troponin measurements during the study period, only the first episode within the study period was eligible. 52% (n=134517) of the patients were admitted to hospital

and so had International Statistical Classification of Diseases and Related Health Problems (ICD) discharge codes. Patients were classified as ACS based on the assigned ICD-10 codes (**Table 4.2**).[7] All other patients, who had 3383 different ICD-10 primary diagnostic codes between them, were classified as non-ACS. The top 10% of ICD-10 codes in non-ACS patients were individually reviewed and classified into diagnostic groups. This represented 83% of all diagnoses in the non-ACS group.

Invasive management was defined as having angiography within 3 months of the peak troponin level. Using a retrospective cohort study design, all patients were followed up until death or censoring in April 2017. Life status was ascertained using routinely collected data on the NHS Spine Application.

All analyses on troponin were performed using the peak troponin level. The distribution of the number of troponin tests measured during the index hospital admission for all patients is shown in **Figure 6.1**. For patients who had a single troponin measurement, the peak troponin was based on this measurement. In the remainder of the patients who had more than one troponin test in the same hospital episode of care, the peak troponin value was defined as the highest of all measurements. For patients with multiple episodes of care for which troponin was tested, the first episode of care was used for all analyses.

Figure 6.1 - Bar chart of the number of troponin tests per patient



Troponin data was treated in four ways. Each centre measured either troponin T or I using either standard or high-sensitivity assays (**Table 3.2**). These tests have unique cut-off points for the 99th percentile of the upper limit of normal (ULN). First, the troponin assays were standardised by scaling the results using the ULN for each troponin assay. Second, the results were dichotomised as being either negative or positive based on the ULN for each assay. Third, I addressed the relationship between troponin level and mortality using bands of troponin level: <½, ½-1, 1-2, 2-5, 5-10, 10-100, 100-1000 and >1000 xULN. Fourth, restricted cubic spline analyses were used to assess the relationship between troponin, on a continuous scale, and mortality.

6.4.2 Statistical analysis

The age distribution of patients who had a troponin measured was compared with that of the general population using the population estimates for

the UK in 2015.[102]

The main analysis of the association of progressively higher troponins with all-cause mortality was performed without making assumptions about the shape of the relationship, i.e. with Cox proportional hazards regression modelling. To assess the association between troponin positivity and all-cause mortality, Cox regression modelling was used. 3-year mortality was assessed and a landmark analysis was performed at 3 months. This involved restricting analyses to patients who survived to 3 months. Corresponding Kaplan-Meier plots were used to display cumulative mortality.

To model non-linear relationships, restricted cubic spline Cox regression analyses were used to model the relationship between mortality and troponin level, using log transformation because of the positive skew of the distribution of troponin values. Four knots were used to model troponin in the spline analyses. Splines were adjusted for patient age, gender, haemoglobin, creatinine, white cell count, C-reactive protein, platelet count, the number of troponin tests during the index hospital admission, family history of cardiovascular disease, current smoker, diabetes mellitus, hypertension, hypercholesterolaemia, heart failure, previous ischaemic heart disease, atrial fibrillation, aortic stenosis, chronic kidney disease, neoplasm and obstructive lung disease. The reference value for restricted cubic spline analyses was a hazard ratio of 1.

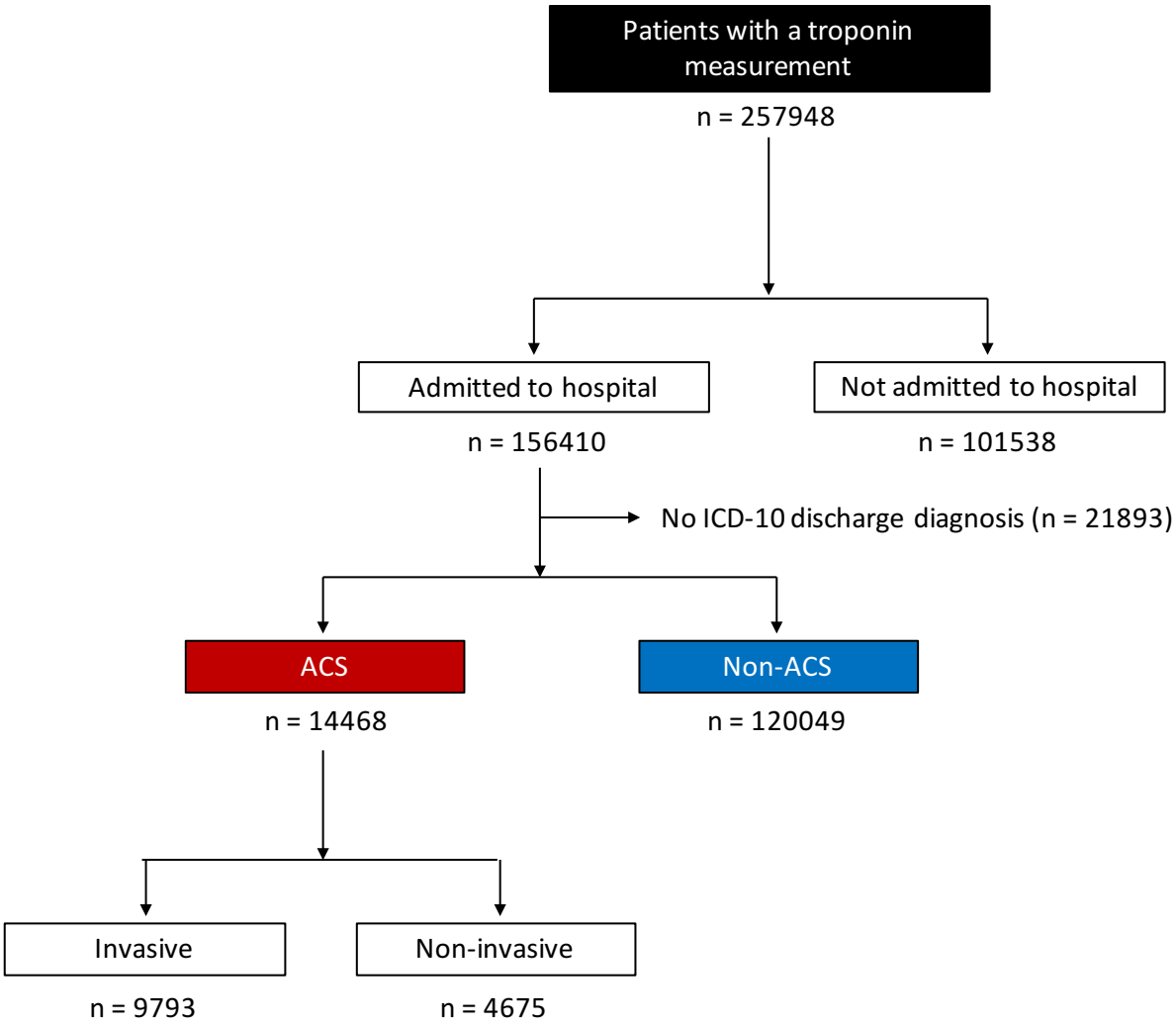
Statistical analyses were performed using SPSS version 24.0 (SPSS, United States) or R 3.5.0 (the R Core Team, Vienna, Austria).

6.5 Results

6.5.1 Baseline population characteristics

A total of 257,948 patients underwent troponin measurement during the study period. Among the 156,410 patients who were admitted to hospital with ICD-10 diagnostic codes, the final recorded diagnosis was ACS in 14,468 patients. A flow diagram of the study cohort is shown in **Figure 6.2**.

Figure 6.2 - Flow diagram of the study cohort in the TROP-RISK study



Every patient who had a troponin measurement during the study period was included. Each patient was eligible on the first occasion they had a troponin measurement during the study period. ACS, acute coronary syndrome; ICD, International Statistical Classification of Diseases and Related Health Problems.

In patients who did not have an ACS diagnosis, the top 10% of ICD-10 codes were reviewed and classified into diagnostic groups (**Table 6.2**). Among these patients, the coded diagnosis was categorised as being cardiovascular in 24%, non-cardiovascular in 45% and a broad range of symptoms in the remaining patients. The most common cardiovascular diagnoses amongst all non-ACS patients were atrial fibrillation/flutter (3%) and heart failure (2%). Pneumonia/lower respiratory tract infection (7%) and malignancy (4%) were the most common non-cardiovascular diagnoses in this non-ACS cohort.

Table 6.2 - Coded diagnoses in non-ACS patients. The top 10% of ICD-10 codes were reviewed in non-ACS patients and classified into diagnostic groups

Symptoms only coded	n (%)	Cardiovascular Disorders	n (%)	Non-cardiovascular disorders	n (%)
Chest pain	19977 (16.64)	Atrial fibrillation / Atrial flutter	3166 (2.64)	Pneumonia / lower respiratory tract infection	7752 (6.46)
Syncope and collapse	3661 (3.05)	Heart failure	2899 (2.41)	Malignancy	4416 (3.68)
Abdominal pain	1642 (1.37)	Pulmonary embolism	1429 (1.19)	Gastrointestinal disorders	5076 (4.23)
Shortness of breath / cough	1307 (1.09)	Chronic ischaemic heart disease	4897 (4.08)	Ischaemic stroke / transient ischaemic attack	3994 (3.33)
Palpitations	1135 (0.95)	Tachyarrhythmia	1450 (1.21)	Other types of infection	3769 (3.14)
Dizziness / giddiness	1033 (0.86)	Bradyarrhythmia	1355 (1.13)	Musculoskeletal disorders	2654 (2.21)
Headache	855 (0.71)	Valvular heart disease	1203 (1.00)	Drug, alcohol, metabolic disorders	2451 (2.04)
Fall	659 (0.55)	Aortic aneurysm	1104 (0.92)	Renal failure	1965 (1.64)
		Angina	1194 (0.99)	Skeletal fracture	1963 (1.64)
		Peripheral vascular disease	805 (0.67)	Other neurological disorders	1857 (1.55)
		Pericardial disease	782 (0.65)	Cerebral haemorrhage (subdural / intracerebral / subarachnoid haemorrhage)	1823 (1.52)
		Hypotension	721 (0.60)	Asthma / chronic obstructive pulmonary disease	1644 (1.37)
		Pulmonary hypertension	648 (0.54)	Other pulmonary disorders	1412 (1.18)
		Bundle branch block	341 (0.28)	Head injury	1133 (0.94)

		Other cardiac arrhythmias	336 (0.28)	Sepsis / Fever	852 (0.71)
		Hypertension	331 (0.28)	Epilepsy / seizure	797 (0.66)
		Cardiac arrest	281 (0.23)	Other haemorrhage	740 (0.62)
		Complication of cardiac device implant	281 (0.23)	Anaemia	644 (0.54)
		Myocarditis	171 (0.14)	Pregnancy-related conditions	195 (0.16)
		Carotid artery disease	169 (0.14)	Other haematological disorders	154 (0.13)
		Dilated cardiomyopathy	175 (0.15)		
		Infective endocarditis	154 (0.13)		
		Cardiac device / vascular graft infection	127 (0.11)		
		Hypertrophic cardiomyopathy	107 (0.09)		
		Other cardiomyopathies	90 (0.07)		
Total	**30269 (25.21)**	**Total**	**24216 (20.17)**	**Total**	**45291 (37.73)**

16.9% (n=20273) of patients had other, less common, ICD-10 codes.

The median age was 65 (IQR 50 to 79) years and 55% were men (**Table 6.3**). **Table 6.4** details the proportion of missing data for each variable. Comparing the age distribution of the 257948 patients in our cohort (**Figure 6.3A**) with that of the general population using WHO population estimates for the UK in 2015 (**Figure 6.3B**), the relative rate of patient presentations who had a troponin measured rose approximately linearly with age, from 20 to 60 years old, and then increased exponentially by approximately 50% per decade of age (**Figure 6.3C**).

The proportion of troponins that were positive, progressively rose with age (**Figure 6.3D**) from 9% in the 18-29 band to 50% in the over 90s. The median positive troponin was 2280 times higher than the median negative troponin, and this relationship was observed across all age groups.

Over a median follow-up of 1198 (IQR 514 to 1866) days, there were

55850 (22%) deaths, of which 31112 (12%) occurred in the first year. The Kaplan-Meier 3-year mortality rate in the cohort was 20%. The mortality rates for each age group are shown in the line graph in **Figure 6.3A**.

Table 6.3 - Characteristics for all patients, troponin positive and troponin negative groups

	All patients (n=257948) **(median (IQR) or number (%))**	Troponin negative (n=180239) **(median (IQR) or number (%))**	Troponin positive (n=77709) **(median (IQR) or number (%))**	**P-value‡**
General demographics				
Age (years)	65 (50 - 79)	60 (46 - 75)	74 (62 - 84)	<0.0001
Male	142718 (55.3)	95551 (53.0)	47167 (60.7)	<0.0001
Ethnicity				
White	147321 (74.5)	103907 (73.4)	43414 (77.3)	<0.0001
South Asian	7169 (3.6)	4823 (3.4)	2346 (4.2)	
Black	19925 (10.1)	14961 (10.6)	4964 (8.8)	
Other	23401 (11.8)	17946 (12.7)	5455 (9.7)	
Haematology / Biochemistry				
Haemoglobin (g/dL)	13.3 (11.9 - 14.6)	13.6 (12.3 - 14.7)	12.6 (10.9 - 14.1)	<0.0001
White cell count (x10^9/L)	8.3 (6.5 - 10.9)	8.0 (6.3 - 10.2)	9.5 (7.2 - 12.7)	<0.0001
Platelet count (x10^9/L)	231 (188 - 280)	234 (194 - 281)	222 (174 - 278)	<0.0001
Creatinine (µmol/L)	77 (65 - 96)	74 (63 - 89)	90 (71 - 126)	<0.0001
Urea (mmol/L)	5.7 (4.4 - 7.9)	5.3 (4.1 - 6.9)	7.2 (5.2 - 11.2)	<0.0001
C-reactive protein (mg/L)	6.0 (2.0 - 29.8)	5.0 (1.6 - 18.0)	14.0 (4.3 - 63.0)	<0.0001
Glucose (mmol/L)	5.7 (5 - 7.3)	5.5 (4.9 - 6.8)	6.3 (5.3 - 8.3)	<0.0001
HbA1c (mmol/mol)	43 (38 - 56)	37 (33 - 42)	38 (34 - 44)	<0.0001
Total cholesterol (mmol/L)	4.5 (3.7 - 5.4)	4.7 (3.9 - 5.5)	4.2 (3.4 - 5.2)	<0.0001
HDL cholesterol (mmol/L)	1.2 (1.0 - 1.5)	1.3 (1.0 - 1.6)	1.1 (0.9 - 1.5)	<0.0001
Triglyceride level (mmol/L)	1.3 (0.9 - 1.9)	1.3 (0.9 - 1.9)	1.2 (0.9 - 1.8)	<0.0001
Echocardiography				
LVEF (%)	55.0 (48.1 - 64.5)	60.0 (55.0 - 65.8)	55.0 (42.3 - 61.8)	<0.0001
LVFS (%)	32.7 (26.5 - 38.0)	33.9 (28.8 - 39.0)	31.0 (23.4 - 37.0)	<0.0001
LVEDd (cm)	4.7 (4.2 - 5.2)	4.6 (4.1 - 5.0)	4.8 (4.3 - 5.3)	<0.0001
LVEDs (cm)	3.1 (2.7 - 3.6)	3.0 (2.6 - 3.4)	3.2 (2.7 - 3.9)	<0.0001

Hospital episode				
Admitted to hospital	134517 (52.1)	80327 (44.6)	54190 (69.7)	<0.0001
Diagnosed with ACS	14468 (5.6)	1647 (0.91)	12821 (16.5)	<0.0001

‡ Comparison between troponin negative and troponin positive groups using Mann-Whitney U test for continuous variables and Chi square test for categorical variables. ACS, acute coronary syndrome, LVEF, left ventricular ejection fraction; LVFS, left ventricular fractional shortening; LVEDd, left ventricular end diastolic diameter; LVEDs, left ventricular end systolic diameter.

Table 6.4 - Characteristics and proportion of missing data for all patients, troponin positive and troponin negative groups

	All patients (n=257948)		Troponin negative (n=180239)		Troponin positive (n=77709)		
	n (median (IQR) or number (% of total patients))	**Total patients (% of all patients)**	**n (median (IQR) or number (% of total patients))**	**Total patients (% of all patients)**	**n (median (IQR) or number (% of total patients))**	**Total patients (% of all patients)**	**P-value‡**
General demographics							
Age (years)	65 (50 - 79)	257852 (99.96)	60 (46 - 75)	180165 (99.96)	74 (62 - 84)	77687 (99.97)	<0.0001
Male	142718 (55.3)	257834 (99.96)	95551 (53.0)	180155 (99.95)	47167 (60.7)	77679 (99.96)	<0.0001
Ethnicity		197816 (76.69)		141637 (78.58)		56179 (72.29)	
White	147321 (74.5)		103907 (73.4)		43414 (77.3)		<0.0001
South Asian	7169 (3.6)		4823 (3.4)		2346 (4.2)		
Black	19925 (10.1)		14961 (10.6)		4964 (8.8)		
Other	23401 (11.8)		17946 (12.7)		5455 (9.7)		
Haematology / Biochemistry							
Haemoglobin (g/dL)	13.3 (11.9 - 14.6)	231265 (89.66)	13.6 (12.3 - 14.7)	164413 (91.22)	12.6 (10.9 - 14.1)	66852 (86.03)	<0.0001
White cell count (x10^9/L)	8.3 (6.5 - 10.9)	231276 (89.66)	8.0 (6.3 - 10.2)	164421 (91.22)	9.5 (7.2 - 12.7)	66855 (86.03)	<0.0001
Platelet count (x10^9/L)	231 (188 - 280)	231185 (89.62)	234 (194 - 281)	164362 (91.19)	222 (174 - 278)	66823 (85.99)	<0.0001
Creatinine (µmol/L)	77 (65 - 96)	231861 (81.89)	74 (63 - 89)	164729 (91.39)	90 (71 - 126)	67132 (86.39)	<0.0001
Urea (mmol/L)	5.7 (4.4 - 7.9)	193688 (75.09)	5.3 (4.1 - 6.9)	135458 (75.15)	7.2 (5.2 - 11.2)	58230 (74.93)	<0.0001
C-reactive protein (mg/L)	6.0 (2.0 - 29.8)	186686 (76.25)	5.0 (1.6 - 18.0)	133884 (74.28)	14.0 (4.3 - 63.0)	62802 (80.82)	<0.0001
Glucose (mmol/L)	5.7 (5 - 7.3)	101332 (39.28)	5.5 (4.9 - 6.8)	69346 (38.47)	6.3 (5.3 - 8.3)	31986 (41.16)	<0.0001

HbA1c (mmol/mol)	43 (38 - 56)	27885 (10.81)	37 (33 - 42)	15655 (8.69)	38 (34 - 44)	12230 (15.74)	<0.0001
Total cholesterol (mmol/L)	4.5 (3.7 - 5.4)	110230 (42.73)	4.7 (3.9 - 5.5)	74788 (41.49)	4.2 (3.4 - 5.2)	35442 (45.61)	<0.0001
HDL cholesterol (mmol/L)	1.2 (1.0 - 1.5)	106818 (41.41)	1.3 (1.0 - 1.6)	72763 (40.37)	1.1 (0.9 - 1.5)	34055 (43.82)	<0.0001
Triglyceride level (mmol/L)	1.3 (0.9 - 1.9)	99804 (38.69)	1.3 (0.9 - 1.9)	66287 (36.78)	1.2 (0.9 - 1.8)	33517 (43.13)	<0.0001
Echocardiography							
LVEF (%)	55.0 (48.1 - 64.5)	31591 (12.25)	60.0 (55.0 - 65.8)	13614 (7.55)	55.0 (42.3 - 61.8)	17977 (23.13)	<0.0001
LVFS (%)	32.7 (26.5 - 38.0)	26054 (10.10)	33.9 (28.8 - 39.0)	13887 (7.70)	31.0 (23.4 - 37.0)	12167 (15.66)	<0.0001
LVEDd (cm)	4.7 (4.2 - 5.2)	46941 (18.20)	4.6 (4.1 - 5.0)	21674 (12.03)	4.8 (4.3 - 5.3)	25267 (32.51)	<0.0001
LVEDs (cm)	3.1 (2.7 - 3.6)	39507 (15.32)	3.0 (2.6 - 3.4)	19145 (10.62)	3.2 (2.7 - 3.9)	20362 (26.20)	<0.0001
Hospital episode							
Admitted to hospital	134517 (52.1)	257948 (100)	80327 (44.6)	180239 (100)	54190 (69.7)	77709 (100)	<0.0001
Diagnosed with ACS	14468 (5.6)	257948 (100)	14468 (5.6)	180239 (100)	12821 (16.5)	77709 (100)	<0.0001

‡ Comparison between troponin negative and troponin positive groups using Mann-Whitney U test for continuous variables and Chi square test for categorical variables. ACS, acute coronary syndrome; LVEF, left ventricular ejection fraction; LVFS, left ventricular fractional shortening; LVEDd, left ventricular end diastolic diameter; LVEDs, left ventricular end systolic diameter.

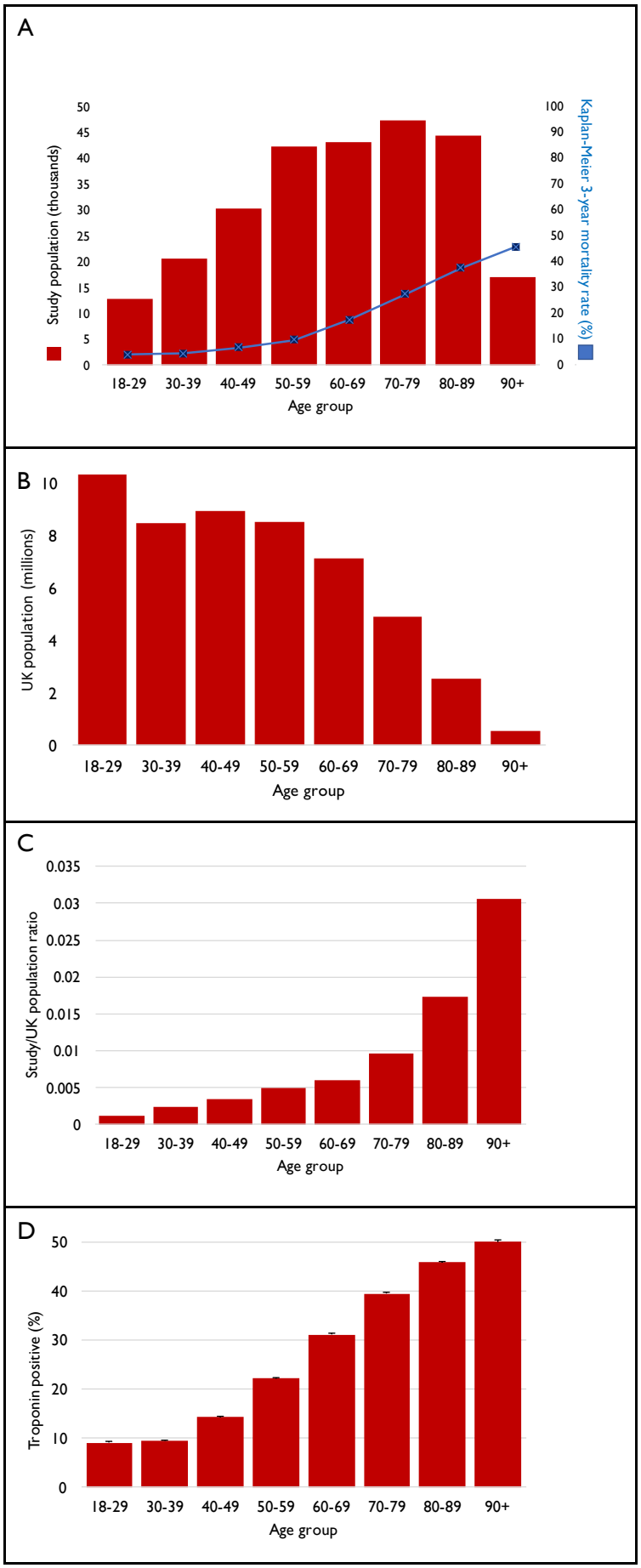
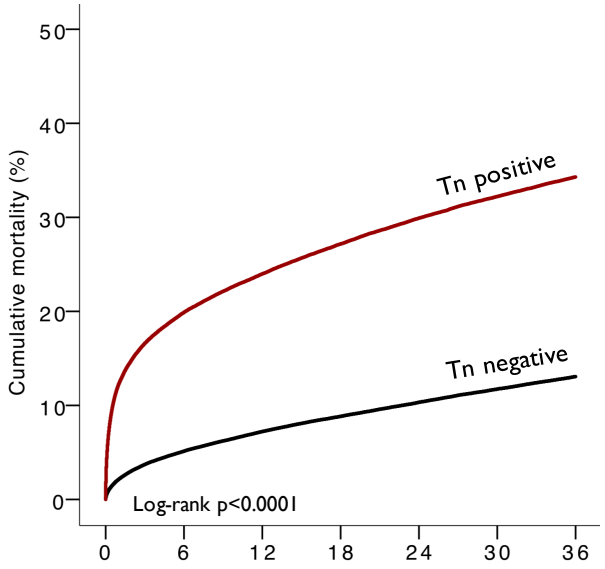


Figure 6.3 - A – Age distribution and 3-year estimated mortality rate in this study; B – Age distribution of the UK population; C – Ratio of study/UK age distribution; D – Troponin positivity across age groups

6.5.2 Relationship between troponin positivity, age and mortality

A positive troponin was associated with an overall 3.2-fold higher mortality hazard (95% CI, 3.1 to 3.2) than a negative troponin over 3 years in the full cohort (Kaplan-Meier plot, **Figure 6.4**; baseline characteristics for the troponin positive and troponin negative groups, **Table 6.4**). For young patients (18-29 years) this was a particularly strong effect, with a mortality hazard ratio of 10.6 (95% CI, 8.5 to 13.3). There was a progressive decline in mortality hazard ratio with age to a value of 1.5-fold (95%, CI 1.4 to 1.6) in the over 90s (**Figure 6.5A**).

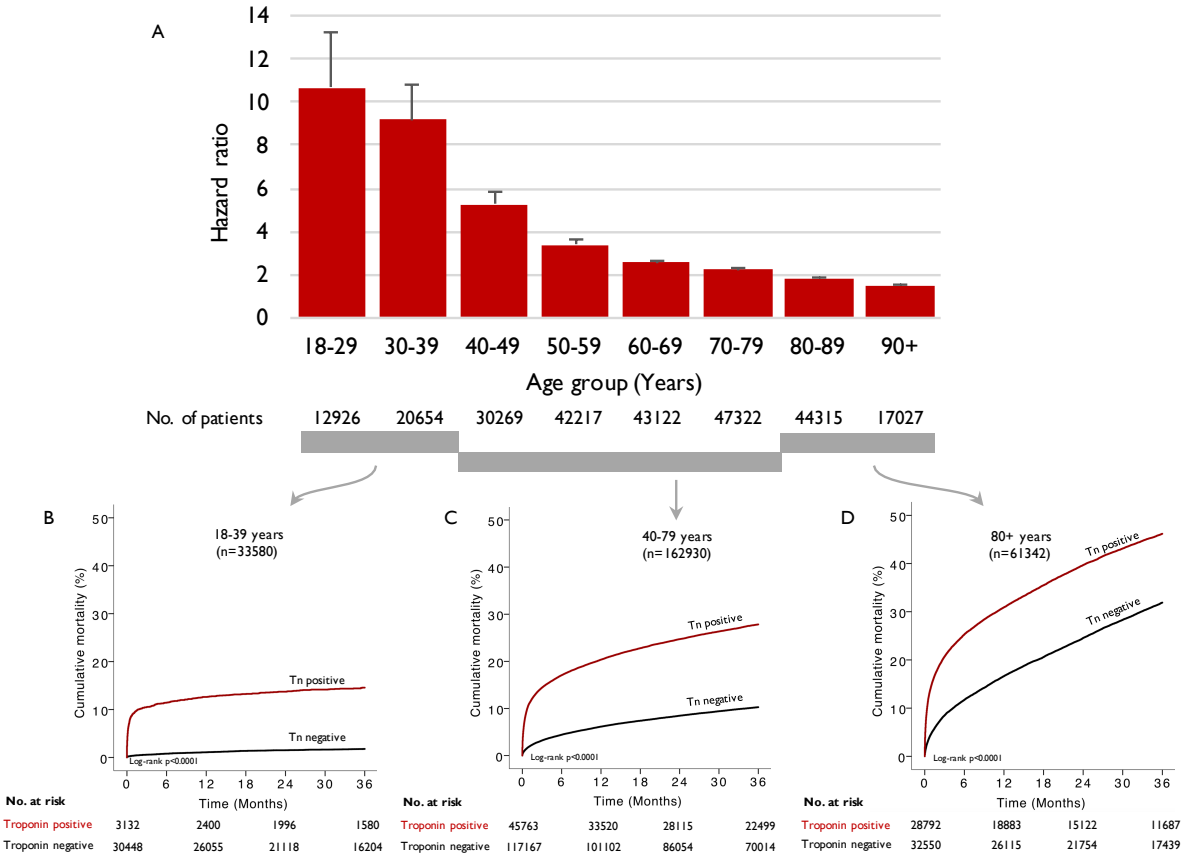
Figure 6.4 - Kaplan-Meier mortality curve by troponin positivity in all patients



No. at risk				
Troponin positive	77709	54823	45246	35777
Troponin negative	180239	153339	128984	103711

Tn, troponin.

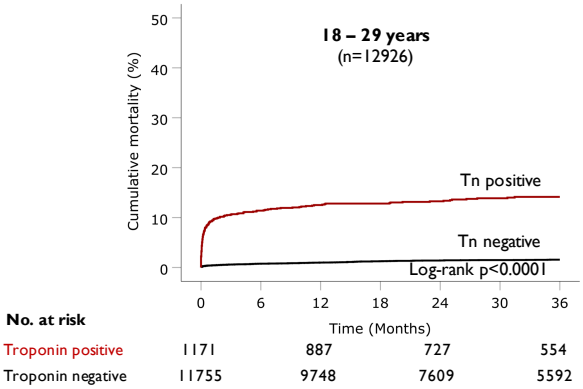
Figure 6.5 - A – Hazard ratios for troponin positive versus negative groups across different age bands for all patients; B–D – Kaplan-Meier mortality curve by troponin positivity in 18-39, 40-79 and 80+ years age bands



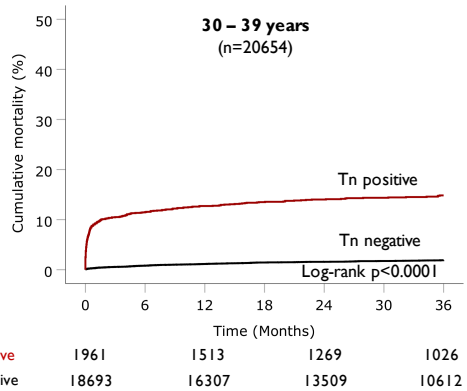
Error bars denote upper 95% confidence interval. Tn, troponin.

Individual survival curves for troponin positive and negative groups are shown for the 18-39, 40-79 and >80 years age groups in **Figure 6.5B-D** and for each age decade in **Figure 6.6**. In absolute terms, the increment in 3-year mortality associated with a positive troponin ranged from 12.8 to 17.5 percentage points across every age bands, with an overall mean increment of 14.8 percentage points across all patients (**Figure 6.5B-D**). Although most of the mortality increment with a positive troponin was within the first three months, even when a landmark analysis was conducted at 3 months (**Figure 6.7**), there was still a perceptible elevation in mortality with a positive troponin which was present in all three age groups, and grew with time.

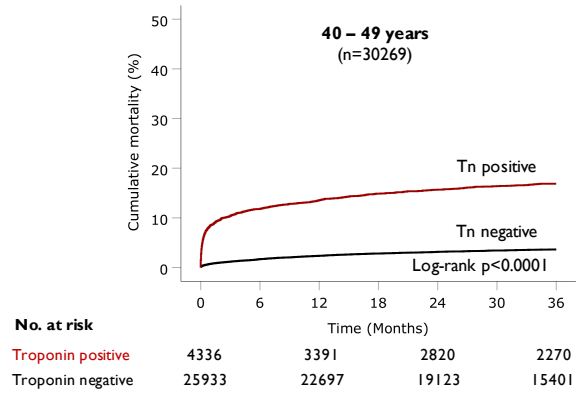
Figure 6.6 - Kaplan-Meier survival curve by troponin positivity in each age decade. Error bars denote upper 95% confidence interval



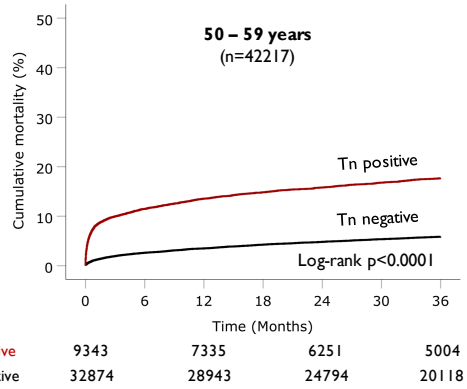
No. at risk				
Troponin positive	1171	887	727	554
Troponin negative	11755	9748	7609	5592



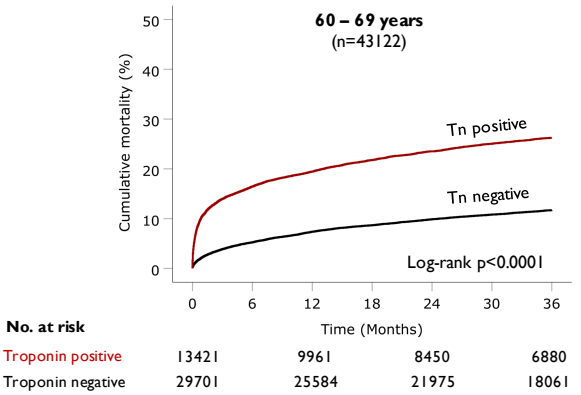
No. at risk				
Troponin positive	1961	1513	1269	1026
Troponin negative	18693	16307	13509	10612



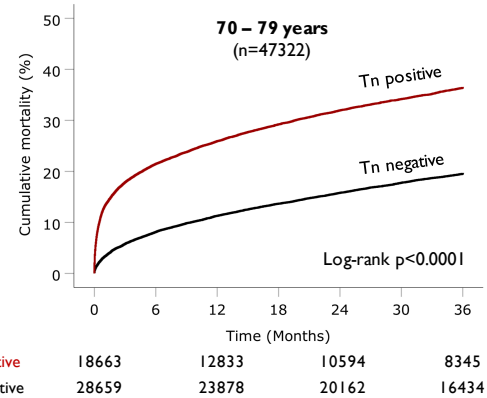
No. at risk				
Troponin positive	4336	3391	2820	2270
Troponin negative	25933	22697	19123	15401



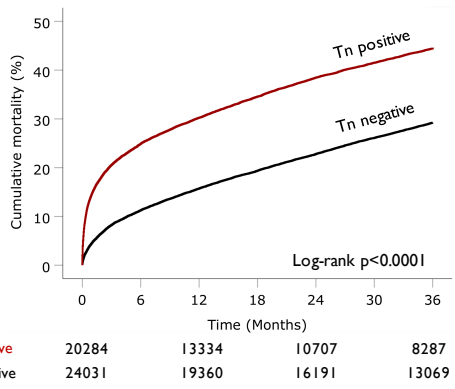
No. at risk				
Troponin positive	9343	7335	6251	5004
Troponin negative	32874	28943	24794	20118



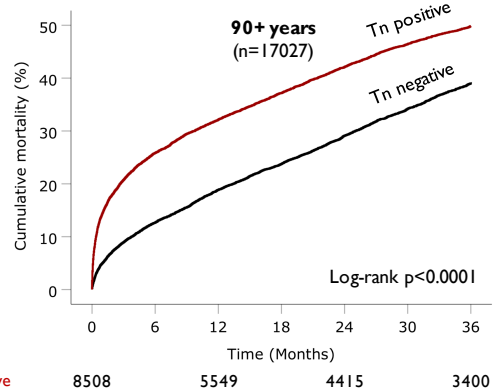
No. at risk				
Troponin positive	13421	9961	8450	6880
Troponin negative	29701	25584	21975	18061



No. at risk				
Troponin positive	18663	12833	10594	8345
Troponin negative	28659	23878	20162	16434



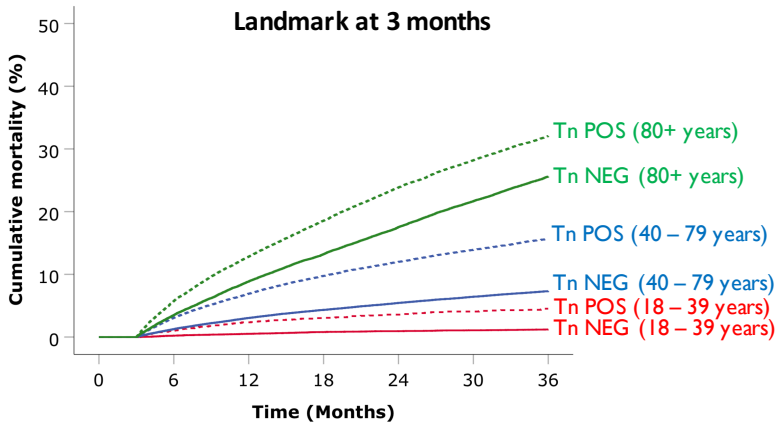
No. at risk				
Troponin positive	20284	13334	10707	8287
Troponin negative	24031	19360	16191	13069



No. at risk				
Troponin positive	8508	5549	4415	3400
Troponin negative	8519	6755	5563	4370

Tn, troponin.

Figure 6.7 - Kaplan-Meier survival curves of 3 month landmark analysis, stratified by troponin positivity in 18-39, 40-79 and 80+ years age groups



No. at risk				
Tn POS (80+)	22859	18883	15122	11687
Tn NEG (80+)	29807	26115	21754	17439
Tn POS (40 – 79)	39171	33520	28115	22499
Tn NEG (40 – 79)	113417	101102	86054	70014
Tn POS (18 – 39)	2802	2400	1996	1580
Tn NEG (18 – 39)	30274	26055	21118	16204

Tn NEG, troponin negative; Tn POS, troponin positive.

6.5.3 Distortion of troponin-mortality link

Although there was indeed progressively increasing mortality up to approximately 5-10 xULN, above this there was an unexpected progressive decline in mortality (**Figure 6.8A**).

Figure 6.8B shows the frequency of deaths in different time windows per 100 patients in each of the troponin bands. They are derived from the cumulative mortality percentage from the Kaplan-Meier curves in **Figure 6.8A**. Notably, the patients with a very high troponin level had high mortality up to 1-month. Beyond that, the ongoing mortality was similar to patients whose troponin values were normal.

Figure 6.8 - A – Three-year Kaplan-Meier cumulative mortality by troponin level for all patients; B – Deaths per 100 patients undergoing troponin testing during each of the time periods shown

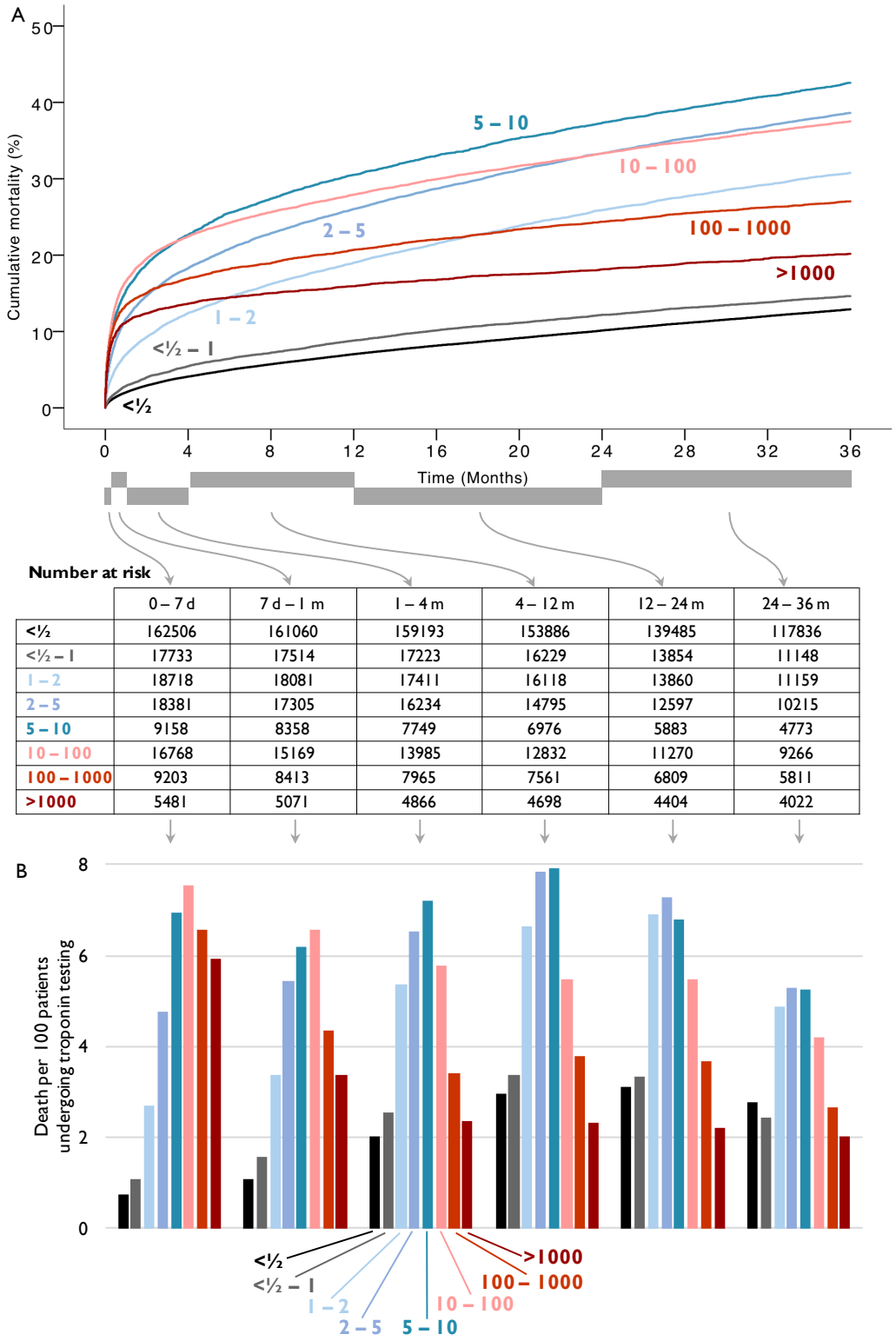
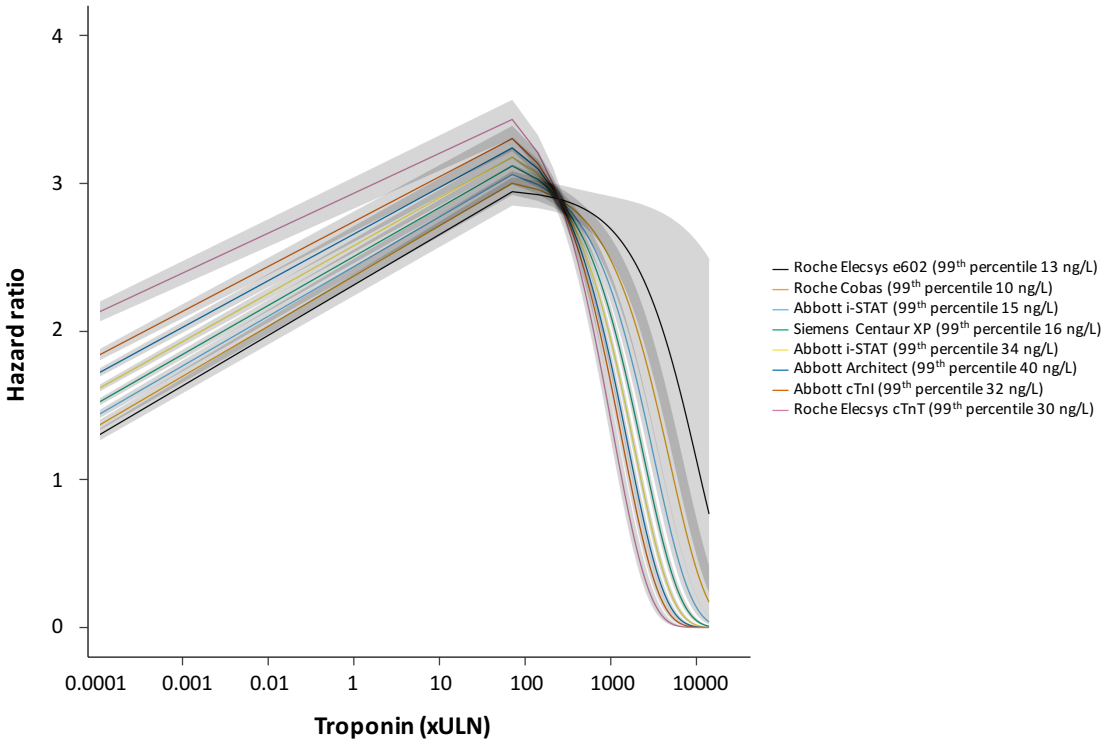


Table 3.2 summarises the details of the different troponin assays used across the different centres. The inverted U-shaped relationship was observed across all troponin assays (**Figure 6.9**) and was also present when using the first troponin level during the index hospital admission.

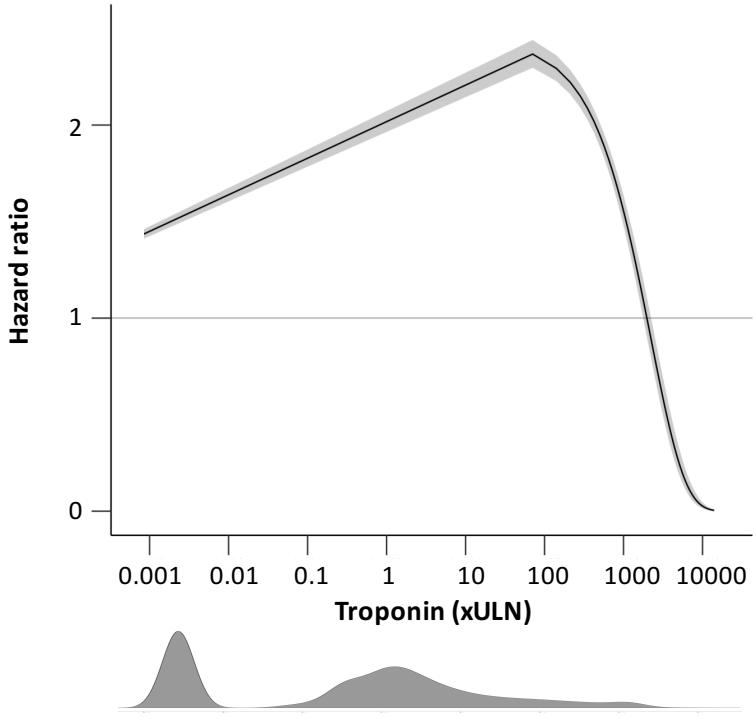
Figure 6.9 - Association between troponin level and the hazard ratio for all-cause mortality using restricted cubic splines stratified by troponin assay



The grey shaded area around the spline curves represents the 95% confidence interval. All curves arise from a single restricted cubic spline analysis using troponin assay as a stratifier. The reference (hazard ratio=1) is the full group of patients (n=247948).

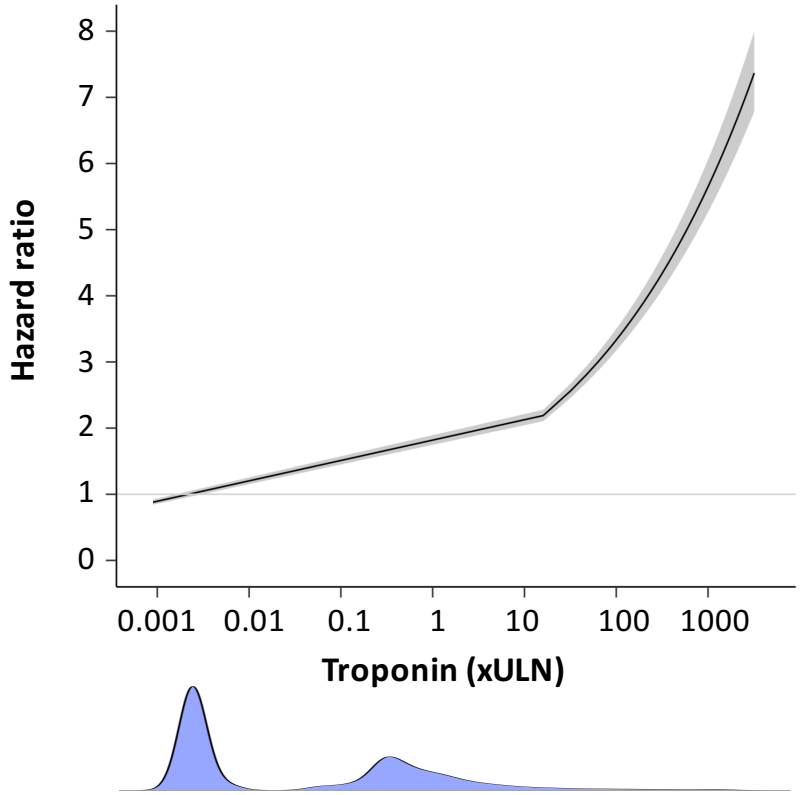
On subgroup analyses, for patients who were and were not admitted to hospital, the inverted U-shaped relationship was only observed in patients who were admitted to hospital (n=156,410) (**Figure 6.10**) with the peak hazard ratio of 2.4 (95% CI, 2.3 to 2.4) at a peak troponin of ~70 xULN. There was a direct and positive relationship between troponin level and hazard ratio across all troponin values for patients not admitted to hospital (n=101,538) (**Figure 6.11**). The Kaplan-Meier 3-year mortality rate was 14% in patients not admitted to hospital, which was lower than the mortality rate of 23% observed in patients admitted to hospital.

Figure 6.10 - Unadjusted association between peak troponin level and the hazard ratio for mortality for all patients admitted to hospital



The grey shaded area around the spline curves represents the 95% confidence interval. The probability distribution of troponin level in all patients (grey shade), is displayed below the x-axis. ACS, acute coronary syndrome. Hazard ratio=1 is the reference value.

Figure 6.11 - Association between peak troponin level and the hazard ratio for all-cause mortality in patients not admitted to hospital (n=101538)

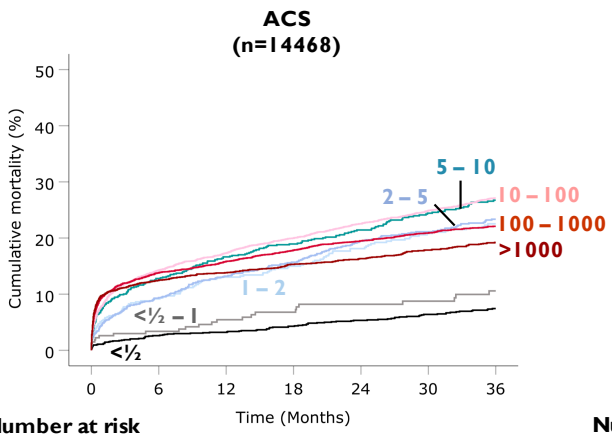


The grey shaded area around the spline curves represents the 95% confidence interval. The probability distribution of troponin level in patients not admitted to hospital is displayed below the x-axis. Hazard ratio=1 is the reference value.

6.5.4 Relationship between troponin level and mortality stratified by ACS diagnosis and invasive management

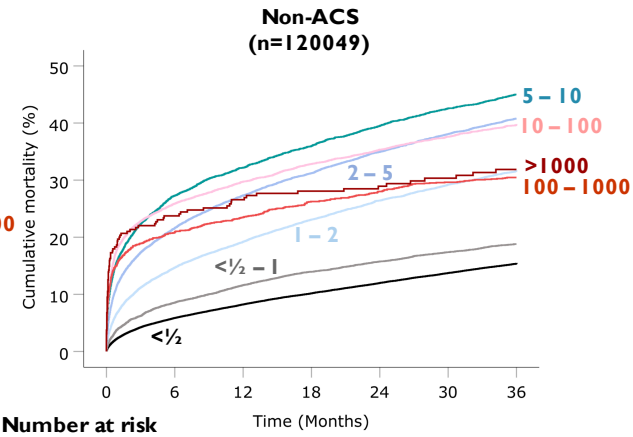
The inverted U-shaped relationship between troponin level and mortality was demonstrated in both ACS (n=14,468) and non-ACS patients (n=120,049) (**Figure 6.12**). In a multivariable model, adjusting for key demographic and clinical factors, the inverted U-shaped pattern between troponin and mortality remained in ACS patients (**Figure 6.13A**), but was a direct positive relationship in non-ACS patients (**Figure 6.13B**).

Figure 6.12 - Three-year Kaplan-Meier cumulative mortality by troponin level for ACS and non-ACS patients



Number at risk

	0–6m	6–12m	12–18m	18–24m	24–30m	30–36m
<½	1379	1331	1283	1216	1109	1034
<½ – 1	268	243	217	198	171	161
1 – 2	569	501	458	423	374	341
2 – 5	1069	944	859	789	704	628
5 – 10	878	746	670	611	544	484
10 – 100	3702	3082	2821	2564	2320	2046
100 – 1000	4248	3559	3267	3009	2733	2451
>1000	2355	1987	1847	1707	1592	1460

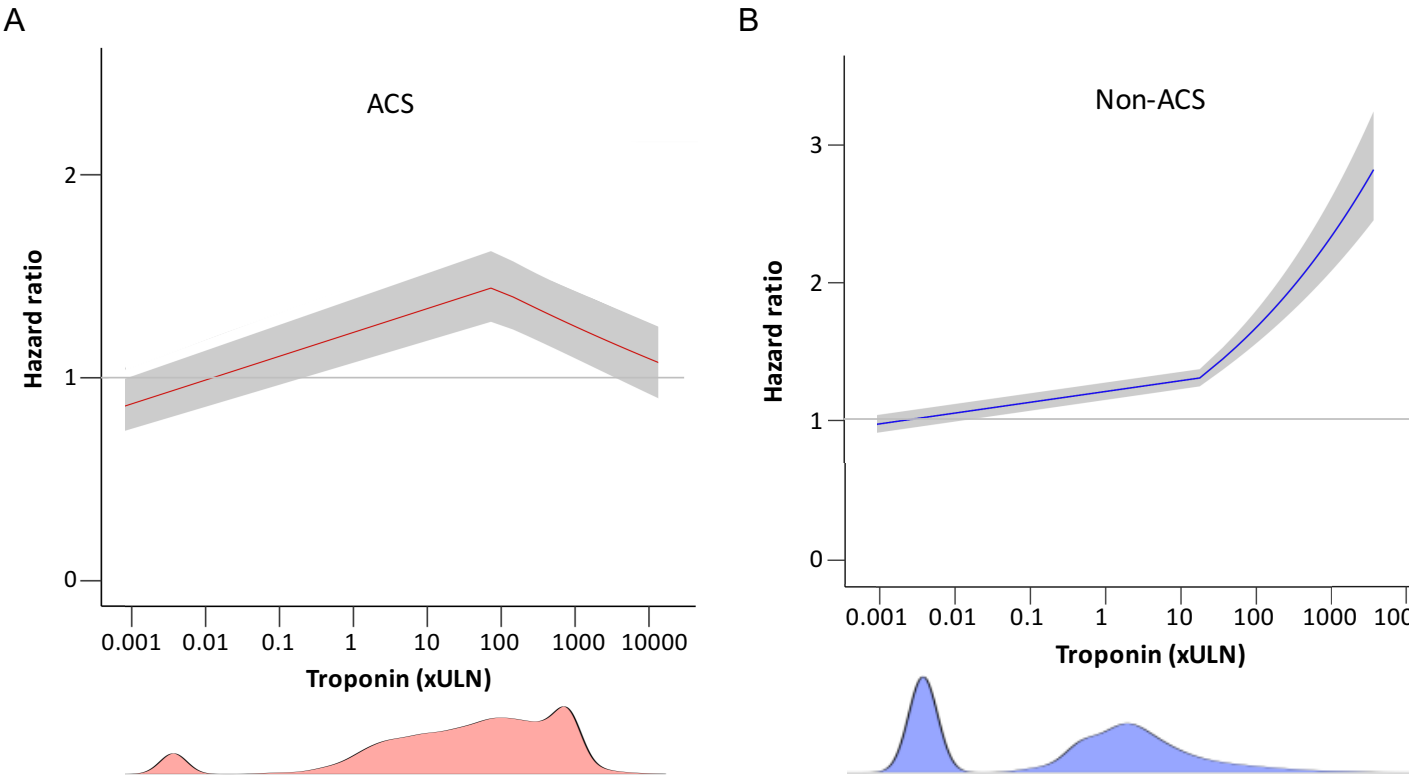


Number at risk

	0–6m	6–12m	12–18m	18–24m	24–30m	30–36m
<½	68525	63387	59307	55079	50274	45515
<½ – 1	10156	8744	7682	6803	6101	5500
1 – 2	13222	10931	9773	8719	7793	6905
2 – 5	12153	9261	8136	7252	6451	5639
5 – 10	5465	3850	3385	3002	2651	2315
10 – 100	8044	5797	5215	4665	4161	3687
100 – 1000	2189	1697	1550	1399	1275	1145
>1000	295	222	201	181	157	142

ACS, acute coronary syndrome.

Figure 6.13 - Multivariable adjusted spline of association between peak troponin level and the hazard ratio for all-cause mortality in (A) ACS and (B) non-ACS subgroups



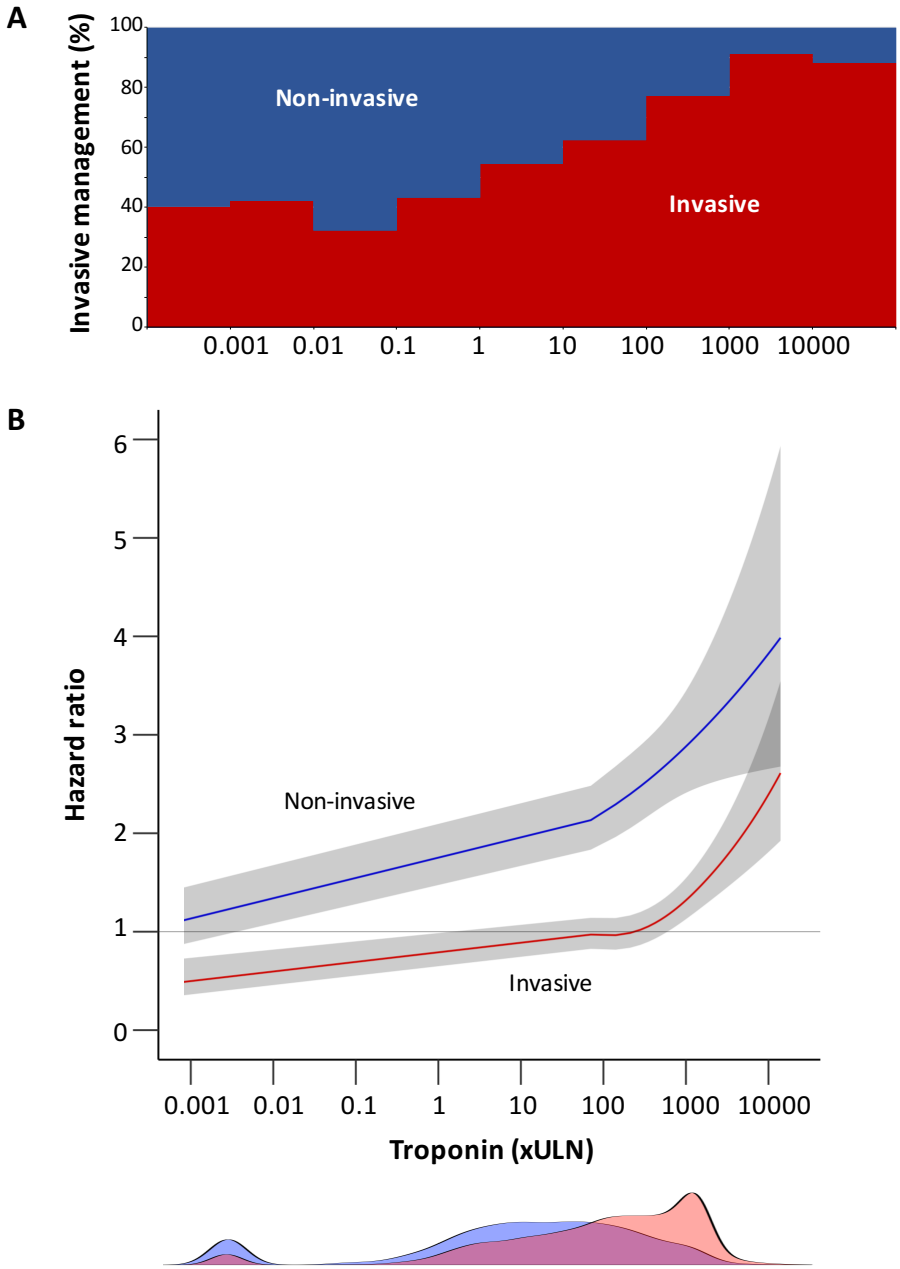
Adjusted for patient age, gender, haemoglobin, creatinine, white cell count, C-reactive protein, platelet count, the number of troponin tests during the index hospital admission, family history of cardiovascular disease, current smoker, diabetes mellitus, hypertension, hypercholesterolaemia, heart failure, previous ischaemic heart disease, atrial fibrillation, aortic stenosis, chronic kidney disease, neoplasm and obstructive lung disease. The grey shaded area around the spline curves represents the 95% confidence interval. Hazard ratio=1 represents the overall ACS group in panel A and the overall non-ACS group in panel B. The probability distribution of troponin level in ACS (red shade) and non-ACS (blue shade) patients is displayed below the x-axis. ACS, acute coronary syndrome.

The rate of invasive management at 3 months was 68% (n=9,793) and 6% (n=7,651) in ACS and non-ACS patients, respectively. The time to revascularisation was 1.3 days (95% CI, 1.2 to 1.4) and 11.7 days (95% CI, 10.7 to 12.6) in ACS patients who underwent PCI and CABG, respectively. Amongst patients who did not have invasive management within 3 months of their first troponin, 4% and 2% of patients had subsequent invasive management during follow-up in the ACS and non-ACS groups, respectively.

In ACS patients, there was a very different rate of invasive management across the spectrum of troponin. The rate was only 39% for troponins below 1 xULN, and 54% between 1 and 10 xULN. Beyond 10 xULN, the rate of invasive management rose to over 90% for greater than 1000 xULN (**Figure 6.14A**). Stratifying ACS patients who were admitted to hospital by invasive management, the restricted cubic spline Cox regression curve showed a progressive increase in mortality within both the invasive and non-invasive strata, even to very high troponin levels (**Figure 6.14B**).

Overall, across all troponin levels, invasive management was associated with lower hazard ratios than non-invasive management in both ACS (**Figure 6.14B**) and non-ACS (**Figure 6.15**) patients.

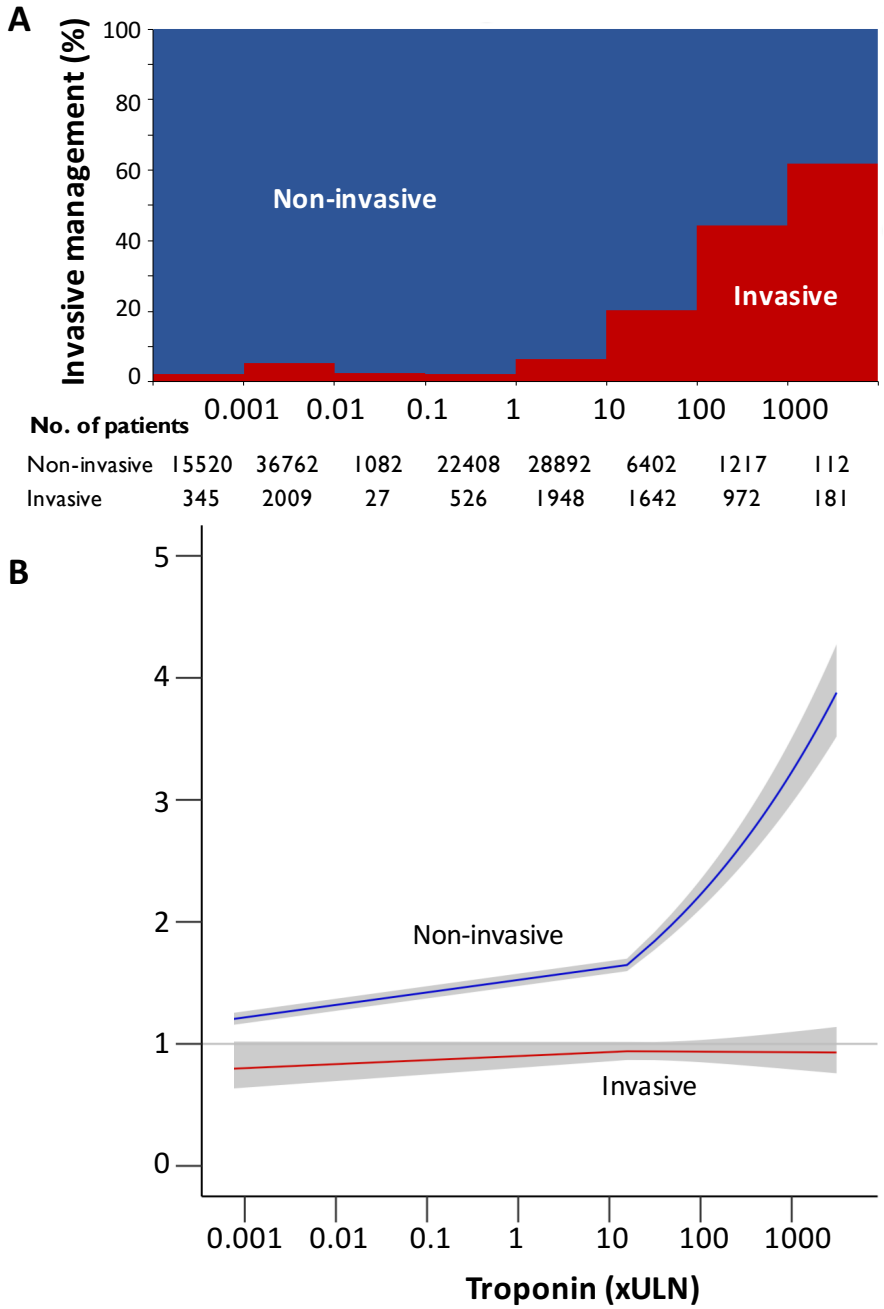
Figure 6.14 - A, Proportion of patients with ACS undergoing invasive management according to troponin level; B, Multivariable adjusted spline of association between peak troponin level and the hazard ratio for all-cause mortality in patients with ACS stratified by invasive versus non-invasive management



Adjusted for patient age, gender, haemoglobin, creatinine, white cell count, C-reactive protein, platelet count, the number of troponin tests during the index hospital admission, family history of cardiovascular disease, current smoker, diabetes mellitus, hypertension, hypercholesterolaemia, heart failure, previous ischaemic heart disease, atrial fibrillation, aortic stenosis, chronic kidney disease, neoplasm and obstructive lung disease. The grey shaded area around the spline curves represents the 95% confidence interval. Both curves arise from a single restricted cubic spline analysis using invasive management as a stratifier. The reference (hazard ratio=1) is the full

group of patients with ACS (n=14468) who did or did not undergo invasive management. The probability distribution of troponin level in invasively managed (red shade) and non-invasively managed (blue shade) patients is displayed at the bottom of the Figure. ACS, acute coronary syndrome.

Figure 6.15 - A, Proportion of patients with non-ACS undergoing invasive management according to troponin level; B, Multivariable adjusted spline of association between peak troponin level and the hazard ratio for all-cause mortality in patients with non-ACS stratified by invasive versus non-invasive management



Adjusted for patient age, gender, haemoglobin, creatinine, white cell count, C-reactive protein, platelet count, the number of troponin tests during the index hospital admission, family history of cardiovascular disease, current smoker, diabetes mellitus, hypertension, hypercholesterolaemia, heart failure, previous ischaemic heart disease, atrial fibrillation, aortic stenosis, chronic kidney disease, neoplasm and obstructive lung disease. The grey shaded area around the spline curves represents the 95% confidence interval. Both curves arise from a single restricted cubic spline analysis using invasive management as a stratifier. The reference (hazard ratio=1) is the full group of patients with non-ACS (n=120049) who did or did not undergo invasive management.

6.6 Discussion

6.6.1 Summary of main findings

This is the first study to address the implications of a raised troponin in a large sample of patients who have had a troponin test for a clinical reason across all ages. Across all age groups, a positive troponin result was found to be clinically meaningful, even if the troponin value was slightly elevated above the normal cut-off. The inclusion of over a quarter of a million patients in this study permitted assessment for the first time of the prognostic power of the troponin value across a wide range of levels, and across the whole spectrum of ages. Even a mildly elevated troponin had prognostic implications in both ACS and non-ACS patients. In those patients admitted to hospital, there was an unexpected inverted U-shaped relationship between troponin level and mortality. For patients with a final non-ACS diagnosis, multivariable adjustment led to the expected direct positive relationship between troponin level and mortality. In the ACS group, the inverted U-shaped relationship remained after multivariable adjustment, however, did revert to a direct positive relationship after stratification of patients with ACS by whether or not they underwent invasive coronary management.

There are often challenges in interpreting troponin results in real-life clinical practice, outside the rigid confines of clinical trials. Troponin levels are measured not only in patients with chest pain, but often also others with a wide variety of presentations. As a result, the troponin test is sometimes positive in isolation, with no other evidence of myocardial ischaemia but with evidence of another clinical diagnosis, such as a urinary tract infection.[43,44] In such circumstances, clinicians may treat the underlying primary diagnosis and not further investigate the possibility of secondary coronary artery disease. There is evidence of substantial under-recognition in hospitalised patients of cardiovascular and other conditions that lead to a fatal myocardial infarction.[103]

Additionally, in elderly patients, there are relatively few finely-stratified

prognostic data, and invasive treatment has not clearly been shown to lead to an improved mortality compared with medical management. With their complication rates from procedures being higher, elderly patients are often managed conservatively.

6.6.2 Relationship between troponin positivity, age and mortality

The hazard ratio for a positive troponin attenuated markedly with age, from 10.6 in the 18-29 year old age category to 1.5 in the 90+ group. The UK Myocardial Ischaemia National Audit Project (MINAP) registry found a raised troponin to be associated with an increased mortality in patients with ACS, including the elderly.[104]

Our study of all-comers, who had a troponin measured, showed that a raised troponin is additive to age in predicting mortality even in the very elderly and showed that young and middle age patients had much higher hazard ratios, particularly the very young. However, because of the high baseline risk in older patients, the absolute increment in 3-year mortality associated with a positive troponin test was remarkably consistent across all ages, at ~15 percentage points. This reflects the fact that the decline in hazard ratio with age is accompanied, not by a waning of importance of troponin levels, but by a rising background mortality in troponin negative patients with age. The absolute estimated 3-year mortality in patients over the age of 80 with a positive troponin was 46%. Almost all of the mortality increment occurred within the first six months (**Figure 6.5B-D**). This pattern of early excess mortality in hospitalised patients with an elevated troponin has also been observed in a 25000 patient study from an American hospital.[105]

Two conclusions can be drawn from this. First, regardless of age, a positive troponin is important and should not be dismissed lightly. Second, since the excess mortality occurs early, it may not be appropriate to follow a conservative wait-and-see strategy.

6.6.3 Distortion of troponin-mortality link

The extent of the association between troponin level and mortality in ACS patients has not been explored using troponin level on a continuous scale. Previous studies had 35 to 170-fold smaller sample sizes.[94,100,106,107] This limited studies to grouping together all patients with troponin values above about 5 xULN or >2.5% of the maximum detected troponin level. They showed a relatively linear association between troponin level and mortality risk. The present study is large enough to be able to analyse high levels of troponin

without grouping, across patient population who had troponin measured for a clinical reason.

There was a rise in mortality from very low troponins to middle-range positive troponins, but above ~70 xULN there was a significant and progressive decline in mortality in patients who were diagnosed with an ACS. Patients with the very highest troponins had similar mortality rates to those with normal troponins after the first month. It is possible that the paradoxical decline in mortality at very high troponins in ACS patients may be driven by a changing case mix as troponin levels increase. Furthermore, the proportion of patients undergoing invasive management rose with higher troponin levels. In keeping with this, when patients with ACS were stratified according to whether or not they underwent invasive management, the inverted U-shaped relationship between troponin level and mortality became a direct positive relationship. Therefore, as patients who underwent invasive management had lower mortality than those who had non-invasive management, and this influence was larger than the statistical effect of higher troponin levels on mortality alone, near the top of the troponin scale, where more patients had invasive management, higher values were associated with lower mortality.

Overall, hazard ratios were lower in patients managed invasively than non-invasively across all troponin levels in ACS patients. Angiography, followed by PCI or CABG, when appropriate, improves clinical outcomes, including mortality in ACS[108] and the findings from the present study may reflect a substantial improvement in prognosis secondary to revascularisation in appropriately selected patients.

6.6.4 Non-ACS patients

This study presents troponin data on the largest cohort of patients who did not have an eventual diagnosis of an ACS. Due to their smaller sample sizes, previous studies have been unable to categorise the upper troponin levels more finely and have reported a progressive association in the lower troponin bands.[100,101,109] Troponin elevation in non-ACS patients may be a marker of diminished organ-level reserve due to global comorbidity rather than suggestion of unstable coronary disease. Consistent with this interpretation, Campbell and colleagues reported the cause of death in such patients to be acute myocardial infarction from obstructive coronary artery disease in only 3% of patients.[101]

6.6.5 Limitations

This study had a retrospective design with data extracted from electronic healthcare records and subject to the limitations of this approach with difficulty in accounting for all potential confounding factors, including clinical factors associated with troponin measurement, such as patient symptoms or ECG findings, and medication treatment.

The timing of troponin testing was determined clinically, rather than standardised to a trial protocol. While troponin is commonly checked in patients suspected of ACS, data on whether patients had cardiac symptoms, such as chest pain, were unavailable. Although data on the reason for troponin measurement for individual patients was not available, the study includes an important population as often in modern medical practice, decisions are made based on results of tests initiated by others. This study helps to inform a clinician in the real-life position of receiving a troponin result.

There was no information available on whether there was differential receipt of evidence-based cardiac care across the age-groups. Finally, all-cause mortality was reported because this is incontrovertible. It was unknown whether the cause of death was related to cardiac pathology.

6.7 Conclusions

The TROP-RISK study is the first analysis using data from the contributing BRCs in the NIHR HIC and presents data from over a quarter of a million patients in routine practice who undergo troponin measurement for all clinical reasons. In young patients, a positive troponin was associated with a high mortality risk. Elderly patients who had a troponin measured had a very high mortality when the troponin is negative, and even higher when positive. A positive troponin is therefore clinically meaningful in all age groups, with the excess mortality with a positive troponin being concentrated in the first 3 months.

Assembling sufficiently large sets of real-world healthcare data can inform us on new patterns of disease that are difficult to define in single centre studies or multicentre trials. Our analysis uncovered an unexpected inverted U-shaped relationship between troponin and mortality. This may be driven by the changing case mix of patients as troponin level increases, with a higher proportion undergoing invasive management.

7. Late translational cardiovascular research - epidemiology of cardiovascular disease in clinically relevant sub-groups

7.1 TROP-AF study

In this Chapter, I further demonstrate how similar epidemiological studies can be performed in clinically relevant subgroups. I investigate the relationship between troponin level, coronary angiography, and all-cause mortality in patients presenting to hospital with atrial fibrillation.

The rationale for answering this particular research question is summarised in **Table 7.1**.

Table 7.1 - Rationale for TROP-AF study

PICO Element	Description	Rationale
Population/ Patient/ Problem	Patients with atrial fibrillation in whom troponin testing has been performed	The NIHR HIC dataset is troponin centric, as all patients were only included in the dataset if they had a troponin measured. A troponin test is commonly performed by accident and emergency in patients presenting with AF. Whilst these patients are captured in the dataset, it is possible that I am not capturing some patients with AF who did not undergo troponin testing.
Intervention	Angiography	Invasive coronary angiography is the gold standard investigation for diagnosing coronary artery disease. This may be followed by coronary revascularisation in the form of PCI or CABG.
Comparison	No angiography	Data on whether or not a patient underwent coronary angiography with or without revascularisation was available for all patients.
Outcome	All-cause mortality	All-cause mortality is the outcome of interest to most patients. Furthermore, all-cause mortality is the only outcome where the diagnosis is free from bias.

7.2 Abstract

Background

Patients presenting to hospital with atrial fibrillation (AF), usually uncontrolled, often undergo troponin measurement, but interpretation of the result is often impeded by uncertainty of its clinical importance. The relationship between troponin level, coronary angiography and all-cause mortality was investigated in patients presenting to hospital with AF across multiple centres.

Methods and Results

The National Institute of Health Research HIC dataset was used to identify patients admitted to 5 centres in the UK, between 2010 and 2017, who had a primary diagnosis of AF. The peak troponin result (highest measured) was scaled as a multiple of the upper limit of normal (xULN) for the troponin assays used across the site.

3,121 patients were included in the study. The median duration of follow-up was 1462 (IQR 929 to 1975) days, during which there were 586 (18.8%) deaths. The multivariable mortality adjusted hazard ratio associated with a positive troponin (value above ULN) was 1.20 (95% CI 1.01 to 1.43, $p<0.05$). Higher troponin values were associated with a higher risk of mortality. A maximum hazard ratio of 2.6 (95% CI 1.9 to 3.4) was reached at a troponin level of approximately 250 xULN.

The relationship between troponin levels and odds of coronary angiography was exponential. The mortality risk was 36% lower in patients who underwent coronary angiography compared to patients who did not (adjusted hazard ratio 0.61, 95% CI 0.42 to 0.89, $p=0.01$).

Conclusion

Raised troponin was associated with increased mortality risk in patients presenting to hospital with AF. Invasive management was associated with a lower hazard ratio, raising the possibility that the troponin release in AF may be related to CAD, which can be managed with coronary revascularisation.

7.3 Introduction

Patients admitted to hospital with atrial fibrillation (AF), the most common amongst tachyarrhythmias[110], often have a troponin mesured,[111,112] usually to assist with the diagnosis of acute coronary syndrome (ACS) manifesting as AF.[113] However, there is usually uncertainty over the clinical importance of troponin values in the context of AF. Both the diagnostic and prognostic value of troponin measurement in ACS[106] and heart failure[114] is established. However, the prognostic value of troponin in the context of AF is not well understood, where troponin elevation may be related to a rapid ventricular rate and the mechanical effects of fibrillation on the atria rather than due to underlying coronary artery disease (CAD). Previous observational analysis of patients presenting to hospital with AF have shown an association between the troponin value and clinical outcomes, however, these studies did not assess mortality. In clinical practice, relatively small troponin elevations in patients with AF are usually ignored and do not routinely result in further investigations for CAD.

In this study, the relationship between troponin level and all-cause mortality was explored in patients admitted to hospital with AF. There was alos an assessment of the pattern of referral for coronary angiography in relation to troponin level, and whether this had an impact on mortality in these patients.

7.4 Methods

7.4.1 Study design and participants

The NIHR HIC database was used to identify patients admitted to hospital with a primary diagnosis of AF who underwent at least one troponin measurement. Patients were excluded if they had a concomitant secondary diagnosis of AF. Diagnoses were established from International Statistical Classification of Diseases (ICD-10) discharge codes.[7] Eligible patients were followed up using routinely collected data, until death or censoring in April 2017.

All analyses on troponin were performed using the peak (highest) troponin level recorded for the patient. If only a single troponin was tested, the peak troponin was based on this measurement. Troponin data was treated in two ways. First, the results were dichotomised as being either positive or negative. Second, troponin was treated on a continuous scale by standardising the troponin results which were measured using different troponin assays. Scaling was performed by calculating the ratio of the observed troponin result divided by the ULN for the relevant troponin assay.

To account for outpatient procedures, patients were categorised as having angiography or revascularisation (percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG)) if performed within 3 months of the peak troponin level.

In this retrospective cohort study, all patients were followed up until death or censoring in April 2017. Mortality was ascertained using the NHS Spine Application.

7.4.2 Statistical Analysis

Patients were grouped in to those who did or did not undergo angiography. Baseline characteristics were compared between the two groups by Mann-Whitney U test or Chi-square test for continuous and categorical data, respectively. Cox regression modelling was used to assess the relationship between troponin and all-cause mortality. Continuous troponin values were log transformed because of the positive skew of values.

Using Martingale residuals, non-linearity was detected when assessing the relationship between the log hazard and all continuous variables (age, creatinine level, haemoglobin level, platelet count, white cell count and troponin level). To model non-linear relationships, restricted cubic splines were used, used four knots. Splines were adjusted for patient demographics, blood test results, cardiovascular risk factors and other comorbidities. Subgroup analyses were performed in angiography and no angiography subgroups. Kaplan-Meier curves were plotted according to whether or not patients underwent angiography.

Statistical analyses were performed using R 3.5.0 (the R Core Team, Austria).

7.5 Results

7.5.1 Summary of main findings

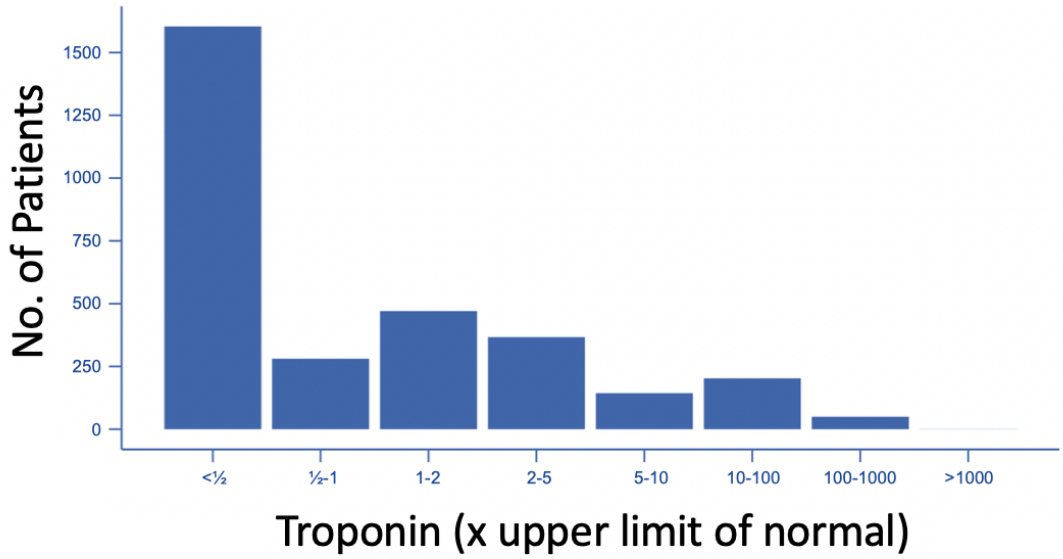
A total of 3,121 patients were admitted across the five sites with a primary diagnosis of AF and a troponin measurement. The baseline characteristics of the patients are shown in **Table 7.2**. The mean age was 73 years (95% confidence interval (CI): 62-82) and 55.7% of patients were male. The majority of these patients' (60.4%) had a peak troponin result that was normal (<1 xULN) (**Figure 7.1**).

Table 7.2 - Baseline characteristics of patients in TROP-AF study

	Patients with primary presentation of AF (n=3121)
Demographic characteristics	
Age (year)	73 (62 – 82)
Male	1738 (55.7)
Haematology and biochemistry results	
C-reactive protein (mg/dL) (n=2796)	5.0 (2.0 – 14.9)
Creatinine (µmol/L) (n=3086)	82 (69 – 100)
Haemoglobin (g/dL) (n=3075)	13.8 (12.5 – 15.0)
Platelet count (x10^9/L) (n=3071)	226 (187 – 274)
Troponin (xULN)	0.5 (0.003 – 2.0)
White cell count (x10^9/L) (n=3075)	8.2 (6.6 – 10.2)
Cardiovascular risk factors	
Diabetes mellitus	355 (11.4)
Hypercholesterolemia	448 (14.4)
Hypertension	1062 (34.0)
Cardiovascular disease	
Aortic stenosis	53 (1.7)
Heart failure	302 (9.7)
Previous myocardial infarction	341 (10.9)
Other comorbidities	
Malignancy	207 (6.6)
Obstructive lung disease	146 (4.7)

Figures represent median (interquartile range) or value (percentage). Numbers in parentheses (n=) indicates the number of patients who had data available for the relevant variable. ULN, 99th percentile of the upper limit

Figure 7.1 - Bar chart of numbers of patients according to troponin level



ULN, 99th percentile of the upper limit of normal.

7.5.2 Relationship between troponin level and coronary angiography

216 (6.9%) patients underwent coronary angiography, of which 78 (36.1%) subsequently underwent coronary revascularisation (PCI (93.6%), CABG (2.6%) or both (3.8%)). 39 (1.2%) of patients had a secondary diagnosis of ACS. The majority of coronary angiograms (89.8%) and 43.6% of revascularisation procedures occurred within 3 days of the peak troponin level. Baseline characteristics are displayed in **Table 7.3**. The patient demographic and clinical factors that were associated with having a coronary angiography are shown in **Figure 7.2** and **Table 7.4**.

Table 7.3 - Baseline characteristics of patients who did and did not undergo angiography

	Angiography (n=216)	No angiography (n=2905)	‡P-value
Demographic characteristics			
Age (year)	73.5 (65.3 – 79.0)	73.0 (63.0 – 83.0)	0.47
Male	144 (66.7)	1594 (54.9)	0.001
Haematology and biochemistry results			
C-reactive protein (mg/dL)	6.1 (2.03 – 16.5)	5.0 (1.9 – 14.8)	0.08
Creatinine (µmol/L)	84.0 (73.3 – 100.8)	81.0 (69.0 – 100.0)	0.04
Haemoglobin (g/dL)	13.9 (12.5 – 15.0)	13.8 (12.5 – 15.0)	0.78
Platelet count (x10^9/L)	224 (182 – 270)	226 (187 – 275)	0.53
Troponin (xULN)	1.4 (0.003 – 5.6)	0.5 (0.003 – 2.0)	<0.0001
White cell count (x10^9/L)	8.5 (6.8 – 10.6)	8.2 (6.6 – 10.2)	0.21
Cardiovascular risk factors			
Diabetes mellitus	32 (14.8)	323 (11.1)	0.12
Hypercholesterolemia	41 (19.0)	407 (14.0)	0.06
Hypertension	74 (34.3)	988 (34.0)	0.94
Cardiovascular disease			
Aortic stenosis	8 (3.7)	45 (1.5)	0.03
Heart failure	27 (12.5)	275 (9.5)	0.15
Previous myocardial infarction	65 (30.1)	276 (9.5)	<0.0001
Other comorbidities			
Malignancy	7 (3.2)	200 (6.9)	0.03
Obstructive lung disease	12 (5.6)	134 (4.6)	0.50

Figures represent median (interquartile range) or value (percentage). Numbers in parentheses (n=) indicates the number of patients who had data available for the relevant variable. ULN, 99th percentile of the upper limit of normal. ‡ Comparison between angiography and no angiography groups using Mann-Whitney U test for continuous variables and Chi-square test for categorical variables.

Figure 7.2 - Multivariable odds ratio of undergoing coronary angiography for continuous variables

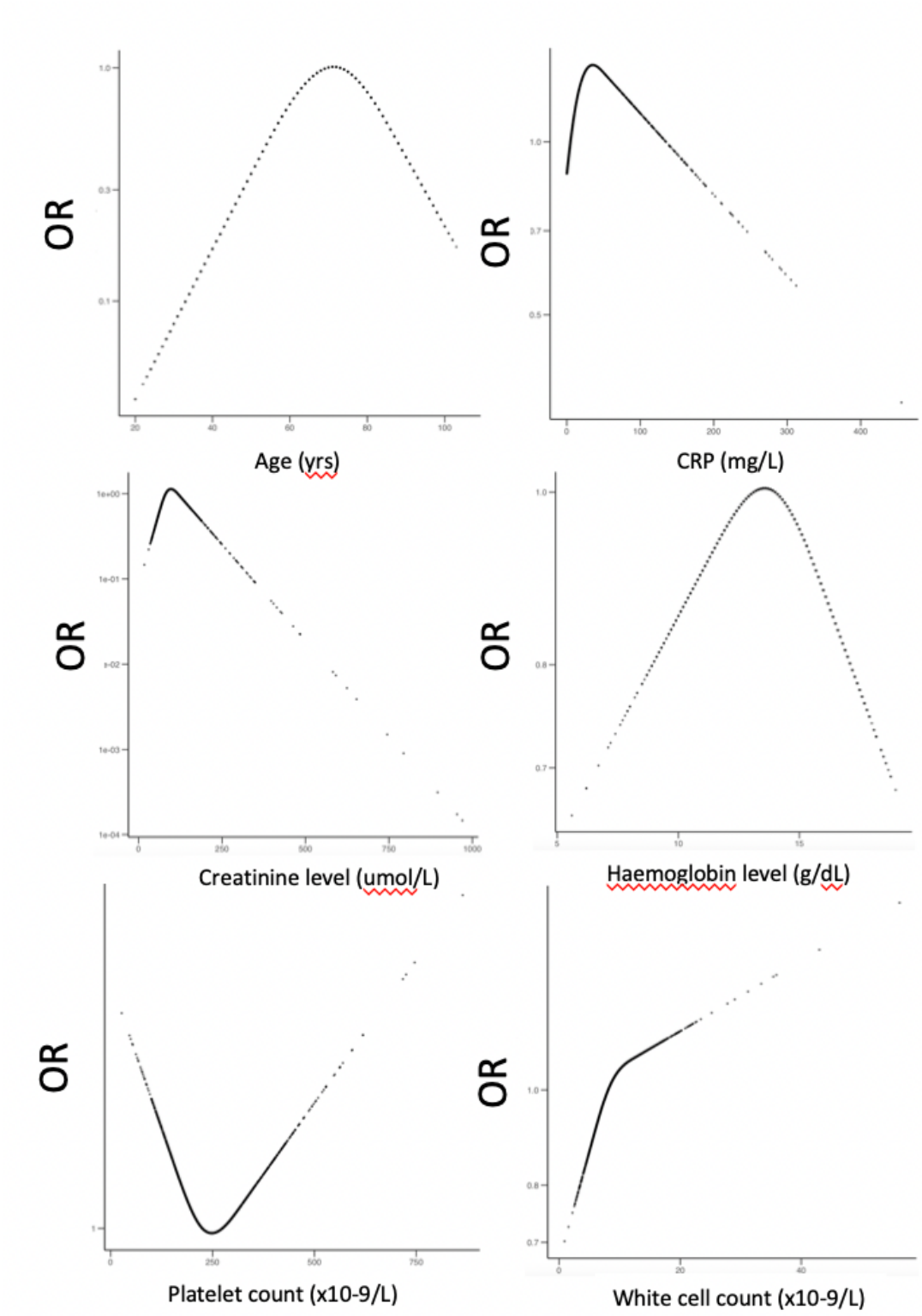


Table 7.4 - Multivariable odds ratio of undergoing coronary angiography for categorical variables

	Odds ratio (95% CI)	P-value
Male (vs female)	1.6 (1.2 – 2.1)	0.004
Diabetes mellitus	1.1 (0.7 – 1.7)	0.68
Hypercholesterolaemia	1.1 (0.8 – 1.7)	0.52
Hypertension	0.8 (0.6 – 1.2)	0.29
Aortic stenosis	2.1 (0.9 – 4.6)	0.07
Heart failure	1.1 (0.7 – 1.7)	0.67
Previous myocardial infarction	3.7 (2.6 – 5.2)	<0.0001
Malignancy	0.4 (0.2 – 0.9)	0.02
Obstructive lung disease	0.9 (0.5 – 1.6)	0.64
Positive troponin	1.5 (1.2 – 2.1)	0.003

Estimates compared to not having the disease, unless otherwise stated.

There was a non-linear relationship between troponin level and odds of undergoing coronary angiography (**Figure 7.3B**). Across normal troponin levels (<1 xULN) there was no association with increasing likelihood of coronary angiography. Above the ULN (>1 xULN), there was a direct positive, linear relationship between troponin level and odds of angiography, until 5 xULN, after which there was an exponential distribution. There was a direct positive relationship between odds ratio of revascularisation and troponin level (**Figure 7.4**).

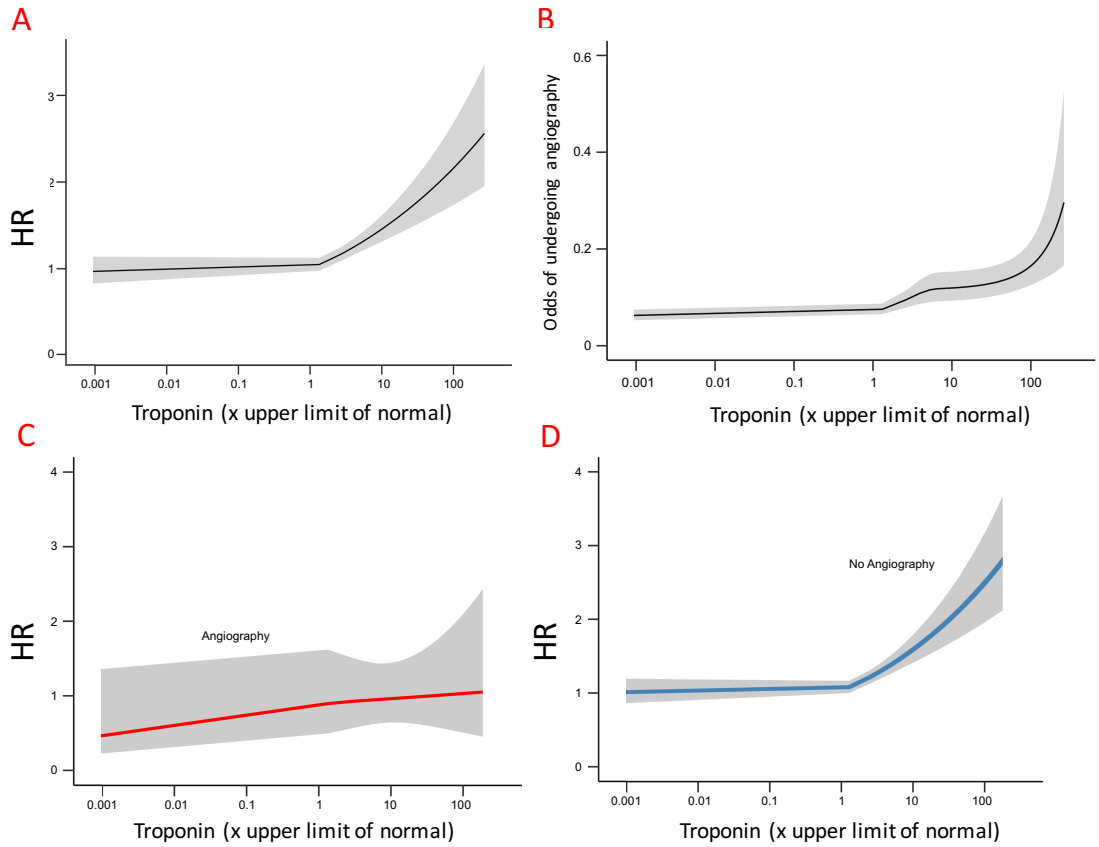
7.5.3 Relationship between troponin level and mortality

There were a total of 586 (18.8%) deaths over a median follow-up of 48.1 months (IQR 30.5 to 64.9). The adjusted mortality hazard ratio associated with a positive troponin was 1.20 (95% CI 1.01 to 1.43, $p<0.05$). The relationship between continuous troponin level and mortality is shown in **Figure 7.3A**. While there was no significant relationship between troponin level and mortality for troponin levels below 1.3 xULN, at higher troponin levels, there was a positive relationship.

Figures 7.3C and 7.3D display the relationship between troponin level and mortality in patients with AF who underwent coronary angiography (**Figure 7.3C**) and patients who did not (**Figure 7.3D**). In patients who underwent angiography, there was no significant relationship between troponin level and mortality (**Figure 7.3C**). Alternatively, in those patients who did not undergo angiography, there was a significant relationship with troponin levels above the

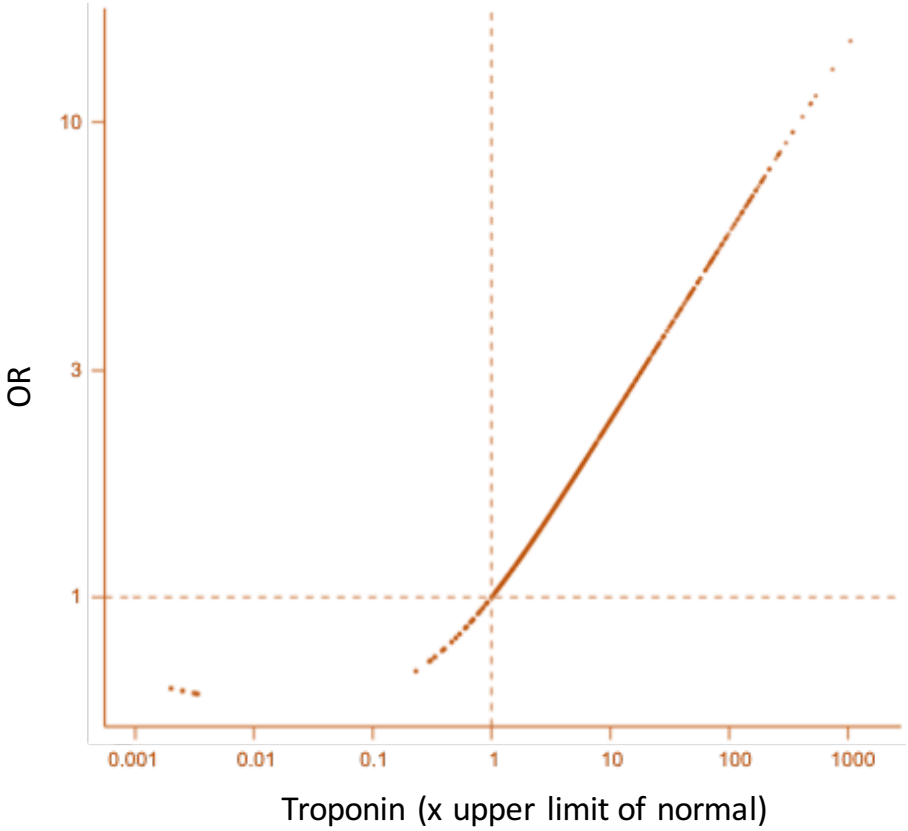
ULN being associated with mortality (**Figure 7.3D**).

Figure 7.3 - Multivariate* restricted cubic spline modelling of (A) association between troponin level and hazard ratio; (B) association between troponin level and odds of coronary angiography; association between troponin level and hazard ratio in angiography (C) and no angiography (D) subgroups



**adjustment for age, gender, C-reactive protein, creatinine, haemoglobin, platelet count, white cell count, diabetes mellitus, hypercholesterolaemia, hypertension, aortic stenosis, heart failure, previous myocardial infarction, malignancy and obstructive lung disease. The shaded area denotes the 95% confidence interval. ULN, 99th percentile of the upper limit of normal.*

Figure 7.4 - Odds ratio of undergoing coronary revascularisation according to troponin level

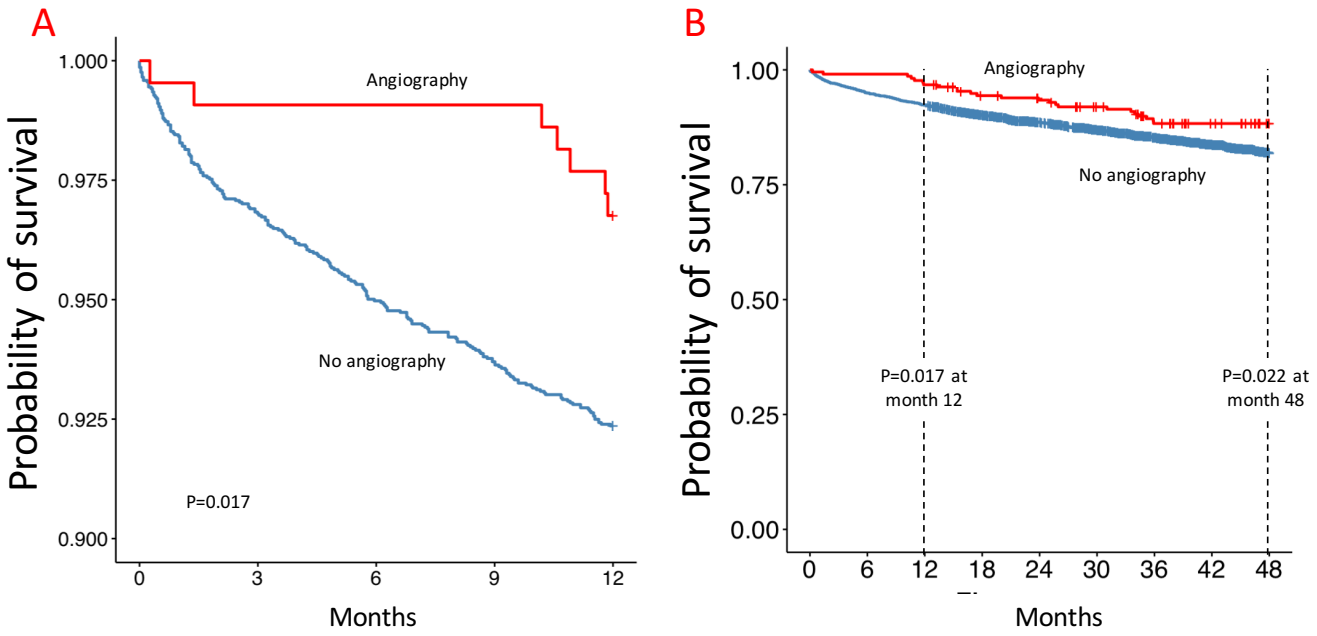


The figure shows the odds ratios of undergoing coronary revascularisation according to troponin level, where the comparator troponin level value is of 1 xULN, which is marked with the red dotted lines. ULN, 99th percentile of the upper limit of normal.

Kaplan-Meier analysis showed survival was lower in patients who did not undergo angiography, in the short-term (p=0.02, **Figure 7.5A**), and in the long-term at 4 year follow-up (p=0.02, **Figure 7.5B**). On multivariable Cox regression analysis, undergoing angiography was associated with a 39% lower mortality (hazard ratio 0.61, 95% CI 0.42 to 0.89, p=0.01). For patients who had a coronary angiogram, there was a non-significant trend towards lower mortality in patients who underwent coronary revascularisation (hazard ratio 0.36, 95% CI 0.12 to 1.10, p=0.07).

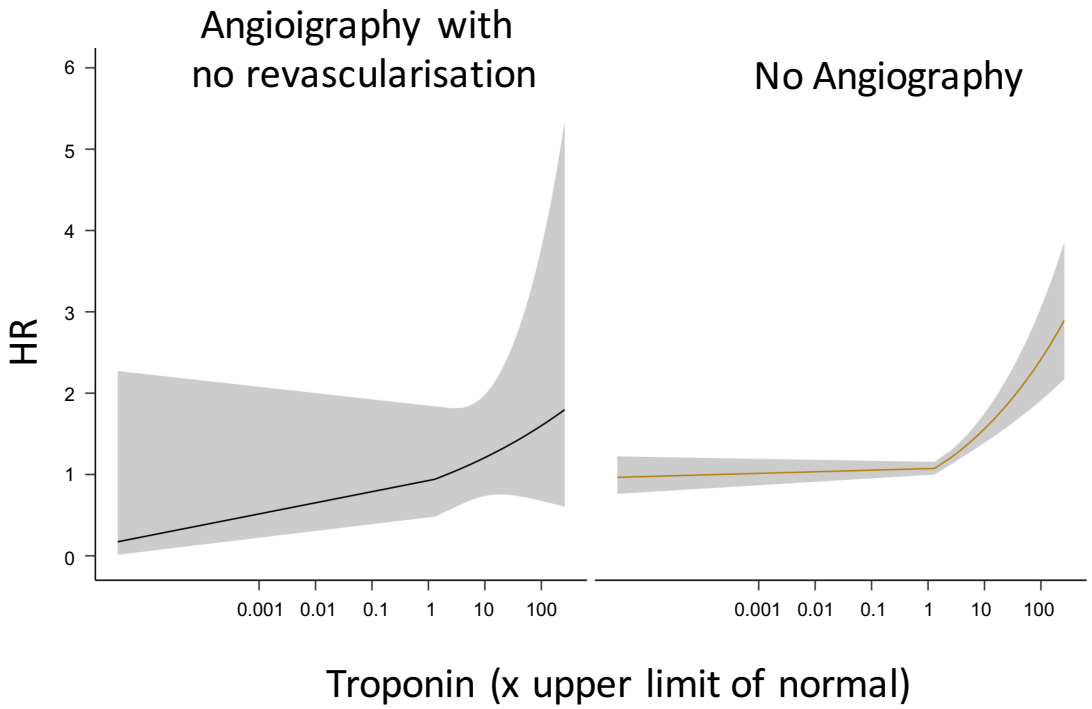
Figure 7.6 shows that a rising troponin level was associated with a higher mortality risk in both the angiography (with no revascularisation) and no angiography groups, although the confidence intervals in the former group were relatively wide.

Figure 7.5 - Kaplan-Meier survival curves according to angiography status over (A) 12 months and (B) 48 months follow-up



Tick marks denote censored events. Survival curves compared using log-rank statistic

Figure 7.6 - Multivariate* restricted cubic spline modelling of association between troponin level and hazard ratio for patients who underwent angiography without revascularisation (left) and those who did not undergo angiography (right)



**adjustment for age, gender, C-reactive protein, creatinine, haemoglobin, platelet count, white cell count, diabetes mellitus, hypercholesterolaemia, hypertension, aortic stenosis, heart failure, previous myocardial infarction, malignancy and obstructive lung disease. The shaded area denotes the 95% confidence interval. 99th percentile of the upper limit of normal.*

7.6 Discussion

7.6.1 Summary of main findings

This study investigated the association between troponin level, coronary angiography and mortality in patients presenting to hospital with AF. In 3,121 patients with AF, there was a significant positive non-linear relationship between troponin levels above the ULN and mortality after adjustment for key patient demographic and clinical factors. Interestingly, even low level troponin levels that were above the ULN were associated with a higher mortality risk.

7.6.2 Non-linear relationship between troponin and mortality

Following an inflection near ULN, there was a direct positive relationship between troponin level and mortality in all patients. This inflection point supports the use of the 99th percentile in determining the 'normal' range of troponin values. Although this does not mean all patients presenting with atrial fibrillation who have a troponin level >1 xULN should undergo a coronary angiogram, it suggests that threshold for the troponin level beforer investigating for coronary artery disease may be lower than in current practice.

7.6.3 Coronary angiography is performed at higher troponin levels

A coronary angiogram was performed in less than 7% of patients who presented to hospital with AF who had a troponin measured. Within the normal range for troponin (< 1x ULN), higher troponin levels were not associated with an increased odds of having a coronary angiogram. Above 1x ULN, there was a positive, exponential relationship between troponin level and odds of undergoing a coronary angiogram. Despite this, even at the highest troponin levels, less than 50% of patients had an angiogram. This finding is consistent with routine clinical practice; small troponin elevations in patients presenting to hospital with AF are not usually considered by clinicians to be a useful marker of CAD. Additionally, even at very high troponin values, alternative explanations are often invoked.

7.6.4 Coronary angiography alters the troponin-mortality relationship

In patients who had a coronary angiogram, troponin levels above 1 xULN were associated with higher mortality, however, this relationship was weak with and not statistically significant. In patients who were managed non-invasively, without a coronary angiogram, there was a direct, positive relationship between higher troponin levels (> 1 xULN) and higher mortality risk. Coronary angiography was associated with a 36% lower mortality across the full spectrum of troponin values. One possibility is that we may be referring a relatively low risk group of patients for coronary angiography compared to those treated medically. A coronary angiogram performed in patients with higher troponin levels is more likely to reveal physiologically important CAD that is amenable to revascularisation. Coronary revascularisation was performed in 36.1% of patients who had an angiogram and there was a trend towards a lower mortality with revascularisation, although not statistically significance. A clinical trial is necessary to investigate the utility of routinely investigating for CAD in patients with AF who have a raised troponin.

7.6.5 Relationship with existing evidence

The present study is the largest observational study to assess the relationship between troponin and mortality in patients admitted to hospital with a primary diagnosis of AF. Previous studies have investigated the association between troponin level and outcome such revascularisation, stroke and cardiovascular death.[112,115] However, all these outcomes risk inaccuracy and bias to varying extents. All-cause mortality is the only outcome which is not subject to such bias.

Conti et al. prospectively enrolled 3,627 patients who presented with AF within 48 hours, and offered coronary angiography if the troponin level was elevated above the ULN.[113] Positive troponin values (>1 xULN) was associated with the presence of CAD with coronary angiography, revascularisation and increased likelihood of adverse cardiovascular events but mortality was not measured. Furthermore, troponin was dichotomised as being either positive or negative based on the ULN. The present study has shown that higher troponin levels are associated with a higher risk of death, indicating that the degree of troponin elevation as a continuous variable is important to account for when make clinical decisions on management. Supporting this finding, in a retrospective study of 200 patients, troponin dichotomised as being positive or negative (at the ULN) was poor at predicting the presence of CAD, whereas the actual peak troponin level, on a continuous scale, did predict CAD.[116]

7.6.6 Mechanisms of troponin release in AF

Troponin T and I bind to tropomyosin in the intracellular sarcomeric contraction complex.[117] During an ACS event, obstruction of the coronary arteries results in myocyte necrosis, which results in the release of troponin into the circulation, with larger infarcts releasing more troponin in to the bloodstream and associated with a higher risk of death.[106] However, an abnormal troponin result may confer a worse prognosis in several other settings, such as heart failure, pulmonary embolism or sepsis.[114]

There is a dispute regarding the cause of circulating troponin in AF, which is one of the reasons why clinicians usually have a relatively low concern related to troponin elevation in patients with AF.[117] One mechanism disputed is a type 2 myocardial infarction diagnosis, where there is ventricular myocyte death (and troponin elevation) during rapidly conducted AF in those patients with pre-existing CAD. While rate control can reduce troponin leak[118],[119] troponin release has also been shown to occur at normal ventricular rates. Furthermore, in patients who do have CAD, it is not always physiologically

significant, which makes treatment decisions on the role of revascularisation, uncertain.

The study was unable to elucidate a mechanism for troponin release nor its association with a worse prognosis, however, the findings raise the possibility that the troponin release in AF may mediated by CAD.

7.6.7 Limitations

Due to the retrospective nature of this study, there was difficulty ini accounting for all potential confounding factors. There may be inaccuracies with the data extracted from electronic health records, which may introduce bias. For example, in clinical practice, larger troponin rises may be more likely to prompt a primary diagnosis of ACS on discharge, rather than AF, even if their presentation was with AF. Furthermore, the results cannot be generalised to all patients with AF, as it only included those who had a troponin measurement, which potentially alters the overall risk profile. Furthermore, data on the sub-type of AF, such as paroxysmal or new-onset AF, was not available, which prevented analysis by chronicity of AF. Furthermore, the role of stress testing was not available in the dataset.

7.7 Conclusions

In patients presenting to hospital with AF, elevated troponin levels at any level were associated with an increased mortality risk. Coronary angiography is rarely performed in this cohort of patients, unless the troponin elevation is large, in which case, coronary angiography was associated with lower mortality. The question is raised as to whether troponin elevation in patients presenting to hospital with AF is mediated by CAD that may be responsive to coronary revascularisation.

8. Discussion

The overall Discussion of this thesis is divided into two main sections. Firstly, I will discuss the overall translational implications of the results of the studies discussed in Chapters 4-7. Secondly, I will discuss the challenges faced by with using hospital EHR for research and the future direction of the NIHR HIC.

8.1 Translational implications

8.1.1 CRP-RISK study

The data from the CRP-RISK study emphasises the need for further prospective work that explores the inflammation pathway as a therapeutic target to address the still unanswered mortality gap in patients presenting with suspected ACS. We know that recurrent major adverse CV events in those undergoing PCI[120], or presenting with ACS is still unacceptably elevated, and shown to be as high as 20% at 3-years[121] despite optimal medical therapy.

Barriers to translation to clinical practice:

In addition to the limitations stated in Section 4.6.5, one of the key barriers to translation to clinical practice is the lack of an effective therapeutic treatment to reduce the mortality risk in these patients with low-grade inflammation.

Facilitator to translation to clinical practice:

The results from this study have the potential to influence future trial design. There is a current body of work, led by Dr Ramzi Khamis, which involves designing a trial which uses hsCRP to select patients at high cardiovascular risk, and randomising them to a novel targeted treatment.

8.1.2 SENIOR-NSTEMI study

The SENIOR-NSTEMI study strengthens the evidence for an invasive approach to managing elderly patients who have had an NSTEMI. Ideally, decision making for clinicians should be driven by the results from RCTs. However, the results of SENIOR-RITA, the ongoing RCT designed to answer this question, are not set to be reported until at least 2029.[85] In the meantime,

clinical decision making needs to be made based on the best available evidence. The SENIOR-NSTEMI study results suggest that invasive management of elderly patients may be the better strategy in those who could be managed either invasively or non-invasively (conservatively).

Barriers to translation to clinical practice:

One of the concerns regarding this study is confounding by indication, where ACS patients with worse prognosis were more likely to receive non-invasive management. Attempts were made to control for this by adjusting results by over 70 confounding factors. I also excluded ACS patients in propensity score strata that had little to no deaths in either the invasive or non-invasive group. Further work is required to determine if this methodology leads to similar effect estimates when attempting to emulate trials that have already been reported.

Facilitator to translation to clinical practice:

The results of this study were shared with the care of the elderly leads at Imperial College Healthcare NHS Trust and a protocol to promote invasive management for the management of elderly patients with NSTEMI is being devised as a collaboration between the cardiology and the elderly medicine departments. The protocol will be implemented in the form of a quality improvement project.

The paper has been cited in the Up-To-Date clinical guidelines, suggesting that that an invasive approach may be the better strategy in elderly patients who could be managed either invasively or conservatively.

As the ‘gold standard’ for evaluating the effectiveness of interventions, we eagerly await the results of the SENIOR-RITA trial, which should facilitate the translation of my findings to clinical practice, should the direction of effect remain consistent.

8.1.3 TROP-RISK study and TROP-AF study

In the TROP-RISK and TROP-AF studies, regardless of age, a raised troponin level was an important prognostic factor. Furthermore, as the excess mortality occurs very early, a wait and see strategy may not be appropriate in these patients. Another key finding from both the TROP-RISK and TROP-AF studies is the relatively high mortality risk associated with troponin elevation in

patients without ACS. In these non-ACS patients, there was a mortality benefit in patients who underwent invasive management.

Barriers to translation to clinical practice:

One of the main findings in this study which has scope to be translated to clinical practice is the potential for invasive management to reduce mortality risk in non-ACS patients with raised troponin values. In clinical practice, there is still no consensus on the best approach to investigate non-ACS patients with elevated troponins. Whilst I only considered invasive management as a therapeutic strategy, other less invasive interventions should also be considered, such as medications and non-invasive cardiovascular imaging. There is currently no large multi-centre trial which has investigated this hypothesis which is a barrier to translation of my clinical findings to clinical practice.

Facilitator to translation to clinical practice:

Clinical trials are warranted to clarify the role of investigating and treating CAD in non-ACS patients with an elevated troponin who had a troponin measured for any clinical reason. Patients will be randomised to either usual practice or a combination of primary prevention measures (aspirin 75mg once daily, atorvastatin 80mg once daily[122]) and a cardiology review.

Whilst we await the results of such a trial, I am currently working with the EHR team at Imperial College Healthcare NHS Trust to create a flag on the patient records which highlights the mortality risk associated with a particular troponin result in both ACS and non-ACS patients. This will help provide clinicians with insight on the high mortality risk in patients with elevated troponin levels, especially in non-ACS patients. This may lead to a behavioural change in the way clinicians interpret the results of elevated troponin levels, especially in patients with a non-ACS diagnosis. This may be initiated as part of a multicentre trial, where a site either has this EHR flag turned on or off, with the primary outcome being mortality.

8.2 Challenges and limitations

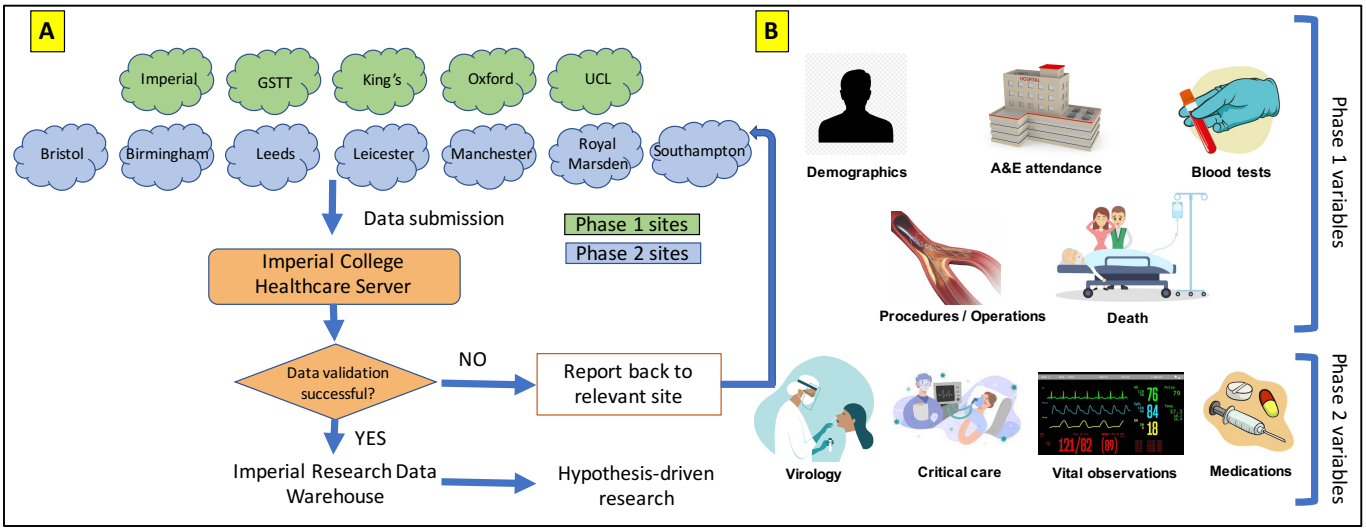
While this thesis has demonstrated that the NIHR HIC infrastructure can be used to aggregate longitudinal EHR data across multiple centres to create datasets for research, challenges still remain.

8.2.1 Population coverage

The exemplar NIHR HIC cardiovascular dataset captured 257,948 patients across five NHS Trusts in England (Phase 1 sites; **Figure 8.1**). The broad demographic of patients in this cohort (**Table 6.3**) allowed stratification of outcomes among different demographic groups. For example, in the TROP-RISK Study, the association between troponin level and mortality was evaluated in different age groups. Furthermore, there was a large enough cohort of elderly patients with NSTEMI in the dataset to assess the effect of invasive management in this under investigated group of patients. Further broadening the demographic case mix of patients in the NIHR HIC cardiovascular cohort will further improve the generalisability of the findings to different age groups, genders and ethnicities.

Future direction:
Data collection for the NIHR HIC Cardiovascular Theme is a continuous process, and new NHS Trusts (Phase 2 sites; **Figure 8.1**) have signed the NIHR HIC data sharing agreement and will contribute data towards future NIHR HIC datasets.

Figure 8.1 - NIHR Health Informatics Collaborative data flow



GSTT, Guy's and St Thomas' NHS Foundation Trust; UCL, University College London Hospitals NHS Foundation Trust

8.2.2 Cardiovascular disease coverage

As the data are primarily collected for clinical care, there will be

differences in clinical practise between clinicians and clinical sites. For example, there may be differences in departmental protocols on the panel of blood tests requested for a patient presenting to hospital with chest pain. These issues can lead to heterogeneous patterns of missing data between the submitting sites. With these challenges in mind, it is important that future data models and research studies generate views of the data associated with the lowest risk of selection bias. Ethical restrictions prevent data from being extracted from 'all' patients in the EHR. It is important that the selection criteria for patient inclusion in the data extracts are well defined. The selection criteria for the first dataset included patients who had a troponin requested in order to capture those with suspected ACS.

Future direction:
For the second cardiovascular dataset, in an attempt to capture more patients with suspected and/or confirmed cardiovascular disease, the entry criteria for the dataset will be all patients who had a troponin, brain natriuretic peptide (BNP) or echocardiogram requested or were under the care of the cardiology team as an inpatient or outpatient.

8.2.3 Structured data

This thesis involved extraction of structured data, which are stored in defined fields within a patient's EHR. **Figure 8.1** summarises the broad headings for the data variables used for the first cardiovascular dataset as part of this thesis (Phase 1 variables). Future datasets will involve extraction of even more granular level data (Figure 1, Phase 2 variables). This will broaden the scope of the research questions that can be answered using the big data extracts from the NIHR HIC. For example, inclusion of data on medications allows assessment of the effect of different medications on outcomes in cohorts of patients where RCT evidence is lacking (discussed further in **Section 8.3.1**).

Future direction:
For the second cardiovascular dataset, the NIHR HIC infrastructure will be developed to allow capture of additional fields of structured data.

8.2.4 Unstructured data

Extraction of even more granular level data can address some of the limitations of the studies presented in this thesis. For example, I had difficulty in accounting for all potential confounding factors, including clinical factors prompting troponin measurement, such as patient symptoms, echocardiography or electrocardiogram findings, and medication treatment.

This thesis has established the methods for extracting 'structured' healthcare record data into a tabular format for big data analysis. Useful clinical data may however be stored in unstructured free text which may more difficult to extract and use for statistical analyses compared to extracting and using discrete, structured data. For example, clinical notes and many imaging reports remain embedded in unstructured free text.

8.2.4.1 Natural language processing and artificial intelligence

As patient identifiers may be contained within unstructured text, they cannot be transmitted for processing at the central site.[123] As a result, there was limited scope to use natural language processing for the current dataset.

Similarly, images may have patient details embedded in them, with the risk of breaching patient confidentiality, should they be transmitted for research purposes. Having access to echocardiography images, as well as images from other cardiovascular imaging modalities, linked with structured data from the HIC, will improve the potential to use artificial intelligence and image processing techniques for prognostic studies.

There is a dedicated workstream on processing unstructured clinical data:

- Extraction of relevant clinical information from unstructured clinical notes and imaging reports using natural language processing
- Removal of patient identifiers from cardiac images

There is a synergistic relationship between data and artificial intelligence, where artificial intelligence is useless without data, and data leverages artificial intelligence for augmented predictive analytics.[124] Future projects are currently in development which utilise these more advanced analytic methods.

Future direction:
There is a workstream to develop methods for text anonymisation and data extraction from unstructured text and cardiovascular images.

8.3 Future work

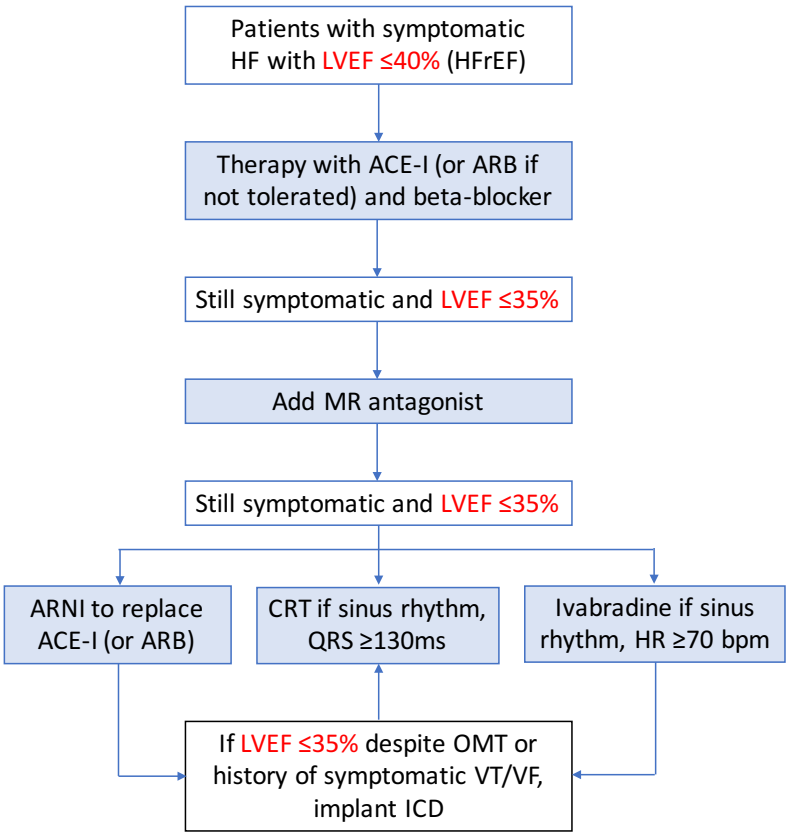
8.3.1 Emulation of clinical trials

As discussed earlier in this Chapter, the next Cardiovascular dataset will capture a wider cohort of patients with cardiovascular disease. The selection criteria will include patients with suspected heart failure (HF), characterized by the request of an echocardiogram or a BNP blood test. Both an echocardiogram and BNP are useful diagnostic methods for patients with possible HF.[123]

Left ventricular ejection fraction (LVEF) is one of the most commonly reported single measures of left ventricular (LV) systolic function and is used to distinguish between HF with reduced ejection fraction (EF) (HFrEF; LVEF ≤ 40%), HF with mid-range EF (HFrEF; LVEF 40-50%), and HF with preserved EF (HFpEF; LVEF ≥ 50%).[125-127] HFpEF and HFrEF seem to have different aetiological and epidemiological profiles. Patients with HFpEF are older, more often female, and more commonly have a history of atrial fibrillation or hypertension, while a history of myocardial infarction is less common.[128,129] The characteristics of patients with HRmrEF seem to be between those with HFrEF and HFpEF but further studies are required to characterise this population in more detail.[130]

With evidence from several randomised controlled trials (RCTs), clinical practice guidelines articulate clear strategies for management of patients with HFrEF, including beta-blocker (BB), mineralocorticoid receptor (MR) antagonist, angiotensin-converting enzyme inhibitor (ACEI) or angiotensin receptor blocker (ARB) therapy (**Figure 8.2**).[125-127] Treatments for patients with HFrEF have shown no consistent mortality benefit in patients with HFpEF.[125,131] In addition to addressing any comorbidities, symptomatic control with diuretic therapy is the main treatment approach recommended in HFpEF.[125]

Figure 8.2 - Therapeutic algorithm for patient with symptomatic heart failure with reduced ejection fraction (HFrEF) based on current European and American guidelines



ACE-I, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor neprilysin inhibitor; CRT, cardiac synchronisation therapy; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; HR, heart rate; ICD, implantable cardioverter defibrillator; LVEF, left ventricular ejection fraction; MR, mineralocorticoid receptor; OMT, optimal medical therapy; VF, ventricular fibrillation; VT, ventricular tachycardia.

ESC guidelines recommend that patients with heart failure with mid-range ejection fraction (HFmrEF) (LVEF 40-50% and a diagnosis of HF) are treated in a similar way to those with HRpEF.[125] The justification was that the positive trial results in HFrEF came from patients with low LVEFs (≤40%) and the neutral results of HFpEF studies were from patients with a LVEF as low as 40%, thus including patients with HFmrEF. While there are no RCTs solely in patients with HFmrEF, stratified analyses of previous RCTs that included some patients with HFmrEF, suggest that HFmrEF may behave more like HFrEF than HFpEF (**Table 8.1**).[132-134] Clinicians sometimes use

BB, MRA, ACEI or ARB to treat HFmrEF,[135] because HFmrEF may be a milder form of HFrEF.

Table 8.1 - Treatment effect for cardiovascular mortality, all-cause mortality and heart failure hospitalisation according to HFrEF, HFmrEF and HFpEF categories in RCTs

Trial	Drug comparison	LVEF in HFrEF range				LVEF in HFmrEF range				LVEF in HFpEF ran		
		LVEF (%)	N	HR (95% CI)	P-value	LVEF (%)	N	HR (95% CI)	P-value	LVEF (%)	N	HR (95% CI)
Cardiovascular mortality												
11 IPD RCT meta-analysis[132]	Beta-blocker vs Placebo	**≤40%**	13354	0.72 (0.66–0.80)	<0.0001	**40 – 49%**	570	0.48 (0.24–0.97)	0.04	**≥50%**	241	1.77 (0.61–5.14)
TOPCAT[133]	Spironolactone vs Placebo					**45 – 49%**	197	0.46 (0.23 – 0.94)	0.03	**≥50%**	1569	0.80 (0.60 - 1.06)
All-cause mortality												
TOPCAT[133]	Spironolactone vs Placebo					**45 – 49%**	197	0.58 (0.34 – 0.99)	0.05	**≥50%**	1569	0.89 (0.69 - 1.14)
Heart failure hospitalisation												
CHARM[134]	Candesartan vs Placebo	**≤40%**	4323	0.77 (0.68 – 0.86)	<0.0001	**40 – 49%**	1322	0.72 (0.55 – 0.95)	0.02	**≥50%**	1953	0.91 (0.74 - 1.13)

HR, hazard ratio; HFmrEF, heart failure with mid-range ejection fraction; HFpEF, hear failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; IPD, individual patient data; LVEF, left ventricular ejection fraction; RCT, randomised controlled trial.

While a new, adequately designed and powered RCT is desirable to confirm the benefits of modern HF therapies in patients with HFmrEF, there are a number of limitations to potential trial evidence. In addition to RCTs being expensive, their feasibility to support clinical decision making is limited for long-term outcomes or to estimate the effect of treatment in particular patient subgroups. The availability of big data, combined with computational and analytic developments, provide an approach to emulate a target trial assessing the prognostic benefit of HF therapies in HFmrEF.

I will emulate a target assessing the benefit of BB, ACEI, ARB or MRA, as a single agent or in combination therapy, for reducing mortality

risk, the number of HF admissions and improving LV function in a subgroup of patients with HF and preserved or mildly reduced LV function (defined by the new parameter(s).

Future direction:
Utilise the statistical methodology developed in Chapter 5 to emulate clinical trials that answer clinical questions with limited RCT evidence

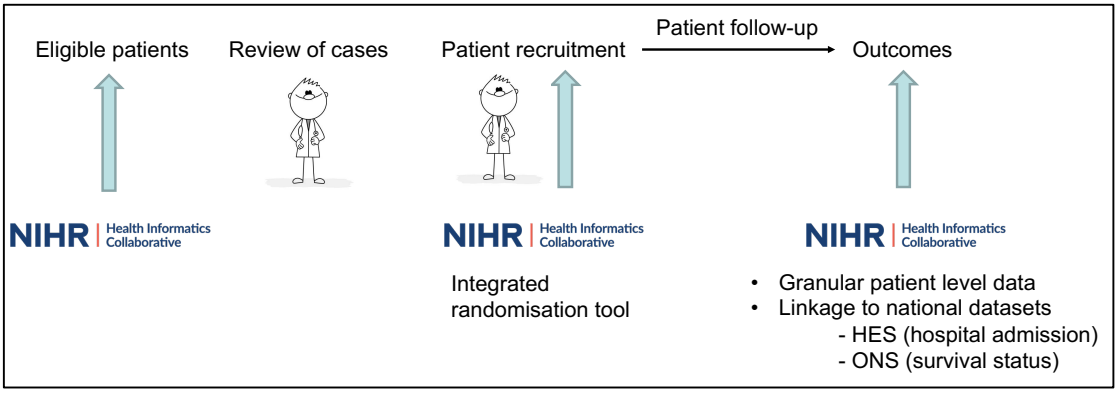
8.3.2 The relationship between observational data and randomised controlled trials

I have demonstrated the utility of retrospect assessments of patient records and I am currently finessing the architecture of the HIC to allow prospective evaluation of patients. Through linkage of the HIC dataset with the ONS and HES national datasets, we will be able to assess outcomes with no loss-to-follow-up for trial participants.

There is scope to develop the architecture of the HIC to augment clinical trial design (**Figure 8.3**):

- Using the data feeds to structured data for automated screening of eligible patients based on inclusion and exclusion criteria
- Using a built in randomisation tool to randomise eligible patients to randomisation arms
- Automated prospective primary outcome assessment. The national data feeds of hospitalisation and mortality data linked to the HIC data will considerably reduce the rate of loss-to-follow-up, unless participants emigrate from the UK.

Figure 8.3 – Using big data to augment clinical trial design



Furthermore, after completion of an RCT, as clinical guidelines are updated, the NIHR HIC big data model can be used to assess successful uptake of an effective intervention. In addition, the data can be used to show whether the intervention produces similar effect estimates in the "real word" to those observed in the clinical trial.[136]

8.3.3 Federation of data

Data federation takes a different approach to data integration to data consolidation. Rather than bringing all the data together into a local data warehouse, federation leaves the data from each Trust where it is and provides a view of the data using a virtual approach. Data federation offers a means for querying data from multiple systems as if it resides in a single data store, therefore, eliminating the need for a data warehouse with fewer steps for potential failure. One of the key benefits of data federation is real-time data access, which is important to answer research questions in a disrupted and volatile healthcare system or changing disease status, such as during the COVID-19 pandemic. This will require an upgrade of the IT infrastructure for the programme with the correct enterprise-grade tools.

8.3.4 Level of research expertise for analysing big data

Collection of large datasets on cardiovascular disease via the NIHR HIC is not only a cost-effective method of data collection but also allows novel analyses to be performed, giving further insight into cardiovascular disease. There is a continuous rise in the number of publications associated with big data in healthcare.[137] It is important for clinicians to develop the skills in using quantitative methods to fully exploit big data for research. There are currently limited options for clinicians to obtain high quality training in data curation and big data analytics beyond a higher education degree.

Future direction:
I have designed a teaching programme targeted towards clinicians interested in developing skills in data curation and the big data analytics used in this thesis. I will work with the data science team at Imperial College London to generate the teaching exercise to deliver the programme. The module headings are outlined underneath.

<u>Data Exploration</u>
Importing data

Scripts
Data visualisation
Data workflow
Data transformation
Exploratory data analysis
Basic SQL training (select, create, group by, where, case when, joins, etc.)

Data Wrangling
Tibbles
Data clean-up
Pivoting
Separating and uniting
Missing values
Mutating joins
Filtering joins
Strings
Factors
Handling dates and times
File structuring – embedding with R to make workflow easier

Programming
Pipes
Functions
Vectors
Iteration

Modelling
Comparing 2 groups
Comparing 3 or more groups
Survival analysis

Communication
Graphics
Saving your graphics

Coding Standards
SNOMED-CT
ICD-10
Other standards

8.4 Conclusion

Developing large datasets on cardiovascular disease via the NIHR HIC platform is not only a cost-effective approach for collecting data but also allows researchers to address a number of novel research questions, with an overarching aim of providing further insight into cardiovascular disease in the UK. Whilst the NIHR HIC has overcome a number of challenges to make routinely collected clinical data available for translational research across multiple sites, further development of the programme is ongoing to expand the number of centres contributing data and the scope of the data fields collected.

9. Academic achievements during this thesis

9.1 Grants

2022 – 2026	BHF 4-year MRes/PhD Studentship Programme (2022 - 2025 intake)	Fully funded with consumable allowance £20K/year	Big data and mechanistic evaluation of left heart failure (Co-supervisor)
2022 – 2022	NIHR Imperial Biomedical Research Centre	£25,000	NIHR HIC Cardiovascular Theme
2020 onwards	British Heart Foundation / NIHR	COVID-19 UK Flagship Project	Collaborative research on cardiovascular medicine and COVID-19 using high-resolution clinical data
2020 – 2022	British Heart Foundation (FS/20/18/34972)	£150,707	Using unstructured echocardiography and electrocardiography data from multicentre electronic health record systems for big data analytics
Aug 2019	European Society of Cardiology	€850	Travel grant for the European Society of Cardiology Congress 2019
2018 – 2019	NIHR Imperial Biomedical Research Centre (P74354)	£78,379 Principal Applicant	The National Health Informatics Collaboration Cardiovascular (NHIC-CV) Project

9.2 Awards and prizes

Feb 2021	**Best Presentation Prize**	National Heart and Lung Institute Away Day 2021
Aug 2019	**Blockbuster from the Young Award**	European Society of Cardiology Congress 2019
Jun 2019	**'Best of the Best' Acute Coronary Syndrome Prize**	British Cardiovascular Science Conference 2019
Mar 2019	**Runner-up President's Medal: Cardiology Section**	Royal Society of Medicine Cardiology 2018
Aug 2018	**Best Abstract Prize**	European Society of Cardiology Congress 2018
Jun 2018	**Best Oral Presentation Prize**	NIHR Clinical Academic Training Office (CATO) Annual Research Symposium 2018
Jun 2018	**Best Lay Abstract Prize**	NIHR Clinical Academic Training Office (CATO) Annual Research Symposium 2018
Nov 2017	**Runner-up Young Cardiologist of the Year Award**	Imperial Valve and Cardiovascular Course 2017

9.3 Publications

Benedetto U, Sinha S, Mulla A, Glampson B, Davies J, Panoulas V, Gautama S, Papadimitriou D, Woods K, Elliott P, Hemingway H, Williams B, Asselbergs FW, Melikian N, Krasopoulos G, Sayeed R, Wendler O, Baig K, Chukwuemeka A, Angelini GD, Sterne JAC, Johnson T, Shah AM, Perera D, Patel RS, Kharbanda R, Channon KM, Mayet J, **Kaura A**. Implications of elevated troponin on time-to-surgery in non-ST elevation myocardial infarction (NIHR Health Informatics Collaborative: TROP-CABG study). *Int J Cardiol*. 2022:S0167-5273(22)00586-1.

Kaura A, Trickey A, Shah ASV, Benedetto U, Glampson B, Mulla A, Mercuri L, Gautama S, Costelloe CE, Goodman I, Redhead J, Saravanakumar K, Mayer E, Mayet J. Comparing the longer-term effectiveness of a single dose of the Pfizer-BioNTech and Oxford-AstraZeneca COVID-19 vaccines across the age spectrum. *EClinicalMedicine*. 2022;46:101344.

Sau A, **Kaura A**, Ahmed A, Patel KHK, Li X, Mulla A, Glampson B, Panoulas V, Davies J, Woods K, Gautama S, Shah AD, Elliott P, Hemingway H, Williams B, Asselbergs FW, Melikian N, Peters NS, Shah AM, Perera D, Kharbanda R, Patel RS, Channon KM, Mayet J, Ng FS. Prognostic Significance of Ventricular Arrhythmias in 13 444 Patients With Acute Coronary Syndrome: A Retrospective Cohort Study Based on Routine Clinical Data (NIHR Health Informatics Collaborative VA-ACS Study). *J Am Heart Assoc*. 2022 Mar 15;11(6):e024260.

Kaura A, Hartley A, Panoulas V, Glampson B, Shah ASV, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Thursz MR, Elliott P, Hemmingway H, Williams B, Asselbergs FW, O'Sullivan M, Lord GM, Trickey A, Sterne JA, Haskard DO, Melikian N, Francis DP, Koenig W, Shah AM, Kharbanda R, Perera D, Patel RS, Channon KM, Mayet J, Khamis R. Mortality risk prediction of high-sensitivity C-reactive protein in suspected acute coronary syndrome: A cohort study. *PLoS Med*. 2022;19(2):e1003911.

Kaura A, Mayet J, Manisty C. Sharpening focus through wider collaboration: evolving heterogeneity in the bi-directional relationship between cardiovascular disease and COVID-19. *Eur Heart J*. 2022;43(11):1121-1123.

Glampson B, Brittain J, **Kaura A**, Mulla A, Mercuri L, Brett SJ, Aylin P, Sandall T, Goodman I, Redhead J, Saravanakumar K, Mayer EK. Assessing COVID-19 Vaccine Uptake and Effectiveness Through the North West London Vaccination Program: Retrospective Cohort Study. *JMIR Public Health Surveill.* 2021;7(9):e30010.

Kaura A, Sterne JAC, Trickey A, Mayet J. Managing NSTEMI in older patients - Authors' reply. *Lancet*. 2021;397(10272):371-372.

Benedetto U, Dimagli A, **Kaura A**, Sinha S, Mariscalco G, Krasopoulos G, Moorjani N, Field M, Uday T, Kendal S, Cooper G, Uppal R, Bilal H, Mascaro J, Goodwin A, Angelini G, Tsang G, Akowuah E; on the behalf of UK Aortic group. Determinants of outcomes following surgery for type A acute aortic dissection: the UK National Adult Cardiac Surgical Audit. *Eur Heart J*. 2021;ehab586.

Benedetto U, Sinha S, Dimagli A, Cooper G, Mariscalco G, Uppal R, Moorjani N, Krasopoulos G, **Kaura A**, Field M, Trivedi U, Kendall S, Angelini GD, Akowuah EF, Tsang G; UK Aortic Group. Decade-long trends in surgery for acute Type A aortic dissection in England: A retrospective cohort study. *Lancet Reg Health Eur*. 2021;7:100131.

Kaura A, Sterne JAC, Trickey A, Abbott S, Mulla A, Glampson B, Panoulas V, Davies J, Woods K, Omigie J, Shah AD, Channon KM, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs FW, O'Sullivan M, Lord GM, Melikian N, Johnson T, Francis DP, Shah AM, Perera D, Kharbanda R, Patel RS, Mayet J. Invasive versus non-invasive management of older patients with non-ST elevation myocardial infarction (SENIOR-NSTEMI): a cohort study based on routine clinical data. *Lancet*. 2020;396(10251):623-634.

Kaura A, Arnold A, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Lord G, Melikian N, Lefroy D, Francis D, Shah AM, Kharbanda R, Perera D, Patel R, Mayet J. Prognostic significance of troponin level in 3,121 patients presenting with atrial fibrillation (TROP-AF Study). *J Am Heart Assoc*. 2020;9:e013684.

Kaura A, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Lord G, Melikian N, Kharbanda R, Shah AM, Perera D, Patel R, Francis DP, Mayet J. Mortality association of troponin level and age in over 250000 consecutive patients undergoing troponin measurement: cohort study across five UK acute centres (NIHR Health Informatics Collaborative TROP-RISK study). *BMJ*. 2019;367:I6055.

9.4 Presentations

9.4.1 International meetings

European Society of Cardiology Congress 2022, Barcelona, Spain (Poster)	**Kaura A**, Sterne JAC, Trickey A, Mulla A, Glampson B, Davies J, Woods K, Panoulas V, Shah AD, Patel RS, Kharbanda R, Shah AM, Perera D, Channon KM, Mayet J. **Developing informatics infrastructure to curate datasets using electronic health record data from five hospitals for translational cardiovascular research**	*Eur Heart J* 2022
European Society of Cardiology Congress 2022, Barcelona, Spain	**Kaura A**, Samuel NA, Roddick AJ, Glampson B, Mulla A, Davies J, Woods K, Patel RS, Shah AM, Perera D, Channon KM, Shah ASV, Mayet J. **The prognostic significance of troponin level in patients with malignancy (NIHR Health Informatics Collaborative TROP-MALIGNANCY**	*Eur Heart J* 2022

(Poster)	**study)**	
European Society of Cardiology Congress 2022, Barcelona, Spain (Oral)	**Kaura A**, Roddick RJ, Samuel NA, Mulla A, Glampson B, Davies J, Woods K, Kharbanda R, Patel RS, Shah AM, Perera D, Channon KM, Mayet J. **The association between troponin level and mortality in patients admitted to hospital with acute stroke (NIHR Health Informatics Collaborative TROP-STROKE study)**	*Eur Heart J* 2022
American Heart Association, Boston, US (Oral)	Sinha S, Benedetto U, Mulla A, Glampson B, Panoulas VF, Gautama S, Davies J, Papadimitriou D, Woods K, Elliott P, Hemmingway H, Williams B, Asselbergs FW, Melikian N, Krasopoulos G, Sayeed R, Wendler O, Baig K, Chukwuemeka A, Angelini G, Sterne J, Johnson T, Shah AM, Perera D, Patel R, Kharbanda R, KM Channon, Mayet J, **Kaura A**. **Implications of Elevated Troponin on Time-to-Surgery in Non-ST Elevation Myocardial Infarction (NIHR Health Informatics Collaborative: Trop-CABG Study)**	*Circulation 2021;144 (Suppl_1), A14169-A14169*
European Society of Cardiology Congress 2019, Paris, France (Oral)	**Kaura A**, Hartley A, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Francis DP, Koenig W, Shah AM, Kharbanda R, Perera D, Patel RS, Mayet J, Khamis R. **hsCRP predicts mortality beyond troponin in 102,337 patients with suspected acute coronary syndrome (CRP-RISK study)**	*Eur Heart J* 2019; 40 (Suppl 1): ehz748.0127
European Society of Cardiology Congress 2019, Paris, France (Oral)	**Kaura A**, Sterne J, Mulla A, Panoulas V, Glampson B, Davies J, Woods K, Omigie J, Melikian N, Francis DP, Kharbanda R, Shah AM, Perera D, Patel RS, Mayet J. **Invasive versus medical management for non-ST elevation myocardial infarction in the elderly (SENIOR-NSTEMI study)**	*Eur Heart J* 2019; 40 (Suppl 1): ehz746.0113 **Blockbuster from the Young Award**
European Society of Cardiology Congress 2019,	**Kaura A**, Davies J, Panoulas V, Glampson B, Mulla A, Woods K, Omigie J, Shah AD, Melikian N, Francis DP, Kharbanda R, Perera D, Shah AM, Patel RS, Mayet J.	*Eur Heart J* 2019; 40 (Suppl 1): ehz745.0753

Paris, France (Poster)	**Supporting big data research in cardiovascular medicine using routinely-collected data**	
European Society of Cardiology Congress 2019, Paris, France (Poster)	**Kaura A**, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Melikian N, Kharbanda R, Perera D, Shah AM, Patel RS, Francis DP, Mayet J. **Troponin level and mortality risk in an unselected population of over 250,000 patients (TROP-RISK study)**	*Eur Heart J* 2019; 40 (Suppl 1): ehz745.0452
American College of Cardiology Congress 2019, New Orleans, USA (Poster)	**Kaura A**, Mulla A, Panoulas V, Benjamin G, Davies J, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs FW, O'Sullivan M, Kharbanda R, Lord GM, Melikian N, Patel R, Perera D, Shah A, Francis DP, Mayet J. **A propensity matched analysis of invasive versus conservative management of elderly patients with non-ST elevation myocardial infarction (SENIOR-NSTEMI Study)**	*J Am Coll Cardiol* 2019; 73 (Suppl 1): 1262
American College of Cardiology Congress 2019, New Orleans, USA (Poster)	**Kaura A**, Arnold A, Panoulas V, Glampson B, Davies J, Woods K, Mulla A, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs FW, Kharbanda R, Lord GM, Melikian N, Patel R, Perera D, Shah A, Lefroy D, Francis DP, Mayet J. **Clinical importance of troponin level in 3,121 Patients presenting with atrial fibrillation (AF-TROP Study)**	*J Am Coll Cardiol* 2019; 73 (Suppl 1): 410
American College of Cardiology Congress 2019, New Orleans, USA (Moderated poster)	**Kaura A**, Hartley A, Panoulas V, Benjamin G, Davies J, Woods K, Mulla A, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs FW, Kharbanda R, Lord GM, Melikian N, Patel R, Perera D, Shah A, Francis DP, Koenig W, Mayet J, Khamis R. **hsCRP predicts mortality beyond troponin in 102,337 patients with suspected acute coronary syndrome in the UK National Institute for Health Research CRP-RISK study**	*J Am Coll Cardiol* 2019; 73 (Suppl 1): 10

American College of Cardiology Congress 2019, New Orleans, USA (Poster)	**Kaura A**, Panoulas V, Glampson B, Davies J, Woods K, Mulla A, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs FW, O'Sullivan M, Kharbanda R, Lord GM, Melikian N, Patel R, Perera D, Shah A, Francis DP, Mayet J. **Unexpected inverted U-shaped relationship between troponin level and mortality explained by revascularization in both patients with and without acute coronary syndrome (TROP-RISK Study)**	*J Am Coll Cardiol* 2019; 73 (Suppl 1): 1086
European Society of Cardiology Congress 2018, Munich, Germany (Poster)	**Kaura A**, Panoulas V, Glampson B, Mulla A, Shah AD, Channon K, Kharbanda R, Melikian N, Patel R, Perera D, Shah A, Francis D, Mayet J. **Inverted-U-shaped relationship between troponin level and mortality in over 250,000 patients across five centres (NHIC Troponin Study)**	*Eur Heart J* 2018; 39 (Suppl 1): 548
European Society of Cardiology Congress 2018, Munich, Germany (Oral)	Panoulas V, **Kaura A**, Glampson B, Mulla A, Shah AD, Channon K, Kharbanda R, Melikian N, Patel R, Perera D, Shah A, Francis D, Mayet J. **Prognostic value of a positive troponin across the age spectrum in over a quarter of a million patients (NHIC Troponin Study)**	*Eur Heart J* 2018; 39 (Suppl 1): 219-220

9.4.2 National meetings

British Cardiovascular Society Conference 2022, Manchester, UK (Oral)	**Kaura A**, Samuel NA, Roddick AJ, Glampson B, Mulla A, Davies J, Woods K, Patel RS, Shah AM, Perera D, Channon KM, Shah ASV, Mayet J. **Prognostic significance of troponin in patients with malignancy (NIHR HEALTH INFORMATICS COLLABORATIVE TROP-MALIGNANCY STUDY)**	*Heart* 2022
British Cardiovascular Society Conference	**Kaura A**, Roddick RJ, Samuel NA, Mulla A, Glampson B, Davies J, Woods K, Kharbanda R, Patel RS, Shah AM, Perera D, Channon KM, Mayet J.	*Heart* 2022

2022, Manchester, UK (Poster)	**Association between troponin and mortality in acute stroke (NIHR HEALTH INFORMATICS COLLABORATIVE TROP-STROKE STUDY)**	
British Cardiovascular Society Conference 2022, Manchester, UK (Poster)	**Kaura A**, Sterne JAC, Trickey A, Mulla A, Glampson B, Davies J, Woods K, Panoulas V, Shah AD, Patel RS, Kharbanda R, Shah AM, Perera D, Channon KM, Mayet J. **Developing informatics infrastructure to curate datasets using electronic health record data from five NHS hospitals for translational cardiovascular research**	*Heart* 2022
British Cardiovascular Society Conference 2019, Manchester, UK (Oral)	**Kaura A**, Arnold A, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Lord G, Melikian N, Lefroy D, Francis D, Shah AM, Perera D, Kharbanda R, Patel R, Mayet J. **The prognostic implication of troponin level in over 3000 patients presenting with atrial fibrillation (NIHR Health Informatics Collaborative AF-TROP Study)**	*Heart* 2019; 105 (Suppl 6): A26-A27
British Cardiovascular Society Conference 2019, Manchester, UK (Oral)	**Kaura A**, Sterne J, Mulla A, Panoulas V, Glampson B, Davies J, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Lord G, Melikian N, Francis D, Perera D, Shah A, Kharbanda R, Patel R, Mayet J. **Invasive versus medical management of elderly patients with non-ST elevation myocardial infarction (NIHR Health Informatics Collaborative SENIOR-NSTEMI study)**	*Heart* 2019: 105 (Suppl 6): A48-A49 **'Best of the Best' Acute Coronary Syndrome Prize**
British Cardiovascular Society Conference 2019, Manchester, UK (Poster)	**Kaura A**, Hartley A, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Haskard D, Lord G, Melikian N, Francis D, Koenig W, Perera D, Shah AM, Kharbanda R, Patel R, Mayet J, Khamis R. **The role of high-sensitivity C-reactive protein in**	*Heart* 2019; 105 (Suppl 6): A120-A121

	predicting mortality beyond troponin in over 100,000 patients with suspected acute coronary syndrome (NIHR Health Informatics Collaborative CRP-RISK Study)	
British Cardiovascular Society Conference 2019, Manchester, UK (Poster)	**Kaura A**, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Lord G, Melikian N, Kharbanda R, Shah AM, Perera D, Patel R, Francis D, Mayet J. **The prognostic implication of a positive troponin across the age spectrum in a quarter of a million patients with suspected acute coronary syndrome (NIHR Health Informatics Collaborative TROP-RISK Study)**	*Heart* 2019; 105 (Suppl 6): A121
British Cardiovascular Society Conference 2019, Manchester, UK (Poster)	**Kaura A**, Panoulas V, Glampson B, Davies J, Mulla A, Woods K, Omigie J, Shah AD, Channon K, Weber JN, Thursz MR, Elliott P, Hemingway H, Williams B, Asselbergs F, O'Sullivan M, Lord G, Melikian N, Kharbanda R, Shah AM, Perera D, Patel R, Francis D, Mayet J. **The relationship between troponin level and mortality in an unselected population of over 250,000 patients with suspected acute coronary syndrome (NIHR Health Informatics Collaborative TROP-RISK study)**	*Heart* 2019; 105 (Suppl 6): A59

9.5 Invited academic talks

20/10/2021	**Why research is so important in medicine?** - 30 minutes interactive lecture - 600 attendees worldwide - https://www.youtube.com/watch?v=o7XLBgIitZg	Elsevier Ltd: Elsevier Live Student Edition
22/06/2021	**Transforming health and care: Why health and care data matters** - NHS England patient education video - https://www.youtube.com/watch?v=0I8AxND03Bg&t=9s	NHS England
09/06/2021	**New ways to investigate cardiovascular disease** - 1 hour interactive lecture	Imperial College Academic

	- 140 attendees	Health Science Centre
18/09/2020	**Career in Cardiology 2020** **Making the Transition to Academic Cardiology – Academic Clinical Fellowships**	British Cardiovascular Society

9.6 Formal student supervision

Jan – May 2022	**BSc Cardiovascular Sciences Research Project 2022** Clinical Importance of Troponin in Patients Presenting with Atrioventricular Block (NIHR Health Informatics Collaborative BRADY-TROP STUDY)	Imperial College London
Jan – Apr 2021	**BSc Cardiovascular Sciences Research Project 2021** The Impact of Ethnicity on Management and Outcomes of patients with Acute Coronary Syndrome (NIHR Health Informatics Collaborative ETHNICITY-ACS Study)	Imperial College London *Grade: Distinction*
Jan – Apr 2021	**BSc Cardiovascular Sciences Research Project 2021** The prognostic significance of cardiac troponin in a secondary care population presenting with acute exacerbations of chronic obstructive pulmonary disease (NIHR Health Informatics Collaborative TROP-COPD Study)	Imperial College London *Grade: Distinction*
Jan – Apr 2021	**Medicine Year 4 Student Elective** Impact of age on the association between lipid profile and mortality (NIHR Health Informatics Collaborative Cholesterol Risk Study)	Bart's and the London *Grade: Distinction*
Jan – Apr 2020	**BSc Cardiovascular Sciences Research Project 2020** Prognostic significance of troponin level in patients with pulmonary embolism: A cohort study using big data (NIHR Health Informatics Collaborative TROP-PE Study)	Imperial College London *Grade: Distinction*

Jan – Apr 2019	**Medicine Year 5 Student Elective 2019** Association between troponin level and mortality in acute stroke: a retrospective cohort study (NIHR Health Informatics Collaborative TROP-STROKE Study)	King's College London
Jan – Apr 2019	**BSc Cardiovascular Sciences Research Project 2019** Prognostic significance of troponin level in patients with malignancy: A cohort study using big data	Imperial College London *Grade: Distinction*

Bibliography

1 Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J*. 2018; **39**:1481-1495.

2 Rapsomaniki E, Timmis A, George J, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet*. 2014; **383**:1899-1911.

3 Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. *N Engl J Med*. 2010; **362**:2155-2165.

4 Pakhomov SS, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Epidemiology of angina pectoris: role of natural language processing of the medical record. *Am Heart J*. 2007; **153**:666-6673.

5 Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PLoS One*. 2012; **7**:e30412.

6 Lependu P, Iyer SV, Bauer-Mehren A, et al. Pharmacovigilance using Clinical Text. *AMIA Jt Summits Transl Sci Proc*. 2013; **2013**:109.

7 World Health Organization (WHO). International Statistical Classification of Diseases and Related Health Problems. 10th revision. Available from: 2016. https://icd.who.int/browse10/2016/en

8 Data saves lives: reshaping health and social care with data. Department of health and Social Care. 2022. https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data (Accessed December 24,2022).

9 Goldacre B, Morley J. Better, Broader, Safer: Using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care. Department of Health and Social Care. 2022. https://www.goldacrereview.org/ (Accessed Dec 24, 2022).

10 *EU General Data Protection Regulation (GDPR):* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

11 *Data Protection Act 2018, c. 12*. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (Accessed December 21 2022).

12 Medicines & Healthcare products Regulatory Agency. CPRD. Available from: https://cprd.com/data

13 Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015; **44**:827-836.

14 Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, Myles P. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019; **48**:1740-1740g.

15 Medicines & Healthcare products Regulatory Agency. CPRD Aurum Frequently asked questions (FAQs). Available from: https://cprd.com/sites/default/files/2022-02/CPRD%20Aurum%20FAQs%20v2.2.pdf

16 NHS Digital. Hospital Episode Statistics (HES). Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

17 Boyd A, Cornish R, Johnson L, et al. Understanding Hospital Episode Statistics (HES). London, UK: CLOSER; 2017. Available from: https://www.closer.ac.uk/wp-content/uploads/ CLOSER-resource-understanding-hospital-episode-statistics-2018.pdf

18 NHS Data Model and Dictionary. OPCS Classification of Interventions and Procedures. Available from: https://www.datadictionary.nhs.uk/supporting_information/opcs_classification_of_interventions_and_procedures.html

19 Medicines & Healthcare products Regulatory Agency. Hospital Episode Statistics (HES) Accident & Emergency and CPRD primary care data Documentation (set 21). Available from:

https://cprd.com/sites/default/files/2022-02/Documentation_HES_AE_set21.pdf

20 Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol*. 2017; **46**:1093-1093i.

21 Medicines & Healthcare products Regulatory Agency. Hospital Episode Statistics (HES) Admitted Patient Care (APC) Data Dictionary. Available from: https://cprd.com/sites/default/files/2022-02/Data_Dictionary_HES_APC.pdf

22 Medicines & Healthcare products Regulatory Agency. Medicines & Healthcare products Regulatory Agency. Available from: https://cprd.com/sites/default/files/2022-02/Documentation_HES_OP_set21.pdf

23 NICOR. National Cardiac Audit Programme (NCAP) Annual Report 2021. Available from: https://www.nicor.org.uk/wp-content/uploads/2021/10/NCAP-Aggregate_-Report_2021_FINAL.pdf

24 Wilkinson C, Weston C, Timmis A, Quinn T, Keys A, Gale CP. The Myocardial Ischaemia National Audit Project (MINAP). *Eur Heart J Qual Care Clin Outcomes*. 2020; **6**:19-22.

25 NICOR. National Cardiac Audit Programme Myocardial Ischaemia National Audit Project (MINAP). 2021 Summary Report. Available from: https://www.nicor.org.uk/wp-content/uploads/2021/10/MINAP-Domain-Report_2021_FINAL.pdf

26 Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ*. 2013; **346**:f2350.

27 National Institute for Cardiovascular Outcomes Research (NICOR). MINAP datasets. Available from: https://www.nicor.org.uk/national-cardiac-audit-programme/datasets/

28 Office for National Statistics. Mortality statistics in England and Wales QMI. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/mortalitystatisticsinenglandandwalesqmi

29 Medicines & Healthcare products Regulatory Agency. CPRD linked data. Available from: https://cprd.com/cprd-linked-data

30 Denaxas S, Gonzalez-Izquierdo A, Direk K, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc*. 2019; **26**:1545-1559.

31 Denaxas SC, George J, Herrett E, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol*. 2012; **41**:1625-1638.

32 Bottle A, Cohen C, Lucas A, et al. How an electronic health record became a real-world research resource: comparison between London's Whole Systems Integrated Care database and the Clinical Practice Research Datalink. *BMC Med Inform Decis Mak*. 2020; **20**:71.

33 THIN (The Health Improvement Network). (2020). THIN: The Health Improvement Network.

34 What is QResearch? at: http://www.qresearch.org/SitePages/What%20Is%20QResearch.aspx (Accessed 17 December 2022).

35 Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: Demographics, chronic disease prevalence and mortality rates. *J Inov Health Inform*. 2011; **19**:251-255.

36 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015; **12**:e1001779.

37 Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186**:1026-1034.

38 National Institute for Health Research. Health Informatics Collaborative. Available from: https://hic.nihr.ac.uk

39 Office for Health Improvement and Disparities. Cardiovascular disease prevention: applying All Our Health. Available from: **h**ttps://www.gov.uk/government/publications/cardiovascular-disease-prevention-applying-all-our-health/cardiovascular-disease-prevention-applying-all-our-health

40 British Heart Foundation. UK Factsheet January 2022. Available from: https://www.bhf.org.uk/-/media/files/research/heart-statistics/bhf-cvd-statistics-uk-factsheet.pdf?la=en&rev=231fa9066a64451e83b107aa63461a37&hash=1CECE96BC260AB985C5BC119F07267E43825D891

41 Kotecha T, Rakhit RD. Acute coronary syndromes. *Clin Med (Lond)*. 2016; **16**:s43-s48.

42 Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med*. 2005; **352**:1685-1695.

43 Agewall S, Giannitsis E, Jernberg T, Katus H. Troponin elevation in coronary vs. non-coronary disease. *Eur Heart J*. 2011; **32**:404-411.

44 Hamm CW, Giannitsis E, Katus HA. Cardiac troponin elevations in patients without acute coronary syndrome. *Circulation*. 2002;**106**:2871-2872.

45 Chapman AR, Adamson PD, Mills NL. Assessment and classification of patients with myocardial injury and infarction in clinical practice. *Heart* 2017; **103**:10-18.

46 Wildi K, Gimenez MR, Twerenbold R, et al. Misdiagnosis of Myocardial Infarction Related to Limitations of the Current Regulatory Approach to Define Clinical Decision Values for Cardiac Troponin. *Circulation*. 2015; **131**:2032-2040.

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

48 Burns EM, Rigby E, Mamidanna R et al. Systematic review of dis- charge coding accuracy. J Public Health (Bangkok) 2012; **34**:138–148.

49 CAPITA. The Quality of Clinical Coding in the NHS. 2014. http://www.chks.co.uk/userfiles/files/The_quality_of_clinical_coding_in_the_NHS.pdf (Accessed 24 December 2022).

50 Liuzzo G, Biasucci LM, Gallimore JR, et al. The prognostic value of C-reactive protein and serum amyloid a protein in severe unstable angina. *N Engl J Med*. 1994; **331**:417-424.

51 Berk BC, Weintraub WS, Alexander RW. Elevation of C-reactive protein in "active" coronary artery disease. *Am J Cardiol*. 1990; **65**:168-1672.

52 Koenig W, Sund M, Frohlich M, et al. C-Reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. *Circulation*. 1999; **99**:237-242.

53 Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. N Engl J Med. 1997;336(14):973-9.

54 Koenig W, Khuseyinova N, Baumert J, Meisinger C. Prospective study of high-sensitivity C-reactive protein as a determinant of mortality: results from the MONICA/KORA Augsburg Cohort Study, 1984-1998. *Clin Chem*. 2008; **54**:335-342.

55 Brunetti ND, Troccoli R, Correale M, Pellegrino PL, Di Biase M. C-reactive protein in patients with acute coronary syndrome: correlation with diagnosis, myocardial damage, ejection fraction and angiographic findings. *Int J Cardiol*. 2006; **109**:248-256.

56 Morrow DA, Rifai N, Antman EM, et al. C-reactive protein is a potent predictor of mortality independently of and in combination with troponin T in acute coronary syndromes: a TIMI 11A substudy. Thrombolysis in Myocardial Infarction. *J Am Coll Cardiol*. 1998; **31**:1460-1465.

57 Seifarth H, Schlett CL, Lehman SJ, et al. Correlation of concentrations of high-sensitivity troponin T and high-sensitivity C-reactive protein with plaque progression as measured by CT coronary angiography. *J Cardiovasc Comput Tomogr*. 2014; **8**:452-458.

58 Fracassi F, Niccoli G, Vetrugno V, et al. Optical coherence tomography and C-reactive protein in risk stratification of acute coronary syndromes. *Int J Cardiol*. 2019; **286**:7-12.

59 Wensley F, Gao P, Burgess S, et al. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *Br Med J*. 2011; **342**:d548.

60 Kaptoge S, Di Angelantonio E, Pennells L, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. *N Engl J Med*. 2012; **367**:1310-1320.

61 Ridker PM, Everett BM, Thuren T, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med*. 2017; **377**:1119-1131.

62 Ridker PM, MacFadyen JG, Everett BM, Libby P, Thuren T, Glynn RJ. Relationship of C-reactive protein reduction to cardiovascular event reduction following treatment with canakinumab: a secondary analysis from the CANTOS randomised controlled trial. *Lancet*. 2018; **391**:319-328.

63 Ridker PM, Everett BM, Pradhan A, et al. Low-Dose Methotrexate for the Prevention of Atherosclerotic Events. *N Engl J Med*. 2019; **380**:752-762.

64 Tardif JC, Kouz S, Waters DD, et al. Efficacy and Safety of Low-Dose Colchicine after Myocardial Infarction. *N Engl J Med*. 2019; **381**:2497-2505.

65 Nidorf SM, Fiolet ATL, Mosterd A, et al. Colchicine in Patients with Chronic Coronary Disease. *N Engl J Med*. 2020; **383**:1838-1847.

66 Ridker PM, Devalaraja M, Baeres FMM, et al. IL-6 inhibition with ziltivekimab in patients at high atherosclerotic risk (RESCUE): a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet*. 2021; **397**:2060-2069.

67 Mani P, Puri R, Schwartz GG, et al. Association of Initial and Serial C-Reactive Protein Levels With Adverse Cardiovascular Events and Death After Acute Coronary Syndrome: A Secondary Analysis of the VISTA-16 Trial. *JAMA Cardiol*. 2019; **4**:314-320.

68 Ries W, Torzewski J, Heigl F, et al. C-Reactive Protein Apheresis as Anti-inflammatory Therapy in Acute Myocardial Infarction: Results of the CAMI-1 Study. *Front Cardiovasc Med*. 2021; **8**:591714.

69 Stone GW, Selker HP, Thiele H, et al. Relationship Between Infarct Size and Outcomes Following Primary PCI: Patient-Level Analysis From 10 Randomized Trials. *J Am Coll Cardiol*. 2016; **67**:1674-1683.

70 Swerdlow DI, Holmes MV, Kuchenbaecker KB, et al. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet*. 2012; **379**:1214-1224.

71 Cook NR, Paynter NP, Eaton CB, et al. Comparison of the Framingham and Reynolds Risk scores for global cardiovascular risk prediction in the

multiethnic Women's Health Initiative. *Circulation*. 2012; **125**:1748-1756, s1-11.

72 Myocardial Ischaemia National Audit Project. *How the NHS cares for patients with heart attack*. *Annual Public Report April 2016–March 2017*. London: NICOR, 2018.

73 Office for National Statistics. *Overview of the UK population: July 2017*. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/july2017

74 Avezum A, Makdisse M, Spencer F, et al. Impact of age on management and outcome of acute coronary syndrome: observations from the Global Registry of Acute Coronary Events (GRACE). *Am Heart J.* 2005; **149**:67-73.

75 Granger CB, Goldberg RJ, Dabbous O, et al. Predictors of hospital mortality in the global registry of acute coronary events. *Arch Intern Med.* 2003; **163**:2345–2353.

76 Bagnall AJ, Goodman SG, Fox KA, et al. Influence of age on use of cardiac catheterization and associated outcomes in patients with non-ST-elevation acute coronary syndromes. Am J Cardiol 2009;**103**:1530–1536

77 Zaman MJ, Stirling S, Shepstone L, et al. The association between older age and receipt of care and outcomes in patients with acute coronary syndromes: a cohort study of the Myocardial Ischaemia National Audit Project (MINAP). *Eur Heart J* 2014;**35**:1551–1558.

78 Sinclair H, Kunadian V. Coronary revascularisation in older patients with non-ST elevation acute coronary syndromes. *Heart.* 2016; **102**:416-424.

79 Gierlotka M, Gąsior M, Tajstra M, et al. Outcomes of invasive treatment in very elderly Polish patients with non-ST-segment-elevation myocardial infarction from 2003-2009 (from the PL-ACS registry). *Cardiol J.* 2013; **20**:34-43.

80 Bauer T, Koeth O, Jünger C, et al. Effect of an invasive strategy on in-hospital outcome in elderly patients with non-ST-elevation myocardial infarction. *Eur Heart J.* 2007; **28**:2873-2878.

81 Devlin G, Gore JM, Elliott J, et al. GRACE Investigators. Management and 6-month outcomes in elderly and very elderly patients with high-risk non-ST-

elevation acute coronary syndromes: The Global Registry of Acute Coronary Events. *Eur Heart J.* 2008; **29**:1275-1282.

82 Suissa S. Immortal time bias in pharmaco-epidemiology*. Am J Epidemiol.* 2008; **167**:492-499.

83 Giobbie-Hurder A, Gelber RD, Regan MM. Challenges of guarantee-time bias. *J Clin Oncol.* 2013; **31**:2963-2969.

84 Luce BR, Kramer JM, Goodman SN, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med.* 2009; **151**:206-209.

85 ClinicalTrials.gov. *The British Heart Foundation SENIOR-RITA Trial (SENIOR-RITA).* Available from: https://clinicaltrials.gov/ct2/show/NCT03052036

86 Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016; **183**:758-764.

87 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.

88 Amsterdam EA, Wenger NK, Brindis RG, et al. 2014 AHA/ACC Guideline for the Management of Patients with Non-ST-Elevation Acute Coronary Syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2014; **64**: e139–e22

89 Tegn N, Abdelnoor M, Aaberge L, et al. After Eighty study investigators. Invasive versus conservative strategy in patients aged 80 years or older with non-ST-elevation myocardial infarction or unstable angina pectoris (After Eighty study): an open-label randomised controlled trial. *Lancet.* 2016; **387**: 1057-1065.

90 Savonitto S, Cavallini C, Petronio AS, et al. Italian Elderly ACS Trial Investigators. Early aggressive versus initially conservative treatment in elderly patients with non-ST-segment elevation acute coronary syndrome: a randomized controlled trial. *JACC Cardiovasc Interv.* 2012; **5**:906-916.

91 Bach RG, Cannon CP, Weintraub WS, et al. The effect of routine, early invasive management on outcome for elderly patients with non-ST-segment elevation acute coronary syndromes. *Ann Intern Med.* 2004; **141**:186-195.

92 Damman P, Clayton T, Wallentin L, et al. Effects of age on long-term outcomes after a routine invasive or selective invasive strategy in patients presenting with non-ST segment elevation acute coronary syndromes: a collaborative analysis of individual data from the FRISC II - ICTUS - RITA-3 (FIR) trials. *Heart.* 2012; **98**:207-213.

93 Gnanenthiran SR, Kritharides L, D'Souza M, et al. Revascularisation compared with initial medical therapy for non-ST-elevation acute coronary syndromes in the elderly: a meta-analysis. *Heart.* 2017; **103**:1962-1969.

94 Lindahl B, Diderholm E, Lagerqvist B, Venge P, Wallentin L; FRISC II (Fast Revascularization during InStability in CAD) Investigators. Mechanisms behind the prognostic value of troponin T in unstable coronary artery disease: a FRISC II substudy. *J Am Coll Cardiol.* 2001; **38**:979-986.

95 Ibanez B, James S, Agewall S, et al. ESC Scientific Document Group. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC*). Eur Heart J.* 2018; **39**:119-177.

96 O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013; **61**:e78-e140.

97 de Lemos JA. Increasingly sensitive assays for cardiac troponins: a review. *JAMA.* 2013; **309**:2262-2269.

98 Shah AS, Newby DE, Mills NL. High sensitivity cardiac troponin in patients with chest pain. *BMJ.* 2013; **347**:f4222.

99 Chapman AR, Lee KK, McAllister DA, et al. Association of high-sensitivity cardiac troponin I concentration with cardiac outcomes in patients with suspected acute coronary syndrome. *JAMA.* 2017; **318**:1913-1924.

100 Campbell AR, Rodriguez AJ, Larson DM, et al. Resource utilization and outcome among patients with selective vs. nonselective troponin testing. *Am Heart J.* 2018; **199**:68-74.

101 Roos A, Bandstein N, Lundbäck M, Hammarsten O, Ljung R, HolzmannMJ. Stable high-sensitivity cardiac troponin T levels and outcomes in patients with chest pain. *J Am Coll Cardiol.* 2017; **70**:2226-2236.

102 Office for National Statistics (ONS). Population Estimates for UK, England and Wales, Scotland and Northern Ireland. 2018. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland.

103 Asaria P, Elliott P, Douglass M, et al. Acute myocardial infarction hospital admissions and deaths in England: a national follow-back and follow-forward record-linkage study. *Lancet Public Health.* 2017; **2**:e191-e201.

104 Myint PK, Kwok CS, Bachmann MO, Stirling S, Shepstone L, Zaman MJ. Prognostic value of troponins in acute coronary syndrome depends upon patient age. *Heart.* 2014; **100**:1583-1590.

105 Winchester DE, Kline K, Estel C, Mahtta D, Taasan S, Peacock FW. Associations between cardiac troponin, mortality and subsequent use of cardiovascular services: differences in sex and ethnicity. *Open Heart.* 2018; **5**:e000713.

106 Antman EM, Tanasijevic MJ, Thompson B, et al. Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. *N Engl J Med.* 1996; **335**:1342-1349.

107 Goldstein SA, Newby LK, Cyr DD, et al. Relationship between peak troponin values and long-term ischemic events among medically managed patients with acute coronary syndromes. *J Am Heart Assoc.* 2017;**6**:e005334.

108 Fox KA, Clayton TC, Damman P, et al. FIR Collaboration. Long-term outcome of a routine versus selective invasive strategy in patients with non-ST-segment elevation acute coronary syndrome a meta-analysis of individual patient data. *J Am Coll Cardiol.* 2010; **55**:2435-2445.

109 de Lemos JA, Drazner MH, Omland T, et al. Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population. *JAMA.* 2010; **304**:2503-2512.

110 Kirchhof P, Benussi S, Kotecha D, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J*. 2016; **37**:2893-2962.

111 Lip GY, Frison L, Halperin JL, Lane DA. Identifying patients at high risk for stroke despite anticoagulation: a comparison of contemporary stroke risk stratification schemes in an anticoagulated atrial fibrillation cohort. *Stroke*. 2010; **41**:2731-2738.

112 Hijazi Z, Oldgren J, Andersson U, et al. Cardiac biomarkers are associated with an increased risk of stroke and death in patients with atrial fibrillation: a Randomized Evaluation of Long-term Anticoagulation Therapy (RE-LY) substudy. *Circulation*. 2012; **125**:1605-1616.

113 Conti A, Mariannini Y, Canuti E, et al. Role of masked coronary heart disease in patients with recent-onset atrial fibrillation and troponin elevations. *Eur J Emerg Med*. 2015; **22**:162-169.

114 Eggers KM, Lind L, Ahlström H, et al. Prevalence and pathophysiological mechanisms of elevated cardiac troponin I levels in a population-based sample of elderly subjects. *Eur Heart J*. 2008; **29**:2252-2258.

115 Hijazi Z, Wallentin L, Siegbahn A, et al. High-sensitivity troponin T and risk stratification in patients with atrial fibrillation during treatment with apixaban or warfarin. *J Am Coll Cardiol*. 2014; **63**:52-61.

116 Alghamry A, Hanna J, Pelecanos A, et al. Predictors of significant coronary artery disease in atrial fibrillation: Are cardiac troponins a useful measure. *Int J Cardiol*. 2016; **223**:744-749.

117 Costabel JP, Burgos LM, Trivi M. The Significance Of Troponin Elevation In Atrial Fibrillation. *J Atr Fibrillation*. 201; **9**:1530.
.

118 Ulimoen SR, Enger S, Norseth J, et al. Improved rate control reduces cardiac troponin T levels in permanent atrial fibrillation. *Clin Cardiol*. 2014; **37**:422-427.

119 Lasek-Bal A, Gąsior Z, Kowalewska-Twardela T, Urbanek T. New-onset atrial fibrillation in patients with elevated troponin I levels in the acute phase of stroke. *Int J Cardiol*. 2015; **195**:210-211.

120 Kalkman DN, Aquino M, Claessen BE, et al. Residual inflammatory risk and the impact on clinical outcomes in patients after percutaneous coronary interventions. *Eur Heart J*. 2018; **39**:4101-4108.

121 Stone GW, Maehara A, Lansky AJ, et al. A prospective natural-history study of coronary atherosclerosis. *N Engl J Med*. 2011; **364**:226-235.

122 Frank L J Visseren, François Mach, Yvo M Smulders, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur J Prev Cardiol.* 2022; **29**:5-115.

123 Smith DA, Wang T, Freeman O, et al. National Institute for Health Research Health Informatics Collaborative: development of a pipeline to collate electronic clinical data for viral hepatitis research. *BMJ Health Care Inform*. 2020; **27**:e100145.

124 Kodera S, Akazawa H, Morita H, Komuro I. Prospects for cardiovascular medicine using artificial intelligence. *J Cardiol*. 2022; **79**:319-325.

125 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**:2129-2200.

126 Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013; **62**:e147-e239.

127 Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACCF/ACC/HFSA focused update on new pharmacological therapy for heart failure: an update of the 2013 ACCF/AHA guidelines for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Heart Failure Society of America. *J Am Coll Cardiol.* 2017; **70**:776-803.

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

129 Meta-analysis Global Group in Chronic Heart Failure (MAGGIC). The survival of patients with heart failure with preserved or reduced left ventricular ejection fraction: an individual patient data meta-analysis. *Eur Heart J.* 2012; **33**:1750-1757.

130 Lam CSP, Solomon SD. The middle child in heart failure: heart failure with mid-range ejection fraction (40–50%). *Eur J Heart Fail.* 2014; **16**:1049-1055.

131 Zheng S, Chan FT, Nabeebaccus AA, Shah AM, McDonagh T, Okonko DO, Ayis S. Drug treatment effects on outcomes in heart failure with preserved ejection fraction: a systematic review and meta-analysis. *Heart.* 2018; **104**:407-415.

132 Cleland JGF, Bunting KV, Flather MD, et al. Beta-blockers for heart failure with reduced, mid-range, and preserved ejection fraction: an individual patient-level analysis of double-blind randomized trials. *Eur Heart J.* 2018; **39**:26-35

133 Solomon SD, Claggett B, Lewis EF, et al. Influence of ejection fraction on outcomes and efficacy of spironolactone in patients with heart failure with preserved ejection fraction. *Eur Heart J.* 2016; **37**:455-462.

134 Lund LH, Claggett B, Liu J, et al. Heart failure with mid-range ejection fraction in CHARM: characteristics, outcomes and effect of candesartan across the entire ejection fraction spectrum. *Eur J Heart Fail.* 2018; **20**:1230-1239.

135 Maggioni AP, Anker SD, Dahlstro¨m U, et al. Are hospitalized or ambulatory patients with heart failure treated in accordance with European Society of Cardiology guidelines? Evidence from 12 440 patients of the ESC Heart Failure Long-Term Registry. *Eur J Heart Fail.* 2013; **15**:1173-1184.

136 Mayo CS, Matuszak MM, Schipper MJ, Jolly S, Hayman JA, Ten Haken RK. Big Data in Designing Clinical Trials: Opportunities and Challenges. *Front Oncol*. 2017;**7**:187.

137 Dash S, Shakyawar SK, Sharma M, et al. Big data in healthcare: management, analysis and future prospects. *J Big Data.* 2019; **6:**54.

Appendix A

R Code

This Appendix contains the R code to generate propensity scores and the steps involved in creating propensity score groups (Table 5.6).

Name	Brief description
1_propensity_scoring	Runs stepwise logistic regression to create propensity score model and adds variable to the dataset for probabilities ROC curve plot

```
#Propensity Scoring
#Sort the package installation
#-------------------------------------------------------------------
-----------------------------------
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/e1071_1.7-4.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/VIM_6.1.0.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/survMisc_0.5.5.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/laeken_0.5.1.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/KMsurv_0.1-5.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/ranger_0.12.1.zip", repos = NULL, type =
"win.binary")

install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/survminer_0.4.8.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/km.ci_0.5-2.zip", repos = NULL, type =
"win.binary")

install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/simputation_0.2.6.zip", repos = NULL, type =
"win.binary")
```

```
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/vcd_1.4-8.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/xtable_1.8-4.zip", repos = NULL, type =
"win.binary")

install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/survey_4.0.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/tableone_0.12.0.zip", repos = NULL, type =
"win.binary")
install.packages("S:/Business Intelligence - Covid Analytics
Project/R Packages/labelled_2.7.0.zip", repos = NULL, type =
"win.binary")

#-----------------------------------------------------------------
-------------------------------------
#Ignore warning of 'Rtools is required to build R packages but is
not currently installed. Please download and install the
appropriate version of Rtools before proceeding'

library(mice)
library(survival)
library(survminer)
library(VIM)
library(tidyverse)
library(simputation)

library(dplyr)
library(broom)
library(tableone)
library(Hmisc)
library(rms)
library(pROC)
library(ggplot2)
library(purrr)
#-----------------------------------------------------------------
-------------------------------------
#Propensity score analysis - GLM: binomial with logit link and
Splines
#-----------------------------------------------------------------
-------------------------------------
#filter analysis dataset to not died

analysis_dataset %>%
```

```
filter(DIEDwithin3daysFROMpeakTROP==0) %>%count()

#new 3294
#-----------------------------------------------------------------
-------------------------------------
#Generating splines for continuous variables
#-----------------------------------------------------------------
-------------------------------------
#library(Hmisc)
#library(rms)
#library(dplyr)

#check<-rcs(analysis_dataset$haemoglobin_result, 3)


agesp<- as.data.frame(rcspline.eval(analysis_dataset$ageyears,
nk=3, inclx=TRUE))
analysis_dataset<- agesp %>%
rename(agesp1 = x,
agesp2= V2) %>%
cbind(analysis_dataset)


creatininesp<-
as.data.frame(rcspline.eval(analysis_dataset$creatinine_result,
nk=3, inclx=TRUE))
analysis_dataset<- creatininesp %>%
rename(creatininesp1 = x,
creatininesp2= V2) %>%
cbind(analysis_dataset)

haemoglobinsp<-
as.data.frame(rcspline.eval(analysis_dataset$haemoglobin_result,
nk=3, inclx=TRUE))
analysis_dataset<- haemoglobinsp %>%
rename(haemoglobinsp1 = x,
haemoglobinsp2= V2) %>%
cbind(analysis_dataset)

plateletcountsp<-
as.data.frame(rcspline.eval(analysis_dataset$plateletcount_result,
nk=3, inclx=TRUE))
analysis_dataset<- plateletcountsp %>%
rename(plateletcountsp1 = x,
plateletcountsp2= V2) %>%
cbind(analysis_dataset)
```

```
potassiumsp<-
as.data.frame(rcspline.eval(analysis_dataset$potassium_result,
nk=3, inclx=TRUE))
analysis_dataset<- potassiumsp %>%
rename(potassiumsp1 = x,
potassiumsp2= V2) %>%
cbind(analysis_dataset)


sodiumsp<-
as.data.frame(rcspline.eval(analysis_dataset$sodium_result, nk=3,
inclx=TRUE))
analysis_dataset<- sodiumsp %>%
rename(sodiumsp1 = x,
sodiumsp2= V2) %>%
cbind(analysis_dataset)


whitecellsp<-
as.data.frame(rcspline.eval(analysis_dataset$whitecellcount_result
, nk=3, inclx=TRUE))
analysis_dataset<- whitecellsp %>%
rename(whitecellsp1 = x,
whitecellsp2= V2) %>%
cbind(analysis_dataset)


analysis_dataset$log_spt<-
log(analysis_dataset$Standardised_peak_troponin)

log_spt_sp<- as.data.frame(rcspline.eval(analysis_dataset$log_spt,
nk=3, inclx=TRUE))
analysis_dataset<- log_spt_sp %>%
rename(lsptsp1 = x,
lsptsp2= V2) %>%
cbind(analysis_dataset)

#-----------------------------------------------------------------
-----------------------------------
#Create individual hospital variables
#-----------------------------------------------------------------
-----------------------------------

table(analysis_dataset$brcname)


analysis_dataset %>%
```

```
mutate(hosp_gstt = if_else(brcname=="GSTT", 1,0),
hosp_ichnt = if_else(brcname=="ICHNT", 1,0),
hosp_oxfor = if_else(brcname=="OXFOR", 1,0),
hosp_uclh = if_else(brcname=="UCLH", 1,0)) -> analysis_dataset


xtabs(~hosp_uclh + Invasive_at_3_daysFROMpeakTROP,
alive_analysis_dataset)

table(analysis_dataset$hosp_gstt)
table(analysis_dataset$hosp_ichnt)
table(analysis_dataset$hosp_oxfor)
table(analysis_dataset$hosp_uclh)

#-----------------------------------------------------------------
-------------------------------------
#Check any NAs as would could screw up modelling for specific
variables
#Need to look at missingness patterns before modelling as can bias
#We may also have correlated terms not needed in the
model/colinearity
#-----------------------------------------------------------------
-------------------------------------
analysis_dataset %>%
summarise_all(funs(sum(is.na(.)))) %>%
t(.) %>% View()

#None

#-----------------------------------------------------------------
-------------------------------------
contVars <- c("ageyears","crp_result", "creatinine_result",
"haemoglobin_result","plateletcount_result", "potassium_result",
"sodium_result", "whitecellcount_result",
"Standardised_peak_troponin",
"frailty_score")


catVars <- c("Male_sex", "hosp_gstt", "hosp_ichnt","hosp_oxfor",
"hosp_uclh", "diabetes", "FhxCHD", "Hypercholesterolaemia",
"Hypertension", "Smoking",
"Aortic_aneurysm", "Angina", "aortic_stenosis",
"atrial_fibrillation", "Cardiac_arrest", "Low_grade_AV_block",
"Complete_AV_block",
"HF", "PVD", "Previous_MI", "SVT", "VFib", "VT", "AKI", "CKD",
"UTI", "Urinary_catheterisation", "Urinary_incontinence",
"Interstitial_lung_disease",
```

```
"Obstructive_Lung_Disease", "Asthma", "Pneumonia",
"Pulmonary_embolism", "Respiratory_Failure", "Ischaemic_stroke",
"TIA", "Haemorrhagic_stroke",
"SAH", "SDH", "Parkinsons", "Alcohol_excess", "Anxiety",
"Bipolar", "Delirium", "Depression", "Dementia", "Gastric_ulcer",
"GI_haemorrhage", "Constipation", "Faecal_incontinence",
"Liver_disease", "Arthritis", "Fracture", "Osteoporosis",
"Malignant_neoplasm",
"Benign_neoplasm", "Leukaemia", "Lymphoma", "Sepsis",
"Inflammatory_disorder", "frailty_score", "Fall", "Hearing",
"Visual",
"Assistance", "Weight_loss")


analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0) -> alive_analysis_dataset

for (i in catVars){
print(i)
print(table(alive_analysis_dataset[, i], useNA='always'))

}

#For some variables there are very few events may want to consider
not including
#table(analysis_dataset[, "GI_haemorrhage"], useNA='always')

for (i in contVars){
print(i)
print(summary(alive_analysis_dataset[, i]))

}


#-----------------------------------------------------------------
-------------------------------------
#Logistic regression model
#-----------------------------------------------------------------
-------------------------------------
analysis_dataset$brcname<-
as.factor(analysis_dataset$brcname)#there were modelling errors
when including separate as binary

analysis_dataset %>% filter(DIEDwithin3daysFROMpeakTROP==0) ->
analysis_dataset_alive
```

```
#hosp_gstt+ hosp_ichnt+hosp_oxfor+hosp_uclh

model_formula_rcs <- "Invasive_at_3_daysFROMpeakTROP~
rcs(ageyears, 3) + Male_sex + brcname + rcs(creatinine_result, 3)
+
rcs(haemoglobin_result, 3) + rcs(plateletcount_result, 3) +
rcs(potassium_result, 3) + rcs(sodium_result, 3) + rcs(log_spt, 3)
+
rcs(whitecellcount_result, 3) + diabetes + FhxCHD +
Hypercholesterolaemia +
Hypertension + Smoking + Aortic_aneurysm + Angina +
aortic_stenosis +
atrial_fibrillation + Cardiac_arrest + Low_grade_AV_block +
Complete_AV_block +
HF + PVD + Previous_MI + SVT + VFib + VT + AKI + CKD + UTI +
Urinary_catheterisation +
Urinary_incontinence + Interstitial_lung_disease +
Obstructive_Lung_Disease +
Asthma + Pneumonia + Pulmonary_embolism + Respiratory_Failure +
Ischaemic_stroke +
TIA + Haemorrhagic_stroke + SAH + SDH + Parkinsons +
Alcohol_excess + Anxiety +
Bipolar + Delirium + Depression + Dementia + Gastric_ulcer +
GI_haemorrhage+ Constipation+ Faecal_incontinence + Liver_disease
+ Arthritis +
Fracture + Osteoporosis + Malignant_neoplasm + Benign_neoplasm +
Leukaemia+Lymphoma +
Sepsis + Inflammatory_disorder + frailty_score + Fall + Hearing +
Visual + Assistance + Weight_loss"



#-----------------------------------------------------------------
-------------------------------------
prop_model <- rms::lrm(formula = as.formula(model_formula_rcs),
data = analysis_dataset_alive, #filter data to those who have not
died
x = T,
y = T,
maxit=10000,
tol=1e-10) #need maxit=1000 as the MLE may not converge in a
number of steps

#-----------------------------------------------------------------
-------------------------------------
#Backwards elimination using p value of 0.2
```

```
backwards<- rms::fastbw(prop_model, rule = "p", sls=0.2,
type='individual')
#use type = individual to get similar behaviour to step
#https://stat.ethz.ch/pipermail/r-help/2010-February/228073.html


#Returns the variables which have been retained in the model
keptvars<- backwards$names.kept
#cat(noquote(keptvars))

#Can use the following to make it less manual for the final model
but will still need to put the RCS in

formula <- ""

for (i in backwards$names.kept) {
formula<- paste0(formula, " + ",i)
}

formula <- paste0("Invasive_at_3_daysFROMpeakTROP ~ ",
substr(formula,4,nchar(formula)))

#-----------------------------------------------------------------
-------------------------------------
#Final model formula for the propensity score
#-----------------------------------------------------------------
-------------------------------------
propen_formula <- "Invasive_at_3_daysFROMpeakTROP ~ Male_sex +
brcname + rcs(haemoglobin_result, 3) + rcs(plateletcount_result,3)
+ rcs(log_spt, 3) + FhxCHD + Hypercholesterolaemia + Hypertension
+ Smoking + Angina + aortic_stenosis + atrial_fibrillation +
Cardiac_arrest + HF + PVD + Previous_MI + VT + AKI +
Interstitial_lung_disease + Obstructive_Lung_Disease + Pneumonia +
Ischaemic_stroke + Anxiety + Bipolar + Dementia + GI_haemorrhage +
Arthritis + Osteoporosis + Malignant_neoplasm + Sepsis +
Inflammatory_disorder + Fall + Visual"

propen_model<- rms::lrm(formula = as.formula(propen_formula),
data = analysis_dataset_alive, #filter data to those who have not
died
x = T,
y = T,
maxit=1000)
#-----------------------------------------------------------------
-------------------------------------
#Predicted probabilities for the whole dataset
```

```
#turn off scientific notation or it will print 10 to the minus
numbers which isnt helpful!
options(scipen=999)

analysis_dataset$propscore2<- predict(propen_model,
newdata=analysis_dataset, type="fitted.ind")

#-----------------------------------------------------------------
-------------------------------------
#car::vif(propen_model)
#Test of multicollinearity
#Rule of thumb vif score over 5 potential prob, anything over 10
needs to be remedied
#drop problematic varaible or create index of closely related
variables
#-----------------------------------------------------------------
-------------------------------------
#Keep check of prognostic model AUC to check it not doing
something crazy
#-----------------------------------------------------------------
-------------------------------------
#subset to those who did not die within 3 days to check the prop
score model
analysis_dataset_nodeath <- subset(analysis_dataset,
DIEDwithin3daysFROMpeakTROP == 0)

#library(pROC)

ROC <-
roc(analysis_dataset_nodeath$Invasive_at_3_daysFROMpeakTROP,
analysis_dataset_nodeath$propscore2)

plot(ROC, col = "red")

auc(ROC)
#0.8301
#This AUC seems reasonable
#-----------------------------------------------------------------
-------------------------------------
```

2_propensity_groupings_plot — Creates propensity score groups for table 1, HRs, probability density function plot (figure 2)

```
#Propensity groupings and plot
```

```
#-----------------------------------------------------------------
-------------------------------------
#Create a vector of percentiles you want, these correspond to
table 1
p<- c(0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99)

#Use quantile function to apply these percentiles to propscore2
percentiles<- as.data.frame(quantile(analysis_dataset$propscore2,
p))

#-----------------------------------------------------------------
-------------------------------------
#Create propensity score groups based on the above output
#-----------------------------------------------------------------
-------------------------------------

analysis_dataset %>%
mutate(propsgroup2 = case_when(
propscore2 < percentiles[1,] ~ 1,
propscore2>=percentiles[1,] & propscore2<percentiles[2,] ~ 2,
propscore2>=percentiles[2,] & propscore2<percentiles[3,] ~ 3,
propscore2>=percentiles[3,] & propscore2<percentiles[4,] ~ 4,
propscore2>=percentiles[4,] & propscore2<percentiles[5,] ~ 5,
propscore2>=percentiles[5,] & propscore2<percentiles[6,] ~ 6,
propscore2>=percentiles[6,] & propscore2<percentiles[7,] ~ 7,
propscore2>=percentiles[7,] & propscore2<percentiles[8,] ~ 8,
propscore2>=percentiles[8,] & propscore2<percentiles[9,] ~ 9,
propscore2>=percentiles[9,] ~10))-> analysis_dataset

#prop score is calculated using the whole dataset including those
who died within 3 days of peak troponin
#looking at Adams code and discussion with Amit and Jonathan.
#-----------------------------------------------------------------
-------------------------------------
#summarise proportions
#-----------------------------------------------------------------
-------------------------------------
analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0) %>%
group_by(propsgroup2) %>%
summarise(n=n()) %>%
mutate(percent = (n / sum(n))*100) %>%
as.data.frame() -> propensity_proportions

propensity_proportions
#-----------------------------------------------------------------
-------------------------------------
```

```
#Summarise propensity groups for those who did not die in 3 days
of peak trop
#0 is non-invasive, 1 is invasive
#-----------------------------------------------------------------
-----------------------------------

analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0) %>%
group_by(propsgroup2, Invasive_at_3_daysFROMpeakTROP) %>%
summarise(num_patients=n(), num_deaths = sum(lifeststusall)) %>%
pivot_wider(names_from = Invasive_at_3_daysFROMpeakTROP,
values_from = c(num_patients, num_deaths)) %>%
#mutate(total = `0` + `1`) %>%
rename(num_patients_noninvasive=num_patients_0,
num_patients_invasive=num_patients_1,
num_deaths_noninvasive= num_deaths_0,
num_deaths_invasive= num_deaths_1) %>%
replace(is.na(.), 0) %>%
mutate(num_deaths_percent_noninvas =
round(num_deaths_noninvasive*100/num_patients_noninvasive,1),
num_deaths_percent_invas =
round(num_deaths_invasive*100/num_patients_invasive, 1)) %>%
as.data.frame() %>%
replace(is.na(.), 0) -> table_1_output

table_1_output$death_noninvas_percent<-
paste0(table_1_output$num_deaths_noninvasive, " (",
table_1_output$num_deaths_percent_noninvas, ")")
table_1_output$death_invas_percent<-
paste0(table_1_output$num_deaths_invasive, " (",
table_1_output$num_deaths_percent_invas, ")")
#-----------------------------------------------------------------
-----------------------------------
percentiles %>%
rename(prop_upper_limit = `quantile(analysis_dataset$propscore2,
p)`) %>%
add_row( prop_upper_limit=1.00000000) %>%
arrange(desc(prop_upper_limit)) %>%
mutate(prop_upper_limit = round(prop_upper_limit, 3))->
tmp_percentile

#Change order so matches Amit's word document
#-----------------------------------------------------------------
-----------------------------------
table_1_output %>%
select(propsgroup2,
num_patients_invasive,
```

```
death_invas_percent,
num_patients_noninvasive,
death_noninvas_percent) %>%
arrange(desc(propsgroup2))-> table_1_output_slim

#-----------------------------------------------------------------
-------------------------------------
table_1_output_slim_comp<- cbind(tmp_percentile,
table_1_output_slim)
table_1_output_slim_comp %>% View()
#-----------------------------------------------------------------
-------------------------------------
write.csv(table_1_output_slim_comp,
"table_1_output_01_08_2021.csv")
#-----------------------------------------------------------------
-------------------------------------
```

```
#-----------------------------------------------------------------
-------------------------------------
#Hazard Ratios for Table 1
#-----------------------------------------------------------------
-------------------------------------
analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0)-> analysis_restrict_1
#-----------------------------------------------------------------
-------------------------------------
# HR for each propensity group for invasive versus non invasive
(minus those who died)
#-----------------------------------------------------------------
-------------------------------------
#So this is slightly different data each time, subsets of
propensity score group
#Can automate this using purrr

#library(purrr)

analysis_restrict_1 %>%
select(propsgroup2, lifeststusall, Invasive_at_3_daysFROMpeakTROP,
PEAKTntoCENSOR, brcname) %>%
split(.$propsgroup2) %>% # from base R
```

```
map(~ coxph(Surv(PEAKTntoCENSOR,lifeststusall) ~
Invasive_at_3_daysFROMpeakTROP + strata(brcname), data = .) ) %>%
map(summary) -> univariate_cox

#-----------------------------------------------------------------
-------------------------------------
options("scipen")

original1 = NULL
for (i in 1:10) {
new <- as.data.frame(round(univariate_cox[[i]][["conf.int"]], 2))
new$prop_group<- i
new$hr_ci<- paste0(new$`exp(coef)`, " (", new$`lower .95`, "-",
new$`upper .95`, ")")
original1 <- rbind(original1, new)
}

#-----------------------------------------------------------------
-------------------------------------
original1 %>%
select(prop_group, hr_ci) -> univariate_results

rownames(univariate_results) <- c()
#-----------------------------------------------------------------
-------------------------------------
saveRDS(univariate_results, "table_1_HR_01_08_2021.rds")
#-----------------------------------------------------------------
-------------------------------------


#-----------------------------------------------------------------
-------------------------------------
#Cox Regression overall for table 1
#-----------------------------------------------------------------
-------------------------------------
analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0 & (propsgroup2>1 &
propsgroup2<8))-> analysis_overall_hr
#-----------------------------------------------------------------
-------------------------------------
Surv(analysis_overall_hr$PEAKTntoCENSOR,
analysis_overall_hr$lifeststusall) -> surv_o

cox_overall <- coxph(surv_o ~ Invasive_at_3_daysFROMpeakTROP +
strata(brcname),
data = analysis_overall_hr)

summary(cox_overall) #output provides HR CIs
```

```
confint(cox_overall) #coefficient CIs
exp(confint(cox_overall)) #Also HR CIs
#-----------------------------------------------------------------
-------------------------------------
#You can output as a table for further processing/automation using
Broom
#-----------------------------------------------------------------
-------------------------------------
#library(broom)

cox_overall %>%
tidy %>%
mutate(
estimate=exp(estimate),
conf.low=exp(conf.low),
conf.high=exp(conf.high)
) %>%
select(term, estimate, starts_with("conf")) %>%
as.data.frame() -> overallhazard

#-----------------------------------------------------------------
-------------------------------------
saveRDS(overallhazard, "table_1_overallhazard_01_08_2021.rds")
#-----------------------------------------------------------------
-------------------------------------
```

```
#-----------------------------------------------------------------
-------------------------------------
#Figure 2
#Combined histogram and probability density function of the
propensity score for
#those who had both non invasive and invasive management who
survived after 3 days of peak troponin concentration
#-----------------------------------------------------------------
-------------------------------------
#setup:
#Make management variable as a factor for ggplot
```

```
analysis_dataset<-analysis_dataset %>%
mutate(management = case_when(Invasive_at_3_daysFROMpeakTROP ==1 ~
"Invasive",Invasive_at_3_daysFROMpeakTROP == 0 ~ "Non-invasive"),
management=factor(management, c("Invasive", "Non-invasive")))

#-----------------------------------------------------------------
-------------------------------------
#line for mean propensity score
#Dont necessarily need to add this but I include for completion
means <- analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0) %>%
group_by(management) %>%
summarise(grp.mean = mean(propscore2))

#-----------------------------------------------------------------
-------------------------------------
#For the plot
#windowsFonts(Arial=windowsFont("TT Arial"))
#dev.off()
#library(ggplot2)

analysis_dataset %>%
filter(DIEDwithin3daysFROMpeakTROP==0) %>%
ggplot(aes(x=propscore2)) -> p

p + geom_rect(aes(xmin=-0.05, xmax=0.036, ymin=0, ymax=Inf),
fill="#D3D3D3", alpha=0.2)+ #lower limit for the propensity score
group 5 to <10
geom_rect(aes(xmin=0.931, xmax=1.02, ymin=0, ymax=Inf),
fill="#D3D3D3", alpha=0.2)+
geom_histogram(aes(y = ..density.., fill=management), alpha=0.4,
color="black", position = "identity", bins = 30)+
geom_density(aes(color = management), size=0.5, show.legend =
FALSE) +
geom_vline(xintercept =c(0.036, 0.931), linetype = "dotted",
size=1) +
#geom_vline(data = means, aes(xintercept = grp.mean, color =
management), linetype = "dashed", show.legend = FALSE)+ #this is
if you want to add the means
xlab("Propensity Score") + ylab("Probability Density Function")+
#geom_text(aes(x=0.02, y=-0.05, label="0.02"), size=3,
family="Arial")+
#geom_text(aes(x=0.08, y=3, label="Cutoff point\n for inclusion"),
size=3, family="Arial", fontface="plain")+
#geom_text(aes(x=0.82, y=3, label="Cutoff point\n for inclusion"),
size=3, family="Arial")+
#annotate("text", x=0.1165, y=-0.05, label="0.12", size=3)+
```

```
#annotate("text", x=0.9346, y=-0.05, label="0.93", size=3)+
annotate("text", x=0.13, y=2.7, label="Cutoff point\n for
inclusion", size=3)+
annotate("text", x=0.82, y=2.7, label="Cutoff point\n for
inclusion", size=3)+
annotate("text", x=-0.02, y=2.7, label="<1st\n
percentile\nexcluded", size=3)+
annotate("text", x=0.98, y=2.7, label=">90th\n
percentile\nexcluded", size=3)+
#geom_label(x = 0.036, y = -0.05, label = "0.036", label.padding
=unit(0.1, "lines"))+
#scale_fill_manual(values = c("#00AFBB", "#E7B800")) +
scale_color_manual(values = c("#00AFBB", "#E7B800"))
theme_classic() +
theme(legend.position="bottom", legend.box = "horizontal") +
theme(legend.title = element_blank())+
scale_fill_discrete(labels=c("Invasive (1934)", "Non-invasive
(1360)"))

#-----------------------------------------------------------------
-------------------------------------
#Take splines of propensity score variable (before restricting the
data)
#The order of when the splines are developed will matter
#-----------------------------------------------------------------
-------------------------------------

propscore2sp<-
as.data.frame(rcspline.eval(analysis_dataset$propscore2, nk=3,
inclx=TRUE))

analysis_dataset<- propscore2sp %>%
rename(propscore2sp1 =x,
propscore2sp2= V2) %>%
cbind(analysis_dataset)

#-----------------------------------------------------------------
-------------------------------------
```