New-onset atrial fibrillation in critically ill patients: risk factors, treatments, and outcomes



Jonathan Bedford

Worcester College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

Acknowledgements

This dissertation would not have been possible without the help of many people. I would like to express my deep gratitude to my supervisors Professors Peter Watkinson, David Clifton, and Gary Collins for their kind and expert supervision. Their enthusiasm for research, and their rigorous scientific approach are examples to follow.

I would like to acknowledge my friends and colleagues in the Critical Care Research Group who provided invaluable feedback, support, and inspiration along the way. Their encouragement and camaraderie made this journey enjoyable.

I am also grateful to the faculty and staff of the Nuffield Department of Clinical Neurosciences for their assistance. This work wouldn't have been possible without my funding body the National Institute for Health and Care Research

Finally, I would like to express my deepest appreciation to my family for their love and support throughout my academic journey. Their encouragement and understanding have been invaluable, and I am forever grateful for their belief, patience, and sacrifices.

Funding

This work was supported by an NIHR Doctoral Research Fellowship (NIHR300224).

Abstract

New-onset atrial fibrillation (NOAF) is defined as atrial fibrillation (AF) occurring in a patient with no history of chronic or paroxysmal AF. It is a common arrhythmia in critically ill patients. Existing tools to estimate the risk of a patient developing NOAF in a general intensive care unit (ICU) have significant limitations. We know little about the haemodynamic changes associated with NOAF in patients in an ICU. The best treatments to be compared in randomised trials are unknown. Studies investigating the survival impact of NOAF in critically ill patients are limited by size or methodology.

In this thesis, an international Delphi study examining the risk factors for NOAF in critically ill patients is undertaken. Using these systematically-identified variables, a NOAF prediction model is developed using international data. This model outperforms a comparable model in an assessment of external validity using multi-centre UK data. A user-friendly tool is built to estimate NOAF risk using 10 readily available clinical variables.

Outcomes associated with first-line treatments for ICU-acquired NOAF using a propensity-weighted analysis are then investigated. Digoxin therapy is found to be associated with inferior rate control compared with amiodarone, with calcium channel blocker therapy being associated with inferior rhythm control compared with amiodarone. Apparent mortality differences between beta blockers and amiodarone in unadjusted analyses are no longer evident when cardiovascular stability at the time of AF onset is taken into account. This finding has now been included in American national guidelines.

The immediate haemodynamic impact of NOAF in the ICU is subsequently explored. This work demonstrates that NOAF is associated with a significant increase in heart rate and reduction in systolic blood pressure. Finally, a competing risks analysis of the association of NOAF with hospital mortality is presented. This study reveals that NOAF is associated with both an increased duration of hospital stay and an increased rate of in-hospital death. NOAF lasting over 30 minutes was associated with increased hospital mortality.

This thesis argues that ICU-acquired NOAF is associated with poor outcomes. It highlights optimal therapies that may be compared in randomised trials. Lastly, it suggests that its prediction is possible. Targeted prevention strategies in high-risk patients combined with optimal treatment has the potential to improve outcomes for a vulnerable cohort.

Related publications

This thesis aggregates and extends some of the content published throughout the course of my doctoral studies. Related publications are listed below by thesis chapter.

Chapter 1

Bedford, J. P., Ede, J. & Watkinson, P. J. Triggers for new-onset atrial fibrillation in critically ill patients. *Intensive & Critical Care Nursing* **67** (2021).

O'Bryan, L. J., Redfern, O. C., Bedford, J. P., *et al.* Managing new-onset atrial fibrillation in critically ill patients: A systematic narrative review. *BMJ Open* **10,** e034774 (2020).

Chapter 3

Bedford, J. P., Garside, T., Darbyshire, J. L., *et al.* Risk factors for new-onset atrial fibrillation during critical illness: A Delphi study. *Journal of the Intensive Care Society* **23,** 414–424 (2022).

Chapter 5

Bedford, J. P., Drikite, L., Corbett, M., *et al.* Pharmacological and non-pharmacological treatments and outcomes for new-onset atrial fibrillation in ICU patients: The CAFE scoping review and database analyses. *Health Technology Assessment* **25,** 1–174 (2021).

Bedford, J. P., Johnson, A., Redfern, O., *et al.* Comparative effectiveness of common treatments for new-onset atrial fibrillation within the ICU: Accounting for physiological status. *Journal of Critical Care* **67,** 149–156 (2022).

Bedford, J. P., Redfern, O., Johnson, A., *et al.* Circadian variation in new-onset atrial fibrillation in patients in ICUs. *Journal of Critical Care* **67,** 1–2 (2022).

Drikite, L., Bedford, J. P., O'Bryan, L., *et al.* Treatment strategies for new onset atrial fibrillation in patients treated on an intensive care unit: a systematic scoping review. *Critical Care* **25,** 1–12 (2021).

Chapter 6

Bedford, J. P., Johnson, A., Redfern, O., *et al.* Comparative effectiveness of common treatments for new-onset atrial fibrillation within the ICU: Accounting for physiological status. *Journal of Critical Care* **67,** 149–156 (2022).

Bedford, J. P., Gerry, S., Hatch, R. A., *et al.* Hospital outcomes associated with new-onset atrial fibrillation during ICU admission: A multicentre competing risks analysis. *Journal of Critical Care* **60,** 72–78 (2020).

Chapter 7

Bedford, J. P., Ferrando-Vivas, P., Redfern, O., *et al.* New-onset atrial fibrillation in intensive care: epidemiology and outcomes. *European Heart Journal. Acute Cardiovascular Care* (2022).

Garside, T., Bedford, J. P., Vollam, S., *et al.* Increased long-term mortality following new-onset atrial fibrillation in the intensive care unit: A systematic review and meta-analysis. *Journal of Critical Care* **72** (2022).

Conference presentations

Bedford, J. P., Redfern, O. C., Clifton, D. A., Collins, G. S., Watkinson, P. J. Prediction of new-onset atrial fibrillation in the intensive care unit. Canada Critical Care Forum, Toronto, Canada (2022).

Bedford, J. P. Precision, prediction, and the need for quality data. Atrial Fibrillation in Acute and Critical Care Symposium, Toronto, Canada (2023).

Funding

This work was supported by an NIHR Doctoral Research Fellowship (NIHR300224).

Contents

List of Abbreviations **viii**

1 Introduction **1**
1.1 Atrial fibrillation 1
1.2 AF during critical illness 5
1.3 Motivation 8
1.4 Objectives 10
1.5 Contribution 12

2 Data and methods **14**
2.1 Data 14
2.2 Statistical methods 20

3 A Delphi study to identify candidate NOAF risk factors **41**
3.1 Background 41
3.2 Methods 42
3.3 Results 45
3.4 Discussion 46

4 Development of a NOAF prediction model **59**
4.1 Background 59
4.2 Methods 62
4.3 Results 75
4.4 Discussion 89

5 Characteristics and treatment of NOAF **100**
5.1 Background 100
5.2 Methods 103
5.3 Results 107
5.4 Discussion 112

6 Short-term outcomes after NOAF **121**
6.1 Background . . . 121
6.2 Methods . . . 124
6.3 Results . . . 127
6.4 Discussion . . . 131

7 Conclusions **136**
7.1 Thesis overview . . . 136
7.2 Parallel / future work . . . 141
7.3 Concluding remarks . . . 145

Appendices

A The resting membrane potential **148**

B A NOAF prediction model using admission-only data **150**
B.1 Introduction . . . 150
B.2 Methods . . . 150
B.3 Results . . . 152
B.4 Discussion . . . 155

C Supplementary figures **157**

D Supplementary tables **176**

References **188**

List of Abbreviations

aCHR . . . . . Adjusted cause-specific hazard ratio
aHR . . . . . Adjusted hazard ratio
AF . . . . . Atrial fibrillation
AFACS . . . . Atrial fibrillation after cardiac surgery
AKI . . . . . Acute kidney injury
APD . . . . . Action potential duration
APDR . . . . . Action potential duration restitution
APACHE . . . Acute physiology and chronic health evaluation
ARIC . . . . . Atherosclerosis risk in communities
AUROC . . . . Area under the receiver operating characteristic
AV . . . . . Atrioventricular
BP . . . . . Blood pressure
CANS . . . . . Cardiac autonomic nervous system
CCB . . . . . Calcium channel blocker
CCHIC . . . . Critical Care Health Informatics Collaborative
CCU . . . . . Coronary care unit
CHARGE . . Cohorts for Heart and Aging Research in Genomic Epidemiology
CI . . . . . Confidence interval
CITL . . . . . Calibration-in-the-large
COPD . . . . . Chronic obstructive pulmonary disease
CREDES . . . Conducting and reporting Delphi Studies
CRP . . . . . C polysaccharide-reactive protein
CSH . . . . . Cause-specific hazard
CSHR . . . . . Cause-specific hazard ratio
CSRU . . . . . Cardiac Surgery Recovery Unit
DBP . . . . . Diastolic blood pressure
ECG . . . . . Electrocardiogram
E_m . . . . . Membrane potential
ECG . . . . . Electrocardiogram
FiO_2 . . . . . Fraction of inspired oxygen
GAM . . . . . Generalised additive model
GBT . . . . . Gradient-boosted tree
HES . . . . . Hospital Episode Statistics
HIC . . . . . Health Informatics Collaborative

HR . . . . . Hazard ratio
HRV . . . . . Heart rate variability
I_f . . . . . Pacemaker current
I_{K1} . . . . . Inward rectifier potassium current
I_{Kr} . . . . . Rapid delayed rectifier K^+ current
I_{Ks} . . . . . Slow delayed rectifier K^+ current
I_{Kur} . . . . . Ultra-rapid delayed rectifier K^+ current
ICU . . . . . Intensive care unit
IECV . . . . . Internal–external cross-validation
INR . . . . . International Normalised Ratio
IPT . . . . . Inverse probability of treatment
IV . . . . . Intravenous
LASSO . . . . Least Absolute Shrinkage and Selection Operator
LOESS . . . . Locally estimated scatterplot smoothing
LOS . . . . . Length of stay
LR . . . . . Logistic regression
MAP . . . . . Mean arterial pressure
MI . . . . . Myocardial infarction
MAR . . . . . Missing at random
MIMIC . . . . Medical Information Mart for Intensive Care
ML . . . . . Machine learning
mV . . . . . Millivolt
MWR . . . . . Multiwavelet re-entry
NIHR . . . . . National Institute for Health and Care Research
NHS . . . . . National Health Service
NLR . . . . . Neutrophil:lymphocyte ratio
NOAF . . . . . New-onset atrial fibrillation
ONS . . . . . Office for National Statistics
OSA . . . . . Obstructive sleep apnoea
pAF . . . . . Paroxysmal atrial fibrillation
PEEP . . . . . Positive end-expiratory pressure
PH . . . . . Proportional hazards
PICRAM . . . Post-Intensive Care Risk-adjusted Alerting and Monitoring
RCT . . . . . Randomised controlled trial
ROC . . . . . Receiver operating characteristic
ROS . . . . . Reactive oxygen species
SA . . . . . Sino-atrial
SOFA . . . . . Sequential organ failure assessment
SBP . . . . . Systolic blood pressure
SDH . . . . . Subdistribution hazard
SDHR . . . . . Subdistribution hazard ratio
SpO_2 . . . . . Peripheral capillary oxygen saturation

STROBE . . . Strengthening the Reporting of Observational Studies in Epidemiology
TrAFFIC . . . Tracking Atrial Fibrillation aFter ICU
TRIPOD . . . Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
TTE . . . . . Transthoracic echocardiography
UK . . . . . United Kingdom
USA . . . . . United States of America
VIS . . . . . Vasoactive-inotropic score
$\mathbf{V_{rest}}$ . . . . . Resting membrane potential
WCC . . . . . White cell count

Introduction

1.1 Atrial fibrillation

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, affecting 3.3% of the United Kingdom (UK) population[1]. Prevalence increases with age, rising from 0.1% in those under 55 years of age to 9.0% in those aged above 80 years[2]. Men tend to develop AF five to ten years earlier than women on average. However, the lifetime cumulative risk becomes similar later in life, with a prevalence of around a quarter in both sexes in those who survive to 90 years of age[3,4].

AF is characterised by a degeneration of the atrial action potential resulting in a change from organised atrial electrical activity to low-amplitude rapid fibrillatory waves. The resulting ventricular rhythm is irregular, and often fast due to the disordered conduction of these fibrillatory waves through the atrioventricular (AV) node[5]. The electrocardiogram (ECG) in AF is therefore characterised by a lack of discrete P waves, low-amplitude fibrillatory waves, and an irregularly irregular ventricular rhythm[6].

AF is associated with a four- to five-fold increase in the risk of stroke[7]. AF-related embolism is predominately attributed to atrial thrombus formation resulting from reduced atrial flow velocity and impaired contractility. However, thromboembolism risk likely represents a more complex interplay of changes in blood flow, the vessel wall, and blood constituents[8]. Additional physiological changes associated with co-existent comorbidities act synergistically to further promote a hypercoagulable state. For example, concomitant left ventricular dysfunction or hypertension confer an additive thromboembolism risk through increased stasis, endothelial dysfunction, and changes in blood coagulation dynamics[9,10]. In anticoagulant-naive patients with AF, the prevalence of atrial thrombus is around 10 to 20%[11]. Anticoagulation

reduces the risk of atrial thrombus to around 3%[12] and reduces stroke risk by approximately two-thirds[13].

Further to thromboembolism risk, AF is also associated with a doubling of the risk of dementia and a tripling of the risk of heart failure[14]. After accounting for age and comorbidities, AF is associated with a 50% to 90% increased risk of death over a 40 year follow-up period[15]. AF also carries a significant economic burden, costing the NHS between 1.5 and 2.5 billion pounds per year (0.9% to 1.6% of NHS expenditure)[16].

1.1.1 Mechanisms for AF initiation

The atrial action potential

Human atria have features that may contribute to a propensity to arrhythmogenesis. The atrial action potential morphology is more triangular and has a less prominent plateau phase than its ventricular counterpart (Figure 1.1a). The atrial action potential duration (APD) is highly variable, but tends to be shorter than the ventricular APD[17]. Atrial myocytes share depolarising currents with ventricular myocytes, including the fast sodium current I_{Na} and the and L-type Ca^{2+} current. However, certain differences exist, including the presence of the ultrarapid delayed rectifier current (I_{Kur}), which accelerates atrial repolarisation[18]. Additionally, the transient outward K^+ current (I_{to}) is more abundant in atrial than ventricular myocytes, contributing to rapid early repolarisation and the shorter atrial APD[19]. The atrial resting membrane potential (E_m) varies between -65 and -80 millivolts (mV) and is more depolarised than that in the human ventricle[20].

Electropathology

The concept of electropathology suggests that many pathophysiological processes lead to molecular changes in atrial tissue, which in turn precipitate structural changes. These structural changes then result in electrical conduction disorders[5]. The severity of electropathology correlates with AF risk. Contributing mechanisms

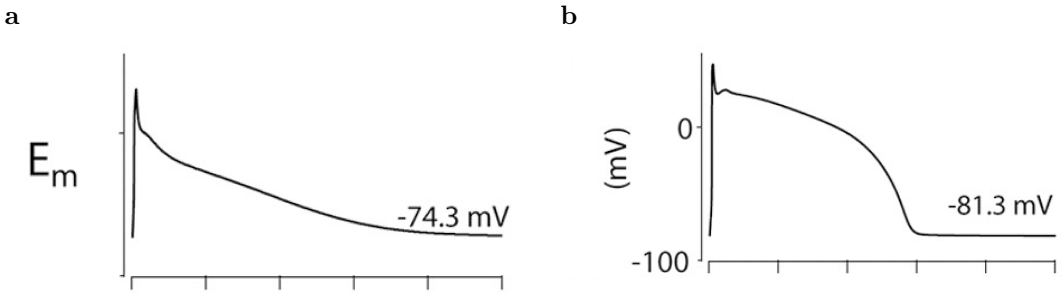


Figure 1.1: Action potentials in a) atria and b) ventricles demonstrating the higher V_{rest} and slower repolarisation of the atrial action potential. Reproduced with permission[21] from Amos *et al.* 1996[20]. mV = millivolts, E_m = membrane potential.

include disordered protein homeostasis and cellular stress signalling secondary to certain comorbidities[22,23].

With increasing electropathology, the electrical properties of atria exhibit increasing regional heterogeneity. This includes differences between the left and right atria[24], within one atrium[25], and between the endocardium and epicardium[26]. There are also important dynamic properties exhibited during atrial repolarisation and conduction that are affected by electropathology. The APD changes with the diastolic interval—the time between the end of the T wave and the beginning the next QRS complex[27]. The relationship between the preceding diastolic interval and the subsequent APD is described by APD restitution. Shorter diastolic intervals result in shorter APDs[28]. The atrial action potential's ability to adjust to changes in rate is in many ways protective, reducing the risk of conduction block if rate increases, and preventing early re-excitation when the rate slows. However, electropathology is associated with heterogeneity in APD restitution[17].

Spontaneous AF initiation often results from focal ectopy originating within the pulmonary veins[29]. Such pulmonary vein "triggers" are thought to be primarily responsible for paroxysmal AF episodes, and pulmonary vein isolation has become a common AF ablation strategy[30]. In the context of electropathology, focal ectopy may result in large oscillations in atrial APD ("APD alternans") which can be sufficiently discordant to precipitate disorganisation in wavefront propagation, ultimately initiating AF. The tendency to these marked oscillations can be inferred

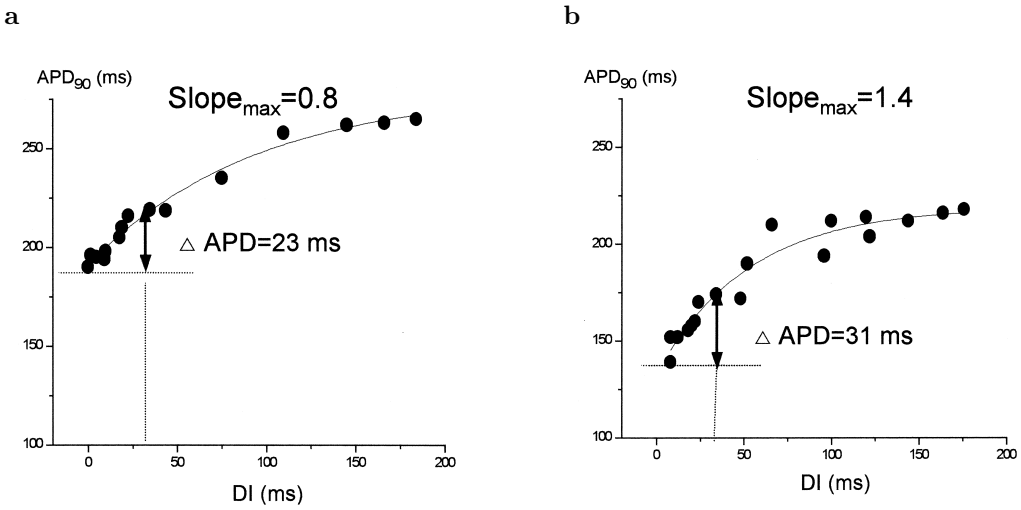


Figure 1.2: Action potential duration restitution curves in a) control subjects and b) patients with paroxysmal atrial fibrillation. DI = diastolic interval. APD_{90} = action potential duration at 90% repolarisation. Reproduced with permission[32] from Kim *et al.* 2002[31].

from the slope of the APD restitution curve, where APDs are plotted against the preceding diastolic interval (Figure 1.2a). Where the slope is steeper, APD will change more for any given change in diastolic interval (Figure 1.2b). APD restitution slopes of greater than one are associated with AF[31], and slopes greater than one near the pulmonary veins enable premature atrial contractions to initiate AF[28].

1.1.2 Mechanisms for AF maintenance

The mechanisms that underlie the maintenance of AF are less clear than those which may initiate it. The conventional paradigm is that AF maintenance relies on atrial electrical and structural remodeling where underlying atrial fibrosis provides a substrate for persistent spatially disorganised electrical activity[33]. The exact mechanisms of sustained AF are the subject of ongoing study and there are two potentially competing hypotheses: focal drivers and multiwavelet re-entry (MWR).

The focal driver hypothesis posits that a stationary driver generates waves that propagate to the surrounding tissue in a non-uniform fashion. The nature of the drivers is contested but may include *rotors* (where spiral waves rotate around an origin—Figure 1.3a) and *focal sources* (where activation spreads away from an

origin—Figure 1.3b). Others assert that endo-epicardial (longitudinal) dissociation—consisting of lines of conduction block running parallel to the atrial musculature—is responsible for the appearance of apparent drivers[34] (Figure 1.3c).

In contrast, the MWR hypothesis refers to self-sustaining short-lived spiral waves, creating functional re-entrant circuits that do not require the presence of a focal driver[35] (Figure 1.3d). The MWR hypothesis does not rely on discrete areas of atrial tissue which may harbour focal sources. Consequently, these wavelets may self-terminate on encountering a non-conducting tissue boundary. AF duration in this hypothesis is therefore a function of the rate of wavelet creation and extinction. This hypothesis may explain observations such as why AF is more difficult to treat in patients with larger atria (i.e. because there is more atrial space for wavelets to propagate and AF is therefore less amenable to treatment).

The different proposed mechanisms of AF maintenance are not necessarily mutually exclusive. Focal drivers and MWR may coexist, and continuously interact with one another. The degree to which a focal driver and the surrounding chaotic activation interact depends on the relative characteristics of the structural and functional re-entrant substrates[36]. There is no clear consensus on the mechanisms of AF maintenance. Further work is necessary to investigate these, and it is likely that high-fidelity mapping tools will be required to clarify the complex electrical conduction disorders involved[37].

1.2 AF during critical illness

New-onset AF (NOAF), defined as AF occurring in a patient with no known history of chronic or paroxysmal AF[38] is common in patients admitted to an intensive care unit (ICU). Studies detecting NOAF through chart review suggest it occurs in around 7% of patients admitted to a general ICU, compared to around 14% in studies using continuous monitoring[39]. Incidence is higher in certain subgroups such as those with septic shock where it approaches 25%[40].

NOAF may have immediate deleterious haemodynamic effects but these have not been studied in detail[41]. It is also possibly associated with thromboembolic

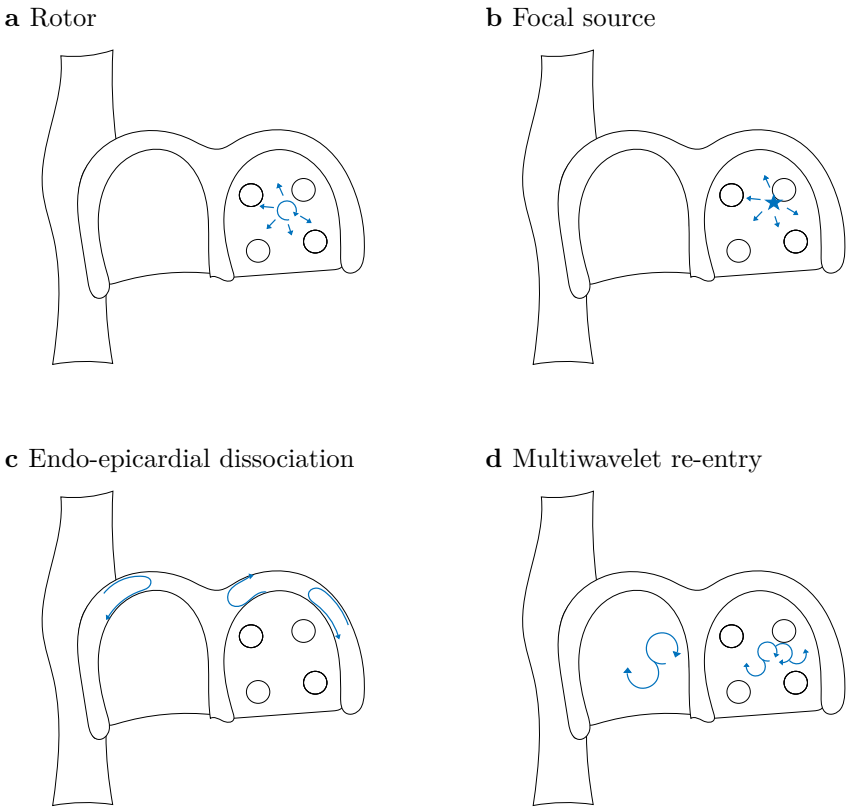


Figure 1.3: Proposed mechanisms of atrial fibrillation. a) Rotor: spiral waves of electrical activity circulate around an origin in the atrial tissue; b) Focal source: Rapid, uncontrolled activation spreads from a single point or a few points in the atria; c) Endo-epicardial dissociation: Differences in conduction between the endocardial and epicardial layers result in unstable electrical activity across the atria; and d) Multiwavelet re-entry: Numerous electrical wavelets meandering through the atria, constantly splitting and recombining, creating functional re-entrant circuits. Image: Author's own.

events, but again evidence is limited[42,43]. It is associated with increased ICU and hospital mortality[40,44,45], and healthcare costs[46]. Despite its impact, little is known about modifiable risk factors[38]. Furthermore, optimal treatment of NOAF during an ICU admission is unclear[39].

1.2.1 Mechanisms promoting AF during critical illness

In critically ill patients, multiple mechanisms encourage electropathology, promoting NOAF. Atrial remodelling can occur rapidly in the context of critical illness[47]. This remodelling alters the mechanical structure and electrical function of the atria

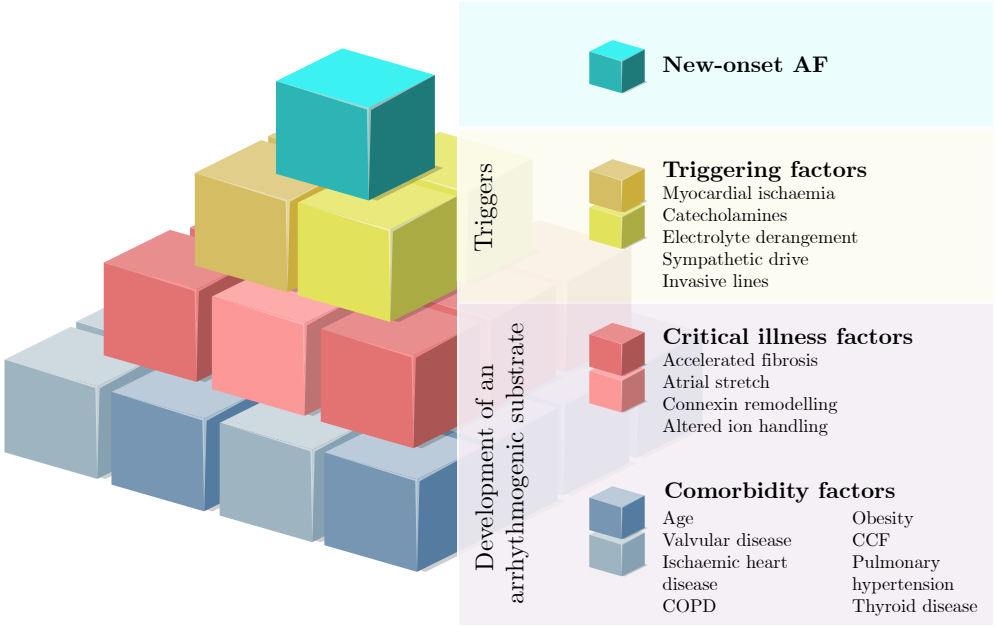


Figure 1.4: Potential mechanisms for new-onset atrial fibrillation (NOAF) during ICU admission. An arrhythmogenic substrate (a structurally and electrically remodelled myocardium) can develop prior to acute illness or surgery due to ageing and comorbidities. Further accelerated remodelling can then occur during critical illness or in the perioperative period. Electrical remodelling (changes in ion handling and connexin expression) occurs alongside structural remodelling (atrial stretch and fibrosis). Factors such as high sympathetic drive, fluid overload and systemic inflammation promote remodelling such that on arrival to the ICU, there is an elevated predisposition to NOAF. Arrhythmogenic potential will depend on the combined effects of comorbidities and acute illness-related factors on the myocardium and will vary considerably from patient to patient. Each factor is not uniquely required, but adds to an accumulated risk. Exposure to triggering events may then initiate AF through multiple mechanisms. AF is sustained due to underlying remodelled tissue. CCF = congestive cardiac failure, COPD = Chronic obstructive pulmonary disease, AF = atrial fibrillation. Reproduced with permission[49] from Bedford *et al.* 2021[48].

(the substrate) such that its propensity for development and maintenance of AF is increased (an arrhythmogenic substrate). Inflammation, electrolyte disturbances, and sympathetic overactivity contribute significantly to this remodelling. Moreover, factors such as hypoxaemia and alterations in haemodynamics can further exacerbate the risk of NOAF[48]. These processes collectively create a fertile environment for the development of NOAF in critically ill patients. These mechanisms are summarised in Figure 1.4 and explored further in Chapter 3.

Prevention of ICU-acquired AF

AF in critically ill patients differs from AF in the general population with respect to risk factors, incidence, and clinical course[50,51]. In the community, AF is associated with chronic conditions such as hypertension and diabetes[2]. However, the development of NOAF on the ICU is more strongly linked with acute factors than pre-existing conditions[40,50,52]. Modifiable risk factors may include serum electrolyte concentrations, an inflammatory state, invasive line position and the use or avoidance of certain medications, but this requires more study. There are therefore no evidence-based preventative treatments for patients in general ICUs.

Prevention of NOAF after cardiothoracic surgery has been investigated, highlighting some pharmacological and non-pharmacological interventions that appear to reduce risk[53]. However, interventions used after cardiothoracic surgery are unlikely to be directly transferable to the general adult ICU patient as the underlying mechanisms of AF development differ[54].

To understand what is known about risk factors for ICU-acquired NOAF, I performed a systematic review of relevant studies prior to commencing my doctoral studies[38]. This was the first evidence synthesis of the risk factors for NOAF in the general adult ICU population. Previous systematic reviews focused on patients with sepsis[40] or provided no evidence synthesis[55]. These reviews also included studies where patients with a chronic history of AF were included in the NOAF group.

1.3 Motivation

The studies identified in the systematic review were mostly single-centre and small, with half containing fewer than 100 patients in the AF group[38]. Although several potential modifiable risk factors were flagged, none of the studies were adequate to clearly determine whether an association exists. Furthermore, existing studies of NOAF risk factors often use variables collected at ICU admission as predictors, but not data collected during a patient's stay[56,57].

As NOAF antecedents have been poorly studied, we do not have reliable prediction tools that could be used to focus attention on avoiding NOAF in those at risk. Developing a tool to identify patients at high risk of developing NOAF is vital for future studies investigating prevention strategies to allow stratification, sample enrichment and efficient study design.

Management guidelines for AF exist but the underlying data from which they are derived are not from patients treated on an ICU[58–60]. NOAF in patients on an ICU differs from AF in studies informing current guidance in the causes of rhythm disturbance[44,52], effectiveness of treatments applied[61], relative risks of those treatments[62,63] and long-term outcomes[44,64]. As a result, these general guidelines are not useful in critically ill patients, where the risks and benefits associated with different treatment options remain unclear.

Prior non-randomised studies have adjusted for admission characteristics when comparing treatment groups, but not for physiological status after NOAF[65–68]. As the haemodynamic changes caused by NOAF may affect initial treatment choice, failure to adjust for these may confound the association between treatment and outcomes[69].

Understandably, there is significant variation within and between units in the management of this common problem[70,71], demonstrating the need for research to understand the effects of current practice, from which future trials can be designed.

Organised atrial activity is important for ventricular filling and cardiac output[72]. NOAF causes a reduction in cardiac output in non-ICU patients[73]. The haemodynamic effects may be more severe in patients admitted to an ICU, as they are already unwell. However, the severity and time course of the haemodynamic effects of NOAF have not been studied in this setting. Understanding how NOAF affects patients' haemodynamics while critically unwell is important. It may improve patient care by arming healthcare staff with an awareness of anticipated changes in patient status, for example changes in blood pressure or vasoactive medication use after AF onset. Understanding these haemodynamic changes will also increase

awareness of what a "normal" change after AF looks like, helping to highlight patients who may require extra interventions.

Existing findings from studies investigating the impact of NOAF on patient outcomes are susceptible to immortal time bias. Not accounting for the time-varying nature of NOAF can result in an under-estimation of associated mortality risk (expanded further in Section 6.1.2). Furthermore, NOAF in the ICU is commonly transient[74], yet it is unclear whether its impact on patients is related to how long the AF lasts.

1.4 Objectives

This thesis is structured to follow the chronology of a patient's ICU admission.

It begins with a summary of the data sources and methods used in this thesis. Risk factors are then explored in detail in the context of a consensus exercise to identify key NOAF predictors. These risk factors are then used to develop a NOAF prediction model. This is followed by an analysis of optimal treatments once NOAF occurs. I then outline the haemodynamic changes associated with NOAF, and subsequent impact on in-hospital mortality.

The objectives are as follows:

1: To systematically identify risk factors for ICU-acquired NOAF by
   (a) expanding the list of known and potential risk factors for NOAF from my existing literature review using a Delphi process, and
   (b) assessing the associations between these predictors and the development of NOAF.

2: To develop and validate a NOAF prediction model by
   (a) determining which of the variables identified in objective 1 to include in a logistic regression model to estimate the probability of NOAF using standard approaches,
   (b) investigating non-linear machine learning approaches to developing a predictive model,
   (c) assessing the relative performance of the linear and non-linear models, focusing on discrimination and calibration, and
   (d) assessing the generalisability of the models by assessing their performance in an external validation cohort.

3: To compare the use and effectiveness of NOAF treatments by
   (a) exploring which drugs are used to treat ICU-acquired NOAF, and
   (b) analysing rate control, rhythm control, and hospital mortality associated with common NOAF treatments after balancing admission and post-NOAF, pre-treatment covariates across treatment groups.

4: To describe the haemodynamic changes associated with ICU-acquired NOAF by
   (a) using descriptive statistics to identify the changes in haemodynamic variables after the onset ICU-acquired NOAF, and
   (b) using multilevel regression methods to determine the impact of NOAF on these parameters while accounting for repeated measures.

5: To investigate the outcomes associated with ICU-acquired NOAF by
   (a) performing multivariable survival analysis to determine the impact of NOAF on hospital survival, and
   (b) analysing the effect of AF duration on hospital outcomes.

The outline of this thesis, and how it relates to the above objectives, is summarised in Figure 1.5.

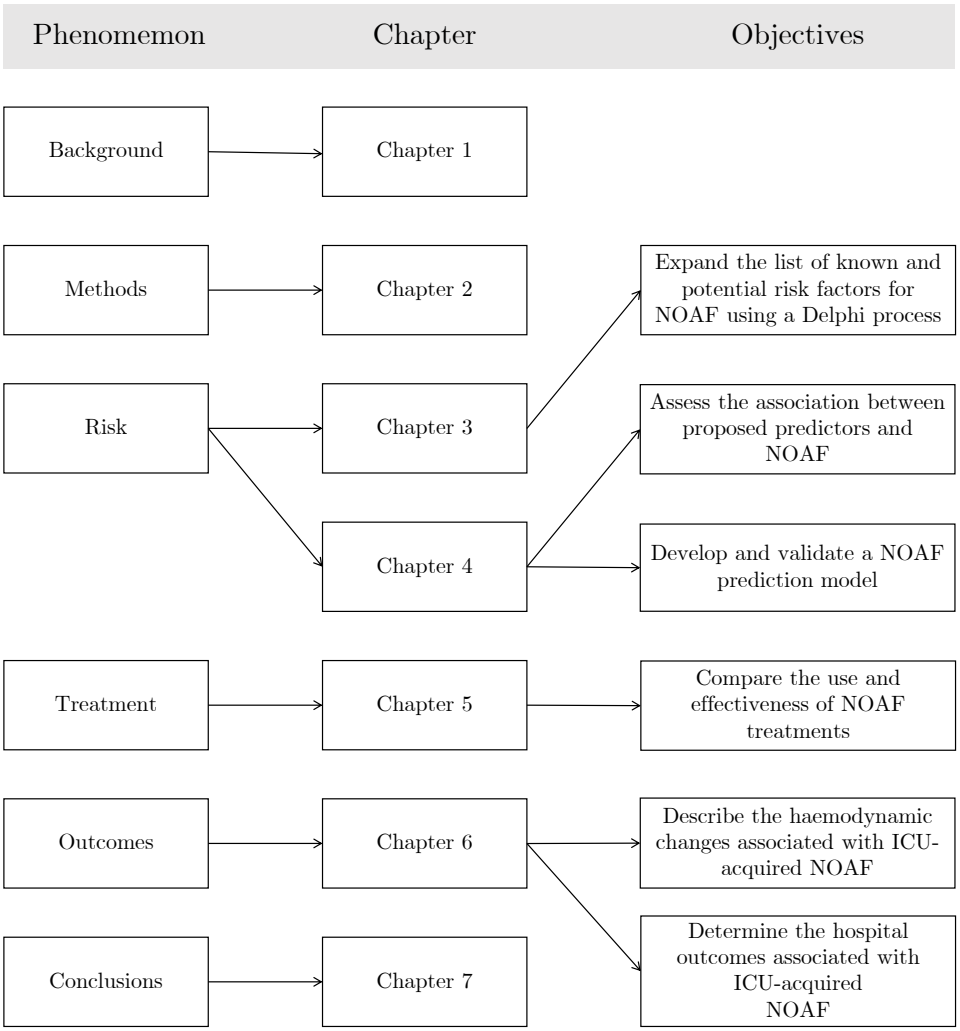


Figure 1.5: Structure of thesis.

1.5 Contribution

I confirm that the work presented in this thesis is my own. Where content has been derived from my published work, this is detailed below, with footnotes in the relevant chapters. Where images have been reproduced from others' work, this has been done so with permission, with relevant licensing cited at the point of use.

This thesis is a collaborative endeavor. The role of collaborators are detailed below, alongside my personal involvement. Other than these contributions, I completed all work independently, with guidance from my supervisors and mentors where appropriate.

Chapter 3

Segments of the introduction, methods, results, and discussion sections of this chapter have previously been published[75]. I was first author of this publication, authored the text, and created the figures that are reproduced herein. Content reproduced with permission[49].

Chapter 5

My work in this chapter formed part of a wider National Institute for Health and Care Research (NIHR) study (The CAFE Study)[76]. A colleague Dr Alistair Johnson extracted and cleaned data from the MIMIC-III database. I worked with a colleague (Laura Drikite) to undertake a scoping review of NOAF treatments to identify candidate treatments for analysis which was overseen by an expert in scoping reviews (Dr Mark Corbett). Segments of the introduction, methods, results, and discussion sections of this chapter have previously been published[76–78]. I was first author of these publications, authored the text, and created the figures that are reproduced herein. Content reproduced with permission[32].

Chapter 6

For Section 6.2.3, Dr Oliver Redfern provided initial R code samples to guide my learning in quantile regression analysis. Segments of the introduction, methods, results, and discussion sections of this chapter have previously been published[77,79]. I was first author of these publications, authored the text, and created the figures that are reproduced herein. Content reproduced with permission[32].

Data and methods

2.1 Data

Work contained in this thesis uses four databases of routinely collected healthcare data. Data sources were selected based on their appropriateness to address the range of my research questions.

To review data quality and comparability, distributions of key vital signs in each data source were graphically assessed and are shown in Appendix C.

2.1.1 The PICRAM database

The PICRAM database[80] is used in Chapters 4, 5, and 6. It comprises data from 13,882 admissions to three general ICUs in the UK: The John Radcliffe Adult ICU, Oxford; The Churchill ICU, Oxford; and The Royal Berkshire Hospital ICU, Reading from 2008 to 2015. Of these, 1,612 are re-admissions (non-index admissions), leaving 12,270 index admissions representing individual patients. The PICRAM database contains no children (aged less than 16 years) nor patients cared for by cardiac surgical teams, therefore all 12,270 patients remained for further study-specific exclusions and analysis detailed in Chapters 4, 5 and 6.

In 2008, only data from patients from the John Radcliffe Adult ICU is available, with the addition of data from the Churchill ICU in 2009. From 2010, data from patients admitted to all three sites is available in the PICRAM database. The number of patients by ICU admission year is shown in Figure 2.1. Plots of key demographic variables (age, sex, ICU and hospital length of stay (LOS), and ICU and hospital mortality) in the PICRAM database are shown in Figure C.1..

Legal and ethical basis for use

The PICRAM study was granted ethical approval by the National Research Ethics Service Committee Oxford (ref: 11/SC/0440) and the National Information Governance Board (ref: ECC 7–05(f)/2011). Use of patient data without informed consent was approved by the National Information Governance Board on 2nd February 2012 (ref: ECC 7-05(f)/2011). The use of this database for my Doctoral studies was approved by the University of Oxford Critical Care Research Group Data Access Committee on 20th September 2019.

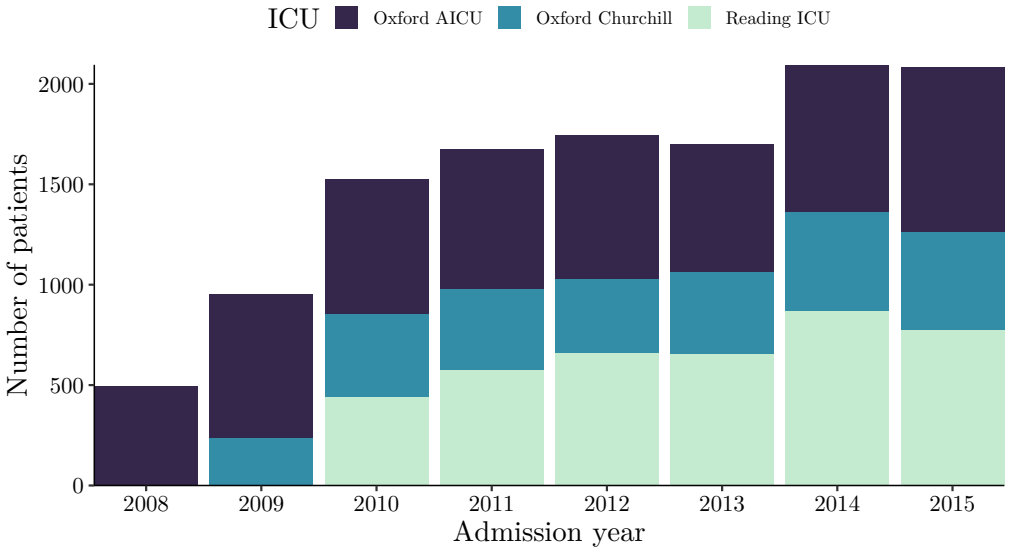


Figure 2.1: Bar chart of patient numbers by year group: PICRAM database. ICU = intensive care unit, AICU = adult ICU.

2.1.2 The MIMIC-III database

The MIMIC-III database[81] is used in Chapters 5 and 6. It comprises data from 61,532 admissions to critical care units at a tertiary care hospital in Boston, United States of America (USA) between 2001 and 2012. Of these, 15,056 are repeated (non-index) admissions, leaving 46,476 index admissions representing individual patients. Unlike the PICRAM database, MIMIC-III includes data from children and patients cared for by a cardiac surgical team. After excluding non-adult patients, patients cared for by a cardiac surgical team, and those without any chart data,

22,684 patients across five ICUs remained for further detailed exclusion and analysis in Chapter 5. These exclusion steps are outlined by ICU in Table 2.1.

To increase anonymity, exact patient admission year is not available, but rather grouped into one of two clusters: 2001 to 2007, and 2008 to 2012. The number of patients by ICU admission year group is shown in Figure 2.2. A graphical summary of patient demographics is shown in Figure C.2. Distributions of key vital signs are shown in Figure C.6.

Legal and ethical basis for use

The MIMIC-III database and associated use of patient data was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, Massachusetts) and the Massachusetts Institute of Technology (Cambridge, Massachusetts).

Table 2.1: MIMIC-III patient numbers after initial exclusions

Unit	Admissions	Index admissions	Index adult admissions	Non-cardiac	Valid data
CCU	7726	5674	5674	1364	1296
CSRU	9312	7612	7611	715	642
MICU	21088	13650	13649	12822	12558
NICU	8100	7870	0	0	0
SICU	8891	6372	6372	5901	5795
TSICU	6415	5298	5291	2464	2393
Total	61532	46476	38597	23266	22684

CCU = Coronary care unit, CSRU = Cardiac surgery recovery unit, MICU = Medical Intensive Care Unit (ICU), NICU = Neonatal ICU, SICU = Surgical ICU, TSICU = Trauma surgical ICU.

2.1.3 The MIMIC-IV database

The MIMIC-IV database[82] is used in Chapter 4 for prediction model development. It provides more recent patient data than MIMIC-III, comprising data from patients admitted to the same centre, but between 2008 and 2019 (rather than between 2001 and 2012). Due to the included dates, there is some currently undefined overlap in patient data between MIMIC-III and MIMIC-IV. Therefore these two databases are not combined nor analysed together in this thesis.

The MIMIC-IV database comprises data from 76,540 adult admissions. Of these, 23,390 are repeated (non-index) admissions, leaving 53,150 index admissions

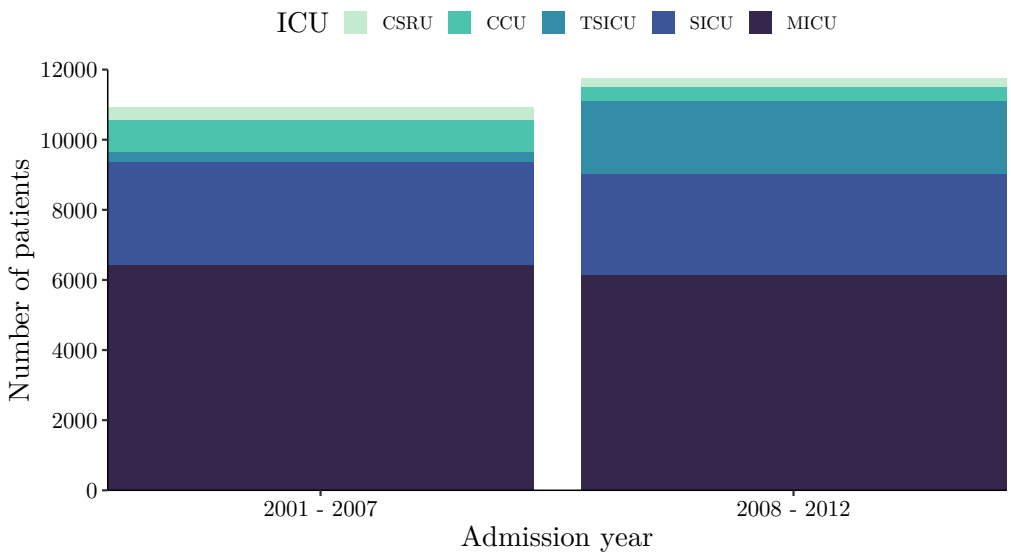


Figure 2.2: Bar chart of patient numbers by year group: MIMIC-III database. Exact admission year is not available in the MIMIC-III database due to deliberate obfuscation to reduce data identifiability. CSRU = Cardiac surgery recovery unit, CCU = Coronary care unit, TSICU = Trauma surgical intensive care unit (ICU), SICU = Surgical ICU, MICU = Medical ICU.

representing individual patients. For the MIMIC-IV analysis, I used location-dependent exclusions, excluding patients cared for in service-specific ICUs. This contrasts with the MIMIC-III analysis, where I included outliers on cardiac surgical units. I took this decision to ensure that my prediction models were developed using data from—and therefore more applicable to—patients admitted to general ICUs.

After excluding non-adult patients, patients admitted to a cardiac surgical / neurosurgical unit, and those without any chart data, 32,529 patients remained for further detailed exclusion and analysis in Chapter 4.

The number of patients by ICU admission year group is shown in Figure 2.3. A graphical summary of patient demographics is shown in Figure C.3. Distributions of key vital signs are shown in Figure C.7.

Legal and ethical basis for use

The MIMIC-IV database and associated use was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center, Boston, Massachusetts and the Massachusetts Institute of Technology, Cambridge, Massachusetts (2001P001699).

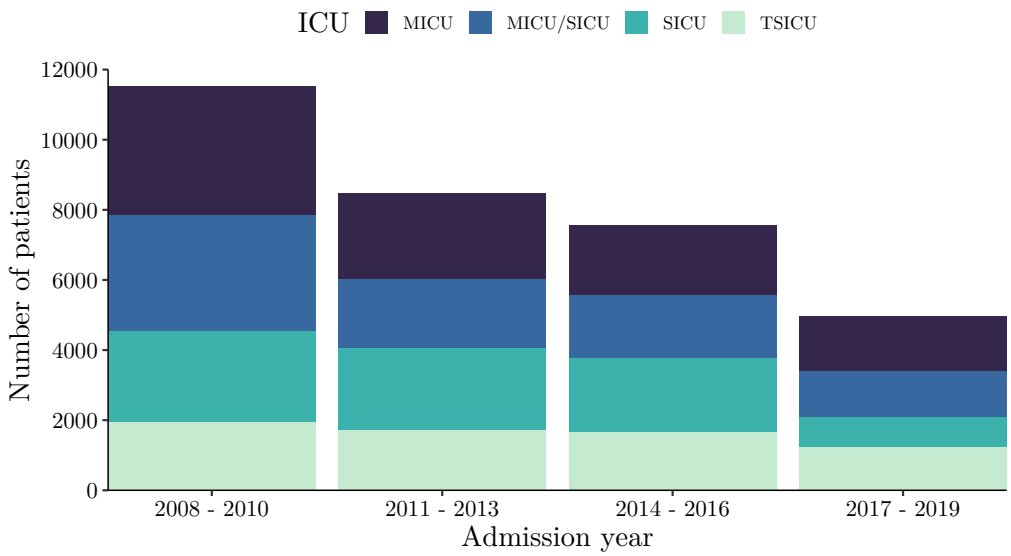


Figure 2.3: Bar chart of patient numbers by year group: MIMIC-IV database. Exact admission year is not available in the MIMIC-IV database due to deliberate obfuscation to reduce data identifiability.

2.1.4 The CCHIC database

The NIHR Health Informatics Collaborative (HIC) data resource represents a collaboration between five National Health Service (NHS) Trusts and their associated Biomedical Research Centres. The collaboration comprises University College Hospitals NHS Foundation Trust, Imperial College Healthcare NHS Trust, Cambridge University Hospitals NHS Foundation Trust, Oxford University Hospitals NHS Foundation Trust, and Guy's and St. Thomas' NHS Foundation Trust. The Critical Care Health Informatics Collaborative (CCHIC) database is used for external validation of the NOAF prediction model developed in Chapter 4.

The CCHIC database[83] was created with retrospectively collected data from the ICU clinical information systems in these centres from 2014 to 2018. The Internal Collaborator Release 2018-09-14 used in this work contained no data from Oxford University Hospitals NHS Foundation Trust and therefore had no crossover with the PICRAM database.

The CCHIC database contains 36,741 episodes, where an episode is defined as a continuous period of level 2/3 care in a single physical location. The database contains data relating to 33,844 "spells", defined as a continuous period of level

2/3 care, regardless of physical location. Depending on data source, these spells may have reconciled multiple care "episodes" in the case where a new episode was started for each patient as they transitioned between physical ICU sites, e.g. from an ICU to an HDU, or between two ICUs within the same hospital. A graphical summary of patient demographics is shown in Figure C.4.

Legal and ethical basis for use

The legal basis for transferring data was provided to those who created the CCHIC dataset under section 251 of the National Health Service Act 2006 (Confidentiality Advisory Group reference 14/CAG/1001). Ethical approval was granted by a Health Research Authority Research Ethics Committee (14/LO/1031). Use of the CCHIC database (Internal Collaborator Release 2018-09-14) for this work was registered and approved by the CCHIC scientific advisory group in September 2018.

2.2 Statistical methods

In this section, a summary of the statistical methods used in this thesis is provided. Details of how these methods are applied to achieve the research objectives are presented within their respective chapters.

2.2.1 Estimating NOAF risk using predictive modelling

Prediction modelling is concerned with predicting an outcome on the basis of some variables or predictors. The relationship between the predictors (x) and the outcome (Y) consists of a systematic component $f(x)$ and a random component ϵ. f is an unknown function which can be estimated using various model fitting techniques using development data. There are myriad forms of f employed across different modelling techniques. The random component is assumed to be independent of x and has mean zero in the model. Many models can be therefore be expressed as $Y = f(x) + \epsilon$. By building clinical prediction models we are attempting to find a good approximation of $f(x)$ in order to make clinically meaningful predictions.

Modelling techniques are often classified into traditional statistical models and "machine learning" (ML) models. The boundary between the two is unclear[84]. Historically, while statistical models typically focused on understanding relationships and providing measures of association, ML models often favoured heuristics and useful approximations[85]. As such, a stated limitation of ML models is their lower comparative interpretability. However, modern ML techniques can provide measures of confidence in associations and many are highly interpretable, making the statistical versus ML model distinction less useful.

In Chapter 4, a logistic regression (LR) model, a gradient-boosted tree (GBT), and a neural network are developed to predict the risk of NOAF in patients admitted to an ICU. Model types were selected for their established use and efficacy in outcome prediction in the ICU[86–88]. The performance of these models are compared. These model types are outlined below.

Traditional statistical models

Logistic regression

Data with a binary outcome can be analysed with LR models. These predict the probability that the dependent variable $Y = 1$ given the values of the independent variables. The logistic (or inverse logit) function $f(x) = \frac{1}{1+e^{-x}}$ can take any value of x and maps it to a corresponding y value between 0 and 1. The logit function $y = log(\frac{x}{1-x})$ is the inverse of the logistic function and maps inputs ranging from zero to one to any value ranging from $-\infty$ to $+\infty$. An LR model is a linear regression of the logit of the response; replacing x with P (probability) an additive LR model (no interactions) has the form:

$$log(\frac{P}{1-P}) = \beta_0 + \beta_1 X_i + \cdots + \beta_k X_k$$

where:

P	is the probability that the dependent variable will equal one
$X1, X2, \ldots, X_k$	are the independent variables
β_0	is the model intercept
$\beta_1 \ldots, \beta_k$	are the coefficients estimated by the model

An LR model in most software is fit analytically using a maximum likelihood estimation, e.g. Iteratively Reweighted Least Squares. Inferences can be drawn from model coefficients about the magnitude and direction of associations with outcome.

Machine learning techniques

The term "machine learning" encompasses a wide variety of modelling approaches. The following section outlines two techniques used in this thesis considered under the ML umbrella—GBTs and neural networks—with a brief introduction into how these models are fit.

Gradient descent

Where linear statistical models are often optimised analytically e.g. using a least-squares approach, many ML models are fit using "gradient descent". Gradient descent is an iterative algorithm that is used to find the minimum of a function, typically a loss function in the context of ML. The loss function measures prediction error, and the goal is to minimise this error. Two or more derivatives of the same

function are known as a gradient. This gradient can be used to descend to the lowest point in a loss function, minimising error, thus *gradient descent*. The magnitude of the gradient is proportional to the rate of change of the function.

The gradient descent algorithm begins by initialising the parameters of the model to some starting values. It then computes the gradient of the loss function with respect to the model parameters, and takes a step in the direction of the negative gradient. The step size is calculated as the slope multiplied by the learning rate η. The learning rate therefore controls the size of the step, and is user-defined. If the steps are too small, a large number of iterations may be required before arriving at the minimum, possibly converging on a local minima. If the steps are too large, the algorithm may overshoot, resulting in oscillations around the minimum. New parameter values are then calculated by subtracting the step size from the previous parameter values. The algorithm repeats this process, iteratively updating the parameters until the improvements become negligibly small or a set number of iterations have been completed.

Gradient boosting

While gradient descent is an algorithm for finding a set of parameters that optimises a loss function, gradient boosting is a technique for building an ensemble of "weak" (slightly better than a random guess) models such that the predictions of the ensemble minimise the loss function. Gradient boosting can be thought of as an iterative functional gradient descent algorithm. That is, algorithms that optimise a global loss function by iteratively choosing a function that points in the negative gradient direction.

Decision tree models

A decision tree prediction model is fitted by creating a tree-like model of decision splits to generate predictions. The model starts by considering the entire dataset, and then splits it into smaller subsets using a process called recursive partitioning. In recursive partitioning, the data is successively split by criteria (decisions) on a single variable which best splits the data. The top of the tree is called the root and

the tree then grows down from the root splitting at each internal node, terminating with "leaf" nodes, which represent predictions.

There are two main types of decision tree models: classification trees, in which the outcomes are categorical and the terminal nodes represent predicted classes, and regression trees, in which the outcomes are continuous measures and the terminal nodes represent predicted values.

Gradient-boosted trees

GBTs are a decision-tree-based modelling method. They are an ensemble method, meaning multiple individual tree-based models are combined. Using gradient-boosting, multiple weak decision trees are iteratively generated. They are often simple, having few branches or only one, resulting in a "decision stump". Each subsequent tree is fitted to reduce the error of the prior ensemble.

The complexity of GBT models can be modulated by adjusting the number of trees in the ensemble, and the depth of those trees. Increasing the number and depth of trees can increase the model's ability to capture complex patterns in the data. As complexity increases, the model becomes more computationally intensive, which can slow down the training process. Additionally, a larger number of trees can increase the risk of overfitting (see below).

Neural networks

A neural network is a ML model that is composed of interconnected units which process and transmit information using weighted connections. A neural network has one or more input nodes (the covariate / feature values), one or more output nodes (often a predicted risk or class), and layers of nodes between the input and output nodes known as "hidden layers".

Neural networks are composed of building blocks known as neurons. In a neuron, each input value is multiplied by a *weight*. Weights are analogous to model coefficients in a linear model. A summation function then combines the weighted inputs, and a *bias* is added to the summed value, the bias being analogous to a linear model intercept. Summation functions are often a simple weighted sum, but

may be more complex. This sum is then passed to an *activation function*, acting as the x axis coordinate. The corresponding y axis value from the activation function can then be fed into the next hidden layer, incorporated into the output value.

Optimal parameters (weights and biases) are estimated during model fitting using a technique known as backpropagation. Backpropagation works backwards from the output layer to the input layer. Weights and biases are adjusted at each layer in proportion to the error that was propagated from the previous layer. The adjustments are usually made though gradient descent.

The complexity of a neural network can be increased by increasing the number of hidden layers, and the number of nodes per hidden layer. Other hyperparameters that can be tailored include the activation function, the learning rate (see Section 2.2.1), and regularisation strategies to avoid overfitting (see Section 2.2.1).

Improving model generalisability

The more complex a model becomes, the more accurately it can capture the underlying relationships in the development data. However, as model complexity increases, the more it will tend to capture noise in the development data rather than population-level associations between variables. Such models are then likely to underperform in external data and are said to have "overfitted". Conversely, overly simple models may fail to capture true relationships, resulting in lower predictive performance ("underfitting"). Therefore, there is a balance to be struck between building a model that is too simple and building a model that is too complex. This is known as the Bias-Variance trade off.

Variance refers to the amount that f changes when different development data is used. Small changes to the development data can result in large changes to f with high variance (complex) methods. Bias is a measure of how much the predictions made by the model differ from the true values. Methods that make simple assumptions about the underlying distribution of the data (e.g. linear regression) are often high bias and low variance approaches.

The bias-variance trade-off refers to this tension between two conflicting goals: building a model that accurately captures the underlying relationships in the data (low bias) and building a model that is not sensitive small fluctuations in the development data (low variance).

Regularisation is the term used to encompass the multiple methods available to limit complexity and reduce overfitting when building a predictive model. There are numerous regularisation techniques. I outline those used in this thesis here.

Adjusting for model optimism

One common technique used to adjust traditional statistical models after model development is known as shrinkage. This technique uses bootstrap resampling[i] or cross-validation to calculate model "optimism"—a measure of how its apparent performance on the development data would be reduced when tested on unseen data. Optimism calculated from the resampled calibration slope (see below) can then be used to "shrink" (reduced in magnitude) model coefficients resulting in an "optimism-adjusted" model.

I use this technique to develop the admission-only model in Appendix B.

L1 and L2 regularisation

Other regularisation techniques include L1 (also known as Least Absolute Shrinkage and Selection Operator—LASSO) and L2 (also known as Ridge) regularisation. In these techniques, as with uniform shrinkage, coefficients from a linear model are shrunk by a penalisation factor.

In L1 regularisation, the penalty term is equal to the absolute sum of the coefficients. This penalty term is modulated by a shrinkage parameter λ. When $\lambda = 0$, this equates to no penalty, and the model coefficients will be unchanged. As λ increases, model coefficients will be shrunk, and can be shrunk to zero, i.e. the predictor is removed from the model. L1 regularisation can therefore be used as a

[i]Bootstrap sampling is a technique where samples are repeatedly drawn with replacement from a dataset to create multiple simulated samples. This allows for estimation of a statistic's distribution without relying on traditional parametric assumptions.

variable selection technique. Figure 2.4a shows the coefficients of the variables in a demonstration model. As λ increases, more variables' coefficients are shrunk to zero.

Similar to L1 regularisation, L2 regularisation puts a constraint on the coefficients by introducing a penalty factor. However, while LASSO regression uses the absolute sum, L2 regularisation uses the square of the coefficient values. This penalty term is again modulated by a shrinkage parameter. In L2 regularisation, coefficient values are shrunk, but never to 0 (Figure 2.4b). L2 regularisation alone cannot therefore be used as a variable selection technique.

L1 and L2 regularisation can be combined in a technique known as elastic net regularisation. Here, a parameter often known as α sets the degree of mixing between ridge regression and LASSO. When $\alpha = 0$, the elastic net does the former, and when $\alpha = 1$, it does the latter. Values of α between those extremes will give a result that is a blend of the two penalisation methods (Figure 2.4c).

I use this technique to regularise the LR model in Chapter 4.

Methods to regularise neural networks

There are many ways to regularise neural networks. Complexity can be limited by the network architecture, e.g. limiting the number of nodes per hidden layer, and the number of hidden layers. Another method is the use of *weight decay*. Weight decay is a regularisation technique that imposes a penalty on weights proportional to the sum of the squares of the weights of the network. In this regard, it could be viewed as a form of L2 regularisation.

I assess the performance of different network architectures, with additional regularisation through weight decay to optimise a neural network model in Chapter 4.

Methods to regularise gradient-boosted trees

As with neural networks, GBTs can be regularised through various techniques. The architecture can be made more simple, for example limiting the number of trees, or the depth of those trees. Other hyperparameters can also be adjusted, for example, the *minimum child weight*—a hyperparameter controlling the minimum number of observations required in a leaf node before a split is considered. A

larger minimum child weight can help to reduce the complexity of the model and improve its generalisation ability, but it can also reduce the model's ability to capture complex patterns in the data.

Sampling of the data at each step can also be used to regularise a model during development. This may involve subsampling rows of data, or subsampling columns (predictors). This sampling may occur per tree, or for each new tree level, or new split. Where a few of the predictors are highly correlated with the outcome, these predictors may always end up forming the first few splits at the root of any tree. The model may then tend to ignore the less dominant predictors. Subsampling predictors reduces this effect and allows the ensemble model to incorporate information from less dominant—though important—predictors.

These regularisation techniques are combined to optimise a GBT model in Chapter 4.

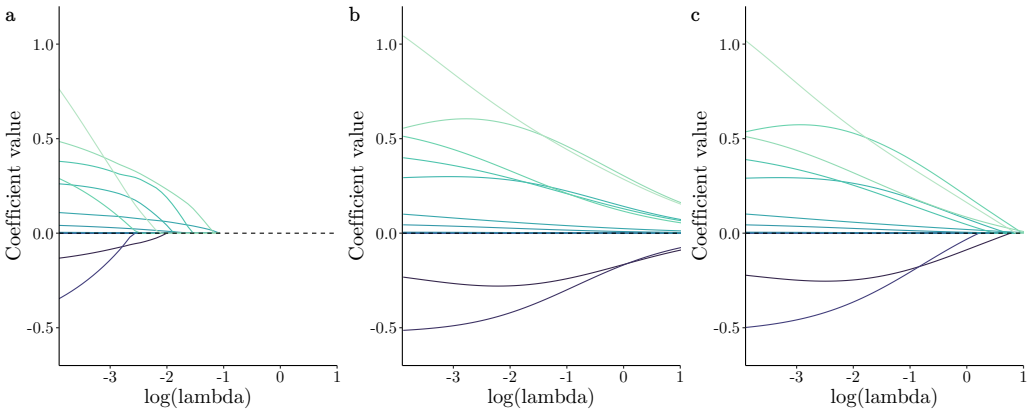


Figure 2.4: Examples using dummy data of a) ridge; b) LASSO; and c) elastic net regularisation with a mixing proportion $\alpha = 0.1$. Each line represents a different variable in the model, whose coefficients are shrunk towards zero as lambda increases.

Methods to assess model performance

In Chapter 4, discrimination, calibration, and decision curve analysis are assessed for three NOAF prediction models developed using penalised LR, GBTs, and neural networks. Their performance is compared to one published comparator model.

Where discrimination is a measure of whether a model gives higher risk estimates for patients who go on to have the event than patients who do not, calibration measures how well predicted risks align with observed risks. Decision curve analysis incorporates the consequences of the decisions made on the basis of a model. These concepts are expanded upon below.

Discrimination

Discrimination measures a model's ability to separate patients who will go on to have an outcome, and those who will not. A useful model will predict a higher risk of an event in a patient that goes on to have the event than one who does not.

The concordance statistic (C-statistic) is a commonly used metric of discrimination. The C-statistic is the probability that a randomly selected patient who developed the outcome of interest has a higher predicted risk than a patient who did not develop the outcome. It is therefore a measure of concordance between model-based risk estimates and observed events. The C-statistic ranges from 0 to 1, with 0.5 representing a model with no more predictive power than a random guess. A C-statistic of 1 suggests that every patient who developed the outcome would have a higher predicted risk than all patients who did not develop the outcome.

In a model predicting binary events, the C-statistic is equivalent to the area under the receiver-operating characteristic (ROC) curve (AUROC). The ROC curve is a plot of the sensitivity or true positive rate against the false positive rate across a range of risk thresholds.

The C-statistic is a summary of a model's discriminative ability, however it does not describe sensitivity or specificity at different, clinically meaningful risk thresholds. This information can be obtained from a labelled ROC curve. Furthermore, the C-statistic provides no insight into whether the overall magnitude of risk is predicted accurately. This aspect is measured by a model's calibration metrics, detailed below.

Calibration

Clinical decisions are based on predicted risks, so these risks should be reliable. A model may have good discrimination, but provide unreliable risk estimates, either

for all patients or those in specific risk groups. For example, a model with perfect discrimination may still produce predicted risks that are systemically too high. If a treatment is suggested for patients with a predicted risk greater than $x\%$, use of this model would result in overtreatment.

Calibration can be visually assessed by dividing the study population into risk groups based on the predicted probability of an event. The mean predicted probability of each group can then be plotted against the observed proportion of events for that group. In addition, a smooth curve can be plotted by regressing the outcome on the predicted probability of the outcome from the fitted prediction model using locally estimated scatterplot smoothing (LOESS)[89].

Whether the "spread" of predicted probabilities is appropriate can be assessed with the calibration slope β_c[90]. This metric is calculated as the β coefficient of a calibration model where the log odds of predictions are used as the sole predictor of the binary outcome[91]. A model with overly extreme predictions (too low and too high) will have a calibration slope $\beta_c < 1$. It may provide reliable risks for patients at medium risk, but may, for example, suggest a near-100% risk of an outcome when the observed risk in that group is 75%, or predict a near-0% risk in patients where the event is observed in 25% (Figure 2.5; turquoise line). Conversely models may produce overly modest predictions (not high enough nor low enough), demonstrated by a calibration slope $\beta_c > 1$ (Figure 2.5; blue line). Well-calibrated models may therefore be more clinically useful than poorly calibrated models with lower discrimination.

Overall calibration α_c ("calibration-in-the-large") compares the average predicted risk with the overall event rate[92]. This metric is calculated as the intercept of the calibration model outlined above with the calibration slope fixed at $\beta_c = 1$. Where observed proportions are higher than the predicted probabilities, calibration-in-the-large $\alpha_c > 0$, and vice versa (Figure 2.5; green and purple lines, respectively).

Models may be poorly calibrated in external data for many reasons. Common reasons include developing a model in a population with different characteristics to that in which it is tested, e.g. disease incidence or patient demographics.

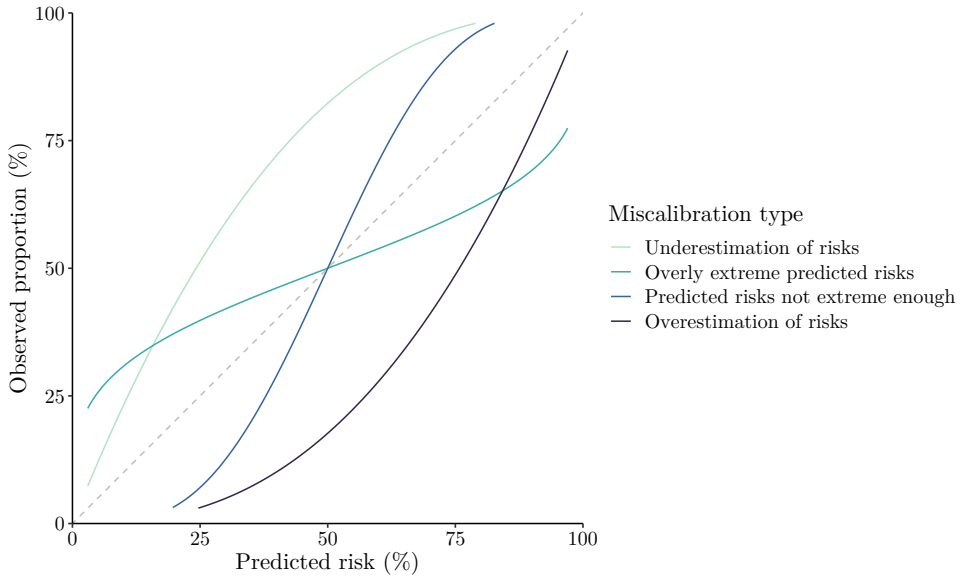


Figure 2.5: Examples of different types of miscalibration. Poorly calibrated models may systemically underestimate (purple) or overestimate (green) risks. Models may also produce estimates of risk that are too extreme (blue) or too modest (turquoise).

Decision curve analysis

Discrimination and calibration do not always fully explain which model is optimal. Decision curve analysis goes beyond these metrics to incorporate the trade-off between finding true positive cases and the cost of identifying false positive non-cases across a range of risk thresholds.

Decision curves plot a threshold probability on the x axis against "net benefit" on the y axis. The x axis can be considered to represent a spectrum of preference from most concern about a disease or event (considering intervention at a low predicted probability of event) to most concern about an intervention (only considering intervention at a high predicted probability of event). Net benefit is calculated across this range of threshold probabilities and is defined as the minimum probability of disease at which further intervention would be warranted[93]. Net benefit is the true positive rate minus the false positive rate with the latter given a weighting factor related to the relative benefits and harms of each. The weighting factor equates to the odds at the threshold probability. Choosing a 10% risk threshold (equivalent to an odds of 9:1) would therefore suggest it is nine times more important

to identify a true positive case than to identify a false positive non-case. Net benefit can be calculated as:

$$NB = TPR - FPR \times \frac{P}{1-P}$$

where:

NB	is the net benefit
TPR	is the true positive rate
FPR	is the false positive rate
P	is the threshold probability

Lower risk thresholds result in higher net benefit, as more true positive cases will be identified. However, appropriate intervention thresholds should be chosen based on weighing the harm of missing a condition versus the harm of an unnecessary intervention, and are disease and patient-specific. Once an appropriate threshold has been identified, the decision curve can be used to see if a model has higher net benefit than others at that threshold value. Net benefit monotonically decreases toward zero with increasing threshold probability. However, especially where events are sparse, empirical estimates of net benefit may be locally stable or increase (known as artifacts). Decision curves are often smoothed to remove these artifacts[94].

The unit of net benefit is true positives. Figure 2.6 shows net benefit curves for two different models, along with lines representing "intervene on all" and "intervene on nobody", as these can be viable options in certain clinical scenarios. The y intercept of the "intervene on all" strategy equates to the prevalence of the condition in the evaluation data. At a risk threshold of 0%, no cases would be missed. Therefore, with a case prevalence of 12% as in the example in Figure 2.6, all 12 patients in 100 would be correctly identified (true positives). The x-intercept (the threshold probability where net benefit is $NB = 0$) for the "intervene on all" strategy is also equal to the case prevalence. Model 2 has a net benefit of roughly 3% at a risk threshold of 20%. This suggests Model 2 would identify three true positives for every 100 patients in the target population. This compares to Model 1's net benefit of less than 1% at the same threshold value.

If concern about missing a disease case is higher, then a lower risk threshold can be used. At a threshold of around 6%, Model 1 has a net benefit of roughly

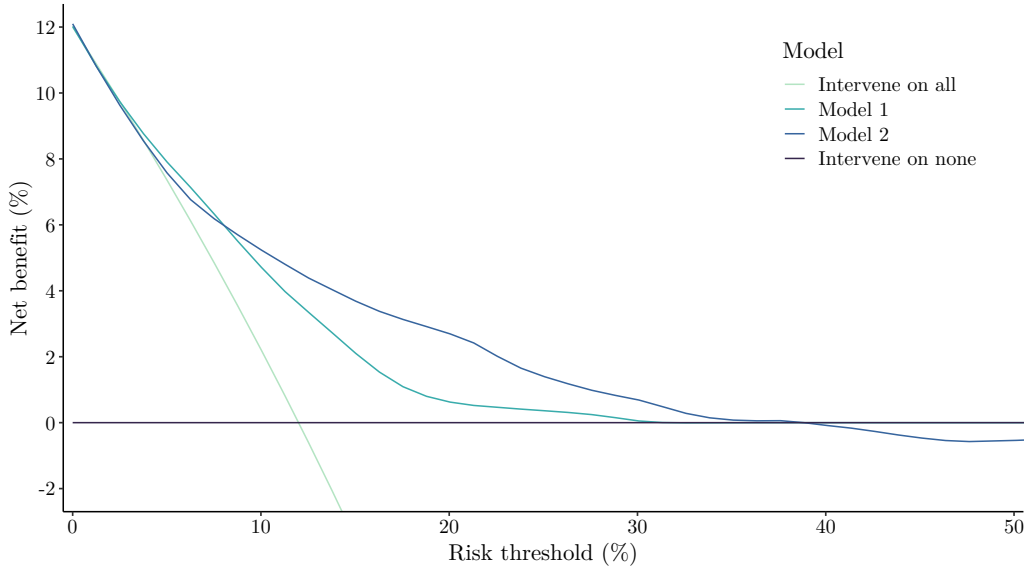


Figure 2.6: Decision curve analysis for two hypothetical models predicting the same outcome. Model 2 has a higher net benefit compared to model 1, as well in comparison to the clinical default strategies of "intervene on all" or "intervene on none", over much of the range of threshold probabilities. Choosing a 20% risk threshold (odds of 4:1) suggests it is four times more important to identify a true positive case than a false-positive non-case. At this threshold, using model 2 to decide which patients on whom to intervene would lead to better clinical outcomes than using model 1, as model 2 has higher net benefit. If the intervention is simple and non-invasive, or if the outcome has more serious consequences, a lower threshold may be considered in order to identify more true positive cases. In this case, at a threshold of 5 to 8%, model 1 should be adopted, or the "intervene on all" strategy at even lower risk thresholds.

8%, slightly higher than that of model 2, and its use might be considered if this risk threshold was chosen. At lower thresholds in this example, the net benefit approaches that of simply treating all patients. Both models have higher net benefit than treating all patients across most of the range of risk thresholds. Notably, at risk thresholds above 40%, Model 2 has negative net benefit. Use of this model at this threshold might be considered harmful, as it will result in an unacceptably high false positive rate (and subsequent unnecessary interventions) versus true positive rate. At thresholds with negative net benefit, it would be better to adopt the "intervene on nobody" strategy.

2.2.2 Estimating treatment effects using propensity weighting

In Chapter 5, I compare the use and effectiveness of different treatments for ICU-acquired NOAF. Patient characteristics including comorbidities and physiological status are likely to influence both treatment choice and outcome. It is therefore important to account for these differences during the analysis of treatment effect.

Many approaches can be used to address confounding arising from systematic differences between patients treated with different interventions. One approach is to use multivariable models of the outcome. In this method, confounding variables are included as effect modifiers in a model producing an estimate of the adjusted association of an intervention with the outcome. Correctly specifying this model is challenging, particularly where there are many treatment arms, or many confounding variables[95].

Propensity score methods do not rely on modelling the outcome and as such can be easier to specify correctly. A propensity score is the conditional probability of being assigned a particular treatment given a patient's observed confounders[96]. A propensity score can be produced in many ways. A common approach is to use an LR model with treatment status as the outcome, and the confounding variables as covariates. A key feature of propensity scores is that patients with equivalent propensity scores should have comparable distributions of observed confounders. This approach seeks to approximate the condition of covariate balance that randomisation achieves through chance.

Propensity scores can be used during statistical analysis in several ways including matching and weighting. In the case of matching in a two-arm study, a treated patient is matched to a control subject with the same or similar PS. Matching can result in a reduction in sample size where not all patients from one group can be matched to patients in the other group. The effect estimate is also only generalisable to patients similar to those who were successfully matched, which may only be a small subset of the initial cohort.

Weighting can also be used, which avoids the sample size reduction often seen with matching. In inverse probability of treatment (IPT)-weighted analysis, patients are given different weights in the analysis inversely proportional to their probability of being treated with an intervention. Individuals who received an "unexpected" (according to the measured confounders) treatment, are up-weighted and vice-versa, such that a "pseudo-population" is created where covariate distributions are the same and there is no longer an association between the measured confounders and treatment choice.

To illustrate this concept, a hypothetical study is given here. In this example, 25 patients were given an intervention (group A—intervention), and 25 patients did not receive the intervention (group B—control). If patients given the intervention tended to be younger, comparative outcomes between the two groups would be confounded by this age difference. Assuming age was the only confounding variable influencing treatment choice, a propensity score (probability of treatment) could be estimated by, for example, building an LR model with age as a single predictor. This model could be used to estimate the probability of treatment with the intervention given a patient's age. IPT weights could then be generated for each patient. The weights are calculated as the inverse of the estimated propensity score for treated patients ($w(x) = \frac{1}{p(x)}$) and the inverse of one minus the estimated propensity score for control patients ($w(x) = \frac{1}{1-p(x)}$). These weights are often normalised such that the mean weight per treatment group is $\bar{w} = 1$.

The example study above is illustrated in Figure 2.7. The mean age group A is 47 years, whereas the mean age of control patients is 66 years (Figure 2.7a). After assigning weights (visualised by proportional unit size in Figure 2.7b), the mean age in this weighted pseudo-population is now more similar. Statistical analysis of the outcome in the balanced pseudo-population can then be undertaken, providing estimates that are generalisable to the entire population from which the observed sample was taken. The treatment effect obtained after applying IPT weights is referred to as the average treatment effect.

In Chapter 5, I use weighted Cox regression (Section 2.2.3) to determine the average treatment effect for the primary and secondary outcomes to adjust for measured confounding between treatment groups. Producing correctly specified weights across multiple treatment groups is more challenging than in cases of two treatment arms. For example, upweighting of a particular patient in treatment group A to balance covariate distributions with treatment group B may worsen balance with covariate distributions in group C. Furthermore, weights that require an underlying prediction model are at risk of misspecification which can result in substantial bias in estimating the treatment effect. Given these challenges and limitations, I used energy balancing weights to balance covariate distributions across treatment groups. These are defined as the weights which minimise the weighted "energy distance" across treatment groups. The energy distance is a measure based on powers of the multidimensional Euclidean distance[97]. The distributional balancing property of these weights allows for robust empirical performance without relying on modeling assumptions for a propensity score (Figure 2.7c and 2.7d).

2.2.3 Estimating NOAF hospital outcomes using Survival analysis

In Chapter 5, I use survival analysis (time to event analysis) to estimate the effects of different NOAF treatments on patient outcomes. Similarly, in Chapter 6, I use survival analysis to analyse the in-hospital mortality associated with ICU-acquired NOAF.

Survival analysis relates to events that occur over time. Key features of time-to-event data are that survival times (a generic term representing time to any event) are always positive, and that censoring may occur, where a patient has not yet experienced the event of interest within the follow-up time. Survival times can be modelled using the survival function $S(t)$. If T is an unknown continuous random variable with $P(T > 0) = 1$ (a positive value) and represents time until an event, the survival function describes the probability of "surviving" (being event-free) past a specific time point t: $S(t) = P\{T > t\}$.

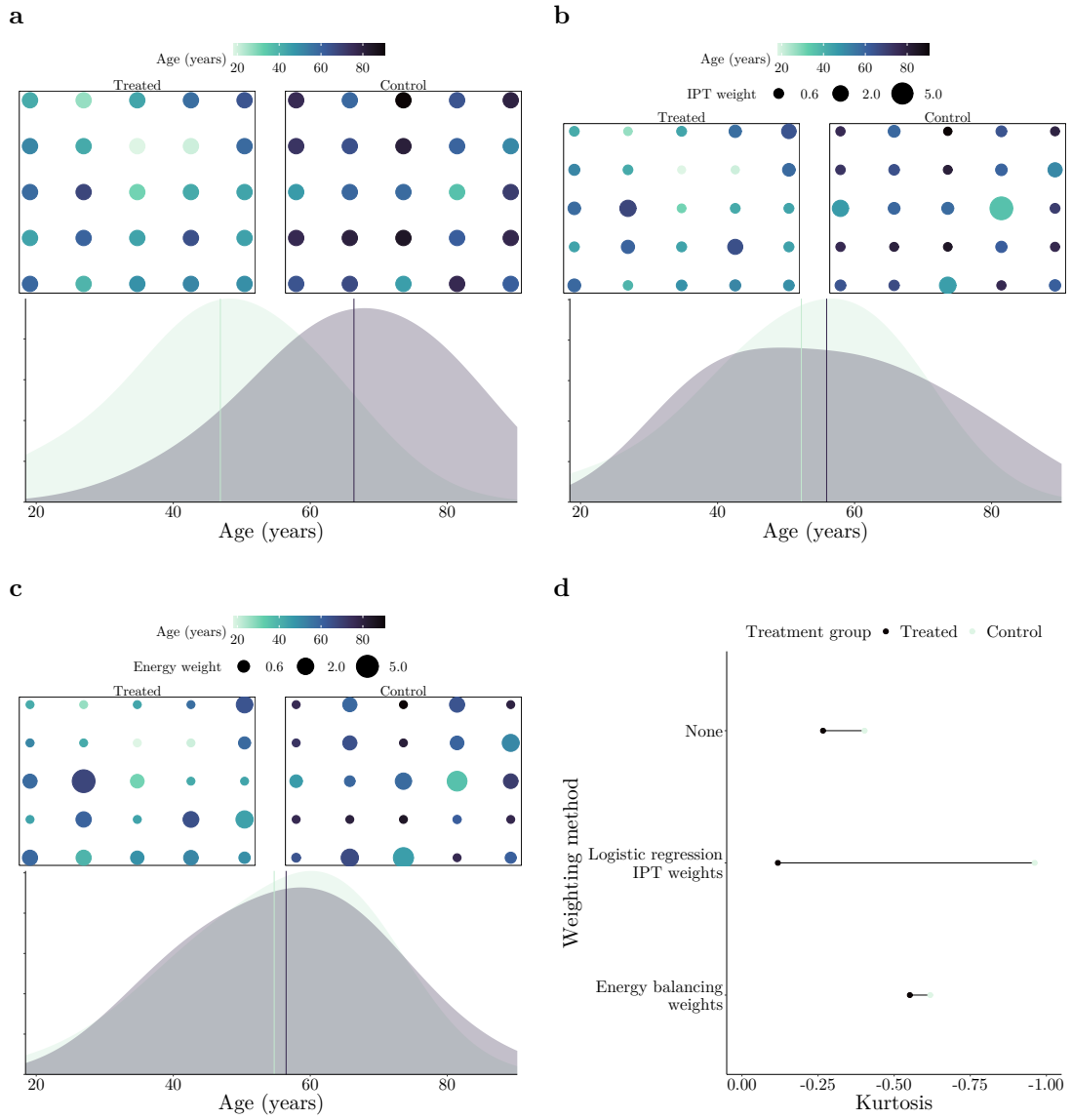


Figure 2.7: IPT weighting example: Distribution of patient ages in group A (light green) and group B (purple) in a hypothetical non-randomised study. Vertical lines represent mean ages by group. Patients in group B are, on average, older. a) In the unweighted original sample there is clear imbalance in mean age between treatment groups; b) in the weighted pseudopopulation using IPT weights generated from an LR model, older patients in group A are upweighted, while in group B, younger patients are upweighted, balancing the weighted means across the two groups. However, this is done at the expense of clear distributional imbalance; c) In the weighted pseudopopulation using energy balancing weights, both sample means and *distribution* are balanced; and d) plot showing preservation of original kurtosis after energy weighting.

The hazard function describes the instantaneous rate of occurrence of the event of interest in subjects who are still at risk of the event: $h(t) = P\{T = t|T \geq t\}$. The hazard function describes the probability that an event time T occurs at a specific time t, given that it had not happened prior to time t. In a setting in which the outcome of interest was death, the hazard function at a given point in time would describe the instantaneous rate of death in subjects who were alive at that point in time. This represents the probability of dying at time t given that you have lived this long. However, the probability of dying at *exactly* time t is zero, so the probability of dying in the next instant following t is considered, given that you have lived to time t. Here, the meaning of instant is a very small interval of time, length Δ:

$$P(t < T < t + \Delta|T > t)$$

This clearly depends on the length of the time interval Δ, so the conditional probability is expressed *relative* to the length of time Δ:

$$\frac{P(t < T < t + \Delta|T > t)}{\Delta}$$

To consider the idea of an infinitesimally small interval, let $\Delta \to 0$. In the absence of competing risks, the hazard function is therefore defined as:

$$h(t) = \lim_{\Delta \to 0} \frac{P(t < T < t + \Delta|T > t)}{\Delta}$$

The Cox model

The Cox model describes the relative effects of covariates on the hazard of an outcome. It is a linear regression of the logarithm of the hazard on the variables $X_1 \ldots X_k$. The covariates act multiplicatively on the hazard at any point in time. The Cox model can be used to estimate the regression coefficient(s), denoted β_i below, without modelling the baseline hazard function. This assumes the proportional hazards assumption holds, which states that the hazard ratio of a covariate of interest is constant over time. The hazard ratio represents the change in hazard associated with a one-unit change in the covariate of interest (or the change in

hazard associated with the presence of a binary covariate). If the hazard ratio is greater than one, as the value of the covariate increases, the hazard increases[98]. The hazard ratio for a covariate of interest can be adjusted to control for other covariates that may influence the outcome, producing an adjusted hazard ratio (aHR). This can be achieved by including additional covariates in the Cox model, or by analysing the effect of a covariate in matched or weighted cohorts (see Section 2.2.2 above).

The Cox proportional hazards model has the form:

$$h_i(t) = h_0(t) \times e^{\beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik}}$$

where:

$h_i(t)$	is the hazard function for individual i
$h_0(t)$	is the baseline hazard function
$X_{i1} \dots X_{ik}$	are the covariate values for individual i
$\beta_1 \dots \beta_k$	are the regression coefficients

The hazard function itself may vary over time, but the effect of a covariate on that hazard function should be constant to fulfill the proportional hazards assumption. This assumption can be checked in different ways. In Chapter 6 I used a plot of scaled Schoenfeld residuals over time for this purpose. For a single covariate of interest, Schoenfeld residuals represent the difference between the observed and the expected covariate values for each individual at each failure time. A non-zero slope in this plot is suggestive of a violation of the proportional hazards assumption.

Competing risks analysis

To this point, I have outlined survival analysis applied to a single event of interest. However, multiple outcomes may be possible. For example, in a study with the primary outcome of in-hospital death, hospital discharge alive is an alternate outcome. In this example, hospital discharge alive precludes the possibility of in-hospital death, so should not be handled through censoring.

In the presence of competing risks, cause-specific hazard functions, and sub-distribution hazard functions can be used.

Cause-specific hazards

The cause-specific hazard function *for a given event of interest* is the instantaneous rate of that event in subjects who have had no events of any type. The analysis involves treating individuals who have experienced a competing event as censored observations. The cause-specific hazard ratio (CSHR) can be interpreted numerically as the relative change in the rate of a particular event per unit increase in covariate value. However, in the presence of competing events, this CSHR cannot be interpreted in relation to the incidence or risk of the event, because the covariate may also alter the rates of other competing events, which will alter the cumulative incidence of the event of interest.

Subdistribution hazard

Where the focus is on rates, a cause-specific hazard model is appropriate, as it models the effect of covariates on the rate of a particular outcome in patients who have not had any event. This analysis is appropriate when investigating aetiological associations[99]. By contrast, where patient risk and prognosis is of most concern, a subdistribution hazard model is appropriate. Commonly, covariates are associated with an increased *rate* (per unit time) of all outcomes of interest. For example, diabetes is associated with an increased rate of cardiovascular mortality, and non-cardiovascular mortality. If diabetes doubles the *rate* of both outcomes, the CSHRs of diabetes for each outcome will equal two. However, the *probability* for a patient of each outcome is less than doubled, due to the competing risk of the alternate outcome. Failure to account for competing risks with an appropriate model can therefore result in estimates of the probability of the occurrence of the outcome that are biased upwards. Conversely, where a covariate is associated with an increased CSHR for one event, but a reduced CSHR of a second event, the CSHRs will underestimate the association of the covariate with the probability of each event.

While the cause-specific hazard function analyses event rate *for a given event of interest* in patients who are event-free, the *subdistribution hazard* function[100] reflects the event rate *for a given event of interest* in patients who have not experienced

the event of interest. Curiously this means that patients who have experienced a competing event remain in the risk set. The subdistribution hazard model allows the estimation of the effects of covariates on the cumulative incidence of a particular event, in the presence of competing events. Questions of prognosis and individual risk (e.g. regardless of the direct association, whether a patient who experience NOAF during an ICU admission is more likely to die in hospital) are better addressed using subdistribution hazard-based methods[101]. The subdistribution hazard ratio (SDHR) estimates the effects of a variable on the cumulative incidence of an outcome in the presence of competing risks. It constitutes a composite measure of risk that jointly evaluates the aetiological association of a risk factor with the outcome of interest and of the risk factor with all competing outcomes[100]. The SDHR does not necessarily estimate the *magnitude* of a covariate effect on the incidence of an outcome in the way that a hazard ratio from a Cox model does. However, when the event of interest is relatively infrequent, the magnitude of the SDHR does approximate the covariate effect on the risk of the event[102].

Both CSHRs and SDHRs are used in Chapter 6 to assess the aetiologic and prognostic association of NOAF on hospital mortality in the presence of the competing risk of hospital discharge alive.

A Delphi study to identify candidate NOAF risk factors

3.1 Background

Given the published morbidity and mortality associated with ICU-acquired NOAF, the identification of modifiable risk factors is important. These may suggest targets for reducing NOAF risk which could be tested in RCTs. Whether optimising AF prevention during an ICU admission would improve patient outcomes remains unknown. Furthermore, predictors with strong biological plausibility can be used to develop NOAF prediction models. Identification of patients at high risk of developing NOAF is vital for future studies investigating prevention strategies to allow stratification, sample enrichment and efficient study design.

Many risk factors for developing NOAF during critical illness have been identified, though mostly through small, single-centre studies[38]. These risk factors include demographic and comorbidity-related factors, along with acute factors such as disease severity[40,103,104]. The evidence for modifiable predictors is scarce as potential targets for prevention have not been studied in detail[38].

In prognostic factor research, variables are often studied because they are readily available. Systematic reviews are helpful to identify risk factors evident in the available literature. However, they can only identify those factors that have already been studied. They therefore have limited value in identifying direction for future study.

An expert consensus can help provide informed direction for future study, particularly in the context of inconclusive findings, or where potential gaps in understanding exist. The Delphi process is an established method to develop expert consensus[105].

The key principles behind the Delphi method are that group consensus is more valid than individual opinion, and that structured communication is an effective way to address complex problems[106]. The Delphi process involves using structured or semi-structured rounds of individual questioning to a panel of experts[107].

Some advantages of using the Delphi method include:

1. Validity of results: A systematic process involving multiple experts supports the aim of ensuring that the final results are valid and representative of the views of the broader target community.
2. Transparency and accountability: The Delphi consensus method is a structured and recorded approach that allows for transparency and accountability in the process of decision-making.
3. Understanding of different viewpoints: By including panel members from different backgrounds, a Delphi consensus exercise can provide valuable insights into different perspectives and opinions.

The aim of this study was to gather expert consensus to identify important risk factors for NOAF during critical illness[i]. These may be used to guide future research into NOAF prevention. Specifically, I aimed to generate a list of variables with strong face validity with which to build a NOAF prediction model.

3.2 Methods

I used the Conducting and REporting DElphi Studies (CREDES) guidance to plan and report my Delphi process[108]. The study protocol is available online[109]. I used an online survey tool (SurveyMonkey) for data collection and analysis. The Delphi process is outlined in figure 3.1.

[i]Segments of the introduction, methods, results, and discussion sections of this chapter have previously been published[75]. Content reproduced with permission[49].

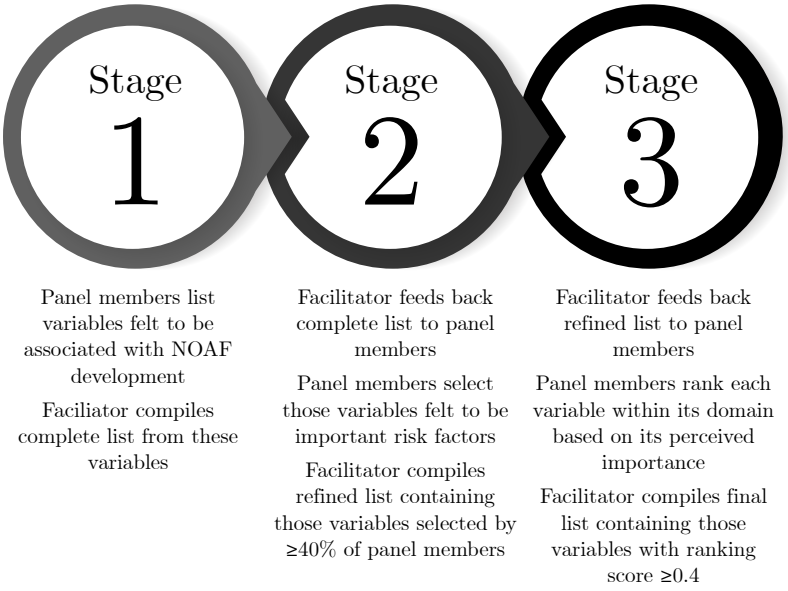


Figure 3.1: Flow diagram of Delphi process.

3.2.1 Panel selection

I identified potential participants from the author lists of articles included in our systematic review of risk factors for ICU-acquired NOAF[38]. I also identified experts through national research networks. I selected participants predominantly representing professional groups that are directly involved with patient care[106,110]. The final panel included general intensive care physicians, cardiac intensive care physicians and cardiologists. To assess panel expertise, I performed a search of the Medline database for publications by the panel members focusing on AF. I used the Scopus database[111] to present H-indices demonstrating author-level publication metrics while maintaining anonymity. This Delphi study was conducted entirely through electronic contact and data capture.

3.2.2 Round one

Panel members were asked to individually list factors associated with the development of NOAF in critically ill patients. Potential protective factors were also permitted. Panel members were offered six categories drawn from the literature in

line with my systematic review (demographics; comorbidities; disease and disease-severity-related variables; investigation results; interventions; and vital signs) to provide structure to their answers.

3.2.3 Round two

The complete list of variables generated by the panel during round one was then provided to each panel member. Each panel member was asked to select variables they felt are associated with NOAF in critically ill patients. The list was then refined by only retaining variables selected by at least 40% of panel members[112].

3.2.4 Round three

Variables that met retention criteria in round two were fed back to the Delphi panel. I asked panel members to rank each variable by perceived strength of association with NOAF development during critical illness. Variables were ranked within their respective domains (for example demographic factors and comorbidity factors were ranked independently).

3.2.5 Average ranking

A normalised average ranking score was then calculated to determine which answers in each domain were most preferred by the panel[113]. Possible scores ranged from zero to one. The answer choice with the largest average ranking score per domain represented the most preferred choice.

The ranking score for each variable was calculated as follows:

$$\frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + \cdots + x_n w_n}{\sum x \times n_c}$$

where:

$x_1 \ldots x_n$	represents the number of votes for rank choices $1 \ldots n$
w	represents the weight of ranked position
$\sum x$	represents the total number of votes for the variable
n_c	represents the total number of candidate variables within domain

Weights are the inverse of respondents' chosen ranks. Position 1 receives the highest weight which equals the number of rank choices e.g. in a question with 3 candidate variables, $w_1 = 3$. The lowest rank choice has a weight $w = 1$.

Variable inclusion

To produce a final list of variables felt to be important predictors of NOAF, I retained variables with an average ranking score of 0.4 or higher. The average ranking score ensured variable exclusion was not inevitable i.e. all variables within a domain could be carried through to the final list given sufficient spread of votes.

3.3 Results

3.3.1 Participants

I approached 22 experts. Of these, 12 agreed to contribute to the Delphi process. All 12 respondents were active clinical researchers based in Europe, North America and South America. In a search of the Medline electronic database, 110 publications relating to AF written by members of the panel were identified. H-indices of panel members ranged from 2 to 29 (mean = 14.4, median = 14).

I first contacted panel members in November 2019, sent information for round one in December 2019, and the final list of variables was confirmed in July 2020.

3.3.2 Round one

The panel responded with 126 distinct candidate risk factors across the six suggested domains. The largest domain contained variables relating to investigation results.

3.3.3 Round two

All 12 of the initial participants responded in round two. I generated a refined list of 102 candidate risk factors after excluding variables not selected by at least 40% of panel members.

3.3.4 Round three

All 12 of the initial participants responded in round three. I produced a final list of 75 candidate risk factors after excluding those variables with an average ranking score of less than 0.4. These factors represent the most important variables that the panel of

experts considered to be risks for developing NOAF during critical illness. Age, sex, and ethnicity were identified as important demographic variables. The comorbidity-related variables with the highest ranking were atrial dilatation, valvular heart disease, and congestive cardiac failure. Disease or disease severity-related variables were also highlighted, with cardiac surgery, septic shock, and sepsis being the highest ranked. Hypokalaemia was ranked the most important investigation-related variable, followed by high left atrial pressure and the presence of atrial ectopy. All three of the highest ranking intervention variables were IV medication infusions, namely catecholamines, aminophylline, and salbutamol. The physiological variables felt to be most important were tachycardia, fever, and heart rate variability. The full list of ranked results for demographic, comorbidity, disease, investigation, intervention and physiological domains are displayed in Tables D.1 to D.6, respectively.

These variables were then refined into a final list of candidate variables for model development. This process is outlined in Section 4.2.4.

3.4 Discussion

In this study, I aimed to systematically identify potential risk factors for NOAF during critical illness to guide future studies that might aim to modify the incidence of NOAF. The feasibility of remote participation facilitated international involvement. The wide range of expertise across the panel provided a broad range of perspectives, maximising the benefits of the group consensus model. The Delphi process thereby provided an expert consensus on candidate variables with strong biological plausibility and face validity.

The expert panel identified several factors across a range of categories. While some factors have been extensively studied in the ICU population, some, for example the use of preventative medications, await investigation in this setting.

The Delphi process facilitated opportunities to include risk factors for NOAF in the context of critical illness that have not been formally evaluated in clinical studies. Many risk factors identified by the panel were variables with evidence of correlation with NOAF either in the community or during critical illness. Variables

that were proposed which do not have a strong evidence base as causal risk factors have physiologically plausible mechanisms of action that could increase risk for NOAF in critically ill patients. The findings could be used to guide studies of preventative interventions in high-risk patients.

These results are used to inform the prediction model development detailed in Chapter 4. Building on the general concepts outlined in Chapter 1, the potential arrhythmogenic mechanisms of the identified risk factors are explored further below.

3.4.1 Non-modifiable risk factors

Demographic factors

The demographic factors highlighted in this study are consistent with those associated with increased risk of AF in patients in the community, such as increasing age[2]. The underlying pathophysiological changes associated with this increased risk are generally well understood. It is associated with atrial fibrosis and changes in ionic currents independent of fibrosis[114,115]. These changes affect normal conduction pathways, promoting non-uniform conduction, electrical dissociation and re-entry circuits, leading to arrhythmia[115–117]. Increasing age is consistently found to increase NOAF risk in patients in the ICU[79].

Male sex is a recognised risk factor for AF in patients in the community[2]. This is also true of NOAF associated with critical illness[39]. Sex appears to modify how other risk factors affect AF risk, with obesity increasing risk more in men than in women[118]. Increasing height also increases AF risk and may relate to increasing atrial size[119]. Sex-dependent ion channel expression[120] and atrial refractory period duration[121] may also contribute.

Ethnicity was identified as an important predictor. Community-based studies reveal that, despite having higher rates of AF risk factors, patients of African descent have a lower risk of being diagnosed with AF compared to white patients[122,123]. However, more recent studies using ambulatory monitoring suggest AF rates may be similar across racial groups, suggesting race-dependent differential clinical detection of AF[124].

Comorbidity

The comorbidity-related factors felt to be most strongly associated with NOAF in the critically ill were related to heart disease, consistent with evidence in the community population. Atrial dilatation is a strong predictor of AF in the community[125], and is thought to be due to mechanisms associated with atrial stretch. Atrial stretch causes a decrease in atrial effective refractory period and an increase in AF inducibility in animal studies[126]. Stretch-activated ion channels may be responsible for these changes[127]. This association has been demonstrated in patients admitted to an ICU, with left atrial diameter increasing NOAF risk[79,128,129].

Atrial wall thickness (as a marker of atrial hypertrophy) correlates with the degree of atrial fibrosis, independent of atrial volume[130]. It also correlates with areas demonstrating complex fractionated atrial electrograms, implicated in the development and maintenance of AF[131]. Atrial volume may be estimated by transthoracic echocardiography (TTE), but assessment of atrial wall thickness is more difficult. While TTE has been used for this purpose[132], the spatial resolution tends to be inadequate[133]. Transoesophageal echocardiography, cardiac computed tomography, and cardiac magnetic resonance imaging can all be used to provide measures of atrial hypertrophy[133], though their use is unlikely to be feasible for routine risk stratification for patients on an ICU. If it could be accurately measured non-invasively, atrial wall thickness could become a useful marker of NOAF risk in critically ill patients.

Conditions leading to right heart pressure overload including pulmonary hypertension, obstructive sleep apnoea (OSA) and chronic obstructive pulmonary disease (COPD) ranked highly. OSA may predispose to AF due to atrial stretch[134] and large variations in afterload. Furthermore, vagal activation secondary to negative tracheal pressure shortens the atrial refractory period and promotes AF[135,136]. Pulmonary hypertension increases AF risk through increased sympathetic activity[137] and chronic right atrial pressure overload causing atrial stretch[138]. These findings are supported by studies in critical care demonstrating increased NOAF risk in patients with OSA, COPD or pulmonary hypertension[44,139,140].

Each of these comorbidity-related factors may be exacerbated in the intensive care setting. For example, excessive fluid resuscitation increasing left atrial stretch and thus increasing NOAF risk.

Disease / disease severity factors

Of the factors relating to disease or disease severity, sepsis and measures of organ failure were ranked highest. Sepsis is an established risk factor for NOAF[39], with systemic inflammation playing a role in the initiation and maintenance of AF (see below). High C polysaccharide-reactive protein (CRP) values were also ranked highly, consistent with these other measures of systemic inflammation.

Inflammation is associated with increased AF risk

Outside critical illness, inflammatory infiltrates are found in atrial myocytes of patients with AF[141]. These are not present in those in sinus rhythm. Serum concentrations of pro-inflammatory cytokines such as interleukin 6, interleukin 8 and tumour necrosis factor are also more elevated in those patients with AF than those without[142,143]. Higher baseline CRP concentrations are predictive of NOAF in the community[144], and CRP values in hospital after myocardial infarction correlate with AF risk during the hospital stay[145].

During the development of critical illness, systemic inflammation increases the risk of AF[52]. Inflammatory mediators encourage myocardial myofibroblast differentiation and subsequent accelerated fibrosis[146]. Moreover, inflammatory cascades and bacterial endotoxins can alter ion channel function, resulting in shortened APDs and increased risk of arrhythmia[147]. Furthermore, reactive oxygen species (ROS) are elevated during critical illness and increase atrial ectopic activity[148]. ROS also promote an arrhythmogenic atrial substrate by altering the function of connexins responsible for intercellular electrical coupling[149]. In patients admitted to an ICU, a significant increase in CRP concentration has been noted prior to NOAF[150], and higher concentrations are associated with increased NOAF risk[52].

These data suggest systemic inflammation may play a role in promoting AF. This assertion is supported by observations that reducing inflammation can reduce

AF risk or promote cardioversion; a randomised study of a five-month tapered course of methylprednisolone versus placebo administered to outpatients after electrical cardioversion resulted in recurrence rates of 10% versus 50%, respectively[151]. However, the results of this single-centre study suggest a profound clinical effect which has not to date been replicated. Serum CRP concentrations were predictive of AF recurrence in this cohort. Prophylactic corticosteroid treatment during cardiac surgery has also been shown to reduce post-operative NOAF[152]. Pivotal randomised controlled trials (RCTs) of steroids in sepsis in general ICUs did not report comparative arrhythmia rates[153,154], or reported only life-threatening arrhythmias[155]. However, in patients on an ICU with septic shock, an observational study associated low-dose hydrocortisone with a lower risk of developing AF[156].

AF promotes inflammation

While available evidence suggests that systemic inflammation promotes NOAF, the development of AF may itself lead to further inflammation. In animal models, rapid atrial pacing leads to increased serum CRP concentrations[157]. Furthermore, in patients with paroxysmal AF, serum CRP concentrations are higher during periods of AF[142], with longer durations of AF associated with higher CRP concentrations[158]. CRP concentrations gradually reduce after successful cardioversion[159]. These observations complicate the interpretation of correlation studies of inflammation and AF.

Cardiac surgery, particularly that involving the mitral valve, is a well-established risk factor for post-operative AF[160]. Despite the body of research, tools to predict AFACS have proven unreliable and the topic remains a research priority[161].

3.4.2 Modifiable factors

The study also identified potentially modifiable factors, such as physiological variables, electrolyte concentrations, interventions, and preventative medications. Many of these, such as target serum potassium and magnesium concentrations, are commonly used to reduce risk and treat AF in the absence of strong evidence for this practice (see below). Other measures such as preventative medications

(see below) have demonstrated utility in non-ICU populations and are worth exploring in future studies.

Physiological status/vital signs

The interaction between physiological variables, disease processes, and underlying comorbidities is complex and worth examining. Fluid balance and increasing weight were felt to be important predictors of NOAF risk. Increasing fluid balance may be a surrogate for atrial stretch and is associated with increased NOAF risk in critically ill patients[129]. This association may not be causal. It may instead reflect that more unwell patients (who often receive larger volumes of fluid) are at higher NOAF risk.

Hypotension is likely to be associated with the development of NOAF when in the context of vasodilation or cardiomyopathy related to systemic inflammation. Independent from systemic inflammation, hypotension may impair tissue perfusion including that of myocardial conduction tissue[162]. Decreased tissue oxygen delivery also stimulates endogenous catecholamines, further increasing the risk of arrhythmias[163].

Investigation results

Hypokalaemia was considered the most important variable in the domain relating to investigation results. Critical illness often co-exists with disturbances in fluid and electrolyte homeostasis[164]. Disorders of potassium and magnesium are common. Mechanistic and clinical evidence suggests that these electrolyte disturbances may increase NOAF risk.

Hypokalaemia and atrial arrhythmias

The ion that contributes most significantly to E_m is K^+(see Appendix A)[18]. An intracellular $[K^+]$[ii] around 30 times that of the extracellular concentration is maintained by the activity of Na^+/K^+ ATPase[165]. The high selective permeability to K^+ generates E_m close to that of the K^+ equilibrium potential (E_K = -89 mV with a extracellular $[K^+]$ of 5 mmol/L and an intracellular $[K^+]$ of 140 mmol/L).

[ii]Square brackets are used in this thesis to represent the concentration of a substance in solution.

This negative resting E_m stabilises atrial and ventricular myocytes during diastole, reducing the likelihood of spontaneous action potentials and subsequent premature extrasystoles[166].

In states of K^+ depletion, extracellular $[K^+]$ is reduced disproportionately over intracellular $[K^+]$. The importance of changes in extracellular $[K^+]$ is demonstrated in the Nernst equation (see Appendix A) as applied to potassium: $E_m = 61.5log_{10}\frac{[K^+]_o}{[K^+]_i}$. The ratio of $[K^+]_o$ to $[K^+]_i$ is a key determinant of E_m. Because $[K^+]_i$ changes very little, small reductions in $[K^+]_o$ can result in E_m becoming significantly more negative. This hyperpolarisation leads to a greater stimulus being required to generate an action potential. This mechanism explains the reduced excitability of neurons in the context of hypokalaemia. In cardiac muscle, however, hyperexcitability is observed. In ventricular myocytes, one mechanism by which hypokalaemia promotes ectopic activity is through inhibiting the activity of Na^+/K^+ ATPase. This results in intracellular accumulation of Na^+, in turn limiting the extrusion of Ca^{2+} by the Na^+/Ca^{2+} exchanger. A buildup of intracellular Ca^{2+} reduces repolarisation reserve and promotes afterdepolarisations. Afterdepolarisations are secondary depolarisations occurring during the repolarisation phase of the cardiac cycle[167] that could otherwise induce triggered activity and AF[168]. Hypokalaemia also results in a prolonged APD. This is due to the repolarising outward K^+ currents being reduced by the lower extracellular $[K^+]$. The longer APD allows recovery of L-type Ca^{2+} channels, increasing the risk of early afterdepolarisations[169].

The mechanisms underlying hypokalaemia-induced arrhythmias in atrial myocytes are less clear. While Ca^{2+} overload and subsequent early afterdepolarisations remains a mechanism in tubulated atrial cells, ectopy in untubulated atrial myocytes may be due to reactivation of the fast I_{Na}. This reactivation is enabled by a rapid action potential in the presence of I_{Kur}, and hyperpolarisation of the resting membrane potential during hypokalaemia further increasing Na^+ channel availability[170].

Low serum $[K^+]$ is associated with a higher risk of developing AF in patients in the community[171]. Baseline measurements of serum $[K^+] < 3.85$ mmol/L were associated with increased AF risk. This association appeared strongest in people

with a history of myocardial infarction. Hypokalaemia has also been linked to AF in patients undergoing cardiac surgery. In this setting, patients with serum $[K^+]$ $\leq$ 3.9 mmol/L were at higher risk of developing AF than those with $[K^+] \geq 4.4$ mmol/L[172]. These findings are supported by another study in patients undergoing cardiac surgery where a $[K^+] < 3.5$ mmol/L was associated with a higher risk of perioperative AF[173]. Hypokalaemia has also been associated with NOAF in patients post-myocardial infarction and has led to guidelines suggest targeting a serum $[K^+]$ in the upper range of normal in these patients[174]. This practice is often applied to the general critically ill population. However, there is little evidence to date demonstrating an association between $[K^+]$ and AF risk in patients on a general ICU. A serum $[K^+] < 3.5$ mmol/L was associated with an increased risk of arrhythmia requiring treatment when compared to a reference range of 4.0 to 5.0 mEq/L in both a cardiac ICU and medical and surgical ICU cohort[175]. However, arrhythmia type was not specified in the study. Current practices of electrolyte supplementation[70] in general ICUs are supported by evidence from high-risk subgroups such as cardiac surgery or post-myocardial infarction. Hypokalaemia appears to have stronger mechanistic and clinical associations with ventricular rather than atrial arrhythmias. The balance of risk of potentially unnecessary electrolyte supplementation, often involving concentrated solutions administered via a central vein, is unclear.

In addition to hypokalaemia, the rate of change of potassium level was considered to be an important risk factor for NOAF. Increasing potassium variability is associated with poorer outcomes in critically ill patients[176,177]. Electrolyte supplementation may induce arrhythmias through rapid changes in extracellular concentration, although this requires further study[176]. These findings support further investigation to determine the optimal target for patients' serum $[K^+]$, and supplementation using considered dosing.

Hypomagnesaemia and atrial arrhythmias

The panel also identified hypomagnesaemia as a risk factor for NOAF. Mg^{2+} is the second most abundant intracellular cation after K^+ and is distributed principally

between bone, muscles and soft tissues. There is a disparity between intracellular and serum concentrations of Mg^{2+} with normal extracellular and intracellular concentrations ranging from 0.7 to 1.0 mmol/L and 5 to 20 mmol/L, respectively. Equilibrium between these compartments occurs slowly. Serum [Mg^{2+}] levels may therefore not be reflective of true body stores[178]. This disparity provides plausibility to Mg^{2+} supplementation in the context of normal serum [Mg^{2+}].

Extracellular Mg^{2+} has a direct inhibitory effect on inward Ca^{2+} currents[179]. In solution, Mg^{2+} binds water more tightly that other common physiological cations. It surrounds itself with two hydration shells and its radius when hydrated is around four-hundred times larger than when dehydrated. The single hydration shell formed by Ca^{2+} increases its radius by only twenty-five fold[180]. Despite the chemical similarities between Mg^{2+} and Ca^{2+}, magnesium's large hydration shell results in very different physiological behaviours. For a Ca^{2+} ion to fit into a channel, it sheds its hydration shell. However, the two shells surrounding a Mg^{2+} ion are bioenergetically more difficult to shed. Mg^{2+} therefore has a blocking effect at Ca^{2+} channels as it cannot readily traverse the pore[181–183]. An IV infusion of Mg^{2+} increases the atrial refractory period and prolongs sinus node recovery time and AV node conduction time[184]. Intracellular magnesium appears to play a role in maintaining the plateau phase of the cardiac action potential[185].

A serum [Mg^{2+}] < 0.73 mmol/L was associated with the development of AF in a longitudinal, community-based cohort[186]. The association with AF risk was not linear and was most marked in those in the lowest quartile of serum [Mg^{2+}]. However, residual confounding from chronic illnesses or malnutrition that predisposes to both hypomagnesaemia and AF could not be excluded. Clinical studies have shown that intravenous (IV) magnesium can augment rate control in AF and facilitate maintenance of sinus rhythm[187]. Hypomagnesaemia lowers the threshold for digoxin-related arrhythmias[188]. In a small study, dietary magnesium restriction was found to increase the risk of AF in women[189]. Furthermore, intracellular Mg^{2+} deficiency determinted from atrial appendage specimens was associated with AF after cardiac surgery (AFACS)[190].

Perioperative magnesium administration reduces the incidence of AFACS[53,191]. It has rate-control properties in NOAF in hospitalised patients[192] and may promote cardioversion when high serum concentrations are targeted[193,194]. However, it is suggested that routine magnesium supplementation in undifferentiated ICU patients has no effect on arrhythmia rates[195] and hyper- rather than hypomagnesaemia is associated with adverse outcomes[196]. To date, there is no evidence supporting an optimal target serum [Mg^{2+}].

Interventions

Of the interventions, administration of sympathomimetic medications such as catecholamines, bronchodilators, aminophylline and vasopressin were ranked highest. Adrenergic stimulation from catecholamines and bronchodilators enhances automaticity and may promote atrial re-entry circuits[197] (see below). Further to phosphodiesterase inhibition, aminophylline is an antagonist at the adenosine A_1 receptor. Reducing adenosine's cardiac sympatholytic function adds to tachyarrhythmia risk[198].

Beta blockers and amiodarone therapy were judged to be important modifiers of NOAF risk, as was vasopressin. Vasopressin induces vasoconstriction through action on V_1 receptors on vascular smooth muscle[199]. It is recommended as a catecholamine-sparing agent in the management of distributive shock[200]. By reducing the requirement for catecholamines, it decreases the stimulation of arrhythmogenic myocardial β_1 receptors and thereby reduces arrhythmia risk[201]. This protective effect has been demonstrated in patients during critical illness in a systematic review of studies comparing catecholamines alongside vasopressin versus catecholamines alone[202].

The autonomic nervous system and AF risk

The autonomic nervous system is divided into three main components: sympathetic, parasympathetic, and enteric. The cardiac autonomic nervous system (CANS) regulates vascular tone, contractility, and cardiac electrophysiology[197]. The ganglion cells of the CANS are located either outside the heart (extrinsic) or inside the pericardial sac (intrinsic)[203]. The intrinsic CANS is mainly composed of sympathetic

and parasympathetic nerves as well as ganglionated plexi (clusters of autonomic ganglia). These nerves and plexi are found epicardially, and along the great vessels[204]. The posterior left atrium and pulmonary veins in particular are richly innervated by sympathetic and parasympathetic fibres[205].

Adrenergic activation may promote arrhythmogenesis through several mechanisms. I_{K1} provides a diastolic outward current that limits the spontaneous depolarisation due to enhanced pacemaker activity or delayed afterdepolarisations[206]. Multiple mechanisms including I_{K1} contribute to orderly and rapid repolarisation, providing a degree of redundancy in what has been termed "repolarisation reserve"[207].

Automaticity is enhanced by reduced I_{K1}, which can result from α-adrenergic stimulation[208], or increased I_f, produced by beta-adrenergic activation[209]. Furthermore, sympathetic activation augments and prolongs the Ca^{2+} transient current, loading myocytes with excess Ca^{2+}[203].

The parasympathetic nervous system also plays a role in arrythmogenesis, particularly when in combination with sympathetic activity[210]. Concurrent parasympathetic activity shortens the pulmonary vein APD. The abbreviated APD contrasts with the longer Ca^{2+} transient induced by sympathetic activity, creating an environment promoting afterdepolarisations. This sympathetically-driven arrhythmogenesis is demonstrated by the observation that AF is elicited by the non-selective beta agonist isoprenaline[211].

Heart rate variability (HRV) was highlighted as a potential predictor. Normal HRV is a feature of the healthy cardiovascular system. It is controlled by a balance between both branches of the autonomic nervous system and is reduced in states of autonomic dysfunction. Reduced HRV predicts NOAF in patients in the community[212]. In hospitalised patients, reduced HRV has been noted to precede ischaemic cardiac events and to be predictive of multi-organ dysfunction[213]. It is not clear whether HRV may be predictive of new-onset arrhythmias in critically ill patients.

In critical care environments, the use of beta-adrenergic medication is associated with the development of NOAF[38,214]. Furthermore, endogenous sympathetic activation during critical illness likely contributes to the increased NOAF risk seen with higher illness severity[201].

Invasive interventions

ICU-specific interventions, such as the placement of pulmonary artery catheters and central venous catheters were also considered important predictors. Two RCTs of pulmonary artery catheters did not compare rates of NOAF between intervention and control groups; however, both reported dysrhythmias during insertion in 3% of patients[215,216]. AF may be induced by myocardial irritation from malpositioned central catheters[217] or by guidewires inserted into the atrium during catheter insertion[218]. It remains unclear whether central venous catheters pose an increased NOAF risk when well positioned.

3.4.3 Preventative measures

The expert consensus in this study puts forward a role for preventative pharmacological therapies. A recent systematic review found that the use of perioperative beta blockers was associated with a reduced risk of AF after cardiac surgery[219]. Beta blocker prophylaxis in the general critical care population has not been studied. Timing of re-initiation of beta blockers for patients who were previously taking them may be important, with early reinstatement after non-cardiac surgery appearing to be protective[220].

Perioperative amiodarone prophylaxis prevents NOAF after cardiac surgery[53]. This intervention is safe, generally well-tolerated, and may be justified given the high risk of NOAF in this cohort[221]. Further study is required as to whether amiodarone prophylaxis may be of benefit to high-risk patients in the general critical care cohort.

Limitations

A Delphi process can be a valuable addition to quantitative data, allowing some degree of qualitative evaluation to better assess results in context. Delphi results will be largely subjective and are not supported by a robust quantitative measure developed from primary data. There is therefore an element of trust associated with the presentation of Delphi results in the context of an anonymous panel. Commonly described limitations of the method include a lack of common guidelines for definition of consensus and for panel selection.

For this study I selected panellists purposively to include individuals with experience relevant to the identification and treatment of AF in the critically ill population. Panellist independence was retained throughout, and at no point were individual panellists made known to each other. At each round consensus was reported in line with pre-determined criteria. The response rate was relatively low (55%) and responses were not acquired from all geographical regions invited. This may act as a source of bias and limit the generalisability of these findings. Variable domains were introduced to provide structure to responses; while it was possible for all variables to be retained within a domain given sufficient spread of votes, a variable's score may have been influenced by the domain in which it was categorised. Lastly, there was overlap between some variables within and across domains. Certain similar variables were retained uncollated as per pre-specified methods to avoid facilitator influence on the results.

Development of a NOAF prediction model

4.1 Background

4.1.1 AF prediction in the community

Prediction of NOAF risk in patients in the community may be useful to identify individuals for monitoring or for enrolment in future prevention trials. Several prediction models have been developed, including the Framingham Heart Study model[222], the Atherosclerosis Risk in Communities (ARIC) model[223], and the CHARGE-AF (The Cohorts for Heart and Aging Research in Genomic Epidemiology AF) score[224]. Direct comparison is hindered by the sparsity of external validation studies[225]. The performance of these models is generally limited. For example, while the CHARGE-AF model performed modestly in development data (C-statistic $C = 0.77$, 95% confidence interval (CI) 0.75 – 0.78), it performed poorly during external validation in two external datasets ($C = 0.66$, 95% CI 0.63 – 0.70 and $C = 0.71$, 95% CI 0.66 – 0.75, respectively)[224]. Calibration was only assessed after recalibrating the model to the external validation data. More recently, a time-varying neural network model based on UK general practice data was found to outperform the CHARGE-AF model ($C = 0.83$ versus $C = 0.73$, respectively), highlighting the potential of machine learning models for AF prediction[226].

4.1.2 Prediction of ICU-acquired NOAF

Common variables used in community models include age, presence of diabetes, left ventricular hypertrophy, and coronary artery disease[222–224,226]. Risk factors for developing AF in the community differ substantially from those of critically ill patients. While age and comorbidities continue to contribute to risk[227], many acute factors including the severity of acute illness and interventions such as vasoactive

drugs also play a significant part[38]. Furthermore, prediction of AF during critical illness would be useful over time frames of hours to days, whereas community risk models are built on data spanning years. The differences in patient population, time frame, and risk factors mean community AF risk models are not suitable for the prediction of AF in acute illness.

Reliable prediction of NOAF in patients admitted to an ICU would allow enrichment of trials of preventative strategies. Identification of risk factors may provide insight into the physiological mechanisms underlying arrhythmogenesis.

At the outset of this work, one usable model to predict NOAF in critically ill patients had been published. This model was developed using data from 1,722 patients with sepsis from two ICUs in the Netherlands[52]. Of these 1,722, 418 (24%) developed NOAF. It was unclear how candidate predictors were selected for inclusion in the model. The final model contained nine predictors. Certain predictors were poorly-defined, increasing the risk of inter-user variability. For example, the presence of an immunocompromised state included any of: acquired immunodeficiency syndrome, prior use of high dose corticosteroids, current use of immunosuppressive drugs (unspecified), current use of antineoplastic drugs (unspecified), recent hematologic malignancy, or any documented humoral or cellular deficiency.

This model has been subject to an external validation study on 7,844 patients with sepsis admitted to an ICU using data from the MIMIC-III database[228]. The external validation study found that the model discriminated poorly ($C = 0.598$, 95% CI 0.587 – 0.609).

One other study investigated the use of a support vector machine—a type of machine learning model—to predict NOAF using single-centre data recorded at ICU admission. This study included patients with pre-existing AF. Furthermore, the model was not subject to any external validation, and only internally validated (cross-validation) performance metrics were reported. The authors provided no means for other investigators to use their model. Its discriminative performance was limited (cross-validation $C = 0.73$).

In the past year, one further ICU-acquired NOAF prediction model has been published[229]. This study used a larger sample of 18,518 patients admitted to an ICU in one centre in Amsterdam, of whom 2,374 (12.8%) developed AF. While this study offered considerable improvement in sample size over the study by Klein Klouwenberg *et al.*[52], it did not account for dynamic variable values, instead using mean values over the course of a patient's ICU stay. The data for model derivation did not include a history of AF, so although patients admitted to the ICU in AF were excluded, many may have had pre-existing paroxysmal AF. Calibration of this model was not assessed, and no external validation was undertaken. Furthermore, because this model used mean values as predictors, a user would need to compute these values over a patient's stay for all 10 dynamic predictors in order to use the model. These variables included some that are unlikely to be available for all patients, for example the activated partial thromboplastin time and the prothrombin time. While practices vary, these coagulation tests are not necessarily routine in general ICUs. Indeed, recent work from the UK, Canada, and Australia suggests routine clotting profiles are generally felt to be unnecessary so this data may not be available in these settings[230]. Another important included predictor was ST segment height, which is not routinely calculated. These limitations significantly limit the models utility.

One model predicting NOAF in patients with sepsis (not ICU-specific) has also been published in the past year[231]. The study was developed using data from 1,568 patients in China, of whom 167 developed NOAF. It is unclear how the candidate variables were selected. The final model requires an International Normalised Ratio (INR) and a plasma [fibrinogen] for each patient to provide a prediction. As with the Ortega-Martorell *et al.* model[229], these results are unlikely to be routinely available. Furthermore, considerable risk was conferred by dopamine use in this model. Dopamine is not recommended as a first-line vasoactive medication[200]. Its use resulted in poorer outcomes and higher arrhythmia rates than noradrenaline in a large RCT[232], a finding supported in a meta-analysis of 32 trials[233]. Roughly one third of patients who developed NOAF in the development data of this model received dopamine. However, dopamine remains in guidelines only for selected

patients at low risk of arrhythmia[200]. The small development sample size and requirement for likely unavailable blood results markedly limits the external validity and utility of this prediction model. Its utility is also likely to be limited in units where an important predictor (dopamine) is rarely used.

The development of these models is an example of why systematic candidate variable selection processes should be undertaken, as the unavailability or limited use of predictors in other healthcare settings may have been identified prior to local model development. Available NOAF prediction models are healthcare-system-specific, and where tested, do not generalise well to external data. Due to these limitations, it is important to build on this work to develop a generalisable, clinically-useful tool based on the best available evidence for NOAF risk factors. Machine learning techniques should also be assessed in NOAF prediction as they may allow for consideration of complex interactions between predictors that may improve prediction performance over pre-specified interactions.

I aimed to use systematically-identified candidate NOAF predictors from Chapter 3 to develop a user-friendly and reliable NOAF prediction model for use in ICUs. I aimed to compare machine learning approaches with a traditional statistical model.

4.2 Methods

4.2.1 Study design and setting

I performed a retrospective observational study combining two large intensive care databases: PICRAM[80] and MIMIC-IV v1.0[82] (see Sections 2.1.1 and 2.1.3, respectively). Study consent and approval details are outlined in Chapter 2. The analysis plan was published in advance[234] and this work is reported according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines[235]. The final prediction model was externally validated using the CCHIC database (see Section 2.1.4).

4.2.2 Participants

All adult ($\geq$ 16 years) patients admitted to one of the study ICUs were included. For patients admitted more than once to an ICU, their first admission was used.

I excluded patients:

- admitted to a service-specific ICU,
- with an ICU LOS of less than three hours,
- with pre-existing arrhythmias, or
- with significant arrhythmia in the first three hours of arrival to ICU, as a surrogate for pre-existing undocumented arrhythmia

Patients with pre-existing arrhythmias were identified by the presence of documented arrhythmia or medications prescribed with an indication of arrhythmia management in their medical history.

4.2.3 Exposures and outcomes

Routine practice in all ICUs included in the databases was to display three-lead ECG continuously, with heart rhythm recorded by the bedside nurse at regular intervals. AF was defined as the documentation of AF lasting for 30 seconds or more[236]. A documented heart rhythm was assumed to persist until the next identifiable rhythm was recorded. NOAF was defined as AF occurring after three hours of admission in a patient with no documented pre-existing arrhythmia.

4.2.4 Predictors

Candidate predictors were generated from the international Delphi exercise outlined in Chapter 3. Initially, disease severity scores and clinical syndromes were distilled to their component variables. Where possible, clinical categories were converted to continuous variables to improve statistical power, for example using patient weight and heart rate rather than dichotomous labels of obesity and tachycardia, respectively[237]. where adequate variable data was not available, or where variables could not be characterised reliably, these variables were excluded. Categorisation of ethnicity resulted in unbalanced groups, with low numbers developing NOAF in

certain racial categories. During cross-validation, groups were further split by ICU resulting in low or no outcomes in certain groups, leading to unstable model estimates. Ethnicity was therefore not explicitly coded as a predictor. The representation of variables identified in the Delphi process in model development is shown in Table D.7.

This process resulted in a final list of 33 variables for prediction model development. To use dynamic variables (those updated throughout the patients' ICU stay), previous minimum, maximum, and mean values for each patient were also incorporated, resulting in a total of 44 variables. While aggregate values may impact on practical utility of a model, they were included where found to be of sufficient importance (see Section 4.2.9). Ratios were log-transformed. Candidate predictors are shown in Table 4.1.

In model types not inherently accounting for non-linear associations, this was accounted for using restricted cubic splines for 10 predictors: age, heart rate, temperature, mean arterial pressure (MAP), and blood measures of potassium, sodium, magnesium, haemoglobin, white cell count (WCC), and pH.

In preliminary exploration of the data, serum CRP values had high missingness ($>$ 80% of all time periods—see below). Neutrophil:lymphocyte ratio was therefore used as a measure of systemic inflammation in place of CRP[238].

4.2.5 Sample size

Sample size calculations were estimated based on an LR model, as sample size calculations for machine learning methods are less well established[239]. Assuming an outcome prevalence of 7.5%[77], 64 parameters (44 predictors, of which 10 required three parameters for restricted cubic splines), and $C = 0.80$, the minimum sample size for new model development was estimated at 6206 patients[240,241]. The number of patients within the databases used in this study far exceeded this (see Section 4.2.1), suggesting a sample size large enough for model development.

Table 4.1: Candidate NOAF predictors

Variable domain	Variable	Data type
Demographics	Age	Continuous
	Sex	Dichotomous
	Weight	Continuous
	Day of ICU stay	Continuous
Comorbidities	Congestive cardiac failure	Dichotomous
	COPD	Dichotomous
	Ischaemic heart disease	Dichotomous
	Hypertension	Dichotomous
Vital signs	Heart rate^{min,max,mean}	Continuous
	Temperature^{max}	Continuous
	Peripheral oxygen saturation	Continuous
	Mean arterial pressure^{min,max}	Continuous
	Presence of atrial ectopic beats	Dichotomous
Interventions	Noradrenaline infusion rate^{max}	Continuous
	Vasopressin use	Dichotomous
	Beta blocker administration	Dichotomous
	Mechanical ventilation	Dichotomous
	Positive end expiratory pressure	Continuous
	Fraction of inspired O_2^{max}	Continuous
	Presence of a central venous catheter	Dichotomous
	Renal replacement therapy	Dichotomous
Laboratory results	Serum potassium concentration^{min,mean}	Continuous
	Serum magnesium concentration	Continuous
	Serum sodium concentration	Continuous
	Blood pH	Continuous
	Serum urea concentration^{max}	Continuous
	Serum creatinine concentration	Continuous
	Serum lactate concentration	Continuous
	Serum albumin concentration	Continuous
	Haemoglobin concentration	Continuous
	Platelet count	Continuous
	White cell count	Continuous
	Neutrophil:lymphocyte ratio	Continuous log

In models incorporating dynamic data, previous minimum, maximum, and mean values (not including current value) for certain variables were included, denoted by superscript text. ICU = intensive care unit, COPD = chronic obstructive pulmonary disease.

4.2.6 Univariable associations of admission variables with NOAF risk

I used logistic regression (LR) models to analyse the unadjusted associations between candidate predictors and NOAF during the ICU stay. For this analysis, the first available value in the initial 24 hours of ICU admission was used, excluding post-AF values. Restricted cubic splines allowed non-linear associations to be modelled.

Due to the confounding association of acute kidney injury (AKI) with both serum electrolyte concentrations and NOAF[38,242], the associations between serum [Mg^{2+}] and [K^+] and NOAF were adjusted for renal function. NOAF risk was

plotted against each variable of interest to visualise any association.

Confidence intervals around the univariable associations were derived based on the Wald method. For each predictor, the confidence interval for its estimated effect was constructed as $\beta \pm z \times SE(\beta)$ where β represents the point estimate of the predictor's effect, $SE(\beta)$ is its standard error, and z is the critical value from the standard normal distribution corresponding to the 95% confidence level. The standard errors were derived from the estimated variance-covariance matrix of the parameter estimates.

As an additional analysis of NOAF risk associated with serum $[Mg^{2+}]$ and $[K^{+}]$, I used dynamic data allowing patients' electrolyte values to vary over the ICU stay. Patient-level random sampling was performed to assess the risk associated with different serum electrolyte concentrations throughout the ICU stay with NOAF. In each resample, each patient contributed a single electrolyte value, and an LR model was built, predicting the onset of AF over the 24-hour period after the electrolyte measurement. Resampling was performed 100 times. This allowed an estimation of the association between dynamic serum $[Mg^{2+}]$ and $[K^{+}]$ with NOAF, rather than only values recorded at admission.

4.2.7 Data structure

For prediction modelling, dynamic patient data was grouped into non-overlapping three-hour periods. This was done in order to: 1) incorporate repeated measurements e.g. vital signs that change over time; 2) reduce bias that would be introduced by patients who had more frequent observations contributing more data than those with less frequent observations; and 3) keep data size and computation time feasible. Variable values were updated as they became available. Each three-hour period had a look-forward window of 24 hours within which NOAF may occur. This window did not include the immediately subsequent three-hour period.

To reduce the influence of rare extreme values, continuous variables were Winsorised at the 0.1 and 99.9% percentiles, other than age which was Winsorised at the 99.9% percentile only[243].

4.2.8 Missing data

I analysed missingness and plotted the missingness of variables (those with $> 1\%$ missing) over time. In clinical practice, certain observations and results are not available immediately on admission as time is needed for laboratory processing or initiation of physiological monitoring. Missingness was therefore likely to be largest early in the patient stay. For univariable analyses of NOAF risk, only complete cases of the variable of interest were used. For dynamic variables, values were assumed to persist until the next recording of that variable, at which point the variable value was updated (the "last observation carried forward" method). This was done to reflect how the resulting models would be used in clinical practice (a clinician using the most recent value available to them).

Where no data was available within a three-hour period, nor available to be carried forward, those data points were imputed. In developing statistical models, multiple imputation allows for pooling of model coefficients across imputations, accounting for within- and between-imputation variance. However, with models such as neural networks and GBTs (see Section 2.2.1), there is currently no equivalent of pooling coefficients, a technique demonstrated in Appendix B. Given the expected low proportion of missing data, I therefore opted pragmatically to use a single imputation of the data for model development. Missing data was imputed using a joint modelling approach based on a multivariate normal distribution[244,245]. The imputation sub-model included all variables considered for the multivariable prediction model along with the outcome variable[246].

This method assumes data were "missing at random" (MAR). This assumption states that the probability of a missing value depends only on the observed data, i.e. the missing data mechanism is ignorable given the observed data. Imputation methods rely on the MAR assumption because they use the observed data to estimate the missing data. If the missing data mechanism were dependent on the unobserved values i.e. missing not at random, then the observed data alone would not contain enough information to correctly estimate the missing data. While the MAR assumption can be difficult to verify, it can be supported by a well-specified

imputation model including key variables which may explain missingness. I visually inspected the distributions of imputed data by plotting density curves of observed data versus imputed data for each variable.

4.2.9 Model development

I developed an LR model, a GBT model, and a neural network. To account for changing risk during an ICU stay, I incorporated dynamic variables. I used variable values updated throughout the patient's ICU stay (described above) to predict the risk of NOAF over the subsequent 24-hour window. Each model was fitted to the entire dataset, incorporating each patients' longitudinal data. Hyperparameters were optimised using grid searches, detailed below.

Variable selection

To improve clinical utility and reduce the risk of overfitting, I aimed to produce a parsimonious model. For each model type, I performed variable selection through recursive feature elimination, where variables are sequentially removed based on feature importance until the cross-validated model performance reduced below a threshold in comparison to the a model comprising all candidate predictors. Hyperparameter optimisation was nested within each step of variable elimination.

The variable selection procedure was followed for each of the three model types, outlined below:

1. For each ICU-level fold with one held-out ICU (outer cross-validation):
    a. Train the model and optimise hyperparameters using *all predictors* on the training data (6 ICUs)
    b. Generate predictions in the held-out ICU (patient level data)
    c. Rank variables by importance[i]
    d. For each subset of numbers of variables S_i
        i. Keep the S_i most important variables
        ii. Train the model on the training data (6 ICUs) with S_i predictors (ICU-level inner cross-validation)
        iii. Generate predictions in the held-out ICU from the outer cross-validation
2. Calculate the average performance across the seven held-out ICUs
3. Select i predictors from the model with the fewest predictors with a C-statistic within 1% of the model comprising all candidate predictors
4. Fit the final model on the full dataset using i predictors selected in the previous step

In the above procedure models were *trained* using all data including repeated measures. To present risks and performance metrics at the patient level, models were *tested* on patient-level data. To achieve this, I performed random sampling to select a single observation set per patient[250]. Each patient was resampled five times with replacement to obtain a spread of their data over their ICU admission. Therefore performance statistics were averaged over 35 cross-validation folds (five per patient with seven ICU-level folds).

Model training

I used the technique described above (Section 4.2.9) to train the different models adjusting them over a range of architectures, hyperparameters, or penalty terms

[i]Variables were ranked using different methods depending on model type. Given the multiple nested cross-validation procedures, these methods were necessarily less intensive and faster than calculation of Shapley values (see Section 4.2.12), though potentially less robust. For the LR model, importance was based on the t-statistic of model parameters from centred and scaled covariates. For the neural network, importance was based on the network weights[247,248]. For the GBT model, importance was based on the "gain" statistic, a measure of the reduction in loss brought by a predictor to the model when the data is split based on that predictor[249].

as appropriate.

A penalised LR was developed, incorporating L1 and L2 regularisation as outlined in Section 2.2.1. Models of different predictor numbers were trained over a range of values for the penalty term (λ) and for the mixing percentage between L1 and L2 regularisation (α). For the GBT models, training was conducted over a range of hyperparameters, including learning rate, number of trees, maximum depth, and minimum child weight (see Section 2.2.1). Artificial neural networks (see Section 2.2.1) of different predictor numbers were trained over a range of architectures and hyperparameters. This included comparing the performance of single-layer fully-connected networks with that of networks of two and three hidden layers. Performance over a range of values for number of nodes per hidden layer, learning rate, and weight decay was also compared.

4.2.10 Measures to assess model performance

Internal validation

I assessed the discriminative ability of each model during the variable selection process above (Section 4.2.9). To obtain an impression of external validity, performance metrics were averaged over sub-models fitted to each permutation of six ICUs, tested on a single hold-out ICU (internal-external cross-validation)[251].

4.2.11 Selection of the final model

A single final model was selected from the three model types. Selection was based on an overall assessment of discrimination and number of variables required.

4.2.12 Model explainability

Three techniques were used to increase the explainability of the final model, expanded upon below. First, I generated Shapley values to explore variable importance at the patient level, and average values over the dataset to assess overall variable importance. I then calculated the Friedman's H-statistic for each variable to

explore model interactions. Lastly, I used partial dependence analysis to explore the relationship with predictor values and the outcome.

Variable importance using Shapley values

Shapley values are a method for interpreting the importance of individual variables in a prediction model. They are a concept derived from cooperative game theory, where each variable is assigned a score based on its contribution to the overall prediction made by the model[252]. The calculation of Shapley values involves determining the marginal contribution of each feature to the overall prediction made by the model, taking into account the interactions between the features. This allows us to attribute a specific value to each feature, which can be useful for understanding the overall behavior of the model.

Given the current set of feature values, the contribution of a feature value to the difference between the actual prediction and the mean prediction is the estimated Shapley value. There are 2^{k} permutations of variables ("coalitions"), where k is the number of variables in the model[253]. Estimating Shapley values is therefore computationally intensive. This can be addressed by sampling m random permutations, where m is considerably less than 2^{k}. This increases the variance of the estimated Shapley values versus the exact Shapley value. As m increases, the variance decreases and the estimated Shapley value more closely estimates the exact Value. I estimated the appropriate size m by plotting the standard deviation of Shapley values over a range of m for an individual instance[254].

Like many other permutation-based interpretation methods, the Shapley value method suffers from inclusion of unrealistic data instances when variables are correlated. Variable values are sampled from the variable's marginal distribution. This is valid where variables are independent. When variables are dependent, then variable values that are unlikely or impossible may be sampled[255].

Variable interaction strength

An interaction in a model occurs when the effect of one feature on the outcome depends on the value of another variable. For example, the effect of heart rate on NOAF risk may vary depending on the age of the patient. If a model makes a prediction based on two variables, we can decompose the prediction into four terms: a constant term, a term for the first feature, a term for the second feature and a term for the interaction between the two features. The interaction between two features is therefore the change in the prediction that occurs by varying the features after considering the individual variable effects.

To investigate model interactions, I calculated the Friedman's H-statistic. This is a measure of the strength of the relationship between a pair of variables and the outcome. It is calculated by measuring the difference between the observed partial dependence function and the decomposed partial dependence function without interactions. The H-statistic is $H = 0$ if there is no interaction and $H = 1$ if all of the variance of the partial dependence function is explained by the interaction. It was estimated for each variable with regards to 1) all other features, to obtain a measure of overall interaction strength, and 2) with each other feature individually, to obtain a measure of interaction strength for each two-way interaction.

Partial dependence plots

Partial dependence plots show how the *average* model predictions change over a range of variable values. To generate a partial dependence plot exploring a single variable, data for each patient is replicated multiple times, with other variables held constant while the variable under investigation v_1 is varied over a plausible range or the range observed within the dataset. Model predictions are then generated over the range of v_1, estimating how varying v_1 would affect NOAF risk for that patient. This method provides an *individual conditional expectation*. Predictions are then averaged over all patients for each value of v_1 to produce the partial dependence plot for that variable. More than one variable can be varied per patient

to explore interactions between variables, showing how average predictions change over different combinations of two or more variables.

I plotted partial dependence plots for each continuous variable to show how the *average* model predictions change over a range of values, and superimpose this on a sample of individual conditional expectation curves demonstrating how changes in the variable affects the risk of individual patients in the dataset. I explored variable interactions by plotting risk heatmaps, displaying how predictions change over different combinations of two variables.

4.2.13 External validation

The CCHIC database was used to externally validate the final model (see Section 2.1.4). Data preparation, covariates and outcome definition were the same as that in the development data.

Predictors

All predictors were available in the external validation dataset other than the presence of atrial ectopic beats. After establishing the relatively low importance of this predictor during model development (see Section 4.2.12), I deemed this dataset valid for external validation purposes. Values for this predictor were randomly imputed as a binary variable at the baseline prevalence found in the development dataset. I undertook a sensitivity analysis where model discrimination was assessed across 100 imputations of the missing variable to assess the effects on model performance.

Performance

Performance of the final model was assessed in the external validation data, evaluating model discrimination, calibration, and net benefit (see below). Confidence intervals around C-statistics and calibration metrics were estimated using bootstrap resampling, calculating bias-corrected intervals from 100 bootstrap samples each containing one observation set per patient.

Discrimination

I calculated model sensitivity and specificity at different risk thresholds and labelled ROC curves for the final model and comparator model, calculating the C-statistic. External validation performance was compared with a comparator model: an LR model developed in two ICUs in the Netherlands, outlined above (Section 4.1.2). To compare C-statistic values between the final model and the comparator model, I computed a Wald statistic, dividing the observed difference in C-statistic values by its standard error, and comparing this to the standard normal distribution in order to report a p-value[256]. The standard error was derived from the bootstrap resamples of the data.

Calibration

I evaluated calibration for the final model and the comparator model in the external CCHIC dataset. To promote a fair comparison, I recalibrated the comparator model to the development data using logistic recalibration. With this method, the observed outcome is logistically regressed against the predicted probabilities of the machine learning model. This is a two-step process that requires a calibrating model that maps the predicted probabilities of the machine learning model to the observed outcome probabilities.

Such recalibration does not change the ordering of the samples, since the logistic function is monotonic. Therefore, the C-statistic is not affected by this technique (see Section 2.2.1).

As the final model was fitted to longitudinal data, the predicted risks are initially a function of the number of three-hour time periods, rather than reflective of risks at a patient level. The model was therefore recalibrated on patient-level resamples of the training data. Predicted and observed risk was calculated over 10 risk groups using deciles in the external validation data. To examine calibration across the range of predicted values, I added a LOESS curve to the calibration plots[89]. I presented calibration curves for the comparator model before and after

recalibration to the development data. I calculated the calibration slope and calibration-in-the-large, as outlined in Section 2.2.1.

Decision curve analysis was then performed for the final model and the comparator model to assess the comparative net benefit over a range of risk thresholds (see Section 2.2.1).

4.2.14 Development of an online prediction tool

I used the *Shiny* R package[257] to develop an interactive web application which provides a predicted NOAF risk based on user-entered covariate values using the final model.

I created the user interface for the prediction application using a combination of R code and HTML elements. This included creating input fields for predictors and a button for the user to submit the covariate values and generate the predicted risk.

The application contained code to accept the user-entered predictor values and use them as inputs to the final prediction model. The model was then used to generate a predicted risk, which was first recalibrated to training data using logistic recalibration as outlined above, then displayed to the user. The application produces both a 24-hour risk, and a cumulative NOAF risk over seven days. To aid interpretability of covariate contributions to risk, Shapley values are displayed graphically for each covariate.

Input validation was incorporated to ensure that only plausible values were accepted by the app. To promote further external validation, I also included a feature where a user can upload tabular data from multiple patients e.g. an external dataset. The application then appends a predicted risk for each row of data.

4.3 Results

4.3.1 Participants

The MIMIC-IV database contained 76,540 adult index ICU admissions. Of these, 28,699 patients fulfilled inclusion criteria (Figure 4.1a). The PICRAM database contained 13,882 adult index ICU admissions. Of these, 10,385 patients fulfilled

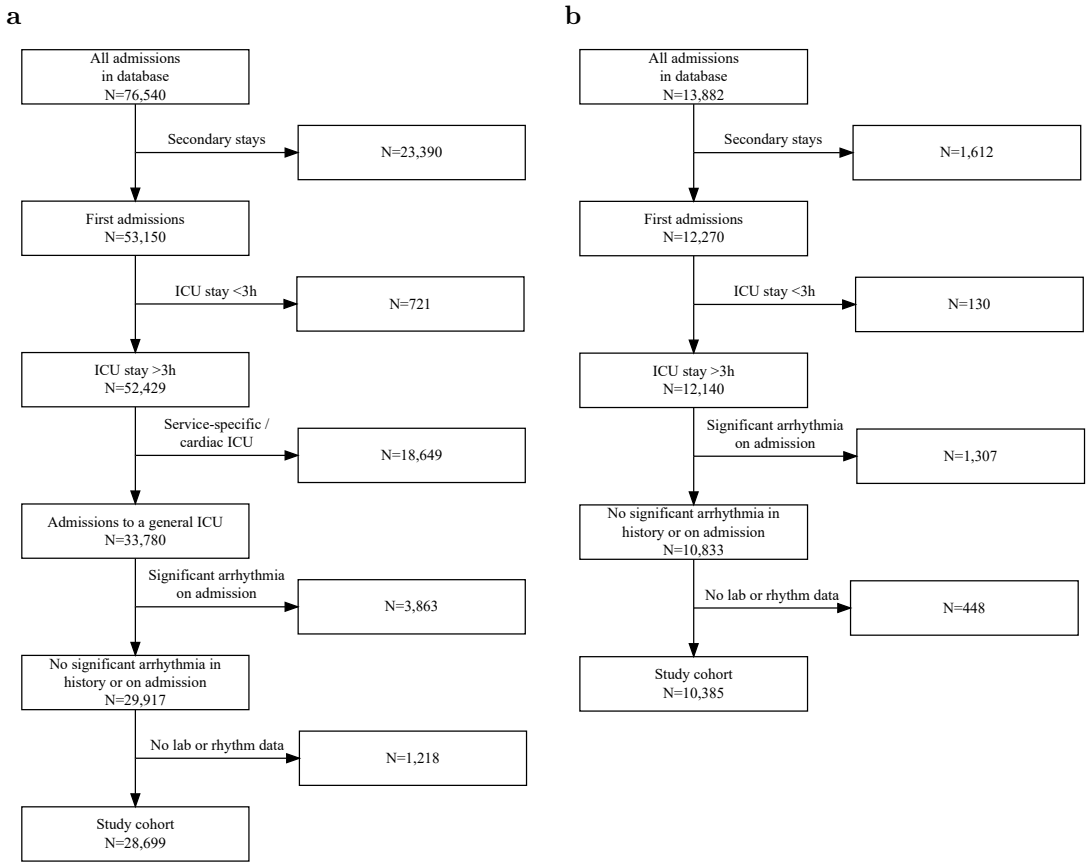


Figure 4.1: Study flow charts for a) the MIMIC-IV database and b) the PICRAM database.

inclusion criteria (Figure 4.1b). Combined, these two sources contained data from 39,084 patients from 2008-2019. Of these, 2797 (7.16%) developed NOAF during the first seven days of their ICU stay.

Patients who developed NOAF tended to be older, had longer ICU and hospital LOS, and higher ICU and hospital mortality compared with patients who did not develop NOAF during their ICU admission (Table D.8). Prior to NOAF, they were a more unwell cohort, with higher rates of invasive mechanical ventilation and central venous catheter placement. On average, they had lower blood pressure and required a higher FiO_2. The number of patients from each study ICU is shown in Figure 4.2 by NOAF status. Proportions of patients within each ICU over time are shown in Figure C.9. More recent data (2017 – 2019) contained only patients from USA ICUs.

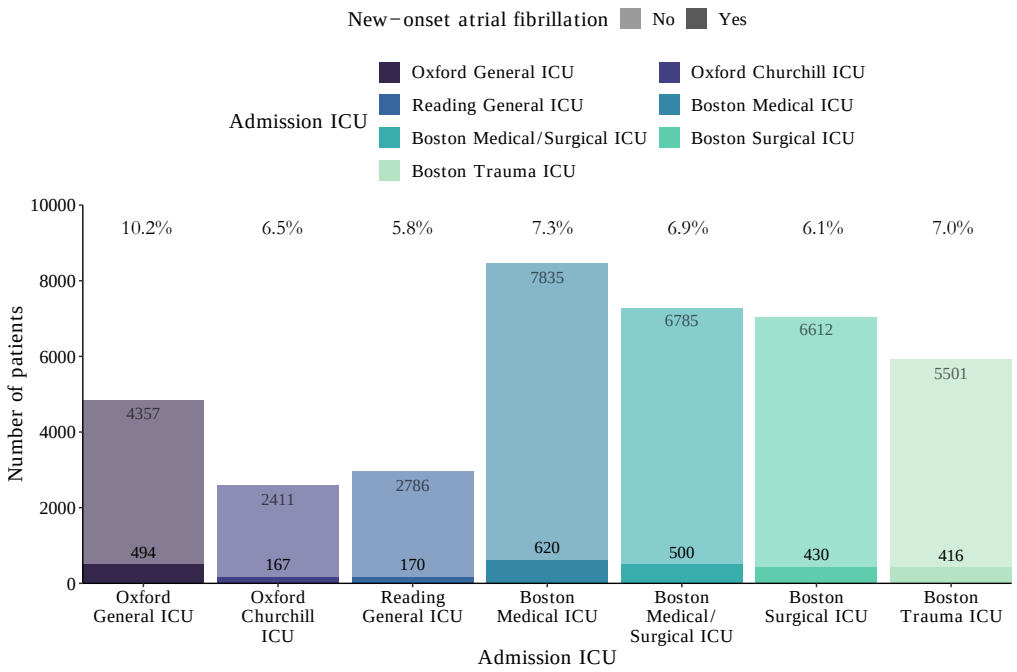


Figure 4.2: Numbers and percentages of patients by new-onset atrial fibrillation (NOAF) status and intensive care unit (ICU). Darker shades represent patients developing NOAF during their stay. Paler shades represent patients who did not develop NOAF. Total bar height is the sum of these two values, representing all included patients per unit, with % NOAF above.

4.3.2 Missing data

Data was incomplete for certain laboratory variables. Missingness decreased during the initial hours of ICU admission, as test results became available (Figure C.10).

The variables with highest missingness were, as a percentage of all three-hour time periods, the lymphocyte:neutrophil ratio (NLR; 43%), serum lactate (24.1%), blood pH (22.3%), and FiO_2 (11.6%). All other variables had a missingness $< 5\%$. Density plots comparing distributions of observed data with imputed data are shown in Figure C.11.

4.3.3 Univariable associations of admission variables with NOAF risk

Multiple admission variables were found to have significant univariable associations with NOAF, including increasing age, increasing positive end-expiratory pressure (PEEP), increasing heart rate, increasing NLR, and increasing serum [urea] (Figures

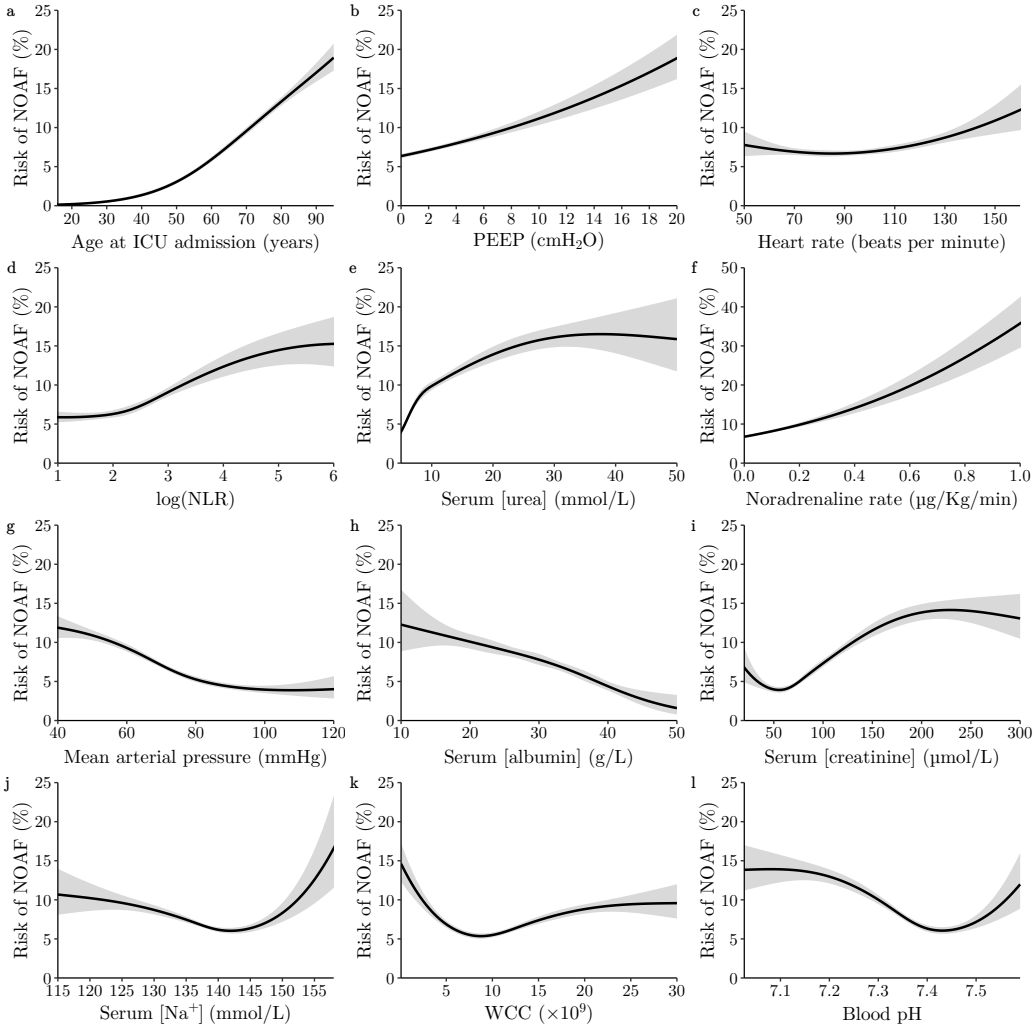


Figure 4.3: Univariable (unadjusted) associations between ICU admission variables and risk of NOAF occurring during entire admission. Shaded areas represent 95% Wald confidence intervals derived from the standard error of model estimates. WCC = white cell count, ICU = intensive care unit, PEEP = positive end-expiratory pressure, NLR = neutrophil:lymphocyte ratio.

4.3a – e, respectively). Noradrenaline infusion rate was also associated with an increasing risk, with an exponential rather than linear association (Figure 4.3f). Certain variables had a negative association, with NOAF risk decreasing with increasing MAP and serum [albumin] (Figures 4.3g and h, respectively). Some variables had a U-shaped relationship with risk, including serum [creatinine], serum [Na^+], WCC, and blood pH.

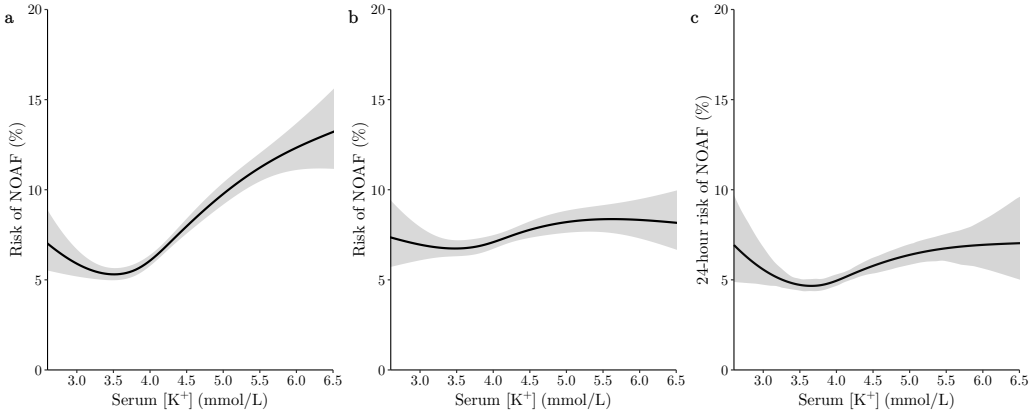


Figure 4.4: Association between serum [K^+] and NOAF risk in a) unadjusted admission values; b) admission values adjusted for renal function; and c) values across the ICU stay adjusted for renal function, assuming a serum [urea] of 8.0 mmol/L and a serum [creatinine] of 80 μmol/L. In a and b, shaded areas represent 95% Wald confidence intervals derived from the standard error of model estimates. In c, shaded area represents 95% bootstrap confidence interval of predictions from models fitted to 100 bootstrap samples of one value per patient, sampled throughout each patient's ICU stay.

Serum potassium concentration

An unadjusted analysis suggested an asymmetric U-shaped relationship between admission serum [K^+] and NOAF risk (Figure 4.4a). After adjusting for renal function, the association of increasing potassium with NOAF risk was not as marked, and the association between hypokalaemia and NOAF was no longer present (Figure 4.4b). However, when incorporating potassium values through the patients' stay, a serum value < 3.5 mmol/L appeared to be associated with increased NOAF risk (Figure 4.4c).

Serum magnesium concentration

As with serum [K^+], admission serum [Mg^{2+}] appeared to have an asymmetric U-shaped association with NOAF risk (Figure 4.5a). After adjusting for renal function, the association of increasing [Mg^{2+}] with NOAF risk was no longer evident. At [Mg^{2+}] > 0.9 mmol/L, NOAF risk reached a plateau. NOAF risk increased significantly at [Mg^{2+}] < 0.7 mmol/L (Figure 4.5b). This was also seen when modelling [Mg^{2+}] throughout ICU stay (Figure (Figure 4.5c).

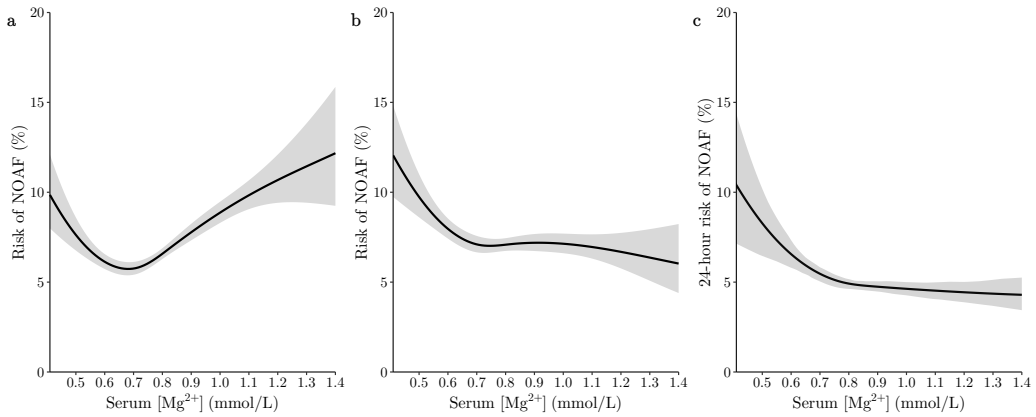


Figure 4.5: Association between serum [Mg^{2+}] and NOAF risk in a) unadjusted admission values; b) admission values adjusted for renal function; and c) values across the ICU stay adjusted for renal function, assuming a serum [urea] of 8.0 mmol/L and a serum [creatinine] of 80 μmol/L. In a and b, shaded areas represent 95% Wald confidence intervals derived from the standard error of model estimates. In c, shaded area represents 95% bootstrap confidence interval of predictions from models fitted to 100 bootstrap samples of one value per patient, sampled throughout each patient's ICU stay.

4.3.4 Model development

NOAF prediction using admission-only data

Performance of an LR model using admission-only data is outlined in Appendix B. Internal-external cross-validation performance was modest. I therefore focus here on models developed with dynamic data.

Penalised logistic regression model

Combinations of L1 and L2 regularisation conferred little performance benefit compared to no penalisation (Figure C.12). The best-performing penalty terms were $\alpha = 0.1$ and $\lambda = 0.00425$. This model required 22 variables to achieve a cross-validation C-statistic $C = 0.828$ (Figure 4.6). When incorporating fewer variables, discrimination was poorer ($C = 0.76$ using 15 variables). Discrimination performance was good ($C > 0.80$) when assessed in each of the seven held-out ICUs (C-statistic range: 0.817 – 0.843; Table 4.2).

Gradient-boosted tree model

Performance changed considerably across hyperparameters, with slower learning rates requiring more trees than faster learning rates, as would be expected (Figure C.13). The final model comprised 140 trees with a maximum depth $d = 4$ and a learning rate $\eta = 0.07$. Resampling hyperparameters were optimised at a column sampling proportion $S_c = 0.5$, and a row sampling proportion $S_r = 0.7$ for each tree. This model required only 10 variables to achieve a cross-validation $C = 0.836$ (Figure 4.6). Discrimination performance was good ($C > 0.80$) when assessed in each of the seven held-out ICUs (C-statistic range: 0.819 – 0.847; Table 4.2).

Neural network model

Across models with different predictor numbers, increasing the number of hidden layers beyond one to two or three did not result in a meaningful performance improvement (12-predictor example single-layer cross-validation $C = 0.817$, 95% CI 0.812 – 0.822; two-layer cross-validation $C = 0.818$, 95% CI 0.814 – 0.822; three-layer cross-validation $C = 0.817$, 95% CI 0.813 – 0.821; Figure C.14). Increasing the number of nodes in a single hidden layer beyond 10 resulted in limited performance improvement (Figure C.15).

The final neural network model was a single-layer fully connected network with 19 predictors, 12 nodes, and a total of 253 weights (Figure C.16). This model required 19 variables to achieve a cross-validation $C = 0.820$ (Figure 4.6). Discrimination performance was good ($C > 0.80$) when assessed in each of the seven held-out ICUs (C-statistic range 0.811 – 0.832; Table 4.2).

4.3.5 The final model

I selected the GBT model for further evaluation given its superior performance in internal-external cross-validation while requiring the fewest predictors.

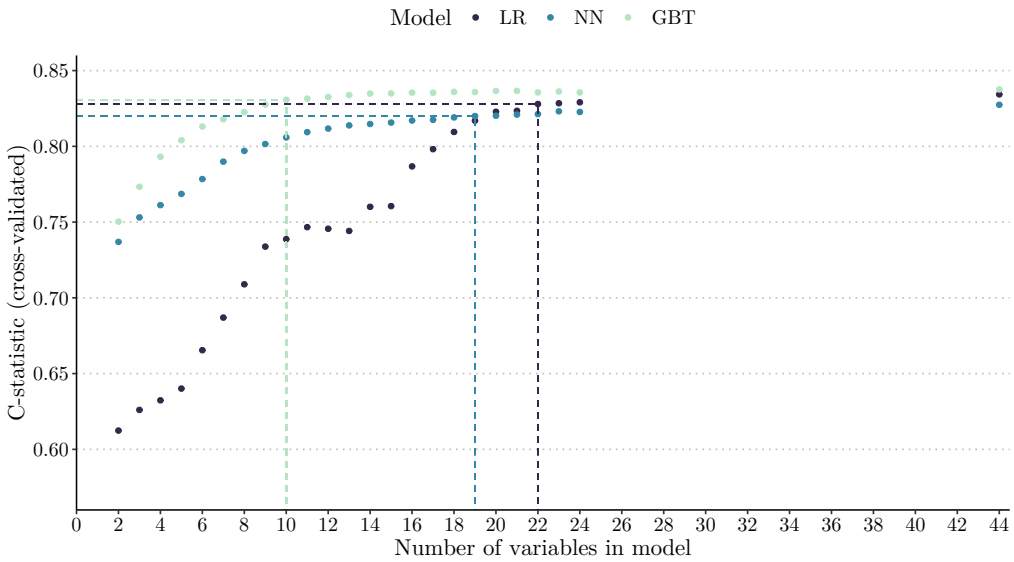


Figure 4.6: Cross-validation C-statistic values by number of variables for the three model types assessed. LR = logistic regression, NN = neural network, GBT = gradient=boosted tree.

Table 4.2: Internal-external cross-validated model C-statistic values.

Held-out ICU	LR	GBT	NN
Oxford AICU	0.821 (0.811 – 0.825)	0.819 (0.809 – 0.826)	0.815 (0.80 – 0.836)
Oxford Churchill ICU	0.828 (0.814 – 0.835)	0.824 (0.811 – 0.831)	0.816 (0.803 – 0.819)
Reading ICU	0.827 (0.824 – 0.839)	0.844 (0.835 – 0.851)	0.818 (0.786 – 0.832)
Boston MICU	0.836 (0.827 – 0.841)	0.836 (0.835 – 0.844)	0.828 (0.813 – 0.840)
Boston MICU/SICU	0.843 (0.838 – 0.850)	0.847 (0.842 – 0.854)	0.832 (0.825 – 0.839)
Boston SICU	0.817 (0.807 – 0.819)	0.818 (0.810 – 0.824)	0.811 (0.794 – 0.822)
Boston TSICU	0.825 (0.818 – 0.835)	0.826 (0.816 – 0.839)	0.820 (0.805 – 0.827)

Statistics presented: C-statistic (95% CI). LR = Logistic regression, GBT = Gradient-boosted tree, NN = neural network, ICU = intensive care unit, AICU = adult ICU, TSICU = Trauma surgical ICU, SICU = Surgical ICU, MICU = Medical ICU.

The final model contained the following 10 predictors:

1. Age
2. Day of ICU stay
3. Heart rate
4. FiO_2
5. Maximum MAP during stay
6. Serum [urea]
7. Maximum [urea] during stay
8. Presence of atrial ectopic beats
9. Noradrenaline rate
10. Presence of a central venous catheter

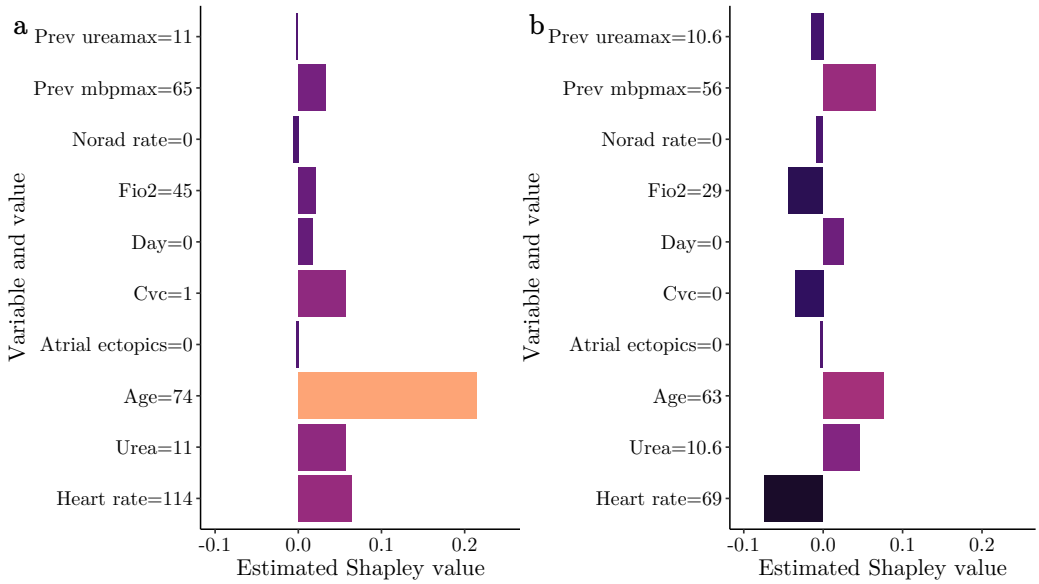


Figure 4.7: Shapley values for two example patients in development data.

Variable importance

The number of m permutations to estimate Shapley values (see Section 4.2.12) was determined visually by plotting the standard deviation of Shapley values over a range of m from a sample of patients (See Figure C.17 for an example using 10 patients for the age and FiO_2 variables). To achieve balance between stability of results and computation time, $m = 100$ was selected.

The mean absolute Shapley values demonstrated a range of variable importance for the 10 included predictors. The predictor with the most importance was patient age (Figure C.18). Example Shapley values for two patients from the development data are shown in Figure 4.7. These plots demonstrate contributions from individual variables to risk. The patient in panel a, for example, has increased risk conferred by an age of 74, an elevated heart rate, and the presence of a central venous catheter. Conversely, certain variable values confer a reduced risk e.g. a heart rate of 69 and an FiO_2 of 29% for the patient in panel b.

Interaction strength

Analysis of interaction strength revealed age as the variable with the most variable-variable interactions (Figure C.19). Analysis of each two-way interaction revealed

that interactions were frequent between multiple variables. Variables including age and heart rate had interactions with most of the other variables, whereas the presence of atrial ectopic beats had limited interactions (Figure C.20).

Partial dependence plots

Partial dependence plots (see Section 4.2.12) demonstrated generally monotonically increasing average NOAF risk associated with the continuous variables, other than day of ICU stay and highest MAP which had a negative association (Figure 4.8). Individual conditional expectation plots revealed inter-patient variability in both predicted baseline risk and how that risk changed over different predictor values.

Examples of two-way partial dependence plots demonstrating interactions are shown in Figure C.21. In these examples, it can be seen how an increasing FiO_2 does not increase AF risk in patients under the age of 50 years, whereas a 66 year-old patient could sit in three different risk groups depending on their FiO_2.

Missing data

The variables included in the final model had low missingness, with 7 / 10 predictors missing in less than 1% of the data points. Only FiO_2 and serum [urea] (and therefore previous maximum serum [urea]) had notable missingness at 11.6% and 4.2%, respectively.

4.3.6 Model performance: External validation

The final GBT model was assessed in previously unseen data from the CCHIC dataset (see Section 2.1.4). Its performance was compared to one published comparator model[52].

CCHIC data

The CCHIC database contained 33,844 adult index ICU admissions. Of these, 29,833 patients fulfilled the inclusion criteria, with 2,686 (9.0%) developing NOAF during their ICU admission.

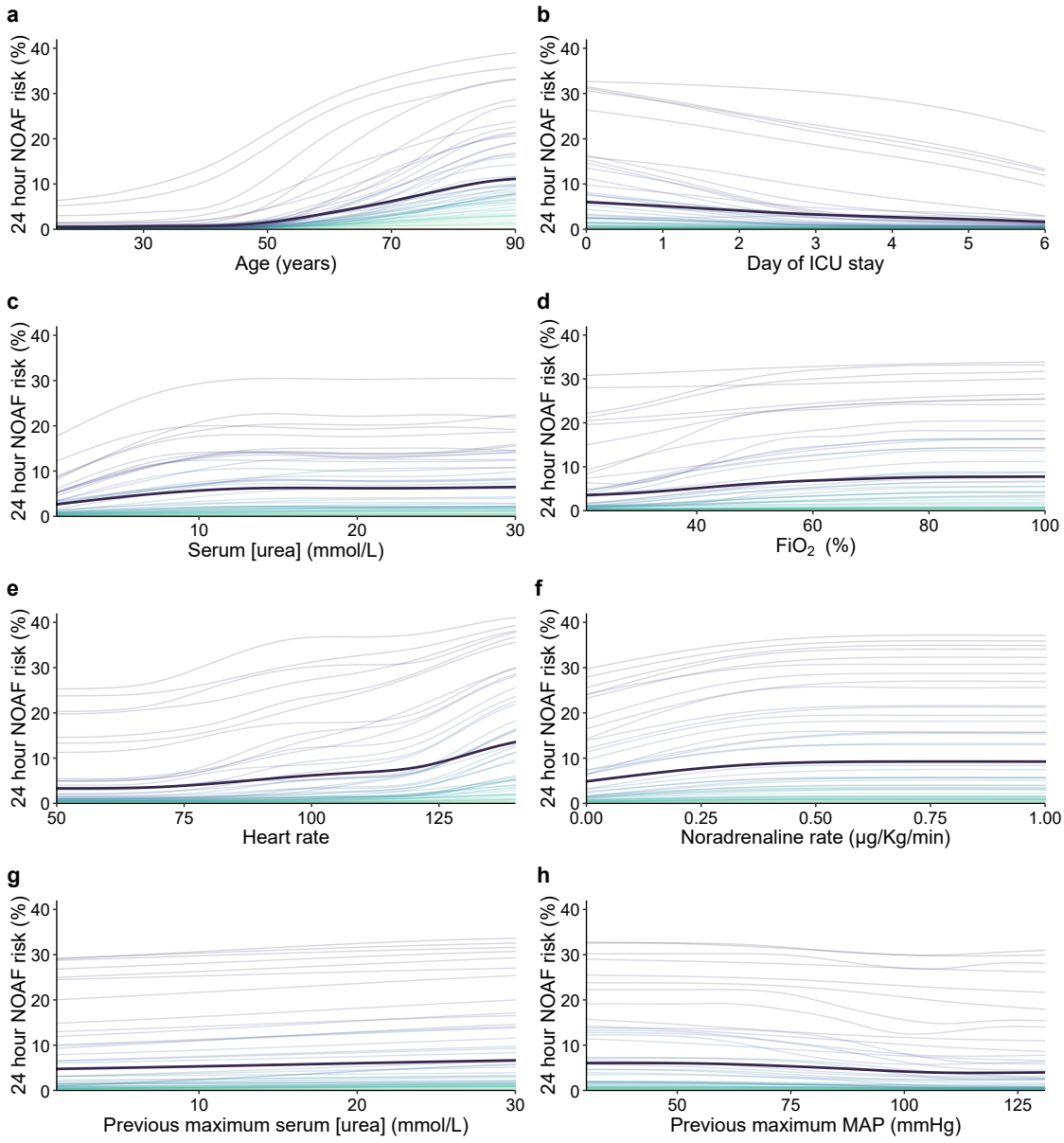


Figure 4.8: Partial dependence plots (black lines) with superimposed individual conditional expectation curves (coloured lines) for each of the eight continuous variables in the final gradient-boosted tree model: a) age; b) day of ICU stay; c) serum [urea]; d) fraction of inspired oxygen (FiO_2); e) heart rate; f) noradrenaline rate; g) previous maximum serum [urea]; and h) previous maximum mean arterial pressure (MAP).

The admission characteristics of included patients are displayed in Table D.9. Patients who developed NOAF tended to be older, had longer ICU and hospital LOS, and higher ICU and hospital mortality compared with patients who did not develop AF during their ICU admission, findings consistent with the development data.

4.3.7 Comparative model performance

Comparison between the GBT model and the comparator model revealed that the GBT model had a significantly better discriminative ability than the LR model (GBT C = 0.812, 95% CI 0.805 – 0.822 versus LR C = 0.786, 95% CI 0.778 – 0.801, p = 0.0003).

External validation calibration plots for the GBT, the comparator model, and the recalibrated (to development data) comparator model are shown in Figure 4.10a, 4.10b, and 4.10c, respectively. Both models were quite well calibrated at lower risks (< 20%), with the comparator model being better calibrated after recalibration than the GBT model at risks < 10%. The GBT model was, however, considerably better calibrated at higher risks, with the LR model predicting increasing risks without any corresponding increase in observed probability. This suggests that the GBT model may be more reliable for making predictions in high-risk patients.

Lastly, the GBT model was found to have a superior net benefit (Figure 4.11) over a range of clinically important risk thresholds. At very low risk thresholds (< 3%), the two models had equivalent net benefit. Both models were more useful than an "intervene on all" or an "intervene on none" strategy.

The tables below (4.3 and 4.4) summarise the impact of using both models to identify high risk patients for randomisation in a preventative trial. Of 1,000 example patients each assessed at one point in their ICU stay, 58 (5.8%) would be expected to develop NOAF over the subsequent 24 hours. Using a risk cutoff of 5.5%, the GBT model would identify 251 patients as high risk, with 40 of these going on to develop AF at 24 hours (a positive predictive value PPV = 15.9%). In order to achieve the same number of true positive cases (40), the comparator model would highlight 300 patients (PPV = 13.4%), an important rise when considering

resource allocation in preventative trials. The GBT model displayed superior PPV across a range of clinically meaningful risk thresholds.

Table 4.3: Model performance in 1000 patients at different risk thresholds: GBT model.

Risk threshold	TP	FN	TN	FP	PPV
2.5%	49	9	595	346	0.12
5%	41	17	715	227	0.15
7.5%	34	25	781	161	0.17
10%	28	30	826	115	0.20
12.5%	23	36	858	84	0.21

TP = true positives; FN = false negatives; TN = true negatives; FP = false positives; PPV = positive predictive value. Each row may not total 1000 due to rounding.

Table 4.4: Model performance in 1000 patients at different risk thresholds: comparator model.

Risk threshold	TP	FN	TN	FP	PPV
2.5%	55	4	385	557	0.09
5%	46	12	589	353	0.12
7.5%	38	21	715	227	0.14
10%	31	28	797	145	0.17
12.5%	23	35	845	97	0.19

TP = true positives; FN = false negatives; TN = true negatives; FP = false positives; PPV = positive predictive value. Each row may not total 1000 due to rounding.

Sensitivity analysis

External validation performance was assessed against multiple imputations of the missing atrial ectopics variable. Model performance did not differ from the main findings, nor vary considerably across the imputations (C = 0.814, 95% CI 0.809 – 0.820).

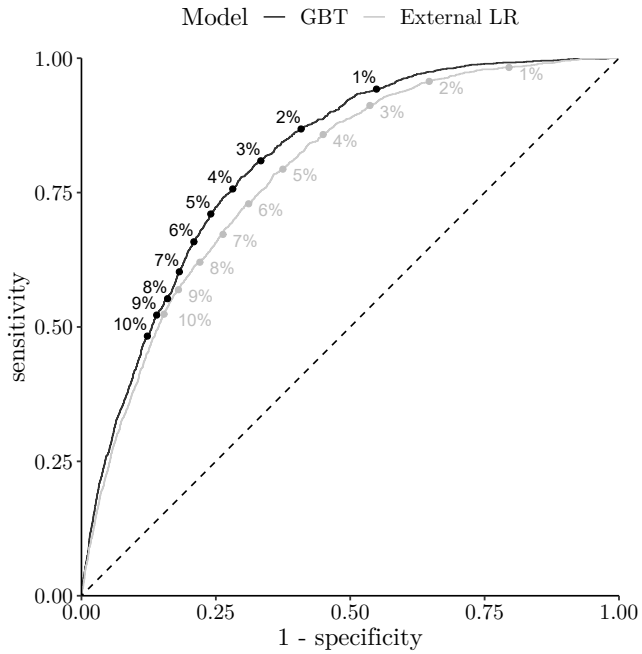


Figure 4.9: External validation ROC curves with associated risk thresholds for the GBT model (black) and the comparator model (grey).

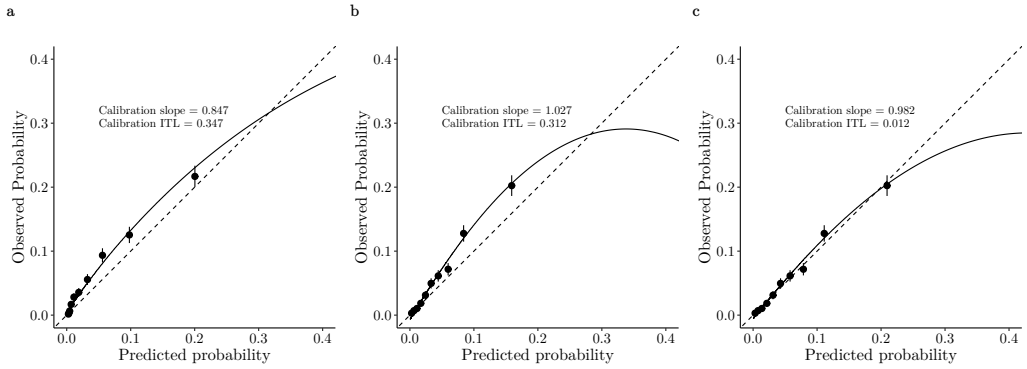


Figure 4.10: External validation calibration plots for a) the GBT model; b) the comparator model; and c) the recalibrated comparator model. ITL = in-the-large.

4.3.8 Development of an online prediction tool

To improve model transparency, I created an online prediction application with tools to promote external validation. The final model was named the *METRIC-AF* prediction model (Machine lEarning arrhyThmia Risk in Intensive Care) which references its 10 predictors, and its quality as a system of measurement or evaluation. The web-based prediction application is available at https://metric-af.shinyapps.io/metric-af/.

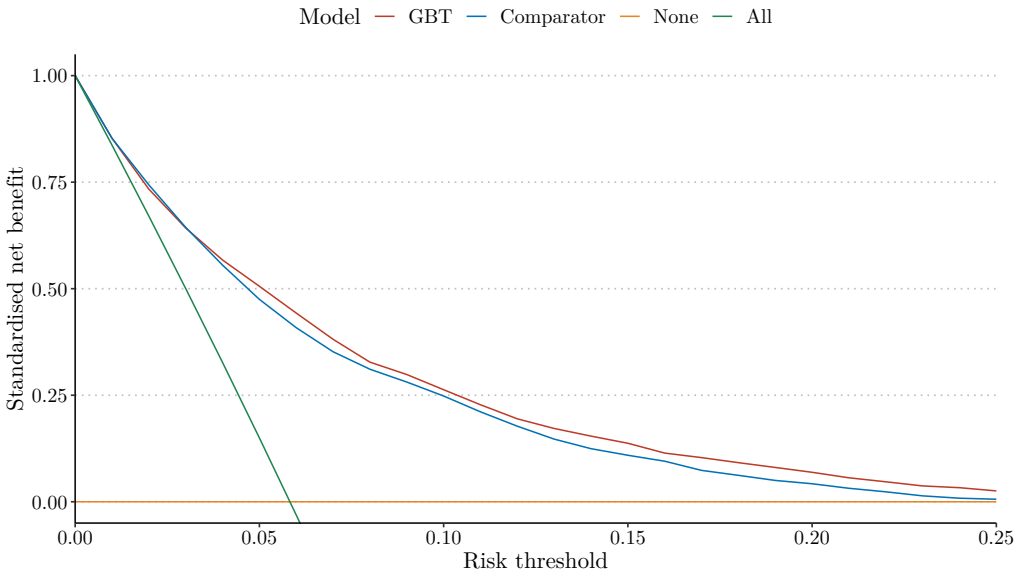


Figure 4.11: External validation decision curves for the GBT (red), and the comparator model (blue).

While publicly available, the web page has not been advertised and the site contains a disclaimer that it should only be viewed by members of the study team. Once the user has entered values for the 10 predictors, they can click an "Estimate risk" button which then displays the estimated 24-hour NOAF risk, Shapley values for predictor contributions, and a bar chart of cumulative NOAF risk over seven days, which assumes the predictor values remain constant (other than day of admission).

To promote independent external validation, the user may also upload their own dataset and the web application contains a function to estimate the NOAF risk for each row of data. The user can then download their data with estimated results appended. Instructions for this process are found at the bottom of the web application page. A screenshot of the METRIC-AF application is shown in Figure 4.12 (table upload instructions not shown).

4.4 Discussion

I present a detailed analysis of associations between candidate risk factors and NOAF in patients admitted to an ICU. I then present a NOAF prediction model developed using data from three centres in the UK and the USA, building on the

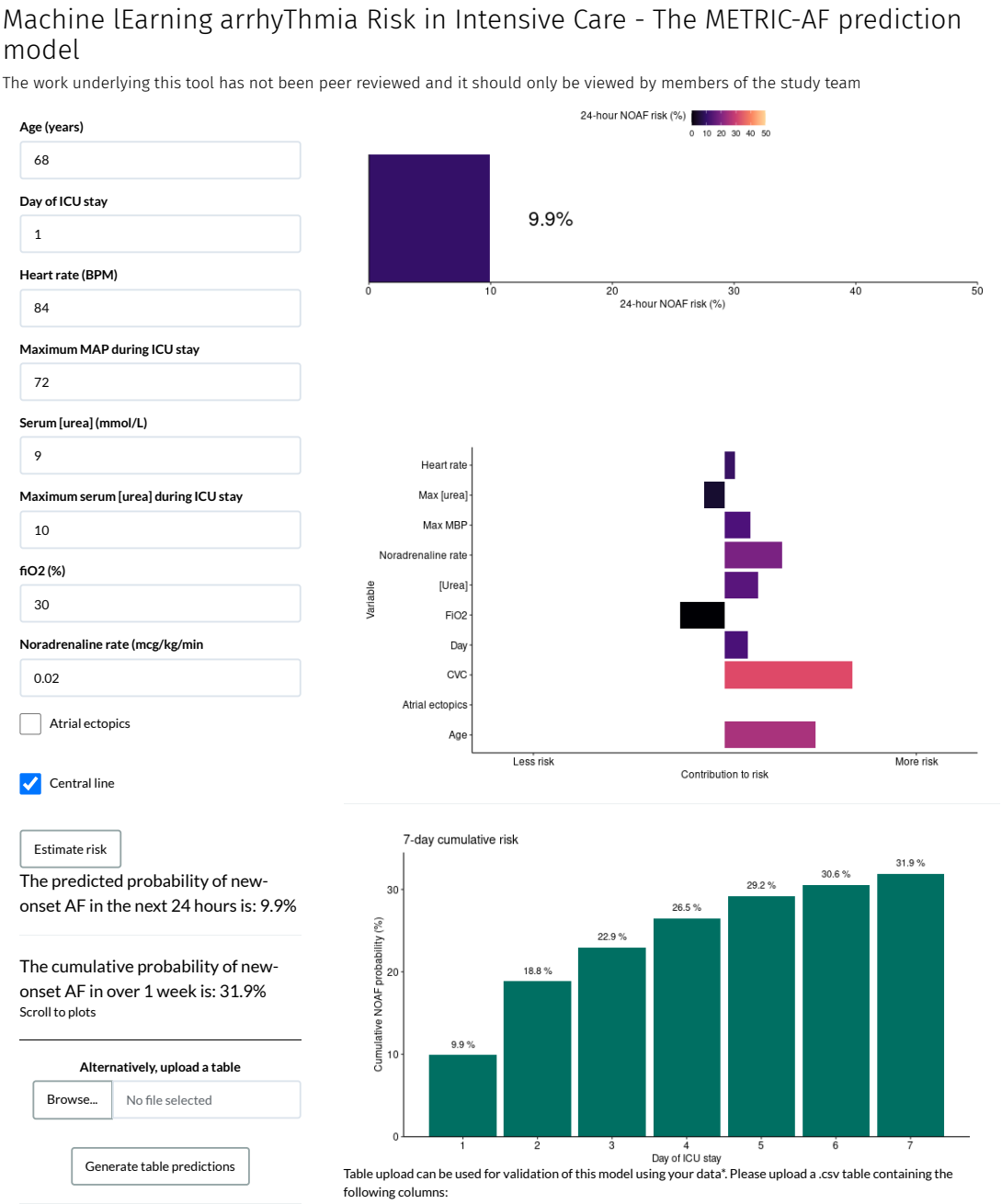


Figure 4.12: A screenshot of the METRIC-AF prediction app. The topmost image displays the patient's 24-hour risk. The Shapley values for each covariate are displayed graphically below this. Finally, a bar chart of cumulative risk over seven days is displayed, assuming covariates other than day of ICU stay are constant.

systematic variable selection of the previous chapter. This model outperforms the current best NOAF prediction model[52] in one assessment of external validity using multi-centre UK data. I provide a user-friendly tool to estimate NOAF risk using 10 readily available clinical variables, increasing transparency and promoting independent external validation.

4.4.1 Univariable associations with NOAF risk

I found evidence supporting associations of multiple variables with NOAF risk. This analysis demonstrated non-linear functional forms of these associations. While many of these associations have been demonstrated in previous literature[44,52,258], they have not been studied in this degree of detail. For example, WCC had a U-shaped relationship with NOAF risk. While WCC has been previously implicated in NOAF risk, this has only been analysed in a linear fashion[52]. My analysis reveals that a low WCC may be more strongly associated than a high WCC, consistent with a previous finding of immunocompromise as a NOAF risk factor[52]. I also found U-shaped unadjusted association with blood pH and serum [sodium], which have not been previously shown.

Similarly, while the association of NOAF risk with vasoactive medications has been previously shown, this has been either as a function of number of vasoactive medications[44], or as a binary variable[52] with no assessment of risk over a range of infusion rates. I demonstrate an exponential association where the increase in NOAF risk from a noradrenaline infusion rate of 0.30 to 0.35 μg/Kg/min is higher than with a rate increase from 0.10 to 0.15 μg/Kg/min.

Evidence for the association of hypokalaemia with NOAF was limited. In unadjusted analysis, a U-shaped association was evident, with higher and lower serum [K^+] being associated with increased NOAF risk. After adjustment for renal function, a serum [K^+] $<$ 3.5mmol/L appeared to be associated with a slight increased NOAF risk. This finding is supported by one retrospective multi-ICU study demonstrating increased risk of arrhythmia at serum [K^+] $<$ 3.5 mmol/L. I found NOAF risk also increased with increasing [K^+]. Klein Klouwenberg *et al.*[52]

similarly found a U-shaped association between serum [K^+] and NOAF risk, evident in their multivariable model. It is likely that raised serum [K^+] is a marker of disease severity and renal impairment, both of which confer an increased NOAF risk during critical illness[227]. There is some mechanistic plausibility behind an association between hypokalaemia and NOAF (see Section 3.4.2), though this is less clear than in ventricular arrhythmias. Furthermore, an association is likely to be masked by the early recognition and treatment of hypokalaemia in the emergency department or theatres made possible by point-of-care blood tests. A causal analysis of hypokalaemia and NOAF risk was beyond the scope of this chapter.

Unadjusted analysis of serum [Mg^{2+}] revealed a U-shaped association with NOAF risk. An increased NOAF risk with higher serum [Mg^{2+}] would be at odds with current mechanistic understanding and some clinical evidence (Section 3.4.2). After accounting for confounding introduced by the elevated [Mg^{2+}] seen in AKI (a factor associated with increased NOAF risk[227]), only low [Mg^{2+}] were associated with increased NOAF risk. While no causality can be attributed, NOAF risk appeared to increase at serum [Mg^{2+}] < 0.7 mmol/L. This finding is supported by evidence of NOAF prevention in patients undergoing cardiac surgery using prophylactic magnesium supplementation[53,191]. In the general ICU, Wilson *et al.* undertook an instrumental variable analysis finding that magnesium supplementation was associated with a small (3%) relative reduction in NOAF risk in a single centre[259]. A UK-wide survey demonstrated the common practice (60% of respondents) of targeting serum [Mg^{2+}] ≥ 1.0 mmol/L in patients who have developed NOAF[70]. The optimal serum [Mg^{2+}] for NOAF prophylaxis is not known, although based on these findings, a target of ≥ 1.0 mmol/L may be unnecessary. Magnesium targets should be further investigated in an RCT of patients with high NOAF risk.

4.4.2 Prediction model development

I developed and compared three predictive models, using a large international dataset which allowed ICU-level cross-validation to mitigate against overfitting. I demonstrated that a GBT model was superior to the other two models in terms of

best performance with the fewest variables. My final model uses objective, readily available clinical variables. The variables are not subject to differences in definitions across professional societies and countries, seen with, for example, comorbidity diagnoses. The model requires only one blood test result, increasing portability over other models which require less-often-performed blood tests. All 10 variables should be easily available in most ICUs in developed nations. The model is designed such that it can be used at any point during a patient's admission, using the most recent clinical values (with the addition of previous best blood pressure and highest serum urea concentration) to predict NOAF risk over the subsequent 24 hours.

Good discrimination was noted across all three models during internal-external cross-validation. However, the GBT model was selected due to superior parsimony. While model complexity and variable numbers need not be limited where a model is integrated into a hospital's clinical information system, this integration is time-consuming and would likely limit the model's clinical utility. Its availability as a web-based application means it does not require the presence of a clinical information system, extending its use to any ICU with a device and an internet connection.

I found evidence of numerous interactions in the final model. Age interacted with many other variables, suggesting that the degree to which changes in other variables, e.g. how FiO_2 affected NOAF risk changed depending on patient age. A simple additive model such as the comparator model that considers each predictor variable in isolation will not account for these complex relationships. In this study, the use of a model accounting for these interactions led to improved predictive discrimination.

The variables included in the final model should be considered. While not necessarily causal, the 10 selected predictors carry considerable information about patient risk. The majority can be thought of as representing illness severity, in the context of patient age and time in the ICU. Key organ failures are represented by cardiovascular: heart rate, highest MAP, and noradrenaline requirement; respiratory: FiO_2; and renal variables: serum urea concentration, and maximum serum urea concentration. Whether the presence of a central line causes AF is unknown. AF may be induced by myocardial irritation from malpositioned central catheters[217] or by

guidewires inserted into the atrium during catheter insertion[218]. It remains unclear whether central venous catheters pose an increased NOAF risk when well positioned, and their presence may simply represent a higher illness severity, associated with NOAF. The presence of atrial ectopics is not found in other ICU scoring systems and likely represents a specific AF risk, rather than a surrogate of illness severity. Atrial ectopic beats can initiate episodes of AF[29] and the inclusion of this variable to a community AF risk scoring system considerably improved its discriminative ability (C-statistic increase from 0.65 to 0.72)[260].

While potentially important risk modifiers, serum $[K^+]$ and $[Mg^{2+}]$ were not included in the final model. Serum urea (along with previous maximum urea) was the only blood result in the parsimonious final model. Chronic kidney disease is associated with increased AF risk in the community, independent of age and other clinical risk factors[261]. There are likely numerous mechanisms underlying this association, including resulting hypertension[262] and activation of the renin-angiotensin-aldosterone system[263,264]. In the ICU, NOAF and AKI share many of the same risk factors, including increasing age, coronary heart disease, illness severity, and use of vasoactive medications[265]. Elevated urea as a NOAF risk factor may be capturing patients with a demonstrated AKI susceptibility, and therefore also at elevated NOAF risk. Furthermore, certain organic compounds that accumulate in the body in states of chronic or acute renal impairment have been linked to cardiovascular disease and arrhythmogenesis. These "uraemic toxins" can precipitate inflammation and generation of reactive oxygen species (ROS)[266]. They have also been shown to lead to atrial hypertrophy and fibrosis in vitro[267]. One uraemic toxin indoxyl sulfate (for which serum urea concentration would be a surrogate) has been shown to increase pulmonary vein and atrial arrhythmogenesis in animal studies[268]. Whether serum urea concentration has a causal association with NOAF is unclear, but this study suggests it is nevertheless an important predictor.

Associations between serum electrolyte concentrations and NOAF risk are biologically plausible and supported by some observational studies, including work

in this chapter. However, neither serum [K^+] nor serum [Mg^{2+}] were of sufficient statistical importance in this data to be included in the final prediction model.

Variables selected were similar to those in the admission model outlined in Appendix B. Of the selected variables in the ML model, eight were candidate variables in the admission model (maximum serum [urea] and day of ICU stay were not appropriate in a model using admission-only data). Of these, six were included in the admission model. This admission model was developed using the current best practice in statistical variable selection. This lends credibility to the variable selection process used in this chapter.

Important insights into suitability of different modelling approaches to multi-centre ICU data were gained during model building. In building the neural network model, the use of additional layers provided no meaningful improvement in performance. Furthermore, performance reached a plateau with the addition of nodes in the single hidden layer. While multi-layer neural networks have been widely used for a variety of prediction tasks, their superiority over single-layer neural networks has not been established[269,270]. In the present study, we compared the performance of a more complex neural network with that of a single-layer neural network for the prediction of arrhythmias. Despite the increased capacity and representational power of the multi-layer neural network, the results showed that the single-layer neural network performed comparably in terms of discriminative ability. The complexity of a model does not reflect its predictive ability. The more complex networks may have overfit the development data, resulting in no further performance improvements beyond a single-layer fully-connected network. These findings demonstrate the importance of well-specified cross-validation to mitigate overfitting. These results have important implications for the development of clinical prediction models, highlighting the need for careful evaluation and comparison of different model architectures during model development.

4.4.3 External model validation

In the external validation study, the final model outperformed the comparator NOAF prediction model in terms of its discriminative ability and net benefit. Analysis of comparative positive predictive value suggested it would more efficiently enrich any preventative study of high-risk patients than its comparator model. Unlike its comparator, each variable is simple (no composite variables) and widely available, demonstrated by low missingness across UK and USA data.

Calibration was similar after recalibrating the comparator model to the development data. The final GBT model demonstrated better calibration at very high risks. Due to case-mix differences, there was a lower incidence of NOAF in the USA ICUs, which may partially explain for the METRIC-AF model's slight underprediction of risks in the UK-only external validation dataset. The comparator model's calibration performance may be in part due to the similar setting of the development and external validation data (both Western Europe). When tested on USA-only data, the comparator model was found to perform poorly in terms of discrimination and calibration[228]. Internal-external validation of the final GBT model suggests good performance in USA-only data. The comparator model tended to overpredict risks in higher-risk groups. This may be due to the linear nature of the model being unable to temper predicted risks as variable values increase.

4.4.4 Strengths and limitations

This work has a number of strengths. Firstly, I used a large dataset comprising international data from seven ICUs of different types, increasing the generalisability of the results. The variables analysed were suggested through a Delphi process to improve model face validity and reduce the probability of spurious associations sometimes highlighted through data-driven approaches[271]. I used the last observation carried forward method to account for missing data in longitudinal variables. This method was chosen to reflect how the model would be used in practice—a clinician would use the most recent values available to them as model inputs. This contrasts with the comparator model which was developed using interpolated data[52]. In

this way, variable values were inferred based on past *and future* values, the latter not being known to the clinician at the time of model use. That model was thus trained on data not available at the time of use. I performed a rigorous and computationally-intensive variable selection process, where models were iteratively re-trained using cross-validation after each variable was eliminated. This method provided performance estimates that incorporated the variation across ICUs due to feature selection. Failure to assess performance externally to the selection process can result in selection bias, resulting in models which perform poorly in new data[272]. I undertook a robust comparison of my model with one previously published comparator model. I used European (UK) data as the external validation dataset. This was done to ensure the comparator model, also fitted to European data, was not disadvantaged by the external validation case mix. Despite the potential disadvantage of 1) different test data case mix, and 2) a missing predictor in the test data, the GBT model outperformed the comparator model.

Models that predict complications during critical illness are limited by the competing risk of death. Many complications share predictors with those of ICU mortality, for example increasing age raising the risk of both NOAF[38] and dying during the ICU admission[273]. Given the shared predictors and direction of association, as NOAF risk increases, so will the risk of dying. For example, 5 of the 10 variables included in the final NOAF prediction model are included in the ICNARC mortality prediction model[274]. While the functional relationship between these variables and the outcome differ, there will be a degree of shared association, limiting the discriminative ability of any NOAF prediction model that includes physiological variables or markers of disease severity. Complex models can however go some way to accounting for this limitation. In my final model, for example, at very high serum [urea] values, NOAF risk stops increasing and reduces slightly. While a physiological explanation is possible, an alternative explanation is that the model is identifying patients at physiological extremes who may in fact die sooner than developing NOAF.

The external validation performance of the final model was limited by the absence of data for one predictor—atrial ectopy. As the presence of atrial ectopics was the least important of the included predictors in the final model, I considered that external validation using this data was valid. As the observed presence of atrial ectopics is informative, the external performance of the final model may exceed that reported in this chapter if tested in data which includes this data. Sensitivity analysis showed consistent performance across multiple imputations of this missing predictor.

The encoding of ethnicity resulted in low numbers of patients in certain racial categories. This led to unstable model estimates during ICU-level cross-validation. Rather than arbitrarily re-code ethnicity into larger groups, it was not used as a predictor. The impact of ethnicity on AF risk is unclear (see Section 3.4.1), and future work should assess the differential performance of the METRIC-AF model on sex- and ethnicity-based subgroups.

Echocardiographic parameters including left ventricular ejection fraction and left atrial diameter are associated with ICU-acquired NOAF in unadjusted analyses[128,275]. These findings are consistent with echocardiographic predictors of NOAF in the general population[125]. While echocardiographic parameters may improve the discriminative ability of a NOAF prediction model, users of the model would need up-to-date echocardiographic data for each patient, prior to AF onset. Inclusion of such variables would therefore limit the clinical utility of a NOAF prediction model. Furthermore, in the community-based Framingham Heart Study model, additional incorporation of echocardiographic measurements to enhance the risk prediction model only slightly improved the C-statistic by around one percent[222]. While echocardiographic parameters may be associated with AF risk, the additional information conferred above other more important predictors may therefore be small. Future studies in ICUs may consider evaluating the additional benefit of including focused echocardiographic data in NOAF prediction.

4.4.5 Future work

While the METRIC-AF model has been subjected to external validation in a large, multi-centre dataset, its predictive performance should be further evaluated. To encourage external validation, I have incorporated a function in the predictive tool which will append predicted risks to any uploaded table containing the relevant predictor values. The performance of the METRIC-AF model should be assessed in patient cohorts spanning diverse demographic subgroups, healthcare systems, and countries.

Comparison against existing community AF prediction models was not undertaken here. Future work should investigate the comparative performance of these models in the ICU setting. Furthermore, more parsimonious iterations of the METRIC-AF model should also be assessed to improve utility in the clinical trial setting.

The METRIC-AF model could be used as a clinical prediction tool to highlight high-risk patients. However, the evidence for preventative strategies is limited[38]. It may, however, be used as a means of trial sample enrichment. For example, in a theoretical study of a NOAF prevention intervention, the METRIC-AF model may be used to highlight high-risk patients for randomisation. Future development work may include assessing the impact of METRIC-AF-generated risk thresholds on the theoretical recruitment rate for such a study. Recruitment rate (reducing with increasing risk thresholds) could be balanced against incidence of AF (increasing with increasing risk thresholds); see Chapter 7. These results could be used to determine the optimal risk cutoff that would result in the highest rate of patients developing NOAF in the study.

Characteristics and treatment of NOAF

5.1 Background

While a deeper understanding of risk prediction and modifiable risk factors may ultimately serve to reduce NOAF risk, optimal treatment once NOAF occurs in the ICU remains an important and poorly-understood aspect that requires further research. International guidelines for the management of AF do not apply to this specific patient population.

5.1.1 Scoping review

As part of a larger NIHR study and team, I took part in a scoping review to identify up-to-date available evidence on NOAF treatments specific to patients treated in an ICU[276] (see "Contributions", Section 1.5). The findings from this scoping review were instrumental in shaping the methodology for a subsequent cohort study.

Scoping review methods

A comprehensive literature review was conducted, encompassing the Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica database (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Web of Science, OpenGrey, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, International Standard Randomised Controlled Trial Number registry (ISRCTN), ClinicalTrials.gov, EU Clinical Trials Register, additional World Health Organization International Clinical Trials Registry Platform (WHO ICTRP), and National Institute for Health Research (NIHR) Clinical Trials Gateway databases in March 2019.

We included studies evaluating treatment or prevention strategies for NOAF or acute anticoagulation in general medical, surgical or mixed adult ICUs. Study details were extracted onto study-specific forms, including population characteristics, intervention and comparator(s), methods addressing confounding, results, and recommendations for future research.

Scoping review findings summary

The evidence base for NOAF management for patients admitted to an ICU was limited. Of the 25 primary studies included, only two were RCTs[65,66]. Many studies identified in our review were non-comparative studies (i.e. lacking a comparator group, n = 12). Only three of the non-randomised comparative studies attempted to control for confounding factors[69,156,277]. In the studies which used more robust approaches, there were nevertheless still concerns about how bias arising from study design or analysis (or both) may have affected their results.

We observed considerable heterogeneity across studies in defining NOAF, treatment doses (e.g. total dose ranging from less than one gram to eight grams for amiodarone)[41,278,279], administration (e.g. bolus or continuous infusion), and time points to assess conversion to sinus rhythm (e.g. within 2[66], 12[65], and 24 hours[41,280]).

The limited evidence from our review[67,281,282] suggested that beta blockers may be equivalent to amiodarone for rhythm control. Although reduced mortality was reported in those who received beta blockers compared to those who received amiodarone[69,282], these findings may be due to residual confounding; beta blockers are often reserved for patients who are the most cardiovascularly stable, as concerns exist that their use way worsen cardiac output or hypotension in unstable patients. Studies that do not adequately account for these differences are therefore likely to find improved outcomes associated with beta blocker use. Despite this, some review articles argued that beta blockers may be a reasonable first-choice treatment due to the current evidence of decreased mortality and improved heart rate control[47,283,284]. Two opinion pieces also favoured the use of beta blockers as the initial pharmacotherapy, given the limited and indirect evidence[63,285]. In contrast,

five reviews discussed amiodarone as a potentially effective treatment, though it was also recognised that amiodarone has potentially significant side-effects[194,286–289].

Calcium channel blockers (CCBs) appeared to be less effective for conversion to sinus rhythm when compared with beta blockers, and associated with more hypotension than amiodarone[65,66].

Comparative evidence for or against electrical cardioversion for patients with ICU-acquired NOAF was not identified. Electrical cardioversion should be considered in patients where AF is contributing to marked haemodynamic instability. NOAF often occurs alongside haemodynamic instability (e.g. during sepsis) but is more likely to be a significant *contributor* to hypotension where ventricular rates are very high or where there is underlying structural heart disease[276].

A scoping review-informed analysis

Studies included in our review were consistent in recommending further research as optimal management strategies have yet to be determined. Findings from previous studies of ICU-acquired NOAF may have been subject to bias by only adjusting for admission variables rather than physiological status at the time of treatment.

Consequently, I undertook a retrospective comparative cohort study investigating the use and effectiveness of different NOAF treatments for patients in the ICU[i]. This analysis aimed to use large granular datasets to account for physiological differences between patient groups at the time of AF onset. To identify treatments for further analysis in the cohort study, I organised an international oversight committee to review the treatments highlighted in our scoping review. Five treatments were identified: IV amiodarone, IV beta blockers, IV CCBs, IV digoxin, and electrical cardioversion (Table 5.1). Each intervention was analysed in an intention-to-treat fashion where treatment groups were determined by the first treatment after AF onset[290].

[i]Segments of the introduction, methods, results, and discussion sections of this chapter have previously been published[76–78]. Content reproduced with permission[32].

Table 5.1: Treatments to be included in the analysis

Treatment group	Treatment
Amiodarone	IV amiodarone
Beta blockers	IV Labetalol
	IV Esmolol
	IV Metoprolol
Calcium channel blockers	IV Diltiazem
	IV Verapamil
Digoxin	IV digoxin
Electrical cardioversion	Electrical cardioversion

Propafenone and ibutilide were identified as candidate therapies in our scoping review, however these were not available in our datasets. IV = intravenous.

5.2 Cohort study methods

5.2.1 Data sources

I performed a retrospective cohort study using data from two large intensive care databases, PICRAM and MIMIC-III v1.4 (see Sections 2.1.1 and 2.1.2). All analyses were performed on each database individually. A combined database analysis was performed for each outcome for interventions present in both databases. This work was reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines[291].

5.2.2 Study population

All adult (≥16 years) patients during their first admission to an ICU were included. Patients cared for by a coronary care or cardiac surgery team were excluded. Some patients ("outliers") on the Coronary Care Unit or the Cardiac Surgery Recovery Unit were included if they were under the care of a non-cardiac team.

I excluded patients with:

- missing hospital outcome data,
- an ICU LOS of less than 24 hours[292],
- documented pre-existing arrhythmias, or
- any recording of significant arrhythmia in the first three hours of arrival to ICU as a surrogate for undocumented pre-existing arrhythmia[74].

Characteristics of NOAF

AF and pre-existing arrhythmias were defined as per Chapter 4, Section 4.2.3. Characteristics of AF episodes were calculated, including the time from ICU admission to AF onset, the time of day of AF onset, and the duration of AF episodes.

The diurnal rhythm of AF onset was investigated using generalised additive models (GAMs) with cyclic cubic regression splines to model the distribution of events over 24 hours. This distribution was compared to a circular uniform distribution using the Watson test. The Watson test translates the principles of rank-based testing to circular data and may be considered the circular equivalent of the Mann-Whitney U test[293]. To account for any "fresh eyes" effect, bootstrap resampling was used to generate a population with a uniform distribution of heart rhythm change documentation times. Specifically, 2000 bootstrap samples were generated, inversely weighted by frequency distribution.

Primary outcomes

The primary outcomes of this study were rate control, rhythm control, and hospital mortality, analysed as time-to-event endpoints. Rate and rhythm control were censored at 24 hours, with mortality censored at 30 days. Time to rate control was defined as the time to a heart rate < 110 beats per minute in the subset of patients with a heart rate of ≥ 110 beats per minute after AF onset[294], and time to rhythm control as the time to first reversion of sinus rhythm[278].

5.2.3 Adjustment for confounding

I performed propensity score-weighted time-to-event analysis to adjust for measured confounding in patient selection between treatment groups. A list of 30 potential confounding variables was generated based on the studies identified in the scoping review outlined in Section 5.1.1 (Table 5.2). Illness severity was measured by the Oxford Acute Severity of Illness Score (OASIS)[273]. This is an admission illness severity score developed to predict ICU and hospital mortality. It consists of age, elective surgery, pre-ICU LOS and seven physiological measurements. The list of confounding variables was reviewed and ratified by members of the CAFE study oversight panel (Table D.10).

To account for confounding, I generated propensity score weights optimised to balance the covariate distributions of the treatment groups[295] using the WeightIt package[296] (see Section 2.2.2). The balance of covariates across weighted groups was assessed by tabulating group means pre- and post-weighting and calculating standardised mean differences.

Weighted Cox survival analysis was used to determine the average treatment effect of NOAF treatments on the outcomes of interest, with amiodarone as the reference group. To account for missing data, I used multiple imputation, generating 20 imputed datasets. To account for the uncertainty in calculated weights and to best estimate standard errors, bootstrapping was used with recalculation of propensity score weights and effect estimates with each bootstrap sample. Multiple imputation was combined with bootstrapping, with 1000 bootstrap samples being obtained from each of 20 imputed datasets[297]. The effect estimates and standard errors from each imputed dataset were then combined using Rubin's rules[298]. All statistical analyses were performed using R Core v4.0.2[299].

5.2.4 Sensitivity analysis

In addition to assigning weights across all treatment groups, I also repeated the analysis using weights generated in successive pairwise drug comparisons versus amiodarone. Unmeasured confounding was explored using array sensitivity

Table 5.2: Confounding variables accounted for in comparative treatment analysis.

Variable domain	Variable
Demographics	Age
	Sex
Comorbidities	Congestive cardiac failure
	COPD (previous or current diagnosis)
	Liver disease (previous or current diagnosis)
	Chronic renal failure
	Thyroid disorders*
	Pre-admission beta blockers
	Pre-admission antipsychotic medication
Admission factors	OASIS
Factors at the time of NOAF treatment	Heart rate[f]
	Blood pressure[f]
	Body temperature[f]
	Presence of vasopressor / inotrope[f]
	Dose of noradrenaline[f]
	Dose of vasopressin[f]
	White cell count[r]
	Serum electrolyte (K^+, Mg^{2+}, Na^+, Ca^{2+}) concentrations[r]
	Serum urea concentration[r]
	Serum creatinine concentration[r]
	Platelet count[r]
	Prothrombin time[r]
	Presence of therapeutic dose anticoagulation[d]
	Presence of bronchodilator therapy[d]
	Mechanical ventilation[t]
	Haemofiltration[p]
	Presence of central venous access[t]

**previous or current diagnosis, or taking relevant medications, [f] = First recording after AF onset, [r] = Most recent value in the ICU prior to AF onset, [d] = On the calendar day of, or the day preceding AF onset, [t] = At the time of NOAF, [p] = within the previous 12 hours. COPD = chronic obstructive pulmonary disease, OASIS = Oxford Acute Severity of Illness Score.*

analyses[300]. This method provides a range of corrected risk estimates based on a range of possible parameters of an unmeasured confounder, with the estimates being treatment effects expressed as HRs. Parameters include the prevalence of said confounding factor among the exposed and unexposed, and the association between the confounder and the outcome.

5.3 Results

5.3.1 Study population

Of the 18,559 patients fulfilling the inclusion criteria in the MIMIC-III database, 1,065 (5.7%) developed NOAF during their ICU admission. Of these, 740 received an intervention of interest (Figure 5.1a).

Of the 8367 patients that fulfilled the inclusion criteria in the PICRAM database, 952 (11.4%) developed NOAF during their ICU admission. Of these, 460 received an intervention of interest (Figure 5.1b). The characteristics of included patients are displayed in Table 5.3.

5.3.2 Characteristics of NOAF in treated patients

Time from ICU admission to NOAF in treated patients was similar between the MIMIC and PICRAM databases (median 40 hours; IQR 21 – 79 hours versus 41 hours; IQR 19 – 75 hours, respectively). Treated patients in the MIMIC-III database had, on average, shorter total durations of AF (median 12 hours; IQR 4 – 37 hours versus 18h; IQR 18 – 44 hours, respectively).

NOAF exhibited a bimodal distribution ($p<0.001$ for circular uniform distribution; Figure 5.2a), with two peaks at 8am and 8pm. The distribution of all heart rhythm documentation appeared uniform (Figure C.22), although documented rhythm changes were not; there appeared to be an increase in all documented rhythm changes around 8am (Figure 5.2b). However, the bimodal AF onset time distribution remained after accounting for the uneven distribution of rhythm change documentation (Figure 5.2c).

5.3.3 NOAF treatment groups

In patients identified in the MIMIC-III database, 94 received amiodarone, 473 received beta blockers (metoprolol, esmolol or labetalol), 144 received CCBs (diltiazem or verapamil), and 29 received electrical cardioversion as their initial NOAF treatment (Table D.11). In the PICRAM database, 344 patients received

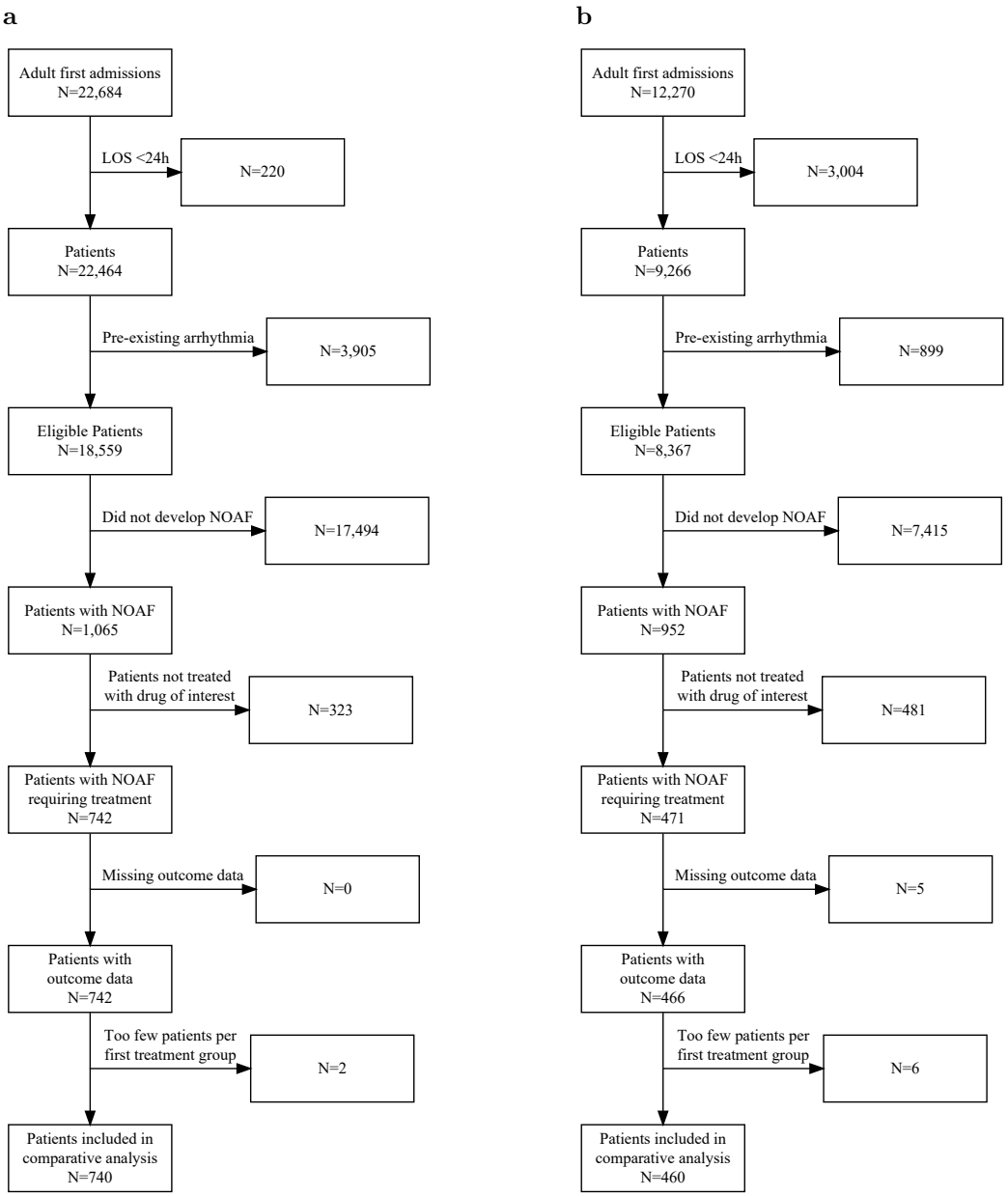


Figure 5.1: Study flow diagrams for a) the MIMIC-III database and b) the PICRAM database.

Table 5.3: Characteristics of included patients: Comparative treatment analysis.

Characteristic	Overall, N = 1,200	MIMIC-III, N = 740	PICRAM, N = 460
Age	72 (64, 80)	74 (64, 82)	70 (63, 77)
Sex			
F	558 (46%)	372 (50%)	186 (40%)
M	642 (54%)	368 (50%)	274 (60%)
COPD	116 (9.7%)	53 (7.2%)	63 (14%)
Dialysis-dependent renal failure	8 (0.7%)	1 (0.1%)	7 (1.5%)
NYHA class III/IV heart failure	2 (0.2%)	0 (0%)	2 (0.4%)
Chronic liver disease	31 (2.6%)	11 (1.5%)	20 (4.3%)
Thyroid disorder	61 (5.1%)	33 (4.5%)	28 (6.1%)
Beta blocker therapy prior to admission	344 (30%)	281 (42%)	63 (14%)
Antipsychotic medication prior to admission	34 (3.0%)	27 (4.0%)	7 (1.5%)
Highest OASIS at three hours	35 (29, 40)	36 (31, 41)	34 (26, 39)
Mechanical ventilation at time of NOAF	586 (49%)	343 (46%)	243 (53%)
Renal replacement therapy during or < 12 hours prior to NOAF	112 (9.3%)	47 (6.4%)	65 (14%)
IV Vasoactive medication at time of NOAF	225 (19%)	101 (14%)	124 (27%)
Therapeutic anticoagulation at time of NOAF	84 (7.0%)	36 (4.9%)	48 (10%)
Central venous catheter at time of NOAF	755 (63%)	429 (58%)	326 (71%)
Bronchodilator therapy on day of, or day preceding NOAF	333 (28%)	258 (35%)	75 (16%)
Serum sodium concentration (mmol/L)	139 (136, 142)	139 (136, 143)	137 (134, 141)
Serum potassium concentration (mmol/L)	4.0 (3.8, 4.4)	4.0 (3.7, 4.4)	4.2 (3.9, 4.5)
Serum magnesium concentration (mmol/L)	0.86 (0.78, 1.00)	0.82 (0.78, 0.95)	0.96 (0.84, 1.12)
Serum urea concentration (mmol/L)	11 (7, 18)	9 (6, 16)	14 (9, 20)
Serum creatinine concentration (μmol/L)	104 (69, 186)	97 (62, 159)	125 (78, 214)
White cell count (10^9/L)	12 (8, 16)	12 (8, 16)	11 (8, 16)
Haemoglobin concentration (g/L)	101 (90, 114)	102 (92, 115)	98 (88, 113)
Platelet count (10^9/L)	181 (117, 265)	190 (123, 283)	166 (109, 234)
Prothrombin time (s)	15.0 (13.6, 18.0)	14.2 (13.1, 16.4)	16.1 (15.0, 18.9)
Systolic blood pressure prior to AF onset	120 (104, 138)	124 (106, 142)	114 (101, 132)
Mean blood pressure prior to AF onset	77 (68, 89)	78 (68, 90)	75 (68, 88)
Heart rate prior to AF onset	95 (83, 109)	92 (81, 101)	102 (88, 118)
Treatment group (by first treatment)			
Amiodarone	438 (36%)	94 (13%)	344 (75%)
Beta blocker	520 (43%)	473 (64%)	47 (10%)
Calcium channel blocker	144 (12%)	144 (19%)	0 (0%)
Digoxin	69 (5.8%)	0 (0%)	69 (15%)
Electrical cardioversion	29 (2.4%)	29 (3.9%)	0 (0%)

Statistics presented: Median (IQR); n (%). Blood results represent most recent result prior to atrial fibrillation (AF) onset. COPD = chronic obstructive pulmonary disease, NYHA = New York Heart Association, OASIS = Oxford Acute Severity of Illness Score, IV = intravenous, NOAF = new-onset atrial fibrillation.

amiodarone, 47 received beta blockers, and 69 received digoxin as their initial NOAF treatment (Table D.12).

Physiological status immediately after NOAF differed across treatment groups. Patients treated with beta blockers tended to have higher blood pressure and lower vasoactive drug use than patients treated with amiodarone in both databases: In the MIMIC-III database the median systolic blood pressure (SBP) was 119 mmHg versus 103 mmHg, the median MAP 80 mmHg versus 72 mmHg, and the proportion of patients on vasoactive medications 9.5% versus 36%; in the PICRAM database the median SBP was 128 mmHg versus 112 mmHg, the median MAP 83 mmHg versus 74 mmHg, and the proportion of patients on vasoactive medications 13% versus 31%. Patients who received electrical cardioversion were the most cardiovascularly

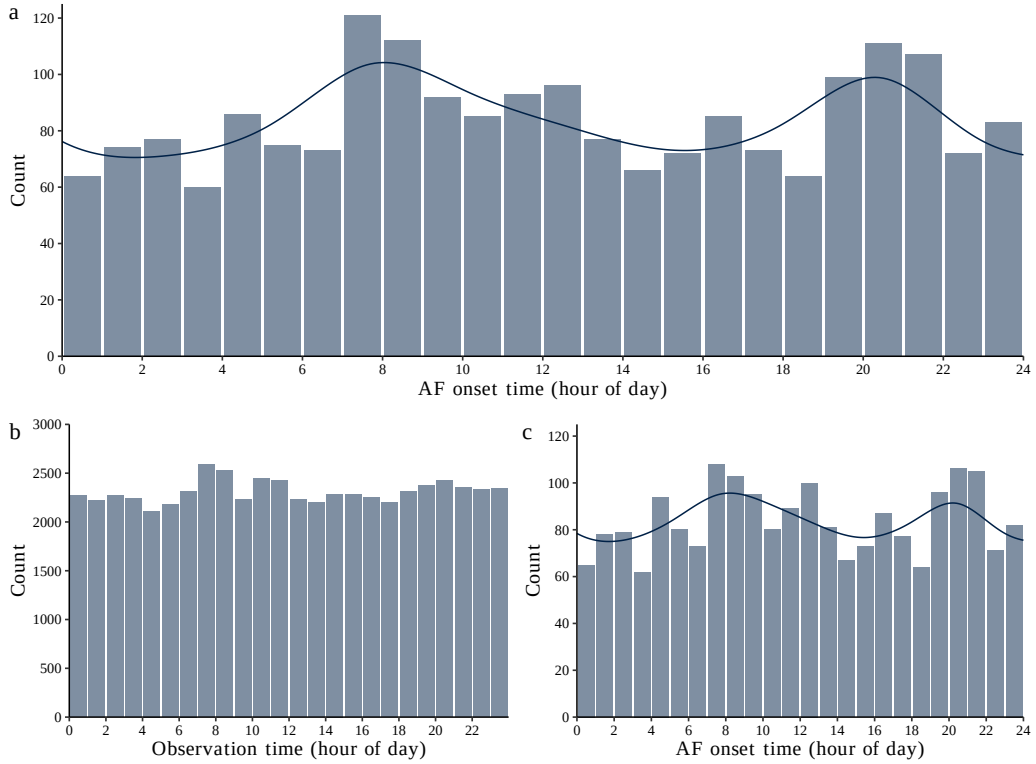


Figure 5.2: a) NOAF times with overlying GAM model; b) Time of day when heart rhythm changes were documented; and c) NOAF times in resampled data accounting for non-uniformity in heart rhythm change documentation.

unstable (median SBP 93 mmHg, median MAP 67 mmHg, proportion of patients on vasoactive medications 45%; Tables D.11 and D.12).

5.3.4 Adjustment for confounding

After propensity score weighting, covariates were well matched across all treatment groups in each database (Figures C.26 and C.27).

5.3.5 Outcomes

Rate control

Unadjusted time to rate control curves for the MIMIC-III and PICRAM databases are shown in Figures C.23a and C.23b, respectively. Rate control occurred immediately in around half of patients after electrical cardioversion. Rate control appeared to occur most slowly in the digoxin group. After adjustment, digoxin

therapy was associated with inferior rate control (aHR 0.56, 95% CI 0.34 – 0.92) in comparison to amiodarone (Table D.14 and Figure 5.3b).

No differences in achieving rate control were observed with beta blockers (aHR 1.09, 95% CI 0.78 – 1.51), CCBs (aHR 0.81, 95% CI 0.55 – 1.19) or electrical cardioversion (aHR 1.59, 95% CI 0.44 – 5.75) versus amiodarone in the MIMIC-III database (Table D.13 and Figure 5.3a), or beta blockers versus amiodarone in the PICRAM database (aHR 0.82, 95% CI 0.48 – 1.42; Table D.14 and Figure 5.3b).

In the combined database analysis of beta blockers versus amiodarone, rate control with beta blockers appeared superior in the unadjusted analysis. However, after adjustment, these differences were no longer evident (aHR 1.15, 95% CI 0.91 – 1.46), consistent with the individual database analyses (Table D.15).

Rhythm control

Unadjusted time to rhythm control curves for the MIMIC-III and PICRAM databases are shown in Figures C.24a and C.24b, respectively. Similar to rate control, rhythm control occurred quickly in around half of patients receiving electrical cardioversion. Rhythm control appeared to occur most slowly in the CCB and digoxin groups. After adjustment, CCBs were associated with inferior rhythm control (aHR 0.59, 95% CI 0.37 – 0.92) versus amiodarone in the MIMIC-III database (Figure 5.3a). No difference was identified in rhythm control with beta blockers (aHR 0.91, 95% CI 0.61 – 1.35) or electrical cardioversion (2.00, 95% CI 0.86 – 4.65) versus amiodarone in the MIMIC-III database.

In the PICRAM database, digoxin was not significantly associated with inferior rhythm control (aHR 0.67, 95% CI 0.41 – 1.09) versus amiodarone. As in the MIMIC-III database, beta blocker therapy was not associated with any significant difference in time to rhythm control (aHR 0.99, 95% CI 0.57 – 1.72; Figure 5.3b).

In the combined database analysis of beta blockers versus amiodarone, rhythm control with amiodarone appeared superior in the unadjusted analysis. However, after adjustment, the difference was no longer evident (aHR 0.85, 95% CI 0.69 – 1.05), consistent with our individual database analyses (Table D.15).

Hospital mortality

Unadjusted survival curves for the MIMIC-III and PICRAM databases are shown in Figures C.25a and C.25b, respectively. Patients who received electrical cardioversion had the lowest survival rate at 30 days. Patients who received beta blockers appeared to have the highest survival rate in both databases. After adjustment, there were no significant differences between beta blockers (aHR 1.03, 95% CI 0.53 – 2.03), CCBs (aHR 1.30, 95% CI 0.61 – 2.76), or electrical cardioversion (aHR 0.96, 95% CI 0.31 – 3.01) in hospital mortality versus amiodarone in the MIMIC-III database (Figure 5.3a). Similarly, no differences were identified between beta blockers (aHR 0.75, 95% CI 0.30 – 1.86) or digoxin (aHR 1.40, 95% CI 0.77 – 2.56) versus amiodarone in hospital mortality in the PICRAM database (Figure 5.3b).

Beta blocker therapy appeared to be associated with reduced mortality in the combined dataset prior to adjustment. However, after adjustment, the difference was no longer evident (aHR 1.03, 95% CI 0.53 – 2.03), consistent with the individual database analyses (Table D.15).

Sensitivity analysis

The findings remained evident in the sensitivity analysis (CCB rhythm control aHR 0.55, 95% CI 0.37 – 0.82, digoxin rate control aHR 0.56, 95% CI 0.36 – 0.88). Furthermore, CCB therapy was associated with an increased risk of reversion to a heart rate $>$ 110 beats per minute (aHR 1.82, 95% CI 1.10 – 3.00; Tables D.16 and D.17).

Array sensitivity analysis suggested that an unmeasured confounder would require high imbalance between treatment groups and extreme association with outcome to bring the observed effect estimate to the null for digoxin's inferior rate control (Table D.18) and CCB's inferior rhythm control (Table D.19).

5.4 Discussion

In this detailed analysis of patients admitted to an ICU, I investigated the characteristics of NOAF and the outcomes associated with first-line treatments. NOAF

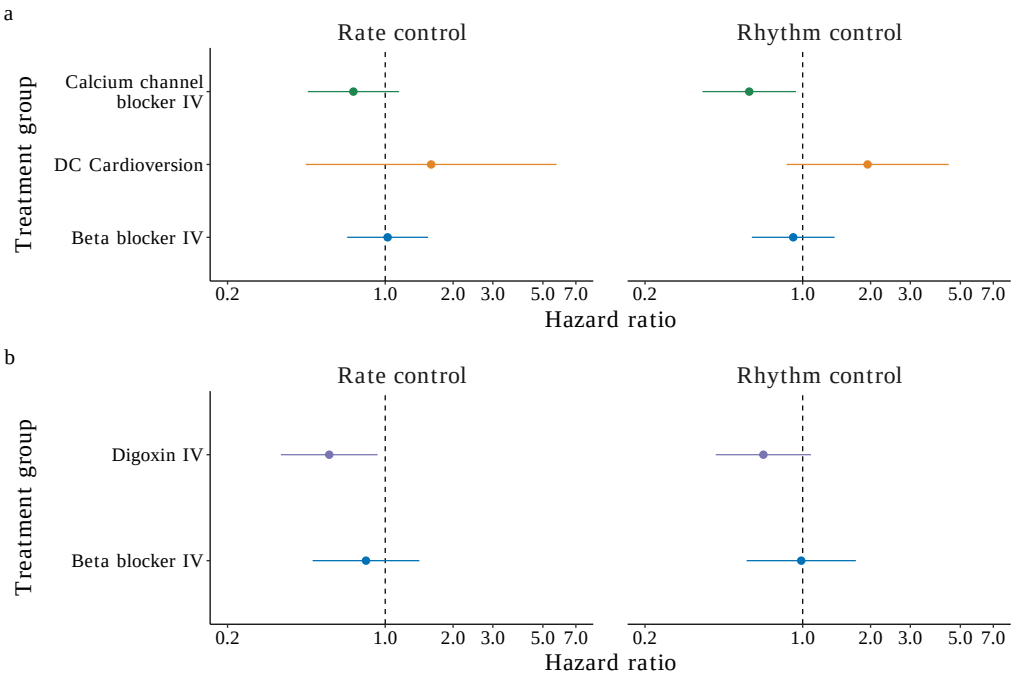


Figure 5.3: Adjusted outcomes associated with NOAF treatment in a) MIMIC-III database and b) PICRAM database.

mostly occurred early in the ICU stay and appeared to exhibit diurnal variation in onset time. Digoxin therapy was associated with inferior rate control compared with amiodarone. Calcium channel blocker therapy was associated with inferior rhythm control compared with amiodarone.

Patients who received beta blockers were less unwell at admission and more stable after AF onset than those treated with amiodarone. Apparent differences in outcomes between beta blockers and amiodarone (reduced mortality and improved rate control but inferior rhythm control) in unadjusted analyses were not found after adjustment for confounding variables including cardiovascular stability at the time of NOAF. These findings have now been cited in American National guidelines[301].

This work highlights the need for large randomised trials of NOAF management in the ICU setting. Further, it suggests amiodarone and beta blockers may be suitable candidate drugs for such a trial. These insights prompted an NIHR call for a randomised trial of amiodarone versus beta blockers.

5.4.1 Characteristics of NOAF in treated patients

NOAF tended to occur early in the ICU stay. This may be partly explained by there being fewer patients to develop NOAF as time passes, as patients are discharged or do not survive. NOAF in patients admitted to an ICU also appeared to have circadian variation, with peaks in the morning and evening. This pattern is consistent across UK and USA populations, and remained after accounting for nursing documentation times. Understanding variation in AF onset time and identifying high-risk periods and susceptible subgroups may inform the timing of future preventative interventions.

Several mechanisms may explain the circadian variation in NOAF during critical illness. Serum catecholamine concentrations increase in the morning after a nocturnal nadir[302]. Subsequent adrenergic stimulation can promote AF through both enhancing automaticity and increasing risk of early and delayed afterdepolarisations (see Section 3.4.2). Beta blocker administration may dampen these changes[303]. Circadian variation in myocyte gene transcription also alters cardiac electrical excitability and metabolism[303]. While these mechanisms may in part account for the morning peak in NOAF, the evening peak is less easily explained. However, the distribution seen in our data is consistent with that reported in patients with paroxysmal AF in the community[304]. In patients admitted to an ICU these peaks may reflect bimodal variation in autonomic regulation[305].

While NOAF was more likely to occur in the morning and evening, the peaks were only around 25% above the baseline rate, suggesting any circadian mechanism had a relatively small effect on the time of onset. The circadian variation in other vital signs in the ICU setting and the potential underlying mechanisms lend credibility to these findings[306]. However, as the peaks occur at times where nursing shift changes occur, the potential impact of a fresh eyes effect should be considered.

5.4.2 Treatment efficacy

To the best of my knowledge, this is the first work to compare NOAF treatment effectiveness in patients on an ICU after balancing baseline and post-AF variables

that may affect treatment decisions. Preference for first-line treatment differed between the databases, suggesting different treatment paradigms between the UK and the USA.

Findings of poorer rhythm control associated with CCB therapy compared with amiodarone are supported by a small randomised study of paroxysmal AF outside ICU which reported a cardioversion proportion of 0% in patients who received verapamil versus 77% in patients who received amiodarone[307]. These findings contrast with one RCT from a medical cardiac ICU which compared amiodarone with CCBs[65] and found no difference in achieving rhythm control. However, this study included only 20 patients per treatment group, making drawing any conclusions difficult.

Digoxin may be selected to reduce heart rate without inducing hypotension. However, it may be less effective during states of increased sympathetic drive and catecholaminergic medication use[308,309]. Its slower onset of action is also recognised[310]. These characteristics may explain the inferior performance in comparison to amiodarone in patients treated on an ICU.

Rate and rhythm control were similar with beta blockers and amiodarone. These findings are consistent with a previous retrospective study of patients admitted to an ICU with new or pre-existing AF that found no difference in time to rate control between metoprolol and amiodarone[61].

In adjusted analyses, there was no significant difference in hospital survival between any of the treatments when compared to amiodarone. In the MIMIC-III database, the unadjusted analysis demonstrated an apparent mortality benefit in the beta blocker treatment group. However, patients in the beta blocker group were younger and less unwell at presentation than the amiodarone group. After developing NOAF, they had higher blood pressures, were less likely to be on vasoactive medications and had lower inflammatory markers. After adjustment, the difference in mortality was no longer evident. These adjusted findings conflict with one large study which suggested a survival benefit with beta blockers over amiodarone therapy in patients with sepsis[69]. However, this study was unable to adjust for

features around the time of NOAF. This work demonstrates that important clinical differences exist between treatment groups after the onset of AF, prior to treatment. Failure to adjust for these factors is likely to result in residual confounding.

5.4.3 Strengths and limitations

This work has several strengths. I performed comprehensive adjustment which included physiological variables after AF onset. There were significant differences in these post-AF variables which have not been adjusted for in previous studies. Furthermore, analysis of routinely collected data over many years provided a sample size large enough to demonstrate differential efficacy in NOAF treatments.

This analysis also has limitations. The study was retrospective and not randomised. While good balance of numerous confounding covariates was achieved prior to assessing treatment efficacy, bias introduced by residual unmeasured confounding may still have occurred. Treatment choices may be more or less appropriate depending on clinical circumstances, and the findings should be taken in this context. Treatment groups were determined by the first treatment after AF onset. Patients may have gone on to receive other therapies, and the effect of any cross-over was not analysed. Future work should assess rates of cross-over by initial treatment. Larger studies would be required to assess the comparative efficacy of different treatment combinations. Although NOAF was not independently verified, documentation of AF in the MIMIC-III database has been shown to be accurate for determining AF onset to within one hour of true onset in independent review of ECG waveforms[311]. Documentation of comorbidities in the MIMIC-III database relied on hospital billing codes which may not have identified all cases.

While this study focused on treated patients, a significant proportion of patients developing NOAF did not receive treatment with one of the study interventions. This group may represent patients who developed only fleeting AF, AF that responded to non-pharmacological interventions e.g. IV fluids or electrolyte supplementation, or patients at the end of their lives. It is also possible that some instances of chart-documented AF in this cohort were erroneous. Future work could consider

analysing the comparative outcomes of patients with NOAF who do not receive rate or rhythm control therapies.

5.4.4 Amiodarone or beta blockers

Given that this work did not identify a difference in treatment effectiveness between amiodarone and beta blocker therapy, it is worth considering the available evidence of their risks and benefits when used in an ICU setting.

Amiodarone

Amiodarone use is associated with many adverse effects including corneal microdeposits, thyroid disease, pulmonary toxicity, and hepatotoxicity[312]. Neuropsychiatric adverse effects can also occur, most commonly tremor and ataxia[313].

Amiodarone and its metabolite monodesethylamiodarone have a long half-life and a high tissue affinity for the lung[314,315]. Indeed, in treated patients, the amiodarone concentration in lung parenchyma exceeds that of the heart[316]. Amiodarone pulmonary toxicity (APT) is likely the result of direct toxic injury and an indirect immunologic injury[317]. Amiodarone treatment causes phospholipid accumulation in pulmonary cells within hours of exposure to therapeutic serum concentrations[318,319]. This has a direct cytotoxic effect on the alveolar-capillary membrane. Amiodarone exposure also results in mitochondrial dysfunction and ROS generation[320]. Affected lung tissue demonstrates infiltration of pneumocytes and the alveolar space by lipid-laden macrophages, septal thickening, and fibrosis. There are also features of an immunologic process including lymphocytic infiltration with features of organising pneumonia[321,322].

APT can occur at any time after treatment is initiated, but those considered at greatest risk are individuals who have received a daily dose of 400 mg or more for more than two months[323]. The risk of APT is roughly 1% per patient-year[324] and increases with higher serum concentrations[325]. However, APT correlates better with the total cumulative dose rather than with the daily dose or serum concentration of amiodarone[326].

Individuals with lung injury appear to be more susceptible to acute APT, which occurs within days of administration[323]. Exposure to high oxygen partial pressures may also speed and exacerbate amiodarone's oxidative insult[327,328]. Case reports of APT occurring less than a week after drug initiation accordingly relate to critically ill patients or patients undergoing thoracic surgery. Many of these cases are supported by histological findings[329–333]. Furthermore, post-mortem studies of patients who received amiodarone for supraventricular tachycardia associated with critical illness have revealed histological changes of widespread lipoid pneumonia, characteristic of APT[334]. One randomised trial of amiodarone for the prevention of post-operative AF was stopped due to higher rates of acute respiratory distress syndrome in patients who received amiodarone[335]. The Society of Thoracic Surgeons guidelines do not recommend its use for patients undergoing pneumonectomy due to potential toxicity risk[336].

While suggestive, these studies do not constitute clear evidence of harm to patients receiving amiodarone during critical illness. A number of small randomised trials of amiodarone for perioperative AF prevention[337,338] and treatment[339,340] have not demonstrated an increase in pulmonary complications in the patients receiving amiodarone. However, such trials often focus on elective cases and exclude those critically ill patients most at risk of APT[341]. In critically ill patients, a combination of systemic inflammation, lung injury due to infection or mechanical ventilation, and high concentrations of oxygen may increase the risk of acute APT[323,342]. While concomitant lung pathology may encourage acute APT, it may also mask it, leading to reduced recognition. Where acute APT is not overt, cellular lung damage may still be present. In patients with limited reserve this effect may impact on outcomes. While the risk of acute APT during critical illness is likely to be small, it appears unlikely to be zero.

Lastly, amiodarone in its IV formulation is acidic and has been associated with phlebitis and extravasation leading to skin necrosis[343,344]. It is therefore recommended to be administered via a central line. In patients without central venous access, this would necessitate an invasive procedure. In patients with existing

central venous access it may delay catheter removal. While amiodarone may be administered enterally, drug absorption via this route is slow and erratic, with highly variable bioavailability even in healthy subjects[345,346]. For these reasons, the intravenous route is preferred for patients in the intensive care unit[71].

Beta blockers

There is no evidence that amiodarone is superior to beta blockers in the management of AF associated with critical illness[347]. Indeed There is some suggestion that beta blockers may be beneficial[69], though as explored above, this may be due to residual confounding. Purported benefits may relate to a better adverse effect profile alongside beneficial changes to haemodynamics, structural and electrical remodeling, and autonomic regulation.

Atrial anti-arrhythmic effects of beta blockers involve both a suppression of pro-arrhythmic effects of catecholamines, and electrophysiological remodeling. Beta blocker-induced remodeling increases the APD and effective refractory period. Adaptation of the APD occurs faster in the atria than in the ventricles, reaching a maximum after roughly one week of treatment[348]. These effects continue beyond cessation of beta blocker treatment and reduce the propensity for initiation and maintenance of re-entrant circuits[349]. After the onset of AF, detrimental electrophysiological remodeling occurs that promotes AF. These changes contrast with those associated with beta blocker treatment and include a reduction in the atrial effective refractory period[350]. Beta blocker treatment may therefore prevent or reduce the detrimental remodeling and reduce the propensity for arrhythmia maintenance.

Acute beta blockade has little effect on atrial electrophysiology in the absence of sympathetic stimulation[197]. However, beta blockers can prevent AF in cases of elevated adrenergic tone such as heart failure[351] or after surgery[352]. This effect suggests anti-adrenergic suppression of the pro-arrhythmic effects of catecholamines plays an important role.

After myocardial infarction, beta blocker treatment improves left ventricular diastolic function[353] and reduces detrimental structural remodeling[354]. In patients

with heart failure, IV beta blocker treatment was associated with an increase in ejection fraction with no adverse effects[355]. Myocardial injury is very common during critical illness[356,357] and the effects of beta blockade in this scenario may confer similar benefits, though this requires more study.

Beta blocker treatment may improve outcomes for patients with sepsis, reducing vasopressor requirements and improving mortality[358]. These effects may involve reduced inflammation[359], reduced myocardial oxygen demand[354] and improved microcirculatory blood flow[358]. These benefits in patients with sepsis appear to occur without significant impact on MAP[360].

The need for comparative research

European surveys suggest that amiodarone remains the most common first-line pharmacotherapy for ICU-acquired NOAF, with beta blockers second[70,71]. This aligns with the UK-based findings in this chapter (though beta blockers were more commonly used in the USA)[69,77]. Outside the ICU, amiodarone is not recommended for chronic AF management by the National Institute for Health and Care Excellence (NICE)[361] and European AF guidance considers amiodarone a "last resort" for rate control, and second line after other treatment options have failed for rhythm control due to the severe side effect profile[236].

Although amiodarone's side-effects are mainly associated with long-term use, acute side-effects occur mostly in critically ill patients[362]. There is a need for central venous access to administer amiodarone to avoid skin toxicity. There is also a known incidence of ICU-administered drugs, including amiodarone, being continued after hospital discharge[363]. Together, this suggests that amiodarone use may constitute an unnecessary risk if beta blockers, which can be administered without central venous access, were found to be equivalently efficacious.

Short-term outcomes after NOAF

6.1 Background

In the previous chapter, the comparative effectiveness of NOAF treatments was explored. This chapter investigates how NOAF itself impacts on patient outcomes, both the immediate haemodynamic impact, and its association with hospital mortality.

6.1.1 The haemodynamic changes associated with AF

Cardiac performance is impaired by AF through several mechanisms. Ventricular diastole is divided into four parts: (i) Isovolumetric relaxation, (ii) the rapid filling phase, (iii) diastasis and (iv) atrial contraction or "kick". During the final phase, the atria contract, which contributes around 20% of the ventricular end-diastolic volume in healthy states, and around 80% flows passively into the ventricles during the rapid filling phase[364]. The contribution of the atrial kick is dependent on the heart rate and cardiac structure. With faster heart rates, ventricular diastolic time reduces, and there is consequently an increasing dependence on the atrial kick for ventricular filling[365]. In patients with impaired ventricular relaxation (diastolic dysfunction), the rapid filling phase and early diastasis are reduced, and these patients may become dependent on the atrial kick for adequate ventricular filling[366]. The loss of organised atrial contraction in AF can therefore lead to heart failure in settings of tachycardia or structural heart disease, particularly in combination.

Even after successful cardioversion from AF, patients may have reduced cardiac output and remain at risk of heart failure—atrial contractile dysfunction persists after cardioversion from AF episodes lasting only one hour[367], and can last for several weeks after cardioversion[368].

An irregular rhythm itself may reduce cardiac output, independent of the loss of atrial kick; an irregular rhythm induced through ventricular pacing reduces cardiac output by around 12% compared to regular ventricular pacing at the same rate. This reduction occurs within two minutes of initiating the irregular rhythm[369]. This reduction may be due to beat-to-beat changes in ventricular filling. Due to the Frank-Starling mechanism, a shorter RR interval (the time elapsed between two successive R-waves on an ECG) will be followed by a lower ejection fraction. This may not be entirely compensated for by the increase in ejection fraction that accompanies longer RR intervals. There may also be beat-to-beat changes in contractility, independent of end-diastolic volume[370].

An irregular rhythm may further reduce cardiac output by inducing changes in neurohumoral factors. Atrial natriuretic peptide induces arterial and venous dilation and inhibition of cardiac sympathetic input[371]. While ventricular pacing increases atrial natriuretic peptide levels, irregular ventricular pacing may be associated with higher concentrations of atrial natriuretic peptide compared to regular ventricular pacing due to greater variation and higher peaks in atrial pressure and wall tension[372,373].

Together, these mechanisms may explain why NOAF is temporally associated with a reduction in cardiac output in non-ICU patients[73]. However, the haemodynamic impact of NOAF in critically ill patients remains poorly understood. Limited data suggests NOAF may precede haemodynamic instability in this context[41], but this requires further study.

6.1.2 Hospital mortality associated with NOAF

These haemodynamic sequelae may partly explain the observed association between ICU-acquired NOAF and increased early mortality[374,375]. However, patients with a greater severity of illness—and consequently a poorer prognosis—are more likely to develop NOAF[51,376]. NOAF may therefore be a marker of disease severity without an independent influence on mortality[50]. The debate over the mortality attributable to NOAF is ongoing, as some studies have not demonstrated an independent

association after correcting for disease severity[377,378]. There is however a growing body of evidence demonstrating an association with ICU[56] and hospital[129,139] mortality after correcting for confounding variables.

NOAF can occur at any point during an ICU admission; analyses that do not take into account this time-varying nature of NOAF can underestimate its association with hospital outcomes. A patient must survive in the ICU for long enough to develop NOAF. Conversely patients who die or are discharged earlier have less opportunity to develop NOAF[379]. This is known as immortal time bias, as a patient in a NOAF cohort was "immortal" prior to developing NOAF[380]. Existing studies have used logistic regression analysis, considering NOAF as a dichotomous rather than time-dependent variable, making them susceptible to immortal time bias. This results in an under-estimation of the mortality risk associated with NOAF[381].

Traditional survival analysis assumes non-informative censoring i.e. the hazard of death remaining unchanged when a censoring event occurs[382]. However, discharged patients have a different risk of death than those that remain in hospital[383,384]. Hospital discharge should therefore not be treated as a censoring event, but rather a competing risk to mortality. Competing risks are events during follow-up that either prevent the primary outcome from occurring, or alter its probability[385]. In the presence of competing risks, appropriate methodology should be used[386] (see Section 2.2.3).

Furthermore, existing studies have not investigated how the duration of NOAF while in an ICU is associated with hospital outcomes.

In this chapter, I aimed to analyse the haemodynamic changes associated with NOAF, and assess the mortality risk associated with NOAF, accounting for the complexities above[i].

[i]Segments of the introduction, methods, results, and discussion sections of this chapter have previously been published[77,79]. Content reproduced with permission[32].

6.2 Methods

6.2.1 Study population

For the analysis of haemodynamic changes associated with NOAF, I used data from the study cohort of treated patients from Chapter 5 for consistency (see Section 5.3.1).

For the analysis of hospital outcomes associated with NOAF, I used the eligibility criteria outlined in section 5.2.2 to identify a cohort of patients who developed NOAF, and a cohort of patients that did not. Patients with an ICU stay of less than 24 hours were also excluded in line with previous similar studies[292]. Only the PICRAM database was used for this analysis. NOAF was defined as per Chapter 4, Section 4.2.3.

6.2.2 Primary outcomes

The primary outcome of this analysis was hospital mortality, accounting for the competing risk of hospital discharge alive. Secondary (haemodynamic) outcomes were changes in heart rate, SBP, and use of vasoactive medications associated with NOAF.

6.2.3 Haemodynamic analysis

To explore the pre-treatment changes in heart rate and SBP during the six hours before and after AF onset, smooth additive quantile regression models[387] were used to fit the 25%, 50%, and 75% quantiles of the distribution of each haemodynamic variable. Quantile regression models how predictors affect specified percentiles of a response variable. This provides a more comprehensive understanding of changes in haemodynamic variables behavior over their entire distribution in relation to time and AF onset. Models allowed for changes in smoothing spline post-AF onset.

I also calculated the proportion of patients receiving vasoactive medications before and after NOAF onset. The vasoactive-inotropic score (VIS) was used to quantify the change in composite dose of vasoactive medications for patients already receiving vasoactive medications prior to AF onset[388].

The VIS is a weighted sum of vasoactive medication doses, calculated as:

$$VIS = D_p + D_b + 100 \cdot A + 10 \cdot M + 10000 \cdot V + 100 \cdot N$$

where:

VIS	is the vasoactive-inotropic score
D_p	is the dopamine dose (μg/kg/min)
D_b	is the dobutamine dose (μg/kg/min)
A	is the adrenaline dose (μg/kg/min)
M	is the milrinone dose (μg/kg/min)
V	is the vasopressin dose (units/kg/min)
N	is the noradrenaline dose (μg/kg/min)

I calculated the number of patients who transitioned from a normal blood pressure (SBP $\geq$ 90 mmHg or MAP $\geq$ 65 mmHg) to a low blood pressure, and vice versa, and those who maintained their current blood pressure category after NOAF. Multilevel linear models were used to test the significance of changes in haemodynamic variables associated with NOAF, allowing for varying intercepts and slopes per patient. Each model included fixed linear segmented regression terms, with a random effect per patient to account for repeated measurements. This method allowed for quantification of haemodynamic changes associated with NOAF, accounting for patients' pre-AF parameters.

6.2.4 Hospital outcomes

Adjustment for confounding

To analyse the association of NOAF with hospital outcomes, I adjusted for 15 admission comorbidities, which, although relatively rare, have been shown to be

independently associated with mortality in patients admitted to an ICU[389]:

1. Liver cirrhosis
2. Portal hypertension
3. Hepatic encephalopathy
4. Severe cardiovascular disease (New York Heart Association Class IV)
5. Severe respiratory disease (permanent shortness of breath with light activity due to pulmonary disease)
6. Home ventilation
7. Chronic renal replacement therapy
8. HIV/AIDS
9. Steroid therapy
10. Congenital immunodeficiency
11. Radiotherapy
12. Chemotherapy
13. Metastatic disease
14. Acute or chronic leukaemia
15. Lymphoma

The analysis was further adjusted for patient sex and the OASIS[273] (see Section 5.2.3). As the OASIS includes age and physiological measurements, these were not included as individual covariates in the analysis. Covariates were included when adjusting, regardless of association with outcome to avoid bias introduced by variable selection methods[390].

In Chapter 5, I adjusted for physiological status prior to AF treatment rather than status at admission. This was done to account for factors influencing AF treatment choice. In this chapter, outcomes were adjusted for admission factors, given their validated association with hospital mortality[273,389].

NOAF as a time-varying covariate

I analysed NOAF as a time-varying covariate to avoid immortal time bias (see above).

Hospital discharge as a competing risk to mortality

Firstly, I generated CSHRs to estimate the aetiological association of NOAF with ICU mortality and on ICU discharge alive. The CSHR can be considered a measure

of aetiological association i.e. the degree with which the occurrence of NOAF is directly associated with hospital mortality rate or rate of hospital discharge[391]. The proportional hazards assumption was checked visually using a graph of the scaled Schoenfeld residuals, along with a smoothing spline (see Section 2.2.3).

Next, I calculated the SDHR to estimate the prognostic association between NOAF and hospital outcomes (see Section 2.2.3).

Sensitivity analysis

Further to the primary model, which was adjusted for sex, comorbidity, and the OASIS, three additional models were built to test the robustness of the main findings. Model 2 included the OASIS only. Model 3 included all features of the primary model plus admission laboratory values (potassium, magnesium, urea, creatinine and lactate concentrations, and white cell count) selected for their association with patient outcomes during critical illness[52,139,392–395]. Model 4 included all features of model 3 plus the use of vasoactive medications.

The primary analysis was also repeated using an alternative NOAF definition potentially relevant to critically ill patients, namely AF lasting for ≥ 1 hour or requiring pharmacological therapy or requiring cardioversion[46,52].

6.3 Results

6.3.1 Study population

The haemodynamic analysis comprised the 740 treated patients from the MIMIC-III database (Figure 5.1a), and the 460 treated patients from the PICRAM database (Figure 5.1b) resulting in a study cohort of 1200 patients. The characteristics of included patients are displayed in Table 5.3.

The hospital outcomes analysis comprised 831 patients who developed NOAF during their admission, and 6,710 patients who did not develop NOAF for comparison. The characteristics of these patients are displayed in table 6.1. Patients who developed NOAF were older, more likely to be male, and were more unwell than patients who did not develop NOAF. Hospital outcome data were not available for

Table 6.1: Characteristics of included patients in the PICRAM-only hospital outcomes analysis.

Variable	Never AF	New-onset AF
n	6710	831
Age (years)	61 [46, 71]	71 [63, 78]
Male sex	3896 (58.1)	508 (61.1)
Severe comorbidity present	995 (14.8)	145 (17.4)
Surgical admission	3043 (45.4)	290 (34.9)
Elective surgery	1615 (24.1)	106 (12.8)
Admission diagnosis		
Medical: cardiovascular / shock	1094 (16.3)	215 (25.9)
Surgical: gastrointestinal	1677 (25.0)	189 (22.7)
Medical: respiratory	1156 (17.2)	173 (20.8)
Surgical: vascular	431 (6.4)	53 (6.4)
Medical: neurological	402 (6.0)	25 (3.0)
Medical: Endocrine, metabolic, poisoning	305 (4.5)	16 (1.9)
Surgical: genitourinary	292 (4.4)	10 (1.2)
Surgical: respiratory	287 (4.3)	21 (2.5)
Medical: gastrointestinal	229 (3.4)	41 (4.9)
Surgical: neurological	198 (3.0)	8 (1.0)
Non-surgical: trauma	198 (3.0)	13 (1.6)
Medical: genitourinary	176 (2.6)	42 (5.1)
Surgical: trauma	116 (1.7)	7 (0.8)
Other	149 (2)	18 (2)
OASIS	33 [24, 41]	42 [33, 49]
Vasopressor use	3086 (46.0)	647 (77.9)
Ventilated in first 24 h	3689 (55.0)	566 (68.1)
Maximum heart rate in first 24 h (BPM)	104 [90, 119]	114 [97, 133]
Minimum MAP in first 24 h (mmHg)	60 [53, 67]	56 [49, 62]
Maximum temperature in first 24 h (°C)	37.1 [36.7, 37.7]	37.2 [36.8, 37.8]
Urine output in first 24 h (ml)	1505 [990, 2300]	1057 [594, 1630]
Plasma potassium concentration at admission (mmol/L)	4.1 [3.7, 4.5]	4.2 [3.7, 4.8]
Plasma magnesium concentration at admission (mmol/L)	0.78 [0.67, 0.91]	0.82 [0.70, 0.94]
Plasma urea concentration at admission (mmol/L)	6.5 [4.6, 10.5]	10.6 [7.0, 17.6]
Plasma creatinine concentration at admission (μmol/L)	82 [60, 130]	113 [76, 196]
WCC at admission ($\times 10^9$/L)	11.7 [8.3, 16.1]	11.6 [7.1, 16.6]
Plasma lactate concentration at admission (mmol/L)	1.5 [1.0, 2.5]	1.8 [1.1, 3.2]

Statistics presented: n (%); Median [25%,75%]. OASIS = Oxford Severity of Illness Score; MAP = mean arterial pressure; WCC = white cell count; BPM = beats per minute.

531 (4.3%) of admissions. Missing hospital outcome data resulted from incomplete linkage of two databases. Linkage was not related to patient characteristics. A comparison of available data from patients with and without missing outcome data did not reveal any systematic differences (Table D.20).

6.3.2 Haemodynamic changes

NOAF was associated with a significant heart rate increase of 21 beats per minute (multi-level model $p < 0.001$; figure 6.1a). It was also associated with a significant reduction in SBP of 5 mmHg ($p < 0.001$; figure 6.1b). This decrease in SBP was despite significant increases in the doses of vasoactive medication after NOAF in those receiving vasoactive medications prior to NOAF (VIS increase of 2.8 ($p = 0.001$; Figure 6.1c).

New hypotension (SBP < 90 or MAP < 65) occurred after NOAF in 209 of the 947 patients with SBP ≥ 90 and MAP ≥ 65 prior to AF onset (22%; Figure 6.2). There was no significant change in the proportion of patients receiving vasoactive medications after NOAF onset (22.0% versus 24.6%, $p = 0.19$).

6.3.3 Hospital outcomes

Overall hospital mortality in the PICRAM hospital outcomes cohort was 14%. This was higher in the NOAF cohort than those without NOAF (34.8% versus 12.2%). The NOAF cohort had an increased hospital rate of death per unit time (unadjusted CSHR 2.58, 95% CI 2.36 – 2.83; adjusted CSHR 1.87, 95% CI 1.70 – 2.05). Patients with NOAF also had a reduced rate of hospital discharge alive per unit time (unadjusted CSHR 0.51, 95% CI 0.48 – 0.53; adjusted CSHR 0.55, 95% CI 0.52 – 0.58). The prolonged hospital stays in patients with NOAF combined with their increased rate of hospital death resulted in a strong prognostic association with dying in hospital despite adjusting for confounding variables (adjusted SDHR 2.26, 95% CI 2.07 – 2.47).

A stacked cumulative incidence curve demonstrating the association between NOAF and both hospital outcomes is shown in figure 6.3. This shows a simultaneously higher rate of hospital mortality and lower rate of hospital discharge alive in the NOAF cohort. Analysis of Schoenfield residuals supported the proportional hazards assumption (proportional hazards test $p = 0.254$ suggesting independence between time and NOAF hazard ratio; Figure C.28).

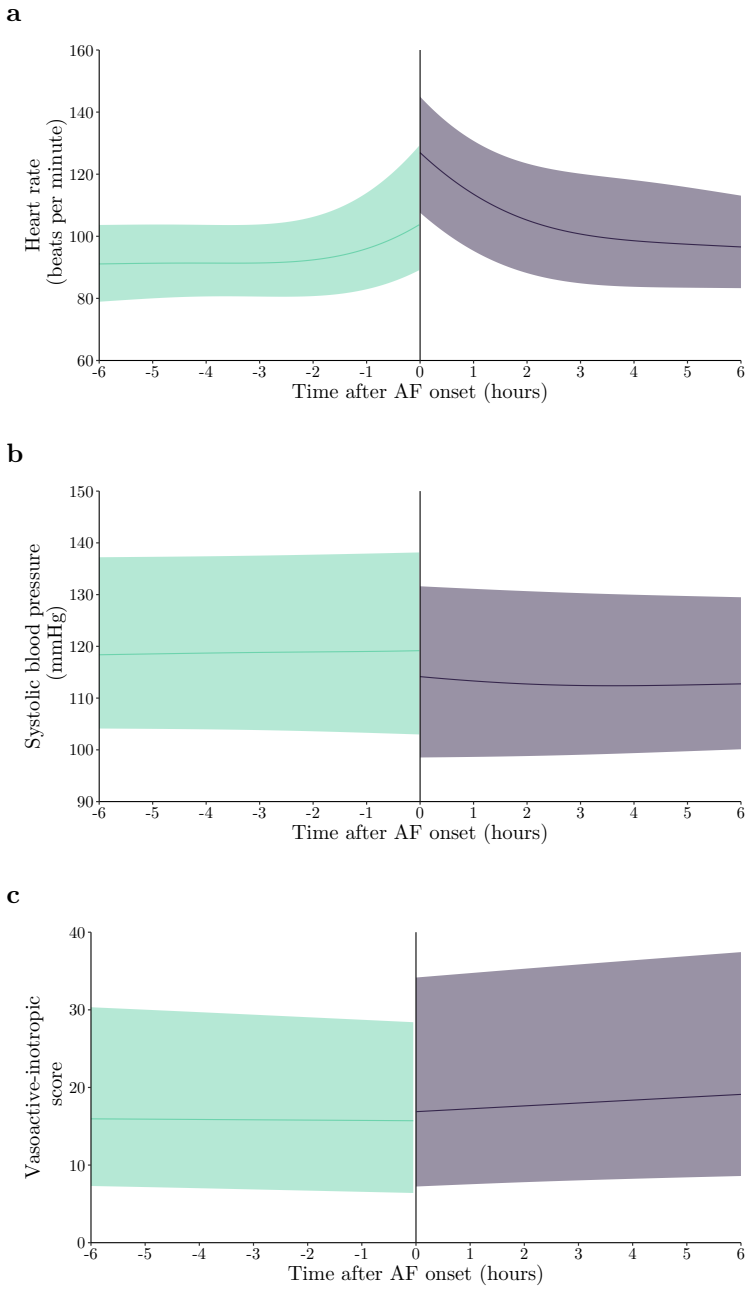


Figure 6.1: Changes in a) heart rate; b) systolic blood pressure; and c) vasoactive-inotropic score associated with atrial fibrillation (AF) onset in the combined MIMIC-III and PICRAM dataset. Vasoactive-inotropic score shown for those patients receiving vasoactive medications prior to AF onset. The solid lines represent the median of the haemodynamic parameter over time. The shaded area encompasses the range between the 25% and 75% quantiles, modelled using quantile regression.

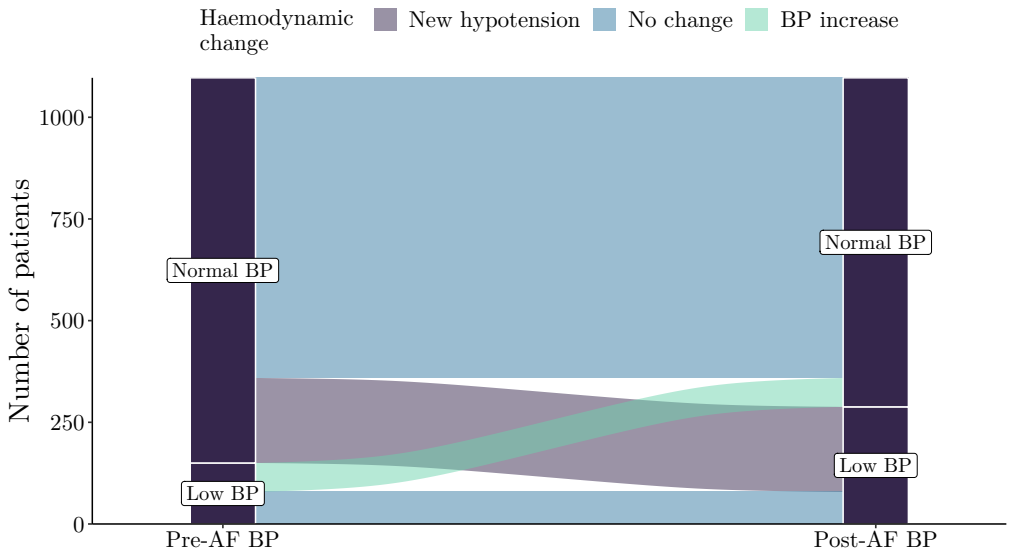


Figure 6.2: Alluvial plot of changes in blood pressure status from normal (systolic blood pressure $\geq$ 90 mmHg or mean arterial pressure $\geq$ 65 mmHg) to low (systolic blood pressure $<$ 90 mmHg or mean arterial pressure $<$ 65 mmHg) and vice versa after AF onset. BP = blood pressure.

Sensitivity analysis

The findings were robust across sensitivity analyses using differing covariates, cohort and NOAF definitions (Tables D.21, D.22 and D.23, respectively).

NOAF of any duration was significantly associated with an increased risk of hospital mortality. After adjustment for confounding factors, NOAF lasting over 30 minutes was significantly associated with an increased risk of hospital mortality (Figure 6.4).

6.4 Discussion

This study quantifies the haemodynamic changes associated with ICU-acquired NOAF, accounting for patients' pre-AF parameters. NOAF during an ICU admission was associated with an increase in heart rate, and a decrease in blood pressure despite an increase in vasoactive medication doses. New hypotension occurred after NOAF in one fifth of patients who went on to be treated with an intervention of interest (Table 5.1). These clinically significant haemodynamic changes associated with NOAF may partly contribute to the observed increased mortality risk.

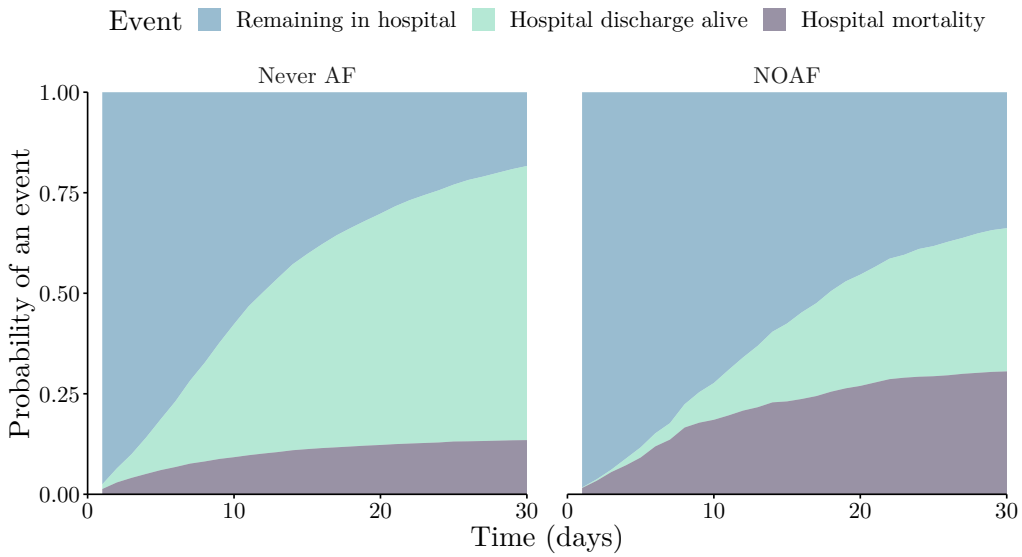


Figure 6.3: Stacked cumulative incidence curves for a) the never atrial fibrillation (AF) cohort and b) the new-onset atrial fibrillation (NOAF) cohort. These curves demonstrate the rates of discharge alive and hospital death for patients who never had AF and patients with NOAF. The cumulative incidence curves for each event are stacked vertically for each cohort. Each stacked graph will reach 1 when every patient has either been discharged or has died.

Patients who developed NOAF on an ICU had a higher probability of in-hospital mortality than those who did not. This association persisted despite adjustment for known confounders and clinically meaningful sensitivity analyses. Aetiologically, NOAF was associated with both an increased mortality rate and reduced rate of discharge per unit time. Patients who developed NOAF on the ICU were therefore exposed to an increased death rate for a longer period, resulting in a significant overall risk of dying in hospital. This study also demonstrates the association of short episodes of NOAF with hospital mortality.

Organised atrial activity contributes to ventricular filling and cardiac output[72]. NOAF precludes these mechanisms and the effects of the loss of atrial contraction on ventricular filling may be compounded by the diastolic dysfunction commonly seen during critical illness[396]. These mechanisms may explain why NOAF is associated with a reduction in cardiac index in non-ICU patients with chronic heart failure[73]. The haemodynamic findings of this work are supported by a study demonstrating that 37% of patients who developed NOAF in an ICU experienced new hypotension

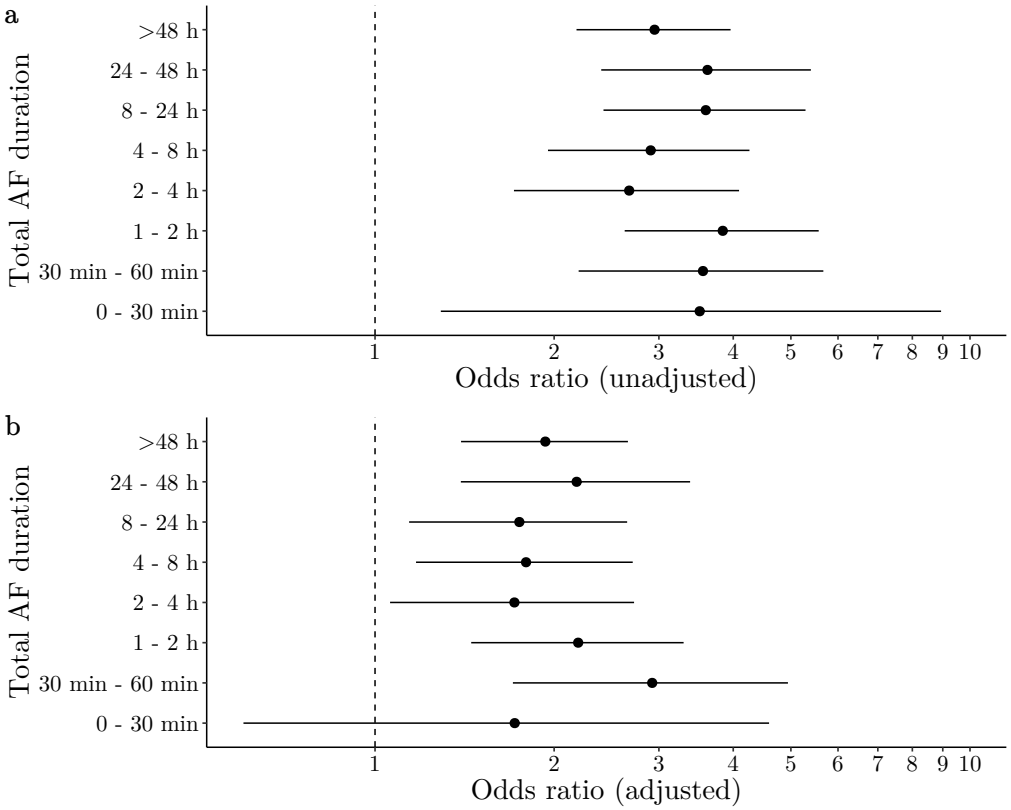


Figure 6.4: Odds ratios for hospital mortality associated with new-onset atrial fibrillation (NOAF) of different durations in a) unadjusted estimates and b) after adjustment for confounding variables. Image reproduced with permission[32] from Bedford *et al.* 2020[79].

(SBP < 90 mmHg) or a $> 50\%$ increase in vasopressor medication[41].

To the best of my knowledge, this work is the largest to date using multivariable survival analysis to assess independent prognostic associations between NOAF and hospital outcomes. Several previous studies have suggested an independent association between NOAF during critical illness and hospital mortality using logistic regression analysis[44,129,258]. However, logistic regression may underestimate the association between NOAF and ICU mortality due to immortal time bias[380,381,397]. I treated NOAF as a time-varying covariate in the primary analysis to mitigate this issue and revealed a stronger association than has been demonstrated in studies using logistic regression. Where the overall probability of an event is low, the SDHR can be interpreted as an odds ratio for the cumulative incidence of an event[102]. Thus acquiring NOAF is associated with an approximate doubling of the odds of dying in

hospital when accounting for its time-varying nature. These results are consistent with the one smaller study investigating ICU outcomes that used survival analysis[52].

I analysed the effect of NOAF on hospital LOS using methods accounting for competing risks[398]. Previous studies report an independent association between NOAF and prolonged hospital LOS[44,56]. As developing NOAF is associated with an increase in the rate of in-hospital death per unit time, the average hospital LOS will be shortened in the NOAF cohort. Studies that do not treat hospital discharge as a competing risk to mortality will therefore underestimate the effect of NOAF on prolonging hospital LOS in those patients that remain alive. NOAF significantly reduced the probability of being discharged from the hospital alive at any given time. This finding may be due to NOAF impacting on discharge decisions.

There are convincing mechanistic explanations (outlined in Section 6.1) to provide biological plausibility to these findings of an independent association of NOAF with adverse hospital outcomes. Any detrimental effect on haemodynamics may have considerable impact in any critically ill patient with minimal physiological reserve. ICU-acquired NOAF may also increase the short-term risk of thromboembolism, potentially affecting hospital outcomes[42,43].

This work revealed an increased mortality risk independently associated with NOAF episodes lasting over 30 minutes. Atrial contractile dysfunction has been demonstrated after AF episodes of one hour[367] and can last for several weeks after cardioversion[368]. This phenomenon is seen after spontaneous, chemical, and electrical cardioversion[399,400] and may occur partly due to impaired cellular calcium handling[401,402]. These findings may be in part explained by these deleterious effects of AF persisting beyond restoration of sinus rhythm in those patients with brief episodes of AF.

The limitations of this work should be considered. It is important to note that this analysis was retrospective, and the development of NOAF in the patients analysed was not independently verified. In the haemodynamic analysis, I included data from invasive and non-invasive (oscillometric) blood pressure measurements. Whether the accuracy of oscillometric blood pressure measurements is affected by

AF is debated[403]. However, a recent analysis of MIMIC-III data showed that AF did not influence the agreement of oscillometric blood pressure to invasive blood pressure in patients admitted to an ICU[404].

Although the dataset provided comprehensive data regarding severe comorbidities, including severe cardiovascular disease, it did not reliably contain information about more common, less severe comorbidities. As a result, the analysis was unable to account for these factors in the hospital outcomes analysis. However, previous research suggests that the impact of these comorbidities on ICU outcome is less significant than physiological measurements[405]. Moreover, the development of NOAF during ICU admission is primarily linked to acute factors rather than pre-existing diagnoses[40,52]. Therefore, it is unlikely that the results would have been significantly altered had more extensive comorbidity data been available.

Regarding the hospital outcomes analysis, it is worth noting that the dataset included patients with missing hospital outcome data. However, as the missing data occurred at random and represented a relatively small proportion of the total data compared to recent publications with similar characteristics[139], it is unlikely that any bias was introduced into the results.

Conclusions

NOAF is a common arrhythmia during critical illness, occurring in around one in seven patients admitted to an ICU[39]. The relationships between potential clinical risk factors and AF onset have not been well studied[38]. Better understanding these associations may help clinicians develop NOAF preventative strategies. Identifying high-risk patients may also help to target these interventions. While a small number of NOAF prediction tools have been published, they are limited in terms of their setting-specific development[231], predictor definitions[52], and practical utility[229].

The optimal treatments of NOAF once it occurs are similarly unclear. Existing research in the ICU is limited, with no large trials undertaken. As such, international guidelines[58,59,406] provide management advice not applicable to the ICU setting.

NOAF has detrimental haemodynamic consequences outside of the ICU setting[73]. These effects may be more severe in ICU patients as they are already unwell. However, the severity and time course of the haemodynamic effects have not been studied. Understanding how NOAF affects patients' blood pressure while critically unwell is important. Creating an understanding of a "normal" response to NOAF may highlight patients who require extra interventions. Lastly, studies of the hospital outcomes associated with ICU-acquired NOAF have methodological limitations which limit interpretation.

7.1 Thesis overview

In this thesis I sought to address these considerable uncertainties, increasing the understanding of NOAF in patients admitted to an ICU.

In Chapter 3, I undertook an international Delphi study to highlight candidate risk factors for use in a NOAF prediction model. The Delphi panel comprised

experts in the field, sharing 110 publications on the subject of AF during critical illness. This work systematically produced a list of variables felt to be important risk modifiers for ICU-acquired NOAF. This allowed for a detailed mechanistic exploration of their association with NOAF, with supporting clinical evidence.

This process provided a list of candidate variables that were widely available in clinical practice, with strong face validity as potential predictors. Developing a model using only predictors based on biological plausibility or clinical evidence reduced the risk of identifying spurious associations in subsequent model development work.

In Chapter 4, I developed a NOAF prediction model which outperformed a comparator model in a large external dataset. Initially, I outlined existing prediction models for NOAF in patients admitted to an ICU. At the outset of my doctoral studies, one prediction model had been published. In the last year, one further ICU-specific model has been published, and another for patients with sepsis (see Section 4.1.2). The more recent models have significant limitations in terms of their derivation and clinical utility. Importantly, several variables in both models are unlikely to be readily available in a general ICU. These studies highlight the importance of identifying candidate variables in a robust and systematic way using subject experts.

The model proposed in my thesis was built using data throughout the first week of the patients' ICU stay. It is therefore valid for use at any time in this window to estimate the risk of NOAF over the subsequent 24-hour period. The variables should be readily available at the bedside, increasing clinical utility. I explored GBTs and neural networks, using ICU-level cross-validation to mitigate the risk of overfitting to the development data. I interrogated the final GBT model, exploring variable-variable interactions, variable importance, and partial dependence curves to visualise predictor-outcome associations.

An online risk-prediction tool is then presented which would allow investigators to generate individual patient risks, or to provide risks for their data, permitting independent external validation. This tool also provides Shapley values, providing some insight into how the model has arrived at its predicted risk based on predictor

values. Clinician confidence in predictive models or decision support systems is fostered by the transparency and explainability of the tool, including in the ICU[407,408]. Providing graphical insights via the web tool, in the context of systematically-derived predictors, may therefore bolster trust among users.

In Chapter 5, I undertook a multi-centre retrospective study of NOAF treatment efficacy. This was the first study to my knowledge where physiological status after AF onset was accounted for, important as it may influence treatment choice. The study used data from multiple databases and balanced admission covariates plus pre-treatment physiological variables across multiple treatment groups. Digoxin therapy was associated with inferior rate control compared with amiodarone. In addition, CCB therapy was associated with inferior rhythm control compared with amiodarone, consistent with one previous study[307]. The study suggested a mortality benefit in the beta blocker treatment group compared to amiodarone in the unadjusted analysis. However, after adjustment, including for physiological variables after AF onset, there was no significant difference in mortality between beta blockers and amiodarone. This suggests previously demonstrated reduced mortality associated with beta blocker therapy may be explained by favourable physiological status prior to treatment. These findings have provided insights into how best to manage NOAF in critically ill patients. Based on this work, the NIHR have suggested a randomised trial of beta blockers versus amiodarone in the treatment of ICU-acquired NOAF, which may further inform treatment strategies.

In Chapter 6, I undertook a comprehensive analysis of the haemodynamic changes associated with NOAF. Previous work investigating the haemodynamic consequences of NOAF in the ICU is very limited. One previous study of 139 patients with NOAF reported new haemodynamic instability in 52 (37%) of patients, but the timing and severity of this instability was not clear. I found that NOAF was associated with an immediate blood pressure drop and increase in vasoactive medication use which persisted over the six hours post-AF onset.

Findings from Chapter 6 demonstrated an increased risk of hospital mortality associated with NOAF, identified using competing risks methodology. This method

estimated the *prognostic* association of NOAF with hospital mortality, accounting for its relationship with reduced rates of hospital discharge per unit time. An analysis that incorporated the time-varying nature of AF onset mitigated against the immortal time bias not accounted for in most previous studies. I revealed a stronger association than has been demonstrated in previous studies using LR, which aligned with one previous study using competing risks methodology[52]. I also found that, after adjustment for confounding factors, NOAF lasting over 30 minutes was associated with an increased risk of hospital mortality. These findings may in part be explained by the haemodynamic impact of NOAF identified above.

7.1.1 Strengths

The thesis presents a thorough exploration of NOAF in ICU patients, covering the patient journey from admission to outcome. Beginning with an investigation into the predictors of NOAF, this research then focused on the optimal treatments, ultimately analysing the associated haemodynamic changes and associated hospital mortality.

Another strength of this work lies in its robust and methodical approach. The use of an international Delphi study for identifying candidate risk factors ensured the inclusion of variables underpinned by strong face validity and biological plausibility. This approach minimised the risk of spurious associations, strengthening the credibility of subsequent model development.

Moreover, the thesis provides a comprehensive analysis of the haemodynamic changes associated with NOAF, a subject area that has been limited in previous research. The discovery of an immediate BP drop and an increase in vasoactive medication use post-NOAF onset adds a new dimension to the understanding of the haemodynamic impact of NOAF.

In addition, the adoption of competing risks methodology provided a more robust estimation of the prognostic association of NOAF with hospital mortality than studies using LR.

The development of the NOAF prediction model in this thesis explored machine learning techniques. The use of these techniques paired with ICU-level cross-validation aimed to limit overfitting, enhancing the model's external validity. Furthermore, the construction of an online risk-prediction tool incorporating elements of explainability extended the clinical utility of this work.

7.1.2 Limitations

While this thesis provides valuable insights into NOAF in critically ill patients, its limitations should be considered when interpreting the findings.

My work so far has predominantly involved retrospective analysis of routinely collected data. While this has strengths including scale and granularity, it is limited by missing data and lack of prospectively validated data labels e.g. heart rhythm.

The Delphi study conducted in chapter 3 had a relatively small sample of experts in the field. While efforts were made to ensure that a diverse range of experts were included in the panel, it is possible that some important risk factors may have been missed due to a lack of representation from certain specialties or regions.

The study of treatment effectiveness conducted in chapter 5 was limited by its observational design. While extensive efforts were made to balance admission covariates plus pre-treatment physiological variables across multiple treatment groups, there may still be unmeasured confounding variables that could influence treatment choice and patient outcomes.

While the development data used in Chapter 4 was large, the resultant prediction model may not be transferable to all ICU settings as the sites included may not represent the full spectrum of such units. Lastly, it is important to note that the findings of this thesis are based on data from critically ill patients and are not generalisable to other populations.

7.2 Parallel / future work

7.2.1 Longer-term outcomes after ICU-acquired NOAF

I have been involved in related collaborative work not detailed in this thesis. I was first author of a retrospective cohort study of 4615 patients with NOAF admitted to ICUs across England between 2009 and 2016 undertaken as part of a separate NIHR-funded project[409]. We analysed patient records from the Risk-II database of routinely collected data from the Case Mix Programme (CMP) national clinical audit of adult intensive care[410] linked to Hospital Episode Statistics (HES) for England[411] and the Office for National Statistics (ONS) mortality databases.

We found that patients who developed NOAF in the ICU were also at a higher risk of subsequent hospitalisation with AF (adjusted CSHR 5.86; 95% CI 5.33 – 6.44), stroke (adjusted CSHR 1.47; 95% CI 1.12 – 1.93), and heart failure (adjusted CSHR 1.28; 95% CI 1.14 – 1.44) over the subsequent five years independent of pre-existing comorbidities compared with controls. Following this work, I made significant contributions to a systematic review and meta-analysis[412]. This work focused on studies of longer-term outcomes after ICU-acquired NOAF. We found that in patients who develop NOAF in an ICU, both 90-day and 1-year mortality are increased in comparison to those who do not develop NOAF. We also demonstrated an increased risk of thromboembolic events after hospital discharge in these patients. Together, this suggests that much of the NOAF associated with critical illness does not resolve with treatment of the underlying condition.

Our work provides the evidence to support current guidelines stating that patients developing NOAF secondary to a potentially "reversible" condition should receive careful follow-up[58] countering the notion in previous guidelines that successful treatment of the underlying condition often eliminates AF[413]. This work suggests patients who develop NOAF during an ICU admission are likely to benefit from cardiological surveillance and follow-up after hospital discharge.

7.2.2 Anticoagulation for NOAF during ICU admission

Multiple risk stratification scores have been developed to assess bleeding risk in patients with AF in the community. Stroke and bleeding risk factors overlap, but the balance of risks is such that patients with high stroke risk and high bleeding risk still benefit from anticoagulation[414]. High bleeding risks in the general population should therefore not generally result in withholding anticoagulation. These scores can however allow for identification of modifiable risk factors for bleeding and closer monitoring of high-risk patients.

Patients who develop NOAF during sepsis are at higher risk of in-hospital and post-discharge stroke and death compared to those who do not, despite adjusting for confounders[45,64]. Those with NOAF are at higher risk of thromboembolic complications than patients with pre-existing AF[45]. The individual risk of thromboembolic events in patients who develop NOAF during an ICU admission is poorly understood, but the work above (see Section 7.2.1) has highlighted an association with post-discharge stroke risk[409,412]. Risk assessment scores have not been developed or validated in the ICU population. These scores are unlikely to reflect the risk of thromboembolic events in critically ill patients due to prothrombotic changes associated with critical illness. The CHA_2DS_2-VASc score has been shown once to be predictive of thromboembolic event risk in the ICU setting although with poor sensitivity and predictive value[42].

Similarly, bleeding risk assessment scores have been developed and validated in community populations with different risk profiles to patients on an ICU. While community-based bleeding risk scores are chiefly comprised of chronic comorbidities, the bleeding risk in patients on an ICU is likely to be more related to acute factors such as illness severity, systemic inflammation, type and location of surgery, nutritional status, invasive devices, and acute coagulopathy[63,415,416]. Bleeding risk in critically ill patients with NOAF is higher than patients in the community, with one study demonstrating that 9% of patients who received systemic anticoagulation had significant bleeding warranting cessation of anticoagulation and at least one blood transfusion[41].

After our scoping review it remained unclear whether or when to administer therapeutic anticoagulation for stroke prevention in patients in an ICU with NOAF[276]. Limited evidence suggested that bleeding risks outweigh the increased risk of thromboembolism while in the ICU[41,417], but optimal timing of anticoagulation is unknown. Two review articles[284,418] proposed a patient-centred approach to only administer anticoagulants in patients with high risk of arterial thromboembolic events. Notably, 64% of respondents of a UK-wide survey reported that they would *not* use anticoagulant therapy in critically ill patients with NOAF[70]. Reluctance to initiate anticoagulation may be due to the uncertainty of the risk-benefit balance due to lack of clear evidence. While in the community, anticoagulation can be recommended for almost all patients with AF, the balance of risks in patients on an ICU is likely to be more complex and dynamic. Modified risk scores which incorporate such complexities may therefore inform the management of critically ill patients with NOAF. In the absence of these scores, a pragmatic approach may be to initiate anticoagulation at the time of hospital discharge, thereby avoiding the bleeding risks associated with medical illness and invasive interventions.

The Tracking Atrial Fibrillation aFter ICU (TrAFFIC) study, outlined below, will improve our understanding of post-discharge AF recurrence and other complications. This may inform the development of risk stratification tools or allow the assessment of the applicability of current tools to the ICU NOAF population. This may ultimately lead to the identification of patients with a sufficiently high risk of future thromboembolism to warrant consideration of anticoagulation during their ICU or hospital stay.

7.2.3 Recurrence of AF after ICU discharge

It is not known whether patients who develop an episode of NOAF which resolves during an ICU admission are at an increased risk of recurrent AF, and how this risk relates to the characteristics of their AF episode. It is unclear whether NOAF developing during an ICU admission predisposes to future episodes of AF and thromboembolism as it does after other types of hospital admission[419,420]. Similarly,

it is unknown whether AF persists in patients discharged from ICU with ongoing AF, or whether it mainly resolves while still in hospital or during the recovery period at home. This information could inform appropriate follow-up and management of this vulnerable group of patients.

The TrAFFIC study is a prospective observational study of ECG monitoring after ICU admission complicated by NOAF[421]. ECG data is collected from wearable monitors while the patient is on the ward, and again three months post-hospital discharge. I developed the study protocol with a co-investigator, and am currently managing the study, recruiting patients, and reviewing ECG data to confirm AF.

The TrAFFIC study is currently being run in two ICUs in Oxford, with expansion to other sites planned for the near future, including internationally. Expanding the study will diversify patient demographics and healthcare systems. This will enhance the generalisability of the findings, making them applicable to a broader patient population. I hope this multinational expansion will foster collaboration promoting a more effective approach to managing this complex and prevalent issue in critical care.

7.2.4 Trial enrichment using individualised risk prediction

Individual risk prediction is a tool that could enhance the efficiency of trials of preventative treatments in the ICU. By identifying high-risk patients, resource allocation can be optimised by focusing preventative measures where they are most likely to be effective. Reliable risk stratification would avoid low-risk patients being unnecessarily exposed to these preventative interventions.

Youden's J statistic is a summary statistic from ROC analysis which estimates one measure of a test's optimal threshold[422]. If a risk prediction model were being used to risk stratify patients, this method could be employed to identify a useful risk cut-off for study enrollment. With equal weight given to sensitivity and specificity, Youden's J is calculated as $J = Se + Sp - 1$, where Se is the model sensitivity and Sp is the model specificity at different thresholds. The risk threshold that maximises J is a candidate for the optimal cut-point. Depending on the trial, it may be more important to not miss participants (high sensitivity), or conversely

to minimise unnecessary recruitment (high specificity). Sensitivity and specificity having an equal weight may therefore not be appropriate. In this case a weighted Youden's J can be calculated. The weight (r) accounts for the prevalence of the condition (π) and the relative cost of a false negative classification compared to a false positive classification (ζ). The optimal cut-point can then be determined[423]. Where $\pi = 0.5$ and $\zeta = 1$ (false positive and false negative predictions being equivalent), $r = 1$ and the equation is equivalent to the unweighted J index. Where high sensitivity is of primary interest, ζ is increased, reducing r and the relative numerical importance of specificity.

The weighted Youden's J can be calculated as:

$$J = Se + (r \times Sp) - 1$$

where:

r is the weight, calculated as $\frac{1-\pi}{\zeta \times \pi}$
π is the prevalence of the condition
ζ is the relative cost of a false negative classification

Researchers could use these methods with the NOAF prediction model detailed in Chapter 4 to identify risk thresholds based on their NOAF prevalence and feasible enrollment rates. Figure 7.1 shows weighted Youden's J values over different risk thresholds and different values of ζ when using the final GBT NOAF prediction model from Chapter 4. Researchers could select thresholds based on desired outcome proportions which could then feed into power calculations. I hope to build on my work in this thesis and develop a tool for researchers looking to enrich trials of preventative interventions using these methods.

7.3 Concluding remarks

The work presented in this thesis has investigated key aspects of NOAF acquired during an ICU admission. A NOAF prediction model developed using systematically-identified predictors has the potential to enrich future trials of prevention studies. A

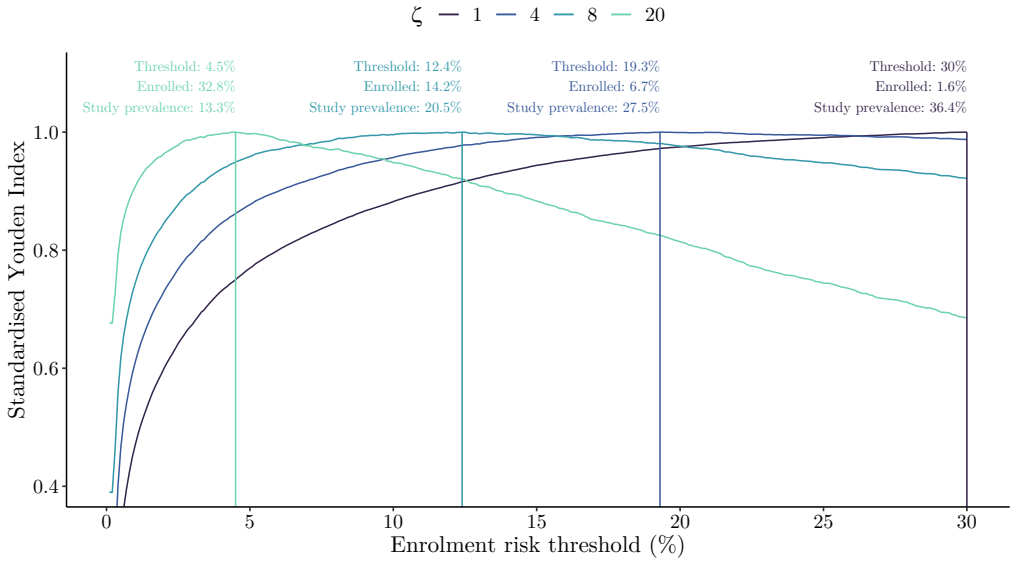


Figure 7.1: Weighted Youden's J values from the final gradient-boosted tree model developed in Chapter 4 using the development data. J values are calculated over different risk thresholds and different values of ζ, representing the relative cost of a false negative classifications. Threshold values represent the recruitment risk threshold. As the cost of false negative classifications (ζ) increases, higher sensitivity is prioritised, lowering the risk threshold. At lower thresholds, a higher percentage of all patients would be enrolled at the expense of a lower prevalence in the study cohort of the condition being predicted.

detailed analysis of comparative NOAF treatment efficacy has increased our understanding of optimal drug choices and paved the way for an upcoming randomised trial. By exploiting the granular nature of ICU data, the haemodynamic changes associated with ICU-acquired NOAF have for the first time been revealed. Lastly, a competing risks analysis accounting for the time-varying nature of NOAF revealed the considerable prognostic association with increased hospital mortality.

Appendices

The resting membrane potential

The plasma membrane plays an important role in establishing resting E_m in cardiac cells[18]. It acts as a selective barrier, controlling the passage of different ions based on its permeability to each type. The movement of ions across the membrane is also influenced by the ionic composition of the intracellular and extracellular fluid. The generation and magnitude of E_m results from the unequal distribution of charged particles across the plasma membrane.

A membrane potential can be demonstrated with two chambers containing different concentrations of potassium chloride (KCL), separated by a membrane permeable to K^+ ions. K^+ ions move from the chamber with higher concentration to the chamber with lower concentration, transferring positive charge in the process. As a result, one chamber becomes more positively charged, while the other loses positive charge, creating an electrical gradient. This electrical force opposes the ion movement caused by the concentration difference. The concentration difference will continue to drive ions from one chamber to the other until the electrical gradient fully opposes the movement of ions driven by the concentration difference. At this point, the electrical and concentration differences will balance—while K^+ ions continue to move, their net movement ceases as the number leaving one chamber equals those arriving at the other. This results in an equilibrium potential: the voltage at which the electrical difference precisely matches the concentration difference between the chambers.

The equilibrium potential (E_{ion}) can be calculated using the Nernst equation as: $E_{ion} = \frac{RT}{zF} ln \frac{[ion]_{out}}{[ion]_{in}}$ where:

E	is the equilibrium potential of ionic species (V)
R	is the gas constant (8.314 $J.K^{-1}mol^{-1}$)
T	is the absolute temperature (K)
z	is the valence of the ion
F	is the Faraday constant (96,485 C/mol^{-1})
ln	is the logarithm to base e
$[ion]_{\text{out}}$	is the extracellular concentration of the ion
$[ion]_{\text{in}}$	is the intracellular concentration of the ion

This equation can be partially solved assuming a temperature of 37° Celcius, becoming $E_{ion} = \frac{61.5}{z} log_{10} \frac{[ion]_{out}}{[ion]_{in}}$.

The inward rectifyer current

The ion that contributes most significantly to E_m is K^+. Provided K^+ channels are open, K^+ will diffuse out of the cell down the concentration gradient, driving the membrane potential toward E_K (around -89 mV). The current with the largest contribution to E_m is I_{K1} (the inward rectifyer current) and relies on a certain type of K^+ channel known at the inward rectifier (K_{IR}).

Resting E_m is maintained by the outward flow of K^+ ions through open K_{IR} channels. This outward current is opposed by smaller inward "leak" currents of ions through Na^+ and Ca^{2+} channels. This inward leak current prevents E_K from being reached.

A NOAF prediction model using admission-only data

B.1 Introduction

This appendix outlines the initial model development process which preceded the work in Chapter 4. In this appendix, I develop a prediction model using admission-only data.

B.2 Methods

B.2.1 Data sources and participants

This analysis used patient data from the PICRAM and MIMIC-IV databases as in Chapter 4 (see sections 4.2.1 and 4.2.2 for details).

Initially, I developed a logistic regression model using admission-only data. To account for missing data I used multiple imputation, generating 10 imputed datasets. To develop a parsimonous model, retaining the most important predictors, I used stepwise backwards elimination using the Wald test p-value based on pooled estimates at each stage[424]. First, a model incorporating all predictors was fitted in each imputed dataset. Model parameter estimates were then pooled into a single model using Rubin's rules. A single variable was then removed based on the Wald test p-value threshold, using the median p-value method[425]. A model incorporating $p-1$ predictors was then fit to each imputed dataset with subsequent pooling and single variable elimination. This procedure was repeated until no further variables were removed, resulting in a final pooled model. I selected a p-value threshold for variable elimination which accounted for sample size.

I allowed for non-linear associations using restricted cubic splines for the following variables: age, serum potassium concentration, serum magnesium concentration, serum sodium concentration, blood pH, blood white cell count, haemoglobin concentration, heart rate, temperature, and MAP.

B.2.2 Predictors

Predictors are listed in Table 4.1. For this admission-only model, the dynamic features (previous maximumum, minimum etc. were not used). I used the first recorded pre-AF variable value. If no value was available pre-AF and in the first 24 hours, this was treated as missing.

B.2.3 Internal validation

I used bootstrapping with nested multiple imputation to assess the internal validity of the admission data logistic regression model. This was a computationally intensive process where the variable selection process above is repeated within bootstrap resamples. I created custom R functions based on the *psfmi* R package, updated to allow parallel processing. The internal validation procedure is outlined below:

1. Generate 100 bootstrap samples from the original data, some of which is missing.
2. Within each bootstrap resample:
    a Generate 10 imputed datasets (imputed bootstrap datasets).
    b Perform stepwise backward elimination using pooled estimates as with the main model, resulting in a single model.
    c Calculate the average performance of this model across the 10 imputed bootstrap datasets.
    d Calculate the average performance of this model across the 10 imputed *original* datasets.
    e Calculate the difference (optimism) in model performance measures between the bootstrap and original datasets.
3. Repeat step 2 over each bootstrap resample and average the optimism estimates across the resamples.

B.2.4 Internal-external cross-validation

The entire model-building process was repeated with all but one of the ICUs used for model development, with the held-out ICU used for validation. This process was repeated, on each occasion omitting one ICU to ascertain a measure of external validation performance (internal external cross-validation—IECV).

B.3 Results

Participants and characteristics are shown in Chapter 4 Section 4.3.1. Missing data are explored in detail in Chapter 4 Section 4.3.2.

B.3.1 Pooled model

After the multiple imputation variable selection process, the pooled model consisted of 11 predictors:

1. Age
2. Weight
3. Congestive cardiac failure
4. Mechanical ventilation
5. Presence of atrial ectopics
6. Serum [urea]
7. Serum [albumin]
8. Serum [sodium]
9. Presence of a central venous catheter
10. FiO_2
11. Heart rate

Apparent performance

The apparent C statistic of the pooled model using only admission variables was $C = 0.786$; 95% CI 0.778 – 0.793.

B.3.2 Internal validation

After the internal validation procedure, the optimism-adjusted C statistic was $C = 0.783$. The optimism-adjusted calibration slope S was $S = 0.981$ (Table B.1).

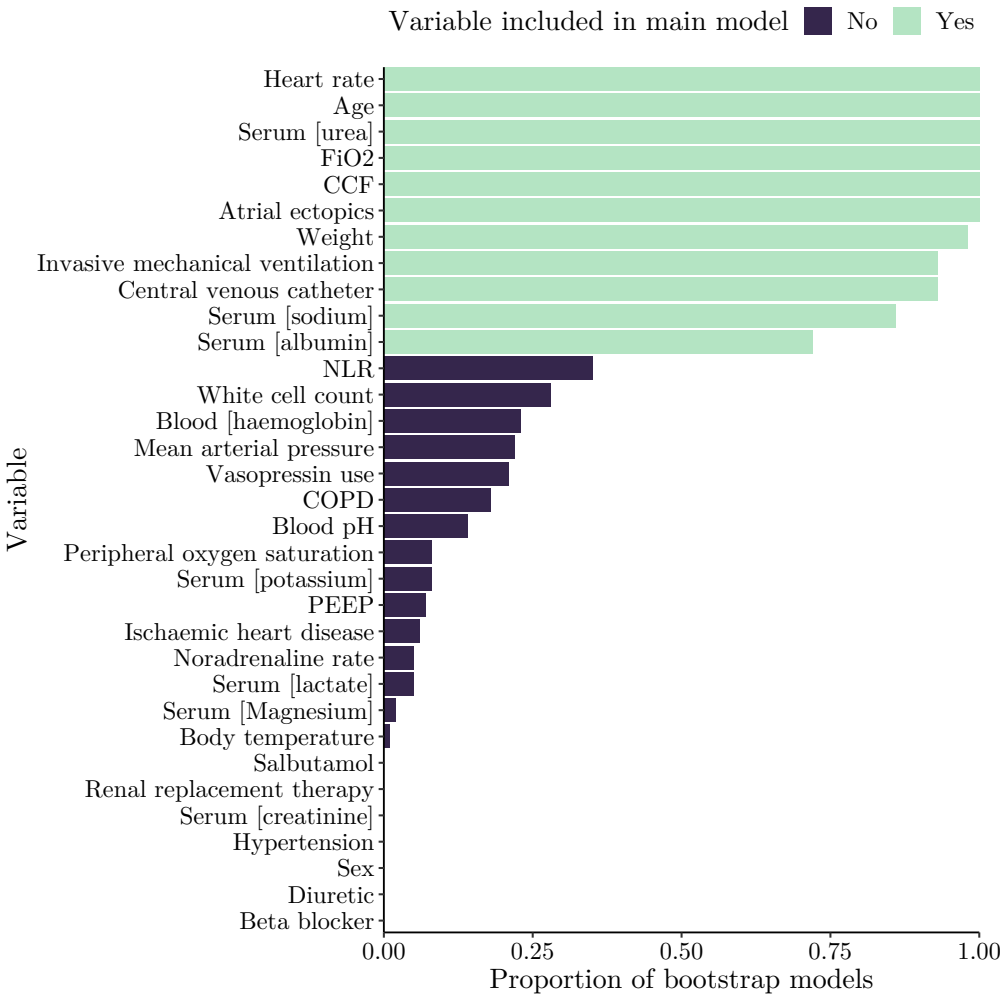


Figure B.1: Variables selected as a proportion of bootstrap resamples. CCF = Congestive cardiac failure; FiO_2 = Fraction of inspired oxygen; NLR = Neutrophil:lymphocyte ratio; COPD = Chronic obstructive pulmonary disease; PEEP = positive end-expiratory pressure.

Table B.1: Internal validation performance statistics

	Original	Bootstrap apparent	Bootstrap test	Optimism	Corrected
C statistic	0.786	0.788	0.785	0.003	0.783
Calibration slope	1.000	1.000	0.981	0.019	0.981

Variable selection

The internal validation process demonstrated good stability in variable selection. Of the 11 predictors in the pooled model, 6 were selected in 100% of bootstrap resamples. Of the variables selected in the initial pooled model, the least frequently selected in the bootstrap resamples was serum [albumin], selected in 72% (Figure B.1).

B.3.3 Model coefficients

The model coefficients before and after shrinkage are shown in Table D.24. The model equation after shrinkage is displayed below.

$$
\begin{aligned}
\text{Prob}\{\text{NOAF}\} &= \frac{1}{1+\exp(-X\beta)}, \text{ where} \\
X\hat{\beta} &= -6.0397 + 0.0059 \times \text{Weight} + 0.3391 \times \text{Mechanical ventilation} \\
&\quad +0.0317 \times \text{Urea} + 0.6580 \times \text{Atrial ectopics} \\
&\quad -0.0157 \times \text{Albumin} + 0.3036 \times \text{cvc} \\
&\quad +0.0095 \times \text{FiO2}(\%) + 0.4966 \times \text{CCF} \\
&\quad +0.0827 \times \text{Age} - 0.00001 \times (\text{Age} - 34)_+^3 \\
&\quad +0.00003 \times (\text{Age} - 62)_+^3 - 0.00002 \times (\text{Age} - 83)_+^3 \\
&\quad -0.0240 \times \text{Sodium} + 0.00003 \times (\text{Sodium} - 132)_+^3 \\
&\quad -0.00007 \times (\text{Sodium} - 138)_+^3 + 0.00004 \times (\text{Sodium} - 143)_+^3 \\
&\quad +0.0054 \times \text{Heart rate} + 0.000003 \times (\text{Heart rate} - 69)_+^3 \\
&\quad -0.000006 \times (\text{Heart rate} - 92)_+^3 \\
&\quad +0.000002 \times (\text{Heart rate} - 121)_+^3 \\
\text{and } (x)_+ &= x \text{ if } x > 0, \text{ otherwise } (x)_+ = 0
\end{aligned}
$$

B.3.4 Internal-external cross-validation

Discriminative performance was consistent across the seven IECV folds and similar to that estimated during internal validation (Table B.2).

The frequency that variables were selected during each of the seven IECV model development folds is shown in Table B.3. Of the eleven variables selected in the main model, nine were selected in all seven IECV folds, with two (serum [albumin] and mechanical ventilation] being selected six times. Four variables not selected in the final model (COPD, NLR, PEEP and blood pH) were selected in only one IECV fold.

Table B.2: Internal-external validation performance of admission model.

Held-out ICU	C-statistic (95% CI)	Calibration slope (95% CI)
Oxford General ICU	0.774 (0.753 – 0.794)	1.02 (0.90 – 1.13)
Oxford Churchill ICU	0.768 (0.730 – 0.802)	1.10 (0.90 – 1.31)
Reading General ICU	0.778 (0.744 – 0.809)	1.13 (0.93 – 1.34)
Boston Medical ICU	0.785 (0.768 – 0.801)	0.92 (0.84 – 1.01)
Boston Medical/Surgical ICU	0.791 (0.773 – 0.808)	1.03 (0.92 – 1.13)
Boston Surgical ICU	0.776 (0.755 – 0.796)	0.95 (0.84 – 1.05)
Boston Trauma ICU	0.784 (0.763 – 0.803)	0.92 (0.82 – 1.03)

Table B.3: Variables selected across seven internal-external validation folds.

Variable	Times selected (n / 7)
Atrial ectopics	7
CCF	7
CVC	7
FiO_2	7
Serum [urea]	7
Age	7
Serum [sodium]	7
Heart rate	7
Weight	7
Serum [albumin]	6
Mechanical ventilation	6
COPD	1
NLR	1
PEEP	1
Blood pH	1

CCF = Congestive cardiac failure; CVC = Congestive cardiac failure; FiO_2 = Fraction of inspired oxygen; NLR = Neutrophil:lymphocyte ratio; COPD = Chronic obstructive pulmonary disease; PEEP = positive end-expiratory pressure.

B.4 Discussion

This model using admission variables to predict AF over an ICU admission performed less well than one published model[52]. The published model used variables which were updated on a day-by-day basis to predict AF over the subsequent day, suggesting that incorporating dynamic variables may provide improved prediction performance.

This appendix followed current best practices for prediction model development

for a binary outcome. This included nested multiple imputation and variable selection within bootstrap resampling during internal validation[426]. Model coefficients were pooled after multiple imputation, with variable elimination occurring after each pooling step.

I used internal-external cross-validation to assess heterogeneity in model performance across ICUs. Repeating the variable selection process outlined above within each internal-external validation fold revealed how reliably different variables were selected when individual ICUs were left out of the model development process. The results suggested consistency in variables, with 9/11 variables being selected in every IECV fold, and the remaining 2 being selected in 6 of 7 folds. Of those variables not selected in the final model, the maximum number of IECV folds in which they were selected was 1/7.

The process of nested multiple imputation and variable selection within bootstrap resampling, and repeating these processes to perform IECV is computationally intensive. It represents the state-of-the-art in statistical model development but is currently rarely performed in published prognostic model development studies[427].

Supplementary figures

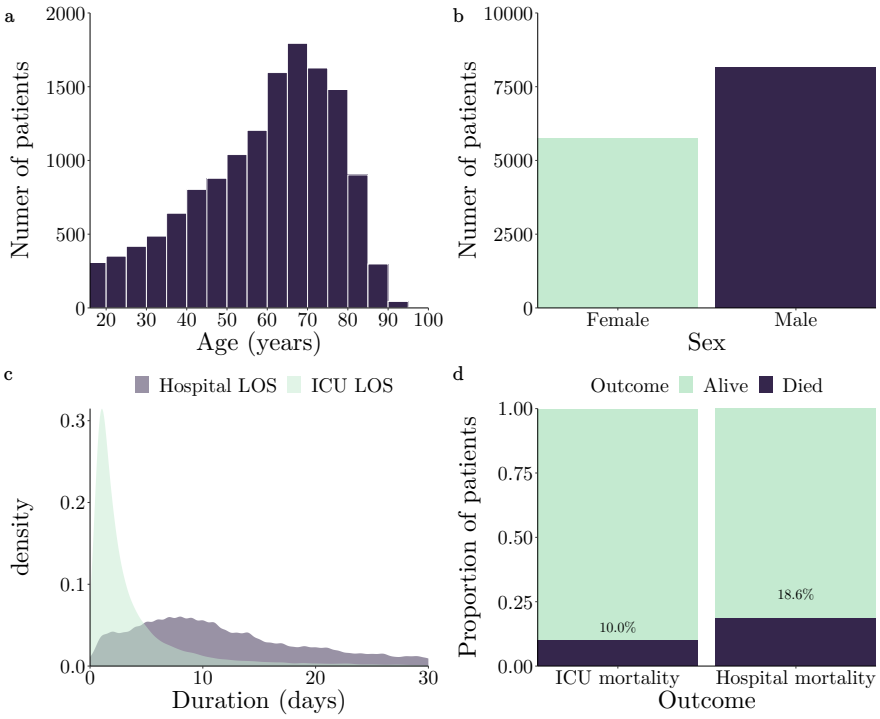


Figure C.1: Demographics Summary: PICRAM Database. a) bar chart of age; b) bar chart of sex; c) density plot of ICU and hospital length of stay; and d) bar chart of patient ICU and hospital mortality.

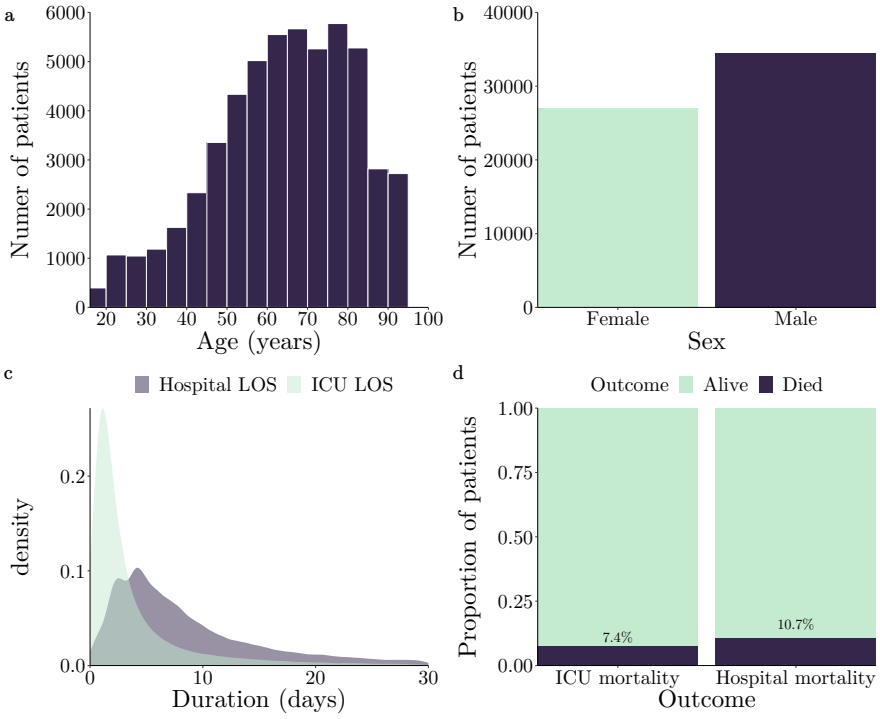


Figure C.2: Demographics Summary: MIMIC-III Database. a) bar chart of age; b) bar chart of sex; c) density plot of ICU and hospital length of stay; and d) bar chart of patient ICU and hospital mortality.

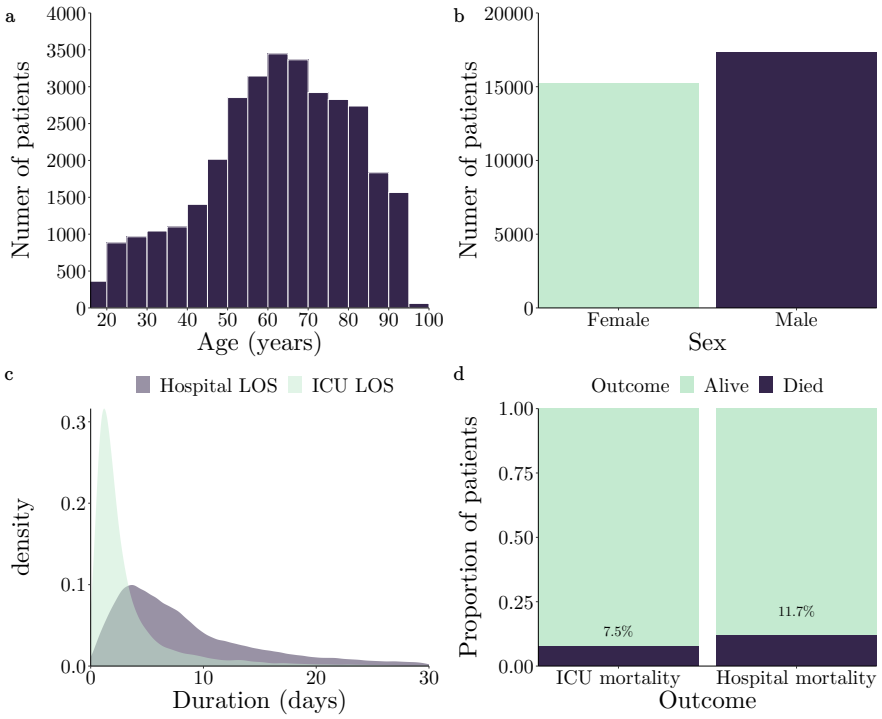


Figure C.3: Demographics Summary: MIMIC-IV Database. a) bar chart of age; b) bar chart of sex; c) density plot of ICU and hospital length of stay; and d) bar chart of patient ICU and hospital mortality.

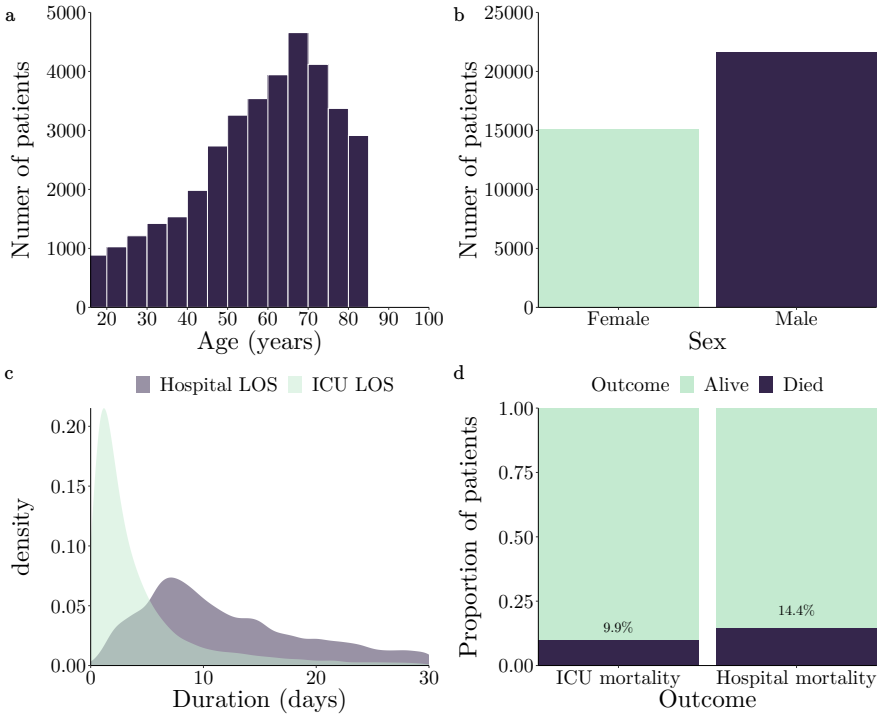


Figure C.4: Demographics Summary: CCHIC Database. a) bar chart of age; b) bar chart of sex; c) density plot of ICU and hospital length of stay; and d) bar chart of patient ICU and hospital mortality.

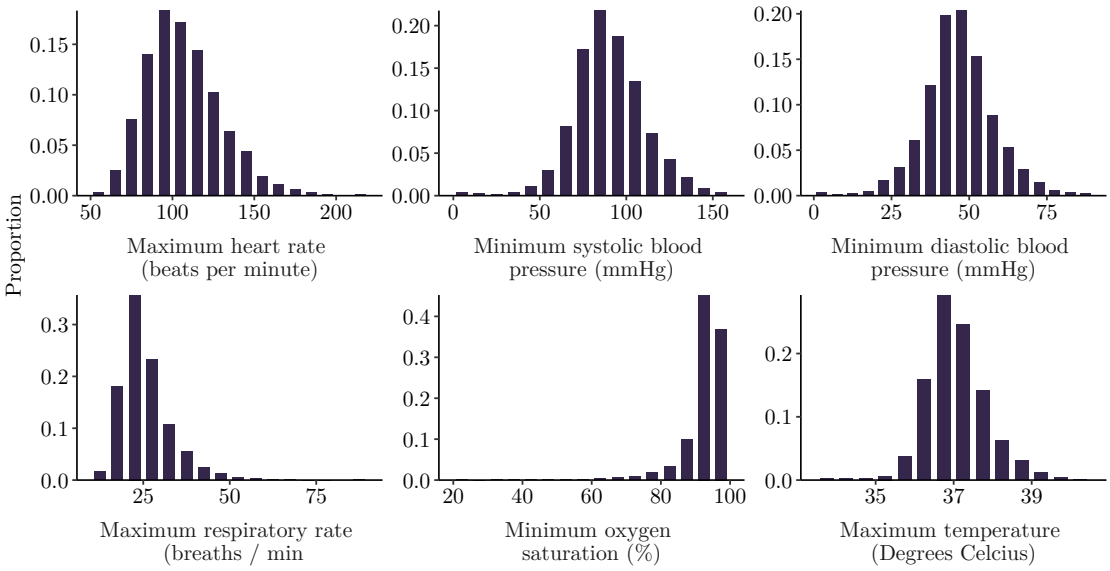


Figure C.5: Histograms of key vital sign distributions at admission (first 24 hours) from the PICRAM database.

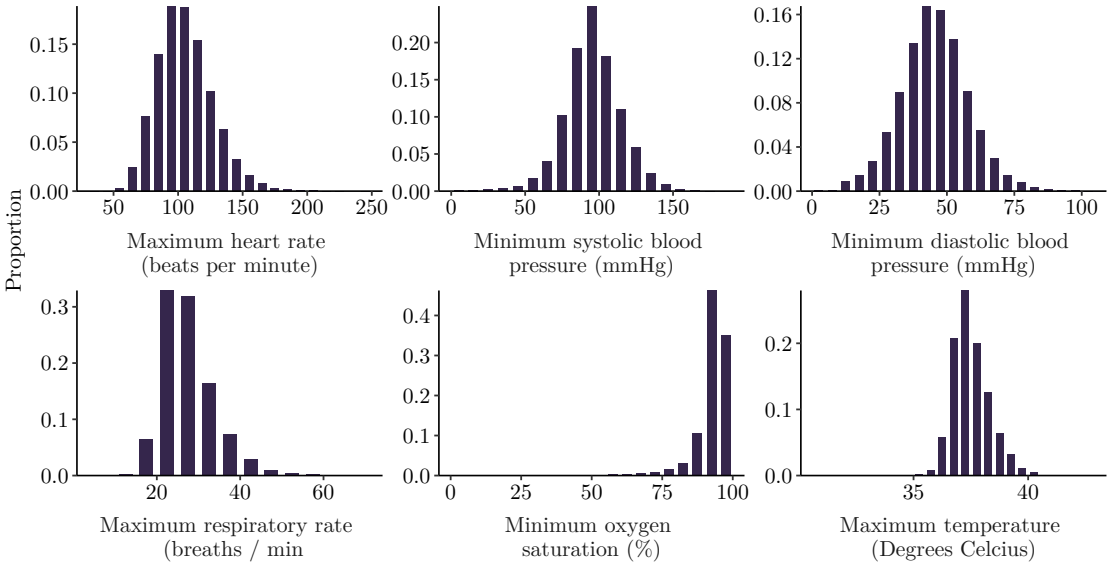


Figure C.6: Histograms of key vital sign distributions at admission (first 24 hours) from the MIMIC-III database.

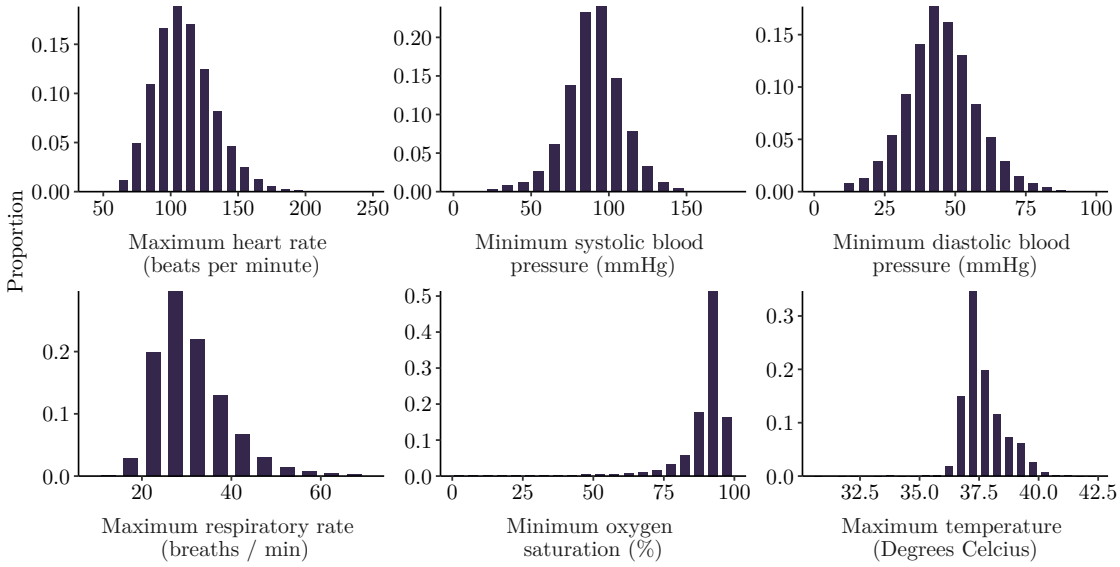


Figure C.7: Histograms of key vital sign distributions at admission (first 24 hours) from the MIMIC-IV database.

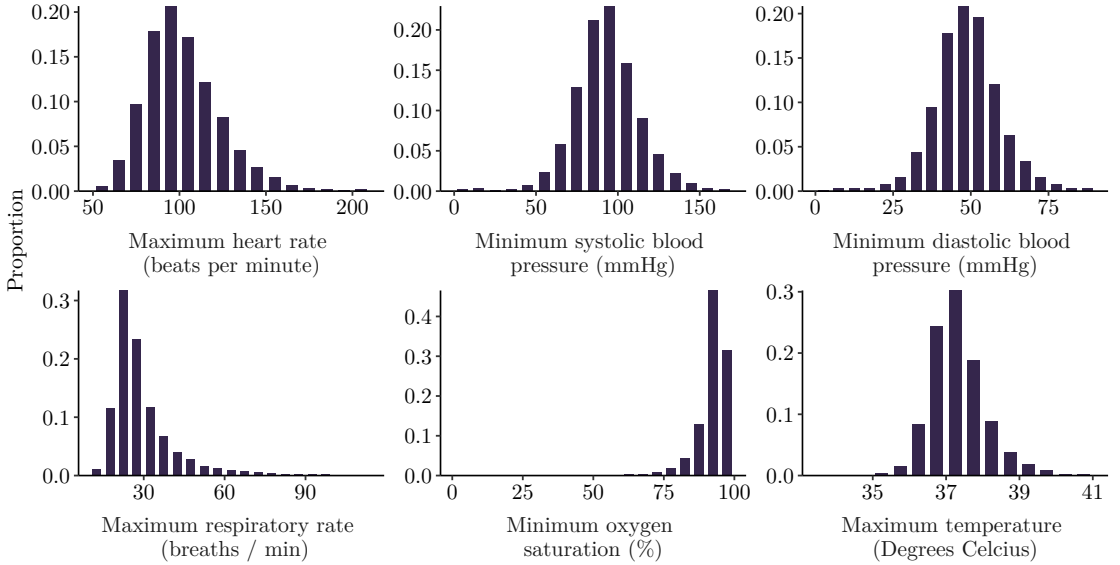


Figure C.8: Histograms of key vital sign distributions at admission (first 24 hours) from the CCHIC database.

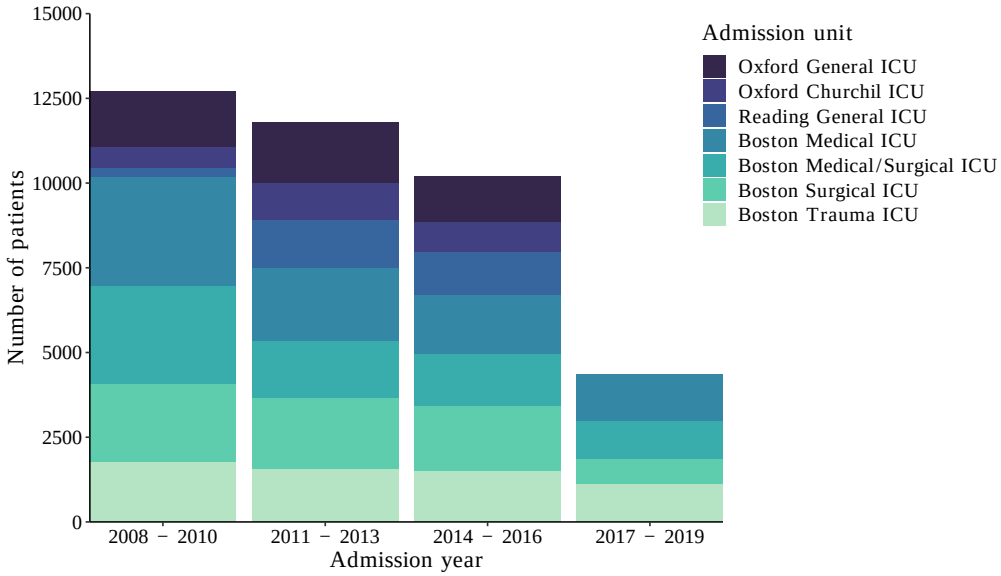


Figure C.9: Numbers of patients by admission year and ICU in the combined PICRAM and MIMIC-IV model development dataset.

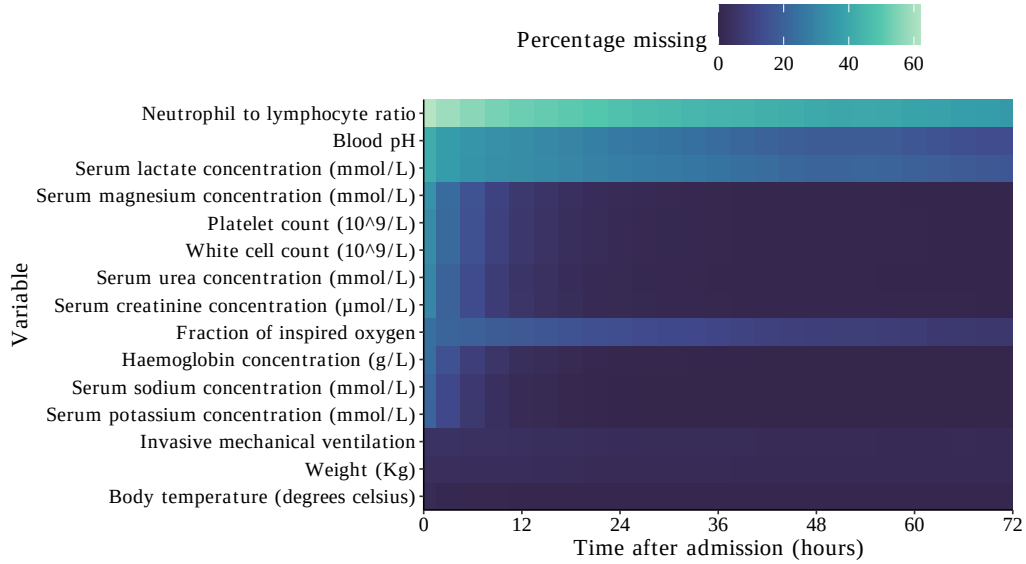


Figure C.10: Variable missingness over time. Data shown for variables with missingness greater than 1% in the first 3h of admission. All other variables were <1% missing.

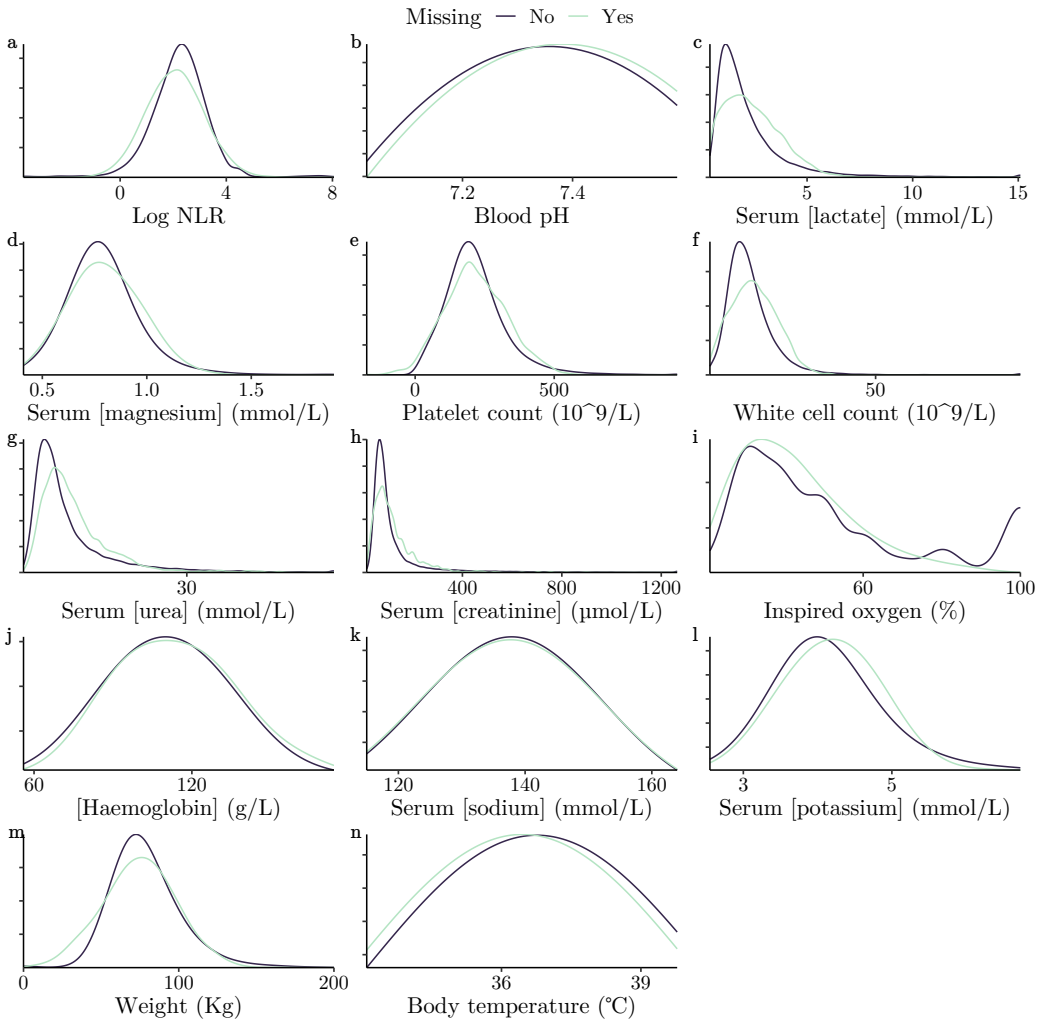


Figure C.11: Density plots comparing distributions of observed data with imputed data. The imputed curve (turquoise) plots the density of the mean imputation over the m datasets. The purple distributions are the those of the observed data. NLR = Neutrophil:lymphocyte ratio.

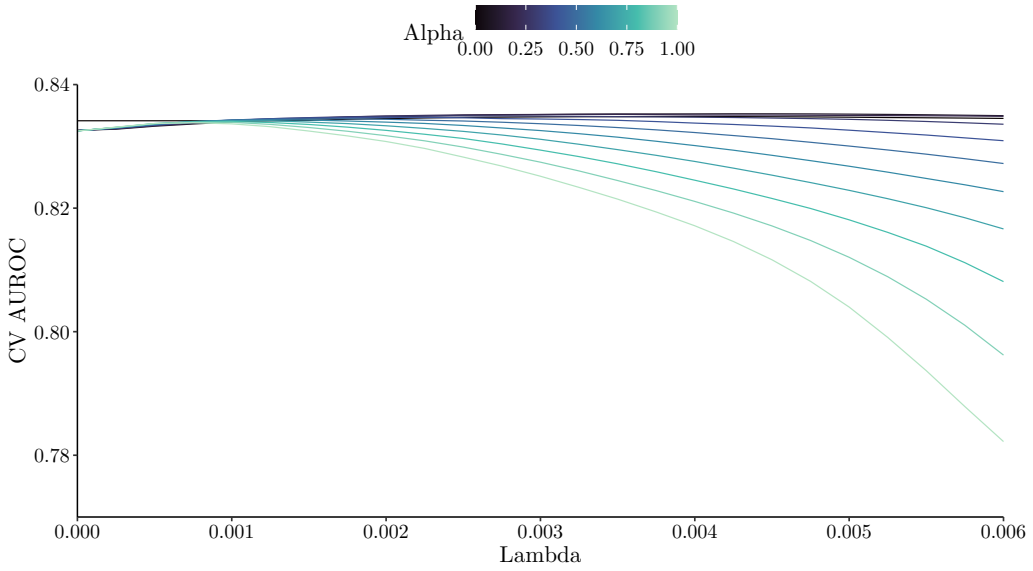


Figure C.12: Model training: Penalised LR model. Cross-validation AUROC values in the development data across a range of alpha and lambda values for a demonstration model comprising 23 predictors.

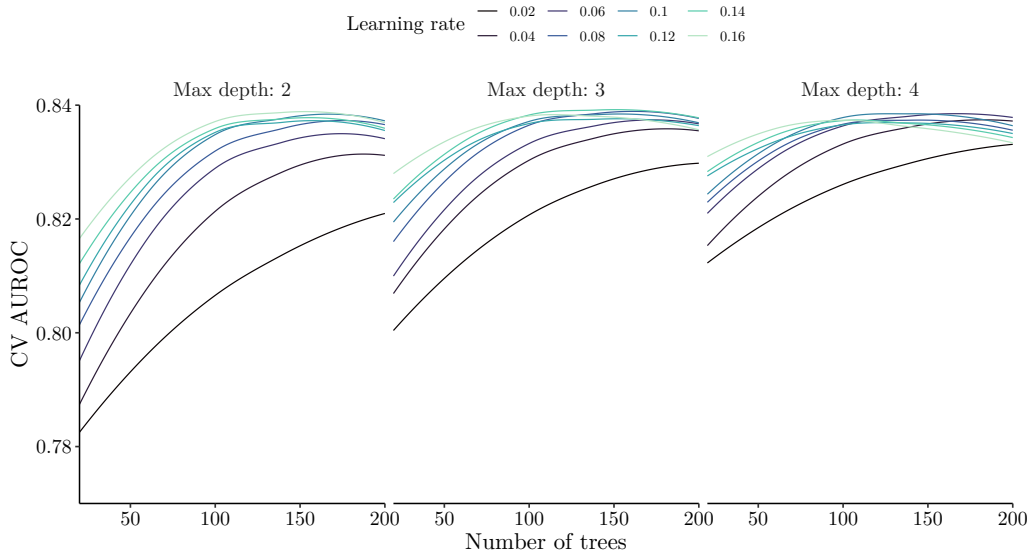


Figure C.13: Model training: Gradient-boosted tree model. Cross-validation AUROC values in the development data across a range of tree numbers, tree depths, and learning rates for a demonstration model comprising 23 predictors.

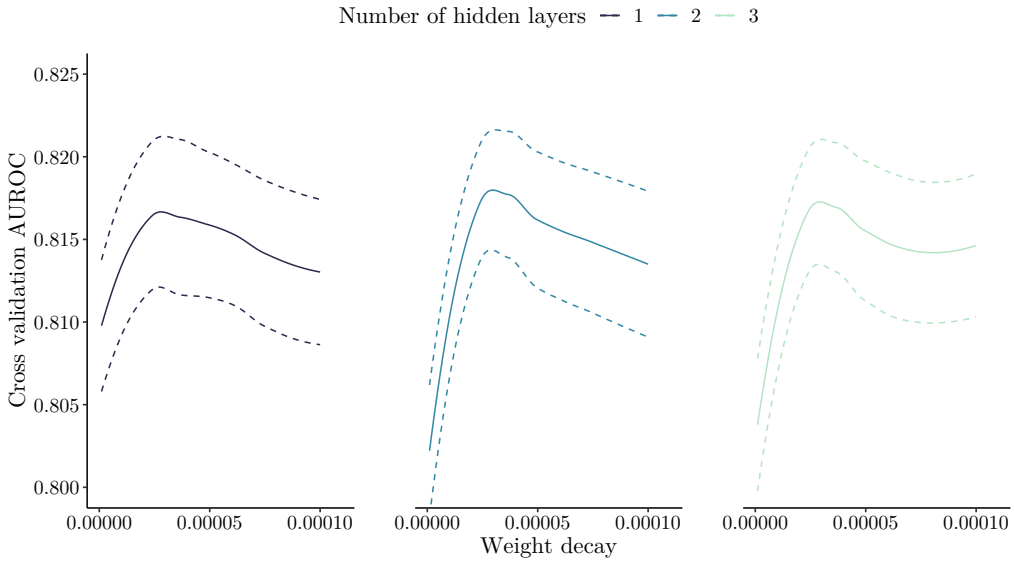


Figure C.14: Cross-validation AUROC values in the development data across different network architecture and weight decay from a demonstration model using 12 predictors. Dashed lines represent 95% CI from resamples.

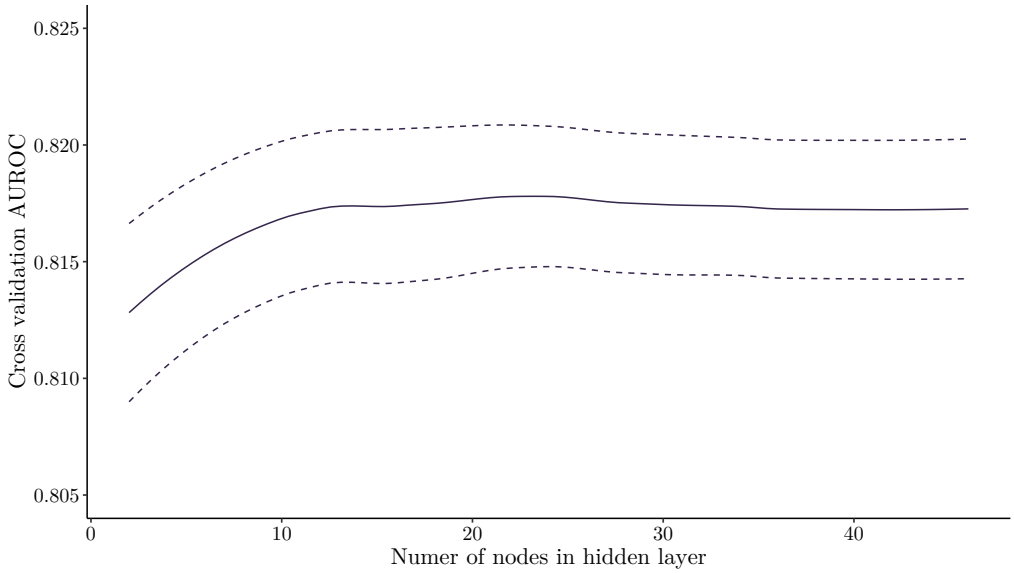


Figure C.15: Model training: Neural network number of nodes per layer. Cross-validation AUROC values in the development data across a range of layer sizes in a single-layer network from a demonstration model using 12 predictors. Dashed lines represent 95% CI from resamples.

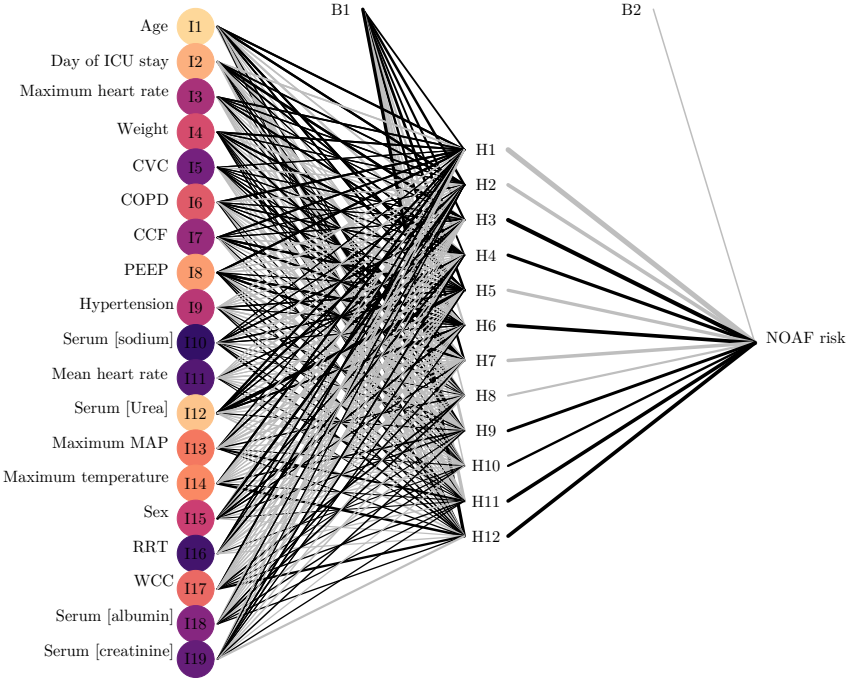


Figure C.16: Visualisation of neural network. I1 – I19 = Each input node; H1 – H12 = each node of the single hidden layer; B1 and B2 = Bias. Weights and biases are visualised by black lines = positive weight, grey line = negative weight. Weight and bias magnitudes are visualised by proportional line thickness (thicker line = higher magnitude). CVC = Central venous catheter; COPD = chronic obstructive pulmonary disease; PEEP = positive end-expiratory pressure; MAP = mean arterial pressure; RRT = renal replacement therapy; WCC = white cell count.

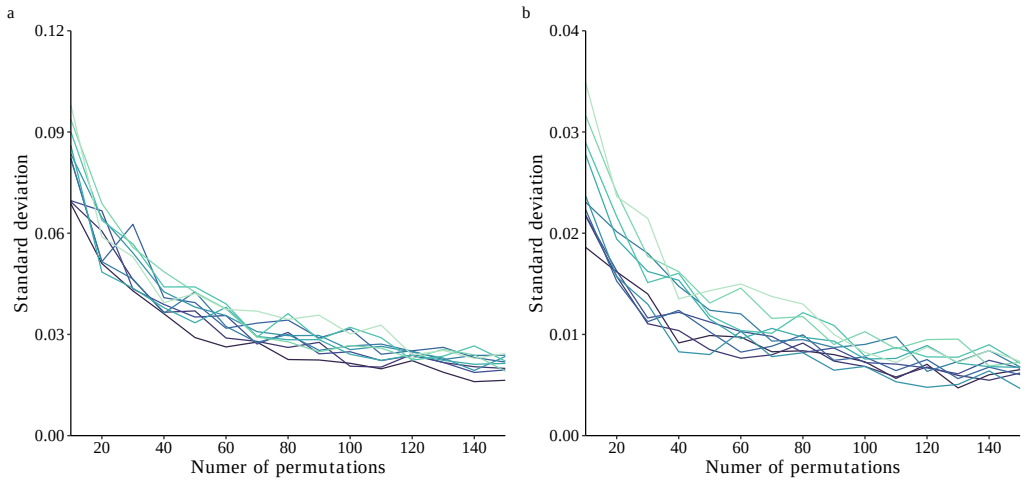


Figure C.17: Standard deviation of Shapley values for a) age and b) FiO_2 over increasing permutations.

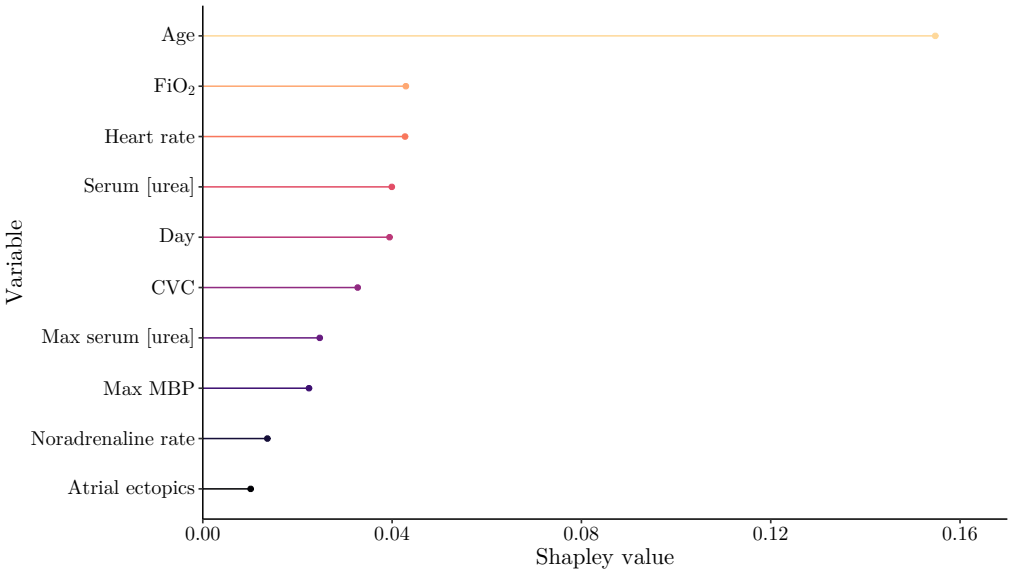


Figure C.18: Mean absolute Shapley values for each variable across all development data.

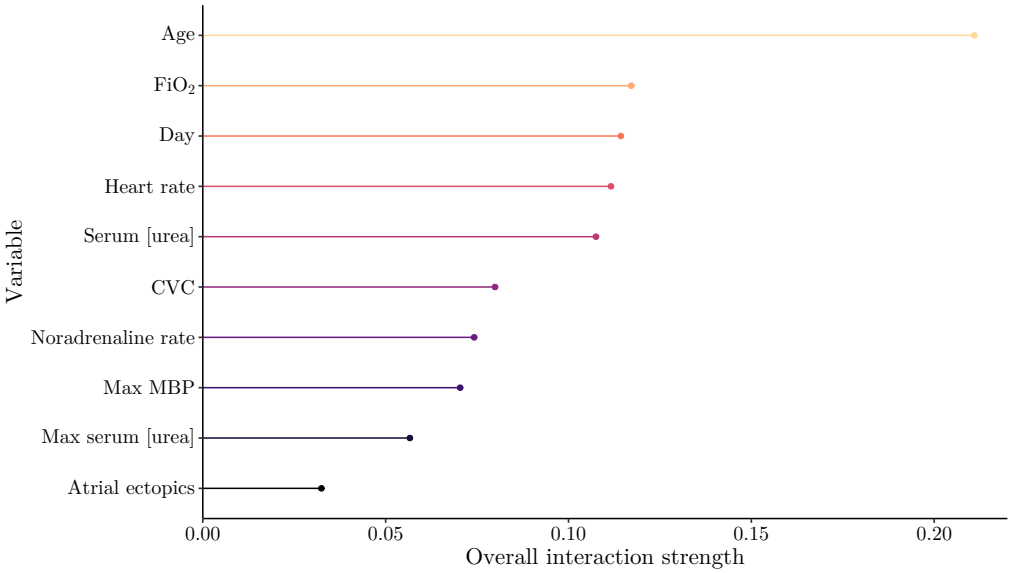


Figure C.19: Interaction strength (H-statistic) for each predictor with all other predictors.

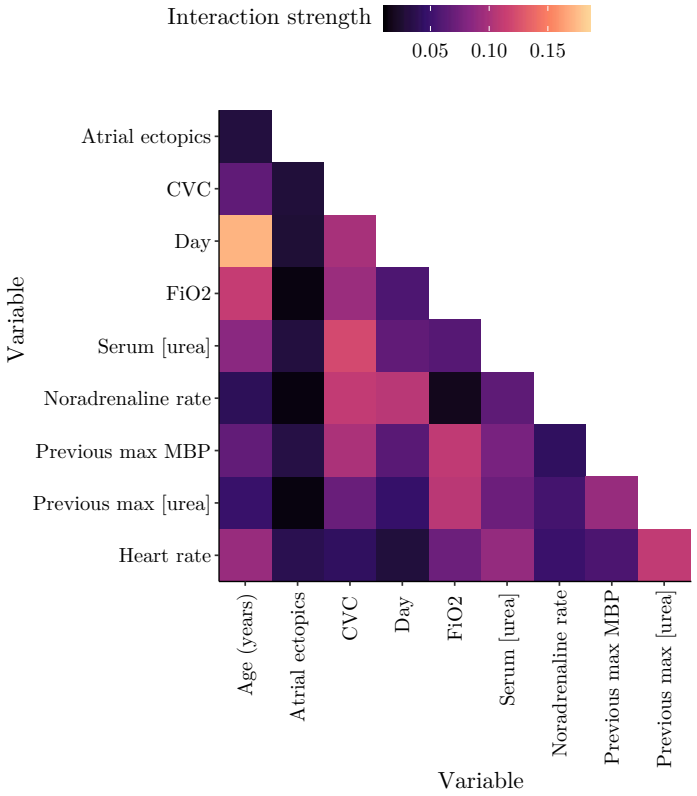


Figure C.20: Interaction strength (H-statistic) of each two-way interaction between each predictor.

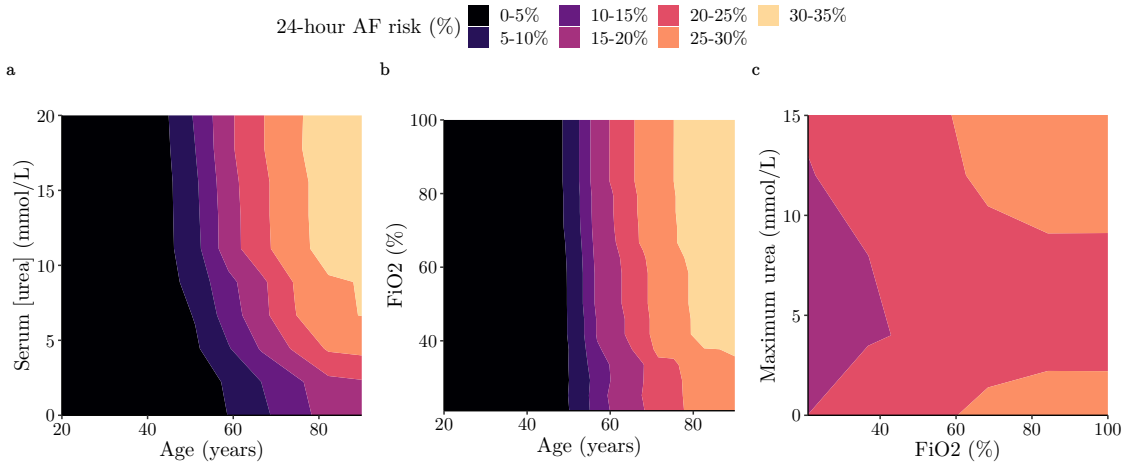


Figure C.21: Examples of two-way interactions in GBT model: a) Age versus serum [urea]; b) age versus FiO_2; and c) FiO_2 versus maximum serum [urea].

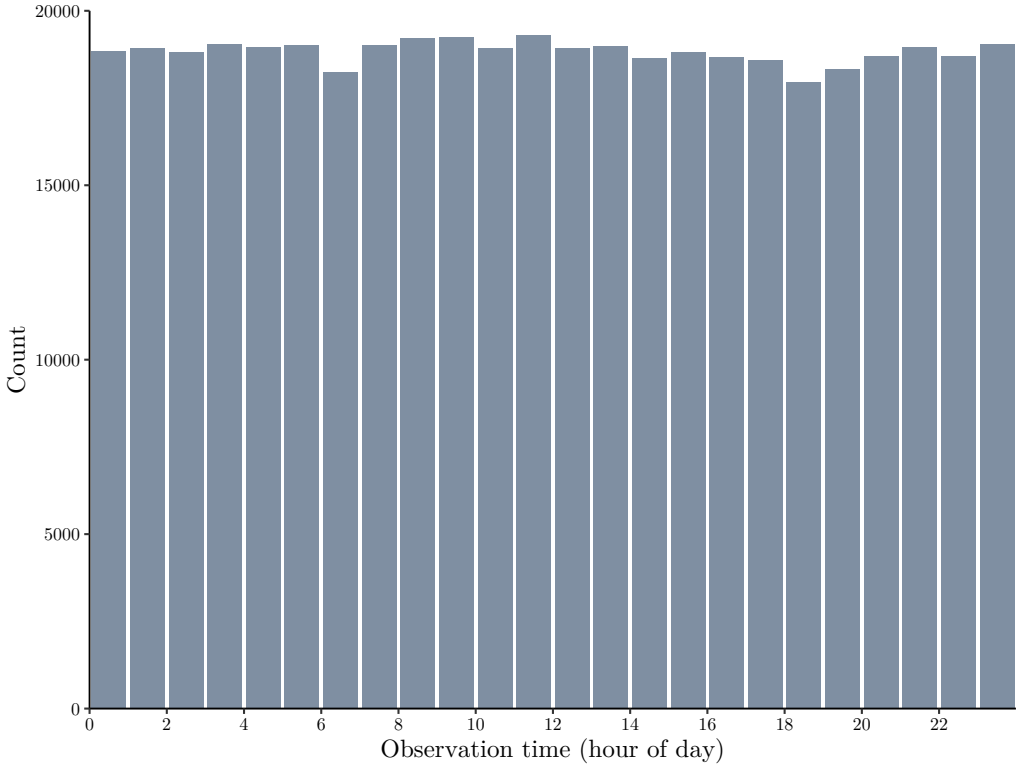


Figure C.22: Frequency of all heart rhythm documentation by time of day.

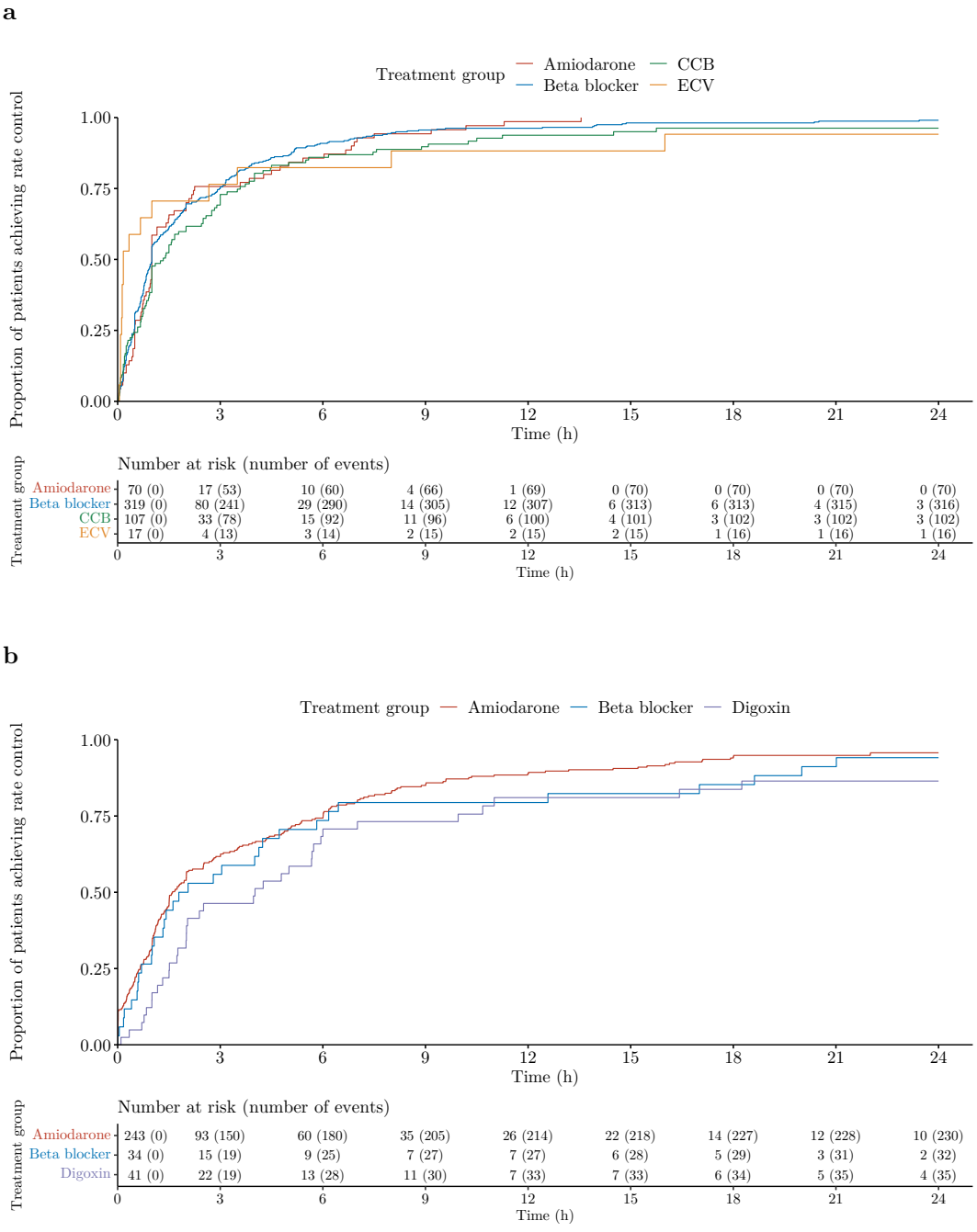


Figure C.23: Kaplan-Meier curves of time to first rate control by treatment group in a) MIMIC database and b) PICRAM database. CCB = Calcium channel blocker, ECV = Electrical cardioversion.

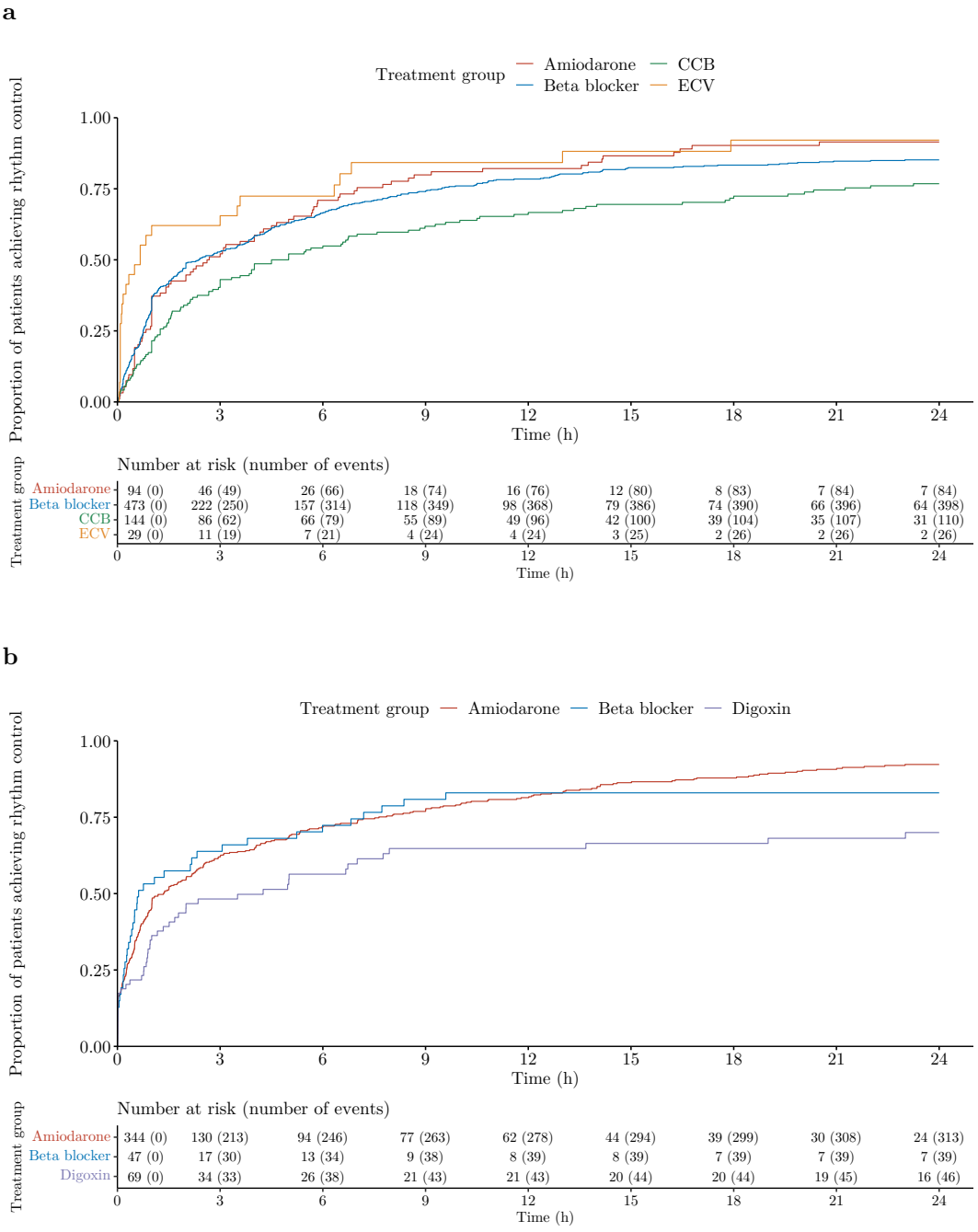
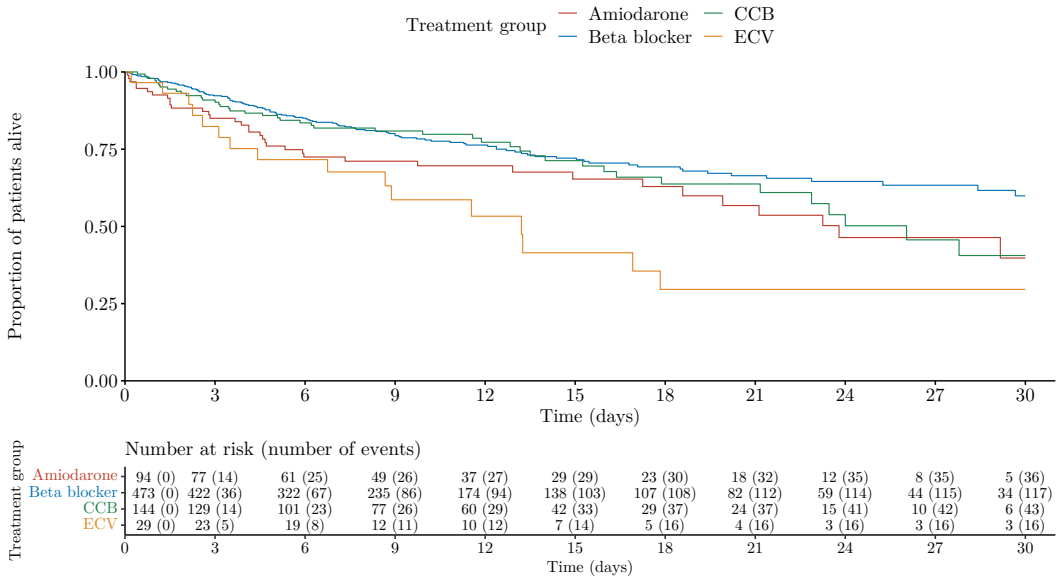


Figure C.24: Kaplan-Meier curves of time to first rhythm control by treatment group in a) MIMIC database and b) PICRAM database. CCB = Calcium channel blocker, ECV = Electrical cardioversion.

a



b

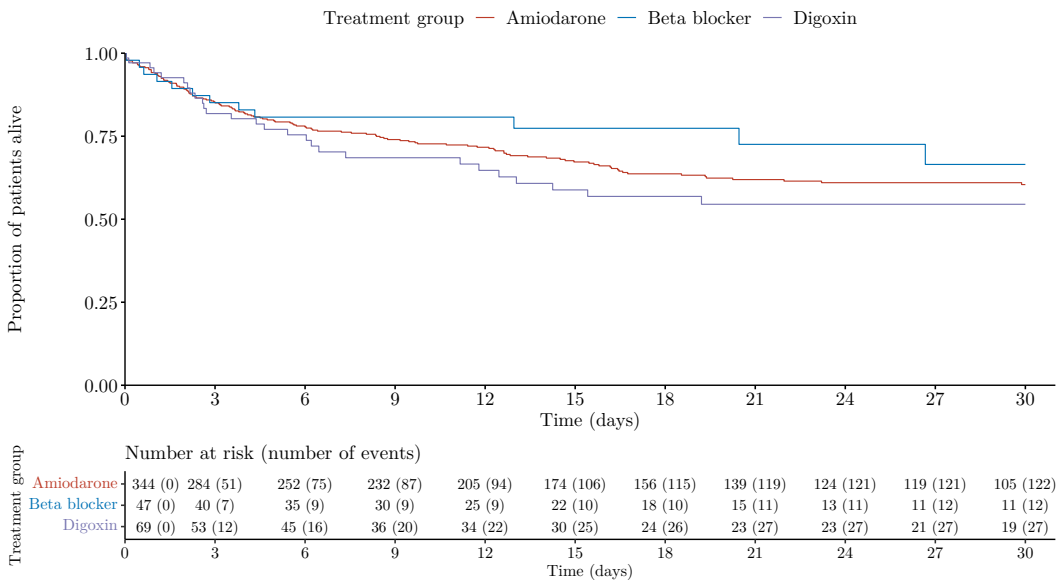


Figure C.25: Kaplan-Meier survival curves by treatment group in a) MIMIC database and b) PICRAM database. CCB = Calcium channel blocker, ECV = Electrical cardioversion.

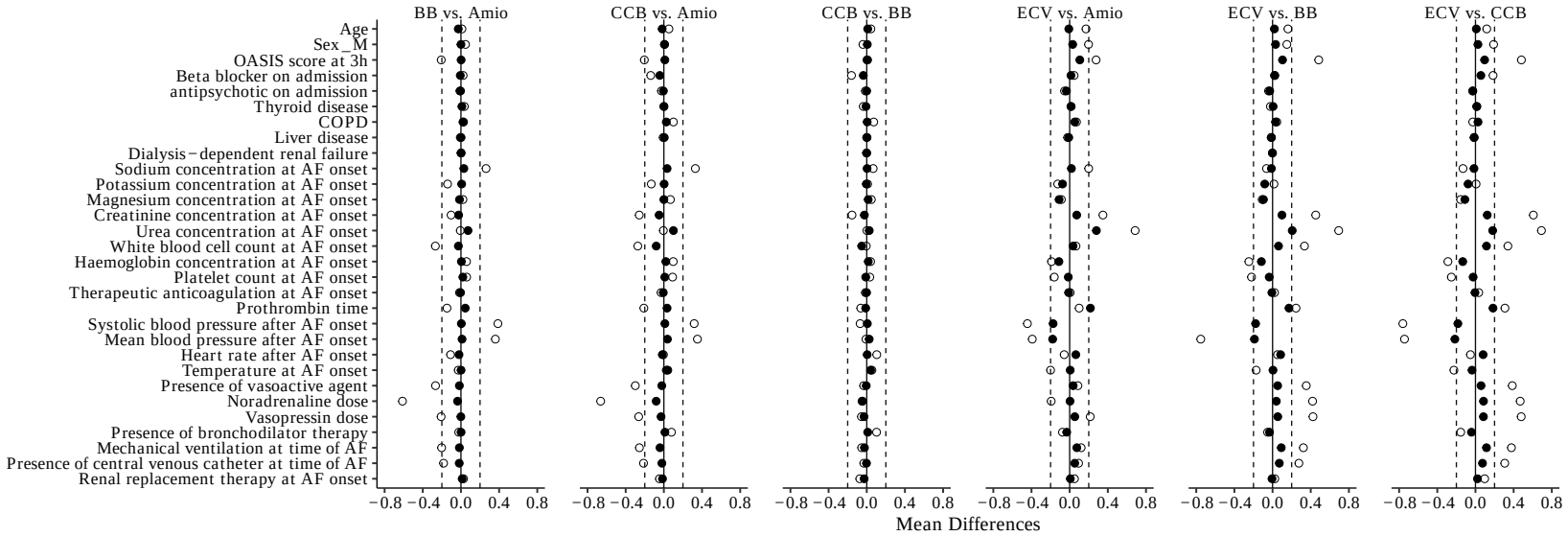


Figure C.26: Standardised mean differences (SMD) for each confounder before and after weighting in the MIMIC-III database. Filled circles = unweighted; open circles = weighted; Dashed vertical lines correspond to SMD = 0.2. AF = atrial fibrillation, COPD = chronic obstructive pulmonary disease, Amio = amiodarone, BB = beta blockers, CCB = calcium channel blocker, ECV = electrical cardioversion.

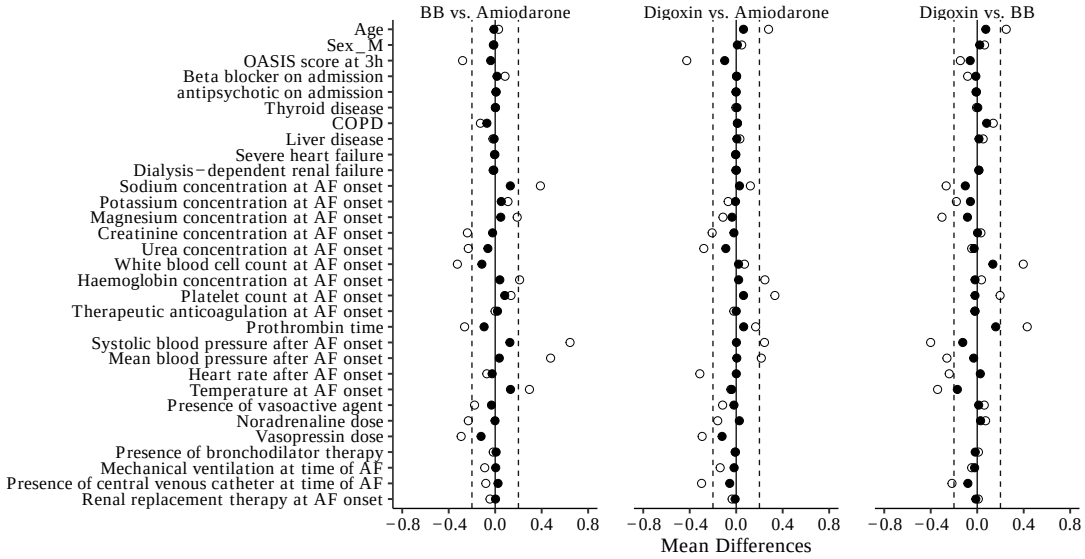


Figure C.27: Standardised mean differences (SMD) for each confounder before and after weighting in the PICRAM database. Filled circles = unweighted; open circles = weighted; Dashed vertical lines correspond to SMD = 0.2. AF = atrial fibrillation, BB = beta blocker, COPD = chronic obstructive pulmonary disease.

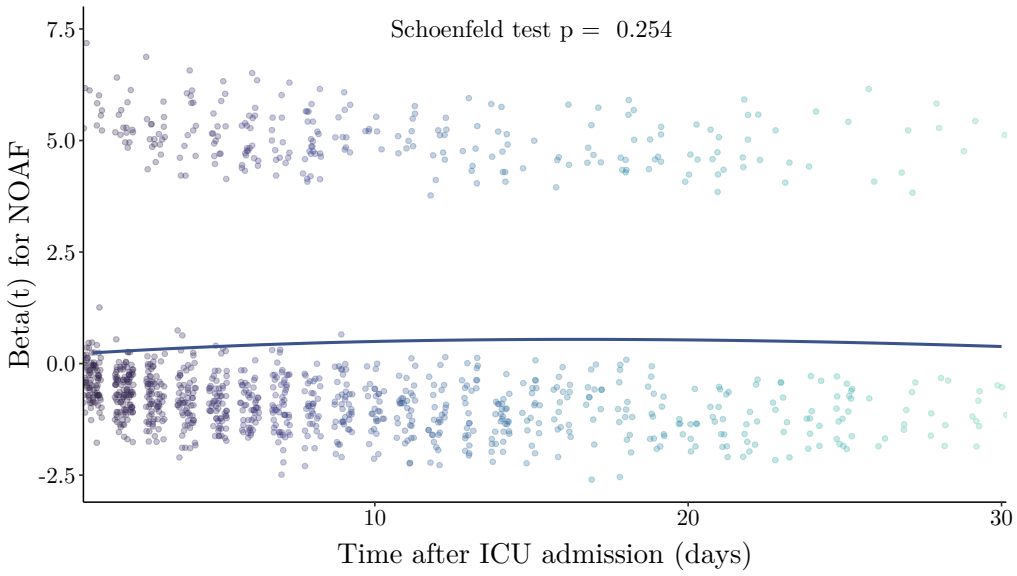


Figure C.28: Schoenfeld residuals of cause-specific Cox model of NOAF mortality association.

Supplementary tables

Table D.1: Ranked demographic variables considered to be important risk factors for NOAF.

Overall rank	Variable	Ranking score	Rank range
1	Age	1	1-1
2	Sex	0.55	2-3
3	Ethnicity	0.46	2-3

Table D.2: Ranked comorbidity variables considered to be important risk factors for NOAF.

Overall rank	Variable	Ranking score	Rank range
1	Atrial dilatation	0.92	1 - 7
2	Valvular heart disease	0.87	1 - 11
3	Congestive cardiac failure	0.78	1 - 20
4	Atrial hypertrophy	0.7	2 - 22
5	Reduced left ventricular ejection fraction	0.67	2 - 22
6	Ischaemic heart disease	0.62	1 - 20
7	Diastolic dysfunction	0.58	2 - 21
8	Pulmonary hypertension	0.57	4 - 21
9	Prior elevated ProBNP	0.56	1 - 21
10	Obstructive sleep apnoea	0.52	2 - 20
11	Hypertension	0.51	2 - 22
12	Previous stroke	0.5	5 - 22
13	Diabetes mellitus	0.47	4 - 17
14	Obesity / High BMI	0.47	1 - 22
15	Chronic obstructive pulmonary disease	0.44	7 - 22
16	Thyroid disease	0.44	6 - 21
17	Chronic kidney disease	0.42	5 - 22

BNP = Brain natriuretic peptide, BMI = body mass index.

Table D.3: Ranked disease / disease severity variables considered to be important risk factors for NOAF.

Overall rank	Variable	Ranking score	Rank range
1	Cardiac surgery	0.9	1 - 11
2	Septic shock	0.85	1 - 11
3	Sepsis	0.76	2 - 17
4	APACHE score	0.73	1 - 13
5	Daily SOFA score	0.67	2 - 14
6	Number of organs in failure	0.62	3 - 15
7	Hospital-acquired pneumonia	0.61	1 - 17
8	Oesophagectomy	0.58	1 - 20
9	Thoracic surgery	0.57	2 - 21
10	Community-acquired pneumonia	0.52	3 - 18
11	Acute coronary syndrome	0.51	4 - 17
12	Thyroid storm	0.5	1 - 18
13	ARDS	0.46	4 - 21
14	Bacteraemia	0.43	6 - 21
15	Endocarditis	0.42	3 - 20

APACHE = Acute physiology and chronic health evaluation, SOFA = sequential organ failure assessment, ARDS = acute respiratory distress syndrome.

Table D.4: Ranked investigation-related variables considered to be important risk factors for NOAF.

Overall rank	Variable	Ranking score	Rank range
1	Hypokalaemia	0.72	1 - 14
2	High left atrial pressure	0.71	1 - 19
3	ECG: Atrial ectopy	0.68	2 - 21
4	High CRP	0.63	2 - 17
5	Low ejection fraction	0.63	1 - 17
6	Atrial ectopic burden	0.63	1 - 20
7	High right atrial pressure	0.62	2 - 19
8	CRP rising	0.6	4 - 17
9	Fluid balance in 6h prior to AF	0.58	1 - 21
10	Rate of change of plasma potassium	0.57	1 - 19
11	Rising proBNP	0.55	2 - 19
12	Acidaemia	0.53	1 - 21
13	Hypomagnesaemia	0.52	3 - 20
14	ECG: P mitrale	0.49	3 - 20
15	Diastolic dysfunction	0.47	2 - 21
16	High WCC	0.44	2 - 21
17	Troponin value	0.42	3 - 21
18	Admission proBNP	0.42	1 - 20
19	Deranged thyroid function	0.4	4 - 21

BNP = Brain natriuretic peptide, CRP = C-reactive protein, ECG = electrocardiogram, WCC = white cell count.

Table D.5: Ranked intervention-related variables considered to be important risk factors for NOAF.

Overall rank	Variable	Ranking score	Rank range
1	Catecholamine administration	0.94	1 - 6
2	Aminophylline administration	0.66	1 - 10
3	Salbutamol administration	0.66	2 - 10
4	Vasopressin administration	0.62	2 - 14
5	Beta blocker withdrawal	0.61	1 - 15
6	Pulmonary artery catheter	0.56	2 - 13
7	Central venous catheter location	0.53	3 - 14
8	Amiodarone administration	0.52	1 - 16
9	Renal replacement therapy	0.5	2 - 15
10	Beta blocker administration	0.47	2 - 16
11	Endotracheal intubation	0.46	2 - 16
12	Diuretic administration	0.44	4 - 15
13	Mechanical ventilation	0.43	2 - 16
14	Levothyroxine administration	0.43	1 - 16
15	Central administration of electrolyte solution	0.41	3 - 15

Table D.6: Ranked physiological variables considered to be important risk factors for NOAF.

Overall rank	Variable	Ranking score	Rank range
1	Tachycardia	0.8	1 - 7
2	Fever	0.77	1 - 7
3	Heart rate variability	0.68	1 - 7
4	Hypoxia	0.56	2 - 7
5	Systemic hypotension	0.55	1 - 7
6	Increasing weight during admission	0.54	1 - 8

Table D.7: Variables identified through Delphi exercise and representation in prediction model development.

Delphi variable	Variable included during model development	Comment
Age	Age	
Sex	Sex	
Ethnicity	-	Sparse data in certain categories*
Atrial dilatation	-	Data not available
Valvular heart disease	-	Data not available
Congestive cardiac failure	Congestive cardiac failure	
Atrial hypertrophy	-	Data not available
Reduced LVEF	Congestive cardiac failure	
Ischaemic heart disease	Ischaemic heart disease	
Diastolic dysfunction	-	Data not available
Pulmonary hypertension	-	Data not available
Prior elevated ProBNP	-	Data not available
Obstructive sleep apnoea	-	Data not available
Hypertension	Hypertension	
Previous stroke	-	Data not available
Diabetes mellitus	-	Data not available
Obesity / High BMI	Weight	
COPD	COPD	
Thyroid disease	-	Data not available
Chronic kidney disease	-	Data not available
Cardiac surgery	-	Non-cardiac surgery cohort
Septic shock	Serum lactate concentration, MAP	Compound concept
Sepsis	Serum lactate concentration	Compound concept
APACHE score	Age, temperature, MAP, pH, heart rate, serum sodium concentration, serum potassium concentration, serum creatinine concentration, white cell count, haemoglobin concentration, serum albumin concentration, FiO_2	Compound score
Daily SOFA score	FiO_2, MV, platelet count, serum creatinine concentration	Compound score

Table D.7 continued from previous page

Delphi variable	Variable included during model development	Comment
Number of organs in failure	-	Compound score
Hospital-acquired pneumonia	-	Data not available
Oesophagectomy	-	Sparse data
Thoracic surgery	-	Non-thoracic surgery cohort
Community-acquired pneumonia	-	Data not available
Acute coronary syndrome	-	Data not available
Thyroid storm	-	Data not available
ARDS	Peripheral oxygen saturation, PEEP, FiO_2	Compound concept
Bacteraemia	-	Data not available
Endocarditis	-	Data not available
Hypokalaemia	Serum potassium concentration	
High left atrial pressure	-	Data not available
ECG: Atrial ectopy	Presence of atrial ectopic beats	
High CRP	Neutrophil:lymphocyte ratio	Sparse data
Low ejection fraction	Congestive cardiac failure	
Atrial ectopic burden	Presence of atrial ectopic beats	
High right atrial pressure	-	Data not available
CRP rising	Neutrophil:lymphocyte ratio	Sparse data
Fluid balance in 6 h prior to AF	-	Data not available
Rate of change of plasma potassium	Serum potassium concentration	
Rising proBNP	-	Sparse data
Acidaemia	Blood pH	
Hypomagnesaemia	Serum magnesium	
ECG: P mitrale	-	Data not available
Diastolic dysfunction	-	Data not available
High white cell count	White cell count	
Troponin value	-	Sparse data
Admission proBNP	-	Sparse data
Deranged thyroid function	-	Sparse data
Catecholamine administration	Noradrenaline infusion rate	
Aminophylline administration	-	Sparse data
Salbutamol administration	-	Sparse data
Vasopressin administration	Vasopressin use	
Beta blocker withdrawal	Beta blocker administration	
Pulmonary artery catheter	-	Sparse data
Central venous catheter location	Central venous catheter	
Amiodarone administration	-	Sparse data
Renal replacement therapy	Renal replacement therapy, serum creatinine concentration, serum urea concentration	
Beta blocker administration	Beta blocker administration	
Endotracheal intubation	Mechanical ventilation	
Diuretic administration	-	
Mechanical ventilation	Mechanical ventilation	
Levothyroxine administration	-	Sparse data
Central electrolyte administration	Serum electrolyte concentrations	
Tachycardia	Heart rate	
Fever	Temperature	
Heart rate variability	-	Data not available
Hypoxia	Peripheral oxygen saturation, PEEP, FiO_2	
Systemic hypotension	Mean arterial pressure	
Increasing weight during admission	Weight	

**Leading to unstable estimates (see results and discussion). APACHE = Acute Physiology and Chronic Health Evaluation, ARDS = Acute Respiratory Distress Syndrome, BMI = Body Mass Index, COPD = Chronic Obstructive Pulmonary Disease, CRP = C-Reactive Protein, ECG = Electrocardiogram, FiO_2 = Fraction of Inspired Oxygen, LVEF = left ventricular ejection fraction, MAP = Mean Arterial Pressure, MV = Mechanical Ventilation, PEEP = Positive End-Expiratory Pressure, proBNP = Pro B-Type Natriuretic Peptide, SOFA = Sequential Organ Failure Assessment.*

Table D.8: Characteristics of included patients: Combined PICRAM and MIMIC-IV model development dataset.

Characteristic	**Never AF**, N = 36,287	**NOAF**, N = 2,797
Age (years)	61 (48, 73)	74 (65, 82)
Sex		
Female	16,579 (46%)	1,241 (44%)
Male	19,708 (54%)	1,556 (56%)
Ethnicity		
Asian	5,623 (15%)	560 (20%)
Black	3,167 (8.7%)	151 (5.4%)
Hispanic	1,111 (3.1%)	37 (1.3%)
Other	2,137 (5.9%)	119 (4.3%)
Unknown	4,176 (12%)	334 (12%)
White	20,073 (55%)	1,596 (57%)
Weight (Kg)	76 (64, 90)	76 (65, 91)
COPD	6,674 (18%)	737 (26%)
Hypertension	16,359 (45%)	1,675 (60%)
Congestive Cardiac Failure	4,360 (12%)	796 (28%)
Ischaemic heart disease	6,063 (17%)	839 (30%)

Table D.8 continued: Characteristics of included patients: Combined PICRAM and MIMIC-IV model development dataset.

Characteristic	**Never AF**, N = 36,287	**NOAF**, N = 2,797
Admission year		
2008 - 2010	11,810 (33%)	906 (32%)
2011 - 2013	10,971 (30%)	832 (30%)
2014 - 2016	9,504 (26%)	712 (25%)
2017 - 2019	4,002 (11%)	347 (12%)
Invasive mechanical ventilation	10,427 (30%)	1,080 (40%)
Presence of central venous catheter	6,473 (18%)	761 (27%)
Fraction of inspired oxygen	0.41 (0.33, 0.60)	0.50 (0.37, 0.81)
Serum sodium concentration (mmol/L)	138.0 (135.2, 141.0)	137.7 (134.4, 140.3)
Serum potassium concentration (mmol/L)	4.0 (3.7, 4.5)	4.2 (3.8, 4.7)
Serum magnesium concentration (mmol/L)	0.78 (0.70, 0.86)	0.78 (0.70, 0.91)
Serum urea concentration (mmol/L)	6 (4, 10)	10 (6, 16)
Serum creatinine concentration (μmol/L)	80 (62, 115)	106 (71, 177)
Serum albumin concentration (g/L)	31 (25, 36)	28 (23, 33)
White cell count (10^9/L)	11 (8, 15)	12 (8, 16)
Serum C-reactive protein concentration (mg/L)	34 (9, 97)	77 (24, 145)
Haemoglobin concentration (g/L)	111 (95, 126)	106 (92, 122)
Platelet count (10^9/L)	203 (147, 269)	191 (132, 257)
Prothrombin time (s)	14.2 (12.4, 16.2)	15.5 (13.3, 18.8)
Serum lactate concentration (mmol/L)	1.70 (1.10, 2.60)	1.90 (1.30, 3.10)
Blood pH	7.37 (7.31, 7.42)	7.34 (7.27, 7.40)
Neutrophil count (10^9/L)	8 (4, 12)	9 (4, 14)
Lymphocyte count (10^9/L)	0.82 (0.38, 1.36)	0.67 (0.31, 1.15)
Neutrophil to lymphocyte ratio	10 (5, 17)	12 (6, 22)
Peripheral oxygen saturation (%)	96.0 (94.0, 99.0)	95.0 (93.0, 98.0)
Mean arterial pressure (mmHg)	71 (62, 82)	66 (58, 76)
Heart rate (beats/min)	92 (79, 107)	93 (80, 109)
Respiratory rate (breaths/min)	21 (18, 26)	23 (19, 28)
Body temperature (degrees celsius)	36.8 (36.4, 37.2)	36.7 (36.3, 37.1)
Presence of atrial ectopic beats	1,116 (3.1%)	232 (8.3%)
ICU length of stay (days)	1.7 (0.9, 3.3)	4.5 (2.4, 8.8)
ICU mortality	2,312 (6.4%)	550 (20%)
ICU length of stay (days)	8 (4, 14)	12 (7, 21)
Hospital mortality	3,853 (11%)	834 (30%)
Admission ICU		
Oxford AICU	4,357 (12%)	494 (18%)
Oxford Churchill	2,411 (6.6%)	167 (6.0%)
Reading ICU	2,786 (7.7%)	170 (6.1%)
Medical Intensive Care Unit (MICU)	7,835 (22%)	620 (22%)
Medical/Surgical Intensive Care Unit (MICU/SICU)	6,785 (19%)	500 (18%)
Surgical Intensive Care Unit (SICU)	6,612 (18%)	430 (15%)
Trauma SICU (TSICU)	5,501 (15%)	416 (15%)

Statistics presented: median (IQR); n (%). AF = atrial fibrillation, NOAF = new-onset AF, ICU = intensive care unit, AICU = adult ICU.

Table D.9: Characteristics of included patients: CCHIC external validation dataset

Characteristic	Never AF, N = 27,147	New-onset AF, N = 2,686
Age	60 (44, 69)	69 (64, 80)
Sex		
F	11,518 (42%)	927 (35%)
M	15,628 (58%)	1,759 (65%)
Ethnicity		
Asian	1,461 (5.4%)	105 (3.9%)
Black	2,296 (8.5%)	118 (4.4%)
Other	1,404 (5.2%)	80 (3.0%)
Unknown	3,752 (14%)	308 (11%)
White	18,234 (67%)	2,075 (77%)
Weight (Kg)	70 (60, 90)	80 (60, 90)
CCF	65 (0.2%)	13 (0.5%)
Mechanical ventilation	8,135 (30%)	1,353 (50%)
CVC	9,667 (36%)	1,598 (59%)
FiO_2	0.40 (0.28, 0.51)	0.50 (0.40, 0.60)
SpO_2	96.0 (94.0, 99.0)	96.0 (93.0, 98.0)
MAP	72 (64, 83)	66 (60, 75)
Heart rate	88 (76, 102)	88 (76, 103)
Temperature (℃)	36.6 (36.1, 37.1)	36.5 (36.0, 37.0)
Serum sodium concentration (mmol/L)	138.8 (136.2, 141.0)	138.4 (135.4, 140.6)
Serum potassium concentration (mmol/L)	4.2 (3.8, 4.6)	4.5 (4.1, 4.9)
Serum urea concentration (mmol/L)	6 (4, 9)	8 (5, 12)
Serum creatinine concentration (μmol/L)	80 (60, 116)	96 (74, 150)
White cell count (10^9/L)	12 (8, 16)	12 (8, 16)
Serum CRP concentration (mg/L)	46 (10, 101)	78 (29, 168)
Haemoglobin concentration (g/L)	108 (93, 124)	104 (90, 119)
Platelet count (10^9/L)	199 (143, 267)	160 (113, 233)
Blood pH	7.36 (7.32, 7.41)	7.35 (7.30, 7.39)
Serum lactate (mmol/L)	1.50 (1.07, 2.38)	1.71 (1.20, 2.64)
ICU length of stay (days)	2 (1, 5)	5 (3, 9)
ICU mortality	2,123 (7.8%)	527 (20%)
Hospital length of stay (days)	12 (7, 24)	16 (10, 31)
Hospital mortality	3,128 (12%)	681 (26%)

Values expressed as median (IQR) or n (%). M = male; F = female; CCF = Congestive cardiac failure; CVC = central venous catheter; MAP = mean arterial pressure; Hb = Haemoglobin; FiO_2 = fraction of inspired oxygen; SpO_2 = peripheral oxygen saturation; CRP = C-reactive protein; ICU = intensive care unit.

Table D.10: Expert panel member details.

Name	Role	Institution
Professor Peter Watkinson	Associate professor of intensive care medicine	University of Oxford
Dr Jonathan Bedford	Clinical Research Fellow	University of Oxford
Dr Andrew Walden	Consultant in intensive care medicine	Intensive care unit, Royal Berkshire Hospital
Professor Ben O'Brien	Professor of Perioperative Medicine	St Bartholomew's Hospital & Barts Heart Centre
Dr Kim Rajappan	Consultant in cardiology	Oxford University Hospitals NHS Foundation Trust
Ian Taylor	Lay representative	N/A
Cathy Taylor	Lay representative	N/A

Table D.11: Characteristics of included patients by treatment group: MIMIC-III database.

Characteristic	Overall, N = 740	amiodarone, N = 94	Beta blocker, N = 473	Calcium channel blocker, N = 144	DC cardioversion, N = 29
Age	74 (64, 82)	76 (63, 83)	73 (64, 83)	73 (65, 81)	77 (69, 85)
Sex					
F	372 (50%)	51 (54%)	234 (49%)	77 (53%)	10 (34%)
M	368 (50%)	43 (46%)	239 (51%)	67 (47%)	19 (66%)
COPD	53 (7.2%)	3 (3.2%)	28 (5.9%)	19 (13%)	3 (10%)
NYHA class III/IV heart failure	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Dialysis-dependent renal failure	1 (0.1%)	0 (0%)	1 (0.2%)	0 (0%)	0 (0%)
Chronic liver disease	11 (1.5%)	2 (2.1%)	7 (1.5%)	2 (1.4%)	0 (0%)
Thyroid disorder	33 (4.5%)	2 (2.1%)	27 (5.7%)	3 (2.1%)	1 (3.4%)
Beta blocker therapy prior to admission	281 (42%)	38 (46%)	193 (45%)	37 (28%)	13 (52%)
Antipsychotic medication prior to admission	27 (4.0%)	4 (4.8%)	19 (4.4%)	4 (3.1%)	0 (0%)
Highest OASIS at three hours	36 (31, 41)	38 (32, 43)	36 (31, 40)	36 (30, 40)	40 (33, 43)
Mechanical ventilation at time of NOAF	343 (46%)	60 (64%)	206 (44%)	55 (38%)	22 (76%)
Renal replacement therapy during or 12h prior to NOAF	47 (6.4%)	5 (5.3%)	38 (8.0%)	1 (0.7%)	3 (10%)
IV Vasoactive medication at time of NOAF	101 (14%)	34 (36%)	45 (9.5%)	9 (6.2%)	13 (45%)
Therapeutic anticoagulation at time of NOAF	36 (4.9%)	6 (6.4%)	23 (4.9%)	5 (3.5%)	2 (6.9%)
Central venous catheter at time of NOAF	429 (58%)	69 (73%)	261 (55%)	75 (52%)	24 (83%)
Bronchodilator therapy on day of, or day preceding NOAF	258 (35%)	33 (35%)	155 (33%)	62 (43%)	8 (28%)
Plasma sodium concentration (mmol/L)	139 (136, 143)	138 (136, 141)	140 (137, 143)	140 (137, 143)	138 (136, 143)
Plasma potassium concentration (mmol/L)	4.0 (3.7, 4.4)	4.0 (3.7, 4.5)	3.9 (3.6, 4.3)	4.0 (3.7, 4.3)	4.1 (3.7, 4.4)
Plasma magnesium concentration (mmol/L)	0.82 (0.78, 0.95)	0.86 (0.78, 0.95)	0.82 (0.78, 0.95)	0.86 (0.78, 0.95)	0.82 (0.78, 0.91)
Plasma urea concentration (mmol/L)	9.3 (6.1, 16.1)	10.0 (6.4, 15.4)	8.9 (5.7, 15.7)	8.9 (6.1, 15.4)	19.8 (9.5, 23.8)
Plasma creatinine concentration (μmol/L)	97 (62, 159)	97 (71, 168)	97 (62, 159)	88 (62, 139)	159 (86, 270)
White cell count (x109 / L)	11.8 (8.5, 16.1)	13.6 (8.8, 19.7)	11.6 (8.8, 15.5)	11.3 (7.5, 16.3)	13.1 (10.5, 16.3)
Haemoglobin concentration (g/L)	102 (92, 115)	100 (88, 113)	104 (92, 115)	101 (93, 116)	99 (91, 111)
Platelet count (x109 / L)	190 (123, 283)	179 (91, 258)	190 (129, 286)	205 (137, 291)	161 (111, 219)
Prothrombin time (s)	14.2 (13.1, 16.4)	15.2 (13.7, 17.8	14.2 (13.1, 16.3)	14.0 (12.9, 15.6)	15.0 (13.6, 17.5)
Systolic blood pressure after AF onset	116 (100, 135)	103 (93, 122)	119 (104, 140)	115 (97, 132)	93 (88, 111)
Mean blood pressure after AF onset	78 (67, 90)	72 (63, 81)	80 (69, 92)	76 (66, 88)	67 (61, 78)
Heart rate after AF onset	122 (104, 137)	124 (110, 139)	121 (102, 136)	124 (110, 141)	123 (98, 147)

Values expressed as median (IQR) or n (%). AF = Atrial Fibrillation; COPD = Chronic Obstructive Pulmonary Disease; NYHA = New York Heart Association; OASIS = Oxford Acute Severity of Illness Score; NOAF = New-Onset Atrial Fibrillation; IV = Intravenous.

Table D.12: Characteristics of included patients by treatment group: PICRAM database.

Characteristic	Overall, N = 460	Amiodarone, N = 344	Beta blocker, N = 47	Digoxin, N = 69
Age	70 (63, 77)	69 (63, 77)	70 (64, 76)	75 (65, 81)
Sex				
F	186 (40%)	141 (41%)	20 (43%)	25 (36%)
M	274 (60%)	203 (59%)	27 (57%)	44 (64%)
COPD	63 (14%)	51 (15%)	1 (2.1%)	11 (16%)
NYHA class III/IV heart failure	2 (0.4%)	2 (0.6%)	0 (0%)	0 (0%)
Dialysis-dependent renal failure	7 (1.5%)	6 (1.7%)	0 (0%)	1 (1.4%)
Chronic liver disease	20 (4.3%)	14 (4.1%)	1 (2.1%)	5 (7.2%)
Thyroid disorder	28 (6.1%)	21 (6.1%)	3 (6.4%)	4 (5.8%)
Beta blocker therapy prior to admission	63 (14%)	44 (13%)	10 (21%)	9 (13%)
Antipsychotic medication prior to admission	7 (1.5%)	5 (1.5%)	1 (2.1%)	1 (1.4%)
Highest OASIS at three hours	34 (26, 39)	34 (27, 40)	34 (22, 38)	30 (25, 36)
Mechanical ventilation at time of NOAF	243 (53%)	192 (56%)	22 (47%)	29 (42%)
Renal replacement therapy during or 12h prior to NOAF	65 (14%)	52 (15%)	5 (11%)	8 (12%)
IV Vasoactive medication at time of NOAF	124 (27%)	105 (31%)	6 (13%)	13 (19%)
Therapeutic anticoagulation at time of NOAF	48 (10%)	37 (11%)	5 (11%)	6 (8.7%)
Central venous catheter at time of NOAF	326 (71%)	262 (76%)	32 (68%)	32 (46%)
Bronchodilator therapy on day of, or day preceding NOAF	75 (16%)	57 (17%)	7 (15%)	11 (16%)
Plasma sodium concentration (mmol/L)	137 (134, 141)	137 (134, 141)	139 (136, 144)	138 (135, 140)
Plasma potassium concentration (mmol/L)	4.2 (3.9, 4.5)	4.2 (3.9, 4.5)	4.1 (4.0, 4.6)	4.2 (3.9, 4.4)
Plasma magnesium concentration (mmol/L)	0.96 (0.84, 1.12)	0.95 (0.84, 1.14)	1.01 (0.92, 1.16)	0.92 (0.82, 1.08)
Plasma urea concentration (mmol/L)	13.6 (8.8, 19.5)	13.8 (9.5, 20.1)	12.1 (7.8, 17.7)	13.2 (8.2, 18.5)
Plasma creatinine concentration (μmol/L)	125 (78, 214)	134 (78, 224)	108 (70, 151)	112 (84, 185)
White cell count (x109 / L)	11.1 (7.7, 16.3)	11.1 (7.5, 16.2)	10.6 (7.6, 13.4)	12.0 (9.5, 16.8)
Haemoglobin concentration (g/L)	98 (88, 113)	97 (87, 111)	103 (94, 112)	101 (90, 116)
Platelet count (x109 / L)	166 (109, 234)	163 (105, 231)	192 (118, 247)	180 (136, 237)
Prothrombin time (s)	16.1 (15.0, 19.0)	16.2 (15.0, 19.0)	15.5 (14.4, 17.0)	16.6 (15.0, 19.4)
Systolic blood pressure after AF onset	116 (99, 131)	112 (97, 128)	128 (108, 156)	119 (103, 135)
Mean blood pressure after AF onset	75 (67, 86)	74 (67, 85)	83 (73, 94)	78 (70, 88)
Heart rate after AF onset	127 (107, 147)	128 (107, 149)	125 (110, 146)	120 (98, 140)

Values expressed as median (IQR) or n (%). AF = Atrial Fibrillation; COPD = Chronic Obstructive Pulmonary Disease; NYHA = New York Heart Association; OASIS = Oxford Acute Severity of Illness Score; NOAF = New-Onset Atrial Fibrillation; IV = Intravenous.

Table D.13: Unadjusted and adjusted hazard ratios compared with amiodarone for each outcome: MIMIC-III database.

Outcome	Treatment group	Unadjusted	Adjusted
Rate control	Beta blocker	1.03 (0.81 – 1.30)	1.09 (0.78 – 1.51)
	Calcium channel blocker	0.83 (0.62 – 1.12)	0.81 (0.55 – 1.19)
	Cardioversion	1.01 (0.39 – 2.62)	1.59 (0.44 – 5.75)
Rhythm control	Beta blocker	0.91 (0.73 – 1.12)	0.91 (0.61 – 1.35)
	Calcium channel blocker	**0.65 (0.50 – 0.84)**	**0.59 (0.37 – 0.92)**
	Cardioversion	1.45 (0.82 – 2.57)	2.00 (0.86 – 4.65)
Reversion to AF	Beta blocker	1.18 (0.79 – 1.78)	1.37 (0.67 – 2.78)
	Calcium channel blocker	1.44 (0.90 – 2.31)	1.73 (0.78 – 3.84)
	Cardioversion	1.80 (0.83 – 3.90)	1.01 (0.28 – 3.71)
Reversion to heart rate* $\geq$ 110	Beta blocker	1.08 (0.80 – 1.46)	0.95 (0.59 – 1.52)
	Calcium channel blocker	1.44 (1.00 – 2.07)	1.61 (0.93 – 2.79)
	Cardioversion	0.67 (0.30 – 1.53)	0.93 (0.36 – 2.42)
Hospital mortality	Beta blocker	**0.64 (0.44 – 0.93)**	1.03 (0.53 – 2.03)
	Calcium channel blocker	0.77 (0.50 – 1.20)	1.30 (0.61 – 2.76)
	Cardioversion	1.56 (0.86 – 2.83)	0.96 (0.31 – 3.01)

*Values expressed as hazard ratio (95% CI). AF = atrial fibrillation, *beats per minute. Significant associations in bold.*

Table D.14: Unadjusted and adjusted hazard ratios compared with amiodarone for each outcome: PICRAM database.

Outcome	Treatment group	Unadjusted HR	Adjusted HR
Rate control	Beta blocker	0.85 (0.57 – 1.27)	0.82 (0.48 – 1.42)
	Digoxin	**0.64 (0.45 – 0.92)**	**0.56 (0.34 – 0.92)**
Rhythm control	Beta blocker	0.95 (0.64 – 1.40)	0.99 (0.57 – 1.72)
	Digoxin	**0.57 (0.41 – 0.81)**	0.67 (0.41 – 1.09)
Reversion to AF	Beta blocker	0.79 (0.50 – 1.27)	0.84 (0.42 – 1.65)
	Digoxin	1.21 (0.78 – 1.89)	1.32 (0.71 – 2.47)
Reversion to heart rate* $\geq$ 110	Beta blocker	0.94 (0.58 – 1.52)	0.88 (0.43 – 1.79)
	Digoxin	1.41 (0.91 – 2.19)	1.14 (0.63 – 2.09)
Hospital mortality	Beta blocker	0.74 (0.40 – 1.38)	0.75 (0.30 – 1.86)
	Digoxin	1.21 (0.79 – 1.86)	1.40 (0.77 – 2.56)

*Values expressed as hazard ratio (95% CI). AF = atrial fibrillation, *beats per minute. Significant associations in bold.*

Table D.15: Unadjusted and adjusted hazard ratios compared with amiodarone for each outcome: Combined databases.

Outcome	Treatment group	Unadjusted results	Adjusted results
Rate control	Beta blocker	**1.32 (1.12 – 1.55)**	1.15 (0.91 – 1.46)
Rhythm control	Beta blocker	**0.81 (0.71 – 0.93)**	0.85 (0.69 – 1.05)
Reversion to AF	Beta blocker	**0.68 (0.55 – 0.85)**	0.75 (0.54 – 1.03)
Reversion to heart rate* $\geq$ 110	Beta blocker	1.16 (0.95 – 1.42)	1.02 (0.76 – 1.37)
Hospital mortality	Beta blocker	**0.77 (0.61 – 0.98)**	1.06 (0.73 – 1.52)

*Values expressed as hazard ratio (95% CI). AF = atrial fibrillation, *beats per minute. Significant associations in bold.*

Table D.16: Sensitivity analysis. Adjusted hazard ratios compared with amiodarone for each outcome: MIMIC-III database.

Outcome	Beta blockers	CCBs	ECV
Hospital mortality	1.09 (0.63 – 1.89)	1.32 (0.70 – 2.49)	1.08 (0.50 – 2.36)
Rate control	1.04 (0.75 – 1.46)	0.73 (0.49 – 1.09)	1.04 (0.34 – 3.19)
Rhythm control	0.90 (0.66 – 1.23)	**0.55 (0.37 – 0.82)**	1.70 (0.84 – 3.41)
Reversion to AF	1.55 (0.86 – 2.81)	1.79 (0.91 – 3.53)	1.60 (0.62 – 4.10)
Reversion to heart rate* $\geq$ 110	1.10 (0.70 – 1.72)	**1.82 (1.10 – 3.00)**	0.55 (0.20 – 1.56)

*Values expressed as hazard ratio (95% CI). AF = Atrial Fibrillation, CCBs = Calcium channel blockers, ECV = electrical cardioversion, *beats per minute. Significant associations in bold.*

Table D.17: Sensitivity analysis. Adjusted hazard ratios compared with amiodarone for each outcome: PICRAM database.

Outcome	Beta blockers	Digoxin
Hospital mortality	0.73 (0.29 - 1.8)	1.33 (0.77 - 2.29)
Rate control	0.88 (0.51 - 1.52)	**0.56 (0.36 - 0.88)**
Rhythm control	1.07 (0.59 - 1.94)	0.66 (0.42 - 1.03)
Reversion to AF	0.86 (0.42 - 1.77)	1.31 (0.74 - 2.31)
Reversion to heart rate* $\geq$ 110	0.98 (0.47 - 2.01)	1.13 (0.65 - 1.96)

Values expressed as hazard ratio (95% CI). AF = Atrial Fibrillation. Significant associations in bold.

Table D.18: Array sensitivity analysis: Digoxin and rate control.

Association	Prevalence in digoxin group	Prevalence in amiodarone group
0.15	0.70	0.30
0.20	0.75	0.30
0.25	0.80	0.30
0.30	0.85	0.30
0.35	0.90	0.30
0.40	0.95	0.30
0.45	1.00	0.30

Examples of minimum required confounder-outcome association and confounder prevalence combinations required for an unmeasured confounder to bring the effect estimate of digoxin on rate control to the null, assuming a prevalence of the unmeasured confounder of 30% in the reference (amiodarone) group.

Table D.19: Array sensitivity analysis: CCBs and rhythm control.

Association	Prevalence in CCB group	Prevalence in amiodarone group
0.15	0.65	0.30
0.20	0.70	0.30
0.25	0.75	0.30
0.30	0.80	0.30
0.35	0.85	0.30
0.40	0.90	0.30
0.45	0.95	0.30
0.50	1.00	0.30

Examples of minimum required confounder-outcome association and confounder prevalence combinations required for an unmeasured confounder to bring the effect estimate of CCBs on rhythm control to the null, assuming a prevalence of the unmeasured confounder of 30% in the reference (amiodarone) group. CCB = calcium channel blocker.

Table D.20: Hospital outcomes analysis: Comparison of characteristics between patients with and without missing outcome data.

Variable	Not missing	Missing	p-value
N	11757	531	
Admission age	64 (49 , 74)	62 (46 , 74)	0.17
Male sex	6910 (58.8)	504 (56.0)	0.32
OASIS	33 (24 , 42)	33 (25 , 40)	0.56

Values expressed as Median (IQR). OASIS = Oxford Acute Severity of Illness Score

Table D.21: Sensitivity analysis: Adjustment for confounding variables.

Model	SDHR (95% CI)
M1 (primary analysis): Sex, OASIS score and comorbidity-adjusted	2.04 (1.79–2.32)
M2: OASIS score-adjusted	2.07 (1.82–2.35)
M3: Sex, OASIS score, comorbidity and laboratory data-adjusted	1.94 (1.70–2.21)
M4: Sex, OASIS score, comorbidity, laboratory data and vasopressor use-adjusted	1.85 (1.63–2.11)

SDHR = subdistribution hazard ratio; OASIS = Oxford Acute Severity of Illness Score.

Table D.22: Sensitivity analysis: Alternative cohort definitions by ICU length of stay.

ICU LOS	SDHR (95% CI)
>24 h (primary analysis)	2.04 (1.79–2.32)
>48 h	2.22 (1.93–2.56)
>72 h	2.29 (1.97–2.67)

SDHR = subdistribution hazard ratio; ICU = intensive care unit; LOS = length of stay.

Table D.23: Sensitivity analysis: Alternative definition of NOAF*.

Model	SDHR (95% CI)
M1 (primary analysis): Sex, OASIS score and comorbidity-adjusted	1.97 (1.72 – 2.25)
M2: OASIS score-adjusted	2.00 (1.75 – 2.28)
M3: Sex, OASIS score, comorbidity and laboratory data-adjusted	1.88 (1.64 – 2.15)
M4: Sex, OASIS score, comorbidity, laboratory data and vasopressor use-adjusted	1.79 (1.56 – 2.04)

**Atrial fibrillation lasting for ≥1h or requiring pharmacological therapy or cardioversion. SDHR = subdistribution hazard ratio; OASIS = Oxford Acute Severity of Illness Score.*

Table D.24: Admission model coefficients before and after shrinkage.

Variable	Coefficient	Coefficient after shrinkage
Intercept	-6.1139	-6.0397
Weight (Kg)	0.0060	0.0059
Mechanical ventilation	0.3455	0.3391
Serum [urea]	0.0323	0.0317
Atrial ectopics	0.6705	0.6580
Serum [albumin]	-0.0160	-0.0157
Central venous catheter	0.3094	0.3036
FiO_2	0.0097	0.0095
CCF	0.5060	0.4966
Age	0.0843	0.0827
Age'	-0.0273	-0.0268
Serum [sodium]	-0.0244	-0.0240
Serum [sodium]'	0.0033	0.0032
Heart rate	0.0055	0.0054
Heart rate'	0.0082	0.0080

References

1. Adderley, N. J. *et al.* Prevalence and treatment of atrial fibrillation in UK general practice from 2000 to 2016. *Heart* **105,** 27–33 (2019).
2. Benjamin, E. J. *et al.* Independent risk factors for atrial fibrillation in a population-based cohort: The Framingham Heart Study. *JAMA - Journal of the American Medical Association* **271,** 840–844 (1994).
3. Chung, S. C. *et al.* Atrial fibrillation epidemiology, disparity and healthcare contacts: a population-wide study of 5.6 million individuals. *The Lancet Regional Health - Europe* **7,** 100157 (2021).
4. Magnussen, C. *et al.* Sex differences and similarities in atrial fibrillation epidemiology, risk factors, and mortality in community cohorts: Results from the biomarcare consortium (Biomarker for cardiovascular risk assessment in Europe). *Circulation* **136,** 1588–1597 (2017).
5. Brundel, B. J. *et al.* Atrial fibrillation. *Nature reviews. Disease primers* **2,** 16017 (2022).
6. Jaïs, P. *et al.* A focal source of atrial fibrillation treated by discrete radiofrequency ablation. *Circulation* **95,** 572–576 (1997).
7. Wolf, P. A., Abbott, R. D. & Kannel, W. B. Atrial fibrillation as an independent risk factor for stroke: The framingham study. *Stroke* **22,** 983–988 (1991).
8. Watson, T., Shantsila, E. & Lip, G. Y. Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. *The Lancet* **373,** 155–166 (2009).
9. Zannad, F. & Stoltz, J. F. Blood rheology in arterial hypertension. *Journal of Hypertension* **10,** S69–S78 (1992).
10. Loh, E. *et al.* Ventricular dysfunction and the risk of stroke after myocardial infarction. *New England Journal of Medicine* **336,** 251–257 (1997).
11. Zhan, Y. *et al.* Assessment and management of the left atrial appendage thrombus in patients with nonvalvular atrial fibrillation. *Canadian Journal of Cardiology* **34,** 252–261 (2018).
12. Lurie, A. *et al.* Prevalence of left atrial thrombus in anticoagulated patients with atrial fibrillation. *Journal of the American College of Cardiology* **77,** 2875–2886 (2021).
13. Aguilar, M. I. & Hart, R. Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks. *Cochrane Database of Systematic Reviews* **2005,** CD001927 (2005).
14. Benjamin, E. J. *et al.* Prevention of atrial fibrillation. Report from a national heart, lung, and blood institute workshop. *Circulation* **119,** 606–618 (2009).

15. Benjamin, E. J. *et al.* Impact of atrial fibrillation on the risk of death: The Framingham Heart Study. *Circulation* **98,** 946–952 (1998).
16. Burdett, P. & Lip, G. Y. Atrial fibrillation in the UK: Predicting costs of an emerging epidemic recognizing and forecasting the cost drivers of atrial fibrillation-related costs. *European Heart Journal - Quality of Care and Clinical Outcomes* **8,** 187–194 (2022).
17. Franz, M. R., Jamal, S. M. & Narayan, S. M. The role of action potential alternans in the initiation of atrial fibrillation in humans: A review and future directions. *Europace* **14,** 58–64 (2012).
18. Varró, A. *et al.* Cardiac Transmembrane Ion Channels and Action Potentials: Cellular Physiology and Arrhythmogenic Behavior. *Physiological Reviews* **101,** 1083–1176 (2021).
19. Van Wagoner, D. R. Electrophysiological remodeling in human atrial fibrillation. *PACE - Pacing and Clinical Electrophysiology* **26,** 1572–1575 (2003).
20. Amos, G. J. *et al.* Differences between outward currents of human atrial and subepicardial ventricular myocytes. *Journal of Physiology* **491,** 31–50 (1996).
21. John Wiley and Sons. *Licence number 5377030115223* 2022.
22. Brundel, B. J. *et al.* Induction of heat shock response protects the heart against atrial fibrillation. *Circulation Research* **99,** 1394–1402 (2006).
23. Yan, J. *et al.* C-Jun N-terminal kinase activation contributes to reduced connexin43 and development of atrial arrhythmias. *Cardiovascular Research* **97,** 589–597 (2013).
24. Caballero, R. *et al.* In humans, chronic atrial fibrillation decreases the transient outward current and ultrarapid component of the delayed rectifier current differentially on each atria and increases the slow component of the delayed rectifier current in both. *Journal of the American College of Cardiology* **55,** 2346–2354 (2010).
25. Wang, Z., Fermini, B. & Nattel, S. Delayed rectifier outward current and repolarization in human atrial myocytes. *Circulation Research* **73,** 276–285 (1993).
26. Michowitz, Y. *et al.* Electrophysiological differences between the epicardium and the endocardium of the left atrium. *PACE - Pacing and Clinical Electrophysiology* **34,** 37–46 (2011).
27. Ozimek, M., Żebrowski, J. J. & Baranowski, R. Information flow between heart rhythm, repolarization, and the diastolic interval series for healthy individuals and LQTS1 patients. *Frontiers in Physiology* **12** (2021).
28. Narayan, S. M. *et al.* Repolarization and activation restitution near human pulmonary veins and atrial fibrillation initiation. A mechanism for the initiation of atrial fibrillation by premature beats. *Journal of the American College of Cardiology* **52,** 1222–1230 (2008).
29. Haïssaguerre, M. *et al.* Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *New England Journal of Medicine* **339,** 659–666 (1998).
30. Jaïs, P. *et al.* Catheter ablation versus antiarrhythmic drugs for atrial fibrillation: The A4 study. *Circulation* **118,** 2498–2505 (2008).

31. Kim, B. S. *et al.* Action potential duration restitution kinetics in human atrial fibrillation. *Journal of the American College of Cardiology* **39,** 1329–1336 (2002).

32. *Elsevier Permissions* `https://www.elsevier.com/about/policies/copyright/permissions`.

33. Lau, D. H. *et al.* Novel mechanisms in the pathogenesis of atrial fibrillation: Practical applications. *European Heart Journal* **37,** 1573–1581 (2016).

34. Allessie, M. A. *et al.* Electropathological substrate of long-standing persistent atrial fibrillation in patients with structural heart disease longitudinal dissociation. *Circulation: Arrhythmia and Electrophysiology* **3,** 606–615 (2010).

35. Moe, G. K. On the multiple wavelet hypothesis of atrial fibrillation. *Arch Int Pharmacodyn Ther* **140,** 183–188 (1962).

36. Carrick, R. T. *et al.* Competitive drivers of atrial fibrillation: The interplay between focal drivers and multiwavelet reentry. *Frontiers in Physiology* **12,** 292 (2021).

37. Hansen, B. J. *et al.* Maintenance of atrial fibrillation: are reentrant drivers with spatial stability the key? *Circulation: Arrhythmia and Electrophysiology* **9** (2016).

38. Bedford, J. P. *et al.* Risk factors for new-onset atrial fibrillation on the general adult ICU: A systematic review. *Journal of Critical Care* **53,** 169–175 (2019).

39. Wetterslev, M. *et al.* New-onset atrial fibrillation in adult critically ill patients: a scoping review. *Intensive Care Medicine* **45,** 928–938 (2019).

40. Kuipers, S., Klouwenberg, K. M. & Cremer, O. L. Incidence, risk factors and outcomes of new-onset atrial fibrillation in patients with sepsis: A systematic review. *Critical Care* **18,** 688 (2014).

41. Kanji, S. *et al.* Epidemiology and management of atrial fibrillation in medical and noncardiac surgical adult intensive care unit patients. *Journal of Critical Care* **27,** 1–326 (2012).

42. Champion, S. *et al.* CHADS2 and CHA2DS2-VASc scores can predict thromboembolic events after supraventricular arrhythmia in the critically ill patients. *Journal of Critical Care* **29,** 854–858 (2014).

43. Clayton, B. *et al.* Risk of thromboembolism in patients developing critical illness-associated atrial fibrillation. *Clinical Medicine, Journal of the Royal College of Physicians of London* **18,** 282–287 (2018).

44. Moss, T. J. *et al.* New-onset atrial fibrillation in the critically ill. *Critical Care Medicine* **45,** 790–797 (2017).

45. Walkey, A. J. *et al.* Incident stroke and mortality associated with new-onset atrial fibrillation in patients hospitalized with severe sepsis. *JAMA - Journal of the American Medical Association* **306,** 2248–2255 (2011).

46. Fernando, S. M. *et al.* New-onset atrial fibrillation and associated outcomes and resource use among critically ill adults - A multicenter retrospective cohort study. *Critical Care* **24,** 15 (2020).

47. Bosch, N. A., Cimini, J. & Walkey, A. J. Atrial Fibrillation in the ICU. *Chest* **154,** 1424–1434 (2018).

48. Bedford, J. P., Ede, J. & Watkinson, P. J. Triggers for new-onset atrial fibrillation in critically ill patients. *Intensive & Critical Care Nursing* **67** (2021).

49. *Green Open Access: SAGE's Archiving and Sharing Policy* `https://uk.sagepub.com/en-gb/eur/journal-author-archiving-policies-and-re-use`.

50. Seguin, P. *et al.* Atrial fibrillation in trauma patients requiring intensive care. *Intensive Care Medicine* **32,** 398–404 (2006).

51. Seguin, P. *et al.* Incidence and risk factors of atrial fibrillation in a surgical intensive care unit. *Critical Care Medicine* **32,** 722–726 (2004).

52. Klein Klouwenberg, P. M. C. *et al.* Incidence, predictors, and outcomes of new-onset atrial fibrillation in critically ill patients with sepsis a cohort study. *American Journal of Respiratory and Critical Care Medicine* **195,** 205–211 (2017).

53. Arsenault, K. A. *et al.* Interventions for preventing post-operative atrial fibrillation in patients undergoing heart surgery. *Cochrane Database of Systematic Reviews* **2013,** CD003611 (2013).

54. Seguin, P. & Launey, Y. Atrial fibrillation is not just an artefact in the ICU. *Critical Care* **14,** 182 (2010).

55. Yoshida, T. *et al.* Epidemiology, prevention, and treatment of new-onset atrial fibrillation in critically ill: A systematic review. *Journal of Intensive Care* **3,** 19 (2015).

56. Carrera, P. *et al.* Epidemiology and outcome of new-onset atrial fibrillation in the medical intensive care unit. *Journal of Critical Care* **36,** 102–106 (2016).

57. Duarte, P. A. *et al.* Factors associated with the incidence and severity of new-onset atrial fibrillation in adult critically ill patients. *Critical Care Research and Practice* **2017,** 1–8 (2017).

58. January, C. T. *et al.* 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: A report of the American college of Cardiology/American heart association task force on practice guidelines and the heart rhythm society. *Journal of the American College of Cardiology* **64,** e1–e76 (2014).

59. Kirchhof, P. *et al.* 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *European Heart Journal* **37,** 2893–2962 (2016).

60. The National Institute for Health and Care Excellence. *Atrial fibrillation: diagnosis and management* 2021. `https://www.nice.org.uk/guidance/ng196`.

61. Moskowitz, A. *et al.* Management of atrial fibrillation with rapid ventricular response in the intensive care unit: A secondary analysis of electronic health record data. *Shock* **48,** 436–440 (2017).

62. Darwish, O. S. *et al.* Challenges of anticoagulation for atrial fibrillation in patients with severe sepsis. *Annals of Pharmacotherapy* **47,** 1266–1271 (2013).

63. Walkey, A. J., Hogarth, D. K. & Lip, G. Y. Optimizing atrial fibrillation management from ICU and beyond. *Chest* **148,** 859–864 (2015).

64. Walkey, A. J. *et al.* Long-term outcomes following development of new-onset atrial fibrillation during sepsis. *Chest* **146,** 1187–1195 (2014).

65. Delle Karth, G. *et al.* Amiodarone versus diltiazem for rate control in critically ill patients with atrial tachyarrhythmias. *Critical Care Medicine* **29,** 1149–1153 (2001).

66. Balser, J. R. *et al.* B-adrenergic blockade accelerates conversion of postoperative supraventricular tachyarrhythmias. *Anesthesiology* **89,** 1052–1059 (1998).

67. Mieure, K. D. *et al.* A comparison of intravenous rate control agents for new onset atrial fibrillation with rapid ventricular rate. *32nd Annual Scientific Sessions of the Heart Rhythm Society, Heart Rhythm. San Francisco, CA United States.* **8,** S216–S217 (2011).

68. Brown, M. *et al.* Impact and treatment success of new-onset atrial fibrillation with rapid ventricular rate development in the surgical intensive care unit. *Journal of Surgical Research* **229,** 66–75 (2018).

69. Walkey, A. J. *et al.* Practice patterns and outcomes of treatments for atrial fibrillation during sepsis a propensity-matched cohort study. *Chest* **149,** 74–83 (2016).

70. Chean, C. S. *et al.* Current practice in the management of new-onset atrial fibrillation in critically ill patients: A UK-wide survey. *PeerJ* **2017,** e3716 (2017).

71. Labbé, V. *et al.* A survey on the management of new onset atrial fibrillation in critically ill patients with septic shock. *Journal of Critical Care* **61,** 18–20 (2021).

72. Samet, P., Bernstein, W. & Levine, S. Significance of the atrial contribution to ventricular filling. *The American Journal of Cardiology* **15,** 195–202 (1965).

73. Pozzoli, M. *et al.* Predictors of primary atrial fibrillation and concomitant clinical and hemodynamic changes in patients with chronic heart failure: a prospective study in 344 patients with baseline sinus rhythm. *Journal of the American College of Cardiology* **32,** 197–204 (1998).

74. Yoshida, T. *et al.* The impact of sustained new-onset atrial fibrillation on mortality and stroke incidence in critically ill patients: A retrospective cohort study. *Journal of Critical Care* **44,** 267–272 (2018).

75. Bedford, J. P. *et al.* Risk factors for new-onset atrial fibrillation during critical illness: A Delphi study. *Journal of the Intensive Care Society* **23,** 414–424 (2022).

76. Bedford, J. P. *et al.* Pharmacological and non-pharmacological treatments and outcomes for new-onset atrial fibrillation in ICU patients: The CAFE scoping review and database analyses. *Health Technology Assessment* **25,** 1–174 (2021).

77. Bedford, J. P. *et al.* Comparative effectiveness of common treatments for new-onset atrial fibrillation within the ICU: Accounting for physiological status. *Journal of Critical Care* **67,** 149–156 (2022).

78. Bedford, J. P. *et al.* Circadian variation in new-onset atrial fibrillation in patients in ICUs. *Journal of Critical Care* **67,** 1–2 (2022).

79. Bedford, J. P. *et al.* Hospital outcomes associated with new-onset atrial fibrillation during ICU admission: A multicentre competing risks analysis. *Journal of Critical Care* **60,** 72–78 (2020).

80. Johnson, A. E. *et al. Physiological trajectory of patients pre and post ICU discharge* in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014* (IEEE, 2014), 3160–3163.

81. Johnson, A. E. *et al.* MIMIC-III, a freely accessible critical care database. *Scientific Data* **3,** 160035 (2016).

82. Johnson, A. *et al.* MIMIC-IV v1.0. *Physionet* (2021).

83. Harris, S. *et al.* Critical Care Health Informatics Collaborative (CCHIC): Data, tools and methods for reproducible research: A multi-centre UK intensive care database. *International Journal of Medical Informatics* **112,** 82–89 (2018).

84. Bzdok, D., Altman, N. & Krzywinski, M. Points of significance: Statistics versus machine learning. *Nature Methods* **15,** 233–234 (2018).

85. Bzdok, D. Classical statistics and statistical learning in imaging neuroscience. *Frontiers in Neuroscience* **11** (2017).

86. Gutierrez, G. Artificial Intelligence in the Intensive Care Unit. *Critical Care* **24** (2020).

87. Holmgren, G. *et al.* Artificial neural networks improve and simplify intensive care mortality prognostication: A national cohort study of 217,289 first-time intensive care unit admissions. *Journal of Intensive Care* **7** (2019).

88. Li, K. *et al.* Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree. *Medicine* **100,** E25813 (2021).

89. Austin, P. C. & Steyerberg, E. W. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in Medicine* **33,** 517–535 (2014).

90. Stevens, R. J. & Poppe, K. K. Validation of clinical prediction models: what does the "calibration slope" really measure? *Journal of Clinical Epidemiology* **118,** 93–99 (2020).

91. Steyerberg, E. W. & Vergouwe, Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal* **35,** 1925–1931 (2014).

92. Van Calster, B. *et al.* Calibration: The Achilles heel of predictive analytics. *BMC Medicine* **17,** 1–7 (2019).

93. Vickers, A. J., van Calster, B. & Steyerberg, E. W. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research* **3,** 1–8 (2019).

94. Capogrosso, P. & Vickers, A. J. A systematic review of the literature demonstrates some errors in the use of decision curve analysis but generally correct interpretation of findings. *Medical Decision Making* **39,** 493–498 (2019).

95. Brookhart, M. A. *et al.* Propensity score methods for confounding control in nonexperimental research. *Circulation: Cardiovascular Quality and Outcomes* **6,** 604–611 (2013).

96. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55 (1983).

97. Székely, G. J. & Rizzo, M. L. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference* **143,** 1249–1272 (2013).

98. Hernán, M. A. The hazards of hazard ratios. *Epidemiology* **21,** 13–15 (2010).

99. Schuster, N. A. *et al.* Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *Journal of Clinical Epidemiology* **122,** 42–48 (2020).

100. Fine, J. P. & Gray, R. J. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94,** 496 (1999).

101. Noordzij, M. *et al.* When do we need competing risks methods for survival analysis in nephrology? *Nephrology Dialysis Transplantation* **28,** 2670–2677 (2013).

102. Austin, P. C. & Fine, J. P. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine* **36,** 4391–4400 (2017).

103. Walkey, A. J. *et al.* Atrial fibrillation among Medicare beneficiaries hospitalized with sepsis: Incidence and risk factors. *American Heart Journal* **165,** 949–955 (2013).

104. Ambrus, D. B. *et al.* Risk factors and outcomes associated with new-onset atrial fibrillation during acute respiratory distress syndrome. *Journal of Critical Care* **30,** 994–997 (2015).

105. Dalkey, N. & Helmer, O. An experimental application of the Delphi method to the use of experts. *Management Science* **9,** 458–467 (1963).

106. Jones, J. & Hunter, D. Qualitative Research: Consensus methods for medical and health services research. *British Medical Journal* **311,** 376 (1995).

107. Mahajan, V., Linstone, H. A. & Turoff, M. The Delphi Method: Techniques and Applications. *Journal of Marketing Research* **13,** 317 (1976).

108. Jünger, S. *et al.* Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine* **31,** 684–706 (2017).

109. Bedford, J. P. *Protocol for "Risk factors for new-onset atrial fibrillation during critical illness: a Delphi study"* 2019. `https://ora.ox.ac.uk/objects/uuid:d7702074-c311-42b4-b1f8-3bfb164dc73d`.

110. Eubank, B. H. *et al.* Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. *BMC Medical Research Methodology* **16,** 56 (2016).

111. Elsevier. *Scopus* `https://www.scopus.com/`.

112. O'Neill, S., Watson, P. J. & Barry, S. A Delphi study of risk factors for Achilles tendinopathy — opinions of world tendon experts. *International Journal of Sports Physical Therapy* **11,** 684–697 (2016).

113. Wang, D. & Li, T. Weighted consensus multi-document summarization. *Information Processing and Management* **48,** 513–523 (2012).

114. Anyukhovsky, E. P. *et al.* Cellular electrophysiologic properties of old canine atria provide a substrate for arrhythmogenesis. *Cardiovascular Research* **54,** 462–469 (2002).

115. Anyukhovsky, E. P. *et al.* Age-associated changes in electrophysiologic remodeling: A potential contributor to initiation of atrial fibrillation. *Cardiovascular Research* **66,** 353–363 (2005).

116. Kalman, J. M., Kumar, S. & Sanders, P. Markers of collagen synthesis, atrial fibrosis, and the mechanisms underlying atrial fibrillation. *Journal of the American College of Cardiology* **60,** 1807–1808 (2012).

117. Lee, G. *et al.* Epicardial wave mapping in human long-lasting persistent atrial fibrillation: Transient rotational circuits, complex wavefronts, and disorganized activity. *European Heart Journal* (2014).

118. Piccini, J. P. *et al.* Differences in clinical and functional outcomes of atrial fibrillation in women and men: Two-year results from the ORBIT-AF Registry. *JAMA Cardiology* **1,** 282–291 (2016).

119. Abhayaratna, W. P. *et al.* Left atrial size. Physiologic determinants and clinical applications. *Journal of the American College of Cardiology* **47,** 2357–2363 (2006).

120. Ambrosi, C. M. *et al.* Gender differences in electrophysiological gene expression in failing and non-failing human hearts. *PLoS ONE* **8** (2013).

121. Bidoggia, H. *et al.* Sex differences on the electrocardiographic pattern of cardiac repolarization: Possible role of testosterone. *American Heart Journal* **140,** 678–683 (2000).

122. Alonso, A. *et al.* Incidence of atrial fibrillation in whites and African-Americans: The Atherosclerosis Risk in Communities (ARIC) study. *American Heart Journal* **158,** 111–117 (2009).

123. O'Neal, W. T. *et al.* Differential impact of risk factors in blacks and whites in the development of atrial fibrillation: the reasons for geographic and racial differences in stroke (REGARDS) study. *Journal of Racial and Ethnic Health Disparities* **4,** 718–724 (2017).

124. Heckbert, S. R. *et al.* Differences by race/ethnicity in the prevalence of clinically detected and monitor-detected atrial fibrillation: MESA. *Circulation: Arrhythmia and Electrophysiology* **13** (2020).

125. Vaziri, S. M. *et al.* Echocardiographic predictors of nonrheumatic atrial fibrillation: The Framingham Heart Study. *Circulation* **89,** 724–730 (1994).

126. Ravelli, F. & Allessie, M. Effects of atrial dilatation on refractory period and vulnerability to atrial fibrillation in the isolated Langendorff-perfused rabbit heart. *Circulation* **96,** 1686–1695 (1997).

127. Bode, F. *et al.* Gadolinium decreases stretch-induced vulnerability to atrial fibrillation. *Circulation* **101,** 2200–2205 (2000).

128. Liu, W. C. *et al.* Prognostic impact of restored sinus rhythm in patients with sepsis and new-onset atrial fibrillation. *Critical Care* **20,** 373 (2016).

129. Shaver, C. M. *et al.* Atrial fibrillation is an independent predictor of mortality in critically ill patients. *Critical Care Medicine* **43,** 2104–2111 (2015).

130. Nakatani, Y. *et al.* Left atrial wall thickness is associated with the low-voltage area in patients with paroxysmal atrial fibrillation. *Journal of Interventional Cardiac Electrophysiology* **58,** 315–321 (2020).

131. Wi, J. *et al.* Complex fractionated atrial electrograms related to left atrial wall thickness. *Journal of Cardiovascular Electrophysiology* **25,** 1141–1149 (2014).

132. López-Candales, A. Is the presence of interatrial septal hypertrophy a marker for atrial fibrillation in the elderly? *American Journal of Geriatric Cardiology* **11,** 399–403 (2002).

133. Whitaker, J. *et al.* The role of myocardial wall thickness in atrial arrhythmogenesis. *Europace* **18,** 1758–1772 (2016).

134. Orban, M. *et al.* Dynamic changes of left ventricular performance and left atrial volume induced by the Mueller maneuver in healthy young adults and implications for obstructive sleep apnea, atrial fibrillation, and heart failure. *American Journal of Cardiology* **102,** 1557–1561 (2008).

135. Linz, D. *et al.* Negative tracheal pressure during obstructive respiratory events promotes atrial fibrillation by vagal activation. *Heart Rhythm* **8,** 1436–1443 (2011).

136. Prystowsky, E. N. *et al.* Enhanced parasympathetic tone shortens atrial refractoriness in man. *The American Journal of Cardiology* **51,** 96–100 (1983).

137. Velez-Roa, S. *et al.* Increased sympathetic nerve activity in pulmonary artery hypertension. *Circulation* **110,** 1308–1312 (2004).

138. Grapsa, J. *et al.* The association of clinical outcome with right atrial and ventricular remodelling in patients with pulmonary arterial hypertension: Study with real-time three-dimensional echocardiography. *European Heart Journal Cardiovascular Imaging* **13,** 666–672 (2012).

139. Arrigo, M. *et al.* New-onset atrial fibrillation in critically ill patients and its association with mortality: A report from the FROG-ICU study. *International Journal of Cardiology* **266,** 95–99 (2018).

140. Goodman, S., Shirov, T. & Weissman, C. Supraventricular arrhythmias in intensive care unit patients: Short and long-term consequences. *Anesthesia and Analgesia* **104,** 880–886 (2007).

141. Chen, M. C. *et al.* Increased inflammatory cell infiltration in the atrial myocardium of patients with atrial fibrillation. *American Journal of Cardiology* **102,** 861–865 (2008).

142. Marcus, G. M. *et al.* Intracardiac and extracardiac markers of inflammation during atrial fibrillation. *Heart Rhythm* **7,** 149–154 (2010).

143. Patel, P. *et al.* Update on the association of inflammation and atrial fibrillation: Clinical review. *Journal of Cardiovascular Electrophysiology* **21,** 1064–1070 (2010).

144. Aviles, R. J. *et al.* Inflammation as a risk factor for atrial fibrillation. *Circulation* **108,** 3006–3010 (2003).

145. Aronson, D. *et al.* Relation of C-reactive protein and new-onset atrial fibrillation in patients with acute myocardial infarction. *American Journal of Cardiology* **100,** 753–757 (2007).

146. Friedrichs, K., Klinke, A. & Baldus, S. Inflammatory pathways underlying atrial fibrillation. *Trends in Molecular Medicine* **17,** 556–563 (2011).

147. Aoki, Y. *et al.* Role of ion channels in sepsis-induced atrial tachyarrhythmias in guinea pigs. *British Journal of Pharmacology* **166,** 390–400 (2012).

148. Karam, B. S. *et al.* Oxidative stress and inflammation as central mediators of atrial fibrillation in obesity and diabetes. *Cardiovascular Diabetology* **16** (2017).

149. Dobrev, D. *et al.* Postoperative atrial fibrillation: mechanisms, manifestations and management. *Nature Reviews Cardiology* **16,** 417–436 (2019).

150. Meierhenrich, R. *et al.* Incidence and prognostic impact of new-onset atrial fibrillation in patients with septic shock: A prospective observational study. *Critical Care* **14,** R108 (2010).

151. Dernellis, J. & Panaretou, M. Relationship between C-reactive protein concentrations during glucocorticoid therapy and recurrent atrial fibrillation. *European Heart Journal* **25,** 1100–1107 (2004).

152. Ho, K. M. & Tan, J. A. Benefits and risks of corticosteroid prophylaxis in adult cardiac surgery a dose-response meta-analysis. *Circulation* **119,** 1853–1866 (2009).

153. Keh, D. *et al.* Effect of hydrocortisone on development of shock among patients with severe sepsis the HYPRESS randomized clinical trial. *JAMA - Journal of the American Medical Association* **316,** 1775–1785 (2016).

154. Venkatesh, B. *et al.* Adjunctive glucocorticoid therapy in patients with septic shock. *New England Journal of Medicine* **378,** 797–808 (2018).

155. Gordon, A. C. *et al.* Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: The VANISH randomized clinical trial. *JAMA - Journal of the American Medical Association* **316,** 509–518 (2016).

156. Launey, Y. *et al.* Impact of low-dose hydrocortisone on the incidence of atrial fibrillation in patients with septic shock: A propensity score-inverse probability of treatment weighting cohort study. *Journal of Intensive Care Medicine* **34,** 238–244 (2019).

157. Shiroshita-Takeshita, A. *et al.* Prednisone prevents atrial fibrillation promotion by atrial tachycardia remodeling in dogs. *Cardiovascular Research* **69,** 865–875 (2006).

158. Watanabe, T. *et al.* C-Reactive protein elevation predicts the occurrence of atrial structural remodeling in patients with paroxysmal atrial fibrillation. *Heart and Vessels* **20,** 45–49 (2005).

159. Kallergis, E. *et al.* The role of the post-cardioversion time course of hs-CRP levels in clarifying the relationship between inflammation and persistence of atrial fibrillation. *Heart* **94,** 200–204 (2008).

160. Burrage, P. S. *et al.* New-onset atrial fibrillation in adult patients after cardiac surgery. *Current Anesthesiology Reports* **9,** 174–193 (2019).

161. O'Brien, B. *et al. Predicting AF after Cardiac Surgery - the PARADISE Score A Clinical Prediction Rule for Post-operative Atrial Fibrillation in Patients Undergoing Cardiac Surgery* `https://fundingawards.nihr.ac.uk/award/NIHR131227`.

162. Von Knorring, J. *et al.* Cardiac arrhythmias and myocardial ischemia after thoracotomy for lung cancer. *The Annals of Thoracic Surgery* **53,** 642–647 (1992).

163. Schomig, A. *et al.* Release of endogenous catecholamines in the ischemic myocardium of the rat. Part A: Locally mediated release. *Circulation Research* (1984).

164. Lee, J. W. Fluid and electrolyte disturbances in critically ill patients. *Electrolyte and Blood Pressure* **8,** 72–81 (2010).

165. Skou, J. C. The influence of some cations on an adenosine triphosphatase from peripheral nerves. *Biochimica et biophysica acta* **23,** 394–401 (1957).

166. Weiss, J. N., Qu, Z. & Shivkumar, K. Electrophysiology of hypokalemia and hyperkalemia. *Circulation: Arrhythmia and Electrophysiology* **10** (2017).

167. Qu, Z. *et al.* Early afterdepolarizations in cardiac myocytes: Beyond reduced repolarization reserve. *Cardiovascular Research* **99,** 6–15 (2013).

168. Burashnikov, A. & Antzelevitch, C. Reinduction of atrial fibrillation immediately after termination of the arrhythmia is mediated by late phase 3 early afterdepolarization-induced triggered activity. *Circulation* **107,** 2355–2360 (2003).

169. Pezhouman, A. *et al.* Molecular basis of hypokalemia-induced ventricular fibrillation. *Circulation* **132,** 1528–1537 (2015).

170. Tazmini, K. *et al.* Hypokalemia promotes arrhythmia by distinct mechanisms in atrial and ventricular myocytes. *Circulation Research* **126,** 889–906 (2020).

171. Krijthe, B. P. *et al.* Serum potassium levels and the risk of atrial fibrillation: The Rotterdam Study. *International Journal of Cardiology* **168,** 5411–5415 (2013).

172. Auer, J. *et al.* Serum potassium level and risk of postoperative atrial fibrillation in patients undergoing cardiac surgery. *Journal of the American College of Cardiology* **44,** 938–939 (2004).

173. Wahr, J. A. *et al.* Preoperative serum potassium levels and perioperative outcomes in cardiac surgery patients. *Journal of the American Medical Association* **281,** 2203–2210 (1999).

174. Antman, E. M. *et al.* ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction; A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology* **44,** E1–E211 (2004).

175. Phillips, C. T. *et al.* Association of hypokalemia with an increased risk for medically treated arrhythmias. *PLoS ONE* **14** (2019).

176. Hessels, L. *et al.* The relationship between serum potassium, potassium variability and in-hospital mortality in critically ill patients and a before-after analysis on the impact of computer-assisted potassium control. *Critical Care* **19,** 4 (2015).

177. Engelhardt, L. J. *et al.* Association between potassium concentrations, variability and supplementation, and in-hospital mortality in ICU patients: a retrospective analysis. *Annals of Intensive Care* **9,** 100 (2019).

178. Elin, R. J. Magnesium: The fifth but forgotten electrolyte. *American Journal of Clinical Pathology* **102,** 616–622 (1994).

179. Hagiwara, S. & Takahashi, K. Surface density of calcium ions and calcium spikes in the barnacle muscle fiber membrane. *The Journal of general physiology* **50,** 583–601 (1967).

180. Maguire, M. E. & Cowan, J. A. Magnesium chemistry and biochemistry. *BioMetals* **15,** 203–210 (2002).

181. Saris, N. E. L. *et al.* Magnesium: An update on physiological, clinical and analytical aspects. *Clinica Chimica Acta* **294,** 1–26 (2000).

182. Jahnen-Dechent, W. & Ketteler, M. Magnesium basics. *CKJ: Clinical Kidney Journal* **5,** 3–14 (2012).

183. Lansman, J. B., Hess, P. & Tsien, R. W. Blockade of current through single calcium channels by Cd2+, Mg2+, and Ca2+s: Voltage and Concentration Dependence of Calcium Entry into the Pore. *Journal of General Physiology* **88,** 321–347 (1986).

184. DiCarlo, L. A. *et al.* Effects of magnesium sulfate on cardiac conduction and refractoriness in humans. *Journal of the American College of Cardiology* **7,** 1356–1362 (1986).

185. Matsuda, H., Saigusa, A. & Irisawa, H. Ohmic conductance through the inwardly rectifying K channel and blocking by internal Mg2+. *Nature* **325,** 156–159 (1987).

186. Khan, A. M. *et al.* Low serum magnesium and the development of atrial fibrillation in the community: The framingham heart study. *Circulation* **127,** 33–38 (2013).

187. Onalan, O. *et al.* Meta-analysis of magnesium therapy for the acute management of rapid atrial fibrillation. *American Journal of Cardiology* **99,** 1726–1732 (2007).

188. Seller, R. H. The role of magnesium in digitalis toxicity. *American Heart Journal* **82,** 551–556 (1971).

189. Nielsen, F. H. *et al.* Dietary magnesium deficiency induces heart rhythm changes, impairs glucose tolerance, and decreases serum cholesterol in post menopausal women. *Journal of the American College of Nutrition* **26,** 121–132 (2007).

190. Reinhart, R. A. *et al.* Myocardial magnesium: Relation to laboratory and clinical variables in patients undergoing cardiac surgery. *Journal of the American College of Cardiology* **17,** 651–656 (1991).

191. Miller, S. *et al.* Effects of magnesium on atrial fibrillation after cardiac surgery: A meta-analysis. *Heart* **91,** 618–623 (2005).

192. Ho, K. M., Sheridan, D. J. & Paterson, T. Use of intravenous magnesium to treat acute onset atrial fibrillation: A meta-analysis. *Heart* **93,** 1433–1440 (2007).

193. Moran, J. L. *et al.* Parenteral magnesium sulfate versus amiodarone in the therapy of atrial tachyarrhythmias: A prospective, randomized study. *Critical Care Medicine* **23,** 1816–1824 (1995).

194. Sleeswijk, M. E. *et al.* Clinical review: Treatment of new-onset atrial fibrillation in medical intensive care patients: A clinical framework. *Critical Care* **11,** 233 (2007).

195. Joseph, T. T. *et al.* Behavioural patterns of electrolyte repletion in intensive care units: lessons from a large electronic dataset. *Scientific Reports* **8,** 1–9 (2018).

196. Homer, L. E. *et al.* Dysmagnesaemia and critical illness outcomes. *Intensive Care Medicine Experimental* **3,** A499 (2015).

197. Workman, A. J. Cardiac adrenergic control and atrial fibrillation. *Naunyn-Schmiedeberg's Archives of Pharmacology* **381,** 235–249 (2010).

198. Barnes, P. J. Theophylline. *American Journal of Respiratory and Critical Care Medicine* **188,** 901–906 (2013).

199. Sharman, A. & Low, J. Vasopressin and its role in critical care. *Continuing Education in Anaesthesia, Critical Care and Pain* **8,** 134–137 (2008).

200. Evans, L. *et al.* Surviving Sepsis Campaign: International guidelines for management of sepsis and septic shock. *Critical Care Medicine* **49,** E1063–E1143 (2021).

201. Dünser, M. W. & Hasibeder, W. R. Sympathetic overstimulation during critical illness: Adverse effects of adrenergic stress. *Journal of Intensive Care Medicine* **24,** 293–316 (2009).

202. McIntyre, W. F. *et al.* Association of vasopressin plus catecholamine vasopressors vs catecholamines alone with atrial fibrillation in patients with distributive shock a systematic review and meta-Analysis. *JAMA - Journal of the American Medical Association* **319,** 1889–1900 (2018).

203. Shen, M. J. & Zipes, D. P. Role of the autonomic nervous system in modulating cardiac arrhythmias. *Circulation Research* **114,** 1004–1021 (2014).

204. Armour, J. A. Functional anatomy of intrathoracic neurons innervating the atria and ventricles. *Heart Rhythm* **7,** 994–996 (2010).

205. Tan, A. Y. *et al.* Autonomic Innervation and Segmental Muscular Disconnections at the Human Pulmonary Vein-Atrial Junction. Implications for Catheter Ablation of Atrial-Pulmonary Vein Junction. *Journal of the American College of Cardiology* **48,** 132–143 (2006).

206. Varró, A. & Baczkó, I. Cardiac ventricular repolarization reserve: A principle for understanding drug-related proarrhythmic risk. *British Journal of Pharmacology* **164,** 14–36 (2011).

207. Roden, D. M. Taking the "idio" out of "idiosyncratic": predicting torsades de pointes. *PACE - Pacing and Clinical Electrophysiology* **21,** 1029–1034 (1998).

208. Braun, A. P., Fedida, D. & Giles, W. R. Activation of α1-adrenoceptors modulates the inwardly rectifying potassium currents of mammalian atrial myocytes. *Pflügers Archiv European Journal of Physiology* **421,** 431–439 (1992).

209. Difrancesco, D. The role of the funny current in pacemaker activity. *Circulation Research* **106,** 434–446 (2010).

210. Amar, D. *et al.* Competing autonomic mechanisms precede the onset of postoperative atrial fibrillation. *Journal of the American College of Cardiology* **42,** 1262–1268 (2003).

211. Crawford, T. *et al.* Clinical value of noninducibility by high-dose isoproterenol versus rapid atrial pacing after catheter ablation of paroxysmal atrial fibrillation. *Journal of Cardiovascular Electrophysiology* **21,** 13–20 (2010).

212. Agarwal, S. K. *et al.* Cardiac autonomic dysfunction and incidence of atrial fibrillation: results from 20 years of follow-up. *Journal of the American College of Cardiology* **69,** 291–299 (2017).

213. Passariello, G. *et al.* Effect of autonomic nervous system dysfunction on sudden death in ischemic patients with anginal syndrome died during electrocardiographic monitoring in Intensive Care Unit. *Minerva Anestesiologica* **73,** 207–212 (2007).

214. Bosch, N. A., Cohen, D. M. & Walkey, A. J. Risk factors for new-onset atrial fibrillation in patients with sepsis: A systematic review and meta-analysis. *Critical Care Medicine* **47,** 280–287 (2019).

215. Harvey, S. *et al.* Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): a randomised controlled trial. *The Lancet* **366,** 472–477 (2005).

216. Rhodes, A. *et al.* A randomised, controlled trial of the pulmonary artery catheter in critically ill patients. *Intensive Care Medicine* **28,** 256–264 (2002).

217. Golamari, R. *et al.* Atrial fibrillation induced by peripherally inserted central catheters. *Baylor University Medical Center Proceedings* **33,** 83–84 (2020).

218. Lee, T. Y. *et al.* Incidence and risk factors of guidewire-induced arrhythmia during internal jugular venous catheterization: Comparison of marked and plain J- wires. *Journal of Clinical Anesthesia* (1996).

219. Blessberger, H. *et al.* Perioperative beta-blockers for preventing surgery-related mortality and morbidity in adults undergoing cardiac surgery. *Cochrane Database of Systematic Reviews* (2019).

220. Khanna, A. K. *et al.* Early resumption of β blockers is associated with decreased atrial fibrillation after noncardiothoracic and nonvascular surgery a cohort analysis. *Anesthesiology* **129,** 1101–1110 (2018).

221. Zebis, L. R. *et al.* Practical regimen for amiodarone use in preventing postoperative atrial fibrillation. *Annals of Thoracic Surgery* **83,** 1326–1331 (2007).

222. Schnabel, R. B. *et al.* Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *The Lancet* **373,** 739–745 (2009).

223. Chamberlain, A. M. *et al.* A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] Study). *American Journal of Cardiology* **107,** 85–91 (2011).

224. Alonso, A. *et al.* Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *Journal of the American Heart Association* **2** (2013).

225. Nadarajah, R. *et al.* Prediction of incident atrial fibrillation in community-based electronic health records: a systematic review with meta-analysis. *Heart* **108,** 1020–1029 (2022).

226. Hill, N. R. *et al.* Predicting atrial fibrillation in primary care using machine learning. *PLoS ONE* **14,** e0224582 (2019).

227. Bosch, N. A. *et al.* New-onset atrial fibrillation as a sepsis-defining organ failure. *Annals of the American Thoracic Society* **16,** 1332–1334 (2019).

228. Rucci, J. M. *et al.* External validation of a risk score for daily prediction of atrial fibrillation among critically ill patients with sepsis. *Annals of the American Thoracic Society* **19,** 697–701 (2022).

229. Ortega-Martorell, S. *et al.* Development of a risk prediction model for new episodes of atrial fibrillation in medical-surgical critically ill patients using the AmsterdamUMCdb. *Frontiers in Cardiovascular Medicine* **9** (2022).

230. Musca, S. *et al.* Routine coagulation testing in intensive care. *Critical Care and Resuscitation* **18,** 213–217 (2016).

231. Li, Z. *et al.* Development and validation of a predictive model for new-onset atrial fibrillation in sepsis based on clinical risk factors. *Frontiers in Cardiovascular Medicine* **9,** 2276 (2022).

232. De Backer, D. *et al.* Comparison of dopamine and norepinephrine in the treatment of shock. *New England Journal of Medicine* **362,** 779–789 (2010).

233. Avni, T. *et al.* Vasopressors for the treatment of septic shock: Systematic review and meta-analysis. *PLoS ONE* **10** (2015).

234. Bedford, J. P. *et al. Statistical analysis plan: Development and validation of prediction models for new-onset atrial fibrillation in patients admitted to an intensive care unit* 2021. `https://ora.ox.ac.uk/objects/uuid:5546f6f3-e19c-4793-9843-7f036a5940a2`.

235. Collins, G. S. *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *British Medical Journal* **350,** g7594 (2015).

236. Hindricks, G. *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *European Heart Journal* (2020).

237. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *British Medical Journal* **332,** 1080 (2006).

238. Zahorec, R. Ratio of neutrophil to lymphocyte counts—rapid and simple parameter of systemic inflammation and stress in critically ill. *Bratislavské lekárske listy* **102,** 5–14 (2001).

239. Guo, Y. *et al.* Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. *BMC Bioinformatics* **11,** 1–19 (2010).

240. Riley, R. D. *et al.* Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine* **38,** 1276–1296 (2019).

241. Ensor, J., Martin, E. & Riley, R. *pmsampsize: Calculates the Minimum Sample Size Required for Developing a Multivariable Prediction Model. R package version 1.0.3.* 2020. `https://cran.r-project.org/package=pmsampsize`.

242. Randall, R. E. *et al.* Hypermagnesemia in renal failure. Etiology and toxic manifestations. *Annals of internal medicine* **61,** 73–88 (1964).

243. Mathiszig-Lee, J. F. *et al.* Highlighting uncertainty in clinical risk prediction using a model of emergency laparotomy mortality risk. *npj Digital Medicine* **5,** 1–8 (2022).

244. Honaker, J., King, G. & Blackwell, M. Amelia II: A program for missing data. *Journal of Statistical Software* **45,** 1–47 (2011).

245. Huque, M. H. *et al.* A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology* **18,** 1–16 (2018).

246. Moons, K. G. *et al.* Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology* **59,** 1092–1101 (2006).

247. Garson, G. D. Interpreting neural-network connection weights. *Artificial Intelligence Expert* **6,** 46–51 (2018).

248. Gevrey, M., Dimopoulos, I. & Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* **160,** 249–264 (2003).

249. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* **29,** 1189–1232 (2001).

250. Jarvis, S. W. *et al.* Are observation selection methods important when comparing early warning score performance? *Resuscitation* **90,** 1–6 (2015).

251. Riley, R. D. *et al.* External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges. *British Medical Journal* **353** (2016).

252. Štrumbelj, E. & Kononenko, I. An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research* **11,** 1–18 (2010).

253. Mitchell, R. *et al.* Sampling permutations for shapley value estimation. *Journal of Machine Learning Research* **23** (2022).

254. Samiei, S. *et al.* Addressing stability in classifier explanations. *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021,* 1920–1927 (2021).

255. Mase, M., Owen, A. B. & Seiler, B. Explaining black box decisions by Shapley cohort refinement. *ArXiv* **ArXIV:1911** (2019).

256. Pepe, M. S., Longton, G. & Janes, H. Estimation and comparison of receiver operating characteristic curves. *Stata Journal* **9,** 1–16 (2009).

257. Chang, W. *et al. shiny: Web Application Framework for R* 2022. `https://shiny.rstudio.com/`.

258. Arrigo, M. *et al.* Cardiovascular events after ICU discharge in patients with new-onset atrial fibrillation: A report from the FROG-ICU study. *International Journal of Cardiology* **270,** 203 (2018).

259. Wilson, M. G. *et al.* Clinician preference instrumental variable analysis of the effectiveness of magnesium supplementation for atrial fibrillation prophylaxis in critical care. *Scientific Reports* **12,** 1–7 (2022).

260. Dewland, T. A. *et al.* Atrial ectopy as a predictor of incident atrial fibrillation: A cohort study. *Annals of Internal Medicine* **159,** 721–728 (2013).

261. Alonso, A. *et al.* Chronic kidney disease is associated with the incidence of atrial fibrillation: The atherosclerosis risk in communities (ARIC) Study. *Circulation* **123,** 2946–2953 (2011).

262. Parikh, N. I. *et al.* Cardiovascular disease risk factors in chronic kidney disease: Overall burden and rates of treatment and control. *Archives of Internal Medicine* **166,** 1884–1891 (2006).

263. Siragy, H. M. & Carey, R. M. Role of the intrarenal renin-angiotensin-aldosterone system in chronic kidney disease. *American Journal of Nephrology* **31,** 541–550 (2010).

264. Li, D. *et al.* Effects of angiotensin-converting enzyme inhibition on the development of the atrial fibrillation substrate in dogs with ventricular tachypacing - Induced congestive heart failure. *Circulation* **104,** 2608–2614 (2001).

265. Jiang, Y. J. *et al.* Risk factors, clinical features and outcome of new-onset acute kidney injury among critically ill patients: a database analysis based on prospective cohort study. *BMC Nephrology* **22,** 1–11 (2021).

266. Tumur, Z. & Niwa, T. Indoxyl sulfate inhibits nitric oxide production and cell viability by inducing oxidative stress in vascular endothelial cells. *American Journal of Nephrology* **29,** 551–557 (2009).

267. Lekawanvijit, S. *et al.* Does indoxyl sulfate, a uraemic toxin, have direct effects on cardiac fibroblasts and myocytes? *European Heart Journal* **31,** 1771–1779 (2010).

268. Chen, W. T. *et al.* The uremic toxin indoxyl sulfate increases pulmonary vein and atrial arrhythmogenesis. *Journal of Cardiovascular Electrophysiology* **26,** 203–210 (2015).

269. Tranchevent, L. C., Azuaje, F. & Rajapakse, J. C. A deep neural network approach to predicting clinical outcomes of neuroblastoma patients. *BMC Medical Genomics* **12** (2019).

270. Yu, H. *et al.* Architectures and accuracy of artificial neural network for disease classification from omics data. *BMC Genomics* **20,** 1–12 (2019).

271. Calude, C. S. & Longo, G. The deluge of spurious correlations in big data. *Foundations of Science* **22,** 595–612 (2017).

272. Ambroise, C. & McLachlan, G. J. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 6562–6566 (2002).

273. Johnson, A. E., Kramer, A. A. & Clifford, G. D. A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. *Critical Care Medicine* **41,** 1711–1718 (2013).

274. Ferrando-Vivas, P. *et al.* Development and validation of the new ICNARC model for prediction of acute hospital mortality in adult critical care. *Journal of Critical Care* **38,** 335–339 (2017).

275. Makrygiannis, S. S. *et al.* Incidence and predictors of new-onset atrial fibrillation in noncardiac intensive care unit patients. *Journal of Critical Care* **29,** 1–697 (2014).

276. Drikite, L. *et al.* Treatment strategies for new onset atrial fibrillation in patients treated on an intensive care unit: a systematic scoping review. *Critical Care* **25,** 1–12 (2021).

277. Walkey, A. J. *et al.* Practice patterns and outcomes associated with use of anticoagulation among patients with atrial fibrillation during sepsis. *JAMA Cardiology* **1,** 682–690 (2016).

278. Mitrić, G. *et al.* Variable use of amiodarone is associated with a greater risk of recurrence of atrial fibrillation in the critically ill. *Critical Care* **20,** 90 (2016).

279. Mayr, A. J. *et al.* High-dosage continuous amiodarone therapy to treat new-onset sapraventricular tachyarrhythmias in surgical intensive care patients: An observational study. *Wiener Klinische Wochenschrift* **116,** 310–317 (2004).

280. Gerlach, A. T. *et al.* Diltiazem versus amiodarone for new-onset atrial arrhythmias in non-cardiac post surgical patients: A cohort study. *Critical Care and Shock* **11,** 94–106 (2008).

281. Matsumoto, Y. *et al.* Comparison of the short acting beta1 blocker landiolol with amiodarone for pharmacological cardioversion for atrial fibrillation in intensive care unit patients. *2015 Annual Meeting of the International Anesthesia Research Society, IARS 2015. Honolulu, HI United States.* **120,** S109 (2015).

282. Balik, M. *et al.* Propafenone for supraventricular arrhythmias in septic shock—Comparison to amiodarone and metoprolol. *Journal of Critical Care* **41,** 16–23 (2017).

283. Rehberg, S. *et al.* Landiolol for managing atrial fibrillation in intensive care. *European Heart Journal, Supplement* **20,** A15–A18 (2018).

284. Sibley, S. & Muscedere, J. New-onset atrial fibrillation in critically ill patients. *Canadian Respiratory Journal* **22,** 179–182 (2015).

285. Trohman, R. G. Atrial fibrillation in the critically ill: Common sense for a common problem. *Critical Care Medicine* **36,** 1681–1682 (2008).

286. Balik, M. *et al.* Management of arrhythmia in sepsis and septic shock. *Anaesthesiology Intensive Therapy* **49,** 419–429 (2017).

287. Keller, M. & Meierhenrich, R. New onset atrial fibrillation in patients with sepsis. *Anaesthesist* **66,** 786–794 (2017).

288. Schwartz, A. *et al.* Cardiac arrhythmias in a septic ICU population: A review. *The Journal of Critical Care Medicine* **1,** 140–146 (2015).

289. Shapiro, W. & Schulze, K. An algorithm for the practical management of the new onset of atrial fibrillation in patients admitted to the ICU. *ICU Director* **3,** 179–184 (2012).

290. Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology* **183,** 758–764 (2016).

291. Von Elm, E. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine* **147,** 573–577 (2007).

292. Palmer, E. *et al.* The association between supraphysiologic arterial oxygen levels and mortality in critically ill patients a multicenter observational cohort study. *American Journal of Respiratory and Critical Care Medicine* **200,** 1373–1380 (2019).

293. Landler, L., Ruxton, G. D. & Malkemper, E. P. Advice on comparing two independent samples of circular data in biology. *Scientific Reports* **11** (2021).

294. Van Gelder, I. C. *et al.* Lenient versus strict rate control in patients with atrial fibrillation. *New England Journal of Medicine* **362,** 1363–1373 (2010).

295. Huling, J. D. & Mak, S. Energy balancing of covariate distributions. *arXiv* **2004** (2020).

296. Noah Greifer. *Package 'WeightIt': Weighting for Covariate Balance in Observational Studies* 2020. `https://ngreifer.github.io/WeightIt/`,.

297. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1,** 54–75 (1986).

298. Campion, W. M. & Rubin, D. B. Multiple imputation for nonresponse in surveys. *Journal of Marketing Research* **26,** 485 (1989).

299. R Core Team. *R: A Language and Environment for Statistical Computing* Vienna, Austria, 2020. `http://www.r-project.org`.

300. Schneeweiss, S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and Drug Safety* **15,** 291–303 (2006).

301. Chyou, J. Y. *et al.* Atrial fibrillation occurring during acute hospitalization: A scientific statement from the American Heart Association. *Circulation* **147,** E676–E698 (2023).

302. Dodt, C. *et al.* Plasma epinephrine and norepinephrine concentrations of healthy humans associated with nighttime sleep and morning arousal. *Hypertension* **30,** 71–76 (1997).

303. Black, N. *et al.* Circadian rhythm of cardiac electrophysiology, arrhythmogenesis, and the underlying mechanisms. *Heart Rhythm* **16,** 298–307 (2019).

304. Viskin, S. *et al.* Circadian variation of symptomatic paroxysmal atrial fibrillation; Data from almost 10,000 episodes. *European Heart Journal* **20,** 1429–1434 (1999).

305. Hjalmarson, A. *et al.* Differing circadian patterns of symptom onset in subgroups of patients with acute myocardial infarction. *Circulation* **80,** 267–275 (1989).

306. Davidson, S. *et al.* Day-to-day progression of vital-sign circadian rhythms in the intensive care unit. *Critical Care* **25,** 156 (2021).

307. Noc, M., Stajer, D. & Horvat, M. Intravenous amiodarone versus verapamil for acute conversion of paroxysmal atrial fibrillation to sinus rhythm. *The American Journal of Cardiology* **65,** 679–680 (1990).

308. Ang, E. L. *et al.* Placebo controlled trial of xamoterol versus digoxin in chronic atrial fibrillation. *British Heart Journal* **64,** 256–260 (1990).

309. Gritsenko, D. *et al.* Amiodarone versus digoxin for rate control in critically ill patients with rapid atrial fibrillation or flutter. *Journal of Emergency and Critical Care Medicine* **2,** 63–63 (2018).

310. DiDomenico, R. J. Should digoxin continue to be used for the management of atrial fibrillation? *Canadian Journal of Hospital Pharmacy* **70,** 391–394 (2017).

311. Ding, E. Y. *et al.* Novel method of atrial fibrillation case identification and burden estimation using the MIMIC-III electronic health data set. *Journal of Intensive Care Medicine* **34,** 851–857 (2019).

312. Vassallo, P. & Trohman, R. G. Prescribing amiodarone: An evidence-based review of clinical indications. *JAMA - Journal of the American Medical Association* **298,** 1312–1322 (2007).

313. Orr, C. F. & Ahlskog, J. E. Frequency, characteristics, and risk factors for amiodarone neurotoxicity. *Archives of Neurology* **66,** 865–869 (2009).

314. Seki, S. *et al.* Amiodarone increases the accumulation of DEA in a human alveolar epithelium-derived cell line. *Biological and Pharmaceutical Bulletin* **31,** 1449–1452 (2008).

315. Gross, S. A. *et al.* Amiodarone and desethylamiodarone toxicity in isolated hepatocytes in culture. *Proceedings of the Society for Experimental Biology and Medicine* **190,** 163–169 (1989).

316. Brien, J. F. *et al.* Distribution of amiodarone and its metabolite, desethylamiodarone, in human tissues. *Canadian Journal of Physiology and Pharmacology* **65,** 360–364 (1987).

317. Camus, P., Colby, T. V. & Rosenow, E. C. Amiodarone pulmonary toxicity. *Drug-Induced and Iatrogenic Respiratory Disease* **93,** 240–259 (2010).

318. Martin, W. J. & Howard, D. M. Amiodarone-induced lung toxicity: In vitro evidence for the direct toxicity of the drug. *American Journal of Pathology* **120,** 344–350 (1985).

319. Martin, W. J. *et al.* Mechanism of phospholipidosis in amiodarone pulmonary toxicity. *Journal of Pharmacology and Experimental Therapeutics* **251,** 272–278 (1989).

320. Nicolescu, A. C. *et al.* Direct mitochondrial dysfunction precedes reactive oxygen species production in amiodarone-induced toxicity in human peripheral lung epithelial HPL1A cells. *Toxicology and Applied Pharmacology* **227,** 370–379 (2008).

321. Akoun, G. M. *et al.* Amiodarone-induced hypersensitivity pneumonitis. Evidence of an immunological cell-mediated mechanism. *Chest* **85,** 133–135 (1984).

322. Fan, K. *et al.* Amiodarone-associated pulmonary fibrosis. Evidence of an immunologically mediated mechanism. *Chest* **92,** 625–630 (1987).

323. Camus, P., Martin, W. J. & Rosenow, E. C. Amiodarone pulmonary toxicity. *Clinics in Chest Medicine* **25,** 65–75 (2004).

324. Connolly, S. J. Effect of prophylactic amiodarone on mortality after acute myocardial infarction and in congestive heart failure: Meta-analysis of individual data from 6500 patients in randomised trials. *The Lancet* **350,** 1417–1424 (1997).

325. Rotmensch, H. H. *et al.* Steady-state serum amiodarone concentrations: Relationships with antiarrhythmic efficacy and toxicity. *Annals of Internal Medicine* **101,** 462–469 (1984).

326. Wolkove, N. Amiodarone pulmonary toxicity. *Canadian Respiratory Journal* **16,** 43–48 (2009).

327. Saussine, M. *et al.* Postoperative acute respiratory distress syndrome; A complication of amiodarone associated with 100 percent oxygen ventilation. *Chest* **102,** 980–981 (1992).

328. Duke, P. K. *et al.* Acute oxygen induced amiodarone pulmonary toxicity after general anesthesia. *Anesthesiology* **75,** A228–A228 (1991).

329. Borders, C. W. *et al.* A rare case of acute diffuse alveolar hemorrhage following initiation of amiodarone: A case report. *Military Medicine* **177,** 118–120 (2012).

330. Lapinsky, S., Mullen, J. B. & Balter, M. S. Accumulation pulmonaire rapide de phospolipides induite par l'amiodarone intraveineux. *Canadian Journal of Cardiology* **9,** 322–324 (1993).

331. Boriani, G. *et al.* Short-term onset of fatal pulmonary toxicity in a patient treated with intravenous amiodarone for post-operative atrial fibrillation. *International Journal of Cardiology* **159,** e1–e4 (2012).

332. Baumann, H. *et al.* Rapid onset of amiodarone induced pulmonary toxicity after lung lobe resection – A case report and review of recent literature. *Annals of Medicine and Surgery* **21,** 53–57 (2017).

333. Handschin, A. E. *et al.* Acute amiodarone-induced pulmonary toxicity following lung resection. *Respiration* **70,** 310–312 (2003).

334. Donaldson, L. *et al.* Acute amiodarone-induced lung toxicity. *Intensive Care Medicine* **24,** 626–630 (1998).

335. Van Mieghem, W. *et al.* Amiodarone and the development of ARDS after lung surgery. *Chest* **105,** 1642–1645 (1994).

336. Fernando, H. C. *et al.* The society of thoracic surgeons practice guideline on the prophylaxis and management of atrial fibrillation associated with general thoracic surgery: Executive summary. *Annals of Thoracic Surgery* **92,** 1144–1152 (2011).

337. Tisdale, J. E. *et al.* A randomized trial evaluating amiodarone for prevention of atrial fibrillation after pulmonary resection. *Annals of Thoracic Surgery* **88,** 886–895 (2009).

338. Lanza, L. A. *et al.* Low-dose oral amiodarone prophylaxis reduces atrial fibrillation after pulmonary resection. *Annals of Thoracic Surgery* **75,** 223–230 (2003).

339. Berry, M. F., D'Amico, T. A. & Onaitis, M. W. Use of amiodarone after major lung resection. *Annals of Thoracic Surgery* **98,** 1199–1206 (2014).

340. Bobbio, A. *et al.* Postoperative outcome of patients undergoing lung resection presenting with new-onset atrial fibrillation managed by amiodarone or diltiazem. *European Journal of Cardio-thoracic Surgery* **31,** 70–74 (2007).

341. Ashrafian, H. & Davey, P. Is amiodarone an underrecognized cause of acute respiratory failure in the ICU? *Chest* **120,** 275–282 (2001).

342. Quinlan, G. J. *et al.* Plasma hypoxanthine levels in ARDS: Implications for oxidative stress, morbidity, and mortality. *American Journal of Respiratory and Critical Care Medicine* **155,** 479–484 (1997).

343. Brady Boyce, B. A. & Yee, B. H. Incidence and severity of phlebitis in patients receiving peripherally infused amiodarone. *Critical Care Nurse* **32,** 27–34 (2012).

344. Russell, S. J. & Saltissi, S. Amiodarone induced skin necrosis. *Heart* **92,** 1395 (2006).

345. Pourbaix, S. *et al.* Absolute bioavailability of amiodarone in normal subjects. *Clinical Pharmacology and Therapeutics* **37,** 118–123 (1985).

346. Latini, R., Tognoni, G. & Kates, R. E. Clinical Pharmacokinetics of Amiodarone. *Clinical Pharmacokinetics* **9,** 136–156 (1984).

347. O'Bryan, L. J. *et al.* Managing new-onset atrial fibrillation in critically ill patients: A systematic narrative review. *BMJ Open* **10,** e034774 (2020).

348. Raine, A. E. & Vaughan Williams, E. M. Adaption to prolonged β-blockade of rabbit atrial, Purkinje, and ventricular potentials, and of papillary muscle contraction. Time-course of development of and recovery from adaption. *Circulation Research* **48,** 804–812 (1981).

349. Kharche, S. R. *et al.* Effects of human atrial ionic remodelling by β-blocker therapy onmechanisms of atrial fibrillation: A computer simulation. *Europace* **16,** 1524–1533 (2014).

350. Workman, A. J., Kane, K. A. & Rankin, A. C. Cellular bases for human atrial fibrillation. *Heart Rhythm* (2008).

351. Bolger, A. P. *et al.* Neurohormonal activation and the chronic heart failure syndrome in adults with congenital heart disease. *Circulation* **106,** 92–99 (2002).

352. Norhayati, M. N. *et al.* Metoprolol for prophylaxis of postoperative atrial fibrillation in cardiac surgery patients: systematic review and meta-analysis. *BMJ open* **10,** e038364 (2020).

353. Poulsen, S. H., Jensen, S. E. & Egstrup, K. Effects of long-term adrenergic β-blockade on left ventricular diastolic filling in patients with acute myocardial infarction. *American Heart Journal* **138,** 710–720 (1999).

354. Doughty, R. N. *et al.* Effects of carvedilol on left ventricular remodeling after acute myocardial infarction: The CAPRICORN Echo Substudy. *Circulation* **109,** 201–206 (2004).

355. DasGupta, P., Broadhurst, P. & Lahiri, A. The effects of intravenous carvedilol, a new multiple action vasodilatory β-blocker, in congestive heart failure. *Journal of Cardiovascular Pharmacology* **18,** S12–S16 (1991).

356. Frencken, J. F. *et al.* Myocardial injury in patients with sepsis and its association with long-term outcome. *Circulation: Cardiovascular Quality and Outcomes* **11** (2018).

357. Lim, W. *et al.* Etiology of troponin elevation in critically ill patients. *Journal of Critical Care* **25,** 322–328 (2010).

358. Morelli, A. *et al.* Microvascular effects of heart rate control with esmolol in patients with septic shock: A pilot study. *Critical Care Medicine* **41,** 2162–2168 (2013).

359. Ibrahim-Zada, I. *et al.* Inhibition of sepsis-induced inflammatory response by β1-adrenergic antagonists. *Journal of Trauma and Acute Care Surgery* **76,** 320–328 (2014).

360. Liu, P. *et al.* The influence of esmolol on septic shock and sepsis: A meta-analysis of randomized controlled studies. *American Journal of Emergency Medicine* **36,** 470–474 (2018).

361. The National Institute for Health and Care Excellence. Atrial fibrillation: diagnosis and management [NICE Guideline No. 196] (2021).

362. Skroubis, T. *et al.* Amiodarone-induced acute lung toxicity in an ICU setting. *Acta Anaesthesiologica Scandinavica* **49,** 569–571 (2005).

363. Leelathanalerk, A. *et al.* Evaluation of deprescribing amiodarone after new-onset atrial fibrillation in critical illness. *American Journal of Medicine* **130,** 864–866 (2017).

364. Namana, V. *et al.* Clinical significance of atrial kick. *QJM* **111,** 569–570 (2018).

365. Carroll, J. D. *et al.* Dynamics of left ventricular filling at rest and during exercise. *Circulation* **68,** 59–67 (1983).

366. Galderisi, M. Diastolic dysfunction and diastolic heart failure: Diagnostic, prognostic and therapeutic aspects. *Cardiovascular Ultrasound* **3** (2005).

367. Louie, E. K. *et al.* 'Stunning' of the left atrium after spontaneous conversion of atrial fibrillation to sinus rhythm: Demonstration by transesophageal Doppler techniques in a canine model. *Journal of the American College of Cardiology* **32,** 2081–2086 (1998).

368. Manning, W. J. *et al.* Impaired left atrial mechanical function after cardioversion: Relation to the duration of atrial fibrillation. *Journal of the American College of Cardiology* **23,** 1535–1540 (1994).

369. Daoud, E. G. *et al.* Effect of an irregular ventricular rhythm on cardiac output. *American Journal of Cardiology* **78,** 1433–1436 (1996).

370. Gosselink, A. T. *et al.* Left ventricular beat-to-beat performance in atrial fibrillation: Contribution of frank-starling mechanism after short rather than long RR intervals. *Journal of the American College of Cardiology* **26,** 1516–1521 (1995).

371. Ellenbogen, K. A., Thames, M. D. & Mohanty, P. K. New insights into pacemaker syndrome gained from hemodynamic, humoral and vascular responses during ventriculo-atrial pacing. *The American Journal of Cardiology* **65,** 53–59 (1990).

372. Noll, B. *et al.* Cardiac natriuretic peptide hormones during artificial cardiac pacemaker stimulation and left heart catheterization. *The Clinical Investigator* **70,** 1057–1060 (1992).

373. Geny, B. *et al.* Atrial natriuretic factor secretion: a role for atrial systolic ejection force? *European journal of applied physiology and occupational physiology* **72,** 440–444 (1996).

374. Artucio, H. & Pereira, M. Cardiac arrhythmias in critically ill patients: Epidemiologic study. *Critical Care Medicine* **18,** 1383–1388 (1990).

375. Brathwaite, D. & Weissman, C. The new onset of atrial arrhythmias following major noncardiothoracic surgery is associated with increased mortality. *Chest* **114,** 462–468 (1998).

376. Knotzer, H. *et al.* Tachyarrhythmias in a surgical intensive care unit: A case-controlled epidemiologic study. *Intensive Care Medicine* **26,** 908–914 (2000).

377. Gupta, S., Tiruvoipati, R. & Green, C. Atrial fibrillation and mortality in critically ill patients: A retrospective study. *American Journal of Critical Care* **24,** 336–341 (2015).

378. Makrygiannis, S. S. *et al.* New-onset atrial fibrillation and clinical outcome in non-cardiac intensive care unit patients. *Australian Critical Care* **31,** 274–277 (2018).

379. Walkey, A. J. & McManus, D. When rhythm changes cause the blues: New-onset atrial fibrillation during sepsis. *American Journal of Respiratory and Critical Care Medicine* **195,** 152–153 (2017).

380. Shintani, A. K. *et al.* Immortal time bias in critical care research: Application of time-varying cox regression for observational cohort studies. *Critical Care Medicine* **37,** 2939–2945 (2009).

381. Ho, A. *et al.* Understanding immortal time bias in observational cohort studies. *Anaesthesia* **68,** 126–130 (2013).

382. Austin, P. C., Lee, D. S. & Fine, J. P. Introduction to the analysis of survival data in the presence of competing risks. *Circulation* **133,** 601–609 (2016).

383. Wolkewitz, M. & Sommer, H. Evaluating mortality in an intensive care unit requires extended survival models. *Critical Care* **18,** 415 (2014).

384. Resche-Rigon, M., Azoulay, E. & Chevret, S. Evaluating mortality in intensive care units: Contribution of competing risks analyses. *Critical Care* **10,** R5 (2005).

385. Gooley, T. A. *et al.* Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Statistics in Medicine* **18,** 695–706 (1999).

386. Wolkewitz, M. *et al.* Risk factors for the development of nosocomial pneumonia and mortality on intensive care units: Application of competing risks models. *Critical Care* **12,** R44 (2008).

387. Fasiolo, M. *et al.* Fast calibrated additive quantile regression. *Journal of the American Statistical Association* (2020).

388. Koponen, T. *et al.* Vasoactive-inotropic score and the prediction of morbidity and mortality after cardiac surgery. *British Journal of Anaesthesia* **122,** 428–436 (2019).

389. Harrison, D. A. *et al.* Ensuring comparisons of health-care providers are fair: development and validation of risk prediction models for critically ill patients. *Health Services and Delivery Research* **3,** 1–132 (2015).

390. Altman, D. G. *et al.* Reporting recommendations for tumor marker prognostic studies (REMARK): Explanation and elaboration. *BMC Medicine* **10,** 51 (2012).

391. Feakins, B. G. *et al.* Standard and competing risk analysis of the effect of albuminuria on cardiovascular and cancer mortality in patients with type 2 diabetes mellitus. *Diagnostic and Prognostic Research* **2** (2018).

392. Bakker, J. *et al.* Blood lactate levels are superior to oxygen-derived variables in predicting outcome in human septic shock. *Chest* **99,** 956–962 (1991).

393. Pirracchio, R. *et al.* Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *The Lancet Respiratory Medicine* **3,** 42–52 (2015).

394. Brivet, F. G. *et al.* Acute renal failure in intensive care units - Causes, outcome, and prognostic factors of hospital mortality: A prospective, multicenter study. *Critical Care Medicine* **24,** 192–198 (1996).

395. Augusto, J. B. *et al.* Predictors of de novo atrial fibrillation in a non-cardiac intensive care unit. *Revista Brasileira de Terapia Intensiva* **30,** 166–173 (2018).

396. Landesberg, G. *et al.* Diastolic dysfunction and mortality in severe sepsis and septic shock. *European Heart Journal* **33,** 895–903 (2012).

397. Lu *et al.* New-onset atrial fibrillation can be falsely associated with increased length of stay in ICU due to immortal time bias. *Critical Care* **24,** 1–3 (2020).

398. Brock, G. N. *et al.* How to handle mortality when investigating length of hospital stay and time to clinical stability. *BMC Medical Research Methodology* **11** (2011).

399. Mattioli, A. V. *et al.* Clinical and echocardiographic features influencing recovery of atrial function after cardioversion of atrial fibrillation. *American Journal of Cardiology* **82,** 1368–1371 (1998).

400. Shapiro, E. P. *et al.* Transient atrial dysfunction after conversion of chronic atrial fibrillation to sinus rhythm. *The American Journal of Cardiology* **62,** 1202–1207 (1988).

401. Sun, H. *et al.* Cellular mechanisms of atrial contractile dysfunction caused by sustained atrial tachycardia. *Circulation* **98,** 719–727 (1998).

402. Sanders, P. *et al.* Reversal of atrial mechanical stunning after cardioversion of atrial arrhythmias: Implications for the mechanisms of tachycardia-mediated atrial cardiomyopathy. *Circulation* **106,** 1806–1813 (2002).

403. Pagonas, N. *et al.* Impact of atrial fibrillation on the accuracy of oscillometric blood pressure monitoring. *Hypertension* **62,** 579–584 (2013).

404. Li, Y. *et al.* Impact of atrial fibrillation on the accuracy of oscillometric blood pressure monitoring in ICU patients from a large real-world database. *Journal of Hypertension* **Publish Ah** (2023).

405. Harrison, D. A. *et al.* A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. *Critical Care Medicine* **35,** 1091–1098 (2007).

406. *Atrial fibrillation: Management. Clinical guideline [CG180]* 2014. `https://www.nice.org.uk/guidance/cg180`.

407. Tonekaboni, S. *et al.* What clinicians want: Contextualizing explainable machine learning for clinical end use. *Proceedings of Machine Learning Research* (2019).

408. Schwartz, J. M. *et al.* Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: Qualitative descriptive study. *JMIR Human Factors* **9** (2022).

409. Bedford, J. P. *et al.* New-onset atrial fibrillation in intensive care: epidemiology and outcomes. *European Heart Journal. Acute Cardiovascular Care* (2022).

410. Harrison, D. A., Brady, A. R. & Rowan, K. Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: The Intensive Care National Audit and Research Centre Case Mix Programme Database. *Critical Care* **9,** 99–111 (2004).

411. Herbert, A. *et al.* Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology* **46,** 1093–1093 (2017).

412. Garside, T. *et al.* Increased long-term mortality following new-onset atrial fibrillation in the intensive care unit: A systematic review and meta-analysis. *Journal of Critical Care* **72** (2022).

413. Fuster, V. *et al.* ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the European Society of Cardiology Committee for practice. *Circulation* **114** (2006).

414. Friberg, L., Rosenqvist, M. & Lip, G. Y. Net clinical benefit of warfarin in patients with atrial fibrillation: A report from the swedish atrial fibrillation cohort study. *Circulation* **125,** 2298–2307 (2012).

415. Granholm, A. *et al.* Predictors of gastrointestinal bleeding in adult ICU patients: a systematic review and meta-analysis. *Intensive Care Medicine* **45,** 1347–1359 (2019).

416. Arnold, D. M. *et al.* Bleeding during critical illness: A prospective cohort study using a new measurement tool. *Clinical and Investigative Medicine* **30** (2007).

417. Slavik, R. *et al.* Patterns of practice for the acute management of new-onset atrial fibrillation in critically-ill intensive care unit patients. *Critical Care Medicine* **31,** A137 (2003).

418. Labbé, V. *et al.* Should we administrate anticoagulants to critically ill patients with new onset supraventricular arrhythmias? *Archives of Cardiovascular Diseases* **108,** 217–219 (2015).

419. Frost, L. *et al.* Incident stroke after discharge from the hospital with a diagnosis of atrial fibrillation. *American Journal of Medicine* **108,** 36–40 (2000).

420. McIntyre, W. F. *et al.* Atrial fibrillation recurrence in patients with transient new-onset trial fibrillation detected during hospitalization for noncardiac surgery or medical illness. *Annals of Internal Medicine* **176,** 1299–1307 (2023).

421. *Tracking Atrial Fibrillation After Intensive Care Admission (TrAFFIC)* `https://clinicaltrials.gov/ct2/show/NCT05229211`.

422. Youden, W. Index for rating diagnostic tests. *Cancer* **3,** 32–35 (1950).

423. Perkins, N. J. & Schisterman, E. F. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology* **163,** 670–675 (2006).

424. Wood, A. M., White, I. R. & Royston, P. How should variable selection be performed with multiply imputed data? *Statistics in Medicine* **27,** 3227–3246 (2008).

425. Eekhout, I., Van De Wiel, M. A. & Heymans, M. W. Methods for significance testing of categorical covariates in logistic regression models after multiple imputation: Power and applicability analysis. *BMC Medical Research Methodology* **17,** 1–12 (2017).

426. Schomaker, M. & Heumann, C. Bootstrap inference when using multiple imputation. *Statistics in Medicine* **37,** 2252–2266 (2018).

427. Damen, J. A. *et al. Prediction models for cardiovascular disease risk in the general population: Systematic review* 2016.