Risk prediction following cardiac surgery

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Biology, Medicine and Health.

2019

Samuel H Howitt

School of Medical Sciences

Division of Cardiovascular Sciences

Contents

74,808 words

Tables

Chapter Eleven

Abbreviations

AF – atrial fibrillation

AKI - Acute kidney injury

AKICS - Acute Kidney Injury after Cardiac Surgery

APACHE - Acute Physiology and Chronic Health Evaluation

API - application programming interface

ASB – assisted spontaneous breathing

AUC - area under the Receiver Operator Characteristic curve

BiPAP – bi-level positive airway pressure

BP – blood pressure

CABG - coronary artery bypass graft

CASUS - Cardiac Surgery Score CICU - Cardiac Intensive Care Unit

CI – confidence interval

CNS – central nervous system

CPB - cardiopulmonary bypass

CRD – Centre for Reviews and Dissemination

CSV – Comma separated values

CT – computed tomography

CVA – cerebrovascular accident

CVP - central venous pressure

CVVH – continuous venovenous haemofiltration

DARE - Database of Abstract of Reviews of Effects

ECG – electrocardiography

ECMO – extracorporeal membrane oxygenation

EPR - electronic patient record

ETT - endotracheal tube

EuroSCORE - European System for Cardiac Operative Risk Evaluation

EWS – Early warning score

F_iO_2 – fraction of inspired oxygen saturations

GCS – Glasgow coma scale

HIS – Hospital information service

HL - Hosmer-Lemeshow

HL7 – Health level seven

HR – Hazards ratio

IABP – intra-aortic balloon pump

ICNARC - Intensive Care National Audit and Research Centre

ICU - Intensive Care Unit

ICURS - Intensive Care Unit Risk Stratification Score

IQR – interquartile range

IT – Information technology

KDIGO - Kidney Disease International Global Outcomes

LODS – Logistic organ dysfunction score

LOS – length of stay

LVAD – Left ventricular assist device

MAP – mean arterial pressure

MCS – mechanical circulatory support

MFT – Manchester University Hospitals NHS Foundation Trust

MODS - Multiple Organ Dysfunction Score

NA – Not applicable

NG – Nasogastric

NHS – National Health Service

O:E – observed to expected

OR – Odds ratio

PAR - pressure adjusted heart rate

PC – pressure controlled

PEEP – Positive end-expiratory pressure

PICOS - Population, Intervention, Comparison, Outcomes, Setting

PLOS – prolonged length of stay

PMV - prolonged mechanical ventilation

PO_2 – partial pressure of oxygen

POCD – postoperative neurocognitive dysfunction

PPI - patient and public involvement

RACE - Rapid Clinical Evaluation

RASS – Richmond agitation and sedation score

ROC – receiver operator characteristic

RRT - renal replacement therapy

SAPS - Simplified Acute Physiology Score

sCr – serum creatinine concentration

sd – standard deviation

SIRS - Systemic Inflammatory Response Syndrome

SMR – standardised mortality rate

SOFA - (Sepsis Related) Sequential Organ Failure Assessment Score

SpO_2 – pulse oximetry oxygen saturations

SR – sinus rhythm

STS - Society for Thoracic Surgery

TOE - transoesophageal echocardiography

TT - tracheostomy tube

TTE - transthoracic echocardiography

UHSM - University Hospital of South Manchester

UK - United Kingdom of Great Britain and Northern Ireland

UO - urine output

VA ECMO – Venoarterial extracorporeal membrane oxygenation

VAD – ventricular assist device

VC – volume controlled

VGNW - Vascular Governance NorthWest

THESIS ABSTRACT
Samuel H Howitt
The University of Manchester
Degree Title: Doctor of Philosophy (PhD)
Thesis Title: Risk prediction following cardiac surgery
March 2019

Objectives: Around 1000 patients die each year in the UK due to complications suffered following cardiac surgery. Early identification of those at risk of specific complications would allow targeted interventions aimed at reducing the harm caused by those complications. However, commonly used risk models only quantify overall mortality risk for groups of patients based on analyses of pre- and intra-operative data. These models do not predict specific complications and are unable to update risk estimates as postoperative events unfold. This thesis aims to advance understanding of postoperative risk prediction following cardiac surgery by assessing the performance of existing risk prediction tools in this population and developing novel risk models.

Methods: Postoperative physiological monitoring data, blood test results, medication administration data and demographics for over 3000 patients were cleaned, analysed using computerised processing algorithms and entered into a comprehensive database. The database was used to validate three mortality models, the Sepsis-3 diagnostic criteria and the KDIGO AKI criteria. A novel dynamic Bayesian model which analyses an individual's urine output to predict their risk of severe oliguria was developed and validated. Finally, the potential usefulness of potassium and magnesium concentrations when predicting atrial fibrillation (AF) was assessed.

Results: While the logistic Cardiac Surgery Score (logCASUS), Rapid Clinical Evaluation (RACE) and Sequential Organ Failure Assessment (SOFA) score all discriminated well between survivors and those who died, calibration of the models was inadequate. The Sepsis-3 criteria identified patients at increased risk of adverse outcomes. The KDIGO staging criteria were poorly calibrated, overestimating the risk associated with mild oliguria following cardiac surgery. The Bayesian urine output model discriminated excellently between those who did and did not go on to suffer severe oliguria and was well calibrated. Postoperative potassium and magnesium concentrations were similar for those who did and did not suffer AF.

Conclusion: The clinical usefulness of existing risk stratification methods has been assessed and weaknesses identified. It has been demonstrated that serum potassium and magnesium concentrations are unlikely to be useful when predicting AF following cardiac surgery. A novel approach to modelling urine output has been described and validated. The model's performance should be assessed in other settings and then the clinical usefulness of the model could be assessed in clinical trials. The methodology described in this thesis should be replicated to improve postoperative prediction of other complications following cardiac surgery.

<u>Declaration</u>

The development of the dynamic Bayesian model included in chapter ten of this thesis forms part of the dissertation that will be submitted by Jordan Oakley to Durham University for his Master's degree in statistics. As stated in chapter ten, the model was developed by Jordan Oakley, guided by the author of this thesis. The author of this thesis assessed the model's performance and was first author of the manuscript which forms the basis of chapter ten. No other portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

<u>Copyright statement</u>

For Janet, Emily, Sarah and Thomas

Preface

The author of this thesis gained the degree of MBChB (European Option) from the University of Manchester in 2007. Following the completion of Foundation Training he entered Speciality Training in Anaesthesia in 2009. Having been awarded Fellowship of the Royal College of Anaesthetists in 2012 he took time Out of Training for Research between ST5 and ST6. The work presented in this thesis was conducted during this period which represents the author's first formal experience in research.

Rationale for presentation in the journal format

This research project was conducted relatively early in my career as an anaesthetist. It was important to me that I developed skills during this period that would be useful to me for the rest of my career. In particular, I wanted to develop my scientific writing and gain experience of the peer review publication process. This approach was agreed during discussions with my supervisory team. The manuscripts produced all relate to the common theme of risk prediction following cardiac surgery but equally stand alone as individual studies. The systematic review forms the basis of the thesis' introduction. The manuscript describing the validation of RACE, logCASUS and SOFA scores as well as those validating risk stratification according to sepsis status and stage of acute kidney injury are included as results chapters. They are followed by two manuscripts concerning the development of new risk prediction tools for use following cardiac surgery. The first discusses the development and validation of a novel model to identify those at risk of severe oliguria and the second investigates the potential usefulness of serum electrolyte concentrations when identifying those at risk of atrial fibrillation. Four manuscripts have been published already and one is in the revision process following submission to Critical Care Medicine. The last manuscript is in the final stages of co-author review and will shortly be submitted for peer review publication.

SECTION ONE: INTRODUCTION

Structure and design of thesis

This thesis comprises four sections containing chapters which are either journal articles or thesis subsections. Where the chapter is a journal article, the formatting required by the journal which published the article or the journal to which the manuscript has been submitted is maintained. Where the chapter is a thesis subsection, it serves to provide background information, expand discussion of methodologies employed in this research programme or discuss the thesis' principal findings.

SHH was first author of all journal articles presented in this thesis. As first author, SHH was responsible for the design of each study, the collection and cleaning of relevant data, the conduct of statistical analyses and the writing of the manuscripts. Co-authors are detailed at the start of each relevant thesis chapter alongside their role in the production of that manuscript. The main roles of co-authors were to guide the study design, to develop the Bayesian models described in chapter ten and to guide manuscript presentation. All chapters which are thesis subsections were written by SHH.

Section one contains three chapters. The first is a thesis subsection which provides a summary of existing risk prediction in critical care and cardiac surgery. It also gives background information on the subjects of cardiac surgery and in particular the postoperative care delivered to cardiac surgery patients. The second chapter is a systematic review of existing postoperative risk prediction models used in cardiac surgery. This was published in the Journal of Cardiothoracic and Vascular Anaesthesia. The third chapter is a thesis subsection which summaries the background to this thesis and details the thesis aims.

Section two contains detailed descriptions of the methodology used in this thesis. It comprises three chapters which are all thesis subsections. The first chapter details the design of the research programme and the collection of data analysed in this thesis. The second chapter describes the cleaning of the data and the identification of relevant endpoints. The third chapter discusses the statistical methodology used when conducting analyses for the studies included in the thesis.

Section three contains five chapters written in journal article format. The first chapter is a study which validates three mortality models identified during the systematic review. This was published in the Thoracic and Cardiovascular Surgeon. The second chapter describes the

validation of the use of the Sepsis-3 criteria for the stratification of risk related to sepsis following cardiac surgery. This was published in the British Journal of Anaesthesia. The third chapter which was published in BMC nephrology validates the use of the Kidney Disease Improving Global Outcomes acute kidney injury guidelines for the stratification of risk related to renal dysfunction following cardiac surgery. This chapter concludes with recommendations regarding future work developing models to stratify risk related to renal dysfunction. The fourth chapter is a study which describes and validates a novel model which predicts severe oliguria in patients who have undergone cardiac surgery. This manuscript has been submitted to Critical Care Medicine and is currently in the revision process. The fifth chapter is a study investigating the potential roles of serum electrolyte concentrations in the prediction of atrial fibrillation following cardiac surgery. This chapter informs future work which will aim to develop a model to identify patients at risk of atrial fibrillation following cardiac surgery.

The final section comprises one chapter which is a thesis subsection discussing the findings of the research programme. This section summarises the conclusions drawn from the work carried out and makes recommendations for future research.

Chapter One: General Introduction

1.1. Risk prediction in cardiac surgery and critical care – an overview.

Identification of patients at risk of poor outcomes has always been key to the practice of medicine. Before the development of formalised risk stratification tools, risk assessment was based on information identified from the patient's history and physical examination. This information was considered by the clinician in the context of the clinician's own medical knowledge and clinical experience to arrive at a diagnosis and an associated prognosis or expected outcome. The prognosis and its accuracy are vitally important to allow the patient and their loved ones to understand their condition and how it is likely to affect them. The prognosis also informs the selection of appropriate treatment options.

If the prognosis is entirely dependent on the knowledge and experience of the treating clinician there will be variation in the assessment of illness severity and the estimation of the likelihood of poor outcomes. Risk stratification tools formalise the assessment of the risk of poor outcomes by combining the known effects of risk factors present to provide an overall assessment of the severity of the patient's condition. Simpler tools assign points where known risk factors are present, allowing the patients at greatest risk (i.e. those with the greatest number of points) to be identified. Logistic models go further and provide an estimate of the risk of adverse outcomes given the presence of combinations of known risk factors. The clinical usefulness of the risk estimates for clinicians treating individual patients is subject to a number of limitations which are discussed later in this section. However, if the risk estimates produced by such models are validated in multicentre datasets they can be used to benchmark the performance of institutions and clinicians. Benchmarking is a means of comparing of outcomes between institutions to assess the quality of the care they provide. The most common comparison made is between mortality rates. Analysing the quality of care delivered by comparing crude mortality rates is not helpful as the number of deaths in an institution would vary markedly depending on the premorbid physiological condition of the patients being treated. During benchmarking, to compensate for differences in patient condition before treatment, risk prediction models are used to create estimates of expected mortality risk for groups of patients based upon the presence of known risk factors. Observed outcomes are then compared with predictions and the ratio of observed deaths to predicted deaths is termed the standardised mortality ratio (SMR). In good institutions the observed mortality rate will be below the rate predicted by the model and the SMR will be less than one. In cardiac surgery the need for accurate risk predictions to allow benchmarking of patient outcomes was a major driver for model development. A landmark report into the high incidence of unexpected deaths in children undergoing cardiac surgery at the

Bristol Royal Infirmary in the 1990s stressed the importance of monitoring the performance of surgeons and institutions.(1) As a result, SMRs are now published for all cardiac surgeons and institutions providing cardiac surgery in UK.(2) Various models have been described to calculate the predictions used for benchmarking in cardiac surgery; the most widely used include the various iterations of the European System for Cardiac Operative Risk Evaluation (EuroSCORE) (3-5) and Society for Thoracic Surgery (STS) scores.(6-9)

Similarly, many Intensive Care unit (ICU) risk scores including the various Acute Physiology And Chronic Health Evaluation (APACHE) scores (10-12) and Simplified Acute Physiology (SAPS) scores (13-15) have been developed over the past 35 years.(16) These scores were originally designed to quantify the pre-treatment risk levels of different patient cohorts to allow benchmarking of ICU performance. However, when calculated daily, they may also provide an updated assessment of risk which can be analysed to show response to treatment and an indication of the clinical progress. Serial use of risk prediction scores provides similar information to the use of scores such as the (Sepsis-Related) Sequential Organ Failure Assessment Score (SOFA) score which was designed specifically to track a patient's clinical progress.(17)

Risk estimates created using logistic regression modelling need to be interpreted with caution when treating individual patients. Such models are developed through analysis of large registries to identify risk factors associated with poor outcomes. The models calculate expected risk, taking into account physiological parameters and details from the medical history. Crucially, the risk predictions are for groups of patients with similar risk scores, not for individuals. Within a group of 100 patients with a score that is associated with a mortality risk of 5%, five would be expected to die and 95 would be expected to survive. Importantly, the model cannot distinguish between survivors and those who will die. This is a severe limitation of existing models; no model predicts outcomes in a manner which is suitable for providing more than a rough context when making treatment decisions for individuals.(16)

This thesis explores the potential use of the vast amounts of data recorded on patients on the CICU to quantify each individual's risk of adverse outcomes. Rather than focussing solely on mortality, specific adverse outcomes such as renal dysfunction and arrhythmia will be considered. The rationale for such an approach is that identification of increased risk of a particular complication is more clinically useful as it allows interventions targeted at the prevention of (or reduction in the harm caused by) the complication. In order to allow prompt, successful interventions, risk will need to be assessed more frequently than the once daily assessment described for existing models and should make predictions for individuals rather than groups of patients.

1.1.1. Cardiac Surgery – The ideal setting for early model development?

Risk models are most often developed through the retrospective analyses of datasets. Within the dataset, the outcomes to be predicted are identified and patients can then be grouped according to whether or not they suffered the outcome. Candidate predictor variables can then be studied with the aim of identifying differences between those who do and do not suffer the outcome being predicted.

It is prudent to start model development in datasets in which interpatient variation is limited to increase the likelihood of identifying associations which may be hidden by confounders in more heterogeneous datasets. Any associations identified can then be tested in more heterogeneous datasets at a later time.(18) Postoperative cardiac surgery patients are particularly well suited to the early stage of model development. Cardiac surgery patients have similar risk profiles; they all have cardiac disease requiring surgery and (except for those undergoing an emergency procedure) they have all been declared fit enough to undergo the surgery. Patients undergo one of a limited number of procedures and their care is typically managed by a small group of clinicians. Unfortunately, the risk of complications following cardiac surgery is relatively high as outlined in chapter four of this thesis. As they are relatively common, complications are usually manged according to protocols which ensure all patients receive evidence-based treatments. As a result of the high incidence of complications, models can be developed in datasets with smaller sample sizes than would be required if the rates of complications were lower. Finally, cardiac surgery patients are almost uniformly subjected to intensive monitoring for the majority of their critical care stay. As a consequence, a vast number of high quality measurements are made as part of routine care. If those frequently recorded measurements could be preserved they are likely to be of high enough resolution to allow detailed analyses to be performed and to allow frequent updates of the risk estimates. Moreover, as these data are recorded routinely any findings from this research programme are likely to have a real world impact.

Before discussing risk prediction further, the remainder of this introduction provides a context to the thesis by briefly discussing cardiac surgery and summarising the principals of post-operative care provided on the ICU.

1.2. Summary of the history of cardiac surgery

Cardiac surgery has changed dramatically since the earliest procedures were described in the late 19th century.(19)   Until the middle of the 20th century, cardiac surgery was largely limited to closed heart procedures and was reserved for patients who would die in the near future without surgical intervention due to a high risk of operative mortality.(19)   In the early and mid-20th century the number of described surgical procedures increased and surgeons performed the earliest open-heart repairs using hypothermia amongst other techniques to protect the patient from the effects of hypoxia during the procedure.(20) A major milestone was reached in 1953 when John Gibbon performed the surgical closure of an atrial septal defect whilst supporting the patient using the first cardiopulmonary bypass (CPB) equipment.(21) Although he struggled to reproduce his results after initial success, his work led to the development of more sophisticated CPB equipment which allowed surgeons to achieve more reproducible success.(22) Following the widespread adoption of CPB in the 1970s, it was possible to perform open heart procedures which took longer to perform and mortality began to fall as more procedures were undertaken. (23, 24)

Today over 35,000 cardiac surgery operations are performed each year in the UK alone. Coronary artery bypass grafting (CABG) is the most common surgical procedure performed. This is followed in order of frequency by aortic valve replacement and mitral valve replacement or repair.   The mortality rate associated with cardiac surgery has continued to fall and is currently around 3%.(25) The vast majority of deaths occur in patients who develop serious complications after cardiac surgery while recovering on the Cardiac Intensive Care Unit (CICU).(2) Some of these complications, such as renal failure, are associated with mortality rates of up to 60%.(26-31) Patients who survive complications often require specialist treatments and have prolonged CICU stays occupying beds that would otherwise be used for other elective surgical patients at a cost of >£150million/year to the NHS (National Health Service).(26, 27, 32-42)

1.3. Management of patients on the Cardiac Intensive Care Unit (CICU) following cardiac surgery

In order to provide a context to the complications which occur following cardiac surgery it is important to understand the postoperative management of patients. This subsection discusses routine care following cardiac surgery and the key complications which may develop.

Patients who have undergone major cardiac surgery are routinely admitted to a critical care environment after their surgery.(43) The majority of cardiac surgery patients are managed according to similar principles and pass through the same milestones as they recover from their operation.(44)

Patients are routinely transferred to the CICU sedated, intubated and mechanically ventilated(43) and are treated according to Intensive Care Society and The European federation of Critical Care Nursing Association's standards governing staffing levels. These patients are all classified as Level 2 or 3 patients who are therefore nursed with nurse:patient ratios of 1:1 or 1:2.(45, 46) The patients are subjected to continuous physiological monitoring which is complemented by nursing observations which are usually obtained hourly. Parameters monitored continuously include fraction of inspired oxygen, capnography, electrocardiography (ECG), intra-arterial and central venous blood pressures and oxygen saturation via pulse oximetry.(43) Nursing staff record urine and drain output hourly and document temperature and neurological assessment every 4 hours. Arterial blood gas analyses are performed on an ad hoc basis but usually at least every 4 hours and laboratory-based biochemical and haematological analyses are performed daily.(47) Cardiac output was previously assessed almost exclusively using thermodilution techniques via a pulmonary artery catheter however a in a recent survey of current practice it was found that the majority of units use transthoracic echocardiography (TTE) or transoesophageal echocardiography (TOE) to investigate low cardiac output states. Pulmonary artery catheters are still employed, particularly in cases in which pressures in the right side of the heart need to be measured.(47)

Critical care management aims to normalise physiology while anticipating and detecting early signs of any developing complications. If such signs are detected measures are taken to prevent or reduce the harm caused by any complications while treating the cause of the complication.

1.3.1. Management of the respiratory system

Protocols vary between institutions but patients are not usually extubated until they are able to follow commands, have a good respiratory pattern, good gas exchange and absence of metabolic and electrolyte derangement as evidenced by arterial blood gas analyses, and a normal body

temperature.   Hypothermia or respiratory failure necessitate prolonged sedation and ventilation until successfully treated or corrected.(48) Some patients display good respiratory parameters while sedated but become agitated with an associated deterioration in respiratory function as sedation is lightened.   In this scenario, extubation is usually postponed as it is possible that once the stimulus of the endotracheal tube is removed the patients will not be alert enough to protect their airway and maintain adequate spontaneous ventilation.

1.3.2. Respiratory complications

The two major respiratory complications after cardiac surgery are failed extubation and prolonged mechanical ventilation.

Failed extubation

Around 4% of patients who have undergone cardiac surgery will fail to maintain their oxygenation when the endotracheal tube is removed despite having fulfilled all extubation criteria.(49) Such patients are said to have "failed extubation" and require reintubation and reinstitution of mechanical ventilation followed by a subsequent repeat attempt at extubation.(50) Even among patients who have undergone straight forward, off-pump CABG and been assessed as appropriate for fast track extubation, 2.5% fail extubation.(51) In cardiac surgery patients, failed extubation is associated with a mortality of up 13-40% and an increase in CICU length of stay of 6 days.(49, 50) Much of the increased risk may be related to the contamination of the lungs with oropharyngeal or gastric contents during the processes of extubation and reintubation or collapse of alveoli that occurs during ineffective self-ventilation. Identifying those at increased risk of failed extubation prior to removal of the endotracheal tube could reduce the harm caused by failed extubation but no reliable methods for doing so have been described.

Prolonged mechanical ventilation

The second important respiratory complication is prolonged mechanical ventilation (PMV). In most modern fast track protocols for low risk patients, extubation is expected to occur within 8 hours of CICU admission. The definition of PMV varies in the literature; it is described in different studies as ventilation for longer than 24h(30, 52), 48h(53), 72h(54) or 96h(55). While the definition of PMV varies, it is widely accepted that longer periods of ventilation are associated with worse outcomes in terms of mortality, development of other complications and length of hospital stay.(52-54) While the nature of the associations remains unclear, PMV causes increased trauma to the respiratory tract and increases the risk of pneumonia.(56) PMV (except when delivered via a tracheostomy) generally requires sedation. As well as its adverse effect on

neurological outcomes, prolonged sedation also worsens other outcomes through causing immobility and cardiovascular suppression.(57) It also prevents usual feeding and is associated with increased risk of acute kidney injury.(58)

1.3.3. Management of the cardiovascular system

Cardiovascular management on CICU, as in all ICU settings, aims to optimise oxygen delivery to the vital organs while reducing the workload and oxygen demand of the heart. This is achieved through assessment and optimisation of intravascular fluid volume status (preload) and haemoglobin concentration, optimisation of systemic vascular resistance (afterload) and control of heart rate and myocardial contractility.(43)

However, surgery on the heart and the use of CPB present specific stresses to the cardiovascular system resulting in increased risks of complications that are different to other surgical settings. Commonly encountered cardiac complications of open heart surgery include the onset of arrhythmias and cardiac tamponade which reduce cardiac output leading to decreased organ perfusion. In addition the cardiovascular system may be stressed by dysfunction of the circulatory system due to vasoplegia secondary to CPB and hypovolaemia secondary to haemorrhage or fluid shifts.

1.3.4. Cardiovascular complications - cardiac

Low cardiac output state

Following cardiac surgery, around 2-4 % of patients will suffer from low cardiac output syndrome.(59, 60) This syndrome is usually diagnosed by the requirement for mechanical or pharmacological support of cardiac output and carries a substantial burden in terms of mortality risk (24-38%) and increases the mean length of post-operative stay by 9 days(59).

Treatment of low cardiac output state aims to improve oxygen delivery using pharmacological therapies to preserve organ function while the precipitating causes are treated or the myocardium recovers from the physical and biochemical insult of surgery and reperfusion injury. Where pharmacological treatment is not sufficient, advanced treatments such intra-aortic balloon pumps or mechanical circulatory support (MCS) may be employed.(61, 62) The most common reversible cardiac causes of low cardiac output are arrhythmias and cardiac tamponade.

Arrhythmias

The term "arrhythmias" covers a range of cardiac rhythm disturbances of different frequencies with varied associated consequences. Arrhythmias with a ventricular rate greater than 100 beats

per minutes are termed tachyarrhythmias. The most frequent tachyarrhythmia after cardiac surgery is atrial fibrillation/flutter (AF) which occurs in up to 30% of patients.(63-65) Factors that have been shown to be associated with the onset of AF include electrolyte imbalance, reperfusion injury, sepsis and post-operative hypovolaemia with associated tachycardia.(66) For most of those who develop AF, normal sinus rhythm will be restored through chemical or, more rarely, electrical cardioversion. However AF, even when transient, is associated with increased lengths of CICU and hospital stay, the development of further complications and increased short and long term mortality.(67, 68) For some patients AF will persist and they will require lifelong anticoagulation to mitigate the associated risk of stroke.(69)

Clinicians attempt to reduce the risk of arrhythmias by correction of predisposing factors such as hypovolaemia, hypoxaemia, electrolyte disturbances and sepsis. Where AF occurs, arrhythmogenic agents such as beta-adrenoceptor agonists should be discontinued where possible and if the arrhythmia persists class II or III antiarrhythmic agents should be administered. Most commonly, beta-blockers or amiodarone are administered.(70)

There is evidence for the administration of anti-arrhythmics perioperatively to reduce the risk of post-operative AF, particularly that those already taking beta-blockers should continue to do so.(71, 72) However trials have repeatedly suggested that risk stratification prior to making a decision to prescribe prophylactic anti-arrhythmic agents is needed to maximise the benefits of AF prevention while minimising the frequency of side effects in patients who would not have gone on to develop AF without the prophylactic treatment.(73, 74)

Other tachyarrhythmias are much less common after cardiac surgery (sustained ventricular tachycardia has an incidence of around 1%). However, they carry a risk of mortality of up to 50%.(75, 76) They are often related to existing structural abnormalities of the heart or ischaemic events although they may also result from administration of medication to treat other arrhythmias.(77) Treatments include class Ia, Ib and III anti-arrhythmics.

Any tachyarrhythmia associated with syncope, shock, ischaemic pain or congestive cardiac failure should be treated with immediate direct current cardioversion.(70)

Bradyarrhythmias are less frequent and are usually caused by atrioventricular block or sick sinus syndrome.(77) Bradyarrhythmias are relatively more common after valve surgery during which the conducting system of the heart is more likely to be disturbed. Most surgeons routinely insert temporary epicardial pacing wires in such cases. In 1-4% of cardiac surgery patients a permanent pacemaker is required because the rhythm does not return to its pre-operative state.(78)

Cardiac tamponade

Bleeding within the surgical field after cardiac surgery can result in a collection of fluid around the heart. The fluid may compress the low pressure cardiac chambers (the atria and right ventricle) and drastically reduce cardiac output.   The incidence of re-operation for cardiac tamponade is around 3%.(37, 79) Early studies showed increased mortality (from 4.2% to 10.7%) and incidence of prolonged mechanical ventilation (from 8.6% to 24%) in cardiac surgery patients who required re-sternotomy for bleeding when compared with those who did not.(80, 81) However, a more recent study found that resternotomy *per se* was not associated with worse outcomes. Rather it was delayed recognition and treatment of tamponade that was associated with increased rates of mortality, renal failure, prolonged mechanical ventilation and increased length of stay on CICU. (37)

Classically, cardiac tamponade presents as Beck's triad of low arterial blood pressure, distended neck veins and muffled heart sounds.(82) However on CICU after cardiac surgery, cardiac tamponade is usually diagnosed following a period of decreasing cardiac output despite adequate fluid resuscitation. A fall in arterial blood pressure accompanied by a rise in central venous pressure and signs of global hypoperfusion such as hyperlactataemia develop.   Automated monitoring algorithms could monitor the physiological parameters described above together with the dose of vasoactive medications being administered to identify patients developing tamponade before their physiology becomes markedly deranged. TTE or TOE can identify fluid around the heart or compression of the atria and right ventricle.(83) The treatment of symptomatic tamponade involves resternotomy, evacuation of blood or haematoma and haemostasis. The sternum may be left open to prevent recurrence of tamponade if re-accumulation of fluid is considered to be likely.

1.3.5. Cardiovascular complications - non-cardiac

Hypovolaemia

Fluid management plays a key role in management of patients following cardiac surgery. Where hypovolaemia is suspected, volume expansion is achieved through the use of balanced crystalloid solutions, colloid solutions (synthetic and human albumin solution) and blood products. Fluid therapy is routinely guided by response of central venous pressure (CVP) and arterial blood pressure to fluid challenges.(47) The response of cardiac output (measured using thermodilution or indicator dilution using pulmonary artery catheters) to fluid challenges was traditionally

considered to be the gold standard guide to fluid therapy. More recently TTE or TOE has become the most used method for assessing the filling of cardiac chambers and contractility of the myocardium when unable to assess fluid volume status in more complex patients.(47)

Vasoplegia

Up to 50% of cardiac surgery patients require vasopressor or inotropic medication postoperatively (47) and relative hypovolaemia secondary to vasoplegia occurs in up to 20% (84). Vasoplegia occurs secondary to a systemic inflammatory response triggered by passage of the patient's blood through the CPB machine.(85)

1.3.6. Other serious complications

Renal

Acute Kidney Injury

There are multiple mechanisms through which renal injury may occur during and after cardiac surgery. Causes are commonly classified as pre-, intra- or post-renal. Pre-renal causes result in decreased renal perfusion during CPB or the post-operative period. This hypoperfusion may be caused by altered renal blood flow or globally decreased cardiac output. Where cardiac output is decreased this is most commonly a consequence of arrhythmias, a decrease in myocardial contractility or decreased preload secondary to absolute or functional hypovolaemia. Functional hypovolaemia due to excessive vasodilation may itself be caused by a systemic inflammatory response to CPB or infection. Intra-renal causes of renal damage include the administration of nephrotoxic drugs such as intravenous contrast, antibiotics and antihypertensives, haemolysis secondary to the use of CPB machines and endogenous toxins released as part of the stress response to surgery. Post renal causes include urinary retention and urinary catheter obstruction.(86)

Acute kidney injury (AKI) is diagnosed and classified according to the recently devised Kidney Disease International Global Outcomes(KDIGO) Acute Kidney Injury Work Group criteria.(87) The system's classification criteria include threshold values for serum creatinine concentration and hourly urine output. The most severe category (Stage 3) is also diagnosed if a patient requires renal replacement therapy (RRT). Around 3% of cardiac surgery patients will suffer AKI requiring RRT (86) with an associated mortality rate of up to 60% and an average increase in length of stay of 13 days.(29, 88, 89) Even mild AKI is associated with both increased mortality and healthcare

costs (88, 90) with risk of mortality correlating with severity of AKI.   Early identification of those at risk of developing severe renal dysfunction could allow remedial measures to be taken to reduce the severity of the AKI they suffer.

Neurological

Neurological complications after cardiac surgery include stroke and neurocognitive disturbance, the most common example of which is delirium.

Stroke

Stroke was found to occur in 2.6% of cardiac surgery patients who underwent surgery in the UK between 2004 and 2008.(91) This figure is in keeping with those quoted in studies in other populations.(92, 93) Post-operative stroke carries a mortality risk of 22% and extends hospital stay by, on average, 7 days.(94) Surgical intra-operative strategies to reduce the incidence of post-operative stroke include imaging the ascending aorta prior to cannulation to avoid dislodging debris from atheromatous plaques and the use of de-airing techniques and filtration devices to remove emboli.(95) Physiological considerations include maintenance of normal haematocrit and blood glucose concentration together with relative normotension and mild hypothermia during CPB.(96) Perioperatively, TOE should be used to ensure there is no aberrant communication between the systemic and pulmonary circulations and that there are no clots in the atria prior to surgery and that cardiac chambers have been de-aired prior to restoration of circulation. (95, 97) In the post-operative period to reduce the incidence of stroke, hyperthermia and hypotension are avoided while blood glucose concentration is tightly controlled. Atrial fibrillation is treated early as it is a recognised risk factor for stroke.(96)

Delirium

Delirium is a syndrome of acute onset characterised by inattention, impaired consciousness and disordered cognition which has a fluctuating course.(98) Following cardiac surgery delirium occurs in 15-50% of patients.(92, 98-100) It is associated with increased risk of long term mortality, an average ten-day increase in length of hospital stay and in increased risk of need for discharge to a nursing home.(100-102) Identified risk factors for post-operative delirium include intra-operative normothermia, prolonged duration of CPB and post-operative mechanical ventilation together with the administration of high doses of fentanyl.(99, 103, 104) Strategies employed on the CICU

to reduce the risk of delirium include early cessation of sedative medications and avoidance or prompt treatment of known risk factors such as hypotension, hypoxaemia and infection. Delirium screening is widespread and facilitates early detection and treatment to reduce its adverse effects on outcomes.(105)

Postoperative neurocognitive dysfunction

Cognitive function may decline after any form of surgery(106) and postoperative neurocognitive dysfunction (POCD) occurs in up to 50% of cardiac surgery patients.(107) POCD is usually defined as a reproducible decline in performance when undertaking various tasks designed assess neurocognitive function but attempts to study POCD are limited by the lack of a standardised definition. POCD has been linked to many adverse outcomes including increased mortality, length of hospital stay, and likelihood of early retirement.(107, 108)

Although the causes of POCD remain unclear, proposed mechanisms by which neuronal damage leading to POCD may occur are chiefly related hypoperfusion of the brain, inflammation and the adverse effects of anaesthetic agents.

Absolute hypoperfusion may be global, such as that which occurs during periods of hypotension in which blood pressure is out of the range in which cerebral autoregulation of blood flow can be achieved (mean arterial pressure<60mmHg). This may be a particularly important factor in patients with pre-existing cerebrovascular disease where flow may be limited by atheroma. There are also many possible causes of localised, absolute hypoperfusion such as microemboli from atheroma disturbed during surgery or air bubbles introduced into the circulation during open heart surgery.(109) Relative hypoperfusion may occur when the metabolic demand of the brain is exceeds it supply of oxygen. Avoidance of hyperthermia is therefore an important consideration during the conduct of cardiac anaesthesia and recovery on CICU.(110)

The inflammatory response to surgery has been suggested as a cause of POCD through possible detrimental effects on cerebral autoregulation. However, no treatment proposed to reduce the effect of inflammation on the brain has proven to be effective. (111-113)

CPB was thought to be a likely cause of POCD as it is conceivable that its use during cardiac surgery would lead to increased delivery of emboli and inflammatory cytokines to the brain. However, when comparing rates of POCD in on-pump and off-pump cardiac surgery no clear link between CPB and POCD has been elucidated.(114, 115)

The use of drugs associated with post-operative delirium such as fentanyl has not been shown to be linked to the development of POCD suggesting that the effects of such drugs on cognition are

temporary.(116) The effects of anaesthetic agents on POCD are equivocal. Interestingly in non-cardiac surgery patients the incidence of POCD has been shown to be the same in patients undergoing general and regional anaesthesia implying that the use of general anaesthetic agents does not contribute to the risk of POCD.(117)

Until further causes of POCD are identified prevention centres on optimisation of cerebral perfusion and oxygenation, reduction in cerebral metabolic demand during surgery and avoidance of emboli.

Gastrointestinal

The rate of major gastrointestinal complications after cardiac surgery is around 2%.(118-120) The most common complications are paralytic ileus, mesenteric ischaemia and gastrointestinal haemorrhage(120) some of which are associated with mortality rates of up to 67%.(121) Studies have found that pre-and postoperative markers of poor cardiovascular function and a prolonged CPB time are the most significant risk factors for postoperative gastrointestinal complications. An association between previous history of gastrointestinal disease has been found in some studies but not in others.(118-120)

Prevention of gastrointestinal complications centres on early mesenteric feeding, maintaining oxygen delivery through optimising mesenteric blood flow and avoidance of hypo and hypercoagulable states.

Infective

Nosocomial post-operative infections occur in up to 14% of CICU patients. The most common infective complications of cardiac surgery include wound infections, line (particularly central venous catheter) infections and pneumonia.(122) Outcomes vary according to the site of infection and the general condition of the patient at the time of infection. The average risk of death in patients with infective complications has been estimated at 25% (123) whereas candidaemia carries a mortality risk of 40-80%.(124, 125)

Risk factors for infection include smoking, diabetes mellitus and complicated clinical course as indicated by prolonged operation times, prolonged ICU admission or the need for intra-aortic balloon pump counter pulsation.(123, 126)

Prevention of infection on CICU is a massive area but centres on infection prevention through measures such as antibiotic prophylaxis, blood glucose control, hand hygiene, single patient

equipment use, and aseptic conduct of invasive procedures.(126) Specific infections such as ventilator acquired pneumonia and central venous catheter infections are minimised through the use of care bundles aimed at reduced the incidence of risk factors for infection.(126-128) Treatment involves administration of antibiotics which must be targeted to the suspected or proven pathogen and continued for an adequate duration.   Treatment of sepsis focuses on support of organ systems while the immune system and antibiotics fight the infection.

1.4. "Failure to rescue" following cardiac surgery

Studies have shown that institutions with the worst patient outcomes following cardiac surgery don't just experience more complications; they also fail to rescue more of the patients who suffer complications.(129) "Failure to rescue" occurs when the lack of appropriate remedial treatment leads to increased severity of the complication or other related complications.(130)   Causes of this failure to rescue can be divided into two main categories; i) failing to recognise complications promptly, and ii) failing to manage complications effectively.(131) Risk prediction models have the potential to improve the quantification of the risk of specific complications. In doing so, these models can reduce the incidence of failure to rescue by allowing identification of impending complications before they become established. Moreover, if these models are validated in multicentre studies, they can standardise the assessment of risk across multiple institutions and improvements in risk detection in these centres.

Chapter Two: Risk models that utilise postoperative patient monitoring data to predict outcomes in adult cardiac surgery; a systematic review (Published journal article)

Chapter type: Journal article

Status: Published

Submission Date: 25/05/16

Published online: 04/10/16

Reference: Howitt SH, Grant SW, Riding DM, Malagon I, McCollum CN. Risk Models That Use Postoperative Patient Monitoring Data to Predict Outcomes in Adult Cardiac Surgery: A Systematic Review. *J Cardiothorac Vasc Anesth* 2017;31(5):1865-1877.

Contributions: SHH conceived the plan for the review, designed the study, conducted the search, identified relevant articles, analysed relevant articles and drafted the manuscript. SWG, IM and CNM all developed the plan for the review. DM identified relevant articles. SWG contributed to the production of a final draft. All authors reviewed and approved the final manuscript.

2.1. Rationale for selecting models which analyse postoperative variables.

This chapter is in the form of a systematic review of risk prediction models which have been validated for use following cardiac surgery. This review sought to identify all models which had been validated in the prediction of outcomes identified in chapter one. The manuscript focuses on models which analyse postoperative variables to estimate risk because this thesis aims to develop models which provide continuously updated risk estimates. Such models will depend on the analyses of postoperative data to provide the risk updates.

2.2. Abstract

Background

Preoperative risk prediction models are used to provide patients with information on perioperative mortality and to risk-adjust surgical outcome analyses. However, risk estimates from preoperative models may become increasingly unrealistic after surgery as they cannot take into account postoperative events. A number of risk models that utilise postoperative data have been developed or validated for adult cardiac surgery but none has been widely adopted. The objective of this review was to identify all such risk prediction models and discuss their uses and limitations.

Methods

A systematic review of the literature was undertaken with Medline, EMBASE and the Cochrane Library searched to identify relevant papers. Identified studies were assessed with regards to model discrimination, model calibration and clinical validity.

Results

The search identified 1649 publications. 86 met the inclusion criteria from which 14 validated models were identified. Eight models were originally designed for use in general intensive care units but subsequently validated for use following cardiac surgery. Six models were designed specifically for cardiac surgery patients. Most models that demonstrated good statistical performance were designed for clinical benchmarking purposes. No validated model provides predictions for specific complications or patient deterioration more frequently than once daily.

Conclusions

This review has identified a number of risk prediction models that utilise postoperative data and have been validated for the prediction of outcomes after adult cardiac surgery. The lack of adoption of these models may be due to variations in patient monitoring protocols and the inability of existing models to guide clinical decision making for individual patients. The risk scores identified are likely to be useful for assessing Cardiac Intensive Care Unit performance, informing discussions with patients or relatives and allocating resources. Future research to develop and validate predictive models that utilise postoperative data to produce frequent estimates of risk for specific patient outcomes may be of benefit.

2.3. Introduction

The most commonly used risk prediction tools in European adult cardiac surgery are the EuroSCORE models.[1, 2] These models use preoperative patient data to predict postoperative mortality. They play a vital role in preoperative clinical decision making, informed consent and performance monitoring. However, they have limited clinical value in subsequent patient management as the predicted risk cannot be modified by the occurrence of significant postoperative events or the patient's response to those events. Consequently, risk estimates may become unrealistic as postoperative events unfold.

Currently, adult cardiac surgery carries a mortality risk of 2-3%.[3, 4] This risk is significantly higher in those who develop postoperative complications. Respiratory [5, 6] and renal failure [7, 8] following cardiac surgery are associated with mortality rates of up to 18% and 60% respectively. Models that identify patients at risk of such complications could reduce morbidity and mortality by alerting clinicians to those who would benefit from early, targeted interventions.

A number of risk prediction models that utilise postoperative data have been developed or validated for use in adult cardiac surgery. Some models calculate risk based on the initiation of treatments or the occurrence of events in the postoperative period.[9-11] These models may provide updated risk estimates that guide staff and resource allocation and may also inform discussions with patients and their relatives. However, they often only demonstrate increased risk once end organ damage has occurred and remedial measures have been taken. Accordingly, they are of limited use in the early identification of those at risk and may not enable timely administration of preventative treatment. Their usefulness for benchmarking may be limited by interinstitutional variation in initiation of treatments according to local protocols. Models based on postoperative physiological monitoring data are potentially better suited to these tasks. Such models share similarities with Early Warning Score (EWS) models[12], which have been widely adopted to identify ward-based patients at risk of clinical deterioration based on analyses of physiological values including heart rate, respiratory rate, oxygen saturation, blood pressure, temperature and conscious level. Despite widespread adoption of EWS models on other wards and the availability of vast amounts of patient monitoring data in the ICU setting following cardiac surgery, no risk model based on patient monitoring data following cardiac surgery has been widely adopted. The objective of this review was to identify all validated risk models which use postoperative patient monitoring data to predict outcomes in adult cardiac surgery. Clinical validity and statistical performance were evaluated to explore possible reasons for the lack of adoption.

<u>2.4. Methods</u>

2.4.1. Literature search and study eligibility

The Database of Abstract of Reviews of Effects (DARE) and PubMed Health databases were searched using the terms "cardiac surgery" or "coronary artery bypass" or "valve "and "risk prediction" or "model" for papers published since 2009 and revealed no existing Cochrane, Centre for Reviews and Dissemination (CRD) or PubMed Health registered reviews. A subsequent search of the Cochrane library, EMBASE and MEDLINE databases from inception to 2015 was performed using the Population, Intervention, Comparison, Outcomes, Setting (PICOS) framework (see Appendix). Two "readers" (SHH and DMR) independently screened the titles and abstracts to select potentially eligible studies. The full text of potentially eligible manuscripts was assessed by both readers independently. Studies were eligible if they reported the validation of a risk prediction model using postoperative patient monitoring data to predict outcomes after adult cardiac surgery. In addition to the validation study, the article that first described the validated model was identified and reviewed for details concerning model development. There were no restrictions on study design. Only studies presented in English were analysed.

2.4.2. Data extraction and quality assessment

Data were extracted from the eligible manuscripts by SHH and included first author's name, year of publication, study design, sample size and population characteristics. For studies describing the development of a risk prediction model information extracted included; statistical model used, factors included in the model, model outcomes and method of validation. For articles describing the validation (internal or external) of a risk prediction model in cardiac surgery patients information extracted included; the quality of the study, statistical performance of the model and characteristics of the validation cohort.

When assessing the models, three main aspects of their performance were considered: discrimination, calibration and clinical validity. Discrimination was usually assessed using the area under the Receiver Operator Characteristic curve (AUC).[13] An AUC of 0.5 represents discrimination between patients who experience an outcome and those who do not, that is no better than chance. An AUC of 1.0 represents perfect discrimination, with values >0.7 generally accepted to indicate adequate discrimination, and >0.8 considered good.[14-16]

Calibration, or how closely the predicted risk matches the observed risk, can be assessed using a variety of different methods. The Hosmer-Lemeshow (HL) test was most commonly used. A high HL χ^2 value with a low associated p value suggests that there is a significant difference between predicted risk and observed outcomes across sub-groups of the cohort.[17] Other calibration

measures included the Brier and the R^2 score. Brier score values approaching zero represent good calibration. The R^2 score is used for continuous outcomes e.g. length of stay, with a value of 1 indicating perfect fit. Clinical validity was assessed considering the quality of the study design, the methodology and the reporting.

2.5. Results

A total of 86 relevant studies were identified. A flow chart to describe the manuscript selection process is shown in Figure 2-1.
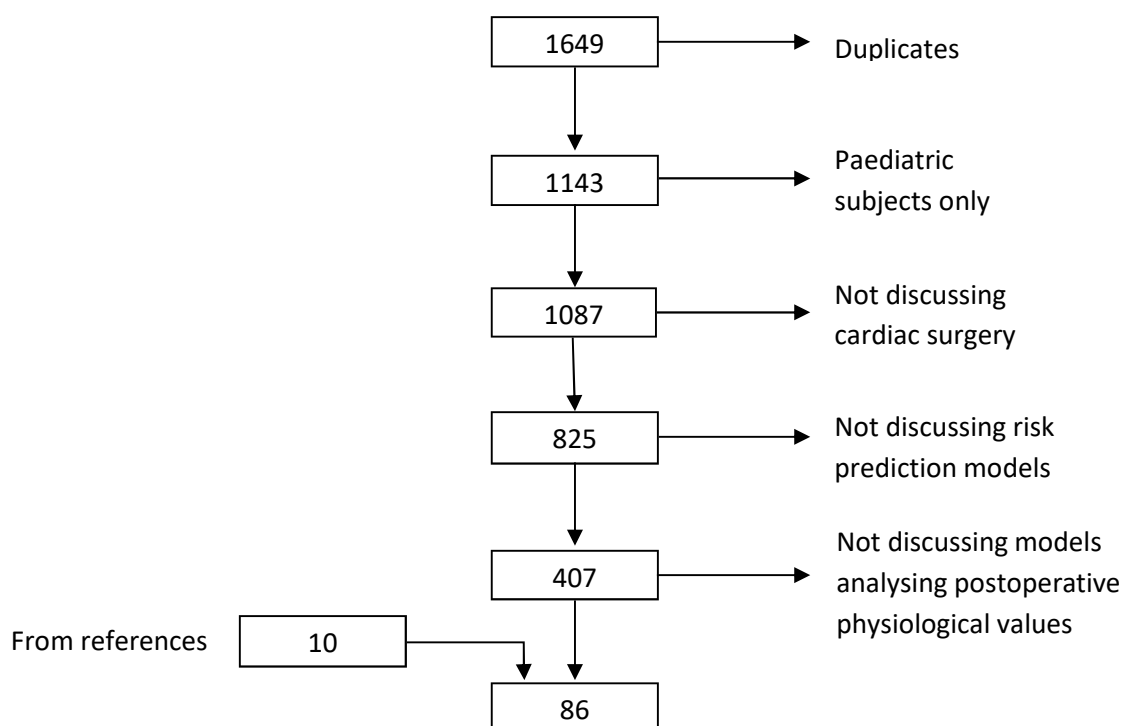
Figure 2-1 - Manuscript selection for review

Amongst these there were 14 risk models which had been validated for use in cardiac surgery patients (Table 2-1). Eight of these models were initially developed using data from general ICU populations with half of these developed using cohorts from which cardiac surgery patients were excluded. Six models were developed using only patients who had undergone cardiac surgery. Most models were developed using logistic regression but expert opinion, Bayesian modelling and Gaussian processes were also utilised. (Table 2-1)

Table 2-1 - Models validated for predicting outcomes following cardiac surgery

Model	Author	Year	Country	Development method	Design cohort	Cardiac surgery validation	Outcomes predicted	No. of physiological parameters
APACHE-II[18]	Knaus	1985	USA	Logistic regression	Excluded cardiac	External	Perioperative, ICU and 30 day Mortality; LOS-ICU ; Prolonged mechanical ventilation	5
APACHE-III[19]	Knaus	1991	USA	Logistic regression	Excluded cardiac	External	Hospital mortality; LOS-ICU; Treatment costs	5
SAPS-II[20]	Le Gall	1993	12 countries	Logistic regression	Excluded cardiac	External	Hospital and ICU Mortality; Prolonged mechanical ventilation	3
MODS[21]	Marshall	1995	Canada	Logistic regression	Surgical ICU	External	Mortality	5
SOFA[22]	Vincent	1996	16 countries	Expert Opinion	General ICU	External	Hospital and ICU Mortality; LOS-ICU	3
LODS[23]	Le Gall	1996	12 countries	Logistic regression	Excluded cardiac	External	Hospital and ICU mortality	5
ICURS[16]	Higgins	1997	USA	Logistic regression	Mixed cardiac	External	Hospital Mortality; Composite morbidity	4
SAPS-3[24]	Moreno	2005	35 countries	Logistic regression	General ICU	External	Hospital and ICU mortality	2
CASUS[25]	Hekmat	2005	Germany	Logistic regression	Mixed cardiac	Internal/ External	30 day and ICU mortality	5
Biagioli[26]	Biagioli	2006	Italy	Bayesian	CABG	Internal	Composite morbidity	2
AKICS[8]	Palomba	2007	Brazil	Logistic regression	Mixed cardiac	Internal	AKI	2
ICNARC[27]	Harrison	2007	UK	Logistic regression	General ICU	External	Perioperative mortality	7
Salamonsen[28]	Salamonsen	2008	Australia	Linear regression	CABG	Internal	LOS-ICU	3
Meyfroidt[29]	Meyfroidt	2011	Belgium	Gaussian process	Mixed cardiac	Internal	LOS-ICU	13*

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery, CABG – Coronary Artery Bypass Graft, LOS-ICU – Length Of Stay on the Intensive Care Unit, AKI – Acute Kidney Injury
* Included multiple statistical values for parameters including means, variances and cumulative totals

The overall quality of these studies was good (Table 2-2). The main limitation was a failure to clearly describe how missing data were handled. Occasionally, preoperative patient characteristics were not included, but in these studies composite measures of patient co-morbidity such as the mean EuroSCORE were usually provided.

<u>Table 2-2 – Validation studies-quality</u>

Validation study and Year	Models Validated	Patient selection criteria detailed	Consecutive Patients Studied	Preop health status well described	Patient demographics well described	Data collection	Handling of missing data	Outcome measures	Validation method	Validation group size
Becker 1995[30]	APACHE-III	Yes	Yes	No	Yes	Prospective	Patients excluded	Mortality and ICU-LOS	External	2,435
Higgins 1997[16]	ICURS	Yes	Yes	Yes	Yes	Prospective	Not discussed	Mortality Morbidity	Internal	2125
Kern 2001[31]	SAPS-II, APACHE-II	Yes	Yes	No	Yes	Prospective	Not discussed	Prolonged Mechanical Ventilation	External	687
Ceriani 2003[32]	SOFA	Yes	Yes	Yes	Yes	Not specified	Not discussed	Mortality	External	218
Serrano 2005[33]	ICURS	Yes	Yes	No	Yes	Prospective	Not discussed	Prolonged Mechanical Ventilation	External	569
Hekmat 2005[25]	APACHE-II,MODS	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	1057
	CASUS	Yes	Yes	No	Yes	Prospective	None missing	Mortality	Internal	1057
Patila 2006[34]	SOFA	Yes	Yes	Yes	Yes	Prospective	Not discussed	Mortality	External	857
Biagioli 2006[26]	locally customised ICURS, Biagioli	Yes	Yes	Yes	Yes	Prospective	Not discussed	Composite Morbidity	Internal	350
Gomes 2007[35]	SOFA	Yes	Yes	Yes	Yes	Not specified	Not discussed	Mortality	External	1458
Palomba 2007[8]	ICURS	Yes	Not specified	Yes	Yes	Prospective	Not discussed	AKI	External	603
	AKICS	Yes	Not specified	Yes	Yes	Prospective	Not discussed	AKI	Internal	215
Salamonsen 2008[28]	Salamonsen	Yes	Not specified	Yes	Yes	Prospective	Patients excluded	LOS-ICU	Internal	117

Table 2-2 – Validation studies-quality (cont.)

Validation study and Year	Models Validated	Patient selection criteria detailed	Consecutive Patients Studied	Preop health status well described	Patient demographics well described	Data collection	Handling of missing data	Outcome measures	Validation method	Validation group size
Hekmat 2010[36]	CASUS	Yes	Yes	No	Yes	Prospective	None missing	Mortality	Internal	3801
Doerr 2011[37]	CASUS, SOFA, SAPS-II APACHE-II	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	2801
Meyfoidt 2011[29]	Meyfroidt	Yes	Yes	No	No	Not specified	Imputed	LOS-ICU	Internal	499
Heldwein 2011[38]	LODS	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	2801
Badreldin 2012[39]	SOFA, CASUS	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	2801
Badreldin 2012[15]	SOFA	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	2801
Doerr 2012[40]	CASUS	Yes	Yes	Yes	Yes	Prospective	None missing	Mortality	External	4054
Doerr 2014[41]	SAPS-II, SAPS -III	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	5207
Ariyaratnam 2015[3]	APACHE-II, ICNARC	Yes	Yes	Yes	Yes	Prospective	Not discussed	Mortality	External	1646
Exarchopoulos 2015[42]	APACHE-II, SAPS-II, SOFA, CASUS	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	150
Tsaousi 2015[43]	APACHE-II, SOFA	Yes	Yes	No	Yes	Prospective	None missing	Mortality	External	1058

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery, LOS-ICU – Length Of Stay on the Intensive Care Unit, AKI – Acute Kidney Injury

Five of the 14 models included purely postoperative variables, four included preoperative and postoperative variables and five models included intraoperative, preoperative and postoperative variables. The variables used by the validated models are detailed in Table 2-3. The organ system most commonly assessed using patient monitoring data was the cardiovascular system. Many models simply include the mean arterial pressure while some depend on knowledge of cardiac output. Others use the composite measure of Pressure Adjusted Heart Rate which is based on heart rate, central venous pressure and mean arterial pressure as shown below.

$$\text{Pressure adjusted heart rate} = \frac{\text{Heart rate} \times \text{central venous pressure}}{\text{mean arterial pressure}}$$

The respiratory system was most commonly assessed using the ratio of arterial partial pressure of oxygen to inspired oxygen concentration. The renal system was assessed using blood test results rather than urine output in all but four models. Temperature was measured in five models.

The statistical performance of the ten models validated for prediction of mortality is shown in Table 2-4. The statistical performance of the models validated for the prediction of morbidity is shown in Table 2-5. Morbidity outcomes predicted included prolonged ICU stay, prolonged ventilation, acute kidney injury (AKI) and composite morbidity. A number of models were developed and validated for both mortality and morbidity. APACHE-II, SAPS-II, SOFA, ICURS and CASUS were validated in multiple patient cohorts. These all showed good discrimination in multiple studies with AUCs > 0.75. Of those validated in multiple studies, SOFA and CASUS scores consistently demonstrated the best combinations of AUCs >0.8 and p values > 0.05 for the HL χ^2 test in external validation cohorts.

Table 2-3 -Variables included in each model

Model	Pre-operative	Intra-operative	Postoperative physiological monitoring	Other Postoperative	Timing of capture
APACHE-II[18]	Age, Chronic Disease Status, type of admission	-	PaO_2/F_iO_2, Temp, MAP, RR	Blood tests: pH, WCC, K^+, Na^+, Hct, Cr GCS, FiO_2	Worst value recorded each day (originally within first 24hours)
APACHE-III[19]	Age, Previous surgery, Gender, Comorbidities	Number of grafts and vessels used. Urgency	HR, MAP, Temp, RR , A-a gradient, UO	Blood tests: Hct, WCC, Cr, Na^+, Albumin, Bilirubin, glucose, BUN, PaO_2	Worst value recorded within first 24hours
SAPS-II[20]	Age, Chronic Disease Status, Type of Admission	-	PaO_2/ F_iO_2, UO	Blood tests: Ur, Cr, WCC, K^+, Na^+, HCO_3^- GCS	Worst value recorded each day (originally within first 24hours)
MODS[21]	-	-	PaO_2/ F_iO_2, PAR	Blood tests: Bilirubin, Cr, Platelets GCS	Worst value recorded each day
SOFA[22]	-	-	PaO_2/ F_iO_2, MAP	Blood Tests: Cr, Bilirubin, Platelets, Vasopressor use, GCS	Worst value recorded each day
LODS[23]	-	-	PaO_2/ F_iO_2, HR, systolic BP, UO	Blood tests: WCC, Ur, Cr, Bilirubin, PT, Platelets, GCS	Worst value recorded each day (originally within first 24hours)
ICURS[16]	Age, Comorbidities, Albumin	CPB time Need for IABP after CPB	A-a gradient, HR, CI,	Blood tests: HCO_3^-	On arrival to ICU
SAPS-3[24]	Age, Comorbidities, Reason for Admission, Pre-admission events	Site of surgery	Temp, HR	Blood tests: Bilirubin, Cr, WCC, pH, Platelets GCS, FiO2, requirement for mechanical ventilation	Within 1 hour of admission
CASUS[25]	-	-	PaO_2/ F_iO_2, PAR	Blood tests: Cr, Bilirubin, lactate, Platelets. Neurological state, Requirement for IABP or VAD	Worst value recorded each day

Table 2-3 -Variables included in each model [cont.]

Model	Pre-operative	Intra-operative	Postoperative physiological monitoring	Other Postoperative	Timing of capture
Biagioli[26]	Age, Weight, Comorbidities, Cr, Requirement for IABP	Type of surgery Duration of CPB	DO_2I,	Blood tests: WCC Requirement for IABP	Within 3 hours of admission
ICNARC[27]	Age, diagnostic category, source of admission, CPR before admission	-	HR, systolic BP, Temp, RR, PaO_2/F_iO_2, UO,	pH, Ur, Cr, Na, WCC, GCS	Within 24 hours of admission
Salamonsen[28]	-	-	MAP, CVP, CI	Blood tests: HCO_3^- Requirement for IABP Cumulative adrenaline and noradrenaline doses	Average values over first four hours on ICU
Meyfroidt[29] *	Comorbidities, Pre-admission events	-	Multiple derived from BP, RR, F_iO_2, SpO_2, PAP, PEEP, HR, CVP, SPAP, UO, Drain Output, CO, Temp	Blood tests Medication	First four hours of admission
AKICS[8]	Age, Cr, Glucose, type of surgery, comorbidities	Duration of CPB	CO, CVP		On ICU admission

*see http://www.kuleuven.be/licm/ml/gpdischarge1.html for details of modelled variables

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery
A-a gradient – alveolar arterial gradient, Albumin – serum albumin concentration, Bilirubin - serum bilirubin concentration, BP – blood pressure, BUN – blood urea nitrogen, CI – Cardiac index, CO – cardiac output, CPB – cardiopulmonary bypass, Cr – serum creatinine concentration, CVP – central venous pressure, DO_2I – oxygen delivery index, F_iO_2– fraction inspired oxygen, GCS –Glasgow Coma Scale, Glucose – serum glucose concentration, Hct – haematocrit, $HCO3^-$ – serum bicarbonate concentration, IABP – Intra-aortic balloon pump, K^+ – serum potassium concentration, lactate – serum lactate concentration, MAP – mean arterial pressure, Na^+ – serum sodium concentration, PAR – pressure adjusted heart rate (HRxCVP/MAP), PaCO2 – arterial partial pressure of carbon dioxide, PaO2 – arterial partial pressure of oxygen, PAP – Peak airway pressure, PCWP – pulmonary capillary wedge pressure, PEEP – positive end-expiratory pressure, pH – blood pH, Platelets – Platelet Count, PT – Prothrombin time, RR – respiratory rate, SPAP – systolic pulmonary artery pressure, SPO_2 – oxygen saturation (pulse oximetry), Temp – temperature, UO – urine output, Ur – serum urea concentration, VAD – ventricular assist device, VT – Tidal Volume, WCC – White Cell Count

Table 2-4 - Studies validating models in the prediction of mortality in cardiac surgery

Model	Author	Year	Country	Validation cohort (n)	Measure of calibration*	Measure of discrimination*
APACHE-II	Hekmat[25]^	2005	Germany	Mixed cardiac (1057)	HL χ^2=6.6‡	AUC=0.89
	Doerr[37]	2011	Germany	Mixed cardiac (2801)	HL χ^2=30.6 (p<0.001)	AUC=0.87
	Ariyaratnam[3]	2015	UK	Mixed cardiac (1646)	HL χ^2=16.2 (p=0.001)	AUC=0.65
	Exarchopoulos[42]	2015	Greece	Mixed cardiac (150)	HL χ^2=10.9 (p=0.20)	AUC=0.82
	Tsaousi[43]	2015	Greece	Mixed cardiac (1058)	HL χ^2=7.4 (p=0.49)	AUC=0.86
APACHE-III	Becker[30]	1995	USA	Mixed cardiac (2435)	R^2=0.22	AUC=0.85
SAPS-II	Doerr[37]	2011	Germany	Mixed cardiac (2801)	HL χ^2=17.15 (p=0.03)	AUC=0.89
	Doerr[41]	2014	Germany	Mixed cardiac (5207)	HL χ^2=57.8 (p=0.000)	AUC=0.88
	Exarchopoulos[42]	2015	Greece	Mixed cardiac (150)	HL χ^2=5.1 (p=0.75)	AUC=0.80
MODS	Hekmat[25]^	2005	Germany	Mixed cardiac (1057)	HL χ^2=5.7‡	AUC=0.90
SOFA	Ceriani[32]	2003	Italy	Mixed cardiac (218)		AUC=0.71
	Patila[34]#	2006	Finland	Mixed cardiac (855)		AUC=0.78
	Gomes[35]	2007	Brazil	Mixed cardiac (1458)		AUC=0.74
	Doerr[37]	2011	Germany	Mixed cardiac (2801)	HL χ^2=6.75 (p=0.56)	AUC=0.91
	Badreldin[39]	2012	Germany	Mixed cardiac (2801)	HL χ^2=6.75 (p=0.56)	AUC=0.91
	Badreldin[15]	2012	Germany	Mixed cardiac (2801)	HL χ^2=14.9 (p=0.06)	AUC=0.88
	Exarchopoulos[42]	2015	Greece	Mixed cardiac (150)	HL χ^2=2.9 (p=0.57)	AUC=0.76
	Tsaousi[43]	2015	Greece	Mixed cardiac (1058)	HL χ^2=4.8 (p=0.58)	AUC=0.86
LODS	Heldwein[38]	2011	Germany	Mixed cardiac (2801)	HL χ^2=6.4 (p=0.49)	AUC=0.93
ICURS	Higgins[16]	1997	USA	Mixed cardiac (2125)	Good HL χ^2‡	AUC=0.85
	Gomes[35]	2007	Brazil	Mixed cardiac (1458)		AUC=0.77
SAPS-3	Doerr[41]	2014	Germany	Mixed cardiac (5207)	HL χ^2=15.2 (p=0.056)	AUC=0.89
CASUS	Hekmat[25]^	2005	Germany	Mixed cardiac (1104)	HL χ^2=5.1‡	AUC=0.96
	Hekmat[36]^	2010	Germany	Mixed cardiac (3801)	HL χ^2=7.0‡	AUC=0.95
	Doerr[23]	2011	Germany	Mixed cardiac (2801)	HL χ^2=14.0 (p=0.05)	AUC=0.97
	Badreldin[43]	2012	Germany	Mixed cardiac (2801)	HL χ^2=14.0 (p=0.05)	AUC=0.97
	Doerr[37]	2012	Germany	Mixed cardiac (4054)	O/E ratio=0.63	AUC=0.97
	Exarchopoulos[42]	2015	Greece	Mixed cardiac (150)	HL χ^2=2.2 (p=0.89)	AUC=0.89
ICNARC	Ariyaratnam[3]	2015	UK	Mixed cardiac (1646)	HL χ2 =9.10 (p=0.33)	AUC=0.85

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, O/E ratio –- ratio of observed to expected outcomes, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery, HL – Hosmer Lemeshow, AUC - Area under the receiver operating characteristic curve.

*if calculated on multiple days the value on the day of the best AUC is shown, ‡ p values not supplied, # only investigated maximum SOFA score, ^ if multiple similar samples of patients were studied in the same paper the values for the biggest sample are shown

Table 2-5 - Studies validating models in the prediction of morbidity in cardiac surgery

Model	Author	Year	Country	Validation cohort (n)	Measure of calibration*	Measure of discrimination*
Length of ICU stay						
APACHE-III	Becker[30]	1995	USA	Mixed Cardiac (2435)	R^2=0.08	
Salamonsen	Salamonsen[28]	2008	Australia	CABG (117)	R^2=0.38	
Meyfroidt	Meyfroidt[29]	2011	Belgium	Mixed cardiac (499)	Good HL χ2 (p=0.38) Brier 0.18	AUC=0.76
Composite morbidity						
ICURS	Higgins[16]	1997	USA	Mixed cardiac (2125)	Good HL χ^2‡	AUC=0.76
	Biagioli[26]	2006	Italy	CABG (740)	Poor HL χ^2‡	AUC=0.82
Biagioli	Biagioli[26]	2006	Italy	CABG (350)	Good HLχ^2 (p=0.35)	AUC=0.70
Acute Kidney Injury						
ICURS	Palomba[8]	2007	Brazil	Mixed cardiac (603)		AUC=0.70
AKICS	Palomba[8]	2007	Brazil	Mixed cardiac (215)	HL χ2 good (p=0.24)	AUC=0.79
Prolonged Mechanical Ventilation						
SAPS-II	Kern[31]	2001	Germany	Mixed cardiac (687)		AUC=0.90
APACHE-II	Kern[31]	2001	Germany	Mixed cardiac (687)		AUC=0.88
ICURS	Serrano[33]	2005	Spain	CABG (569)	HL χ^2= 12.1 (p=0.10)	AUC=0.68

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, ICURS – Intensive Care Unit Risk Stratification Score, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery

CABG, Coronary Artery Bypass Graft, HL – Hosmer Lemeshow, AUC - Area under the receiver operating characteristic curve
*if calculated on multiple days the value on the day of the best AUC is shown
‡ p values not supplied

2.5.1. Models developed for general ICU and validated in cardiac surgery patients

Acute Physiology and Chronic Health Evaluation II (APACHE II)

In 1985 Knaus et al. developed the APACHE II score[18] from the original APACHE score.[44] APACHE-II estimates the risk of mortality for ICU patients using data including patient age, co-morbidity and an Acute Physiologic Score (APS) based on the most abnormal values of 12 physiological variables recorded during the first 24 hours of ICU admission. Cardiac surgery patients were excluded from the model's development dataset.

A 2001 study by Kern et al. demonstrated that APACHE-II discriminated well when predicting prolonged mechanical ventilation in 687 cardiac surgery patients.[31] In 2005 Hekmat et al. demonstrated that APACHE-II scores calculated daily for 1057 cardiac surgery patients performed well, with postoperative day 3 scores best predicting 30 day mortality.[25] In 2011, Doerr et al. conducted similar analyses using the records of 2801 cardiac surgery patients.[37] When predicting ICU mortality APACHE-II showed adequate discrimination for each postoperative day but calibration was only adequate on the first postoperative day. Mean and worst APACHE-II scores for each patient were also used to generate mortality predictions with the mean APACHE-II score demonstrating the best discrimination and calibration.

Exarchopoulos and colleagues demonstrated that APACHE-II scores at ICU admission successfully predicted 30 day mortality in 150 cardiac surgery patients.[42] Similarly Tsaousi et al. demonstrated that ICU admission APACHE-II score successfully predicted in-hospital mortality in 1058 cardiac surgery patients.[43] However, in a UK study, Ariyaratnam et al. found admission APACHE-II scores predicted perioperative mortality poorly.[3]

Acute Physiology and Chronic Health Evaluation III (APACHE-III)

APACHE-III was developed using data from 17,440 patients from 40 hospitals.[19] The same physiological variables included in APACHE-II were measured in the first 24 hours of admission, together with urine output and four additional blood analyses. The final model included 17 physiological variables which combined to create the APS. Compared with APACHE-II, APACHE-III assigns greater weight to extremely abnormal values. The APS is combined with chronic disease status and age to produce the final APACHE-III score. As with APACHE-II, cardiac surgery patients were not included in the development cohort.

A model including APS from APACHE-III, patient information and surgery type was validated in 2435 coronary artery bypass graft (CABG) patients.[30] This discriminated well when predicting

hospital mortality for groups of patients, but in individuals the APS scores correlated poorly with mortality, length of ICU stay and treatment costs.

Simplified Acute Physiology Score II (SAPS-II)

The SAPS-II model was developed using data from 137 centres across 12 countries over a six month period in 1991-1992.[20] SAPS-II was designed for general ICUs. Cardiac surgery patients were excluded. Similarly to the APACHE scores, this model also used the worst recorded value for each variable during the first 24 hours of admission.

The ability of daily SAPS-II scores to predict 30-day mortality after cardiac surgery was also assessed in Doerr's 2011 study.[37] Discrimination was found to be good but the model was poorly calibrated in this group of patients. Derived variables such as maximum and mean SAPS–II score showed excellent discrimination and calibration. The same author subsequently analysed mortality predictions for 5207 cardiac surgery patients (including the initial 2801). The calibration of daily SAPS-II scores was inadequate, but again discrimination was acceptable.[41]

Kern et al. also assessed the ability SAPS-II to predict prolonged mechanical ventilation after cardiac surgery, reporting good discrimination but without commenting on calibration.[31] Exarchopoulos et al. found that admission SAPS-II score performed well when predicting 30 day mortality in a study of 150 cardiac surgery patients.[42]

Multiple Organ Dysfunction Score (MODS)

In 1995 Marshall et al. described the MODS as a tool to grade the severity of organ dysfunction in patients admitted to a Canadian surgical ICU between 1988 and 1990.[21]   The score was developed in order to measure patients' progress on a daily basis during ICU stay and used data from 336 patients to grade dysfunction in 6 major organ systems.

In 2005 and 2010 Hekmat et al. published validation studies in which MODS was calculated daily in two cohorts of 384 and 1057 cardiac surgery patients.[25, 36] MODS had good discriminatory abilities with some variation depending on the day on which the score was calculated. Calibration was reported as acceptable, although p values for the HL χ^2 test were not supplied.

The (Sepsis-Related) Sequential Organ Failure Assessment Score (SOFA)

SOFA score was developed in 1996 to standardise the assessment of a patient's progress on the ICU during a septic episode.[22] Designed by an expert committee, it grades the dysfunction of each organ system depending on the most abnormal value recorded for parameters chosen to represent those systems. Daily scores for each organ system can be compared separately with previous values or combined into a total score to reflect the overall patient progress.

In 2003 a team from Italy calculated SOFA scores for the first 10 postoperative days in cardiac surgery patients who stayed more than 96 hours in ICU. [32] The worst daily SOFA score, the maximum SOFA score possible when combining the worst scores for each organ system (regardless of the days on which they were recorded) and the difference between these two values and the first day SOFA score were calculated. All four derivatives of the SOFA score demonstrated good discrimination with the worst daily score demonstrating the best performance. In 2006 Patila et al. prospectively calculated the SOFA score for 857 cardiac surgery patients.[34] The maximum SOFA score during the first 3 days demonstrated acceptable discrimination for mortality with the overall maximum postoperative SOFA performing slightly better. A 2007 study analysed the association between the day 1 SOFA score and hospital mortality for 1458 cardiac surgery patients and found that the score had acceptable discrimination.[35]

SOFA scores calculated on each of the first six postoperative days, as well as mean and maximum SOFA scores showed good calibration and discrimination in Doerr's study.[37] In a subsequent analysis of the same data, predictions for 30 day mortality made using daily SOFA scores, the maximum SOFA score and the mean of all SOFA scores recorded throughout ICU admission were compared with predictions made using the mean of all daily SOFA scores up to that point.[15] Daily SOFA scores and their derivatives all demonstrated good discrimination.

In Exarchopoulos' validation study, the SOFA score demonstrated acceptable discrimination and calibration when predicting 30 day mortality.[42] Tsaousi et al. studied the accuracy of in-hospital mortality predictions made using day one SOFA scores, maximum and mean SOFA scores and the difference between maximum SOFA and the daily SOFA score. Day one SOFA demonstrated good discrimination but was outperformed by the other SOFA derivatives.[43]

Logistic Organ Dysfunction Score (LODS)

The LODS was developed by Le Gall et al. in 1996.[23] It aimed to predict hospital mortality using a subset of the same database used to develop the SAPS-II score. The LODS uses the worst values recorded during the first 24 hours of ICU admission for 12 variables. Cardiac surgery patients were again excluded. In 2011 Heldwein et al. showed that daily LODS scores could be used to predict mortality in cardiac surgery patients,[38] with the best discrimination observed on the third postoperative day.

Simplified Acute Physiology Score 3 (SAPS-3)

The SAPS-3 score was developed using data from 21336 patients from 309 ICUs across 35 countries [24, 45] including 1657 cardiac surgery patients. Variables were selected using a combination of expert opinion and regression modelling. They included existing measures for the classification of illness and physiological instability measured within the first hour of ICU admission. The model is formed of 20 variables, including those reflecting the geographical location of the institution in which it is being used. The total SAPS-3 score is reduced by 6 points for cardiac surgery patients to reflect the greater use of vasoactive drugs and the frequency of abnormal postoperative physiology in these patients. In 2014 Doerr et al. compared SAPS-3 with SAPS-II in 5207 cardiac surgery patients.[41] They calculated the scores on the first six postoperative days and found that SAPS-3 outperformed SAPS-II but was not adequately calibrated when predicting ICU mortality.

Intensive Care National Audit and Research Centre model (ICNARC)

In 2007 Harrison et al. published the ICNARC model,[27] developed using data from 216,626 patients admitted to 163 general ICUs in the UK between 1995 and 2003. The score includes the worst values for 12 variables, six of which were physiological. Cardiac surgery patients were included in the development cohort. In 2015 Ariyaratnam et al. validated the ICNARC model on 1646 cardiac surgery patients in a UK centre and found that it performed well in terms of discrimination and calibration.[3]

2.5.2. Models designed specifically for cardiac surgery

Intensive Care Unit Risk Stratification Score (ICURS)

In 1997 Higgins et al. produced the ICURS based on pre-, intra- and postoperative data recorded on admission to ICU after cardiac surgery for 2440 patients.[16]   Separate logistic regression models to predict in-hospital mortality and composite morbidity (defined in terms of specific measures of organ dysfunction) were developed. Eight variables were included in the mortality model and 13 in the morbidity model.

ICURS discriminated well in prospective validation sets, and calibration was reported as good. In 2005 Serrano validated ICURS' ability to predict the duration of mechanical ventilation. ICURS performed best when predicting ventilation lasting more than 48 hours, but discrimination was below the acceptable threshold.[33] In 2006 Biagioli et al. studied the predictions generated by an ICURS model developed using Higgin's methods in their own development cohort. In a

separate validation group of 350 cardiac surgery patients this customised model performed poorly.[26] In 2007 Palomba et al. used the ICURS scores of 603 cardiac surgery patients to predict the development of mild AKI with acceptable discrimination.[8]

Cardiac Surgery Score (CASUS)

The Cardiac Surgery Score (CASUS) was developed by Hekmat et al. in 2005 to produce daily 30 day mortality estimates for cardiac surgery patients.[25] The development dataset included 384 patients who underwent cardiac surgery requiring cardiopulmonary bypass followed by admission for >24 hours to ICU. The model based predictions on the most abnormal daily values of 10 variables.

The score was validated in two groups of 1057 and 1104 patients and performed consistently well. In 2010, a subsequent validation using data from 3801 patients, which included the 1104 from the 2005 paper, revealed good discrimination and calibration. CASUS performed best on day 1 and worst on day 5.[36]

Daily CASUS scores, together with mean and maximum CASUS scores, were validated for 30 day mortality prediction at a different German centre in 2011 and were found to perform consistently well over the first six postoperative days.[37] Maximum and mean CASUS scores demonstrated superior discrimination and satisfactory calibration. The same data were used to show that CASUS outperformed SOFA in ICU mortality prediction.[39] The same year a further comparison of CASUS with the new logistic CASUS based on 4054 patients (including the 2801 previously analysed in other studies) was performed.[40] Although discrimination was good, calibration was found to be poor. CASUS was validated in the Exarchopoulos study and demonstrated good discrimination and calibration on the first postoperative day.[42] Log-CASUS[40] and Rapid Clinical Evaluation (RACE)[46], both based on CASUS, performed well in development sets but are yet to be validated themselves.

Biagioli Model

In 2006 Biagioli et al. produced a risk model for cardiac surgery using a Bayes linear approach.[26] The authors trained their model to predict morbidity using data for a range of predictor variables taken from a group of 740 patients undergoing CABG surgery. The final model included pre- and intraoperative data combined with white cell count and oxygen delivery index measured within 3 hours of ICU admission. In a validation set of 350 patients, the model had good discrimination and calibration and outperformed models created using logistic regression.[26]

Salamonsen Model

In 2008 Salamonsen et al. produced a risk model designed to predict which patients undergoing CABG would not be ready for discharge from ICU within their "fast-track" schedule (<12 hours).[28] Pre-, intra- and postoperative variables were used to develop a multiple linear regression model to predict length of stay on the ICU. The model was validated in 117 patients. The R^2 value for the validation set was poor and the 95% confidence intervals for predicted lengths of stay of 4 and 12 hours spanned 29 and 70 hours respectively. Consequently, the authors concluded that their model was not useful.

Meyfroidt Model

In 2011 Meyfroidt *et al.* collected a range of admission, medication, laboratory and physiological data from the first 4 hours of ICU admission for 461 cardiac surgery patients. They used these data to train Gaussian process models to perform two tasks:[29] (i) a classification task to predict whether patients would be discharged from ICU on day 2, (ii), a regression task designed to predict the actual day of ICU discharge. Data for five physiological variables were averaged across 40 minute segments and these averaged values were included in the final model. The models were tested on a validation cohort of 499 patients and were able to adequately identify patients likely to be discharged on day 2 but were less successful when predicting the day of discharge.

Acute Kidney Injury after Cardiac Surgery (AKICS) model

In 2007 Palomba et al. developed and validated a model to predict mild AKI in patients following cardiac surgery.[8] The model was based on eight variables, two of which were postoperative physiological variables. It performed well when validated in 215 patients.

2.6. Discussion

This systematic review has identified 14 validated risk models that utilise postoperative patient monitoring data to predict outcomes after adult cardiac surgery. The most commonly validated predictions were for mortality, but the prediction of composite morbidity, ICU length of stay, and specific morbidity outcomes were also tested. Of the fourteen models, eight were developed on non-cardiac surgery patients but have subsequently been validated in cardiac surgery and six were developed specifically for cardiac surgery.

Postoperative risk prediction models may be useful for performing three main tasks after cardiac surgery. The first is resource allocation. Future operating lists and staffing levels may be adjusted according to the predicted length of stay or mortality rates (used as a surrogate for severity of

illness) of patients present on the ICU. Secondly, for benchmarking institutional performance where risk estimates can be used to generate standardised predictions for mortality rates against which observed outcomes can be measured. Finally, with caution, risk models may be used to inform clinical decision making and discussions with patients and their relatives.

The models identified estimate the risk of adverse outcomes for groups of patients with similar scores. They state the proportion of a group of patients with similar risk scores that would be expected to suffer the outcome. This information may provide a context to clinical decision making and prognostic discussions. Moreover, changes in the predicted risk over time or in response to treatment may give an indication of a patient's progress. However, it should be acknowledged that the scores cannot identify whether or not an individual patient will suffer the outcome.

The majority of models with good discrimination and calibration identified in this review are those which calculate 30-day mortality risk daily based upon the worst value for each parameter in each 24 hour period. Although models which predict mortality are potentially useful for benchmarking and resource allocation they are of limited use in guiding real-time treatment decisions. The prediction of specific complications or patient deterioration after cardiac surgery would be much more relevant to the treating clinicians. Such an approach would allow targeted treatment to prevent or reduce the impact of these developing complications.[24] Our review identified only the AKICS score as being capable of predicting acute kidney injury while APACHE-II and SAPS-II successfully predicted prolonged ventilation. Secondly, these scores are calculated retrospectively once the worst values in a 24 hour period are known; by the time increased risk is detected the complication may be established.[18-23,25,27] Derivative scores such as the mean or maximum value for validated scores over a number of days show even better predictive power.[15, 34, 37] However, due to their retrospective nature these scores also have little value in the day to day treatment of patients. Importantly, serial scores and their aforementioned derivatives are not independent of the quality of care provided by the ICU; poor care will lead to poor scores. Serial scores should not therefore be used to produce mortality predictions against which observed mortality rates are measured when benchmarking ICU performance. However, trends in serial risk scores could be compared to identify institutions where predicted mortality increases during specific periods of the postoperative stay. For example, it may be found that an institution's predicted mortality risk increases around day 3 or 4 in an institution which does not have robust procedures in place to allow early detection and effective management of sepsis.

A number of models provide a snapshot of risk using data obtained within the first four hours of ICU admission following cardiac surgery.[8, 16, 24, 26, 27, 29] This may be the most appropriate

time to estimate risk for the purposes of benchmarking ICU performance. However, these models cannot reliably guide resource allocation or clinical decision making after the initial period on ICU as their predictions may become inaccurate as postoperative events unfold. Some authors validated these models as daily assessment tools to be calculated using the worst scores for each 24 hours with acceptable statistical performance.[25,37] While statistical performance may be good, as with scores designed for serial use, the predictions are obtained too late to influence patient management and the effect of the quality of ICU care on the scores themselves precludes their use for ICU benchmarking.

Models that would be of most benefit in clinical decision making would utilise up to date clinical information and provide continuously updated predictions, however none of the models identified utilises real-time patient monitoring data. The majority of identified models require the most abnormal value for each parameter over a given period and categorise continuous variables according to the degree of abnormality.   This approach sacrifices predictive accuracy to improve the ease of use and minimise the need for computing power. With recent developments in computing more ambitious approaches may be possible. The model developed by Meyfroidt utilising Gaussian processes does use computerised analyses of a large number of data points.[29] However, even this model analysed average values calculated for 40 minute periods rather than continuous data.

This review has demonstrated that models developed for use in general ICU patients such as the SOFA, SAPS-II and APACHE-II scores may be applied successfully to cardiac surgery patients.[25, 26, 30, 32, 34-43].   This is despite their developers' excluding cardiac surgery patients from development datasets due to their low observed mortality when compared with other patient groups with similar levels of physiological derangement.[18]

Despite the good performance of general ICU models, there may be advantages to using cardiac surgery specific scores. Firstly, a number of risk factors included in general ICU models such as metastatic cancer and liver cirrhosis are largely irrelevant, as they usually contraindicate cardiac surgery. In addition, there are many significant differences in care protocols between cardiac and general ICU's. For example, the CASUS developers noted that the conscious level of patients is routinely decreased in the early postoperative period secondary to sedation. Therefore, they introduced a 'neurological state score' which was quicker and easier to calculate than the Glasgow Coma Scale and decreased the impact of appropriately low conscious level on risk estimates. They also recognised the need to correct risk scores for artificially normal physiological values which are only present as a direct consequence of supportive treatments frequently used following cardiac surgery, such as mechanical cardiovascular support or renal replacement

therapy.[25] As a result, despite general ICU models demonstrating good statistical performance, a cardiac surgery specific model may be preferred by clinicians.

There are however, a number of key limitations of the cardiac surgery specific models identified in this review which are likely to explain their limited adoption. First, unlike widely used pre-operative cardiac surgery risk models [1, 2] and the models developed for general ICUs, most cardiac surgery models have been based on data from single centres. This approach optimises data quality and completeness for model development but may lead to concerns about the application of models to different populations. For example, the Biagioli, ICURS, Meyfroidt and AKICS models require cardiac output measurement using a Swann-Ganz catheter which is not routinely used in all cardiac surgery centres.[47] The Meyfroidt model also contains variables derived from entropy measurements. These values describe the variation within a patient's physiological data, but monitoring equipment capable of producing these values may not be available in all ICUs. Similarly, when initiation of specific treatments e.g. intra-aortic balloon counterpulsation, are used as surrogates for severity of physiological derangement, local practices can affect the validity of these surrogate variables. The cardiovascular component of the SOFA score is based on the administration of vasoactive medication using specific protocols (such as dopamine being administered before noradrenaline to treat hypotension). In many centres clinicians will know that these patterns of drug administration are not followed and this may lead to diminished confidence in the SOFA score despite reports of good performance in multiple studies.[15, 32, 34, 35, 36, 39-43]

2.7. Conclusion

Risk prediction models based on preoperative data have real value when advising patients on their decision whether to undergo surgery and when assessing the performance of cardiothoracic units. However, postoperative models identified in this review have the key advantage of being updated throughout a patient's admission. If they are used to produce risk estimates at the time of admission to ICU, they may be used to assess the quality of the ICU care in isolation from the pre- and intraoperative events. Models which produce daily risk estimates deliver updated predictions which enable optimisation of resource allocation planning in cardiac surgery units. As described in this review, most of the models make predictions which are accurate enough to perform these two tasks. SOFA and CASUS are the most extensively validated scores and use readily available postoperative variables to produce their risk estimates. This combination of ease of calculation and accuracy defines them as the most appropriate postoperative scores identified

in this study. Their discriminatory power is beyond that displayed by preoperative scores such as EuroSCORE and EuroSCORE II. [4, 48] With caution, these scores may also be used to inform discussions with patients and their relatives and provide a broad context for clinical decision making.

However, no existing model provides estimates for the risk of specific complications for individuals with sufficient accuracy and frequency to reliably guide specific clinical decisions. This is probably the main reason why such models have not achieved widespread adoption into clinical practice.

Technological developments have the potential to improve risk prediction after cardiac surgery. In future, computerised models designed to calculate risk much more frequently could provide contemporaneous risk estimates. The most useful models would predict specific complications early enough to allow clinicians time to intervene to prevent the complications occurring or, where that is not possible, reduce their impact. The ideal model would analyse physiological variables and not the current treatments, thus avoiding the pitfall of interinstitutional variation in management protocols. Variables could be selected from the huge amount of post-cardiac surgery data available on the ICU according to the specific outcome being predicted. The accuracy of such models may be improved by advances in computing which enable real-time analysis of raw monitoring data rather than categorical "worst values" recorded over a given time period. Analyses of changes in, rather than absolute values of, an individual's physiological variables may allow identification of those at increased risk of clinical deterioration before arbitrary thresholds for abnormality are reached and end organ damage occurs.

Search strategy details

Embase Search

("heart surg*" OR "cardi* surg*" OR coronary adj3 bypass OR "coronary graft" OR "CABG" OR (valv* adj3 (rep* OR surg*))).ti,ab OR *HEART SURGERY/ OR *CORONARY ARTERY BYPASS GRAFT/ OR *MITRAL VALVE REPLACEMENT/ OR *MITRAL ANNULOPLASTY/ OR *HEART TRANSPLANTATION/ OR *VALVULOPLASTY/ OR *CORONARY ARTERY BYPASS SURGERY/

AND

(morbidity OR mortality OR "renal failure" OR "renal replacement" OR "kidney injury" OR arrhythmia OR bleeding OR resternotomy OR "re-sternotomy" OR "respiratory failure" OR fail* adj3 extubation OR fibrillation OR death OR length of stay OR (renal AND replacement AND therapy) OR (prolonged adj3 ventilation) OR fibrillation).ti,ab OR SURGICAL MORTALITY/ OR KIDNEY FAILURE/ OR RENAL REPLACEMENT THERAPY/ OR REOPERATION/ OR POSTOPERATIVE COMPLICATION/ OR HEART TAMPONADE/ OR MORBIDITY/ OR LENGTH OF STAY/ OR DEATH/ OR HEART ARRHYTHMIA/ OR HEART ATRIUM FIBRILLATION/

AND

("intensive care" OR "critical care").ti,ab OR INTENSIVE CARE/

AND

(Predict* OR realtime OR "statistical model" OR "regression model" OR algorithm OR "risk stratification" OR "early identification").ti,ab OR CLINICAL DECISION MAKING/ OR DECISION SUPPORT SYSTEM/ OR MEDICAL DECISION MAKING/ OR COMPUTER SYSTEM/ OR PREDICTION AND FORECASTING/ OR *RISK ASSESSMENT/

Medline Search

("heart surg*" OR "cardi* surg*" OR "coronary artery bypass" OR "coronary bypass" OR "coronary graft" OR "CABG" OR (valv* adj3 (replac* OR repair OR surg*))).ti,ab OR exp *CARDIAC VALVE ANNULOPLASTY/ OR exp *CORONARY ARTERY BYPASS/ OR *CARDIAC SURGICAL PROCEDURES/ OR *HEART TRANSPLANTATION/ OR *HEART VALVE PROSTHESIS/

AND

(morbidity OR mortality OR("renal failure" OR "renal replacement" OR arrhythmia* OR bleeding OR resternotomy OR "re-sternotomy" OR "respiratory failure" OR fail* adj3 extubation OR death.ti,ab OR "kidney injury" OR prolonged adj3 ventilation OR fibrillation OR (failed AND extubation) OR "length of stay".ti,ab OR MORBIDITY/ OR MORTALITY/ OR HOSPITAL MORTALITY/ OR RENAL INSUFFICIENCY/ OR *ACUTE KIDNEY INJURY/ OR *RENAL REPLACEMENT THERAPY/ OR *REOPERATION/ OR *POSTOPERATIVE COMPLICATIONS/ OR *CARDIAC TAMPONADE/ OR *RESPIRATORY INSUFFICIENCY/ OR *DEATH OR *ARRHYTHMIAS, CARDIAC/ OR *ATRIAL FIBRILLATION/ OR *ATRIAL FLUTTER/ OR RENAL REPLACEMENT THERAPY/ OR RENAL DIALYSIS/ or HEMOFILTRATION/ or TRACHEOSTOMY/ OR LENGTH OF STAY/

AND

("intensive care OR "critical care")ti,ab OR *CRITICAL CARE/ OR *INTENSIVE CARE/

AND

(Predict* OR realtime OR "statistical model" OR "regression model" OR algorithm OR "risk stratification" OR "early identification").ti,ab OR *DECISION MAKING, COMPUTER-ASSISTED/ OR *DECISION SUPPORT SYSTEMS, CLINICAL/ OR *COMPUTER SYSTEMS/

Cochrane Library Search

"Cardi* surg*" OR CABG OR "Coronary Artery Bypass" OR "Heart surg*" OR "coronary graft" OR "Coronary bypass" OR Valv* adj3 (replac* or repair or surg*) OR (MeSH descriptor: [Coronary Artery Bypass] explode all trees) OR (MeSH descriptor: [Thoracic Surgery] explode all trees) OR (MeSH descriptor: [Cardiac Valve Annuloplasty] explode all trees) OR (MeSH descriptor: [Heart Valve Prosthesis Implantation] explode all trees) OR (MeSH descriptor: [Cardiac Surgical Procedures] explode all trees)

AND

morbidity OR mortality OR "renal failure" OR "renal replacement" OR arrythmia OR bleeding OR resternotomy OR "re-sternotomy" OR "respiratory failure" OR fail* adj extubation OR "kidney injury" OR death OR "length of stay" OR prolonged adj3 ventilation OR (MeSH descriptor: [Morbidity] explode all trees) OR (MeSH descriptor: [Mortality] explode all trees)OR (MeSH descriptor: [Renal Insufficiency] explode all trees) OR (MeSH descriptor: [Acute Kidney Injury] explode all trees)OR (MeSH descriptor: [Renal Replacement Therapy] explode all trees)OR (MeSH descriptor: [Postoperative Complications] explode all trees)OR (MeSH descriptor: [Reoperation]

explode all trees)OR (MeSH descriptor: [Respiratory Insufficiency] explode all trees) OR (MeSH descriptor: [Cardiac Tamponade] explode all trees) OR (MeSH descriptor: [Death] explode all trees) OR (MeSH descriptor: [Arrhythmias, Cardiac] explode all trees) OR(MeSH descriptor: [Tracheostomy] explode all trees) OR (MeSH descriptor: [Length of Stay] explode all trees)

AND

realtime OR "statistical model" OR "regression model" OR algorithm OR "risk prediction" OR "risk stratification" OR "early identification" OR (MeSH descriptor: [Decision Making, Computer-Assisted] explode all trees) OR (MeSH descriptor: [Decision Support Techniques] explode all trees)

AND

"intensive care" OR "critical care" OR (MeSH descriptor: [Critical Care] explode all trees)

2.9. References

1.   Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16(1):9-13.
2.   Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardiothorac Surg* 2012;41(4):734-744; discussion 744-735.
3.   Ariyaratnam P, Loubani M, Biddulph J, Moore J, Richards N, Chaudhry M, et al. Validation of the intensive care national audit and research centre scoring system in a UK adult cardiac surgery population. *Journal of Cardiothoracic and Vascular Anesthesia* 2015;29(3):565-569.
4.   Grant SW, Hickey GL, Dimarakis I, Trivedi U, Bryan A, Treasure T, et al. How does EuroSCORE II perform in UK cardiac surgery; an analysis of 23 740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. *Heart (British Cardiac Society)* 2012;98(21):1568-1572.
5.   Hosseinian L, Chiang Y, Itagaki S, Polanco A, Rhee A, Chikwe J. Earlier versus later tracheostomy in patients with respiratory failure after cardiac surgery in the United States. *J Cardiothorac Vasc Anesth* 2014;28(3):488-492.
6.   Bailey ML, Richter SM, Mullany DV, Tesar PJ, Fraser JF. Risk factors and survival in patients with respiratory failure after cardiac operations. *The Annals of thoracic surgery* 2011;92(5):1573-1579.
7.   Thakar CV, Worley S, Arrigain S, Yared J-P, Paganini EP. Influence of renal dysfunction on mortality after cardiac surgery: Modifying effect of preoperative renal function. *Kidney Int* 2005;67(3):1112-1119.
8.   Palomba H, de Castro I, Neto AL, Lage S, Yu L. Acute kidney injury prediction following elective cardiac surgery: AKICS Score. *Kidney Int* 2007;72(5):624-631.
9.   LaPar DJ, Gillen JR, Crosby IK, Sawyer RG, Lau CL, Kron IL, et al. Predictors of operative mortality in cardiac surgical patients with prolonged intensive care unit duration. *Journal of the American College of Surgeons* 2013;216(6):1116-1123.
10.   Vohra HA, Goldsmith IR, Rosin MD, Briffa NP, Patel RL. The predictors and outcome of recidivism in cardiac ICUs. *European Journal of Cardio-Thoracic Surgery* 2005;27(3):508-511.
11.   Grothusen C, Attmann T, Friedrich C, Freitag-Wolf S, Haake N, Cremer J, et al. Predictors for long-term outcome and quality of life of patients after cardiac surgery with prolonged intensive care unit stay. *Interventional Medicine and Applied Science* 2013;5(1):3-9.
12.   Morgan RJM WF, Wright MM. . An early warning scoring system for detecting developing critical illness. *Clin Intensive Care* 1997(8):100.
13.   Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143(1):29-36.
14.   Angus DC, Clermont G, Kramer DJ, Linde-Zwirble WT, Pinsky MR. Short-term and long-term outcome prediction with the Acute Physiology and Chronic Health Evaluation II system after orthotopic liver transplantation. *Critical Care Medicine* 2000;28(1):150-156.
15.   Badreldin A, Elsobky S, Lehmann T, Brehm BB, Doenst T, Hekmat K. Daily-mean-SOFA, a new derivative to increase accuracy of mortality prediction in cardiac surgical intensive care units. *Thoracic and cardiovascular surgeon* 2012;60(1):43-50.
16.   Higgins TL, Estafanous FG, Loop FD, Beck GJ, Lee JC, Starr NJ, et al. ICU admission score for predicting morbidity and mortality risk after coronary artery bypass grafting. *Annals of Thoracic Surgery* 1997;64(4):1050-1058.
17.   Hosmer DW, Lemeshow S. Applied logistic regression: Wiley; 1989.
18.   Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13(10):818-829.
19.   Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. THe apache iii prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100(6):1619-1636.
20.   Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *Jama* 1993;270(24):2957-2963.

21.   Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Crit Care Med* 1995;23(10):1638-1652.

22.   Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22(7):707-710.

23.   Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, et al. The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. *Jama* 1996;276(10):802-810.

24.   Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005;31(10):1345-1355.

25.   Hekmat K, Kroener A, Stuetzer H, Schwinger RH, Kampe S, Bennink GB, et al. Daily assessment of organ dysfunction and survival in intensive care unit cardiac surgical patients. *Annals of Thoracic Surgery* 2005;79(5):1555-1562.

26.   Biagioli B, Scolletta S, Cevenini G, Barbini E, Giomarelli P, Barbini P. A multivariate Bayesian model for assessing morbidity after coronary artery surgery. *Critical Care* 2006;10(3).

27.   Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model*. *Critical Care Medicine* 2007;35(4):1091-1098.

28.   Salamonsen MR, Bailey MJ, Salamonsen RF. Can ICU stay be predicted accurately enough for fast-tracking cardiac surgical patients? *Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine* 2008;10(3):194-201.

29.   Meyfroidt G, Guiza F, Cottem D, De W, Van K, Aerts JM, et al. Computerized prediction of intensive care unit discharge after cardiac surgery: development and validation of a Gaussian processes model. *BMC Medical Informatics & Decision Making* 2011;11.

30.   Becker RB, Zimmerman JE, Knaus WA, Wagner DP, Seneff MG, Draper EA, et al. The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery by-pass surgery. *J Cardiovasc Surg (Torino)* 1995;36(1):1-11.

31.   Kern H, Redlich U, Hotz H, Von C, Grosse J, Konertz W, et al. Risk factors for prolonged ventilation after cardiac surgery using APACHE II, SAPS II, and TISS: Comparison of three different models. *Intensive care medicine* 2001;27(2):407-415.

32.   Ceriani R, Mazzoni M, Bortone F, Gandini S, Solinas C, Susini G, et al. Application of the sequential organ failure assessment score to cardiac surgical patients. *Chest* 2003;123(4):1229-1239.

33.   Serrano N, Garcia C, Villegas J, Huidobro S, Henry CC, Santacreu R, et al. Prolonged intubation rates after coronary artery bypass surgery and ICU risk stratification score. *Chest* 2005;128(2):595-601.

34.   Patila T, Kukkonen S, Vento A, Pettila V, Suojaranta-Ylinen R. Relation of the Sequential Organ Failure Assessment Score to Morbidity and Mortality After Cardiac Surgery. *Annals of Thoracic Surgery* 2006;82(6):2072-2078.

35.   Gomes RV, Tura B, Mendonca HTH, Almeida LAL, Rouge A, Matos PMP, et al. A first postoperative day predictive score of mortality for cardiac surgery. *Annals of Thoracic & Cardiovascular Surgery* 2007;13(3):159-164.

36.   Hekmat K, Doerr F, Kroener A, Heldwein M, Bossert T, Badreldin AM, et al. Prediction of mortality in intensive care unit cardiac surgical patients. *European Journal of Cardio-Thoracic Surgery* 2010;38(1):104-109.

37.   Doerr F, Badreldin AM, Heldwein MB, Bossert T, Richter M, Lehmann T, et al. A comparative study of four intensive care outcome prediction models in cardiac surgery patients. *Journal of cardiothoracic surgery* 2011;6.

38.   Heldwein MB, Badreldin AM, Doerr F, Lehmann T, Bayer O, Doenst T, et al. Logistic Organ Dysfunction Score (LODS): a reliable postoperative risk management score also in cardiac surgical patients? *Journal of cardiothoracic surgery* 2011;6.

39.   Badreldin AM, Doerr F, Ismail MM, Heldwein MB, Lehmann T, Bayer O, et al. Comparison between Sequential Organ Failure Assessment score (SOFA) and Cardiac Surgery Score (CASUS) for mortality prediction after cardiac surgery. *Thoracic & Cardiovascular Surgeon* 2012;60(1):35-42.

40.   Doerr F, Badreldin AM, Bender EM, Heldwein MB, Lehmann T, Bayer O, et al. Outcome prediction in cardiac surgery: the first logistic scoring model for cardiac surgical intensive care patients. *Minerva anestesiologica* 2012;78(8):879-886.

41.   Doerr F, Badreldin AM, Can F, Bayer O, Wahlers T, Hekmat K. SAPS 3 is not superior to SAPS 2 in cardiac surgery patients. *Scandinavian cardiovascular journal : SCJ* 2014;48(2):111-119.

42.   Exarchopoulos T, Charitidou E, Dedeilias P, Charitos C, Routsi C. Scoring Systems for Outcome Prediction in a Cardiac Surgical Intensive Care Unit: A Comparative Study. *American Journal of Critical Care* 2015;24(4):327-334.

43.   Tsaousi GG, Pitsis AA, Ioannidis GD, Pourzitaki CK, Yannacou-Peftoulidou MN, Vasilakos DG. Implementation of EuroSCORE II as an adjunct to APACHE II model and SOFA score, for refining the prognostic accuracy in cardiac surgical patients. *Journal of Cardiovascular Surgery* 2015;56(6):919-927.

44.   Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981;9(8):591-597.

45.   Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med* 2005;31(10):1336-1344.

46.   Badreldin AM, Doerr F, Bender EM, Bayer O, Brehm BR, Wahlers T, et al. Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices. *European Journal of Cardio-Thoracic Surgery* 2013;44(6):992-997; discussion 997-998.

47.   Bignami E, Belletti A, Moliterni P, Frati E, Guarnieri M, Tritapepe L. Clinical practice in perioperative monitoring in adult cardiac surgery: is there a standard of care? Results from an national survey. *Journal of Clinical Monitoring and Computing* 2015:1-19.23.

48.   Chalmers J, Pullan M, Fabri B, McShane J, Shaw M, Mediratta N, et al. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. *Eur J Cardiothorac Surg* 2013;43(4):688-694.

Chapter Three: Summary of introduction and thesis aims

3.1. Summary of introduction

The first section of this thesis has outlined the current state of postoperative risk prediction following cardiac surgery.   The systematic review identified models which may be useful when making decisions concerning resource allocation and when benchmarking the performance of critical care units following cardiac surgery. Some of the models identified could also be useful when discussing patient progress in terms of change in mortality risk estimate.   The review identified that while models and scoring systems designed for use in general ICU populations perform adequately in cardiac surgery patients, models designed specifically for cardiac surgery patients may perform slightly better. Two models encountered during the systematic review, the Rapid Clinical Evaluation (RACE) Score and the logistic Cardiac Surgery Risk score (logCASUS) have been developed from the original CASUS score. Unlike the original CASUS score neither of these models has been externally validated.

The main limitation of the models identified during the systematic review is the clinical usefulness of the predictions they make.   All identified models predict risk for groups of patients with similar risk scores rather than for individual patients. Within a group of 100 patients with a mortality risk of 5%, five would be expected to die, but the model cannot distinguish between the 95 survivors and the five who will die. The frequency of the risk estimates was also a limitation with most models requiring data collected over a 24-hour period. None of the models identified provided up to date estimates of the risk of a specific complication in a manner which might allow intervention to prevent the complication occurring or reduce its severity.

3.2. Thesis aims and questions

3.2.1. Aims

The overarching goal of this thesis is to advance knowledge in the field of risk prediction modelling following cardiac surgery. This goal will be achieved by accomplishing the following three aims:

1) To collate and clean data from various sources to create a large, reliable dataset of "real world data" for patients who have undergone cardiac surgery at Wythenshawe Hospital.

2) To validate existing risk prediction models and risk stratification tools in UK cardiac surgery patients. This will include validating models identified in the systematic review which have previously not been validated in UK cardiac surgery patients. The thesis will also validate the use of other risk stratification tools designed for use in the general hospital population specifically in cardiac surgery patients. Assessment of existing models will allow greater understanding of their strengths and weaknesses which will inform the future development of novel risk prediction models. The risk prediction tools to be investigated will be defined in the research questions outlined in the following subsection.

3) To develop risk prediction models which predict complications in a manner which would allow intervention to prevent their occurrence or reduce the harm the complications cause. Methods used should be able to run in real-time identifying when risk of adverse outcomes increases in order to provide clinical useful warnings to clinicians.

Using data collected from cardiac surgery patients from Wythenshawe Hospital (part of Manchester University Hospitals Foundation Trust) I aim to answer the following questions during this thesis.

1) How well do the logCASUS, RACE and SOFA scores predict ICU-mortality following adult cardiac surgery in a tertiary cardiothoracic centre in the UK?

2) What is the incidence of sepsis as defined by the Sepsis-3 criteria in patients who have undergone adult cardiac surgery at a tertiary cardiothoracic centre in the UK? Are the Sepsis-3 criteria useful in the stratification of risk of adverse outcomes in patients who have undergone non-transplant adult cardiac surgery in a tertiary cardiothoracic centre in the UK?

3) Are outcomes different for patients who are diagnosed with the same stage of acute kidney injury using different criteria within the Kidney Disease Improving Global Outcomes Acute Kidney Injury guidelines following adult in a tertiary cardiothoracic centre in the UK?

4) Is it possible to analyse postoperative urine output data to identify patients at risk of poor outcomes related to renal dysfunction following adult cardiac surgery in a tertiary cardiothoracic centre in the UK?

5) Are serum potassium and magnesium concentrations relevant to the prediction of new onset atrial fibrillation following adult cardiac surgery in a tertiary cardiothoracic centre in the UK?

SECTION TWO: METHODS

<u>Introduction</u>

The data used throughout this programme of research was collected from patients who had undergone cardiac surgery at Wythenshawe Hospital (part of Manchester University Hospitals NHS Trust). All data were collected as part of the Vascular Governance NorthWest (VGNW) database which is based at Manchester University Hospitals NHS Trust. Amendments to the ethical approvals previously granted to the VGNW project were submitted to and approved by the Nation Research Ethics Service at Haydock.

Data were collected, collated and rigorously cleaned by SHH using the R Studio statistical software package.(132) Collection of the data used in this project is described in chapter four. Algorithms were developed which analysed the cleaned data to identify outcomes which were identified as relevant by the Patient and Public Involvement group convened to guide the project. Details of the methods used during the production of final dataset are described in chapter five. Cleaned, anonymised data from the VGNW database were released to research partners at Durham University.

The statistical methods used in this programme of research are discussed in chapter 6. Statisticians from Durham University guided SHH's statistical analyses for some of the manuscripts and performed the statistical modelling described in chapter 10 where more complex Bayesian methodology was employed. Specific contributions are detailed on the first page of each chapter that is based on a journal article.

Chapter Four: Project Design and Data collection

4.1. Considerations regarding project design

This programme of research was based on retrospective analysis of data routinely recorded for patients who underwent cardiac surgery at Wythenshawe hospital between January 2013 and November 2017. The use of this methodology has key advantages.

1) Data used in this research programme is "real world" data. Prospective studies such as randomised controlled trials often involve extra resources and record data specifically for the study. Conclusions drawn from such studies may not be reproducible when the same methodology is used outside the study itself where resources to record some variables are not available or data quality is not as high. Risk prediction tools validated or developed during this research programme will analyse data that are available as part of routine patient care. They will not therefore, suffer the decrease performance sometimes seen when models developed using data collected specifically for research are subsequently used to analyse real-world data.(133)

2) As the data were all recorded routinely, it is less likely that any selection bias was introduced. Patients did not have to undergo extra procedures or monitoring to participate in the study so patient motivation and health beliefs were unlikely to affect participation.

3) As data were recorded routinely, the gathering of data was relatively cheap both financially and in terms of time and resources. This allowed the collection of data on a larger number of patients than would have been possible if interventions were being performed or non-routinely collected variables were measured.

There were also disadvantages to this methodology.

1) Compared with randomised control trial methodology, the observational design of the programme's analyses limited the ability to control variables which may have potentially confounded the studies' results. Potential confounding variables can be adjusted for during statistical analyses but as inclusion criteria were relatively lax it is possible that some confounders were unaccounted for during this research programme. However, this risk was mitigated by the large number of patients included in the study which facilitated robust statistical adjustment to remove confounders.

2) The data recorded may be of lower quality because it was not recorded specifically for research purposes. All analysed data were recorded by clinicians whose main aim was to

provide a clinical record of the patient's treatment on CICU. Therefore, data quality is likely to be below that which could have been obtained using dedicated, prospective data collection for specifically for research purposes.

In summary, dedicated prospective data collection may result in higher quality data which may allow analyses of the research data to achieve greater accuracy. However, when transferred to "real-world" situations where data quality is lower, methods which performed well in the higher quality research data might fail. The aims of this thesis focus on the real world application of risk estimation. Therefore, the use of routinely gathered data was more appropriate.

4.2. Data sources

Data analysed for this thesis was obtained from five main sources at Wythenshawe Hospital:

1) The hospital's clinical governance database
2) The hospital's perfusion database
3) The hospital's pathology laboratory database
4) The Draeger Innovian electronic patient record (EPR) on CICU
5) The Draeger Infinity bedside physiological monitors on CICU

Details of data available from each source are detailed in Tables 4-1 to 4-10.

It was not possible to record data from the Draeger Infinity bedside physiological monitors until July 2016. Reasons for the delay in data capture are detailed in the Discussion section of this thesis.

4.2.1. Clinical governance database

The first source of data identified was the Dendrite clinical governance database. This database is used to provide data for the Society for Cardiac Surgery's cardiac surgery audit project. It contains demographic, pre-operative, operative and outcome data for all cardiac surgery patients (Table 4-1). While all variables shown in Table 4-1 were extracted from the Dendrite database, unfortunately the quality of data recorded in some fields was too poor for all of the data to be useful. Consequently data concerning variables such as height, weight, length of stay on CICU, Readmission to CICU, Time to Extubation, Reoperation were not extracted from the Dendrite database. Data regarding these variables were captured from the Innovian patient record instead.

Interrogation of the Dendrite database allowed the identification of all adult patients who had undergone any cardiac surgery after January 2013. As well as providing data summarised in Table

4-1, the Dendrite database also contained the patients' hospital numbers which were used to extract all relevant entries from the Draeger Innovian EPR.

Table 4-1 - Data obtained from the Dendrite Clinical Governance Database

Variable	Frequency recorded
Date of Birth	Once
Gender	Once
Admission Date	Once per admission
Operation Start	Once per operation
Operation End	Once per operation
Urgency	Once per operation
Procedure details	Once per operation
Hospital Discharge Date	Once per admission
Logistic EuroSCORE	Once per operation
Discharge status	Once per admission
Preoperative cardiac rhythm	Once per operation
Preoperative Renal Support	Once per operation

4.2.2. Perfusion database

The duration of CPB for each surgical procedure were obtained from the hospital's perfusion database. Height and weight data were recorded consistently well in this database as these variables are used by perfusionists to determine target flow rates during CPB. Therefore, height and weight data were also extracted from the perfusion database as shown in Table 4-2.

Table 4-2 - Data obtained from the perfusion database

Variable	Frequency recorded
CPB time	Once per operation
Height	Once per operation
Weight	Once per operation

4.2.3. Blood analyses databases

The pathology laboratory database

There were two sources of blood test results analysed in this programme. The first was the hospital's pathology laboratory database. As part of routine care, one set of blood samples are sent for testing in the pathology laboratory for each patient on each day of their ICU admission. The routine set of analyses requested for patients on CICU includes a full blood count and clotting screen performed by the haematology laboratory and urea and electrolyte concentrations

(including magnesium), liver function tests and C-reactive protein assay performed by the biochemistry laboratory (Table 4-3). Where indicated more frequent sampling was performed and similarly, for patients who were stable but had ongoing requirements for CICU care, the frequency of sampling was reduced. Results of preoperative blood tests performed at the preoperative clinic were also obtained from the pathology laboratory database. Rarely, in life threatening emergencies preoperative blood samples were not taken as the patient was taken directly to the operating theatre. Where possible the most recent preoperative and all postoperative values for blood analyses recorded during a patient's CICU admission were obtained.

Table 4-3 - Data obtained from the Pathology Laboratory database

Variable	Frequency recorded (routine care)
Sodium concentration	Daily
Potassium concentration	Daily
Urea concentration	Daily
Creatinine concentration	Daily
Albumin concentration	Daily
Bilirubin concentration	Daily
C-Reactive Protein concentration	Daily
Magnesium concentration	Daily
Haemoglobin concentration	Daily
White Cell Count	Daily
Platelet Count	Daily
Activated Partial Thromboplastin Time	Daily
Prothrombin Time	Daily
International Normalised Ratio	Daily
Fibrinogen Assay	Ad Hoc

Blood gas analyses

The second source of blood test results were the Gemstar point of care blood gas analysis machines on CICU. Blood gas analyses (Table 4-4) were performed at the point of care; the samples were not sent to the pathology laboratory. Blood gas analysis results were automatically transmitted to the Innovian EPR from where they were obtained for analysis. Blood gas samples were typical taken every four hours on CICU but this schedule was amended according to clinical judgement.

Table 4-4 - Data obtained from the Gemstar blood gas analyses

Variable	Frequency recorded (routine care)
Arterial pH	At least four hourly
Arterial Partial Pressure of Carbon Dioxide	At least four hourly
Arterial Partial Pressure of Oxygen	At least four hourly
Mixed Venous Oxygen Saturation	Ad hoc
Lactate concentration	At least four hourly
Chloride concentration	At least four hourly
Bicarbonate concentration	At least four hourly
Base Excess	At least four hourly

4.2.4. Draeger Innovian electronic patient record (EPR)

The majority of data analysed in this research program was extracted from the Draeger Innovian EPR. The EPR is a software suite which allows the documentation of a clinical record to facilitate an integrated multidisciplinary approach to postoperative care. Data are stored in a number of tabs within the EPR as highlighted by the red rings in Figure 4-1. The subsections which follow describe the data contained within each tab to allow an understanding of how the data analysed in this research program were collected. All data that are displayed in these tabs in the user interface of the EPR are also transferred to and stored within the Draeger report server database. Data analysed during this research programme were extracted using SQL queries of this server using Crystal Reports software. This methodology allowed huge quantities of data to be extracted efficiently and presented in comma separated values (.csv) spreadsheet files which were then processed in R studio software suite.
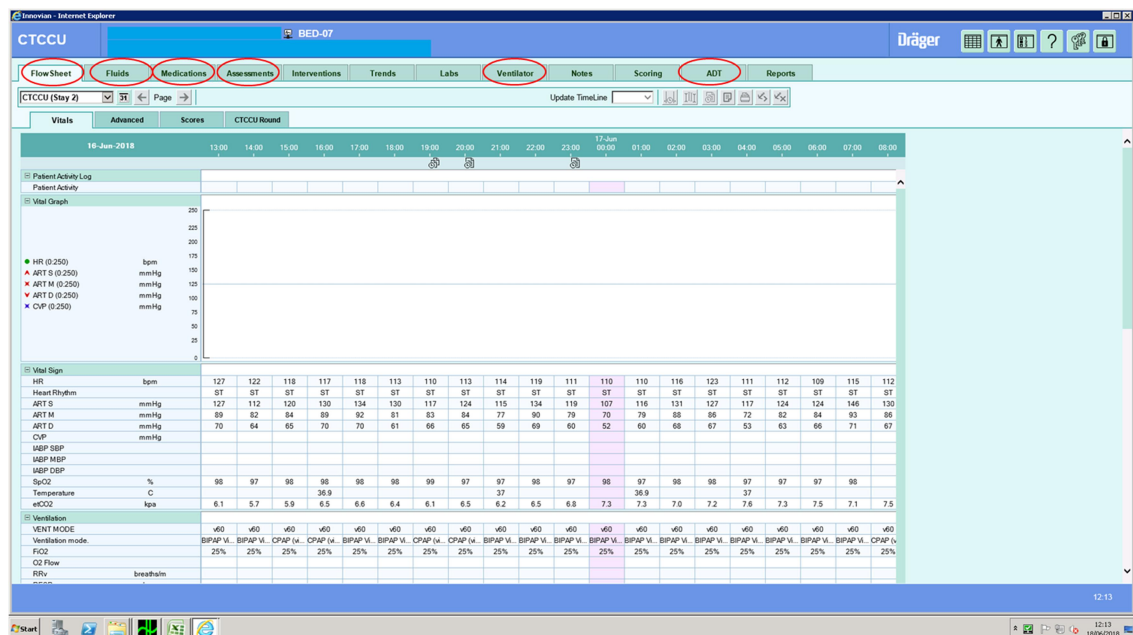
Figure 4-1 - The tabs within the Draeger Innovian electronic record

The ADT Tab

The ADT tab is populated by the nurse admitting the patient to the CICU. Most of the data are pulled automatically from the Hospital Information System (HIS) using the "Get HIS" function but some fields such as height and weight are entered manually (Table 4-5).

Table 4-5 - Data obtained from the Innovian EPR – ADT Tab

Variable	Frequency recorded
Date of Birth	Once
Gender	Once
Time and Date of arrival on CICU	Once
Time and Date of discharge from CICU	Once
Height	Once
Weight	Once

The FlowSheet Tab

The Flowsheet tab documents physiological measurements recorded for each patient. The tab chiefly contains fields which populate automatically from the patient monitoring equipment connected to the patient's bed space via the Health level seven (HL7) interface. For these fields (e.g. Heart rate, blood pressures) values must be verified by the nursing staff before they are saved. Additional data fields are completed manually by clinicians at the bedside (Table 4-6). Such fields contain data which are not measured by monitoring devices that have a connection to the EPR (e.g. thermometers, gas flow meters) or clinical assessments which are performed by the nurses such as sedation scores (Figure 4-2).

Figure 4-2 - The Draeger Innovian Flowsheet Tab

Table 4-6 - Data obtained from the Innovian EPR – Flowsheet Tab

Variable	Frequency recorded (routine care)	Input method
Heart Rate	Hourly	Pulled via HL7 from monitor
Heart Rhythm	Hourly	Inputted by nursing staff
Systolic Blood Pressure	Hourly	Pulled via HL7 from monitor
Diastolic Blood Pressure	Hourly	Pulled via HL7 from monitor
Mean Arterial Pressure	Hourly	Pulled via HL7 from monitor
Central Venous Pressure	Hourly	Pulled via HL7 from monitor
Temperature	Four Hourly	Inputted by nursing staff
Richmond Agitation and Sedation Scale score	Hourly	Inputted by nursing staff
Glasgow Coma score	Hourly	Inputted by nursing staff
Respiratory rate	Hourly	Inputted by nursing staff
Oxygen saturations	Hourly	Pulled via HL7 from monitor
Fraction of inspired oxygen	Hourly	Inputted by nursing staff
Intra-aortic balloon pressures	Hourly	Inputted by nursing staff
Mechanical circulatory support flows	Hourly	Inputted by nursing staff

The Assessments Tab

Data recorded in the assessments tab are entered manually by the clinician caring for the patient. The tab consists of a number of checklists which contain information regarding the treatments being delivered (Table 4-7). Data from fields contained within this tab were used to identify patients who received specific treatments and helped pinpoint the times such treatments were initiated and completed.

Table 4-7 - Data obtained from the Innovian EPR – Assessments Tab

Variable	Frequency recorded
Endotracheal tube insertion date	Daily
Tracheostomy insertion date	Daily
VasCath insertion date	Daily
LVAD clamps available	Daily
Intra-aortic balloon pump insertion date	Daily

LVAD – Left ventricular assist device

The Fluids and Medications Tabs

These tabs are considered together as data from both tabs are combined to calculate the fluid intake and output for the patients (Table 4-8). All data in these tabs are inputted by the nursing staff at the bedside.

The fluids tab (Figure 4-3) details all fluids (including blood products) administered to the patient through any route. The type of fluid and the volume administered each hour is recorded. Hourly fluid outputs including urine, drain and nasogastric aspirates are also recorded in this tab.

Figure 4-3 - The Draeger Innovian Fluids Tab

The medications tab (Figure 4-4) contains the name of each drug administered, the dose administered and the volume of fluid the drug is contained within when it is administered. The total volumes of all fluids and medications recorded across the fluids and medications tabs are determined hourly and total hourly input totals are displayed in the fluids tab. Similarly, all fluid output is summed to give hourly fluid output. The main limitation to the data recorded in the Innovian EPR is the lack of data concerning oral medication. Only intravenous medication data are recorded electronically; oral medication administration is only recorded on paper charts at the patient's bedside. While the majority of medication is given intravenously, particularly in the early postoperative stay, for those patients who experienced prolonged CICU admissions, more medication is given orally. Beta-blockers and other antihypertensives are generally given orally throughout the CICU stay and the lack of data regarding administration of these drugs is one of the main weaknesses of this dataset.
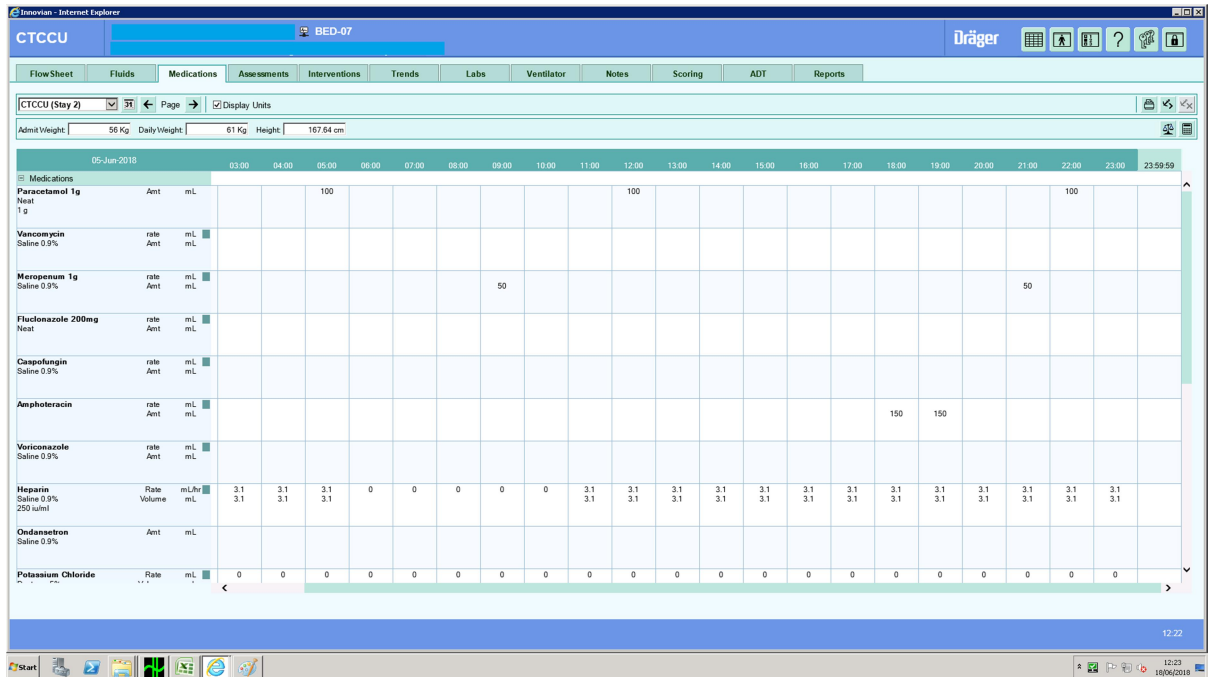
Figure 4-4 - The Draeger Innovian Medications Tab

Table 4-8 - Data obtained from the Innovian EPR – Fluids and Medications Tabs

Variable	Frequency recorded
Total fluid in	Hourly
Total fluid out	Hourly
Urine output	Hourly
Vomit/NG output	Hourly
Drain output	Hourly
Gastric aspirates	Hourly
Fluid Type	Hourly
Fluid Volume	Hourly
Blood product transfusion	Hourly
NG feed input	Hourly
Name of drug administered intravenously	Ad hoc
Dose of drug administered intravenously	Ad hoc

NG – Nasogastric

Ventilator Tab

The ventilator tab (Figure 4-5) is populated automatically from the ventilator via an HL7 connection. If this connection fails, there is the option for clinicians to manually enter data on an hourly basis. A new column of variables is populated whenever there is a change of settings or a marked change in measured values.   Variables extracted from this tab are detailed in Table 4-9.
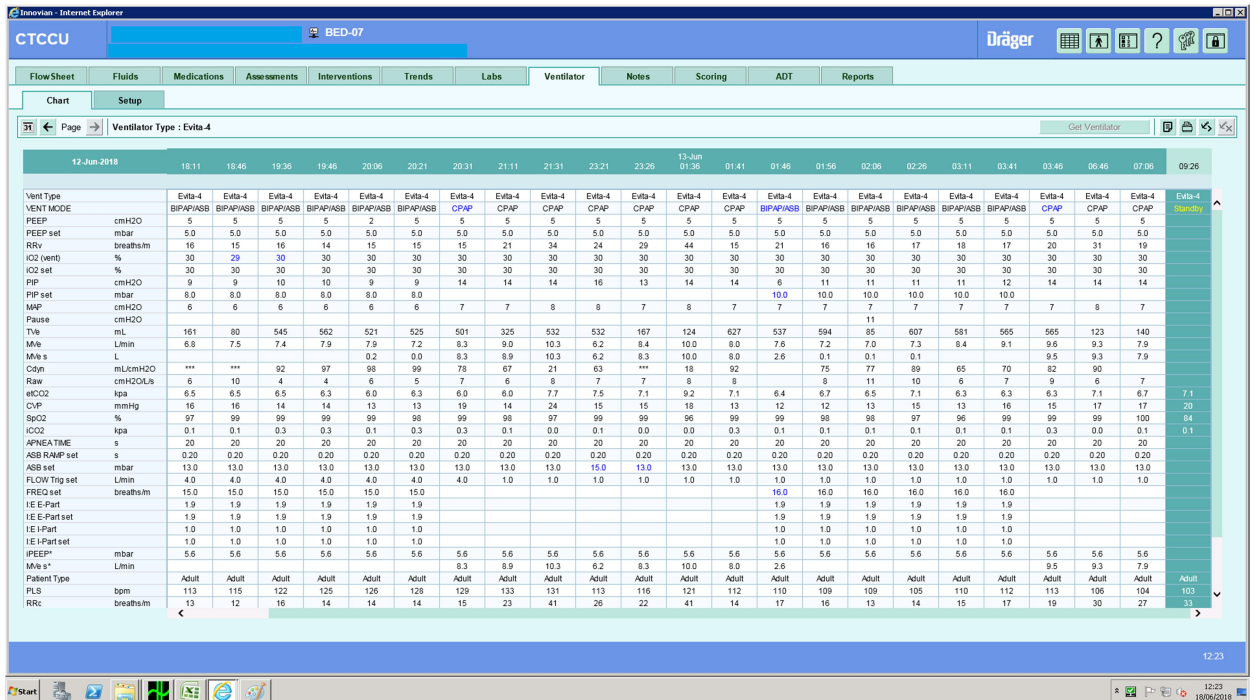
Figure 4-5 - The Draeger Innovian Ventilator Tab

Table 4-9 - Data obtained from the Innovian EPR – Ventilator Tabs

Variable	Frequency recorded
Mode of ventilation	Upon significant change typically multiple times per hour
Respiratory Rate	Upon significant change typically multiple times per hour
Peak inspiratory pressure	Upon significant change typically multiple times per hour
PEEP	Upon significant change typically multiple times per hour

PEEP – Positive end-expiratory pressure

4.2.5. Draeger Infinity bedside patient monitors

Patients in the CICU are monitored continuously using Draeger Infinity bedside monitors. The layout of the monitor's output and an example of the parameters recorded is shown in Figure 4-6. The bedside monitors display the previous 9 seconds of data continuously; the traces sweep from

left to right and then start again on the left and overwrite the previous images. The recorded traces are analysed by the monitor and used to determine the discrete variables which populate the Flowsheet tab (e.g. heart rate, blood pressures etc.) as discussed in the previous subsection. All waveform data are also transmitted to a report server where it can be viewed in real-time by Draeger's PatientWatch software (Figure 4-6). PatientWatch allows a clinician to view the output from the monitors at any bedside remotely via the hospital's intranet network. However, there is no capability to save data either on the monitor itself or using the PatientWatch software.

Figure 4-6 - The Draeger Infinity bedside patient monitor output screen

For this project it was necessary to record and analyse the waveform traces in order to identify information contained within the traces which could help identify patients at risk of complications. In order to record the waveforms the bespoke application programming interface (API) shown in Figure 4-7 was developed from a generic diagnostics API provided by Draeger. This modified API was designed and built by staff from Rinicare Ltd who acted as a research partner on this project. Using the API it is possible to record all monitoring data by "listening" to the report server through which data were flowing (Figure 4-8) before it was deleted. The principal limitation of the API is that data can only be recorded prospectively by telling the API to listen to the output from the monitor at a bed space. Thus for this project the Innovian EPR was accessed multiple times every day to determine which beds were occupied by cardiac surgery patients.

Once the appropriate bed spaces were identified, the API was executed and the .csv were produced and saved in a patient specific file. Where a patient moved to a different bed within the CICU, all data recorded from the different bed spaces for that same individual were collated in one patient folder. The variables recorded using the API and the resolution of the traces recorded are shown in Table 4-10.
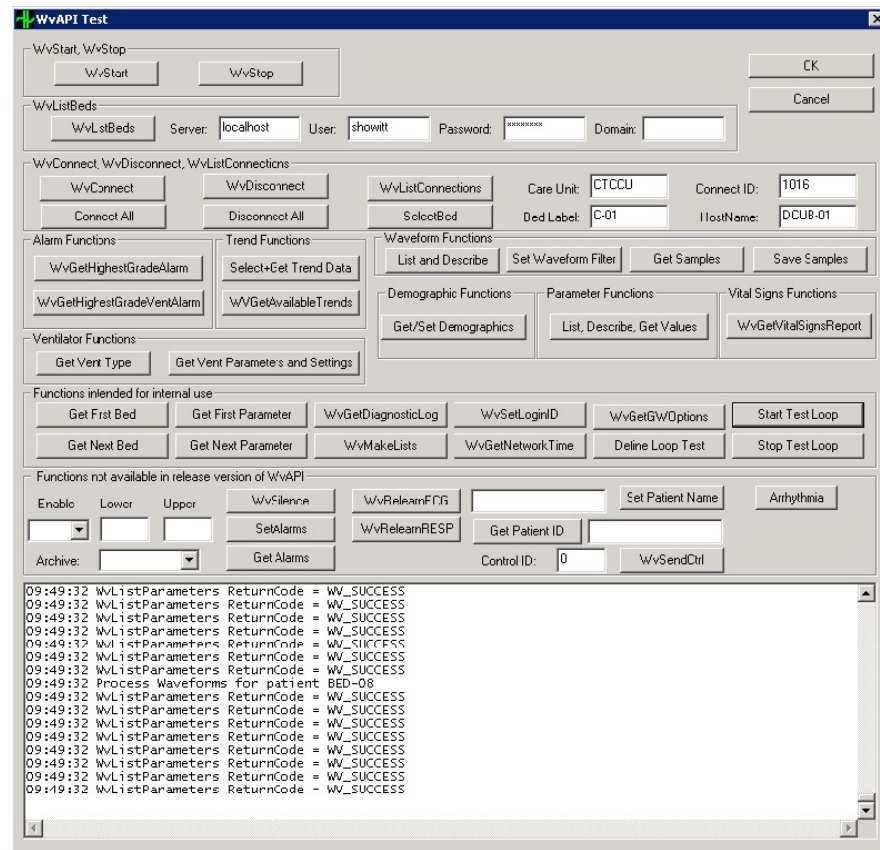
Figure 4-7 - The Application programming interface used to capture data from the Gateway report server
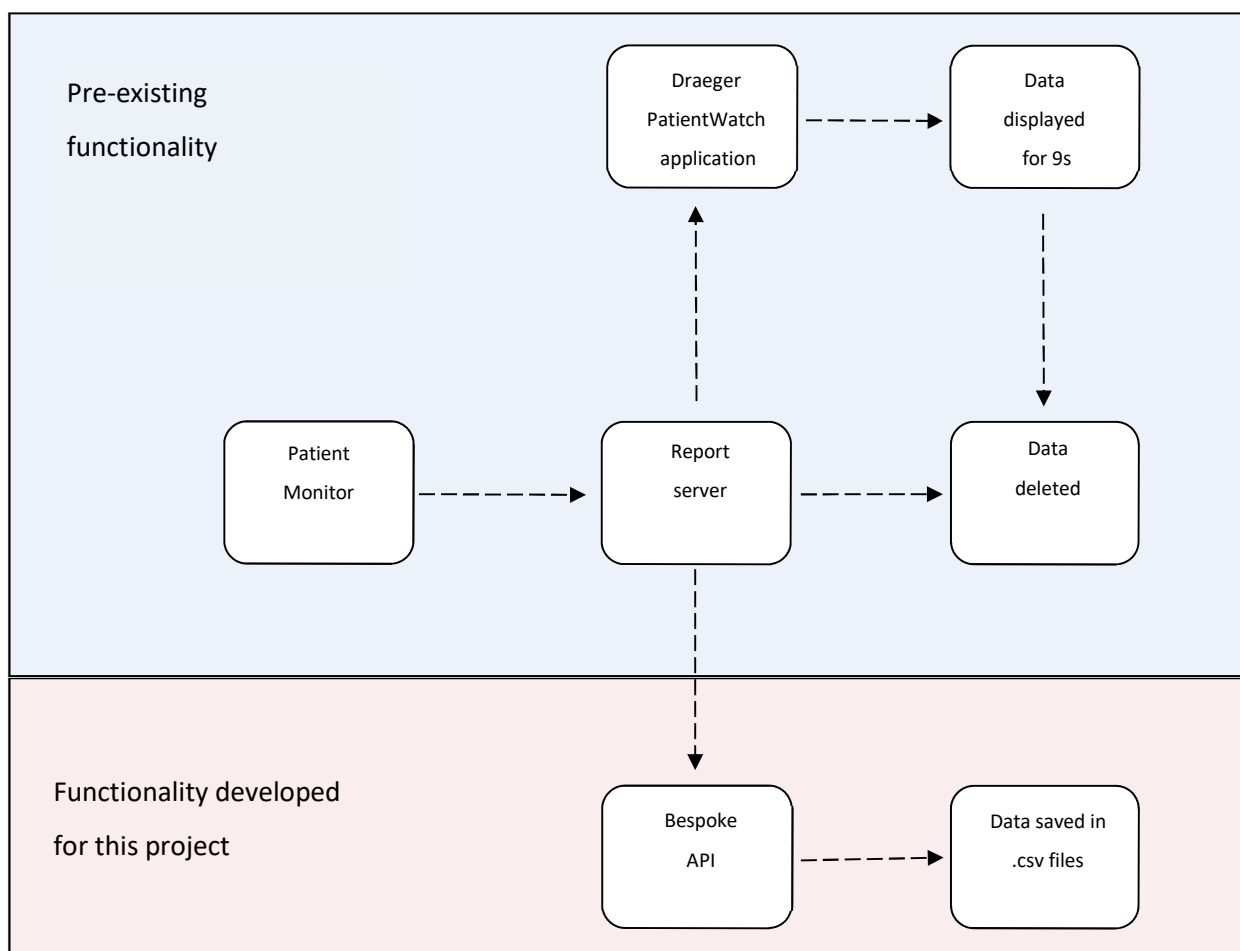
Figure 4-8 - Data flow for the waveform traces recorded by the bedside monitors

Table 4-10 - Data obtained from the bedside patient monitor

Variable	Resolution (Hz)
ECG trace (Leads I,II and III)	200
Pulse plethysmography trace	100
Arterial pressure waveform trace	100
Central venous pressure waveform trace	100
Ventilator Airway pressure trace	50
Ventilator volume trace	50
Pulse oximetry oxygen saturations	6

Data recorded prospectively from continuous monitors were combined with data collected retrospectively from the other data sources as detailed in the next chapter.

4.3. Selection of important outcomes (assisted by the patient and public involvement group)

It was important to ensure that this project targeted the prediction of outcomes that were important to patients and clinicians. As discussed in chapter two, models should predict these complications in a manner which would allow interventions to be performed to prevent the complication or reduce its consequences. The major complications which occur following cardiac surgery were identified in the literature report for this research programme. A patient and public involvement (PPI) group was convened to gain a patient's perspective regarding the importance of different clinical outcomes. Complications were ranked in order of importance by the PPI group and the results of this process are detailed in Table 4-11 alongside the frequencies of the complications obtained from the literature search.

Table 4-11 - Major complications following cardiac surgery and their frequencies. Ordered as ranked by importance to the members of the Patient and Public involvement group

Complication	Frequency
Mortality	3% (2)
Return to theatre for post-operative bleeding	3% (37, 80)
Renal replacement therapy (including dialysis)	2% (29, 134)
Acute kidney injury	50% (135, 136)
Post-operative infection	5% (56, 137)
Respiratory failure requiring reintubation and ventilation	5% (49, 50, 138)
Tracheostomy	2% (26)
Prolonged mechanical ventilation (>24 hours)	8% (26, 49, 50)
Readmission to CICU	7% (139, 140)
Prolonged CICU stay (>24 hours for CABG, >48 hours for valve surgery)	15% (141, 142)
Cardiac arrhythmias	18% (143)

Subsequent discussions with the PPI group informed the choice of complications this thesis would aim to predict. The complications selected were mortality, sepsis, AKI (including need for RRT) and cardiac arrhythmias. Cardiac arrhythmia was selected despite its low initial ranking because of its relatively high frequency. This increased the likelihood of developing successful risk prediction algorithms for this complication. Moreover, clinicians could provide prophylactic

antiarrhythmic medication to those at highest risk of cardiac arrhythmias meaning that a prediction model which stratified risk of arrhythmias would have a large clinical impact. Unfortunately due to delays in the appointment of a postdoctoral researcher at the department of statistics at Durham University discussed below, the arrhythmia model was not completed in time to be included as part of this thesis.

4.4. Ethical approvals

Although all data analysed during this research programme was recorded routinely during clinical care, approvals were obtained from the Research and Development department at Manchester University Hospitals NHS Trust (2017CD007) and the National Research Ethic Service – Haydock (09/H1010/2+5) to store and analyse the data.   The chief considerations were to ensure security of the data and to anonymise the data prior to sharing it with research partners. As all data were recorded during routine clinical care it was agreed that verbal consent to participation in the research project was adequate for all prospectively patients enrolled after the ethical approval had been granted. For similar reasons the ethics committee allowed historical data recorded prior to the start of this project to be added to the database.

4.5. Collaboration with Durham University for statistical analyses.

This research programme involves the development of novel Bayesian models to analyse the vast amounts of data recorded on the CICU to identify trends in a patient's own data indicative of increased risk of complications. The analyses are complex and require collaboration with experts in the field of Bayesian statistics. Prof McCollum had previously worked on a pilot study which preceded this programme with a team of statisticians at Durham University led by Prof Goldstein and Dr Caiado. This pilot study analysed heart beat variation on ECGs to identify when the heart rate was becoming unpredictable. A decrease in stability of the patient's heart rate was found in 18 patients who died on CICU within hours of their death but was not identified in 40 healthy controls.(144) As part of this research programme the British Heart Foundation funded the salary cost of a post-doctoral researcher in the department of our collaborators at Durham University. Under the guidance of Dr Caiado and Prof Goldstein the researcher was to conduct the complex Bayesian analyses require to produce the novel risk prediction models.

4.5.1. Delays related to the collaboration

Unfortunately the statistician who was appointed to the post originally did not perform adequately and was ultimately dismissed. The work performed in the six months that the researcher spent on the project resulted in no useful output and this delayed the analysis of data and design of models. Fortunately, Dr Caiado and her colleague Jordan Oakley were able to continue to work on the urine output model described in chapter ten of this thesis but work using continuous monitoring data stalled. In February 2018, Ben Lopez was appointed to the post-doctoral researcher post and has produced work of the highest quality.

Despite the delays, this thesis still presents the validation of a novel urine output model which employs Bayesian analyses to identify patients at risk of severe oliguria based on trends within their own previous urine output values. Work continues on a model to identify those at risk of arrhythmia based on changes in their ECG traces relative to their own previous waveforms but this work was not ready for inclusion in this thesis.

Chapter Five: Data cleaning

This chapter summarises the steps involved in cleaning the data and the rationale for performing the cleaning process in this manner. All data were obtained in the comma separated value (.csv) file format from the sources described in the previous section. Each file was loaded into R Studio and cleaned using code written in the programming language R (R foundation for statistical computing). All data points were assigned to the appropriate patient episodes using the steps detailed in the subsections of this chapter. Once assigned appropriately, all variables were cleaned using semi-automated algorithms in a reproducible way. The cleaning steps are described in detail in later sections but in summary there were three key parts to the initial cleaning process.

1. The format of all entries in each field was standardised to allow fair comparison of values. E.g. F_iO_2 of "21%" and ".21" were both recoded as "0.21", ventilator mode of "Bipap/ASB" and "bIPAP+asb" were both recoded as "BiPAP/ASB".

2. Cleaned variables were examined and values which fell outside the range of possible values were excluded.

3. Semi-automated algorithms were used to identify incidences of missing data and to generate "manual lookup" files. These files were used to guide manual inspection of the case notes which was then performed using the Draeger Innovian user interface. Missing data identified during this lookup process was entered into the "manual lookup" files which were then loaded back into R Studio and used to fill in the data gaps identified in the original dataset.

4. In general, suspicious data points were identified and flagged as suspicious. Such entries could then be omitted from analyses but were distinguished from missing data.

Data was cleaned in two distinct datasets. First, all data recorded between January 2013 and May 2015 was processed. Later, all data collected between July 2016 and November 2017 was processed in a separate batch using the original cleaning code. This approach allowed the analyses for chapters seven, eight and nine to be performed while data were still being recorded in the second batch. As the data processing was predefined and semi-automated, the output files from each batch were equivalent and data from both batches was subsequently combined for the analyses performed in chapters ten and eleven. Details of the specific cleaning steps required for each stage of the data cleaning process are provided in the following subsections. Subsequent analyses of the dataset for specific studies required additional data processing steps which are detailed at the start of each relevant results chapter.

5.1. Creating the initial patient index

Key considerations

The goal of this section was to create a Patient Index which classified all episodes of data recording ensuring that every data point was be assigned to the correct episode. Data from the Dendrite database, Fluids, Ventilator and FlowSheet Tabs were loaded into the workspace in RStudio. Data were cleaned and divided into episodes related to each Hospital Admission, Operation Number and CICU Admission number (Figure 5-1). The start and endpoints of each episode were determined.

The first classification of episodes was performed by hospital admission. As our data collection period was over four years long, some patients were admitted to hospital multiple times during the project. The first time each patient was treated on CICU following cardiac surgery they were assigned an "Admission Number" of 1. Where a patient was then discharged home but readmitted to hospital and treated on CICU following cardiac surgery for a second time, the "Admission Number" was increased to 2. Each subsequent readmission resulted in further increase in Admission Number.

The second classification was performed according to "Operation Number." Postoperative data recorded following the first cardiac surgery operation of any hospital admission were assigned an "Operation Number" of 1. Data recorded following a subsequent operation during the same hospital admission were assigned an "Operation number" of 2. Subsequent operations resulted in a further increase in Operation Number.

The final element of the unique identifier for each episode was the "CICU Admission Number." Data from each patient's first CICU admission for each "Operation Number" were assigned a "CICU Admission Number" of 1. Where patients were discharged from CICU to the ward and subsequently readmitted to CICU without having had further surgical intervention the data recorded were assigned a "CICU Admission Number" of 2. Subsequent readmissions resulted in further increases in the "CICU Admission Number". Importantly, if the readmission to CICU was preceded by further surgery, the "Operation Number" increased by 1 and the "CICU Admission Number" reset to 1 as it was considered to be the first CICU admission related to that operation.

The unique identifier for each patient episode was made by combining Patient Id, Admission Number, Operation Number and CICU Admission Number as shown in Figure 5-1.
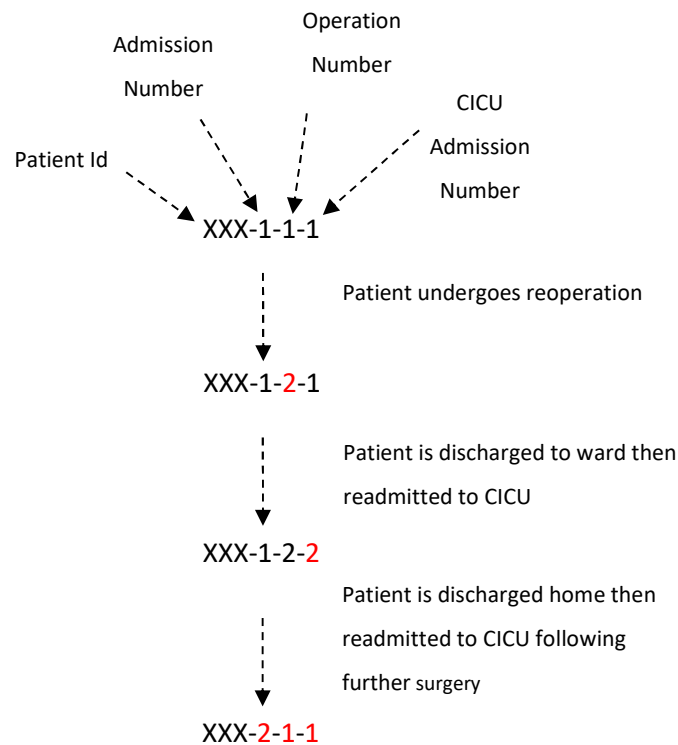
Figure 5-1 - Unique identifier structure

Cleaning Steps

Step 1.   Load FlowSheet, Dendrite, Fluids, Fluids summary and Ventilator sheets.

Step 2.   Standardise all dates and times to the posix format.

Step 3.   Assign each operation an Admission Number based on Hospital Admission date/time recorded in dendrite.

Step 4.   Assign Operation Number to each operation recorded in Dendrite.

Step 5.   Extract key parameters from Dendrite data frame creating Patient Index data frame.

Step 6.   Remove operations recorded in Dendrite which occurred outside the study period using OpStart field.

Step 7.   Remove any duplicated entries (entries within the same sheet which contained identical data) in any sheets.

Step 8.   Remove empty readings at the end of fluids where discharge has not been actioned.

Step 9.   Merge FlowSheet and Fluids with PatientIndex to make Check data frame.

Step 10.   Ensure that only data recorded after each operation but before each patient's next operation or hospital discharge are labelled with the relevant episode label.

Step 11.   Create a field which reflects the time interval between consecutive entries in Check.

Step 12. Where the interval between consecutive entries for a patient is >3 hours assign all subsequent data a new CICU Admission Number.

Step 13. Find first and last ITU entry date/times for each episode

Step 14. Rationalise Patient Index1 by selecting one entry for each ITU admission with fields: Id, AdmitDateTime, OpStart, FirstITUEntry, LastITUEntry, Admission, AdmissionOp and ITUAdmission

Step 15. Verify the list of patient Ids against the zlog of prospectively gathered continuous data collected (where applicable). Identify typographical error made when entering Patient identifiers and correct these errors.

5.2. Identifying reoperation not recorded in the Dendrite database and times where intubation status is unclear after automated analysis of EPR data

Key considerations

It was found that a number of "re-operations" were absent from the Dendrite database. Often these procedures were performed out of hours or in an emergency and these factors may have led to the lapses in data entry into the Dendrite database. In order to identify missed operations all incidences in which a break in monitoring data of > 1 hour occurred were identified and manually investigated through examination of the case notes section of the EPR. This exercise was combined with a second lookup task which is required to accurately categorise each patient's intubation status (whether invasive mechanical ventilatory support was in place) at each time point of their admission. The tasks were conducted at the same time to prevent repetition when examining the case notes section of the EPR and to reduce the total amount of time required to extract relevant information. The intubation statuses looked up during this process were then inputted back into the dataset during the production of the ventilated episode output files (Stage 4).

The steps detailed below describe how the intubation status was determined using a reproducible algorithm using data recorded in the Ventilator, FlowSheet and Assessment tabs of the EPR. Once the automated classification of intubation status had been performed, the periods for which intubation status was unclear were identified and added to the lookup file. Periods for which no data were recorded in FlowSheet fields used to determine intubation status for longer than 1 hour were also included as they may have signified absence from the CICU during a missed re-operation.

Occasionally, the mode of ventilation label pulled automatically from the ventilator into the patient monitor and subsequently into the EPR was a label which is only assigned by ventilators in

the operating theatre. Where one of these modes of ventilation was recorded, this time point was also added to the lookup file as it was likely to signify reoperation. Once the case notes section of the EPR had been examined and relevant data regarding intubation status and missed reoperations had been extracted, the Dendrite data frame was amended to ensure its completeness.

The cleaning steps described in section 5.1 were then repeated to produce a completed patient index.

<u>Cleaning Steps</u>

Step 16. Make FlowSheet$Intubated by categorising entries in FlowSheet$VENT.M into: "ETT or TT", "def No ETT or TT", "Unsure if FM, ETT or TT and Def TT". This step involves identifying all strings entered in the VENT.M field of FlowSheet which will only be entered in the EPR if a patient is intubated (e.g."BIPAP/ASB") and strings which will only be entered if a patient was not intubated (e.g. "Nasal High Flow"). Ambiguous entries in the Ventilation Mode field (e.g. "CPAP") resulted in a provisional "Unsure if FM, ETT or TT and Def TT" status.

Step 17. Assign definitive status to "Unsure" entries if possible after categorising strings entered in FlowSheet$VentilationMode

Step 18. Merge FlowSheet and Ventilator data frames to make FlowVent data frame.

Step 19. Where VENT.M contains strings similar to "CPAP" and a flow rate is present in the O2 flow field then this CPAP must be provided by the Draeger CPAP bellows. As this equipment is only provided via a facemask, the intubation status can be set "def No ETT or TT".

Step 20. If we are unsure because there are no data in VENT.M and Ventilation Mode fields assign and intubation status of "no data"

Step 21. If we are unsure at a given time point but we were sure within the previous 60 minutes use the previous value. This was needed because the VENT.M and Ventilation.mode fields from Flowsheet (which are verified by nurses) are only populated hourly. However, entries from the Ventilator Tab may be made many times each hour. Consequently, in between two known statuses obtained from FlowSheet one hour apart there may be multiple ambiguous entries from the Ventilator Tab's X.VENT.MODE field. It is important to note that an entry occurring in the Ventilator Tab between hourly entries made in FlowSheet which clearly identified the intubation status would still result in a change in classification of intubation status; only ambiguous entries were ignored. It is also important to note that this step extends a previously known intubation status a maximum of 1 hour into the future.

Step 22. If there are no readings after a period of no data label this as "End Readings No data." This classification is made using the "intubated2" field.

Step 23. Generate a lookup file for tracheostomy timings based on entries in FlowSheet and/or assessments tab. In this step any patient for whom there is any indication of the presence of a tracheostomy being present including "Trachy Mask" string recorded in the

VENT.M field or a string present in the $DateTracheostomyInserted field of the Assessments data frame is identified.

Step 24.   Complete manual lookup using the case notes section of the EPR to identify the date of tracheostomy insertion and removal.

Step 25.   Merge his file back into FlowVent data frame.

Step 26.   Assign "def TT" status to all times between tracheostomy insertion and decannulation. This updated intubation status is labelled Intubation3.

Step 27.   Where ventilator is off at the end of an admission and we are unsure of intubation status assign a status of "def no ETT or TT" between last entry and last "known" intubation status. This only corrects ambiguous entries. A patient would not leave CICU while on CPAP through an ETT.

Step 28.   If a no data/unsure entry is in the same hour as a definite intubation status pull that status forwards to replace the unknown for a maximum of 1 hour. This is done for the same reason outline in step 21.

Step 29.   Write the file identifying all instances where we have no knowledge of intubation status for >1 hour due to a complete lack of data (which might represent absence from the unit during re-operation). Also included in this file are all periods lasting >1 hour where intubation status is unclear due to ambiguous entries or PC or VC in ventilation mode.

Step 30.   Lookup all periods identified in the file produced in step 29, by examining manually the case notes section of the EPR.

Step 31.   Amend the Dendrite data frame appropriately to include all newly identified operations.

Step 32.   Reload all files and rerun steps 1-15 to produce a complete Patient Index

5.3. Ensuring data from all tabs is assigned to relevant patient episodes and make initial episode summary fields.

Key considerations

Once all re-operations which had been missed initially had been included in the Dendrite data frame and the updated PatientIndex1 data frame had been produced, final amendments to the PatientIndex1 data frame were made to ensure consistency and remove data which should not be included in the final dataset. Initial summary statistics were also produced calculating the time between consecutive CICU admissions and lengths of stay on the CICU.

Failed discharge from ICU was defined as readmission within 48 hours of CICU discharge. In cases where a patient left CICU but only to undergo reoperation before returning CICU this was not classified as a failed discharge.   Finally, attributes such as weight and gender were added to the patient index file and the procedure details fields were standardised.

Step 33.   Amend lastITUentry where required to ensure removal of any data identified during the lookup process as being recorded erroneously due to a malfunction of the ventilator data streaming.

Step 34.   Remove those pts for whom no Innovian data are identified (these few patients were not admitted to CICU) or where age <18

Step 35.   Look for single entry admissions and remove them. Manual case note inspection revealed that these entries were caused in Step 12 when patients who were close to discharge were monitored less frequently than 3 hourly. LastITUentry was corrected for each of these episodes.

Step 36.   Where reoperation occurs, ensure Flowsheet data from during the second intraoperative period are not labelled as coming from the subsequent CICU admission. This is not an issue for other tabs as only data which populate automatically from the monitors were entered. No data would be manually entered into fluids or assessments tabs during a theatre episode. All ventilator data pulled from a Primus ventilator (used exclusively in theatres) were removed from the ventilator data frame.

Step 37.   Find time between CICU discharge and readmission for each patient where applicable

Step 38.   If discharge is followed within 2 hours by another operation and another CICU admission then assign the value "Surgery" to the Failed DC 48 hrs field.

Step 39.   If no surgery occurs within 2 hours of discharge and time to next admission is <48Hrs this is classed as a failed discharge.

Step 40.   Determine the Cumulative LOS on CICU. This is all time for which admissions are only separated due to reoperation

Step 41.   TotalITUTime= total time on ICU during each hospital admission and includes readmission to CICU following initial discharge to the ward.

Step 42.   Eliminate those not in PatientIndex from dendrite and attach operation details to each operation in PatientIndex1.

Step 43.   Calculate age at time of operation

Step 44.   Check for excessively long/short operations and correct any by amending operation start and end time accordingly using information from the EPR case notes

Step 45.   Standardise the names of operations performed. This is particularly relevant for reoperation where the operation details were entered manually during the lookup exercise in step 20.

Step 46.   Where the primary Procedure details field is blank substitute in data from "additional cardiac surgery" and "other non-cardiac surgery" fields into ProcDetails field

Step 47.   Calculate 30day and ITU mortality

Step 48.   Check which repeat operations were for bleeding and make finalreoperation sheet

Step 49.   Load weights and CPB times from the Perfusion database

Step 50.   Where they are missing look in Dendrite data frame

Step 51.   Make all weights data frame and manually lookup missing weights in Innovian.

Step 52.   Load these as allweightsdone.   Add height and weight to patientindex1 from this.

Step 53.   Add gender to PatientIndex 1 from dendrite

Step 54.   Impute weight (gender specific where weight is missing).

5.4. Producing the Ventilated Episodes output file

<u>Key considerations</u>

The aim of this stage was to finalise the classification of the intubation status of all patients at every point of their CICU stay. Initially the same steps 16-28 were rerun. This automated process was able to confidently identify the intubation status for the majority of all patients' CICU stays. Where there was any uncertainty concerning data that had been obtained, manual examination of the case notes section of the EPR in step 30 was used to confirm the intubation status. Subsequent cleaning steps then identified all incidences where the intubation status changed and labelled extubation and reintubation events appropriately. The time between extubation and reintubation was also determined together with the total time spent ventilated. All periods of ventilation via a tracheostomy identified in step 24 were also labelled.

It was identified that rarely the automated classification system had incorrectly classified an intubation status due to erroneous data entries. To ensure all such episodes were corrected all EPR records were manually examined for all patients who were identified as experiencing a period of intubation which lasted less than 4 hours (or less than two hours if it was a first episode). The records of all patients who suffered reintubation within 48 hours of extubation were also identified.

The output file produced in this section was a data frame which detailed times of extubation and intubation for all patients. Summary statistics including length of ventilation and time to reintubation were also recorded along with time of insertion or removal of tracheostomy tube.

<u>Cleaning Steps</u>

> Step 55. Rereun 17-30 to produce best estimates of intubation status
> Step 56. Merge in Vent Changes from FinalLookUp file
> Step 57. Manually correct any missing intubation statuses
> Step 58. Create admitted tubed column
> Step 59. Create tubed during admission column to mark reintubation
> Step 60. Create admitted extubated column
> Step 61. Create extubated during admission to mark extubation
> Step 62. Create date tubed for all incidences of intubation
> Step 63. Create dateextubated for all incidences of extubation
> Step 64. Create tubed episodes
> Step 65. Manually check and correct entries where patients arrived without ETT in situ
> Step 66. Check how long there was between extubation and any subsequent reintubation
> Step 67. Highlight where extubation was followed within 48hrs by reintubation (failed extubation) – Reintubated48hrs
> Step 68. Calculate duration of ventilated episode

Step 69.   Calculate cumulative time ventilated

Step 70.   Insert trachyin and trachyout status into tubed episodes

Step 71.   Merge tubed episodes with relevant fields from patient index to make tubedepsfinal

Step 72.   Manually inspect and correct any failed extubations and any intubation episodes lasting <4 hours if not first episode or <2hours if first episode

5.5. Cleaning data from flowsheet tab

Key considerations

The variables within this data frame represented a large proportion of the predictor variables derived from physiological monitor outputs. Many of the fields were of the free text format and with data from over 3500 patients being analysed there were are large number of typographical errors.   In addition to incorrect spelling, character strings were present in some fields which were supposed to contain numerical data. The initial steps in this cleaning section were to remove characters from numeric fields and then to format the fields as numerical data. All string variables were recoded to ensure standardisation. For example, "Sinus", "SR", "sr" and "Sr" entries in the Heart Rhythm field were all re-classified as "SR". A label was added to all blood pressure readings where the values were not physiologically probable. This label was applied to values that were too high or low to be compatible with life together with blood readings which were likely to have been recorded from severely damped waveform traces. Similarly, impossible values were removed from all fields pertaining to blood gas analyses and temperature and S_pO_2 measurements.

The presence of intra-aortic balloon pump pressure readings or pump flow rates which implied the presence of intra-aortic balloon pumps or mechanical circulatory support (MCS) respectively were extracted and used to generate a lookup file. The EPR records of all patients with data entered in these fields or data entered in fields from the "Assessment" tab associated with treatments were manually examined. Following this examination the time of instigation and cessation of the therapies was recorded and every recorded blood pressure was classified as having been recorded in the presence or absence of MCS.

Using this information it was possible to identify patients whose arterial blood pressure traces appeared damped due to the presence of MCS. Having identified these patients, revised labels of "reliable" arterial blood pressures were produced. All values which were physiologically possible taking into account whether or not MCS was present were labelled as reliable.

Blood gas samples were typically labelled as containing arterial or venous blood but where this labelling was absent an algorithm was devised to determine the source of the sample depending on the absolute value of the PO_2 and the S_PO_2 measurement recoded at the time the blood sample was taken. Where ambiguity remained case notes within the EPR were examined manually and often it was found that clinicians had stated the source of the sample in this section. The classifications identified on analyses of case notes were then copied into the Flowsheet data frame and all PO_2 blood gas samples taken from arterial samples were labelled to distinguish them from PO_2 values measured in venous samples.

Fields for Glasgow Coma Scale(145) score and Richmond Agitation-Sedation Scale (146)scores were standardised by removing characters and reclassifying any GCS <3 as 3. The Central Nervous Systems score as proposed by Hekmat et al(147) was also calculated.

Cleaning Steps

> Step 73.   Recode HR by removing symbols
> Step 74.   Clean rhythm by standardising rhythm into one of 12 rhythms
> Step 75.   Find first occurrence of AF
> Step 76.   Where AF was not in preop rhythm field of dendrite create first new AF field in PatientIndex
> Step 77.   Clean blood pressure removing words and symbols
> Step 78.   Where BP is blank label it missing
> Step 79.   Where Pulse pressure is <25 and mean is >100, MAP <30 or MAP>250 label erroneous
> Step 80.   Clean pH standardising full stops
> Step 81.   If <6.8 put 6.8
> Step 82.   If pH<6 or >7.8 label as not possible
> Step 83.   If lactate <0 or >40 label as not possible
> Step 84.   Find all incidences of IABP being recorded in flowsheet or assessments tab
> Step 85.   Manually look up insertion and removal times for all those with an IABP – IABPLookUP
> Step 86.   Merge IABP details into flowsheet
> Step 87.   Make a field which determines if IABP is present at every row according to IABP.in and IABP.out times
> Step 88.   Recode temperature removing parentheses, letters, commas etc.
> Step 89.   Set temp <15 or >42 to not possible
> Step 90.   Clean S_PO_2 removing words and symbols and letters
> Step 91.   Set S_PO_2 >100 to not possible and <50 to unreliable-low
> Step 92.   Clean pO_2 removing letters etc.
> Step 93.   Where sample type is not recorded determine if sample was venous/arterial. If PO_2<4 – venous, ifPO_2 <5 and S_PO_2>85 venous, if PO_2>7.5 arterial otherwise lookup manually
> Step 94.   Merge ABG lookup into flowsheet

Step 95.   Create ART.PO2 from all ABGS where sample was arterial

Step 96.   Clean F_iO_2 removing words and symbols

Step 97.   Clean RASS recoding non-specified entries

Step 98.   Clean GCS removing words

Step 99.   Create CNS score as per Hekmat N, S, C

Step 100.   Identify all incidences of VAD flows recorded in flowsheet, VAD recorded in dendrite and clamps available being present in Assessments

Step 101.   Manually check ECMO/VAD on and off times

Step 102.   Merge ecmovadlookupinto flowsheet

Step 103.   Make fields describing whether for each timestamp ECMO, short-term VAD or long term VAD or Any of the above were present.

Step 104.   Where ECMO or VAD was present overwrite any bloodpressure erroneous==Yes entries with No as the values may have been accurate.

Step 105.   Make reliable ART.S, ART.M, and ART.D for all BP where BPerroneous ==No

5.6 Cleaning fluids and medication tabs

Key Considerations

In this cleaning stage, data from the Fluids tab were divided into the appropriate patient episodes using the time for which the data were entered and the first and last CICU entries for each episode recorded in the Patient Index data frame.

Urine output data were occasionally entered at multiple points during the same hour. Where this occurred the values were summed, creating the Urine60 field which represented the total urine output for each hour. Any entries indicative of the use of renal replacement therapy were also extracted at this point and all such incidences were later verified manually. A key part of this stage was the averaging of urine output where no value was recorded. Where urine output was recorded as "0" this value was preserved. However, where an hourly urine output was blank, this was considered to represent an absence of recording not an absence of urine output. In clinical practice urine output is recorded on the hour. Often nursing staff are too busy to note the urine output at a given hour, rather than enter a number measured 15 minutes late, the recording is often omitted and the total value recorded at the next measurement time.   Therefore, where blank entries were observed, the next recorded value of urine output was divided equally amongst the preceding blank hours.

Step 106. Merge fluids with patient index

Step 107. Assign data to correct episodes using first and last ITU entries

Step 108. Check for filter use using CVVH fields

Step 109. Combine UO into hourly segments

Step 110. Where 0 is entered preserve it

Step 111. Where NA is entered divide any subsequent UO by that number of hours missing a value

Step 112. Label where the patient is in their last 6, 12 and 24 hours on ITU

5.7. Cleaning blood test results

Key considerations

In this stage blood results obtained from the pathology laboratory database were each assigned to the appropriate patient episode. For each episode, where available, the most recent preoperative blood test results were also obtained.   All postoperative blood results were obtained by searching the pathology database for all blood tests sent for CICU and pairing the results with the patient identifiers contained in patient index. It was found that some samples had been sent to the lab labelled with the incorrect location and so were missed by the initial search. A cleaning code was therefore written to identify all days for which there were no postoperative blood results for a given patient. Using this list, the pathology database was searched manually using the relevant patient identifiers and all missing blood results were added to the appropriate data frames. A similar procedure was performed to ensure that all preoperative blood results were also identified.

Cleaning Steps

Step 113. Load Biochem and haemCoag

Step 114. Make a data frame containing each day for each patient between first firstITU entry and last lastITU entry

Step 115. Pair this with the blood results to identify days with no blood results

Step 116. Manually check this wasn't due to a location error

Step 117. Merge preop bloods with Patient index

Step 118. Manually check all operations have a preop sample

Step 119. Manually check all multiple ops have been updated appropriately

Step 120. Merge preop with post op bloods – link is NewId

Step 121. Only keep bloods entries where the postop blood came from within the first and last entries

<u>5.8. Identifying AKI</u>

<u>Key considerations</u>

The Kidney Disease Improving Global Outcomes (KDIGO) Acute Kidney Injury (AKI) guidelines(87) stratify the severity of AKI in adults using hourly urine output measurements and serum creatinine concentrations. In order to determine the onset of each stage of AKI suffered by each patient on CICU, data from the blood results and fluids data frames were analysed together.

Serum creatinine concentrations

First for each patient, every postoperative creatinine concentration was compared with the pre-operative value for that episode. Where the postoperative concentration exceeded a threshold to diagnose a stage of AKI the timestamp of the postoperative result was noted. Secondly, a code which created a moving 48 hour window was used to identify all occasions when the postoperative creatinine value increased by more than 26.5 micromol/L within 48 hours (this is one of the KDIGO stage 1 AKI criteria). Finally, all incidences when creatinine increased above the absolute threshold for stage 3 AKI (353.3 micromol/L) were also identified.

Hourly urine output

All hourly urine output values produced for each patient during step 111 were divided by the weight of the patient to give an hourly urine output value per kilo. For a very small number of patients the patient weight was not recorded. Where this was the case the imputed weight derived in step 34 was used instead. Next, code was written which analysed the urine output data in moving windows of six, twelve and 24 hours. Where urine output was below the thresholds described in the KDIGO guidelines for the required time period, the date and time at which this happened were recorded. This process resulted in a record which contained the first incidence where each criterion contained within the KDIGO guidelines was fulfilled by each patient. These incidences were analysed in combination to give the time the onset of each stage of AKI experienced by all patients. Occasionally AKI was diagnosed by urine output through analysis of a period during which the original record contained a blank value. Where this occurred the EPR was examined manually to determine whether there was any reason to doubt the diagnosis. If it was found that the blanks represented a note inserted into the record to describe urine output lost into the bed or toilet, the diagnosis was disregarded and the rest of the record examined to ensure any later onset of AKI by urine out was recorded instead. All incidences where anuria lasted > 12 hours (stage 3 AKI) were also examined. Where this occurred in a

patient who did not have a catheter in situ and the patient went on to pass urine, the oliguria was labelled as "Anuric no catheter."

<u>Cleaning Steps</u>

Step 122.   Check where creat has increased 1.5 x preop baseline
Step 123.   Create time interval
Step 124.   Flag where postopcreat has gone up by 26micromol/l within 48 hours
Step 125.   Check where creat increased 2x preopbaseline
Step 126.   Check where creat is 3 x baseline or goes >353.3 micromoll-1
Step 127.   Divide hourly UO by weight
Step 128.   Where UO is below 0.5ml/kg for 6 hours set stage1 hourly outputs to Yes
Step 129.   Where UO is below 0.5ml/kg for 12 hours set stage2 hourly outputs to Yes
Step 130.   Where UO is below 0.3ml/kg for 24 hours set stage3 hourly outputs to Yes
Step 131.   Where UO is 0 for 12 hours set stage3Anuria to Yes
Step 132.   Find all those patients with diagnoses made where NAs were present in the time during which AKI was diagnosed
Step 133.   Ensure the AKI wasn't diagnosed when urine was discarded/lost by accident
Step 134.   Assign the reason for AKI diagnosis
Step 135.   Work out which criteria for AKI were fulfilled at any time
Step 136.   Check UO related diagnosis was not due to lost UO where this occurred after the creat diagnosis
Step 137.   Manually check all instances of AKI where AKI3 was reached without RRT being started have been correctly assigned
Step 138.   Note where AKI 3 anuria occurred in healthy people with no catheter – Anuric no cath

5.9. Finalisation and anonymization

<u>Key considerations</u>

Haematological blood results (including those identified in step 118 were linked to the relevant patient episode). Cleaned versions of variables including age, urgency, procedure details and CPB time were the transferred to the Patient Index file. Finally, all patient identifiers were replaced by unique ID numbers. All data transferred from VGNW for analysis were non identifiable. A "key" file was stored on the VGNW database to allow a link back to the original data.

<u>Cleaning Steps</u>

Step 139.   Clean haem assigning a preop and all available PostOp haem and coag to each episode
Step 140.   Add in age and CPB time and urgency
Step 141.   Add in Proc details cleaned from dendrite
Step 142.   Anonymise every data frame and write the output .csv files.

Chapter Six: Statistical Methods

During this research programme associations between predictor variables and outcomes were tested using different statistical methods. Generally, initial comparisons were made using univariable tests and then the effect of potential confounding variables was adjusted for using multivariable analyses.(148) This thesis subsection discusses the rationale for the statistical analyses employed.

6.1. Univariable analyses

Univariable analyses compare outcomes between patients grouped by one variable e.g. sepsis status. Outcomes compared during this research programme include binary outcomes such as ICU-mortality, continuous outcomes such as length of stay and time-to-event analyses such as 2-year survival.

6.1.1. Continuous outcomes

When analysing continuous outcomes, before a choice of statistical test was made, the normality of the data was assessed. If the outcome data followed a normal (parametric) distribution, tests such as the Student's t-test (149) were used to compare the outcomes of the groups. Such analyses compare the mean values in each group and assume that the outcomes are normally distributed. If data are not normally distributed (non-parametric) these assumptions are not valid and can cause inappropriate conclusions to be drawn.(150, 151) Non parametric data were therefore analysed using tests which do not assume the normality of the data being analysed. Non-parametric tests used in this thesis include the Mann Whitney U test(152) (also known as the Wilcoxon rank-sum test) and the Kruskal-Wallis(153) test.

6.1.2. Binary outcomes

The choice of univariable test used to analyse the association between a predictor variables and a binary outcome was determined by the expected number of outcomes in the smallest group of patients analysed. Although a large number of patients were included in this programme, some of the groups used in the analyses were relatively small. Where the expected number of outcomes for the smallest group was less than ten, Fisher's exact test (154) was employed(151). Where the smallest expected number of outcomes was ten or more the Chi square (155) test was used.

Time to event analyses were conducted using Kaplan-Meier plots(156) which provide a graphical representation of survival taking into account loss to follow up during a study. The statistical significance of differences in survival rates was analysed using the log-rank test.(157)

6.2. Multivariable analyses to adjust for confounders

Multivariable analyses were used to adjust for the effect of known and suspected risk factors for adverse outcomes on the observed outcomes. Controlling for all possible confounders is likely to result in overfitting of any model, leading to results which are not reproducible in other populations.(158) Therefore, only the most important known confounders were adjusted for. The confounders controlled for during this research programme were the CPB time, type of operation performed and logistic EuroSCORE(5). The logistic EuroSCORE is an extensively-validated risk score which quantifies the risk of operative mortality following cardiac surgery based on pre-and intraoperative variables.(159, 160) Prolonged CPB has similarly been shown to be associated with increased mortality risk and prolonged length of stay.(161, 162) The operation type (valve surgery or not valve surgery) was controlled for during analysis of postoperative AF as heart valve surgery is an important risk factor for the development of postoperative AF.(163)

6.2.1. Continuous outcomes

To control for the confounding effects of other variables on continuous outcomes such as length of stay, linear regression modelling was used. Multiple linear regression allows the size of the effect of a one unit change in each variable on the outcome to be estimated. The expected value of Y (Ŷ) can be calculated by summing the products of each predictor variable's value and beta coefficient.

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 ... + b_i X_i$$

Where $\hat{Y}$ is the expected outcome value, b_0 is the value of Y when all predictor variables are equal to 0, b_1 to b_i are the beta coefficients of given predictor variables and X_1 to X_i are the values of those predictor variables. Using this equation Y can be predicted based on the known values and beta coefficients of the predictor variables.

6.2.2. Binary outcomes

For binary outcomes, confounding predictor variables are adjusted for using multivariable logistic regression modelling. This approach is similar to the multiple linear regression described above. However as the outcome is binary rather than continuous, instead of estimating the value of the outcome variable, the equation estimates the natural logarithm of the odds (also known as the logit function) of the binary outcome occurring.

$$\ln(odds) = ln\left(\frac{P}{1-P}\right) = b_0 + b_1 X_1 + b_2 X_2 + \ ... \ + \ b_i X_i$$

Where P is the probability of the binary outcome occurring, b_1 to b_i are the beta coefficients for each of the predictor variables whose values are given by X_1 to X_i.

It is possible to quantify the effect of a change in each variable on the odds of the outcome occurring. This is done using the odds ratio which, for binary predictors, is the odds of the outcome occurring if the predictor variable is present divided by the odds the outcome occurring if the predictor variable is not present. Through algebra it can be shown that the odds ratio for each predictor variable may be found by raising e to the power of the variable's beta coefficient in the equation above.(164) The odds ratio for each variable takes into account the effects of other predictor variables on the likelihood of the outcome occurring. It shows the effect on the odds of the outcome occurring if the chosen predictor changes but all other predictor variables remain constant. In other words, the odds ratios are a means of displaying the effect size related to a change in given predictor adjusted for the confounding effect of other variables. For this reason they are properly called adjusted odds ratios.

During the development of logistic regression models, Wald tests(165) were performed to assess the likelihood that the beta-coefficient for each variable was in fact 0. These probabilities are displayed as p values for each variable included in any model. It is widely accepted that a p value of <0.05 (which represents a probability of less than 5% that the outcome is not related to the predictor variable) is sufficient to conclude that the relationship is statistically significant.

6.2.3. Time-to-event outcomes

For time to event outcomes, Cox proportional hazards regression analyses were performed. These are similar to the multivariable logistic regression models. The hazard rate at any time can be found using the following equation.

$$h(t) = h_0(t)\exp(b_1 X_1 + b_2 X_2 + \cdots + b_i X_i)$$

Where *h(t)* is the hazard rate at time (*t*), h_0 is the hazard at time 0 when all predictor variables X_1 to X_i are set to 0 and b_1-b_i and X_1-X_i are the beta co-efficients and values of predictor variables. Similarly to logistic regression, *e* raised to the power of the coefficient for a given variable gives the hazard ratio for that variable. The hazard ratio shows the change in hazard rate which occurs where one variable changes and all other remain constant. It is therefore another means of displaying the effect size related to a change in a given predictor adjusted for confounders. As with logistic regression, when producing Cox proportion Hazards models in R Studio the probability that the coefficient is 0 is calculated and displayed as a p value for that variable.

6.3. Bayesian analyses

Chapter ten of this thesis describe the development and validation of a Bayesian model. Bayesian analyses are an alternative to regression analyses which also predict and outcome based on multiple dependent variables. A full description of Bayesian modelling is well beyond the scope of this thesis but a brief summary of the methodology used is presented below.

Bayesian analyses centre on the use of information already known (priors), to modify inferences drawn from the analyses of sample data (likelihood) to make revised predictions (posterior estimates). The analyses are based on the Bayes' theorem which states that.

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$$

Where A and B are events with a probability not equal to 1.

In this thesis the Bayesian methodology was used to predict future values for urine output for each individual patient based on that individual's previously observed urine output values and prior knowledge of usual patterns seen in the dataset as a whole.

6.4. Statistical evaluation of model performance

While logistic regression is mainly used to adjust for confounders in this thesis, its main use in medicine is to produce estimates expected frequencies of outcomes based on the presence of known risk factors. Logistic regression models developed using one sample of patients may not perform as well in a different sample of patients. It is therefore important to externally validate models before using them in clinical settings where the patients being analysed and the institutions in which they are being treated are different to those of the development dataset. In chapter seven of this thesis the performance of previously described models is assessed in a process known as external validation. This subsection discusses the methods used when assessing performance of risk prediction models.

There are two main characteristics to be considered when measuring model performance; discrimination and calibration. Discrimination is the ability of the model to identify which patients within a cohort are at the highest risk of an outcome. Calibration is a measure of how closely the predictions made by a model match observed outcomes in the cohort of patients being studied. Assessments of discrimination and calibration are combined with an assessment of the model's clinical usefulness to give an overall evaluation of model performance.

6.4.1. Discrimination

Discrimination is generally assessed through the analysis of receiver operating characteristic (ROC) curve plots. The output of a logistic regression model is a probability that an outcome will occur. A perfect model would assign a probability of 1 to patients who went on to suffer an outcome and 0 to those who did not. It would, therefore, have a false positive rate of 0 and a true positive rate of 1. A model which performed no better than chance would exhibit an equal number of true positive and false positive. For real world models the threshold for classifying a predicted probability to be positive can be set anywhere between 0 and 1. A ROC curve is formed by plotting the rate of true positive test results against the rate of false positive test results across the full range of possible threshold values. If the threshold value is moved closer to 1 there will be fewer false positives at the expense of identifying fewer true positives and vice versa. The ROC plot provides a graphical representation of this trade-off between the false positive rate and the true positive rate.(166) The area under the receiver operating characteristic curve (AUC) is also known as the c-statistic and is a measure of the model's ability to identify which patients are more likely than others to experience an outcome. The AUC for a perfect model would be 1 whereas that of a model performing no better than random chance would have an AUC of 0.5.

6.4.2. Calibration

Calibration is assessed by comparing outcomes predicted by the model with the observed outcomes in groups of patients within a validation cohort. In this thesis three indicators of calibration were used. The simplest is the ratio of observed to expected outcomes (O:E) ratio. The total number of expected outcomes is found by summing the predicted probabilities of the outcome for all patients. The total number of observed outcomes in the group is then divided by the expected number of outcomes. Perfect calibration would results in an O:E ratio of 1. A model which under-predicted risk would result in an O:E ratio > 1 whereas a model which over-predicted risk would result in and O:E ratio of <1.(167)

The main weakness of using overall O:E ratios is that the model's calibration may be good overall but may be poor for those with particularly high or low predicted risk. Calibration plots and the Hosmer-Lemeshow test (168) were therefore used to test risk across the full range of predicted probabilities. Calibration plots are formed by plotting the observed risk of outcome against the predicted risk of outcome for equally sized groups of patients grouped according to their predicted risk. Perfectly calibrated models would result in a plot where the observed and expect risk are identical in each group with the plotted points forming a straight line which has a gradient of 1 and passes though the origin. In reality these plots can show groups of patients for whom the model over or under predicts risk.(169) The Hosmer-Lemeshow test divides the patients into groups (normally ten) according to their predicted risk. The observed risk is then compared with the predicted risk for each group and a probability of there being no difference between the predictions and the observations is calculated. In this test a p value of >0.05 is generally accepted to represent predictions that are accurate as it indicates a high probability that any difference between predictions and observations across the groups is due to chance alone. The Hosmer-Lemeshow test results are dependent on the sample size and number of groups used in the analysis. However the main disadvantages of the test are that the test's output gives no indication of whether a model under or over-predicts risk and does not quantify the calibration error.(170)

The usefulness of a model is dependent on its calibration and its discriminative ability in the context of the purpose for which it is being used. A model which is used to identify those at greatest risk within a sample population in order to target resources appropriately can be useful even where calibration is poor if it discriminates well between those who will and will not suffer an outcome. In this scenario the accuracy of the estimated risk is not as relevant. However, if the risk predictions are to be used to risk adjust outcome data then the calibration of the model must be good. With caution, well calibrated risk predictions may also be used to inform patients and

their relatives of risk based on known predictor variables. However, predictions made by logistic regression models are only accurate for groups of patients not individuals. In a group of 100 patients who all had a predicted mortality risk of 40% (assuming a perfect model) 40 people would die. However, each individual within that high risk group will either survive or die and the model cannot tell those who will die from those who will survive.

SECTION THREE: RESULTS

Data analysed for this thesis were collected and processed in two main batches due to a delay in the initiation of waveform data recording from the Draeger Infinity bedside physiological monitors. Reasons for this delay are detailed in the Discussion section of this thesis.

The first batch of data was collected for patients who underwent cardiac surgery at Wythenshawe Hospital between January 2013 and May 2015. This batch contained data from 2284 patients and data cleaning began on this batch in June 2015. Data from all sources detailed in section four of this thesis except for the Draeger Infinity Bedside monitors were collected for these patients. After collection of waveform traces from the bedside monitors was implemented (July 2016) data from all sources were recorded for a further 1318 patients admitted to CICU between July 2016 and November 2017.

The results chapters of this thesis were written as journal articles. The first three chapters of this section only analysed data recorded between January 2013 and May 2015. In chapters ten and eleven data from both cohorts was combined to give a final dataset containing data from 3602 patients. Depending on the outcomes being studied in each chapter and often at the request of peer-reviewers, certain patient groups were excluded from different studies. Table 6-1 therefore details which patients were included in each of the studies presented in the results section of this thesis.

Table 6-1 – Patient selection for inclusion in each study contained within this results section

Relevant thesis chapter	Patient groups excluded	Patients included in chapter's analyses
Patients admitted between January 2013 and May 2015 (n=2284)		
Chapter Seven: Validation of existing mortality models	Patients whose initial CICU stay was not long enough to allow a full dataset to be collected (n=29)	2255
Chapter Eight: Incidence and outcomes of sepsis	Patients who underwent cardiac transplantation (n=54)	2230
Chapter Nine: Validation of KDIGO AKI criteria	Patients who required RRT preoperatively (n=7) and those for whom preoperative creatinine concentrations were not available (n=10)	2267
Full patient cohort - Admitted between January 2013 and November 2017 (n=3602)		
Chapter Ten: Development of novel urine output model	Patients who received mechanical circulatory support or transplantation (n=228) and patients undergoing eligible surgery but who required RRT preoperatively (n=4)	3370
Chapter Eleven: Electrolytes and AF	Patients who received mechanical circulatory support or transplantation (n=228) and patients undergoing eligible surgery but who suffered AF preoperatively (n=306)	3068

CICU – Cardiac Intensive Care Unit, KDIGO – Kidney Disease Improving Global Outcomes, AKI – Acute Kidney Injury, RRT – Renal Replacement Therapy, AF – Atrial Fibrillation/Flutter

Chapter Seven: Validation of Three Postoperative Risk Prediction Models for Intensive Care Unit Mortality after Cardiac Surgery (Published journal article)

Status: Published

Submission Date: 14/08/2017

Published online: 09/01/2018

Contributions: SHH conceived the plan for and designed the study, conducted the analyses, and drafted the manuscript. SWG, IM and CNM all developed the plan for the study. CC and MG guided the statistical analyses and the development of the model used to impute missing PAR values. SWG contributed to the production of a final draft. All authors reviewed and approved the final manuscript.

7.1. Additional data processing for this manuscript.

This study required the daily assessment of patient data to find the most abnormal value for each variable included in three post-operative risk prediction models. Once the worst values for each variable were identified they were entered into the models to calculate a risk score as described in the original papers. A number of additional data cleaning steps were required to adapt the dataset produced as described in Chapter five of this thesis for the analyses required in this study.

Firstly, the hospital's radiology database was interrogated and all incidences of computed tomography (CT) scans of the head performed on patients who were present on CICU were noted. This information was required because the presence of suspect or proven cerebrovascular accident (CVA) was a variable included in the logCASUS and RACE scores.

Secondly, it was identified that the central venous pressure (CVP) values were absent for many patients towards the end of their CICU stay. This lack of data was caused by the central venous catheter being removed in healthy patients prior to discharge to the ward. The CVP was required to calculate the pressure adjusted heart rate (PAR) variable used in the logCASUS model. To compensate for the missing CVP values a generalised linear model was developed to predict the PAR for patients with missing CVP data using heart rate and arterial blood pressure alone.

Next, a unique identifier was then created for each patient-day on the CICU. Using this unique identifier it was possible to find the most abnormal parameter for each value for each patient on each day as per the original model descriptions. For the logCASUS model (171) the worst daily value for each parameter was simply multiplied by the beta coefficient described in the original paper. For the RACE score (172) the worst value was categorised according to the thresholds described in the original paper. This categorical variable was then multiplied by the beta coefficient stated in the original paper.   Both the logCASUS and RACE score were calculated in this way for each day of the first week of each patient's first ICU admission.   The SOFA score (17) was calculated for each patient on a daily basis for the same period using the threshold values described in the original paper.

The discrimination of each model was assessed as described in the manuscript. The calibration was assessed for the logCASUS and RACE scores but not for SOFA as SOFA was not a logistic score. Finally all models were recalibrated using a subset of the patients in the cohort and the new calibration was tested in the remaining patients within the dataset.

Additional cleaning steps

Step 1. Load in all patient data.

Step 2. Select only data from the first seven days of each patient's first post-operative CICU admission.

Step 3. Create Truncated Day variables and merge with patient ID to create unique PtAndDay identifiers.

Step 4. Identify the worst daily Glasgow Coma Scale scores and calculate the worst daily CNS scores.

Step 5. Identify intubation status at the time each ABG is analysed (for SOFA score).

Step 6. Calculate daily urine output (for SOFA score).

Step 7. Identify all dose of vasopressors administered each hour for each patient.

Step 8. Create the CICU episode day number variable (required for logCASUS and RACE)

Step 9. Calculate PAR and impute missing values where CVP was unavailable and therefore prevented PAR calculation using usual methodology.

Step 10. Where F_iO_2 is not recorded use last recorded value.

Step 11. Identify worst daily value for presence of intra-aortic balloon pump.

Step 12. Identify worst daily value for presence of mechanical circulatory support.

Step 13. Identify maximum daily PAR.

Step 14. Identify minimum daily ratio of arterial to inspired oxygen concentrations. (P/F ratio)

Step 15. Identify minimum daily mean arterial blood pressure.

Step 16. Identify maximum daily creatinine concentration.

Step 17. Identify maximum daily bilirubin concentration.

Step 18. Identify maximum daily lactate concentration.

Step 19. Identify minimum daily platelet count.

Step 20. Identify worst daily intubation status.

Step 21. Identify worst daily value for presence of renal replacement therapy.

Step 22. Identify worst daily score for vasopressor use (SOFA score).

Step 23. Update worst daily score for central nervous system according to CT head data.

Step 24. Manually inspect the EPR to determine values for all variables if there is no daily score and update fields accordingly.

Step 26. Where no variable is present use the last known value. If no previous value is present use the median value for all patients.

Step 27. Calculate log CASUS score using the worst daily values and beta coefficients for each variable.

Step 28. Calculate RACE score using the worst daily values and beta coefficients for each variable.

Step 29. Calculate SOFA score using the worst daily values and beta coefficients for each variable.

Step 30. Merge in mortality data.

Step 31. Create ROC curves for all three tests.

Step 32. Create three new linear models which use the daily score calculated by each of the three models as the sole predictor of mortality.

Step 33. Recalibrate models by recalculating the mortality risk for each patient each day based on their risk score and the mortality events observed in this patient cohort

Step 34. Plot calibration curves for original and recalibrated calibrated models.

Step 35. Perform Hosmer-Lemeshow tests for original and recalibrated calibrated models.

Step 36. Calculate observed: expected mortality ratios for original and recalibrated calibrated models.

7.2. Abstract

Background

A number of cardiac surgery risk prediction models based on postoperative data have been developed. However, unlike preoperative cardiac surgery risk prediction models, postoperative models are rarely externally validated or utilised by clinicians. The objective of this study was to externally validate three postoperative risk prediction models for 30-day mortality after cardiac surgery.

Methods

The logistic Cardiac Surgery Scores (logCASUS), Rapid Clinical Evaluation (RACE) and Sequential Organ Failure Assessment (SOFA) scores were calculated over the first seven postoperative days for consecutive adult cardiac surgery patients between January 2013 and May 2015. Model discrimination was assessed using receiver operating characteristic curve analyses. Calibration was assessed using the Hosmer-Lemeshow (HL) test, calibration-plots and observed to expected ratios. Models were locally recalibrated .

Results

2255 patients were included with an ICU mortality rate of 1.8%. Discrimination for all three models on each postoperative day was good with areas under the receiver operating characteristic curve of >0.8. Generally, RACE and logCASUS had better discrimination than SOFA. Calibration of the RACE score was better logCASUS, but the ratios of observed to expected mortality for both were generally <0.65. Locally recalibrated SOFA, logCASUS and RACE models all performed well.

Conclusion

All three models demonstrated good discrimination for the first seven days after cardiac surgery. After recalibration, logCASUS and RACE scores appear to be most useful for daily risk prediction after cardiac surgery. If appropriately calibrated, postoperative cardiac surgery risk prediction models have the potential to be useful tools after cardiac surgery.

7.3. Introduction

Preoperative cardiac surgery risk prediction models such as the EuroSCORE models [1-3] have been widely adopted and extensively studied. The primary purpose of such models is to allow the risk of mortality to be estimated prior to intervention. The risk estimates can be used to inform clinical decision making when considering intervention and to risk-adjust surgical outcome data on an 'intention to treat' basis.

However, once an intervention has taken place preoperative models become less useful because intraoperative and postoperative events that may affect risk are not taken into account. Following intervention, models which include risk factors from the intra- and postoperative periods may be more useful for estimating risk and could aid postoperative clinical decision making and allow benchmarking of Cardiac Intensive Care Unit (CICU) performance.

A number of potentially useful models which analyse postoperative data have been developed. Some models designed for use in general intensive care unit (ICU) patients also accurately predict mortality after cardiac surgery,[4-14] with the Sequential Organ Failure Assessment (SOFA) score generally demonstrating the best performance.[7, 15] The Cardiac Surgery Risk Score (CASUS) and its derivatives are examples of models designed specifically for use following cardiac surgery. The CASUS model has been validated in Germany using data from multiple institutions [4,6,11,12,16] and in a study of 150 patients in Greece.[17] The derivative Logistic Cardiac Surgery Risk Score (logCASUS)[18] and Rapid Clinical Evaluation (RACE)[19] models which calculate ICU mortality risk have not been externally validated.

Despite a number of models being available, few are utilised in clinical practice. This lack of adoption may be due to the absence of comparative external validation studies in contemporary cohorts. The objective of this study was therefore to validate the logCASUS, RACE and SOFA scores for the prediction of ICU mortality in cardiac surgery patients. The performance of serial daily scores for each model was also assessed.

7.4. Patients and methods

Prospectively collected data for consecutive adult patients admitted to the CICU at our institution following cardiac surgery between 1st January 2013 and 31st May 2015 were analysed. Our institution is a tertiary adult cardiac surgery centre and our case-mix includes patients undergoing cardiac transplantation and mechanical circulatory support. As in the original studies which described the logCASUS[18] and RACE[19] scores, only data from each patient's first CICU admission after cardiac surgery were included. Patients whose CICU admissions were too short to

allow calculation of risk using the models were excluded. The primary outcome for the study was ICU mortality.

7.4.1. Data collection, validation and cleaning

Preoperative patient characteristics and postoperative outcome data were collected from the clinical governance database which is compiled by clinicians and validated by database managers. Postoperative data from the patients' CICU admissions were obtained from the electronic patient record. Results of blood analyses were obtained from the pathology laboratory database and data concerning postoperative cerebrovascular accidents were obtained from the radiology database. As described in the original studies [15,18,19] the most abnormal value for each variable recorded on each day was entered into the models. Data from all four sources were collated and cleaned using reproducible algorithms in R Studio (R Foundation for Statistical Computing).[20]

All data were entered into the Vascular Governance North West database and managed according to the protocol and ethical approvals governing this database. As data were pseudonymised prior to analysis the Research Ethics Committee concluded that ethical approval for these analyses was not necessary.

7.4.2. Missing data

Where a variable was not measured on a given day, the patient's most recent postoperative value was used to calculate the risk score. Except for bilirubin, substituted blood test data were required in <3% of risk score calculations. Previous postoperative bilirubin concentrations were not available for 8.3% of daily calculations and therefore missing values were substituted using the nearest subsequent value for that patient. Where the above substitutions were not possible due to a complete absence of data for a given patient, the median value for that parameter in all patients was imputed. Bilirubin was imputed in this way for 6.2% of patients but other variables were only imputed for 0.1% of patients.

For the logCASUS score, calculation of the pressure adjusted heart rate (PAR) (which combines information from heart rate, central venous pressure (CVP) and mean arterial pressure) was not possible for 7% of score calculations. This was most commonly because the central venous catheter had been removed before CICU discharge. To address missing CVP data, a logistic regression model was developed using data from patients for whom data were complete. This model was then used to calculate a modelled PAR. There were no missing outcome data.

7.4.3. Statistical analyses

Central tendency of variables is described using mean and standard deviation where the distribution was parametric and median and interquartile range where the distribution was non-parametric.

The logCASUS, RACE and SOFA scores were calculated for each patient on a daily basis for postoperative days one to seven. The discrimination of all scores for the prediction of ICU mortality was assessed using the area under the receiver operating characteristic curve (AUC). De Long's method for calculating AUC variance was used for the calculation of AUC 95% confidence intervals.[21] AUC values of ≥0.7 were considered acceptable and values of ≥0.8 were considered good.

The calibration of logCASUS and RACE ICU mortality estimates was assessed using the ratio of observed outcomes to predicted outcomes (O:E ratio), the Hosmer Lemeshow (HL) test and calibration plots. A high HL χ^2 value with a low p value suggests that there is a significant difference between predicted risk and observed outcomes.[22] The calibration plots illustrate how the mean predicted probability of ICU mortality compares with the observed incidence of ICU mortality for five equally sized groups based on the ranked predicted risks calculated by the model. Calibration of the original SOFA score could not be evaluated because it is a non-logistic score.

A sub-group analysis excluding patients who underwent cardiac transplantation or initiation of mechanical circulatory support was also performed. Finally, local recalibration of the models was performed. Data were divided randomly into two equally sized datasets; a training dataset and an evaluation dataset. Each model was fully recalibrated using data for each variable from the training dataset. The calibration and discrimination of each recalibrated model was then tested in the evaluation dataset.

7.5. Results

Patient characteristics

There were 2284 consecutive patients who met the inclusion criteria. 29 patients were excluded because their admission to CICU was too short to allow calculation of the risk scores. The mean (sd) age of the patients was 65.8 (11.8) years and 27.3% were female. The most common procedure was isolated coronary artery bypass graft surgery (53.3%). Additional patient and operative characteristics are shown in Table 7-1. The overall ICU mortality rate was 2.0%. The ICU mortality rate in the final validation cohort was 1.8%.

Table 7-1 - Patient characteristics in the validation cohort

Variable	Total n =2,255
Age, mean (sd), years	65.8 (11.8)
Female Gender, n(%)	616 (27.3)
Height, mean (sd), cm	170 (9.2)
Weight, mean (sd), Kg	81 (15.8)
Surgery	
Isolated CABG, n(%)	1202 (53.3)
Isolated Valve, n(%)	477 (21.2)
Isolated Aortic, n(%)	23 (1.0)
Combined cardiac procedures, n(%)	397 (17.6)
Cardiac transplantation, n(%)	51 (2.3)
Mechanical circulatory support, n(%)	41 (1.8)
Other, n(%)	64 (2.8)
Urgency	
Elective/Scheduled, n(%)	1309 (58.0)
Urgent, n(%)	884 (39.2)
Emergency/Salvage, n(%)	62 (2.7)
CPB Duration, median(Interquartile range), minutes	101.0 (80.0-130.0)
Logistic EuroSCORE, median (Interquartile range)	4.0 (2.1-7.7)

CABG=Coronary Artery Bypass Graft, CPB=Cardiopulmonary Bypass

7.5.1. Model performance on the first postoperative day

The variables included in each model are detailed in Table 7-2. A day-by-day description of the levels of risk predicted by the models is shown in the Appendix (Table 7-5). All three models demonstrated good discrimination when calculated on the first postoperative day (Figure 7-1a). The AUC for the RACE and logCASUS scores were the same at 0.94 (95%CI 0.91-0.97) for both. The AUC for the SOFA score was 0.91 (95%CI 0.86-0.96). The HL tests, together with the comparison of the O:E ratios and calibration plots implied poor calibration of both logistic models (Table 7-3). As seen in Figure 7-1b, predictions were least accurate for those patients who had the highest predicted risk. Sub-group analysis demonstrated no significant effect on model performance when patients undergoing cardiac transplantation or initiation of mechanical circulatory support were excluded.

Table 7-2 - Risk factors and variables included in the analysed models

System	logCASUS	RACE	SOFA
Cardiovascular	PAR	MAP	MAP
	Lactate concentration	Lactate concentration	Vasoactive medication
	Ventricular assist device	Ventricular assist device	
	Intra-aortic balloon pump	Intra-aortic balloon pump	
Respiratory	P_aO_2/F_iO_2 ratio	Intubation	P_aO_2/F_iO_2 ratio
Coagulation	Platelet count	Platelet count	Platelet count
Hepatobiliary	Bilirubin concentration	Bilirubin concentration	Bilirubin concentration
Renal	Creatinine concentration	Creatinine concentration	Creatinine concentration
	Renal replacement therapy	Renal replacement therapy	Urine output
Central Nervous	Neurological state score	Neurological state score	Glasgow coma scale
Other	Postoperative day number	Postoperative day number	

PAR=Pressure adjusted heart rate=heart rate x central venous pressure/mean arterial pressure, MAP=mean arterial pressure, P_aO_2=arterial partial pressure of oxygen, F_iO_2=inspired fraction of oxygen.
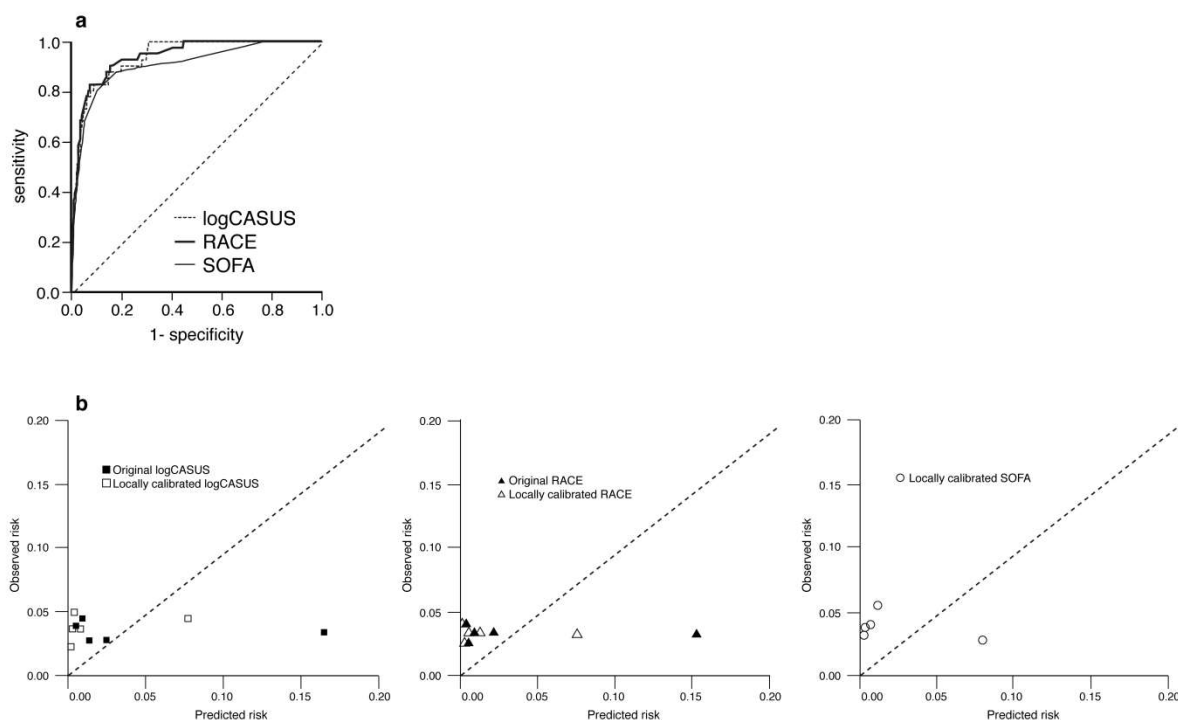
Figure 7-1 (a) Receiver Operating Characteristic (ROC) curves for the validated models on the first postoperative day. (b) Calibration plots for the original logCASUS and RACE models and recalibrated logCASUS, RACE and SOFA models on the first postoperative day.

The dashed line represents the line of perfect calibration.

7.5.2. Serial scores

The daily measures of discrimination and calibration for the models are shown in Table 7-3. LogCASUS and RACE scores calculated daily up to day seven of the postoperative CICU admission demonstrated good discrimination. The AUC of the SOFA score was generally lower than those of the cardiac surgery-specific scores in the early postoperative period but the difference reduced towards the end of the first postoperative week. Calibration plots, HL test and O:E ratios suggested poor calibration for logCASUS. Calibration for RACE was better but remained suboptimal.

Table 7-3 - Daily performance of the original models for ICU mortality

Day No (N)	Number of events	logCASUS AUC (95% CI)	HLχ^2 (p value)	O:E ratio	RACE AUC (95% CI)	HLχ^2 (p value)	O:E ratio	SOFA AUC (95% CI)
1(2255)	41	**0.94 (0.91-0.97)**	43.81 (<0.01)	0.41	0.94 (0.91-0.97)	22.84 (<0.01)	0.52	**0.91 (0.86-0.96)**
2(1705)	36	0.92 (0.85-0.98)	23.36 (<0.01)	0.52	0.93 (0.88-0.99)	5.04 (0.75)	0.87	0.90 (0.84-0.96)
3(957)	27	0.93 (0.90-0.96)	51.77 (<0.01)	0.34	**0.94 (0.91-0.98)**	14.42 (0.07)	0.59	0.89 (0.82-0.96)
4(607)	21	0.93 (0.89-0.96)	40.99 (<0.01)	0.35	0.90 (0.85-0.95)	12.54 (0.12)	0.57	0.91 (0.86-0.96)
5(399)	21	0.88 (0.80-0.95)	36.12 (<0.01)	0.41	0.87 (0.81-0.94)	9.39 (0.31)	0.63	0.88 (0.82-0.95)
6(273)	17	0.89 (0.83-0.95)	31.12 (<0.01)	0.39	0.87 (0.79-0.95)	13.94 (0.08)	0.60	0.86 (0.78-0.93)
7(183)	16	0.84 (0.73-0.95)	27.18 (<0.01)	0.45	0.85 (0.76-0.94)	12.42 (0.13)	0.62	0.88 (0.81-0.96)

AUC=Area under the receiver operation characteristic, HL=Hosmer-Lemeshow, O:E ratio=observed expected ratio. The best AUCs are given in bold

7.5.3. Local recalibration

The AUC, O:E ratios and HL test results for the recalibrated models' performance in the evaluation dataset are detailed in Table 7-4. The analyses of recalibrated model performance were limited to the first five days as the training dataset only contained seven patients who died after being on CICU for more than 5 days. Local recalibration of models using the training dataset generally resulted in marginal improvement in discrimination for all scores. The calibration of the recalibrated logCASUS model was adequate on every day. The HL tests for RACE and SOFA suggested adequate calibration on every day except day 5. Calibration plots for the original and recalibrated models are shown in Figure 7-1b. Full details of the recalibrated models are provided in Tables 7-6 -7-8.

Table 7-4 - Daily performance of the models for ICU mortality in the evaluation dataset following local recalibration

Day No (N)	Number of events	LogCASUS AUC (95% CI)	HLχ^2 (p value)	O:E ratio	RACE AUC (95% CI)	HLχ^2 (p value)	O:E ratio	SOFA AUC (95% CI)	HLχ^2 (p value)	O:E ratio
1(1127)	18	0.96 (0.93-0.99)	2.79 (0.94)	0.85	0.96 (0.92-0.99)	4.77 (0.78)	0.82	0.93 (0.87-0.99)	5.75 (0.67)	0.76
2(850)	15	**0.96 (0.94-0.99)**	5.15 (0.74)	0.67	**0.97 (0.95-0.99)**	4.24 (0.83)	0.72	**0.93 (0.88-0.98)**	3.29 (0.91)	0.73
3(461)	12	0.92 (0.85-0.99)	5.23 (0.73)	0.85	0.94 (0.88-1.00)	3.11 (0.93)	0.80	0.92 (0.88-0.97)	3.60 (0.89)	0.76
4(303)	11	0.93 (0.88-0.97)	2.28 (0.97)	1.14	0.90 (0.83-0.97)	10.11 (0.25)	0.92	0.87 (0.79-0.95)	14.23 (0.08)	0.96
5(195)	11	0.92 (0.86-0.98)	4.39 (0.82)	1.26	0.80 (0.62-0.98)	22.14 (<0.01)	1.26	0.87 (0.80-0.95)	18.46 (0.02)	1.14

AUC=Area under the receiver operation characteristic, HL=Hosmer-Lemeshow, O:E ratio=observed expected ratio. Best AUCs in bold

7.6. Discussion

This study has validated the performance of the logCASUS, RACE and SOFA scores in a cohort of 2255 patients from a tertiary cardiac centre in the UK. The observed ICU mortality (1.8%) is in line with that for all cardiac surgery in the UK [23]. This is despite the cohort including 62 patients who underwent emergency/salvage procedures and 41 who underwent instigation of mechanical circulatory support. In these groups ICU mortality was 21.0% and 24.3% respectively. In the remaining patients the ICU mortality rate was 1.2%.

All models demonstrated good discrimination throughout the first postoperative week with discrimination declining slightly towards the end of the week. Both the logCASUS and RACE scores demonstrated poor calibration in our cohort and significantly over-predicted risk. The poor calibration demonstrated by the RACE and CASUS models may be due to a similar calibration drift effect to that observed with preoperative risk models due to improvements in care and clinical outcomes over time.[24] Assuming that our findings of poor calibration of the original models are replicated elsewhere, the models would need to be recalibrated before the risk estimates could be clinically useful.[25] After recalibration in our training cohort, all models demonstrated improved calibration as assessed by the HL test but under-predicted risk in lower risk groups and over-predicted risk in the highest risk group. Overall, the logCASUS was slightly better calibrated that the SOFA and RACE scores.

This study represents the first external validation of the logCASUS and RACE models and the first validation of the SOFA score in UK cardiac surgery. We utilised contemporary data from a tertiary cardiac centre with excellent clinical results and undertook a comprehensive assessment of model performance. As with any clinical study, there were missing data but the proportion in this study was low. The variables with the most missing data were PAR (required for logCASUS calculation only) and serum bilirubin. A clinically robust approach to handling missing data was adopted.

A potential limitation of the study is that it is based on data from a single centre and includes relatively few outcomes. The division of our dataset for development and evaluation of locally recalibrated versions of the models exacerbated this problem and so we limited the evaluation of the recalibrated models to the first five postoperative days. When the number of outcomes is low, validation results and performance statistics need to be interpreted with caution. Potential options to increase the sample size would be to expand the number of participating centres or increase the timeframe for data collection. The first option is not feasible at present due to a lack of UK centres currently collecting the necessary data in a suitable format. Although the second

option is feasible, this would introduce temporal issues related to changes in practice and performance over time that could affect the results.[25]

The SOFA score was the first of the validated models to be developed and was designed based on expert consensus in 1996 to assess the progress of patients suffering from sepsis.[15] It has since been validated as a prediction tool for adverse outcomes in general ICU patients[26] and also specifically in patients who have undergone cardiac surgery.[7, 10] It grades the function of six organ systems using a five point scale for each with totals ranging between zero and 24 (Tables 7-2 and 7-6).

The logistic logCASUS[18] and RACE[19] scores were developed by the same team as the additive CASUS score[4] using data from a single centre. The RACE score was designed as a user friendly version of logCASUS. Unlike SOFA, these scores were developed exclusively for patients who have undergone cardiac surgery. They both include use of mechanical circulatory support and intra-aortic balloon pump counterpulsation which are more common in cardiac surgery patients than the general ICU population. Both scores grade neurological status using a scale which reduces the impact of the low conscious level expected immediately after cardiac surgery and adjust the predicted risk based on the time since surgery (Tables 7-7 and 7-8).

During the recalibration process we were able to identify which variables within each model were significantly associated with ICU mortality. We found that variables which can be controlled by physicians such as MAP and PAR were not significantly associated with outcome. Conversely interventions which may affect those parameters such as use of mechanical ventilation, renal replacement therapy, or mechanical circulatory support were shown to have significant and relatively large effects on risk. Serum creatinine and lactate concentrations and the platelet count were the most significant of the blood analyses assessed (Tables 7-6 – 7-8).

Despite generally superior discriminatory ability compared to preoperative models,[27] none of the validated models has been widely adopted in clinical practice. Possible reasons for this include problems with the ability to easily calculate the scores, a perceived lack of clinical utility or validity and inadequate external validation studies. Clinical utility of the scores is limited by the fact that they are inherently retrospective as they rely on the worst value obtained over a 24-hour period. Importantly, the cardiac surgery-specific scores are designed to be used only during a patient's first admission to CICU. In addition, in these models the first postoperative day is the reference for the "ICU day" variable, i.e. the beta-coefficient for postoperative day one is zero. The models do not provide an "ICU day" coefficient for the operative day and consequently cannot produce risk estimates until the first postoperative day. As a result of these limitations in model design,

the logCASUS and RACE risk scores could not be calculated for 29 patients for whom the first ICU episode finished on the operative day. This is clinically relevant because for 24 of these patients, the short initial admission was due to reoperation for bleeding and the ICU mortality rate in this group was 17%. Although this did not significantly change the overall ICU mortality rate (2.0% versus 1.8%) the inability to assess risk in these patients is a limitation of the models.

Despite limitations, the models studied could potentially be utilised for three main purposes. All models discriminate well between patients at high and low risk of mortality therefore clinicians could use the scores to identify patients with the highest risk amongst those present on CICU and to target resources including staff allocations accordingly. Secondly, if validated in multicentre-studies the models could be utilised for the risk-adjustment of CICU benchmarking data in a similar way to that in which preoperative models have been used to risk-adjust surgical outcome data. [28] Utilising models that include postoperative variables to generate risk predictions allows risk estimates to be modified by the occurrence of intra- and early postoperative events. Such estimates would be better suited for the assessment of CICU performance. If the models were to be used for this purpose risk predictions should be made early in the CICU stay because scores calculated later are likely to be influenced by the quality of care already received on the CICU.

Comparison of serial scores recorded in different institutions could however be used determine possible deficiencies in care. If it was noted that for one institution, risk predictions made at a certain point in the postoperative stay generally increased in a manner which differed from predictions made in other institutions at a similar time-point, the protocols and procedures in use during that period of the CICU stay could be analysed to identify possible improvements.

Lastly, if the models are appropriately calibrated, daily scores could be used to assess patient progress, assist clinical decision making and to inform discussions with patients and their relatives by providing the most up to date assessment of patient risk possible. Although predictions will never be completely accurate for individual patients, they may be used as a guide.

In conclusion, all three models showed good discrimination when used during the first postoperative week after cardiac surgery. In their original forms the cardiac surgery specific models were poorly calibrated, particularly in patients with the highest risk, but all three models could be recalibrated using local data. Further research into optimising postoperative models to maximise their clinical utility is required if they are going to be widely adopted.

7.7 Appendices

Table 7-5 - The proportion of patients with low, medium and high predicted ICU mortality risk on each postoperative day

Model	Estimated ICU mortality risk	Percentage of Patients with stated level of mortality risk (%)						
		Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
LogCASUS	<2%	89.4	87.9	82.6	78.5	72.8	70.6	65.9
	2-10%	7.6	6.8	10.0	14.5	16.4	18.4	17.6
	>10%	3.0	5.3	7.4	6.9	10.8	11.0	16.5
RACE	<2%	86.7	88.4	82.6	77.6	72.8	75.7	61.5
	2-10%	10.0	7.9	10.4	14.9	16.4	11.8	18.7
	>10%	3.3	3.8	6.9	7.6	10.8	12.5	19.8
SOFA	<2%	82.7	81.6	75.3	75.6	72.3	66.9	65.9
	2-10%	13.3	13.5	18.0	16.2	16.9	23.5	18.7
	>10%	4.0	4.8	6.7	8.3	10.8	9.6	15.4

Table 7-6 - Beta coefficients for the recalibrated SOFA score when predicting ICU mortality

System	Factor	Estimate	95% CI		P value
	(Intercept)	-6.34	-9.34	-4.53	<0.01 *
Respiratory	PF ratio ≥400mmHg	0.00			-
	PF ratio <400mmHg	0.63	-1.11	3.58	0.56
	PF ratio <300mmHg	0.29	-1.36	3.20	0.79
	PF ratio <200mmHg and MV	1.60	-0.12	4.54	0.14
	PF ratio <100mmHg and MV	1.96	0.07	4.96	0.08
Coagulation	Platelets >150 x10^9 L^{-1}	0.00			-
	Platelets 100-150 x10^9 L^{-1}	0.75	0.09	1.43	0.03
	Platelets 50-99 x10^9 L^{-1}	1.20	0.50	1.91	<0.01 *
	Platelets 20-49 x10^9 L^{-1}	2.25	1.04	3.43	<0.01 *
	Platelets ≤20 x10^9 L^{-1}	1.36	-2.26	4.72	0.41
Liver	Bilirubin <1.2 mgdL^{-1}	0.00			-
	Bilirubin 1.2-1.9 mgdL^{-1}	0.24	-0.41	0.87	0.46
	Bilirubin 2.0-5.9 mgdL^{-1}	0.28	-0.39	0.93	0.41
	Bilirubin 6.0-11.9 mgdL^{-1}	1.73	0.18	3.21	0.02 *
	Bilirubin ≥12 mgdL^{-1}	15.94	-50.10	-	0.98
CNS	GCS 15	0.00	-	-	-
	GCS 13-14	-0.57	-2.03	0.50	0.36
	GCS 10-12	0.22	-1.99	1.68	0.81
	GCS 6-9	1.01	-0.93	2.44	0.22
	GCS <6	1.25	1.11	2.25	0.02 *
Renal	Creatinine <1.2 mgdL^{-1}	0.00			-
	Creatinine 1.2-1.9 mgdL^{-1}	0.31	-0.52	1.07	0.44
	Creatinine 2.0-3.4 mgdL^{-1}	1.83	1.16	2.50	<0.01 *
	Creatinine 3.5-4.9 mgdL^{-1} **or** UO < 500ml day^{-1}	0.46	2.30	2.08	<0.01 *
	Creatinine ≥5.0 mgdL^{-1} **or** UO <200ml day^{-1}	2.02	1.26	2.79	<0.01 *
Cardiovascular	MAP ≥70mmHg	0.00			-
	MAP < 70mmHg	0.57	-0.46	1.75	0.30
	Dopamine ≤5μgkg^{-1}min^{-1} or Dobutamine	1.34	0.55	2.85	<0.01 *
	Dopamine >5μgkg^{-1}min^{-1} or Adrenaline ≤0.1μgkg^{-1}min^{-1} or Noradrenaline ≤0.1μgkg^{-1}min^{-1}	0.34	-0.62	1.48	0.52
	Dopamine >15μgkg^{-1}min^{-1} or Adrenaline >0.1μgkg^{-1}min^{-1} or Noradrenaline >0.1μgkg^{-1}min^{-1}	1.43	0.44	2.60	<0.01 *

PF ratio - ratio of arterial partial pressure of oxygen to fraction inspired oxygen; GCS – Glasgow Coma Scale; UO – urine output; *=p value <0.05

Table 7-7 - Beta coefficients for the recalibrated logCASUS score when predicting ICU mortality

Factor	β-coefficient				β-coefficient in original study
	Estimate	95%CI		P value	
(Intercept)	-5.05	-6.72	-3.43	<0.01*	-5.65
Minimum daily PF ratio (mmHg)	-0.01	-0.03	0.01	0.47	0.00
Maximum Daily Creatinine (mgdL^{-1})	0.01	0.00	0.01	<0.01*	0.31
Renal Replacement Therapy - No	0.00			-	0.00
Renal Replacement Therapy - Yes	1.39	0.79	2.01	<0.01*	0.41
Maximum Daily Bilirubin (mgdL^{-1})	0.01	-0.00	0.02	0.10	0.22
Maximum Daily PAR	0.01	-0.02	0.04	0.42	0.06
Maximum Daily Lactate (mmolL^{-1})	0.26	0.11	0.41	<0.01*	0.21
IABP – No	0.00			-	0.00
IABP - Yes	0.55	-0.08	1.15	0.08	0.68
VA ECMO or VAD -No	0.00				0.00
VA ECMO or VAD - Yes	1.28	0.62	1.93	<0.01*	2.30
Minimum Daily Platelets(x10^{9}L^{-1})	-0.01	-0.02	-0.00	<0.01*	0.00
Neurological state - Normal	0.00			-	0.00
Neurological state - Confused	0.39	-0.70	1.35	0.45	0.47
Neurological state - Sedated	0.81	0.08	1.56	0.03*	0.70
Neurological state - Focal Neurological Damage	3.01	2.00	3.98	<0.01*	1.47
ICU Day 1	0.00			-	0.00
ICU Day 2	0.31	-0.46	1.09	0.43	0.01
ICU Day 3	0.25	-0.62	1.11	0.57	0.84
ICU Day 4	0.24	-0.75	1.19	0.62	1.04
ICU Day 5	0.44	-0.59	1.43	0.39	1.26

PF ratio = ratio of arterial partial pressure of oxygen to fraction inspired oxygen, PAR – Pressure adjusted Heart Rate; IABP – Intra-aortic balloon pump; VA ECMO – venoarterial extracorporeal membrane oxygenation; VAD – ventricular assist device.*=p value <0.05

Table 7-8– Beta coefficients for the recalibrated RACE score when predicting ICU mortality

Factor	β-coefficient				β-coefficient in original study
	Estimate	95% CI		P value	
(Intercept)	-6.41	-7.38	-5.53	<0.01*	-6.23
Intubated - No	0.00			-	0.00
Intubated - Yes	1.80	0.88	2.71	<0.01*	0.66
Creatinine <1.2 mgdL^{-1}	0.00			-	0.00
Creatinine 1.2-4.0 mgdL^{-1}	0.55	0.01	1.10	0.04*	0.74
Creatinine >4.0 mgdL^{-1}	-12.4	-	29.14	0.98	0.88
Renal Replacement Therapy - No	0.00			-	0.00
Renal Replacement Therapy - Yes	1.62	1.01	2.26	<0.01*	0.57
Bilirubin <1.2 mgdL^{-1}	0.00			-	0.00
Bilirubin 1.2-7.0 mgdL^{-1}	0.19	-0.41	0.80	0.53	0.60
Bilirubin >7.0 mgdL^{-1}	1.96	0.10	4.17	0.05	2.01
MAP >70mmHg	0.00			-	0.00
MAP 50-70mmHg	0.34	-0.31	1.04	0.32	0.45
MAP < 50mmHg	0.60	-2.04	2.46	0.59	0.91
Lactate <2.1 mmol L^{-1}	0.00			-	
Lactate 2.1-8.0 mmol L^{-1}	0.34	-0.24	0.93	0.25	0.80
Lactate >8.0 mmol L^{-1}	1.86	0.35	3.33	0.01*	2.40
IABP – No	0.00			-	0.00
IABP – Yes	0.50	-0.12	1.10	0.11	0.74
VA ECMO or VAD - No	0.00			-	0.00
VA ECMO or VAD - Yes	1.31	0.64	1.99	<0.01*	1.90
Platelets >120 x 10^{9}L^{-1}	0.00			-	0.00
Platelets 51-120 x 10^{9}L^{-1}	0.43	-0.18	1.04	0.16	0.54
Platelets <51 x 10^{9}L^{-1}	1.85	0.73	2.93	<0.01*	0.94
Neurological State - Normal	0.00			-	0.00
Neurological State - Confused	0.20	-0.99	1.22	0.72	0.44
Neurological State - Sedated	-0.62	-1.65	0.45	0.24	0.75
Neurological State - Cerebral Damage	2.27	1.14	3.38	<0.01*	1.25
ICU Day 1	0.00			-	0.00
ICU Day 2	0.60	-0.18	1.38	0.13	0.01
ICU Day 3	0.54	-0.33	1.40	0.22	0.84
ICU Day 4	0.44	-0.56	1.40	0.37	1.04
ICU Day 5	0.49	-0.55	1.47	0.34	1.26

MAP – mean arterial pressure; IABP – Intra-aortic balloon pump; VA ECMO – venoarterial extracorporeal membrane oxygenation; VAD – ventricular assist device. * = p value<0.05

Recalibration was performed by making new logistic regression models using data from the training dataset. The models all predicted ICU mortality in the training dataset using variables described in the original studies. Where variables had been categorised in the original studies, the same thresholds for categorisation were used in the recalibration process. For some variables, despite containing data from over 1000 patients, the training dataset did not contain enough patients in each category for each variable to allow a sensible beta coefficient for that category to be generated. This can be seen in particular with the coefficient for >12mg/dl in the SOFA score and creatinine >4mg/dL in the RACE score.

7.8. References

[1]   Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16:9-13.

[2]   Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. Eur Heart J 2003;24:881-2

[3]   Nashef SA, Roques F, Sharples LD et al. EuroSCORE II. Eur J Cardiothorac Surg 2012;41:734-744; discussion 744-735.

[4]   Hekmat K, Kroener A, Stuetzer H, et al. Daily assessment of organ dysfunction and survival in intensive care unit cardiac surgical patients. Ann Thorac Surg 2005;79:1555-1562.

[5]   Howitt SH, Grant SW, Riding DM, Malagon I, McCollum CN. (in press) Risk models that utilise postoperative patient monitoring data to predict outcomes in adult cardiac surgery; a systematic review. J Cardiothorac Vasc Anesth.

[6]   Hekmat K, Doerr F, Kroener A et al. Prediction of mortality in intensive care unit cardiac surgical patients. Eur J Cardiothorac Surg 2010;38:104-109.

[7]   Doerr F, Badreldin AM, Heldwein MB et al. A comparative study of four intensive care outcome prediction models in cardiac surgery patients. J Cardiothorac Surg.2011;6. DOI:10.1186/1749-8090-6-21

[8]   Ceriani R, Mazzoni M, Bortone F et al. Application of the sequential organ failure assessment score to cardiac surgical patients. Chest 2003;123:1229-1239.

[9]   Patila T, Kukkonen S, Vento A, Pettila V, Suojaranta-Ylinen R. Relation of the Sequential Organ Failure Assessment Score to Morbidity and Mortality After Cardiac Surgery. Ann Thorac Surg 2006;82:2072-2078.

[10]   Gomes RV, Tura B, Mendonca HT et al. A first postoperative day predictive score of mortality for cardiac surgery. Ann Thorac Cardiovasc Surg 2007;13:159-164.

[11]   Badreldin A, Elsobky S, Lehmann T, Brehm BB, Doenst T, Hekmat K. Daily-mean-SOFA, a new derivative to increase accuracy of mortality prediction in cardiac surgical intensive care units. Thorac Cardiovasc Surg 2012;60:43-50.

[12]   Doerr F, Badreldin AM, Can F, Bayer O, Wahlers T, Hekmat K. SAPS 3 is not superior to SAPS 2 in cardiac surgery patients. Scand Cardiovasc J 2014;48:111-119.

[13]   Ariyaratnam P, Loubani M, Biddulph J et al. Validation of the intensive care national audit and research centre scoring system in a UK adult cardiac surgery population. J Cardiothorac Vasc Anesth 2015;29:565-569.

[14]   Heldwein MB, Badreldin AM, Doerr F et al. Logistic Organ Dysfunction Score (LODS): a reliable postoperative risk management score also in cardiac surgical patients? J Cardiothorac Surg.2011;6. DOI:10.1186/1749-8090-6-110

[15]   Vincent JL, Moreno R, Takala J et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22:707-710.

[16]   Badreldin AM, Doerr F, Ismail MM, et al. Comparison between Sequential Organ Failure Assessment score (SOFA) and Cardiac Surgery Score (CASUS) for mortality prediction after cardiac surgery. Thorac Cardiovasc Surg 2012;60(1):35-42.

[17]   Exarchopoulos T, Charitidou E, Dedeilias P, Charitos C, Routsi C. Scoring Systems for Outcome Prediction in a Cardiac Surgical Intensive Care Unit: A Comparative Study. Am J Crit Care 2015;24:327-334.

[18]   Doerr F, Badreldin AM, Bender EM et al. Outcome prediction in cardiac surgery: the first logistic scoring model for cardiac surgical intensive care patients. Minerva Anestesiol 2012;78:879-886.

[19]   Badreldin AM, Doerr F, Bender EM et al. Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices. Eur J Cardiothorac Surg 2013;44:992-997.

[20]   R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing 2015.

[21]   DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Bio-metrics 1988;44:837-45.

[22]   Hosmer DW, Lemeshow S. Applied logistic regression(2nd Edition). New York:Wiley; 2000;147-155.

[23]   Bridgewater B, Grant SW, Hickey GL, et al. National Adult Cardiac Surgery Audit Report: National Institute for Cardiovascular Outcomes Research; 2012 2012.

[24]   Hickey GL, Grant SW, Murphy GJ et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. Eur J Cardiothorac Surg 2013;43:1146-1152.

[25]   Grant SW, Grayson AD, Jackson M et al. Does the choice of risk-adjustment model influence the outcome of surgeon-specific mortality analysis? A retrospective analysis of 14637 patients under 31 surgeons. Heart 2008;94:1044-1049.

[26]   Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001;286:1754-1758.

[27]   Grant SW, Hickey GL, Dimarakis I et al. How does EuroSCORE II perform in UK cardiac surgery; an analysis of 23 740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. Heart 2012;98:1568-1572.

[28]   Grant SW, Hickey GL, Cosgriff R et al. Creating transparency in UK adult cardiac surgery data. Heart 2013;99:1067-1068.

Chapter Eight: Incidence and outcomes of sepsis after cardiac surgery as defined by the Sepsis-3 guidelines (Published journal article)

Status: Published

Submission Date: 28/07/2017

Published online: 24/11/2017

Reference: Howitt SH, Herring M, Malagon I, McCollum CN, Grant SW. Incidence and outcomes of sepsis after cardiac surgery as defined by the Sepsis-3 guidelines. *Br J Anaesth* 2018; 120: 509-16

Author contributions:

Study Design: S.H.H., I.M., C.N.M., S.W.G.

Data Collection and cleaning: S.H.H., M.H.

Data analysis: S.H.H., M.H., I.M., C.N.M., S.W.G.

First draft of manuscript: S.H.H.

Revision of Manuscript: S.H.H., M.H., I.M., C.N.M., S.W.G.

Rationale for inclusion of this chapter in the thesis

The Sepsis-3 guidelines were developed to aid the early diagnosis of sepsis in a range of health care settings. They define sepsis and provide criteria for its diagnosis. While the guidelines were not designed specifically to be used on CICU following cardiac surgery, this chapter set out to examine whether they could identify those at risk of adverse complications within this cohort. If an association between sepsis according to the new definitions and poor outcomes could be demonstrated then the first instance where the Sepsis-3 criteria were fulfilled could be used as an endpoint. This endpoint could then form the target of future risk predictions models which aim to predict sepsis in cardiac surgery patients.

8.1. Additional data processing for this manuscript.

The cleaning code used to conduct the analyses for this manuscript built upon the code created during the production of the paper which validated the SOFA score in the previous chapter. The daily SOFA scores for each patient were analysed and where a rise of ≥2 in the SOFA score was detected the date and time of the SOFA rise was recorded. The code produced an index file of all times when the SOFA score had increased by ≥2. The clinical case notes and medication records within the EPR were the examined manually and each incidence of SOFA rise was inspected to identify whether clinical suspicion of infection, initiation of antimicrobial therapy or microbiologically proven infection had been recorded. Where suspected or proven infection coincided with a rise of ≥2 in the SOFA score, sepsis with suspected or proven infection was diagnosed. Length of CICU stay, 30-day survival and 2-year survival were compared for those with sepsis, SOFA rise in the absence of infection and those without a SOFA rise ≥2.

<u>Additional cleaning steps</u>

Step 1. Load in all the data from the validation paper.

Step 2. Compare every SOFA score for each patient with all previous values.

Step 3. Identify all days where the SOFA score rises by ≥2.

Step 4. Manually lookup all incidences of a SOFA rise ≥2 in the EPR identifying suspected or proven infection, source of the infection and causative organism where infection was proven.

Step 5. Collate all data regarding infections, sepsis and outcomes.

Step 6. Compare associations between sepsis category and outcomes using univariable analyses.

Step 7. Adjust for CPB time and logistic EuroSCORE using multivariable logistic regression comparing each group with "No sepsis" group.

Step 8. Create Table 8-1.

Step 9. Compare outcomes for sepsis with proven infection with those for suspected infection

Step 10. Compare-2 year survival rates between the different groups using Kaplan-Meir plots, Log rank test and Cox proportional hazards regression.

8.2. Summary (Abstract)

Background

The Sepsis-3 guidelines diagnose sepsis based on organ dysfunction in patients with either proven or suspected infection. The objective of this study was to assess the incidence and outcomes of sepsis diagnosed using these guidelines in patients on the Cardiac Intensive Care Unit (CICU) after cardiac surgery.

Methods

Daily Sequential Organ Failure Assessment (SOFA) scores were calculated for 2230 consecutive adult cardiac surgery patients between January 2013 and May 2015. Patients with a rise in SOFA score of ≥2 and suspected or proven infection were identified. The length of CICU stay, 30-day mortality and 2-year survival were compared between groups. Multivariable linear regression, multivariable logistic regression and Cox proportional hazards regression were used to adjust for possible confounders.

Results

Sepsis with suspected or proven infection was diagnosed in 104 (4.7%) and 107 (4.8%) patients respectively. After adjustment for confounding variables, sepsis with suspected infection was associated with an increased length of CICU stay of 134.1 (95%CI 99.0-168.2) hours ($p<0.01$) and increased 30-day mortality risk (odds ratio 3.7, 95%CI 1.1-10.2, p=0.02). Sepsis with proven infection was associated with an increased length of CICU stay of 266.1 (95%CI 231.6-300.7) hours ($p<0.01$) and increased 30-day mortality risk (odds ratio 6.6, 95%CI 2.6-15.7, $p<0.01$).

Conclusions

Approximately half of sepsis diagnoses were based on proven infection and half on suspected infection. Patients diagnosed with sepsis using the Sepsis-3 guidelines have significantly worse outcomes after cardiac surgery. The Sepsis-3 guidelines are a potentially useful tool in the management of sepsis following cardiac surgery.

Word count 250

8.3. Introduction

The Sepsis-3 guidelines were introduced in 2016 and define sepsis as organ dysfunction in the presence of proven or suspected infection.[1] In the critical care setting, organ dysfunction identified using the Sequential Organ Failure Assessment (SOFA) score (Table 8-1)[2] replaces the Systemic Inflammatory Response Syndrome (SIRS) as the means by which the adverse physiological effects of infection are identified.[3] According to the new guideline, suspected or proven infection with proven organ dysfunction (defined as an increase of ≥2 in SOFA score) results in the diagnosis of sepsis. In previous definitions, suspected infection could only result in a diagnosis of suspected sepsis until infection was proven on microbiological culture.[3 4]

There are limited published data on the frequency of sepsis following cardiac surgery. Previous studies often limited their investigations to patients with positive microbiological cultures from specific sites such as the wound or the respiratory tract. Such studies used previous definitions of sepsis and identified sepsis in 0.5% -2% of cardiac surgery patients.[5-7] In these studies, sepsis was associated with mortality rates in the range of 17%-79%. As organ dysfunction is frequent following cardiac surgery due to the inflammatory response to surgery and cardiopulmonary bypass, Sepsis-3 criteria could potentially diagnose sepsis in patients with transient organ dysfunction due to surgery and coincidental minor infection.

The objective of this study was to ensure that adoption of the Sepsis-3 guidelines is appropriate for patients undergoing cardiac surgery. To achieve this objective we have assessed the incidence of sepsis as defined by the new guidelines and also investigated whether diagnosis with proven or suspected infection influences short and mid-term clinical outcomes.

Table 8-1 -The SOFA score[2]

SOFA score	1	2	3	4
Respiration PaO_2 / FiO_2 (kPa)	<53.3	<40.0	<26.7 and mechanical ventilation	<13.3 and mechanical ventilation
Coagulation Platelets $x10^3 \mu l^{-1}$	< 150	< 100	< 50	< 20
Liver Bilirubin ($\mu mol\ L^{-1}$)	20-32	33-101	102-204	> 204
Cardiovascular Hypotension	MAP <70mmHg	dopamine ≤ 5µgkg-1min-1 or dobutamine (any dose)	dopamine > 5µg kg-1min-1 or adrenaline ≤ 0.1µg kg-1min-1 or noradrenaline ≤ 0.1µg kg-1min-1	dopamine > 15µg kg-1min-1 or adrenaline > 0.1µg kg-1min-1 or noradrenaline > 0.1µg kg-1min-1
Central nervous system Glasgow Coma Score	13-14	10-12	6-9	<6
Renal Creatinine ($\mu mol\ L^{-1}$) or urine output (mlday^{-1})	106-177	178-309	310-442 < 500	> 442 < 200

PaO_2, arterial partial pressure of oxygen; FiO_2, fraction of inspired oxygen; MAP, mean arterial pressure

8.4. Methods

8.4.1. Patients and Data collection

All relevant clinical and monitoring data were collected prospectively from consecutive adult patients admitted to the Cardiac Intensive Care Unit (CICU) after cardiac surgery at University Hospital of South Manchester (UHSM) between January 2013 and May 2015. Patients undergoing cardiac transplantation were excluded from the study.

Data were collected for the duration of the patients' first CICU admission following cardiac surgery from three sources. i) Patient demographics, preoperative morbidity and outcome data were collected from the hospital's clinical governance database. ii) Physiological variables, medication data, and case note data regarding the suspicion or diagnosis of infection were collected from the electronic patient record (EPR). iii) Haematology and biochemistry results together with all microbiology reports were collected from the hospital's pathology database. Hourly recordings of physiological variables, medication administrated and all available biochemical and haematological results were cleaned using cleaning algorithms in R Studio (R Foundation for statistical computing)[8].

Daily SOFA scores (Table 8-1) were calculated for each patient using the most abnormal value recorded for each variable on each day.[2] For all patients who experienced a SOFA score increase of ≥2, the clinical notes were examined to identify suspected or proven infection. Proven infection was confirmed by microbiological cultures (excluding isolated c.albicans-positive sputum cultures, mixed growth urine samples or screening swabs which indicated colonisation). Infection was classified as suspected if antibiotics other than those given as standard prophylaxis were administered or suspicion of infection was documented in the clinical notes section of the EPR. All indicators of suspected or proven infection recorded within 24 hours of the day of the SOFA ≥2 rise were included to ensure that no suspected or proven infection was missed.

8.4.2. Missing Data

Where blood analyses necessary for calculation of daily SOFA scores were missing, the last known appropriate result recorded for that patient was substituted. Bilirubin concentrations were not routinely measured for low risk patients, so there were 340 occasions (5.2% of all SOFA score calculations) when there was no bilirubin level available for the SOFA score calculation. In 257 of these cases a bilirubin level subsequently measured for that patient was used. In the remaining 83 calculations (total of 39 patients), SOFA was calculated incorporating the median bilirubin

concentration for all patients. In the one patient with no available creatinine level and two with no available platelet count, the median for the missing variable was used. All other data were complete.

8.4.3. Statistical analysis

Normally distributed data were described using the mean and standard deviation; data with non-parametric distributions were described using the median and interquartile range (IQR). Outcome measures were length of CICU stay (hours), 30-day mortality (defined as death due to any cause within the first 30 days after cardiac surgery) and 2-year survival. The relationship between sepsis and length of CICU stay was analysed using the Kruskal-Wallis test as length of stay was not normally distributed. Univariate analyses of the relationship between sepsis and 30-day mortality were conducted using Fisher's exact test due to the low observed mortality rate. Two year survival rates were compared using the log-rank test.

Logistic EuroSCORE and cardiopulmonary bypass (CPB) time were adjusted for using linear regression analyses for the multivariable length of CICU stay analysis. Multivariable logistic regression analyses were performed to adjust for the effect of logistic EuroSCORE on 30-day mortality. The low number of deaths prevented the inclusion of additional confounders in these analyses.[9] Finally, Cox proportional hazards analyses were performed to adjust for the influence of logistic EuroSCORE and CPB time, on 2-year survival. The Logistic EuroSCORE[10] is an extensively validated preoperative risk prediction model for perioperative mortality that includes patient co-morbidities, variables reflecting cardiac function and operative risk factors. It demonstrates good discriminative ability for UK cardiac surgery.[11]

Data collection was approved by the National Research Ethics Service–Haydock as part of the Vascular Governance Northwest Project (09/H 1000 10/2+5) and all analyses were performed using R.[8]

8.5. Results

During the study period, 2230 patients were admitted to CICU after cardiac surgery. The mean (range) age was 66.1 (18-93) years and the majority of patients were men (1615, 72.4%). Full patient characteristics for the study population are shown in Table 8-2. Median length of CICU stay (IQR) was 48.8 (40.1-93.0) hours. Overall 30-day mortality was 1.5% and 2 year survival was 93.0%. SOFA rises of ≥2 were identified on 710 occasions in 323 patients. A total of 573 patients were discharged from the CICU on the first postoperative day. In these patients only one SOFA score was available preventing the calculation of a difference between the daily SOFA scores. As a result these patients were classified as not suffering sepsis during the CICU admission.

Table 8-2 - Patient characteristics

Variable	All (n=2230)	Unable to calculate SOFA rise (n=573)	No SOFA rise (n=1334)	SOFA rise > 2 but no infection (n=112)	Sepsis Suspected Infection (n=104)	Sepsis Proven Infection (n=107)
Age, Mean (range), years	66.1 (18-93)	63.2 (23-86)	66.9 (19-93)	69.6 (28-87)	65.5 (29-89)	68.6 (18-91)
Female Gender, n (%)	615 (27.6)	99(17.3)	418 (31.3)	36 (32.1)	31 (29.8)	31 (29.0)
Height, mean (SD), cm	169.6 (9.2)	171.5 (8.7)	168.9 (9.3)	169.1 (8.2)	169.5 (8.8)	168.0 (10.6)
Weight, mean (SD), Kg	81.0 (15.9)	81.7 (14.1)	81.7 (16.6)	81.6 (14.4)	81.6 (15.2)	80.1 (17.8)
Type Surgery						
Isolated CABG, n (%)	1214 (54.4)	479(83.6)	619 (46.4)	46 (41.1)	31 (29.8)	39 (36.4)
Isolated Valve, n (%)	482 (21.6)	37 (6.5)	367(27.5)	26 (23.2)	26 (25.0)	26 (24.3)
Isolated Aortic, n (%)	23 (1.0)	2 (0.3)	13 (1.0)	2 (1.8)	4 (3.8)	2 (1.9)
Combined cardiac procedures, n (%)	404 (18.1)	19 (3.3)	296 (22.2)	30 (26.7)	32 (30.8)	27 (25.2)
Other, n (%)	107(4.8)	36 (6.3)	39 (2.9)	8 (7.1)	11 (10.6)	13 (12.1)
Urgency						
Elective/Scheduled, n (%)	1324 (59.3)	317 (55.3)	823 (61.7)	61 (54.5)	62 (59.6)	61 (57.0)
Urgent, n (%)	842 (37.8)	244 (42.5)	488 (36.5)	45 (40.1)	29 (27.9)	36 (33.6)
Emergency/Salvage, n (%)	64 (2.9)	12 (2.1)	23 (1.7)	6 (5.4)	13 (12.5)	10 (9.3)
Duration of CPB, median (interquartile range), minutes	100.0 (79.0-128.0)	87.0 (69-105.0)	104.0 (84.0-134.0)	101.0 (84.0-135.5)	118.0 (94.0-165.0)	108.0 (84.0-147.0)
Logistic EuroSCORE, median (Interquartile range)	4.0 (2.1-7.7)	2.3 (1.4-3.5)	4.6 (2.4-8.4)	5.1 (3.3-11.8)	7.5 (3.1-19.7)	7.8 (3.6-15.1)
Requirement for Renal Replacement Therapy, n (%)	107 (4.8)	7 (1.2)	24 (1.8)	10 (8.9)	26 (25.0)	40 (37.4)
Mechanical Ventilation > 72 hours n (%)	147 (6.6)	-	37 (2.8)	7 (6.3)	29 (27.9)	50 (46.7)

CABG, coronary artery bypass graft; CPB, cardiopulmonary bypass

8.5.1. Sepsis

The Sepsis-3 criteria for sepsis were met by 211 (9.5%) of the 2230 patients. Sepsis with suspected infection occurred in 104 patients (4.7%) and sepsis with proven infection was demonstrated in 107 (4.8%). The respiratory tract was the most frequent source of both proven (72.1%) and suspected infection (55.4%). Other sources of infection are shown in Table 8-3.

Table 8-3 - Suspected or proven sources of infection in those diagnosed with sepsis

Suspected source	Suspected infection*	Proven Infection*
Not specified (antibiotics started)	51	NA
Unknown	10	NA
Respiratory	148	96
Abdominal/Gastrointestinal	6	-
Wound	6	8
Genitourinary	3	5
Bacteraemia/catheters	6	24
Endocarditis/myocarditis	9	-
Dental	1	-

* Multiple sites were implicated in many patients

The median length of CICU stay (IQR) was 145.2 (114.5-261.7) hours for those with sepsis due to suspected infection, 211.5 (117.2-478.1) hours for those with sepsis due to proven infection, and 47.0 (28.8-72.6) hours for those without sepsis (p <0.01 for both). After controlling for the logistic EuroSCORE and CPB time using linear regression modelling, patients with sepsis had significantly longer CICU stays than those without. The increase in length of CICU stay (95% CI) was 134.1 (99.0-168.2) hours for those with suspected infection and 266.1 (231.6-300.7) hours for those with proven infection (p<0.01 for both). The linear regression model is detailed in the Appendix.

To ensure the length of stay analysis was not skewed by the 573 patients discharged on the first post-operative day, a sensitivity analysis using the same linear regression model on data taken exclusively from patients with two or more daily SOFA scores (n=1657) was performed. The

increase in length CICU of stay (95%CI) attributed to a diagnosis of sepsis with suspected infection in this subgroup remained significant at 135.7 (99.1-172.3) hours; the increase related to proven infection was 265.8 (229.7-301.9) hours (p<0.01 for both).

The 30-day mortality was 6.6% for those who suffered sepsis compared with 1.0% for those who did not (p<0.01). The mortality rates for sepsis with suspected infection (5.8%) and sepsis with proven infection (7.5%) were both significantly higher than the rate of 1.0% for those without sepsis (p<0.01 for both). After adjusting for pre-and intraoperative confounders using the logistic EuroSCORE (full model detailed in the Appendix), the odds ratio associated with sepsis was 3.7 (95%CI 1.1-10.2, p=0.02) for suspected infection and 6.6 (95% CI 2.6-15.7, p <0.01) for proven infection.

Among those who suffered from sepsis, the main differences between survivors and those who died were that those who died had a higher median logistic EuroSCORE (16.7 vs 6.7, p<0.01) and were less likely to have undergone isolated CABG or valve surgery (21.4% vs 60.4%, p=0.01). Rates of renal replacement therapy (71.4% vs 28.4%, p<0.01) and prolonged mechanical ventilation (92.9% vs 33.5%, p<0.01) were higher in non-survivors than survivors.

Among patients with sepsis the 2-year survival was 87.5% for those with suspected infection and 73.8% for those with proven infection compared with 94.3% for those without sepsis (p<0.01 for both). As seen in Figure 8-1 the greatest difference in mortality rates was seen in the first 12 postoperative months. A second log-rank analysis which included only those patients alive one year post surgery showed a smaller difference in the rates of survival to two years between those who had suffered sepsis (98.1%) and those who had not (96.0%) which was no longer statistically significant(p=0.06).
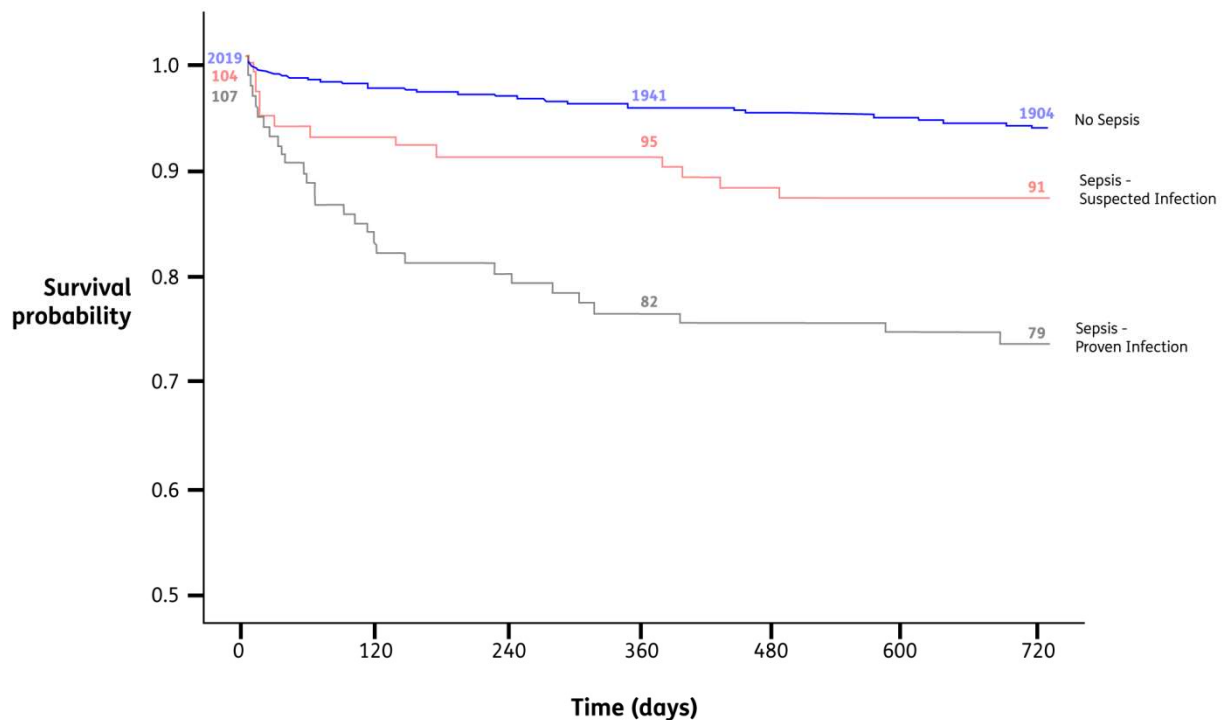
Figure 8-1 - Two-year survival according to sepsis status

The confounding effects of preoperative logistic EuroSCORE and CPB time on 2-year survival were adjusted for using Cox proportional hazards regression. For those with sepsis due to suspected infection the hazard ratio was 1.1 and the effect on survival was not statistically significant (95%CI 0.5-2.4, p=0.76). However, for sepsis due to proven infection compared with those without sepsis the hazard ratio was 3.6 (95%CI 2.2-5.9 p<0.01). The model is detailed in the Appendix.

8.5.2. SOFA rise ≥2 in the absence of sepsis

112 patients developed a SOFA rise ≥2 in the absence of proven or suspected infection. The median length of CICU stay (IQR) for these patients was 83.2 (48.5-124.9) hours. This was significantly shorter than the median CICU stay of 211.5 hours for both those with sepsis due to proven infection and 145.2 hours for those with suspected infection (p<0.01 for both). On multivariable analysis (full model detailed in the Appendix), a SOFA rise ≥2 without sepsis was associated with a statistically insignificant difference in length of stay of 6.9 hours (95%CI -28.2-41.9, p=0.70).

The 30-day mortality rate for those with a SOFA rise ≥2 in the absence of sepsis was 2.7%. This was higher than that for patients with lesser increases in SOFA scores and lower than that for

those with sepsis and suspected or proven infection but none of these differences was statistically significant. (p=0.10, p=0.32 and p=0.13 respectively). After adjusting for pre-and intraoperative confounders using the logistic EuroSCORE (full model detailed in the Appendix), a SOFA rise≥2 in the absence of sepsis was not significantly associated with 30-day mortality (odds ratio 2.1 (95% CI 0.5-6.2, p =0.23).   The 2-year survival rate for patients who suffered a SOFA rise ≥2 without sepsis was 91.1%. This was not significantly different to the rate of 94.3% for patients with stable or small rises (<2) in the SOFA score (p=0.13) neither was it significantly different from the 87.5% in those with suspected infection (p=0.38). It was however, significantly higher than the 73.8% for those with proven infection and sepsis (p<0.01).

8.5.3. Septic Shock

Of the 211 patients diagnosed with sepsis, 159 patients (75.4%) met criteria relating to serum lactate concentration and use of vasopressors compatible with a diagnosis of septic shock. For this subgroup, median (IQR) length of CICU stay was 193.2 (139.5-364.0) hours, thirty day mortality was 8.8%, and the 2-year survival rate was 76.7%. All of these results were significantly worse than for patients with sepsis who did not suffer septic shock (p<0.02 for all).

Table 8-4 - Patient outcomes

Sepsis Status	Subgroup	n (%)	30-day mortality, n (%)	Median ICU stay, hours (interquartile range)	2 year survival, n (%)
No Sepsis	**All patients**	**2019 (90.5)**	**20 (1.0)**	**47.0 (28.8-72.6)**	**1904 (94.3)**
	SOFA rise not calculable	573 (25.7)	4 (0.7)	22.6 (20.0-25.3)	553 (96.5)
	SOFA rise < 2	1334 (59.8)	13 (1.0)	52.6 (45.4.-86.2)	1248 (93.6)
	SOFA rise ≥ 2 without infection	112 (5.0)	3 (2.7)	83.2 (48.5 -124.9)	102 (91.1)
Sepsis	**All patients**	**211 (9.5)**	**14 (6.6)***	**176.0 (115.7-404.6)***	**170 (80.5)***
	Suspected infection	104 (4.7)	6 (5.8)*	145.2 (114.5-261.7)*	91 (87.5)*
	Proven infection	107 (4.8)	8 (7.5)*	211.5 (117.2-478.1)*	79 (73.8)*

* indicates p value of <0.05 when compared with the frequency of the outcome in the No Sepsis group during univariate analyses

8.6. Discussion

This is the first study to validate the Sepsis-3 guidelines in a cohort of patients after cardiac surgery. The new guidelines allow patients with only suspected infection to be diagnosed with sepsis. This may be one reason why the incidence of sepsis following cardiac surgery in this study (9.5%) is higher than that reported in studies which only included those with infection proven by microbiological culture.[5-7 12] The frequency of sepsis with positive cultures (4.8%) was also higher than most previous studies of sepsis after cardiac surgery. However, the majority of the previous studies only including specific sources of infection such as the wound or respiratory tract.[5 6 12 13] Despite the higher incidence of sepsis observed in our cohort, the 30-day mortality of 6.6% for patients with sepsis was lower than that found in previous studies in cardiac surgery.[5 7 14 15]

Sepsis based on the Sepsis-3 guidelines was a significant risk factor for adverse outcomes in our cohort. 30-day mortality risk increased 6-fold in patients who met the Sepsis-3 criteria. Patients who suffered sepsis also had significantly longer CICU stays compared with patients who did not. Overall 2-year survival rates were lower for patients with sepsis, although our secondary analysis including only those who survived to 1 year illustrates that most of the impact of sepsis on mortality risk appears to be observed in the first 12 months. This relatively short term effect on risk is different to that reported in patients from general ICUs which detected impact on survival in the longer term.[16] This difference may be due to the cohort of patients included in our study. In the cardiac surgery patients studied, the organ dysfunction that triggered the diagnosis of sepsis often progressed to organ failure in patients who had already been physiologically stressed by their surgery. 31% of those with sepsis required renal replacement therapy and 37% required prolonged mechanical ventilation. Patients either recovered from these critical complications or died as a result of them within a relatively short period.

Studies in general ICU populations found that when Sepsis-3 criteria and SIRS-related sepsis criteria were applied to the same patients with suspected infection, the Sepsis-3 criteria identified fewer patients than SIRS-related criteria. They also showed that patients identified by the Sepsis-3 criteria were likely to suffer worse outcomes.[17 18] However, the variables included in the SIRS criteria are influenced by the inflammatory response to major surgery as well as treatments such as mechanical ventilation, patient warming and perioperative beta-blockade which are frequently employed following cardiac surgery. Importantly, unlike the SOFA-related definitions, the SIRS criteria cannot recognise the effects of interventions on the absolute values of these parameters. Moreover, in our cohort, 88% of patients fulfilled the criteria for SIRS postoperatively. Therefore, using SIRS-related criteria to diagnose sepsis would have led to the vast majority of suspected infections resulting in a diagnosis of sepsis, even where inflammation did not progress above the

postoperative baseline.   Consequently, the Sepsis-3 definitions seem to provide the most appropriate means for detecting sepsis after cardiac surgery. This may be also true for patient groups who require critical care treatments after undergoing other types of major surgery or suffering from conditions which result in non-infective, inflammatory responses such as pancreatitis or severe burns.

A significant proportion of the patients diagnosed with sepsis in this study went on to meet the criteria for septic shock. The length of CICU stay, 30-day mortality and 2-year survival rate associated with septic shock were significantly worse than those for patients who suffered sepsis without septic shock. Further work to identify patients at the highest risk of developing septic shock would therefore be of clinical importance.

Approximately half of the patients diagnosed with sepsis in our study had proven infection and half had suspected infection; a proportion similar to that documented in the general ICU population.[19] The new Sepsis-3 guidelines include suspicion of infection to ensure all patients with sepsis were identified and treated early.[1] This analysis demonstrates that patients with suspected infection suffered outcomes more similar to those with proven infection than to those with no sepsis. While outcomes for patients with proven infection were worse than for those with suspected infection, it is clearly appropriate to adopt Sepsis-3 as this improves the early identification of patients at a high risk of adverse short term outcomes.

Limitations

The new Sepsis-3 criteria identify worsening organ function by the change in the daily SOFA score rather than absolute values. The guidelines state that the baseline SOFA should be assumed to be 0 unless a patient has "pre-existing organ dysfunction (acute or chronic)". Following cardiac surgery, the mean day one SOFA score was 5.4 and over 90% of patients had a day one SOFA score >2. Consequently, assuming a baseline SOFA score of 0 for patients undergoing this major surgery would be inappropriate. As a result, we required two daily postoperative scores in order to calculate the change in SOFA score and were therefore unable to diagnose sepsis before the second postoperative day. 26% of our patients were discharged on the first post-operative day and therefore could not be classified as having sepsis. The median CICU stay in these early discharge patients was 22.6 hours and their 30-day mortality was only 0.7%. Although sepsis would be unusual on the first day following cardiac surgery we have performed a sensitivity analysis excluding these patients and the conclusions from the analysis are unchanged.

The small 30-day mortality rate in the study (34 deaths) prevented the inclusion of additional confounders into the logistic regression analyses for 30-day mortality.[9] Logistic EuroSCORE was

chosen to be the sole confounder entered into the model as it was considered the most clinically relevant variable.

As with any observational study there were missing data. The incidence of missing data was however very low and SOFA scores were calculated using imputed bilirubin values for less than two percent of patients. In three of these patients creatinine concentration or platelet count were also imputed. A further potential limitation of this study is that it was conducted at a single centre. Although our centre performs almost all aspects of adult cardiac surgery, a larger, multicentre study would be the optimal method to validate these findings further.

Case notes were only examined for patients who suffered a SOFA rise ≥ 2. Therefore, this study was not able to characterise outcomes for those who suffered proven or suspected infection in the absence of significant organ dysfunction.

8.7. Conclusion

This is the first study exploring how Sepsis-3 criteria influence the diagnosis of sepsis in cardiac surgery patients. Patients with sepsis due to both proven and suspected infection suffered prolonged CICU stays and increased 30-day mortality justifying the adoption of Sepsis-3 guidelines in cardiac surgery.

Declaration of interests

No author has a conflict of interest.

Funding

8.8. Appendix

Tables describing multivariable adjustment models

Table 8-5 - Linear regression model for length of CICU stay accounting for effects of confounders

Variable	Beta-coefficient (hours)	95% Confidence Interval		p value
Intercept	2.5	-14.0	19.1	0.76
Sepsis - Proven Infection	266.1	231.6	300.7	<0.01
Sepsis -Suspected Infection	134.1	99.0	168.2	<0.01
Logistic EuroSCORE (per point)	2.4	1.5	3.3	<0.01
Cardiopulmonary Bypass Time (per minute)	0.5	0.4	0.6	<0.01

Table 8-6 - Linear regression model for length of CICU stay accounting for effects of confounders in those who stayed long enough for 2 or more SOFA scores to be calculated

Variable	Beta-coefficient (hours)	95% Confidence Interval		p value
Intercept	28.3	7.9	48.6	<0.01
Sepsis - Proven Infection	265.8	229.7	301.9	<0.01
Sepsis - Suspected Infection	135.7	99.1	172.3	<0.01
Logistic EuroSCORE (per point)	2.1	1.1	3.1	<0.01
Cardiopulmonary Bypass Time (per minute)	0.3	0.2	0.5	<0.01

Table 8-7 - Logistic Regression model for 30-day mortality

Variable	Beta coefficient	Odds Ratio	95% Confidence Interval		p value
Intercept	-5.2				<0.01
Sepsis - Proven Infection	1.9	6.6	2.6	15.7	<0.01
Sepsis - Suspected Infection	1.3	3.7	1.1	10.2	0.02
Logistic EuroSCORE (per point)	0.0	1.1	1.0	1.1	<0.01

Table 8-8 - Cox Proportional Hazards Ratio Model for 2 year non-survival

Variable	Hazard Ratio	95% Confidence Interval		p value
Sepsis – Proven Infection	3.6	2.2	5.9	<0.01
Sepsis – Suspected Infection	1.1	0.5	2.4	0.76
Logistic EuroSCORE (per point)	1.0	1.0	1.0	0.02
Cardiopulmonary Bypass Time (per minute)	1.0	1.0	1.0	<0.01

Table 8-9 - Linear regression model for length of CICU stay investigating significance of a SOFA rise ≥2 in the absence of sepsis

Variable	Beta-coefficient (hours)	95% Confidence Interval		p value
Intercept	6.3	-11.4	24.1	0.48
SOFA rise ≥2 without sepsis	6.9	-28.2	41.9	0.70
Logistic EuroSCORE (per point)	3.6	2.7	4.6	<0.01
Cardiopulmonary Bypass Time (per minute)	0.6	0.4	0.7	<0.01

Table 8-10 - Logistic Regression Model for 30-day mortality investigating significance of a SOFA rise ≥2 in the absence of sepsis

Variable	Beta-coefficient	Odds ratio	95% Confidence Interval		p value
Intercept	-4.9				<0.01
SOFA rise ≥2 without sepsis	0.7	2.1	0.5	6.2	0.23
Logistic EuroSCORE (per point)	0.1	1.1	1.0	1.1	<0.01

8.9. References

1. Singer M, Deutschman CS, Seymour C, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016; **315**: 801-10

2. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; **22**: 707-10

3. Bone RC, Balk RA, Cerra FB, et al. DEfinitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. the accp/sccm consensus conference committee. american college of chest physicians/society of critical care medicine. *Chest* 1992; **101**: 1644-55

4. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive care medicine* 2003; **29**: 530-8

5. Michalopoulos A, Stavridis G, Geroulanos S. Severe sepsis in cardiac surgical patients. *European Journal of Surgery* 1998; **164**: 217-22

6. Fowler VG, O'Brien SM, Muhlbaier LH, Corey GR, Ferguson TB, Peterson ED. Clinical Predictors of Major Infections After Cardiac Surgery. *Circulation* 2005; **112**: I-358-I-65

7. Oliveira DC, Oliveira Filho JB, Silva RF, et al. Sepsis in the postoperative period of cardiac surgery: problem description. *Arq Bras Cardiol* 2010; **94**: 332-6, 52-6

8. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing 2015

9. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**: 1373-9

10. Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J* 2003; **24**: 881-2

11. Hickey GL, Grant SW, Murphy GJ, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg* 2013; **43**: 1146-52

12. Chen LF, Arduino JM, Sheng S, et al. Epidemiology and outcome of major postoperative infections following cardiac surgery: risk factors and impact of pathogen type. *American journal of infection control* 2012; **40**: 963-8

13. Sahu M, Siddharth B, Choudhury A, et al. Incidence, microbiological profile of nosocomial infections, and their antibiotic resistance patterns in a high volume Cardiac Surgical Intensive Care Unit. *Annals of Cardiac Anaesthesia* 2016; **19**: 281-7

14. Michalopoulos A, Geroulanos S, Rosmarakis ES, Falagas ME. Frequency, characteristics, and predictors of microbiologically documented nosocomial infections after cardiac surgery. *Eur J Cardiothorac Surg* 2006; **29**: 456-60

15. Falagas ME, Rosmarakis ES, Rellos K, Michalopoulos A, Samonis G, Prapas SN. Microbiologically documented nosocomial infections after coronary artery bypass surgery without cardiopulmonary bypass. *Journal of Thoracic and Cardiovascular Surgery* 2006; **132**: 481-90

16. Winters BD, Eberlein M, Leung J, Needham DM, Pronovost PJ, Sevransky JE. Long-term mortality and quality of life in sepsis: a systematic review. *Crit Care Med* 2010; **38**: 1276-83

17. Donnelly JP, Safford MM, Shapiro NI, Baddley JW, Wang HE. Application of the Third International Consensus Definitions for Sepsis (Sepsis-3) Classification: a retrospective population-based cohort study. *The Lancet Infectious diseases* 2017; **17**: 661-70

18. Sterling SA, Puskarich MA, Glass AF, Guirgis F, Jones AE. The Impact of the Sepsis-3 Septic Shock Definition on Previously Defined Septic Shock Patients. *Crit Care Med* 2017; **45**(9):1436-1442

19. Vincent JL, Sakr Y, Sprung CL, et al. Sepsis in European intensive care units: results of the SOAP study. *Crit Care Med* 2006; **34**: 344-53

Chapter Nine: The KDIGO acute kidney injury guidelines for cardiac surgery patients in critical care: a validation study (Published journal article)

Status: Published

Submission Date: 17/01/2018

Published online: 25/06/2018

Reference: Howitt SH, Grant SW, Caiado C, et al. The KDIGO acute kidney injury guidelines for cardiac surgery patients in critical care: a validation study. BMC nephrology 2018; 19: 149

Conference Presentation: Audio-visual presentation to the Society for Cardiothoracic Surgery, Glasgow, UK, March 2018

Authors' contributions

SHH designed the project, acquired and cleaned the data, performed the analyses and wrote the first draft of the manuscript. SWG designed the project and revised the first draft. CC supervised the statistical design of the project and the performance of the analyses. EC and DK both verified and cleaned the data collected from the clinical databases. IK, IM and CM all designed the project and revised the drafts. All authors read and approved the final manuscript.

9.1. Additional data processing for this manuscript.

This manuscript compares outcomes of patients who met different criteria for the same stage of AKI while on CICU following cardiac surgery. Before such analyses could be performed the cleaning code needed to be modified from that used to create the original output files. In the original cleaning code only the new onset of each stage of AKI was labelled. For this study, if the AKI stage was diagnosed by urine output but the diagnosis of AKI had been made using data which contained blank values checks were made to ensure the diagnosis was correct (Steps 132 and 133 of the cleaning code). For the manuscript which forms the basis of this chapter, the first time each specific criterion for each level of AKI was also analysed. Therefore the EPR was reviewed manually for all incidences where a urine output criterion for any stage of AKI were met using blank urine output entries regardless of whether or not that stage of AKI had already been diagnosed by creatinine criteria.

Additional cleaning steps

> Step 1. Load in all the data from the original dataset.
>
> Step 2. Label the first time each criterion was met for each level of AKI for each patient.
>
> Step 3. Identify incidences where urine output criteria are met for the first time where a blank hourly value was present in the hours used to make the diagnosis.
>
> Step 4. Manually lookup all incidences where this occurred and if it is due to loss of urine in to the toilet/bed etc. remove the label and check it should not be applied later in that patient's admission.
>
> Step 5. Label those who suffer AKI 3 due to anuria when not catheterised and go on to void large volumes as "Anuric no catheter."
>
> Step 6. Import Dendrite and PatientIndex data frames from original dataset.
>
> Step 7. Match each outcome to each patient.
>
> Step 8. Remove those with preoperative RRT and those with no preoperative creatinine concentration.
>
> Step 9. Create the summary characteristics table. (Table 9-2)
>
> Step 10. Disregard all incidences where urine output or creatinine criteria were only met after RRT had been established.
>
> Step 11. Conduct univariable analyses of progression to RRT, LOS and 30-day mortality by criteria met for AKI-1.
>
> Step 12. Conduct univariable analyses of progression to RRT, LOS and 30-day mortality by criteria met for AKI-2.

Step 13. Conduct multivariable analyses of progression to RRT, LOS and 30-day mortality by criteria met for AKI-1 controlled for logistic EuroSCORE.

Step 14. Conduct multivariable analyses of progression to RRT, LOS and 30-day mortality by criteria met for AKI-2 controlled for logistic EuroSCORE.

Step 15. Conduct univariable analyses comparing LOS and 30-day mortality between those with no AKI and AKI-1 by UO alone.

Step 16. Conduct multivariable analyses comparing LOS and 30-day mortality between those with no AKI and AKI-1 by UO alone for logistic EuroSCORE.

Step 17. Add in 2 – year survival data from hospital clinical governance database.

Step 18. Conduct univariable analyses comparing 2-year survival between those with no AKI and AKI-1 by UO alone

Step 19. Conduct univariable analyses 2-year survival by criteria met for AKI-1.

Step 20. Conduct univariable analyses 2-year survival by criteria met for AKI-2.

Step 21. Conduct multivariable analyses comparing 2-year survival between those with no AKI and AKI-1 by UO alone for logistic EuroSCORE.

Step 22. Conduct multivariable analyses of 2-year survival by criteria met for AKI-1, controlled for logistic EuroSCORE.

Step 23. Conduct multivariable analyses of 2-year survival by criteria met for AKI-2, controlled for logistic EuroSCORE.

Step 24. Construct a variable showing maximum AKI level reached and reason for reaching that level of AKI for each patient.

Step 25. For comparative analyses between groups, plot 2 year survival Kaplan Meier curve according to the maximum level of AKI reached and criteria met for that level.

9.2. Abstract

Background: The Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury (AKI) guidelines assign the same stage of AKI to patients whether they fulfil urine output criteria, serum creatinine criteria or both criteria for that stage. This study explores the validity of the KDIGO guidelines as a tool to stratify the risk of adverse outcomes in cardiac surgery patients.

Methods: Prospective data from consecutive adult patients admitted to the cardiac intensive care unit (CICU) following cardiac surgery between January 2013 and May 2015 were analysed. Patients were assigned to groups based on the criteria they met for each stage of AKI according to the KDIGO guidelines. Short and mid-term outcomes were compared between these groups.

Results: A total of 2267 patients were included with 772 meeting criteria for AKI-1 and 222 meeting criteria for AKI-2. After multivariable adjustment, patients meeting both urine output and creatinine criteria for AKI-1 were more likely to experience prolonged CICU stay (OR 4.9, 95%CI 3.3-7.4, p<0.01) and more likely to require renal replacement therapy (OR 10.5, 95%CI 5.5-21.9, p<0.01) than those meeting only the AKI-1 urine output criterion. Patients meeting both urine output and creatinine criteria for AKI-1 were at an increased risk of mid-term mortality compared to those diagnosed with AKI-1 by urine output alone (HR 2.8, 95%CI 1.6-4.8, p<0.01). Patients meeting both urine output and creatinine criteria for AKI-2 were more likely to experience prolonged CICU stay (OR 16.0, 95%CI 3.2-292.0, p<0.01) or require RRT (OR 11.0, 95%CI 4.2-30.9, p<0.01) than those meeting only the urine output criterion. Patients meeting both urine output and creatinine criteria for AKI-2 were at a significantly increased risk of mid-term mortality compared to those diagnosed with AKI-2 by urine output alone (HR 3.6, 95%CI 1.4-9.3, p<0.01).

Conclusions: Patients diagnosed with the same stage of AKI by different KDIGO criteria following cardiac surgery have significantly different short and mid-term outcomes. The KDIGO criteria need to be revisited before they can be used to stratify reliably the severity of AKI in cardiac surgery patients. The utility of the criteria also needs to be explored in other settings.

348/350 words

Keywords: Acute kidney injury; Cardiac surgery, Critical care

9.3. Background

Acute Kidney Injury (AKI) occurs in up to 50% of patients following cardiac surgery.[1, 2] Even in its mildest form, AKI is associated with increased mortality and prolonged Critical Care Unit stay.[3-6]   AKI requiring renal replacement therapy (RRT) occurs in 2-5% of patients following cardiac surgery and is associated with mortality of up to 60%.[1, 7, 8] The Kidney Disease: Improving Global Outcomes (KDIGO) AKI guidelines were designed to standardise the criteria for AKI based on serum creatinine and urine output (Table 9-1).[9] Patients are assigned the same stage of AKI regardless of which criteria (urine output, serum creatinine or both) for that stage are met. However, concerns have been raised that the guidelines' urine output criteria are poorly calibrated.[10]

Table 9-1 - KDIGO criteria for diagnosis of AKI in adults[9]

Stage of AKI	Serum Creatinine	Urine output
1	1.5–1.9 times baseline OR ≥0.3 mg/dl (≥26.5 µmol/l) increase within 48 hours	<0.5 ml/kg/h for 6–12 hours
2	2.0–2.9 times baseline	<0.5 ml/kg/h for ≥12 hours
3	≥3.0 times baseline OR Increase in serum creatinine to ≥4.0 mg/dl (≥353.6 µmol/l) OR Initiation of renal replacement therapy	<0.3 ml/kg/h for ≥24 hours OR Anuria for ≥12 hours

Each stage of AKI is diagnosed when any of the criteria for that stage of AKI are met.

Studies validating the KDIGO guidelines following cardiac surgery have frequently stratified patient risk based on serum creatinine alone as reliable urine output data are difficult to collect.[1, 11-13] Studies that have had access to urine output data have tended to be relatively small and have disagreed on the importance of urine output when identifying those at risk of adverse outcomes.[2, 14, 15] A recent study with access to urine output data after cardiac surgery did demonstrate that patients with AKI diagnosed on oliguria alone had increased long-term mortality but due to the relatively small sample size they were unable to assess the importance of urine output within each AKI level.[6]

The objective of this study was to validate the KDIGO guidelines for AKI by assessing the outcomes of patients meeting different criteria for each stage of AKI after cardiac surgery.

9.4. Methods

9.4.1. Data

Data from 2,284 consecutive patients admitted to the cardiac intensive care unit (CICU) following cardiac surgery at Wythenshawe Hospital (part of Manchester University NHS Foundation trust) were collected prospectively between January 2013 and May 2015. Wythenshawe Hospital is a tertiary centre for adult cardiac surgery, cardiothoracic transplantation and mechanical circulatory support as a bridge to cardiac transplantation or recovery. Patients requiring RRT preoperatively and those with no preoperative creatinine values were excluded as shown in Figure 9-1. Patients who received mechanical circulatory support were excluded from length of stay (LOS) analyses as their CICU stay was prolonged while awaiting definitive treatment. All data were collected as part of the Vascular Governance North West (VGNW) database and processed according this project's protocols and ethical approvals.

Serum creatinine concentration was usually measured daily and all available results were extracted from the hospital's pathology laboratory database. Our institution's laboratory measures creatinine using techniques based on Jaffe chemistry with a total imprecision of <6%. Every creatinine value for each patient was analysed and both relative and absolute increases in creatinine were used to classify AKI stages according to the KDIGO criteria (Table 9-1). The relative increases were calculated using the most recently recorded preoperative level as the baseline value. Urine output was recorded hourly on the CICU electronic patient record. Where the hourly value was recorded as none or zero, this value was accepted whereas when no value was entered for a given hour the next volume of urine recorded was divided equally by the number of blank hours prior to this recording. Whenever urine output fell below the thresholds in

the KDIGO criteria, the time and appropriate stage of AKI was recorded. The need for RRT and postoperative LOS on CICU were identified from the electronic patient record. Serum creatinine concentration and urine output measurements recorded after initiation of RRT were not included in analyses as both are influenced heavily by RRT itself.

The hospital clinical governance database recorded 2-year all-cause mortality and the preoperative comorbidity, urgency and complexity of surgery as measured by the logistic EuroSCORE.[16] Prolonged LOS was defined as a CICU stay longer than 120 hours for cardiac transplant patients or >72 hours for all other patients.

9.4.2. Statistical Analyses

Patients were assigned to groups based on the stages of AKI they reached according to the KDIGO guidelines. Within the groups that reached each AKI stage, patients were categorised as either i) meeting the urine output criteria ii) meeting the serum creatinine criteria or iii) meeting both urine output and serum creatinine criteria. Rates of prolonged LOS, RRT and 2-year mortality for those who did not develop AKI were compared with those for patients diagnosed with AKI-1 by urine output alone. Analyses within groups of patients meeting different combinations of criteria for each stage of AKI were then performed. The null hypothesis was that outcomes would be similar between patients diagnosed with the same stage of AKI based on the different KDIGO criteria.

Univariable analyses of categorical outcomes were performed using the chi-square test or Fisher's exact test in the event of sparse data. The logistic EuroSCORE [16] which calculates mortality risk for cardiac surgery based on 13 preoperative variables (including preoperative renal function) and four operative variables was used to adjust for surgical risk in a multivariable logistic regression model. The logistic EuroSCORE has been shown to have adequate discriminatory ability in UK cardiac surgery.[17] The results of the multivariable analyses are detailed in the appendix.

Univariable and multivariable analyses of long term mortality rates were performed using the log-rank test and Cox proportional hazards regression modelling respectively. Data cleaning and statistical analyses were conducted using R (R Foundation for statistical computing).[18]

9.5. Results

Data from 2284 patients were available. Seven patients who required RRT preoperatively and ten patients with no preoperative creatinine values were excluded leaving 2,267 patients for the analysis (Figure 9-1). Patient characteristics are shown in Table 9-2. There were 1448 patients who did not develop AKI during their CICU stay. A total of 819 (36.1%) developed AKI and 147 (6.5%) required RRT. There were 177 (7.8%) patients who died within two years of surgery. Of the 1448 patients who did not develop AKI, 255 (15.5%) had a prolonged LOS on CICU and the 2-year mortality rate for this group was 3.9%.
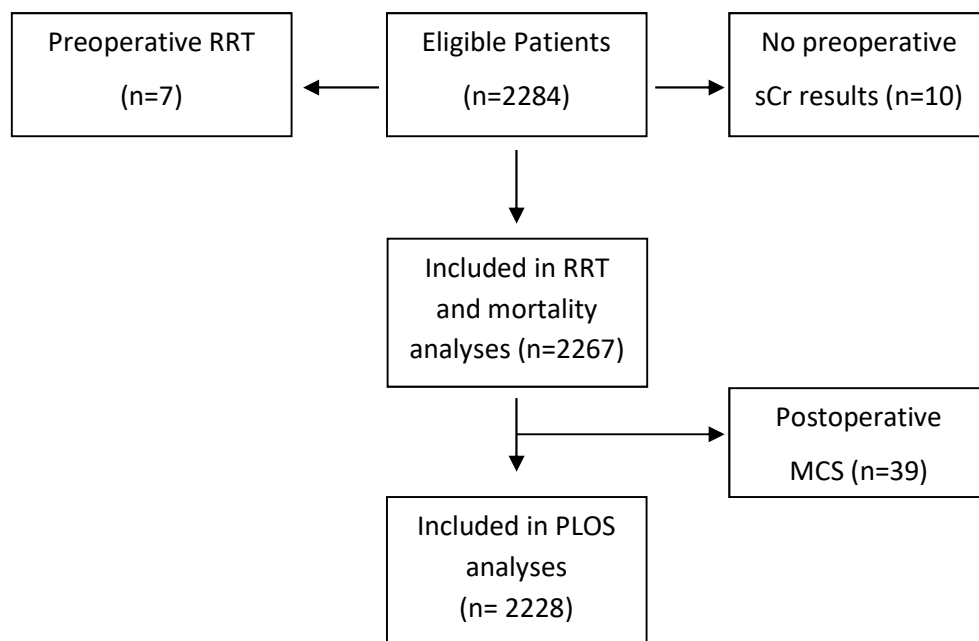
Figure 9-1 - Flow chart for inclusion of patients in analyses.

RRT = renal replacement therapy, sCR = serum creatinine result, MCS = mechanical circulatory support, PLOS = prolonged length of stay

Table 9-2 - Characteristics of patients admitted to the cardiac intensive care unit following cardiac surgery

Characteristic	All (n=2267)	AKI-1UO (n=370)	AKI-1 sCr (n=192)	AKI-1 Both (n=210)	AKI-2 UO (n=97)	AKI-2 sCr (n=92)	AKI-2 Both (n=33)
Age, mean (sd) , years	65.9 (11.6)	67.3 (10.5)	67.6 (12.7)	68.5 (11.8)	67.7 (10.9)	66.7(12.3)	64.4 (12.3)
Female gender, %	27.2	30.8	28.1	26.7	34.0	27.2	24.2
Weight, mean (sd), Kg	81.5 (15.8)	89.4 (16.6)	75.3 (14.0)	87.2 (17.7)	96.5 (17.3)	80.1 (16.3)	89.1 (18.0)
Logistic EuroSCORE, median (Interquartile range)	4.0 (2.1-7.7)	4.4 (2.5 -7.7)	7.7 (3.5-18.1)	6.3 (2.8-13.0)	4.3 (2.5-7.6)	9.7 (4.2-18.4)	5.6 (2.3-15.6)
Operation, n (%)							
CABG	1211 (53.4)	177 (47.8)	56 (29.2)	92 (43.8)	49 (50.5)	28 (30.4)	10 (30.3)
Valve	477 (21.0)	93 (25.1)	50 (26.0)	39 (18.6)	26 (26.8)	23 (25.0)	8 (24.2)
CABG and Valve	301 (13.3)	60 (16.2)	41 (21.4)	39 (18.6)	13 (13.4)	12 (13.0)	5 (15.2)
Aortic	122 (5.4)	19 (5.1)	16 (8.3)	16 (7.6)	4 (4.3)	11 (12.0)	3 (9.1)
Cardiac Transplantation	53 (2.3)	5 (1.4)	19 (9.9)	10 (4.8)	1 (1.0)	8 (8.7)	4 (12.1)
MCS	39 (1.7)	6 (1.6)	8 (4.2)	6 (2.9)	2 (2.1)	4 (4.3)	2 (6.1)
Other – minor	20 (0.9)	3 (0.8)	1 (0.5)	5 (2.4)	0 (0.0)	3 (3.3)	1 (3.0)
Other – major	44 (1.9)	7 (1.9)	1 (0.5)	3 (1.4)	2 (2.1)	3 (3.3)	0 (0)
Urgency, n (%)							
Elective	1321(58.3)	217 (58.6)	99 (51.6)	110 (52.4)	56 (57.7)	42 (45.7)	14 (42.4)
Urgent	890 (39.3)	145 (39.2)	83 (43.2)	89 (42.4)	38 (39.2)	43 (46.7)	17 (51.5)
Emergency	40 (1.8)	7 (1.9)	8 (4.2)	7 (3.3)	3 (3.1)	5 (5.4)	1 (3.1)
Salvage	16 (0.7)	1 (0.3)	2 1.0)	4 (1.9)	0 (0.0)	2 (2.2)	1 (3.1)
CPB time, median (Interquartile range), minutes	101.0 (80.0-130.0)	100.0 (81.0-133.2)	129.0 (97.0-182.0)	107.5 (83.3-132.8)	99.0 (75.5-120.0)	112.0 (90.0-155.0)	123.0 (81.0-157.0)

UO - urine output; sCr - serum creatinine; CABG - coronary artery bypass graft; MCS - mechanical circulatory support, CPB - cardiopulmonary bypass

Acute Kidney Injury Stage 1 (urine output only) vs no AKI

AKI-1 was diagnosed in 772 (34.1%) patients (Table 9-3) with 370 (47.9%) of these patients meeting only the urine output criterion (AKI-1-UO). As AKI-1-UO patients had the best outcomes these patients were compared with the no AKI group. On univariable analysis, the rate of prolonged LOS for AKI-1-UO (39.6%) was significantly higher than for patients without AKI (p<0.01). There were 22 (5.9%) AKI-1-UO patients who died within 2-years although this was not statistically significantly higher than the 2-year mortality rate in the no AKI group (p=0.10). On multivariable analysis adjusted for the logistic EuroSCORE the risk of prolonged LOS for AKI-1-UO was higher (OR 2.8, 95%CI 2.2-3.6, p<0.01) but the mortality risk within the first two years was not significantly higher (HR 1.4, 95%CI 0.9-2.3, p=0.18) than for those without AKI.

Acute Kidney *Injury Stage 1*

Of the other patients diagnosed with AKI-1, 192 (24.9%) met only the serum creatinine concentration criteria (AKI-1-sCr) and 210 (27.2%) met both urine output and creatinine criteria (AKI-1-both). Details of the outcomes for these groups are shown in Table 9-3. On univariable analysis, rates of prolonged LOS and RRT were significantly higher for AKI-1-sCr than for AKI-1-UO patients (p<0.01 for both).   The 2-year mortality rate for AKI1-sCr was also significantly worse than that for AKI-1-UO (p<0.01). Outcomes for those with AKI-1-both were worse still with prolonged LOS and RRT rates significantly worse than those for the AKI-1-sCr group (p<0.02 for both). The 2-year mortality rate for AKI-1-both was higher than that for AKI-1-sCr but this difference did not achieve statistical significance (p=0.09). The 2-year mortality rate for AKI-1-both was however significantly higher than that for AKI-1-UO (p<0.01).

Table 9-3 - Influence of urine output and serum creatinine criteria for AKI-1 on outcomes

Criteria met	N (%)	Progressed to RRT, N(%)	Prolonged LOS N(%)	2-year mortality, N(%)
Urine output alone	370	16 (4.3)	145 (39.6)	22(5.9)
Creatinine alone	192	28 (14.6) *	119 (64.7)*	24 (12.5)*
Urine output AND Creatinine	210	58 (27.6) *†	155 (76.0)* †	39 (18.6)*

AKI - acute kidney injury, RRT - renal replacement therapy, LOS - length of stay.
* = p values for comparison with urine output alone group <0.01 on univariable analysis
†= p values for comparison with serum creatinine alone group <0.02 on univariable analysis

On multivariable analysis adjusted for the logistic EuroSCORE, compared with AKI-1-UO, those with AKI-1-sCr had higher risks of prolonged LOS (OR 2.6, 95%CI 1.7-3.9, p<0.01) and RRT (OR 3.2, 95%CI 1.5-7.2, p<0.01). Similarly compared with AKI-1-UO, AKI-1-both was associated with even greater risks of prolonged LOS (OR 4.9, 95% CI 3.3-7.4, p<0.01) and RRT (OR 10.5, 95%CI 5.5-21.9, p<0.01). Mortality risk within the first two years following surgery was greater for AKI-1-both than that for AKI-1-UO (HR 2.8, 95% CI 1.6-4.8, p<0.01) but the smaller difference in mortality risk between AKI-1-sCr and AKI-1-UO over the same period (HR 1.4, 95%CI 0.7-2.7) was not statistically significant (p=0.29).

Acute Kidney Injury Stage 2

There were 222 (28.8%) patients with AKI-1 who progressed to AKI-2. In 97 (43.7%) of these patients, AKI-2 was based on urine output alone (AKI-2-UO), in 92 (41.4%) it was based on serum creatinine concentration alone (AKI-2-sCr) and in 33 (14.7%) diagnosis was based on both criteria (AKI-2-both). Outcomes for these groups are shown in Table 9-4. On univariable analysis, the rates of prolonged LOS and RRT for AKI-2-sCr were significantly higher than for AKI-2-UO (p≤0.01 for both) but the difference in 2-year mortality was not statistically significant (p=0.20). Again, the rates of prolonged LOS and RRT for AKI-2-both were significantly higher than for AKI-2-sCr (p≤0.02 for both) but the difference in 2-year mortality rates was not significantly different (p=0.16). 2-year mortality in the AKI-2-both group was however significantly higher than that in the AKI-2-UO group (p=0.01).   On multivariable analysis, compared with AKI-2-UO, AKI-2-sCr was associated

with increased risks of prolonged LOS (OR, 2.1 95%CI 1.0-4.4, p=0.04) and RRT (3.2, 95%CI 1.4-7.7, p<0.01). Similarly compared with AKI-2-UO, AKI-2-both carried even greater risk of prolonged LOS (OR 16.0, 95% CI 3.2-292.0, p<0.01) and RRT (OR 11.0, 95%CI 4.2-30.9, p<0.01). Mortality during the first two postoperative years was significantly higher for AKI-2-both than AKI-2-UO (HR 3.6, 95% CI 1.4-9.3, p<0.01) but the difference in mortality risk between AKI-2-sCr and AKI-2-UO was not statistically significant (HR 1.5, 95%CI 0.6-3.5, p=0.40).

Table 9-4 - Influence of urine output and serum creatinine criteria for AKI-2 on outcomes

Criteria met	N (%)	Progressed to RRT, N(%)	Prolonged ICU LOS, N(%)	2-year mortality, N(%)
Urine output alone	97	11 (11.3)	58 (61.1)	10 (10.3)
Creatinine alone	92	30 (32.6) *	70 (79.5)*	15 (16.3)
Urine output AND Creatinine	33	20 (60.6) *†	30 (97.0)* †	9 (27.3)*

AKI - acute kidney injury, RRT - renal replacement therapy, LOS – length of stay.

* = p values for comparison with urine output alone group ≤0.01 on univariable analysis
†= p values for comparison with serum creatinine alone group ≤0.02 on univariable analysis

Acute Kidney Injury Stage 3

AKI-3 by KDIGO criteria was diagnosed in 173 patients. In 47 (27.2%) of these patients criteria for AKI-1 or AKI-2 had not been met before RRT was started almost immediately after surgery. There were 26 (15.0%) patients who met AKI-3 criteria based on urine output or creatinine without needing renal replacement therapy. In 16 of these patients, AKI-3 was based on a high serum creatinine levels with preserved urine output. For nine AKI-3 patients, anuria was observed for 12 hours but only because the patient was not catheterised prior to discharge to the ward. 157 (85%) patients diagnosed with AKI-3 suffered a prolonged LOS. The small numbers within the subgroups precluded further analyses for AKI-3.

Comparative survival analysis

To allow a visual comparison of survival between AKI stages rather than simply within AKI stages, patients were grouped according to the urine output and serum creatinine criteria met for the maximum stage of AKI attained by the patient (up to and including AKI-2). The group of patients who went straight to AKI-3 due to early initiation of RRT is also shown. The survival curves for these groups are shown in Figure 9-2 which demonstrates clear overlap of the 2-year mortality risk for AKI-1 and AKI-2 patient subgroups.

Number at risk

AKI Level	0	120	240	360	480	600	720
No AKI	1448	1434	1430	1418	1414	1406	1391
AKI-1 UO	301	293	290	288	285	285	283
AKI-1 sCr	151	144	141	140	139	135	133
AKI-1 Both	98	92	87	85	85	83	83
AKI-2 UO	97	94	93	90	88	88	87
AKI-2 sCr	92	83	81	78	78	78	77
AKI-2 Both	33	27	25	25	24	24	24
Straight to AKI-3	47	29	29	27	27	27	27
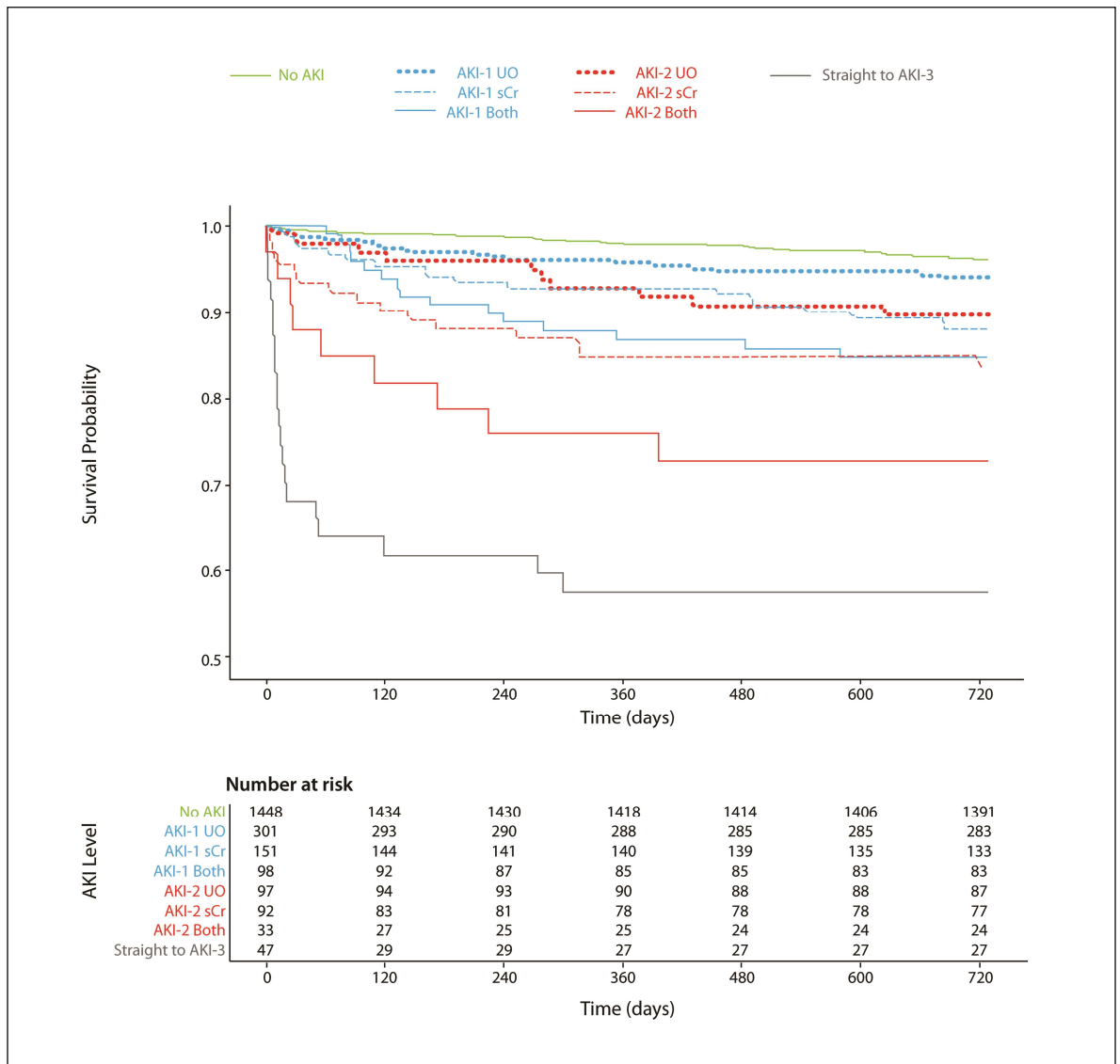
Figure 9-2 - Kaplan Meier plots stratified according to the KDIGO criteria met for the maximum stage of AKI attained up to AKI-2.

A separate group of patients who went straight to AKI-3 is shown for comparison. AKI – Acute Kidney Injury, UO – urine output, sCr – serum creatinine.

9.6. Discussion

This represents the first study to explore the associations between different criteria for multiple stages of KDIGO-defined AKI and morbidity and mortality after cardiac surgery. The frequency of AKI in our population was broadly similar to that previously reported in cardiac surgery patients.[1, 19, 20] RRT was more frequent than reported elsewhere [1, 6-8, 14] possibly as a substantial number of high risk tertiary referral, transplant and mechanical circulatory support patients are referred to our centre.

Univariable analyses identified stepwise increases in the risk of prolonged LOS, RRT and 2-year mortality within each stage of AKI where patients met urine output criteria, creatinine criteria or both. The increases in risk of poor outcomes were confirmed on multivariable analysis. This trend was previously reported in general ICU patients [21] and in cardiac surgery patients [6] although in the latter study the differences were not statistically significant.

In this study, when comparing outcomes for patients diagnosed with the same stage of AKI, all patients who met any criteria for that stage were included. This approach was chosen to improve the clinical relevance of our findings. In practice, clinicians cannot know whether a patient with AKI-1 will subsequently develop AKI-2 until the higher stage is diagnosed or the patient has left CICU without meeting AKI-2 criteria. As a consequence of this approach, some patients appear in both the AKI-1 and AKI-2 analyses. However, for comparison of survival outcomes between patients suffering different stages of AKI, each patient was placed retrospectively into a subgroup according to the maximum AKI stage reached during their CICU stay to avoid double counting.

The KDIGO AKI guidelines were developed using expert opinion at a conference of the Kidney Disease: Improving Global Outcomes group.[9] The guidelines drew on previous diagnostic criteria such as AKIN[22] and RIFLE[23] using urine output and creatinine to grade severity of renal dysfunction. The guidelines referenced several studies which demonstrated that patients with increasing stages of AKI had greater subsequent need for RRT and increased mortality risk. These findings were used to support the stratification of AKI within the KDIGO guidelines. However, none of the validation studies cited assessed the risk associated with AKI diagnosed by hourly urine output measurement.[24-30] The authors acknowledged that the urine output thresholds were less well substantiated than those related to serum creatinine concentration. Indeed, they noted that "the influence of urinary output criteria on AKI staging needs to be further investigated" and these concerns have been echoed in subsequent work validating the guidelines in their entirety.[10] The calibration of urine output thresholds might be particularly poor for

cohorts such as those who have undergone major surgery for whom oliguria may be an appropriate response to surgery.

Many studies assessing the KDIGO guidelines in cardiac surgery patients used creatinine criteria alone to classify AKI and explore its association with subsequent adverse outcomes.[12, 20] Those which have included urine output disagreed on whether UO diagnosed by AKI was associated with adverse outcomes.[2, 6, 14] The current study contains more than three times the number of participants as the largest similar study into this topic in cardiac surgery patients. It is also the first to explore differences between with rates of adverse outcomes in those who meet different criteria within the same AKI stage for both AKI-1 and AKI-2.

Our findings support conclusions drawn from earlier studies that the risk of adverse outcomes associated with AKI diagnosed by UO alone is relatively low.[2, 14]. In many cases the oliguria which resulted in diagnoses of AKI-1 and AKI-2 by urine output alone was almost certainly an appropriate physiological response to the stress of surgery. It is understandable that physiologically appropriate oliguria was associated with a prolonged LOS on CICU, but equally it is unsurprising that such oliguria was not associated with significantly increased 2-year mortality rates. In contrast, the risk of adverse outcomes in patients who met both urine output and creatinine criteria for AKI-1 or AKI-2 was markedly higher. Patients who suffered AKI-1 meeting both criteria had worse 2-year survival than those who suffered AKI-2 by urine output alone. This is highly relevant to treating clinicians; in these instances an AKI-1 diagnosis may be falsely reassuring. To ensure that the direct relationship between increasing AKI stages and risk of adverse outcomes is maintained, the AKI classification criteria may need to be adjusted to reduce the importance placed on isolated oliguria and increase the risk attributed to fulfilling both urine output and creatinine criteria. While we have studied patients undergoing cardiac surgery, similar findings may be reproduced in other cohorts, particularly those undergoing other types of major surgery and this should be the focus of further investigation.

Study Limitations

This study is based on consecutive data from a large tertiary cardiac surgery centre in the UK. The data utilised have been rigorously cleaned using reproducible algorithms. There were very few missing data and very few cases excluded from the analysis. Postoperative urine output and serum creatinine concentration data were available every patient. Creatinine was measured on the first postoperative day for 2265 of the 2267 patients (99.7%) and while this proportion did decline gradually as the length of admission increased, even on the seventh postoperative day, 220 of the 240 (91.7%) patients who remained on CICU had a daily creatinine measurement. Urine output data were available for all patients throughout their admission.

Survival data were extracted from our hospital's clinical governance database which is populated automatically by data from the NHS digital spine database. This approach is potentially less robust than using a dedicated follow-up process due to potential time delays in updating the status of the patient after death. However, as at least three months elapsed between the end of the follow up period for all patients and the extraction of survival data, this is unlikely to have affected the results.

We are confident that we have controlled for most relevant preoperative and intraoperative differences between the groups in our regression analyses by using the well validated logistic EuroSCORE. Although most known confounders including pre-operative renal function were included as variables in the logistic EuroSCORE we cannot guarantee that all potential preoperative confounders have been adjusted for. Differences in postoperative management can have profound effects on urine output and serum creatinine. In addition, variations in post-operative management could influence the outcome measures utilised in this study. Adjusting for variations in post-operative management was beyond the scope of this study and is unlikely to have influenced the results as at the time of the study there were no institutional guidelines regarding management of AKI which would have led to systematic differences in treatments between the patient groups.

This study aimed to test the KDIGO criteria across the most heterogeneous sample possible. The cohort included those undergoing cardiac transplantation and those receiving mechanical circulatory support. Such patients are known to have greater risks of AKI, RRT and other adverse outcomes than the general cardiac surgery population. Patients from these higher risk groups made up 46%(n=21) of those who started RRT prior to any urine output or creatinine criteria for AKI being met and this may have contributed to the high mortality rate associated in the "straight to AKI-3" group (Figure 9-2).

9.7. Conclusions

The current KDIGO guidelines assign the same AKI stage to patients with markedly different risks of adverse outcomes after cardiac surgery. This study identified a stepwise increase in the frequency of poor outcomes following cardiac surgery where urine output criteria, creatinine criteria or both criteria for each AKI stage were met. Moreover, 2-year mortality for some subgroups within AKI-1 clearly overlapped that of some subgroups within AKI-2. The KDIGO criteria need to be revisited before they can be used to reliably stratify the severity of AKI in cardiac surgery patients on the critical care unit. The applicability of these criteria to other

patient groups within the critical care unit, particularly those in whom oliguria may be an appropriate physiological response also requires further evaluation.

9.8. Appendix

Tables describing multivariable adjustment models

Table 9-5 - Multivariable logistic regression model for PLOS in group of patients with no AKI or AKI-1-UO

Variable	Odds Ratio	95% confidence interval		P value
(Intercept)	0.15	0.12	0.18	<0.01
Logistic EuroSCORE	1.06	1.04	1.08	<0.01
AKI-1-UO	2.80	2.16	3.64	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-1-UO – acute kidney injury stage 1 by urine output

Table 9-6 - Cox proportional hazards regression model for 2-year mortality in group of patients with no AKI or AKI-1-UO

Variable	Hazard ratio	95% confidence interval		P value
Logistic EuroSCORE	1.05	1.03	1.06	<0.01
AKI-1-UO	1.41	0.85	2.34	0.18

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-1-UO – acute kidney injury stage 1 by urine output

Table 9-7 - Multivariable logistic regression model for PLOS in group of patients with AKI-1

Variable	Odds ratio	95% confidence interval		P value
(Intercept)	0.51	0.40	0.66	<0.01
Logistic EuroSCORE	1.03	1.01	1.05	<0.01
AKI-1-UO	0.00	-	-	-
AKI-1-sCr	2.58	1.74	3.85	<0.01
AKI-1-both	4.85	3.25	7.35	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-1-UO – acute kidney injury stage 1 by urine output, AKI-1-sCr– acute kidney injury stage 1by serum creatinine concentration, AKI-1-both – acute kidney injury stage 1 by urine output and serum creatinine concentration.

Table 9-8 - Multivariable logistic regression model for RRT in group of patients with AKI-1

Variable	Odds ratio	95% confidence interval		P value
(Intercept)	0.03	0.01	0.05	<0.01
Logistic EuroSCORE	1.03	1.01	1.04	<0.01
AKI-1-UO	0.00	-	-	-
AKI-1-sCr	3.19	1.46	7.23	<0.01
AKI-1-both	10.54	5.51	21.93	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-1-UO – acute kidney injury stage 1 by urine output, AKI-1-sCr– acute kidney injury stage 1by serum creatinine concentration, AKI-1-both – acute kidney injury stage 1 by urine output and serum creatinine concentration.

Table 9-9 - Cox proportional hazards regression model for 2-year mortality in group of patients with AKI-1

Variable	Hazard ratio	95% confidence interval		P value
Logistic EuroSCORE	1.03	1.02	1.05	<0.01
AKI-1-UO	0.00	-	-	-
AKI-1-sCr	1.42	0.75	2.70	0.29
AKI-1-Both	2.81	1.64	4.82	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-1-UO – acute kidney injury stage 1 by urine output, AKI-1-sCr– acute kidney injury stage 1by serum creatinine concentration, AKI-1-both – acute kidney injury stage 1 by urine output and serum creatinine concentration.

Table 9-10 - Multivariable logistic regression model for PLOS in group of patients with AKI-2

Variable	Odds ratio	95% confidence interval		P value
(Intercept)	1.26	0.77	2.08	0.26
Logistic EuroSCORE	1.03	0.99	1.08	0.14
AKI-2-UO	0.00	-	-	-
AKI-2-sCr	2.10	1.03	4.37	0.04
AKI-2-both	15.99	3.16	292.04	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-2-UO – acute kidney injury stage 2 by urine output, AKI-2-sCr– acute kidney injury stage 2 by serum creatinine concentration, AKI-2-both – acute kidney injury stage 2 by urine output and serum creatinine concentration.

Table 9-11 - Multivariable logistic regression model for RRT in group of patients with AKI-2

Variable	Odds ratio	95% confidence interval		P value
(Intercept)	0.11	0.05	0.22	<0.01
Logistic EuroSCORE	1.01	0.98	1.04	0.62
AKI-2-UO	0.00	-	-	-
AKI-2-sCr	3.18	1.40	7.65	<0.01
AKI-2-both	11.00	4.17	30.94	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-2-UO – acute kidney injury stage 2 by urine output, AKI-2-sCr– acute kidney injury stage 2 by serum creatinine concentration, AKI-2-both – acute kidney injury stage 2 by urine output and serum creatinine concentration.

Table 9-12 - Cox proportional hazards regression model for 2-year mortality in group of patients with AKI-2

Variable	Hazard ratio	95% confidence interval		P value
Logistic EuroSCORE	1.04	1.02	1.06	<0.01
AKI-2-UO	0.00	-	-	-
AKI-2-sCr	1.46	0.60	3.51	<0.01
AKI-2-Both	3.58	1.38	9.30	<0.01

EuroSCORE- European System for Cardiac Operative Risk Evaluation; AKI-2-UO – acute kidney injury stage 2 by urine output, AKI-2-sCr– acute kidney injury stage 2 by serum creatinine concentration, AKI-2-both – acute kidney injury stage 2 by urine output and serum creatinine concentration.

9.9. References

1.   Machado MN, Nakazone MA, Maia LN. Prognostic value of acute kidney injury after cardiac surgery according to kidney disease: improving global outcomes definition and staging (KDIGO) criteria. *Plos One* 2014;9(5):e98028.

2.   Lagny MG, Jouret F, Koch JN, Blaffart F, Donneau AF, Albert A, et al. Incidence and outcomes of acute kidney injury after cardiac surgery using either criteria of the RIFLE classification. *BMC Nephrology* 2015;16:76.

3.   Dasta JF, Kane-Gill SL, Durtschi AJ, Pathak DS, Kellum JA. Costs and outcomes of acute kidney injury (AKI) following cardiac surgery. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2008;23(6):1970-1974.

4.   Duthie FA, McGeehan P, Hill S, Phelps R, Kluth DC, Zamvar V, et al. The Utility of the Additive EuroSCORE, RIFLE and AKIN Staging Scores in the Prediction and Diagnosis of Acute Kidney Injury after Cardiac Surgery. *Nephron Clinical practice* 2014;128(1-2):29-38.

5.   Elmistekawy E, McDonald B, Hudson C, Ruel M, Mesana T, Chan V, et al. Clinical impact of mild acute kidney injury after cardiac surgery. *Annals of Thoracic Surgery* 2014;98(3):815-822.

6.   Petaja L, Vaara S, Liuhanen S, Suojaranta-Ylinen R, Mildh L, Nisula S, et al. Acute Kidney Injury After Cardiac Surgery by Complete KDIGO Criteria Predicts Increased Mortality. *J Cardiothorac Vasc Anesth* 2017;31(3):827-836.

7.   Thakar CV, Worley S, Arrigain S, Yared J-P, Paganini EP. Influence of renal dysfunction on mortality after cardiac surgery: Modifying effect of preoperative renal function. *Kidney Int* 2005;67(3):1112-1119.

8.   Mangano CM, Diamondstone LS, Ramsay JG, Aggarwal A, Herskowitz A, Mangano DT. Renal dysfunction after myocardial revascularization: risk factors, adverse outcomes, and hospital resource utilization. The Multicenter Study of Perioperative Ischemia Research Group. *Annals of Internal Medicine* 1998;128(3):194-203.

9.   Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practical Guideline for Acute Kidney Injury in *Kidney Inter* 2012. p. 1-138.

10.   Ralib A, Pickering JW, Shaw GM, Endre ZH. The urine output definition of acute kidney injury is too liberal. *Critical Care* 2013;17(3):R112.

11.   Kristovic D, Horvatic I, Husedzinovic I, Sutlic Z, Rudez I, Baric D, et al. Cardiac surgery-associated acute kidney injury: risk factors analysis and comparison of prediction models. *Interactive cardiovascular and thoracic surgery* 2015;21(3):366-373.

12.   Birnie K, Verheyden V, Pagano D, Bhabra M, Tilling K, Sterne JA, et al. Predictive models for kidney disease: improving global outcomes (KDIGO) defined acute kidney injury in UK cardiac surgery. *Critical Care* 2014;18(6):606.

13.   Machado MN, Nakazone MA, Maia LN. Acute kidney injury based on KDIGO (Kidney Disease Improving Global Outcomes) criteria in patients with elevated baseline serum creatinine undergoing cardiac surgery. *Revista Brasileira de Cirurgia Cardiovascular : órgão oficial da Sociedade Brasileira de Cirurgia Cardiovascular* 2014;29(3):299-307.

14.   McIlroy DR, Argenziano M, Farkas D, Umann T, Sladen RN. Incorporating oliguria into the diagnostic criteria for acute kidney injury after on-pump cardiac surgery: impact on incidence and outcomes. *J Cardiothorac Vasc Anesth* 2013;27(6):1145-1152.

15.   Karim HM, Yunus M, Saikia MK, Kalita JP, Mandal M. Incidence and progression of cardiac surgery-associated acute kidney injury and its relationship with bypass and cross clamp time. *Ann Card Anaesth* 2017;20(1):22-27.

16.   Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J* 2003;24(9):881-882.

17.   Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg* 2013;43(6):1146-1152.

18.   R Core Team. R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing 2015.

19.   Jorge-Monjas P, Bustamante-Munguira J, Lorenzo M, Heredia-Rodríguez M, Fierro I, Gómez-Sánchez E, et al. Predicting cardiac surgery–associated acute kidney injury: The CRATE score. *Journal of critical care* 2016;31(1):130-138.

20.   Bastin AJ, Ostermann M, Slack AJ, Diller G-P, Finney SJ, Evans TW. Acute kidney injury after cardiac surgery according to Risk/Injury/Failure/Loss/End-stage, Acute Kidney Injury Network, and Kidney Disease: Improving Global Outcomes classifications. *Journal of critical care* 2013;28(4):389-396.

21.   Kellum JA, Sileanu FE, Murugan R, Lucko N, Shaw AD, Clermont G. Classifying AKI by Urine Output versus Serum Creatinine Level. *Journal of the American Society of Nephrology* 2015;26:2231-2238.

22.   Mehta RL, Kellum JA, Shah SV, Molitoris BA, Ronco C, Warnock DG, et al. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Crit Care* 2007;11.

23.   Bellomo R, Ronco C, Kellum JA, Mehta RL, Palevsky P. Acute renal failure–definition, outcome measures, animal models, fluid therapy and information technology needs: the Second

International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care* 2004;8.

24.   Ali T, Khan I, Simpson W, Prescott G, Townend J, Smith W, et al. Incidence and outcomes in acute kidney injury: a comprehensive population-based study. *J Am Soc Nephrol* 2007;18(4):1292-1298.

25.   Joannidis M, Metnitz B, Bauer P, Schusterschitz N, Moreno R, Druml W, et al. Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the SAPS 3 database. *Intensive Care Med* 2009;35(10):1692-1702.

26.   Bagshaw SM, George C, Dinu I, Bellomo R. A multi-centre evaluation of the RIFLE criteria for early acute kidney injury in critically ill patients. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2008;23(4):1203-1210.

27.   Uchino S, Bellomo R, Goldsmith D, Bates S, Ronco C. An assessment of the RIFLE criteria for acute renal failure in hospitalized patients. *Crit Care Med* 2006;34(7):1913-1917.

28.   Thakar CV, Christianson A, Freyberg R, Almenoff P, Render ML. Incidence and outcomes of acute kidney injury in intensive care units: a Veterans Administration study. *Crit Care Med* 2009;37(9):2552-2558.

29.   Lassnigg A, Schmidlin D, Mouhieddine M, Bachmann LM, Druml W, Bauer P, et al. Minimal changes of serum creatinine predict prognosis in patients after cardiothoracic surgery: a prospective cohort study. *J Am Soc Nephrol* 2004;15(6):1597-1605.

30.   Hoste EAJ, Clermont G, Kersten A, Venkataraman R, Angus DC, De Bacquer D, et al. RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. *Critical Care* 2006;10(3):R73-R73.

Chapter Ten: A novel patient-specific model for predicting severe oliguria; development and comparison with KDIGO acute kidney injury classification (submitted journal article)

Chapter type: Journal article

Status: Undergoing 2nd peer review by Critical Care Medicine following revisions

Submission Date: 17/01/2018

Reference: Howitt SH, Oakley J, Grant SW, Caiado C, Goldstein M, Malagon I, McCollum CN. Development of a novel patient-specific model for predicting severe oliguria and associated adverse outcomes.

Author contributions, SH devised the study, collected and cleaned the data and guided the development and validation of the model. SH performed the statistical analyses of the model's performance and wrote the first draft. JO and CC developed the Bayesian risk prediction model and ran the model in the validation cohort under the guidance of MG. SWG revised the first draft and guided the evaluation of model performance. IM and CM revised the first draft. All authors approved the final manuscript.

Rationale for inclusion of this study in the thesis

The previous chapter identified key limitations of the existing KDIGO classification of AKI according to urine output. In particular the guidelines' criteria classify a large number of patients as suffering AKI-1 by urine output and in our study this group experienced relatively good outcomes. The aim of this chapter was to devise a novel method of analysing urine output to better identify patients at the risk of adverse outcomes. As discussed within this chapter, a stricter urine output criterion of 0.3ml/kg/hour for six hours (severe oliguria) was identified in the literature. However, potentially serious renal damage may occur while waiting for this stricter criterion to be achieved. In order to mitigate this risk, this chapter attempts to use Bayesian forecasting to identify those who would suffer severe oliguria in advance of it happening. This was achieved through collaboration with statisticians at Durham University. The discrimination and calibration of the model's risk predictions at multiple time points are tested. Outcomes of patients with high predicted risk of oliguria are then analysed.

10.1. Additional data processing for this manuscript.

This chapter describes the development of a novel risk prediction model which identifies patients at risk of severe oliguria based on the analysis of their previous urine output. The cleaned urine output data were created in steps 106-11 of the main cleaning code described above. The development of the model was performed by Jordan Oakley under the supervision of Camila Caiado. Classification of the model was then analysed by the author of this thesis.

Step 1. Remove all data for patients who underwent mechanical circulatory support, cardiac transplantation or preoperative RRT.

Step 2. Supply the urine output/kg data to the team of statisticians at Durham University.

Step 3. Determine the severe oliguria status for each patient for each hour of their CICU admission.

Step 4. Compare predictions made by the model for each patient at 12, 24, 36, 38 and 72 hours with outcomes observed within the next 12 hours using ROC curves.

Step 5. Determine the O:E ratio for severe oliguria prediction using the model.

Step 6. Produce calibration plots for the model's predictions at each time point.

Step 7. Select the subgroup of patients who were classified as high risk of urine output.

Step 8. Select the subgroup of patients who were always classified as low risk of urine output.

Step 9. Determine how often frusemide was given to each patients in each risk group.

Step 10. Determine which patients fulfilled the AKI-1 UO criterion and the time at which the criterion was met.

Step 11. Select only data recorded during the first CICU admission for each patient.

Step 12. Group patients according to whether they met the AKI-1 UO criterion and whether the model identified them as high risk.

Step 13. Determine the incidence of RRT which occurred following model high-risk classification or fulfilment of AKI-1 UO criterion.

Step 14. Where onset of RRT occurred before high risk prediction or AKI-1 UO classify this as a false negative prediction.

Step 15. Compare rates of RRT for those classified as low and high risk by the model.

Step 16. Compare rates of prolonged length of stay for those classified as low and high risk by the model.

Step 17. Compare rates of hospital mortality for those classified as low and high risk by the model.

Step 18. Perform logistic regression analysis to account for confounders when using model classification to predict subsequent RRT.

Step 19. Perform logistic regression analysis to account for confounders when using model classification to predict prolonged length of stay.

Step 20. Perform logistic regression analysis to account for confounders when using model classification to predict hospital mortality.

Step 21. Calculate the time difference between high risk classification and onset of severe oliguria.

Step 22. Determine the sensitivity, specificity, positive predictive value and negative predictive value of the model's classification and the AKI-UO classification when identifying those at risk of RRT.

Step 23. Create the precision recall curves for prediction of severe oliguria at 12, 24, 36, 38 and 72 hours.

Step 24. Repeat analyses from steps 4-6 using the outcome of severe oliguria occurring within the next 6 hours only (sensitivity analysis).

10.2. Abstract

Objective

The KDIGO urine output criteria for acute kidney injury (AKI) lack specificity for identifying patients at risk of adverse renal outcomes. The objective was to develop a method to analyse hourly urine output values to identify those at risk of developing severe oliguria.

Design

This was a retrospective cohort study utilising prospectively collected data.

Setting

A cardiac intensive care unit in the UK.

Patients

Patients undergoing cardiac surgery between January 2013 and November 2017

Measurement and main results

Patients were randomly assigned to development (n=981) and validation (n=2389) datasets. A patient-specific, dynamic Bayesian model was developed to predict future urine output on an hourly basis. Model discrimination and calibration for predicting severe oliguria (<0.3ml/kg/hr for 6 hours) occurring within the next 12 hours were tested in the validation dataset at multiple time points. Patients with a high-risk (probability of severe oliguria >0.8) were identified and their outcomes were compared with those for low-risk patients and for patients who suffered AKI based on KDIGO urine output criteria.

Model discrimination was excellent at all time points (AUC >0.9 for all). Calibration of the model's predictions was also excellent. After adjustment using multivariable logistic regression, patients in the high-risk group were more likely to require renal replacement therapy (OR 10.4, 95%CI 5.9-18.1), suffer prolonged hospital stay (OR 4.4, 95% CI 3.0-6.4) and die in hospital (OR 6.4, 95%CI 2.8-14.0) (p<0.001 for all). Outcomes for those identified as high-risk by the model were significantly worse than for patients who met the KDIGO urine output criterion for AKI.

Conclusions

This novel, patient-specific model accurately identifies patients at increased risk of severe oliguria. Classification according to model predictions outperformed the KDIGO urine output criteria. As the new model identifies patients at risk before severe oliguria develops it could potentially facilitate intervention to improve patient outcomes.

10.3. Introduction

Acute kidney injury (AKI) is defined and stratified by the KDIGO AKI guidelines(1) and occurs in up to 75% of patients in general intensive care units (2, 3) and up to 30% of patients following cardiac surgery.(4) The KDIGO guidelines stratify the severity of AKI based on serum creatinine concentration and urine output. Studies in both cardiac surgery and general ICU patients have shown that the guidelines' creatinine criteria successfully identify patients with increased risk of prolonged length of stay, short-term mortality and long term mortality.(3, 5-8) However, there is less agreement about the value of the guidelines' urine output criteria which define AKI as urine output below 0.5ml/kg/hr for more than 6 hours. Most large studies were unable to obtain enough urine output data to adequately assess the importance of the urine output criteria in the prediction of adverse outcomes.(3, 7, 8) Some smaller studies demonstrated that calibration of the KDIGO urine output thresholds may be inadequate by showing that patients diagnosed with AKI by urine output alone had relatively good outcomes compared with those who also met the guideline's serum creatinine criteria.(2, 9-11) Ralib et al demonstrated that a urine output threshold of 0.3ml/kg/hr for 6 hours (severe oliguria) was more closely associated with adverse outcomes in general ICU patients.(9) However, use of this threshold rather than the 0.5ml/kg/hr for 6 hours threshold specified in the KDIGO stage 1 definitions could lead to adverse patient outcomes related to the 6 hours of marked oliguria required before risk stratification could occur. Dynamic Bayesian modelling (12, 13) has been used in related settings (14, 15) and could provide a solution to this problem by identifying those at greatest risk of severe oliguria early enough to allow treatment to be administered. The objective of this study was to develop and validate a patient-specific dynamic Bayesian model which could run in real time to predict the risk of developing severe oliguria. Associations between those at a high predicted risk of severe oliguria and adverse outcomes were investigated and outcomes in this high-risk group were compared with those patients who met existing KDIGO urine output criteria for AKI.

10.4. Materials and Methods

10.4.1. Data

Prospectively collected data from adult patients admitted to the cardiac intensive care unit (CICU) following cardiac surgery between January 2013 and November 2017 were analysed. Patients receiving mechanical circulatory support (MCS) or cardiac transplantation were excluded. Patients who received renal replacement therapy (RRT) preoperatively were also excluded.

Hourly urine output values and their timings together with the timing of any decision to initiate RRT were extracted from the electronic patient record.   Only urine output data recorded before the initiation of RRT was analysed.   Outcome data was collected from the hospital's clinical governance database. All data was cleaned and stored in the Vascular Governance NorthWest (VGNW) database, handled according to the database's ethical approvals and pseudonymised prior to analysis. As data was pseudonymised prior to analysis, the Research Ethics Committee concluded that ethical approval for these analyses was not necessary.   All data cleaning and analysis was performed using R (R Foundation for statistical computing).(16)

10.4.2. Model development

Eligible patients were randomly assigned to either model development or model validation datasets in a ratio of 1:2.5 to ensure a development group of around 1000 patients. A dynamic linear model was developed using data from the development dataset. The model analysed each patient's own hourly urine output values and then from the 6th hour on CICU predicted the individual's urine output for the next 6 hours. Calculations were repeated on an hourly basis allowing the model to produce updated predictions throughout the CICU stay as each new measurement became available. The probability of the next 6 hours' urine output being below 0.3ml/kg/hr was calculated using Bayesian forecasting.   The model applied weightings to the contributions of urine output values according to how recent they were with the most recent values deemed the most relevant. This allowed the forecast to update quickly in response to changing trends. During model development, it was identified that for the majority of patients the trajectory of urine output was relatively stable. However, a reproducible trend was identified in the urine output of those who developed oliguria. The model's performance was tested with and without the inclusion of the trend term and the trend term was found to improve model performance. A detailed description of model development is described in the appendix.

10.4.3. Model validation (statistical analyses)

It was recognised that for a subgroup of patients, the model could potentially provide inappropriate reassurance to clinicians. The model could correctly predict a low risk of severe oliguria occurring within 6 hours in patients whose urine output remained >0.3ml/kg/hr for the next 6 hours but who went on to suffer severe oliguria soon afterwards. The validation analyses therefore tested the model's ability to identify which patients would suffer severe oliguria (UO

<0.3ml/kg/hr for 6 hours) within 12 hours of the prediction. Risk predictions made during the last 12 hours of a patient's admission were disregarded as it was not possible to confirm if severe oliguria subsequently occurred following discharge from CICU. For completeness, performance of the model when predicting severe oliguria limited to the six hours following predictions was also assessed with full results in the appendix. The validation studies were described according to the criteria outlined in the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.(17)

Discrimination (the ability to distinguish those who would suffer severe oliguria from those who would not) was assessed using Receiver Operator Curve (ROC) analyses. The 95% confidence intervals for the area under the curves (AUC) were calculated using DeLong's method.(18) Due to the low incidence of severe oliguria, precision recall curves were also used to assess the impact of the large proportion of true negative results on the model's performance in the ROC analyses.(19) Calibration (how well predicted risk matched observed outcomes) was assessed using the ratio of observed to expected outcomes (O:E ratio) and calibration plots.(20) The calibration plots were used to illustrate the agreement between predicted and observed and risk of severe oliguria for patients assigned to twenty evenly sized groups according to their predicted risk.

It is anticipated that in clinical practice clinicians are likely to interpret the model's continually updated predictions for the risk of severe oliguria rather than a binary high/low risk classification. However, to allow comparison of the model's predictions with the existing categorical KDIGO classification, patients were assigned to either a high-risk or a low-risk group. Patients for whom the probability of severe oliguria reached >0.8 during their stay were arbitrarily classified as high-risk and those who did not were classified as low-risk. This relatively high threshold was selected a priori as the aim was to produce a classification with a high specificity. Associations between this classification and postoperative RRT, prolonged length of stay (PLOS) and hospital mortality were tested using univariable and multivariable analyses. Outcomes for patients grouped according to classification by the model and the KDIGO criterion were also compared. PLOS was defined as a hospital stay >10 days. If RRT was initiated within three hours of CICU admission, the patient was excluded from the analyses as case note analyses revealed that all of these decisions to start RRT had been made during surgery before the patient arrived on CICU. If the decision to initiate RRT was made before a high-risk classification the RRT was considered to have been administered to a low risk patient. Univariable analyses were performed using the Chi Square test or Fisher's exact test in the event of sparse data. Multivariable logistic regression was used to adjust for the confounding effects of pre- and perioperative variables associated with adverse outcomes using the extensively validated logistic EuroSCORE model.(21, 22) Cardiopulmonary

bypass (CPB) time was used as a surrogate marker to adjust for intra-operative procedure complexity.

The sensitivity, specificity, positive predictive value and negative predictive value of classification by the new model based on the arbitrary threshold of 0.8 for the identification of those at risk of subsequent RRT were calculated. These values were compared with equivalent values obtained when classifying patients according to i) the KDIGO UO criterion (UO <0.5ml/kg/hr for 6 hours) and ii) observed severe oliguria (UO<0.3ml/kg/hr for 6 hours).

10.4.4. Missing data

Where hourly urine output was recorded as "0" this value was used. Where hourly values were blank, the next recorded urine output was divided by the number of hours that had elapsed since the previous reading and this value was substituted for the blank values. Where this imputation resulted in urine output lower than the 0.5ml/kg for 6 hours the cases notes were examined and the urine output entries verified through entries in the nursing notes. Where weight was missing, the value was imputed using the median weight for a patient of that gender.

10.5. Results

In total 3,602 patients were admitted to CICU following cardiac surgery, 228 were excluded as they underwent cardiac transplantation or received MCS and four patients were excluded as they received RRT preoperatively. Of the eligible 3370 patients, 981 were randomly assigned to the development group and the remaining 2389 patients were assigned to the validation group. The patient characteristics of each group are shown in Table 10-1. Patient weight was missing for 13 (1.3%) and 23 (1.0%) patients in the development and validation cohorts respectively.

Table 10-1 - Patient Characteristics

Characteristic	Development group (n=981)	Validation group (=2389)
Age, mean (sd) , years	66.4 (11.2)	66.7 (10.9)
Female gender, %	279 (28.2)	660 (27.6)
Weight, mean (sd), Kg	82.2 (15.9)	81.8 (16.4)
Logistic EuroSCORE, median (Interquartile range)	3.8 (2.1-7.4)	3.7 (2.0-7.0)
Operation, n (%)		
CABG	544 (55.5)	1394 (58.4)
Valve	227 (23.1)	505 (21.1)
CABG and Valve	125 (12.7)	337 (14.1)
Aortic	65 (6.6)	118 (5.0)
Other – minor	3 (0.3)	5 (0.2)
Other – major	17 (1.7)	30 (1.3)
Urgency, n (%)		
Elective	574 (58.5)	1380 (57.8)
Urgent	395 (40.3)	958(40.1)
Emergency	9 (0.9)	44 (1.8)
Salvage	3 (0.3)	7 (0.3)
CPB time, median (Interquartile range), minutes	102.0 (81.0-129.0)	102.0 (82.0-129.0)

CABG – coronary artery bypass grafting, CPB – cardiopulmonary bypass

In the validation cohort, 2088 (87.4%) patients suffered at least one hour of urine output below 0.3ml/kg/h. There were 197 (8.2%) patients who experienced severe oliguria and 89 (3.7%) patients who required RRT. In total, 4942 (2.8%) hourly urine output entries were missing and these values were imputed using the methods described in the previous section. A total of 19 (0.8%) patients received RRT within three hours of arrival on CICU and these patients were excluded from the RRT analyses. PLOS was observed in 589 (24.7%) patients and 36 (1.5%) died prior to hospital discharge. There were no missing outcome data.

10.5.1. Predicting severe oliguria

The AUCs for the prediction of severe oliguria within the next 12 hours for predictions made at 12, 24, 36, 48 and 72 hours are shown in Figure 10-1. At each time point the AUC for the predictions was >0.9 representing excellent discrimination between those who did and did not go on to suffer severe oliguria within the next 12 hours.
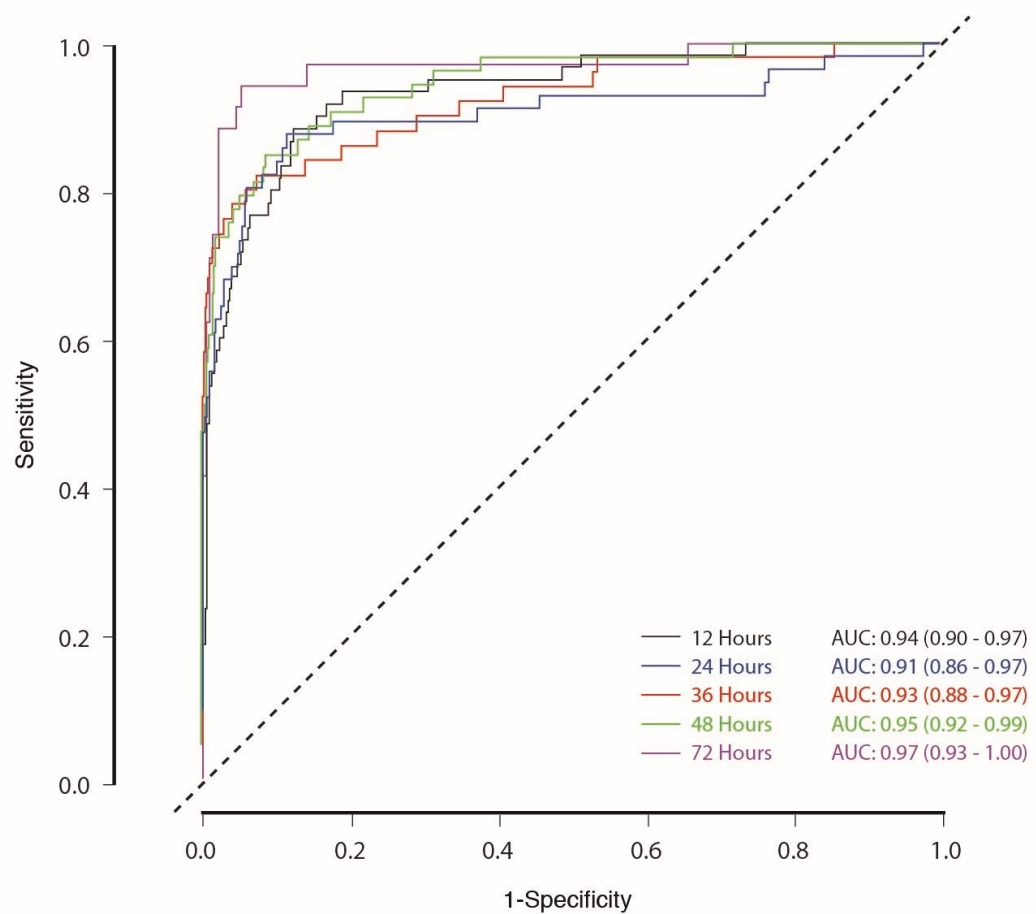
Figure 10-1 - Receiver operating characteristic curves for the prediction of severe oliguria (<0.3ml/kg/hr for 6 hours) during the next 12 hours following predictions made by the model at 12, 24, 36, 48 and 72 hours.

As illustrated by Figure 10-2 and the O:E ratios detailed in Table 10-2, calibration was also excellent.
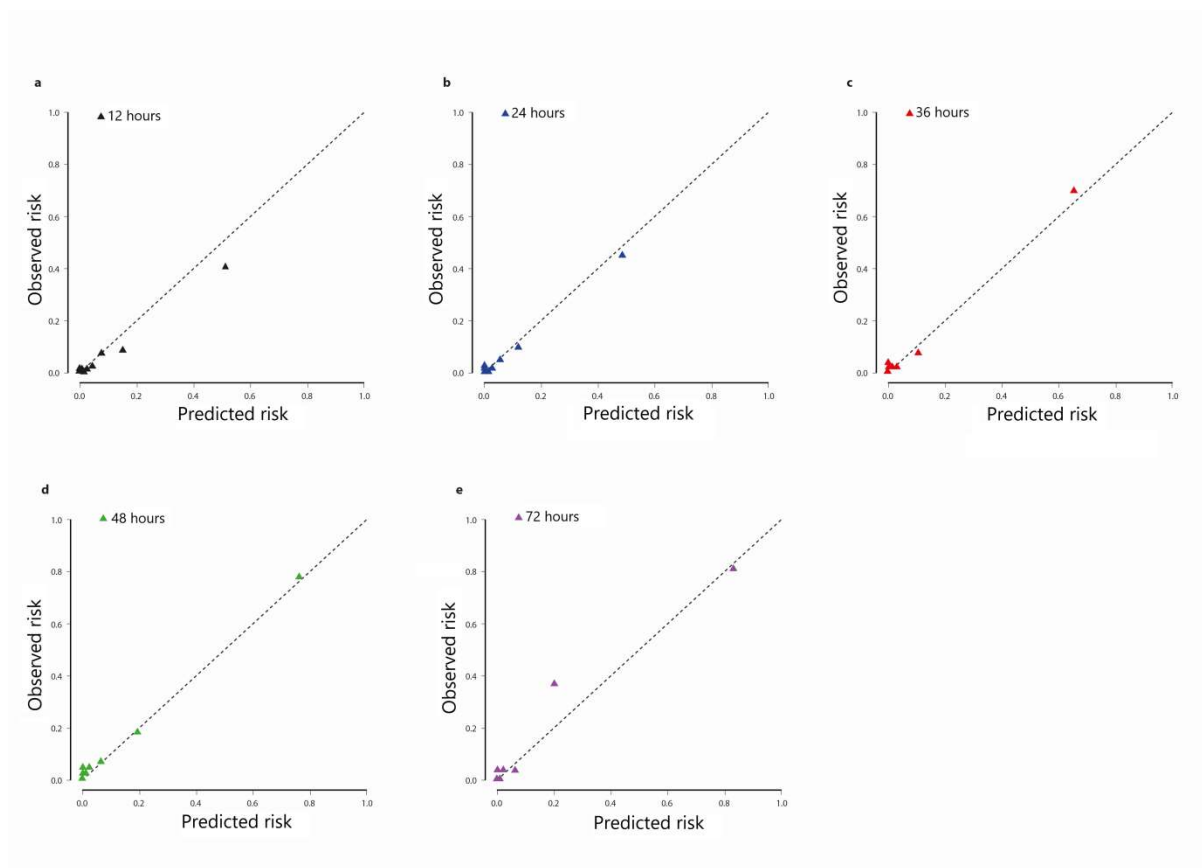
Figure 10-2 - Calibration plots for the Bayesian model's prediction of severe oliguria (0.3ml/kg/hr for 6 hours) during the next 12 hours at time points a)12 hours, b)24 hours, c)36 hours, d)48 hours and e)72 hours. Patients were split into groups according to predicted risk. For each of the twenty groups, mean observed risk is plotted against mean predicted risk.

Table 10-2 - Comparison of observed outcomes and model's predictions for severe oliguria occurring within 12 hours

Time point (number of patients still on CICU)	Observed severe oliguria within 12 hours, n(%)	Predicted severe oliguria within 12 hours, n(%)	O:E ratio
12 hours (1947)	61 (3.1)	82 (4.2)	0.74
24 hours (1694)	57 (3.4)	61 (3.6)	0.93
36 hours (1137)	51 (4.5)	44 (3.9)	1.16
48 hours (909)	54 (5.9)	48(5.3)	1.13
72 hours (545)	35 (6.4)	30 (5.6)	1.15

CICU – coronary artery bypass grafting, severe oliguria – urine output <0.3ml/kg/hr for 6 hours, O:E ratio – ratio of observed to expected outcomes

The precision recall curves (Figure 10-3 in the appendix) illustrate the trade-off between ensuring that every patient who will go on to suffer AKI is identified and that the number of false positives is minimised. As shown in Figure 10-3, as recall (also known as sensitivity) approaches 1 the Precision (positive predictive value) falls. This effect was most pronounced for predictions made in the first 24 hours.

Table 10-6 of the appendix describes the model's performance when predicting severe oliguria occuring within 6 hours of prediction. Discrimination was consistently better than when predicting severe oliguira occuring within 12 hours following predictions but risk was consistenty overestimated. Across the five time points analysed there were 258 incidences where a patient developed severe oliguria within 12 hours of predictions, however on 109 occasions severe oliguria only developed between 7 and 12 hours after prediction.

10.5.2. Classification task

In the validation dataset 158 patients experienced a probability of severe oliguria >0.8 and were assigned to the high-risk group. The remaining 2231 patients were assigned to the low-risk group. Outcomes for these two groups are reported in Table 10-3.

Table 10-3 - Outcome of patients according to classification by the Bayesian model

Group	RRT, n(%)	PLOS, n(%)	Mortality, n(%)
High-risk (n=158)	29(18.4)*	93 (58.9) *	14 (8.9) *
Low-risk (n=2231)	41(1.8%)	496 (22.2)	22 (1.0)

* p<0.001 when compared to low-risk classification by the model
RRT = renal replacement therapy, PLOS = prolonged length of stay in hospital

High-risk patients experienced increased rates of subsequent RRT, PLOS and hospital mortality compared with those classified as low-risk (P<0.001 for all outcomes). On multivariable analysis, high-risk classification was associated with increased risk of RRT (OR 10.4, 95%CI 5.9-18.1), PLOS (OR 4.4, 95% CI 3.0-6.4) and hospital mortality (OR 6.4, 95%CI 2.8-14.0) (p<0.001 for all outcomes). The multivariable models used for risk adjustment are shown in the Appendix (Tables 10-7 to 10-9). The median (IQR) time from high-risk classification to the onset of severe oliguria of 3.0 (0.0-4.0) hours

The KDIGO urine output criterion identified 628 patients (26.3%) as suffering AKI by urine output. The outcomes for patients classified according to the model's predictions and the KDIGO criterion are compared in Table 10-4. Outcomes for those classified as being at high risk by the model and those meeting the KDIGO criteria were not compared directly as some patients would have been included in both groups.

Table 10-4 - Outcomes for patients grouped according to risk level as determined analysis of urine output by KDIGO-AKI guideline and the Bayesian model.

Group	n (%)	RRT, n (%)	PLOS, n (%)	Hospital mortality, n(%)
Low-risk by model and no KDIGO AKI	1725 (72.2)	15 (0.9)	320 (18.6)	10 (0.6)
Low-risk by model but KDIGO AKI	506 (21.2)	18 (3.6)	176 (34.8)	12 (2.4)
High-risk by model but no KDIGO AKI	36 (1.5)	3 (8.3)	30(83.3)	3 (8.3)
High-risk by model and KDIGO AKI	122 (5.1)	26 (21.3)	73 (59.8)	11 (9.0)

KDIGO =Kidney Disease Improving Global Outcomes, UO = urine output, AKI = Acute Kidney Injury, PLOS = prolonged length of stay in hospital, RRT = renal replacement therapy

Patients who met the KDIGO urine output criterion for AKI but were classified as low-risk by the model (n=506) experienced rates of RRT (3.6%), PLOS (34.8%) and mortality (2.4%) which were significantly lower than the risks for those classified as high-risk by the Bayesian model ($p<0.001$ for all). When used to predict future RRT requirement, the Bayesian model classification achieved greater specificity and positive predictive value (but lower sensitivity) than the KDIGO AKI criterion. The performance of the dynamic Bayesian model was almost identical to that achieved by classification according to actual observed oliguria. (Table 10-5)

Table 10-5 - Performance of the Bayesian model, existing KDIGO AKI-UO criterion and severe oliguria when identifying those at risk of RRT.

Classification Method	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
AKI-UO	0.74	0.75	0.08	0.99
Model	0.41	0.94	0.18	0.98
Severe oliguria	0.41	0.94	0.18	0.98

Severe oliguria = observed UO <0.3ml/kg for 6 hours, AKI-UO = observed UO <0.5ml/kg for 6 hours, RRT = renal replacement therapy

10.6. Discussion

This patient-specific dynamic Bayesian model was developed and validated in separate cohorts which together contained high quality, prospectively-gathered data for over 3000 patients. The number of patients and observations included in the study were considered large enough to justify the split sample approach of the study. As shown in Table 10-1, the characteristics of patients in each group were similar.

The model successfully identified patients at risk of severe oliguria demonstrating excellent discrimination and calibration at each time point. Outcomes were significantly worse for patients deemed to be at a high-risk (probability >0.8) of severe oliguria than for those assigned to the low-risk group. Those identified as high-risk by the model also suffered worse outcomes than those who only met the KDIGO urine output criterion for AKI.

In clinical practice, classification into high and low-risk groups based on an arbitrary threshold is unlikely to be necessary and significantly diminishes the usefulness of the model.   Rather, patient monitoring software would analyse the individual's urine output data in real-time and display predictions of hourly urine output for the next six hours alongside the probability that these value would fall below a chosen threshold. The use of threshold values for hourly urine output is established in all widely used AKI classifications (1, 23, 24) and is therefore well understood by the majority of clinicians. This information, together with the trend of risk for that patient would inevitably be much more useful to a treating clinician than knowledge of the patient's risk group based on an arbitrarily dichotomised risk classification. In this study, a threshold was used to dichotomise the patients to allow the comparison of outcomes observed in patients classified as high and low-risk by the model. The categorisation also allowed comparison of outcomes between patients classified as high-risk by the model and patients who met the existing KDIGO AKI criteria. The threshold used for the classification exercise was deliberately high at 0.8 to reduce the number of false positive high-risk classifications which are a weakness of the existing KDIGO AKI classification.(2, 9, 10) As a result a large subgroup (n=506) met the KDIGO AKI criterion but were classified as low-risk by the model. Outcomes for these patients were significantly better than for the group classified as high-risk by the model suggesting that for a large proportion of those who meet the KDIGO urine output criterion, the risk of adverse outcomes is actually relatively low.

The significant increase in risk of adverse outcomes found to be associated with a predicted or observed fall in urine output to < 0.3ml/kg/hr for 6 hours is similar to that found in general ICU patients(9) and justifies the selection of this threshold in this study. Risk stratification was not

significantly improved when classification was made according to observed rather than predicted severe oliguria. The main advantage of using the dynamic Bayesian model is that it provides reliable, early warnings of impending severe oliguria before it occurs, allowing time to deliver treatments with the aim of preventing severe oliguria and its consequences. Even if a warning were only raised when a probability of 0.8 for severe oliguria was reached - as in our classification exercise – this would allow interventions aimed at preserving renal function. In reality patients for whom risk of severe oliguria is increasing are likely to be reviewed before a probability of 0.8 is reached, affording even more time for intervention.

Clinical use of a urine output screening protocol which employs this dynamic Bayesian model is perfectly feasible because although mathematically complex, the model is computationally inexpensive and can run on standard computers or tablets available at the bedside. The model uses the trend of urine output rather than comparison of point values against arbitrary thresholds. The progressive decline in urine output towards the defined threshold of 0.3ml/kg is intuitively more relevant than the occurrence of a point value below an arbitrary "normal". Indeed, over 85% of those classified as low-risk suffered at least one hour of urine output below 0.3ml/kg/hr but this group had excellent outcomes. As the only data required by the model are patient weight and hourly urine output values, the model should be transferrable across all patients on critical care units. In this study we chose to calculate the probability of urine output dropping below 0.3ml/kg/hr but this threshold could be altered to suit different patient cohorts. Under these circumstances the model could be useful across a range of settings, alerting clinicians to the risk of urine output dropping below a threshold they consider to be clinically significant.

While these results are encouraging, analyses of urine output alone cannot identify all patients at risk of adverse outcomes related to renal dysfunction. Indeed, 41 patients received RRT despite being classified as low-risk because their urine output was maintained around or above 0.3ml/kg/hr. Analysis of the EPR for these patients, identified deranged biochemistry (elevated urea and/or creatinine concentrations, hyperkalaemia or metabolic acidosis) (n=31), fluid overload (n=16), hyperlactataemia (n=4) and sepsis (n=1) as the indications for RRT initiation. In addition, while the novel model accurately predicted severe oliguria, less than 20% of those who suffered severe oliguria went on to require RRT.

Currently, creatinine concentration performs a key role in the identification of those at risk of adverse outcomes related to renal dysfunction. The existing KDIGO(1) creatinine criteria - which are shared by the AKIN and RIFLE guidelines (23, 24) - have been shown to stratify risk accurately in both cardiac surgery patients (25, 26) and the general inpatient population (27, 28). Similarly, recent advances in the use of biomarkers have been shown to enable the early identification of

those at increased risk of adverse outcomes related to renal dysfunction (29-31). Moreover, the combination of biomarkers and serum creatinine analyses increases the accuracy of patient risk classification. (30, 31) Future work should focus on integrating the novel analysis of urine output described in this study with other physiological variables measured in real-time together with biomarker and serum creatinine results to optimise the early detection of deranged renal physiology.

Most patients in this study received interventions such as fluid challenges aimed at normalising urine output. A total of 488 (20.4%) patients also received diuretics during their ICU admission. Data on the success of such interventions has not been investigated as part of this study but is likely to be of value as part of future work. The impact of treatment on urine output was a reason for selecting a 12-hour window when validating the model's predictions for severe oliguria. In some patients, a response to treatment triggered by observed low urine output was successful leading to fewer patients than predicted developing severe oliguria. In a subgroup of patients the response to intervention was transitory, causing the urine output to rise briefly above the 0.3ml/kg/hr threshold before falling back to a level consistent with severe oliguria. There were 109 incidences identified in which a patient suffered severe oliguria between 7 and 12 hours following predictions. Therefore, although the model was designed to predict six hours of urine output values it was considered clinically more relevant to assess the performance of the model based on the selected threshold over a longer period of time.

Limitations

The unbalanced nature of the data had the potential to make the AUC statistics seem overly impressive. Indeed, precision recall curve analyses showed that the excellent discrimination identified on ROC curve analyses of predictions made at 12 and 24 hours was influenced by the large proportion of patients who did not suffer severe oliguria and whom the model correctly identified as being at low risk of oliguria. However, this effect was less significant for predictions made after this time.

The development of this model benefited from being conducted in a group of patients undergoing cardiac surgery in one institution where the risk of complications is well known but the single centre design could limit transferability across other health care settings. The methodology developed will therefore need to be validated in different patient groups and in different institutions. With appropriate development, it could easily be applied to all intensive care unit patients. The ability of the model to improve patient outcomes through early recognition of impending severe oliguria should then be tested.

10.7. Conclusions

This dynamic Bayesian model, which analyses hourly urine output values, can be used to accurately predict the risk of severe oliguria occurring within the next 12 hours. Classification according to the model's predictions was shown to outperform the current method for screening patient urine output; the KDIGO AKI criteria. Crucially, the use of dynamic Bayesian modelling allows those at high-risk to be identified before they suffer a prolonged period of severe oliguria and in time to offer treatment. The model requires no additional information other than hourly urine output values and the patient's weight, can be easily run by computers routinely available at the bedside and provides an output that is easily interpreted by the clinical team. Before widespread adoption, the model requires validation in a range of critical care units and across the full range of critical care patients. The effect of implementation of the model in clinical practice on outcomes should then be assessed.

10.8. References

1.   Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practical Guideline for Acute Kidney Injury in *Kidney Inter* 2012. 1-138.
2.   Kellum JA, Sileanu FE, Murugan R, et al. Classifying AKI by Urine Output versus Serum Creatinine Level. *Journal of the American Society of Nephrology* 2015;26:2231-2238.
3.   Joannidis M, Metnitz B, Bauer P, et al. Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the SAPS 3 database. *Intensive Care Med* 2009;35(10):1692-1702.
4.   Vives M, Wijeysundera D, Marczin N, et al. Cardiac surgery-associated acute kidney injury. *Interactive cardiovascular and thoracic surgery* 2014;18(5):637-645.
5.   Hansen MK, Gammelager H, Jacobsen CJ, et al. Acute Kidney Injury and Long-term Risk of Cardiovascular Events After Cardiac Surgery: A Population-Based Cohort Study. *J Cardiothorac Vasc Anesth* 2015.
6.   Machado MN, Nakazone MA, Maia LN. Prognostic value of acute kidney injury after cardiac surgery according to kidney disease: improving global outcomes definition and staging (KDIGO) criteria. *Plos One* 2014;9(5):e98028.
7.   Dasta JF, Kane-Gill SL, Durtschi AJ, et al. Costs and outcomes of acute kidney injury (AKI) following cardiac surgery. *Nephrol Dial Transplant* 2008;23(6):1970-1974.
8.   Thakar CV, Christianson A, Freyberg R, et al. Incidence and outcomes of acute kidney injury in intensive care units: a Veterans Administration study. *Crit Care Med* 2009;37(9):2552-2558.
9.   Ralib A, Pickering JW, Shaw GM, et al. The urine output definition of acute kidney injury is too liberal. *Critical Care* 2013;17(3):R112.
10.   Howitt SH, Grant SW, Caiado C, et al. The KDIGO acute kidney injury guidelines for cardiac surgery patients in critical care: a validation study. *BMC Nephrol* 2018;19(1):149.
11.   Lagny MG, Jouret F, Koch JN, et al. Incidence and outcomes of acute kidney injury after cardiac surgery using either criteria of the RIFLE classification. *BMC Nephrol* 2015;16:76.
12.   Pole A, West M, Harrison J. Applied Bayesian forecasting and time series analysis. New York: Chapman and Hall; 1994.
13.   West M, Harrison PJ, Migon HS. Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association* 1985;80(389):73-83.
14.   Hickey GL, Grant SW, Caiado C, et al. Dynamic prediction modeling approaches for cardiac surgery. *Circ Cardiovasc Qual Outcomes* 2013;6(6):649-658.
15.   Smith AFM, West M. Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter. *Biometrics* 1983;39(4):867-878.
16.   R Core Team. R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing 2015.
17.   Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and ElaborationThe TRIPOD Statement: Explanation and Elaboration. *Annals of Internal Medicine* 2015;162(1):W1-W73.
18.   DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837-845.
19.   Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *Plos One* 2015;10(3):e0118432.
20.   Crowson CS, Atkinson EJ, Therneau TM. Assessing Calibration of Prognostic Risk Scores. *Statistical methods in medical research* 2016;25(4):1692-1706.
21.   Roques F, Michel P, Goldstone AR, et al. The logistic EuroSCORE. *Eur Heart J* 2003;24(9):881-882.
22.   Hickey GL, Grant SW, Murphy GJ, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg* 2013;43(6):1146-1152.

23.   Mehta RL, Kellum JA, Shah SV, et al. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Crit Care* 2007;11.

24.   Bellomo R, Ronco C, Kellum JA, et al. Acute renal failure–definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care* 2004;8.

25.   Machado MDN, Miranda RC, Takakura IT, et al. Acute kidney injury after on-pump coronary artery bypass graft surgery [Spanish;Portuguese;English] Lesion renal aguda post-revascularizacion del miocardio con circulacion extracorporea. *Arquivos Brasileiros de Cardiologia* 2009;93(3):230-235+242-247+247-252.

26.   Birnie K, Verheyden V, Pagano D, et al. Predictive models for kidney disease: improving global outcomes (KDIGO) defined acute kidney injury in UK cardiac surgery. *Critical Care* 2014;18(6):606.

27.   Fujii T, Uchino S, Takinami M, et al. Validation of the Kidney Disease Improving Global Outcomes Criteria for AKI and Comparison of Three Criteria in Hospitalized Patients. *Clinical Journal of the American Society of Nephrology* 2014.

28.   Cheng X, Wu B, Liu Y, et al. Incidence and diagnosis of Acute kidney injury in hospitalized adult patients: a retrospective observational study in a tertiary teaching Hospital in Southeast China. *BMC Nephrol* 2017;18(1):203.

29.   Banos V, Betbese A, Garcia-Alvarez M, et al. Urinary neutrophil gelatinase-associated lipocalin as an early predictor of acute kidney injury in cardiac surgery patients. *European journal of anaesthesiology* 2013;30.

30.   Kashani K, Al-Khafaji A, Ardiles T, et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. *Critical Care* 2013;17(1):R25-R25.

31.   Egal M, de Geus HR, Groeneveld AB. Neutrophil Gelatinase-Associated Lipocalin as a Diagnostic Marker for Acute Kidney Injury in Oliguric Critically Ill Patients: A Post-Hoc Analysis. *Nephron* 2016;134(2):81-88.
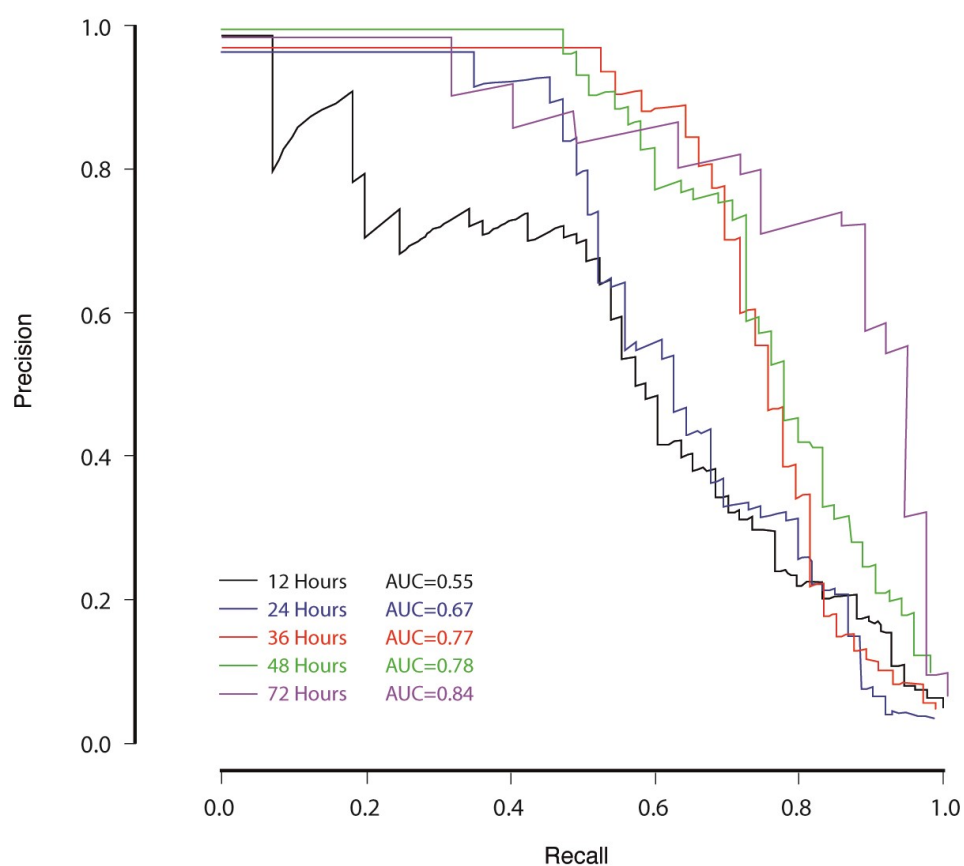
Figure 10-3 - Precision recall curves for the prediction of severe oliguria (<0.3ml/kg/hr for 6 hours) during the next 12 hours following each prediction made by the model at 12, 24, 36, 48 and 72 hours.

Table 10-6 - Performance of models when predicting severe oliguria occurring with the next 6 hours

Time point (number of patients still on CICU)	AUC (95% CI)	Observed severe oliguria within 6 hours	Predicted severe oliguria within 6 hours	O:E ratio
12 hours	0.98 (0.96-0.99)	21	90	0.23
24 hours	0.98 (0.97-0.99)	30	61	0.49
36 hours	0.99 (0.98-1.00)	34	49	0.69
48 hours	0.99 (0.98-1.00)	36	49	0.73
72 hours	0.99 (0.98-1.00)	38	31	0.92

CICU-cardiac intensive care unit, AUC – area under the curve, severe oliguria – urine output <0.3ml/kg/hr for six hours

Logistic regression models used to adjust for confounders during multivariable analyses

Table 10-7 - Logistic regression model for prediction of Renal replacement therapy

Variable	Beta coefficient	Odds ratio	95% CI for Odds ratio	P value
Intercept	-4.53			<0.001
Model high-risk classification	2.34	10.36	5.86-18.07	<0.001
Logistic EuroSCORE	0.04	1.03	1.01-1.06	<0.001
CPB time (minutes)	0.00	1.00	1.00-1.01	0.34

CI – confidence interval, CPB – cardiopulmonary bypass

Table 10-8 - Logistic regression model for prediction of prolonged length of stay

Variable	Beta coefficient	Odds ratio	95% CI for Odds ratio	P value
Intercept	-2.41			<0.001
Model high-risk classification	1.48	4.38	2.99-6.44	<0.001
Logistic EuroSCORE	0.07	1.08	1.06-1.09	<0.001
CPB (minutes)	0.01	1.00	1.00-1.01	<0.001

CI – confidence interval, CPB – cardiopulmonary bypass

Table 10-9 - Logistic regression model for prediction of hospital mortality

Variable	Beta coefficient	Odds ratio	95% CI for Odds ratio	P value
Intercept	-6.13			<0.001
Model high-risk classification	1.86	6.44	2.82-13.98	<0.001
Logistic EuroSCORE	0.03	1.03	1.00-1.05	0.06
CPB (minutes)	0.01	1.01	1.00-1.01	<0.001

CI – confidence interval, CPB – cardiopulmonary bypass

Model development

For the Bayesian analysis done throughout this paper we use a transformed second order polynomial dynamic linear model, with an unknown constant observational variance, V_t, and an unknown evolution variance , W_t. When forecasting, since many trends are exponentially decreasing and also to avoid negative values in the prediction intervals, we decide to use a log transform of the response (from investigation into the data it is noted that there is no difference between anuria and the very low urine outputs and hence we replace zero urine outputs by the minimum non-zero urine output) and the model is given by

$$\log(Y_t) = \mu_t + \nu,$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{1t,}$$
$$\beta_t = \beta_{t-1} + \omega_{2t.}$$

Where Y_t is the predicted urine output, μ_t is the level and β_t is the rate of change in level at time t. Here $\nu_t \sim N(0, V_t)$ (normally distributed with zero-mean and covariance matrix V_t) and $\omega_t \sim N(0, W_t)$ (normally distributed with zero-mean and covariance matrix W_t), where the unknown observational variance and an unknown evolution variance are modelled using information discounting where

$$V_t^{-1}|D_t \sim Ga(\delta^* \frac{n_t}{2}, \delta^* \frac{d_t}{2}) \quad W_t = \begin{pmatrix} \omega_{\alpha t} & \omega_{\mu \beta t} \\ \omega_{\mu \beta t} & \omega_{\beta t} \end{pmatrix}$$

and Ga represents the Gamma distribution.

The discount factors allow for "information loss" in the observation and system evolution equations respectively. In other words, urine output from 6 hours ago is less useful in modelling the patients next urine outputs compared to the most recent urine output. The discount factors (like all other parameters) were chosen to minimise the number of false negatives so that a low risk classification can be trusted.

The dynamic model that we described above is updated using algorithms defined by West et al (13) and the starting prior $(\theta_0|D_0) \sim N(m_0, C_0)$ which describes our initial beliefs about the system is, for a second order polynomial model, a bivariate normal distribution. We have

$$\theta_0 \,|\, D_0 \sim N\left[\begin{pmatrix} m_0 \\ C_0 \end{pmatrix}, \begin{pmatrix} C_{\mu 0} & 0 \\ 0 & C_{\beta 0} \end{pmatrix} \right],$$

where D_t refers to the information that we have available to us at time t. This initial prior represents our initial beliefs about the expected urine output for patients entering CICU and is required to initiate the system evolution recurrence relations. Using the notation that we have defined, the recursive procedure that is used to provide the forecasts is,

Updating Recurrence Relationships

$$\boldsymbol{\theta}_t | D_t \sim t_{n_t}(\boldsymbol{m}_t, \boldsymbol{C}_t),$$

$$\phi_t | D_t \sim Ga\left(\frac{n_t}{2}, \frac{d_t}{2}\right)$$

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t,$$

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' Q_t,$$

$$\boldsymbol{W}_t = (\delta^{-1} - 1) \boldsymbol{G} \boldsymbol{C}_t \boldsymbol{G}'$$

$$e_t = Y_t - f_t,$$

$$n_t = \delta^* n_{t-1} + 1$$

$$d_t = \delta^* d_{t-1} + S_{t-1} e_t^2 / Q_t$$

$$S_t = d_t / n_t$$

$$\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t / Q_t.$$

Updating Forecasting Distributions

$$\boldsymbol{\theta}_{t+k} | D_t \sim t_{\delta^* n_t}(\boldsymbol{a}_{t+k}, \boldsymbol{R}_{t+k}),$$

$$Y_{t+k} | D_t \sim t_{\delta^* n_t}(f_{t+k}, Q_{t+k}),$$

$$\boldsymbol{a}_{t+k} = \boldsymbol{G} \boldsymbol{a}_{t+k-1},$$

$$\boldsymbol{R}_{t+k} = \boldsymbol{G} \boldsymbol{R}_{t+k-1} \boldsymbol{G}' + \boldsymbol{W}_{t+k},$$

$$f_{t+k} = \boldsymbol{F}_{t+k}' \boldsymbol{a}_{t+k},$$

$$Q_{t+k} = \boldsymbol{F}_{t+k}' \boldsymbol{R}_{t+k} \boldsymbol{F}_{t+k} + k S_{t+k},$$

$$\boldsymbol{a}_t = \boldsymbol{m}_t,$$

$$\boldsymbol{R}_t = \boldsymbol{C}_t.$$

Chapter Eleven: Are serum potassium and magnesium concentrations associated with atrial fibrillation following cardiac surgery? (journal article)

Status: Final manuscript undergoing final co-author review

Submission Date: pending

Conference Presentation: Audio-visual presentation to the European Society for Cardiothoracic surgery. Milan, Italy October 2018

Authors: Dr Samuel H. Howitt MBChB, FRCA, Mr Stuart W. Grant PhD, MRCS , Dr Ignacio Malagon PhD, FRCA, FFICM, Professor Charles McCollum MD, FRCS

Author contributions, SH devised the study, collected and cleaned the data and performed the statistical analyses. SWG revised the first draft and guided the statistical analyses. IM and CM revised the first draft.

Rationale for inclusion of this study in the thesis

This research project aims to deliver a risk prediction model which analyses postoperative parameters to identify patients at increased risk of atrial fibrillation (AF) following cardiac surgery. The first stage of this process was to examine whether postoperative concentrations of potassium and magnesium were likely to be useful when predicting postoperative AF. To do this the postoperative potassium and magnesium concentrations for patients who did and did not develop AF were compared.

11.1. Additional data processing required for this manuscript.

Step 1. Remove patients who received cardiac transplantation or mechanical circulatory support

Step 2. Select data only from the first postoperative CICU admission

Step 3. Subclassify surgery into "valve surgery" and "not valve surgery"

Step 4. Exclude patients who were known to suffer from AF preoperatively.

Step 5. Limit dataset to events occurring within 72 hours of CICU admission

Step 6. For those who suffered AF limit the dataset to before the occurrence of AF.

Step 7. Determine the mean and minimum potassium concentrations for those who suffered AF (overall and for 12 hours before AF).

Step 8. Determine the mean and minimum potassium concentrations for those who did not suffer AF.

Step 9. Compare the rates of profound hypokalaemia (<3.5mmol/l) in those who did and did not suffer AF (Chi square test)

Step 10. Compare mean potassium concentrations for those who did and did not develop AF (Student's t-test).

Step 11. Merge in logistic EuroSCORE and CPB time data.

Step 12. Perform logistic regression analysis to test association between AF and [K+]<4.5mmol/L or controlling for surgery type, logistic EuroSCORE and CPB time.

Step 13. Determine difference between preoperative and mean [K+] in the 12 hours before onset of AF for those who suffered AF ([ΔK+]).

Step 14. Determine difference between preoperative and mean [K+] for those who did not suffer AF ([ΔK+]).

Step 15. Compare [ΔK+] for those who did and did not suffer AF (Student's t-test).

Step 16. Determine the mean and minimum magnesium concentrations for those who suffered AF (overall and for 24 hours before AF).

Step 17. Determine the mean and minimum magnesium concentrations for those who did not suffer AF.

Step 18. Compare the rates of profound hypomagnesaemia (<0.7mmol/l) in those who did and did not suffer AF (Chi square test)

Step 19. Compare mean magnesium concentrations for those who did and did not develop AF (Student's t-test).

Step 20. Perform logistic regression analysis to test association between AF and [Mg2+]<1.0mmol/L controlling for surgery type, logistic EuroSCORE and CPB time.

Step 21. Identify all incidences of electrolyte replacement therapy within the first 72 hours.

Step 22. Limit the dataset to the onset of AF for those who suffered AF.

Step 23. For those who did not suffer AF limit the dataset to the median time of AF onset in the AF group.

Step 24. Determine the number of electrolyte replacement doses given to patients.

Step 25. Compare the number of doses of electrolytes replacement therapy given to those who did and did not suffer AF (Wilcoxon rank sum test).

Step 26. Perform logistic regression analyses to test association between AF administration of electrolyte replacement therapy controlling for surgery type, logistic EuroSCORE and CPB time.

Step 27. Calculate median time from last potassium and magnesium measurement to onset of AF.

Step 28. Divide postoperative stay for those who did not suffer AF into 12 and 24 hour "blocks" for potassium and magnesium analyses respectively.

Step 29. Determine the mean potassium and magnesium concentrations for each postoperative "block."

Step 30. Determine whether the mean concentrations were above the classification thresholds (4.5mmol/L for potassium, 1.0mmol/L for magnesium) for the majority of the postoperative "blocks".

Step 31. Where the electrolyte classification for the majority of "chunks" is different to the classification of the overall mean concentrations, change to the classification to that of the majority of "blocks".

Step 32. Repeat analyses in Steps 12 and 20 using the revised classification of postoperative electrolyte concentrations.

11.2. Abstract

Introduction

Potassium and magnesium are frequently administered following cardiac surgery to reduce the risk of atrial fibrillation (AF); although the evidence for this practice is unclear. This study was designed to evaluate the relationship between serum potassium and magnesium levels and AF following cardiac surgery.

Methods

Prospectively collected data for all patients who underwent cardiac surgery between January 2013 and November 2017 was analysed. Cardiac rhythm was assessed using continuous ECG monitoring in 3068 patients on the cardiac intensive care unit and associations between serum potassium and magnesium concentrations and post-operative AF were assessed using univariable and multivariable analyses. The association between electrolyte replacement therapy and AF was also analysed.

Results

AF developed within 72 hours of cardiac surgery in 545(17.8%) of the 3068 patients. After adjusting for logistic EuroSCORE, operation type and cardiopulmonary bypass time, mean serum potassium concentration<4.5mmol/L was associated with an increased risk of AF (OR 1.4 (95%CI 1.3-2.7, $p<0.001$). Mean magnesium concentration <1.0mmol/L was not associated with an increased risk of AF (OR 0.82, 0.65-1.03, $p=0.09$) but the administration of magnesium increased the risk of developing AF (OR 1.55, 1.28-1.88, $p <0.01$).

Conclusions

Maintaining a serum potassium concentration>4.5mmol/L following cardiac surgery may reduce the incidence of postoperative AF. Low magnesium levels were not associated with postoperative AF but magnesium replacement conferred an increased risk of AF.

11.3. Introduction

Postoperative atrial fibrillation or flutter (AF) occurs in 18%-30% of patients undergoing cardiac surgery,(1-3) and is associated with prolonged stay in hospital,(4) increased healthcare costs(5) and increased short(6) and long term mortality risk.(7) The plasma concentrations of potassium and magnesium are thought to be important factors in the development of AF.(8) The administration of potassium replacement therapy to prevent postoperative AF is a common practice and is based on cardiac myocyte electrophysiology,(9) and supported by studies in the literature.(10, 11) Magnesium replacement is also frequently used in the prevention and treatment of AF. It is administered to both increase the response to potassium supplementation(12) and to lower the risk of AF directly.(8, 13)

Recently, Lancaster et al reported that higher concentrations of magnesium and potassium increased the risk of AF following cardiac surgery.(14) Hoekstra et al also failed to identify a benefit from potassium administration with the target of achieving higher postoperative serum potassium concentrations in this setting.(15) Most studies focused on predicting postoperative AF failed to included postoperative electrolyte concentrations or replacement therapy.(16-19) The objective of this study was to explore the role of serum potassium and magnesium concentrations and the administration of electrolyte replacement therapy in the development of AF following cardiac surgery.

11.3. Methods

Prospectively gathered data from 3068 patients admitted to the cardiac intensive care unit (CICU) following cardiac surgery between January 2013 and November 2017 were collected and analysed as part of a major study on complications following cardiac surgery funded by the British Heart Foundation. Patients undergoing transplantation or requiring mechanical circulatory support and those diagnosed with atrial fibrillation or atrial flutter preoperatively were excluded.   Any patient for whom no post-operative serum potassium and magnesium levels were available were also excluded.

Hourly cardiac rhythm assessments performed by the treating clinicians examining the continuous ECG traces recorded by the Draeger Infinity bedside monitors (Draeger) and all potassium concentrations measured by the Gem5000 (Instrumentation Laboratory) point of care blood gas analysers were extracted from the electronic patient record (EPR) for the first 72 hours on CICU following cardiac surgery. All administrations of intravenous potassium or magnesium replacement therapy were also extracted from the EPR along with the date and time of administration. Preoperative magnesium and potassium concentrations for each patient and

post-operative magnesium levels over the same time period were extracted from the hospital's pathology laboratory database together. All laboratory analyses were performed using the Architect C1600 (Abbott) analysers use for routine biochemical assays in our institution.

The primary question was whether potassium concentrations < 4.5mmol/L or magnesium concentrations <1.0mmol/L were associated with postoperative AF. The threshold value for potassium (4.5mmol/L) was chosen as this had been the threshold used in previous studies (10, 20) and it is also the target potassium concentration in our institution. The magnesium threshold of 1.0mmol/L was chosen based on our institution's treatment protocol as there is a general lack of consensus for a threshold value in the literature. For those who suffered AF, the mean potassium concentration in the twelve hours immediately prior to onset of AF was calculated. For those who did not suffer AF the overall mean potassium concentration was calculated. Similar magnesium concentration values were calculated but as magnesium concentrations were routinely measured daily, the mean magnesium concentration for those who developed AF was calculated over 24 rather than 12 hours. To ensure that the overall classification of the mean electrolyte concentrations was appropriate, a sensitivity analysis was conducted in which the patients' CICU stays were divided into portions of 12 hours for potassium concentrations and 24 hours for magnesium concentrations. The mean value of each electrolyte during each block was calculated and classified using the same thresholds (4.5mmol/L for potassium and 1.0mmol/L for magnesium). All patients whose overall classification was not the same as that observed in the majority of their discrete blocks were identified and reassigned according to the group in which their electrolyte concentrations fell during the majority of their blocks. Analyses were then repeated using the revised classifications.

Further analyses compared the distributions of the concentrations of these electrolytes in those who did and did not suffer AF. To test whether lower target thresholds might be useful the incidences of potassium concentrations <3.5mmol/L and magnesium concentrations <0.7mmol/L for those who suffered AF were compared with those for patients who did not. To test whether change relative to preoperative concentration was relevant we compared the difference between pre-and postoperative potassium concentrations (Δ [K^+]) in those who did and did not suffer AF. For those who developed AF, Δ [K^+] was defined as the difference between a patient's mean potassium concentration during the 12 hours before the onset of AF and the preoperative concentration. For those who did not suffer AF Δ [K^+] was defined as the difference between the patient's mean postoperative potassium concentration and the preoperative potassium concentration over 72 hours. Preoperative magnesium concentrations were not assessed as they were not routinely measured.

Finally, we compared the electrolyte replacement therapy administered before the onset of AF for those who developed AF with that administered to those who did not develop AF. To allow a fair comparison, for those who did not develop AF only doses administered before the median time of onset of AF in the AF group were counted. Outcome data were extracted from the hospital's clinical governance database. All data were pseudonymised prior to analysis. Data collection, cleaning and storage were performed according to the ethical and R&D approvals for the Vascular Governance NorthWest database.

11.3.1. Statistical Analyses

Univariable analyses of differences between proportions were performed using Chi square tests except for the comparison of the proportion of patients experiencing a magnesium concentration below 0.7mmol/L. This comparison was performed using Fisher's exact test due to the small number of outcomes. Electrolyte concentrations were compared using the Student's t-test as they were found to be normally distributed. Univariable comparisons of the number of doses of electrolyte replacement therapy administered were made using the Wilcoxon rank sum test as these variables were not normally distributed.

The associations between postoperative AF and potassium concentrations below 4.5mmol/L and magnesium concentrations <1.0mmo/L and were assessed using multivariable logistic regression analyses adjusting for type of surgery, cardiopulmonary bypass (CPB) time and preoperative co-morbidity (using the logistic EuroSCORE.(21) These analyses were repeated for the sensitivity analyses using the revised classifications as described above. Similar multivariable models were also used to assess whether electrolyte replacement therapy was associated with AF. All analyses were conducted using R (R foundation for statistical computing).(22)

11.4. Results

During the study period 3602 cardiac surgery patients were admitted to CICU. 228 were excluded as they underwent cardiac transplantation or required mechanical circulatory support and 306 were excluded as they were in AF before admission to CICU. A total of 3068 eligible patients (mean [sd] age was 66.1 [11.0] years) were identified of whom 19 were excluded from the potassium analyses due to a lack of postoperative potassium values within the specific time period (within the 12 hours preceding AF for those who suffered AF or during the CICU stay for those who did not). Similarly, 112 were excluded from the magnesium analyses due to an absence of postoperative magnesium concentrations within specified time period. Preoperative potassium values were missing from a further 45 patients so the ΔK^+ analyses included only 3004 patients.(Figure 11-1)
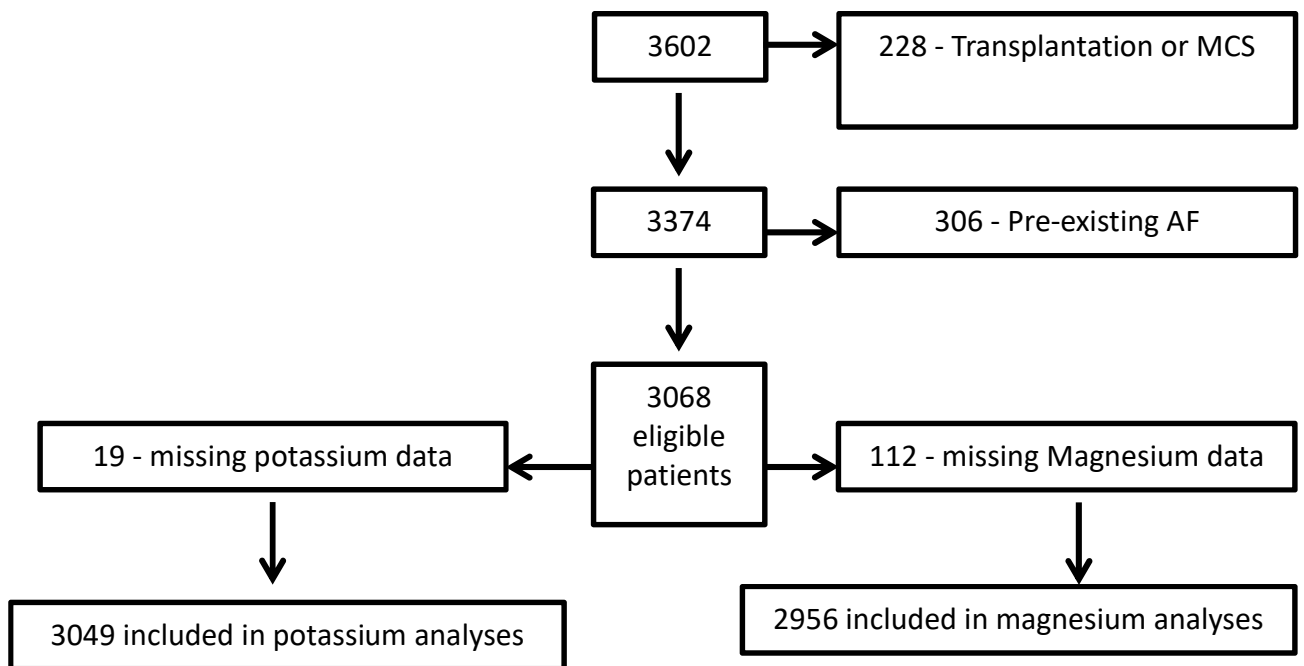
Figure 11-1 - Flow diagram showing selection of eligible patients

The most common procedure performed was isolated CABG which was performed for 1867 (60.8%) patients and the median length of CICU stay was 46.3 (25.2-69.2) hours. Detailed characteristics of the patients are shown in Table 11-1.

Table 11-1 - Patient characteristics

Characteristic	AF (n=545)	No AF (n=2523)
Age, mean (sd) , years	70.1 (9.9)	65.3 (11.0)
Female gender, %	152 (27.9)	680 (27.0)
Weight, mean (sd), Kg	82.5 (15.9)	81.9 (16.2)
Logistic EuroSCORE, median (Interquartile range)	5.3 (2.9-9.7)	3.1 (1.7-6.2)
Operation, n (%)		
CABG	268 (49.2)	1599 (63.4)
Valve	127 (23.3)	454 (18.0)
CABG and Valve	103 (18.7)	308 (12.2)
Aortic	43 (7.9)	121 (4.8)
Other – minor	5 (0.9)	41 (1.6)
Urgency, n (%)		
Elective	333 (61.1)	1399 (55.4)
Urgent	198 (36.3)	1080 (42.8)
Emergency	13 (2.4)	36 (1.4)
Salvage	1 (0.2)	8 (0.3)
CPB time, median (Interquartile range), minutes	107.0 (83.0-138.5)	100.0 (80.0-126.0)

AF – Atrial fibrillation/flutter, CABG – coronary artery bypass graft, CPB – cardiopulmonary bypass

A total of 545 patients (17.8%) developed AF within 72 hours of CICU admission. The median (IQR) time to onset of AF was 39.0 (29.2-51.0) hours. The median (IQR) time from last potassium concentration measurement to onset of AF was 2.0 (1.0-3.0) hours. The median (IQR) time from last recorded magnesium concentration to onset of AF was 12.0 (4.5-18.0) hours.

11.4.1. Primary analyses of electrolyte concentrations

As seen in Table 11-2, in the 12 hours preceding onset of AF, 274 patients (51.6%) experienced a mean potassium concentration <4.5mmol/L. Of those who did not suffer AF, 1057 (42.0%) experienced a mean potassium concentration <4.5mmol/L during the first 72 hours of their CICU stay (p<0.001).

Table 11-2 - Proportion of patients who did and did not suffer AF who experienced low electrolyte concentrations

Group	$[K^+]$ <4.5mmol/L N(%)	$[Mg^{2+}]$ <1.0mmol/l N(%)	P value
AF	274 (51.6)	145 (30.2)	<0.001
No AF	1057 (42.0)	973 (39.3)	<0.001

$[K^+]$ – potassium concentration; $[Mg^{2+}]$ – magnesium concentration

On multivariable analysis (detailed in the appendix), potassium concentration <4.5mmol/L (OR 1.41, 95% CI 1.15-1.72), increased logistic EuroSCORE (OR 1.04, 95% CI 1.03-1.05) and valve surgery (OR 1.46, 95% CI 1.17-1.83) were associated with the development of AF (p<0.001 for all). Of those who did not suffer AF but did experience an overall mean potassium concentration of ≥4.5mmol/L, 125 (8.6%) experienced 12 hourly mean potassium concentrations <4.5mmol/L for the majority of their twelve hour blocks. Of those who did not suffer AF and did experience an overall mean potassium concentration of <4.5mmol/L, 33 (3.1%) experienced 12 hourly mean potassium concentrations ≥4.5mmol/L for the majority of their twelve hour blocks. When logistic regression analyses were repeated using potassium concentrations classified according to the mean values of the majority of the twelve hour blocks, potassium concentration <4.5mmol/L (OR 1.23, 95% CI 1.0-1.50, p=0.04), increased logistic EuroSCORE (OR 1.04, 95% CI 1.03-1.05, p<0.001) and valve surgery (OR 1.46, 95% CI 1.17-1.83, p<0.001) were associated with the development of AF.

Only 145 (30.3%) patients who suffered AF experienced a magnesium concentration <1.0mmol/L in the 12 hours prior to onset of AF compared with 973 (39.3%) of those who did not suffer AF (p<0.001). After multivariable adjustment, magnesium concentration <1.0mmol/L (OR 0.82, 95% CI 0.65-1.03) was not found to be associated with an increase in the developing of AF (p=0.09). Details of the multivariable analyses are included in the Appendix.

Of those who did not suffer AF but experienced an overall mean magnesium concentration of ≥1.0mmol/L, 37 (2.5%) experienced 24 hourly mean magnesium concentrations <1.0mmol/L for the majority of their 24 hour blocks. Of those who did not suffer AF and experienced an overall mean magnesium concentration of <1.0mmol/L, 27 (2.8%) experienced 24 hourly mean magnesium concentrations ≥1.0mmol/L for the majority of their 24 hour block. When the logistic regression analyses were repeated using magnesium concentrations classified according to the mean values for the majority of the twelve hour blocks, magnesium concentration <1.0mmol/L was not associated with the development of AF (OR 0.81, 95% CI 0.64-1.01, p=0.06).

Full details of the sensitivity analyses are also displayed in Tables 11-6 and 11-7 of the appendix.

11.4.2. Secondary analyses of electrolyte concentrations

Of those who suffered AF, 31 (5.8%) experienced a potassium concentration below 3.5mmol/L prior to the onset of the arrhythmia compared with 149 (5.9%) of those who did not develop AF (p=1.0). A magnesium concentration below 0.7mmol/L was observed in one (0.2%) patient who went on to develop AF and 20 (0.9%) of those who did not (p=0.23).

The overall mean [sd] potassium concentrations recorded before the onset of AF (4.63mmol/L [0.30]) was for clinical purposes identical to the mean potassium concentration for those who did not develop AF (4.58mmol/L [0.26]). However, the small difference (0.05mmol/L) was statistically significant (p<0.001). Univariable comparisons between the mean potassium concentration recorded in the 12 hours before the onset of AF and potassium concentrations recorded in those who did not suffer AF are shown in Table 11-3. Minimum potassium concentrations for those who did not suffer AF occurred relatively early in the postoperative period; the median time from CICU admission to the minimum potassium concentration recorded was 17.8 hours (IQR 2.0-32.8).

Table 11-3 - Comparisons of electrolyte concentrations for those who did and did not develop AF

| AF group | | No AF group | | |
Derivative	Mean (sd) concentration mmol/L	Mean (sd) concentration mmol/L	Derivative	p value
Mean [K$^+$] in the 12 hours before onset of AF	4.50 (0.35)	4.58 (0.26)	Mean[K$^+$] in first 72 hours	<0.001
		3.96 (0.36)	Minimum[K$^+$] in first 72 hours	<0.001
Δ [K$^+$] (Mean [K$^+$] in the 12 hours before onset of AF – Preoperative [K$^+$])	0.21 (0.48)	0.26 (0.41)	Δ [K$^+$] (Mean [K$^+$] in first 72 hours - Preoperative [K$^+$])	0.05
Mean [Mg^{2+}] in the 24 hours before onset of AF	1.09 (0.26)	1.05 (0.21)	Mean [Mg^{2+}] in first 72 hours	<0.001
		0.95 (0.18)	Minimum [Mg^{2+}] in first 72 hours	<0.001

[K$^+$] – potassium concentration; [Mg^{2+}] – magnesium concentration

As shown in Table 11-3, those who suffered AF exhibited a smaller rise in mean potassium concentration relative to the preoperative value (Δ [K$^+$]) than those who did not. For both groups the mean postoperative potassium concentration was greater than the preoperative value.

The mean [sd] of all magnesium concentrations recorded before onset of AF (1.17mmol/L [0.27]) was 0.12mmol/L higher than the mean magnesium concentration for those who did not develop AF (1.05mmol/L [0.21], p<0.001). Univariable comparisons between the mean magnesium concentration recorded in the 24 hours before the onset of AF and magnesium concentrations recorded in those who did not suffer AF are also shown in Table 11-3. The minimum magnesium concentration recorded in those who did not suffer AF was observed slightly later than for potassium with a median time from ICU admission of 34.8 hours (IQR 14.8-40.4) but this was still before the median onset time of AF in the AF group.

11.4.3. Electrolyte replacement therapy

Potassium replacement therapy was administered to 2551 (83.2%) patients and magnesium to 1240 (40.4%) patients. The median (IQR) number of doses of potassium (20mmol IV) and

magnesium replacement therapy (20mmol IV) administered during the first 72 hours of ICU admission were 3 (1-5) and 0 (0-1) respectively.

As seen in Figure 11-2, patients who developed AF received a similar number of potassium doses to those who did not (median 2.0 vs 2.0, p=0.70). 139 (25.5%) of those who suffered AF received magnesium compared with 470 (18.6%) of those who did not (p<0.001). These findings were confirmed on multivariable regression analyses controlling for logistic EuroSCORE, CPB Time, and valve surgery. There was no association between potassium administration and AF (OR 1.01, 95% CI 0.97-1.05, p=0.56) but AF was more likely in those who received a higher number of doses of magnesium replacement therapy (OR 1.55, 95%CI 1.28-1.88, p<0.001). Details of the multivariable analyses are displayed in Tables 11-8 and 11-9 of the appendix.
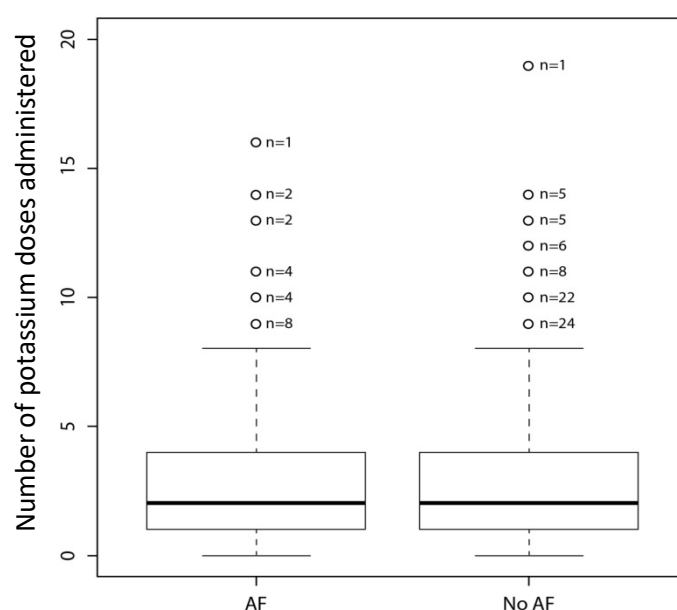
Figure 11-2 - Boxplot illustrating the administration of potassium replacement therapy to those who did and did not develop AF.

11.5. Discussion

The frequency of postoperative AF identified in this study is at the lower end of the ranges reported in the literature.(1-3, 18) This may be because the study was truncated at 72 hours in order to avoid comparing electrolyte concentrations from the immediate postoperative period, where risk of AF is highest, with concentrations measured much later in a prolonged CICU stay. It is also possible that our postoperative treatment protocols, including routine beta blocker prophylaxis and replacement of potassium to a target concentration of 4.5mmol/L may have favourably impacted the frequency of AF. This target resulted in an overall mean potassium concentration for all patients of >4.5mmol/L during the first postoperative 72 hours. Differences in potassium concentrations between groups who did and did not suffer AF may have been larger had the high target not been in place.

Investigating the potential impact of high-normal target electrolyte thresholds such as the potassium thresholds of 4.5mmol/L described by Auer(10) and Hoekstra(20) was the primary objective of this study. A potassium concentration below 4.5mmol/L was associated with an increase in the risk of AF, whereas the impact of a magnesium concentration below 1.0 mmol/L was the opposite of that expected. During their first 72 hours on CICU, the group who did not suffer AF experienced minimum potassium and magnesium concentrations which were, on average, lower than those recorded in the 12 hours before AF onset in those who suffered the arrhythmia. The median times at which the minimum potassium and magnesium concentrations were recorded were both earlier than the median AF onset time for the AF group. Despite suffering lower electrolyte concentrations at times when AF was most likely to develop, such patients remained in sinus rhythm.   It is therefore clear that electrolyte concentrations alone do not adequately explain the risk of AF and that models aiming to identify those at risk of AF are unlikely to rely heavily on the analyses of such electrolyte concentrations. However, electrolyte optimisation may still prevent AF developing in some patients. The benefits of maintaining a potassium concentration above 4.5mmol/L in particular were confirmed in this study. It is possible that Hoekstra's study failed to identify a benefit of maintaining higher potassium concentrations because the mean plasma concentration achieved in their higher target concentration group was only 4.33mmol/L and this was only 0.11mmol/L higher than the mean concentration achieved in their lower target concentration group (4.22mmol/L).(20)

Our secondary analyses demonstrated that use of low normal thresholds for potassium (3.5mmol/L) and magnesium (0.7mmol/L) concentrations would be unhelpful as few patients experience these low concentrations on our CICU.

The trend between greater magnesium replacement doses (together with higher magnesium concentrations) and increased risk of AF supports similar findings from Lancaster et al.(14) It may be that magnesium was supplemented more aggressively when clinicians suspected an increased risk of AF.   The prophylactic administration of magnesium to patients with normal magnesium concentrations should therefore be the subject of a prospective randomised controlled trial.

Existing models designed to predict AF following cardiac surgery are largely based on preoperative patient characteristics.(16-19) However, only postoperative risk factors can be modified once the patient has undergone surgery and as such these risk factors are more important in clinical care on the CICU. Where postoperative risk factors are analysed by existing risk scores, the postoperative variables included are mainly based on the administration of medications or therapies which have since been studied in more detail.(19) Consequently, some of the medications included as variables, such as beta blockers, are widely administered as routine care following cardiac surgery as part of treatment protocols.

Limitations

In order to maximise the number of potassium concentration measurements included in the study, we used point of care measurements from the CICU blood gas analysers. This achieved a median time interval between the latest potassium concentration and the onset of AF of only 2.0 hours.   Concerns have been raised previously about the accuracy of electrolyte concentration measurements made by point of care analysers at extremes of expected ranges. However, their accuracy has been demonstrated repeatedly to meet internationally recognised calibration targets.(23-25) As all post-operative potassium readings were measured using the same instrument, they can probably be relied upon to the extent required by this study design.

The retrospective nature of this study necessarily limited analyses to the data available from our EPR which only includes medications administered intravenously. As our cardiac surgery protocol includes the routine administration of beta blockade daily using a pre-printed prescription chart unless actively omitted by treating physicians it is unlikely that the prescription of beta blockers influenced our results. Although oral electrolyte replacement data was not available to us, intravenous replacement is almost universal following cardiac surgery in patients on our CICU. The retrospective design also precluded the strict standardisation of patient management. In particular, while our institution's protocol only aimed to keep magnesium concentrations >1.0mmol/L it is likely that unless a patient was experiencing hypermagnesaemia, additional magnesium would be administered to any patient whom clinicians identified as being at increased risk of developing AF.

The large number of patients in the study allowed the inclusion of preoperative logistic EuroSCORE, CPB time and surgery type into the logistic regression models used to adjust for potential confounders. Unfortunately, high quality data on other possible confounders such as cross-clamp time were not available.

The mean electrolyte concentrations recorded during a shortened period immediately before onset of arrhythmia were analysed for those who suffered AF as this was thought to best represent the levels of electrolytes present at that time. As no event regarding arrhythmia onset was available in the control group, comparisons were made with mean and minimum values recorded in those who did not suffer AF over the first 72 hours of their admission. This issue was also experienced by previous retrospectives studies such as that by Lancaster et al.(14) The sensitivity analyses in which we assigned the patients to groups based on the mean electrolyte concentrations for each 12 hour block of the admission aimed to address this issue. The similarity of the results observed in our primary analyses and the associated sensitivity analyses indicate that inappropriate classification of electrolyte concentrations is unlikely to have affected our findings.

11.6. Conclusion

This study confirms that maintaining potassium concentrations above 4.5mmol/L in the early postoperative period contributes to preventing AF following cardiac surgery. Replacement of magnesium was unexpectedly associated with increased risk of developing AF suggesting that magnesium replacement therapy in particular should be the subject of a prospective randomised controlled trial.

11.7. Appendix

Table 11-4 - Details of the multivariable logistic regression model showing impact of potassium concentration on risk of postoperative AF

Variable	Odds ratio (95% CI)	P value
Logistic EuroSCORE	1.04 (1.03-1.05)	<0.001
CPB Time (hours)	1.00 (1.00-1.00)	0.862
$[K^+]$<4.5mmol/L	1.41 (1.15-1.72)	<0.001
Valve surgery	1.46 (1.17-1.83)	<0.001

Table 11-5 - Details of the multivariable logistic regression model showing impact of magnesium concentration on risk of postoperative AF

Variable	Odds ratio (95% CI)	P value
Logistic EuroSCORE	1.04 (1.03-1.05)	<0.001
CPB Time (hours)	1.00 (1.00-1.00)	0.581
$[Mg^{2+}]$ < 1.0mmol/L	0.82 (0.65-1.03)	0.093
Valve surgery	1.43 (1.13-1.81)	0.003

Table 11-6 - Details of the multivariable logistic regression model showing impact of potassium concentration on risk of postoperative AF (sensitivity analysis)

Variable	Odds ratio (95% CI)	P value
Logistic EuroSCORE	1.04 (1.03-1.05)	<0.001
CPB Time (mins)	1.00 (1.00-1.00)	0.83
$[K^+]$ <4.5mmol/L	1.23 (1.00-1.50)	0.04
Valve surgery	1.46 (1.17-1.83)	<0.001

In the sensitivity analysis, for those who did not suffer AF, the mean potassium concentration classification (<4.5mmol/L or ≥4.5mmol/L) of the majority of the 12 hour blocks was entered into the model.

Table 11-7 - Details of the multivariable logistic regression model showing impact of magnesium concentration on risk of postoperative AF (sensitivity analysis)

Variable	Odds ratio (95% CI)	P value
Logistic EuroSCORE	1.04 (1.03-1.05)	<0.001
CPB Time (mins)	1.00 (1.00-1.00)	0.59
$[Mg2^+] < 1.0$ mmol/L	0.81 (0.64-1.01)	0.06
Valve surgery	1.43 (1.13-1.81)	0.002

In the sensitivity analysis, for those who did not suffer AF, the mean magnesium concentration classification (<1.0mmol/L or ≥1.0mmol/L) of the majority of the 24 hour blocks was entered into the model.

Table 11-8 - Details of the multivariable logistic regression model showing impact of potassium replacement therapy on risk of postoperative AF

Variable	Odds ratio (95% CI)	P value
Logistic EuroSCORE	1.04 (1.03-1.05)	<0.001
CPB Time (hours)	1.00 (1.00-1.00)	0.679
Potassium doses	1.01 (0.97-1.05)	0.564
Valve surgery	1.44 (1.15-1.80)	0.001

CPB – cardiopulmonary bypass

Table 11-9 - Details of the multivariable logistic regression model showing impact of magnesium replacement on risk of postoperative AF

Variable	Odds ratio (95% CI)	P value
Logistic EuroSCORE	1.04 (1.03-1.06)	<0.001
CPB Time (hours)	1.00 (1.00-1.00)	0.738
Magnesium doses	1.55 (1.28-1.88)	<0.001
Valve surgery	1.47 (1.18-1.84)	0.001

CPB – cardiopulmonary bypass

11.8. References

1.   LaPar DJ, Speir AM, Crosby IK, et al. Postoperative atrial fibrillation significantly increases mortality, hospital readmission, and hospital costs. *Annals of Thoracic Surgery* 2014;98(2):527-533; discussion 533.
2.   Tran DTT, Perry JJ, Dupuis J-Y, et al. Predicting New-Onset Postoperative Atrial Fibrillation in Cardiac Surgery Patients. *Journal of cardiothoracic and vascular anesthesia* 2015;29(5):1117-1126.
3.   Hosokawa K, Nakajima Y, Umenai T, et al. Predictors of atrial fibrillation after off-pump coronary artery bypass graft surgery. *British journal of anaesthesia* 2007;98(5):575-580.
4.   Todorov H, Janssen I, Honndorf S, et al. Clinical significance and risk factors for new onset and recurring atrial fibrillation following cardiac surgery - a retrospective data analysis. *BMC Anesthesiol* 2017;17(1):163.
5.   Vaporciyan AA, Correa AM, Rice DC, et al. Risk factors associated with atrial fibrillation after noncardiac thoracic surgery: Analysis of 2588 patients. *Journal of Thoracic and Cardiovascular Surgery* 2004;127(3):779-786.
6.   Attaran S, Shaw M, Bond L, et al. Atrial fibrillation postcardiac surgery: a common but a morbid complication☆ ☆ ☆. *Interactive cardiovascular and thoracic surgery* 2011;12(5):772-777.
7.   Tulla H, Hippelainen M, Turpeinen A, et al. New-onset atrial fibrillation at discharge in patients after coronary artery bypass surgery: short- and long-term morbidity and mortality. *Eur J Cardiothorac Surg* 2015.
8.   Raiten JM, Ghadimi K, Augoustides JGT, et al. Atrial Fibrillation After Cardiac Surgery: Clinical Update on Mechanisms and Prophylactic Strategies. *Journal of cardiothoracic and vascular anesthesia* 2015;29(3):806-816.
9.   Bharucha DB, Kowey PR. Management and prevention of atrial fibrillation after cardiovascular surgery. *Am J Cardiol* 2000;85(10A):20D-24D.
10.   Auer J, Weber T, Berent R, et al. Serum potassium level and risk of postoperative atrial fibrillation in patients undergoing cardiac surgery. *J Am Coll Cardiol* 2004;44(4):938-939; author reply 939.
11.   Wahr JA, Parks R, Boisvert D, et al. Preoperative serum potassium levels and perioperative outcomes in cardiac surgery patients. Multicenter Study of Perioperative Ischemia Research Group. *Jama* 1999;281(23):2203-2210.
12.   Whang R, Whang DD, Ryan MP. Refractory potassium repletion. A consequence of magnesium deficiency. *Arch Intern Med* 1992;152(1):40-45.
13.   Miller S, Crystal E, Garfinkle M, et al. Effects of magnesium on atrial fibrillation after cardiac surgery: a meta-analysis. *Heart* 2005;91(5):618-623.
14.   Lancaster TS, Schill MR, Greenberg JW, et al. Potassium and Magnesium Supplementation Do Not Protect Against Atrial Fibrillation After Cardiac Operation: A Time-Matched Analysis. *Ann Thorac Surg* 2016;102(4):1181-1188.
15.   Hoekstra M, Hessels L, Yeh L, et al. Computer guided low-normal versus high-normal potassium control and its effect on atrial fibrillation in 1225 thoracic ICU patients-the grip-compass trial. *Intensive care medicine* 2013;39.
16.   Cameron MJ, Tran DTT, Abboud J, et al. Prospective External Validation of Three Preoperative Risk Scores for Prediction of New Onset Atrial Fibrillation After Cardiac Surgery. *Anesthesia & Analgesia* 2018;126(1):33-38.
17.   Mariscalco G, Biancari F, Zanobini M, et al. Bedside tool for predicting the risk of postoperative atrial fibrillation after cardiac surgery: the POAF score. *J Am Heart Assoc* 2014;3(2):e000752.
18.   Ferreira AF, F AS, Moreira R, et al. Postoperative Atrial Fibrillation After Coronary Artery Bypass Grafting Surgery. *Rev Port Cir Cardiotorac Vasc* 2017;24(3-4):129.
19.   Mathew JP, Fontes ML, Tudor IC, et al. A multicenter risk index for atrial fibrillation after cardiac surgery. *Jama* 2004;291(14):1720-1729.

20.   Hoekstra M, Hessels L, Rienstra M, et al. Computer-guided normal-low versus normal-high potassium control after cardiac surgery: No impact on atrial fibrillation or atrial flutter. *Am Heart J* 2016;172:45-52.

21.   Roques F, Michel P, Goldstone AR, et al. The logistic EuroSCORE. *Eur Heart J* 2003;24(9):881-882.

22.   R Core Team. R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing 2015.

23.   Jain A, Subhan I, Joshi M. Comparison of the point-of-care blood gas analyzer versus the laboratory auto-analyzer for the measurement of electrolytes. *International Journal of Emergency Medicine* 2009;2(2):117-120.

24.   Allardet-Servent J, Lebsir M, Dubroca C, et al. Point-of-Care Versus Central Laboratory Measurements of Hemoglobin, Hematocrit, Glucose, Bicarbonate and Electrolytes: A Prospective Observational Study in Critically Ill Patients. *Plos One* 2017;12(1):e0169593.

25.   Jose RJ, Preller J. Near-patient testing of potassium levels using arterial blood gas analysers: can we trust these results? *Emerg Med J* 2008;25(8):510-513.

SECTION FOUR: DISCUSSION

Chapter Twelve: General Discussion

This thesis has been presented in the journal format and so the findings of each chapter have been outlined within the manuscripts included in the results section. The research programme has answered the research questions posed at the start of this thesis and those answers will be summarised in this section. A narrative description of problems encountered during the data collection phase of the project will be discussed as these problems affected the progression of the project and the studies performed. The key strengths and limitations of the methodology used within the programme of research overall and within each results chapter will also be discussed. The relevance of this research programme's findings to the postoperative care of patients who have undergone cardiac surgery will be explored. Finally, taking into account the findings of this thesis, as well as insights gained during the collection and analysis of the dataset, recommendations for future work and the potential impact of this research programme within cardiac surgery and critical care as a whole will be summarised.

12.1. Key findings

The first stated aim of this research project was to create a dataset suitable for the validation of existing risk stratification systems and the development of novel risk prediction models. Although this task was more difficult than anticipated, the complete dataset has been collated and cleaned and important outcomes have been identified and labelled. The dataset contains millions of data points and is cleaner and better annotated than any dataset present in publicly available biobanks such as PhysioNet. This high quality dataset allowed the analyses described within this thesis to be conducted and will be used to develop more risk prediction models following the submission of this thesis.

Chapter seven detailed the validation of the Rapid Clinical Evaluation (RACE) and logistic Cardiac Surgery Risk Score (logCASUS) alongside the SOFA score. This was the first validation of the RACE and logCASUS scores, and the first validation of the SOFA score in cardiac surgery patients in the UK. All models discriminated well between those who would go on to survive to CICU discharge and those who would not. Consequently, all three models could be used to identify patients at the greatest risk on the critical care unit and guide the allocation of clinical resources accordingly. For example, those with the highest levels of risk could be seen first on ward rounds or their beds could be moved closer to the main doctors' station or sisters' office. The daily calculation of the

scores was also validated, confirming the potential for such scores to be used on a daily basis to provide an updated assessment of mortality risk for patients on the CICU.

The RACE and logCASUS score are logistic scores i.e. the models' outputs are estimates of risk for groups of patients. The calibration of these scores was found to be poor in the dataset used for this research programme. Reasons for the poor calibration are discussed in detail in chapter seven but are likely to include:

i)   differences in management between institutions used to create and validate the models,

ii)   improvements in treatments related to technological advances and the incorporation of the findings of studies shown to improve outcomes into routine care. This phenomenon is known as calibration drift and has previously been reported in preoperative models used in cardiac surgery.(173)

Recalibration of the models was performed using a training subset and this improved their calibration when tested in a separate testing cohort. However, calibration was still suboptimal reflecting the major limitation when attempting to predict mortality; mortality is a rare complication which has become increasingly infrequent over time. Models which predict mortality therefore require calibration and validation in vast datasets before the accuracy of their predictions can be relied upon.

Chapter eight quantified the incidence of sepsis in patients who had undergone cardiac surgery and described the first validation of the Sepsis-3 criteria in this setting. The incidence of sepsis following cardiac surgery identified using the new definition was higher than previously described when older definitions were used to define sepsis. Importantly, physiological derangement in the presence of either "suspected" or "proven" infection qualified as sepsis and was associated with poor outcomes. This is highly relevant because older definitions had required microbiologically proven infection in order for the diagnosis of sepsis to be made. Around half of suspected infections identified in this study were never microbiologically proven. Older definitions would have effectively excluded a large number of patients who were identified using the new definition. The fact that sepsis with suspected infection carried increased risk of poor outcomes supported the decision made by the Sepsis-3 team to include suspected infection within the guideline. The study also answered questions about the applicability of the Sepsis-3 criteria to cardiac surgery. As the Sepsis-3 criteria rely on the identification of physiological derangement using the SOFA score it was conceivable that the physiological derangement due to cardiac surgery and cardiopulmonary bypass could mask physiological deterioration due to sepsis. The study demonstrated even after the substantial, non-infective physiological insult of cardiac

surgery, the Sepsis-3 criteria identified those with the worst outcomes. Consequently, the use of the Sepsis-3 criteria for cardiac surgery patients can be recommended.

Chapter nine investigated the outcomes of patients who fulfilled different criteria for various stages of acute kidney injury as defined in the KDIGO guidelines.(87) The study showed that patients who were diagnosed with the same stage of AKI by different criteria experienced different rates of adverse outcomes. AKI diagnosed due to low urine output alone carried less risk than AKI diagnosed by creatinine criteria. Patients meeting both urine output and creatinine criteria were at the greatest risk of adverse outcomes. Moreover, patients diagnosed with the supposedly less severe AKI-1 by both urine output and creatinine actually experienced worse outcomes that those diagnosed with the supposedly more severe AKI-2 by urine output alone. These findings are important because when the guidelines were written the authors recognised that the thresholds used within the urine output classification were arbitrarily defined. Indeed they stated that "the influence of urinary output criteria on AKI staging needs to be further investigated." Alongside similar studies in both cardiac surgery patients (136) and the general ICU population (174, 175), the findings of this chapter provide evidence that recalibration of the urine output thresholds contained within the KDIGO AKI guidelines is required. The findings also informed the decision to design the alternative method for the stratification of risk according to urine output following cardiac surgery described in chapter ten of this thesis.

The novel model described in chapter ten identifies patients at greatest risk of severe oliguria (urine output <0.5ml/kg/hr for 6 hours) based on analyses of their own previous urine output. The discrimination of the model was excellent throughout the first 72 hours although when calibration was tested, the model overestimated risk in the first 24 hours. The unpredicted recovery in urine output observed within the early postoperative period may be explained by positive response to treatments and natural resolution of the stress response to surgery. To test the associations between the risk of severe oliguria and adverse outcomes, the predicted risk levels were categorised as high or low risk. Patients at high risk of severe oliguria were found to have the greatest risk of adverse outcomes. Outcomes for those who were identified as high risk by the model were worse than outcomes for those who fulfilled KDIGO urine output criteria for AKI but remained classified as low risk by the model. The model's predictions were so accurate that high predicted risk of severe oliguria was found to identify those who would go on to require RRT equally as well as classification according to observed severe oliguria. The fact that increased risk can be identified before the severe oliguria occurs is clinically important as it provides additional time for intervention to deliver preventative treatments.

Chapter eleven was designed to inform the development of risk prediction models to predict the onset of atrial fibrillation following cardiac surgery. Serum electrolyte concentrations have been associated with the onset of arrhythmias including atrial fibrillation.(176-178) This study found that patients who developed AF were more likely to have experienced a mean potassium concentration <4.5mmol/L than those who did not. The study concluded that maintaining potassium concentrations >4.5mmol/L may reduce the risk of developing AF. However, most patients who did not suffer AF experienced potassium concentrations below the mean potassium concentration recorded in the AF group prior to the onset of the arrhythmia. This is important because it suggests that monitoring postoperative potassium concentrations to predict postoperative atrial fibrillation is unlikely to be successful. This finding will inform future work on the development of models for this purpose. Magnesium replacement therapy was associated with the development of AF. This may be due to the administration of magnesium as prophylaxis in patients who a clinician deemed to be at high risk of AF (e.g. those experiencing multiple premature atrial ectopic beats) and this finding should be explored in a randomised controlled trial.

12.2. Narrative review of problems encountered during the data collection phase of this thesis

12.2.1. Delays related to information technology (IT) infrastructure

The extraction of data from the clinical governance database, the perfusion database and the hospital pathology databases was relatively was straightforward. As described in chapter 5, manual checks were required to ensure the completeness of data from these sources. However, the vast majority of data was obtained reliably and quickly with no need for particular upgrades to any IT infrastructure.

Conversely, data extraction from the Draeger Innovian EPR and bedside monitors was much more complicated and resulted in delays to the project. After gaining ethical approval for the research programme, a substantial amount of funding was spent on the acquisition and installation of a dedicated Draeger Infinity Gateway report server. This server was acquired to safeguard the clinical data on the live clinical server. If searches and data extraction exercises were performed on the live server, the increased workload could have impaired function of the EPR on the ICU. There was also a small potential risk of corrupting the live data while accessing it to generate reports. Once the Gateway server was installed, data from the live clinical server were transferred to the report server. Data searches and extraction could then be performed on the

report server without any effect on the live clinical server. Unfortunately the setup of the server ran into problems.

The transfer of the EPR data to the report server was performed relatively quickly. This allowed the extraction of data from the EPR to be performed relatively early in the research project. Data collected from the EPR were therefore cleaned and ready for analysis early enough to allow the work presented in chapters seven to ten to be performed.

Unfortunately the recording of waveform data from the Draeger infinity bedside monitors was much more complicated. Prior to registering for the thesis, enquiries were made to Draeger and it was understood that the waveform data would be exported from the monitors via HL7 protocol and saved onto the report server. When the Gateway server was delivered (four months after registration on the PhD programme) it was discovered that the HL7 export only supported a frequency of 6Hz. Clearly, no waveform could be reproduced using data of such low resolution. Consequently, a whole new approach to access the data was devised.   The solution required access to an extra piece of Draeger software, the Win API, which was kindly installed free of charge after a delay of a further six months. During this period other compatibility issues were identified and overcome during the installation of the Gateway report server. Fortunately no EPR data were lost due to these delays. The data were all still recorded onto the live clinical server. Once the Gateway server was available all data were transferred over and accessed retrospectively.

The Win API software was to be used to interrogate the report server and record the data passing through the report sever as shown in Figure 4-8. Unfortunately, the software was unable to perform as intended in its original format. Fortunately, software engineers from our research collaborator in the BHF funded programme, Rinicare Ltd, were able to re-engineer the software to make it fit for our purpose. This process took an additional three months and identified a final problem with the hospital's IT infrastructure. Some of the switches which were transferring data from the live server to the report server did not have enough bandwidth to transmit all of the data contained within the waveform files.   Identifying the source of the problem and upgrading the switches took a further three months. In total, collection of the waveform data was delayed by 14 months. Unfortunately as shown in figure 4-8, until the API was functioning there was no mechanism to prevent waveform data being deleted after it had been displayed on the bedside monitor.

12.2.2. The WannaCry malware attack in May 2017

In May 2017 the WannaCry malware attack infected a number of NHS computers throughout the UK.(179) In response to the potential risk of harm posed to the hospital trust all computers and servers running Windows XP were shut down for a prolonged period. Unfortunately, this included the bedside monitors on the CICU. Consequently all data recording was stopped for a 10 day period. This unfortunate event enforced the division of data into two subsets; that collected before the shutdown and that collected afterwards. This unexpected event resulted in the loss of data from around 30 patients who were admitted to the CICU during the period when the electronic records were shut down. However the enforced division of the dataset allowed the early subset of live data to be analysed while collection of data continued in the second subset. As the early subset was available earlier than expected, the processing of the ECG output files began ahead of time. This was useful because the removal of noise and identification of key components within the waveform was a complicated process. When the final dataset (containing both the early and late cohorts) was produced, techniques developed on the early dataset could be transferred easily to the complete dataset.

12.3. Strengths and limitations of this thesis

This subsection discusses the key strengths and weaknesses of the methodology used in this research programme to provide context for its findings. Firstly, considerations pertaining to the research programme as a whole are discussed. Subsequently, strengths and limitations of analyses performed within specific chapters are discussed.

12.3.1. Data quality

The methodology of data capture for the studies presented in this thesis is a key strength of the project. As outlined in the introduction, cardiac surgery patients are an ideal group in whom to conduct risk prediction analyses. The patients have broadly similar clinical backgrounds; they all have cardiovascular disease but are fit enough to survive major surgery. Patients all undergo one (or occasionally more than one) procedure performed by one of a small number of specialist surgeons. All patients treated in this institution are managed on the same critical care unit and are at risk of a well described range of complications. This homogeneity increases the chance of identifying associations within the dataset.(18) Moreover, physiological monitoring data are reproducibly collected during and following cardiac surgery using dedicated monitoring devices.(47) The overall quality of the data analysed for this thesis was therefore high. This was not because the data were recorded specifically for research but because the data were "real world" data that were clinically important. The clinical necessity of high quality data resulted in the high quality of the collected dataset. Nevertheless, data quality was not perfect and the reasons for and consequences of lapses in the quality of data are discussed below.

Erroneous data

"Real world" data contain erroneous information despite the best efforts of those involved. The vast majority of data points analysed in this research program were entered into the Draeger Innovian EPR. Therefore, the EPR-derived data contained the most errors. The reasons for the presence of erroneous data within a database are often related to the method by which data entry occurs. Many of the parameters contained within the dataset were entered manually into the EPR or the clinical governance database by clinicians or administrators. These variables included measured physiological parameters such as urine output and temperature as well as oxygen, medication and fluid administration. Wherever manual data entry occurs there is potential for introduction of errors.(180) In the database used in this thesis, implausible manually entered values were identified using automated algorithms during data cleaning. Errors encountered included "wrong field" errors where a parameter was recorded in the box related to

a different parameter. The other common type of error was the typographical error in which the entry in the correct box contained a spelling mistake or inappropriate punctuation.

Where "wrong field" errors were identified, erroneously entered data were removed so that erroneous data were effectively ignored. Data that had been entered into an inappropriate field were not reassigned to another field as there was an element of uncertainty as to field in which the values were supposed to be recorded. Consequently, "wrong field" errors resulted in the loss of data entered into the wrong field.

Typographical errors were corrected wherever possible using reproducible algorithms. For example, a temperature of "36..7" °C or a ventilatory mode of "BiPAPABS" was easily corrected to "36.7" °C or "BiPAP/ASB" respectively. However, clearly implausible entries which were likely to have been entered by mistake but which resulted in ambiguous entries e.g. a GCS of "18" were removed.

Many parameters were recorded automatically from patient monitors rather than requiring manual input. This approach minimises the risk of "wrong field" and typographical input errors. However, the use of automated entry introduces the risk of recording values which have been measured when equipment was not recording appropriately. For example, where artefacts were present on the ECG the monitor's algorithms may have calculated a heart rate inaccurately. Similarly, if medication was being administered through a central venous catheter lumen which is also being used to transduce central venous pressure, inaccurately elevated central venous pressures may have been recorded.   To mitigate the risks associated with automated data collection clinicians are required to manually verify values entered automatically into the EPR. Despite this failsafe, clearly erroneously recorded values were encountered. In particular, for central venous pressure, a large number of implausibly high values were encountered. Where physiologically improbable values were recorded a "flag" was attached to the reading. During analyses flagged readings were excluded.

While the treatment of erroneous data during the cleaning process resulted in data loss, as the cleaning codes developed were semi-automated, they were reproducible and are unlikely to have introduced bias into the analyses. Erroneous readings which remained within the physiologically appropriate range could not be identified and will have remained within the data analysed.

Missing data

Missing data were also identified as an issue across the majority of data sources used to construct the final dataset. It was apparent that some episodes of re-operation had been omitted from the clinical governance database. This may be because these procedures tended to be performed as

emergencies and many such procedures were performed out of hours. As reoperation is an important event as well as a risk factor for other adverse outcomes a robust cleaning code was created to identify any missed trips to theatre. Once identified, missed trips were added manually to the final database.

Some blood test results were also missed during the automated data capture. This usually occurred because the patient's location (which was used to extract all bloods results for patients on CICU) was entered incorrectly. To remedy these omissions, cleaning code was created to identify every day for which a patient's results were not present. A list of potentially missing data was then used to guide manual interrogation of the clinical records and wherever a previously unidentified value was identified this was entered into the dataset.

Due to the complex nature of the surgery performed and the critical care required postoperatively, physiological values are usually recorded hourly for all patients receiving ICU care following cardiac surgery. Occasionally, where patients were fit for ward care but were not discharged for logistical reasons the frequency of observations was reduced to four hourly. Urine output was particularly well recorded as almost all patients were catheterised at the time of surgery making hourly urine output measurement straightforward. There is a well-established practice of leaving the value blank if no recording was made and only entering "0" where no urine was produced rather than when the output was simply not measured. The analyses performed within this thesis were therefore performed using high quality and frequently recorded data points. Occasionally, missing data did cause a problem. For example, central venous pressure values were unavailable after the central venous catheter had been removed. However, missing data were always handled in a reproducible way as described in the methodology of each results chapter. It is therefore unlikely that missing data had a significant effect on the results obtained.

12.3.2. Study location

All analyses were performed on data that were collected from one site. Patients were managed according to local protocols and even where management was individualised, decisions were made by a small number of clinicians. Consequently, differences in postoperative treatment methods were a relatively small. This is important because differences in postoperative management represent a potential confounding variable within this dataset. Patients were monitored according to standardised protocols so there was also little variation in the availability of data for different patients.

Importantly, treatments and outcomes observed in this single centre dataset may not be representative of care delivered across different sites. As a result, any associations between

independent variables and outcomes identified during this thesis may not be replicated in other centres. This potential problem is exacerbated by the case-mix encountered at Wythenshawe Hospital. As a quaternary referral centre, the institution receives patients from all over the North West of England and Wales. The unit performs cardiac transplantation and provides mechanical circulatory support which are specialist treatments which may not be available elsewhere. The patients at Wythenshawe hospital may therefore be more unwell than the general cardiac surgery population across the UK as a whole and may receive treatments that are not available elsewhere. The cardiac surgery specific scores validated in chapter seven were designed using datasets which included patients who had undergone cardiac transplantation and those receiving mechanical circulatory support. Therefore, such patients were included in the validation study in chapter seven. To increase transferability of findings across UK cardiac surgery centres, patients who underwent transplantation surgery or received mechanical circulatory support were excluded from analyses in chapter eight, ten and eleven. Patients who underwent transplantation and mechanical circulatory support were included in the study comparing outcomes for patients who developed AKI in chapter nine. However, the prolonged length of stay outcome was categorised taking into account whether transplantation was performed or mechanical circulatory support was initiated.

Finally, all patients included in this study had undergone cardiac surgery. While such patients have much in common with other patient groups on the critical care unit it cannot be assumed that the findings of this thesis will be confirmed in studies involving different patient groups.

12.3.3. Sample size
Despite the datasets used for each study containing over 2200 patients the numbers of adverse outcomes, particularly mortality, were low. The low number of outcomes affected the analyses used in each of the results chapters. The main effect of the low number of outcomes was the limitation of the number of confounding variables for which adjustment could be made. It is widely accepted rule that there should be at least five and ideally ten outcomes in the dataset for each predictor variable included in a logistic regression model.(181, 182) Consequently, logistic regression analyses adjusting for the effect of confounders on mortality rates in this thesis were limited to the inclusion of one or two confounders. To mitigate this limitation, the logistic EuroSCORE(5) was used widely throughout this thesis to control for overall pre-operative morbidity and surgical complexity. The Logistic EuroSCORE has been extensively validated.(173, 183-185) While calibration of the score has decreased overtime, resulting in over-prediction of risk, the discriminative ability of the model remains good. Risk adjustment using the logistic EuroSCORE is therefore statistically appropriate. Effects of sample size on the specific studies

reported in specific chapters are discussed in the relevant results chapters. Any attempt to increase in the size of the dataset would require either a multicentre approach or a prolonged study duration. The consequences of these strategies are also discussed below in the section focused on future work.

12.4. Specific considerations for particular studies

The three models validated in chapter seven predict mortality following cardiac surgery. Mortality following cardiac surgery is low and has been falling in recent years. Consequently, despite including over 2000 patients, only 41 deaths were observed in the dataset analysed. This limited the quality of the recalibration that could be performed as reliable recalibration is dependent on the dataset containing an adequate number of outcomes.

In chapter eight, the use of the Sepsis-3 criteria were modified slightly. In the original guideline it was stated that the baseline SOFA score should be assumed to be zero and changes should be measured relative to this assumed baseline. The study showed that the more than 90% of cardiac surgery patients had SOFA score >2 on the first postoperative day. Therefore, it was decided to use the day one score as a baseline for each patient. This meant that a change in SOFA score could only be assessed from day two onwards. This precludes the identification of sepsis before the second postoperative day. While it would be unusual for a patient, particularly a patient undergoing elective surgery, to develop sepsis on the first postoperative day, this remains a limitation of the study.

The design of the new risk prediction model for the analysis of urine output described in chapter ten was subject to a particular limitation. It was not possible to adjust risk estimates for severe oliguria taking into account the administration of fluids or medication or the use of renal replacement therapy. The model has been shown to work well for cardiac surgery patients on our critical care unit. However, interventions may be performed slightly differently in our institution than at other centres and results may therefore not replicate across other centres where different approaches to interventions to correct low urine output are made. The calibration of the model in the cardiac surgery cohort was best after the initial 24 hours. The calibration in the first 24 hours may have been influenced by the success of clinical interventions aimed at improving urine output in this early postoperative period. The overestimation of risk in the early stages may also have been due to recovery from the major but temporary physiological insults of cardiac surgery and cardiopulmonary bypass. For patients on the general ICU who have not experienced the same surgical insult, the models predictions may be more reliable in the early stages of their ICU stay. Performance of the model in such populations should be investigated.

In chapter 11 it was demonstrated that the association between serum electrolyte concentrations and the onset of atrial fibrillation after cardiac surgery was not straightforward. The analyses in this chapter were based on the classification of patients into groups who did and did not develop atrial fibrillation during the postoperative period. This classification was made based upon the heart rhythm recorded in the electronic patient record by doctors and nurses treating the patient. While there is no reason to believe the classifications to be inaccurate it was not possible to verify that heart rhythm diagnoses were correct.

12.5. Recommendations for future research

12.5.1. Overall plan for research programme

As stated previously the work described within this thesis lays the foundation for future research using the dataset that has been created. The continuation of the BHF-funded project will ensure that the dataset continues to be expanded using protocols designed as part of this thesis. The project's collaborators at Durham University will continue to develop risk prediction models guided by the author of this thesis and the supervisory team. Models developed using the dataset created for this thesis will require validation in other cardiac surgery cohorts. Transferability across different critical care populations should then also be confirmed in further validation studies. Finally, clinical usefulness of models developed should be tested in clinical trials. It is envisaged that during such trials outcomes for groups of patients treated by clinicians who have access to the model's predictions will be compared with outcomes for groups treated by clinicians who do not.

Rinicare Ltd, a collaborator in this project, will develop software which can run the risk prediction models that are developed in real-time to provide up-to-date estimates of patient risk at the bedside. Such software would be used in the trials mentioned above. If such trials are successful, it is anticipated that within five to ten years a software suite will be available as an "add on" to existing software produced by the major providers of patient monitoring equipment.

12.5.2. Recommendations based on the work presented in specific chapters

While the models validated in chapter seven discriminated well, they should be recalibrated in a larger cohort of patients before their predictions can be considered to be accurate. Successful recalibration would require a dataset much larger than that used in this thesis. Consideration should be given to the purpose of recalibration before selecting the means of increasing the size of the dataset. If the predictions are to be used to guide treatments or inform discussions with

relatives then the recalibration should be performed using a larger dataset from the institution in which the model will be used. Collecting such a large dataset from even the largest single centre would take many years. Consequently, such a dataset is likely to be susceptible to the effects of calibration drift. If the models are to be used to benchmark performance of cardiothoracic critical care units, the models should be calibrated using a large dataset gathered from across multiple centres. If this is the case predictions will not be as accurate when used in an individual institution as they would if the recalibration were performed using data from that institution alone. The discrimination of all three models was adequate to recommend their use for stratification of risk within the critical care unit. Future trials could investigate whether targeting resources according to predicted mortality risk is associated with an improvement in clinical outcomes.

The validation of the use of the Sepsis-3 criteria to identify those with sepsis discussed in chapter eight should encourage further studies to validate the use of the criteria across different patient groups in multiple institutions. The validated definition defines the onset of sepsis as the time when infection is suspected in the presence of an increase >2 in the SOFA score. This endpoint point could be used to identify patients who suffer sepsis and the time at which a diagnosis of sepsis is made. Further work could then seek to identify trends in the physiological data of these patients leading up to the diagnosis. Novel risk prediction algorithms could then be designed to recognise these trends and identify those at greatest risk of developing sepsis.

Chapter nine concludes by recommending that the KDIGO AKI guidelines be revisited to modify the significance attached to isolated oliguria. Such a recalibration would make the system more clinically useful by reducing the number of patients who are classified as suffering AKI by urine output alone; the vast majority of whom have very good outcomes. As discussed in the previous subsection, the model developed in chapter ten should be applied to other patient groups in other settings to ensure its transferability. The use of the model should be the subject of a clinical trial in which outcomes for a group patients treated by clinicians who have access to the model's urine output predictions are compared with outcomes for group treated by clinicians who only have access to raw urine output data and KDIGO AKI stage classification. Importantly, this chapter demonstrated the feasibility of analyses where a patient's own parameters are analysed and modelled to predict future values. Further work will aim to transfer this approach to other physiological parameters. Work already underway by the team working on this project will aim to use a similar approach to develop a risk prediction model to identify those at highest risk of developing atrial fibrillation after cardiac surgery. The work described in chapter eleven of this thesis demonstrated that analyses of serum electrolyte concentrations are unlikely to be useful

for this purpose. Rather, the model will analyse the ECG data collected as part of this project with the aim of identifying subtle differences in the ECG waveforms compared with the modelled predictions which indicate increased risk of subsequent atrial fibrillation. The surprising association between magnesium replacement and onset of postoperative AF should form the basis of a randomised controlled trial.

12.6 Conclusions

A comprehensive dataset has been created for patients undergoing cardiac surgery at Wythenshawe Hospital. Analyses performed using this dataset validated three risk scores for ICU mortality as well as the Sepsis-3 criteria for the identification of those at risk of poor outcomes related to sepsis. Analyses of the performance of the KDIGO-AKI guidelines provided evidence that the criteria used to stratify AKI according to urine output require recalibration before they can reliably be applied to cardiac surgery patients. An alternative means of analysing urine output to better identify those at risk of adverse outcomes related to renal dysfunction has been proposed and validated. Finally, work that will inform the development of models to predict atrial fibrillation has been presented. The established research programme will continue with the overall aim of producing a patient monitoring software suite which can alert clinicians to increasing risk of various complications in individual patients.

REFERENCES

1.   Walshe K, Offen N. A very public failure: lessons for quality improvement in healthcare organisations from the Bristol Royal Infirmary. *Quality in Health Care* 2001;10(4):250.
2.   Bridgewater B, Grant SW, Hickey GL, et al. National Adult Cardiac Surgery Audit Report: National Institute for Cardiovascular Outcomes Research; 2012 2012.
3.   Nashef SA, Roques F, Michel P, et al. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16(1):9-13.
4.   Nashef SA, Roques F, Sharples LD, et al. EuroSCORE II. *Eur J Cardiothorac Surg* 2012;41(4):734-744; discussion 744-735.
5.   Roques F, Michel P, Goldstone AR, et al. The logistic EuroSCORE. *Eur Heart J* 2003;24(9):881-882.
6.   Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development. *Ann Thorac Surg* 2018;105(5):1411-1418.
7.   Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 1-Coronary Artery Bypass Grafting Surgery. *Annals of Thoracic Surgery* 2009;88(1 SUPPL.):S2-S22.
8.   O'Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 2-Isolated Valve Surgery. *Annals of Thoracic Surgery* 2009;88(1 SUPPL.):S23-S42.
9.   Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 Cardiac Surgery Risk Models: Part 3-Valve Plus Coronary Artery Bypass Grafting Surgery. *Annals of Thoracic Surgery* 2009;88(1 SUPPL.):S43-S62.
10.   Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13(10):818-829.
11.   Knaus WA, Wagner DP, Draper EA, et al. THe apache iii prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100(6):1619-1636.
12.   Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981;9(8):591-597.
13.   Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984;12(11):975-977.
14.   Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *Jama* 1993;270(24):2957-2963.
15.   Metnitz PG, Moreno RP, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med* 2005;31(10):1336-1344.
16.   Zimmerman JE, Kramer AA. A history of outcome prediction in the ICU. *Curr Opin Crit Care* 2014;20(5):550-556.
17.   Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22(7):707-710.
18.   Lynch JG. On the External Validity of Experiments in Consumer Research. *Journal of Consumer Research* 1982;9(3):225-239.
19.   Weisse AB. Cardiac Surgery: A Century of Progress. *Texas Heart Institute Journal* 2011;38(5):486-490.
20.   Lewis FJ, Taufic M. Closure of atrial septal defects with the aid of hypothermia; experimental accomplishments and the report of one successful case. *Surgery* 1953;33(1):52-59.
21.   Gibbon JH. The development of the heart-lung apparatus. *The American Journal of Surgery* 1978;135(5):608-619.

22.   Dewall RA, Gott VL, Lillehei CW, et al. A simple, expendable, artificial oxygenator for open heart surgery. *Surg Clin North Am* 1956:1025-1034.

23.   Miller DJ. A short history of myocardial revascularization. In: Jr MD, editor. The practice of coronary artery bypass surgery. New York: Plenum Medical Book Co; 1982.

24.   Baldwin JC EJ, Kopf GS. Heart Sugery. In: Zaret BL MM, Cohen LS., editor. Yale University School of Medicine Heart Book: William Morrow & Co; 1992. p. 313-330.

25.   SCTS. SCTS Blue Book Online - Cardiac Surgery Audit Data. . 2018 [cited 11/06/2018]Available from: http://www.bluebook.scts.org/

26.   Hosseinian L, Chiang Y, Itagaki S, et al. Earlier versus later tracheostomy in patients with respiratory failure after cardiac surgery in the United States. *J Cardiothorac Vasc Anesth* 2014;28(3):488-492.

27.   Bailey ML, Richter SM, Mullany DV, et al. Risk factors and survival in patients with respiratory failure after cardiac operations. *Ann Thorac Surg* 2011;92(5):1573-1579.

28.   Esteban A, Ferguson ND, Meade MO, et al. Evolution of mechanical ventilation in response to clinical research. *Am J Respir Crit Care Med* 2008;177(2):170-177.

29.   Mehta RH, Grab JD, O'Brien SM, et al. Bedside tool for predicting the risk of postoperative dialysis in patients undergoing cardiac surgery. *Circulation* 2006;114(21):2208-2216; quiz 2208.

30.   Serrano N, Garcia C, Villegas J, et al. Prolonged intubation rates after coronary artery bypass surgery and ICU risk stratification score. *Chest* 2005;128(2):595-601.

31.   Mangano CM, Diamondstone LS, Ramsay JG, et al. Renal dysfunction after myocardial revascularization: Risk factors, adverse outcomes, and hospital resource utilization. *Annals of Internal Medicine* 1998;128(3):194-203.

32.   Thille AW, Harrois A, Schortgen F, et al. Outcomes of extubation failure in medical intensive care unit patients. *Crit Care Med* 2011;39(12):2612-2618.

33.   Seymour CW, Martinez A, Christie JD, et al. The outcome of extubation failure in a community hospital intensive care unit: a cohort study. *Crit Care* 2004;8(5):R322-327.

34.   Manns B, Doig CJ, Lee H, et al. Cost of acute renal failure requiring dialysis in the intensive care unit: clinical and resource implications of renal recovery. *Crit Care Med* 2003;31(2):449-455.

35.   Dasta JF, Kane-Gill SL, Durtschi AJ, et al. Costs and outcomes of acute kidney injury (AKI) following cardiac surgery. *Nephrol Dial Transplant* 2008;23(6):1970-1974.

36.   Rewa O, Bagshaw SM. Acute kidney injury[mdash]epidemiology, outcomes and economics. *Nat Rev Nephrol* 2014;10(4):193-207.

37.   Karthik S, Grayson AD, McCarron EE, et al. Reexploration for bleeding after coronary artery bypass surgery: Risk factors, outcomes, and the effect of time delay. *Annals of Thoracic Surgery* 2004;78(2):527-534.

38.   Graf J, Muhlhoff C, Doig GS, et al. Health care costs, long-term survival, and quality of life following intensive care unit admission after cardiac arrest. *Crit Care* 2008;12(4):R92.

39.   Echahidi N, Pibarot P, O'Hara G, et al. Mechanisms, prevention, and treatment of atrial fibrillation after cardiac surgery. *J Am Coll Cardiol* 2008;51(8):793-801.

40.   Kheterpal S, Tremper KK, Heung M, et al. Development and validation of an acute kidney injury risk index for patients undergoing general surgery: results from a national data set. *Anesthesiology* 2009;110(3):505-515.

41.   Nava S, Gregoretti C, Fanfulla F, et al. Noninvasive ventilation to prevent respiratory failure after extubation in high-risk patients. *Crit Care Med* 2005;33(11):2465-2470.

42.   Lei Q, Chen L, Zhang Y, et al. Predictors of prolonged mechanical ventilation after aortic arch surgery with deep hypothermic circulatory arrest plus antegrade selective cerebral perfusion. *J Cardiothorac Vasc Anesth* 2009;23(4):495-500.

43.   Stephens RS, Whitman GJ. Postoperative Critical Care of the Adult Cardiac Surgical Patient. Part I: Routine Postoperative Care. *Crit Care Med* 2015;43(7):1477-1497.

44.   Probst S, Cech C, Haentschel D, et al. A specialized post-anaesthetic care unit improves fast-track management in cardiac surgery: a prospective randomized trial. *Critical Care* 2014;18(4):468.

45.   EfCCNa. Position Statement on Workforce Requirements within European Critical Care Nursing 2007 [cited 2015 08/07/2015]Available from: http://www.efccna.org/images/stories/publication/2007_ps_workforce_requirements_neu.pdf

46.   ICS. Core standards for intensive care units. 2013 [cited 2015 08/07/2015]Available from: http://www.ficm.ac.uk/sites/default/files/Core%20Standards%20for%20ICUs%20Ed.1%20%282013%29.pdf

47.   Bignami E, Belletti A, Moliterni P, et al. Clinical practice in perioperative monitoring in adult cardiac surgery: is there a standard of care? Results from an national survey. *Journal of Clinical Monitoring and Computing* 2015:1-19.

48.   Pezawas T, Rajek A, Plöchl W. Core and skin surface temperature course after normothermic and hypothermic cardiopulmonary bypass and its impact on extubation time. *European journal of anaesthesiology* 2007;24:20-25.

49.   Yazdanian F, Azarfarin R, Aghdaii N, et al. Cardiac variables as main predictors of endotracheal reintubation rate after cardiac surgery. *J Tehran Heart Cent* 2013;8(1):42-47.

50.   Engoren M, Buderer NF, Zacharias A, et al. Variables predicting reintubation after cardiac surgical procedures. *Ann Thorac Surg*;67(3):661-665.

51.   Dorsa AG, Rossi AI, Thierer J, et al. Immediate extubation after off-pump coronary artery bypass graft surgery in 1,196 consecutive patients: feasibility, safety and predictors of when not to attempt it. *J Cardiothorac Vasc Anesth* 2011;25(3):431-436.

52.   Suematsu Y, Sato H, Ohtsuka T, et al. Predictive risk factors for delayed extubation in patients undergoing coronary artery bypass grafting. *Heart and Vessels* 2000;15(5):214-220.

53.   Kern H, Redlich U, Hotz H, et al. Risk factors for prolonged ventilation after cardiac surgery using APACHE II, SAPS II, and TISS: Comparison of three different models. *Intensive care medicine* 2001;27(2):407-415.

54.   Murthy SC, Arroliga AC, Walts PA, et al. Ventilatory dependency after cardiovascular surgery. *Journal of Thoracic & Cardiovascular Surgery* 2007;134(2):484-490.

55.   Rajakaruna C, Rogers CA, Angelini GD, et al. Risk factors for and economic implications of prolonged ventilation after cardiac surgery (Structured abstract). In: Journal of Thoracic and Cardiovascular Surgery; 2005. p. 1270-1277.

56.   He S, Chen B, Li W, et al. Ventilator-associated pneumonia after cardiac surgery: a meta-analysis and systematic review. *J Thorac Cardiovasc Surg* 2014;148(6):3148-3155 e3141-3145.

57.   Anger KE, Szumita PM, Baroletti SA, et al. Evaluation of Dexmedetomidine Versus Propofol-based Sedation Therapy in Mechanically Ventilated Cardiac Surgery Patients at a Tertiary Academic Medical Center. *Critical Pathways in Cardiology* 2010;9(4):221-226.

58.   Engoren M, Buderer NF, Zacharias A. Long-term survival and health status after prolonged mechanical ventilation after cardiac surgery. *Crit Care Med* 2000;28(8):2742-2749.

59.   Algarni KD, Maganti M, Yau TM. Predictors of Low Cardiac Output Syndrome After Isolated Coronary Artery Bypass Surgery: Trends Over 20 Years. *Ann Thorac Surg* 2011;92(5):1678-1684.

60.   Maganti MD, Rao V, Borger MA, et al. Predictors of Low Cardiac Output Syndrome After Isolated Aortic Valve Surgery. *Circulation* 2005;112(9 suppl):I-448-I-452.

61.   Cakir H, Uncu H, Gur O, et al. Risk factors for intraaortic balloon pump use in coronary artery bypass surgery. *Adv Clin Exp Med* 2014;23(2):253-257.

62.   Miceli A, Duggan SM, Capoun R, et al. A clinical score to predict the need for intraaortic balloon pump in patients undergoing coronary artery bypass grafting. *Ann Thorac Surg* 2010;90(2):522-526.

63.   Frost L, Jacobsen CJ, Christiansen EH, et al. Hemodynamic predictors of atrial fibrillation or flutter after coronary artery bypass grafting. *Acta anaesthesiologica Scandinavica* 1995;39(5):690-697.

64.   Hosokawa K, Nakajima Y, Umenai T, et al. Predictors of atrial fibrillation after off-pump coronary artery bypass graft surgery. *British journal of anaesthesia* 2007;98(5):575-580.

65.   Mayr A, Knotzer H, Pajk W, et al. Risk factors associated with new onset tachyarrhythmias after cardiac surgery--a retrospective analysis. *Acta anaesthesiologica Scandinavica* 2001;45(5):543-549.

66.   Patel D, Gillinov MA, Natale A. Atrial Fibrillation after Cardiac Surgery: Where are we now? *Indian Pacing and Electrophysiology Journal* 2008;8(4):281-291.

67.   Attaran S, Shaw M, Bond L, et al. Atrial fibrillation postcardiac surgery: a common but a morbid complication. *Interactive cardiovascular and thoracic surgery* 2011;12(5):772-777.

68.   Mathew JP, Parks R, Savino JS, et al. Atrial fibrillation following coronary artery bypass graft surgery: Predictors, outcomes, and resource utilization. *Journal of the American Medical Association* 1996;276(4):300-306.

69.   Dunning J, Nagarajan DV, Amanullah M, et al. What is the optimal anticoagulation management of patients post-cardiac surgery who go into atrial fibrillation? *Interactive cardiovascular and thoracic surgery* 2004;3(3):503-509.

70.   Frendl G, Sodickson AC, Chung MK, et al. 2014 AATS guidelines for the prevention and management of perioperative atrial fibrillation and flutter for thoracic surgical procedures. *J Thorac Cardiovasc Surg* 2014;148(3):e153-e193.

71.   Crystal E, Connolly SJ, Sleik K, et al. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: A meta-analysis. *Circulation* 2002;106(1):75-80.

72.   Skiba MA, Pick AW, Chaudhuri K, et al. Prophylaxis Against Atrial Fibrillation after Cardiac Surgery: Beneficial Effect of Perioperative Metoprolol. *Heart, Lung and Circulation* 2013;22(8):627-633.

73.   Auer J, Weber T, Berent R, et al. A comparison between oral antiarrhythmic drugs in the prevention of atrial fibrillation after cardiac surgery: The Pilot Study of Prevention of Postoperative Atrial Fibrillation (SPPAF), a randomized, placebo-controlled trial. *American Heart Journal* 2004;147(4):636-643.

74.   Barnes BJ, Kirkland EA, Howard PA, et al. Risk-Stratified Evaluation of Amiodarone to Prevent Atrial Fibrillation After Cardiac Surgery. *Annals of Thoracic Surgery* 2006;82(4):1332-1337.

75.   Topol EJ, Lerman BBL, Baughman KL, et al. De novo refractory ventricular tachyarrhythmias after coronary revascularization. *The American Journal of Cardiology* 1986;57(1):57-59.

76.   Sapin PM, Woelfel AK, Foster JR. Unexpected ventricular tachyarrhythmias soon after cardiac surgery. *Am J Cardiol* 1991;68(10):1099-1100.

77.   Peretto G, Durante A, Limite LR, et al. Postoperative arrhythmias after cardiac surgery: incidence, risk factors, and therapeutic management. *Cardiol Res Pract* 2014;2014:615987.

78.   Brodell GK, Cosgrove D, Schiavone W, et al. Cardiac rhythm and conduction disturbances in patients undergoing mitral valve surgery. *Cleve Clin J Med* 1991;58(5):397-399.

79.   Mehta RH, Sheng S, O'brien SM, et al. Reoperation for bleeding in patients undergoing coronary artery bypass surgery: incidence, risk factors, time trends, and outcomes. *Circulation: Cardiovascular Quality and Outcomes* 2009;2(6):583-590.

80.   Moulton MJ, Creswell LL, Mackey ME, et al. Reexploration for bleeding is a risk factor for adverse outcomes after cardiac operations. *J Thorac Cardiovasc Surg* 1996;111(5):1037-1046.

81.   Unsworth-White J, Herriot A, Valencia O, et al. Resternotomy for bleeding after cardiac operation: A marker for increased morbidity and mortality. *Annals of Thoracic Surgery* 1995;59(3):664-667.

82.   Beck CS. TWo cardiac compression triads. *Journal of the American Medical Association* 1935;104(9):714-716.

83.   Price S, Prout J, Jaggar SI, et al. 'Tamponade' following cardiac surgery: terminology and echocardiography may both mislead. *European Journal of Cardio-Thoracic Surgery* 2004;26(6):1156-1160.

84.   Byrne JG, Leacche M, Paul S, et al. Risk factors and outcomes for 'vasoplegia syndrome' following cardiac transplantation. *Eur J Cardiothorac Surg* 2004;25(3):327-332.

85.   Omar S, Zedan A, Nugent K. Cardiac vasoplegia syndrome: pathophysiology, risk factors and treatment. *Am J Med Sci* 2015;349(1):80-88.

86.   Vives M, Wijeysundera D, Marczin N, et al. Cardiac surgery-associated acute kidney injury. *Interactive cardiovascular and thoracic surgery* 2014;18(5):637-645.

87.   Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practical Guideline for Acute Kidney Injury in *Kidney Inter* 2012. p. 1-138.

88.   Mangano CM, Diamondstone LS, Ramsay JG, et al. Renal dysfunction after myocardial revascularization: risk factors, adverse outcomes, and hospital resource utilization. The Multicenter Study of Perioperative Ischemia Research Group. *Annals of Internal Medicine* 1998;128(3):194-203.

89.   Thakar CV, Worley S, Arrigain S, et al. Influence of renal dysfunction on mortality after cardiac surgery: Modifying effect of preoperative renal function. *Kidney Int* 2005;67(3):1112-1119.

90.   Ng SY, Sanagou M, Wolfe R, et al. Prediction of acute kidney injury within 30 days of cardiac surgery. *Journal of Thoracic and Cardiovascular Surgery* 2014;147(6):1875-1883.

91.   Abah U, Large S. Stroke prevention in cardiac surgery. *Interactive cardiovascular and thoracic surgery* 2012;15(1):155-157.

92.   Miyazaki S, Yoshitani K, Miura T, et al. Predictors of postoperative derilium and stroke after off-pump coronary artery bypass surgery. *Journal of Neurosurgical Anesthesiology* 2008;20(4).

93.   Aboyans V, Labrousse L, Lacroix P, et al. Predictive factors of stroke in patients undergoing coronary bypass grafting: statins are protective. *European Journal of Cardio-Thoracic Surgery* 2006;30(2):300-304.

94.   Engelman DT, Cohn LH, Rizzo RJ. Incidence and predictors of tias and strokes following coronary artery bypass grafting: report and collective review. *The heart surgery forum* 1999;2(3):242-245.

95.   George L. Review of intraoperative TOE. *Australas J Ultrasound Med* 2009;12(2):30-36.

96.   Engelman RM, Engelman DT. Strategies and Devices to Minimize Stroke in Adult Cardiac Surgery. *Seminars in Thoracic and Cardiovascular Surgery* 2015;27(1):24-29.

97.   J. A, M. S. Brain injury after adult cardiac surgery. *Acta anaesthesiologica Scandinavica* 2004;48(1):4-19.

98.   Brown CH. Delirium in the Cardiac Surgical Intensive Care Unit. *Current Opinion in Anaesthesiology* 2014;27(2):117-122.

99.   Norkiene I, Ringaitiene D, Kuzminskaite V, et al. Incidence and risk factors of early delirium after cardiac surgery. *BioMed Research International* 2013;2013.

100.   Mangusan RF, Hooper V, Denslow SA, et al. Outcomes associated with postoperative delirium after cardiac surgery. *Am J Crit Care* 2015;24(2):156-163.

101.   Andrejaitiene J, Sirvinskas E. Post-cardiac surgery delirium risk factors and clinical outcome. *Heart surgery forum* 2011;14:S97-S98.

102.   Gottesman RF, Grega MA, Bailey MM, et al. Delirium after coronary artery bypass graft surgery and late mortality. *Ann Neurol* 2010;67(3):338-344.

103.   Burkhart CS, Dell-Kuster S, Gamberini M, et al. Modifiable and nonmodifiable risk factors for postoperative delirium after cardiac surgery with cardiopulmonary bypass. *Journal of cardiothoracic and vascular anesthesia* 2010;24(4):555-559.

104.   Boodhwani M, Rubens FD, Wozny D, et al. Predictors of early neurocognitive deficits in low-risk patients undergoing on-pump coronary artery bypass surgery. *Circulation* 2006;114(SUPPL. 1):I461-I466.

105.   van den Boogaard M, Pickkers P, van der Hoeven H, et al. Implementation of a delirium assessment tool in the ICU can influence haloperidol use. *Critical Care* 2009;13(4):R131-R131.

106.   McDonagh DL, Mathew JP, White WD, et al. Cognitive Function after Major Noncardiac Surgery, Apolipoprotein E4 Genotype, and Biomarkers of Brain Injury. *Anesthesiology* 2010;112(4):852-859.

107.   Newman MF, Kirchner JL, Phillips-Bute B, et al. Longitudinal Assessment of Neurocognitive Function after Coronary-Artery Bypass Surgery. *New England journal of medicine* 2001;344(6):395-402.

108.   Steinmetz MDJ, Christensen PDKarl B, Lund PDT, et al. Long-term Consequences of Postoperative Cognitive Dysfunction. *Anesthesiology* 2009;110(3):548-555.

109.   Lombard FW, Mathew JP. Neurocognitive Dysfunction Following Cardiac Surgery. *Seminars in Cardiothoracic and Vascular Anesthesia* 2010;14(2):102-110.

110.   van Harten AE, Scheeren TWL, Absalom AR. A review of postoperative cognitive dysfunction and neuroinflammation associated with cardiac surgery and anaesthesia. *Anaesthesia* 2012;67(3):280-293.

111.   Mathew JP, Mackensen GB, Phillips-Bute B, et al. Randomized, double-blinded, placebo controlled study of neuroprotection with lidocaine in cardiac surgery. *Stroke* 2009;40(3):880-887.

112.   Mathew JP, White WD, Schinderle DB, et al. Intraoperative magnesium administration does not improve neurocognitive function after cardiac surgery. *Stroke* 2013;44(12):3407-3413.

113.   Nagels W, Demeyere R, Van Hemelrijck J, et al. Evaluation of the neuroprotective effects of S(+)-ketamine during open-heart surgery. *Anesth Analg* 2004;98(6):1595-1603, table of contents.

114.   Cheng DC, Bainbridge D, Martin JE, et al. Does off-pump coronary artery bypass reduce mortality, morbidity, and resource utilization when compared with conventional coronary artery bypass? A meta-analysis of randomized trials. *Anesthesiology* 2005;102(1):188-203.

115.   Liu YH, Wang DX, Li LH, et al. The effects of cardiopulmonary bypass on the number of cerebral microemboli and the incidence of cognitive dysfunction after coronary artery bypass graft surgery. *Anesth Analg* 2009;109(4):1013-1022.

116.   Silbert BS, Scott DA, Evered LA, et al. A comparison of the effect of high- and low-dose fentanyl on the incidence of postoperative cognitive dysfunction after coronary artery bypass surgery in the elderly. *Anesthesiology* 2006;104(6):1137-1145.

117.   Rasmussen LS, Johnson T, Kuipers HM, et al. Does anaesthesia cause postoperative cognitive dysfunction? A randomised study of regional versus general anaesthesia in 438 elderly patients. *Acta anaesthesiologica Scandinavica* 2003;47(3):260-266.

118.   Mierdl S, Meininger D, Dogan S, et al. Abdominal complications after cardiac surgery. *Annals of the Academy of Medicine Singapore* 2001;30(3):245-249.

119.   Soos P, Schmack B, Weymann A, et al. Perioperative risk factors for serious gastrointestinal complications treated by laparotomy after cardiac surgery using cardiopulmonary bypass. *Critical Care* 2011;15.

120.   Vohra HA, Farid S, Bahrami T, et al. Predictors of survival after gastrointestinal complications in bypass grafting. *Asian Cardiovascular and Thoracic Annals* 2011;19(1):27-32.

121.   Eris C, Yavuz S, Yalcinkaya S, et al. Acute mesenteric ischemia after cardiac surgery: an analysis of 52 patients. *ScientificWorldJournal* 2013;2013:631534.

122.   Lola I, Levidiotou S, Petrou A, et al. Are there independent predisposing factors for postoperative infections following open heart surgery? *Journal of cardiothoracic surgery* 2011;6.

123.   Pawar M, Mehta Y, Kapoor P, et al. Central venous catheter-related blood stream infections: incidence, risk factors, outcome, and associated pathogens. *Journal of Cardiothoracic & Vascular Anesthesia* 2004;18(3):304-308.

124.   Forstner C, Lassnigg A, Tobudic S, et al. A prospective analysis of invasive candidiasis following cardiac surgery: Severity markers are predictive. *Journal of Infection* 2013;66(6):528-535.

125.   Pasero D, de FG, Rana NK, et al. Candidemia after cardiac surgery in the intensive care unit: An observational study. *Interactive cardiovascular and thoracic surgery* 2011;12(3):374-378.

126.   Inweregbu K, Dave J, Pittard A. Nosocomial infections. *Continuing Education in Anaesthesia, Critical Care & Pain* 2005;5(1):14-17.

127.   Bion J, Richardson A, Hibbert P, et al. 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. *BMJ quality & safety* 2013;22(2):110-123.

128.   O'Grady NP, Murray PR, Ames N. Preventing Ventilator-Associated Pneumonia: Does the Evidence Support the Practice? *JAMA : the journal of the American Medical Association* 2012;307(23):2534-2539.

129.   Reddy HG, Shih T, Englesbe MJ, et al. Analyzing "failure to rescue": Is this an opportunity for outcome improvement in cardiac surgery? *Annals of Thoracic Surgery* 2013;95(6):1976-1981.
130.   Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* 2009;361(14):1368-1375.
131.   Pasquali SK, He X, Jacobs JP, et al. Evaluation of failure to rescue as a quality metric in pediatric heart surgery: An analysis of the STS congenital heart surgery database. *Annals of Thoracic Surgery* 2012;94(2):573-580.
132.   R Core Team. R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing 2015.
133.   Kang H. Appropriate design of research and statistical analyses: observational versus experimental studies. *Korean Journal of Anesthesiology* 2013;65(2):105-107.
134.   Thakar CV, Arrigain S, Worley S, et al. A clinical score to predict acute renal failure after cardiac surgery. *J Am Soc Nephrol* 2005;16(1):162-168.
135.   Machado MN, Nakazone MA, Maia LN. Prognostic value of acute kidney injury after cardiac surgery according to kidney disease: improving global outcomes definition and staging (KDIGO) criteria. *Plos One* 2014;9(5):e98028.
136.   Lagny MG, Jouret F, Koch JN, et al. Incidence and outcomes of acute kidney injury after cardiac surgery using either criteria of the RIFLE classification. *BMC Nephrol* 2015;16:76.
137.   Jonkers D, Elenbaas T, Terporten P, et al. Prevalence of 90-days postoperative wound infections after cardiac surgery. *European Journal of Cardio-Thoracic Surgery* 2003;23(1):97-102.
138.   Rady MY, Ryan T, Starr NJ. Perioperative determinants of morbidity and mortality in elderly patients undergoing cardiac surgery. *Critical Care Medicine* 1998;26(2):225-235.
139.   Joskowiak D, Kappert U, Rudolph A, et al. Readmission to intensive care unit after cardiac surgery - A single center experience with 5983 patients. *Thoracic and cardiovascular surgeon* 2011;59.
140.   Boeken U, Minol J, Assmann A, et al. Intensive care unit readmission after cardiac surgery: Predictors and consequences. *Interactive cardiovascular and thoracic surgery* 2012;15.
141.   Messaoudi N, De J, Stockman B, et al. Prediction of prolonged length of stay in the intensive care unit after cardiac surgery: The need for a multi-institutional risk scoring system. *Journal of Cardiac Surgery* 2009;24(2):127-133.
142.   Ettema RGA, Peelen LM, Kalkman CJ, et al. Predicting prolonged intensive care unit stays in older cardiac surgery patients: A validation study. *Intensive care medicine* 2011;37(9):1480-1487.
143.   Lapar DJ, Speir AM, Crosby IK, et al. Postoperative atrial fibrillation significantly increases mortality, hospital readmission, and hospital costs. *Annals of Thoracic Surgery* 2014;98(2):527-533.
144.   Caiado CCS, Hickey, G.L., Grant, S.W., Goldstein, M., Markarian, G., McCollum, C., Bridgewater, B. Heart Online Uncertainty and Stability Estimation. . In: Bissell JJea, editor. Tipping Points: Modelling Social Problems and Health: Wiley; 2015.
145.   Teasdale G, Jennett B. ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS: A Practical Scale. *The Lancet* 1974;304(7872):81-84.
146.   Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med* 2002;166(10):1338-1344.
147.   Hekmat K, Kroener A, Stuetzer H, et al. Daily assessment of organ dysfunction and survival in intensive care unit cardiac surgical patients. *Annals of Thoracic Surgery* 2005;79(5):1555-1562.
148.   Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology From Bed to Bench* 2012;5(2):79-83.
149.   Student. THE PROBABLE ERROR OF A MEAN. *Biometrika* 1908;6(1):1-25.
150.   Jamart J. Statistical tests in medical research. *Acta Oncol* 1992;31(7):723-727.
151.   Bellolio MF, Serrano LA, Stead LG. Understanding statistical tests in the medical literature: which test should I use? *International Journal of Emergency Medicine* 2008;1(3):197-199.

152.   Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann Math Statist* 1947;18(1):50-60.
153.   Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* 1952;47(260):583-621.
154.   Fisher Robert A. Chapter 4. In: Statistical Methods for Research Workers. 13 ed. Edinburgh: Oliver and Boyd; 1958.
155.   Pearson K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1900;50(302):157-175.
156.   Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 1958;53(282):457-481.
157.   Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966;50(3):163-170.
158.   J. WS, D. OJ. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* 2012;3(2):260-267.
159.   Hickey GL, Grant SW, Caiado C, et al. Dynamic prediction modeling approaches for cardiac surgery. *Circ Cardiovasc Qual Outcomes* 2013;6(6):649-658.
160.   Zingone B, Pappalardo A, Dreas L. Logistic versus additive EuroSCORE. A comparative assessment of the two models in an independent population sample. *European Journal of Cardio-Thoracic Surgery* 2004;26(6):1134-1140.
161.   Salis S, Mazzanti VV, Merli G, et al. Cardiopulmonary bypass duration is an independent predictor of morbidity and mortality after cardiac surgery. *J Cardiothorac Vasc Anesth* 2008;22(6):814-822.
162.   Madhavan S, Chan SP, Tan WC, et al. Cardiopulmonary bypass time: every minute counts. *The Journal of cardiovascular surgery* 2018;59(2):274-281.
163.   Greenberg JW, Schuessler RB, Lancaster TS, et al. Postoperative atrial fibrillation following cardiac surgery: a persistent complication. *European Journal of Cardio-Thoracic Surgery* 2017;52(4):665-672.
164.   C. SJ. Logistic Regression: A Brief Primer. *Academic Emergency Medicine* 2011;18(10):1099-1104.
165.   Wald A. Sequential Tests of Statistical Hypotheses. *Ann Math Statist* 1945;16(2):117-186.
166.   Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine* 2013;4(2):627-635.
167.   Vivian V, Stéphane R, Françoise C-C, et al. How to evaluate the calibration of a disease risk prediction tool. *Statistics in Medicine* 2009;28(6):901-916.
168.   Hosmer DW, Lemeshow S. Applied logistic regression: Wiley; 1989.
169.   Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in Medicine* 2014;33(3):517-535.
170.   Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model†. *European Journal of Cardio-Thoracic Surgery* 2018:ezy180-ezy180.
171.   Doerr F, Badreldin AMA, Bender EM, et al. Outcome prediction in cardiac surgery: The first logistic scoring model for cardiac surgical intensive care patients. *Minerva anestesiologica* 2012;78(8):879-886.
172.   Badreldin AM, Doerr F, Bender EM, et al. Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices. *European Journal of Cardio-Thoracic Surgery* 2013;44(6):992-997; discussion 997-998.
173.   Hickey GL, Grant SW, Murphy GJ, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg* 2013;43(6):1146-1152.
174.   Kellum JA, Sileanu FE, Murugan R, et al. Classifying AKI by Urine Output versus Serum Creatinine Level. *Journal of the American Society of Nephrology* 2015;26:2231-2238.

175.   Ralib A, Pickering JW, Shaw GM, et al. The urine output definition of acute kidney injury is too liberal. *Critical Care* 2013;17(3):R112.
176.   Auer J, Weber T, Berent R, et al. Serum potassium level and risk of postoperative atrial fibrillation in patients undergoing cardiac surgery. *J Am Coll Cardiol* 2004;44(4):938-939; author reply 939.
177.   Krijthe BP, Heeringa J, Kors JA, et al. Serum potassium levels and the risk of atrial fibrillation: The Rotterdam Study. *International Journal of Cardiology* 2013;168(6):5411-5415.
178.   Wahr JA, Parks R, Boisvert D, et al. Preoperative serum potassium levels and perioperative outcomes in cardiac surgery patients. Multicenter Study of Perioperative Ischemia Research Group. *Jama* 1999;281(23):2203-2210.
179.   O'Dowd A. Major global cyber-attack hits NHS and delays treatment. *BMJ* 2017;357.
180.   Goldberg SI, Niemierko A, Turchin A. Analysis of Data Errors in Clinical Research Databases. *AMIA Annual Symposium Proceedings* 2008;2008:242-246.
181.   Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49(12):1373-1379.
182.   Vittinghoff E, McCulloch CE. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology* 2007;165(6):710-718.
183.   Bhatti F, Grayson AD, Grotte G, et al. The logistic EuroSCORE in cardiac surgery: how well does it predict operative risk? *Heart* 2006;92(12):1817-1820.
184.   Toumpoulis IK, Anagnostopoulos CE, DeRose JJ, et al. European system for cardiac operative risk evaluation predicts long-term survival in patients with coronary artery bypass grafting. *Eur J Cardiothorac Surg* 2004;25(1):51-58.
185.   Kieser TM, Rose MS, Head SJ. Comparison of logistic EuroSCORE and EuroSCORE II in predicting operative mortality of 1125 total arterial operations †. *European Journal of Cardio-Thoracic Surgery* 2016;50(3):509-518.