UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Academic Unit of Primary Care and Population Sciences

Prediction modelling in acute hospital care: a case study of acute kidney injury

By

Dr Luke Eliot Hodgson

Thesis for the degree of Doctor of Medicine

October 2018

ABSTRACT i

Background: acute kidney injury (AKI) is a global healthcare priority, with strategies to improve associated outcomes desirable. Prediction models or clinical prediction rules (CPRs), utilise multiple predictors to provide objective estimates of future risk and their use has been suggested in the field of AKI. A rapid expansion in derived CPRs has rarely been followed by external validation, with even fewer models undergoing impact analysis on patient outcomes. Using e-alert systems to identify AKI has become commonplace in the UK, though few studies have assessed their efficacy. Combining an in-hospital e-alert system with an AKI CPR has not previously been described.

Methods: a systematic review of CPRs to predict hospital-acquired AKI (HA-AKI) in acute hospital settings was performed, followed by an external validation of one – the AKI prediction score (APS). Thirdly an impact analysis study using a controlled, before-and-after design on acute medical admissions to two adult non-specialist hospital sites was conducted (2014-16). At admission, the CPR highlighted patients at risk of HA-AKI in conjunction with an e-alert which triggered care bundles of interventions. Primary outcome was incident HA-AKI using a difference-in-differences analysis. Secondary outcomes in those developing HA-AKI included: in-hospital mortality, AKI progression, intensive care unit (ICU) escalation and effects on process measures. Patients with established community-acquired AKI (CA-AKI) were also highlighted only at the intervention site.

Results:

I. A systematic review found 53 CPR studies, the majority in specialised areas; of the 11 in general hospital settings five had external validation. Significant shortcomings in design and reporting were found.
II. External validation of the APS model found modest discrimination with area under the receiver operating characteristic curves (AUROCs) ranging 0.65-0.71 and acceptable calibration plots.
III. An impact analysis combining the CPR with an AKI e-alert found a reduction in HA-AKI at the intervention site (odds ratio, OR 0·990 (0·981-1·000), P=0·049). Of cases who developed HA-AKI mortality significantly reduced on unadjusted (27·46% pre vs 21·67% post intervention, OR 0·731 (0·560-0·954) P=0·021) and difference-in-differences analysis (OR 0·924 [95% CI 0·858-0·996] P=0·038). Process measures significantly improved at the intervention site. In contrast no outcome improvements were found in patients presenting with CA-AKI following introduction of an AKI e-alert.

Conclusions: Few validated AKI CPRs have been described in general hospital settings. An external validation of one model, the APS, led on to an impact analysis of this CPR. This innovative study utilised IT to integrate the CPR and an AKI e-alert in an acute hospital setting. The major findings were a reduction in de novo hospital-acquired AKI and reduced mortality in those who developed AKI. Future research should assess if these findings are generalisable and sustainable. Integration with primary care IT and the employment of biomarkers to apply precision care are further avenues following on from this case study.

Publications, Presentations ii

Parts of this work have been published or presented:

LE Hodgson, BD Dimitrov PJ Roderick, R Venn, LG Forni. A systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. BMJ Open 2017;7:e016591. doi: 10.1136/bmjopen-2017-016591.

LE Hodgson, BD Dimitrov PJ Roderick, R Venn, LG Forni. Predicting AKI in Emergency Admissions: An external validation study of the acute kidney injury prediction score (APS). BMJ Open 2017 Mar 8;7(3):e013511. doi: 10.1136/bmjopen-2016-013511.

LE Hodgson, PJ Roderick, R Venn, GL Yao, BD Dimitrov, LG Forni. The ICE-AKI Study: **I**mpact analysis of a **C**linical prediction rule and **E**lectronic AKI alert in general medical patients. PLoS One. 2018; 13(8): e0203183.

Work relating to the thesis has also been delivered as a podium presentation at the European Society of Intensive Care Medicine (ESICM), annual conference, Vienna September 2017:
LE Hodgson, D Hargreaves, G Chater, J. de Carvalho, R Venn, L Forni. Impact analysis of an integrated electronic alert for those at risk or with acute kidney injury: effects on process measures.

Contents

Contents Page iii

List of figures iv

List of tables v

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Appendices vi

Appendices iv

Academic Thesis: Declaration Of Authorship vii

I, Luke Hodgson

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Prediction modelling in acute hospital care: a case study of acute kidney injury

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

LE Hodgson, BD Dimitrov PJ Roderick, R Venn, LG Forni. A systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. BMJ Open 2017;7:e016591. doi: 10.1136/bmjopen-2017-016591.
LE Hodgson, BD Dimitrov PJ Roderick, R Venn, LG Forni. Predicting AKI in Emergency Admissions: An external validation study of the acute kidney injury prediction score (APS). BMJ Open 2017 Mar 8;7(3):e013511. doi: 10.1136/bmjopen-2016-013511.
LE Hodgson, PJ Roderick, R Venn, GL Yao, BD Dimitrov, LG Forni. The ICE-AKI Study: Impact analysis of a Clinical prediction rule and Electronic AKI alert in general medical patients. PLoS One. 2018; 13(8): e0203183.

Signed: .....

Date: .....

Acknowledgements viii

This thesis is dedicated first to the memory of Borislav Dimitrov, who sadly passed away early in 2017, who provided me with endless insights into statistics whilst we bounced an array of research ideas between us over a coffee or two. Second to Bethan my partner and my children Lois and Soren for their support and understanding throughout. Finally my sincere thanks go to my supervisors and friends Professors Paul Roderick and Lui Forni and Dr Richard Venn for their invaluable guidance from beginning to end of this thesis, though the research road will continue.

List of abbreviations ix

ACEi – Angiotensin-converting enzyme inhibitors
ADQI - Acute Dialysis Quality Initiative Group
ADH - antidiuretic hormone
ATP - Adenosine triphosphate
AKI – Acute kidney injury
AKIN – Acute kidney injury network
ALT – Alanine aminotransferase
ARB – Angiotensin receptor blockers
ARF – Acute renal failure
ASA – American Society of Anesthesiologists Physical status grading used in pre-operative assessment
AST – Aspartate transaminase
ATN – acute tubular necrosis
AVPU – scale of consciousness best response: **A**lert, responds to **V**oice, **P**ain, **U**nresponsive.
AUC/AUROC – Area under the receiver operating characteristic curve
BMI – Body mass index
BP – Blood pressure
CA-AKI – Community-acquired AKI
Ca^{2+} - Serum Calcium
CI-AKI – Iodinated contrast AKI
CKD – Chronic kidney disease
COPD - Chronic obstructive pulmonary disease
CCF – Congestive cardiac failure
CKD-EPI – CKD Epidemiology collaborative equation
COX – Cyclo-oxygenase
CrCl - creatinine clearance
CRP – C-reactive protein
D – Derivation study
Da - Daltons
DBP – Diastolic Blood Pressure
ECMO - Extracorporeal membrane oxygenation
eGFR – estimated glomerular filtration rate
ESRD – end-stage renal disease
EDTA - Ethylenediaminetetraacetic acid
EV – External validation study
HA-AKI – Hospital-acquired AKI
Hb - Haemoglobin
HbA1C – glycated haemoglobin (A1c) Marker of long-term glucose control
HCO_3 – serum Sodium Bicarbonate
HES - hospital episode statistics
H-L – Hosmer-Lemeshow goodness-of-fit test (Calibration statistic)
HR – Heart rate (beats per minute)
HTN – Hypertension
ICD - International Classification of Diseases
ICU – Intensive Care Unit
IDMS - isotope dilution-mass spectrometry
IHD – Ischaemic heart disease
IV – Internal Validation study
K^+ - Serum Potassium
KDIGO – Kidney disease improving global outcomes (Stage 1-3 AKI defined by magnitude of SCr rise or fall in urine output)
LOS – length of stay
LVAD – Left ventricular assist device
LVEF – Left ventricular ejection fraction
MAP – Mean arterial pressure
MDRD – Modification of diet in renal disease equation

Mg^{2+} - serum Magnesium

MI – Myocardial infarction

Na^{+} – Serum sodium

NCEPOD
NICE - National institute for health and care excellence
NSAID – Non-steroidal anti-inflammatory drug

NYHA – New York Heart Association Classification for heart failure (I-IV)

PVD – Peripheral vascular disease

RCT - randomised controlled trial
RBF - Renal blood flow

RR – respiratory rate (breaths per minute)

RRT – Renal replacement therapy
RIFLE – Risk, Injury, failure, loss of kidney function

SCr – serum creatinine

Systolic Blood Pressure
TBW - total body water
TRIPOD – Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis. Checklist in derivation (37 points – 1 point for each recommended item reported).

TRIPOD Study types

Type 1a: Development only

Type 1b: Development and validation using resampling

Type 2a: Random split-sample development and validation,

Type 2b: Non-random split-sample development and validation

Type 3: Development and validation using separate data

Type 4: Validation only.

VD - volume of distribution
WCC – White cell count

WRF – 'worsening renal failure' (defined by individual study)

Chapter 1 - Introduction

1.1 Acute Kidney Injury (AKI)

1.1.1 An introduction to AKI

Acute kidney injury (AKI) is a syndrome,(1) characterised by a rapid decrease in renal excretory function, with the accumulation of products of nitrogen metabolism, decreased urine output (not always), accumulation of metabolic acids, potassium and phosphate.(2, 3) The term AKI replaced acute renal failure (ARF) in part to emphasise a continuum of injury exists long before sufficient loss of excretory kidney function can be measured with standard laboratory tests ie serum creatinine (SCr).(2, 4) William Heberden in 1802 gave the first description of ARF, then termed “ischuria renalis.”(5) During the First World War the syndrome was named “war nephritis,”(6) and was largely forgotten until Bywaters and Beall’s seminal paper on crush syndrome at the time of the Second World War.(7) Acute tubular necrosis (ATN) was the term used to describe this clinical entity, based on patchy necrosis of renal tubules seen at autopsy. In clinical practice, the terms ATN and ARF were used interchangeably for years and until recently, a precise biochemical definition for ARF was lacking. With no consensus on diagnostic criteria, a 2002 survey revealed at least 35 definitions of ARF described in the scientific literature.(8) Such knowledge undoubtedly stimulated more recent efforts to provide consensus definitions for both research and clinical practice.(3)

1.1.2 AKI is a research priority

AKI has long been recognised as an important marker of the gravity of an underlying disease but was not felt to be directly responsible for outcomes.(9, 10) However, recent studies have demonstrated AKI is independently associated with mortality, morbidity, increased length of hospital stay, occurrence or progression of chronic kidney disease (CKD) and the need for renal replacement therapy (RRT) including dialysis and transplantation, and health care costs.(11-14) However the absence of a consensus definition contributed to a marked heterogeneity of published studies hindering comparison between patients and groups and extrapolation of studies.(4, 8) In 2005 the American Society of Nephrology Renal Research Report highlighted significant knowledge gaps relating to AKI from its natural history, spectrum, risk factors and underlying aetiology. Data on long-term outcomes and the influence on CKD was particularly limited.(15) The US Centre for Disease Control 2008 report found that between 1984 and 2005 much of the observed increase in reported kidney disease was a result of increases in hospitalisation for AKI.(16) They highlighted a need for research to determine the causes for this and examine the risk for associated CKD.

Both the Acute Dialysis Quality Initiative (ADQI) Group, who proposed a graded classification system known as the RIFLE (Risk Injury Failure Loss End Stage Renal Disease) criteria(17) and the Acute Kidney Injury Network (AKIN)(18) aimed to standardise the definition of AKI to facilitate research and collaboration. In 2012 KDIGO (Kidney Disease: Improving Global Outcomes) proposed a single definition of AKI for practice, research and public health, combining the RIFLE and AKIN definitions with some modifications.(3) KDIGO AKI is defined as an acute (hours to days) increase in serum creatinine (SCr) or a reduction in urine volume (table 1.1 definitions).(3)

AKIN staging		RIFLE		KDIGO		Urine output (all definitions)
Stage	**SCr**	**Class**	**SCr or GFR**	**Stage**	**SCr**	
1	Increase ≥26.5 µmol/l or increase 150-200% baseline	**Risk**	Increase SCr 150-200% baseline or GFR decrease >25%	**1**	Increase ≥26.5 µmol/l or increase 150-200% baseline	<0.5 ml/kg/hr >6 hrs
2	Increased >200-300% baseline	**Injury**	SCr >200-300% or GFR decreased >50%	**2**	Increase 200-300% baseline	<0.5 ml/kg/hr >12 hrs
3	Increased >300% baseline, or ≥354 µmol/l & acute increase ≥44 µmol/l or on RRT	**Failure**	SCr >300%, or SCr >354 µmol/l & acute rise >44 µmol/l or GFR decreased >75%	**3**	Increase >300% baseline, or ≥354 µmol/l or on RRT	<0.3 ml/kg/hr >24 hrs or anuria >12 hrs
		Loss	Persistent acute renal failure = complete loss of kidney function >4 weeks			
		End-stage	ESRD >3 months			

Table 1.1 AKI definitions. For AKIN, the increase in SCr must occur in <48 hours. For RIFLE, AKI should be abrupt (1–7 days) and sustained (>24 hours). ESRD – end-stage kidney disease, RRT – renal-replacement therapy, SCr – serum creatinine. Adapted from AKIN, RIFLE, KDIGO definitions.(3, 17, 18)

1.1.3 The purpose for this research

AKI is now recognised to be a global healthcare priority.(19, 20) Research in AKI prevention and improved management are significant areas identified by healthcare leaders. Though much research has highlighted AKI risk factors, few studies have investigated targeted interventions to recognise risk, prevent AKI development and improve management of those who have met AKI criteria in a general hospital setting, outside the intensive care unit. This research aims to study the utility of prognostic prediction models in the field by first reviewing currently available models in general and specialised hospitalised populations. Secondly, an external validation of one model will be performed. Finally an impact analysis study will investigate the effects of a complex healthcare intervention, including the use of a prediction model and an AKI e-alert, on patient outcomes.

1.2 The Kidney – anatomy and function

1.2.1 Introduction

The nephron is the functioning unit of the kidney, made up of a glomerulus and tubule system. Each kidney has over a million nephrons. The glomerulus is made up of the afferent arteriole, bringing blood from the systemic circulation, a tuft of capillaries with drainage through the efferent arteriole. Blood is filtered across the glomerular basement membrane into the space in Bowman's Capsule to form an ultrafiltrate of plasma which then passes through the tubules and is subjected to selective secretion and reabsorption of electrolytes and other solutes. Through this process the kidney maintains water, electrolyte and acid base balance. The urine produced is concentrated according to prevailing physiological conditions and is excreted via the ureters into the bladder. The kidneys are a crucial regulator of body homeostasis. Primarily the kidney regulates fluid and electrolyte balance, acid/base balance, and excretes metabolic waste products. However, the kidney also produces hormones and has some metabolic functions (table 1.2).

Functions of the kidney
1. Maintenance of electrolyte balance
2. Maintenance of fluid balance
3. Maintenance of Acid/Base balance
4. Elimination of metabolic waste products including urea and creatinine.
5. Endocrine & Metabolic: production of Erythropoetin, Renin and Calcitriol; metabolism of Insulin

Table 1.2 – summary of kidney functions.

1.2.2 Measuring kidney function – glomerular filtration rate (GFR)

The glomerular filtration rate (GFR), the rate in unit time at which fluid is filtered by the glomerulus, is accepted as the best index of kidney

function.(21) What constitutes normal GFR is debatable with two of the most widely cited studies dating back to the 1950s and 1960s.(22) A more recent study found when adjusted for body surface area healthy females had a mean GFR of 108.7 ±17.5 ml/min and males 106.1 ±15.8 ml/min per 1.73 m^2.(23) The glomerular capillary hydrostatic pressure provides the driving force for filtration and is dependent on cardiac output (the kidneys receive around 25% of the cardiac output) and circulating volume. Oncotic or colloid osmotic pressure from plasma proteins within the filtered blood and the hydrostatic pressure in Bowman's space oppose this pressure. The net filtration pressure is the difference between these opposing pressures (figure 1.1).(24) GFR is also determined by the glomerular filtration coefficient reflecting the permeability and surface area of the basement membrane across which fluid is filtered. GFR and its change over time are vital to detection of kidney disease and its severity.(25) In AKI the GFR is usually decreased compromising the functional capabilities of the kidney.

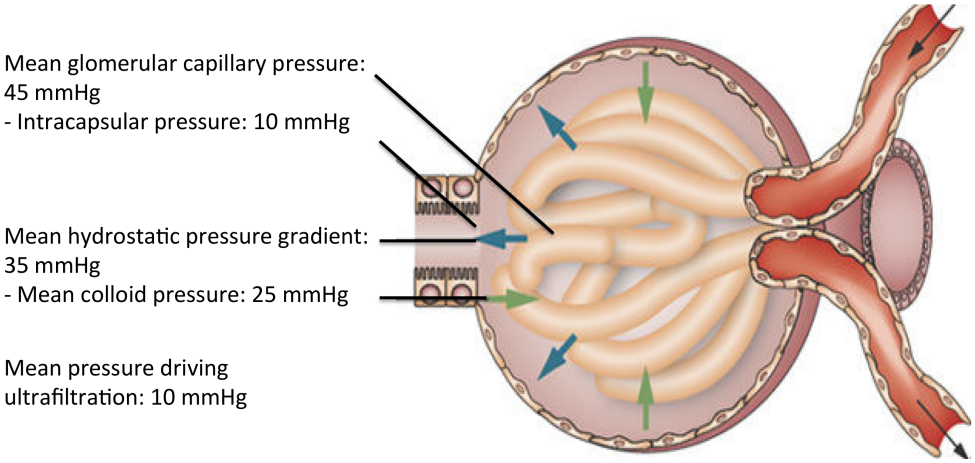


Figure 1.1 Glomerular haemodynamics adapted from Prowle et al 2010.(26)

An ideal endogenous marker of renal filtration function would exhibit the following characteristics:

- constant rate of production,
- water soluble,
- no protein binding,
- freely filtered at glomerulus,
- no tubular secretion or reabsorption,
- no extrarenal metabolism or elimination and,
- an assay for measurement which is precise, reliable, rapid, inexpensive, and widely available.

Unfortunately, to date no such perfect endogenous marker has been described. Renal clearance is based on the premise that the rate by which a substance is removed from the plasma by the kidney must be equal to the rate of excretion into the urine. Knowing the urine and plasma concentrations of a substance and the urine flow rate, allows calculation of clearance which may be conceptualised as the virtual

volume of plasma from which this substance is completely removed during a time period.(27) Inulin is the gold standard filtration marker, however, the requirement for an intravenous infusion and a difficult chemical assay makes it impractical for clinical use.(27) Although there are other established exogenous serum markers to estimate GFR such as iothalamate, ethylenediaminetetraacetic acid and iohexol, their use is expensive and impractical for routine use.(28) Thus, in clinical practice, measurement of serum creatinine (SCr) and urea remain the mainstay for determining changes in renal function.

1.2.3 Creatinine

Creatinine was named by Liebig in 1847. In 1926 Poul Brandt Rehberg made the assumption that creatinine was exclusively excreted through glomeruli, and proposed that by measuring creatinine in urine and blood, a glomerular filtration rate could be derived from the clearance rate of creatinine.(29) In popular use since the 1940s, Creatinine remains the laboratory hallmark of AKI in daily practice, in part because of its convenience and low cost.(30-32) Creatinine is an amino acid compound with a molecular weight of 113 daltons (Da), derived from the conversion of creatine (Cr) to phosphorylcreatine in skeletal muscle and subsequent liver metabolism of creatine (summary in figure 1.2). In health creatinine is released into the plasma at a relatively constant rate, is freely filtered by the glomerulus, and is not reabsorbed or metabolised by the kidney. Clearance of SCr is the most widely used means for estimating GFR and SCr levels generally have an inverse relationship to GFR.(33) SCr together with a urine collection can be used to estimate GFR by measuring creatinine clearance (CrCl) while several equations have been developed to estimate GFR based on SCr alone. However, as a filtration marker SCr has multiple limitations (figure 1.3 and table 1.3 which also highlights the limitations of urea).(2) Firstly, baseline SCr (often not known)(34) can be affected by a host of factors including age, ethnicity, sex and dietary intake. In acute illness the release of creatinine into the serum can be variable including because of reduced production. In a

study of patients with glomerulonephritis, when compared with inulin, creatinine grossly and unpredictably overestimated GFR.(35) Several drugs (e.g. trimethoprim and H_2 receptor blockers),(36) impair creatinine secretion causing a transient and reversible increases in SCr. In diabetic ketoacidosis, increased serum concentration of acetoacetate can cause interference with selected assays and present a falsely elevated SCr referred to as the Jaffe reaction.(37) SCr does not reflect real-time changes in GFR or always a genuine acute injurious process;(28) as GFR falls, fractional hypersecretion of creatinine in glomerular disease increases as the disease worsens;(35, 38, 39) due to the nonlinear relation between concentrations of SCr and GFR, the former becomes abnormal only when the latter decreases by more than 50% and does not reflect dynamic changes in filtration rates.(40, 41) This time taken to accumulate may delay the diagnosis of AKI and, hence, delay appropriate supportive and therapeutic interventions. For these reasons SCr has been termed an 'imperfect gold standard'.(42)

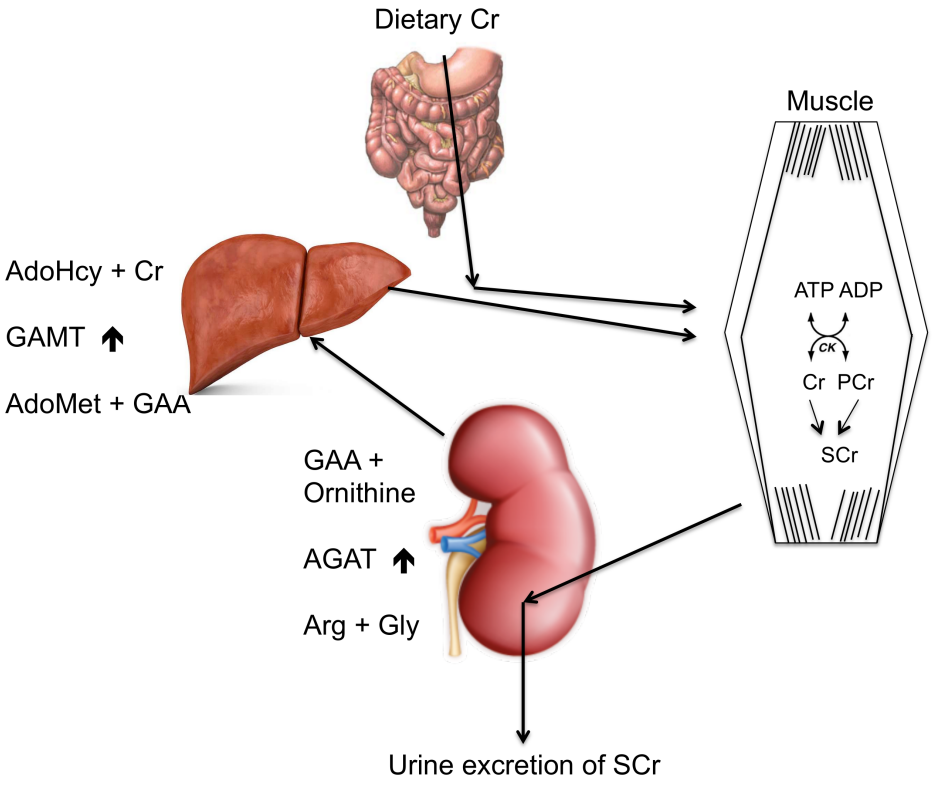


Figure 1.2 Major routes of Creatine metabolism. The majority of Cr is in muscular tissues with daily demand met by intestinal absorption of dietary Cr or biosynthesis. The

first step of Cr biosynthesis probably occurs mainly in the kidney, whereas the liver is likely to be the principal organ accomplishing methylation of GAA to Cr. Muscular Cr and PCr are non-enzymatically converted at an almost steady rate (~2% of total Cr per day) to serum Creatinine (SCr), which diffuses out of the cells and is excreted by the kidneys into the urine. AdoHcy -*S*-adenosyl-L-homocysteine, AdoMet – *S*-adenosyl-L--methionine, Arg – Arginine, - Cr – Creatine, GAA - Guanidinoacetic acid, GAMT - glycine amidinotransferase, Gly -Glycine , PCr – phosphorylcreatine SCr – serum Creatinine.

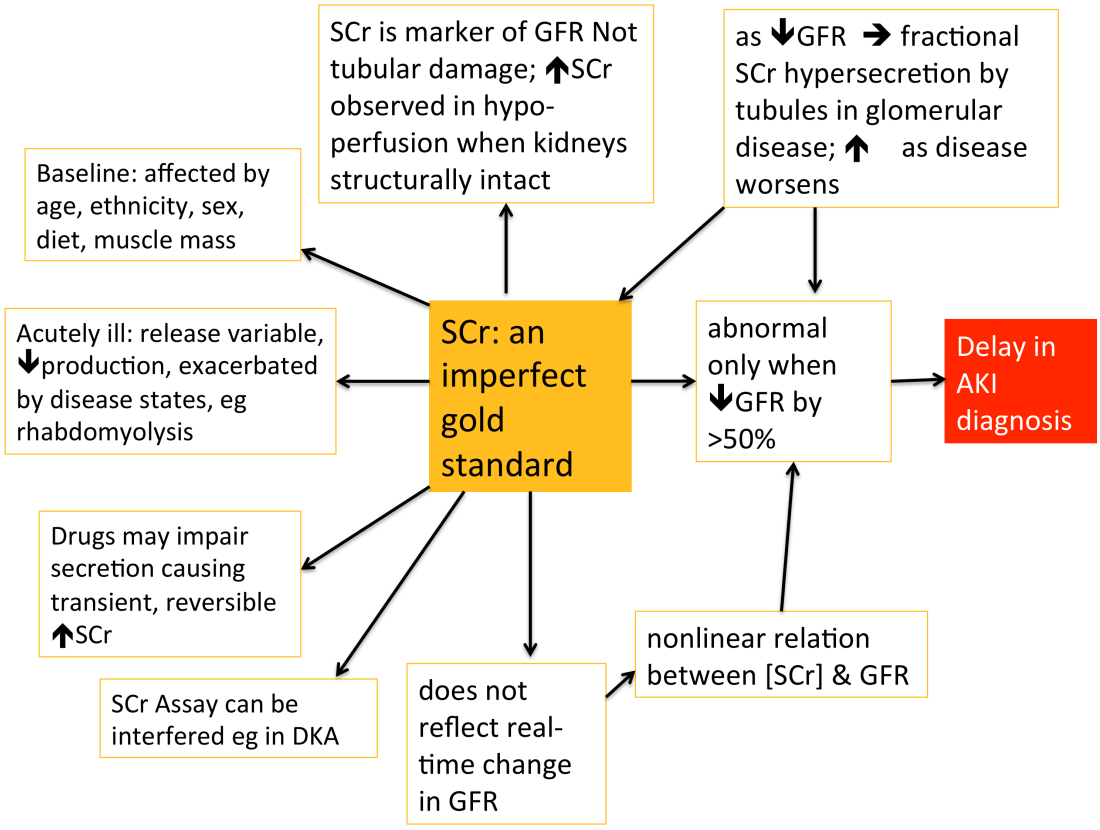


Figure 3. Creatinine the imperfect gold standard. GFR – glomerular filtration rate, SCr – serum creatinine.

	Increase	Decrease
Creatinine	Male	Female
	Younger age	Older age
	Large muscle mass	Protein restriction (renal disease, liver disease)
	Ethnicity	Vegetarian diet
	Ingestion of cooked meat	Muscle wasting (neuromuscular diseases, malnutrition)
	Jaffe reaction (ketotic states, hyperglycemia)	Amputation
	Drugs (cimetidine, trimethoprim)	Jaffe reaction (hyperbilirubinemia)
	Vigorous exercise	
Urea	Decreased circulating volume	Volume expansion
	Increased dietary protein	Pregnancy
	Critical illness (fever, trauma, burns, sepsis)	SIADH
	Gastrointestinal bleeding	Dietary protein restriction
	Drugs (corticosteroids, tetracyclines)	Liver disease

Table 1.3 - Causes of alterations in creatinine and urea to consider when using these markers as a measure of renal function. SIADH – syndrome of inappropriate Antidiuretic Hormone secretion.

1.2.4 Creatinine kinetics and relevance to AKI

There is a well-described inverse correlation between CrCl and SCr. Thus if all determinants of SCr concentration are kept constant except for CrCl, then a 50% reduction in CrCl from baseline will lead at steady state to a 100% increase in SCr, irrespective of baseline function. Steady state, however, may not be reached for several days after an episode of AKI.(43) Waikar and Bonventre simulated creatinine kinetics after AKI in the setting of normal baseline kidney function and CKD stages 2-4 (stage 2 GFR 60-89 mL/min, stage 3 GFR 30-59, stage 4 GFR 15-29).(44) They found that the percentage changes in SCr after severe AKI are highly dependent on baseline kidney function. Twenty-four hours after a 90%

reduction in CrCl, rise in SCr was 246% with normal baseline kidney function, 174% in stage 2 CKD, 92% in stage 3 CKD, and only 47% in stage 4 CKD. In contrast, absolute increase was nearly identical across the spectrum of baseline function. Time to reach a 50% increase in SCr was directly related to baseline kidney function: normal baseline took 4 hours whilst stage 4 CKD took 27 hours. Again in contrast, the time to reach a 44.2 μmol/L in SCr was virtually identical after a >50% reduction in CrCl.

Alterations in the volume of distribution (VD) of creatinine can in turn alter SCr concentration. Though studies have suggested that the VD of SCr is roughly equivalent to total body water (TBW)(45), among critically-ill patients (e.g. following resuscitation for sepsis, burns and pancreatitis), increase in TBW can reach >10% within 72 hours.(46) In critically-ill patients with AKI, Macedo and colleagues calculated cumulative fluid balance and computed a fluid-adjusted SCr concentration reflecting the effect of VD during the development phase of AKI.(47) They demonstrated that dilution of SCr may lead to underestimation of the severity of AKI and increases the time required to identify its development. However, this approach has been criticised by others, who have argued that this fails to account for the actual kinetics of creatinine excretion.(48) A study by Pickering *et al,* also demonstrated using a model combining creatinine kinetic and volume kinetics that fluid resuscitation leads to an underestimation of both SCr concentration and AKI severity.(49)

1.2.5 Urea

Measurement of urea marked the beginning of efforts to quantify renal function with its isolation by Rouelle in 1773 and subsequent use as a diagnostic test by Strauss in 1903.(32) Urea is a low molecular weight (60 Da) water-soluble, by-product of protein and amino acid metabolism

produced by the liver, used as a serum marker of uraemic solute retention and elimination. In chronic hemodialysis, the degree of urea clearance correlates with outcome and is used to model hemodialysis adequacy. Acute and large rises in serum urea concentration are characteristic of the development of the uraemic syndrome and retention of a large variety of toxins. In a review of 857 publications 90 such compounds were described.(50) The accumulation of urea itself is believed to predispose to adverse bio-chemical, and physiologic and metabolic effects, such as increased oxidative stress, altered function of co-transport pathways important in regulation of intracellular electrolytes and water, and alterations in immune function.(51, 52) Furthermore, retention of uraemic toxins may contribute to secondary organ dysfunction.(53) Similar to SCr, urea levels exhibit a nonlinear and inverse relationship with GFR. However, the use of urea levels to estimate GFR is problematic due to numerous extrarenal factors that influence its endogenous production and renal clearance independent of GFR (table 1.3). The rate of urea production is not constant with serum values modified by a high protein intake, critical illness (sepsis, burns, trauma), gastrointestinal hemorrhage (due to absorption of degraded blood products), level of hydration and drug therapy, such as corticosteroids. Of note patients with chronic liver disease may have near normal values for both SCr (decreased production due to decreased hepatic creatine synthesis, increased tubular creatinine secretion, or loss of skeletal muscle mass)(54) and urea (decreased production and protein restriction) despite severely reduced GFR and impaired kidney function. The rate of renal clearance of urea is also not constant. An estimated 40–50% of filtered urea is passively reabsorbed by proximal and distal renal tubular cells and in states of decreased effective circulating volume there is enhanced resorption of sodium and water in the proximal renal tubular cells along with a corresponding increase in urea resorption. Consequently, serum urea may increase out of proportion to changes in SCr and be under-representative of GFR.

Overall urea concentration is a poor measure of GFR, being produced at variable rates, affected by a number of disease states and undergoes renal tubular reabsorption.(32) Furthermore it does not represent real-time changes in GFR and requires time to accumulate. Likewise, urea does not reflect true “acute” kidney injury. As such, reliance on urea could lead to potential delays in diagnosis of acute changes to GFR or detection of AKI. As a result of these limitations it is not generally used in the setting of AKI, nor is not used in any of the GFR estimating equations in current use.

1.2.6 Oliguria

The English physician Heberden used the definition of oliguria associated with acute renal failure over 200 years ago.(55) Though part of consensus definitions of AKI oliguria is often neglected, in part due to the large number of patients outside the ICU setting with no urinary catheter to accurately measure urine output. Furthermore, though urine output is an important clinical marker it is not renal specific.(56, 57) Indeed a reduction in output may be an appropriate physiological response during periods of hypovolaemia, fasting, post-operatively, following pain or trauma.(58) For example, as a response to hypovolaemia, antidiuretic hormone (ADH) is released leading to oliguria. A study nearly 80 years ago demonstrated that up to 60 hours fasting resulted in a concentrated urine of less than 20 ml/hour,(59) supported by a more recent study of oliguric on-call junior doctors.(60) As such oliguria may be misleading, reflecting normal physiological responses or a transient haemodynamic disturbance, not necessarily indicative of glomerular or tubular injury. Moreover, severe tubular injury may not be apparent initially as oliguria given that concentrating ability is disturbed.(55) Under such circumstances, urine flow reflects GFR and oliguria will not be apparent until GFR has fallen significantly. AKI Studies have suggested that oliguria in isolation is associated with the least elevation in mortality risk.(61) However, in combination with elevations in SCr, oliguria is

associated with the highest mortality.(56) This combination probably represents stress and impairment of renal function coupled with a significant fall in GFR manifesting as oligoanuria.(56)

1.2.7 Beyond creatinine and oliguria

Due to limitations of existing markers of renal injury, there has been much interest in the detection and validation of new biomarkers for AKI to replace or complement SCr.(57, 62-64) Such biomarkers vary in anatomical origin, physiological function, release time, kinetics, and distribution (see table 1.4).(65) Both serum and urine biomarkers have been described.

Biomarker	Reflects
Cystatin C	Glomerular filtration
Albuminuria	Glomerular integrity
IGFBP-7 and TIMP2	Tubular stress
NGAL, KIM-1, NAG, L-FAB	Tubular damage
IL-18	Intra-renal inflammation

Table 1.4 – renal biomarkers. IGFBP-7 - Insulin-like growth factor binding protein-7, IL-18 - interleukin-18, KIM-1 - kidney injury molecule-1, L-FAB - liver fatty acid-binding protein, NAG - N-acetyl-β-d-glucosaminidase, NGAL - neutrophil gelatinase-associated lipocalin, TIMP2 - tissue inhibitor metalloproteinase 2.

Biomarkers may allow detection of subtle changes in renal function prior to SCr rises.(66) Of note, patients who are biomarker-positive, without SCr elevations appear to have a greater risk of complications, a longer stay in hospital and higher mortality.(67) Indeed international guidance have suggested that biomarkers be utilised in combination with traditional markers to better define and characterise AKI.(66) To date few of these tests are in routine hospital use however.

1.3 AKI Significance

1.3.1 Epidemiology of AKI

The epidemiology of AKI has been difficult to ascertain in part due to the variety of definitions in use prior to RIFLE, AKIN and KDIGO.(3, 17, 18) Furthermore, a recent review found the long-term prognosis following an episode of AKI varies, depending not just on cause and clinical setting, but may also be explained by underlying pre-AKI and post-AKI renal function rather than the AKI episode itself.(68) Using the KDIGO definition one in five adults worldwide admitted to hospital either have AKI at presentation, or develop AKI during their hospital admission.(69) However, widely differing rates of AKI are reported across a spectrum of emergency and elective admissions. For example following general surgery incidence has been reported to be around 1%,(70, 71) whilst amongst critically ill patients it can be as high as 70%, with an in-hospital mortality of 50%.(69, 72-76) Cardiac surgery is one of the areas in which AKI has been extensively studied, with incidence rates of 3–30%,(77-80) and up to 5% of patients requiring RRT, associated with significant acute mortality.(81) In this speciality many of the studies preceded consensus definitions. One recent UK study by Bernie et al using KDIGO criteria suggested an overall incidence of 23% with 4% progressing to AKI stage 3. In this study 30-day mortality increased with severity: 2% in those stage 1, 4% stage 2 and 28% with stage 3 AKI, compared to 0.5% in those without AKI.(82) In the field of Liver transplantation requirement for RRT has been reported in up to 17% of transplant recipients.(83) In studies of general hospitalised populations using KDIGO criteria incidence of hospital-acquired AKI ranged 7-10%.(84-86)

Current understanding of the epidemiology of AKI is largely based on studies of patients who developed hospital-acquired AKI (HA-AKI). However, the majority of cases probably arise from the community (CA-AKI),(84, 87-91) with volume depletion the most commonly identified

cause.(88) Interestingly, recent studies demonstrate a significant proportion of CA-AKI patients are never admitted.(92, 93) The study by Johnson and colleagues described the potential role of linked hospital and community clinical records to improve recognition as well as definitions of AKI. In this population based study over the course of a year 0.8% met AKI criteria with over two-thirds of cases having their first alert in the community, with fewer than half of such cases requiring hospital admission.(93)

Disentangling whether a patient has community-acquired (CA-AKI), hospital acquired AKI (HA-AKI), a combination, or indeed neither, can be difficult for a number of reasons. First, on admission to hospital a previous SCr result may not be available or may be from a number of months ago, making baseline establishment problematic. Second, as rises in SCr may lag significantly behind an insult, an injury may have already been sustained prior to admission, but not initially identified as AKI.(30) Third, a patient may have a degree of community injury and then over the course of initial hospital management be exposed to further insult, for example, through iatrogenic interventions or medications. Fourth recognition of AKI relies on repeat testing and a proportion of cases may never come to medical attention, or have only a single blood sample or have two blood samples spaced far enough apart to miss AKI criteria if renal function has recovered and creatinine values have fallen. A number of studies suggest that whilst not all community cases are admitted to hospital, a significant number are, with significant potential to harness data linkage between these two settings. Such real time data could aid establishment of baseline values and potentially improve recognition and track the trajectory of renal function at a population level.(92, 93)

Worldwide, in developed and developing regions there are both similarities and differences in incidence, cause, pathophysiology, and

public health implications of AKI.(19) For example in rural areas of developing countries, AKI is commonly a consequence of community-acquired diseases such as diarrhoea, dehydration, infectious diseases and animal venoms. Furthermore, under-reporting of AKI is a major problem in understanding the true knowledge of its impact in many parts of the world, especially in developing countries.

Large epidemiologic investigations have indicated that the incidence of AKI is increasing.(94) For example, a US national study of cases of dialysis-requiring AKI found that from 2000 to 2009, incidence increased from 222 to 533 cases per million person-years.(95) The authors reported that temporal changes in the population distribution of age, race, and sex, trends of sepsis, acute heart failure, and receipt of cardiac catheterisation and mechanical ventilation accounted for only about a third of the observed increase. In this study the total number of deaths associated with dialysis-requiring AKI rose from 18,000 to nearly 39,000 over the time period. Of particular relevance in developed countries with ageing populations, a number of studies have described an age-dependent relationship with AKI.(96, 97) This is attributed to both anatomic and physiologic changes in the ageing kidney and to comorbidities (i.e. CKD, CCF, hypertension, cardiovascular disease),(98) that may require procedures and/or medications that act as stressors and alter renal haemodynamics or are directly nephrotoxic.(99)

In the critically ill with AKI survival remains poor even after key developments in the supportive care afforded to such patients including protective lung ventilation(100) and improved survival from sepsis.(101) Although RRT is the mainstay of supportive treatment for advanced AKI, this therapy is potentially harmful and not readily available in all settings and global regions.(102) Indeed different resuscitation strategies for sepsis in two recent large multi-centre international randomised controlled trials (RCTs) showed no difference in AKI incidence.(103) This may relate, in part, to changes that have occurred in the typical ICU patient, who is often older with more comorbid illness (including CKD) with

diminished physiological reserve.(28) However, another explanation is the limited capacity of conventional markers of kidney function to detection early injury to the kidney. This may partly explain why therapeutic interventions showing promise in animal models have proven disappointing in human trials,(104, 105) where a delay in diagnosis from time of insult to recognition of AKI may have attenuated the opportunity for early successful intervention before established injury occurs. AKI is thought to also contribute to the CKD population by permanently reducing renal function in people with and without prior CKD.(106) For these reasons AKI is regarded as a priority for clinical practice, policy and research.(19) Indeed, in 2015 The Lancet and the International Society of Nephrology in a joint commission issued a call to arms - the Global 0by25 initiative - to reduce mortality from preventable and avoidable AKI.(20) This commission put forward five 'Rs' (risk assessment, recognition, response, renal support, and rehabilitation) as specific areas to target to improve patient outcomes. As most cases of AKI are attributable to simple causes such as volume depletion, hypotension, and exposure to nephrotoxic medications,(107) attention has shifted from treatment to prevention, early detection, and proactive management to avoid further damage.(102)

1.3.2 Aetiology

1.3.2.1 Traditional concepts

A report by Bywaters and Beal in 1941 linking crush injury to acute impairment of renal function is arguably the starting point for modern medicine's discussion of AKI.(108) Traditionally ARF was split into pre-renal (a decline in glomerular filtration due to renal hypoperfusion), intrinsic (structural damage to the kidney including acute nephritis), and post-renal (obstructive) causes.(109) Mechanical obstruction of urinary flow and intrinsic renal aetiologies categorised by the structure primarily affected (i.e. tubular, glomerular, interstitial, or vascular) are relatively well understood mechanisms, with the latter having a clear histopathological

rationale. However, the majority of AKI is not due to obstruction or intrinsic disease. The term ‘pre-renal’ was used to convey the notion of a systemic or distant insult subsequently resulting in renal dysfunction. For example, an acute myocardial infarction (MI) leading to poor cardiac output or acute haemorrhage leading to hypovolaemia. However, the pathophysiology of AKI is rarely so simply explained, with probably multiple acute factors combing in an augmented fashion to affect a vulnerable host.(110) A heterogeneous group of hospitalised patients are affected by AKI, usually with pre-existing risk factors such as age, diabetes, heart failure, CKD, liver disease and cancer.(3, 111) Acute insults, may include sepsis, volume depletion including haemorrhage and dehydration, surgery, nephrotoxic drugs and contrast agents.

The mechanism of pre-renal AKI had previously been postulated to result from global renal ischaemia, cellular damage and ATN. However, this paradigm has been challenged.(112) ATN describes a form of intrinsic AKI due to severe, persistent hypoperfusion of the kidneys not thought to be associated with histopathological changes, therefore not classified as intrinsic AKI and expected to resolve within days. The term is conceptually flawed implying as it does a degree of certainty from taking a history, performing an examination and doing blood/urine tests that no histopathological tubular injury is present with a renal biopsy rarely performed.(113) Furthermore, pre-renal azotaemia and ATN imply that AKI does not represent a continuum of injury.

Given that the leading clinical conditions associated with AKI, namely sepsis, major surgery, heart failure and hypovolaemia,(114) are all associated with shock, AKI has frequently been attributed to ischaemia on the basis of macro-haemodynamic changes. Consistent morphological findings in ischaemic AKI are apoptotic cell death in distal and proximal tubules,(115, 116) and vascular congestion, endothelial damage, and leukocyte accumulation in peritubular capillaries.(117, 118) An increasing body of evidence suggests AKI can occur in the absence of overt signs of hypoperfusion, global or regional.(110) A complex series of alterations

are thought to take place in the renal tubular system. These include tubular dynamics such as activation of tubuloglomerular feedback, which may even include increased blood flow.(119) Tubular metabolism alterations including depletion of Adenosine triphosphate (ATP) lead to the generation of reactive oxygen molecules and impaired calcium sequestration.(119) The structural response of the tubule cell is multi-faceted and includes loss of cell polarity and brush borders, cell death and proliferation. Surviving tubule cells have a remarkable ability to regenerate and proliferate following AKI.(120) This regeneration involves the appearance of dedifferentiated epithelial cells, followed by up-regulation of genes that encode for a variety of growth factors cells, expression of differentiation factors and finally undergo re-differentiation.(121)

1.3.2.2 New paradigms - septic AKI and organ cross-talk

1.3.2.2.1 Septic AKI

Sepsis, defined as life-threatening organ dysfunction caused by a dysregulated host response to infection, is a frequent cause of AKI amongst hospitalised patients.(122) Though sepsis-induced AKI is increasingly recognised to be more complex than a simple pre-renal insult, understanding the mechanisms of sepsis-induced AKI phenotype have been limited by the lack of pathologic specimens available to study as a consequence of the inherent risks of performing renal biopsies in this patient group. A consistent observation in human and animal studies is the presence of:

- diffuse microcirculatory flow abnormalities,(123)
- inflammation and,(124)
- adaptive cellular responses to injury.(125)

Current evidence suggests the origin of most cases of sepsis-induced AKI is multifaceted, rather than the result of an individual insult.(110) Despite controversy about the role of global renal blood flow (RBF), both animal models and human studies have shown that the occurrence of

sepsis-induced AKI is not exclusive of decreased RBF states (126), and indeed can develop in the setting of increased RBF.(127, 128) Interestingly in one study of patients following cardiac arrest, only a small proportion developed significant AKI, unless post-arrest cardiogenic shock was present.(129) One proposed theory suggests sepsis-induced AKI reflects the early manifestation of an adaptive response of the tubular cells to a danger signal.(110) Inflammation and microvascular dysfunction both characterise and amplify this signal and in response, mitochondria within tubular cells orchestrate a metabolic downregulation and reprioritisation, favouring cell survival processes (such as maintenance of membrane potential and cell cycle arrest), at the expense of "kidney function" (i.e. tubular absorption and secretion of solutes).(110, 130)

1.3.2.2.2 Organ cross-talk

Increasing evidence has suggested that the effects of metabolic and fluid disturbances seen in AKI cases may be immunosuppressive and have adverse effects on distant organs – so-called organ cross-talk.(131-133) Though full explanations remain to be elucidated as to why AKI is an independent contributor to mortality,(134) the adverse effects of complex organ interactions may contribute.(135) In animal models, there is a clear causal effect: injuring or removing the kidneys results in deleterious systemic effects and distant organ dysfunction. Potential mechanisms include dysfunctional inflammatory cascades, oxidative stress, activation of pro-apoptotic pathways, differential molecular expression, and leukocyte trafficking.(134) Almost all the major organs have now been found to be affected in AKI cases as cause of the AKI, impairment due to the effects of AKI or a combination of the two (figure 1.4 and appendix table A1.1 for summary).(134, 136, 137)

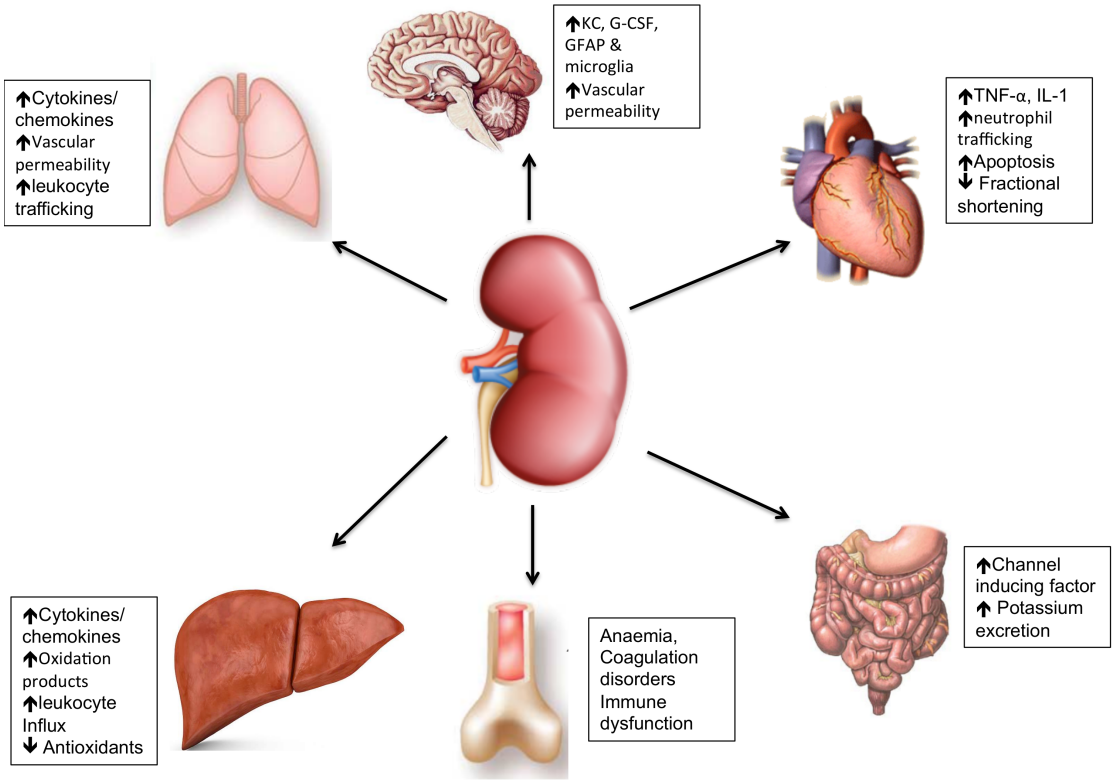


Figure 14. Organ cross-talk in AKI in part adapted from Scheel et al.(136) G-CSF – granulocyte colony-stimulating factor, GFAP – glial fibrillary acidic protein, GI – gastrointestinal, GSH – glutathione, IL – interleukin, KC – keratinocyte-derived chemokine, TNF – tumour necrosis factor.

1.3.3 Morbidity, mortality and origin of AKI

Despite a well known association between AKI and mortality, Levy et al's study in 1996 was perhaps the first to separate the effect of ARF on outcome from the effect of an underlying disease process complicated by ARF. After adjusting for differences in comorbidity, ARF was associated with an odds ratio of dying of 5.5.(138) Mortality figures vary in the literature depending on AKI definition and population studied with contemporary work using consensus definitions suggest a short-term mortality of around 9-34% in those hospitalised.(139-141) A UK study using hospital episode statistics (HES) suggested AKI has an in-patient mortality of 28%.(142) In one recent study thirty-day mortality was similar for HA-AKI and CA-AKI, but significantly lower in cases not admitted (24.2%, 20.2% and 2.6%, respectively).(92) Five-year mortality was high

in all groups, but followed a similar pattern (67.1%, 64.7% and 46.2%).

A substantial body of evidence suggests even small acute changes in kidney function are associated with both an increase in mortality and the development or worsening of existing of CKD with its associated morbidity and costs.(12, 106, 143-145) For example, incident CKD rates of up to 30% after a period of AKI following cardiac surgery have been reported.(146-151) A meta-analysis of 13 studies found pooled incidences of CKD and end stage renal failure (ESRF) post-AKI were 25.8/100 person-years and 8.6/100, respectively.(106) Compared to patients without AKI, adjusted hazard ratios (HRs) were 8.8 for developing CKD, 3.1 for ESRF and 2.0 for mortality.

Both Cruz et al in the ICU setting and Uchino in all admissions found a graded increased risk of mortality with worsening RIFLE class.(152, 153) Increased mortality has also been reported when oliguria in addition to a rise in SCr has been present.(56, 152) Coca's 2010 study of post-operative AKI found that for each AKI stage, longer duration of AKI was associated with higher mortality. Of further note mortality rate for those with severe AKI of short duration was much lower than the mortality associated with stage 1 AKI but of longer duration. They postulated that duration could discriminate those with true renal injury and those with only 'prerenal' AKI.

The link between AKI and CKD has been the subject of a multiple studies,(154-157) though demonstration of a clear association does not necessarily confer causation. Epidemiological studies often struggle to identify accurate pre-morbid and post-AKI renal function in order to precisely interpret long-term data. For example, in retrospective studies follow-up data may be missing or may have been captured at times of intercurrent illness, hence blunt endpoints, such as dialysis dependence

or mortality, are used. In addition, SCr and the derived eGFR, despite their limitations, are often the only markers of renal function used. Critical illness in particular, may be associated with significant decreases in SCr through many potential mechanisms and these changes may persist through to hospital discharge hence confounding assessment of renal function.(158) Moreover, elevated SCr levels at hospital discharge may represent pre-existing CKD rather than non-recovery, depending on the completeness of data availability on prior renal function.

Several studies have reported lower mortality in patients with AKI who had pre-existing CKD.(111, 159, 160) The reason for this finding is not clear though for example, Khosla et al postulated process of care for those with CKD may be differ such as access to nephrology.(159) Patients with CKD may also require a lesser burden of acute illness to reach the same level of AKI. One study found no difference in outcomes when adjustment was made for acute illness severity.(161) The threshold definition of AKI within the CKD cohort may also be different with a study by Broce et al finding in an adjusted model the odds ratio for in-hospital mortality became significant with a creatinine increment of >0.2mg/dl in those with a baseline eGFR >60mls/min, whilst in those with an eGFR <30mls/min this was not significant until an increment of >0.5mg/dl was reached.(162)

1.3.4 Economic consequences

Most cost estimates of AKI have come from single-centre studies of academic hospitals.(163) For example, in a 2005 study using KDIGO SCr criteria, AKI was associated with an average $7,082 increase in costs, ranging $5,400 for KDIGO stage 1 to $27,300 for KDIGO stage 3.(12) Kerr *et al*, using routine data for NHS England, after sensitivity analyses, estimated in-patient costs related to AKI ranged £894,193,943 to £1,153,732,733, or approximately £1,100 per episode. This amount was

just over 1% of the NHS budget in 2010-11.(142) Silver *et al*, used International Classification of Diseases (ICD) codes to ascertain AKI, known to identify patients with more severe AKI.(164) Estimated in-patient costs related to AKI in the United States ranged from $5.4-24 billion.(165) In this study, additional costs attributed to dialysis-requiring AKI ranged from $11,016 to $42,077. Using a Markov model to estimate long-term costs arising from CKD and ESRD in patients who have had AKI relative to a matched group without AKI, Kerr *et al*, found that post-discharge care cost the NHS an additional £179 million. Further research has been called for in this area to better quantify the costs associated with post-discharge AKI care.(163)

1.3.5 Deficits in AKI management and interventions to address these concerns

Significant deficits in the recognition and subsequent management of patients who have developed AKI have been shown.(166, 167) The 2009 National Confidential Enquiry into Patient Outcome and Death (NCEPOD) report suggested poor management of patients who died with AKI in 40% of cases.(166) In England the National Institute for Health and Care Excellence (NICE) guidance for AKI has emphasised risk assessment to be integral to improving care. In 2014 NHS England issued a patient safety alert mandating the flagging of patients with AKI in hospital and in primary care. Studies have reported the impact of electronic hospital systems flagging patients with AKI,(141, 168-170) but not those at risk of developing AKI. Two of the studies found a reduction in mortality in the subgroup of patients who had a care bundle completed.(141, 169) However, a recent large well conducted US single centre RCT did not detect any clinical benefit in alerting patients with hospitalised AKI to the clinical care team, though the authors argued this may have been because minor AKI is an artefact of multiple testing of an inherently variable substance (SCr).(170) Additionally, it is likely that the issue of an alert in isolation is unlikely to be effective without an effector arm, for

instance simple accessible guidance, education and effective treatment. Finally, it is possible that the study was open to cross-contamination, whereby there were improvements in recognition and care of non-alerted AKI cases.(170) There has been much recent interest in highlighting patients at risk of AKI or at an earlier stage of the evolution of AKI at which point interventions, or avoidance of harms, could have most benefit.

Though change in SCr is the most common way AKI is diagnosed, there are a number of issues using this at scale electronically in healthcare systems:

- Optimal measurement using enzymatic assay and calibration traceable to a reference method based on isotope dilution-mass spectrometry (IDMS) standard and the presence of chromogens that interfere with measurement,(171)
- relevance of e-alert systems using estimated baseline and how baseline is defined, with a significant number of cases lacking a prior SCr,
- differentiating community from hospital-acquired cases,
- management of multiple alerts in the same patient,
- the significance of small changes in SCr in patients with low weight or pre-existing CKD,
- influence of fluid balance / dilution / volume of distribution and,
- biological variation in production, for example in sepsis.(172)

All the above can contribute to false positive and false negative alerts.(173) However, despite these limitations a pragmatic AKI algorithm has been suggested for use within the NHS.(174) Indeed this has been shown to work reasonably well,(175) though in individual cases will fall short of expert assessment. Future developments would ideally include comprehensive data linkage between hospital and community settings to

aid with baseline establishment, early correct identification of the setting of the AKI and the significance of this setting.

In summary AKI is now a global health priority, with measures to reduce associated mortality urgently called for. The Global 0by25 initiative commission put forward five ‘Rs’ (risk assessment, recognition, response, renal support, and rehabilitation) as specific areas to target to improve patient outcomes.(20)

1.4 Prediction Models

1.4.1 Background

For millennia, there has been a desire to predict the future. Though it is widely believed that most complex systems are computationally irreducible, with their future time evolution inherently unpredictable,(176) this has not prevented the application of the scientific method to the prediction of novel phenomena.(177) For example, the prediction of Neptune’s existence from calculations of perturbations in the orbit of Uranus by Leverrier,(178) the deviation of light by the sun's gravitation field by Einstein and the helical structure of DNA by Watson and Crick based on predictions by Pauling and Bragg.(179) Over a century ago Abbe(180) and Bjerknes(181) proposed that the laws of physics could be used to forecast the weather by integrating the governing partial differential equations, starting from the observed current weather.(182) Today complex forecasting is employed across society from the weather,(183) to financial markets,(184) to attempts to beat the bookie in football betting.(185)

In medicine prognosis has remained central to the art of clinical decision-making from the time of Hippocrates.(186) A clinician uses multiple predictors to estimate the probability of an outcome to support

subsequent management. Accurate identification of characteristics associated with an outcome may be crucial when comparing treatments, counseling individuals or designing a clinical trial.(187) However because of the biases of subjective estimates, physicians' ability to correctly predict outcomes is inconsistent and flawed.(188-192)

Prediction research investigates the ability of variables, to predict the presence, or absence, of a specific diagnosis (diagnostic), or future outcome (prognostic).(186, 193) Aetiological research aims to explain whether an outcome can reliably be attributed to a particular risk factor with adjustment for other causal factors - confounders. Prognostic research using a multivariable approach with positive or negative associations between exposures, termed variables or predictors and outcomes derived to predict the risk of a future outcome. All variables potentially associated with the outcome, not necessarily causally, can be considered in a prognostic study. Every causal factor is a predictor but not every predictor is a cause. Furthermore, not all causal factors end up being included in the most valid predictor model. To guide individual prognostication, analysis and reporting of prognostic studies should focus on absolute risk estimates of outcomes given combinations of predictor values. Relative risk estimates, including odds and hazard ratios, have no direct meaning or relevance to prognostication in practice. In contrast the calibration and discrimination of a multivariable model are highly relevant to prognostic research but meaningless in aetiological research.(186) Multivariable regression analysis quantifies the relation between two or more predictors and an outcome.(194) By combining independent predictors and assigning relative weighting a prediction model, also called a clinical prediction rule (CPR) or risk score can be produced. Such models are not meant to replace the role of the health professional,(195-198) rather, they are intended to help clinicians make decisions by providing objective estimates of probability as a supplement to other relevant information. Model performance is assessed in the original

development (or derivation) data set, with validation recommended internally and externally. Quantifying the effect of using a prediction model on patient and physician behaviours and outcomes is termed impact analysis. Though prognostic research has grown in prominence, evidence of impact on clinical outcomes has to date remained scarce.(199)

1.4.2 A brief history of prediction models

A pubmed search reveals the number of articles with the term ‘clinical prediction rule’ have rapidly increased since the turn of the century (figure 1.5). In the field of cardiovascular disease alone CPR articles increased from just over 100 a year in 2000 to more than 500 by 2012.(200)

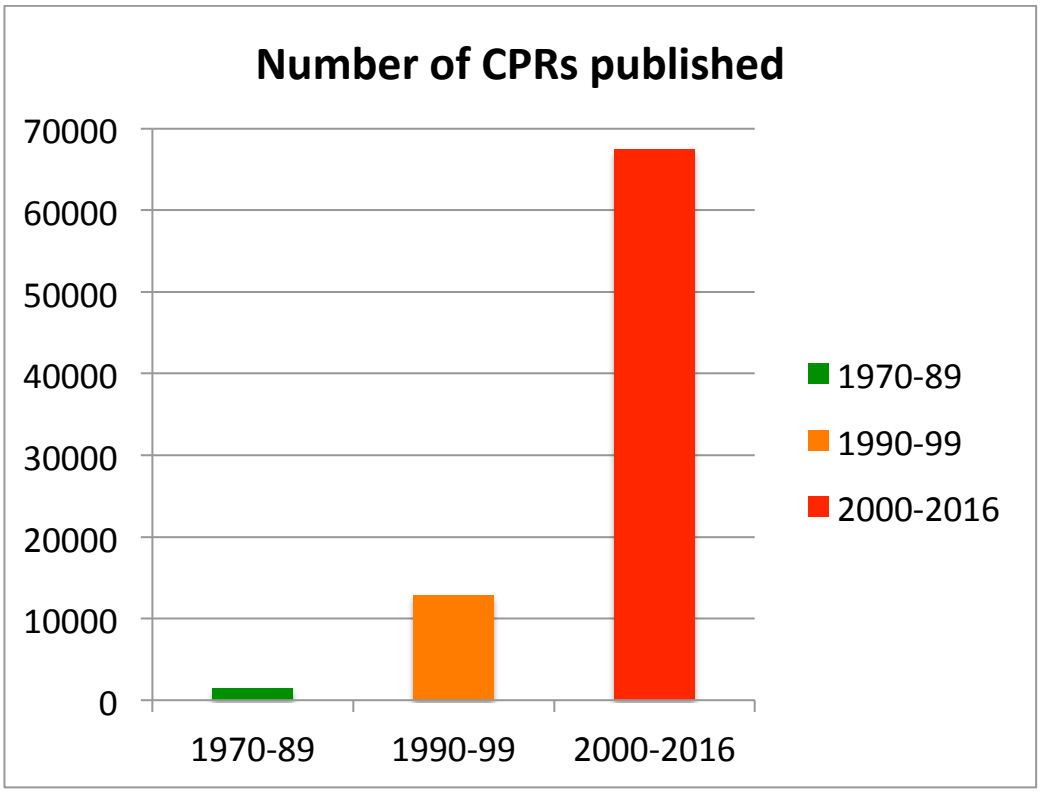


Figure 1.5 – CPR (Clinical prediction rule) studies increasing over time.

1950s	1970s	1980s	1990s	2000s	2010 and beyond
Apgar Score	Framingham Cardiovascular Strep throat - Centor	ICU Mortality - APACHE	Ottawa, Canadian MSK scores	Stroke risk - ABCD (TIA) - $CHADS_2$ (AF) Pneumonia – CURB65 Fracture risk - FRAX	Big data, Machine learning

Table 1.5 – Prediction models across the decades. AF – atrial fibrillation, ICU – intensive care unit, MSK – musculoskeletal, TIA – transient ischaemic attack.

In 1952 Dr Virgina Apgar devised arguably the earliest enduring prediction model which bears her name (table 1.5 highlights significant models over time).(201) Validated in 1958,(202) it remains in worldwide use, providing a rapid method of assessing the newborn infant. As with any good model, the Apgar is parsimonious, comprising heart rate, respiratory effort, muscle tone, reflex irritability, and colour. As with most scores there are limitations with inclusion of subjective components and can be influenced for example by medications and prematurity.(203) Two decades later the Framingham cardiovascular score was published,(204) and despite not having excellent prediction in all populations,(205) is one of the most widely used predictors of cardiovascular risk. In the same decade there was great interest in developing scores to differentiate a viral sore throat from streptococcal infection, to limit antibiotic use and need for investigations, whilst not missing a potentially serious, treatable infection. The most widely used of these scores, the Centor score, combines clinical exam (tonsillar exudates, swollen tender nodes, fever) with history (no cough). In the original description the score outperformed clinicians who tended to over rely on clinical examination.(206)

In the 1980s many complex scoring systems were described to predict ICU mortality, utilising multiple parameters, the most well-known being the APACHE (acute physiology and chronic health evaluation).(207) However, such scores are time consuming – for example the APACHE IV

score is said to take 37 minutes to calculate.(208) In the 1990s the Canadian group led by Stiell published a number of musculoskeletal (MSK) rules designed models with high negative predictive value,(209) to reduce unnecessary investigations.(209-212)

In the 2000s prediction models were described to improve stratification of patients at risk of stroke. First in patients presenting with transient ischaemic attack (TIA), to avoid unnecessary admission to hospital or premature discharge the ABCD score,(213) with subsequently updating is probably the most widely used.(214) Secondly, in patients with atrial fibrillation the $CHADS_2$ score was derived to influence the use of anticoagulation, with subsequent updates.(215) The CURB-65 score was one of a number of Pneumonia associated mortality scores combining age, a blood parameter (urea) and physiological parameters (confusion, tachypnea and hypotension).(216) This has subsequently been used by guidelines to influence place of care (consideration for ICU, or at the other extreme discharge home) and antibiotics (intravenous, dual or single agent).(217) In the ageing Western population, the need for accurate assessment of fracture risk led to the development of the FRAX score (fracture risk assessment tool) through the WHO.(218) Finally, the 2000s also witnessed the evolution of CPRs combining physiological observations as track and trigger systems in the acute hospital environment, to influence frequency of observations and seniority of personnel involved. For example the NEWS based on the ViEWS score,(219) is recommended by the Royal College of Physicians to standardise the physiological assessment of patients in hospital and has recently been updated.(220, 221)

The current decade has seen interest in embedding prediction models into the complex electronic technology increasingly central to healthcare practice. Furthermore potentially powerful techniques such as machine

learning, whereby computers improve through experience, are increasingly being reported, with examples including prediction of death after surgery,(222) or a MI.(223) Such tools are of great interest to multinational companies for example Google® and their healthcare research spin-off Deepmind.®

1.4.3 Indications and shortcomings of prediction models in stratified medicine

Stratified medicine refers to the targeting of treatments according to characteristics shared by subgroups of patients.(224, 225) Prediction models may offer advantages over pure clinical acumen. For example, statistical models can accommodate more factors than the human brain is capable of taking into consideration.(226) When given identical data a statistical model consistently gives the same result whereas human clinical judgment results in both inconsistency and disparity, especially with less experienced clinicians.(188-192, 227) Overconfident physicians tend to underestimate mortality risk, whereas those who lack self-confidence tend to overestimate mortality.(192) A number of prediction models have been shown to be more accurate than clinical judgment alone.(189, 228-230) Prognostic models can estimate absolute risk of an outcome for an individual; those with the highest absolute risk may derive the largest absolute benefit from treatment, when the treatment effect expressed in relative terms is the same for all.(224) For example, in the field of AKI in an acute clinical environment, with finite resources, such as staffing, this could allow restricted or personalised management to those who will benefit the most or who are at highest risk of an outcome. An accurate prediction model thus offers a number of potential applications:

1. Inform the individual patient about prognosis,
2. provide estimates of prognosis as a guide for ordering additional tests such as an expensive renal biomarker and/or therapies,
3. health technology evaluation: prognostic estimates derived

with/without results of a test to measure incremental prognostic information provided by that test,(231)

4. estimate the effect of a single factor such as a treatment on prognosis in an observational study with many uncontrolled confounding factors; to estimate the effect of the factor, efforts to control the simultaneous effects of the confounding variables must be attempted,
5. evaluation of a factor in a trial or cost effectiveness evaluation: e.g. decision analytic models require valid estimates of prognosis under different scenarios, based on treatment with/without knowledge of the predictor,
6. RCT design: who to randomise, the randomisation process using prognostic factors and adjusting for case mix when analysing results including health economics,(232)
7. test for differential therapeutic benefit or to estimate the clinical benefit for an individual in a clinical trial, taking into account the fact that low-risk patients must have less absolute benefit and,(233)
8. back translation: models that predict differential treatment response incorporation in practice guidelines tests may help define the disease.

Premature implementation of stratified medicine approaches however, may be harmful if people who might otherwise benefit from treatment are denied access. It is therefore crucial that the research that underpins this approach is robust. Unfortunately a number of methodological challenges and research deficiencies are well known to exist in the field.(234-236) A small proportion of prediction models have made their way into established clinical practice. Poor clinical uptake may be explained by five themes that will be explored in this thesis using AKI prediction as the case study:

1. Models are frequently not user-friendly and fail to take into account the continual, dynamic way in which clinicians gather

information,(226)

2. large number of models are available, with comparisons difficult to interpret,(234, 235)
3. a tacit acknowledgment clinicians do not know how to take advantage of such tools,(226)
4. the crucial importance of prognostication, that may explain reluctance for relinquish control of decision-making to a prediction model and,
5. lack of external validation and impact analysis leaving questions around generalisability and efficacy.

1.4.4 Stages of prediction research

Prediction model research can be broadly split along three lines: development (or derivation), validation and impact analysis (table 1.6). The final step is implementation into clinical practice.

Development	Identify predictors and assign relative weights Estimate performance: calibration and discrimination, potential for optimism (internal validation) and adjust for overfitting
Validation	Test performance in new participants Narrow: same institution or researchers at a later time Broad: other institutions, populations or wider inclusion criteria
Impact	Quantifying model impact on decision making, patient outcome and costs

Table 1.6 – 3 stages of prediction model research

1.4.5 Model Development

1.4.5.1 Introduction

Though there is no consensus on the ideal method for model development, a growing body of guidance can help avoid well-known design shortcomings.(237-240) Logistic and Cox regression are most frequently used for short-term and long-term outcomes, respectively.(238, 241, 242) Firstly, the relationships of interest, outcome and potential predictors should be specified, which should follow from clinical experience and previous studies.(243) Royston and colleagues advocate a number of preliminary steps to minimise bias and maximise reproducibility (summary table 1.7).(238) Of course there are other important considerations including assessing robustness of the model to influential observations and outliers, interaction between predictors, adjusting the final model for over-fitting and exploring reproducibility.

Objective(s)	Clearly defined
Sample	Include those at risk who require prognostication
Design	Prospective cohort enables optimal data collection Retrospective studies using cohorts already assembled for other reasons allow longer follow-up and are quicker, often at the expense of data quality
Predictors	Clinically relevant parameters e.g. demographics, socio-economic, history, examination, test results, previous treatment and functional state Define, standardised, reproducible and available at time when model intended to be used.(244, 245) Subjective predictors risk studying the observer's predictive ability
Data quality	Strategy to deal with missing values, handling continuous predictors and selecting important predictors
Outcome(s)	Selecting measure(s) of model performance relevant to patients; define period and methods of measurement; measure blinded to potential predictors to prevent information bias
Sample size	If number of predictors larger than outcome events, there is a risk of overestimating performance Suggested for each candidate predictor at least 10 events required.(243, 246-248)
Model performance	Selecting measures of performance including discrimination and calibration

Table 1.7 - Preliminary steps for model derivation.

1.4.5.2 Candidate predictors

Predictors already reported as prognostic would usually be candidates for assessment. For example, in AKI prediction, evaluating chronic co-morbidities including CKD would be sensible. Studies often measure more predictors than can easily be used and runs the risk of having too few events per predictor.(248) Predictors that are highly correlated can be excluded beforehand.(238) Frequently derivation studies employ univariate analysis to select for candidates for multivariate analysis: this implies multiple testing, underestimation of standard errors and P-values, a limited power to select diagnostically important predictors, and an unstable selection of predictors, resulting in overfitting.(194, 249, 250) Furthermore, this technique risks rejection of candidate predictors that may well be significant on multivariate analysis.(243, 251) Predictors need to be available at the time the model would be employed, for

example in AKI prediction at hospital admission not every test will be available, some would not be routinely requested, nor would a diagnosis necessarily have been made.

1.4.5.3 Quality of data and strategies to deal with missing data

Judgement is required to guide data quality assessment, with consistent measurement of predictors and outcomes. Most studies have missing predictor or outcome data, yet this is rarely discussed or statistically addressed.(193, 252, 253) Of particular relevance in AKI prediction is when previous renal function is missing. A complete-case analysis removes participants with missing data, which may be a reasonable strategy when few observations are missing.(240) However, this reduces sample size and may lead to biased results, with included cases potentially unrepresentative of the whole cohort.(254) Where a candidate predictor has a lot of missing data it may be excluded as this problem is likely to recur.(238) Rather than omitting individuals imputation techniques can be employed, with simple methods including use of the median or mean, however, this is biased and inefficient when predictors are correlated.(255) Kuhfeld decribed a general imputation method allowing predictors to be non-linearly related to one another, modified by Harrell.(243) Multiple imputation is preferred as single imputation, by estimating standard errors and p values too small, can increase chance findings.(254, 256) This technique creates multiple copies of the data set, with missing values replaced by imputed values drawn from their predicted distribution by using the observed data.(257, 258) Standard statistical analyses can be applied on each imputed data set, then combined to produce an overall estimate of each regression coefficient or model performance measure.(237)

Maximum information should be extracted from predictors; thus when there is a choice of a categorical or a continuous variable, the latter is preferred.(243, 259) For ordered categorical variables, such as stage of

CKD, collapsing of categories may be required. Harrell et al, suggested types of interactions that have frequently been found to be important in predicting outcomes and can be pre-specified (appendix table A1.2). No consensus exists on what constitutes the best method for selecting variables for inclusion.(238) Investigating a concise number of clinically sensible variables can minimse overfitting and aid reproducibility.(240, 243) A full model approach and stepwise selection methods have their own strengths and weaknesses (see appendix table A1.2 and text for further discussion).(194, 238, 240, 243, 260)

1.4.5.5 Sample size

In developing predictions based on 100 patients, it would be foolish to divide the patients into 50 subgroups and quote the average outcome for each subgroup, yet such instances with models using up to 50 variables in small populations exist in the prediction literature!(243) To enhance accuracy, variables must be limited or the model simplified unless the sample is large. Harrell et al, described data reduction methods and regression modelling strategies to yield reliable models,(187) and proposed a 'rule of thumb' to have predictive discrimination that validates on a new sample, no more than one variable per ten events should be examined in derivation.(243) This has been supported by subsequent researchers.(246-248)

1.4.5.6 Internal validation

Internal validation quantifies optimism using the original study sample. A popular approach is to randomly split the training data in two parts: one to develop and another to measure performance. However, this approach is weak and inefficient because not all available data are used to develop the model.(250) A more sophisticated approach may be to apply shrinkage techniques such as bootstrapping (figure 1.6) and cross-validation (appendix table A1.3.(237, 261, 262)

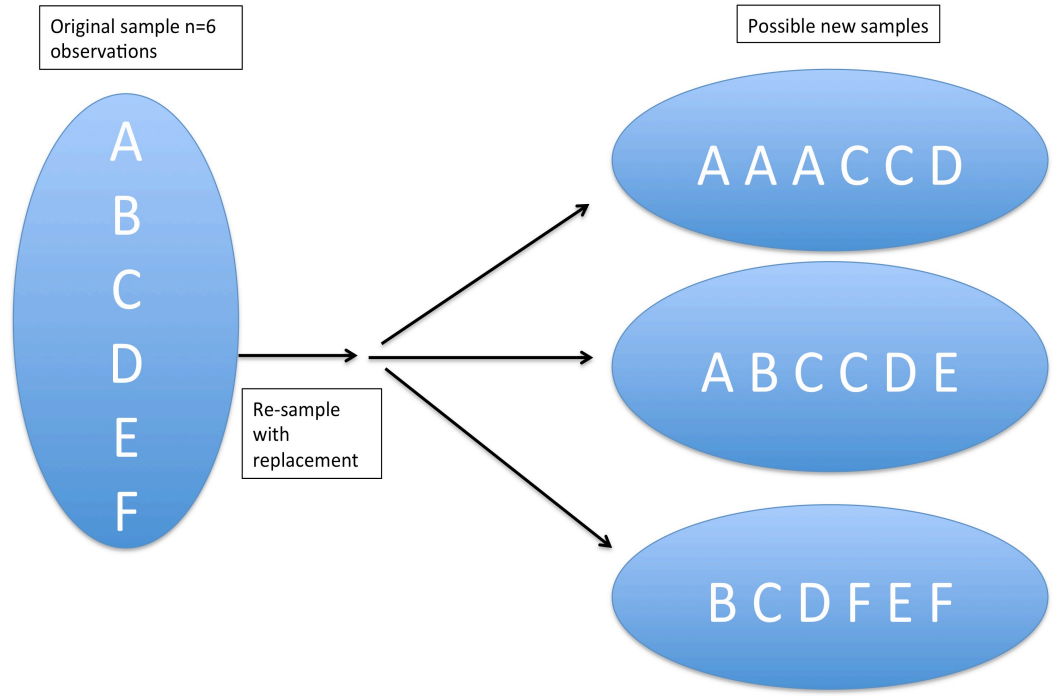


Figure 1.6 – schematic of bootstrapping for internal validation.

1.4.6 Statistics to assess performance

1.4.6.1 Discrimination

Performance of a logistic regression model is assessed by calibration and discrimination.(238) Discrimination is a measure of how well a model can separate those who do and do not have the outcome of interest.(263) If predicted values for cases are all higher than for non-cases, the model would be said to discriminate perfectly, even if the predicted risk does not match the proportion with disease. The area under the receiver operating characteristic curve (AUROC) allows assessment of predictive value over various cut-offs of probability of outcome, derived from plotting sensitivity vs. 1-specificity.(264) AUROC is used as a measure of discrimination related to a rank correlation between predicted and observed outcomes. The AUROC value represents the probability that a randomly selected

pair of cases, one with and one without the outcome, the case with the outcome will be ranked higher than the case without the outcome. The denominator is the number of patients with disease, multiplied by the number without disease. AUROC ranges from 0.5 (no discrimination, no better than chance) to 1.0 (perfect discrimination). In external validation the AUROC is sensitive to sample size and incidence of predicted outcome, requiring large sample sizes for adequate performance and comparison assessments.(265-267)

Because the AUROC focuses solely on accuracy without incorporating information on consequences it cannot tell us whether a model is worth using, or if more than model is being assessed, which is best.(268) For example, when a false-negative result is much more harmful than a false-positive result, a model with much greater specificity but slightly lower sensitivity than another would have a higher AUROC, but would be a poorer choice for clinical use.(268) The AUROC has a number of other drawbacks (discussion in appendix),(263) that must be recognised when performing prediction research in complex diseases or syndromes such as AKI with multiple causes and patient groups affected, where the outcome has yet to take place. In the case of an AKI prediction this is crucial to understand and consider when assessing: is the model robust, will validation succeed, what is the scope for improvement and if implemented, will the predictions offer sufficient utility to help improve outcomes? It is clear clinically that attempting to predict an outcome such as AKI, defined by a magnitude rise in SCr, is a complex process. Thus, accurately estimations of risk and classification into risk strata, may be the most pragmatic achievable goals.(269)

1.4.6.2 Calibration

Calibration refers to the degree of agreement between observed and predictive probabilities, and is often assessed by testing for lack-of-fit.(270, 271) Whilst discrimination measures a predictor's ability to

separate patients with different responses, calibration refers to the extent of bias; the degree of under or overestimation across the range and may reflect the degree of generalisability of the model. Calibration can be assessed by forming subgroups eg by deciles, ordered by expected probability and check for bias by comparing with observed responses. When the average predicted risk within subgroups of a prospective cohort, for example, matches the proportion that actually develops disease, we say a model is well calibrated.(263)

Investigators often use the Hosmer-Lemeshow (H-L) goodness-of-fit test for calibration of a prediction model.(271) This compares estimated-to-observed likelihood of outcome for groups of patients. In a well-fitting model estimated likelihood will be similar to observed likelihood and the H-L P value non significant (>0.05). However limitations of this test are well documented,(267, 272, 273) with non-significance found in small datasets and the opposite in large datasets. The later phenomenon was explained at length by Kramer and Zimmermann using large intensive care populations.(267) Furthermore, the test does not convey the indication of magnitude or direction of any miscalibration, hence there is a growing preference for calibration plots.(237)

When calibration is reported graphically predicted outcome probabilities, on the x-axis, are plotted against observed outcome frequencies, on the y-axis, commonly by tenths of predicted risk. This can be augmented by a smoothed (lowess) line over the predicted probability range.(240, 274) A plot displays direction and magnitude of model miscalibration across a probability range which can be combined with estimates of the calibration slope and intercept.(275) The calibration slope proposed by Cox can be calculated to quantify miscalibration,(276) being the regression coefficient β in a logistic model with the linear predictor as the only covariate: observed mortality = α + β linear predictor.(277) A well-calibrated model shows predictions lying on or around the 45° line of the calibration plot.

Perfect calibration shows a slope of 1 and intercept of 0, although caveats have been identified recently for example by Vach.(278) Calibration plots tend to show good calibration in the data set from which they were developed. They may be accompanied by a test for calibration intercept equals 0 and slope equals 1.(279) Comparing predicted versus observed outcome probabilities may also be in tabular form, commonly by deciles of predicted risk. Finally calibration plots may also be evaluated in relation to key predictors, for example, Bannister et al examined calibration regarding age and gender.(280)

Though calibration has been overlooked,(263) there will always be a trade-off between discrimination and calibration when assessing prediction models. Indeed it has been shown that a perfectly calibrated model, in which the predicted risk equals the observed risk for all subgroups, cannot achieve an AUROC equal to 1 in usual settings.(281) With the assumption of a uniform distribution of risk in the population it has been calculated that the maximum AUROC is 0.83 and this upper limit varies with the distribution of risk in the population.(282) Internal calibration refers to agreement between observed and predicted probabilities in the sample in which the model was developed. Lack of internal calibration is related to issues of lack of model fit and misspecification of the fitted logistic regression model.(274) External calibration refers to agreement between observed and predicted probabilities and may be suboptimal if the original model was overfitted, especially with small sample sizes, or when a model is tested on a different population with a different underlying hazard. For example, applying the Framingham model predicting cardiovascular disease in two countries or ethnic groups may require re-calibration by changing the intercept.(204, 205) There are a number of methods for model updating including updated regression coefficients, for example, using the slope of the calibration plot of the original model in the validation set, reporting the re-estimated intercept, or the estimated regression coefficients of the

model, including new predictors.(237)

The appendix discusses some of the more recent measures proposed to allow researchers to go beyond the AUROC when either comparing models, updating an existing model or adding a biomarker to a model. To date such techniques remain rarely employed with the majority of studies only reporting discrimination with or without measures of calibration.

1.4.7 External Validation

It is not enough to demonstrate good model performance on the development sample only, because most show optimistic results, even after corrections such as bootstrapping.(283, 284) Randomly splitting a single data set into development and validation data sets is frequent but is a weak, inefficient form of validation, because not all available data are used to develop the model.(250, 285) If the data set is large enough, a temporal split is a stronger approach and can be considered intermediate between internal and external validation.(237) Bootstrapping provides nearly unbiased estimates of predictive accuracy, however, only sampling variability is considered with no change in the patient population.(284) In external validation predictions are made using the original model and compared with those observed. Such studies differs from the original temporally or spatially, sometimes with different definitions, measurements, population cohorts or even the outcome event (table 1.8).(242, 283)

Unfortunately external validation studies are scarce (286) and when performed, the models tend to show reduced accuracy,(243, 284, 286) often due to a combination of statistical and clinical factors.(199, 242, 284) For example, a prediction model for bacterial infection in children presenting with fever performed disappointingly when applied to patients from another hospital in a later period, which the authors concluded could have been because of flaws in derivation with too few events in relation to

predictors tested and insufficient correction for optimism.(287) A second example pertains to the EuroSCORE (European System for Cardiac Operative Risk Evaluation), developed in Europe to predict mortality in cardiac patients.(288) When validated in Australia predictions were poorly calibrated.(289) The authors felt the reasons were likely to be multifactorial including different surgical indications, co-morbidities and the timeframe between derivation and validation. A further reason suggested for reduced performance in validation is such studies often include fewer events producing populations that may be different due to random variation.(285) It has been suggested that validation should contain at least 100 events and 100 non-events to detect substantial change in accuracy with 80% power.(265)

	Example
Temporal or Spatial	Rural, urban populations
Data	Different measurements, definitions
Care setting	Secondary vs. Primary Care
Participant	Race - White Caucasian vs. Afro-Caribbean
Outcome	Mortality vs. Length of stay

Table 1.8 – types of external validation.

When a validation study shows disappointing results, researchers often reject the original model and develop a new one from their own data.(290) However, the new model often also has limitations and multiple models for the same outcome create confusion. For example, there are over 100 published models for predicting long-term outcome in patients with neurotrauma.(291) Building on previous knowledge is a principle well recognised and used in intervention studies (e.g. meta-analyses of randomised trials) yet to date in the predictive research field this is rarely enacted. An alternative, recommended solution is to update existing models including the addition of biomarkers,(292) A number of techniques have been suggested in the literature for updating (appendix

table A1.4).(237, 271, 290, 292-294) For example, if a model was derived where the outcome had a higher prevalence, calibration of the rule in a lower prevalence area may be poor as a result of systematically too high predicted probabilities. By adjusting only the intercept of the original prediction model, poor calibration can be improved.(242, 292)

Validation like derivation requires assessment of discrimination and calibration and follow reporting guidance.(237) There is no consensus on acceptable performance, however, even moderately performing models are likely to do better than clinicians' own assessments.(295, 296) For example, the Framingham risk model has an AUROC of little over 0.70, yet is widely used. Even if performance is less good than in derivation, the model may still be useful and requires both contextualisation and clinical judgment.(297) Of note more complex models tend to give overoptimistic predictions when validation is attempted.(298) Finally, clinically valuable risk scores splitting population into for example three prognostic groups is often thought useful and validation can investigate whether the observed proportions of events were similar in these groups and whether separation in outcome was maintained.(242)

1.4.8 Impact analysis and implementation

Prognostic models are developed to provide objective estimates of outcome probabilities, with an underlying assumption such information can improve clinician decision making and consequently outcomes.(294) However, outcomes can improve only when the information provided by the model changes behaviour and management decisions.(186, 286) Studying these effects is termed an impact analysis. Despite the proliferation of prognostic models, there remains a dearth of such studies with more called for.(199, 293)

A number of study designs can be employed summarised in table 1.9 and further discussed in chapter 4 of this thesis.

Design	Explanation
RCT	Strong methodology, expensive, time-consuming
Cluster, stepped-wedge	Clusters of centres randomly allocated a time period with the intervention
Before-after	Prone to underlying time-dependent trends in outcomes
One-off impact	Outcome measured in alternating periods when prediction is present or absent
Before-after with control using difference in differences	Uses comparison group experiencing same trends and external drivers, but not exposed to intervention

Table 1.9 impact analysis designs. RCT – randomised controlled trial.

1.4.9 Why are prediction models not more widely used?

Prediction models are widely recommended within practice guidelines to inform decision-making.(299, 300) The use of high quality models has been shown to benefit patients, physicians, health care systems and reduce costs.(209, 211, 301, 302) However, most models suffer from significant methodological and reporting shortcomings, one explanation for a lack of implementation.(186, 237, 253, 293, 303, 304) For example, a systematic review of 71 studies published in 2008 in high impact journals described widespread limitations, including lack of validation and impact analysis.(193) Only with full and clear reporting of information can risk of bias and potential usefulness be adequately assessed. To address

these issues the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) initiative developed a set of recommendations and a checklist for the reporting of studies developing, validating, or updating a prediction model.(237) Whether such guidance has permeated the field of AKI research will form part of the presented thesis. In addition to shortcomings in model design, the lack of impact studies in many fields warrants caution in implementation into practice. For example a recent Cochrane systematic review of cluster-RCTs of interventions identified 15 studies suitable for inclusion.(305) The review concluded tailored implementation can be effective, but the effect is variable and tends to be small to moderate and that more research in the field was justified. Indeed there are a number of impact studies that demonstrated no benefit for example in falls prevention,(306) community-acquired pneumonia,(307) and induced abortion care.(308) This knowledge emphasises the need to rigorously assess models prior to widespread clinical implementation. A systematic review by Kawamoto and colleagues included 70 impact analysis studies and found four independent predictors of success (figure 1.7) Of 32 systems possessing all four features, 30 (94%) significantly improved clinical practice.

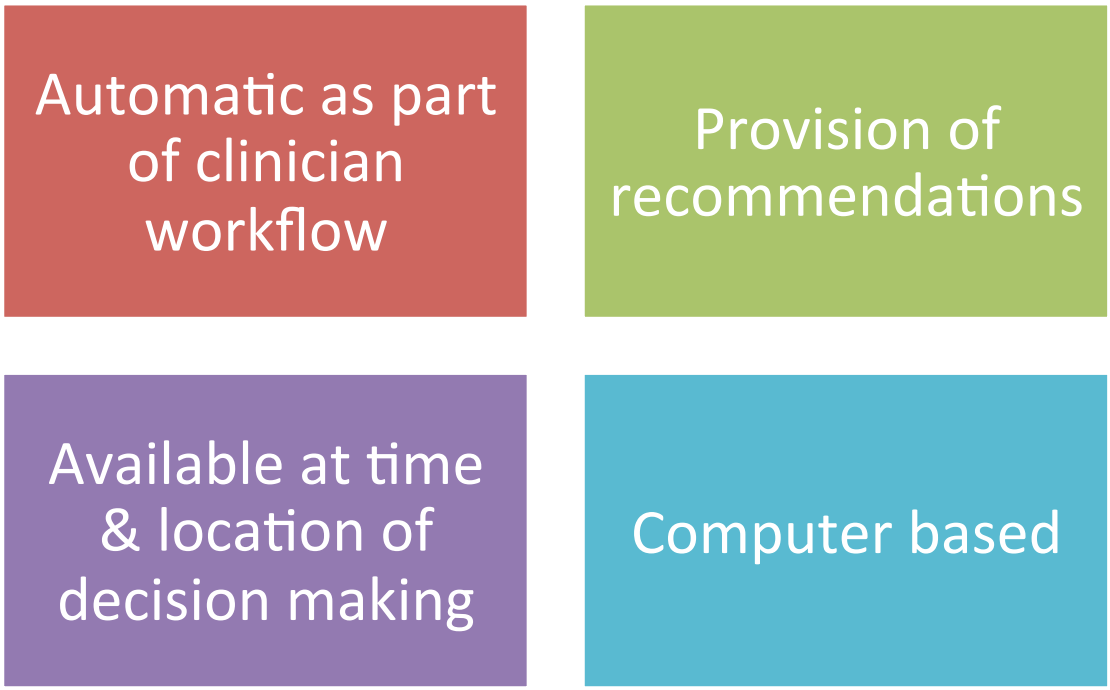


Figure 1.7 – Key factors to enable successful Impact of prediction models (from Kawamoto 2005) (309)

Some models have been widely accepted such as the Framingham cardiovascular risk score,(204) despite not necessarily having excellent prediction in all populations.(205) This may be because of high face validity, with clinicians trusting the model to guide practice rather than their own experience.(186) Other successfully implemented models such as the Ottowa Canadian MSK rules all arose from the same clinical researchers who gained international respect for their work, including impact analysis, which could further encourage uptake.(310-312) The Nottingham prognostic index for breast cancer is easy to use, maintains prediction in long-term follow-up and has been widely validated and revised over time.(313) There is a relative paucity of studies looking at the acceptability of prediction models though the researchers behind the Ottawa rules have produced an instrument to measure this (appendix table A1.5).(314) Table 1.10 summarises enablers and barriers to model implementation discussed.

Enablers	Barriers
Evidence of benefit, experienced research group – Ottawa MSK rules	Lack of evidence of benefit
Computer based decision support	Requires manual input and/or calculation
Ease of use at bedside, delivered at time & location of decision making	Complexity
Face validity – Framingham cardiovascular risk	Poor calibration at individual level
Automatic provision of decision support as part of clinician workflow	Ambiguous terms
Recommendations not just assessments	Changes in practice over time

Table 1.10 – Summary of enablers and barriers to implementation of prediction models. MSK – musculo-skeletal.

1.5 Summary and study aims

This introduction has described the increasing burden of AKI, its multifactorial nature and challenges both in recognising those at risk and implementing strategies to improve outcomes when AKI develops. The role of prediction modelling holds promise in this area, however, the majority of research studies to date have only described derivation of a model. For scientific rigour and clinical application it is imperative that external validation and impact analysis studies be performed. Information technology holds great promise in healthcare, with integration of prediction models one avenue of interest that has yet to be extensively explored. Finally future studies are likely to increasing investigate the effects of incorporating biomarkers with prediction models to further risk stratify patient populations.

The body of this thesis aims to explore prognostic prediction in the acute care setting through a case study of AKI. Building on the opening chapter that has set out the methodological considerations in the field the following chapters will cover:

1. A systematic review of the literature of hospital-acquired AKI prediction models, with an in-depth appraisal of models derived in general hospital settings,
2. external validation of one prediction model - the acute kidney injury physiological score (APS) and,
3. implementation and impact analysis of the APS as part of a complex electronic healthcare intervention.

Chapter 2. A systematic review of prognostic prediction models for acute kidney injury (AKI)

2.1 Summary of the chapter

Multiple prognostic prediction models across hospital settings have been published, to provide clinicians with objective evidence to identify those at highest risk of developing AKI. This systematic review searched Medline, EMBASE & Web of Science using recommended filters and appraised available models in the field of HA-AKI. 53 prediction model studies were found, over half in cardiac surgery (n=15) or contrast-induced AKI (n=15) with over 2 million patient episodes; 15 of the 53 models have external validation. A further 21 studies have been published purely as external validation studies of one of the 53 models. Only 11 models were described in general medicine and/or non-specialist surgical populations (474,478 patient episodes), five externally validated. No impact analysis studies were found.

Outcome definitions were heterogenous with a minority using consensus definitions. The median number of variables was 7 (interquartile range 5-10); most commonly age, CKD, diabetes, and heart failure. Definitions of CKD and establishment of baseline renal function differed markedly, with frequent use of admission SCr as baseline, thus potentially confusing community-acquired AKI with HA-AKI. Of the 53 models, in derivation AUROCs, to predict the outcome ranged 0.66-0.93 (median 0.80). Of 37 studies reporting internal validation AUROCs ranged 0.61–0.95 (median 0.80). No internal validation calibration information was available in 26/53 models. AUROCs dropped considerably in external validation: median 0.72 (range 0.57-0.93). Calibration information was only provided in 25/61 external validations; the Hosmer-Lemeshow p-value was <0.05 in 8 of the 16 studies where reported, suggesting lack of fit. Across the 11 general admission models AUROCs ranged 0.71-0.80 (median 0.75) for

derivation. Excluding studies using non-consensus definitions or those including admission SCr as predictor and/or baseline, left only four studies with AUROCs ranging 0.71-0.74 in derivation (n=3 studies), 0.67-0.76 for internal validation (n=3) and 0.65-0.71 for external validation (n=3). Only one model study presented a calibration plot for both derivation and validation, with re-calibration required in validation.

Conclusions

A large number of AKI prediction models have been described, predominantly in specialist areas. There are few externally validated models for general emergency admissions. Common deficiencies in methodology and statistical reporting were present, with handling of SCr (baseline function and use as a predictor) a particular concern.

2.2 Introduction - prediction models in AKI

Chapter 1 described the urgent need to improve outcomes associated with AKI and the use of prediction models / CPRs have been suggested to enhance identification of high-risk patients and allow preventive strategies to be employed.(17, 315, 316) Early models in the field were derived from small single-centre ICU studies, using risk of mortality associated with acute renal failure as the outcome, rather than development of AKI per se.(317-328) Currently, models for AKI defined as requiring RRT following cardiac surgery are the most robust, however, such events are relatively rare, limiting generalisability. Prediction models using a variety of definitions of renal impairment have been published in other specialist areas including non-cardiac surgery, following intravenous contrast (contrast-induced AKI, CI-AKI) and in general hospital emergency admissions. Systematic reviews have been called for in the field of prediction research.(329, 330) Four reviews have recently been published in the aforementioned specialised fields but these models may lack generalisability (table 2.1).

Study, year	Field	Studies reviewed
Huen *et al, 2012*(331)	Cardiac surgery	7
Silver *et al* 2015(332)	CI-AKI	12
Caragata *et al* 2016(333)	Liver transplantation	7
Wilson et al, 2016(334)	Non-cardiac surgery: general (n=1), liver transplantation (n=3), liver resection (n=2)	6

Table 2.1 Recent systematic reviews of AKI prediction models.

Two areas will be covered by this review: AKI prediction models in specialised fields including cardiac and transplant surgery and CI-AKI; second, models derived in general hospital populations.

2.3 Methods

The CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS)(335) and TRIPOD(237) guidance were followed, to frame the review question, determine details to extract and perform critical appraisal. The CHARMS checklist suggests seven items to focus the review question (table 2.2).

Item	Explanation
1. Study type (prognostic or diagnostic)	Prognostic prediction models
2. Scope	All published prognostic prediction models for development of AKI in a hospital setting; to inform risk stratification & potential uses in decision-making in different patient groups
3. Type of modelling	Model development +/- external validation in independent data; external model validation & model updating, if present
4. Target population	Adult (≥18) Patients in acute hospital environment
5. Outcome predicted	Development of AKI (or equivalent definition, including RRT) after an admission to hospital or procedure (such as surgery), or contrast
6. Time span of prediction	In-hospital development of the outcome
7. Intended moment of using the model	Pre-operatively to predict the risk of post-op AKI or need for RRT; at admission to risk stratify or guide therapy

Table 2.2 - CHARMS checklist for systematic review. AKI – acute kidney injury, RRT – renal replacement therapy.

2.3.1 Research question, rationale and objectives

What are the available prognostic prediction models for the development of HA-AKI in:

1. specialised populations and,
2. non-specialised populations?

AKI is an important independent contributor to mortality and international guidance has highlighted an urgent requirement to improve recognition and risk stratification.(3, 12, 143-145, 316) Prediction models could potentially contribute to this effort, however, as chapter 1 outlined clinical application has been hampered for a number of reasons:

1. the number of potential new predictors and prediction models continuously increases, but few models are externally validated,
2. new studies often find conflicting results on predictors, possibly due to small sample sizes and,(336)
3. reporting of individual studies' methodology and results has been substandard making conclusions and comparisons problematic.(193, 237, 241)

This review aimed to collate all empirical evidence for prediction models in AKI across hospital patient settings, fitting the pre-specified eligibility criteria for inclusion. Explicit, systematic methods were selected with a view to minimising bias and provide reliable findings from which conclusions can be drawn and decisions made.(337, 338)

The study objectives were to systematically search the available literature on multivariable prediction models for AKI (or equivalent definition of acute renal dysfunction) in hospitalised patients and the variables most commonly used. Each prediction model was assessed against published (TRIPOD) recommendations for reporting of prediction models.(237) Further analysis including a risk of bias assessment was performed in the models assessing general hospital populations.

2.3.2 Inclusion and exclusion criteria (table 2.3), identification and screening

Inclusion criteria
Papers in peer-reviewed journals reporting a prognostic multivariable prediction model (scoring system or algorithm) identifying patients who developed AKI, or other measures of renal dysfunction such as requirement for RRT in older studies
Validation studies (and update) of an existing model
Retrospective, prospective and case-control studies
Adults (≥18 years) in a hospital setting including cardiac surgery; non-cardiac surgery, general patients acutely admitted to hospital & patients receiving contrast
Statistical measures of discrimination (AUROC or or c-statistic)
Exclusion criteria
Research involving paediatric patients (<18 years old)
Non-human studies
Case reports & conference abstracts
Only logistic regression without a prediction model
Lack of discrimination statistics eg AUROC (unless model validated elsewhere)
Studies that investigated a single predictor, test, or marker
Studies that investigated only causality between one or more variables & an outcome
Use of patients already with the outcome (eg AKI already present on admission to hospital)
Patients in primary care
Renal transplantation
Novel, not widely available tests, such as biomarkers

Table 2.3 - Inclusion and Exclusion criteria of all prediction studies. AUROC - area under the receiver-operating characteristic curve.

Following the advice of a librarian experienced performing systematic reviews and existing guidance on database search strategies,(339) this review searched Medline, EMBASE and Web of Science databases with no language restrictions from inception to November 2015 and repeated in November 2016 using recommended filters for prediction studies (appendix tables A2.2-5).(340) Several search filters have been developed that can be used to find relevant prediction research studies that combine epidemiological terms related to prediction research and can be combined with disease specific terms to find relevant studies in electronic databases.(340) The sensitive filter for Haynes Broad Filter has been shown to retrieve 98% of clinical prediction rules and was used to maximise retrieval of articles.(341, 342) The terms searched in addition were “acute kidney injury,” “acute renal failure,” “AKI,” “ARF” and “contrast induced nephropathy.”

Titles and abstracts from all three databases were screened, full articles reviewed if thought to be eligible for inclusion and duplicates were removed. Reference lists were examined from retrieved articles, reference literature (systematic reviews, national(316) and international guidance(3)) and the authors own literature files were also analysed. The data extraction form was based on previous systemic reviews,(193, 253) and the CHARMS(335) and TRIPOD checklists(237) (summary table 2.4). Methodological quality of included studies was assessed using the 37 point TRIPOD checklist, a quality assessment tool specifically developed for assessment of prediction models.(237) To assess the influence of study quality on the results, a global score for each study was calculated, consisting of the sum of the scores for each individual criterion (score of 1 for criterion met, score of 0 for each criterion not met, or if it was unclear whether the criterion was met). For validation only studies the maximum score was 31.

It was anticipated that the study outcome, HA-AKI, would vary given the numerous definitions in use prior to KDIGO.(3) Thus, during the search strategy studies were included that recruited cases with a SCr at admission and repeated during a hospital admission to diagnose the outcome. To assess performance discrimination and calibration were recorded. In addition, ease of bedside use and whether the models could be electronically automated - factors known to influence successful uptake - were noted.(309) A quantitative synthesis of the models was not performed, being beyond the scope of review and formal methods for meta-analysis of prediction models are yet to be fully developed.

Summary of Data retrieved
Data source (years, retrospective, prospective; cohort, case-control, trial data)
Participants & setting (eg cardiac surgery, single or multi-centre, country
Primary outcome (and any blinding)
Candidate predictors (definitions; continuous data dichotomised & how selected for modelling?)
Sample size, EPP (including all predictors considered)
Type of model(s) evaluated - derivation, validation (internal, external)
Missing data, number included & excluded (criteria)
Type of model (eg full model approach), shrinkage
Incidence of outcome & mortality data
Variables in the model(s)
Performance: discrimination (AUROC or C-Statistic) & calibration (eg H-L P value, slope/curve), risk groups
Internal & external validation (in same study)
External validation studies with relevant performance measures
Additional resources, funding

Table 2.4 – summary of data extraction. AUROC - area under the receiver-operating characteristic curve, EPP – events per predictor, H-L – Hosmer-Lemeshow.

2.4 Results

2.4.1 Summary results of all retrieved studies

Following exclusions, 53 prediction models were included, with over 2 million patient episodes (figure 2.1).

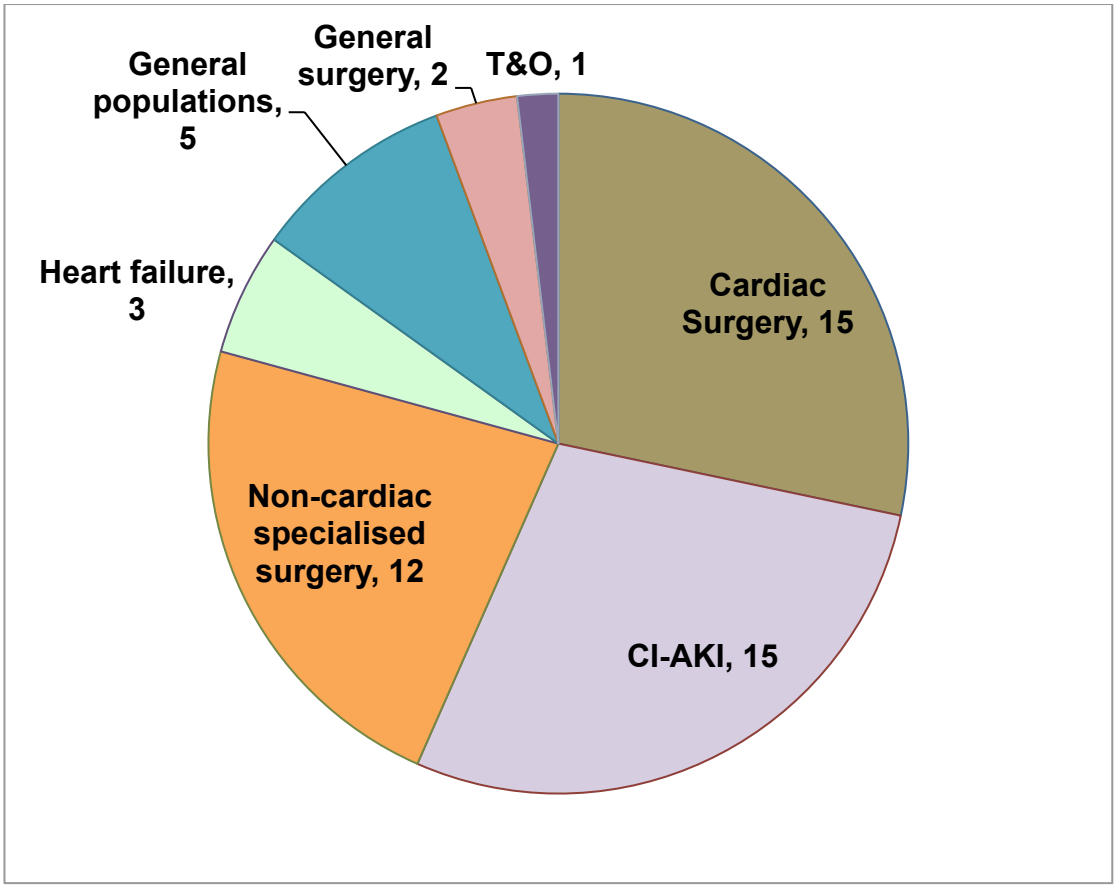


Figure 2.1 – breakdown of the 53 AKI prediction models across the different specialist fields. CI-AKI – contrast-induced AKI, T&O – Trauma and Orthopaedics.

Of the 53 studies 18 had external validation. A further 21 articles described the external validation of one, or more of the models, with a number of models externally validated on multiple occasions. 15 model and 11 validation studies have been published in cardiac surgery alone with 793,653 patient episodes. This compares to a review in 2012 which included seven models.(331) The field of CI-AKI accounted for 23 studies, including 15 prediction models and 1,416,982 patient episodes. A recent systematic review in this field included 12 model studies and four external validation studies.(332) Of 15 non-cardiac surgical models (257,956 patient episodes), ten were transplant studies (seven liver, two cardiac, one lung), two in liver resection, two general surgical cohorts and one in trauma and orthopaedics (T&O). Three models were in the context of patients presenting to hospital with heart failure. Finally, five general population studies at admission to hospital were reported (summary table

2.5). The median TRIPOD reporting score (out of 37pts) in derivation studies was 26 (IQR 25-30) and in those validating an existing model (out of 31pts) was 22 (20-24), suggesting significant shortcomings, increasing risk of bias.

Table 2.5 – summary of AKI prediction models. *Liver (n=7), Heart (n=2), Lung (n=1) transplantation, liver resection (n=2). **Models EV on >1 occasions. AUROC – area under the receiver operating characteristic curve, EV – external validation, IV – internal validation, H-L – Hosmer-Lemeshow test, T&O – trauma and orthopaedics.

	Field (n=)	Derivation AUROC median (range)	IV AUROC (n=)	IV AUROC	IV calibration (n=)	IV H-L	IV H-L <0.05	IV risk ranges	IV Plot/ slope	EV	EV AUROC	EV calibration (n=)	EV H-L	H-L P<0.05	EV risk ranges	EV Plot/ slope
Specialised populations	Cardiac Surgery (n=15)	0.80 (0.66-0.84)	9	0.80 (0.69-0.85)	8	4		5	0	9	0.76 (0.60-0.93)	7	13**	8**	1	0
	CI-AKI (n=15)	0.83 (0.69-0.93)	13	0.82 (0.61-0.95)	10	2		7	1	4	0.60 (0.57-0.86)	5	1		3	1
	*Non-cardiac specialist surgery (n=12)	0.81 (0.70-0.91)	8	0.79 (0.70-0.90)	3	2		1	0	0	-	0				0
General populations	General Surgery (n=2)	0.79 (0.77, 0.80)	1	0.8	1	0		1	0	1	0.66	1	0		1	0
	T&O (n=1)	0.74	1	0.73	1	0		0	1	1	0.7	1				1
	Heart failure (n=3)	0.74 (0.71, 0.76)	1	0.76	1	1		0	0	1	0.65, 0.65	1	0		1	0
	General hospital (n=5)	0.73 (0.72-0.75)	4	0.71 (0.66-0.76)	3	1	1	2	0	2	0.65-0.71, 0.71	2	2			1
	Total = 53	0.80 (0.66-0.93)	37	0.80 (0.61-0.95)	27	10	1	16	2	18	0.72 (0.57-0.93)	17	16	8	6	3

The first cardiac surgery study included patient data from 1987,(148) with 27 models using data preceding RIFLE.(17) The majority of studies were retrospective or retrospectively extracted prospectively collected databases or registries, though the exact design was not always clear. Most model studies were single-centre (n=33/53), with the USA accounting for almost half of the studies. In the model studies, no information on missing data was given in 20 (37%); cases were excluded in 22 (41%) with only three using imputation techniques. A median 24 (IQR 18-40) variables were analysed, though a number of studies only reported those significant on univariate or multivariate analysis.

Outcome was clearly reported, though blinding was only mentioned twice.(343, 344) Blinding of assessment of predictors, was mentioned in two derivation(343, 344) and no validation studies. However, predictor measurement was probably blinded for the outcome in 26 studies, simply because of the prospective design. Continuous variables were dichotomised in a quarter and univariate analysis used to select variables for multivariate analysis in 47/53 studies. A variety of selection techniques were described including backward, forward or both, with six using a full model approach. Only five studies mentioned sample size calculations. Events per predictor (EPP) was calculated for the predictors in the final model and for total number of predictors assessed. For the former, 9/53 studies had <10 EPP. However, from the reported candidate predictors 35 studies had <10 EPP. There was hetereogenity in outcome definitions: 16 studies reported incidence using consensus definitions and 13 used these definitions as the outcome, though time frame measured was variable. When the outcome was requirement for RRT this was predominantly (13/14) in surgical cohorts. No CI-AKI studies used current KDIGO CI-AKI definitions; most commonly a rise in SCr of >44 μmol/L, over a variety of timeframes was employed. The 53 models included 93 different predictor variables, including demographics, past medical or surgical history, procedure information, laboratory parameters, observations and admission diagnosis. The median number of variables included was 7 (IQR 5-10). Age, CKD, diabetes and heart failure were most frequently included. Anaemia was the most frequently employed blood parameter. In the

surgical studies type, urgency, shock and use of an intra-aortic balloon pump (cardiac surgery) were the most frequently included.

Discrimination and calibration are presented in table 2.5. In derivation median AUROC was 0.80 (range 0.66-0.93), with calibration reported in 33 derivations; 27 reporting the H-L statistic and one study used a calibration plot.(345) For internal validation AUROC ranged 0.61–0.95 (median 0.80) where reported (absent in 16/53). Shrinkage techniques such as bootstrapping, discussed in chapter 1, were employed in 15/53 models for internal validation. Where calibration was reported, 16 studies used a risk range or compared outcome in derivation vs validation in different risk groups; 10 reported the H-L statistic, 9/10 with a p-value >0.05 suggesting satisfactory calibration; two studies used a calibration plot.(345, 346) 18 of the 53 models have been externally validated, 10 at least twice, predominantly in cardiac surgery and CI-AKI fields. Median AUROC dropped to 0.72 (range 0.57-0.93). Excluding the cardiac studies where RRT was predominantly the outcome median AUROC in external validations was 0.66. Calibration information was available for 17 models (some more than once) in the form of risk groups (n=5) or H-L statistic (n=16), with half of these reporting a p-value <0.05 suggesting inadequate calibration. A calibration plot was provided in three recent studies,(345, 347, 348) with one suggesting re-calibration was required.(345)

2.4.2 Specialist Prediction Models

2.4.2.1 Cardiac Surgery studies

15 model studies (667,622 patient episodes) were found in the field of cardiac surgery, with a further 11 external validation studies; nine models have been externally validated at least once,(82, 148, 150, 349-354) five more than once.(148, 150, 349-351) Table 2.6 highlights the number of different outcome definitions used.

Outcome	Studies
RRT	(148, 150, 349, 350, 354)
AKI (RIFLE, AKIN or KDIGO) or RRT	(82, 355-357)
Increase in SCr to >176.8 μmol/l if baseline <132.6 μmol/l, or a 50% increase if SCr 132.6-265.2 μmol/l, <7 days post-op	(351)
SCr ≥176.8 μmol/l & increase ≥61.9 μmol/l from baseline, or RRT or evidence of renal failure at autopsy	(344)
duration of AKI defined as an increase ≥26.5μmol/l or ≥50% SCr from baseline	(353)
2 of: increase of SCr to >200μmol/L or doubling of SCr over baseline or RRT	(352)
RRT & composite of doubling of SCr or RRT <2 weeks	(358)
Fall in eGFR to <30 ml/min/1.73m^2	(359)

Table 2.6 - Outcome definition for the 15 Cardiac prediction models.

RRT was required in 1.6% of the 21 derivation or validation studies reporting this outcome. In-patient mortality was 51% (n=1,464, range 14-66%) for patients requiring RRT and 2.5% (n=98,369, range 0.9-2.6%) if not (recorded in 7 studies). In 12 studies where information was limited to patients requiring RRT, mortality was 53.5% (range 12.5-83.3%). Where AKI was the primary outcome, in-hospital mortality in those with AKI ranged 9.8-29.2% (3 studies). The median number of variables used was 9 (8-11), most frequently CKD (n=14), heart failure (n=11), age (n=10) and diabetes (n=10). The large number of models including type of surgery, use of adjuncts, such as intra-aortic balloon pumps and intra-op events, such as inotropes and cardio-pulmonary bypass (CPB) duration, are shortcomings when attempting to implement a prediction model *before* an intervention. Median AUROC in derivation was 0.80 (range 0.66-0.84), 0.80 (range 0.69-0.85) in 9 models with internal derivation and 0.76 (range 0.60-0.88) in external validation studies. Calibration was missing in 7/15 derivation, 7/15 internal validations and 23/39

external validations. In derivation studies median TRIPOD score (out of 37) was 27 (IQR 26-31) and in 11 purely external validation studies (out of 31), median score was 24 (22-25). There were significant shortcomings, for example, no studies gave information on blinding assessment of the predictors or the outcome; only three mentioned sample size, two discussed model updating and missing data was not mentioned in eight. In summary multiple, similar, have been described, most frequently predicting need for RRT, which is rare but has high associated mortality. To date there have been no reports of implementation or impact analysis.

2.4.2.2 Non-cardiac specialist surgery

Ten transplantation populations and two liver resection studies were included with no external validation (table 2.7 summarises outcome definitions). A specific potential benefit of a prediction model in organ transplantation could be tailoring peri-operative immunosuppressive drug strategy according to the degree of (renal) risk. Of the transplant models, six used RRT as the outcome, with an overall incidence of 7% (n=1,935/27,473). In the four liver transplant studies reporting mortality in those requiring RRT mortality was 22% (n=39/178) compared to 4.1% (n=42/1,033) without RRT.(360-363)

Type surgery	Number studies	Outcome
Liver resection	2	RIFLE(364, 365)
Liver Transplantation	7	RRT(360-363) SCr >133 μmol/L & increase >50% or RRT<7 days(366) RIFLE increase SCr <1 month post-op(367) KDIGO 72hrs(368)
Heart Transplantation	2	RRT(369, 370)
Lung Transplantation	1	RRT(371)

Table 2.7 - Non-cardiac specialist surgical studies. RR T- renal replacement therapy, SCr – serum creatinine.

The median number of variables included was 7 (IQR 4-8), most commonly intra-op events, CKD and diabetes. In four liver transplant studies the model for end stage liver disease (MELD) score (incorporating bilirubin, SCr and international normalised ratio) was employed on its own,(362) or in combination with other variables.(361, 363, 367) The study by Sanchez *et al*, included post-op ICU stay as well as combining SCr with the MELD score (which includes SCr) which does not make clinical sense.(361) Median derivation AUROCs 0.81 (range 0.70-0.91) with internal validation median 0.79 (range 0.70-0.90), though was missing from four. Internal validation calibration was presented in three studies (H-L test in two, risk range in one). No specialist surgical models had external validation. Median score on the TRIPOD checklists was 25 (IQR 23-28). Blinding strategies, sample size explanations and model updating were not reported. In summary, a number of AKI prediction models have been described in non-cardiac surgery specialist areas, without external validation. Though these areas are of interest in light of the high-risk cohorts and issues of nephrotoxic medications, no impact analysis studies exist that could help clinicians decide whether implementation is warranted.

2.4.2.3 Contrast-induced AKI (CI-AKI)

AKI following administration of contrast (CI-AKI) has previously been referred to as contrast induced nephropathy,(372) with a recent definition implemented to align with KDIGO definitions.(3) Alterations in renal haemodynamics and tubular toxicity are considered potential factors in the pathogenesis of CI-AKI,(373) which is associated with adverse outcomes.(374, 375) This review retrieved 15 model studies and eight purely externally validations of four existing models,(343, 376-378) using a variety of outcomes (n=808,570 patient episodes). The majority of external validations (n=10) were of the Mehran model.(343) 14 models derived from general percutaneous coronary intervention (PCI) populations, with one using a cohort undergoing PCI for ST-elevation myocardial infarction (STEMI). Of the eight purely external validation studies, six were in the STEMI cohort and two in the general PCI population.

One model used RRT as the primary outcome, required in 0.4% (60/14,624), with an in-patient mortality of 35%.(379) In the general PCI studies incidence of CI-AKI was 6.7% (range 0.75-32.6%, variable definitions). In the six studies reporting mortality, of those who developed CI-AKI 9% died in-hospital vs. 0.5% of patients without the outcome. In the seven STEMI CI-AKI studies the outcome developed in 16% (n=882/5,515); mortality was 21.5% (n=26/121, n=3 studies) who developed CI-AKI vs 0.01% (n=7/652) in those who did not. The most frequently included predictors were: CKD (n=14), diabetes (11), CCF (11), age (10), hypotension or cardiogenic shock (8) and contrast volume or type (7) (table 2.8). A number of studies include predictors related to the procedure, limiting clinical utility, if risk is to be predicted *before* intervention.

Variable	Number of times used
Diabetes	12
CKD	12
CCF	11
Age	9
Hypotension / Cardiogenic Shock	8
Contrast volume/type	7
IABP	6
MI / location	5
Anaemia	5
Urgent/Emergency	4
PVD	4

Table 2.8 – Variables used in CI-AKI Prediction models in order of frequency. Abbreviations - CKD = Chronic kidney disease, CCF = congestive cardiac failure, IABP = Intra-aortic balloon pump, PVD = Peripheral vascular disease, MI = Myocardial infarction.

Median AUROCs were 0.83 (range 0.69-0.93) in derivation, 0.82 (0.61-0.95) internal validation (reported in 13 models) and 0.60 (0.57-0.86) on external validation (n=16). Calibration was not reported in eight derivation and five internal validation studies; of the latter two reported the H-L test, two risk ranges and one a calibration slope; only five external validation reported calibration. TRIPOD checklist median score (out of 37) for the studies with a derivation model was 26 (IQR 22-30) and for validation studies (out of 31 points) was 21 (20-22).

In summary this group represent a heterogenous population, yet as with cardiac surgery, multiple model studies have been reported using similar predictors. A significant reduction in discrimination in the limited number of external validations and frequent absence of calibration are significant limitations. Finally no impact analyses have to date been reported.

Further details of all 42 models studies in specialist areas can be found in appendix tables A2.12-13.

2.4.2.4 Summary of specialist AKI prediction models

Since the first cardiac surgery prediction model in 1997(148) numerous models have been published in this field, CI-AKI and transplantation populations. Despite different populations and excluding some surgical specific predictors, the most frequently included predictors were CKD, diabetes, age and heart failure. Appraisal against recommended reporting suggested shortcomings in the majority of the studies;(237) including reporting of blinding, handling of missing data, sample size, absent calibration and model updating. Validation studies rarely showed a comparison of the distribution of variables with the development data. The majority of surgical studies used RRT as the outcome, though few gave criteria for initiating it; furthermore this requirement is in most populations rare, at the most severe spectrum of AKI. As few used consensus definitions, making comparisons and drawing conclusions around generalisability are problematic.

Even in groups with heterogenous populations discrimination was moderate, including studies with small numbers of events that carry the risk of optimism and indeed in models with multiple external validations, very few maintain discrimination performance. Indeed with the 2005 Thakar cardiac surgery model, the only model with multiple external validations showing discrimination AUROCs $\geq$0.80, performance dropped when predicting AKI by recent consensus definitions used by Heise (0.66),(380) Kim (0.63)(356) and Birnie (0.70).(82) Although a number of models have been externally

validated, none have undergone published impact analysis, nor is there evidence they are used in clinical practice. Table 2.9 summarises the major shortcomings of these models.

Increasingly large number of similar models
Lack of model updating
Reduced performance where externally validated
No impact analysis studies or evidence of implementation
Reporting including methodology and performance

Table 2.9 Summary of shortcomings in specialist AKI prediction models

2.4.3 General models

Of particular interest to the presented thesis were models derived in general hospital populations routinely seen in both specialist and non-specialist hospitals. The remainder of the systematic review is narrowed to include only AKI prediction models in these settings.

2.4.3.1 Methods

This section utilised the search methodology already described and used the same data extraction form for the specialist models. Further information was extracted on definitions of baseline renal function, handling of community AKI cases, whether SCr was used as a predictor in analysis and finally the magnitude and timeframe used to define the outcome. The same performance statistics were retrieved as for the specialist models; in addition ease of bedside use and potential electronic automation were recorded.(309) A global TRIPOD score quantified reporting and furthermore quality (risk of bias) was assessed by piloting a version of PROBAST (Prediction study Risk Of Bias Assessment Tool), nearing completion and ready for piloting when this review was undertaken (Wolff R, Whiting P, Mallett S, et al, personal communication, website: http://s371539711.initial-website.co.uk/probast/). Elements were considered in the following domains: participants, predictors, outcome, sample size, missing data, statistical analysis and overall judgement of bias and applicability.

2.4.3.2 Results

From 14,046 articles identified by the initial search strategy, 254 full articles were reviewed (PRISMA flow chart, figure 2.2). Following the exclusion of the specialised fields models (n=42) left 11 general model studies (n=474,478 patient episodes), in general surgery,(70, 71) T&O,(345) general hospital cohorts (predominantly medicine and surgery),(84-86, 381, 382) and heart failure (summarised in table 2.9 and appendix table A2.6).(383-385) Two further studies were purely external validations.(386, 387) HA-AKI incidence was 7% (21,641 events), though this varied from <1% in the general surgery models,(70, 71) to 28% across the heart failure studies with heterogeneous definitions (timeframe and marker) employed (table 2.10 for definitions, further information in appendix table A2.6). For example, five studies took admission SCr to represent baseline function, potentially confusing CKD, established CA-AKI and emerging AKI.(71, 86, 382, 384, 385) One study produced a model to predict admission AKI as well as HA-AKI at 72 hours, with the former not considered suitable for analysis in this review.(85) In seven of nine studies reporting age, this was significantly higher in the group with the outcome, with eight reporting a mean or median age over 65 years in the outcome group (table 2.10). Mortality was significantly higher in those who developed the outcome in the six studies where data were available (ranging 6-42%).

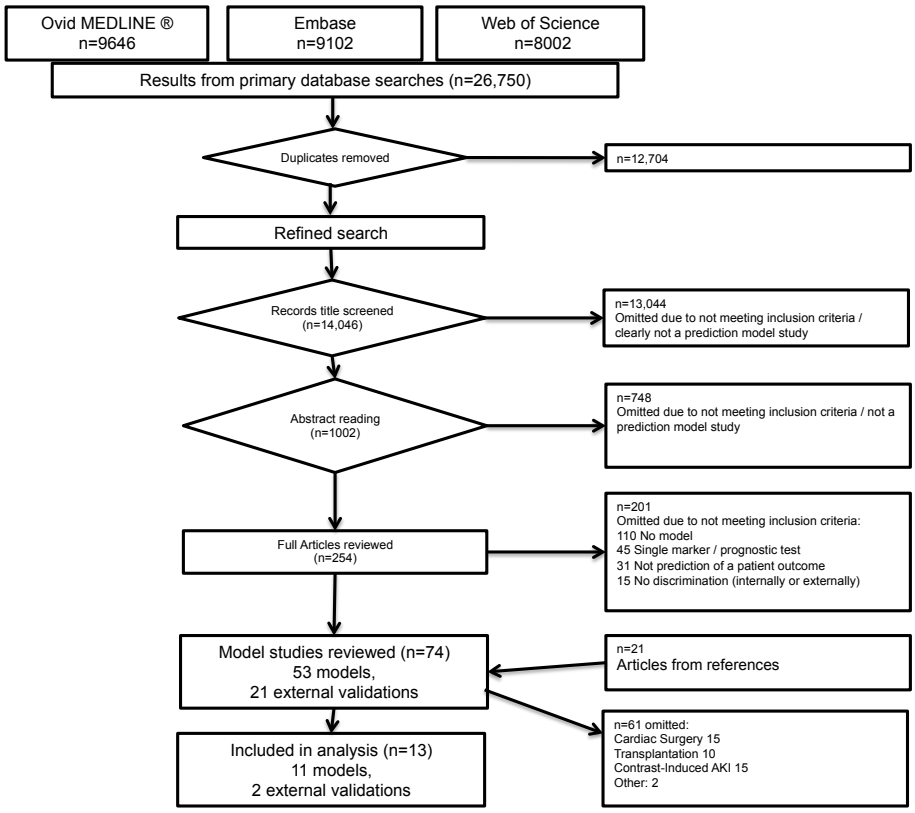


Figure 2.2 PRISMA flow chart for general prediction models.

Table 2.10 Summary of HA-AKI prediction models in general populations.

Population	**General Surgery**		**T&O**	**General (Medical & Surgical)**					**Heart failure**		
Author, year (n=derivation)	**Kheterpal** 2007 (n=14,066)	**Kheterpal** 2009 (n=57,080)	**Bell** 2015 (n=6,220)	**Drawz** 2008 (n=360)	**Matheny** 2010 (n=26,107)	**Koyner** 2016 (n=202,961)	**Bedford** 2016 (n=7,556)	**Forni** 2013 (n=1,867)	**Forman** 2004 (n=1,004)	**Breidthardt** 2011 (n=657)	**Wang** 2013 (n=1,010)
Centres, Design	1, R	121, R	3, R	3, CC	1, R	5, R	3, R	1, P	11, R	1, P	1, R
Age (with outcome)	59	65 (±15)	77 (±11)	67	-	70 (±16)	-	80 (70-86)	68.7	79 (72-85)	73 (67-78)
Age (no outcome)	47	54 (±17)	70 (±16)	63	-	63 (±19)	-	73 (61-81)	66.8	79 (70-85)	71 (63-75)
Outcome predicted	eGFR <50 (<7 days)	↑SCr ≥177μmol/l, RRT (30 day)	KDIGO ↑SCr	↑SCr*	RIFLE ↑SCr	KDIGO ↑SCr (24hr)	KDIGO ↑SCr (72hr)	KDIGO ↑SCr (<7 days)	↑SCr >26.5μmol/l**	↑SCr >26.5μmol/l**	AKIN SCr (<48hr)
Events	121	561	672	120	1,352	17,541	222	95	271	136	341
Mortality with outcome	15%	42%	-	-	-	6%	-	20%	27%	17%	17%
Mortality no outcome	3%^	8%^	-	-	-	1%	-	4%	-	6%	2%
Predictors tested	30	19	11	19	23	29	45	25	29	48	35
Predictors included	7	9	7	7	27	29	12	7	4	3	8
EPP	4	30	61	6	59	605	5	4	9	3	10
Inappropriate handling of SCr	X	X		X	X	X			X	X	X
Derivation AUROC	0.77	0.80	0.74	0.73	0.75	-	-	0.72	-	0.71	0.76
IV AUROC	-	0.80	0.73	0.66	-	0.74	0.67	0.76	-	-	0.76
EV AUROC	0.67	X	0.71	X	X	X	0.71	0.65-0.71#	0.65, 0.65	X	X
Derivation Calibration	RR	RR	Plot	-	H-L P=0.29	-		H-L P=0.96	-	-	H-L P=0.98
IV Calibration	-	RR	Plot	RR	-	-	H-L P=0.04	-	-	-	H-L P=0.13
EV Calibration	RR	-	Plot	-	-	-	H-L P=0.12	H-L P=0.06-0.09, Plot	RR	-	-
TRIPOD items	25	28	34	26	28	24	29	29	26	23	30
Bedside calculation	-	-	-	-	-	-	-	Yes	Yes	Yes	Yes
Electronic Automation	-	-	Yes***	Yes	-	Yes	-	Yes	Yes	Yes	Yes

Design: R - retrospective, P – prospective, CC – case-control, Mortality - In-hospital. AUROC – area under the receiver operating characteristic curve, Plot – Calibration plot, EPP – Events per predictor, EV – external validation, H-L – Hosmer-Lemeshow test, IV – internal validation, RR – risk range, RRT – renal replacement therapy, SCr – serum creatinine, T&O – Trauma & Orthopaedics, TRIPOD – how many of the 37 recommended items were reported. *Increase sCr ≥44μmol/L if baseline SCr of ≤168μmol/L, ≥88μmol/L baseline 177-433μmol/L & ≥133μmol/L baseline >442μmol/L. **During admission, ***Used linked community and hospital data, ^Propensity matched, [#]validations in medicine/surgery with/without baseline SCr.

Area of concern	Description
Missing data	Imputation recommended to avoid bias, rarely described(237, 254)
Definitions of outcome & predictors	No consistent strategy to differentiate CA-AKI from HA-AKI; 2 studies excluded patients with CKD;(70, 381) 5 took admission SCr as baseline; 5 included SCr as predictor despite it forming outcome; Co-morbidities inconsistently defined
Blinding predictors or outcome	Not reported
Sample size	Calculations not described, 6 studies had <10 EPP, increasing risk of overfitting & underfitting(237, 248)
Univariate to select for multivariate analysis	Not recommended, used in 10/11 models(237)
Bootstrapping	Adjustment for optimism rarely described(345, 385)
Calibration plots	Present in 1 model & 1 external validation(237, 345, 387)
External validation & updating	Validation adjusts for optimism, assesses generalisability. but was scarce; updating not described(237)
Newer performance measures	Techniques such as decision curve analysis offer insight into clinical consequences - not described(268)
Use of data linkage	Only one study utilised data linkage(345)

Table 2.11 - Summary of limitations in methodology and reporting. CA-AKI – community-acquired AKI, CKD – chronic kidney disease, EPP – events per predictor, HA-AKI – hospital-acquired AKI, SCr – serum Creatinine.

Study reporting

The median recommended reporting items was 28 (IQR 25-30, out of 37) suggesting significant shortcomings (summary table 2.11, with TRIPOD reporting summarised in appendix table A2.8). Eight studies were retrospective, two prospective and one a case control. Five studies were single-centre; the USA (n=6) and UK (n=3) accounted for the majority. Only three studies used imputation techniques for missing data.(71, 345, 382) Definitions were heterogenous with five using RIFLE,(381) AKIN,(384) or KDIGO criteria for changes in SCr (table 2.10).(84, 85, 345) One study used KDIGO SCr change within a 24-hour timeframe of predictors being measured.(86)

Candidate predictors, model building and sample size

A median of 29 (IQR 19-35) predictors were considered, though frequently studies only reported those significant on univariate or multivariate analysis. Blinding of predictor assessment and outcome was not mentioned. Continuous predictors were dichotomised in four and ten studies used univariate analysis to select for multivariate analysis. Two models employed bootstrapping,(345, 385) with no study reporting sample size calculations. Median number of outcome events was 271 (121-672). All of the studies had >10 EPP *included in the model,* however the EPP was <10 in six studies, when accounting for the total number of candidate predictors assessed.(70, 84, 85, 382, 383, 385) Of 56 predictors a median of 7 (7-12) were included, such as demographics, past history, procedure information, laboratory parameters, physiological observations and hospital admission diagnoses (most common in figure 2.3, full details appendix tables A2.9-10). Only 4 studies included physiological parameters.(84, 86, 382, 385) Seven studies included admission SCr as potential predictor, with five including this in the final model, potentially confusing *prediction* with a *diagnosis* of AKI.(71, 86, 381, 384, 385) Each study's handling of SCr in terms of when a baseline was calculated (prior or at admission) and whether SCr was used as a predictor are summarised in appendix table A2.11.

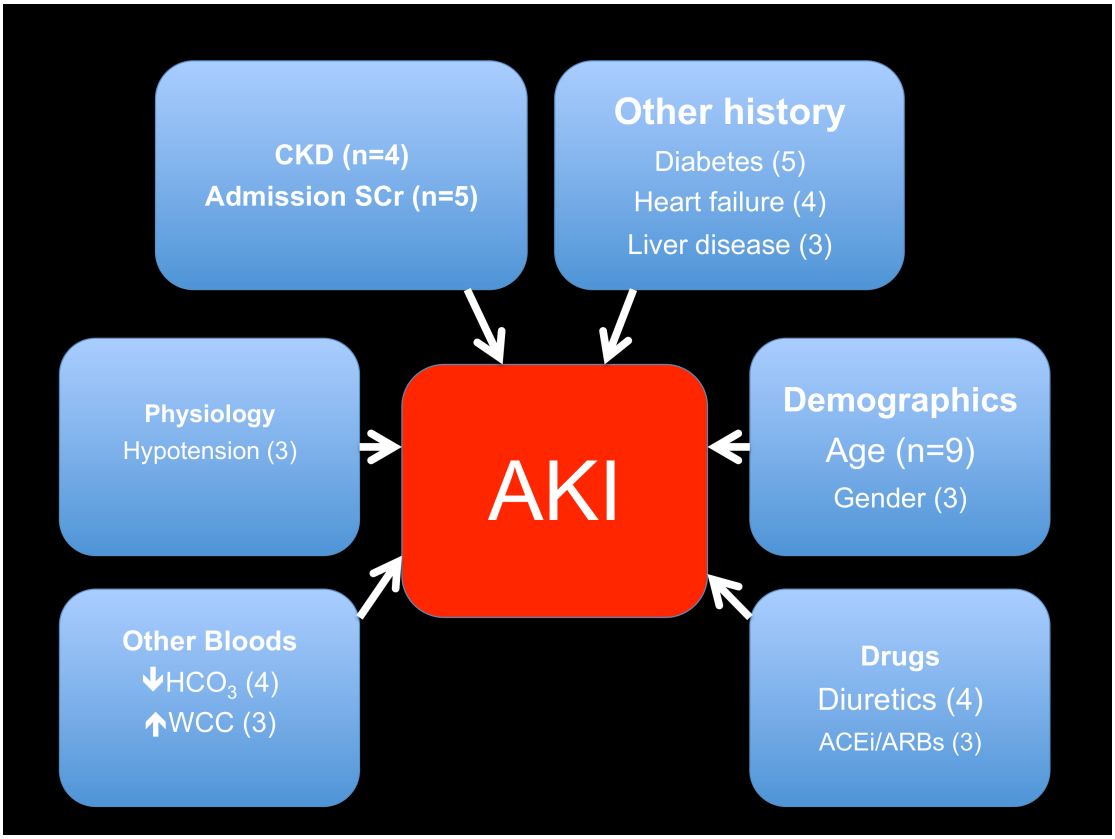


Figure 2.3 ACEi – Angiotensin-converting enzyme inhibitors, ARBs – Angiotensin-receptor blockers, Bloods – laboratory parameters, CKD – chronic kidney disease, ↓HCO_3 – reduced serum bicarbonate, SCr – serum creatinine, ↑WCC - raised white cell count.

Model performance (table 2.10)

Median AUROC was 0.745 (range 0.71-0.80) for derivation (n=8), 0.74 (range 0.66-0.80) for internal validation (n=7) and range 0.65-0.71 for five models externally validated. Excluding studies using non-consensus definitions and those including admission SCr as predictor and/or baseline, left only four studies.(84, 85, 345, 383) with AUROCs ranged 0.71-0.74 in derivation (n=3), 0.67-0.76 for internal validation (n=3) and 0.65-0.71 for external validation (n=3). Only one model study presented a calibration plot for derivation and validation and required re-calibration as the model over-predicted the outcome.(345) The H-L statistic was used in three derivations,(84, 381, 384) and two internal validations.(85, 384) Of those reporting the H-L p-value, only the Bedford internal validation study reported a value <0.05 indicating suboptimal calibration. In this study the model under-predicted AKI in the lowest-risk groups whilst over-predicting risk in the upper risk group.(85)

Five models have been externally validated: on separate populations within the same study;(85, 345) other model studies;(385) or stand alone external validations.(70, 387) One validation provided a calibration plot,(345) one the H-L statistic,(85) and one reported both.(387) In the external validation of the Forni study calibration plots showed agreement at low probability rates whilst at higher rates calibration deviated in the medical cohort.(387) Two of the three surgical models have been externally validated: the Kheterpal model,(70) in a Chinese population (AUROC 0.66),(386) and the UK T&O study used a third centre for external validation.(345) Two of the mixed general population models have external validation,(84, 85) the latter having been derived on medical patients and externally validated in medical and surgical cohorts.(387) The first of the three heart failure studies was externally validated in the subsequent two studies with inferior discrimination (AUROC 0.65 in both validations).(383-385) No model updating or impact analysis were reported.

Quality assessment and risk of bias summary

Quality assessment based on a draft version of the PROBAST tool suggested evidence in 9/11 studies a high risk of bias with significant shortcomings (summary table 2.12). For example, one study used a case-control design which is inappropriate for developing a prediction model as it does not enable calculation of absolute risks and thus yields incorrect estimates of model intercept or baseline hazard.(335) A wide variety of predictors were considered with use of univariate analysis to select for multivariate in 10/11 of the studies. Six studies were potentially underpowered having <10 EPP. Seven of the studies introduced potential bias in handling of renal function and SCr either in failing to establish a reliable baseline renal function, excluding patients with reduced renal function, or employing it as a predictor. Finally, outcome definition frequently varied, in part owing to a number of the studies preceding consensus definitions.

Table 2.12. Risk of bias summary based on PROBAST (Prediction study Risk Of Bias Assessment Tool, permission from Wolff R, personal communication).

Population	**General Surgery**		**T&O**	**General (Medical & Surgical)**					**Heart failure**		
Model, year	**Kheterpal (2007)**	**Kheterpal (2009)**	**Bell (2015)**	**Drawz (2008)**	**Matheny (2010)**	**Koyner (2016)**	**Bedford (2016)**	**Forni (2013)**	**Forman (2004)**	**Breidthardt (2011)**	**Wang (2013)**
Study participants	?	?	+	+	+	?	+	+	?	?	?
Predictors	?	?	?	?	-	?	-	+	-	-	?
Outcome	-	-	+	-	-	-	+	+	-	-	-
Sample size & missing data	-	+	+	-	-	?	?	?	-	-	?
Statistical analysis	-	-	+	-	+	-	-	?	-	-	-
Overall judgement of bias	**-**	**-**	**+**	**-**	**-**	**-**	**-**	**+**	**-**	**-**	**-**
Overall judgement of applicability	-	-	?	-	-	-	+	+	-	-	-
Usability of the model	+	+	+	+	+	+	+	+	+	+	+

Study participants domain - design of the included study, and inclusion and exclusion of its participants; Predictors domain - definition, timing, and measurement of predictors (also assesses whether predictors have not been measured and were therefore omitted from the model); Outcome domain - definition, timing, and measurement of predicted outcomes; Sample size and missing data domain - number of participants in the study and exclusions owing to missing data; Statistical analysis domain - methods (eg appropriate presentation of discrimination and calibration). Red = “high”, Green = “low” or Amber = “unclear” risk of bias.

2.4.4 Discussion

2.4.4.1 Principal findings

In this first systematic review of HA-AKI prediction in general hospital settings, the most common predictors were age, diabetes, CKD, drugs (ACEi/ARBs and diuretics), heart failure, (admission) serum creatinine and bicarbonate. Modest discrimination is unsurprising when attempting at a single time point to predict a future event reflecting diverse aetiologies and affecting heterogeneous patient groups. Significant shortcomings mirror those described elsewhere:(193, 252, 253, 388)

- multiple similar models, rarely externally validated,
- no impact analysis or evidence of clinical implementation,
- incomplete reporting, including calibration with a high risk of bias and,
- little consideration of electronic automation, allowing presentation without additional data input beyond usual clinical care, which influences uptake.(309)

Methodological and reporting shortcomings in the studies included six studies having <10 EPP potentially leading to overfitting, with only three employing multiple imputation to handle missing data which can increase sample size and power (summary table 2.11).(237, 248, 254) Handling of SCr and CKD was of particular concern. First, in part due to a previous lack of a consensus definition, the outcome - HA-AKI - had heterogeneous definitions, in magnitude of SCr rise and time-frame. For example, the Kheterpal study (2009)(71) used a rise in SCr ≥177 μmol/L which has been shown to significantly underestimate rates of AKI when compared with more recent definitions.(389) Koyner et al used a rolling timeframe of 24 hours, whilst others used SCr elevation at any point during admission.(86) Indeed one study produced a separate model to predict AKI at admission to hospital.(85) This was further confused by seven studies inclusion of admission SCr as a potential predictor, with subsequent inclusion in five models, five studies taking admission SCr to represent a patient's baseline and two studies excluding all patients with a reduced admission eGFR from analysis. This

risks confusing prediction and detection of AKI events. Issues with definitions have been described before in systematic reviews of prediction models and should be considered when researchers embark on future studies.(390, 391) A formal risk of bias assessment suggested the majority of studies had domains placing the studies at high risk of bias (table 2.12). Published after TRIPOD, Bell and colleagues' model provides researchers with a good template for adherence to reporting guidance, with a low risk of bias and demonstrates the utility of data linkage (for example between community and hospital), though lack of validation in other populations tempers recommendation for implementation.(345)

2.4.4.2 Strengths and limitations of this review

This review summarises available AKI prediction models in general populations who account for the majority of hospital admissions and AKI cases.(169, 392, 393) Selection following an extensive literature search and critical appraisal.(237, 335) The large number of patient episodes provides important insights into AKI prediction complementing recent reviews in sub-specialist areas, also presented in the first part of this review.(331-334) In-patient mortality in those who developed the outcome ranged 6-42% (in the six studies reporting mortality) emphasising this is a crucial group to promptly identify.

The first limitation is the small number of externally validated models, which tempers recommending a model. Second, though the review aimed to include general populations, caution should be employed, for example, when comparing a model derived on Heart Failure patients to one from an Orthopaedic cohort. However, in many UK hospitals, such populations share similarities being a predominantly elderly demographic with co-morbidities and if one aim of a prediction model is generalisability, a model should be tested in these different fields. Third, as study outcome definitions and handling of SCr (baseline and as predictor of outcome) were heterogenous, model

comparisons are problematic, though recent studies were more likely to use KDIGO SCr change. Fourth, no studies included urine output, probably reflecting the small number of patients outside the ICU environment who have this marker closely monitored. Fifth, TRIPOD recommendations were used as a reporting benchmark, however, the relative importance of individual items and what constitutes an acceptable 'score' is arguable, though a formal PROBAST risk of bias assessment was also carried out providing further insight. The absence of impact analysis limits the recommendation of one model over another. Finally, a meta-analysis was not performed without access to individual participant data (IPD). Expert guidance now exists in this area and offers opportunities to improve the scope of external validation research.(391, 394)

2.4.4.3 Comparison with previous systematic reviews

Both this study and a review of CI-AKI models found pre-existing predictors - age, CKD, diabetes and heart failure - to be the most commonly included.(332) A Cardiac surgery review reported specialty specific predictors in addition to these chronic co-morbidities. A non-Cardiac surgery review (5/6 studies in liver transplantation or resection) reported age, CKD and diabetes in at least two models.(334) Finally, a liver transplantation review highlighted the importance of CKD and unsurprisingly degree of liver dysfunction.(333) The present review found drugs or acute laboratory values frequently included, though only five models included acute physiological parameters. This review and the non-Cardiac surgery review included adherence to recommended TRIPOD reporting with similar shortcomings. Across the other reviews, only in the fields of CI-AKI and Cardiac surgery were external validations reported.(331, 332) Ease of use, including if necessary a calculator and potential for electronic automation were rarely considered. No impact analysis studies have been described.

2.4.4.4 Future directions

Management of HA-AKI presents a significant challenge, that could be helped by robust prediction models to risk stratify, encourage prevention and prompt recognition, key healthcare priorities.(166, 315) Appraisal and synthesis of prediction studies may enable clinicians and policymakers judge model utility however, this is problematic when key study details go unreported.(237) Though much of the AKI literature is on (often assumed) hospital-acquired AKI, the majority of cases arise from the community.(87, 91) Indeed, a recent study demonstrated a significant proportion of such patients are never hospitalised.(92) This review suggests even in HA-AKI, the strongest predictors are pre-existing patient factors. The two laboratory measures frequently included – serum creatinine and bicarbonate – may also reflect a chronic component. It is likely a proportion of cases classed as HA-AKI represent evolving community cases, thus, models using such pre-existing risk factors makes clinical sense. This continuum of harm between community and hospital suggests that a risk prediction model in place at, or even *before* hospital admission, combined with early flagging of those who have met AKI criteria, may be required to improve outcomes.

Electronic linkage of patient records between community and hospital data is desirable to ensure accurate inclusion of predictors: from the community comprehensive data on chronic morbidity and medications, and acute in-hospital laboratory and physiological parameters. This may also enable bedside automation as part of clinical workflow, which as chapter one discussed can encourage beneficial implementation.(309, 395) Acute physiological parameters rarely assessed overall in the models, could be an avenue of future research to improve modest performance at a single time point. As hospitals increasingly employ electronic track and trigger observation systems, this may enable the application of complex statistics (e.g. machine learning) to account for the effects of trends and repeated measures. Risk stratification using chronic comorbidity and medication(s) with trends in physiology, could be further enhanced by measurement of urine output and/or newer biomarkers. To date research has relied on retrospective

databases often only at a single time point. Future research would thus require prospective collection of rich data to achieve accurate prediction modelling demanded by clinicians and patients prior to implementation.

Impact analysis in prediction research is sparse making it difficult to recommend model implementation alongside, or replacing, usual care.(199) Potential areas for impact analysis (summarised in table 2.13) include in specific populations how a model could influence location of peri-operative care of surgical patients or drug and/or contrast dosing in patients with heart failure. Second, in a wider hospital setting, the effects of highlighting those at highest risk to teams (ward, outreach critical care or Nephrology) with an adequate effector arm. This has been demonstrated by existing AKI alerts in *established* AKI where outcome benefit has been limited to patients who had best practice delivered.(141, 170, 396) Third, as healthcare embraces complex technology, the inclusion of physiological (including urine output) or laboratory trends may significantly improve model performance and effects on outcome. Fourth, by identifying a high risk group, further risk stratification could be achieved by employing one of the increasing number of renal biomarkers,(66) or response to an intervention such as a frusemide stress test.(397) Finally, one external validation study found patients high risk on the prediction model who developed AKI had a higher rate of mortality than the low risk group who developed HA-AKI, indicating the model predicts disease severity.(387) This could allow early review of such patients to help inform whether escalation of care may be required, or indeed be appropriate in the increasing number of frail elderly patients admitted to hospitals.

Table 2.13 Potential areas for future Impact analysis of general AKI prediction models.

Population	Impact analysis to inform Clinical use
General Surgery	Peri-operative: haemodynamic targets, place of care, drugs, contrast delivery
Trauma & Orthopaedics	
General Populations	Risk stratification of large populations: e.g. influencing intensity of observations, remote monitoring, application of biomarkers in subgroups at high-risk
Heart failure	Optimise haemodynamic status: diuretic dosing, use/volume of contrast

To conclude this chapter, improving the management of patients to prevent AKI, or reduce associated complications, is a global priority. This review suggests there are few externally validated prediction models to help identify those at risk of AKI across general hospital populations, with the majority of models in specialised fields. Future research should concentrate on model development and validation, combining primary care data with acute electronic patient records (EPR), with subsequent exploration of electronic implementation to enable clinical uptake and impact analysis.

Chapter 3. Predicting AKI in emergency admissions: an external validation study of the acute kidney injury prediction score (APS)

3.1 Chapter summary

Chapter 1 highlighted the importance of HA-AKI and how prediction models may identify those at risk of its' development. Chapter 2 suggested a dearth of validated models in general hospital settings. This third chapter presents an external validation of one model: the **A**cute kidney injury **P**rediction **S**core (APS) in a single UK non-specialist acute hospital (2013-2015, 12,554 episodes). Four cohorts were studied: adult medical and general surgical populations, with and without a known pre-admission baseline serum creatinine (SCr). Model performance and reporting followed expert guidance.

Results: HA-AKI incidence within 7 days (KDIGO change in SCr) was 8.1% (n=409) of medical patients with known baseline SCr, 6.6% (n=141) in those without a baseline, 4.9% (n=204) in surgical patients with baseline and 4% (n=49) in those without. Across the four cohorts AUROCs were: medical with known baseline 0.65 (95% CIs 0.62-0.67) and no baseline 0.71 (0.67-0.75), surgical with baseline 0.66 (0.62-0.70) and no baseline 0.68 (0.58-0.75). For calibration, in medicine and surgical cohorts with baseline SCr, Hosmer-Lemeshow p-values were non-significant, suggesting acceptable calibration. In the medical cohort, at a cut-off of 5 points on the APS to predict HA-AKI, positive predictive value was 16% (13-18%) and negative predictive value 94% (93-94%). Of medical patients with HA-AKI, those with an APS ≥5 had a significantly increased risk of death (28% vs 18%, odds ratio 1.8 [95% CI 1.1-2.9], p =0.015).

Conclusions: this external validation of the APS found moderate discrimination and acceptable calibration to predict HA-AKI. The model may be useful as a severity marker when HA-AKI does occur. Harnessing linked data from primary care may be one way to improve risk prediction.

3.2 Introduction

Chapter 1 revealed a continuum of injury exists before loss of excretory kidney function can be measured with the most commonly employed laboratory test, SCr. This has encouraged investigators to explore the potential role of prediction models to highlight patients before AKI is established.(195-198) Though most prediction research involves derivation studies, external validation is crucial to address overfitting and generalisability.(199, 237, 241, 242, 286) Chapter 2 revealed the majority of HA-AKI models focus on specialist groups such as cardiac surgery, with few in acute medicine or surgery.(84, 85, 381, 382) One such model is the acute kidney injury prediction score (APS), derived in acute medicine patients in a single UK centre which uses a combination of co-morbidities and acute physiological variables derived from a hospital electronic data system (Appendix Table 3A.1 for variables). Discrimination assessed by the AUROC to predict HA-AKI within 7 days was 0.72 in derivation and 0.76 in an internal validation cohort (without a baseline SCr).(84) As this was deemed a promising model in an area of general hospital interest further assessment was thought appropriate.

This chapter describes a study which externally validated the APS in a general medical population with a known baseline SCr. The presence of this baseline accounts for patients with CKD allowing more confident conclusions as to whether acute deterioration in renal function has already occurred at admission: community-acquired AKI (CA-AKI). The study also validated the model in:

I. a general medical population without a known baseline SCr and,
II. general surgical populations with and without baseline SCr.

The presence or not of a baseline SCr remains a conundrum for researchers and clinicians. Those with previous blood tests are more likely to have chronic disease that precipitated previous testing and/or monitoring of renal function. Equally those without a baseline SCr may have undiagnosed disease that would not lead to an algorithm alerting to new AKI. For example, a patient with

a SCr of 250μmol/L would not be flagged as an AKI case by the NHS algorithm, though a clinician in the right context would interpret such a value as probable AKI if acute illness was present. Published reporting guidance was followed.(237)

3.3 Methods

A retrospective observational cohort external validation study of the APS (see Appendix Table 3A.1 for variables and weightings) was performed on the adult medical and surgical units of St Richard's Hospital site of Western Sussex Hospitals NHS Foundation Trust (WSHFT), for the period March 2013 to February 2015. St Richard's is a separate hospital site from Worthing hospital where in 2011 the APS was derived on general medical patients. There was no cross-site contamination of staff, though catchment populations for the two sites, 20 miles apart, are similar. Surgery at the Trust includes general surgery, urology and trauma and orthopaedics, but does not include major trauma, neurosurgery, cardiac, major vascular or transplantation. (See Appendix Table 3A.2 for clinical and demographic values for the derivation and presented external validation cohorts).(84) Ethical approval was given by NHS Research Ethics Committee London - South East (REC reference 13/LO/0884).

At admission, all acute medicine and elderly care, emergency and elective surgery routinely had physiological observations measured and entered via handheld systems into the clinical data software system (Patientrack© Sydney, NSW, Australia). Previous ICD-10 electronically coded history (heart failure, liver disease and diabetes mellitus) were retrieved and CKD was defined as an eGFR <60mls/minute prior to admission using a National algorithm that has been shown to perform well.(175) KDIGO Criteria for AKI was employed (SCr increase of ≥1.5 from the admission value or ≥26.5μmol/l within a rolling 48 hours during the first seven days of the patients admission). The cohort included only the first admission of an individual during the study period, in contrast to the derivation study that included patients with more than one admission. Patients were included if they were ≥18 years of age, stayed

at least one night in the medical or surgical division over the 24 month period and had at least one SCr repeated.
Exclusion criteria:

- patients with AKI on admission (defined using KDIGO change from baseline SCr or absolute SCr value ≥354μmol/l),
- patients moved directly from the A&E department to the Intensive Care Unit (ICU) (as neither area uses the Patientrack© data system),
- aged <18 years,
- obstetrics and gynaecology admissions or,
- discharged without spending a night in hospital.

Patients were followed-up until discharge from hospital, or death during the in-patient spell. The analysis was performed for the patients' first admission where more than one occurred in the study period. As all patients admitted had an APS calculated automatically by the Patientrack© system before any outcome had occurred, there were no missing demographic or physiological data at admission. As past history relied on previous coded events (from attending hospital), potentially such data could have been missing. None of the researchers involved in analysis of the data were involved in the management of the patients. The research team members responsible for data analysis had access only to the fully anonymised individual level data and were blinded to other patient data, as well as to the components of the calculated scores in the hospital information system. Since there is no consensus on how to determine what counts as an adequate sample size in such studies, all available (19,276) hospital cases for the period were included in the analysis.(237)

Following logistic regression analysis, discrimination was assessed by the AUROC,(237) and calibration by the Hosmer-Lemeshow (H-L) test and graphically by plotting predicted probabilities (x-axis) against the observed event rate (y-axis) of the outcome and deriving a linear function (an intercept of zero and a slope of one indicating perfect calibration).(237) Predictive values and likelihood ratios were calculated to further inform on the way

model performance could impact on clinical workload. Following extraction, all data were fully anonymised on Microsoft® Excel® and analyses performed on SPSS® v22.

3.4 Results

From an initial 19,276 patient episodes in medicine including elderly care and surgery over a two-year period 12,554 episodes were analysed (n=7,170 in medicine, n=5,384 in surgery) after excluding cases with CA-AKI (n=782) or with no repeat SCr (n=5,940) (figure 3.1). Over a quarter of patients (n=3,329) had no baseline SCr. Table 3.1 summarises the clinical and demographic data of the groups. Incidence of HA-AKI was:

- Medical patients with a known SCr baseline 8.1%
- Medical patients without a known baseline SCr 6.6%
- Surgical patients with a known baseline SCr 4.9%
- Surgical patients with no known baseline was 4%

The medical cohort were significantly older than the surgical cohort: 78 years (IQR 65-86) vs 67 (51-77), respectively (P≤0.001). There was a higher frequency of morbidity, particularly heart failure in the medical population, who also had longer hospital stays. In-patient mortality was increased in those with HA-AKI across all 4 groups, but in absolute terms this was greatest in medical patients; in those with a baseline SCr mortality was 21.5% in those who developed HA-AKI vs 4.5% in those who did not (p≤0.001).

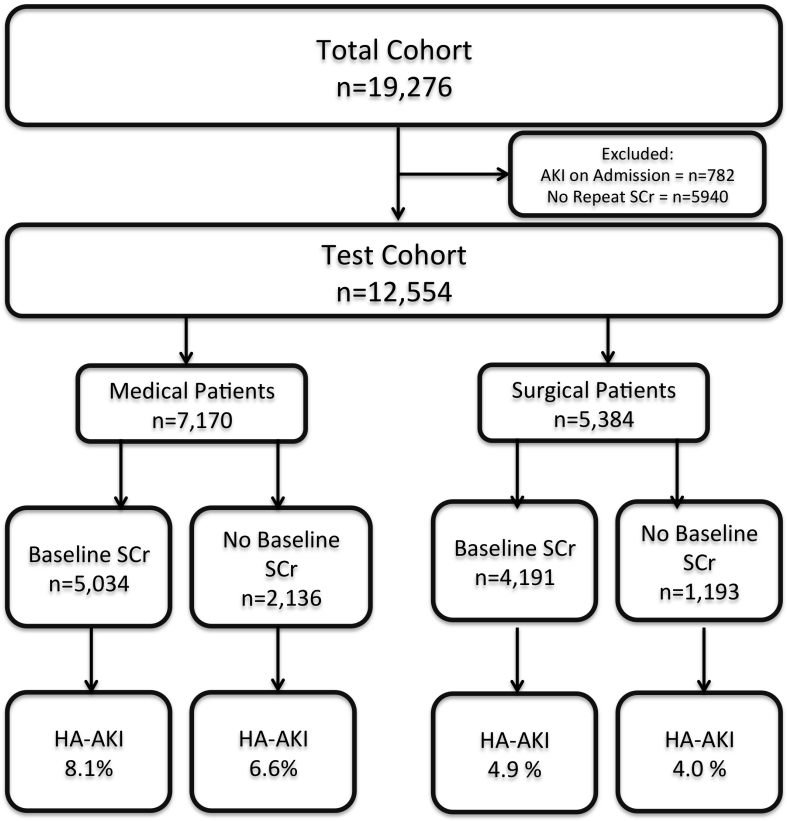


Figure 3.1 Consort study flow chart of external validation study participants. HA-AKI – hospital acquired acute kidney injury. SCr – serum creatinine.

Table 3.1 – Clinical and demographics data of the 4 cohorts.

	Medicine with baseline SCr (n=5,034)			**Medicine no baseline SCr (n=2,136)**			**Surgery with baseline SCr (n=4,191)**			**Surgery no baseline SCr (n=1,193)**		
HA-AKI	HA-AKI 8.1 % (n=409)	No HA-AKI (n=4,625)	P value	HA-AKI 6.6% (n=141)	No HA-AKI (n=1,995)	P value	HA-AKI 4.9% (n=204)	No HA-AKI (n=3,987)	P value	HA-AKI 4% (n=49)	No HA-AKI (n=1,144)	P value
Mortality*	21.5% (n=88)	4.5% (n=209)	**<0.001**	23% (n=32)	4% (n=72)	**<0.001**	5.9% (n=12)	0.4% (n=15)	**<0.001**	8.2% (n=4)	0.9% (n=10)	**0.002**
LOS	12 (6-21)	6 (3-12)	**<0.001**	11.5 (5-20)	5 (2-11)	**<0.001**	8 (4-14)	3 (2-5)	**<0.001**	7 (3-17)	3 (2-6)	**<0.001**
APS	4 (3-5)	3 (2-4)	**<0.001**	3 (3-4)	2 (1-3)	**<0.001**	3 (2-4)	2 (0-3)	**<0.001**	2 (0-3)	0 (0-2)	**<0.001**
Age	84 (77-89)	79 (67-86)	**<0.001**	85 (75-91)	74 (55-85)	**<0.001**	77 (69-84)	69 (56-78)	**<0.001**	69 (53-83)	52 (37-67)	**<0.001**
RR ≥20	28%	23%	**0.034**	24%	19%	0.139	8%	5%	0.063	12%	6%	0.055
<Alert AVPU	1.7%	1%	0.197	0.8%	0	0.619	n=0	0.2%	1	n=0	n=2	1
CKD	29%	10%	**<0.001**				14%	3%	**<0.001**			
Diabetes	25%	21%	0.059	16%	10%	0.066	21%	15%	**0.021**	10%	5%	0.187
Heart failure	37%	18%	**<0.001**	28%	9%	**<0.001**	9%	2%	**<0.001**	n=3	n=8	**0.009**
Liver disease	2%	0.8%	**0.026**	1%	1%	0.185	0.5%	0.4%	0.593	n=0	n=4	1
NEWS	2 (1-4)	1 (0-3)	**<0.001**	2 (0-3)	1 (0-3)	**0.024**	1 (0-2)	1 (0-2)	0.376	1 (0-2)	1 (0-1)	0.209

APS – **A**KI **P**rediction **S**core, AVPU – consciousness scale of best response: Alert, Vocal, Pain, Unresponsive, CKD – chronic kidney disease, HA-AKI – hospital-acquired AKI (within 7 days of admission), LOS – length of stay (days) in hospital, Mortality – in-hospital, NEWS – national early warning score, RR – respiratory rate (per minute)

For the primary analysis in medical patients with a baseline SCr the AUROC was 0.65 (95% CI 0.62-0.67) and the H-L p=0.064. In those patients without a baseline SCr the AUROC was 0.71 (0.67-0.75), H-L P=0.014. In surgical patients with baseline SCr: AUROC 0.66 (95% CI 0.62-0.69) and H-L p=0.093; Surgery without SCr baseline: AUROC 0.67 (0.58-0.75) and HL p=0.664 (figure 3.2). Calibration plots demonstrated agreement at low probability rates whilst at higher rates calibration deviated in the medical cohort, though the number of events were small (figure 3.3). Table 3.2 compares predicted rates of HA-AKI (from the original derivation study cohort) compared to observed rates in the validation medical cohort (with baseline SCr), grouped according to APS admission score.(84) The table shows in the validation cohort 4% of patients scoring 0-2 APS points developed HA-AKI, compared to 28% of cases scoring ≥7 points with an odds ratio of 4.7 (95% CI 3.1-7.2).

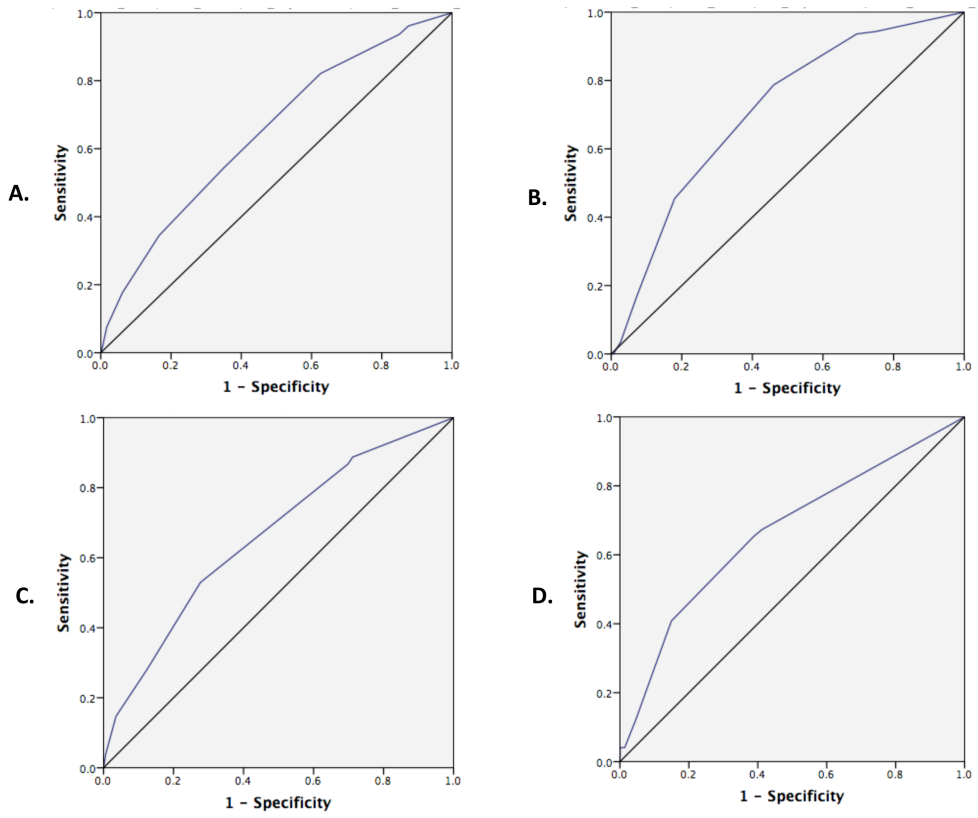


Figure 3.2 Area under the receiver operating characteristic curves for APS to predict HA-AKI. A: medicine with known baseline SCr 0.65 (95% CI 0.62-0.67); B: medicine no baseline SCr 0.71 (95% CI 0.67-0.75); C: Surgery with known baseline SCr 0.66 (95% CI 0.62-0.70); D: surgery without a baseline SCr 0.67 (95% CI 0.58-0.75). SCr – serum creatinine.

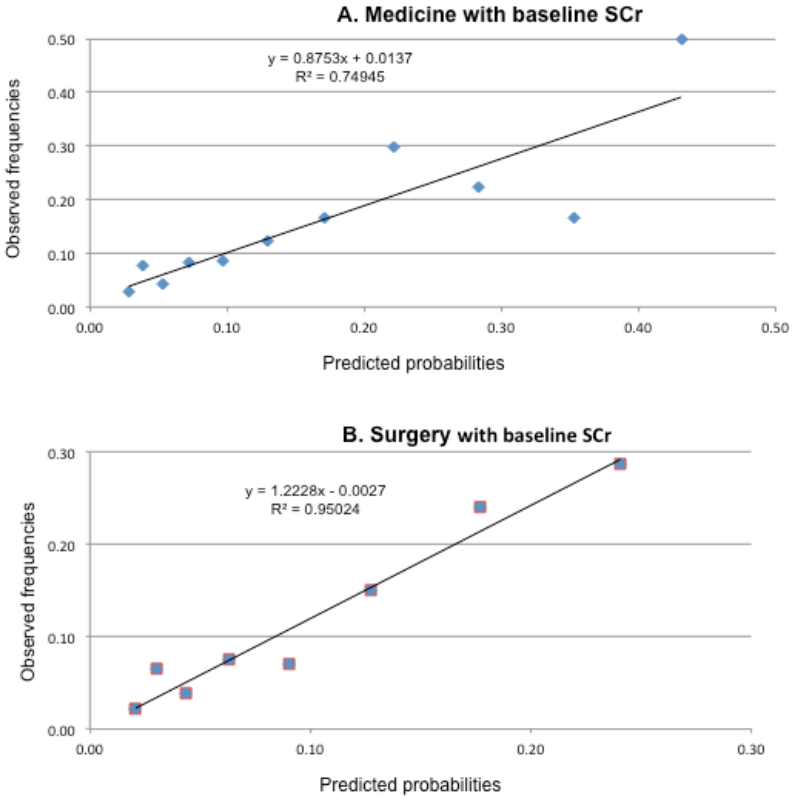


Figure 3.3 Calibration Plots of predicted probabilities vs observed rates of HA-AKI. Predicted probabilities vs observed rates (HA-AKI) at each level of the APS score in the medical (TOP) and surgical (BOTTOM) cohorts with a known baseline creatinine. HA-AKI – hospital-acquired AKI.

APS Points	Validation HA-AKI	Derivation Study HA-AKI
0-2	4%	3%
3-4	8%	7%
5-6	14%	19%
≥7	28%	33%

Table 3.2 – Risk range for HA-AKI by APS in external validation study vs original derivation study. APS - Acute kidney injury prediction score

Further performance at cut-off of 3 and 5 points on the APS for medical patients with a known baseline SCr is presented in table 3.2.

APS	Sensitivity	Specificity	PPV	NPV	PLR	NLR
≥5pts	34% (30-39)	82% (81-83)	16% (13-18)	94% (93-95)	2.1 1.8-2.4	0.79 0.73-0.84
≥3pts	82% (78-86)	37% (36-39)	10% (9-12)	96% (95-97)	1.3 1.25-1.38	0.48 0.39-0.59

Table 3.2 – APS prediction of HA-AKI in external validation. APS - Acute kidney injury prediction score. (95% confidence intervals) PLR – positive likelihood ratio, PPV - positive predictive value, NLR – negative likelihood ratio, NPV - negative predictive value.

In the medical cohort with a known baseline SCr, HA-AKI was associated with significantly higher in-patient mortality using a cut-off of 5 points on the APS: 28% of patients who developed HA-AKI died vs 18% with a score <5 points (odds ratio 1.8 [95% CI 1.1-2.9], p=0.015). This was also found in patients without a baseline SCr (mortality 42% vs 19%, p=0.029).

3.5 Discussion

3.5.1 Summary of main findings

This chapter presents one of the few external validations of a HA-AKI prediction model – the APS – for acute general medical and surgical cohorts with and without a baseline SCr. Moderate discrimination and satisfactory calibration were found across cohorts which is important as those lacking a baseline SCr can be difficult to interpret as to whether AKI is firstly present or not and secondly whether it has already occurred in the community or is a de-novo hospital event. External validation addresses optimism and generalisability.(237, 241, 242, 284, 304) Discrimination, assessed by AUROC analysis was inferior to the derivation study (0.65 in the comparable medical with baseline SCr vs 0.72, though 95% confidence intervals crossed) probably reflecting overfitting and clinical factors such as case-mix.(242, 284,

297) However, the observed rates of HA-AKI (8.1% vs 7.2% in the derivation) and associated in-patient mortality were similar in those with the outcome (20.5% vs 21.5%) and those without the outcome (3.5% vs 4.5%) in derivation and validation cohorts, respectively. The validation cohort were older and had higher rates of diabetes and heart failure whereas in contrast, rates of CKD and liver disease were lower. The two physiological observations - reduced level of consciousness and tachypnea - were also less commonly abnormal in the validation cohort (summary appendix table 3A.2).

Different methods were used in the two studies, for example, to define CKD (eGFR >1 and <6 months prior to admission in derivation vs national baseline in this external validation). It is unclear to what extent this would have affected results, though patients could conceivably have already developed AKI in the community. For example, if the last available SCr was a number of months ago, HA-AKI may be overestimated and CA-AKI underestimated. The validation uses the pragmatic national baseline definition that will be applicable in clinical practice and allows automation in an electronic hospital system. The original APS study had relatively few events per variable, as well as high weighting for two infrequent variables (liver disease and conscious level) that may account for reduced validation performance. In the larger validation population these variables, though still significant, were rarely present. Furthermore, the derivation cohort included patients who had more than one admission, whilst the external validation cohort included only first admission during the study period. Patients with recurrent admissions may have been prone to recurrent AKI episodes, and in particular a prior episode of AKI could be incorporated to improve model prediction. This is an area of future interest.

However, currently employed statistical markers of performance have limitations outlined at length in chapter 1 in this thesis. Correctly predicting a future event is more complex than diagnostics, for example pulmonary embolism pick-up on different imaging modalities, partly due to the time between prediction and outcome, with a significant stochastic element. Accurate risk stratification is often the best that can be achieved. AKI, defined

by a nominal change in SCr, may be the manifestation of a number of acute precipitants such as sepsis and vomiting, affecting a susceptible host for example with heart failure and reduced physiological reserve, making accurate prediction at a single time point with high discrimination, impossible.(269, 398) Given there is no consensus on what constitutes acceptable discrimination, descriptors such as “poor” or "good" to a particular AUROC are often ascribed. However the AUROCs has many shortcomings described in chapter one.(263, 399, 400) Though calibration is often overlooked, it is crucial in risk prediction,(263, 401) and in this validation, calibration plots suggest adequate fit between predicted probabilities and observed frequencies of clinical value. More nuanced alternatives such as decision-analytic measures could help if a prediction model is to be used for making clinical decisions. Lastly other measures of performance may be warranted in specific applications, such as reclassification metrics to gain insight into the value of adding a novel predictor to an established model.(268, 402, 403)

3.5.2 Potential clinical use of the APS

Firstly, using risk ranges, an APS score of 0-2 points (36% of all medical patients with a baseline SCr) was associated with a low risk of HA-AKI (4%), whilst a score ≥7 points had a 28% risk. Secondly, at 5 APS points (18% of patients) PPV was relatively low (16% [13-18%]), however, the high NPV 94% [93-95%] suggests an ability to identify those at low risk (as rule-out) and importantly even if HA-AKI developed patients (with an APS <5 points) were significantly less likely to die than those with a higher APS. Thirdly, with only 7 variables the APS is simple to calculate. Fourthly, all variables can potentially be automatically calculated in hospitals with electronic records containing clinical coding and physiological observation systems. Fifthly, the score in this validation performed with similar discrimination in two other cohorts – general surgery and in patients without a baseline SCr. Finally the score represents an aid memoire to variables associated with HA-AKI, consistently found on systematic review of other AKI prediction models.(332)

3.5.3 Alternative models

Chapter 2 provided a rationale for systematic review to appraise and compare models to inform whether a model is worth further investigating or implementing and highlighting alternatives. Only one other model for general emergency medical and surgical admissions has been externally validated, with an AUROC of 0.67 in derivation and 0.71 in external validation.(85) However, variables (age, previous admissions, admission diagnostic category, 7 laboratory parameters, Charlson co-morbidity index and proteinuria) may be difficult to automate and inclusion of admission diagnoses is a shortcoming if use is required at the earliest opportunity. Two older studies, one a retrospective cohort and the other including 27 variables have not been externally validated.(381, 382) Importantly all three may be difficult to calculate at the bedside. Finally the recent model produced by Koyner has not been externally validated and was designed to predict the outcome on a rolling 24-hour timeframe.(86)

3.5.4 Study strengths and limitations

This large study is one of the few external validations in general populations, who at least in England account for the majority of admissions to hospital and AKI cases.(169, 392, 393) Stringent inclusion criteria (stayed at least one night, repeat SCr performed), would have excluded a large proportion of 'low risk' patients, recognised as such by the clinical team, making the group in question of clinical relevance. Using only the first admission during the study period avoids for the effects of recurrent admissions. It is likely those with multiple attendances would be at higher risk of HA-AKI as well as scoring higher on the APS which takes into account age and (coded) co-morbidities, associated with an increased risk of re-admissions.(404-407) Model updating could assess the effects of a prior episode of AKI, for example. Whilst discrimination was only moderate, this represents a single time point from an outcome that could occur up to 7 days later. If for example, a physiological deterioration occurred following admission (with tachypnoea or reduced conscious level) this would be reflected in a higher APS. Patients without a known baseline SCr pose particular problems: firstly, CKD contributes to the

APS and secondly assessment of whether the AKI was community or hospital-acquired is more difficult. The study demonstrated that in this cohort the APS had similar performance to those with a known baseline SCr.

The single-centre nature of the cohort cautions against generalisability. Also although data were collected in a prospective nature by the electronic hospital record, past medical histories relied on previously coded events on the hospital database, thus they are likely to underestimate disease prevalence. This could be addressed by improved linkage with primary care databases, that could also incorporate information on medications. Moreover, the validation site is in the same geographical area to the derivation site and though overall a more affluent population, bears similarities (elderly population on the South-East Coast of England) that may not make the results widely generalisable. Relying on a baseline SCr that could have last been measured a number of months ago means a proportion considered to have HA-AKI may have already fulfilled criteria for a change in SCr prior to admission, ie have CA-AKI, however this is true in current clinical practice where the NHS algorithm is in use.

3.5.5 Future directions

Further external validation of prediction models, with updating where necessary is desirable.(237) Unfortunately, as AKI is a syndrome reflecting diverse underlying pathophysiological states imposed on a host with chronic disease, any model at a single time-point, is unlikely to predict with high discrimination. Improving models may require the combination of comprehensive medical records with fluctuating physiological trends, mirroring the traditional clinical approach to prognosis, bearing in mind the evidence of enablers of uptake.(309, 408) Drawbacks of using urine output and SCr known to be unpredictably affected by diverse inputs suggest that the employment of more refined biomarkers as both predictors and markers of significant AKI are required.(30) Lastly, no AKI prediction models have undergone impact analysis. As no single intervention has been found to improve outcome in AKI, a model would most likely to be used as a part of a

systematic alert to risk and initiation of enhanced monitoring in the appropriate clinical location with avoidance of iatrogenic harm.

Conclusion

The APS to predict HA-AKI in an external validation study of general medical and surgical patients performed with moderate discrimination and acceptable calibration. The prediction rule could help identify at admission those patients at higher risk of developing HA-AKI, in order to prevent its occurrence and avoid, or at least mitigate associated significant complications. The following chapter will explore the effects of implementation of the model on clinical outcomes by an impact analysis.

Chapter 4 – Impact Analysis

This chapter covers three areas:

1. a background to impact analysis,
2. an overview of clinical decision support systems and their implementation with reference to AKI and,
3. the ICE-AKI Study: **I**mpact analysis of a **C**linical prediction rule and **E**lectronic AKI alert in general medical patients.

4.1 Background to impact analysis

The final stage of evaluation of a prediction model is to test impact in practice for relevant clinical outcomes.(199, 286, 409) Implementation is only warranted if demonstration of effectiveness has been shown including improved outcomes or reduced costs whilst maintaining quality.(294, 409) Few impact studies have been performed,(410) with a review of 434 models, finding only 2.8% had undergone impact analysis.(410) A systematic review in primary care found only 18 studies.(411) Prediction performance of most models deteriorates on external validation,(199, 243, 284, 286) and even if accurate, outcomes may not be improved or the model may be rejected for a number of reasons. First, clinicians' intuitive estimation may be as good, if not better. For example, Sinuff *et al*., found ICU physicians were more accurate than prediction rules for mortality.(188) Second, practical barriers such as cumbersome calculations and concern around litigation may lead to rejection of the models.(409) Even in ICU where models have been available for decades, with access to electronic patient records (EPRs),(412) models are rarely in routine use,(413, 414) perhaps because most are mortality predictors with limitations compromising generalisability.(415-421) Lack of broad validation may lead clinicians, perhaps rightly, to mistrust the probabilities provided.(197, 198, 286, 287). For example, the SUPPORT study showed survival estimates combining a model with a physician's estimate better identified patients with high probabilities of survival or death but this failed to improve the outcome investigated.(422) Application can also be limited by the

inclusion of ambiguous definitions such as “alternative diagnosis just as likely” in the Wells score for deep vein thrombosis,(423) or changes in practice over time, such as use and/or availability of imaging. Finally one study suggested the best predictors of implementation were familiarity acquired during training, confidence in its usefulness, and whether the model was user-friendly.(312)

Ideally in an impact study patients, or units, are randomised to the use or not of the model. Randomisation of individual patients is difficult however, as learning effects will lead to a reduced contrast between the two study groups, resulting in a diluted measured impact.(424) Randomising centres may prevent risk of contamination.(199) Two examples of successful RCT impact studies assessed the Ottawa ankle rule,(425) and in a cluster RCT a pneumonia model.(307) A stepped-wedge or cluster randomised trial is particularly useful for complex interventions with clusters (e.g. hospitals) randomly allocated a time period when they are given the intervention.(426)

Partly because RCTs are expensive and time consuming other approaches are employed such as a before-after with control study. Using this design two studies from the group who derived the Ottawa rules found a positive impact,(210, 427) whilst in a third study from an independent group of researchers found no benefit,(428) suggesting local implementation strategies are key to success.(429, 430) The ‘before-after’ study, using the same care providers is sensitive to temporal changes in, for example, therapeutic approaches. A variant, relatively cheap and easy to implement, is when clinicians are asked to make a decision for an individual before they have been provided with the model’s predicted risk, and subsequently after they have been ‘exposed’ to the model predictions for the same patient.(241) In a ‘one-off ’ impact study the outcome is measured in alternating time periods when the prediction model is, or is not, available. If clinicians in the centre have changed over time, results may be biased. When the outcome of interest is only behaviour or decision making of healthcare professionals, a cross-sectional study may suffice.(431) Clinicians can be randomised to receiving or not receiving predictions, with management decisions then compared. Finally there is the difference-in-differences approach first described in the

economics literature by Card and Krueger,(432) which uses a comparison group experiencing the same trends but is not exposed to the intervention and is described later in this chapter.(433)

4.2 Overview of clinical decision support systems (CDSS) implementation, AKI alerting and care bundles

4.2.1 CDSS in healthcare

Health care delivered often falls short of optimal, evidence based practice.(434, 435) For example, one study found it took an average of five years for guidelines to be adopted into practice.(436) The complexity of delivering health care routinely exceeds the bounds of human cognition, for example the use of medications in patients with renal impairment.(437) Partly as a response to this knowledge clinical decision support systems (CDSS) have increasingly been introduced to provide pertinent information, intelligently filtered or presented at appropriate times, to enhance care and outcomes.(438) Substantial evidence exists that quality of care however, does not improve simply through isolated introduction of electronic health records (EHRs).(439, 440) Indeed the pros and cons of such technology have been debated in the US since the passage of acts to mandate and incentivise adoption.(441) Though a meta-analysis of 148 RCTs on CDSS implementation found only 20% influenced clinical outcomes,(442) there is evidence to suggest CDSS are effective in specific circumstances, for example, for preventive conditions and medication-related decision support.(309, 443) One type of CDSS is the generation of alerts of potential errors, enabling staff to improve therapeutic decision-making.(444)

Alerting is implemented as a safety mechanisms in many industries, with the premise that it is beneficial to indicate a particular action or a situation that exceeds safety thresholds or when implementation of a task may compromise safety.(444) Careful design and implementation is essential to avoid becoming an irritation, even impeding performance.(444) For example one interruptive alert intended to avoid drug interactions with warfarin led to

incorrect omission of important therapy.(445) A second study found an increase in mortality for children transferred into special care after introduction of a computerised physician order entry, which was thought to be due to slowing down clinicians in delivering care.(446) One study found even experienced care givers could not identify the majority of their own ICU critical alarm sounds partly due to the sheer number.(447) Thus, human factors principles in implementation must be taken into account.

4.2.2 CDSS in the field of AKI

AKI alerts are recognised by international ADQI consensus guidelines to represent an opportunity to prompt earlier clinical evaluation, further testing and ultimately intervention, rather than as a diagnostic label.(316, 448) EHRs including CDSS can detect changes in SCr, increase AKI recognition, whilst reducing the time to therapeutic interventions to prevent progression, and improve outcomes.(448) In 2015 ADQI identified 12 automated AKI alert systems, eight in adult hospitals.(448) Delivery included interruptive as well as non-interruptive alerts within electronic systems, paper notification, paging or telephone calls. Few studies have evaluated the sensitivity and specificity of the AKI detection algorithms employed, though the NHS England AKI algorithm has been found to perform well in clinical practice.(175) The optimal algorithm in clinical settings is currently unknown.(448) One significant problem is establishment of baseline renal function, discussed in the preceding chapters. Firstly in patients with prior results those results may be old and a change in renal function may already have occurred well before the index admission. Secondly, as seen in the external validation presented in chapter 3, a proportion of patients admitted will not have a prior SCr to establish a baseline of renal function. Potential surrogates to replace an unknown baseline SCr have several limits. Use of admission SCr will fail to detect AKI present at the time of admission while use of an estimated baseline may ignore pre-existing CKD leading to a high rate of false positive detection of AKI. (449, 450)

4.2.3 Evidence to support CDSS optimal implementation

Key factors can optimise alerting (table 4.1) and these will be discussed in the ICE-AKI study methodology section of this chapter.

Factor	Explanation
Define situation, alert threshold	Minimise number & frequency to avoid fatigue(451, 452) Well-designed easy to acknowledge with corrective actions improve response(444, 453)
Placement, visualisation	Proximity compatibility principle(454) Ensure optimal viewing angle(455)
Prioritisation, habituation	Red & orange: increase hazard perception & priority(456) Minimise number of colours to avoid confusion Shape of alert alters prioritisation.(444) Tiering increases adherence(457) Reduce habituation: avoid excessive alerting(452)
Written information	Signal word draws attention, hazard level(458) Statement about hazard Instruction Consequence if instruction ignored
Mental models	Understand users beliefs and association with processes associated with the intervention.(459)

Table 4.1 – summary of key factors to optimise implementation of alerting.

Many alerts are incorporated as part of CDSSs most commonly medication-related alerting to prevent adverse events and serve as a support aid.(444) However, clinicians often override recommendations, even clinically significant ones.(460-462) Furthermore, excessive alerts results in overload or 'alert fatigue.'(463) However, relatively little evidence is available to explain why systems succeed or fail.(464) In one review of 68 studies, CDSSs improved drug dosing and preventive care but not convincingly for diagnosis, with few studies on patient outcomes.(465). As mentioned in chapter 1 (figure 1.3) a systematic review suggested predictors of improved clinical practice are: computer based, automatic provision as part of clinician workflow, at the time and location of decision making, with recommendations rather than just assessments.(309) The review also suggested performance feedback, sharing recommendations with patients and requesting reasons for not following recommendations contributed to improved outcomes. Another review found that 57% of CDSSs affect practitioner behaviour, whereas only 30% demonstrated a positive effect on patient outcomes.(466) Moxey et al,

reviewed 58 studies and also reported integration into workflows and relevance and timeliness of the messages impacted uptake but also found that availability of hardware, technical support and training were crucial.(467) Systems endorsed by colleagues with minimal perceived threats to autonomy were more likely to be accepted. Carroll et al, studied factors influencing response and concluded prompts around serious issues would be addressed though the most significant factor predicting response was position at the top of the page.(468) Finally, Bates and colleagues, reflecting on their extensive experience with CDSS proposed 'Ten Commandments' to follow summarised in table 4.2(408):

1. Speed is everything	First consideration in a busy clinical environment(469, 470)
2. Anticipate needs, deliver in real time	Information accessible electronically, in isolation, has little effect(471)
3. Fit into user's workflow	eg guideline on a single screen at time of prescribing
4. Little things make a big difference	usability but also making flow hard to ignore suggestions(472)
5. Recognise physicians strongly resist stopping	if alternative not offered
6. Changing direction easier than stopping	
7. Simple interventions work best	eg guideline on a single screen
8. Ask for additional information only if necessary	likelihood of success in implementation inversely proportional to number of extra data elements
9. Monitor impact, get feedback, respond	eg work out barriers, what works
10. Manage and maintain knowledge-based systems	eg updating if changes in guidelines occur.

Table 4.2 Ten Commandments of clinical decision support systems. From Bates et al.(408)

Reflecting on this body of research in CDSS alerting, specific goals of an in-hospital AKI alert include:

1. flag patients with reversible factors to halt progression, speed recovery and reduce complications by intensification of management,
2. identify a cohort who require escalation of care, or decisions over limitations of treatment and,
3. identify a cohort for the use of biomarkers to further risk stratify.

In the single RCT of an AKI alert no difference in outcome was found, however, the effector arm consisted solely of the alert with a link to practice guidance without for instance a care bundle or other specific interventions.(170) Work led by Selby and colleagues in the UK suggested that in patients alerted and having a care bundle promptly completed improved outcomes were found compared to those who did not, but patients were not randomised.(141, 396) In a recent large before-after study in the US, an AKI alert was associated with a small but statistically significant decreased risk of death in those who developed HA-AKI alongside reduced length of stay after adjustment for other factors.(140) Thus from evidence to date, it is not enough to simply alert the presence of AKI, but associated with the initiation of appropriate care may lead to outcome benefit. The second point serves to highlight that a proportion of patients who have developed AKI are critically unwell but may be more appropriate for end of life care rather than active, aggressive medical management. Finally, identification of patients with or at risk of AKI could allow employment of biomarkers to further stratify patients. This may be of particular relevance if the biomarker is expensive or when initial stratification is necessary prior to biomarker employment to minimise false positives results.

4.2.4 Care bundles

A care bundle is a group of three to five evidence-based interventions which, when performed together, have a better outcome than if performed individually.(473) Each intervention should be accepted as good practice and widely applicable. In a number of medical fields outside AKI care bundles have been found to be effective, for example, relating to reductions in catheter-related sepsis in critical care.(474) However, success is influenced by implementation processes such as shaping of knowledge, monitoring compliance and feedback.(475) For example, one behavioural study looking at a sepsis care bundle found themes influencing implementation that could then be addressed such as specific training for staff and permitting collegial challenge.(475)

In the absence of any specific drug to treat AKI, guidelines suggest management is supportive including recognition and management of sepsis, shock and hypovolaemia, avoidance of medication that could contribute or worsen AKI, appropriate investigations and referral to specialists when indicated.(316) However, as outlined in Chapter 1 a number of studies have found such basic care is not delivered systematically.(166, 476, 477) The care bundle approach holds promise to address shortcomings in this field as part of a complex healthcare intervention. Crucially this requires behavioural change on the part of the clinical team that can be influenced by education, appropriate bundle design and feedback including compliance, processes of care and outcomes. Selby and Kolhe found five AKI care bundle studies and found limited data to support their role in improving both processes of care and outcomes though more evidence was called for in the field.(478)

The systematic review presented in Chapter 2 suggested there are few validated, usable prediction models in general hospital settings. Chapter 3 validated the APS in a second general hospital setting and suggested the model compares well with others described to date and uses easily available variables that could subsequently allow assessment of generalisability. The two hospitals where the APS was studied had the available technology to electronically generate the model on patients and their setting provided a natural experiment to explore use of the this model in clinical practice. Pragmatically therefore the APS represented a usable and workable prediction model that could be used in clinical practice and be assessed as part of a complex healthcare intervention. The remainder of this chapter will present an impact analysis study utilising the APS prediction model alongside an AKI alert, both embedded in an electronic CDSS.

4.3 The ICE-AKI Study:

Impact analysis of a **C**linical prediction rule and **E**lectronic **AKI** alert

4.3.1 Summary

Background

Using e-alert systems to highlight community-acquired acute kidney injury (CA-AKI) and hospital-acquired AKI (HA-AKI) has become commonplace in the UK with the aim of triggering early intervention(s) to improve patient outcomes. The effect of combining an e-alert system with an HA-AKI prediction model / clinical prediction rule (CPR) has not been evaluated.

Methods

A controlled, before-and-after study on acute medical admissions to two adult non-specialist hospital sites (one intervention, one control) in the South of England (2014-16) was conducted. At admission, a CPR highlighted patients at higher risk of HA-AKI in conjunction with an e-alert which triggered care bundles of interventions based on current best evidence for practice. The primary outcome for evaluation of the CPR was incident HA-AKI using a difference-in-differences analysis across two ten-month time periods pre- and post- introduction, adjusted for age and co-morbidities. Secondary outcomes in those developing HA-AKI included: AKI progression; intensive care unit (ICU) escalation; and in-hospital mortality.

Findings

30,295 admissions to the two sites were included in the analyses, with a mean age of 74·5 (±17.0) years. Overall, CA-AKI was present in 8·3% (n=2,502) and HA-AKI developed in 7.3% (n=2,040). After intervention difference-in-differences analysis, adjusting for age and co-morbidities and prior admissions, indicated a significant reduction in incidence compared with the control site (OR 0·990 (0·981-1·000), P=0·049). Unadjusted HA-AKI incidence did not change significantly at either site (intervention site, 8·28% pre- and 7·73% post-CPR, odds ratio [OR], 0·928 (95% CI 0·824-1·045) P=0·223; control site 6·55% pre-, 6·67% post-CPR, OR 1·019 [0·888-1·170]

P=0·805). In-hospital mortality in HA-AKI cases significantly reduced at intervention site on difference-in-differences analysis (OR 0·924 [95% CI 0·858-0·996] P=0·038) and unadjusted analysis (27·46% pre- vs 21·67% post-CPR, OR 0·731 (0·560-0·954) P=0·021). Reductions were seen in ICU escalation and AKI progression at the intervention site, without reaching statistical significance. Mortality in those flagged by the AKI risk CPR reduced following the intervention (14% pre vs 11% post, P=0·008). Outcomes for CA-AKI cases did not significantly change. A process measure review suggested significant improvements at the intervention site.

Interpretation
This impact assessment showed that a multi-modal intervention, including an electronically integrated CPR alongside an e-alert for those developing HA-AKI may improve in-hospital outcomes in acute medical admissions. CA-AKI outcomes were not affected by an admission e-alert. The study provides a template for investigating interventions utilising integrated electronically generated prediction modelling. Further studies should assess generalisability and cost effectiveness.

4.3.2 Introduction

Up to 20% of adults admitted acutely to hospital meet AKI KDIGO criteria at, or during their stay, with high associated mortality.(69) As the introductory chapter of this thesis outlined, this may reflect the limited ability of conventional markers to detect renal injury in a timely fashion as well as heterogeneity in aetiology, in turn providing some explanation as to why clinical trials of therapeutic interventions have proved disappointing.(479) Following the NHS patient safety alert in 2014 hospitals were mandated to deploy an electronic alert upon AKI development,(480) however, from evidence to date there is uncertainty around the relative benefits of AKI e-alert systems necessitating ongoing rigorous evaluation.(481, 482) Furthermore, chapters 1 and 2 highlight the dearth of impact analyses of the large number of published prediction models or CPRs.

This impact study investigated the use of the Acute Kidney Injury prediction score (APS), externally validated in chapter 3,(84, 387) with aims including:

(i) in emergency medical admissions, without identifiable AKI at admission, could introduction of an electronically generated AKI CPR reduce the incidence of HA-AKI,

(ii) in concert with the CPR, does an e-alert for new HA-AKI reduce associated complications including in-hospital mortality, escalation to intensive care and maximal AKI stage,

(iii) does identification of a group at high risk of HA-AKI improve their in-hospital outcomes and,

(iv) what is the effect of e-alerting cases of CA-AKI at admission to hospital?

Objectives

To address these aims required fulfillment of a number of objectives:

I. identify baseline rate of HA-AKI & CA-AKI,

II. identify baseline outcomes of patients with HA-AKI & CA-AKI
    a. in-hospital mortality, 7-day mortality,
    b. progression of AKI (eg from KDIGO Stage 1 to 3),
    c. length of stay,
    d. escalation to ICU,

III. compare historical rates at intervention site and control site
    a. in-hospital mortality,
    b. progression of AKI (eg from KDIGO Stage 1 to 3),
    c. length of stay,
    d. escalation to ICU,

IV. evaluate effect of introducing the APS and AKI alert on new HA-AKI and associated outcomes,

V. evaluate introduction of the nationally recommended AKI e-alert system on AKI outcomes for those with AKI on admission
    a. compared with historical rates at intervention site and,

b. compared with control site without the AKI alert,

VI. a process evaluation to explain and understand effects of the interventions.

4.3.3 Methods

4.3.3.1 Study design and participants

A prospective controlled before-after impact analysis of the APS CPR and AKI e-alert was performed on two adult general acute medical units. The two sites, part of the same NHS Trust, were split into intervention and control and the study ran over two ten-month periods (July 2014 – April 2015 and July 2015 – April 2016). There was a bed down period for the intervention of two months (see figure 4.1 for study design and appendix table A3.1 for variables and weightings of the APS CPR). West Sussex is a county in the south of England, bordering East Sussex, Hampshire to the west and Surrey to the north, and to the south the English Channel. Chichester, population 26,795 with 113,794 in the non-metropolitan district (census 2011), in the southwest is the county town and only city, with the coastal settlement of Worthing one of the largest towns (population 104,600).

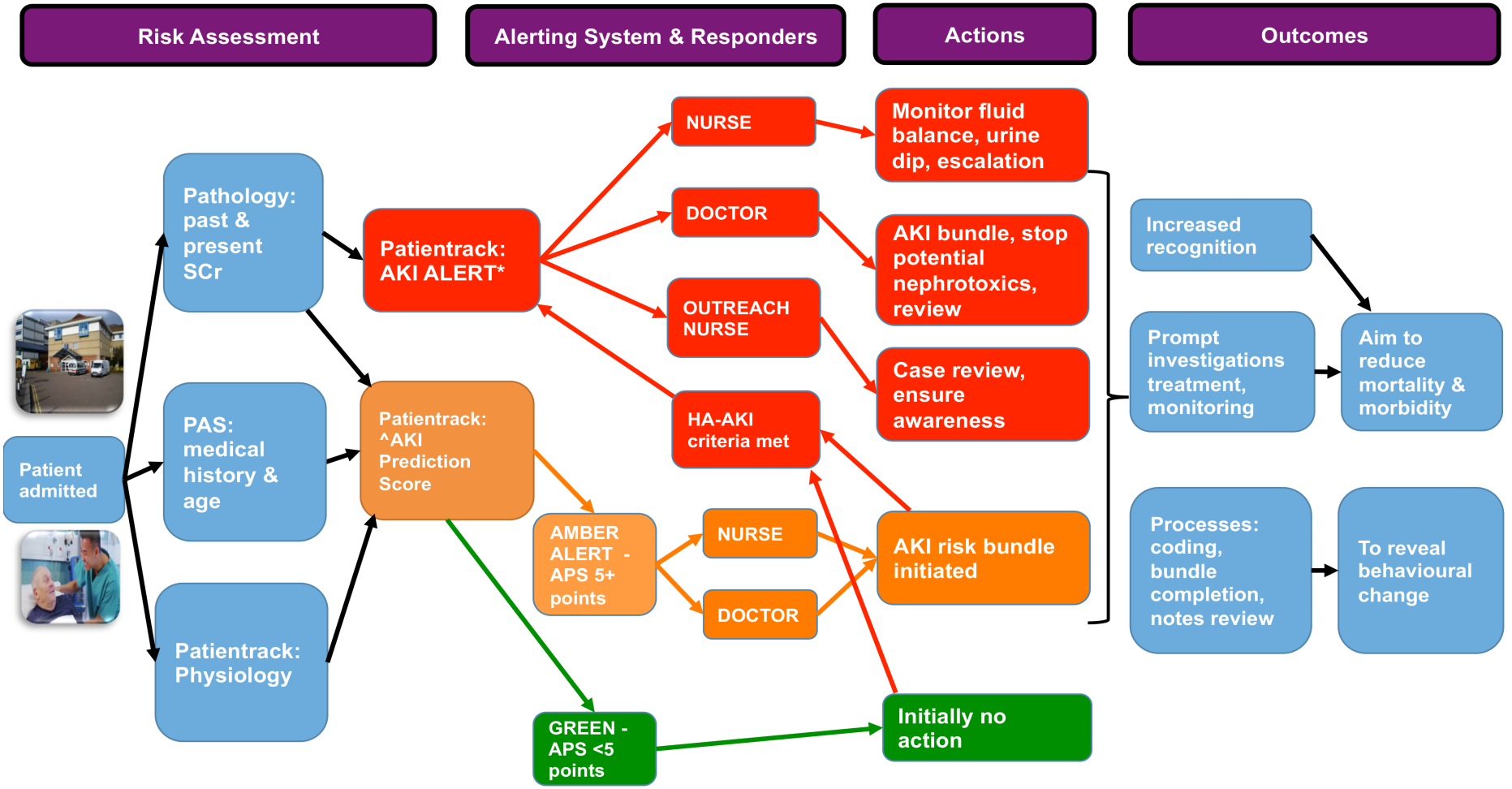

Risk Assessment
Alerting System & Responders
Actions
Outcomes
Patient admitted
Pathology: past & present SCr
PAS: medical history & age
Patientrack: Physiology
Patientrack: AKI ALERT*
Patientrack: ^AKI Prediction Score
NURSE
DOCTOR
OUTREACH NURSE
HA-AKI criteria met
AMBER ALERT - APS 5+ points
NURSE
DOCTOR
GREEN - APS <5 points
Monitor fluid balance, urine dip, escalation
AKI bundle, stop potential nephrotoxics, review
Case review, ensure awareness
AKI risk bundle initiated
Initially no action
Increased recognition
Prompt investigations treatment, monitoring
Aim to reduce mortality & morbidity
Processes: coding, bundle completion, notes review
To reveal behavioural change

Figure 4.1 Study intervention. At Control no traffic light seen. APS – AKI score, RED boxes = AKI (community or hospital-acquired), AMBER = APS ≥5 points – cut-off for flagging patient at risk of AKI, GREEN box – all other patients (APS <5 points). HA-AKI – Hospital-acquired AKI, PAS – Patient administration system, SCr – serum creatinine. Patientrack AKI ALERT* - patients with AKI at admission defined as increase >1.5 times baseline SCr using NHS England Algorithm for calculation of baseline, ^AKI Prediction Score – clinical prediction rule.

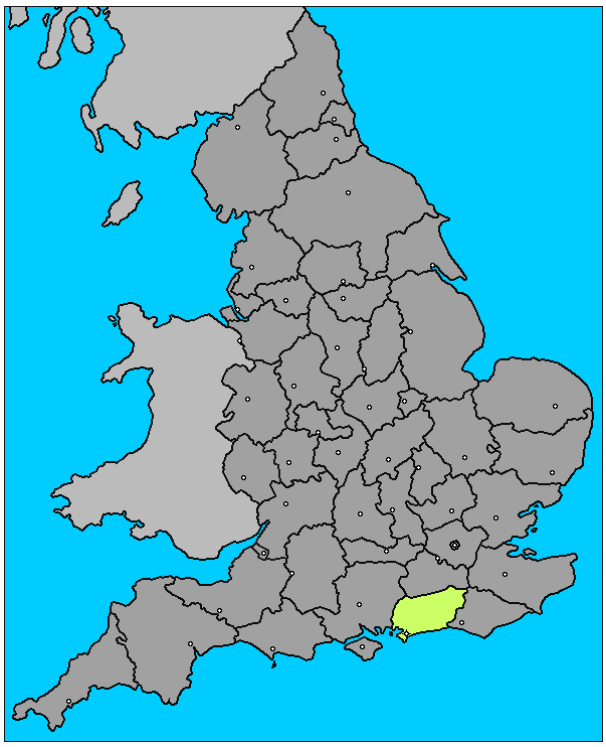


Figure 4.2 Map of England highlighting West Sussex County in Yellow.

Western Sussex Hospitals NHS Foundation Trust (WSHFT) is a non-specialist hospital organisation of 870 beds and serves a population of around 450,000 people across a catchment area covering most of West Sussex. The Trust runs three hospitals: St. Richard's Hospital in Chichester, Southlands Hospital (not a site of acute admissions) in Shoreham-by-Sea and Worthing Hospital in the centre of Worthing. St Richard's and Worthing hospitals provide 24-hour A&E, acute medical care, maternity and children's services. In 2009 the Royal West Sussex and Worthing and Southlands Hospitals NHS trusts merged and in 2013 the Trust became a Foundation Trust, with an annual budget of £400 million and currently employs 6,881 staff. In 2015/16,

the Trust treated 135,792 inpatient and day cases, saw 136,804 patients in A&E and held 585,846 outpatient appointments. There are around 40-60 acute medical admissions per 24-hour period. The two acute sites are 20 miles apart with relatively similar catchment areas, though the control site has a higher rural population and less socio-economic deprivation (table 4.2).(483)

	Worthing	Chichester	National average
Life expectancy	77.8M 82.6F	79.2M 83.5F	78.3M 82.3F
Unemployment	7%	2.8%	7.8%
Very good health	44%	48%	47%
Good health	36%	35%	34%
Incapacity benefits	7%	4%	7%

Table 4.2 - 2011 Census data ONS M – male, F – female.

The APS CPR was originally derived at the intervention site.(84) The design of the study whilst not randomised, offers a number of advantages. Having two sites and historical baseline data allows multiple outcome comparisons (temporal and spatial). Together these can allow for robust conclusions to be drawn. Furthermore the location of the two sites provides a natural experiment with a number of advantages for a study into a complex hospital intervention. The hospitals are relatively well balanced in terms of case mix and age yet there is little cross-contamination of front line staff: Consultant physicians do not work on both sites, junior staff do not rotate between the hospitals and the nursing staff work solely on one site. Ethical approval was given by NHS Research Ethics Committee London - South East (REC reference 13/LO/0884) and the study registered with clinicaltrials.org (NCT03047382).

At admission, all in-patients routinely have physiological observations measured and entered via handheld systems into the clinical data software system (Patientrack© Sydney, NSW, Australia). Previous ICD-10 coded history (heart failure, liver disease and diabetes mellitus), were retrieved and CKD was defined as an eGFR <60mls/minute prior to admission (in those with an available baseline SCr) and baseline SCr was defined using NHS England's National algorithm.(175) KDIGO criteria for AKI was employed using only SCr as few of this patient group have routine urinary catheters placed (SCr increase of ≥1·5 from baseline value or absolute value

≥354μmol/l for CA-AKI, or ≥1·5 from the admission value or ≥26·5μmol/l within a rolling 48 hours during the first seven days of admission for HA-AKI). If no baseline was available, as per national guidance, CA-AKI was assumed absent unless SCr ≥354μmol/l.

4.3.3.2 – The Patient journey and interventions

The study fits MRC definitions for a complex healthcare intervention including: interacting components within the experimental and control sites (both hospitals are in the same Trust and part of the wider NHS); the number of behaviours required by those delivering the intervention; the number of groups targeted (doctors, nurses, healthcare assistants and pharmacy); the number and variability of outcomes; and finally the degree of tailoring the intervention to the individual patient.(484) Implementation of the intervention required integration of technology, education and buy-in from a multidisciplinary team. As already outlined the technology alerted patients with AKI and considered at high risk of developing AKI on the APS prediction model, in multiple electronic locations visible to the multidisciplinary team, whilst also generating electronic care bundles to be completed by the team. Crucially, as with any complex intervention, successful implementation required a number of educational interventions (discussed in 4.3.3.3). These educational interventions were not limited for example to doctors. Furthermore, the outreach nursing team could remotely identify the target group on a daily basis and then work with the ward team to both identify and deliver best practice care.

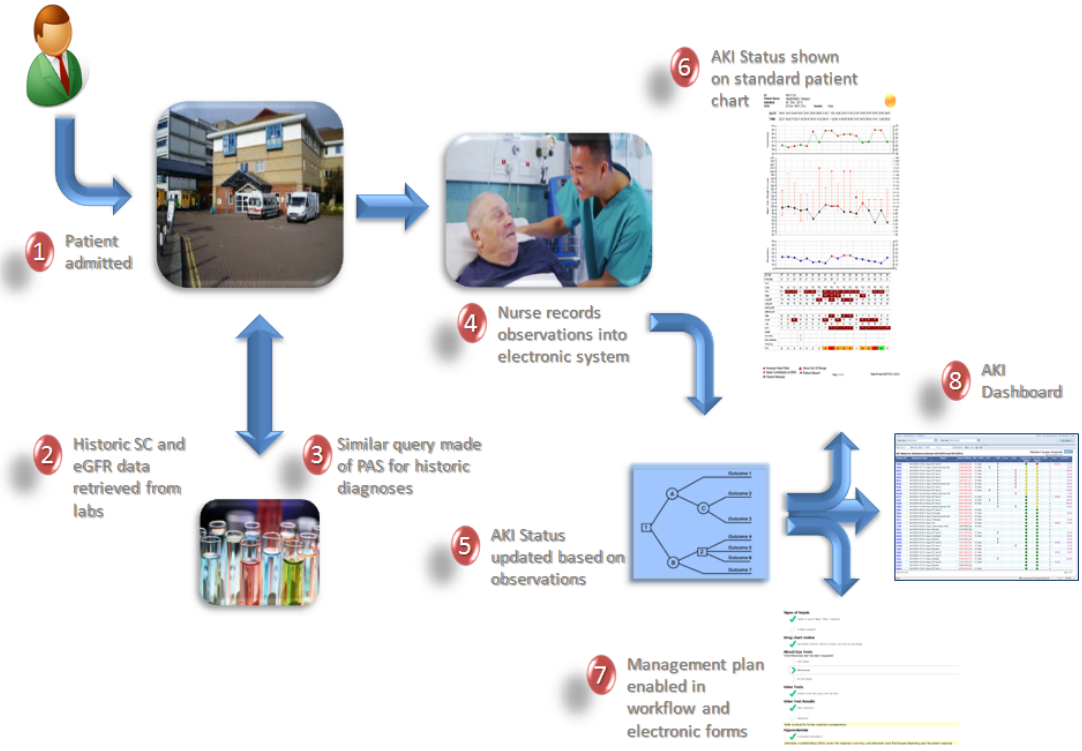


Figure 4.3 - Summary of patient journey and calculation of AKI status and AKI CPR calculation. CPR – clinical prediction rule, PAS – patient administration system SCr – serum creatinine.

The interventions are summarised in figure 4.1. At admission for the purposes of the study those with an APS of ≥5 points were flagged 'at risk of AKI,' (coloured AMBER) in multiple electronic locations and a care bundle generated, requiring *consideration* of all 4 interventions. Those with an APS <5 points were classed GREEN, with no alert. This cut-off was chosen for its relatively high specificity and positive predictive value in derivation and validation work to minimise false-positives that have been shown to limit clinical uptake.(84, 387) At the intervention site, the presence of AKI flagged RED on admission (CA-AKI) as soon as the SCr result was calculated and produced an initial 1-hour bundle with 4 directives, followed by prompts depending on response. At 6 hours and lastly at 24-hours further bundles were recommended (figure 4.4, table 4.3). The same process occurred for those who subsequently developed HA-AKI. Response to the alerts was multidisciplinary involving ward doctors, nurses, health care assistants and pharmacists, together with the critical care outreach nurse team on a daily basis reviewing patients remotely or at the bedside. None of the researchers involved in data analysis were involved in management of the patients and had access only to the fully anonymised individual level data and were blinded to other patient data, as well as components of the calculated APS score.

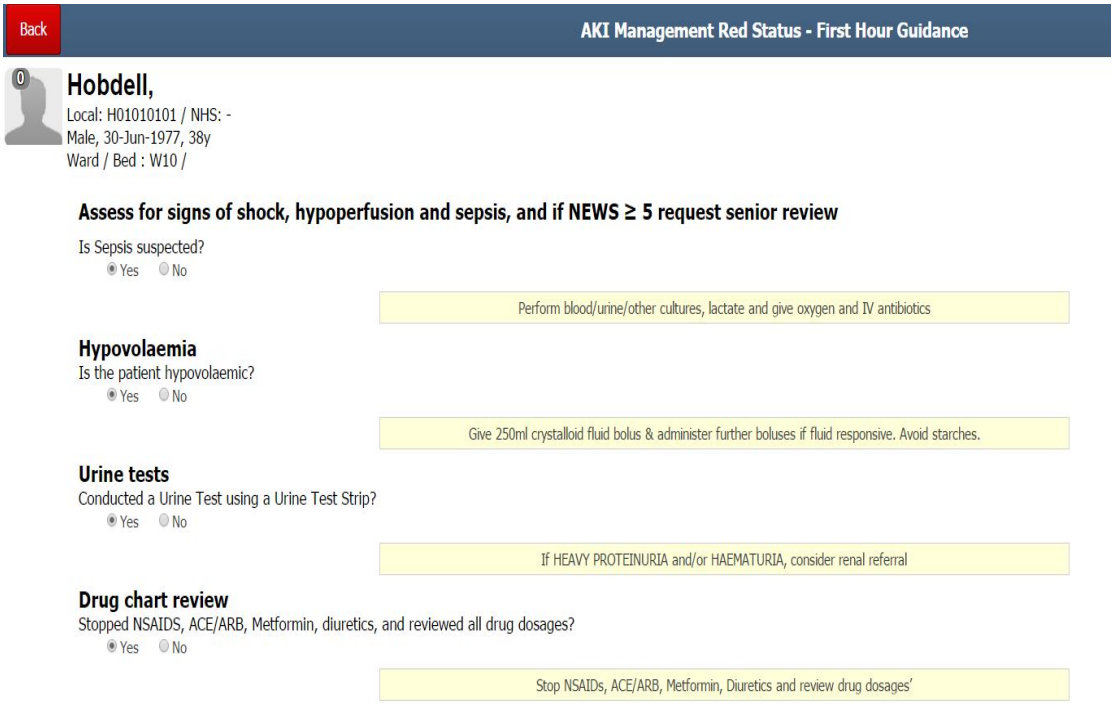


Figure 4.4A First hour AKI bundle for those with AKI.

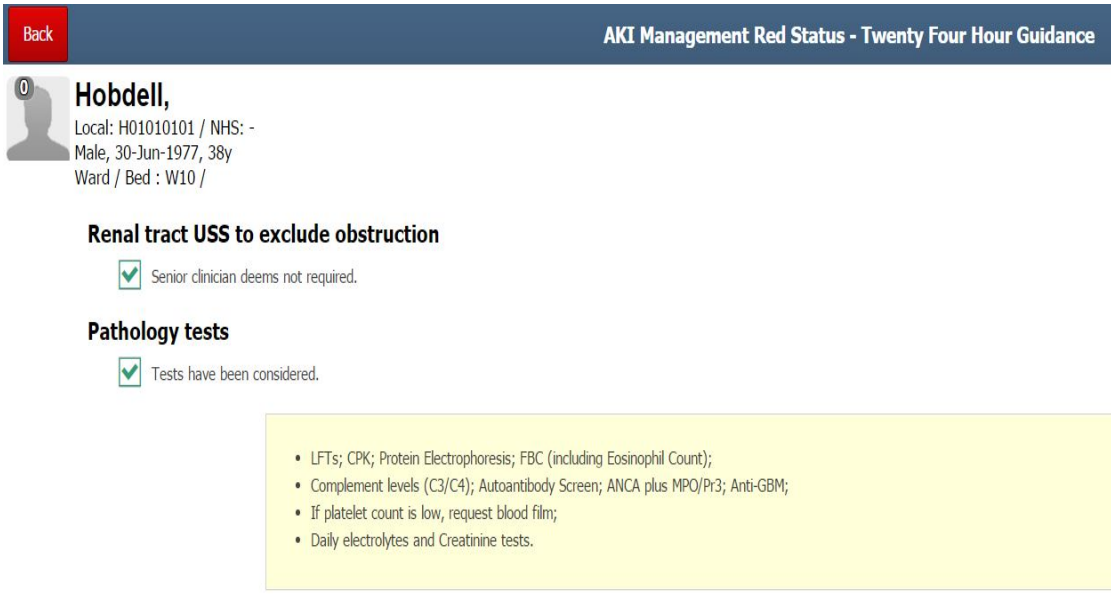


Figure 4.4B - First 24 hours AKI bundle for those with AKI.

Table 4.3 – summary of care bundle directives for patients with AKI - RED and at risk (APS 5) of AKI – AMBER. SCr – serum creatinine.

1, 6 & 24 hour advice & tasks to be considered & submitted	Triggered when KDIGO criteria met: change SCr >1.5 baseline at admission (or if >354µmol/L) or ≥26.5 µmol/l increase on rolling 48hr period in hospital. 1 hour Search for underlying cause of admission – eg. If sepsis suspected send blood cultures, IV antibiotics. Fluid bolus & re-assess, fluid balance chart, daily weights Review drug chart & stop potential nephrotoxics Urine dip 6 hour Re-review fluid status, consider escalation of care 24 hour Where appropriate consider imaging renal tract Consider other specific blood tests: autoimmune, CK. If not imprivng for discussion with Nephrology
Advise review / consider	Search for underlying cause: eg. sepsis suspected: send cultures Assess fluid status & consider fluid bolus, re-assess; fluid balance chart, daily weights Review drug chart & consider stopping potential nephrotoxics Urine dip

In addition to the bundles, the patients were flagged in a number of areas including an AKI tile in the electronic patient summary view, the electronic observation chart and an AKI dashboard highlighted those with AKI and those at risk (figure 4.5-7). The AKI dashboard, including break-down of the APS score was available to all clinical teams at the intervention site, with the critical care outreach team in particular using this modality to access and review patients by the bedside or remotely. Lastly a graph of SCr results could be accessed from a link from the observation chart (figure 4.8).

Figure 4.5 - AKI status highlighted in AKI tile in bottom right of the patient summary view.

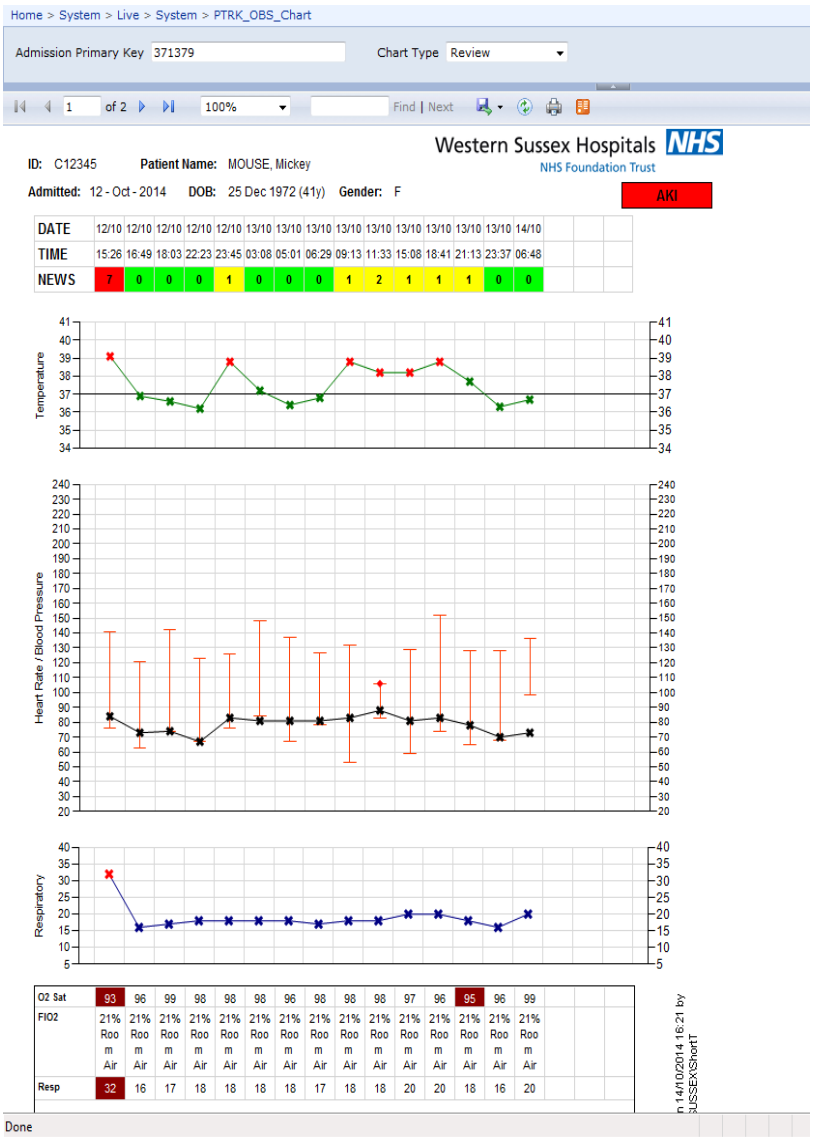


Figure 4.6 - AKI status highlighted in electronic observation chart (top right).

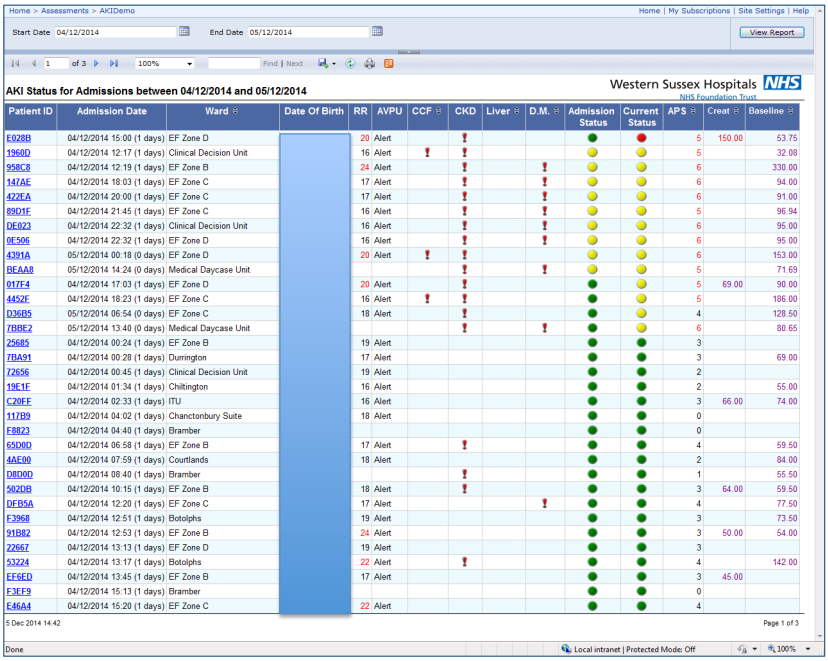


Figure 4.7 - AKI dashboard highlights with AKI and those at risk including the break-down of the APS score.

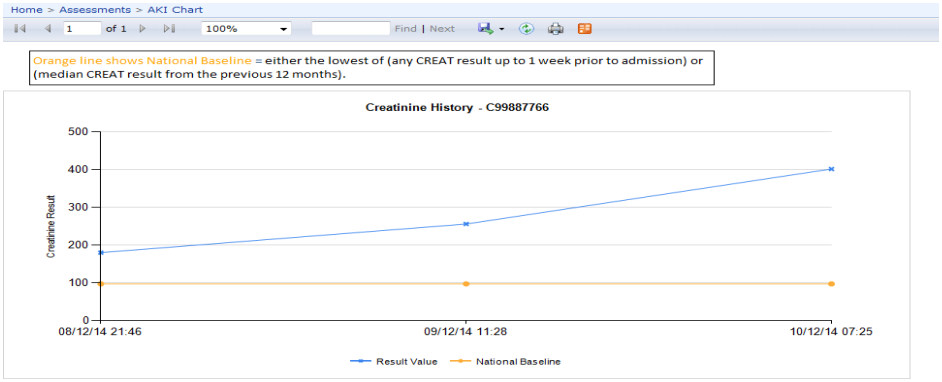


Figure 4.8 - SCr trend graph (including baseline SCr) linked from observation chart. SCr - serum creatinine.

4.3.3.3 Inclusions, exclusions and integration of the technology

Patients were included if ≥18 years of age, stayed at least one night in the medical unit over the 20-month period (2014-2016) and had at least one SCr repeated.

Exclusion criteria:

- patients moved directly from the A&E department to the ICU (as neither area uses the Patientrack© data system),
- aged <18 years,
- non-medical (General Surgery, Trauma & Orthopaedics, Obstetrics and Gynaecology) admissions and,
- discharged without spending a night in hospital.

Patients were followed-up until hospital discharge or in-hospital death. As all patients admitted had an APS calculated automatically by the Patientrack© system before any outcome had occurred, there were no missing demographic or physiological data at admission. Age was defined as that at study entrance. Ethnicity is known to be poorly recorded in routine data and was thus not investigated. Co-morbidities were identified from appropriate coding in the hospital's patient administration system, PAS (Semahelix®). Comorbidities will be defined by people having the relevant ICD for each diagnosis (including diabetes, heart failure and liver disease). Previous CKD was defined by the presence of at least two values of eGFR <60ml/min/1.73m^2 (CKD stage 3-5) at least 3 months apart with no intervening values ≥60ml/min/1.73m^2. In order to avoid classifying people as having CKD who had in fact had two isolated episodes of AKI, an upper limit of 1 year was applied. Medication status was determined from prescription of relevant drugs in the hospital's electronic prescription record and a notes review audit. From coding death, hospitalisations and AKI (ICD N17 code 'Acute kidney failure') were captured.

Barriers to integration and acceptance of CPRs and CDSS and expert guidance were considered prior to incorporation into the computerised clinical workflow,(485) as part of a multi-faceted AKI guideline implementation.(309) The controlled before-after design has been reported in previous impact analysis studies.(210, 427) Around the time of the study NHS England patient safety alert mandated an e-alert for all AKI cases which limited the timeframe

over which the study, a natural experiment, could be performed before the e-alert was introduced at the second (control) site.

4.3.3.4 Specific methodological issues

(i) Defining AKI

Baseline kidney function was defined using the method used for NHS England/Renal Association National AKI detection algorithm, developed from KDIGO.(3, 174) Assessments were made as to whether a previous SCr measure had been taken within a year. A SCr reference value was established depending on whether the previous SCr value was within 0-7 days or >1 week and <1 year, using the lowest value for the former and the median value for the latter. The ratio of the index creatinine value on admission to this reference value was then calculated and used to determine presence and degree of AKI. CKD was defined by at least two eGFR values <60 ml/min/1.73m^2 at least three months apart, without values above 60 ml/min/1.73m^2 in between those values. Both WSHFT hospital laboratories report eGFR calculated using the simplified MDRD equation with creatinine assays using calibration traceable to a standardised reference material - isotope dilution mass spectrometry. It was acknowledged as a limitation that some patients presented with AKI without a reliable baseline SCr on record. Furthermore missing data are an important issue when using routine data; using coding for co-morbidities relied on previous attendance at the Trust and could have underestimated disease prevalence.

(ii) CDSS and care bundle considerations

As explored in this chapter empirical evidence can guide CDSS implementation (summary table 4.1). When considering an alert for risk of AKI, a compromise between sensitivity and specificity was required to avoid overload including alert frequency. Reliance on using the National algorithm,(174) for baseline SCr estimation (median result over a year) is a pragmatic solution compared to the goal standard of expert clinical review, with inevitable instances of false positives or false negatives, with potential for a reduction in response.(451, 452, 463)

The multi-modal intervention aimed to display both those at risk and those with AKI in multiple electronic areas accessed by the clinical staff involved including nurses, health care assistants, junior doctors, pharmacists and critical care outreach staff. To achieve this the status of each patient was clearly displayed using the 'proximity compatibility principle,'(454) taking into account the visual field,(455) was employed during design within the clinical workflow in a number of key areas in the Patientrack observation system: Wardview with an AKI tile (figure 4.5) denoting baseline SCr and AKI stage or risk, MyView linking to a dashboard collating relevant patients (figure 4.8) and at the top of the Observation Chart with a link to a SCr graph (figures 4.7, 4.9). A high proximity task requires information integration, thus for the outreach nursing team in a separate Reports section, all patients in the hospital at risk, or with AKI, could be visualised and sorted by ward or consultant, with SCr results, APS score and its components. The Reports were designed for use for example at the daily elderly care handover meeting, led by a consultant geriatrician, to improve systematic awareness of the patient groups and to encourage compliance with the associated care bundles.

In addition to the flagging of patients, a care bundle of actions were automatically generated. The care bundles were designed using published guidance and previous studies to include widely accepted interventions that would each be deemed necessary to be delivering best practice care.(141, 473, 479) Each care bundle had four components, was designed to be quick to complete and compliance was monitored. Any member of the care team could complete the care bundle. At the time of study implementation a large number of both assessments and alerts were being or had recently been introduced across different platforms and are summarised in table 4.4 This highlights the complexity of healthcare delivery, particularly when a mix of CDSS and paper notes are in concurrent use.

Location	Direct work	Alert	Description
Patientrack	Clinical observations		Observations input onto Patientrack & then made visible on observation chart
Patientrack		NEWS thresholds	Dictates level eg observation frequency, seniority response
		AKI & AKI at risk	Care Bundle
		MUST assessment	Malnutrition screening tool for nurses
		MRSA assessment	Screening for nurses
		VTE assessment	Screening for doctors
		Dementia assessment	Screening for doctors
Patientrack or Paper	Fluid assessment		In/output charting
EPMA	Prescribing		Medications & Fluids
Paper	Medical notes		Proforma for admission notes followed by daily paper input; out/patient notes
Server	Clinic notes		Out-patient notes on server
ICU EPR	All		All Medical notes, Prescribing, Observations
PACS server	Imaging		Access to X-ray, CT, MRI, USS including reports

Table 4.4 Summary of hospital alerts and different systems in which clinician workflow sits.

For patients who had developed AKI the colour red was chosen, with amber for those judged at high risk to maximise hazard perception and priority.(456) The numbers of colours was minimised to avoid confusion.(444) Tiering of alerts was employed to prioritise those who had developed AKI over those who were at higher risk.(457) We aimed to mitigate habituation by minimising rates of false positive alarms, however, inevitably amongst a plethora of other signals over time, maintaining compliance was a concern. Evidence from research into warning labels was considered that resulted in a signal word – AKI – being used to draw attention and communicate level of the hazard.(458) Instruction and hazard statements using explicit terms were employed, for example for AKI cases ‘stop nephrotoxics’ were employed to provide clarity.(486, 487) Finally during the research, national drivers (NHS Patient Safety alert, NICE guidance and CQUIN)(480, 488) led to a rise in prominence of AKI in the hospital environment. These drivers were utilised to

ensure it was clear the area was a Trust priority, further emphasised by employing an AKI Champion, performing ward education and nursing Outreach work.

(iii) Education intervention

Raising awareness of AKI was multi-modal: at the intervention site posters were placed to explain the problem, groups at risk and what the intervention entailed. A suite of educational interventions included outreach critical care team reviews at the bedside with education of nursing, healthcare assistants and junior doctors. Each case of new AKI was reviewed either remotely or at the bedside by the outreach nursing team (composed of senior nurses with intensive care experience) that allowed a continuous improvement and awareness initiative to take place between the outreach nurses and the teams on the ward. Secondly, junior doctors underwent specific training through lectures and e-learning in AKI management and the intervention being introduced and were encouraged by their consultant colleagues to maintain bundle completion. Thirdly on the ward, Pharmacists not previously known to engage with the electronic observation system were shown how to access information and subsequently ensure AKI was highlighted with appropriate prescribing. Finally a diagnosis of AKI was embedded into the discharge proforma raising awareness for primary care, the patient and for subsequent admissions (summary in appendix table A4.1). At the control site only the e-learning module was introduced, whilst cross site senior management emphasised the importance of the topic, cascading generic information from the national drivers.

Prior to initiation of the study a qualitative analysis was planned to reveal behavioural and situational barriers to implementation of the intervention and provide insights for future studies. This would take the form of semi-structured interviews aiming to explore the challenges surrounding implementation of an alert and associated care bundles from the perspective of the multidisciplinary team, and how these could be addressed. Such qualitative methodology can

allow investigators to identify, via in-depth analysis, both personal and contextual factors.

4.3.3.5 Outcomes

The primary outcome to assess the APS CPR was change in incidence of HA-AKI within seven days of hospital admission following intervention when adjusting for baseline differences in age, co-morbidity and prior attendances, between intervention and control sites. Other outcomes assessed separately those who went on to develop HA-AKI including mortality in-hospital and within seven days of admission, maximal increase in AKI KDIGO Stage, increase in SCr from admission, and ICU escalation. These outcomes were also assessed in the subgroup patients of patients highlighted by the CPR at high AKI risk (APS ≥5 AMBER). Outcomes for patients with CA-AKI were assessed to investigate the effect of the AKI alert at admission, only present at the intervention site. Process measures were collected to attempt to explain the effects of the intervention including drug prescribing, completion of the suggested care bundles and a notes review of a randomly selected collection of cases of HA-AKI at both sites, pre and post intervention.

4.3.3.6 Statistical analysis

Descriptive statistics reveal the case mix including age and comorbidities between the two sites and identify the characteristics of people developing CA-AKI and HA-AKI versus those without these outcomes. Comparison was made between the occurrence of such events and to include the clinical identification and coding of AKI pre and post intervention. T-tests compared the difference if continuous data was normally distributed; the Mann-Whitney U test was employed for non-parametric continuous data comparison; χ^2 tests were used to compare binary outcomes.

Generalised estimating equations (GEE) with logistic link function were chosen to estimate the risks of binary outcomes (incident HA-AKI, in-patient mortality, death within seven days after hospital admission, progression of AKI stage, escalation to ICU) between the intervention and control hospitals. GEE,

described by Liang and Zeger, takes into account the clustering effects of multiple hospital admissions for the same individuals.(489) Generalised linear mixed models were employed to analyse continuous variables (change in SCr and hospital length of stay) including individual patients as a random effect.(490) To avoid potential bias due to existing background differences between hospitals and temporal changes, we adopted the difference-in-differences approach by including time variable (before-after) and its interaction with the hospitals.(432, 433, 490-493) All models were adjusted by age and co-morbidities (heart failure, CKD, vascular disease, hypertension, diabetes mellitus and liver disease). The estimates of the intervention with a time period interaction term with associated 95% confidence intervals were reported.

Difference in differences

The Difference-in-differences approach was first described in the economics literature,(432) but has increasingly been employed to assess healthcare interventions. Before-after designs are valid only if there are no underlying time-dependent trends in outcomes unrelated to the policy or healthcare change. If clinical outcomes were already improving, then using a pre-post study would lead to the erroneous conclusion that the policy was associated with better outcomes. The difference-in-differences design addresses this by using a comparison group that is assumed to experience the same trends but is not exposed to the policy change.(433) Outcome after and before the policy are compared between study group and the control group, which allows the investigator to subtract out the background changes in outcomes. Estimates are derived from regression models rather than simple subtraction, allowing adjustment for factors such as patient or hospital characteristics that may differ between the groups.(433) The association between policy implementation and outcomes is estimated by examining the interaction between the pre-post and exposed-unexposed variables.(433) This is important for example in an area such as AKI, with national and local drivers to improve outcomes that could influence a simple before-after study.(480)

The increasing popularity of difference-in-differences in health policy and medicine reflects the credibility of designs and ease of implementation and estimation.(494) An example where this approach has been used in the US was to evaluate the impact of Medicare's bariatric surgery reimbursement being limited to "Centers of Excellence." A study compared complications in Medicare patients (subject to the policy) and commercially insured patients (not subject to the policy) before and after the initiation.(495) Prior studies, using a simple "pre-post" study design found a beneficial impact, but failed to account for secular trends toward improved outcomes.(496) However, using difference-in-differences, similar declines in complications for both groups were found, suggesting no specific beneficial effect of the program.(495) Other recent examples include the impact of health insurance expansions, payment policy, malpractice reform, resident work hour reform, changes in clinical practice, smoking laws and electronic medical record implementation.(497, 498) In the UK, Morris and colleagues used this methodology to assess the impact of centralised stroke services in London and Manchester.(490) Their findings suggested an improvement in London compared to the rest of England when accounting for temporal changes.

The two main assumptions of difference-in-differences analysis are parallel trends and common shocks. The parallel trends assumption states that the trends in outcomes between the treated and comparison groups are the same prior to the intervention.(499) If true, one assumes these parallel trends would continue for both groups even if the program was not implemented. This can be tested empirically by examining trends prior to implementation. In a regression model, this is evaluated by assessing the significance of the interaction term between time and policy exposure in the pre intervention period.(433) If the prior trends are significantly different a difference-in differences analysis would be biased and another comparison group would be required. Secondly a 'shock' is an unexpected event affecting a system. The common shocks assumptions state that any events occurring during or after the time the policy changed will equally affect the treatment and comparison groups. A limitation to implementing difference-in-differences design is the difficulty finding a control group for which these assumptions are met.(433) Of

significant importance, the presented ICE-AKI study used a control site exposed to 'common shocks' nationally, (such as a drive to address AKI care) and locally, being part of the same, Trust any financial and staffing issues were common to both.

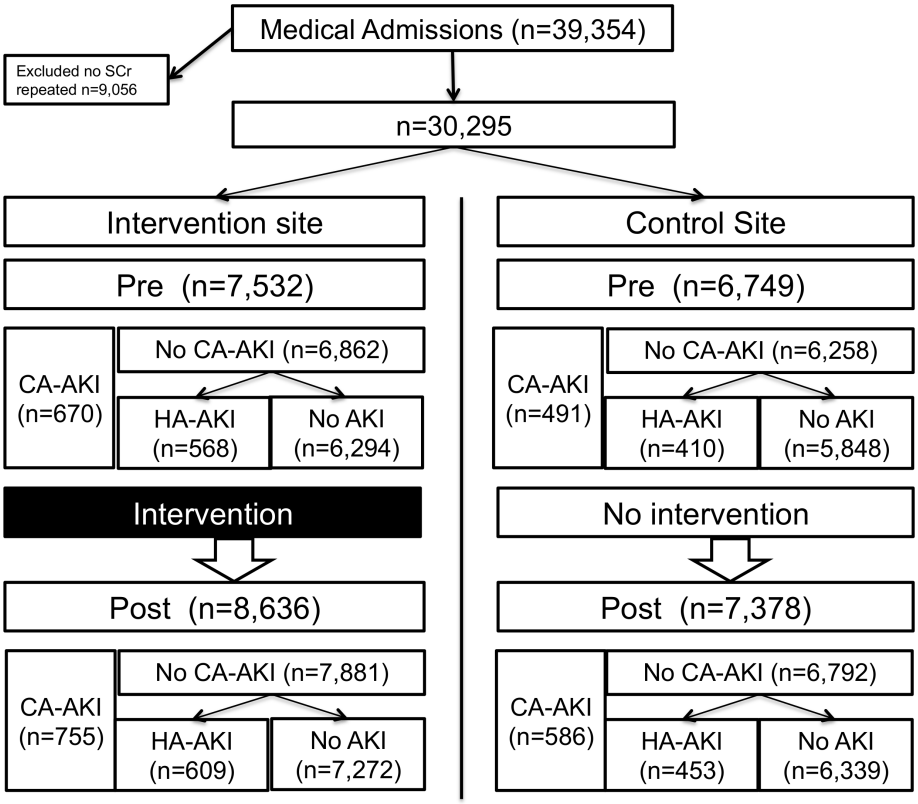
To test whether the groups' case mix did not change differentially over time, we examined for differences between patient characteristics before and after implementation in both groups (table 4.6). The changes and risk of difference events before and after were tested separately for intervention and control hospital. Data were extracted fully anonymised on Microsoft® Excel® and all analyses were conducted using SAS (V9.4®).

4.4 Results

4.4.1 Summary of results

Over the 20-month study period 30,295 admissions were analysed (figure 4.9) with a mean age of 74·5 (standard deviation ±17·0), length of stay 11 days (±13·1) and in-hospital mortality 8·3%. AKI was present on admission (CA-AKI) in 8·3% (n=2,502) and 7·3% (n=2,040) developed HA-AKI. Table 4.5 shows the sites were relatively well matched at admission pre and post intervention, though the intervention site had higher lengths of stay, age, admission tachypnoea and some reported co-morbidities (liver disease, hypertension and vascular disease). Across all medical admissions between periods, at both sites there was no overall change in mortality, whilst length of stay reduced (table 4.6). More patients were escalated to ICU at the intervention site in the second period (P=0·013) though on difference-in-differences analysis there was no significant change in mortality or escalation to ICU.

Figure 4.9 Consort diagram inclusion and exclusions. CA-AKI – Community-acquired AKI, HA-AKI – Hospital-acquired AKI, SCr – serum Creatinine.



4.4.2 Incidence of HA-AKI

Difference-in-differences analysis suggested a reduction in HA-AKI at the intervention site (OR 0.990, 95% CI 0.981-1.000, P=0.049). Unadjusted incidence of HA-AKI did not change significantly at the intervention (8.28% pre, 7.73% post intervention, OR 0.928, 95% CI 0.824-1.045, P=0.223) or control sites (6.55% pre, 6.67% post intervention, OR 1.019, 95% CI 0.888-1.170, P=0.805) (table 4.6).

Table 4.5 Demographics and clinical characteristics of patients pre and post intervention, with p values fro within site changes over time.

	Intervention Site			Control Site		
	Pre (n=7,532)	Post (n=8,636)	P value	Pre (n=6,749)	Post (n=7,378)	P value
Age, mean (SD)	75·1 (±16·9)	74·2 (±17·4)	0·0012	74·6 (±16·7)	74·1 (±16·9)	0·1492
NEWS, median (IQR)	1 (0-3)	1 (0-3)	0·0088	2 (0-3)[#]	1 (0-3)	0·0131
Respiratory rate ≥20	27·2% [#]	25·8%[#]	0·054	20·2%	18·3%	0·003
<Alert on AVPU scale	1·3%	1·3%	0·944	1·2%	1·0%	0·339
CKD	47·7%	46·5%	0·122	49·1%	45·9%	<0.001
Diabetes	22·3%	23·7%	0·045	22·6%	23·2%	0·367
Heart failure	24·2%	24·4%	0·797	23·1%	23·4%	0·705
Liver disease	2·7%[#]	3·1%[#]	0·110	1·9%	2·0%	0·903
Hypertension	62·1%[#]	59·6%[#]	0·001	56·9%	55·1%	0·034
Vascular disease	10·3% [#]	10·3%[#]	0·959	6·5%	7·6%	0·014

IQR – Interquartile range, SD – standard deviation. AVPU – best response: Alert, Vocal, Pain, Unresponsive, CKD – chronic kidney disease (baseline estimated glomerular filtration rate <60mls/min), NEWS – National early warning score, Respiratory rate –breaths/minute. [#] = significant difference (P<0.05) between sites during same period. T-test, Mann Whitney U tests or χ^2.

Table 4.6 - Pre and post intervention results for all included patients.

	Intervention Site			Control Site		
Metric	Before (n=7,532)	After (n=8,636)	P value, OR (95% CI)	Before (n=6,749)	After (n=7,378)	P value, OR (95% CI)
In-patient Mortality	9·82% (n=740)	9·06% (n=782)	0·100, 0·914 (0·822 - 1·016)	7·13% (n=481)	7·10% (n=524)	0·974, 0·996 (0·876 - 1·133)
7-day mortality	4·20% (n=316)	3·61% (n=312)	0·060, 0·856 (0·730 - 1·004)	2·84% (n=192)	2·71% (n=200)	0·645, 0·952 (0·778 - 1·163)
ICU escalation	2·46% (n=185)	3·10% (n=268)	0·013, 1·272 (1·052 - 1·538)	2·44% (n=165)	2·32% (n=171)	0·659, 0·947 (0·762 - 1·176)
Peak SCr rise	11·65 (±35·7)	10·57 (±32·0)	0·045	9·05 (±27·6)	9·22 (±27·4)	0·715
Length of stay	12·31 (±14·7)	11·50 (±13·1)	<0·001	10·69 (±13·4)	9·57 (±10·9)	<0·001
	Before (n=6,862)	After (n=7,881)		Before (n=6,258)	After (n=6,792)	
Incident HA-AKI	8·28% (n=568)	7·73% (n=609)	0·223, 0·928 (0·824 - 1·045)	6·55% (n=410)	6·67% (n=453)	0·805, 1·019 (0·888 - 1·170)

Mean (± standard deviation). HA-AKI – hospital-acquired AKI, ICU – intensive care unit, SCr – serum creatinine.

Metric	P value, odds ratio (95% CI)
In-patient Mortality	0·328, 0·995 (0·985 - 1·005)
7-day mortality	0·240, 0·996 (0·989 -1·003)
ICU escalation	0·255, 1·003 (0·998 - 1·009)
Incident HA-AKI	0·049, 0·990 (0·981-1·000)

Difference in differences estimates Intervention vs Control site.

4.4.3 HA-AKI outcomes (table 4.7)

All outcomes tended to fall at the intervention site and rise at the control site. In cases of HA-AKI, unadjusted in-patient mortality significantly decreased at the intervention site (27·5% pre vs 21·7% post, P=0.021, OR 0·731 [0·560-0·954]). This remained significant when accounting for baseline differences and control comparison using difference-in-differences analysis (P=0·038, OR 0·92 [95% CI 0·858-0·996]) (figure 4.10). A similar reduction in mortality by day seven was found. ICU escalation, progression to stage 3 AKI and maximal increase in SCr reduced at the intervention site, whilst increasing at the control site, without reaching statistical significance on difference-in-differences analysis. Length of stay was not seen to change at either site or over time in this cohort.

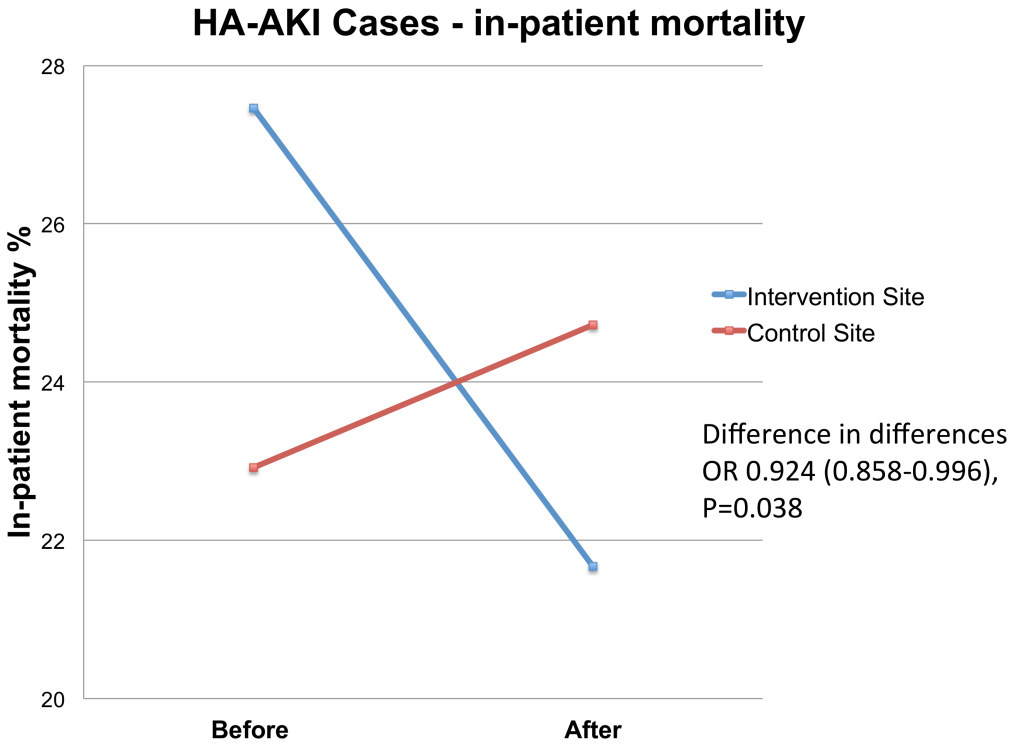


Fig 4.10 – In-patient mortality in cases who developed HA-AKI before and after the intervention.

Table 4.7 Pre and post intervention outcomes for HA-AKI cases.

	Intervention Site			Control Site		
Metric	Before (n=568)	After (n=609)	P value, OR (95% CI)	Before (n=410)	After (n=453)	P value, OR (95% CI)
In-patient mortality	27·46% (n=156)	21·67% (n=132)	0·021, 0·731 (0·560-0·954)	22·92% (n=94)	24·72% (n=112)	0·576, 1·104 (0·807-1·511)
7-day mortality	16·20% (n=92)	10·51% (n=64)	0·004, 0·608 (0·432-0·855)	9·02% (n=37)	12·58% (n=57)	0·101, 1·451 (0·937-2·247)
Stage 3 AKI	8·27% (n=47)	6·57% (n=40)	0·268, 0·779 (0·503-1·208)	4·15% (n=17)	5·52% (n=25)	0·429, 1·350 (0·718-2·538)
ICU escalation	7·39% (n=42)	6·90% (n=42)	0·821, 0·928 (0·595-1·446)	7·07% (n=29)	10·6% (n=48)	0·074, 1·557 (0·962-2·521)
Peak SCr rise	71·0 (±77·4)	66·1 (±76·3)	0·269	64·6 (±61·0)	64·3 (±59·0)	0·939
Length of Stay	14·7 (±15·4)	15·0 (±13·9)	0·708	15·4 (±14·3)	13·7 (±12·9)	0·064

Mean (Standard deviation), HA-AKI – hospital-acquired AKI, ICU – intensive care unit, SCr – serum creatinine, Stage 3 – KDIGO Staging x3 increase SCr.

Metric	P value, Odds Ratio (95% CI) Intervention vs Control
In-patient mortality	0·038, 0·924 (0·858 - 0·996)
7-day mortality	<0·001, 0·907 (0·859 - 0·957)
Max Increase SCr	0·592
Stage 3 AKI	0·319, 0·980 (0·942 -1·020)
ICU escalation	0·107, 0·969 (0·931 -1·007)
Length of Stay	0·194

Difference in differences estimates.

4.4.4. Patients with CA-AKI and those flagged at high risk

There were no significant changes in outcomes in those with CA-AKI at either site (table 4.8 and appendix table A4.2). At the intervention site of those flagged high risk (AMBER alert) 14% developed HA-AKI, vs 5% with no flag (GREEN) (P<0.001). In this group of patients alerted by the APS CPR at high risk (≥5 points) in-patient mortality decreased following intervention (14% pre vs 11% post, P=0·008, OR 0·78 [0·66-0·94]) with no change at control site (table 4.8 and appendix table 4.2).

	Intervention Site			Control Site		
CA-AKI Cases In-patient mortality	Before (n=670) 23%	After (n=755) 23%	OR (95% CI), P-value 1·01 (0·79-1·29), 0·95	Before (n=491) 19%	After (n=586) 17%	OR (95% CI), P-value 0·86 (0·63-1·17), 0.34
AMBER (APS ≥5) In-patient Mortality	Before (n=2,057) 14%	After (n=2,351) 11%	0·78 (0·66-0·94), 0·008	Before (n=1,810) 10%	After (n=1,851) 10%	0·96 (0·78-1·20), 0·74

Table 4.8. In-patient mortality in cases with CA-AKI and in those flagged at high-risk on admission by the CPR (AMBER - APS ≥5). CA-AKI – community-acquired AKI, OR – odds ratio.

4.4.5. Process measure review

Overall electronic bundle compliance, defined as completion and submission was 15% in those at risk and 26% in those with AKI. However, a number of significant process measure improvements were seen at the intervention site vs control. For example, on interrogation of the electronic prescribing system significantly more patients had documentation of stopping potentially nephrotoxic medications with AKI given as the reason as well as stopping ACEi/ARB classes of drugs in those flagged (both P<0·001). In a notes review of 181 patients, a number of measures improved at the intervention site in before-after and between site comparisons, including AKI documentation (P=0·033) (summarised in tables 4.9A/B and appendix table A4.3).

	Intervention site			
	Pre-intervention	Post intervention	P value	OR (95% CI)
Documented AKI	21%	55%	0·002	4·533 (1·763-11·657)
Repeated U&Es	67%	89%	0·021	3·766 (1·218-11·640)
Drugs checked	51%	55%	0·831	1·145 (0·493-2·660)
Fluid Plan	44%	64%	0·087	2·211 (0·935-5·224)
Search for cause	40%	46%	0·666	1·275 (0·544-2·988)
Appropriate special tests	34%	37%	0·822	1·143 (0·468-2·794)
AKI documented on discharge	20%	42%	0·049	2·857 (1·031-7·919)

Table 4.9A Pre vs post intervention summary of process measures from notes review. U&Es – urea and electrolytes.

	Post intervention period			
	Intervention	Control	P value	OR (95% CI)
Documented AKI	53%	30%	0·033	2·710 (1·141-6·434)
Repeated U&Es	89%	72%	0·062	3·019 (0·961-9·487)
Drugs checked	55%	23%	0·004	3·960 (1·573-9·971)
Fluid Plan	64%	54%	0·388	1·522 (0·645-3·589)
Search for cause	46%	21%	0·023	3·148 (1·224-8·095)
Appropriate special tests	37%	16%	0·050	3·048 (1·100-8·441)
AKI documented on discharge	42%	29%	0·322	1·786 (0·665-4·798)

Table 4.9B Comparison of process measures between intervention and control sites after intervention. U&Es – urea and electrolytes.

4.4.6 Qualitative review

Following initial implementation, a thematic analysis was conducted with 20 interviews of the members of the multidisciplinary clinical team.(496) The audio-recordings were uploaded to a secure server and transcribed and then checked by two researchers working with the author. Using constant comparison, a technique derived from grounded theory,(497) transcripts were compared within and between each other aiding the iterative search for themes to be reviewed, defined and named.

Themes identified from the staff interviews were used to inform and optimise implementation of the intervention and be used to inform and provide recommendations for future similar studies. Major themes were elicited. Positive experiences included feelings the intervention was *'a useful tool'* and those interviewed expressed ways through which it affected the care they provide including prioritisation: *'If I see an AKI red alert then my first thought is to always go and check the person's bloods myself'* (pharmacist), and early identification '*Yes, it's a warning and can help to prevent'* (healthcare assistant). '*Well obviously, if we're able to identify them, then we're able to go and review them and then to put in place the right interventions following the bundle'* (specialist nurse). *'So in the morning you can direct yourself as to which patients you are going to see first….it helps prioritise where we go as well so it does.. yes impact on the care we provide because we get to see the right patient first'* (Consultant). '*Yes… its just another visual prompt.. it does get busy especially in this environment its really, really difficult to be able to have eyes on everything so you do use it'* (deputy ward sister)
Effective monitoring was a second theme: '…*it really does.. it gives you details what you should be doing, if you should be cutting out nephrotoxic medications and questioning any of the practices… and so it kind of really focuses your mind into what's best practices and what you should be doing for this person'* (nurse) and *'…better, I think definitely better, because we have to go up to them every hour, we are going there much more to see their output so it helps us to see whether the patient is declining…'* (healthcare assistant)

The traffic light system was thought to be simple and effective providing the right amount of information with one interviewee stating *'Red… you cannot miss it'* (nurse) *and another saying 'Just the traffic light system, it's quite simple'* (Pharmacist) *and 'I think it just gives you the right amount of information to get an idea as to how sick that patient is.'* (Consultant).

However, significant barriers to use were also revealed. The hospital's IT capabilities were found as a problem, along with the availability and access to computers, lack of training and confusion regarding responsibilities.
'I think getting onto the system…and you know just trying to access it all because it is very much having to get onto a computer' (healthcare assistant)
'access to a computer, not having enough computers and chargers' (healthcare assistant), *'I think that it [the system] can be slow to load up'* (Consultant) and *'sometimes it does not work properly or not proper access* (pharmacist).

A number of interviewees thought the care bundles were a '*tick box exercise'* with some confusion related to who should fill them out. Pharmacists thought it more appropriate for nurses and doctors, whilst healthcare assistants thought it a 'nursing job' and nurses thought it was indented for doctors to use and complete:
'I am aware but I can't say that I would look at it' (pharmacist)
'It's for the Doctors to do, so we don't tend to go into it' (deputy sister)
*'[I'm] aware of it but things on it are mainly a nursing job' (*healthcare assistant*)*
'It's a tick box exercise to say right ok yep - that doesn't really prove whether it has or hasn't been done.' (deputy ward sister)
'I don't find it hugely helpful that we have to click on each thing to say that we've done it.' (doctor)

In general, staff reported there could have been more training on the AKI alert system and care bundles and indicated specific training is needed and would be helpful.
'[we had] no training' (pharmacist)

'some online training' (healthcare assistant)
'not enough and ensure that everyone gets the training' (healthcare assistant)
'….there is a need for it [training], potentially useful' (junior doctor)

During the interviews the staff provided a number of ideas that could potentially improve the system and assist future practice and studies.

'Make it more accurate… if renal function improves they (patients) remain RED in the system.' (pharmacist)
'create a link between electronic prescribing and patientrack [the physiological observation system]' (pharmacist)
'patient track could give you a summary of the alerts and what is missing for each patient.' (Consultant)
'why there is patientrack and patientrack reports, tedious to go from one screen to the other to capture the info.' (Consultant)
These suggestions include: functionality (for example if renal function has improved the alert could reflect this), integration with other electronic systems such as prescribing and finally speed (including not having to flip between different screens).

4.5 Discussion

4.5.1 Main findings

This is the first impact analysis study combining an AKI CPR and an AKI e-alert in hospitalised patients. The intervention was associated with a decrease in the incidence of new AKI but, importantly a significant improval in survival in those individuals who developed HA-AKI was also found. This may reflect a synergistic benefit of highlighting those at risk and de novo AKI, implying best practice would be institution of a care bundle in both groups, in tandem with a multimodal service improvement approach. Other outcome measures including escalation of care and progression of AKI were suggested. There was no change in outcome following introduction of an e-alert for patients with established AKI at admission to hospital, indicating the measures introduced did not affect the underlying disease process.

The primary aim of the study was to investigate whether an AKI CPR could prevent HA-AKI. On a difference-in-differences analysis a reduction was achieved. The treatment effect may have been limited due to a lack of specificity in the CPR - in derivation and validation the APS had moderate discrimination (AUROC of 0.65-0.72) and to avoid alert fatigue, only patients with a score ≥5 were alerted to (84, 348). Furthermore, as AKI was only diagnosed through change in SCr, known to lag behind insult (30), it is possible a proportion classed as HA-AKI reflected community-acquired injury. Further important findings are that both in those highlighted as at high risk of developing AKI (AMBER) and in patients who subsequently developed HA-AKI, improved mortality outcomes were found, which may reflect systematic early recognition, with prompt care initiated even without documented evidence of a completed care bundle. Indeed without these alerts many of these patients would not have triggered an urgent response: for example median national early warning score (NEWS) was only 2 in the AMBER group, yet mortality ranged 11-14%. Based on the NEWS alone, such patients would not have been flagged high risk. This study provides evidence that application of best practice can translate into improved outcomes even in a heterogenous group at high risk of complications.

4.5.2 Comparison with other studies and intervention uptake

To date, the only RCT of an AKI e-alert showed no outcome benefit, though the intervention in that study was limited to an electronic link to practice guidelines (170). In two UK studies in patients who had developed AKI a mortality benefit was reported in the 22-26% of those with a completed care bundle (141, 396). A recent large before-after study in the US suggested a significant mortality benefit for the use of an AKI flag (140). Though rarely reported, other CPR impact analysis studies have described successful integration into a CDSS (500, 501).

The majority of studies reporting on the effects of alerts incorporated as part of CDSSs are medication related (444), however, such recommendations are often ignored (460-462). Relatively little evidence explains why some succeed, with a paucity of evidence to demonstrate convincing improvement in patient outcomes (309, 464-466). However, two systematic reviews (309, 467), suggested three independent factors influence success:

- automatic provision of decision support as part of clinician workflow at the time and location of decision-making,
- provision of recommendations rather than just assessments and,
- computer based decision support.

These factors and others were considered during study design, which provides a proof of concept for future investigations of complex healthcare interventions. Alerts could be quickly acknowledged, with a single alert at admission for those at risk, with a further alert on patients who met AKI criteria within the nursing clinical workflow (309, 408). A cut-off on the APS CPR was chosen to minimise false positives to avoid blunting and eventual elimination of responses (452, 463). Optimal visual field positioning and the use of appropriate colours (red and amber) were chosen as they are associated with an increase in hazard perception and prioritisation (454-456, 502). Despite these measures, compliance with electronic submission of care bundles was low ranging 15-26%, similar to previous AKI studies (396, 503, 504). This could have been improved by making the alert interruptive however, this may have increased frustration and alert fatigue (396, 505, 506). Doctors were expected to submit the completed bundle however it was not part of their workflow. However, crucially the intervention was multimodal: highly visible flagging in multiple electronic areas accessed by all members of the clinical team, alongside education, that may have changed group behavior not captured by a formal bundle submission. Whilst establishing conclusive proof of behavioural change is challenging, we found significant improvement in process measures, such as stopping potential nephrotoxic medication.

The qualitative interviews provided important insights into why the intervention may have helped improve outcomes: the technology was visible and frontline multidisciplinary staff were well aware of it and importantly had incorporated its use as part of the daily delivery of clinical care. However, the interviews also revealed shortcomings of implementation including education, roles and responsibilities, lack of time in a busy environment and with the overall infrastructure of the IT system in use in the clinical arena. Finally the interviews suggested a number of potential improvements that future studies and iteration of the technology could address: ongoing education, data linkage and finally further consideration of time as the crucial element in the busy clinical environment.

Study strengths and weaknesses

This natural experiment introducing an electronic alert intervention at one site within a similar geographical area to a control site allows for unrelated change in disease occurrence and practice, with adjustment for confounders and has been described in previous impact studies (210, 427). HA-AKI associated mortality at the two sites (20-26%) was similar to the other UK study showing a mortality benefit (396). The process measure evaluation identified increased documentation, coding of AKI and stopping of nephrotoxic medications at the intervention site, which adds to the main data analysis suggesting both improved recognition and clinician behaviours following intervention (tables S5-6).

The study took place as part of an information technology service improvement intervention that did not allow for a RCT design, with selection bias a concern. Between sites in both periods overall mortality was higher at the intervention site, as was the rate of HA-AKI, which may reflect differences in co-morbidities and socioeconomic factors. However, the before-after controlled design using difference-in-difference analyses is accepted as a way of detecting the effects of an intervention, and to control for confounders and secular trends (507, 508). Outcomes did not improve in the control hospital suggesting other external factors did not significantly influence outcomes seen at the intervention site. However, given the difficulty of eliminating bias, a

single study is unlikely to be definitive and replication and synthesis of evidence across studies is needed to support inferences about effectiveness (484). A lack of power from two sites over ten-month periods could explain why outcomes including ICU escalation and AKI severity, though improving at the intervention site, did not reach statistical significance – though for example in the second period significantly fewer patients with HA-AKI at the intervention site were escalated than at the control site. Generalisability may be limited by the study population on the South-Coast of England, being older (mean age 74) than the English NHS average admitted under general medicine (65 years), though the National average geriatric admission age is 78 (509). Both sites are non-specialist general hospitals and the analysis used medical patients – future studies could examine generalisability to different hospital populations. However, our population probably reflects the majority of acute admissions in the UK given that the mean age of patients with CA-AKI in the RISK study being 75 years (unpublished, personal communication with study authors). The study did not assess whether impact was sustained or whether the intervention could be successfully introduced at the control site – the recommended fourth phase of impact analysis and implementation (485). Other parts of the country such as inner cities would however, have higher variety of ethnic diversity which could be studied. The before-after analysis was by admission, not individual and this approach can have shortcomings due to the effects of single patients having multiple admissions. Also calculation of the APS CPR past history relied on previous coded events (from previously attending hospital) and could therefore have been incomplete. The predictive accuracy of the CPR is could be improved by planned linkage with primary care as well as provide information of GP prescribed drugs, known to be associated with an increased risk of AKI. Finally, the low level of bundle compliance could be investigated further for example with ongoing qualitative work to understand barriers to implementation that cannot be captured in synthesis of quantitative data.

4.5.3 Conclusions

In an impact analysis performed on general medical admissions, introduction of an AKI CPR was associated with a reduction in HA-AKI. In the context of a multi-faceted intervention, including an electronically generated care bundle for patients at risk of AKI using a CPR and e-alert for patients with HA-AKI, a significant reduction in mortality was demonstrated. An e-alert had no affect on outcomes for patients admitted with established AKI from the community. Further studies in the field are warranted to confirm or refute these findings and assess for generalisability to other hospital populations, healthcare systems, and for sustainability.

Chapter 5 – Discussion

5.1 Introduction

This concluding chapter will discuss the principal findings of the thesis, with initial focus on the significance of AKI, issues with recognition and role of prediction models. This will be followed by an outline of the major findings of the systematic review of AKI prediction models and external validation of one such model. The major part of the section will explore the relevance of the impact analysis study performed, with lessons for future research. Strengths and limitations of the research and an outline of future directions will conclude the discussion.

5.1.1 AKI is a global problem and the role of electronically integrated prediction models

Chapter 1 outlined the gathering knowledge of the significance of AKI. Consensus definitions now provide clear diagnostic criteria, raise awareness and form the platform for rapidly increasing research efforts.(3) That there are few reports of interventions to improve outcomes may be due to a combination of factors. First, diagnostic criteria - creatinine and urine output - have limitations and may lag significantly behind insult. Second, AKI is a syndrome rather than a disease, with diverse underlying aetiologies making one intervention unlikely to be globally helpful. Third the silent nature of AKI means evolution may not be promptly identified, unlike, for example, a myocardial infarction presenting with central crushing chest pain.

Prediction models hold promise to systematically improve prediction to allow prevention and/or early recognition of AKI. Across medicine the number of such models has rapidly increased but unfortunately most have shortcomings in design and reporting and very few have gone through the recommended rigorous process of validation and impact analysis prior to implementation.(410) Information technology including CDSS, the role of e-alerting and integration of prediction modelling could be of significant benefit

but are not well described, with one recent review reporting only two impact analysis studies.(500, 501)

5.1.2 Available prediction models – a systematic review

Chapter 2 explored acute hospital AKI prediction models in specialist and general settings. An incredible 53 models have been published mostly in the fields of cardiac surgery, CI-AKI and liver transplantation, frequently using similar predictors, with no impact analysis. General hospital settings account for the majority of admissions and AKI, but only five models in these areas have been externally validated, with at best moderate discrimination and often absent calibration data. As with the specialist models, frequent shortcomings in design and reporting may improve with recent TRIPOD guidance.

5.1.3 External validation of a prediction model

In only the 5th external validation of a general prediction model, chapter 3 explored performance of the AKI Prediction Score (APS). Using published guidance the model was tested in medical and surgical cohorts with and without a baseline creatinine including analysis of calibration through plots. As is frequently described, external validation found reduced discriminatory performance. With more outcome events than the derivation study, it is likely that this study revealed overfitting in the original study though calibration was satisfactory. A deficiency of the validation is the lack of model updating which could be explored in the future, for example with more comprehensive chronic co-morbidity using primary care data linkage and acute physiological changes, incorporating repeated observations. Pragmatically the APS, with robust predictors of AKI that could be generated electronically was chosen to further study through an impact analysis following implementation in clinical practice.

5.1.4 Impact analysis and implementation

The unique impact analysis study provides a proof of concept for future integration of prediction models and e-alerting into hospital CDSS. This

complex healthcare intervention included the use of appropriately designed care bundles incorporated into the electronic healthcare record. To account for well-known drawbacks of non-randomised studies the study employed difference-in-differences methodology. Significant findings included a reduction in incidence of HA-AKI and of great interest a reduction in associated mortality in those who did develop HA-AKI at the intervention site. Importantly prospectively planned process measures were investigated and improvements, for example, in prescribing behaviour provide insight into how, or why, outcomes improved.

From quantitative data including low compliance of electronic bundle submission and qualitative work including staff interviews the intervention met challenges. Firstly, the research took place whilst an increasing number of alert and tasks were introduced, particularly into the nursing workflow. Individually these tasks are straightforward, however, in a busy clinical environment frequent addition of new requirements unsurprisingly led to resistance. Whilst the intervention met recommendations being part of the clinical workflow for nurses,(309) this was not the case for doctors who were expected to fill in the electronic care bundles separate to their paper based workflow. A fully functional electronic record, may have improved concordance. The bundle was not interruptive, however, such alerts can potentially increase frustration and alert fatigue.(505, 506) Furthermore, multiple alert locations coupled with education and multi-disciplinary team involvement provided a comprehensive awareness initiative. For example, the critical care outreach team performed distant monitoring and bedside education that supplemented the ward based clinical team. Information gathered from interviews suggested that the intervention was recognised and deemed useful to the multidisciplinary team in their daily practice. However, shortcomings including lack of education, clarity over roles and responsibilities and problems with access to the existing technology suggested areas to improve. Furthermore using evidence from the literature in the field, to improve future uptake would require a number of changes to current practices summarised in table 5.1.

Switch Clinician work from paper to EHR with care bundles in clinical workflow or
Place responsibility for bundle completion on nurses in whose workflow they currently sit
Improve education and provide feedback on performance / compliance as part of a quality improvement project – limited feedback existed during the study, but progress could be made relaying benefits of the intervention
Link e-prescribing to the AKI alerts to further highlight risk of nephrotoxic medications
Future linkage to primary care to enhance co-morbidity status to improve accuracy

Table 5.1 Potential improvements to the intervention. EHR – electronic health record.

5.2 Strengths and weaknesses of this study

5.2.1 Strengths

This thesis adds new scientific knowledge to the field. The systematic review is the first of its kind in general populations, where the majority of hospital AKI occurs and the external validation study is one of few performed. The impact study, to the author's knowledge the first in the field, provides a template for future studies looking at electronic integration of a CPR with an alert. Its findings of outcome benefit are significant and demand replication. The study addressed a relevant research question in the acute hospital setting, with population, interventions, and outcome measures explicitly described and relevant to the UK and other advanced healthcare systems. The large sample size in a general medical cohort commonly encountered outside teaching hospitals is a significant strength, whilst acknowledging that results require replication and should not be applied to specialist patient groups. The acute medical environment is a challenging area to target and measure improvements in care with research in this field relatively sparse,(510) perhaps owing to the heterogeneity of the patients, lack of a coherent research structure and the acuity of the patients for example making recruitment and consent challenging.

Despite widespread costly implementation of electronic alerting, robust research demonstrating clear outcome benefit is lacking(442) and this study provides important, somewhat conflicting insights. This study is the first to date to describe the effects of the nationally recommended AKI alert compared to a contemporaneous control site without an alert, for patients at

admission with established CA-AKI. This demonstrated no change in outcomes for the alerted group both compared to before implementation and with the control site. A combination of factors probably exist for this finding, summarised in table 5.2, including a proportion with established injury, poor compliance with the intervention, patients having limitations of care in place and insufficient power to detect significant changes when including a heterogenous group. One conclusion would be that expecting an alert to improve outcomes for patients who have already fulfilled AKI criteria, with diverse underlying aetiologies may be impossible, further emphasising the need to improve earlier recognition of those at risk.

Injury established by admission and not affected by an alert at this point
Poor compliance with the intervention
Compared to HA-AKI, larger proportion of patients with 'AKI' represented dehydration with little room to improve outcomes thus diluting any treatment effect
Patients having limitations of care in place
Insufficient power to detect significant changes in sample size

Table 5.2 Factors that could account for lack of effect of an AKI alert for patients with community-acquired AKI. HA-AKI – hospital-acquired AKI.

In contrast, combining a CPR for those at risk, with an AKI alert for new HA-AKI, suggested outcome benefits in the combined group who met HA-AKI criteria within a week of admission. Clinically this makes sense that both patients at risk (who may be sustaining renal injury prior to an elevation in the SCr) and at a clear early stage of AKI produces benefit when compared to the community cohort, a proportion of whom may have sustained the renal injury days prior to admission. A final strength of the study was that the intervention was more comprehensive (education, outreach nursing including electronic monitoring of patients, multidisciplinary input and electronic care bundles) than the only RCT in the field of AKI alerting, where only a link to KDIGO guidance was delivered.

5.5.2 Weaknesses

The study took place in a single hospital Trust, though the two separate hospitals have little interaction at ward level and 'contamination' of staffing

personnel. There are many examples in the literature where positive findings at a single-centre are then not replicated in larger multi-centre RCTs.(101, 511-513) An RCT was not possible within the constraints of the healthcare setting and thus concerns remain over biases associated with interventions lacking full randomisation. However, the difference-in-differences methodology has been applied in a number of similar healthcare wide initiatives and overcomes a number of the shortcomings of a non-randomised study. A time series design with cross-over could in the future assess whether results of the impact study are generalisable. Furthermore, alert fatigue is a well-known phenomenon and sustainability of the intervention should be addressed. A number of the outcomes lacked power to detect a difference in the analysis such as number of cases of HA-AKI patients progressing to AKI Stage 3 or requiring escalation to the ICU. Due to resource constraints, follow-up was limited to the index hospital admission.

Both the process measure results including bundle compliance and qualitative interviews suggested acceptance of the intervention was inconsistent. Though visibility and knowledge of the alerts and care bundles was high, submission of electronic bundles was low, though similar to previous studies in the field. The interviewees also highlighted the need for more comprehensive ongoing education, clarity of roles and rapid access to the IT hardware in the busy clinical environment. Though care bundles were designed to be usable, future directions could include better hardware linkage such as mobile Apps and alerts to minimise the time the clinical team have to spend away from direct care.

Not all causes of AKI have meaningful preventive interventions and grouping all patients together could dilute effects of truly useful timely intervention in subgroups. For example, a general medical cohort on the South coast of England has a large number of nursing home residents who are subsequently palliated. Reversibility of AKI in such patients may be limited and a future study, for example, could aim to recruit cohorts where reversible pathology may be expected. The study was limited to using SCr to diagnose AKI as urine output was not electronically recorded. Longer outcome data, for

example, at 90 days and a year would significantly add to the data – was the outcome benefit sustained in such a high-risk group? Finally in a time of relative resource scarcity, a health economics analysis would provide important insights. Use of technology is potentially expensive and should be rigorously analysed as with other healthcare interventions such as pharmaceutical agents or device.

This complex healthcare intervention involved technology (AKI alerting, prediction model and care bundles) used by a multidisciplinary frontline clinical team. However, consideration should also be made of the study in context and relation to other ongoing strategies at the time to guide future work. The presence of a control site and methodology including difference-in-differences, process measure review and qualitative work, allows some mitigation of external drivers. However, as described in Chapter 1 national and international guidelines and NHS strategies (such as CQUIN incentives) have led to a focus on AKI that was previously lacking. The NHS has encouraged clinical leaders such as nurse specialists and there are multiple online resources (such as think kidneys.nhs.uk awareness campaign) and the UK Renal Registry that have helped to raise the profile of AKI. It is clear from both the limitations of the prediction model, lack of effect of an alert on AKI at hospital admission and limited compliance with the care bundles, that improvement requires ongoing multi-faceted strategies. At the core of this needs to be education of a multidisciplinary team, with appropriate available timely technology to enable best supportive management to be delivered to the wide variety of patients with or at risk of AKI.

5.6 Future research directions

5.6.1 Primary care linkage

Linkage of data between primary and secondary care electronic systems holds great promise for example to allow comprehensive inclusion of co-morbidity, previous blood results and socio-demographics. For example in Denmark all residents have a unique civil registration number, used for all

health-care contacts permitting unambiguous linkage with their national patient registry.(514, 515) This has provided studies with rich data sources for example prospective registration of bacteraemias.(516) In the UK such linkage is rare, though for example in the Hampshire region this has been used to explore community AKI alerting.(93) Such data linkage could improve establishment of baseline renal function, timing of injury and subsequent follow-up of patients. Furthermore a prediction model could then conceivably be used in both settings, for example to allow for earlier alerting of patients who have deteriorated in the community prior to admission.

5.6.2 Big data for external validation of models

One reason for the shortage of external validation studies is a lack of suitable data available. However, increasingly researchers have access to big data, for instance through meta-analyses using individual participant data (IPD),(394, 517) or by analyses of databases containing electronic health records for potentially millions of patients.(518) For example, QRISK2 was developed and validated on databases with over 3 million patient episodes.(519, 520) From the systematic review presented no general HA-AKI models to date have utilised such large datasets which could aid development of robust general models whilst also allowing more personalised risk stratification in clinically relevant subgroups, for example age groups. Beyond big data exploration, it is imperative that research then goes on to assess impact on outcomes, for example in multi-centre cluster RCTs or using a natural experiment as presented in this thesis, to assess the utility of employing prediction models into electronic systems.

5.6.3 Increased complexity - lessons from other scientific fields and the learning health system

Today weather predictions solve a system of nonlinear differential equations at half a billion points per time step, up to weeks ahead, accounting for dynamic, thermodynamic, radiative and chemical processes.(182) This complexity serves as a high bar in human prediction, that healthcare should

strive to emulate. Meteorological predictions enable quantitative assessment of the degree of confidence in a particular forecast and have been shown to produce substantial benefits that outweigh investment costs in research, computing and observational programmes.(521) With the complexity of large multivariate datasets, clear visualisation is important to interpret and communicate the data. NASA Viewpoints is an example of a tool for exploring such multidimensional data, enabling rapid interactive analysis. Originally created for space missions, it has been used in areas ranging from aeronautical engineering to virology and finance.(521)

Employing such complex methodology into healthcare systems would require clinicians and IT expertise to bring together comprehensive healthcare records and dynamic data on physiological and blood parameters in the acute setting. This could enable the electronic record to display risk stratification at multiple time points from primary care to hospital admission with subsequent updating during an in-patient stay. Trends could provide powerful data overcoming limitations of the vast majority of current prediction models derived at a single time point. The learning health system was described a number of years ago as a collaboration between research, information technology and clinical practice with each influencing the other in continuous cycles (summarised figure 5.1).(522) The concept of a continuously learning system was articulated by the Institute of Medicine in the US as a way of showing how evidence informs practice and vice versa in an iterative process of innovation.(523, 524) This concept could hold promise in the field of prediction research where such collaboration is fundamental to successful implementation and subsequent improvements.

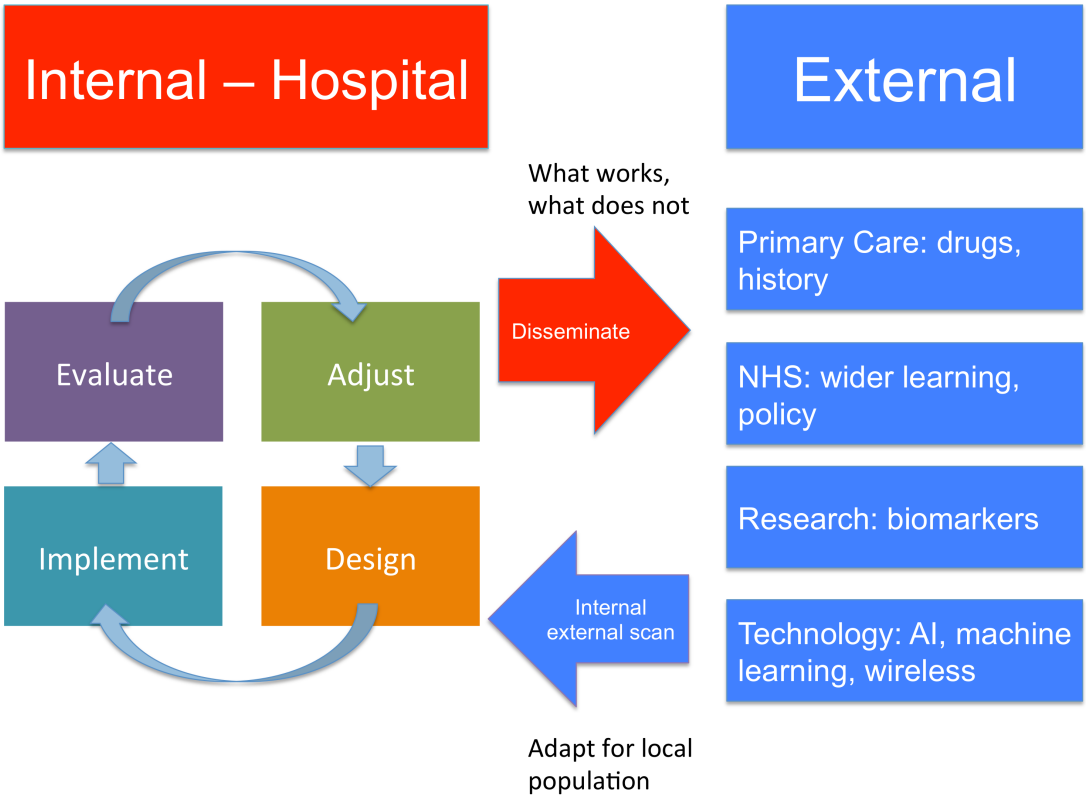


Figure 5.1 – the learning health system – AKI as a case study. AI – artificial intelligence.

The CPRs described in the systematic review all have modest predictive performance and one further avenue of future research would be to use such a prediction model to select a cohort for a further test such as a biomarker to either rule in or rule out the likelihood of a disorder evolving, or predict disease severity. For example, a number of renal biomarkers have been investigated, most commonly in intensive care populations, with few studies in general hospital populations. A CPR with few associated costs, at a sensitive cut-off, which would rule-out most cases, could then identify a group for a biomarker to provide a high positive predictive value.

Finally the impact study provides a template for future studies of electronic integration of alerting and CPRs in other fields such as sepsis. Here early diagnosis and management is also thought to improve outcomes and measures to improve care for such patients has been called for by the WHO.(525) However, this is challenging and identifying high risk groups for

example to perform extra diagnostic testing, such as a polymerase chain reaction (PCR) panel to identify an organism, or early intensive monitoring, is of great interest worldwide. An accurate prediction model alongside electronic integration is one way of investigating this conundrum.

5.7 Final summary

AKI is a significant global healthcare problem and prediction modelling may help recognise patients at risk, enable prompt interventions and subsequently improve outcomes. This thesis critiqued a large number of derived AKI prediction models, usually lacking external validation, with the majority in specialised populations. An external validation of one model in general medical and surgical populations was followed by an impact analysis study. This innovative, unique study utilised healthcare technology to integrate both a prediction model and an AKI alert in an acute hospital setting. The major findings were an associated reduction in de novo hospital-acquired AKI and reduced mortality in those who developed AKI. Future research should investigate whether the results of this study are generalisable and sustainable.

References

1. Kellum JA. Diagnostic Criteria for Acute Kidney Injury: Present and Future. Crit Care Clin. 2015;31(4):621-32.
2. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. Lancet. 2012;380(9843):756-66.
3. KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney Int 2012;Suppl 2(1):1-136.
4. Bellomo R, Kellum J, Ronco C. Acute renal failure: time for consensus. Intensive Care Med. 2001;27(11):1685-8.
5. Eknoyan G, Agodoa L. On improving outcomes and quality of dialysis care, and more. Am J Kidney Dis. 2002;39(4):889-91.
6. Davies FC, Weldon RP. A CONTRIBUTION TO THE STUDY OF " WAR NEPHRITIS.". The Lancet.190(4900):118-20.
7. Bywaters EG, Beall D. Crush Injuries with Impairment of Renal Function. Br Med J. 1941;1(4185):427-32.
8. Kellum JA, Levin N, Bouman C, Lameire N. Developing a consensus classification system for acute renal failure. Curr Opin Crit Care. 2002;8(6):509-14.
9. Barretti P, Soares VA. Acute renal failure: clinical outcome and causes of death. Ren Fail. 1997;19(2):253-7.
10. Kellum JA, Angus DC. Patients are dying of acute renal failure. Crit Care Med. 2002;30(9):2156-7.
11. Coca SG, Yusuf B, Shlipak MG, Garg AX, Parikh CR. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.
12. Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol. 2005;16(11):3365-70.
13. Van Berendoncks AM, Elseviers MM, Lins RL. Outcome of acute kidney injury with different treatment options: long-term follow-up. Clin J Am Soc Nephrol. 2010;5(10):1755-62.
14. Bagshaw SM. Short- and long-term survival after acute kidney injury. Nephrol Dial Transplant. 2008;23(7):2126-8.
15. American Society of Nephrology Renal Research Report. J Am Soc Nephrol. 2005;16(7):1886-903.
16. Control CfD, Prevention. Hospitalization discharge diagnoses for kidney disease--United States, 1980-2005. MMWR Morbidity and mortality weekly report. 2008;57(12):309.
17. 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(4):R204-12.

18. 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(2):R31.
19. Lameire NH, Bagga A, Cruz D, De Maeseneer J, Endre Z, Kellum JA, et al. Acute kidney injury: an increasing global concern. Lancet. 2013;382(9887):170-9.
20. Mehta RL, Cerda J, Burdmann EA, Tonelli M, Garcia-Garcia G, Jha V, et al. International Society of Nephrology's 0by25 initiative for acute kidney injury (zero preventable deaths by 2025): a human rights case for nephrology. Lancet. 2015;385(9987):2616-43.
21. Stevens LA, Schmid CH, Greene T, Li L, Beck GJ, Joffe MM, et al. Factors other than glomerular filtration rate affect serum cystatin C levels. Kidney Int. 2009;75(6):652-60.
22. Delanaye P, Schaeffner E, Ebert N, Cavalier E, Mariat C, Krzesinski JM, et al. Normal reference values for glomerular filtration rate: what do we really know? Nephrol Dial Transplant. 2012;27(7):2664-72.
23. Poggio ED, Rule AD, Tanchanco R, Arrigain S, Butler RS, Srinivas T, et al. Demographic and clinical characteristics associated with glomerular filtration rates in living kidney donors. Kidney International.75(10):1079-87.
24. Prowle JR, Echeverri JE, Ligabo EV, Ronco C, Bellomo R. Fluid balance and acute kidney injury. Nat Rev Nephrol. 2010;6(2):107-15.
25. Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function--measured and estimated glomerular filtration rate. N Engl J Med. 2006;354(23):2473-83.
26. Prowle JR, Echeverri JE, Ligabo EV, Ronco C, Bellomo R. Fluid balance and acute kidney injury. Nat Rev Nephrol. 2010;6.
27. Perrone RD, Madias NE, Levey AS. Serum creatinine as an index of renal function: new insights into old concepts. Clin Chem. 1992;38(10):1933-53.
28. Bagshaw SM, Gibney RT. Conventional markers of kidney function. Crit Care Med. 2008;36(4 Suppl):S152-8.
29. Rehberg PB. Studies on Kidney Function: The Excretion of Urea and Chlorine Analysed according to a Modified Filtration-Reabsorption Theory. Biochemical Journal. 1926;20(3):461-82.
30. Uchino S. Creatinine. Curr Opin Crit Care. 2010;16(6):562-7.
31. Addis T, Barrett E, Menzies JT. A CLINICAL METHOD FOR THE APPROXIMATE DETERMINATION OF SERUM CREATININE CONCENTRATION. J Clin Invest. 1947;26(5):879-82.
32. Swan SK. The search continues—an ideal marker of GFR. Clinical Chemistry; 1997.
33. Stevens LA, Levey AS. Measurement of kidney function. Med Clin North Am. 2005;89(3):457-73.
34. Gaiao S, Cruz DN. Baseline creatinine to define acute kidney injury: is there any consensus? Nephrol Dial Transplant. 2010;25(12):3812-4.

35. Shemesh O, Golbetz H, Kriss JP, Myers BD. Limitations of creatinine as a filtration marker in glomerulopathic patients. Kidney Int. 1985;28(5):830-8.
36. Kemperman FA, Silberbusch J, Slaats EH, Prins AM, Krediet RT, Arisz L. Follow-up of GFR estimated from plasma creatinine after cimetidine administration in patients with diabetes mellitus type 2. Clin Nephrol. 2000;54(4):255-60.
37. Molitch ME, Rodman E, Hirsch CA, Dubinsky E. Spurious serum creatinine elevations in ketoacidosis. Ann Intern Med. 1980;93(2):280-1.
38. Erley CM, Bader BD, Berger ED, Vochazer A, Jorzik JJ, Dietz K, et al. Plasma clearance of iodine contrast media as a measure of glomerular filtration rate in critically ill patients. Crit Care Med. 2001;29(8):1544-50.
39. Doolan PD, Alpen EL, Theil GB. A clinical appraisal of the plasma concentration and endogenous clearance of creatinine. Am J Med. 1962;32:65-79.
40. Bjornsson TD. Use of serum creatinine concentrations to determine renal function. Clin Pharmacokinet. 1979;4(3):200-22.
41. Enger E, Blegen EM. THE RELATIONSHIP BETWEEN ENDOGENOUS CREATININE CLEARANCE AND SERUM CREATININE IN RENAL FAILURE. Scand J Clin Lab Invest. 1964;16:273-80.
42. Waikar SS, Betensky RA, Emerson SC, Bonventre JV. Imperfect gold standards for kidney injury biomarker evaluation. J Am Soc Nephrol. 2012;23(1):13-21.
43. Chiou WL, Hsu FH. Pharmacokinetics of creatinine in man and its implications in the monitoring of renal function and in dosage regimen modifications in patients with renal insufficiency. J Clin Pharmacol. 1975;15(5-6):427-34.
44. Waikar SS, Bonventre JV. Creatinine Kinetics and the Definition of Acute Kidney Injury. Journal of the American Society of Nephrology : JASN. 2009;20(3):672-9.
45. Edwards KD. Creatinine space as a measure of total body water in anuric subjects, estimated after single injection and haemodialysis. Clin Sci. 1959;18:455-64.
46. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, et al. Comparison of two fluid-management strategies in acute lung injury. N Engl J Med. 2006;354(24):2564-75.
47. Macedo E, Bouchard J, Soroko SH, Chertow GM, Himmelfarb J, Ikizler TA, et al. Fluid accumulation, recognition and staging of acute kidney injury in critically-ill patients. Critical Care. 2010;14(3):R82-R.
48. Prowle JR, Leitch A, Kirwan CJ, Forni LG. Positive fluid balance and AKI diagnosis: assessing the extent and duration of 'creatinine dilution'. Intensive Care Med. 2015;41(1):160-1.

49. Pickering JW, Ralib AM, Endre ZH. Combining creatinine and volume kinetics identifies missed cases of acute kidney injury following cardiac arrest. Critical Care. 2013;17(1):R7.
50. Vanholder R, De Smet R, Glorieux G, Argiles A, Baurmeister U, Brunet P, et al. Review on uremic toxins: classification, concentration, and interindividual variability. Kidney Int. 2003;63(5):1934-43.
51. Prabhakar SS, Zeballos GA, Montoya-Zavala M, Leonard C. Urea inhibits inducible nitric oxide synthase in macrophage cell line. Am J Physiol. 1997;273(6 Pt 1):C1882-8.
52. Lim J, Gasson C, Kaji DM. Urea inhibits NaK2Cl cotransport in human erythrocytes. J Clin Invest. 1995;96(5):2126-32.
53. Rabb H, Wang Z, Nemoto T, Hotchkiss J, Yokota N, Soleimani M. Acute renal failure leads to dysregulation of lung salt and water channels. Kidney Int. 2003;63(2):600-6.
54. Sort P, Navasa M, Arroyo V, Aldeguer X, Planas R, Ruiz-del-Arbol L, et al. Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med. 1999;341(6):403-9.
55. Lehner GF, Forni LG, Joannidis M. Oliguria and Biomarkers of Acute Kidney Injury: Star Struck Lovers or Strangers in the Night? Nephron. 2016;134(3):183-90.
56. Kellum JA, Sileanu FE, Murugan R, Lucko N, Shaw AD, Clermont G. Classifying AKI by Urine Output versus Serum Creatinine Level. J Am Soc Nephrol. 2015;26(9):2231-8.
57. Prowle JR, Liu YL, Licari E, Bagshaw SM, Egi M, Haase M, et al. Oliguria as predictive biomarker of acute kidney injury in critically ill patients. Crit Care. 2011;15.
58. Zaloga GP, Hughes SS. Oliguria in patients with normal renal function. Anesthesiology. 1990;72(4):598-602.
59. Chesley LC. UREA EXCRETION AT LOW URINE VOLUMES. THE CALCULATION OF "MINIMAL" UREA CLEARANCES. J Clin Invest. 1938;17(2):119-23.
60. Solomon AW, Kirwan CJ, Alexander ND, Nimako K, Jurukov A, Forth RJ, et al. Urine output on an intensive care unit: case-control study. BMJ. 2010;341:c6761.
61. 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):1692-702.
62. Palevsky PM, Liu KD, Brophy PD, Chawla LS, Parikh CR, Thakar CV, et al. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for acute kidney injury. Am J Kidney Dis. 2013;61(5):649-72.
63. Ostermann M, Joannidis M. Biomarkers for AKI improve clinical practice: no. Intensive Care Med. 2015;41(4):618-22.

64. Ostermann M, Philips BJ, Forni LG. Clinical review: Biomarkers of acute kidney injury: where are we now? Crit Care. 2012;16(5):233.
65. Ostermann M, Joannidis M. Acute kidney injury 2016: diagnosis and diagnostic workup. Critical Care. 2016;20:299.
66. Murray PT, Mehta RL, Shaw A, Ronco C, Endre Z, Kellum JA, et al. Current Use of Biomarkers in Acute Kidney Injury: Report and Summary of Recommendations from the 10(th) Acute Dialysis Quality Initiative Consensus Conference. Kidney international. 2014;85(3):513-21.
67. Haase M, Devarajan P, Haase-Fielitz A, Bellomo R, Cruz DN, Wagener G, et al. The outcome of neutrophil gelatinase-associated lipocalin-positive subclinical acute kidney injury: a multicenter pooled analysis of prospective studies. J Am Coll Cardiol. 2011;57(17):1752-61.
68. Sawhney S, Mitchell M, Marks A, Fluck N, Black C. Long-term prognosis after acute kidney injury (AKI): what is the role of baseline kidney function and recovery? A systematic review. BMJ Open. 2015;5(1):e006497.
69. Susantitaphong P, Cruz DN, Cerda J, Abulfaraj M, Alqahtani F, Koulouridis I, et al. World incidence of AKI: a meta-analysis. Clin J Am Soc Nephrol. 2013;8(9):1482-93.
70. Kheterpal S, Tremper KK, Englesbe MJ, O'Reilly M, Shanks AM, Fetterman DM, et al. Predictors of postoperative acute renal failure after noncardiac surgery in patients with previously normal renal function. Anesthesiology. 2007;107(6):892-902.
71. Kheterpal S, Tremper KK, Heung M, Rosenberg AL, Englesbe M, Shanks AM, 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):505-15.
72. Lafrance JP, Miller DR. Acute kidney injury associates with increased long-term mortality. J Am Soc Nephrol. 2010;21(2):345-52.
73. 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. Nephrol Dial Transplant. 2008;23(4):1203-10.
74. Ostermann M, Chang RW. Acute kidney injury in the intensive care unit according to RIFLE. Crit Care Med. 2007;35(8):1837-43; quiz 52.
75. Bagshaw SM, George C, Bellomo R. Early acute kidney injury and sepsis: a multicentre evaluation. Crit Care. 2008;12(2):R47.
76. Singbartl K, Kellum JA. AKI in the ICU: definition, epidemiology, risk stratification, and outcomes. Kidney Int. 2012;81(9):819-25.
77. Loef BG, Epema AH, Smilde TD, Henning RH, Ebels T, Navis G, et al. Immediate postoperative renal function deterioration in cardiac surgical patients predicts in-hospital mortality and long-term survival. J Am Soc Nephrol. 2005;16(1):195-200.
78. 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. Ann Intern Med. 1998;128(3):194-203.
79. Frost L, Pedersen RS, Lund O, Hansen OK, Hansen HE. Prognosis and risk factors in acute, dialysis-requiring renal failure after open-heart surgery. Scand J Thorac Cardiovasc Surg. 1991;25(3):161-6.
80. Robert AM, Kramer RS, Dacey LJ, Charlesworth DC, Leavitt BJ, Helm RE, et al. Cardiac surgery-associated acute kidney injury: a comparison of two consensus criteria. Ann Thorac Surg. 2010;90(6):1939-43.
81. Ferguson TB, Jr., Hammill BG, Peterson ED, DeLong ER, Grover FL. A decade of change--risk profiles and outcomes for isolated coronary artery bypass grafting procedures, 1990-1999: a report from the STS National Database Committee and the Duke Clinical Research Institute. Society of Thoracic Surgeons. Ann Thorac Surg. 2002;73(2):480-9; discussion 9-90.
82. 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. Crit Care. 2014;18(6):606.
83. Yalavarthy R, Edelstein CL, Teitelbaum I. Acute renal failure and chronic kidney disease following liver transplantation. Hemodial Int. 2007;11 Suppl 3:S7-12.
84. Forni LG, Dawes T, Sinclair H, Cheek E, Bewick V, Dennis M, et al. Identifying the patient at risk of acute kidney injury: a predictive scoring system for the development of acute kidney injury in acute medical patients. Nephron Clin Pract. 2013;123(3-4):143-50.
85. Bedford M, Stevens P, Coulton S, Billings J, Farr M, Wheeler T, et al. Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study. Southampton UK: Queen's Printer and Controller of HMSO.; 2016 Feb.
86. Koyner JL, Adhikari R, Edelson DP, Churpek MM. Development of a Multicenter Ward-Based AKI Prediction Model. Clin J Am Soc Nephrol. 2016;11(11):1935-43.
87. Wonnacott A, Meran S, Amphlett B, Talabani B, Phillips A. Epidemiology and outcomes in community-acquired versus hospital-acquired AKI. Clin J Am Soc Nephrol. 2014;9(6):1007-14.
88. Schissler MM, Zaidi S, Kumar H, Deo D, Brier ME, McLeish KR. Characteristics and outcomes in community-acquired versus hospital-acquired acute kidney injury. Nephrology (Carlton). 2013;18(3):183-7.
89. Wang Y, Cui Z, Fan M. Hospital-acquired and community-acquired acute renal failure in hospitalized Chinese: a ten-year review. Ren Fail. 2007;29(2):163-8.
90. Obialo CI, Okonofua EC, Tayade AS, Riley LJ. Epidemiology of de novo acute renal failure in hospitalized African Americans: comparing community-acquired vs hospital-acquired disease. Arch Intern Med. 2000;160(9):1309-13.

91. Xu X, Nie S, Liu Z, Chen C, Xu G, Zha Y, et al. Epidemiology and Clinical Correlates of AKI in Chinese Hospitalized Adults. Clin J Am Soc Nephrol. 2015;10(9):1510-8.
92. Sawhney S, Fluck N, Fraser SD, Marks A, Prescott GJ, Roderick PJ, et al. KDIGO-based acute kidney injury criteria operate differently in hospitals and the community-findings from a large population cohort. Nephrol Dial Transplant. 2016;31(6):922-9.
93. Johnson M, Hounkpatin H, Fraser S, Culliford D, Uniacke M, Roderick P. Using a linked database for epidemiology across the primary and secondary care divide: acute kidney injury. BMC Med Inform Decis Mak. 2017;17(1):106.
94. Bagshaw SM, George C, Bellomo R. Changes in the incidence and outcome for early acute kidney injury in a cohort of Australian intensive care units. Crit Care. 2007;11(3):R68.
95. Hsu RK, McCulloch CE, Dudley RA, Lo LJ, Hsu CY. Temporal changes in incidence of dialysis-requiring AKI. J Am Soc Nephrol. 2013;24(1):37-42.
96. Coca SG. Acute kidney injury in elderly persons. Am J Kidney Dis. 2010;56(1):122-31.
97. Abdel-Kader K, Palevsky PM. Acute kidney injury in the elderly. Clin Geriatr Med. 2009;25(3):331-58.
98. Chronopoulos A, Cruz DN, Ronco C. Hospital-acquired acute kidney injury in the elderly. Nat Rev Nephrol. 2010;6(3):141-9.
99. Wang X, Bonventre JV, Parrish AR. The aging kidney: increased susceptibility to nephrotoxicity. Int J Mol Sci. 2014;15(9):15358-76.
100. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. N Engl J Med. 2000;342(18):1301-8.
101. Angus DC, Barnato AE, Bell D, Bellomo R, Chong CR, Coats TJ, et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. Intensive Care Med. 2015.
102. Vanmassenhove J, Kielstein J, Jorres A, Biesen WV. Management of patients at risk of acute kidney injury. Lancet. 2017;389(10084):2139-51.
103. Kellum JA, Chawla LS, Keener C, Singbartl K, Palevsky PM, Pike FL, et al. The Effects of Alternative Resuscitation Strategies on Acute Kidney Injury in Patients with Septic Shock. Am J Respir Crit Care Med. 2015.
104. Allgren RL, Marbury TC, Rahman SN, Weisberg LS, Fenves AZ, Lafayette RA, et al. Anaritide in acute tubular necrosis. Auriculin Anaritide Acute Renal Failure Study Group. N Engl J Med. 1997;336(12):828-34.
105. Hirschberg R, Kopple J, Lipsett P, Benjamin E, Minei J, Albertson T, et al. Multicenter clinical trial of recombinant human insulin-like growth factor I in patients with acute renal failure. Kidney Int. 1999;55(6):2423-32.

106. Coca SG, Singanamala S, Parikh CR. Chronic kidney disease after acute kidney injury: a systematic review and meta-analysis. Kidney Int. 2012;81(5):442-8.
107. Mehta RL, Burdmann EA, Cerda J, Feehally J, Finkelstein F, Garcia-Garcia G, et al. Recognition and management of acute kidney injury in the International Society of Nephrology 0by25 Global Snapshot: a multinational cross-sectional study. Lancet. 2016;387(10032):2017-25.
108. Basile DP, Anderson MD, Sutton TA. Pathophysiology of Acute Kidney Injury. Comprehensive Physiology. 2012;2(2):1303-53.
109. Thadhani R, Pascual M, Bonventre JV. Acute renal failure. N Engl J Med. 1996;334(22):1448-60.
110. Gomez H, Ince C, De Backer D, Pickkers P, Payen D, Hotchkiss J, et al. A unified theory of sepsis-induced acute kidney injury: inflammation, microcirculatory dysfunction, bioenergetics, and the tubular cell adaptation to injury. Shock. 2014;41(1):3-11.
111. Pannu N, James M, Hemmelgarn BR, Dong J, Tonelli M, Klarenbach S. Modification of outcomes after acute kidney injury by the presence of CKD. Am J Kidney Dis. 2011;58(2):206-13.
112. Bellomo R, Bagshaw S, Langenberg C, Ronco C. Pre-renal azotemia: a flawed paradigm in critically ill septic patients? Contrib Nephrol. 2007;156:1-9.
113. Langenberg C, Bagshaw SM, May CN, Bellomo R. The histopathology of septic acute kidney injury: a systematic review. Crit Care. 2008;12(2):R38.
114. Uchino S, Kellum JA, Bellomo R, Doig GS, Morimatsu H, Morgera S, et al. Acute renal failure in critically ill patients: a multinational, multicenter study. JAMA. 2005;294(7):813-8.
115. Safirstein RL. Acute renal failure: from renal physiology to the renal transcriptome. Kidney Int Suppl. 2004(91):S62-6.
116. Castaneda MP, Swiatecka-Urban A, Mitsnefes MM, Feuerstein D, Kaskel FJ, Tellis V, et al. Activation of mitochondrial apoptotic pathways in human renal allografts after ischemiareperfusion injury. Transplantation. 2003;76(1):50-4.
117. Bonventre JV, Zuk A. Ischemic acute renal failure: an inflammatory disease? Kidney Int. 2004;66(2):480-5.
118. Molitoris BA, Sutton TA. Endothelial injury and dysfunction: role in the extension phase of acute renal failure. Kidney Int. 2004;66(2):496-9.
119. Devarajan P. Update on mechanisms of ischemic acute kidney injury. J Am Soc Nephrol. 2006;17(6):1503-20.
120. Devarajan P, Mishra J, Supavekin S, Patterson LT, Steven Potter S. Gene expression in early ischemic renal injury: clues towards pathogenesis, biomarker discovery, and novel therapeutics. Mol Genet Metab. 2003;80(4):365-76.
121. Bonventre JV. Dedifferentiation and proliferation of surviving epithelial cells in acute renal failure. J Am Soc Nephrol. 2003;14 Suppl 1:S55-61.

122. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-10.
123. De Backer D, Creteur J, Preiser JC, Dubois MJ, Vincent JL. Microvascular blood flow is altered in patients with sepsis. Am J Respir Crit Care Med. 2002;166(1):98-104.
124. Wang Z, Holthoff JH, Seely KA, Pathak E, Spencer HJ, 3rd, Gokden N, et al. Development of oxidative stress in the peritubular capillary microenvironment mediates sepsis-induced renal microcirculatory failure and acute kidney injury. Am J Pathol. 2012;180(2):505-16.
125. Singer M, De Santis V, Vitale D, Jeffcoate W. Multiorgan failure is an adaptive, endocrine-mediated, metabolic response to overwhelming systemic inflammation. Lancet. 2004;364(9433):545-8.
126. Prowle JR, Ishikawa K, May CN, Bellomo R. Renal blood flow during acute renal failure in man. Blood Purif. 2009;28(3):216-25.
127. Langenberg C, Wan L, Egi M, May CN, Bellomo R. Renal blood flow in experimental septic acute renal failure. Kidney Int. 2006;69(11):1996-2002.
128. Langenberg C, Bellomo R, May C, Wan L, Egi M, Morgera S. Renal blood flow in sepsis. Crit Care. 2005;9(4):R363-74.
129. Chua HR, Glassford N, Bellomo R. Acute kidney injury after cardiac arrest. Resuscitation. 2012;83(6):721-7.
130. Castrop H. Mediators of tubuloglomerular feedback regulation of glomerular filtration: ATP and adenosine. Acta Physiol (Oxf). 2007;189(1):3-14.
131. Cohen G, Haag-Weber M, Horl WH. Immune dysfunction in uremia. Kidney Int Suppl. 1997;62:S79-82.
132. Forni LG, Chawla LS. Biomarkers in cardiorenal syndrome. Blood Purif. 2014;37 Suppl 2:14-9.
133. Feltes CM, Van Eyk J, Rabb H. Distant-organ changes after acute kidney injury. Nephron Physiol. 2008;109(4):p80-4.
134. Grams ME, Rabb H. The distant organ effects of acute kidney injury. Kidney Int. 2012;81(10):942-8.
135. Doi K, Rabb H. Impact of acute kidney injury on distant organ function: recent findings and potential therapeutic targets. Kidney Int. 2016;89(3):555-64.
136. Scheel PJ, Liu M, Rabb H. Uremic lung: new insights into a forgotten condition. Kidney Int. 2008;74(7):849-51.
137. Ologunde R, Zhao H, Lu K, Ma D. Organ cross talk and remote organ damage following acute kidney injury. Int Urol Nephrol. 2014;46(12):2337-45.
138. Levy EM, Viscoli CM, Horwitz RI. The effect of acute renal failure on mortality. A cohort analysis. JAMA. 1996;275(19):1489-94.
139. Jurawan N, Pankhurst T, Ferro C, Nightingale P, Coleman J, Rosser D, et al. Hospital acquired Acute Kidney Injury is associated with increased mortality

but not increased readmission rates in a UK acute hospital. BMC Nephrology. 2017;18:317.
140. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical Decision Support for In-Hospital AKI. J Am Soc Nephrol. 2017.
141. Kolhe NV, Reilly T, Leung J, Fluck RJ, Swinscoe KE, Selby NM, et al. A simple care bundle for use in acute kidney injury: a propensity score-matched cohort study. Nephrol Dial Transplant. 2016;31(11):1846-54.
142. Kerr M, Bedford M, Matthews B, O'Donoghue D. The economic impact of acute kidney injury in England. Nephrol Dial Transplant. 2014;29(7):1362-8.
143. Uchino S, Bellomo R, Bagshaw SM, Goldsmith D. Transient azotaemia is associated with a high risk of death in hospitalized patients. Nephrol Dial Transplant. 2010;25(6):1833-9.
144. Coca SG, Peixoto AJ, Garg AX, Krumholz HM, Parikh CR. The prognostic importance of a small acute decrement in kidney function in hospitalized patients: a systematic review and meta-analysis. Am J Kidney Dis. 2007;50(5):712-20.
145. Wang HE, Muntner P, Chertow GM, Warnock DG. Acute kidney injury and mortality in hospitalized patients. Am J Nephrol. 2012;35(4):349-55.
146. Hobson CE, Yavas S, Segal MS, Schold JD, Tribble CG, Layon AJ, et al. Acute kidney injury is associated with increased long-term mortality after cardiothoracic surgery. Circulation. 2009;119(18):2444-53.
147. Carmichael P, Carmichael AR. Acute renal failure in the surgical setting. ANZ J Surg. 2003;73(3):144-53.
148. Chertow GM, Lazarus JM, Christiansen CL, Cook EF, Hammermeister KE, Grover F, et al. Preoperative renal risk stratification. Circulation. 1997;95(4):878-84.
149. Thakar CV, Liangos O, Yared JP, Nelson DA, Hariachar S, Paganini EP. Predicting acute renal failure after cardiac surgery: validation and re-definition of a risk-stratification algorithm. Hemodial Int. 2003;7(2):143-7.
150. Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP. A clinical score to predict acute renal failure after cardiac surgery. J Am Soc Nephrol. 2005;16(1):162-8.
151. Fortescue EB, Bates DW, Chertow GM. Predicting acute renal failure after coronary bypass surgery: cross-validation of two risk-stratification algorithms. Kidney Int. 2000;57(6):2594-602.
152. Cruz DN, Bolgan I, Perazella MA, Bonello M, de Cal M, Corradi V, et al. North East Italian Prospective Hospital Renal Outcome Survey on Acute Kidney Injury (NEiPHROS-AKI): targeting the problem with the RIFLE Criteria. Clin J Am Soc Nephrol. 2007;2(3):418-25.
153. 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):1913-7.

154. Chawla LS, Eggers PW, Star RA, Kimmel PL. Acute kidney injury and chronic kidney disease as interconnected syndromes. N Engl J Med. 2014;371(1):58-66.
155. Leung KC, Tonelli M, James MT. Chronic kidney disease following acute kidney injury-risk and outcomes. Nat Rev Nephrol. 2013;9(2):77-85.
156. Venkatachalam MA, Griffin KA, Lan R, Geng H, Saikumar P, Bidani AK. Acute kidney injury: a springboard for progression in chronic kidney disease. Am J Physiol Renal Physiol. 2010;298(5):F1078-94.
157. Lo LJ, Go AS, Chertow GM, McCulloch CE, Fan D, Ordonez JD, et al. Dialysis-requiring acute renal failure increases the risk of progressive chronic kidney disease. Kidney Int. 2009;76(8):893-9.
158. Prowle JR, Kolic I, Purdell-Lewis J, Taylor R, Pearse RM, Kirwan CJ. Serum creatinine changes associated with critical illness and detection of persistent renal dysfunction after AKI. Clin J Am Soc Nephrol. 2014;9(6):1015-23.
159. Khosla N, Soroko SB, Chertow GM, Himmelfarb J, Ikizler TA, Paganini E, et al. Preexisting chronic kidney disease: a potential for improved outcomes from acute kidney injury. Clin J Am Soc Nephrol. 2009;4(12):1914-9.
160. Ishani A, Xue JL, Himmelfarb J, Eggers PW, Kimmel PL, Molitoris BA, et al. Acute kidney injury increases risk of ESRD among elderly. J Am Soc Nephrol. 2009;20(1):223-8.
161. Cerda J, Cerda M, Kilcullen P, Prendergast J. In severe acute kidney injury, a higher serum creatinine is paradoxically associated with better patient survival. Nephrol Dial Transplant. 2007;22(10):2781-4.
162. Broce JC, Price LL, Liangos O, Uhlig K, Jaber BL. Hospital-acquired acute kidney injury: an analysis of nadir-to-peak serum creatinine increments stratified by baseline estimated GFR. Clin J Am Soc Nephrol. 2011;6(7):1556-65.
163. Silver SA, Chertow GM. The Economic Consequences of Acute Kidney Injury. Nephron. 2017.
164. Grams ME, Waikar SS, MacMahon B, Whelton S, Ballew SH, Coresh J. Performance and limitations of administrative data in the identification of AKI. Clin J Am Soc Nephrol. 2014;9(4):682-9.
165. Silver SA, Long J, Zheng Y, Chertow GM. Cost of Acute Kidney Injury in Hospitalized Patients. J Hosp Med. 2017;12(2):70-6.
166. National Confidential Enquiry into Patient Outcome and Death. Adding insult to injury: a review of the care of patients who died in hospital with a primary diagnosis of acute kidney injury (acute renal failure) http://www.ncepod.org.uk/2009report1/Downloads/AKI_report.pdf. NCEPOD, 2009.
167. Aitken E, Carruthers C, Gall L, Kerr L, Geddes C, Kingsmore D. Acute kidney injury: outcomes and quality of care. QJM. 2013;106(4):323-32.

168. Thomas M, Sitch A, Dowswell G. The initial development and assessment of an automatic alert warning of acute kidney injury. Nephrol Dial Transplant. 2011;26(7):2161-8.
169. Selby NM, Crowley L, Fluck RJ, McIntyre CW, Monaghan J, Lawson N, et al. Use of electronic results reporting to diagnose and monitor AKI in hospitalized patients. Clin J Am Soc Nephrol. 2012;7(4):533-40.
170. Wilson FP, Shashaty M, Testani J, Aqeel I, Borovskiy Y, Ellenberg SS, et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. Lancet. 2015.
171. Andreev E, Koopman M, Arisz L. A rise in plasma creatinine that is not a sign of renal failure: which drugs can be responsible? J Intern Med. 1999;246(3):247-52.
172. Doi K, Yuen PST, Eisner C, Hu X, Leelahavanichkul A, Schnermann J, et al. Reduced Production of Creatinine Limits Its Use as Marker of Kidney Injury in Sepsis. Journal of the American Society of Nephrology : JASN. 2009;20(6):1217-21.
173. Lin J, Fernandez H, Shashaty MG, Negoianu D, Testani JM, Berns JS, et al. False-Positive Rate of AKI Using Consensus Creatinine-Based Criteria. Clin J Am Soc Nephrol. 2015;10(10):1723-31.
174. England N. Algorithm for detecting Acute Kidney Injury (AKI) based on serum creatinine changes with time http://www.england.nhs.uk/wp-content/uploads/2014/06/psa-aki-alg.pdf: NHS England; 2014.
175. Sawhney S, Fluck N, Marks A, Prescott G, Simpson W, Tomlinson L, et al. Acute kidney injury-how does automated detection perform? Nephrol Dial Transplant. 2015;30(11):1853-61.
176. Chaitin GJ. Information, randomness & incompleteness: papers on algorithmic information theory: World Scientific; 1990.
177. Sornette D. Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(Suppl 1):2522-9.
178. Brown EW. ON THE PREDICTIONS OF TRANS-NEPTUNIAN PLANETS FROM THE PERTURBATIONS OF URANUS. Proceedings of the National Academy of Sciences of the United States of America. 1930;16(5):364-71.
179. Eisenberg D. The discovery of the α-helix and β-sheet, the principal structural features of proteins. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(20):11207-10.
180. Abbe C. The physical basis of long-range weather forecasts. Monthly Weather Review. 1901;29(12):551-61.
181. Bjerknes V. Das Problem der Wettervorhersage: betrachtet vom Standpunkte der Mechanik und der Physik1904.
182. Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. Nature. 2015;525(7567):47-55.

183. Wang Z, Li P, Li L, Huang C, Liu M. Modeling and Forecasting Average Temperature for Weather Derivative Pricing. Advances in Meteorology. 2015;2015:8.
184. Diebold F, Yilmaz K. Better to give than to receive: Predictive directional measurement of volatility spillovers. International Journal of Forecasting. 2012;28(1):57-66.
185. Langseth H, editor Beating the bookie: A look at statistical models for prediction of football matches2013.
186. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
187. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Stat Med. 1984;3(2):143-52.
188. Sinuff T, Adhikari NK, Cook DJ, Schunemann HJ, Griffith LE, Rocker G, et al. Mortality predictions in the intensive care unit: comparing physicians with scoring systems. Crit Care Med. 2006;34(3):878-85.
189. Bandiera G, Stiell IG, Wells GA, Clement C, De Maio V, Vandemheen KL, et al. The Canadian C-spine rule performs better than unstructured physician judgment. Ann Emerg Med. 2003;42(3):395-402.
190. Jacobs LM, Schwartz R. Prospective analysis of acute cervical spine injury: a methodology to predict injury. Ann Emerg Med. 1986;15(1):44-9.
191. McKee TR, Tinkoff G, Rhodes M. Asymptomatic occult cervical spine fracture: case report and review of the literature. J Trauma. 1990;30(5):623-6.
192. Poses RM, McClish DK, Bekes C, Scott WE, Morley JN. Ego bias, reverse ego bias, and physicians' prognostic. Crit Care Med. 1991;19(12):1533-9.
193. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5):1-12.
194. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52(10):935-42.
195. Braitman LE, Davidoff F. Predicting clinical states in individual patients. Ann Intern Med. 1996;125(5):406-12.
196. Feinstein AR. "Clinical Judgment" revisited: the distraction of quantitative models. Ann Intern Med. 1994;120(9):799-805.
197. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Ann Intern Med. 1993;118(3):201-10.
198. Christensen E. Prognostic models including the Child-Pugh, MELD and Mayo risk scores--where are we and where should we go? J Hepatol. 2004;41(2):344-50.
199. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol. 2008;61(11):1085-94.

200. Wessler BS, Lai Yh L, Kramer W, Cangelosi M, Raman G, Lutz JS, et al. Clinical Prediction Models for Cardiovascular Disease: Tufts Predictive Analytics and Comparative Effectiveness Clinical Prediction Model Database. Circ Cardiovasc Qual Outcomes. 2015;8(4):368-75.
201. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg. 1953;32(4):260-7.
202. Apgar V, Holaday DA, James LS, Weisbrot IM, Berrien C. Evaluation of the newborn infant; second report. J Am Med Assoc. 1958;168(15):1985-8.
203. The Apgar score. Pediatrics. 2006;117(4):1444-7.
204. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. Am J Cardiol. 1976;38(1):46-51.
205. Liao Y, McGee DL, Cooper RS, Sutkowski MB. How generalizable are coronary risk prediction models? Comparison of Framingham and two national cohorts. Am Heart J. 1999;137(5):837-45.
206. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. Med Decis Making. 1981;1(3):239-46.
207. 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):591-7.
208. Kuzniewicz MW, Vasilevskis EE, Lane R, Dean ML, Trivedi NG, Rennie DJ, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest. 2008;133(6):1319-27.
209. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med. 1992;21(4):384-90.
210. Stiell IG, Wells GA, Hoag RH, Sivilotti ML, Cacciotti TF, Verbeek PR, et al. Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. JAMA. 1997;278(23):2075-9.
211. Stiell IG, Wells GA, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet. 2001;357(9266):1391-6.
212. Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA. 2001;286(15):1841-8.
213. Rothwell PM, Giles MF, Flossmann E, Lovelock CE, Redgrave JN, Warlow CP, et al. A simple score (ABCD) to identify individuals at high early risk of stroke after transient ischaemic attack. Lancet. 2005;366(9479):29-36.
214. Johnston SC, Rothwell PM, Nguyen-Huynh MN, Giles MF, Elkins JS, Bernstein AL, et al. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. Lancet. 2007;369(9558):283-92.

215. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA. 2001;285(22):2864-70.
216. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377-82.
217. Lim WS, Baudouin SV, George RC, Hill AT, Jamieson C, Le Jeune I, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. Thorax. 2009;64 Suppl 3:iii1-55.
218. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int. 2008;19(4):385-97.
219. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS--Towards a national early warning score for detecting adult inpatient deterioration. Resuscitation. 2010;81(8):932-7.
220. Physicians RCo. National Early Warning Score (NEWS) Standardising the assessment of acute-illness severity in the NHS - Report of a working party July London: 2012.
221. Physicians. RCo. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. London: RCP, 2017.
222. Wang G, Lam KM, Deng Z, Choi KS. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. Comput Biol Med. 2015;63:124-32.
223. Wallert J, Tomasoni M, Madison G, Held C. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. BMC Med Inform Decis Mak. 2017;17(1):99.
224. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013;346:e5793.
225. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. Nat Rev Drug Discov. 2007;6(4):287-93.
226. Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? J Am Coll Cardiol. 2003;42(5):851-3.
227. Gandara E, Wells PS. Diagnosis: use of clinical probability algorithms. Clin Chest Med. 2010;31(4):629-39.
228. Marchese MC. Clinical versus actuarial prediction: a review of the literature. Percept Mot Skills. 1992;75(2):583-94.

229. Gardner W, Lidz CW, Mulvey EP, Shaw EC. Clinical versus actuarial predictions of violence of patients with mental illnesses. J Consult Clin Psychol. 1996;64(3):602-9.
230. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psychol Assess. 2000;12(1):19-30.
231. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247(18):2543-6.
232. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Stat Med. 1986;5(5):421-33.
233. Knaus WA, Harrell FE, Fisher CJ, Jr., Wagner DP, Opal SM, Sadoff JC, et al. The clinical evaluation of new drugs for sepsis. A prospective study design based on survival analysis. JAMA. 1993;270(10):1233-41.
234. Yealy DM, Auble TE. Choosing between clinical prediction rules. N Engl J Med. 2003;349(26):2553-5.
235. Stiell IG, Clement CM, McKnight RD, Brison R, Schull MJ, Rowe BH, et al. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. N Engl J Med. 2003;349(26):2510-8.
236. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA. 1991;266(1):93-8.
237. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-73.
238. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338:b604.
239. Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med. 2003;138(8):644-50.
240. Harrell FJ. Regression Modeling Strategies. New York: Springer; 2001.
241. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
242. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
243. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.
244. van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. J Clin Epidemiol. 2004;57(7):672-82.
245. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer. 1994;69(6):979-85.

246. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA. 1997;277(6):488-94.
247. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol. 1995;48(12):1495-501.
248. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48(12):1503-10.
249. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059-79.
250. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
251. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol. 1996;49(8):907-16.
252. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med. 2011;9:103.
253. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med. 2010;8:20.
254. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59(10):1087-91.
255. Donner A. The Relative Effectiveness of Procedures Commonly Used in Multiple Regression Analysis for Dealing with Missing Values. The American Statistician. 1982;36(4):378-81.
256. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol. 1995;142(12):1255-64.
257. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res. 1999;8(1):3-15.
258. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.
259. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332(7549):1080.
260. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise

variables. British Journal of Mathematical and Statistical Psychology. 1992;45(2):265-82.
261. Steyerberg EW EM, Habbema JD. Application of shrinkage techniques in logistic regression analysis: a case study. . Statistica Neerlandica. 2001;55:76-88.
262. Efron B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. Journal of the American Statistical Association. 1983;78(382):316-31.
263. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115(7):928-35.
264. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
265. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol. 2005;58(5):475-83.
266. Peek N, Arts DG, Bosman RJ, van der Voort PH, de Keizer NF. External validation of prognostic models for critically ill patients required substantial sample sizes. J Clin Epidemiol. 2007;60(5):491-501.
267. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med. 2007;35(9):2052-6.
268. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74.
269. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin Chem. 2008;54(1):17-23.
270. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med. 1997;16(9):965-80.
271. Hosmer DW LS. Applied Logistic Regression. New York: Wiley; 1989. p. 187-215.
272. Bertolini G, D'Amico R, Nardi D, Tinazzi A, Apolone G. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. J Epidemiol Biostat. 2000;5(4):251-3.
273. Marcin JP, Romano PS. Size matters to a model's fit. Crit Care Med. 2007;35(9):2212-3.
274. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med. 2014;33(3):517-35.
275. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Stat Methods Med Res. 2016;25(4):1692-706.
276. Cox DR. Two Further Applications of a Model for Binary Regression. Biometrika. 1958;45(3/4):562-5.

277. Miller ME, Langefeld CD, Tierney WM, Hui SL, McDonald CJ. Validation of probabilistic predictions. Med Decis Making. 1993;13(1):49-58.
278. Vach W. Calibration of clinical prediction rules does not just assess bias. J Clin Epidemiol. 2013;66(11):1296-301.
279. Miller ME, Hui SL, Tierney WM. Validation techniques for logistic regression models. Stat Med. 1991;10(8):1213-26.
280. Bannister CA, Poole CD, Jenkins-Jones S, Morgan CL, Elwyn G, Spasic I, et al. External validation of the UKPDS risk engine in incident type 2 diabetes: a need for new type 2 diabetes-specific risk equations. Diabetes Care. 2014;37(2):537-45.
281. Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. J Clin Epidemiol. 1992;45(1):85-9.
282. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. Biostatistics. 2005;6(2):227-39.
283. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683-90.
284. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515-24.
285. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003;56(5):441-7.
286. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201-9.
287. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol. 2003;56(9):826-32.
288. 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.
289. Yap CH, Reid C, Yii M, Rowland MA, Mohajeri M, Skillington PD, et al. Validation of the EuroSCORE model in Australia. Eur J Cardiothorac Surg. 2006;29(4):441-6; discussion 6.
290. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol. 2008;61(1):76-86.
291. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak. 2006;6:38.
292. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567-86.

293. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381.
294. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.
295. Ivanov J, Borger MA, David TE, Cohen G, Walton N, Naylor CD. Predictive accuracy study: comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery. Ann Thorac Surg. 2000;70(1):162-8.
296. Loeb M, Walter SD, McGeer A, Simor AE, McArthur MA, Norman G. A comparison of model-building strategies for lower respiratory tract infection in long-term care. J Clin Epidemiol. 1999;52(12):1239-48.
297. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453-73.
298. Sauerbrei W. The Use of Resampling Methods to Simplify Regression Models in Medical Statistics. Journal of the Royal Statistical Society Series C (Applied Statistics). 1999;48(3):313-29.
299. Rabar S, Lau R, O'Flynn N, Li L, Barry P. Risk assessment of fragility fractures: summary of NICE guidance. BMJ. 2012;345:e3698.
300. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S49-73.
301. Brand DA, Frazier WH, Kohlhepp WC, Shea KM, Hoefer AM, Ecker MD, et al. A protocol for selecting patients with injured extremities who need x-rays. N Engl J Med. 1982;306(6):333-9.
302. Dunlop MG, Beattie TF, White GK, Raab GM, Doull RI. Guidelines for selective radiological assessment of inversion ankle injuries. Br Med J (Clin Res Ed). 1986;293(6547):603-5.
303. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol. 2008;61(4):331-43.
304. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14:40.
305. Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, et al. Tailored interventions to address determinants of practice. Cochrane Database Syst Rev. 2015;4:CD005470.

306. Meyer G, Kopke S, Haastert B, Muhlhauser I. Comparison of a fall risk assessment tool with nurses' judgement alone: a cluster-randomised controlled trial. Age Ageing. 2009;38(4):417-23.
307. Marrie TJ, Lau CY, Wheeler SL, Wong CJ, Vandervoort MK, Feagan BG. A controlled trial of a critical pathway for treatment of community-acquired pneumonia. CAPITAL Study Investigators. Community-Acquired Pneumonia Intervention Trial Assessing Levofloxacin. JAMA. 2000;283(6):749-55.
308. Foy R, Penney GC, Grimshaw JM, Ramsay CR, Walker AE, MacLennan G, et al. A randomised controlled trial of a tailored multifaceted strategy to promote implementation of a clinical guideline on induced abortion care. BJOG. 2004;111(7):726-33.
309. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.
310. Brehaut JC, Stiell IG, Visentin L, Graham ID. Clinical decision rules "in the real world": how a widely disseminated rule is used in everyday practice. Acad Emerg Med. 2005;12(10):948-56.
311. Graham ID, Stiell IG, Laupacis A, O'Connor AM, Wells GA. Emergency physicians' attitudes toward and use of clinical decision rules for radiography. Acad Emerg Med. 1998;5(2):134-40.
312. Brehaut JC, Stiell IG, Graham ID. Will a new clinical decision rule be widely used? The case of the Canadian C-spine rule. Acad Emerg Med. 2006;13(4):413-20.
313. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat. 1992;22(3):207-19.
314. Brehaut JC, Graham ID, Wood TJ, Taljaard M, Eagles D, Lott A, et al. Measuring acceptability of clinical decision rules: validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. Med Decis Making. 2010;30(3):398-408.
315. Mehta RL, Cerdá J, Burdmann EA, Tonelli M, García-García G, Jha V, et al. International Society of Nephrology's 0by25 initiative for acute kidney injury (zero preventable deaths by 2025): a human rights case for nephrology. The Lancet.
316. National Institute for Health and Care Excellence. Acute Kidney Injury: Prevention, Detection and Management Up to the Point of Renal Replacement Therapy. (Clinical guideline CG169). London: National Clinical Guideline Centre.; 2013 Aug.
317. Rasmussen HH, Pitt EA, Ibels LS, McNeil DR. Prediction of outcome in acute renal failure by discriminant analysis of clinical variables. Arch Intern Med. 1985;145(11):2015-8.
318. Lohr JW, McFarlane MJ, Grantham JJ. A clinical index to predict survival in acute renal failure patients requiring dialysis. Am J Kidney Dis. 1988;11(3):254-9.

319. Schaefer JH, Jochimsen F, Keller F, Wegscheider K, Distler A. Outcome prediction of acute renal failure in medical intensive care. Intensive Care Med. 1991;17(1):19-24.
320. Liano F, Gallego A, Pascual J, Garcia-Martin F, Teruel JL, Marcen R, et al. Prognosis of acute tubular necrosis: an extended prospectively contrasted study. Nephron. 1993;63(1):21-31.
321. Paganini EP, Halstenberg WK, Goormastic M. Risk modeling in acute renal failure requiring dialysis: the introduction of a new model. Clin Nephrol. 1996;46(3):206-11.
322. Chertow GM, Lazarus JM, Paganini EP, Allgren RL, Lafayette RA, Sayegh MH. Predictors of mortality and the provision of dialysis in patients with acute tubular necrosis. The Auriculin Anaritide Acute Renal Failure Study Group. J Am Soc Nephrol. 1998;9(4):692-8.
323. Lins RL, Elseviers M, Daelemans R, Zachee P, Zachee P, Gheuens E, et al. Prognostic value of a new scoring system for hospital mortality in acute renal failure. Clin Nephrol. 2000;53(1):10-7.
324. Mehta RL, Pascual MT, Gruta CG, Zhuang S, Chertow GM. Refining predictive models in critically ill patients with acute renal failure. J Am Soc Nephrol. 2002;13(5):1350-7.
325. Lins RL, Elseviers MM, Daelemans R, Arnouts P, Billiouw JM, Couttenye M, et al. Re-evaluation and modification of the Stuivenberg Hospital Acute Renal Failure (SHARF) scoring system for the prognosis of acute renal failure: an independent multicentre, prospective study. Nephrol Dial Transplant. 2004;19(9):2282-8.
326. Dharan KS, John GT, Antonisamy B, Kirubakaran MG, Jacob CK. Prediction of mortality in acute renal failure in the tropics. Ren Fail. 2005;27(3):289-96.
327. Chertow GM, Soroko SH, Paganini EP, Cho KC, Himmelfarb J, Ikizler TA, et al. Mortality after acute renal failure: models for prognostic stratification and risk adjustment. Kidney Int. 2006;70(6):1120-6.
328. Demirjian S, Chertow GM, Zhang JH, O'Connor TZ, Vitale J, Paganini EP, et al. Model to predict mortality in critically ill adults with acute kidney injury. Clin J Am Soc Nephrol. 2011;6(9):2114-20.
329. Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ. 2001;323(7306):224-8.
330. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144(6):427-37.
331. Huen SC, Parikh CR. Predicting acute kidney injury after cardiac surgery: a systematic review. Ann Thorac Surg. 2012;93(1):337-47.
332. Silver SA, Shah PM, Chertow GM, Harel S, Wald R, Harel Z. Risk prediction models for contrast induced nephropathy: systematic review. BMJ. 2015;351:h4395.

333. Caragata R, Wyssusek KH, Kruger P. Acute kidney injury following liver transplantation: a systematic review of published predictive models. Anaesth Intensive Care. 2016;44(2):251-61.
334. Wilson T, Quan S, Cheema K, Zarnke K, Quinn R, de Koning L, et al. Risk prediction models for acute kidney injury following major noncardiac surgery: systematic review. Nephrol Dial Transplant. 2016;31(2):231-40.
335. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
336. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. Eur J Cancer. 2007;43(17):2559-79.
337. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA. 1992;268(2):240-8.
338. Oxman AD, Guyatt GH. The science of reviewing research. Ann N Y Acad Sci. 1993;703:125-33; discussion 33-4.
339. Morton S, Berg A, Levit L, Eden J. Finding what works in health care: standards for systematic reviews: National Academies Press; 2011.
340. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One. 2012;7(2):e32844.
341. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc. 2001;8(4):391-7.
342. Wilczynski NL, McKibbon KA, Walter SD, Garg AX, Haynes RB. MEDLINE clinical queries are robust when searching in recent publishing years. J Am Med Inform Assoc. 2013;20(2):363-8.
343. Mehran R, Aymong ED, Nikolsky E, Lasic Z, Iakovou I, Fahy M, et al. A simple risk score for prediction of contrast-induced nephropathy after percutaneous coronary intervention: development and initial validation. J Am Coll Cardiol. 2004;44(7):1393-9.
344. Aronson S, Fontes ML, Miao Y, Mangano DT, Investigators of the Multicenter Study of Perioperative Ischemia Research G, Ischemia R, et al. Risk index for perioperative renal dysfunction/failure: critical dependence on pulse pressure hypertension. Circulation. 2007;115(6):733-42.
345. Bell S, Dekker FW, Vadiveloo T, Marwick C, Deshmukh H, Donnan PT, et al. Risk of postoperative acute kidney injury in patients undergoing orthopaedic surgery-development and validation of a risk score and effect of acute kidney injury on survival: observational cohort study. BMJ. 2015;351:h5639.

346. Tsai TT, Patel UD, Chang TI, Kennedy KF, Masoudi FA, Matheny ME, et al. Validated contemporary risk model of acute kidney injury in patients undergoing percutaneous coronary interventions: insights from the National Cardiovascular Data Registry Cath-PCI Registry. J Am Heart Assoc. 2014;3(6):e001380.
347. Tziakas D, Chalikias G, Stakos D, Altun A, Sivri N, Yetkin E, et al. Validation of a new risk score to predict contrast-induced nephropathy after percutaneous coronary intervention. Am J Cardiol. 2014;113(9):1487-93.
348. Hodgson LE, Dimitrov BD, Roderick PJ, Venn R, Forni LG. Predicting AKI in emergency admissions: an external validation study of the acute kidney injury prediction score (APS). BMJ Open. 2017;7(3):e013511.
349. Mehta RH, Grab JD, O'Brien SM, Bridges CR, Gammie JS, Haan CK, et al. Bedside tool for predicting the risk of postoperative dialysis in patients undergoing cardiac surgery. Circulation. 2006;114(21):2208-16; quiz
350. Wijeysundera DN, Karkouti K, Dupuis JY, Rao V, Chan CT, Granton JT, et al. Derivation and validation of a simplified predictive index for renal replacement therapy after cardiac surgery. JAMA. 2007;297(16):1801-9.
351. 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):624-31.
352. Ng SY, Sanagou M, Wolfe R, Cochrane A, Smith JA, Reid CM. Prediction of acute kidney injury within 30 days of cardiac surgery. J Thorac Cardiovasc Surg. 2014;147(6):1875-83, 83 e1.
353. Brown JR, Kramer RS, MacKenzie TA, Coca SG, Sint K, Parikh CR. Determinants of acute kidney injury duration after cardiac surgery: an externally validated tool. Ann Thorac Surg. 2012;93(2):570-6.
354. Rahmanian PB, Kwiecien G, Langebartels G, Madershahian N, Wittwer T, Wahlers T. Logistic risk model predicting postoperative renal failure requiring dialysis in cardiac surgery patients. Eur J Cardiothorac Surg. 2011;40(3):701-7.
355. Heise D, Kunze N, Bürger J, Quintel M, Perl T. Modifying a kidney injury score by including perioperative data Comparison of three predictive scores. Central European Journal of Medicine. 2012;7(4):420-8.
356. Kim WH, Lee SM, Choi JW, Kim EH, Lee JH, Jung JW, et al. Simplified clinical risk score to predict acute kidney injury after aortic surgery. J Cardiothorac Vasc Anesth. 2013;27(6):1158-66.
357. Berg KS, Stenseth R, Wahba A, Pleym H, Videm V. How can we best predict acute kidney injury following cardiac surgery?: a prospective observational study. Eur J Anaesthesiol. 2013;30(11):704-12.
358. Demirjian S, Schold JD, Navia J, Mastracci TM, Paganini EP, Yared JP, et al. Predictive models for acute kidney injury following cardiac surgery. Am J Kidney Dis. 2012;59(3):382-9.

359. Brown JR, Cochran RP, Leavitt BJ, Dacey LJ, Ross CS, MacKenzie TA, et al. Multivariable prediction of renal insufficiency developing after cardiac surgery. Circulation. 2007;116(11 Suppl):I139-43.
360. Rueggeberg A, Boehm S, Napieralski F, Mueller AR, Neuhaus P, Falke KJ, et al. Development of a risk stratification model for predicting acute renal failure in orthotopic liver transplantation recipients. Anaesthesia. 2008;63(11):1174-80.
361. Sanchez EQ, Gonwa TA, Levy MF, Goldstein RM, Mai ML, Hays SR, et al. Preoperative and Perioperative Predictors of the Need for Renal Replacement Therapy After Orthotopic Liver Transplantation. Transplantation. 2004;78(7):1048-54.
362. Romano TG, Schmidtbauer I, Silva FM, Pompilio CE, D'Albuquerque LA, Macedo E. Role of MELD score and serum creatinine as prognostic tools for the development of acute kidney injury after liver transplantation. PLoS One. 2013;8(5):e64089.
363. Kim JM, Jo YY, Na SW, Kim SI, Choi YS, Kim NO, et al. The predictors for continuous renal replacement therapy in liver transplant recipients. Transplant Proc. 2014;46(1):184-91.
364. Slankamenac K, Beck-Schimmer B, Breitenstein S, Puhan MA, Clavien PA. Novel prediction score including pre- and intraoperative parameters best predicts acute kidney injury after liver surgery. World J Surg. 2013;37(11):2618-28.
365. Slankamenac K, Breitenstein S, Held U, Beck-Schimmer B, Puhan MA, Clavien PA. Development and validation of a prediction score for postoperative acute renal failure following liver resection. Ann Surg. 2009;250(5):720-8.
366. Xu X, Ling Q, Wei Q, Wu J, Gao F, He ZL, et al. An effective model for predicting acute kidney injury after liver transplantation. Hepatobiliary Pancreat Dis Int. 2010;9(3):259-63.
367. Park MH, Shim HS, Kim WH, Kim HJ, Kim DJ, Lee SH, et al. Clinical Risk Scoring Models for Prediction of Acute Kidney Injury after Living Donor Liver Transplantation: A Retrospective Observational Study. PLoS One. 2015;10(8):e0136230.
368. Hilmi IA, Damian D, Al-Khafaji A, Planinsic R, Boucek C, Sakai T, et al. Acute kidney injury following orthotopic liver transplantation: incidence, risk factors, and effects on patient and graft outcomes. Br J Anaesth. 2015;114(6):919-26.
369. Boyle JM, Moualla S, Arrigain S, Worley S, Bakri MH, Starling RC, et al. Risks and outcomes of acute kidney injury requiring dialysis after cardiac transplantation. Am J Kidney Dis. 2006;48(5):787-96.
370. Kilic A, Grimm JC, Shah AS, Conte JV, Whitman GJ, Sciortino CM. An easily calculable and highly predictive risk index for postoperative renal failure

after heart transplantation. J Thorac Cardiovasc Surg. 2014;148(3):1099-104; discussion 104-5.
371. Grimm JC, Lui C, Kilic A, Valero V, 3rd, Sciortino CM, Whitman GJ, et al. A risk score to predict acute renal failure in adult patients after lung transplantation. Ann Thorac Surg. 2015;99(1):251-7.
372. Laville M, Juillard L. Contrast-induced acute kidney injury: how should at-risk patients be identified and managed? J Nephrol. 2010;23(4):387-98.
373. Maeder M, Klein M, Fehr T, Rickli H. Contrast nephropathy: review focusing on prevention. J Am Coll Cardiol. 2004;44(9):1763-71.
374. McCullough PA, Adam A, Becker CR, Davidson C, Lameire N, Stacul F, et al. Epidemiology and prognostic implications of contrast-induced nephropathy. Am J Cardiol. 2006;98(6A):5K-13K.
375. Solomon R, Dauerman HL. Contrast-induced acute kidney injury. Circulation. 2010;122(23):2451-5.
376. Bartholomew BA, Harjai KJ, Dukkipati S, Boura JA, Yerkey MW, Glazier S, et al. Impact of nephropathy after percutaneous coronary intervention and a method for risk stratification. Am J Cardiol. 2004;93(12):1515-9.
377. Marenzi G, Lauri G, Assanelli E, Campodonico J, De Metrio M, Marana I, et al. Contrast-induced nephropathy in patients undergoing primary angioplasty for acute myocardial infarction. J Am Coll Cardiol. 2004;44(9):1780-5.
378. Tziakas D, Chalikias G, Stakos D, Apostolakis S, Adina T, Kikas P, et al. Development of an easily applicable risk score model for contrast-induced nephropathy prediction after percutaneous coronary intervention: a novel approach tailored to current practice. Int J Cardiol. 2013;163(1):46-55.
379. Freeman RV, O'Donnell M, Share D, Meengs WL, Kline-Rogers E, Clark VL, et al. Nephropathy requiring dialysis after percutaneous coronary intervention and the critical role of an adjusted contrast dose. Am J Cardiol. 2002;90(10):1068-73.
380. Heise D, Sundermann D, Braeuer A, Quintel M. Validation of a clinical score to determine the risk of acute renal failure after cardiac surgery. Eur J Cardiothorac Surg. 2010;37(3):710-6.
381. Matheny ME, Miller RA, Ikizler TA, Waitman LR, Denny JC, Schildcrout JS, et al. Development of inpatient risk stratification models of acute kidney injury for use in electronic health records. Med Decis Making. 2010;30(6):639-50.
382. Drawz PE, Miller RT, Sehgal AR. Predicting hospital-acquired acute kidney injury--a case-controlled study. Ren Fail. 2008;30(9):848-55.
383. Breidthardt T, Socrates T, Noveanu M, Klima T, Heinisch C, Reichlin T, et al. Effect and clinical prediction of worsening renal function in acute decompensated heart failure. Am J Cardiol. 2011;107(5):730-5.

384. Wang YN, Cheng H, Yue T, Chen YP. Derivation and validation of a prediction score for acute kidney injury in patients hospitalized with acute heart failure in a Chinese cohort. Nephrology (Carlton). 2013;18(7):489-96.
385. Forman DE, Butler J, Wang Y, Abraham WT, O'Connor CM, Gottlieb SS, et al. Incidence, predictors at admission, and impact of worsening renal function among patients hospitalized with heart failure. J Am Coll Cardiol. 2004;43(1):61-7.
386. Xing XZ, Wang HJ, Huang CL, Yang QH, Qu SN, Zhang H, et al. Two acute kidney injury risk scores for critically ill cancer patients undergoing non-cardiac surgery. World J Emerg Med. 2012;3(4):278-81.
387. Hodgson L, Dimitrov, BD, Roderick, PJ, Venn, R, LG, Forni. Predicting AKI in Emergency Admissions: An external validation study of the acute kidney injury prediction score (APS). Bmj Open. 2017;In Press.
388. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ. 2011;343:d7163.
389. Bihorac A, Brennan M, Baslanti TO, Bozorgmehri S, Efron PA, Moore FA, et al. NATIONAL SURGICAL QUALITY IMPROVEMENT PROGRAM UNDERESTIMATES THE RISK ASSOCIATED WITH MILD AND MODERATE POSTOPERATIVE ACUTE KIDNEY INJURY. Critical care medicine. 2013;41(11):10.1097/CCM.0b013e31829860fc.
390. Ensor J, Riley RD, Moore D, Snell KI, Bayliss S, Fitzmaurice D. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. BMJ Open. 2016;6(5):e011190.
391. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356.
392. Hospital Episode Statistics Analysis HaSCIC. Hospital Episode Statistics: Admitted patient care - 2014-15. http://www.hscic.gov.uk/pubs/hes1415. Health and Social Care Information Centre, 2015 25th November 2015. Report No.: 978-1-78386-547-5.
393. Porter CJ, Juurlink I, Bisset LH, Bavakunji R, Mehta RL, Devonald MA. A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. Nephrol Dial Transplant. 2014;29(10):1888-93.
394. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353.
395. Kannry J, McCullagh L, Kushniruk A, Mann D, Edonyabo D, McGinn T. A Framework for Usable and Effective Clinical Decision Support: Experience from the iCPR Randomized Clinical Trial. EGEMS (Wash DC). 2015;3(2):1150.
396. Kolhe NV, Staples D, Reilly T, Merrison D, McIntyre CW, Fluck RJ, et al. Impact of Compliance with a Care Bundle on Acute Kidney Injury Outcomes: A Prospective Observational Study. PLoS One. 2015;10(7):e0132279.

397. Koyner JL, Davison DL, Brasha-Mitchell E, Chalikonda DM, Arthur JM, Shaw AD, et al. Furosemide Stress Test and Biomarkers for the Prediction of AKI Severity. J Am Soc Nephrol. 2015;26(8):2023-31.
398. Coggon DI, Martyn CN. Time and chance: the stochastic nature of disease causation. Lancet. 2005;365(9468):1434-7.
399. 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. Crit Care Med. 2000;28(1):150-6.
400. Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. Eur Radiol. 2015;25(4):932-9.
401. Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. Cancer Invest. 2009;27(3):235-43.
402. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27(2):157-72; discussion 207-12.
403. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38.
404. Raven MC, Billings JC, Goldfrank LR, Manheimer ED, Gourevitch MN. Medicaid Patients at High Risk for Frequent Hospital Admission: Real-Time Identification and Remediable Risks. Journal of Urban Health : Bulletin of the New York Academy of Medicine. 2009;86(2):230-41.
405. Ko M, Lee Y, Chen C, Chou P, Chu D. Prevalence of and Predictors for Frequent Utilization of Emergency Department: A Population-Based Study. Medicine (Baltimore). 2015;94(29):e1205.
406. Longman JM, M IR, Passey MD, Heathcote KE, Ewald DP, Dunn T, et al. Frequent hospital admission of older people with chronic disease: a cross-sectional survey with telephone follow-up and data linkage. BMC Health Serv Res. 2012;12:373.
407. Kirby SE, Dennis SM, Jayasinghe UW, Harris MF. Patient related factors in frequent readmissions: the influence of condition, access to services and patient choice. BMC Health Serv Res. 2010;10:216.
408. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc. 2003;10(6):523-30.
409. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical

decision rules. Evidence-Based Medicine Working Group. JAMA. 2000;284(1):79-84.
410. Keogh C, Wallace E, O'Brien KK, Galvin R, Smith SM, Lewis C, et al. Developing an International Register of Clinical Prediction Rules for Use in Primary Care: A Descriptive Analysis. Annals of Family Medicine. 2014;12(4):359-66.
411. Wallace E, Uijen MJM, Clyne B, Zarabzadeh A, Keogh C, Galvin R, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. BMJ Open. 2016;6(3):e009957.
412. Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. Crit Care Med. 1974;2(2):57-60.
413. Le Gall JR. The use of severity scores in the intensive care unit. Intensive Care Med. 2005;31(12):1618-23.
414. Teres D, Lemeshow S. Why severity models should be used with caution. Crit Care Clin. 1994;10(1):93-110; discussion 1-5.
415. 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):1619-36.
416. 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):2957-63.
417. Cowen JS, Kelley MA. Errors and bias in using predictive scoring systems. Crit Care Clin. 1994;10(1):53-72.
418. 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):1336-44.
419. 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):1345-55.
420. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34(5):1297-310.
421. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med. 2007;35(3):827-35.
422. Knaus WA, Harrell FE, Jr., Lynn J, Goldman L, Phillips RS, Connors AF, Jr., et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med. 1995;122(3):191-203.

423. Wells PS, Anderson DR, Bormanis J, Guy F, Mitchell M, Gray L, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. Lancet. 1997;350(9094):1795-8.
424. Hall LM, Jung RT, Leese GP. Controlled trial of effect of documented cardiovascular risk scores on prescribing. BMJ. 2003;326(7383):251-2.
425. Auleley GR, Ravaud P, Giraudeau B, Kerboull L, Nizard R, Massin P, et al. Implementation of the Ottawa ankle rules in France. A multicenter randomized controlled trial. JAMA. 1997;277(24):1935-9.
426. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. J Clin Epidemiol. 2011;64(9):936-48.
427. Stiell IG, McKnight RD, Greenberg GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. JAMA. 1994;271(11):827-32.
428. Cameron C, Naylor CD. No impact from active dissemination of the Ottawa Ankle Rules: further evidence of the need for local implementation of practice guidelines. CMAJ. 1999;160(8):1165-8.
429. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA. 1995;274(9):700-5.
430. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999;282(15):1458-65.
431. Moons KG. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem. 2010;56(4):537-41.
432. Card D, Krueger AB. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. National Bureau of Economic Research, 1993.
433. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA. 2014;312(22):2401-2.
434. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348(26):2635-45.
435. Kohn LT CJ, Donaldson MS, eds. To err is human: building a safer health system. . Washington, DC: 1999.
436. Lomas J, Sisk JE, Stocking B. From evidence to practice in the United States, the United Kingdom, and Canada. Milbank Q. 1993;71(3):405-10.
437. Chertow GM, Lee J, Kuperman GJ, Burdick E, Horsky J, Seger DL, et al. Guided medication dosing for inpatients with renal insufficiency. JAMA. 2001;286(22):2839-44.
438. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007;14(2):141-5.

439. Linder JA, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. Arch Intern Med. 2007;167(13):1400-5.
440. Zhou L, Soran CS, Jenter CA, Volk LA, Orav EJ, Bates DW, et al. The relationship between electronic health record use and quality of care over time. J Am Med Inform Assoc. 2009;16(4):457-64.
441. Blumenthal D. Launching HITECH. N Engl J Med. 2010;362(5):382-5.
442. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med. 2012;157(1):29-43.
443. Roshanov PS, Misra S, Gerstein HC, Garg AX, Sebaldt RJ, Mackay JA, et al. Computerized clinical decision support systems for chronic disease management: A decision-maker-researcher partnership systematic review. Implementation Science : IS. 2011;6:92-.
444. Phansalkar S, Edworthy J, Hellier E, Seger DL, Schedlbauer A, Avery AJ, et al. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. Journal of the American Medical Informatics Association : JAMIA. 2010;17(5):493-501.
445. Strom BL, Schinnar R, Aberra F, Bilker W, Hennessy S, Leonard CE, et al. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. Arch Intern Med. 2010;170(17):1578-83.
446. Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics. 2005;116(6):1506-12.
447. Cropp AJ, Woods LA, Raney D, Bredle DL. Name that tone. The proliferation of alarms in the intensive care unit. Chest. 1994;105(4):1217-20.
448. James MT, Hobson CE, Darmon M, Mohan S, Hudson D, Goldstein SL, et al. Applications for detection of acute kidney injury using electronic medical records and clinical information systems: workgroup statements from the 15(th) ADQI Consensus Conference. Can J Kidney Health Dis. 2016;3:9.
449. Siew ED, Matheny ME, Ikizler TA, Lewis JB, Miller RA, Waitman LR, et al. Commonly Used Surrogates for Baseline Renal Function Can Impact Acute Kidney Injury Classification and Prognosis. Kidney international. 2010;77(6):536-42.
450. Bagshaw SM, Uchino S, Cruz D, Bellomo R, Morimatsu H, Morgera S, et al. A comparison of observed versus estimated baseline creatinine for determination of RIFLE class in patients with acute kidney injury. Nephrol Dial Transplant. 2009;24(9):2739-44.
451. Bliss JP, Gilson RD, Deaton JE. Human probability matching behaviour in response to alarms of varying reliability. Ergonomics. 1995;38(11):2300-12.

452. Bliss JP, Dunn MC. Behavioural implications of alarm mistrust as a function of task workload. Ergonomics. 2000;43(9):1283-300.
453. Westenskow DR, Orr JA, Simon FH, Bender HJ, Frankenberger H. Intelligent alarms reduce anesthesiologist's response time to critical faults. Anesthesiology. 1992;77(6):1074-9.
454. Wickens CD, Carswell CM. The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1995;37(3):473-94.
455. Sanders AF. Some aspects of the selective process in the functional visual field. Ergonomics. 1970;13(1):101-17.
456. Braun CC, Silver NC. Interaction of signal word and colour on warning labels: differences in perceived hazard and behavioural compliance. Ergonomics. 1995;38(11):2207-20.
457. Paterno MD, Maviglia SM, Gorman PN, Seger DL, Yoshida E, Seger AC, et al. Tiering drug-drug interaction alerts by severity increases compliance rates. J Am Med Inform Assoc. 2009;16(1):40-6.
458. Leonard S, Karnes E, Schneider T. Scale values for warning symbols and words. Trends in Ergonomics/Human Factors V Elsevier, Amsterdam. 1988:669-74.
459. Bostrom A, Atman CJ, Fischhoff B, Morgan MG. Evaluating risk communications: completing and correcting mental models of hazardous processes, Part II. Risk Anal. 1994;14(5):789-98.
460. Nightingale PG, Adu D, Richards NT, Peters M. Implementation of rules based computerised bedside prescribing and administration: intervention study. BMJ : British Medical Journal. 2000;320(7237):750-3.
461. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of Drug Safety Alerts in Computerized Physician Order Entry. Journal of the American Medical Informatics Association : JAMIA. 2006;13(2):138-47.
462. Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS. Physicians' decisions to override computerized drug alerts in primary care. Arch Intern Med. 2003;163(21):2625-31.
463. Ash JS, Berg M, Coiera E. Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors. Journal of the American Medical Informatics Association : JAMIA. 2004;11(2):104-12.
464. Kanouse DE, Kallich JD, Kahan JP. Dissemination of effectiveness and outcomes research. Health Policy. 1995;34(3):167-92.
465. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA. 1998;280(15):1339-46.
466. Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes:

a synthesis of high-quality systematic review findings. J Am Med Inform Assoc. 2011;18(3):327-34.
467. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S-A. Computerized clinical decision support for prescribing: provision does not guarantee uptake. Journal of the American Medical Informatics Association : JAMIA. 2010;17(1):25-33.
468. Carroll AE, Anand V, Downs SM. Understanding why clinicians answer or ignore clinical decision support prompts. Appl Clin Inform. 2012;3(3):309-17.
469. Lee F, Teich JM, Spurr CD, Bates DW. Implementation of physician order entry: user satisfaction and self-reported usage patterns. J Am Med Inform Assoc. 1996;3(1):42-55.
470. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. N Engl J Med. 1976;295(24):1351-5.
471. Maviglia SM, Zielstorff RD, Paterno M, Teich JM, Bates DW, Kuperman GJ. Automating complex guidelines for chronic disease: lessons learned. J Am Med Inform Assoc. 2003;10(2):154-65.
472. Dexter PR, Perkins S, Overhage JM, Maharry K, Kohler RB, McDonald CJ. A computerized reminder system to increase the use of preventive care for hospitalized patients. N Engl J Med. 2001;345(13):965-70.
473. Resar R, Griffin F, Haraden C, Nolan T. Using care bundles to improve health care quality. IHI innovation series white paper. Cambridge, MA: Institute for Healthcare Improvement; 2012. 2014.
474. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med. 2006;355(26):2725-32.
475. Steinmo SH, Michie S, Fuller C, Stanley S, Stapleton C, Stone SP. Bridging the gap between pragmatic intervention design and theory: using behavioural science tools to modify an existing quality improvement programme to implement “Sepsis Six”. Implementation Science. 2015;11(1):14.
476. Cox ZL, McCoy AB, Matheny ME, Bhave G, Peterson NB, Siew ED, et al. Adverse drug events during AKI and its recovery. Clin J Am Soc Nephrol. 2013;8(7):1070-8.
477. Meran S, Wonnacott A, Amphlett B, Phillips A. How good are we at managing acute kidney injury in hospital? Clin Kidney J. 2014;7(2):144-50.
478. Selby NM, Kolhe NV. Care Bundles for Acute Kidney Injury: Do They Work? Nephron. 2016;134(3):195-9.
479. Joannidis M, Druml W, Forni LG, Groeneveld AB, Honore P, Oudemans-van Straaten HM, et al. Prevention of acute kidney injury and protection of renal function in the intensive care unit. Expert opinion of the Working Group for Nephrology, ESICM. Intensive Care Med. 2010;36(3):392-411.

480. Stage 3 'Directive' Patient Safety Alert: 'Standardising the early identification of Acute Kidney Injury'. [press release]. https://www.england.nhs.uk/2014/06/09/psa-aki/, June 2014.
481. Lachance P, Villeneuve P-M, Wilson FP, Selby NM, Featherstone R, Rewa O, et al. Impact of e-alert for detection of acute kidney injury on processes of care and outcomes: protocol for a systematic review and meta-analysis. BMJ Open. 2016;6(5):e011152.
482. Hoste EA, Kashani K, Gibney N, Wilson FP, Ronco C, Goldstein SL, et al. Impact of electronic-alerting of acute kidney injury: workgroup statements from the 15(th) ADQI Consensus Conference. Can J Kidney Health Dis. 2016;3:10.
483. Statistics OfN. 2011 Census General Report for England and Wales https://www.ons.gov.uk/census/2011census/2011censusdata: ONS Statistics OfN; 2011.
484. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new MRC guidance. Journal of epidemiology and community health. 2012;66(12):1182-6.
485. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). BMC Med Inform Decis Mak. 2011;11:62.
486. Wogalter MS, Godfrey SS, Fontenelle GA, Desaulniers DR, Rothstein PR, Laughery KR. Effectiveness of warnings. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1987;29(5):599-612.
487. Laughery KR, Vaubel KP, Young SL, Brelsford Jr JW, Rowe AL. Explicitness of consequence information in warnings. Safety Science. 1993;16(5–6):597-613.
488. <NICE_AKI_Full.pdf>.
489. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986:13-22.
490. Morris S, Hunter RM, Ramsay AIG, Boaden R, McKevitt C, Perry C, et al. Impact of centralising acute stroke services in English metropolitan areas on mortality and length of hospital stay: difference-in-differences analysis. BMJ : British Medical Journal. 2014;349.
491. Ashenfelter OC, Card D. Using the longitudinal structure of earnings to estimate the effect of training programs. National Bureau of Economic Research Cambridge, Mass., USA; 1984.
492. Patel MS, Volpp KG, Small DS, Hill AS, Even-Shoshan O, Rosenbaum L, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients. JAMA. 2014;312(22):2364-73.
493. Rajaram R, Chung JW, Jones AT, Cohen ME, Dahlke AR, Ko CY, et al. Association of the 2011 ACGME resident duty hour reform with general

surgery patient outcomes and with resident examination performance. JAMA. 2014;312(22):2374-84.
494. Angrist JD, Pischke J-S. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. Journal of Economic Perspectives. 2010;24(2):3-30.
495. Dimick JB, Nicholas LH, Ryan AM, Thumma JR, Birkmeyer JD. Bariatric surgery complications before vs after implementation of a national policy restricting coverage to centers of excellence. JAMA. 2013;309(8):792-9.
496. Nguyen NT, Hohmann S, Slone J, Varela E, Smith BR, Hoyt D. Improved bariatric surgery outcomes for Medicare beneficiaries after implementation of the medicare national coverage determination. Arch Surg. 2010;145(1):72-8.
497. McCullough JS, Christianson J, Leerapan B. Do electronic medical records improve diabetes quality in physician practices? Am J Manag Care. 2013;19(2):144-9.
498. Ryan AM, Burgess JF, Jr., Dimick JB. Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. Health Serv Res. 2015;50(4):1211-35.
499. Abadie A. Semiparametric difference-in-differences estimators. The Review of Economic Studies. 2005;72(1):1-19.
500. Boutis K, Grootendorst P, Willan A, Plint AC, Babyn P, Brison RJ, et al. Effect of the Low Risk Ankle Rule on the frequency of radiography in children with ankle injuries. CMAJ. 2013;185(15):E731-8.
501. McGinn TG, McCullagh L, Kannry J, Knaus M, Sofianou A, Wisnivesky JP, et al. Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. JAMA Intern Med. 2013;173(17):1584-91.
502. Poulton EC. Searching for newspaper headlines printed in capitals or lower-case letters. J Appl Psychol. 1967;51(5):417-25.
503. Kanagasundaram NS, Bevan MT, Sims AJ, Heed A, Price DA, Sheerin NS. Computerized clinical decision support for the early recognition and management of acute kidney injury: a qualitative evaluation of end-user experience. Clinical Kidney Journal. 2016;9(1):57-62.
504. McCoy AB, Waitman LR, Gadd CS, Danciu I, Smith JP, Lewis JB, et al. A computerized provider order entry intervention for medication safety during acute kidney injury: a quality improvement report. Am J Kidney Dis. 2010;56(5):832-41.
505. Isaac T, Weissman JS, Davis RB, Massagli M, Cyrulik A, Sands DZ, et al. Overrides of medication alerts in ambulatory care. Arch Intern Med. 2009;169(3):305-11.
506. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ. 2013;346:f657.

507. Benning A, Dixon-Woods M, Nwulu U, Ghaleb M, Dawson J, Barber N, et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. BMJ. 2011;342.
508. Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 2. Study design. Quality and Safety in Health Care. 2008;17(3):163-9.
509. Hospital Episode Statistics, Admitted Patient Care, England - 2012-13, http://content.digital.nhs.uk/hesdata. 2013.
510. Bion JF, Heffner JE. Challenges in the care of the acutely ill. Lancet. 2004;363(9413):970-7.
511. van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, et al. Intensive insulin therapy in critically ill patients. N Engl J Med. 2001;345(19):1359-67.
512. Finfer S, Chittock DR, Su SY, Blair D, Foster D, Dhingra V, et al. Intensive versus conventional glucose control in critically ill patients. N Engl J Med. 2009;360(13):1283-97.
513. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345(19):1368-77.
514. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol. 2014;29(8):541-9.
515. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol. 2015;7:449-90.
516. Nielsen SL, Lassen AT, Gradel KO, Jensen TG, Kolmos HJ, Hallas J, et al. Bacteremia is associated with excess long-term mortality: a 12-year population-based cohort study. J Infect. 2015;70(2):111-26.
517. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.
518. Cook JA, Collins GS. The rise of big clinical databases. Br J Surg. 2015;102(2):e93-e101.
519. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ. 2008;336(7659):1475-82.
520. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ. 2010;340:c2442.
521. Lazo JK, Morss RE, Demuth JL. 300 billion served: Sources, perceptions, uses, and values of weather forecasts. Bulletin of the American Meteorological Society. 2009;90(6):785-98.
522. Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med. 2012;157(3):207-10.

523. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. J Am Med Inform Assoc. 2015;22(1):43-50.
524. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Grossmann C, Powers B, McGinnis JM, editors. Washington DC: National Academy of Sciences.; 2011.
525. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. Recognizing Sepsis as a Global Health Priority - A WHO Resolution. N Engl J Med. 2017;377(5):414-7.
526. Kramer AA, Postler G, Salhab KF, Mendez C, Carey LC, Rabb H. Renal ischemia/reperfusion leads to macrophage-mediated increase in pulmonary vascular permeability. Kidney Int. 1999;55(6):2362-7.
527. Ronco C, Haapio M, House AA, Anavekar N, Bellomo R. Cardiorenal syndrome. J Am Coll Cardiol. 2008;52(19):1527-39.
528. Kelly KJ. Acute renal failure: much more than a kidney disease. Semin Nephrol. 2006;26(2):105-13.
529. Kelly KJ. Distant effects of experimental renal ischemia/reperfusion injury. J Am Soc Nephrol. 2003;14(6):1549-58.
530. Oliver JA, Verna EC. Afferent mechanisms of sodium retention in cirrhosis and hepatorenal syndrome. Kidney Int. 2010;77(8):669-80.
531. Miyazawa S, Watanabe H, Miyaji C, Hotta O, Abo T. Leukocyte accumulation and changes in extra-renal organs during renal ischemia reperfusion in mice. J Lab Clin Med. 2002;139(5):269-78.
532. Golab F, Kadkhodaee M, Zahmatkesh M, Hedayati M, Arab H, Schuster R, et al. Ischemic and non-ischemic acute kidney injury cause hepatic damage. Kidney Int. 2009;75(8):783-92.
533. Park SW, Chen SW, Kim M, Brown KM, Kolls JK, D'Agati VD, et al. Cytokines induce small intestine and liver injury after renal ischemia or nephrectomy. Lab Invest. 2011;91(1):63-84.
534. Serteser M, Koken T, Kahraman A, Yilmaz K, Akbulut G, Dilek ON. Changes in hepatic TNF-alpha levels, antioxidant status, and oxidation products after renal ischemia/reperfusion injury in mice. J Surg Res. 2002;107(2):234-40.
535. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Stat Med. 2004;23(16):2509-25.
536. Altman DG, Andersen PK. Bootstrap investigation of the stability of a Cox regression model. Stat Med. 1989;8(7):771-83.
537. Chatfield C. Model Uncertainty, Data Mining and Statistical Inference. Journal of the Royal Statistical Society Series A (Statistics in Society). 1995;158(3):419-66.

538. van Houwelingen JC, le Cessie S. Logistic Regression, a review. Statistica Neerlandica. 1988;42(4):215-32.
539. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological). 1972;34(2):187-220.
540. Efron B, Tibshirani R. Improvements on Cross-Validation: The .632+ Bootstrap Method. Journal of the American Statistical Association. 1997;92(438):548-60.
541. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med. 1997;16(9):981-91.
542. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology. 1997;8(1):12-7.
543. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882-90.
544. Wang TJ, Gona P, Larson MG, Tofler GH, Levy D, Newton-Cheh C, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. N Engl J Med. 2006;355(25):2631-9.
545. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MSV, et al. Criteria for Evaluation of Novel Markers of Cardiovascular Risk: A Scientific Statement From the American Heart Association. Circulation. 2009;119(17):2408-16.
546. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2(8):e124.
547. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, et al. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008;167(3):362-8.
548. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Validity of prognostic models: when is a model clinically useful? Semin Urol Oncol. 2002;20(2):96-107.
549. Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol. 2010;63(8):938-9; author reply 9.
550. Shapiro AR. The evaluation of clinical predictions. A method and initial application. N Engl J Med. 1977;296(26):1509-14.
551. Mittlbock M, Schemper M. Explained variation for logistic regression. Stat Med. 1996;15(19):1987-97.
552. Choodari-Oskooei B, Royston P, Parmar MK. A simulation study of predictive ability measures in a survival model I: explained variation measures. Stat Med. 2012;31(23):2627-43.
553. Akaike H. A new look at the statistical model identification. IEEE transactions on automatic control. 1974;19(6):716-23.

554. Eriksen BO, Hoff KR, Solberg S. Prediction of acute renal failure after cardiac surgery: retrospective cross-validation of a clinical algorithm. Nephrol Dial Transplant. 2003;18(1):77-81.
555. Englberger L, Suri RM, Li Z, Dearani JA, Park SJ, Sundt TM, 3rd, et al. Validation of clinical scores predicting severe acute kidney injury after cardiac surgery. Am J Kidney Dis. 2010;56(4):623-31.
556. Candela-Toha A, Elias-Martin E, Abraira V, Tenorio MT, Parise D, de Pablo A, et al. Predicting acute renal failure after cardiac surgery: external validation of two new clinical scores. Clin J Am Soc Nephrol. 2008;3(5):1260-5.
557. Di Bella I, Da Col U, Ciampichini R, Affronti A, Santucci A, Fabbri M, et al. [Validation of a new scoring system to predict the risk of postoperative acute renal failure in cardiac surgery]. G Ital Cardiol (Rome). 2007;8(5):306-10.
558. Knapik P, Rozentryt P, Nadziakiewicz P, Polonski L, Zembala M. Retrospective cross-validation of simplified predictive index for renal replacement therapy after cardiac surgery. Interact Cardiovasc Thorac Surg. 2008;7(6):1101-6.
559. Vives M, Monedero P, Perez-Valdivieso JR, Garcia-Fernandez N, Lavilla J, Herreros J, et al. External validation and comparison of three scores to predict renal replacement therapy after cardiac surgery: a multicenter cohort. Int J Artif Organs. 2011;34(4):329-38.
560. Kiers HD, van den Boogaard M, Schoenmakers MC, van der Hoeven JG, van Swieten HA, Heemskerk S, et al. Comparison and clinical suitability of eight prediction models for cardiac surgery-related acute kidney injury. Nephrol Dial Transplant. 2013;28(2):345-51.
561. Wong B, St Onge J, Korkola S, Prasad B. Validating a scoring tool to predict acute kidney injury (AKI) following cardiac surgery. Can J Kidney Health Dis. 2015;2:3.
562. Brown JR, DeVries JT, Piper WD, Robb JF, Hearne MJ, Ver Lee PM, et al. Serious renal dysfunction after percutaneous coronary interventions can be predicted. Am Heart J. 2008;155(2):260-6.
563. Ghani AA, Tohamy KY. Risk score for contrast induced nephropathy following percutaneous coronary intervention. Saudi J Kidney Dis Transpl. 2009;20(2):240-5.
564. Maioli M, Toso A, Gallopin M, Leoncini M, Tedeschi D, Micheletti C, et al. Preprocedural score for risk of contrast-induced nephropathy in elective coronary angiography and intervention. J Cardiovasc Med (Hagerstown). 2010;11(6):444-9.
565. Fu N, Li X, Yang S, Chen Y, Li Q, Jin D, et al. Risk score for the prediction of contrast-induced nephropathy in elderly patients undergoing percutaneous coronary intervention. Angiology. 2013;64(3):188-94.
566. Gurm HS, Seth M, Kooiman J, Share D. A novel tool for reliable and accurate prediction of renal complications in patients undergoing percutaneous coronary intervention. J Am Coll Cardiol. 2013;61(22):2242-8.

567. Victor SM, Gnanaraj A, S V, Deshmukh R, Kandasamy M, Janakiraman E, et al. Risk scoring system to predict contrast induced nephropathy following percutaneous coronary intervention. Indian Heart J. 2014;66(5):517-24.
568. Chen YL, Fu NK, Xu J, Yang SC, Li S, Liu YY, et al. A simple preprocedural score for risk of contrast-induced acute kidney injury after percutaneous coronary intervention. Catheter Cardiovasc Interv. 2014;83(1):E8-16.
569. Gao YM, Li D, Cheng H, Chen YP. Derivation and validation of a risk score for contrast-induced nephropathy after cardiac catheterization in Chinese patients. Clin Exp Nephrol. 2014;18(6):892-8.
570. Liu Y, Liu YH, Chen JY, Tan N, Zhou YL, Li HL, et al. A simple pre-procedural risk score for contrast-induced nephropathy among patients with chronic total occlusion undergoing percutaneous coronary intervention. Int J Cardiol. 2015;180:69-71.
571. Skelding KA, Best PJ, Bartholomew BA, Lennon RJ, O'Neill WW, Rihal CS. Validation of a predictive risk score for radiocontrast-induced nephropathy following percutaneous coronary intervention. J Invasive Cardiol. 2007;19(5):229-33.
572. Sgura FA, Bertelli L, Monopoli D, Leuzzi C, Guerri E, Sparta I, et al. Mehran contrast-induced nephropathy risk score predicts short- and long-term clinical outcomes in patients with ST-elevation-myocardial infarction. Circ Cardiovasc Interv. 2010;3(5):491-8.
573. Ando G, Morabito G, de Gregorio C, Trio O, Saporito F, Oreto G. Age, glomerular filtration rate, ejection fraction, and the AGEF score predict contrast-induced nephropathy in patients with acute myocardial infarction undergoing primary percutaneous coronary intervention. Catheter Cardiovasc Interv. 2013;82(6):878-85.
574. Aykan AC, Gul I, Gokdeniz T, Kalaycioglu E, Turan T, Boyaci F, et al. Is coronary artery disease complexity valuable in the prediction of contrast induced nephropathy besides Mehran risk score, in patients with ST elevation myocardial infarction treated with primary percutaneous coronary intervention? Heart Lung Circ. 2013;22(10):836-43.
575. Kul S, Uyarel H, Kucukdagli OT, Turfan M, Vatankulu MA, Tasal A, et al. Zwolle risk score predicts contrast-induced acute kidney injury in STEMI patients undergoing PCI. Herz. 2015;40(1):109-15.
576. Liu YH, Liu Y, Tan N, Chen JY, Chen J, Chen SH, et al. Predictive value of GRACE risk scores for contrast-induced acute kidney injury in patients with ST-segment elevation myocardial infarction before undergoing primary percutaneous coronary intervention. Int Urol Nephrol. 2014;46(2):417-26.
577. Narula A, Mehran R, Weisz G, Dangas GD, Yu J, Genereux P, et al. Contrast-induced acute kidney injury after primary percutaneous coronary intervention: results from the HORIZONS-AMI substudy. Eur Heart J. 2014;35(23):1533-40.

Appendices

Appendix 1 Introduction

A1.1 - Organ-cross-talk in AKI
A1.2 - Examples of interactions between predictors
A1.3 TRIPOD recommended bootstrap validation method.
A1.4 – Model updating methods.
A1.5 - The Ottawa Acceptability of Decision Rules Instrument.
Statistics in-depth: variable selection, model selection, sample size, internal validation techniques, Discrimination – AUROC limitations; statistics beyond the AUROC – illustrated by the addition of Biomarkers to a model; interactions between predictors

Organ cross-talk

Organ	Effects
Lung	increased vascular permeability, dysregulation of salt and water channels and cytokine induction.(53, 526)
Heart	Increasingly recognised, poorly understood.(527) Cardiac decompensation may affect the kidney via hemodynamic, humoral and immune-mediated pathways, including elevated TNF-α, IL-1, neutrophil trafficking & apoptosis, with reduced fractional shortening. AKI in turn may induce salt and water retention to increases preload whilst it is also speculated that AKI induces endothelial cell activation and cytokine secretion resulting in myocardial damage via neutrophil infiltration and myocyte apoptosis.(528) Though animal models in this field are limited, in one study of wild-type mice, renal ischaemia increased cytokine expression in the heart, caused apoptosis of myocytes, and impaired cardiac function.(529)
Liver	Altered intra-hepatic haemodynamics may decrease GFR in an intrinsically normal kidney.(530) Limited data on kidney-liver interactions, though similar processes that affect the heart and lung have been described.(531-533) Inflammation, apoptotic pathways oxidative stress and decreased antioxidants have all been implicated in experimental AKI.(532-534)
Brain	Raised KC, G-CSF, GFAP & microglia Raised vascular permeability
GI tract	Channel inducing factor Elevated Potassium excretion
Bone Marrow	Anaemia, Coagulation disorders, Immune dysfunction

Table – A1.1 Organ-cross-talk in AKI.

Treatment and severity of disease	Patients with 'mild' disease have little scope for benefit.
Age and risk factors	Older subjects are generally less affected by risk factors, being robust enough to survive with risk factors present.
Age and type of disease	Some diseases are incurable and have the same prognosis regardless of age. Others are treatable or have less effect on younger patients.
Measurement and state of subject during measurement	Cardiac function measured at rest using an echocardiogram may have less predictive value and thus a smaller slope vs. outcome than function measured during stress (eg cardiopulmonary exercise testing).
Calendar time and treatment	Some treatments evolve or their effectiveness improves with training of staff.
Quality and quantity of a symptom	

Appendix Table A1.2 - Examples of interactions between predictors

1. Develop model using full original sample (size n) & determine performance
2. Generate bootstrap sample, by sampling n individuals with replacement from original sample
3. Develop model using bootstrap sample
a. Determine apparent performance (bootstrap performance)
b. Determine performance of bootstrap model in original sample (test performance)
4. Calculate optimism as difference between the two
5. Repeat steps 2-4 >100 times
6. Average estimates of optimism in step 5 & subtract the value from the apparent performance obtained in step 1 to obtain an optimism-corrected estimate of performance.

Appendix Table A1.3 TRIPOD recommended bootstrap validation method.

Number	Updating method	Reason
1	Adjust intercept	Different incidence of the outcome
2	Adjust regression coefficients of predictors & intercept	Regression coefficients of original model overfitted
3	As 2 & extra adjustment regression coefficients for predictors with different strength in validation population	As in method 2, & strength (regression coefficient) of ≥1 predictors may be different in validation population
4	As 2 & stepwise selection of additional predictors	As in method 2, & ≥1 predictors were not included in the original derivation
5	Re-estimation of all co-efficients, using data of the validation population	Strength of all the predictors may be different in the validation population
6	Model 5 & stepwise selection of additional predictors	As in method 5, & ≥1 predictors were not included in the original derivation

Table A1.4 – Model updating methods.

Please indicate your level of agreement with each of the following statements about the rule.	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly agree	Moderately agree	Strongly agree	No opinion Don't know
The rule is easy to use.							
The rule is easy to remember.							
The rule is useful in my practice.							
The wording of the rule is clear and unambiguous.							
My colleagues support use of the rule.							
Patients benefit from use of the rule.							
Using the rule results in improved use of resources.							
Using the rule would increase the chance of lawsuits.							
The evidence supporting the rule is flawed.							
I'm already using another rule or similar strategy.							
The rule does not account for an important clinical cue.							
The environment I work in makes it difficult to use the rule							

Table A1.5 - The Ottawa Acceptability of Decision Rules Instrument from Brehaut et al.(314)

Statistics in-depth

Variable selection

False negative findings can result when a study is not designed with the statistical power to detect a factor truly predictive of differential treatment effect.(236) It has recently been suggested that meta-analyses based on individual participant data (IPD) from multiple trials are performed to address this shortcoming.(224) Dichotomisation of continuous factors, whilst potentially providing clarity for users, reduces power. Instead it is possible to employ statistical methods such as fractional polynomials to screen continuous factors and identify potential interactions with treatment.(535) Inappropriate subgroup analysis can give spurious evidence and because of the analysis of a large number of variables, appropriate correction for multiple statistical testing is required to reduce the risk of chance findings.(236) Analyses may be restricted to just individuals testing positive for a factor, however using this approach means that inferences should be restricted to the selected patients.(224) Lastly evidence of an interaction between a factor and treatment response should ideally be explained by plausible biological reasoning.

Model selection

In a full model approach all candidate variables are included and this may avoid overfitting and selection bias and provide correct standard errors and P values.(240) However, it is often impractical to include all candidates and the full model is not always easy to define. Stepwise selection methods are widely applied to identify a limited number of variables for inclusion. Backward elimination starts with all candidate variables with a nominal significance level, often 5%, chosen in advance. A 1% level usually results in a model with fewer variables whilst significance levels of 10% can result in inclusion of some unimportant variables, as can the full model approach. A sequence of tests is applied to determine removal of a given variable. Backward is preferred to forward selection, which involves building the model from the best candidate predictor, because all correlations between predictors are considered.(238) Stepwise selection methods have been criticised in the literature.(243) Firstly, selected variables may change when a relatively small number of patients is added or removed. Simulations have demonstrated stepwise methods have limited power to select important variables if the data set is small.(194) Conversely, there is a risk that one or more (almost) random covariable(s) are selected, because multiple comparisons are made.(260) Variance

of the coefficients is usually calculated as if the selection was predetermined but ignoring model uncertainty causes an underestimation of the variability of the estimated coefficients in the model.(536) Stepwise selection causes selection bias in the estimated regression coefficients.(194)

Overfitting refers to the phenomenon in which the coefficients of selected variables are biased to more extreme values. The selection bias that occurs means that a regression coefficient is overestimated, because the corresponding predictor is more likely to be significant if its estimated effect is larger rather than smaller.(240) Overfitting leads to worse prediction in independent data and is more likely to occur in small data sets or with weakly predictive variables. In contrast underfitting refers to the average underestimation when all estimated regression coefficients are considered.(537) Careful fitting is essential so that interactions, if present, represent biologic phenomena rather than general lack of model fit.(243) It is usual to include a few strong predictors and several weaker ones.(238) Choice of statistical model is sometimes based on previous distributional examinations, but it is frequently based on maximising how available information is used. Binary and ordinal logistic models are frequently used for discrete completely assessed outcomes(538), and the Cox proportional hazards(539) and parametric survival models are frequently used for censored time-to-event data.(243)

Sample size
Screening based on statistical significance and stepwise variable selection involve multiple comparisons leading to unreliable models.(260) To reduce the number of candidate predictors data reduction techniques such as variable clustering and factor analysis can be used until the number of variables to use as candidates in the regression analysis follows the rule of thumb.(243)

Internal validation techniques
With split-half cross-validation, the model is developed on one randomly drawn half and tested on the other and vice versa. The average is taken as an estimate of performance; the procedure is repeated 10 times, with all subjects having served to test the model. To improve stability of the cross-validation, the whole procedure can be repeated several times, taking new random sub-samples.(250)

The most efficient validation has been claimed to be achieved by computer-intensive re-sampling techniques such as the bootstrap.(540) Bootstrapping replicates the process of sample generation from an underlying population by drawing samples with replacement from the original data set, of the same size as the original data set. Models may be developed in bootstrap samples and tested in the original sample, or in those subjects not included in the bootstrap sample.(262) Steps in performing bootstrapping suggested by TRIPOD are summarised in table A1.3.

Discrimination – AUROC limitations

In clinical practice the AUROC can be well below the theoretical maximum of 1 even with perfect calibration.(263) For example in simulation studies Gail and colleagues suggested with the assumption of uniform distribution of risk, the maximum AUROC is 0.83.(282) In a perfectly calibrated model in a population with an average 10-year risk of 10% with relatively little spread, risk is centred around 10%, and the maximum AUROC is 0.63.(263)

Though it is commonly believed that sensitivity and specificity are properties of a test and are not subject to alteration by disease prevalence, as are the positive and negative predicted values, this has been shown to be false.(541, 542) The AUROC is not the probability that individuals are classified correctly or that a person with a high test score will become a case; it describes how well models rank order cases and non-cases, An example would be if a model assigning all cases a value of 0.53 and all non-cases a value of 0.52 would have perfect discrimination (1.0), although the probabilities assigned may be unhelpful. The actual predicted probabilities do matter, however, in prediction models to be used in clinical practice.

Discrimination is of most interest when classification into groups with or without prevalent disease is the goal, such as in diagnostic testing when comparing two imaging modalities eg in the diagnosis of pulmonary embolism. In prognostic modeling when predictors are assessed future disease status remains to be determined by stochastic process, and can only be estimated as a probability.(398) Measures of discrimination are nonetheless commonly emphasised, which ignores the random nature of the outcome.

In epidemiological studies common choices of effect measures include the odds ratio (OR) or hazard ratio. However, using these measures relating a predictor to an outcome, to have an impact on the curve for an individual measure or score needs to be sizeable, such as 16 per 2 standard deviation units.(543) This size effect is possible with a risk score, but is unlikely to be achievable for most individual measures. For example, individual components of the Framingham score, such as blood pressure and smoking, all have far smaller hazard ratios. It has been argued that for risk prediction, actual or absolute predicted risk, not captured by the AUROC, is of primary clinical interest. As one would expect, the AUROC curve is also insensitive to the impact of adding new predictors to a score or predictive model.(263, 269) For example the impact of a new predictor is lower when other strong predictors are in the model, even when it is uncorrelated with the other predictors. To illustrate this Wang et al, examined a risk score based on multiple plasma biomarkers, with estimates of mortality in the high risk (top 20%) and low risk (bottom 40% of risk scores) groups were 20% and 3%, respectively, indicating important differences in predicted risk. However, incorporation of these biomarkers (multivariate hazard ratio of 4) into a risk function led to little improvement in the c-statistic.(544) In a cohort generally considered low-risk, there may be a small proportion of high-risk individuals. Rank-based measures such as the AUROC do not take such a distribution into account. Differences between two individuals at low risk (e.g. 1.0% vs. 1.1%) have the same impact on the AUROC as two individuals at moderate vs high risk (e.g. 5% vs. 20%) if their differences in rank are the same. Yet clinically it may only be important to separate the latter two cases. Being based on ranks, makes the AUROC less sensitive than measures based on the likelihood or other global measures of model fit, potentially making it a poor choice for the selection of variables to be used in a predictive model.(240)

Statistics beyond the AUROC – illustrated by the addition of Biomarkers to a model

Despite the limitations of using the AUROC in prognostic prediction models, this continues to dominate the discourse when assessing how 'good' a model is. However, a number of other measures have been proposed that could provide clinicians with additional information when assessing potential clinical utility. Studies may be prompted by declining performance of an existing model over time, for example as diagnosis or treatments change, or when a new biomarker is discovered. Rather than developing new models from scratch, it is recommended to consider whether existing models could be improved by recalibration or incorporating a biomarker.(293) The growing interest in biomarkers has stimulated a move to provide standards and guidance in the field. For example in cardiovascular medicine, the American Heart Association(545) emphasises that biomarkers should not be evaluated in isolation for their predictive abilities but rather on their added predictive contribution beyond existing or established predictors requiring a multivariable approach in design, conduct, analyses and reporting.(431) This is particularly important in the 'omics' setting where large numbers of markers are studied, often with each marker tested separately for its association with the outcome. Such analysis runs the danger of false positive findings and does not guarantee a true predictive role of the biomarker beyond established predictors (546). Rigorous testing of benefit is important, for example, to justify cost or when the new test may be invasive and thus potentially associated with harm.

The AUROC can be insensitive to detecting improvements in model performance when a new marker is added to a model that already includes important predictors.(243, 263, 269, 547) Partly as a response, techniques to estimate the added value of predictors have been described. Clinical risk reclassification classifies predicted risk estimates into clinically

relevant categories and cross-classifies these categories.(263) The per cent reclassified can be used as an indication of the clinical impact of a new marker, and will likely vary according to the original risk category. It is obviously important to verify that these individuals are being reclassified correctly, by comparing predicted risks from the models to the crude proportion developing events within each cell, or the observed risk. Pencina et al described the Net reclassification improvement (NRI) as a measure of change in categories.(402) Separate reclassification tables for cases and controls are formed, then the proportions moving up or down categories among cases and controls are examined. The NRI is the difference in proportions moving up and down among cases vs controls.(402) NRI is usually used to quantify if adding a new biomarker to an existing model is beneficial, but can also be used for comparing two models.(402) The Integrated discrimination improvement (IDI) is the estimated improvement in average sensitivity of a model with addition of the new predictor minus the estimated decrease in mean specificity, summarised over all possible risk thresholds.(402). It estimates the magnitude of the probability improvements or worsening between two models over probability thresholds.(237) In contrast to the NRI the IDI does not require subjectively, predefined risk thresholds.

Traditional use of the AUROC or sensitivity/specificity to judge additional value of a biomarker may have limited clinical relevance. For example, how high should the AUROC be to justify clinical use? Or if a new test increased specificity by 10% but decreased sensitivity by 5% compared to a standard test which should be used? Vickers proposed decision curve analysis as a way to inform on clinical consequences, by determining the relationship between a chosen predicted probability threshold and the relative value of a false-positive and a false-negative result to obtain a value of net benefit of using the model at that particular threshold.(268) By explicitly considering clinical consequences of decisions and providing data about the value of a model, decision curves can be used to determine whether or not the model should be employed in practice. For example, if a patient would opt for biopsy if told that a risk of prostate cancer was ≥20%, but not if the risk was <20%, it can be shown that the patient considers harms associated with a missed cancer to be four times greater than the harms associated with an unnecessary biopsy. The ratio of harms is the odds at the probability threshold. This threshold probability can be used to determine both whether a patient is defined as test-positive or negative and to model the clinical consequences of true and false positives using a clinical net benefit function.(268) Vickers et al went on to describe how to use decision curves to correct for overfitting using 10-fold cross-validation, calculate confidence intervals, application to censored data including competing risk and calculation directly from predicted probabilities.

A similar technique to decision curves is the weighted accuracy metric that also uses threshold probability from a model to classify patients as positive or negative and assign a relative weight to the cost of false negatives vs false positives.(548) Pepe and colleagues described the predictiveness curve, to complement risk modeling by assessing the usefulness of a model when applied to the population, displaying essential information about risk not conveyed by the AUROC.(547) The authors proposed that the predictiveness and classification performance of a marker, displayed together in an integrated plot, provide a comprehensive, cohesive assessment of a risk marker or model.

A number of other performance measures may be reported in prediction model studies. For example, the Brier score which is the mean squared error between outcome and prediction originated from the weather forecasting literature,(549) Explained variation (Shapiro's R^2) is the geometric mean of the probability assigned to the event that occurred.(550-552) The associated approximate R^2 statistics (termed the 'sum-of-squares' R^2 and the 'entropy-based' R^2, respectively) are obtained by scaling each measure relative to the value achieved from a null model.(551) Model fit may be assessed by the Akaike Information Criterion (AIC), which penalises the log-likelihood of the model for the number of parameters included.(553) Likelihood-based measures, such as Bayes information criterion, by adjusting for the number of variables in the model, are more sensitive, global measures of model fit than the AUROC.(240) Lower values indicate better fit and a penalty is paid if the number of variables

is increased. If global fit is better for one model but general calibration and discrimination are similar, fit may be better among some individuals, for instance, those at higher risk.

To conclude, arguably of utmost importance for risk prediction is whether a model can accurately stratify individuals into clinically relevant higher or lower risk categories. For clinical use of a new model or expensive biomarker, it is important to know if a higher or lower estimated risk would change treatment decisions for the individual patient, compared to an existing model (or current practice). To assess potential for reclassification, risk can be estimated over a range of values of a biomarker to determine whether it may be important to measure in an individual. For those at low risk, additional information may be minimal, whereas for those at higher risk the impact on risk of disease may be substantial. The estimated risk or predicted values, and how well these predict actual risk, may therefore be a more important aspect of a prognostic model than sensitivity and specificity, on which the AUROC curve is based.

Interactions between predictors
Maximum information should be extracted from predictors and response; thus when there is a choice of a categorical or a continuous variable, the latter is preferred.(243) Expert opinion suggests to avoid turning continuous predictors into dichotomies, as more predictive information is retained.(259) For ordered categorical variables, such as stage of CKD, collapsing of categories may be required. Harrell et al, suggested types of interactions that have frequently been found to be important in predicting outcomes and can be pre-specified (table A1.4 Careful fitting of a model is essential so that interactions, if present, represent biological phenomena rather than general lack of model fit.

Appendix 2 Systematic review

A2.1 - Study inclusion criteria.

Inclusion Criteria • Articles in peer-reviewed journals reporting a prognostic multivariable prediction model (scoring system or algorithm) identifying patients who developed HA-AKI (or other measures of renal dysfunction in older studies) • Validation studies (and updating) of an existing model • Retrospective, prospective and case-control designs • Adults (≥18 years) in general hospital settings • Statistical measures of discrimination (AUROC or c-statistic)
Exclusion Criteria • Patients <18 years old • Cardiac surgery, other specialised surgery (e.g. transplantation), CI-AKI • Non-human studies • Case reports or conference abstracts • Only logistic regression without a prediction model • Lack of discrimination statistics (unless model validated elsewhere) • Studies that investigated a single predictor, test, or marker • Studies that investigated only causality between one or more predictors & an outcome • Use of patients already with the outcome (e.g. AKI present at hospital admission) • Patients in primary care • Novel, not widely available tests, such as biomarkers

AUROC - area under the receiver-operating characteristic curve, CI-AKI – Contrast-Induced AKI, HA-AKI - hospital-acquired-AKI.

A2.2 - Embase Search

LINE	SEARCH TERM
1	(acute AND kidney AND injury).ti,ab
2	AKI.ti,ab
3	(acute AND renal AND failure).ti,ab
4	ARF.ti,ab
5	(contrast AND induced AND nephropathy).ti,ab
6	ACUTE KIDNEY INJURY/
7	OR/1-6
8	predict*.ti,ab
9	PREDICTIVE VALUE OF TESTS/
10	scor*.ti,ab
11	observ*.ti,ab
12	OBSERVER VARIATION/
13	8 OR 9 OR 10 OR 11 OR 12
14	7 AND 13
15	(acute AND kidney AND injury).ti,ab
16	AKI.ti,ab
17	(acute AND renal AND failure).ti,ab
18	ARF.ti,ab
19	(contrast AND induced AND nephropathy).ti,ab
20	ACUTE KIDNEY FAILURE/
21	OR/15-20
22	predict*.ti,ab
23	exp METHODOLOGY/
24	validat*.ti,ab
25	OR/22-24
26	21 AND 25
27	14 AND 26
	Articles: 9102

A2.3 Ovid MEDLINE® search

Line	Search term
1	(acute AND kidney AND injury).ti,ab
2	AKI.ti,ab
3	(acute AND renal AND failure).ti,ab
4	ARF.ti,ab
5	(contrast AND induced AND nephropathy).ti,ab
6	ACUTE KIDNEY INJURY/
7	OR/1-6
8	predict*.ti,ab
9	PREDICTIVE VALUE OF TESTS/
10	scor*.ti,ab
11	observ*.ti,ab
12	OBSERVER VARIATION/
13	OR/8-12
14	7 AND 13
	Articles: 9646

A2.4 - Web of Science search

RESULTS	WEB OF SCIENCE SEARCH
8002	(TS=((acute kidney injury) OR (aki) OR (acute renal failure) OR (arf) OR (contrast induced nephropathy)) AND TS=(predict* OR scor* OR observ* OR validat*)) AND DOCUMENT TYPES: (Article) Refined by: WEB OF SCIENCE CATEGORIES: (UROLOGY NEPHROLOGY OR SURGERY OR CARDIAC CARDIOVASCULAR SYSTEMS OR TRANSPLANTATION OR CRITICAL CARE MEDICINE OR MEDICINE GENERAL INTERNAL OR MEDICAL INFORMATICS OR GASTROENTEROLOGY HEPATOLOGY OR ANESTHESIOLOGY) AND DOCUMENT TYPES: (ARTICLE) AND WEB OF SCIENCE CATEGORIES: (UROLOGY NEPHROLOGY OR SURGERY OR CARDIAC CARDIOVASCULAR SYSTEMS OR TRANSPLANTATION OR MEDICAL INFORMATICS OR CRITICAL CARE MEDICINE OR HEALTH CARE SCIENCES SERVICES OR MEDICINE GENERAL INTERNAL OR MEDICAL LABORATORY TECHNOLOGY OR GASTROENTEROLOGY HEPATOLOGY OR ANESTHESIOLOGY OR EMERGENCY MEDICINE) Indexes=SCI-EXPANDED Timespan=All years

A2.5 - CHARMS checklist and data extracted for systematic review.

Item	Explanation in the Review
1. Type of studies	Prognostic prediction models
2. Scope	Published prognostic prediction models for development of AKI in general hospital settings; to inform risk stratification & potential uses in decision-making in different patient groups
3. Type of studies	Model development +/- external validation in independent data; external model validation & model updating, if present
4. Target population	Adult (≥18) Patients in acute hospital environment
5. Outcome predicted	Development of AKI (or equivalent definition, including RRT) after an admission to hospital or Surgery
6. Time span of prediction	In-hospital development of the outcome
7. Intended moment of using the model	Pre-operatively to predict the risk of post-op AKI or need for RRT; at admission to risk stratify or guide therapy
Summary of Data extracted	
• Data source (years, retrospective, prospective; cohort, case-control, trial data) • Participants & setting (eg cardiac surgery, single or multi-centre, country • Primary outcome (and any blinding) • Candidate predictors (definitions; continuous data dichotomised? & how selected for modelling) • Sample size, EPV (including all predictors considered) • Type of model(s) evaluated - derivation, validation (internal, external) • Missing data, number included & excluded (criteria) • Type of model (eg full model approach), shrinkage • Incidence of outcome & mortality data • Predictors in the model(s) • Performance: discrimination (AUC or C-Statistic) & calibration (eg H-L P value, slope/curve), risk groups • Internal & external validation (in same study) • External validation studies with relevant performance measures • Additional resources, funding	

AKI – acute kidney injury, EPV – events per variable, H-L – Hosmer-Lemeshow goodness-of-fit test, RRT – renal replacement therapy.

A2.6(i) Surgery

Author, Type, TRIPOD detail	Population, Outcome, AKI Definitions, Methods	Outcomes, Predictors & Model Performance	External Validation
Kheterpal 2007 **General surgery** **TRIPOD 1A - Derivation** **(25/37 pts)**	USA single centre, retrospective cohort study (n=65,043). Data collected 2003-6. Mean Age with outcome 59, without outcome 47 (P<0.001). Male with outcome 56%, without outcome 52% (P=0.32). Inclusion: pre-op eGFR (Cockcroft-Gault) ≥80 ml/min; major surgery (≥2 days in-patient). Exclusions (n=49,941): pre-op eGFR <80 (n=5659). cardiac, transplant, urology & ECT, suprarenal aortic cross-clamping; pre-op AKI & IV contrast <7 days post-op, no pre-op SCr (n=6,534). Included: n=15,102. Outcome: reduction of eGFR to ≤50ml/min <7 days post-op. Predictors: 24 pre, 6 intra-op. Collinearity predictors evaluated; bivariate correlation matrix; remaining predictors entered into logistic regression full model fit. Missing data: excluded from full model. After exclusions n=14,066 included. Un-weighted model continuous predictors dichotomised.	Outcome (AKI) in 0.8% (n=121), 0.1% (n=14) required RRT. Propensity matched 30-day mortality with outcome 15% (n=17/118) vs. 2.7% (n=9/352) without. AKI associated with significant increase in 30-day, 60-day, 1-yr mortality. **7 pre-op predictors: age, emergent surgery, liver disease, BMI, high-risk surgery, PVD & COPD.** Weighted c-Statistic 0.77 (95% CIs 0.75-0.79). Un-weighted risk factor scale (cut-off Age >59, BMI ≥32) c-Statistic 0.73 (0.7-0.76). With intra-op: vasopressor dose, infusion & diuretic: AUC 0.79 (0.77-0.81) No calibration statistics.	Xing 2012 - AUC 0.66
Kheterpal 2009 **General surgery** **TRIPOD 2A - Derivation, Validation** **(28/37 pts)**	USA multi-centre (121) retrospective database study (n=152,244). 2005-6. Mean age with outcome 64.8 (±14.8), without 53.5 (±17.3) (P<0.001). Male with outcome 57%, without outcome 39% (P<0.001). Included n=75,952. Random split derivation 75% (n=57,080) & validation (25% n= 18,872). Exclusions (n=76,292): vascular, cardiac, urology, ophthalmology, obstetric, or urologic procedures; day case; pre-op AKI (rapidly increasing azotaemia & SCr ≥265 µmol/L <24h of surgery) or previous RRT (n=1637). Admission SCr taken as baseline, assessed as predictor & included in the model. 'Mild' pre-op renal insufficiency defined SCr 106-168 µmol/L; 'moderate' >177 µmol/L. Outcome: AKI defined as increase SCr ≥177 µmol/L (from pre-op value) or RRT <30 days. Missing data: SPSS assessed impact of imputation. Continuous predictors dichotomised. Collinearity & Pearson correlations evaluated for all 19 preoperative predictors (comorbidities, drugs, type of surgery). Remaining predictors entered into full model fit logistic regression.	Outcome in 1% (n=762/75,952) – n=561 derivation, n=201 in validation sets. Mortality 42% (n=320) in those with outcome vs 8% in a propensity matched group without outcome. **9 predictors (simplified risk index): age ≥56 yr, male, emergency, intraperitoneal surgery, diabetes, CCF, ascites, HTN, mild or moderate pre-op renal insufficiency.** c-Statistic 0.80 (0.79-0.81) in derivation & internal validation cohorts. Calibration: Risk classes reported for derivation vs validation sets.	-

A2.6(ii) Trauma & Orthopaedics			
Author, Type, TRIPOD detail	**Population, Outcome, AKI Definitions, Methods**	**Outcomes, Predictors & Model Performance**	**External Validation**
Bell 2015 **T&O** **TRIPOD 3,4 – Derivation, Internal & EV** **(34/37 pts)**	UK multi-centre (3) retrospective cohort study linking multiple prospectively collected databases (n=15,218). 2005-11. Overall mean age 70.7 (±15.3), with outcome 76.5 (±11.1) without outcome 70.0 (±15.6). Overall Male 37%, with outcome 47%, without 36%. Included: derivation n=6,220 (2 sites) & validation n=4,395 (1 site). Exclusions: missing SCr (n=2,688), RRT, 2nd operation (n=1,915). Outcome: KDIOGO SCr changes <7 days. CKD defined using eGFR from CKD-EPI. Admission SCr taken as baseline if elective admission. Entered 11 candidate predictors (age, sex, CKD (baseline eGFR), diabetes, number drugs, ACEi/ARB, NSAID/COX-2, statin, urgency, ASA grade & deprivation category into Backward/forward multivariable selection. Applied a conservative selection criterion of P<0.15 to limit over-fitting risk. Bootstrapping for IV. To assess robustness sensitivity analyses performed: multiple imputation relaxing & restricting the backward selection removal criterion & adding non-linear & interaction terms. Categorised eGFR.	Outcome (AKI) in 10.8% (n=672) derivation & 6.7% (n=295) validation sets. With AKI adjusted hazard ratio 1.53 (95% CI 1.38-1.70). **7 predictors: age, male, diabetes, number drugs, CKD (eGFR), ACEi/ARBs & ASA.** Risk calculator supplied. Derivation AUC 0.74 (0.72-0.76), Internal validation 0.73. EV 0.70. Risk groups shown. Calibration plot. Calibration suboptimal in validation cohort (over-predicted risk). Re-calibration: correction factor, added to intercept; intercept and regression coefficient index as the only predictor used to transform prognostic index & compute recalibrated probabilities.	Same Study different site AUC 0.70

A2.6(iii) General admissions			
Author, Type, TRIPOD detail	**Population, Outcome, AKI Definitions, Methods**	**Outcomes, Predictors & Model Performance**	**External Validation**
Drawz 2008 **Medicine, surgery, obstetrics** **TRIPOD 2A – Derivation, Internal Validation (26/37 pts)**	USA multi-centre (3), retrospective, case-controlled study (n=180 cases, n=360 controls). 2003. Mean age with outcome 67, controls 63 (P=0.01). Males with outcome 74%, controls 80% (P=0.18). Hospital-acquired AKI (HA-AKI) defined: increase SCr ≥44µmol/L if baseline SCr ≤168µmol/L), ≥88µmol/L baseline 177-433µmol/L & ≥133µmol/L baseline >442µmol/L. Admission SCr presumed to be baseline. Random split derivation (2/3) & internal validation). 'Control' cases – mix of same discharge diagnosis or next patient admitted to clinical team. Inclusions: age ≥18 & normal admission SCr or admission SCr not qualifying as AKI vs known baseline. Exclusions: RRT, no repeat SCr performed. 19 predictors assessed: demographics (age, sex, race), medical history, medications & admission observations & blood parameters (BP, HR, HCO3, urea, SCr, & albumin). Predictors with p value <0.20 univariate analysis entered into multiple logistic regression model. Final model chosen by maximising likelihood ratio, c-statistic & R2 while minimising AIC. Also produced a simplified model, created by categorising continuous variables into quartiles. Cases with missing data excluded. Multiple imputation also performed.	No information on mortality. **7 predictors: age, SBP, HR, HCO_3, urea, albumin & drugs (NSAIDs, ACE-I, ARBs or diuretic).** Derivation c-statistic 0.73. Simplified: HR ≥70/min, HCO_3 (<24 or >30mmol/L), SCr ≥88µmol/L & drugs. Internal validation 0.66. Simplified model HR ≥70, HCO_3 (<24 or >30mmol/L), SCr ≥88 µmol/L & NSAIDs/ACEi/ARBs/Diuretics - C-statistic derivation 0.69, internal validation 0.66. No H-L p-value. Risk range in validation set plotted: 0/1 risk factor = 16% risk HA-AKI, vs 4 risk factors = 62%.	-
Matheny 2010 **General admissions** **TRIPOD 1B – Derivation** **(28/37 pts)**	USA single centre, retrospective cohort study (n=61,179). 1999-2003. No data on mean age (25.6% age >65). Overall males 44.4%. Inclusions: adult admissions ≥2 days (n=26,107). Exclusions: missing data, those with a baseline eGFR <60 (n=11,342), AKI on admission, no SCr available within 48 hrs of admission (n=10,378) or no repeat SCr (n=13,352). Outcome (<30 days post admission): AKI Risk = ≥2 SCr results ≥150% of baseline. AKI Injury = ≥200% baseline. eGFR using MDRD equation. 27 predictors assessed: coded diagnoses (including admission diagnosis), blood parameters (including admission SCr) & drugs following univariate analysis placed in multivariable model. Missing values captured as a separate category. 10-fold cross-validation employed to estimate overfitting.	AKI Risk 5.2% (n=1,352), AKI Injury 2.8% (n=726). No mortality data. **27 predictors: Female, Age, Race, 11 classes of drugs, Contrast, bacterial infection (use of antibiotics), admission SCr, MI, rhabdomyolysis, hepatitis, pancreatitis, ammonia, AST/ALT ratio, thrombocytopenia, leucocytosis, hypercalcaemia, glucose.** AKI Risk: AUC 0.75 (0.73–0.76). H-L P = 0.29. AKI Injury: AUC 0.78 (0.76–0.79), H-L P=0.12. Calibration plotted by deciles.	-

A2.6(iii) General admissions			
Author, Type, TRIPOD detail	**Population, Outcome, AKI Definitions, Methods**	**Outcomes, Predictors & Model Performance**	**External Validation**
Forni 2013 **General medical TRIPOD 2B, 4 – Derivation, Internal & EV** **(29/37 pts)**	UK single centre. Prospective cohort study (n=3,707). 2012. Median (IQR) age with outcome 80 (70-86), without outcome 73 (61-81) (P<0.001). Males with outcome 51% without outcome 49% (P=0.834). Inclusion: medical patients staying >1 night in hospital (n=1,867). Exclusions: RRT, non-medical patients, age <18, AKI on admission (n=184), missing data (n=553). Included n=3,523. Derivation n=1,867. Outcome: AKI (KDIGO SCr change <7 days). CKD defined – eGFR <60 on Pre-admission SCr measured >1 month & <6 months. Internal validation: patients with no previous SCr result, but with a SCr on admission within normal range (defined 80-120µmol/L) (n=1,656). 25 predictors on univariate, If P <0.05 variable entered into multivariable analysis. No missing data information.	Derivation group developed AKI 7% (n=95) – mortality 20% vs 3.5% (n=62) without outcome. In validation cohort n=60 developed AKI. **7 predictors: Age 60-79 (1 point) ≥80 (3 pts), CCF, CKD, Diabetes (2 pts), Liver disease (3 pts), respiratory rate ≥20/min, <alert on AVPU score (3 pts).** Derivation AUC 0.72 (0.66–0.77). H-L P=0.96. Risks plotted. Validation AUC 0.76 (0.71–0.82). No H-L reported.	Hodgson 2017, AUC 0.65-0.71
Bedford 2016 **General admissions TRIPOD 2A, 3 – Derivation, EV** **(29/37 pts)**	UK multi-centre (3), 2011. Retrospective cohort study (n=11,655). Average age and sex not given. Included: derivation n=7,556 admissions & internal validation n=2,514. Exclusions: non-emergency, pre-admission AKI, AKI at admission, obstetrics, patients with no info on AKI at 72 hours. Outcomes: AKI & AKI Stage 2/3. AKI <72 hours, using KDIGO change in SCr. Ordinal logistic regression with univariable analysis for development of multivariable analysis. 45 Predictors included demographics, bloods, prior admissions, co-morbidity. Backwards selection used for retention of statistically significant predictors. Missing data excluded or given own category. 3:1 random split for internal validation. External validation n=1,585, single centre.	Derivation AKI 9.6% (n=241), AKI 2/3: n=40. No mortality data. EV AKI 7.6% (n=120), AKI 2/3 n=12. **12 predictors: age, primary diagnosis, previous hospital admissions, Charlson co-morbidity index score, HbA1C, troponin, proteinuria, baseline eGFR, K^+, WCC, Mg^{2+}, CRP.** IV AUC 0.67 (0.64-0.71) any AKI, 0.68 for AKI 2/3. No derivation AUC H-L P=0.04 any AKI model, P=0.005 for AKI 2/3.	Same study AUC 0.71 (0.63 AKI 2/3). H-L P=0.12 AKI, P=0.14 for AKI 2/3.
Koyner 2016 General admissions TRIPOD 2A Derivation, Internal Validation (24/37 pts)	USA multi-centre (5) Retrospective cohort study (n=269,999). 2008-2013. Mean age with outcome 70 (±16), without outcome 63 (±19) (P<0.001). Males with outcome 49%, without outcome 43% (P<0.001). Included: n=202,961. Exclusions: SCr >354 µmol/L on admission (n=11,305), those without SCr measurement (n=52,508) & AKI prior to arrival on ward (n=3,225). Admission SCr defined as baseline, assessed as predictor & included in model. Outcome: rise SCr as per KDIGO but within 24hrs period. Model included 29 predictors. Continuous predictors modelled using restricted cubic splines with knot placement. Variable importance plot created. Laboratory values & vital signs updated periodically therefore separated into time intervals & logistic regression used for model estimation. Values closest to beginning of that time variable used to predict outcome	AKI 8.6% (n=17,541). Mortality with outcome 6% (n=1031) vs 1% (n=1,419) without **29 predictors: SCr, Urea, HR, anion gap, Urea/SCr, RR, glucose, WCC, K^+, Oxygen Saturations, age, HCO_3, Na^+, temperature, prior ICU, albumin, bilirubin, Ca^{2+}, platelets, time, SBP/ DBP, pulse pressure, sex, AVPU, Alkaline phosphatase, Hb, total protein, AST.** Dsicrimination reported for validation cohort only: AKI AUC 0.74 (0.74-0.74), AKI Stage 3 AUC 0.83 (0.83-0.84) Model including only SCr, BUN & their ratio AUC 0.69 (0.68-0.69).	-

for that interval, if no values available during an interval, most recent value used, if no previous value available, median value across entire cohort imputed. Split derivation (60%) & internal validation (40%) by time.

A2.6(iv)
Heart failure

Author, Type, TRIPOD detail	Population, Outcome, AKI Definitions, Methods	Outcomes, Predictors & Model Performance	External Validation
Forman 2004 **TRIPOD 1B - Derivation, Internal Validation** **(26/37 pts)**	USA multi-centre (11) retrospective cohort study (n=1,009). 1997-8. Overall mean age 67 (±15), with outcome 68.7, without outcome 66.8 (P=0.07). Overall males 51.2%, 52% with outcome 50.9% without outcome. Exclusion (number not given): elective, <2 days, severe aortic stenosis, anticipated transplant, RRT, LVAD, high output failure, age <20, chemotherapy. Excluded n=5 with missing charts. Included n=1004. Outcome: worsening renal function (WRF) - rise SCr >26.5µmol/l during admission. 29 predictors assessed: demographics, history, drugs, symptoms, signs. Unclear method for excluding patients who had AKI at admission. Used admission SCr as baseline & as a predictor. Multivariable Cox regression models, stepwise selection. Bootstrapping for IV. Missing data: predictors missing >15% excluded; categorical data assumed "not present" & separate dummy indicator used if >5% of values missing.	'WRF' 27% (271/1,004). Mortality: risk ratio 7.5 with outcome (number not reported). **4 predictors: CCF, diabetes & BP >160 mmHg (1 point), SCr 132.6-212µmol/l (2 points) & SCr >221µmol/l (3 points).** Risk 'WRF': 0 pts = 10%, 1 = 19%, 2 = 20%, 3 = 30%, 4+ = 53%. 22% of total sample with risk score ≥4 had 53% likelihood WRF vs 10% risk among 12% with risk score 0 points (p<0.001). **No AUC or Calibration statistics.**	Breidthardt 2011 - 0.65 Wang 2013 - 0.65
Breidthardt 2011 **TRIPOD 1A – Derivation, (EV Forman)** **(23/37 pts)**	Swiss multi-centre (3) prospective analysis, with derivation (Basel score) & external validation of Forman score (n=767). 2001-2, 2006-2010. Overall median age 79 (71-85), with outcome 79 (72-85), without outcome 79 (70-85) (P=0.36). Overall males 55%, with outcome 61%, without outcome 54% (P=0.08). Included n=657. Exclusions (n=110): stay <2 days, incomplete SCr. Outcome: WRF = in-hospital increase SCr ≥26.5µmol/L. CKD from eGFR (using MDRD equation) <60 for >3/12 pre-admission. eGFR at admission to hospital included as a predictor. Unclear method for excluding patients who had AKI at admission. 48 predictors assessed on univariate & those with P value <0.05 entered into multivariable analysis. No missing data information. n=223 had blood gas analysis.	Outcome 21% (136/657). In-hospital mortality with outcome 17% (n=23) vs 6% (n=33) without (P <0.01). **3 predictors (n=223): HCO_3 <21 mmol/L, Diuretics, CKD** - AUC 0.71 (0.63-0.79). A computer-based, complex, exponential risk model AUC 0.75 (0.67-0.82). No H-L calibration data. Scores & percentage developing outcome: 0 - 1%, 1 –35%, 2 –27%, 3 – 35%.	-

A2.6(iv)
Heart failure

Author, Type, TRIPOD detail	Population, Outcome, AKI Definitions, Methods	Outcomes, Predictors & Model Performance	External Validation
Wang 2013 **TRIPOD 2A – Derivation, External validation (Forman)** **(30/37 pts)**	China, single centre, retrospective cohort study (n=1,709). 2004-11. Median age with outcome 73 (67-78), without outcome 71 (63-75) (P<0.001). Males 56.6% with outcome, 55% without outcome (P=0.13).	Overall AKI 32% (n=550). Mortality 16.5% (n=91) vs.1.9% (n=22) without AKI (P <0.01). Stay with AKI 14 vs. 11 days without (P <0.01).	
	Inclusion: CCF admission diagnosed by 2 cardiologists using European Society of Cardiology guidelines.	**8 predictors: Age ≥70; ≥3 CCF admissions, systolic BP <90mmHg, Na^+ <130mmol/L, NYHA IV, proteinuria, SCr ≥104 µmol/L & furosemide dose ≥80 mg/day.**	
	Exclusions: age <18, stay <2 days, missing data, hospital transfer, use LVAD, ESRD or RRT & septic or haemorrhagic shock; cardiac op, pacemaker or cardioversion & contrast.	Derivation AUC 0.76 (0.73–0.79) H-L P=0.98. Calibration plots by deciles. Validation 0.76 (0.72-0.8), H-L P=0.13.	-
	Split derivation (60%, n=1010) & validation (40%, n=699).	≥8 points high risk - 55.1% incidence vs. 18% if <8 points No calibration slope.	
	Outcome: AKI (AKIN): increase SCr ≥26.4 µmol/L or ≥50% in <48 hrs. eGFR – MDRD – unclear whether admission SCr was used to estimate baseline eGFR or how patients with AKI at admission were excluded. Admission SCr used as a predictor.	Forman - 0.65 (0.62–0.69). vs Forman score, improvement of 0.11 AUC, (P <0.001(DeLong)(9)	
	35 predictors – those with P value <0.1 on unviariate analysis placed in multivariate analysis (n=932).		

A2.7 – Abbreviations used in **A2.6(i-iv)**

ACEi – Angiotensin-converting enzyme inhibitors
AKI – Acute kidney injury
AKIN – Acute kidney injury network
ALT – Alanine aminotransferase
ARB – Angiotensin receptor blockers
ASA – American Society of Anesthesiologists Physical status grading used in pre-operative assessment
AST – Aspartate transaminase
AVPU – scale of consciousness best response: **A**lert, responds to **V**oice, **P**ain, **U**nresponsive.
AUC/AUROC – Area under the receiver operating characteristic curve
BMI – Body mass index
BP – Blood pressure
CA-AKI – Community-acquired AKI
Ca^{2+} - Serum Calcium
CI-AKI – Iodinated contrast AKI
CKD – Chronic kidney disease
COPD - Chronic obstructive pulmonary disease
CCF – Congestive cardiac failure
CKD-EPI – CKD Epidemiology collaborative equation
COX – Cyclo-oxygenase
CRP – C-reactive protein
D – Derivation study
DBP – Diastolic Blood Pressure
eGFR – estimated glomerular filtration rate
ESRD – end-stage renal disease
EV – External validation study
HA-AKI – Hospital-acquired AKI
Hb - Haemoglobin
HbA1C – glycated haemoglobin (A1c) Marker of long-term glucose control
HCO_3 – serum Sodium Bicarbonate
H-L – Hosmer-Lemeshow goodness-of-fit test (Calibration statistic)
HR – Heart rate (beats per minute)
HTN – Hypertension
ICU – Intensive Care Unit
IHD – Ischaemic heart disease
IV – Internal Validation study
K^+ - Serum Potassium
KDIGO – Kidney disease improving global outcomes (Stage 1-3 AKI defined by magnitude of SCr rise or fall in urine output)
LOS – length of stay
LVAD – Left ventricular assist device
LVEF – Left ventricular ejection fraction
MAP – Mean arterial pressure
MDRD – Modification of diet in renal disease equation
Mg^{2+} - serum Magnesium
MI – Myocardial infarction
Na^+ – Serum sodium
NSAID – Non-steroidal anti-inflammatory agent

NYHA – New York Heart Association Classification for heart failure (I-IV)

PVD – Peripheral vascular disease

RIFLE – Risk, Injury, failure, loss of kidney function

RR – respiratory rate (breaths per minute)

RRT – Renal replacement therapy

SCr – serum creatinine

Systolic Blood Pressure

TRIPOD – Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis. Checklist in derivation (37 points – 1 point for each recommended item reported).

TRIPOD Study types

Type 1a: Development only

Type 1b: Development and validation using resampling

Type 2a: Random split-sample development and validation,

Type 2b: Non-random split-sample development and validation

Type 3: Development and validation using separate data

Type 4: Validation only.

WCC – White cell count

WRF – ‘worsening renal failure’ (defined by individual study)

A2.8 - TRIPOD items reported in 11 general model studies. Red = item in <50% of models.

Title & Abstract		TRIPOD Item description	Reported ?
Title	1	Identify study as developing &/or validating a multivariable prediction model, target population & outcome.	10
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results & conclusions.	9
Introduction			
Background & objectives	3a	Explain medical context (including whether diagnostic or prognostic) & rationale for developing or validating the multivariable prediction model, including references to existing models.	11
	3b	Specify objectives, including whether the study describes development or validation of the model or both.	11
Methods			
Source of data	4a	Describe study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development & validation data sets, if applicable.	11
	4b	Specify key study dates, including start of accrual; end of accrual; & if applicable, end of follow-up.	11
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number & location of centres.	11
	5b	Describe eligibility criteria for participants.	11
	5c	Give details of treatments received, if relevant.	9
Outcome	6a	Clearly define outcome predicted by the prediction model, including how & when assessed.	11
	6b	Report any actions to blind assessment of the outcome to be predicted.	0
Predictors	7a	Clearly define all predictors used in developing or validating the model, including how & when they were measured.	11
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	0
Sample size	8	Explain how the study size was arrived at.	2
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	7
	10a	Describe how predictors were handled in the analyses.	11
Statistical analysis methods	10b	Specify type of model, model-building procedures (including predictor selection) & method for internal validation.	11
	10c	For validation, describe how the predictions were calculated.	6
	10d	Specify all measures used to assess model performance & if relevant, to compare multiple models.	8
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	1
Risk groups	11	Provide details on how risk groups were created, if done.	10
Development vs. validation	12	For validation, identify differences from development data in setting, eligibility criteria, outcome & predictors.	4
Results			
Participants	13a	Describe flow of participants through the study, including number of participants with & without the outcome & if applicable, a summary of the follow-up time. A diagram may be helpful.	11
	13b	Describe characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors & outcome.	9
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	3
Model development	14a	Specify the number of participants & outcome events in each analysis.	1 1
	14b	If done, report the unadjusted association between each candidate predictor & outcome.	1 1
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients & model intercept or baseline survival at a given time point).	8
	15b	Explain how to the use the prediction model.	1 1
Model performance	16	Report performance measures (with CIs) for the prediction model.	8
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	1
Discussion			
Limitations	18	Discuss any limitations of the study (non-representative sample, few events per predictor, missing data).	11
Interpretation	19a	For validation, discuss results with reference to performance in development data & any other validation data.	8
	19b	Give overall interpretation of results, considering objectives, limitations, results from similar studies & other relevant evidence.	11
Implications	20	Discuss potential clinical use of the model & implications for future research.	11
Other information			
Suppl info	21	Provide information about availability of supplementary resources, (study protocol, Web calculator, & data sets).	3
Funding	22	Give the source of funding & role of the funders for the present study.	9

A2.9 – Most common predictors included in the 11 general models

Field	Total	General Surgery		T&O	General					Heart Failure		
Study		Kheterpal 2007	Kheterpal 2009	Bell 2015	Drawz 2008	Matheny 2010	Koyner 2016	Bedford 2016	Forni 2013	Forman 2004	Breidthardt 2011	Wang 2013
Demographics												
Age	**9**	X	x	x	x	x	x	x	x			x
Male/gender	**3**		x	x			x					
Past history												
Diabetes	**5**		x	x				x	x	x		
CKD	**4**			x				x	x		x	
Heart failure	**4**		x						x	x		x
Liver disease	**3**	X				x			x			
Drugs												
Diuretics	**4**				x	x					x	x
ACEi/ARBs	**3**			x	x	x						
Observations												
Hypotension/ Shock	**3**				x		x					x
Bloods												
SCr	**5**		x			x	x			x		x
Bicarbonate	**4**				x	x	x				x	
éWCC	**3**					x	x	x				

ACEi – angiotensin-converting enzyme inhibitor drugs, ARB – Angiotensin 2 receptor blocker drugs, CKD = chronic kidney disease, Bloods – laboratory parameters, SCr – serum creatinine, T&O – Trauma and Orthopaedics, WCC – white cell count.

A2.10 – All predictors included in the 11 general models

Field	General Surgery		T&O	General					Heart Failure			Total
Study	**Kheterpal 2007**	**Kheterpal 2009**	**Bell 2015**	**Drawz 2008**	**Matheny 2010**	**Koyner 2016**	**Bedford 2016**	**Forni 2013**	**Forman 2004**	**Breidthardt 2011**	**Wang 2013**	
Demographics												
Age	x	x	x	x	x	x	x	x			x	9
Male/gender		x	x			x						3
BMI	x											1
Race					x							1
Past history												
Diabetes		x	x				x	x	x			5
CKD			x				x	x		x		4
Heart failure		x						x	x		x	4
Liver disease	x				x			x				3
Hypertension		x										1
PVD	x											1
Ascites		x										1
COPD	x											1
Previous admissions						x (ICU)	x					2
Charlson co-morbidity index							x					1
ASA Grade			x									1

Field	General Surgery		T&O	General					Heart Failure			Total
Study	Kheterpal 2007	Kheterpal 2009	Bell 2015	Drawz 2008	Matheny 2010	Koyner 2016	Bedford 2016	Forni 2013	Forman 2004	Breidthardt 2011	Wang 2013	
Drugs												
Diuretics				x	x					x	x	**4**
ACEi/ARBs			x	x	x							**3**
NSAIDs				x	x							**2**
Contrast					x							**1**
Number of drugs			x									**1**
Other drugs[1]					x							**1**
Observations												
Hypotension / Shock				x		x					x	**3**
Pulse pressure						x						**1**
Hypertension				x					x			**2**
Heart rate[2]				x		x						**2**
Temperature						x						**1**
Respiratory rate[3]						x		x				**2**
O2 saturations						x						**1**
Consciousness[4]						x		x				**2**
Surgery, Other												
Type of surgery	x	x										**2**
Emergency	x	x										**2**
Time						x						**1**

Field	General Surgery		T&O	General					Heart Failure			Total
Study	Kheterpal 2007	Kheterpal 2009	Bell 2015	Drawz 2008	Matheny 2010	Koyner 2016	Bedford 2016	Forni 2013	Forman 2004	Breidthardt 2011	Wang 2013	
Labs, Diagnosis												
Primary diagnosis							x					**1**
SCr		x			x	x			x		x	**5**
Haemoglobin						x						**1**
êPlatelets[5]					x	x						**2**
CRP							x					**1**
↑WCC					x	x	x					**3**
Bacterial infection[6]					x							**1**
MI[7]					x		x					**2**
Rhabdomyolysis[8]					x							**2**
Hepatitis/AST[9]					x	x						**2**
Alk Phosphatase						x						**1**
Bilirubin						x						**1**
Pancreatitis[10]					x							**1**
Bicarbonate				x	x	x				x		**4**
Anion gap						x						**1**
BUN/Cr						x						**1**
Urea				x		x						**2**
↑Ca^{2+}/Ca^{2+}[11]					x	x						**2**
églucose					x	x						**2**
Magnesium							x					**1**
Potassium						x	x					**2**
↓Na^{+}/Na^{+}[12]						x					x	**2**
Albumin				x		x						**2**
Total protein						x						**1**
Proteinuria							x				x	**2**

ACEi – angiotensin-converting enzyme inhibitor drugs, ARB – Angiotensin 2 receptor blocker drugs, ASA grade – American Anesthesiology Association grade, CKD = chronic kidney disease, COPD – Chronic obstructive pulmonary disease, NSAIDs – non-steroidal anti-inflammatory drugs, PVD – peripheral vascular disease. [1]Other drugs: Aminoglycoside, Amphotericin B, Cyclosporine, Acyclovir, Cisplatin, [2]= ≥70/min, [3]= Respiratory rate ≥20/min, [4]= not alert on AVPU scale (best response: Alert, to Voice, Pain, Unresponsive), [5] =platelet count <75% lower limit of normal, [6] =acute use of antibiotics, [7] =elevated CK-MB or Troponin-I or T, [8] = increase CK x5 in absence of myocardial infarction, [9] =peak AST or ALT >400IU/L, [10] =>x3 lipase normal range, [11] Serum Ca^{2+} =>upper limit normal, Na – serum Sodium, [12] =serum Sodium <130mmol/L, *SCr. T&O – Trauma and Orthopaedics. NB Koyner et al did not specify ranges for the predictors included.

A2.11 – Handling of Serum Creatinine and Chronic Kidney Disease in the general models

Author, year (n=derivation cases)	Kheterpal 2007 (n=14,066)	Kheterpal 2009 (n=57,080)	Bell 2015 (n=6,220)	Drawz 2008 (n=360)	Matheny 2010 (n=26,107)	Koyner 2016 (n=202,961)	Bedford 2016 (n=7,556)	Forni 2013 (n=1,867)	Forman 2004 (n=1,004)	Breidthardt 2011 (n=657)	Wang 2013 (n=1,010)
CKD defined			X				X	X	X	X	X
Admission SCr used as baseline		X		X		X			X		X
Omitted cases with reduced GFR	X				X						
Admission SCr assessed as predictor		X		X	X	X			X	X	X
CKD included in model			X				X	X		X	
Admission SCr included in model		X			X	X			X		X

CKD - Chronic Kidney Disease, SCr - Serum Creatinine. Red shading indicates concern over handling of SCr in the study.

A2.12(i) - Cardiac Surgery Studies

Author & setting. Derivation (D) Internal validation (V), External validation (EV) TRIPOD study type & checklist fulfilled	Population, Outcome, AKI Definitions, Methods	Outcome Incidence & Mortality	Risk factors in model & Performance	Externally validated? (AUC)
(148) Chertow *1997* **Cardiac Surgery** **D,V** **TRIPOD 2B** **(26/37 pts)**	USA multi-centre (43), retrospective evaluation of prospectively collected database, 99% males (n=43,642). Exclusions: pre-op serum creatinine (SCr) ≥265µmol/L or active endocarditis (n=537). Excluded all subjects with missing data (number not given). Included n=42,773. April 1987 – March 1994 Prospectively validated on independent sample (n=3,795). April 1994 – December 1994. Outcome: renal replacement therapy (RRT) <30 days (Indication not specified). eGFR: Cockcroft-Gault method. Variables: 35 pre-op & 35 intra-op. Those with P<0.05 on univariate analysis entered into forward & backward multivariable analyses. Bootstrapping (100) employed.	RRT in 1.1% (n=471) in derivation. Mortality in those requiring RRT 63.7% (n=300), vs. 4.3% (n=1819) if not.	10 variables: valvular surgery, eGFR <60, IABP, previous cardiac surgery, NYHA-IV, PVD, LVEF <35%, Pulmonary rales, COPD, Hypotension. AUC 0.76. No H-L P-value. Risk algorithm plotted in table derivation vs validation sets. Internal validation AUC unreported. Used classification tree based on recursive portioning & risk categories.	Fortescue 2000 - 0.72 Eriksen 2003 - 0.71 Thakar 2003 – RRT 0.78 & composite 0.65 Wijeysundera 2007 –0.68-0.70
(150) Thakar (Cleveland) 2005 **Cardiac Surgery** **D,V** **TRIPOD 2A** **(30/37 pts)**	USA single centre, retrospective evaluation of prospectively collected database (n=33,217). 1993-2002. Exclusion (n=1,540): pre-op RRT, heart transplant recipients, ECMO, pre-op tracheostomy or ventilation, implantable cardiac-defibrillator, LVADs or sternal work; missing data. Included n=31,677. Random split derivation (n=15,838) & internal validation (n=15,839). Outcome: RRT. Indications for RRT: uremia, overload, or biochemical abnormalities (clinical judgment). 15 variables assessed. From univariate analysis variable was a candidate if P=<0.10 or near 0.10 & of particular clinical interest. Continuous variables dichotomized. 12 variables forced into model & with backward selection. Bootstrapping (1000) used. Assigned score points to each risk factor using model parameter estimates, multiplied by 2 & rounded to the nearest integer.	RRT in 1.7% (n=539/31,677). No mortality data reported.	13 Variables: Female, CCF, LVEF <35%, Pre-op IABP (2pts), COPD, diabetes, previous surgery, Emergency (2pts), Valve surgery only, CABG & valve (2pts), Other cardiac surgeries (2pts), Pre-op SCr 106-186µmol/L (2pts) or ≥186 µmol/L (5pts). AUC derivation 0.81 (0.78-0.83) & internal validation 0.82 (0.80-0.85). No H-L p-value. Risk categories plotted derivation vs validation.	Englberger 2010 – 0.86 (0.84-0.88) RRT & 0.81 (0.79-0.83) severe AKI. Candela-Toha 2008 – 0.86 (0.81-0.9) Demirjian 2012 0.84 (0.83-0.86) Berg 2013 0.88 (0.84-0.93), Ng 2014 0.75. Kim 2013 0.63-0.65. Bernie 2014 0.7 (0.69-0.72) AKI ; Stage 3 AKI 0.78 (0.75-0.81) Wijeysundera 2007 – 0.8-0.81

				Di Bella 2007 – 0.82 Heise et al – 0.66 (0.65- 0.68) Vives 2011 0.82 (0.79-0.85) Wong 2015 0.61-0.78
(349) Mehta 2006 **Cardiac Surgery** **D,V** **TRIPOD 2B** **(31/37 pts)**	USA multi-centre (>600), retrospective study on a prospectively collected national database (n=449,524). 2002-4. Exclusion (number not given): RRT; missing data: SCr, age, gender, race. Validated in 2nd sample (n=86,009) January – June 2004. Outcome: RRT. Indications not defined. eGFR - MDRD equation. 24 variables studied. Backward stepwise selection. Regression coefficients from simplified model converted into whole integers to create bedside prediction tool. Missing values of risk factors & outcomes defaulted to most common value; missing LVEF estimated at 50%.	RRT in 1.4% (n=6,451) of derivation & 1.6% (n=1376) of validation sets. Mortality with RRT 45% (n=3,522) & without RRT 2.5% (n=13,193).	10 variables: Pre-op SCr, age, race, type surgery, diabetes, shock, NYHA class, lung disease, recent MI & prior surgery. c-statistic 0.83 derivation & validation sets. H-L P=0.07 in validation. Predicted vs observed plotted for validation set.	Wijeysundera 2007 –0.75-0.78 Englberger 2010 – RRT 0.81 (0.78-0.86), severe AKI 0.76 (0.73-0.80) Ng 2014 –0.61 Bernie 2014 AKI 0.74 (0.72-0.76), stage-3 AKI 0.79 (0.77-0.82) Kim 0.60-0.62 Vives 2011 0.76 (0.72-0.79)
(359) Brown 2007 **Cardiac surgery** **DV** **TRIPOD 1B** **(26/37 pts)**	USA mulit-centre (8), retrospective study on a prospectively collected database (n=11,301). 2001-5. Exclusion (n=2938): moderate (eGFR - MDRD) <60) or severe (eGFR <30 ml/min/1.73m^2) renal dysfunction at baseline; those with missing values. Included: n=8,363. Outcome: 'severe renal insufficiency' (eGFR <30) – worst post-op SCr. 29 variables studied. Significant univariate variables (14) put in backward stepwise multi-variable logistic model. Bootstrapping (200) used.	Outcome in 3% (n=229/8,363). In-hospital mortality 26.2% with outcome vs. 0.7% without. RRT 0.37% (n=31).	8 variables: Age (70-74, 75-79 or ≥80), sex, diabetes, PVD, HTN, IABP, white blood count, prior CABG. AUC 0.72 (0.68-0.75). H-L P-value =0.28. Observed & expected deciles of risk highly correlated (R=0.97).	
(350) **Wijeysundera (SRI) 2007** **Cardiac surgery** **D, V, EV** **TRIPOD 3** **(32/37 pts)**	Canadian dual centre, retrospective study of prospectively collected database, Consecutive cases CPB. Simplified renal index (SRI).(n=20,131). Toronto derivation (n=10,751) 1999-2004 & validation at Toronto (n=2,566) 2004-5 & Ottawa (n=6,814) 1999-2003. Validated Cleveland, Chertow & Mehta scores. Exclusions (no number given): pre-op RRT or SCr >300 µmol/L]) & infrequent procedures (eg, heart transplant). Outcome: post-op RRT.	RRT overall 1.7% (n=336/20,131): derivation 1.3% (n=139), Toronto validation 1.8% (n=45) & Ottawa validation	7 variables (0-5pts): pre-op eGFR ≤30 (2 points), diabetes, LVEF ≤40%, previous cardiac surgery, procedure, urgency, & IABP. AUC 0.83. H-L P=0.48. Derivation AUC 0.81 (0.78-0.84), H-L P=0.27. Internal validation 0.78 (0.72-0.84). Derivation vs Validation sets plotted by risk score	Same study external validation 0.78 (0.74-0.81), Candela-Toha 2008 - 0.82 (0.76-0.87) Knapik 2008 - 0.73 (0.62-0.81) Englberger 2010

	RRT decision by nephrologist. Indications: metabolic (acidosis, hyperkalemia), anuria, & overload. From 18 variables in bivariate analysis, 14 (pre & post-op) fitted into multivariable logistic regression model. Bootstrapping (200) used. Age dichotomised. Missing data (<0.5%): imputed; Dichotomous variables assigned most frequent sex-specific value & continuous variables assigned median value. Unknown LVEF considered normal. eGFR: Cockcroft-Gault.	2.2% (n=152). Mortality: derivation 47% (n=65/139) if RRT required vs 1% (n=115/10,612) if not.		RRT - 0.79 (0.77-0.82) & severe AKI 0.75 (0.72-0.77) Ng 2014 – 0.75 Kim 2013 - 0.63-0.62 Vives 2011 - 0.79 (0.76-0.82)
(351)Palomba (AKICS) 2007 Cardiac surgery D,V TRIPOD 2B (25/37 pts)	Brazilian single centre retrospective derivation cohort (2003-5) with prospective validation (2005-6). Exclusions (number not given): age <18 years, >90, emergent surgery, congenital heart disease repair, aortic aneurysm, pre-op SCr >3.0 mg/dl and renal transplant patients. Outcome: increase SCr to >177 µmol/L if baseline <133; or increase 50% if SCr 133-265 µmol/L <7 days post-op. Derivation (n=603) & prospective validation sets (n=215) - patients admitted to ICU post-op. RRT Criteria not defined. eGFR: Cockcroft-Gault. 20 Variables: 13 pre-op, 3 intra-op, 3 early post-op; those significant on univariate placed in multivariable model. Continuous variables dichotomised according to AUC No information on missing data.	Outcome in 11% (n=66) of whom 18% (n=12) required RRT in derivation. In validation outcome in 14% (30/215). Overall mortality 6%; with AKI 39% (n=26); & 82% (n=10) if RRT required.	8 variables: Age >65, pre-op SCr >106 µmol/L, pre-op glucose >7.8mmol/L, CCF, combined surgery, CPB time, low cardiac output & low CVP. Derivation AUC 0.84 (0.78–0.89), H-L P=0.8. Validation AUC 0.85 (0.79–0.90), H-L P=0.24. Risk categories presented in table for validation set.	Kim - 0.66-0.68 Kiers: AKI-D - 0.84, AKI-R 0.73, AKI-I 0.75, AKI-F 0.82
(344) Aronson 2007 Cardiac surgery D,V TRIPOD 2A (31/37 pts)	Multi-centre (70, across 17 countries), patients prospectively recruited into a study (Multicenter Study of Peri-op Ischemia (McSPI) EPI-II Study), with data retrospectively analysed (n=5,436) 1996-2000. Inclusion: >18 yr, CABG (CPB). Exclusions: 371 withdrew; 256 other surgery; 9 incomplete data. Included: n=4801. Derivation cohort (n=2,381) & validation set (n=2,420) split by time. n=47 excluded from validation set (missing data). Composite outcome: post-op SCr ≥177 µmol/L & increase ≥62 µmol/L from baseline, or renal failure (RRT or evidence renal failure at autopsy). 31 variables (26 pre-op, 5 intra-op). Treating physicians blinded to research data. Outcomes pre-specified, defined by protocol, & determined by independent observers blinded to study question. All variables with p-value <0.15 entered into multi-variable models using combination stepwise selection method. Missing data: patients excluded.	Outcome in 4.8% (n=231); 27% (n=63) died vs 1.8% (n=84/4570) without outcome. Outcome associated with significant increased ICU & hospital LOS & cost.	8 variables: age >75, CCF, MI, CKD, Pre-op pulse pressure >40 mm Hg, intra-op multiple inotropes, IABP, CPB time. Derivation AUC 0.84. H-L P=0.84. Validation AUC 0.80. 3 risk groups (low, medium, high) plotted derivation vs validation.	

(354) Rahmanian 2011 **Cardiac Surgery** **D,V** **TRIPOD 2B** **(26/37 pts)**	USA single-centre, retrospectively analysis of a prospectively collected database (n=3,016), 2007-9. Exclusion: pre-op RRT (n=21). Included n=2,559 n=2,511 derivation, n=484 internal validation (2009). Outcome: post-op RRT. RRT decision by nephrologist & surgeon (no criteria given). Variables: 25 pre-op, 5 intra-op. Those significant in univariate analysis (P-value <0.05) included in stepwise multivariate logistic regression. No information on missing data.	RRT in 3.9% (n=98) in derivation set. Mortality RRT group 37.8% (n=37) vs.1.3% (n=32) if no RRT. Outcome validation set not given.	9 variables: Age >75, diabetes, CCF (NYHA 4), PVD, MI (<21 days), AF, Pulmonary HTN, SCr >177 µmol/L, CPB time. AUC 0.83 (0.78-0.86). Validation cohort 0.85 (0.77-0.94). No calibration data.	Kiers AKI-D 0.78, AKI-R 0.65, AKI-I or more 0.67, AKI-F 0.74
(358) Demirjian 2012 **Cardiac surgery** **D, EV** **TRIPOD 1B & 4** **(24/37 pts)**	USA single centre, prospectively collected data retrospectively analysed (n=25,898). Externally validated Cleveland score. 2000-8. Inclusions: 1st surgical episode. Exclusions: ESRD, RRT; use pre-surgical IABP, ECMO variables with missing data ≥10% or prevalence <5%. Outcomes: RRT; composite of doubling SCr or RRT <2 weeks. Initial 47 variables (pre- & intra-op) reduced after exclusions to 31 in combined & 27 in pre-surgical models. eGFR - MDRD. continuous missing data variables replaced with median values. Bootstrapping & shrinkage for validation (1000). Retained variables using backward selection with P value <0.10. Model comparisions based on Akaike information criterion (AIC) minimisation.	Outcome (doubling SCr or RRT) in 4.3% (n=1,113). RRT in 1.7% (n=429). Mortality for RRT 48% (n=206) & doubling of SCr 10% (n=68).	Pre-op (18 variables): female, race, BMI, Pulmonary disease, CCF, Diabetes, HTN, Procedure, previous surgery, emergent, GFR, Albumin, HCO_3, Sodium, Urea, Haemoglobin, Platelets, Bilirubin. Composite: C-Statistic 0.80 (0.78-0.81) H-L P=<0.001. RRT: 0.88 (0.86-0.89) H-L P=0.2. Pre & intra-op (20 variables): Race, BMI, Pulmonary disease, CCF, Diabetes, HTN, emergent surgery, GFR, Albumin, potassium, HCO_3, Sodium, Urea, Haemoglobin, Platelets, Bilirubin, CPB time, intra-surgical transfusions, vasopressor use, urine output. Composite: 0.83 (0.81-0.84) H-L P=0.09. RRT: 0.91 (0.90-0.92), H-L P=0.2. Optimism correction <0.002 with AIC 0.96-0.97 – precluded need for shrinkage.	
(353) Brown 2012 **Cardiac surgery** **D,V** **TRIPOD 3** **(27/37 pts)**	USA single centre, retrospective analysis of prospectively collected database (n=4,987); 2002-7. External validation (n=1,219) separate cohort (six centres, July – December 2007). Exclusion: RRT (n=70), AKI prior, ESRD, renal transplant; missing data (n=86). Derivation included n=4,831. External validation n=1,219. Outcome: *duration* of AKI – increase ≥26.5 µmol/L or ≥50% SCr baseline. eGFR – MDRD. Pre-op (12) intra-op (27) & post-op (5) variables in univariate analysis followed by multivariable model using backwards stepwise approach.	AKI in 39% (n=1,886/4,831) of derivation & 35% (n=427/1,219) of Validation sets.	8 variables: Age, Male, Diabetes, HTN, vascular disease, eGFR <60, transfusion, previous surgery. Derivation c-index 0.66 & validation 0.71. No H-L p-value. Plotted derivation vs validation sets according to risk score.	Berg 2013 - 0.65 (0.63-0.68)
(355) Heise 2012 **Cardiac Surgery**	German, single centre ICU, retrospective analysis (n=1,560). Exclusions: Missing data (n=369).	AKI in 38% (455/1,191) -	3 scores: Pre-op: previous surgery, SCr >97 µmol/L, age, BMI & EURO-Score;	

D,V **TRIPOD 2B** **(26/37 pts)**	Included n=1,191 – n=662 derivation (2004-5), n=529 validation set (2005-6). Outcomes: AKI by SCr AKIN criteria. Variables: 27 pre-op, 8 intra-op, 7 post-op. Dichotomised continuous variables using AUC assessment. Univariate followed by conditional forward multivariate analysis. mode & a non-conditional stepwise mode. Overall model fit statistic assessed by Likelihood Ratio Test.	Derivation, n=253/662 & validation n=202/529. No mortality data.	Pre & intra-op: previous surgery, time, SCr, age, BMI, CPB time, transfusion, RRT during CPB; Pre, intra & post-op: prev surgery, SCr, time, peri-op increase SCr, age, BMI, transfusion, glucose at ICU admission & transfusion <24hrs post-op. Pre-op AUC 0.69 (0.65-0.73) derivation & 0.69 (0.65–0.73) in validation sets. Pre & intra-op AUC: 0.69 (0.65-0.73) derivation 0.69 (0.65–0.73) in validation. Pre, intra & post-op score AUC: 0.73 (0.69-0.77) derivation & 0.73 (0.69-0.76) in validation set. No calibration statistics. Risk groups graphed for validation: score 0-1 24%, 2-3 35%, 4-5 60%, >5 70%.	
(357) Berg 2013 **Cardiac Surgery** **D,V, EV** **TRIPOD 2B & 4** **(26/37 pts)**	Norway, single centre, retrospective analysis of a prospectively collected database (n=5029) 2000-7. Exclusions (n=51): RRT (n=9) & missing SCr (n=42). Included n=4,978. Outcome: 50% increase SCr, increase ≥26.4 µmol/l (maximal post-op value used) or RRT.3 Bootstrapping (400) for internal validation. Calibration curves. 16 pre-op & 6 intra-op variables. Model reduced using limited backwards step-down, keeping variables according to Akaike's Information Criterion. Also validated 3 other prediction scores. Delong's methods used to compare the AUCs. No further information on missing data.	Outcome in 12.7% (n=633), of whom 11.2% (n=71) died. RRT in 65, of whom 37% died (n=24) Mortality 0.8% (n=36) without the outcome.	11 variables: age, BMI, statin use, HTN, PVD, chronic pulmonary disease, haemoglobin, SCr, previous surgery, emergency operation & operation type. AUC 0.82 (0.80-0.84). H-L P=0.17. Estimated shrinkage factor 0.976. Calibration curves plotted for validation. Brown AUC 0.65 (0.63-0.68), H-L P= <0.0001. Antunes(228) - AUC 0.74 (0.72-0.76), H-L P= <0.0001 (model incorporating Age & SCr). Thakar score for Outcome AUC 0.77 (0.75-0.79) & for RRT AUC 0.88 (0.84-0.93), H-L P=<0.0001.	
(356) Kim 2013 **Thoracic Aortic surgery with CPB.** **D,V** **TRIPOD 2A** **(27/37 pts)**	South Korean, single centre, retrospective study (n=799) & external validation of 4 risk scores (Palomba, Wijeysundera, Mehta, Thakar). 1997-2010. Exclusion: RRT (n=12), missing SCr or urine output values (n=44), died <24 hrs post-op (n=6). Included n=737 - random allocation derivation (n=417) & validation (n=320). Outcomes: RIFLE criteria for AKI <7 days post-op & RRT <30 days post-op. 56 variables - 20 in univariate analysis; if p value <0.05 included in multivariate model; backward selection employed. For development of final model, continuous variables categorised according to cut-off point on AUC. Delong's methods used to compare AUCs.	123/417 (29%) developed AKI; 24 (5.8%) required RRT (n=3 died). AKI Mortality 9.9%.	6 variables: Age >60, eGFR <60, pre-op LVEF <55%; Intra-op: Oliguria, diuretic use & surgery duration >7hrs. Derivation AUC 0.74 (0.69-0.79), No H-L reported. Validation AUC 0.74 (0.69-0.80), H-L P=0.99. Risk scores & likelihood ratios (0-5) plotted derivation vs validation sets.	
(82) Birnie 2014 **Cardiac Surgery**	UK, multi-centre, retrospective analysis of a prospectively collected database (n=30,854). 1996-2010.	AKI in 23% (4,892/20,995)	15 variables: Age, male, BMI >35kg/m^2 smokers, NYHA 3/4, diabetes, PVD, HTN, anaemia,	Same study - 0.74 (0.72-0.76),

D,V,EV **TRIPOD 3** **(32/37 pts)**	Validated 4 other risk scores (including Cleveland, Mehta & Ng). Inclusion: ≥16 years, +/- CPB (including thoracic aorta). Exclusions: RRT, kidney transplant patients, deaths in theatre & missing data (n=9,859). Included n=20,995 (Complete case analysis). 2 sites derivation (n=16,527 - n=12,435 one site & 4,092 other site); external validation one site (n=4,468). Outcomes: AKI KDIGO guidelines (change SCr) - Stage-1 - increase ≥26µmol/L or x1.5-1.9 pre-op SCr <7 days; Stage 2: x2-2.9; Stage-3: ≥x3 or increase to ≥354µmol/L, or RRT. Other outcomes included: LOS; in-hospital mortality. Indications RRT: uraemia, overload, or biochemical, according to institutional protocol. 24 pre-op & intra-op Variables first underwent univariate analysis. Calibration plots of observed versus predicted values analysed using linear regression to provide slope & intercept. Used multiple imputations; dichotomised continuous data. eGFR - Cockcroft-Gault. Critical pre-op event = cardiogenic shock, inotropes, ventilation or IABP.	; Stage-3 in 4.4% (n=919). Mortality: no AKI 0.5% (77/16,103), AKI Stage-1 2.2% (74/3,430), Stage-2 3.7% (20/543), Stage-3 28% (258/919). >80% of deaths preceded by AKI.	eGFR, catheter to surgery <24 hours, x3 vessel disease, poor LVEF, emergency or salvage operation & complexity. Any stage AKI AUC 0.73 (0.72-0.74); H-L P=0.49. Stage-3 (added critical pre-op event) - AUC 0.78 (0.75-0.80), H-L P=0.001. Calibration plots graphed.	H-L =0.19. Stage-3 AKI: initial model - 0.78 (0.75-0.80) & more inclusive model - 0.79 (0.76-0.81)
(352) Ng 2014 **Cardiac surgery** **D,V,EV** **TRIPOD 2A** **(28/37 pts)**	Australian, multi-centre (18) retrospective study (n=32,279). Exclusions: pre-op RRT (n=521), missing data SCr (n= 86) & other variables (n=3250). 2001-9. Included n=28,422. Random split derivation (n=17,095) & validation sets (n=11,327). Pre & post-op derivation models proposed & external Validation of the Thakar, Mehta & SRI scores. Outcome: AKI = 2 of: increase SCr to >200 µmol/L; x2 SCr; RRT. eGFR - Cockcroft-Gault. 38 pre-, peri- & early post-op variables. NYHA class (2,713 missing) & RBC transfusions (4,241 missing) not considered. Missing LVEF: left out or estimated. Bootstrapping (1000) employed to select candidate models. Using the creation data set, the AUC, Bayesian information criteria & H-L P value & prediction mean square error were calculated to select a final model.	AKI 5.8% (n=1,642). 30-day mortality rate if AKI developed 17.4% (n=286) vs. 1.6% if no AKI (n=428).	10 pre-op variables: BMI>30, endocarditis, procedure, SCr, diabetes, urgency, eGFR, CCF, age, & cardiogenic shock). AUC 0.77. H-L P=0.06. 17 variables including intra-op: BMI>30, pre-op SCr, diabetes, IABP, endocarditis, CPB time, cardiogenic shock, non-RBC blood products transfused, gender, re-op for bleeding, age, eGFR, urgency, HTN hypercholesterolemia, procedure & pulmonary disease. Validation AUC 0.81. HL P=0.6. In validation with complete data pre-op AUC 0.77 & intra-op AUC 0.83. Risk score range: 4-15 pts 2%, 16-25 7%, 26-44 25%.	Bernie 2014 – AKI - 0.73 (0.71-0.75); Stage-3 AKI - 0.79 (0.76-0.82)
(151) Fortescue 2000 **Cardiac surgery** **EV** **TRIPOD 4** **(25/31 pts)**	USA multi-centre (12), retrospective study on a prospectively collected database (n= 9,498). External validation of Chertow score. 1993-5. Exclusion: baseline SCr >265 µmol/L, valve surgery. Included n=8,797. Outcome: RRT (no criteria). Missing data: binary variables patient removed; categorical	RRT in 1.2% (106/8,797). No mortality data.	Chertow score - AUC 0.72 & H-L P=0.28. Risk groups plotted: Low (0-5pts) 0.5%, Medium (6-10) 0.9%, High (11-15) 2.9%, Very high (>15) 4%.	

	coded missing & eGFR (Cockcroft-Gault) included average of any missing value.			
(149) Thakar 2003 **Cardiac surgery** **EV** **TRIPOD 4** **(22/31 pts)**	USA, single centre, retrospectively analysed a prospectively collected database (n=24,660); External validation of Chertow score. 1993-2000. Exclusion (n=2,071): heart transplants, RRT, ECMO, LVAD, pre-op tracheostomy or ventilation; defibrillator implants; missing data. Included n=22,589. Outcome: AKI defined: RRT or (in 'inclusive definition') a >50% reduction eGFR from baseline. Nor RRT criteria eGFR - Cockcroft-Gault.	RRT in 1.8% (n=412/22,589). Frequency post-op RRT 0.5-15.5% based on risk. No mortality data.	Chertow score - AUC 0.78 for RRT. Using more inclusive definition of AKI - 0.65.	
(554) Eriksen 2003 **Cardiac surgery** **EV** **TRIPOD 4** **(24/31 pts)**	Norway, single-centre, retrospective Case-control analysis (n=2,154). 1995-9. External validation of Chertow score. Exclusion (n=117): RRT, active endocarditis, simultaneous other surgery & pre-op SCr >265 µmol/l. n=38 cases vs Controls: random sample (33%, n=640) of n=2,037 patients not requiring RRT. Outcome: RRT <30 days (no criteria). eGFR - Cockcroft-Gault. Missing data: dichotomous coded absent & average value used for continuous variables.	Post-op RRT 1.9% (n=38). 30-day mortality for patients requiring RRT 45% (n=17) & 2.6% (n=55) without RRT.	Chertow score - AUC 0.71. H-L P=0.40.	
(555) Englberger 2010 **Cardiac surgery** **EV** **TRIPOD 4** **(23/31 pts)**	USA single center retrospective analysis (n=12,999). Thakar, Mehta & Wijeysundera studies validated. 2000-7. Excluded: missing data (n=757), SCr >301 µmol/L (n=35), deaths <24hrs (n=111). Included n=12,096. Outcome: post-op RRT or 'severe AKI' (SCr 177 µmol/L & x2 increase from baseline).	RRT in 2.1% (n=254), & severe AKI in 3.9% (n=467).	Cleveland: AUC 0.86 (0.84-0.88) for RRT & 0.81 (0.79-0.83) for severe AKI; Mehta: 0.81 (0.78-0.86) & 0.76 (0.73-0.80) respectively; Wijeysundera - 0.79 (0.77-0.82) & 0.75 (0.72-0.77). Cleveland H-L P=0.2 RRT severe AKI P=0.7. Mehta H-L P=0.6 for RRT & P=0.4 for severe AKI	
(556) Candela-Toha 2008 **Cardiac surgery** **EV** **TRIPOD 4** **(26/31 pts)**	Spanish, single centre, retrospective analysis of a prospectively collected database (n=1892). 2002-6. Thakar & Wijeysundera scores externally validated. Exclusions (n=112): RRT & renal transplant; intra-op or early (<24 h) post-op deaths & missing data (n=13). Included n=1,780 (Thakar) & n=1,563 (Wijeysundera). eGFR - Cockcroft-Gault. Outcome: RRT. Indications RRT: overload, uraemia or biochemical; nephrologist decision.	RRT 3.7% (n=67) of Thakar & 3.8% of Wijeysundera cohorts. No mortality data given.	Thakar Score - AUC 0.86 (0.81-0.9) H-L P=<0.001. Wijeysundera - AUC 0.82 (0.76-0.87). H-L P=<0.001. Both models re-calibrated with resulting H-L P values of 0.17 (Thakar) & 0.32 (Wijeysundera).	
(557) Di Bella 2007 **Cardiac surgery**	Italian single centre, retrospective study (n=1,642). 2003-6. External validation of Thakar score. Excluded: RRT, ECMO, ICD, ventilated/tracheostomy patients	RRT in 1.3% (n=22) of whom 73%	Thakar AUC 0.82 (0.74-0.90). No H-L P-value. Risk range requiring RRT: 0-2 pts – 0.4%, 3-5 –	

EV **TRIPOD 4** **(21/31 pts)**	& those with missing data. Outcome: post-op RRT (no criteria).	(n=16) died.	1.9%, 6-8 – 13%, 9-10 – 33%.	
(558) Knapik 2008 **Cardiac surgery** **EV** **TRIPOD 4** **(21/31 pts)**	Polish single centre, retrospective external validation of Wijeysundera score (n=1421). 2005-6. Exclusions: RRT or SCr >300 µmol/L, rare procedures (heart transplantation, LVAD). Outcome: post-op RRT (anaesthetist or surgeon decision). RRT indications: metabolic (hyperkalaemia / acidosis), anuria, overload. eGFR - Cockcroft-Gault. No missing data information.	RRT in 2.3% (n=33) of whom 76% (n=25) died.	Wijeysundera score - AUC 0.73 (0.62-0.81). No calibration data.	
(380) Heise 2010 **Cardiac surgery** **EV** **TRIPOD 4** **(25/31 pts)**	German, single centre retrospective external validation of the Thakar score (n=3628). 2002-5. Exclusions (n=120): RRT, LVAD, missing data. Included n=3508. Outcome: post-op RRT. Indications RRT: hyperkalaemia, overload & uraemia. To reduce bias: single physician evaluated all patients regarding presence of risk factors.	RRT in 11% (n=383). No mortality data.	Thakar AUC 0.66 (0.65-0.68) A modified score did not differ significantly – AUC 0.67 (0.66-0.69). No calibration data.	
(559) Vives 2011 **Cardiac surgery** **EV** **TRIPOD 4** **(24/31 pts)**	Spanish multi-centre (24) retrospective external validation of Thakar, Mehta & Wijeysundera scores (n=9,121). 2007. Case control - each centre supplied patients requiring post-op RRT & patients (up to n=50) as controls. Included n=1084 (n=248 RRT vs 836 controls). Exclusions: mortality <48hrs, RRT, minor surgery. Outcome: RRT (no criteria). eGFR - MDRD. No missing data information.	RRT in 248. Mortality 66% (n=164/248) in those requiring RRT vs. 3.8% (n=32/836) if not.	Thakar AUC: 0.82 (0.79-0.85) H-L P=<0.001. Mehta AUC 0.76 (0.72-0.79) H-L P=0.03. Wijeysundera AUC 0.79 (0.76-0.82) H-L P=<0.001.	
(560) Kiers 2013 **Cardiac surgery** **EV** **TRIPOD 4** **(22/31 pts)**	Netherlands, single centre, retrospective validation of 8 prediction scores (n=1,409). 2006-9. Exclusions (n=21): RRT (10), kidney transplant (3) or died <24 h post-surgery (n=6); rare procedure, missing data each n=1. Total cohort n=1,388, Chertow, Thakar, Wijeysundera, Rahmanian, Mehta (n=1,247), Brown (n=763), Palomba (n=1,244), Aronson (n=833). Outcome: post-op RRT <7 days (AKI-D) & AKI RIFLE definitions (SCr - risk AKI-R, injury AKI-I, failure AKI-F). eGFR - Cockcroft–Gault. Independent researcher checked variables to improve validity.	RRT 1.9% (n=27). AKI in 9.3% (n=129), AKI-R 4.7% (n=65), AKI-I 2.3% (n=32) & AKI-F 2.3% (n=32). Mortality with RRT 15% (n=4), without RRT 1.4% (n=20).	AUCs: Thakar: AKI-D 0.93, AKI-R 0.75, AKI-I or more 0.81, AKI-F 0.88. Mehta: 0.85, 0.74, 0.79, 0.84 respectively Wijeysundera: 0.88, 0.70, 0.77, 0.83 respectively Palomba: 0.84, 0.73, 0.75, 0.82 Rahmanian: 0.78, 0.65, 0.67, 0.74. Chertow & Brown AUCs not reported. No calibration data.	
(561) Wong 2015 **Cardiac Surgery**	China, single centre, retrospective case-control external validation - modified Thakar Score (n=2,343). 2007-2011.	AKI 1: 6% (n=142); 2:	Thakar for AKI Stage 1 AUC 0.61 (0.56-0.65), H-L P=0.29,	

EV **TRIPOD 4** **(25/31 pts)**	Exclusions: RRT (n=25), missing SCr (n=2). Included n=2316. Outcome: AKI KDIGO Stages 1-3 or RRT. Stage 1-2 <5 days & Stage 3 & RRT in-hospital. No criteria for RRT. Excluded Insulin or Emergency surgery.	2.6% (n=60); 3: 5.8% (n=134); RRT 5.4% (n=125). Mortality: Any AKI 29% (n=98/336), Stage 1: 14% (n=20); 2: 23% (n=14); 3: 48% (n=64). No AKI 2.5% (n=50/1980)	2 - AUC 0.61 (0.54-0.68) H-L P=0.08, 3 - AUC 0.78 (0.74-0.82), H-L P=0.84. RRT - AUC 0.78.	

A2.12(ii) Non-cardiac surgery studies

Author & setting **Derivation (D)** **Internal validation (V), External validation (EV)** **TRIPOD study type & checklist fulfilled**	**Population, Outcome, AKI Definitions, Methods**	**Results**	**Risk factors in model & Performance**	**Externally validated?**
(361) Sanchez 2004 **Liver Tx** **D** **TRIPOD 1A** **(22/37 pts)**	USA single centre, retrospective analysis of a prospectively collected database (n=766). 1996-2001. Exclusions (n=42): RRT (n=22), liver-kidney transplants (20). Included n=724. Validation cohort n=122 (2002). Outcome: post-op RRT - decision by nephrologists. 22 Variables: 11 pre-op, 9 intra-op & 2 post-op. Those with P <0.05 in univariate analysis included in multivariate logistic regression; cut-off for each variable established by goodness-of-fit model & a subsequent model constructed. No missing data information.	RRT in 87/724 (12%). Mortality if RRT required 8% (n=7/87) vs. 2.7% (n=17/637) without.	4 variables: pre-op SCr >168 µmol/L, pre-op urea >9.6 mmol/L, post-op ICU stay >3 days, & MELD score >21. AUC: 0.91. No calibration. AUC in validation not reported. Validation cohort model predicted 15/20 RRT patients.	
(369) Boyle 2006 **Heart Tx** **D** **TRIPOD 1B**	USA single centre, retrospective analysis of a prospectively collected database (n=774). 1993-2004. Excluded (18): pre-op RRT (n=12), missing data (n=2), previous transplant (n=4). Included n=756. Outcome: in-patient RRT (no criteria).	RRT in 5.8% (44/756). In-hospital mortality 50% (22/44) for RRT vs 1.4% (10/732).	4 variables: pre-op SCr, Albumin, Diabetes & CPB time. AUC 0.74-0.75, H-L P=0.19	

(24/37 pts)	Variables included pre-op (21) demographics, co-morbidities, drugs, surgery type, lab data & intra-op (3). eGFR – MDRD. Missing data: multiple imputations. Univariate analysis P-value <0.50 put in multivariable analysis. Bootstrapping (1000 repetitions, stepwise variable selection). Variables selected on >50% of bootstrap runs selected for final model.			
(360) Rueggeberg 2008 **Liver Tx** **D,V** **TRIPOD 2B** **(25/37 pts)**	German, single centre retrospective study (n=265). Inclusions: 1st operation, ≥18yr. Exclusions (n=27): 2nd op (7), RRT, Age (8), combined (3), transplant pre-study start (9). Included n=238: n=71 derivation (2000-1), n=167 validation (2001-3). Outcome: RRT. Indications for RRT – acidosis, overload, uraemia. No information on missing data. Pre-op (18) and intra-op (13) variables. Univariate variables P=<0.15 put into forward stepwise logistic regression analysis.	RRT in 18% (n=13) of derivation & 14% (n=23) of validation sets. 28 day mortality if required RRT 17% (n=6/36) vs. 1% (n=2/202) without.	6 Variables: Hepatitis B/C, HTN, pre-op MAP, transfusion intra-op, intra-op lactate, MAP intra-op <50mmHg. Derivation AUC 0.91, H-L P=0.99. No validation AUC/Calibration statistics.	
(365) Slankamenac 2009 **Liver resection** **D,V** **TRIPOD 2A** **(28/37 pts)**	Swiss, single centre retrospective database study (n=576). Exclusion: trauma (n=7). Included n=569. 2002-7. Random split derivation (n=380) & validation (n=189) sets. Outcome: RIFLE increase SCr >26.5 µmol/L baseline, or increase >x1.5 <48 hrs post-op, or reduction urine output <0.5ml/kg/hr >6hrs. 7 pre-op variables. CKD = eGFR <60. CVD = IHD, coronary revascularization, cerebral arterial disease or PVD. Used Shrinkage & bootstrapping. No missing data information. Continuous variables dischotomised; used shrinkage. For simplified model used stepwise backward regression & bootstrapping (380). For score regression coefficients transformed into points.	AKI overall 15.1% (n=86/569) – 15.3% (n=58/380) derivation & 14.8% (n=28/189) validation sets. Mortality with AKI 22.5% (n=20/86) vs 0.8% (n=4/483) without (p<0.001). 75% had malignancy - 49% had pre-op chemotherapy.	7 variables (0–22 points): Age (<60, 60-69, ≥70), female, CKD, CVD, diabetes, bilirubin ≥17, ALT ≥35/50. AUC 0.8. Simplified (0-7pts): ALT, CVD, CKD & diabetes - AUC 0.77, H-L P=0.75. Validation - 0.77, H-L P=0.98.	Slankamenac 2013 - 0.60 (0.52–0.69) (same centre)
(366) Xu 2010 **Liver Tx** **DV** **TRIPOD 2B** **(25/37 pts)**	China, single-centre retrospective review of prospectively collected data. Included n=132. Derivation n=102 (2004-5) & prospective validation n=44 (2005-6). Outcome: AKI <7 days post-op: >133 µmol/L & increase >50% or need for RRT. No criteria for RRT. No information on exclusions or missing data or how many variables assessed. Continuous variables dischotomised; 14 variables in univariate analysis with P <0.05 put in forward stepwise multivariate logistic regression analysis.	AKI in 32% (n=33/102), & 10% (n=10) required RRT. 1 year mortality 29.9% vs. 7.7% without AKI (P= 0.001).	2 Variables in pre-op model: SCr & Sodium: AUC 0.77. 4 Combined Pre-op SCr with intra-op hypotension, urine output & use of vasopressor – AUC 0.91, H-L P-value 0.97. Validation AUC: 0.89.	
(364) Slankamenac	Swiss, single centre, retrospective database study (n=576). 2002-7.	Frequency post-op AKI 14.9%	7 variables: Blood transfusions, hepatico-jejunostomy & oliguria (intra-op) & ALT, CVD,	

2013 **Liver resection** **D,EV** **TRIPOD 1B** **(27/37 pts)**	Criteria: RIFLE >26.5 µmol/L SCr or increase >x1.5 <48hrs. Excluded (n=27): incomplete data, trauma. Included n=549. 7 intra-op & pre-op variables. Stepwise backward logistic regression with cross validation & bootstrapping. Shrinkage & a constant factor (Copas factor) calculated to indicate degree of overfitting. Decision curve analysis performed.	(n=82/549). Mortality 23% (n=19/82) vs. 0.8% (n=4/467).	CKD & diabetes (pre-op) - AUC 0.81 (0.76-0.86). H-L P=0.93. Pre-op alone - 0.60 (0.52-0.69). k-fold Cross-validation - 0.79 (0.73–0.84). Observed & predicted risks plotted.	
(362) Romano 2013 **Liver Tx** **D** **TRIPOD 4** **(23/37 pts)**	USA single centre, retrospective study (n=114). Exclusion: RRT (n=22). Included n=92. 2009. Outcome: post-op AKI - SCr increase >26.5 µmol/L <72hrs. No criteria for RRT. 22 pre & intra-op variables investigated; those on univariate analysis (p=<0.10) entered in backward stepwise model. No missing data information.	Outcome in 67% (n=52). RRT n=13. Mortality with outcome 38.5% (n=20). Mortality with RRT 62% (n=8).	MELD (incorporating bilirubin, SCr & INR) AUC - 0.80. H-L P=0.34.	
(363) Kim 2014 **Liver Tx** **D,V** **TRIPOD 1B** **(22/37 pts)**	South Korean single centre, retrospective study (n=173). Exclusions (n=16): RRT (15) or ESRD pre-op, combined transplant (1), age <18. 2008-2011. Included n=157. Outcome: RRT post-op (no criteria). Variables: 10 pre-op: demographics, history, blood parameters, 5 intra-op including transfusions, urine output. Univariate followed by stepwise multivariate analysis. Bootstrapping (200 repetitions) employed. No missing data information.	RRT in 27% (n=42/157). 90 day mortality 43% (n=18) if RRT required vs 1% (n=1) if no RRT.	5 Variables: type of donor, pre-op encephalopathy, MELD score, tumour as indication & intra-op blood loss. Derivation AUC 0.90 (0.85-0.95) No H-L. Validation AUC 0.90 (0.85-0.95); predicted & observed frequencies plotted.	
(371) Grimm 2014 **Lung Tx** **DV** **TRIPOD 2A** **(29/37 pts)**	USA multi-centre retrospective study of a prospectively collected database (n=10,963). 2005-2012. Exclusion (no numbers): age <18, additional organ transplanted. Random split derivation (80%, n=8771) & validation sets (20% n=2192). Outcome: RRT (no criteria given). Missing data: excluded. 18 Variables in univariate analysis P=< 0.2 or clinically relevant entered into multivariable logistic model in a forward & backward fashion.	RRT in 598/10,963 (5.5%). No mortality data.	15 variables: age, race, diagnosis, lung allocation score, BMI, diabetes, eGFR, ventilation, ICU pre-op, ECMO bridge, double lung, Karnofsky performance scale, days waiting, infection <2/52 & ischemia time ≥6 hours. C-statistic 0.71 derivation & 0.73 validation. No H-L reported. Sample weighted bubble plot predicted vs observed r = 0.86.	
(370) Kilic 2014 **Heart Tx** **DV** **TRIPOD 2A** **(31/37 pts)**	USA multi-centre, retrospective analysis of a prospectively collected database (n=14,635). 2000-2010. Random split derivation (80%, n=11,689) & validation sets (20%, n=2,946). No information on exclusion numbers. Excluded: aged <18 or additional organs transplanted or re-do. Outcome: post-op RRT (no criteria).	RRT required in 1128/14,635 (7.7%). No mortality data given.	13 variables: age ≥60, congenital or ‘‘other’’ heart failure cause, eGFR <60 mL/min, bilirubin ≥17 µmol/L, BMI ≥30 kg/m^2, diabetes, ventilation, ICU prior, recent infection, blood transfusion on wait list, donor age ≥30, biatrial anastomosis, & ischemic time ≥4 hours. Derivation AUC 0.70, H-L Chi-square 15.4 – “non significant”	

	Variables in univariate analysis with P=<0.20 entered into multivariable logistic regression model. Continuous data dichotomised. Variables with >25% missing data excluded.		Validation AUC 0.70, H-L Chi-square 3.7 “non significant”. Sample weighted bubble plot predicted vs observed r = 0.91.	
(368) Hilmi 2015 **Liver Tx** **D** **TRIPOD 1A** **(20/37 pts)**	USA single centre, retrospective study of a prospectively collected database (n=543). 2005-9. Exclusions: RRT (n=107), death <24 hours (n=12), fulminant liver failure, live donor graft recipients. Included n=424. Outcome: AKI (KDIGO change SCr) 72 hours post-op. 30 variables including transplant indication, pre-op history, blood parameters & drugs, intra-op data & operation, post-op data. eGFR - abbreviated MDRD. Significant variables univariate analysis entered into forward stepwise regression model. No missing data information.	AKI in 221/424 (52%). Mortality 3/221 (1.4%) with AKI vs. 0/223 without AKI.	4 variables: Female, wt >100 kg, Child-Pugh score & Diabetes. AUC 0.71 (0.66-0.76).	
(386) Xing 2012 **Cancer surgery** **EV** **TRIPOD 4** **(19/31 pts)**	China, dual centre ICU admissions (n=536). 2009-2010. Exclusions: Non-op (n=35), readmissions (n=14) & ICU stay<24 hours (n=70). Included n=449. Outcome: AKIN criteria. No Missing data information.	AKI 10.3% (n=46/449). In-hospital mortality 13% (n=6/46 with AKI vs. 1.5% (n=6/401) without.	Kheterpal AUC 0.66 (0.57-0.74) Plotted Risk range (%): 1pt (6), 2pt (15), 3pt 15() 4pt (33) Abelha AUC 0.51 (0.42–0.59).	

A2.12(iii) CI-AKI Studies

Author & setting Derivation (D) Internal validation (V), External validation (EV) TRIPOD study type & checklist fulfilled	**Population, Outcome, AKI Definitions, Methods**	**Outcome Incidence & Mortality**	**Risk factors in model & Performance**	**Externally validated?**
(379) Freeman 2002 **CI-AKI (PCI)**	USA Multi-centre (8), retrospective analysis of a prospectively collected database (n=16,592). Derivation n=10,729 (1997-9). Exclusions: RRT (n=117), no SCr (n=1019), no contrast dose (n=313), weight (n=48). Included	RRT 0.44% derivation (n=41) & 0.35% (n=19) validation sets.	5 variables: PVD, diabetes, CKD, CCF & cardiogenic shock. c-statistic 0.93. validation set c-statistic 0.89.	

D,V **TRIPOD 2B** **(28/37 pts)**	n=9,242. Validation n=5,863 (1999-2000). Exclusions (n=481): RRT, missing data. Included n=5,382. Outcome: post-procedure RRT (Nephrologist, no indications). CKD = SCr >177 µmol/L. eGFR - Cockcroft & Gault. 11 Variables significant on univariate entered into stepwise multivariate logistic regression.	In-patient Mortality 39% (n=21) RRT vs 1.4% without RRT in derivation. Mortality with RRT 26% in validation.	Risk range in validation set plotted.	
(343) Mehran 2004 **CI-AKI (PCI)** **D,V** **TRIPOD 2A** **(31/37 pts)**	USA single centre, retrospective study, prospectively gathered data (6-year period, dates not given). SCr pre-procedure & 48hrs post (n=8,443). Exclusions (n=86): RRT, contrast <1 week, PCI MI & shock. Outcome: CI-AKI defined = increase ≥25% or ≥44.2 µmol/L 48hrs post-PCI. CKD = SCr ≥133 µmol/L or eGFR <60 (modified MDRD). Bias minimisation: charts reviewed by independent research personnel unaware of study objectives. Random 2:1 split derivation (n=5571) & validation. 16 variables significant on univariate entered into multivariate analysis (n=4,898) - no missing variables. Bootstrapping used.	CI-AKI 13% (729) derivation. CI-AKI (multivariate analysis) 13% (646/4.898) in derivation; 13.9% (386/2,786) Validation. Significant increases in rates of RRT & one-year mortality with increments of risk score (no data given).	8 Variables: Hypotension, IABP, CCF, CKD, age >75, anaemia, diabetes, & contrast media. *c*-statistic 0.69 (SCr) & 0.70 (eGFR). H-L P=0.42-0.43. Validation c-statistic 0.67.	Tziakas 2013 - 0.59 (0.55–0.64) Tziakas 2014 - 0.59 (0.57-0.62) Kul - 0.79 (0.7-0.88) Gao - 0.57 (0.54–0.60) Aykan - 0.68 (0.63-0.74). Sgura - 0.57 (0.52-0.62) Narula 0.60 (0.57–0.63) Lui 2014 - 0.69-0.84. Ando - 0.80 Lui 2015 - 0.80-0.84
(376) **Bartholomew 2004** **CI-AKI (PCI)** **D,V** **TRIPOD 2B** **(22/37 pts)**	USA single centre, retrospective analysis (n=20,479). Exclusion: RRT (n=356), CABG (n=334). Time split derivation n=10,481 (1993-8) & validation n=9,998 (1999-2002) sets. Outcome: >88.4 µmol/L increase SCr baseline. eGFR - Cockcroft & Gault. eGFR 60-89 = mild impairment; <60 at least moderate. 50 variables assessed. No missing data information.	Outcome in 2% (413/20,479). 2.8% (n=293) in derivation vs. 1.2% (n=120) in validation set (p<0.0001). Associated with increased stay & mortality (no data).	8 variables: eGFR <60 ml/min, IABP, urgent or emergency PCI, diabetes, CCF, HTN, PVD, & contrast volume. Derivation c-statistic 0.89. H-L P=0.1. Validation c-statistic & H-L not reported. Risk range in validation set plotted.	Tziakas 2013 - 0.59 (0.54–0.63). Tziakas 2014 - 0.59 (0.56-0.62). Skelding 2007 - 0.86
(377) Marenzi 2004 **CI-AKI (STEMI)** **D** **TRIPOD 1A** **(17/37 pts)**	Italian single centre, retrospective study (n= 218). 2001-3. Inclusion: presenting <12 hrs (or <18 hrs if cardiogenic shock) from symptoms. Exclusions: RRT (n=1), unsuitable for PCI, CABG required (n=2), PCI not indicated (n=7). Included n=208. Outcome: CI-AKI = rise SCr >44.2 µmol/L (time not given). eGFR - Cockcroft-Gault. Indication RRT: oligo-anuria >48 hrs despite >1g furosemide or pulmonary oedema. 24 Variables. No missing data information.	CI-AKI in 19% (40/208). Stay 13 vs. 8 days (p<0.001) & mortality 31% (n=12) vs. 0.6% (n=1) without CI-AKI.	Independent variables: age >75, anterior MI, contrast volume & IABP, time-to-reperfusion, **No AUC.** .	Sgura - 0.57 (0.51-0.62). Narula - 0.62 (0.50–0.65)
(562) Brown 2008	USA multi-centre, retrospective analysis of a prospective	Outcome in 0.7%	7 variables: age ≥80, female, diabetes,	

CI-AKI (PCI) **D,V** **TRIPOD 2B** **(25/37 pts)**	database (n=11,498). 2003-5. Exclusion RRT (n= 357). Included n=11,141. Outcomes: 'serious renal dysfunction' (SRD) = ≥177 µmol/L increase SCr baseline, or ≥50% increase. eGFR – MDRD. 21 variables. Variables in univariate significant at P <0.10 included in multivariate logistic model - backward step approach. Used bootstrapping. No missing data information.	(83/11141). Mortality 19.3% (n=16), vs 0.9% (n=100) without outcome. Stay 13 vs. 3 days without outcome.	CCF, urgent priority & emergent priority, IABP, SCr 115-168 or ≥177 µmol/L. AUC 0.87 (0.82-0.91). H-L P=0.51. Validation AUC 0.84 (0.80-0.89).	
(563) Ghani 2009 **CI-AKI (PCI)** **D,V** **TRIPOD 2A** **(15/37 pts)**	Kuwait, single centre prospective study (n=500). March – May 2005. Exclusions (n=153): RRT, AKI, planned pre or post-procedure RRT, contrast <14 days, complications during PCI. Included n=347 (split n=247 derivation & 100 validation sets). Outcome: CI-AKI = increase SCr >44.2 µmol/L <48 hours. Complete case analysis. 18 variables. Variables in univariate significant included in multivariate logistic model – forward step approach.	CI-AKI in 5.5% (n=13/247) derivation & 5% (n=5) validation. No mortality data.	5 variables: SCr >115 µmol/L, shock, female, multi-vessel PCI, & diabetes. Derivation: H-L P=0.37 (no AUC) AUC in validation 0.61 (no H-L). Risk range in validation set plotted.	
(564) Maioli 2010 **CI-AKI (PCI)** **D,V** **TRIPOD 2A** **(22/37 pts)**	Italian single centre study, retrospective analysis (n=1,384). Excluded (n=166): STEMI (n=119), RRT (n=6), no consent (n=41). 2003-4. Derivation n=1,218. Validation n=502 (with eGFR <60ml/min). eGFR by Cockroft-Gault. Outcome: CI-AKI = SCr increase ≥44.2 µmol/L <5 days. CCF = NYHA III/IV. Continuous variables dichotomised. 14 variables on univariate followed by stepwise multiple logistic regression. Excluded incomplete data (n=70).	CI-AKI in 9.4% (n=114/1218) derivation set. In-hospital mortality 7% (8/114) with CI-AKI vs 0.5% (5/1104) without. OR 17 (95% CI 5-52), P =0.001. CI-AKI in 10.8% (54/502) validation set.	7 variables: age ≥73 (1 pt), diabetes (2pt), LVEF ≤45% (2pt), baseline SCr ≥133 µmol/L (2pt), baseline eGFR ≤44 (2pt), 'post-hydration' SCr ≥ pre-hydration SCr (2pts) & 2nd procedure <72 hrs (3pts). C-statistic 0.85 derivation & 0.82 validation sets. Risk range in validation set plotted.	
(565) Fu 2013 **CI-AKI (PCI)** **D,V** **TRIPOD 2A** **(26/37 pts)**	China, single centre retrospective study, patients aged ≥65 (n =668). 2008-10. Validation: n=277 (same time period). Inclusion: age ≥65, available SCr 48-72hrs post-PCI. Exclusion: incomplete data, malignancy, pulmonary or urinary infections, nephrectomy, CABG or RRT within a week (no numbers reported). Outcome: increase ≥44µmol/L or ≥25% 48-72hrs. eGFR - MDRD. 44 variables. 21 significant on univariate analysis placed into multivariable model.	CI-AKI 15.7% (n=105) of derivation & 16% (n=46) in validation set.	9 variables: eGFR <60, diabetes, LVEF <45%, hypotension, age >70, MI, emergency, anaemia & contrast volume. c statistic 0.79 validation cohort. H-L P-value >0.05. No derivation C-statistic or H-L. Risk range in validation set plotted.	
(378) Tziakas 2013 **CI-AKI (PCI)** **D,V,EV**	Greece single centre, retrospective study of a prospectively collected database (n=509). 2008-10. Exclusions: RRT (n=7), in-hospital death (n=1), CABG (n=2),	CI-AKI in 10% (50/488) of derivation & 14%	5 Variables: CKD, metformin, previous PCI, PVD & contrast volume. c-statistic 0.80 (0.73–0.87). H-L P>0.05.	Validation study 6 centres, 4 countries (same authors) -

TRIPOD 2B & 4 (30/37 pts)	repeat PCI <1 week (n=11). Validated Mehran & Bartholomew scores. Included n=688. n=488 for derivation, n=203 Validation set. Excluded repeat PCI (n=2) RRT (n=1). CI-AKI = ≥25% increase SCr baseline or ≥44.2 µmol/L SCr at 48 hrs. eGFR - Cockcroft–Gault. Variables: 57 demographic, clinical & procedural. Univariate followed by multivariate analysis. Internally validated by bootstrapping. No missing data information.	(28/200) of validation sets. Individual mortality data not given.	Weighted: (CKD 2pts, metformin 2pts, previous PCI 1pt, PVD 2, contrast volume 1) - c-statistic 0.76 (0.72–0.80). H-L P>0.05. No contrast volume c-statistic 0.73 (0.69–0.77) & H-L P>0.05. Bootstrapping corrected 0.75 & without contrast volume - 0.73. Calibration slope 0.88 on bootstrapping. Risk range in validation set plotted. Validation - 0.87 (0.80–0.93) Mehran - 0.59 (0.55–0.64) Bartholomew - 0.59 (0.54–0.63).	0.74 (0.71-0.77).
(566) Gurm 2013 CI-AKI (PCI) **D,V** **TRIPOD 2A** **(32/37 pts)**	USA multi-centre (46), retrospective analysis of a prospectively collected database (n=81,218). 2010-12. Exclusions: RRT (1,897), missing SCr (10,558). Included n=68,573 - split derivation 70% (n=48,001) & validation sets 30% (n= 20,572). Outcome: CI-AKI = rise SCr ≥44µmol/L baseline to peak (<1 week). Secondary outcome: RRT. Variables: 46 including drugs. Random forest regression model trained - determines consensus prediction for each observation by averaging results of individual recursive partitioning tree models. Used network reclassification. Missing data for continuous imputed as median & for categorical variables mode.	CI-AKI in derivation 2.6% (n=1243) & Validation 2.5% (n=505). RRT 0.35% (n=169) & 0.32% (n=66). No mortality data.	46 variables full model: CI-AKI AUC 0.85 (0.84-0.87). RRT 0.88 (0.82-0.93). Reduced (15): indication, PCI status, IHD presentation, shock, CCF <2 weeks, LVEF, Diabetes, Age, weight, height, CK-MB, SCr, Haemoglobin, Trop I, T. - CI-AKI 0.84 (0.82-0.86) & RRT 0.88 (0.82-0.93). Validation: full model - 0.85 (0.84-0.87) & reduced model - 0.84 (0.82-0.86). No H-L. Observed predicted plotted.	
(567) Victor 2014 CI-AKI (PCI) **D,V** **TRIPOD 2A** **(26/37 pts)**	India single centre retrospective study (n=3,152). 2008-11. Exclusions (n=1952): 'non-Indian', missing data, angiogram <14 days of PCI, RRT, AKI pre-PCI, cardiogenic shock, IABP & development PCI related complications. Included n=1,200 (split n=900 derivation & n=300 validation). Outcome = increase SCr ≥44µmol/L or ≥25% at 48hrs. No criteria for RRT. 40 variables; 13 significant in univariate analysis placed in multivariable regression: forward selection.	CI-AKI 9.7% (*n*=87) in derivation & 8.7% (*n*=26) validation sets. RRT in 10.6% (n=12/113). Mortality 6.2% (n=7/113) with CI-AKI vs. n=0 without.	7 variables: contrast load, diabetes with microangiopathy, hypotension, PVD, albuminuria, GFR & anaemia. Derivation AUC 0.93; Validation 0.95. No calibration H-L p-value. Regression plots risks/ incidence derivation vs validation.	
(568) Chen 2014 CI-AKI (PCI) **D,V** **TRIPOD 2A** **(30/37 pts)**	China single centre, retrospective study (n=2,500). Exclusions: RRT, CABG, contrast <14 days. 2009-11. Split derivation (n=1500) & validation (n=1000) sets. Outcome: CI-AKI = increase SCr ≥44µmol/L or ≥25% baseline <5 days post-PCI. Continuous data dichotomised using AUC cut-offs. 20 variables assessed in univariate analysis & those with P	CI-AKI in 16.4% (246/1,500) of derivation & 17.2% (172/1,000) of validation sets.	9 variables: Age ≥70, previous MI, diabetes, hypotension, LVEF ≤40%, anaemia, eGFR<60, HDL <1mmol/L, urgent PCI. *c*-statistic = 0.82 (0.79–0.85). H-L *P*= 0.89. Validation - 0.82 (0.78-0.86).	

	<0.05applied to stepwise multivariate logistic regression model. Bootstrapping used. No information on missing data.		Risk range in validation set plotted.	
Gao (569) 2014 **CI-AKI (PCI)** **D,V,EV** **TRIPOD 2A & 4** **(27/37 pts)**	China single-centre, retrospective analysis (n=3,945). 2005-6, 2010. Exclusions (number not given): missing SCr, RRT, contrast <1 week, or 3 days post, CABG, death <24 hrs; IV hydration. Split derivation (n=2,764) & validation (n=1,181) sets. Included n=2,671 in derivation. Outcome: CI-AKI = ≥44.2µmol/L or ≥25 % baseline <72 hrs. eGFR - MDRD. 24 Variables. 16 Variables with P <0.2 in univariate analysis included in multivariable stepwise logistic regression analysis. No missing data information.	CI-AKI in 4.5 % (173/3852) - 4.6% (n=123) derivation & 4.2% (n=50) validation sets. CI-AKI Mortality 3.4% (n=6) vs 0.3% (n=11) without CI-AKI (P <0.001).	7 Variables: Age >60, HTN, acute MI, CCF, IABP, eGFR (70-89, 50-69, 30-49, <30), contrast volume - AUC 0.76 (0.72–0.80), H-L P=0.50. Validation AUC 0.71 (0.63–0.79), H-L P=0.54. Mehran (n=3,945) -0.57 (0.54–0.60).	
Tsai (346) 2014 **CI-AKI (PCI)** **D,V** **TRIPOD 2A** **(32/37 pts)**	USA mulit-centre, retrospectively analysed prospectively collected database (registry) (n=1,254,089). 2009-11. Exclusions: multiple PCI (32,999), same day discharge (42,029), missing SCr (207,789), RRT (24,260). Included n=947,012. Split derivation (70%, n=662,504) & validation (30%, n=284,508). Outcomes: CI-AKI (AKIN baseline to peak) & RRT. Studied only pre-procedure variables (43) in univariate analysis followed by multivariable logistic regression. Missing data: categorical (<1%) imputed to most common value, continuous variables imputed to median values.	CI-AKI 7.3% (n=48,363) & RRT 0.3% (n=1,988) in derivation & validation sets. Derivation mortality AKI 8.9% (n=4,348) vs 1.1% (n=2,801) without. Mortality not given for validation.	12 Variables CI-AKI: Age, CCF presentation <2 weeks, eGFR, Diabetes, CCF, CVD, NSTEMI, STEMI, Shock, Arrest, Anaemia, IABP. Derivation C-Statistic 0.71 (0.71-0.72) & validation 0.71 (0.71-0.72). Calibration line 1.01. 6 variables for RRT: CCF presentation <2 weeks, GFR, Diabetes, NSTEMI, STEMI, Arrest. C-statistic 0.88. Used integer scoring system. Risk range in validation set plotted.	
(570) Liu 2015 **CI-AKI (chronic total occlusion (CTO) lesions)** **DV, EV** **TRIPOD 2A** **(14/37 pts)**	China, single-centre, retrospectively analysed prospectively collected database (n=728). No exclusion information. Outcome: CI-AKI = increase ≥44µmol/L 48-72hrs. Split derivation (n=494), validation (n=233). 24 variables – those with p <0.20 in univariate analysis, along with other known CI-AKI risk factors entered into multivariate logistic regression by forward stepwise selection. No missing data information.	Outcome CI-AKI in 3.0% (n=15) in derivation set.	3 variables: age ≥75 LVEF <40% & Cr >133 µmol/L. C-statistic 0.79. H-L P value 0.91. Validation C-Statistic 0.86 Mehran AUC 0.80 derivation & 0.843 validation sets. Risk groups: low risk (0 – 0%), moderate (1 - 5.1%), & high risk (≥ 2 - 19.4%) in training. & validation 1%, 9.8%, 10%.	
(571) Skelding 2007 **CI-AKI (PCI)** **EV** **TRIPOD 4** **(21/31pts)**	USA single centre, external validation study of Bartholomew Score (n=5,025) using registry data. 2000-3. Exclusions (n=1,816): CABG, RRT (n=80), lack of consent (n=135); missing SCr (n=1,601). Included n=3,209. Outcome: CI-AKI = increase SCr >88.4 µmol/L (no duration defined). Missing data: multiple imputations.	CI-AKI in 2% (n=61). Stay with CI-AKI 11 vs. 3 days without (*P*= <0.001). In-hospital mortality 6.6% (n=4/61) with CI-	c-statistic = 0.86. No H-L value reported.	

		AKI vs. 1.2% (n=37/3,152) without.		
(572) Sgura 2010 **CI-AKI (STEMI)** **EV** **TRIPOD 4** **(22/31 pts)**	Italian single centre external validation of Mehran & Marenzi scores (n=1046). Consecutive cases, presenting <12 hrs of symptom onset. 2002-8. Exclusions (n=155): RRT (n=18), Cardiogenic shock (n=94), missing data (n=43). Included n=891. Outcome: CI-AKI SCr = increase ≥44 µmol/L or ≥25% baseline <48 hrs. eGFR - MDRD. CKD defined as eGFR <60. Compete case analysis.	CI-AKI in 14.1% (n=126). In-hospital mortality increased with Mehran risk groups – 1% (n=6/562) low risk vs 24% (n=7/29) very high risk (P <0.001).	Mehran AUC 0.57 (0.52-0.62). Marenzi 0.57 (0.51-0.62).	
(573) **Ando 2013** **CI-AKI (STEMI)** **EV** **TRIPOD 4** **(20/31 pts)**	Italian single centre, retrospective external validation of Mehran & AGEF scores (modification of syntax based on age, LVEF & eGFR) (n=507). 2008-11. Inclusion: STEMI <12 hrs. Exclusions (n=26): cancer, chronic inflammatory disease, RRT, monoclonal gammopathy, transplant, contrast exposure <7 days, emergency CABG or STEMI-related mechanical complications & deaths <12 hrs procedure. Included n=481. Outcome: CI-AKI = increase SCr ≥44µmol/L or ≥25% baseline <72 hrs. eGFR - MDRD. No missing data information.	5% (n=25/481) developed CI-AKI.	Mehran AUC 0.80 (0.77-0.84), H-L P=0.77. AGEF AUC 0.88 (0.85-0.91), H-L P=0.25.	
(574) Aykan 2013 **CI-AKI (STEMI)** **EV** **TRIPOD 4** **(20/31 pts)**	Turkey single centre, retrospective validation of Mehran study & Syntax scores (n=497). 2011-12. Inclusion: presenting <12 hrs symptoms undergoing PCI. Exclusions (n=95): RRT, previous MI, PCI, CABG, cancer, recent major surgery & trauma. Included n=402. Outcome: ≥25% increase SCr <72 hrs. eGFR - MDRD. eCKD defined GFR <60.	CI-AKI in 32.6% (n=131).	Mehran - AUC 0.68 (0.63-0.74). Syntax - AUC 0.66 (0.60-0.71).	
Kul (575) 2014 **CI-AKI (STEMI)** **EV** **TRIPOD 4** **(20/31 pts)**	Turkey single centre retrospective validation of Mehran & Zwolle scores (n=314). 2011-12. Exclusions: for CABG, medically managed, previous CABG, eGFR <30, death <48hrs. Outcome: = increase SCr ≥44µmol/L or ≥25% <72 hrs. eGFR – MDRD. No missing data information.	CI-AKI in 12.1% (n=38). Mortality with CI-AKI 23.6% (n=9) vs. 0.3% without (n=1).	Zwolle AUC 0.85 (0.78-0.92) Mehran AUC 0.79 (0.7-0.88) No discrimination reported.	
Liu (576) 2014 **CI-AKI (STEMI)** **EV** **TRIPOD 4** **(20/31 pts)**	China single centre, prospective validation study of Mehran & GRACE scores (n=251 after exclusions). 2010-11. Exclusions (numbers not given): pregnancy, allergy, exposure <7 days, treatment with nephro-protective or nephrotoxic drugs. Outcome: CI-AKI defined as: increase SCr of ≥26.5 or ≥44µmol/L	CI-AKI ≥26.5 - 17% (n=43/251); ≥44µmol/L - 9% (n=22/251), & ≥50% 8%	GRACE AUCs for increase SCr ≥26.5, ≥44µmol/L & ≥50%: 0.72, 0.79 & 0.69 respectively. Mehran: AUCs 0.78, 0.84, 0.69 respectively. No H-L statistic.	

	<48-72 h after contrast exposure, or ≥50% increase. No missing data information.	(n=19/251). Mortality: ≥26.5 - 11.6% (n=5), ≥44µmol/L - 22.7% (n=5), ≥50% 15.8% (n=3). No CI-AKI mortality 2.4% (n=5).		
(577) Narula 2014 CI-AKI (STEMI) EV TRIPOD 4 (21/31pts)	USA multi-centre, retrospective analysis of prospectively collected data from the HORIZONS-AMI trial (n=3,602). 2005-7. Exclusions: prior RRT or missing SCr. Included n=2,968. Outcome: CI-AKI = increase ≥44µmol/L or 25% <48hrs. eGFR - Cockcroft–Gault. Complete case analysis.	CI-AKI in 16% (479/2,968).	Mehran AUC 0.60 (0.57–0.63), Marenzi AUC 0.62 (0.50–0.65). No H-L statistic. Risk range for Mehran: low (12%), moderate (15%), high (24%), very high (37%); Marenzi: 0 - 10%, 1 - 14%, 2 - 21%, 3 - 28%, 4/5 - 50%	
(347) Tziakas 2014 CI-AKI EV TRIPOD 4 (22/31 pts)	Greece multi-centre (6 centres, 4 countries) prospective study (n=2,882). 2010-12. Exclusions: RRT (n=26), deaths <48hrs (n 23), CABG (n=51), missing SCr (n=53) & repeat revascularization (n=40) <1 week. Included n=2,689. Outcome: CI-AKI = increase SCr ≥44µmol/L or ≥25% baseline <48 hrs post-PCI. Validated Tziakas score & additive predictive value of using reclassification tables, net reclassification improvement, & integrated discriminative improvement. eGFR - Cockcroft–Gault. No missing data information.	CI-AKI in 15.7% (423/2,689). No mortality data.	Tziakas 2013 - AUC 0.74, (0.71-0.77). H-L P=0.184. Calibration slope 0.96. Validation analysis of Mehran score – AUC 0.59 (0.57-0.62) & the Bartholomew score – AUC 0.59 (0.56-0.62).	

A2.13 – predictors included in the specialist models

Variable	Cardiac	Non-cardiac specialist surgical	CI-AKI	Total
CKD or admission SCr	14	8	11	33
Diabetes	10	5	12	27
Age	11	2	8	21
CCF	11	1	10	22
Intra-op events	10	7		17
IABP	6		6	12
PVD	6		4	10
HTN	6	1	2	9
Type of Surgery	8	1		9
Hypotension / Cardiogenic Shock			8	8
Gender	4	2	2	8
Previous Cardiac Surgery	8			8
Contrast volume/type			7	7
BMI		3	2	5
MI / location of MI			5	5
Anaemia			5	5
Chronic respiratory disease	4			4
Urgent/Emergency			4	4
2nd procedure			2	2
Recent CCF admission			2	2
CVD		2		2
Bilirubin	1	1		2
time to intervention			1	1
multi-vessel disease			1	1
pre post hydration sCr			1	1
CK-MB, Troponin			1	1
Previous MI			1	1
HDL			1	1
CVD			1	1
INR		1		1
Medication			1	1
Proteinuria			1	1

Appendix 3 – external validation

3A.1. Acute kidney injury prediction score (APS)
3A.2 Summary of key variable between validation (medicine with baseline SCr) and original derivation study.

Appendix Table 3A.1. Acute kidney injury prediction score (APS)

	Points Scored			
	0	**1**	**2**	**3**
Age (years)	**<60**		**60-79**	**≥80**
Respiratory Rate	**<20**	**≥20**		
AVPU Score	**Alert**			**Other**
CKD Stage 3a-5		**Y**		
Heart failure		**Y**		
Diabetes			**Y**	
Liver disease				**Y**

AVPU scale best response (alert, vocal, pain, unresponsive), CKD – chronic kidney disease (eGFR <60mls/min), Respiratory rates - breaths/minute.

Appendix Table 3A.2 Summary of key variable between validation (medicine with baseline SCr) and original derivation study.

	Validation		Derivation Study	
Variable	HA-AKI	No HA-AKI	HA-AKI	No HA-AKI
Age	84 (76-89)	79 (67-86)	80 (70-86)	73 (61-81)
Diabetes	25%	21%	16%	8%
Heart failure	37%	18%	20%	8%
CKD	29%	10%	49%	28%
Liver disease	**2%**	**1%**	**4%**	**2%**
AVPU <Alert	1.7%	1%	4%	1%
RR ≥20	28%	23%	37%	22%

CKD – chronic kidney disease, HA-AKI – hospital-acquired AKI, RR – respiratory rate breaths per minute.

Appendix 4 Impact analysis

Table A4.1 – actions initiated at intervention site vs control site.
Table A4.2A/B – Breakdown AKI at admission, AMBER (APS ≥5 points) and GREEN (APS <5 points): Mortality, escalation and progression of AKI; Demographics, past medical history, physiological observations
Table A4.3 – process changes at the intervention site

Table A4.1 Differences between Intervention & Control sites post intervention

	Intervention Site	Control Site
CA-AKI	• Red flag on Ward view, Observation chart & Patientrack reports • Care bundle of actions • Visible to Intensive Care Outreach team with aim to review <24hrs • Pharmacy awareness • Red flag automatically brought into e-discharge summary	Red flag on Patientrack report
Amber risk group	• Amber flag on Ward view, Observation chart & Patientrack reports • Care bundle of actions • Visible to Intensive Care Outreach team with review where appropriate • Pharmacy awareness	Usual Care
HA-AKI	• Red flag on Ward view, Observation chart & Patientrack reports • Care bundle of actions • Visible to Intensive Care Outreach team with aim to review <24hrs • Pharmacy awareness • Red flag automatically brought into e-discharge summary	
Education	• Mandatory training - E-learning • Training & education on AKI & the CPR to junior doctors (lectures), Nurses, HCAs, Pharmacy (ward-based)	Mandatory training - E-learning

AMBER - score of ≥5 on AKI prediction score (APS), CPR – clinical prediction rule, HCA - health care assistant.

	CA-AKI						AMBER (APS ≥5)						GREEN (APS <5)					
	Intervention Site			Control Site			Intervention Site			Control Site			Intervention Site			Control Site		
	Before (n=670)	After (n=755)	P-value, OR (95% CI)	Before (n=491)	After (n=586)	P-value, OR (95% CI)	Before (n=2,057)	After (n=2,351)	P-value, OR (95% CI)	Before (n=1,810)	After (n=1,851)	P-value, OR (95% CI)	Before (n=4,805)	After (n=5,533)	P-value, OR (95% CI)	Before (n=4,448)	After (n=4,941)	P-value, OR (95% CI)
In-patient mortality	23%	23%*	P=0·95, 1·01 (0·79-1·29)	19%	17%	P=0·34, 0·86 (0·63-1·17)	14%*	11%**	P=0·008, 0·78 (0·66-0·94)	10%	10%	P=0·742, 0·96 (0·78-1·20)	6%*	6%*	P=0·935, 0·99 (0·84-1·16)	5%	5%	P=0·436, 1·08 (0·89-1·31)
7 day mortality	12%	12%*	P=0·87, 0·97 (0·71-1·34)	9%	7%	P=0·219, 0·75 (0·48-1·16)	6%*	4%	P=0·03 0·74 (0·56-0·97)	4%	4%	P=0·489 0·88 (0·63-1·24)	3%*	2%	P=0·437 0·90 (0·70-1·16)	2%	2%	P=0·432, 1·14 (0·83-1·56)
ICU escalation	7%	8%	P=0·92, 1·04 (0·07-1·54)	8%	7%	P=0·484, 0·85 (0·54-1·34)	2%	1%	P=0·08 0·67 (0·42-1·05)	2%	2%	P=1·0, 0·98 (0·59-1·63)	2%	3%*	P<0·001 1·68 (1·31-2·17)	2%	2%	P=0·719, 0·95 (0·71-1·26)
HA-AKI							14%	14%	P=1·0 0·97 (0·82-1·18)	13%	12%	P=0·446, 0·92 (0·76-1·12)	6%*	5%	P=0·103 0·87 (0·73-1·03)	4%	5%	P=0·122, 1·17 (0·96-1·43)
HA-AKI in-patient mortality							31%*	25%	P=0·067 0·71 (0·49-1·01)	28%	27%	P=1·0, 0·98 (0·65-1·49)	24%	19%	P=0·154, 0·74 (0·50-1·11)	17%	22%	P=0·178, 1·41 (0·86-2·31)
HA-AKI 7-day mortality							18%	11%**	P=0·020 0·57 (0·36-0·91)	10%	14%	P=0·113, 1·49 (0·84-2·64)	14%*	10%	P=0·096, 0·64 (0·39-1·07)	8%	11%	P=0·315, 1·49 (0·75-2·94)
HA-AKI ICU escalation							9%	4%**	P=0·027, 0·45 (0·23-0·91)	4%	8%	P=0·105, 2·01 (0·90-4·75)	6%	10%	P=0·127, 1·67 (0·90-3·07)	11%	13%	P=0·549, 1·23 (0·68-2·23)
Increase to AKI stage 3	5%	5%	P=0·6, 0·86 (0·51-1·44)	6%	5%	P=0·486, 0·82 (0·47-1·41)	9%*	6%	P=0·267, 0·68 (0·36-1·26)	3%	5%	P=0·22, 1·98 (0·72-5·44)	8%	7%	P=0·756, 0·90 (0·49-1·66)	6%	6%	P=1·0, 0·98 (0·44-2·22)
Peak increase SCr	0 (0-19)	0 (0-11)	P=0·204	0 (0-10)	0 (0-13)	P=0.744	51 (32-87)	44 (29-74)**	P=0·026	46 (31-78)	51 (34-78)	P=0·265	46 (30-72)	45 (30-76)	P=0.981	47 (32-80)	46 (31-72)	P=0.765

Appendix Table 4.2A - Mortality, escalation and progression of AKI.

	CA-AKI						AMBER (APS 5+)						GREEN (APS <5)					
	Intervention Site			Control Site			Intervention Site			Control Site			Intervention Site			Control Site		
	Before (n=670)	After (n=755)	P-value, OR (95% CI)	Before (n=491)	After (n=586)	P-value, OR (95% CI)	Before (n=2,057)	After (n=2,351)	P-value, OR (95% CI)	Before (n=1,810)	After (n=1,851)	P-value, OR (95% CI)	Before (n=4,805)	After (n=5,530)	P-value, OR (95% CI)	Before (n=4,448)	After (n=4,941)	P-value, OR (95% CI)
Age	80 (70-88)	81 (70-87)	P=0.971	81 (70-87)	80 (71-87)	P=0.848	85 (78-90)	85 (78-90)	P=0.178	84 (78-89)	84 (77-89)	P=0.312	76 (61-85)*	72 (58-85)**	P=0.001	75 (61-84)	75 (60-84)	P=0.620
LOS	9 (4-18)	8 (4-17)	P=0.356	8 (3-16)	8 (3-16)	P=0.823	9 (4-19)*	9 (4-18)*	P=0.346	8 (4-17)	8 (3-15)**	P=0.02	6 (3-15)*	6 (2-14)**	P=0.012	5 (2-12)	5 (2-11)***	P=0.001
CCF	33%	33%	P=0·866, 0·98 (0·79-1·22)	32%	33%	P=0·845, 1·03 (0·80-1·33)	54%	54%	P=0·954, 1·0 (0·89-1·13)	52%	54%	P=0·260, 1·08 (0·95-1·23)	10%	11%	P=0·50 1·05 (0·92-1·19)	10%	11%	P=0·417, 1·06 (0·93-1·21)
DM	33%	35%*	P=0·401, 1·10 (0·88-1·37)	35%	30%**	P=0·05, 0·77 (0·60-0·99)	56%	56%	P=0·692, 1·03 (0·91-1·16)	57%	60%*	P=0·094, 1·12 (0·98-1·28)	7%	8%	P=0·001 1·28 (1·10-1·48)	7%	9%**	P=0·004, 1·25 (1·07-1·45)
LD	4%	5%*	P=0·452, 1·125 (0·76-2·305)	3%	2%	P=0·430, 0·712 (0·33-1·56)	6%	7%	P=0·092, 1·24 (0·97-1·59)	5%	6%	P=0·196, 1·23 (0·91-1·67)	1%*	1%*	P=0·928 1·02 (0·72-1·45)	1%	1%	P=0·462, 0·82 (0·51-1·33)
HTN	72%	72%*	P=0·906, 1·02 (0·81-1·29)	71%	63%**	P=0·008, 0·71 (0·55-0·91)	80%	79%	P=0·43, 0·94 (0·81-1·09)	79%	78%	P=0·472, 0·94 (0·81-1·10)	53%*	49%*	P<0·001 0·87 (0·81-0·94)	47%	46%	P=0·443, 0·97 (0·89-1·05)
Vascular	13%	15%	P=0·492, 1·11 (0·82-1·50)	10%	12%	P=0·436, 1·18 (0·80-1·73)	15%*	16%*	P=0·528, 1·06 (0·90-1·25)	11%	12%	P=0·145, 1·16 (0·95-1·43)	8%*	8%*	P=0·394 0·94 (0·81-1·08)	5%	5%**	P=0·04, 1·22 (1·01-1·48)
CKD	83%	87%	P=0·102, 1·29 (0·96-1·72)	85%	87%	P=0·428, 1·15 (0·82-1·63)	79%	79%	P=0·768, 1·02 (0·89-1·18)	83%*	78%**	P=0·001, 0·75 (0·64-0·89)	30%	27%	P=0·01 0·89 (0·82-0·97)	31%*	29%**	P=0·007, 0·89 (0·81-0·97)
NEWS	2 (1-4)	2 (1-4)	P=0.022	2 (1-4)	2 (1-4)	P=0.796	2 (1-4)	2 (0-4)**	P=0.032	2 (1-4)	2 (0-4)**	P=0.05	1 (0-3)	1 (0-3)	P=0.342	1 (0-3)	1 (0-3)	P=0.109
SBP <90	8%	8%	P=1·0, 1·01 (0·68-1·49)	13%*	10%	P=0·250, 0·79 (0·54-1·15)	3%	3%	P=0·325, 0·83 (0·59-1·18)	5%*	4%*	P=0·285, 0·83 (0·60-1·15)	2%	2%	P=0·78 0·95 (0·73-1·26)	3%*	3%*	P=0·26, 0·87 (0·69-1·10)
RR >20	32%*	30%*	P=0·358, 0·90 (0·72-1·13)	24%	23%	P=0·666, 0·93 (0·71-1·24)	46%*	41%* **	P=0·001, 0·81 (0·72-0·92)	34%	32%	P=0·277, 0·93 (0·81-1·06)	19%*	19%*	P=0·597 1·03 (0·93-1·13)	14%	12%**	P=0·015, 0·86 (0·76-0·97)
<A on AVPU	3%	2%	P=0·397, 0·74 (0·38-1·46)	2%	3%	P=0·686, 1·26 (0·56-2·84)	3%	3%	P=0·671, 1·08 (0·78-1·51)	4%	3%	P=0·52, 0·88 (0·62-1·27)	0.30%	0.20%	P=0·85 0·87 (0·40-1·87)	0%	0%	P=0·249, 0·45 (0·14-1·49)

Appendix Table 4.2B. Demographics, past medical history, physiological observations

Color code red = CA-AKI, Amber – APS ≥5 points flagged at risk of AKI, Green APS <5 points. SCr – serum creatinine. * = significant (P<0.05) difference between sites during the same period, ** = significant (P<0.05) difference between periods (pre vs post intervention) at the same site. Median (interquartile range). HA-AKI – hospital-acquired AKI, ICU – intensive care unit, LOS – length of stay, SCr – serum creatinine, Stage 3 – KDIGO staging x3 increase SCr.

Appendix Table 4.3 – process changes at the intervention site

Prescribing practice*	• At the intervention site significantly more episodes of medications were stopped with AKI as the specific reason vs the control site (n=95 vs 36, P<0·001, OR 2·1 [1·41-3·10]. • Significantly more patients at the intervention site had ACEi/ARBs stopped vs control site (308 vs 146, P<0·001, OR 1·831 [1·493-2·247])
Bundle completion	• For patients at risk of AKI on the APS (AMBER) and those with AKI a care bundle was recommended. Analysis was available for 7 of the 10 months following intervention with 15% of AMBER bundles (256/1,688 episodes) completed and 26% of bundles (258/983 episodes) completed for patients with AKI.
Coding of AKI	• ICD-10 coding of AKI increased at both sites with the intervention site having a larger increase (+21% vs +17%). • Coded AKI mortality at the intervention site significantly reduced (24% vs 20%, P=0·0057) with no significant change at the control site (20% vs 18% P=0·44).
Notes review of processes documented	• Random sample of patients who developed HA-AKI independently audited against recommended standards (documentation, investigations, management) • Significant increase in documentation of AKI in notes and on discharge and medication review at intervention site (see Table)

*Introduction of electronic prescribing in the second period of the study (both sites), allowed for limited interrogation of prescribing behavior.