BMJ Open

Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations

Luke Eliot Hodgson,[1] Alexander Sarnowski,[2] Paul J Roderick,[1] Borislav D Dimitrov,[1] Richard M Venn,[3] Lui G Forni[2,4]

[1]Academic Unit of Primary Care and Population Sciences, Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton, UK
[2]Intensive Care Department, The Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK
[3]Anaesthetics Department, Western Sussex Hospitals NHS Foundation Trust, Worthing Hospital, Worthing, UK
[4]Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

Correspondence to
Dr Luke Eliot Hodgson;
drlhodgson@gmail.com

ABSTRACT

Objective Critically appraise prediction models for hospital-acquired acute kidney injury (HA-AKI) in general populations.

Design Systematic review.

Data sources Medline, Embase and Web of Science until November 2016.

Eligibility Studies describing development of a multivariable model for predicting HA-AKI in non-specialised adult hospital populations. Published guidance followed for data extraction reporting and appraisal.

Results 14 046 references were screened. Of 53 HA-AKI prediction models, 11 met inclusion criteria (general medicine and/or surgery populations, 474 478 patient episodes) and five externally validated. The most common predictors were age (n=9 models), diabetes (5), admission serum creatinine (SCr) (5), chronic kidney disease (CKD) (4), drugs (diuretics (4) and/or ACE inhibitors/angiotensin-receptor blockers (3)), bicarbonate and heart failure (4 models each). Heterogeneity was identified for outcome definition. Deficiencies in reporting included handling of predictors, missing data and sample size. Admission SCr was frequently taken to represent baseline renal function. Most models were considered at high risk of bias. Area under the receiver operating characteristic curves to predict HA-AKI ranged 0.71–0.80 in derivation (reported in 8/11 studies), 0.66–0.80 for internal validation studies (n=7) and 0.65–0.71 in five external validations. For calibration, the Hosmer-Lemeshow test or a calibration plot was provided in 4/11 derivations, 3/11 internal and 3/5 external validations. A minority of the models allow easy bedside calculation and potential electronic automation. No impact analysis studies were found.

Conclusions AKI prediction models may help address shortcomings in risk assessment; however, in general hospital populations, few have external validation. Similar predictors reflect an elderly demographic with chronic comorbidities. Reporting deficiencies mirrors prediction research more broadly, with handling of SCr (baseline function and use as a predictor) a concern. Future research should focus on validation, exploration of electronic linkage and impact analysis. The latter could combine a prediction model with AKI alerting to address prevention and early recognition of evolving AKI.

Strengths and limitations of this study

► This is the first systematic review of prediction models for hospital-acquired acute kidney injury (AKI) in general hospital populations who account for the majority of hospital admissions and AKI cases.

► The models were selected following an extensive literature search; the review followed the latest critical appraisal guidance and assessed validity of the models in terms of risk of bias and applicability, highlighting important shortcomings such as handling of serum creatinine.

► The large number of patient episodes provides important insights into AKI prediction and complements other recent reviews in specialised areas (cardiac surgery, contrast-induced AKI and liver transplantation).

► Lack of access to individual participant data prevented a meta-analysis of the studies, an avenue of future research.

► The small number of externally validated models and absence of impact analysis limit the recommendation and implementation of an individual model.

INTRODUCTION

Acute kidney injury (AKI) is defined as an acute increase in serum creatinine (SCr) or reduction in urine volume.[1] The incidence of AKI is increasing, affecting up to one in five hospitalised adults worldwide.[2] A continuum of injury exists long before sufficient loss of excretory kidney function can be measured with standard laboratory tests (ie, SCr).[3 4] Associated mortality remains high, in part reflecting the severity of the underlying disease, but may also be due to the limitations of conventional markers to detect early injury.[5]

Deficits in recognition and management of patients with AKI[6] have led to practice guidance calling for improved risk assessment, at which point interventions could be most beneficial.[7] One suggested strategy to achieve

this aim is through the implementation of clinical prediction models.[8 9] Though development and validation of AKI prediction models is desirable,[7 10] clinical application in this and other fields has been hampered for a number of reasons:

▶ potential predictors and models continuously increase with new studies often finding conflicting results[11]
▶ substandard reporting of methodology and results make conclusions problematic[12 13]
▶ few general hospital population studies exist; specialist fields (cardiac and transplant surgery and contrast-induced (CI-AKI)) account for the majority of AKI models and all systematic reviews but are unlikely to be generalisable and[14–17]
▶ models rarely enable electronic automation as part of clinical workflow, known to influence uptake.[18]

High-quality systematic reviews of prediction models have been called for.[19] Following recent reporting guidance (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS)[20] and Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD)),[12] this review appraises hospital-acquired AKI (HA-AKI) prediction models in general populations who, in the UK, account for the majority of hospital episodes[21] and AKI cases.[22 23]

METHODS
Published guidance (CHARMS, TRIPOD and Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)) helped frame the review question, data extraction, reporting and appraisal.[12 20 24] The research question was: what are the available prognostic prediction models for the development of HA-AKI in adult general populations? Using explicit, systematic methods to minimise bias and provide reliable findings from which conclusions can be drawn and decisions made,[25 26] the review aimed to collate empirical evidence for AKI prediction models across general hospital settings, fitting prespecified eligibility criteria (online supplementary table 1). Performance was assessed by discrimination and calibration including validation studies. The presence of any impact analysis studies was also investigated. The review aimed to provide recommendations for the most robust, usable models, including the ability to incorporate future electronic data linkage, for example, between the community (primary care) and hospital.

Data sources, study selection and data extraction
We searched MEDLINE, Embase and Web of Science databases (inception to November 2016) using recommended filters (online supplementary tables 2–4).[27 28] Titles and abstracts were screened by two reviewers (LEH and AS), and full articles were reviewed if eligible. Disagreements were resolved by iterative screening rounds. Reference lists from retrieved articles, systematic reviews, national[7] and international guidance[1] and our own literature files

were also analysed. Data extraction and quality assessment were performed by two investigators (LEH and AS) with disagreements resolved by a third reviewer (LGF). A data extraction form was used based on previous reviews and guidance (summary online supplementary table 5).[12 13 20] Items extracted included design (eg, cohort and case–control), population, location, outcome (definition duration of follow-up and blinding of assessment), modelling method (eg, logistic), method of internal validation (eg, bootstrapping), number of participants and events, number and type of predictors, model presentation and predictive performance (calibration and discrimination). The presence of external validation was recorded.

Outcome, model performance and clinical utility
It was anticipated that study outcome, HA-AKI, would vary given the numerous definitions in use prior to Kidney Disease: Improving Global Outcomes (KDIGO) in 2012.[1] Thus, during the search strategy, we included studies with an SCr around admission and repeated during a hospital admission to diagnose HA-AKI. Information was gathered on how a study defined a patients baseline renal function, how community AKI cases were handled, whether SCr was used as a predictor in analysis and finally the magnitude and timeframe used to define the outcome. Discrimination and calibration are the most common methods to assess model performance. Discrimination is usually assessed graphically by the area under the receiver operating characteristic curve (AUROC), representing how well a model separates and ranks patients who experienced the outcome from those who did not. For prediction models, the AUROC, which focuses solely on accuracy, has a number of shortcomings, such as a lack of information on consequences and when used in populations where the outcome prevalence is rare.[29 30] Calibration describes how well predicted results agree with observed results.[12 30] The Hosmer-Lemeshow (H-L) test, despite limitations, is the most commonly used calibration statistic.[31 32] It is also recommended to graphically plot expected and actual outcomes, for example, with a calibration slope.[12] In addition to performance, ease of bedside use and whether the models could be electronically automated—factors known to influence successful uptake—were recorded.[18] 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.

Study quality assessment
A global TRIPOD score for each study was calculated to quantify reporting, consisting of the sum of the scores for each individual item (out of a maximum 37, with a score of 1 for criterion met, score of 0 for each item not met or unclear).[12] As yet there has been no suggested cut-off for what represents a high-quality study, though it would be reasonable to judge that those studies with the most significant gaps in reporting are likely to be

at higher risk of bias. Furthermore, the quality (risk of bias) of each study was assessed by piloting a version of Prediction study Risk Of Bias Assessment Tool (PROBAST), a tool for assessing risk of bias and applicability of prognostic model studies, 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: study participants, predictors, outcome, sample size and missing data, statistical analysis and overall judgement of bias and applicability.
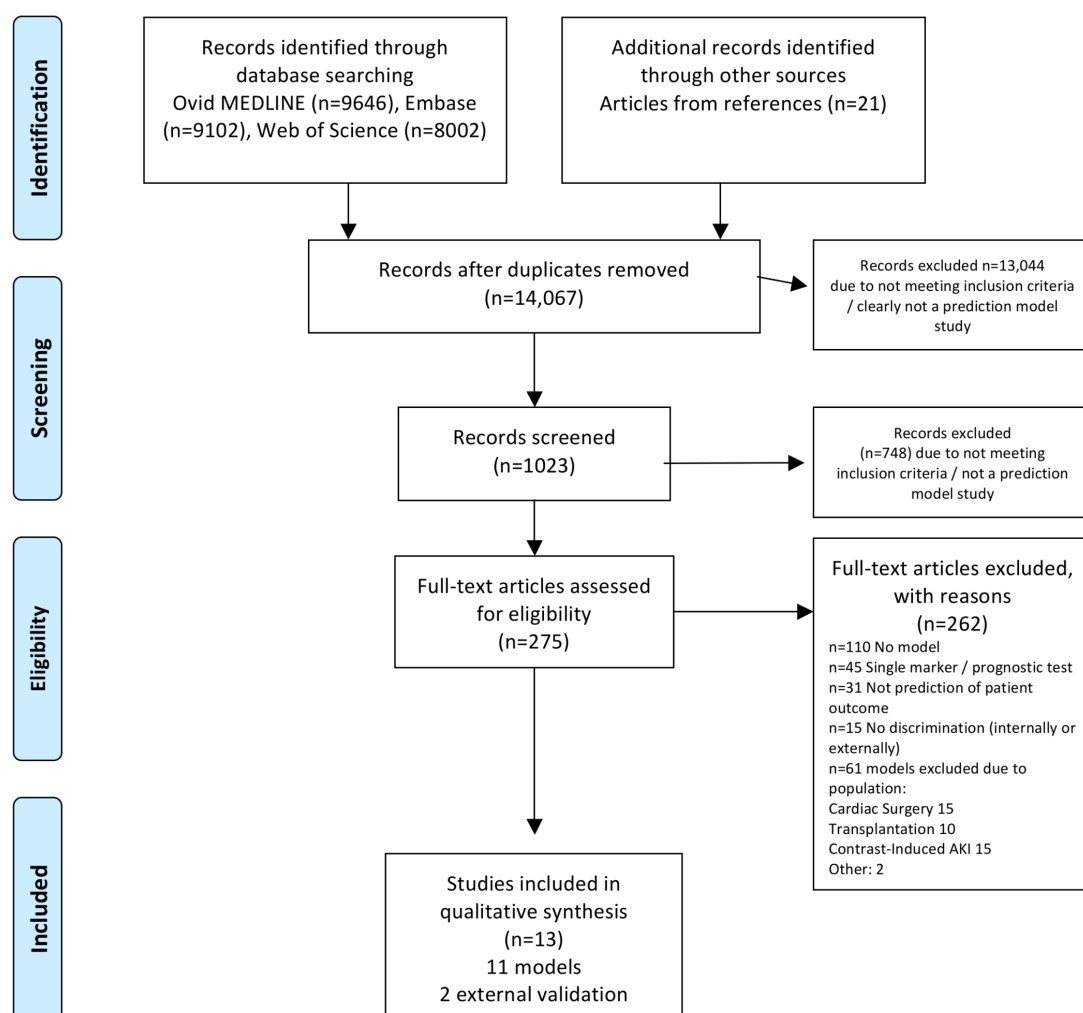
Patient involvement

Patients were not involved in setting the research question, outcome, design and implementation of the study. There are no plans to involve patients in dissemination.

RESULTS

From 14046 articles identified by the search strategy, 254 full articles were reviewed (PRISMA flow chart, figure 1). Specialised fields (predominantly cardiac surgery, transplantation or CI-AKI) accounted for 61 of 74 (82%) of all studies. This review included 11 general model studies (n=474478 patient episodes), in general surgery,[33 34] trauma and orthopaedics (T&O),[35] general hospital cohorts (predominantly medicine and surgery)[36–40] and heart failure (summarised in table 1 and online supplementary etable 6, with abbreviations in supplementary etable 7).[41–43] Two further studies were purely external validations.[44 45] HA-AKI incidence was 7% (21641 events), though this varied from <1% in the general surgery models[33 34] to 28% across the heart failure studies, and heterogeneous definitions (timeframe and marker) were employed (see table 1 for definitions used with further information in online supplementary etable 6).

PRISMA 2009 Flow Diagram

Identification

Records identified through database searching
Ovid MEDLINE (n=9646), Embase (n=9102), Web of Science (n=8002)

Additional records identified through other sources
Articles from references (n=21)

Screening

Records after duplicates removed
(n=14,067)

Records excluded n=13,044 due to not meeting inclusion criteria / clearly not a prediction model study

Records screened
(n=1023)

Records excluded
(n=748) due to not meeting inclusion criteria / not a prediction model study

Eligibility

Full-text articles assessed for eligibility
(n=275)

Full-text articles excluded, with reasons
(n=262)
n=110 No model
n=45 Single marker / prognostic test
n=31 Not prediction of patient outcome
n=15 No discrimination (internally or externally)
n=61 models excluded due to population:
Cardiac Surgery 15
Transplantation 10
Contrast-Induced AKI 15
Other: 2

Included

Studies included in qualitative synthesis
(n=13)
11 models
2 external validation

Figure 1 PRISMA study flow chart.[24] AKI, acute kidney injury; PRISMA, Preferred Reporting Items for Systematic Review and Meta-Analysis.

Table 1 Summary of HA-AKI prediction models

Population	General surgery		T&O	General (medical and surgical)					Heart failure		
Author, year (n=derivation)	Kheterpal et al, 2007[33] (n=14 066)	Kheterpal et al, 2009[34] (n=57 080)	Bell et al, 2015[35] (n=6220)	Drawz et al, 2008[38] (n=360)	Matheny et al, 2010[37] (n=26 107)	Koyner et al, 2016[40] (n=202 961)	Bedford et al, 2016[39] (n=7556)	Forni et al, 2013[36] (n=1867)	Forman et al, 2004[43] (n=1004)	Breidthardt et al, 2011[41] (n=657)	Wang et al, 2013[42] (n=1010)
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 ≥177 µmol/L, RRT (30 day)	KDIGO ↑SCr	↑SCr*	RIFLE ↑SCr	KDIGO ↑SCr (24 hours)	KDIGO ↑SCr (72 hours)	KDIGO ↑SCr (<7 days)	↑SCr >26.5 µmol/l†	↑SCr >26.5 µmol/l†	AKIN SCr (<48 hours)
Events	121	561	672	120	1352	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 reported	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

*Increase sCr ≥44 µmol/L if baseline SCr of ≤168 µmol/L, ≥88 µmol/L baseline 177–433 µmol/L and ≥133 µmol/L baseline >442 µmol/L.
†During admission.
‡Propensity matched.
§Validations in medicine/surgery with/without baseline SCr.
¶Used linked community and hospital data.
AKIN, Acute Kidney Injury Network; AUROC, area under the receiver operating characteristic curve; CC, case–control; eGFR, estimated glomerular filtration rate; EPP, events per predictor; EV, external validation; IV, internal validation; HA-AKI, hospital-acquired acute kidney injury; H-L, Hosmer-Lemeshow test; KDIGO, Kidney Disease: Improving Global Outcomes; mortality, in-hospital; P, prospective; plot, calibration plot; R, retrospective; RIFLE, Risk, Injury, failure, loss of kidney function; RR, risk range; RRT, renal replacement therapy; SCr, serum creatinine; T&O, trauma and orthopaedics; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis.how many of the 37 recommended items were reported.

Table 2 Summary of limitations in methodology and reporting

Area of concern	Description
Missing data	Multiple imputation recommended to avoid bias—rarely described.[12 50]
Definitions of outcome and predictors	No consistent strategy used to differentiate CA-AKI from HA-AKI. Two studies excluded patients with pre-existing CKD[33 37]; five studies took admission SCr as baseline; five included SCr as predictor despite it forming the outcome; comorbidities inconsistently defined: including from admission diagnoses or coded history.
Blinding of predictors or outcome	Not reported.
Sample size	Calculations not described, six studies had <10 EPP. Small sample increases risk of overfitting and underfitting.[12 49]
Univariate to select for multivariate analysis	Technique not recommended, used in 10 of 11 models.[12]
Bootstrapping	Adjust for optimism, without losing information—rarely described.[35 43]
Calibration plots	Important part of model performance,[12] present in only one model and one external validation.[35 45]
External validation and model updating	Validation adjusts for optimism, assesses generalisability but was scarce, while model updating is recommended but not described.[12]
Newer performance measures	Techniques such as decision curve analysis offer insight into clinical consequences—not described.[29]
Use of data linkage	Only one study used data linkage.[35]

AKI, acute kidney injury; CA-AKI, community-acquired AKI; CKD, chronic kidney disease; EPP, events per predictor; HA-AKI, hospital-acquired AKI; SCr, serum creatinine.

For example, five studies took admission SCr to represent a patients baseline, potentially confusing CKD, established and emerging AKI.[34 38 40 42 43] Of note, one study produced a model to predict AKI admission as well as HA-AKI at 72 hours with the former not considered suitable for analysis in this review.[39]

In seven of the nine studies reporting age, this was significantly higher in the group with the outcome, with eight studies reporting a mean or median age over 65 years in the outcome group (table 1). Mortality was significantly higher in those who developed the outcome in the six studies where data were available (ranging 6%–42%). No impact analyses were retrieved.

Study reporting
A median 28 (IQR 25–30) of 37 recommended items were reported, suggesting significant shortcomings (key shortcomings are summarised in table 2 with TRIPOD reporting summarised in online supplementary etable 8). By design, eight studies were retrospective, two were prospective and one was a case control. Five studies were single centre. USA (n=6) and UK (n=3) accounted for the majority of the models. Only three studies used imputation techniques for missing data.[34 35 38] Definitions were heterogeneous (table 1) with five using Risk, Injury, failure, loss of kidney function (RIFLE),[37] Acute Kidney Injury Network (AKIN)[42] or KDIGO criteria for changes in SCr.[35 36 39] One study used KDIGO SCr change within a 24-hour timeframe of predictors being measured.[40]

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 assessment of predictors and study outcome was not mentioned. Continuous predictors were dichotomised in four studies, and 10 studies used univariate analysis to select for multivariate analysis. No models mentioned shrinkage techniques or sample size calculations. Median number of outcome events was 271 (121–672). For statistical power, all of the studies had more than 10 events per predictor (EPP) *included in the model*. However, the EPP was <10 in six studies, when accounting for the total number of candidate predictors assessed.[33 36 38 39 41 43] Of a total of 56 different predictors, a median of 7 (7–12) was included per model, including demographics, history, procedure information, laboratory parameters, physiological observations and hospital admission diagnoses (most common presented in figure 2, full details onlinesupplementary etables 9–10). Only four studies included physiological parameters in their final model.[36 38 40 43] Seven studies included admission SCr as potential predictor with five including this in the final model, thus potentially confusing prediction with a diagnosis of AKI.[34 37 40 42 43] 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 online supplementary etable 11.

Model performance
Median AUROC (or C-Statistic) was 0.745 (range 0.71–0.80) for derivation (eight studies) and 0.74 (range 0.66–0.80) for internal validations reporting discrimination (seven studies). Excluding the studies using non-consensus-based definitions and those including admission
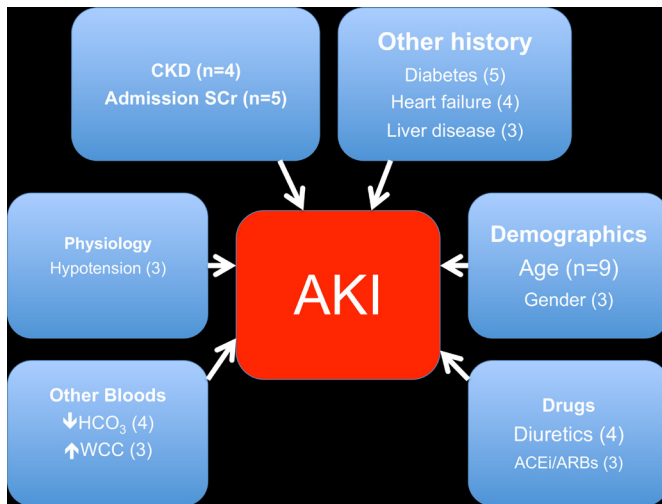
Figure 2 Predictors most frequently included in the 11 HA-AKI prediction models. ACEi, ACE inhibitors; ARBs, angiotensin-receptor blockers; Bloods, laboratory parameters; CKD, chronic kidney disease; HA-AKI, hospital-acquired acute kidney injury; ↓HCO$_3$, reduced serum bicarbonate; SCr, serum creatinine; ↑WCC, raised white cell count.

SCr as predictor and/or baseline, only four studies were left.[35 36 39 41] In these studies, AUROCs ranged 0.71–0.74 in derivation (three studies), 0.67–0.76 for internal validation (three studies) and 0.65–0.71 for external validation (three studies). Only one model study presented a calibration plot for derivation and validation.[35] The H-L statistic was used in three derivations[36 37 42] and two internal validations.[39 42]

Five models have been externally validated: on separate populations within the same study,[35 39] other model studies[43] or stand alone external validations,[33 45] where the AUROCs were moderate, ranging 0.65–0.71. One validation provided a calibration plot,[35] one the H-L statistic[39] and one reported both.[45] In the Bell external validation cohort calibration suggested the model overpredicted the outcome requiring recalibration.[35] In the external validation of the Forni study calibration plots showed agreement at low probability rates, while at higher rates calibration deviated in the medical cohort.[45] Two of the three surgical models have been externally validated: the Kheterpal model,[33] in a Chinese population (AUROC 0.66),[44] and the UK T&O study used a third centre for external validation.[35] Two of five mixed general population models have external validation,[36 39] the latter having been derived on medical patients and externally validated in medical and surgical cohorts.[45] The first of the three heart failure studies was externally validated in the subsequent studies with inferior discrimination (AUROC 0.65 in both validations).[41–43] No model updating was reported.

Quality assessment and risk of bias summary

Quality assessment was based on a draft version of the PROBAST tool. This suggested evidence in 9 of the 11 included studies of a high risk of bias (summarised in table 3) with shortcomings across the major domains

Table 3 Risk of bias summary based on Prediction study Risk Of Bias Assessment Tool (PROBAST) (permission from Wolff R, personal communication)

Population	General surgery		T&O	General (medical and surgical)					Heart failure		
Model, year	Kheterpal et al, 2007[33]	Kheterpal et al, 2009[34]	Bell et al, 2015[35]	Drawz et al, 2008[38]	Matheny et al, 2010[37]	Koyner et al, 2016[40]	Bedford et al, 2016[39]	Forni et al, 2013[36]	Forman et al, 2004[43]	Breidthardt et al, 2011[41]	Wang et al, 2013[42]
Study participants	?	?	+	+	+	?	+	+	?	?	?
Predictors	?	?	?	?	-	?	-	+	-	?	?
Outcome	-	-	+	-	-	-	+	+	-	-	-
Sample size and 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.

of the assessment. 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.[20] 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 less than 10 EPP assessed. 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.

DISCUSSION
Principal findings
In this first systematic review of HA-AKI prediction in general hospital settings, the most common predictors were age, diabetes, CKD, drugs, heart failure, SCr and bicarbonate. Modest discrimination performance of all the models is unsurprising when attempting at a single time point to predict a future event reflecting diverse aetiologies, affecting heterogeneous patient groups. Significant shortcomings mirror those described elsewhere[13 46–48]:

► multiple similar models, rarely externally validated
► no impact analysis or evidence of clinical implementation
► incomplete reporting and
► little consideration of electronic automation (allowing presentation without additional data input beyond usual clinical care), which influences uptake.[18]

Methodological and reporting shortcomings in the studies (summarised in table 2) included six studies having less than 10 EPP potentially leading to overfitting, with only three employing multiple imputation to handle missing data that can increase sample size and power.[12 49 50]

Handling of SCr and CKD was of particular concern in a number of areas. First, in part due to a previous lack of a consensus definition, the outcome in question, HA-AKI, had heterogeneous definitions, both in magnitude of SCr rise and timeframe. For example, the Kheterpal study[34] used a rise in SCr ≥177 µmol/L, which has been shown to significantly underestimate rates of AKI when compared with more recent definitions.[51] Koyner et al[40] used a rolling timeframe of 24 hours while others used SCr elevation at any point during an admission. Indeed one study produced a separate model to predict AKI at admission to hospital.[39] This was further confused in seven studies by the inclusion of admission SCr as a potential predictor (with 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 estimated glomerular filtration rate from their analysis. This risks

confusing prediction and detection of AKI events. Issues with differing definitions have been described before in systematic reviews of prediction models and should be considered when researchers embark on future studies.[52 53]

A formal risk of bias assessment (PROBAST) suggested the majority of studies had domains placing the studies at high risk of bias. 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 (eg, between community and hospital), though lack of validation in other populations tempers recommendation for implementation.[35]

Strengths and limitations of this review
This review summarises the currently available AKI prediction models in general populations who account for the majority of hospital admissions and AKI cases.[21–23] The models were selected following an extensive literature search, and the review employed the most recent critical appraisal guidance and risk of bias assessment.[12 20] The large number of patient episodes provides important insights into AKI prediction complementing other recent reviews in cardiac surgery, CI-AKI, liver transplantation and non-cardiac surgery.[14–17] 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 one model over another. Second, though we 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 (predominantly elderly demographic with comorbidities), 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 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 risk of bias assessment was also carried out (PROBAST) providing further insight into respective study strengths and weaknesses. 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. Expert guidance now exists in this area and offers opportunities to improve the scope of external validation research.[53 54]

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.[15] A cardiac surgery review reported specialty-specific predictors in addition to these chronic comorbidities. A non-cardiac surgery review (five of six studies in liver transplantation or resection) reported age, CKD and diabetes in at least two models.[17] Finally, a liver transplantation review highlighted the importance of CKD and (unsurprisingly) liver dysfunction.[16] The present review found drugs or acute laboratory values frequently included, though only four models included acute physiological parameters. Our study 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.[14 15] Ease of use (including if necessary a calculator) and potential for electronic automation were rarely considered across the models reviewed. No impact analysis studies have been described.

Future directions

Management of HA-AKI presents a significant challenge that could be helped by robust prediction models to risk stratify populations, encourage prevention and promote prompt recognition.[6 10] Appraisal and synthesis of prediction studies may enable clinicians and policy-makers judge model utility; however, this is problematic when key study details are not reported.[12] Though much of the AKI literature is on (often assumed) HA-AKI, the majority of cases arise from the community (community-acquired AKI).[55 56] Indeed, a recent study demonstrated a significant proportion of such patients are never hospitalised.[57] This review suggests even in HA-AKI, the strongest predictors are pre-existing patient factors. The two laboratory measures frequently included—SCr 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 could suggest 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 (chronic morbidity, medication, laboratory and physiological parameters). This may also enable bedside automation as part of clinical workflow, where there is evidence that beneficial implementation can be achieved.[18 58] Acute physiological parameters assessed as predictors in seven studies and subsequently included in only four studies could be an avenue of future research to improve the modest performance of all models at a single time point (admission to hospital) described to date. As hospitals increasingly employ electronic track and trigger observation systems, this may then enable the application of complex statistics (eg, 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. Unfortunately, to date, such research has not been published, with reliance on using retrospective databases often only providing information at a single time point. A future study in this area would thus require prospective collection of rich data, with the aim to achieve accurate prediction modelling demanded by clinicians and patients prior to implementation.

Impact analysis in prediction research is sparse making it difficult to conclude whether a model is worth implementing alongside, or replacing, usual care.[59] This is important as, for example, one study suggested clinical acumen may be superior to prediction models,[60] while another found the combination of a model with clinical acumen was better than either alone.[61] Some impact analyses have suggested benefit, but conclusions are limited due to their rarity and design (mostly before–after without control).[62] There are a number of potential areas for impact analysis and clinical implementation (summarised in table 4). First, in specific populations, a model could influence location of perioperative 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 could be investigated. This has been demonstrated by existing AKI alerts in *established* AKI where outcome benefit has been limited to patients who had best practice delivered.[63–65] Third, as healthcare embraces complex technology, the inclusion of physiological (including urine output) or laboratory trends may be the only way to significantly improve model performance.

Table 4	Potential areas for future impact analysis of AKI prediction models
Population	**Impact analysis to inform clinical use**
General surgery Trauma and orthopaedics	Perioperative: haemodynamic targets, place of care, drugs, contrast delivery
General populations	Risk stratification of large populations: for example, 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

Fourth, a model could identify a high-risk group to be further risk stratified by employing one of the (increasing number of) available renal biomarkers,[66] or response to an intervention such as a frusemide stress test.[67] Finally, one external validation study found those patients high risk on the prediction model who did develop AKI had a higher rate of mortality than the low-risk group who developed HA-AKI, indicating the model predicts disease severity.[45] 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.

To conclude, improving the management of patients to prevent AKI, or reduce associated complications, is a global health priority. This systematic review suggests there are few externally validated prediction models to help identify those at risk of AKI across general hospital populations. Future research should concentrate on validation, utility of additional markers, exploration of electronic implementation to enable clinical uptake and impact analysis.

Twitter First systematic review of AKI prediction models in general hospital populations

Contributors LEH, LGF, RMV, BDD and PJR developed the idea for the study. LEH, AS, LGF, BDD and PJR were involved in the study conception, preliminary literature review and design of the search strategy and the study protocol. LEH, AS and LGF were involved in screening and data extraction of papers. All authors reviewed data extraction output. LEH drafted the manuscript, which was critically reviewed and approved by all authors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int* 2012(Suppl 2):1–136.
2. Susantitaphong P, Cruz DN, Cerda J, *et al*. World incidence of AKI: a meta-analysis. *Clin J Am Soc Nephrol* 2013;8:1482–93.
3. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. *Lancet* 2012;380:756–66.
4. Bellomo R, Kellum J, Ronco C. Acute renal failure: time for consensus. *Intensive Care Med* 2001;27:1685–8.
5. Bagshaw SM, Gibney RT. Conventional markers of kidney function. *Crit Care Med* 2008;36(Suppl 4):S152–8.
6. 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). 2009. NCEPOD http://www.ncepod.org.uk/2009report1/Downloads/AKI_report.pdf
7. 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.
8. Rabar S, Lau R, O'Flynn N, *et al*. Risk assessment of fragility fractures: summary of NICE guidance. *BMJ* 2012;345:e3698.
9. Goff DC, Lloyd-Jones DM, Bennett G, *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.
10. Mehta RL, Cerdá J, Burdmann EA, *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;385:2616–43.
11. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* 2007;43:2559–79.
12. Moons KG, Altman DG, Reitsma JB, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
13. Bouwmeester W, Zuithoff NP, Mallett S, *et al*. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:e1001221–12.
14. Huen SC, Parikh CR. Predicting acute kidney injury after cardiac surgery: a systematic review. *Ann Thorac Surg* 2012;93:337–47.
15. Silver SA, Shah PM, Chertow GM, *et al*. Risk prediction models for contrast induced nephropathy: systematic review. *BMJ* 2015;351:h4395.
16. Caragata R, Wyssusek KH, Kruger P. Acute kidney injury following liver transplantation: a systematic review of published predictive models. *Anaesth Intensive Care* 2016;44:251–61.
17. Wilson T, Quan S, Cheema K, *et al*. Risk prediction models for acute kidney injury following major noncardiac surgery: systematic review. *Nephrol Dial Transplant* 2016;31:231–40.
18. Kawamoto K, Houlihan CA, Balas EA, *et al*. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005;330:765.
19. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 2001;323:224–8.
20. Moons KG, de Groot JA, Bouwmeester W, *et al*. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744.
21. Hospital Episode Statistics Analysis HaSCIC. Hospital Episode Statistics: Admitted patient care - 2014-15. 2015. Health and Social Care Information Centre http://www.hscic.gov.uk/pubs/hes1415
22. Selby NM, Crowley L, Fluck RJ, *et al*. Use of electronic results reporting to diagnose and monitor AKI in hospitalized patients. *Clin J Am Soc Nephrol* 2012;7:533–40.
23. Porter CJ, Juurlink I, Bisset LH, *et al*. A real-time electronic alert to improve detection of acute kidney injury in a large teaching hospital. *Nephrol Dial Transplant* 2014;29:1888–93.
24. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
25. Antman EM, Lau J, Kupelnick B, *et al*. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240–8.
26. Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci* 1993;703:125–34.
27. Geersing GJ, Bouwmeester W, Zuithoff P, *et al*. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One* 2012;7:e32844.
28. Wilczynski NL, McKibbon KA, Walter SD, *et al*. MEDLINE clinical queries are robust when searching in recent publishing years. *J Am Med Inform Assoc* 2013;20:363–8.
29. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–74.
30. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928–35.
31. Hosmer DW, Hjort NL. Goodness-of-fit processes for logistic regression: simulation results. *Stat Med* 2002;21:2723–38.
32. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med* 2007;35:2052–6.
33. Kheterpal S, Tremper KK, Englesbe MJ, *et al*. Predictors of postoperative acute renal failure after noncardiac surgery in patients with previously normal renal function. *Anesthesiology* 2007;107:892–902.
34. Kheterpal S, Tremper KK, Heung M, *et al*. Development and validation of an acute kidney injury risk index for patients undergoing general surgery: results from a national data set. *Anesthesiology* 2009;110:505–15.
35. Bell S, Dekker FW, Vadiveloo T, *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.

36. Forni LG, Dawes T, Sinclair H, *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;123:143–50.

37. Matheny ME, Miller RA, Ikizler TA, *et al*. Development of inpatient risk stratification models of acute kidney injury for use in electronic health records. *Med Decis Making* 2010;30:639–50.

38. Drawz PE, Miller RT, Sehgal AR. Predicting hospital-acquired acute kidney injury–a case-controlled study. *Ren Fail* 2008;30:848–55.

39. Bedford M, Stevens P, Coulton S, *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.

40. Koyner JL, Adhikari R, Edelson DP, *et al*. Development of a Multicenter Ward-Based AKI Prediction Model. *Clin J Am Soc Nephrol* 2016;11:1935–43.

41. Breidthardt T, Socrates T, Noveanu M, *et al*. Effect and clinical prediction of worsening renal function in acute decompensated heart failure. *Am J Cardiol* 2011;107:730–5.

42. Wang YN, Cheng H, Yue T, *et al*. Derivation and validation of a prediction score for acute kidney injury in patients hospitalized with acute heart failure in a Chinese cohort. *Nephrology* 2013;18:489–96.

43. Forman DE, Butler J, Wang Y, *et al*. Incidence, predictors at admission, and impact of worsening renal function among patients hospitalized with heart failure. *J Am Coll Cardiol* 2004;43:61–7.

44. Xing XZ, Wang HJ, Huang CL, *et al*. Two acute kidney injury risk scores for critically ill cancer patients undergoing non-cardiac surgery. *World J Emerg Med* 2012;3:278–81.

45. Hodgson LE, Dimitrov BD, Roderick PJ, *et al*. Predicting AKI in emergency admissions: an external validation study of the acute kidney injury prediction score (APS). *BMJ Open* 2017;7:e013511.

46. Noble D, Mathur R, Dent T, *et al*. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011;343:d7163.

47. Collins GS, Mallett S, Omar O, *et al*. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011;9:103.

48. Mallett S, Royston P, Dutton S, *et al*. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med* 2010;8:20.

49. Peduzzi P, Concato J, Feinstein AR, *et al*. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503–10.

50. Donders AR, van der Heijden GJ, Stijnen T, *et al*. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.

51. Bihorac A, Brennan M, Ozrazgat-Baslanti T, *et al*. National surgical quality improvement program underestimates the risk associated with mild and moderate postoperative acute kidney injury. *Crit Care Med* 2013;41:2570–83.

52. Ensor J, Riley RD, Moore D, *et al*. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. *BMJ Open* 2016;6:e011190.

53. Debray TP, Damen JA, Snell KI, *et al*. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;356:i6460.

54. Riley RD, Ensor J, Snell KI, *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:i3140.

55. Wonnacott A, Meran S, Amphlett B, *et al*. Epidemiology and outcomes in community-acquired versus hospital-acquired AKI. *Clin J Am Soc Nephrol* 2014;9:1007–14.

56. Xu X, Nie S, Liu Z, *et al*. Epidemiology and Clinical Correlates of AKI in Chinese Hospitalized Adults. *Clin J Am Soc Nephrol* 2015;10:1510–8.

57. Sawhney S, Fluck N, Fraser SD, *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;31:922–9.

58. Kannry J, McCullagh L, Kushniruk A, *et al*; *A Framework for Usable and Effective Clinical Decision Support: Experience from the iCPR Randomized Clinical Trial*. . Wash DC: EGEMS, 2015:3. 1150.

59. Toll DB, Janssen KJ, Vergouwe Y, *et al*. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085–94.

60. Sinuff T, Adhikari NK, Cook DJ, *et al*. Mortality predictions in the intensive care unit: comparing physicians with scoring systems. *Crit Care Med* 2006;34:878–85.

61. Brabrand M, Hallas J, Knudsen T. Nurses and physicians in a medical admission unit can accurately predict mortality of acutely admitted patients: a prospective cohort study. *PLoS One* 2014;9:e101739.

62. Fillmore CL, Bray BE, Kawamoto K. Systematic review of clinical decision support interventions with potential for inpatient cost reduction. *BMC Med Inform Decis Mak* 2013;13:135.

63. Wilson FP, Shashaty M, Testani J, *et al*. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. *Lancet* 2015;385:1966–74.

64. Kolhe NV, Reilly T, Leung J, *et al*. A simple care bundle for use in acute kidney injury: a propensity score-matched cohort study. *Nephrol Dial Transplant* 2016;31:1846–54.

65. Kolhe NV, Staples D, Reilly T, *et al*. Impact of Compliance with a Care Bundle on Acute Kidney Injury Outcomes: A Prospective Observational Study. *PLoS One* 2015;10:e0132279.

66. Murray PT, Mehta RL, Shaw A, *et al*. Potential use of biomarkers in acute kidney injury: report and summary of recommendations from the 10th Acute Dialysis Quality Initiative consensus conference. *Kidney Int* 2014;85:513–21.

67. Koyner JL, Davison DL, Brasha-Mitchell E, *et al*. Furosemide Stress Test and Biomarkers for the Prediction of AKI Severity. *J Am Soc Nephrol* 2015;26:2023–31.