BMJ Open

A LASSO-derived clinical score to predict severe acute kidney injury in the cardiac surgery recovery unit: a large retrospective cohort study using the MIMIC database

Tucheng Huang,[1,2,3] Wanbing He,[1,2,3] Yong Xie,[1,2,3] Wenyu Lv,[1,2,3] Yuewei Li,[4] Hongwei Li,[1,2,3] Jingjing Huang,[1,2,3] Jieping Huang,[1,2,3] Yangxin Chen,[1,2,3] Qi Guo ,[1,2,3] Jingfeng Wang[1,2,3]

To cite: Huang T, He W, Xie Y, *et al*. A LASSO-derived clinical score to predict severe acute kidney injury in the cardiac surgery recovery unit: a large retrospective cohort study using the MIMIC database. *BMJ Open* 2022;**12**:e060258. doi:10.1136/bmjopen-2021-060258

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-060258).

TH and WH contributed equally.

Received 15 December 2021
Accepted 13 May 2022

For numbered affiliations see end of article.

Correspondence to
Jingfeng Wang;
wjingf@mail.sysu.edu.cn,
Qi Guo;
guoq69@mail.sysu.edu.cn and
Yangxin Chen;
chenyx39@mail.sysu.edu.cn

ABSTRACT

Objectives We aimed to develop an effective tool for predicting severe acute kidney injury (AKI) in patients admitted to the cardiac surgery recovery unit (CSRU).

Design A retrospective cohort study.

Setting Data were extracted from the Medical Information Mart for Intensive Care (MIMIC)-III database, consisting of critically ill participants between 2001 and 2012 in the USA.

Participants A total of 6271 patients admitted to the CSRU were enrolled from the MIMIC-III database.

Primary and secondary outcome Stages 2–3 AKI.

Result As identified by least absolute shrinkage and selection operator (LASSO) and logistic regression, risk factors for AKI included age, sex, weight, respiratory rate, systolic blood pressure, diastolic blood pressure, central venous pressure, urine output, partial pressure of oxygen, sedative use, furosemide use, atrial fibrillation, congestive heart failure and left heart catheterisation, all of which were used to establish a clinical score. The areas under the receiver operating characteristic curve of the model were 0.779 (95% CI: 0.766 to 0.793) for the primary cohort and 0.778 (95% CI: 0.757 to 0.799) for the validation cohort. The calibration curves showed good agreement between the predictions and observations. Decision curve analysis demonstrated that the model could achieve a net benefit.

Conclusion A clinical score built by using LASSO regression and logistic regression to screen multiple clinical risk factors was established to estimate the probability of severe AKI in CSRU patients. This may be an intuitive and practical tool for severe AKI prediction in the CSRU.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Least absolute shrinkage and selection operator regression and multivariable logistic regression were used to establish a clinical score model.
⇒ The performance of this novel clinical score model in both the primary cohort and validation cohort was evaluated using the area under the receiver operating characteristic curve, calibration curves and decision curve analysis.
⇒ This novel clinical score model might not be suitable for those with a renal failure history.
⇒ External validation of this novel clinical score model was lacking.

INTRODUCTION

Acute kidney injury (AKI), a common complication in patients admitted to the intensive care unit (ICU) worldwide,[1 2] is associated with adverse short-term and long-term prognoses.[3] It has been reported that more than half of patients in the cardiac surgery recovery unit (CSRU) suffer from AKI of some stage,[4] which is associated with high mortality and rehospitalisation rates.[5] The early and rapid diagnosis and treatment of AKI may help reduce mortality and rehospitalisation rates. Although several biomarkers have been used for the early diagnostic and prognostic prediction of AKI,[6 7] the clinical utilisation of these biomarkers has been limited. When the levels of these biomarkers increase, renal injury occurs. Thus, identifying critically ill patients at high risk of AKI is an important part of the overall management of CSRU patients.

Graphical calculation devices, which are presented as a scale or score that incorporate possible risk factors to make clinical prognostic predictions, have become increasingly popular. It has been extensively used to predict the probability of death or recurrence events for a patient with cancer.[8] Recently, some researchers established a clinical prediction model for forecasting the occurrence of AKI in patients undergoing cardiac surgery.[9] However, that small, single-centre study did

not exclude patients with chronic kidney disease and thus probably overestimated the occurrence of AKI; additionally, only logistic regression for variable selection was used. By machine learning, a model was established to predict cardiac surgery-associated AKI, although the sample was small and urine output was neglected.[10] Another study used a convolutional neural network model to predict severe AKI in the ICU, while patients with a previous diagnosis of chronic kidney disease were not excluded.[11]

Least absolute shrinkage and selection operator (LASSO) regression is of great strength for variable selection because it can efficiently address the potential association between covariates, such as collinearity.[12] Accordingly, in this study, we performed LASSO regression to select variables and built a logistic regression model to identify independent risk factors for severe AKI in patients admitted to the CSRU. We aimed to determine the risk factors for severe AKI and develop a clinical score for evaluating the probability that patients undergoing critical cardiac care will acquire severe AKI.

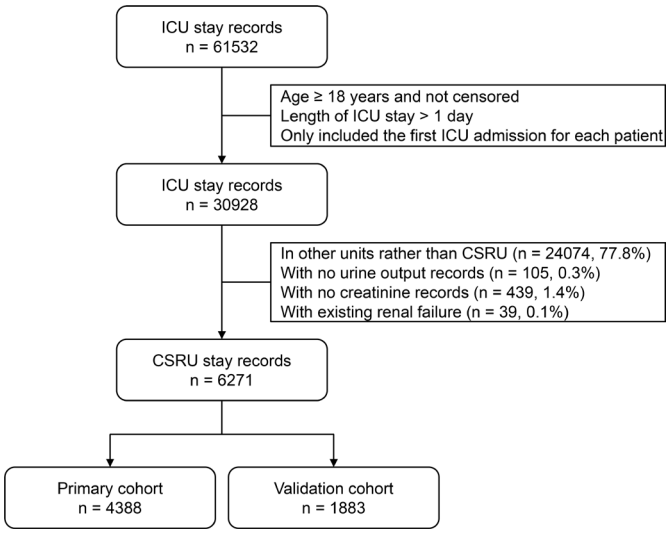

Figure 1 Flow chart of enrolled subjects. A total of 6271 CSRU stay records were enrolled in this study. CSRU, cardiac surgery recovery unit; ICU, intensive care unit.

METHODS

Data source and ethics approval

The data were extracted from the Medical Information Mart for Intensive Care (MIMIC)-III data set. As a large and publicly available database, MIMIC-III comprises the clinical information for 61 532 ICU stay cases between 2001 and 2012. The use of the MIMIC-III database was approved by the review boards of the Massachusetts Institute of Technology and Beth Israel Deaconess Medical Center.[13] Because the information used in the study was from a publicly deidentified database, the informed consent requirement was waived.

Study population

Adult ICU stays longer than 1 day were included. When a patient had multiple ICU admissions, only the first medical record was selected for the study. The exclusion criteria were as follows: patients in units other than the CSRU (n=24 074, 77.8%); patients with no urine output records (n=105, 0.3%); patients with no creatinine data (n=439, 1.4%) and patients with existing renal failure (n=39, 0.1%) (figure 1). During the CSRU stay, all creatinine and urine output records were extracted, and AKI was defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines.[14] Baseline serum creatinine was defined as the lowest creatinine in the past 7 days. Both urine output and serum creatinine criteria were used to identify AKI. Information about renal replacement therapy was not considered in this study. Severe AKI was defined as stage 2 or stage 3 AKI under the KDIGO criteria. Patients in the CSRU were screened, and a total of 6271 patients were included. Chronologically, the first 70% of patients were allocated to the primary cohort, and the last 30% were allocated to the validation cohort. Subsequently, we established a clinical score model by using the primary cohort data and validated the model by using the validation cohort.

Variable extraction

The following variables were extracted.

Demographics: age (years), sex, height (cm), and weight (kg).

Vital signs: heart rate (/min), respiratory rate (/min), temperature (°C, saturation of peripheral oxygen (%), blood glucose level (mg/dL), systolic blood pressure (SBP, mm Hg), diastolic blood pressure (DBP, mm Hg), central venous pressure (CVP, mm Hg) and mean artery pressure (mm Hg). The mean value of vital signs in the 24 hours after admission was included for analysis.

Laboratory tests: white blood cell count (×10^9/L), haemoglobin (g/L), platelets (×10^9/L), chloride (mmol/L), sodium (mmol/L), blood urea nitrogen (mg/dL), bicarbonate (mmol/L), pH, partial pressure of oxygen (pO_2, mm Hg), partial pressure of carbon dioxide (pCO_2, mm Hg), creatinine (mg/dL) and potassium (mmol/L). The values of laboratory tests in the first 24 hours after admission were used for the analysis. In addition, 24-hour urine output was extracted.

Procedures: administration of furosemide, use of sedative, ventilation, vasopressor, cardiopulmonary bypass, coronary artery bypass grafting and left heart catheterisation. The sedative drugs in this study included midazolam, fentanyl andpropofol.

Comorbidities: coronary artery disease, congestive heart failure, atrial fibrillation, stroke, diabetes, renal disease, liver disease, chronic obstructive pulmonary disease and malignancy.

All variables were collected in the initial 24 hours after admission to predict severe AKI as early as possible. The frequency of missing values for each variable was less than 15%. The missing values were filled in by the random forest method using R software.

Table 1 Baseline characteristics of the enrolled subjects in the primary and validation cohorts

	Primary cohort	Validation cohort	P value
n	4388	1883	
Age, years	66.0±12.8	65.9±13.3	0.715
Male	2921 (66.6)	1229 (65.3)	0.332
Weight, kg	83.0±19.1	83.2±20.0	0.785
Heart rate, /min	84.9±10.7	84.6±10.8	0.357
Respiratory rate, /min	17.2±3.1	17.2±3.0	0.914
Glucose, mg/dL	131.2±23.2	132.4±23.2	0.060
SBP, mm Hg	113.3±10.7	113.9±10.8	0.040
DBP, mm Hg	57.1±6.9	57.3±7.0	0.244
CVP, mm Hg	10.6±3.5	10.7±3.6	0.191
Urine output, mL	2075.0 (1480.0–2880.0)	2080.0 (1457.0–2900.0)	0.949
pO_2, mm Hg	314.0 (211.0–383.0)	308.0 (206.0–386.0)	0.168
Sedative	3707 (84.5)	1593 (84.6)	0.905
Ventilation	3836 (87.4)	1642 (87.2)	0.811
Furosemide	675 (15.4)	292 (15.5)	0.901
Atrial fibrillation	1695 (38.6)	754 (40.0)	0.293
Congestive heart failure	1018 (23.2)	442 (23.5)	0.814
Stroke	258 (5.9)	108 (5.7)	0.823
Left heart catheterisation	1288 (29.4)	551 (29.3)	0.942
Severe AKI	2452 (55.9)	1020 (54.2)	0.213

The data are depicted as the mean±SD, the median (IQR) or a number (percentage). Continuous data were compared with Student's t test or the rank-sum test, while categorical data were compared using the χ^2 test.
AKI, acute kidney injury; CVP, central venous pressure; DBP, diastolic blood pressure; pO_2, partial pressure of oxygen; SBP, systolic blood pressure.

Statistical analysis

Continuous variables are denoted as the mean±SD or the median (IQR), whereas categorical variables are expressed as numbers (percentages). Continuous data were compared with Student's t test or the rank-sum test, while categorical data were compared using the χ^2 test.

In this study, LASSO was performed for variable selection. LASSO regression is a compression estimation used to address the collinearity between covariates. When there are several collinear predictors, LASSO selects only one and ignores the others or zeroes out some regression coefficients. Cross-validation was used during LASSO regression, and 1–SE criterion was used to select lambda. Namely, the value of lambda was identified when the cross-validated error was within one SE of the minimum. ORs with 95% CIs, statistics describing the strength of the association between disease and exposure, were calculated by logistic regression, thus estimating the association of independent risk factors with AKI. Finally, a clinical score model was established based on the above analysis, which was further validated with C-indices, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic (ROC) curves, the areas under the ROC curves (AUCs), calibration curves and decision curve analysis. We used 10-fold cross validation to identify the optimal clinical score model. Briefly, the primary cohort was randomly divided into 10 roughly equal-sized groups. One group was taken as a test data set, and the remaining groups were used as a training data set. The model was fitted on the training data set and evaluated on the test data set. After repeating the process 10 times, the optimal model with the best performance was identified.

SPSS software (V.23.0, IBM, NY, USA) and R software (V.3.6.3, R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analysis. The packages used in this study included *missForest*, *glmnet*, *rms*, *pROC*, *caret* and *rmda*. A two-sided $p<0.05$ was considered statistically significant.

Patient and public involvement

Patients and/or the public were not directly involved in this study.

RESULTS

Patients with severe AKI comprised 55.9% (2452/4388) and 54.2% (1020/1883) of the primary and validation cohorts, respectively. No significant difference in the severe AKI rate was observed between the two cohorts

Table 2 Baseline characteristics of the severe AKI and non-severe AKI groups in the primary cohort

	Severe AKI	Non-severe AKI	P value
n	2452	1936	
Age, years	67.4±12.2	64.3±13.3	<0.001
Male	1606 (65.5)	1315 (67.9)	0.094
Weight, kg	86.7±20.2	78.4±16.5	<0.001
Heart rate, /min	85.0±10.8	84.7±10.6	0.475
Respiratory rate, /min	17.2±3.1	17.2±3.0	0.999
Glucose, mg/dL	133.4±23.6	128.6±22.2	<0.001
SBP, mm Hg	112.7±10.5	114.0±10.8	<0.001
DBP, mm Hg	56.5±6.9	57.9±6.9	<0.001
CVP, mm Hg	11.2±3.7	9.8±3.1	<0.001
Urine output, mL	1735.5 (1245.0–2384.3)	2550.0 (1930.0–3355.0)	<0.001
pO_2, mm Hg	309.0 (204.0–379.0)	323.0 (224.0–389.0)	0.009
Sedative	2116 (86.3)	1591 (82.2)	<0.001
Ventilation	2183 (89.0)	1653 (85.4)	<0.001
Furosemide	341 (13.9)	334 (17.3)	0.002
Atrial fibrillation	1074 (43.8)	621 (32.1)	<0.001
Congestive heart failure	673 (27.4)	345 (17.8)	<0.001
Stroke	132 (5.4)	126 (6.5)	0.121
Left heart catheterisation	762 (31.1)	526 (27.2)	0.005

The data are depicted as the mean±SD, the median (IQR) or a number (percentage). Continuous data were compared with Student's t test or the rank-sum test, while categorical data were compared using the χ^2 test
AKI, acute kidney injury; CVP, central venous pressure; DBP, diastolic blood pressure; pO_2, partial pressure of oxygen; SBP, systolic blood pressure.

(p=0.213). Except for SBP (primary cohort, 113.3 mm Hg vs validation cohort, 132.4 mm Hg, p=0.040), no clinical characteristics showed a significant difference between the primary and validation cohorts (table 1).

In the primary cohort, patients with severe AKI were older, had higher weights and had higher blood glucose level than those without severe AKI (p<0.001). SBP and DBP were significantly lower (112.7 mm Hg vs 114.0 mm Hg and 56.6 mm Hg vs 57.9 mm Hg, respectively), while CVP was significantly higher (11.2 mm Hg vs 9.8 mm Hg) in the severe AKI group (p<0.001). Urine output and pO_2 were lower in the severe AKI group (p<0.01). Drug administration was also different, namely, severe AKI patients received sedatives, ventilation and furosemide significantly more often (p<0.001). The stroke prevalence rates were the same, but a higher prevalence of atrial fibrillation, congestive heart failure and left heart catheterisation was observed in severe AKI patients (p<0.05) (table 2).

To confirm the possible risk factors for severe AKI, we performed LASSO regression to select variables. A total of 18 variables were enrolled for further analysis according to the 1–SE criterion (figure 2). Then, we conducted logistic regression analysis based on the LASSO results. A total of 14 variables were shown to be associated with severe AKI (table 3).

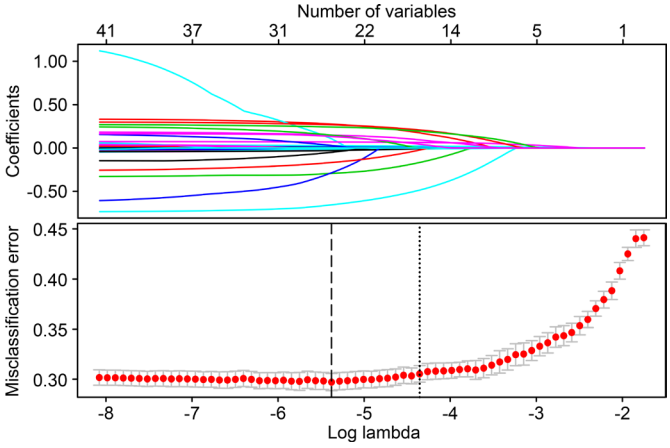

Figure 2 LASSO coefficient profiles of variables and misclassification errors for different models. The upper panel presents the associations between the coefficients of variables and the log lambda value. Each line corresponds to one distinct variable. With increasing log lambda, the coefficient of the variable tended towards 0. The lower panel presents the selection of the applicable model. Vertical lines were drawn at the optimal values by adopting the minimum criteria (dashed line) and the SE of the minimum criteria (dotted line, the 1–SE criteria). In our study, the lambda value was chosen according to the 1–SE criteria. LASSO, least absolute shrinkage and selection operator.

Table 3 Variables in the LASSO regression and multivariate logistic regression models

	LASSO regression	Logistic regression		
Variables	**β**	**β**	**OR (95% CI)**	**P value**
Age	0.011221	0.017	1.017 (1.010 to 1.023)	<0.001
Male	−0.165641	−0.404	0.667 (0.568 to 0.784)	<0.001
Weight	0.023091	0.031	1.032 (1.027 to 1.037)	<0.001
Heart rate	0.000058	0.007	1.007 (1.000 to 1.014)	0.055
Respiratory rate	−0.006347	−0.042	0.959 (0.936 to 0.982)	0.001
Glucose	0.000846	0.002	1.002 (0.999 to 1.005)	0.181
SBP	−0.004721	−0.010	0.990 (0.983 to 0.997)	0.007
DBP	−0.009688	−0.015	0.985 (0.974 to 0.997)	0.011
CVP	0.063826	0.072	1.075 (1.051 to 1.099)	<0.001
Urine output	−0.000603	−0.001	0.999 (0.999 to 0.999)	<0.001
pO_2	−0.000127	−0.001	0.999 (0.998 to 1.000)	0.001
Sedative	0.173715	0.340	1.405 (1.032 to 1.912)	0.031
Ventilation	0.093818	0.189	1.209 (0.862 to 1.694)	0.272
Furosemide	−0.484207	−0.757	0.469 (0.387 to 0.569)	<0.001
Atrial fibrillation	0.193466	0.279	1.322 (1.139 to 1.536)	<0.001
Congestive heart failure	0.207495	0.305	1.357 (1.143 to 1.611)	<0.001
Stroke	−0.021989	−0.254	0.776 (0.580 to 1.038)	0.087
Left heart catheterisation	0.043483	0.166	1.181 (1.014 to 1.376)	0.033

LASSO, least absolute shrinkage and selection operator; SBP, systolic blood pressure; DBP, diastolic blood pressure; CVP, central venous pressure; pO_2, partial pressure of oxygen.

Next, we included the above significant factors to build a clinical score based on the logistic regression model (figure 3). Each level of every variable was assigned a score. By adding the scores for all of the selected variables, the total score was obtained. By checking the number corresponding to the total scores, the probability of severe AKI can be estimated for a given patient.

The C-indices were 0.779 for the primary cohort and 0.778 for the validation cohort. The ROC curves demonstrated that the model had good discriminative ability in

Age, years	10	20	30	40	50	60	70	80	90				
Score	0	2	4	6	8	10	12	14	16				
Sex	Male	Female											
Score	0	5											
Weight, kg	20	40	60	80	100	120	140	160	180	200	220	240	260
Score	0	8	15	23	31	38	46	54	62	69	77	85	92
Respiratory rate, /min	5	10	15	20	25	30	35						
Score	14	11	9	7	5	2	0						
SBP, mmHg	70	80	90	100	110	120	130	140	150	160	170	180	
Score	15	14	13	11	10	8	7	6	4	3	1	0	
DBP, mmHg	20	30	40	50	60	70	80	90	100	110			
Score	15	13	11	10	8	6	5	3	2	0			
CVP, mmHg	0	5	10	15	20	25	30	35					
Score	5	9	14	18	23	27	32	36					
Urine output, mL	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000		
Score	84	76	67	59	51	42	34	25	17	8	0		
pO2, mmHg	0	100	200	300	400	500	600	700	800				
Score	10	9	8	7	5	4	3	1	0				
Sedative	No	Yes											
Score	0	6											
Furosemide	No	Yes											
Score	10	0											
Atrial fibrillation	No	Yes		**Total scores**			125	135	151	161	171	188	197
Score	0	3		**Probability of severe AKI**			0.05	0.10	0.30	0.50	0.70	0.90	0.95
Congestive heart failure	No	Yes											
Score	0	4											
Left heart catheterization	No	Yes											
Score	0	2											

Figure 3 Clinical score for the prediction of severe AKI in CSRU patients. All 14 selected variables, including age, sex, weight, respiratory rate, SBP, DBP, CVP, urine output, pO_2, sedative usage, furosemide, atrial fibrillation, congestive heart failure and left heart catheterisation, were given corresponding points based on their values. The total points of these variables corresponded to the predicted probability of severe AKI in the CSRU. AKI, acute kidney injury; CSRU, cardiac surgery recovery unit; CVP, central venous pressure; DBP, diastolic blood pressure; SBP, systolic blood pressure.

Table 4 Model performance in the primary and validation cohorts

	AUC (95% CI)	Accuracy (95% CI)	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Cut-off value	Cut-off score
Primary cohort	0.779 (0.766 to 0.793)	0.702 (0.688 to 0.715)	0.609	0.820	0.811	0.623	0.566	167.9
Validation cohort	0.778 (0.757 to 0.799)	0.715 (0.694 to 0.735)	0.781	0.637	0.718	0.722	0.065	161.8

AUC, area under the receiver operating characteristic curve.

both the primary cohort (AUC: 0.779, 95% CI: 0.766 to 0.793) and the validation cohort (AUC: 0.778, 95% CI: 0.757 to 0.799) (table 4). Calibration plots showed that the apparent curves were adjacent to the ideal curves in both the primary and validation groups. Finally, decision curve analysis was performed to compare the clinical usability and benefits of the model. The decision curves showed acceptable net benefits across a range of high risks of severe AKI in the primary and validation cohorts (figure 4).

We also evaluated the model performance after excluding the variable of urine output. Without urine output information, the model also showed acceptable discriminative ability in both the primary cohort (AUC: 0.713, 95% CI: 0.698 to 0.728) and the validation cohort (AUC: 0.718, 95% CI: 0.695 to 0.741) (online supplemental table 1). For patients without suffering AKI in the initial 24 hours after admission, the model performed with an AUC of 0.680 (95% CI: 0.651 to 0.709) in the primary cohort and an AUC of 0.673 (95% CI: 0.630 to 0.715) (online supplemental table 2).

DISCUSSION

AKI is a complicated clinical syndrome characterised by reduced urine production and/or rapid increases in serum creatinine.[15] AKI has been reported to be positively associated with short-term mortality in CSRU populations.[5 16] Delayed diagnosis of AKI is an independent risk factor for nosocomial death.[17] Therefore, the early identification of patients at risk for AKI might help to reduce short-term mortality, improve prognosis, and reduce the healthcare burden.

In this study, we extracted the clinical information of 6271 patients from the MIMIC-III database. We identified the following 14 possible risk factors for severe AKI by LASSO regression and logistical regression: age, sex, weight, respiratory rate, SBP, DBP, CVP, urine output, pO_2, sedative use, furosemide, atrial fibrillation, congestive heart failure and left heart catheterisation. Subsequently, a clinical score model was constructed by quantifying the weight of the aforementioned variables. The clinical score model was well fitted, as evaluated by

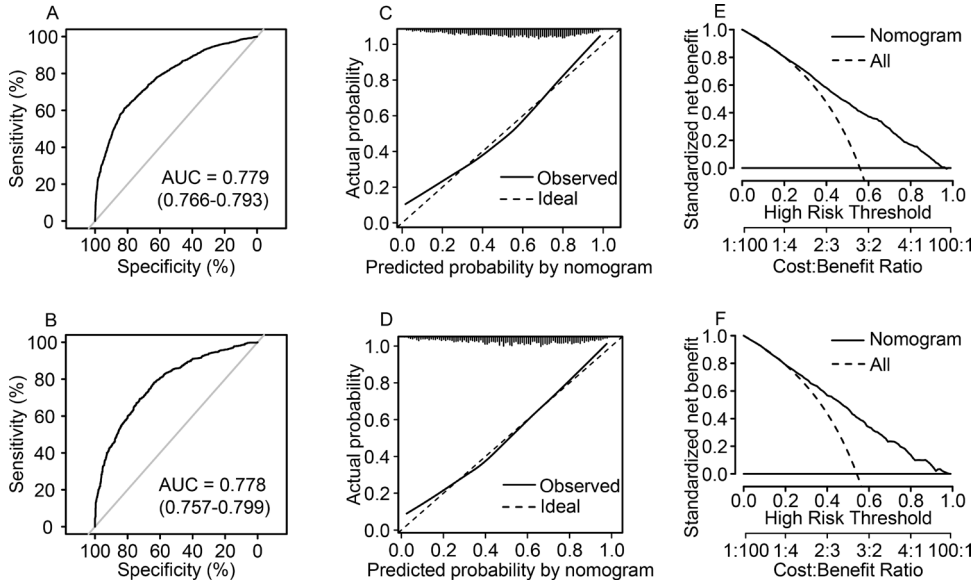

Figure 4 Performance evaluation of the severe AKI prediction model. ROC curves in the primary cohort (A) and validation cohort (B). The AUCs of the model in the primary and validation cohorts were 0.779 and 0.778, respectively. Calibration curves in the primary cohort (C) and validation cohort (D). The observed values were close to the ideal values, indicating a satisfactory forecasting performance of the clinical score model. Decision curve analyses in the primary cohort (E) and validation cohort (F), showing the net benefit from the model. AKI, acute kidney failure; AUC, area under the receiver operating characteristic curve; ROC, receiver operator characteristic curve.

the AUC, calibration curves and decision curve analysis in both the primary and validation cohorts. The model could calculate a severe AKI probability immediately after the initial 24 hours and might help clinicians perform early intervention.

Several scoring systems and prognostic models have been built to predict AKI. Scoring systems such as the Cleveland Clinic Score[18] and the Mehta Score[19] only consider AKI patients requiring dialysis and thus might miss patients with subclinical AKI. Additionally, clinical prediction models have been used to forecast AKI in patients undergoing cardiac surgery[9] or coronary angiography.[20] These studies enrolled both mild and severe AKI patients. Our model was generated from the MIMIC-III database, with a larger sample size and more variables. This study predicted only severe AKI, which might be more attractive for clinical practice. Moreover, the primary cohort and validation cohort were assigned by admission time. According to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement, non-random assignment by time is a stronger design feature for evaluating model performance than random assignment.[21]

LASSO regression is a popular variable selection algorithm for multicollinear data or high-dimensional data.[22] LASSO has been widely used for clinical prediction. For example, via LASSO, researchers have built a clinical model to predict the diagnosis and prognosis of colon cancer.[23] A radiomics signature using LASSO has been developed to evaluate survival in patients with non-small-cell lung cancer.[24] LASSO has been used to predict AKI in patients with haematologic tumours, patients suffering from cardiac surgery or patients hospitalised in the neurosurgical ICU.[12 22 25] In the present study, based on clinical profiles, LASSO was performed to select relevant coefficients from a multitude of variables, simultaneously removing all unrelated variables. Through dimensionality reduction using LASSO, 42 clinical variables were screened down to 14 risk factors, according to the 1–SE criterion.

Among those 14 variables, older age and obesity were independent risk factors for AKI, as indicated by previous investigations.[26 27] Additionally, hypotension has been reported to be associated with new-onset AKI in ICU patients with shock.[28] High CVP, indicating fluid overload, is another factor affecting AKI.[29] Consistent with previous studies, these risk factors were included in the clinical score model and given a weighted score. Reduced urine output is a clinical manifestation of AKI and is also an important factor underlying the poor prognosis of AKI. In this study, decreased urine output was one of the most important predictors of AKI in CSRU patients. Overall, the clinical score model contained 14 variables, more than half of which have been reported to be associated with AKI. In addition, ROC curves, calibration curves and decision curve analysis showed consistent results in both the primary and validation cohorts, showing that the clinical score model could be an effective and reliable tool for predicting the risk of severe AKI.

Several limitations of our study must be noted. First, this study was based on the MIMIC-III database, whose data were collected between 2001 and 2012. Some therapies might not meet the latest guidelines and some newer medicines might not be included. Because of the single-centre nature of the data, the performance of our model might vary when applied to other regions. The potential residual confounding by variables not recorded in this database could not be evaluated. Second, only patients without existing renal failure were included in this study. Thus, this novel score model might not be suitable for those with a renal failure history. Third, missing values were filled by the random forest method, which might lead to biased regression coefficient estimates.[30] Therefore, further studies are needed to verify our model. Fourth, our model was designed to be used immediately after the initial 24 hours of admission, and it may not work for patients who suffer AKI within those initial 24 hours.

CONCLUSION

In conclusion, this study established and validated a novel clinical score by using LASSO regression and logistic regression to screen for multiple clinical risk factors to estimate the probability of severe AKI in CSRU patients. This clinical score model can be an intuitive and reliable predictive tool that might help in individualised clinical decision-making and risk management for severe AKI.

Author affiliations
[1]Department of Cardiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
[2]Guangzhou Key Laboratory of Molecular Mechanism and Translation in Major Cardiovascular Disease, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
[3]Guangdong Provincial Key Laboratory of Arrhythmia and Electrophysiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China
[4]Department of Respiratory Medicine, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China

Acknowledgements We would like to thank the participants, developers and investigators associated with the Medical Information Mart for Intensive Care (MIMIC)-III database.

Contributors TH: conceptualisation, data analysis, writing original draft, writing review and editing. WH: conceptualisation, writing original draft, writing review and editing. YX, WL and YL: writing original draft and data curation. HL, JJH and JPH: literature search and data interpretation. YC, QG and JW: conceptualisation, writing review and editing and data curation. QG accepts full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish. QG was confirmed as guarantor.

Funding This study was supported by grants from the National Natural Science Foundation of China (nos. 82070237, 81870170, 81770229, 81970200), Guangdong Basic and Applied Basic Research Foundation (no. 2020A1515110313), Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory) (no. 2019GZR110406004), Guangzhou Science and Technology Bureau (nos. 201803040010, 201707010206, 202102010007) and Yat-sen Start-up Foundation (no. YXQH202014).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study involves human participants. The data were extracted from the Medical Information Mart for Intensive Care (MIMIC)-III data set. As a large and publicly available database, MIMIC-III, comprises the clinical information for 61532 ICU stay cases between 2001 and 2012. The use of the MIMIC-III database was approved by the review boards of the Massachusetts Institute of Technology and Beth Israel Deaconess Medical Center. Since the information used in the study was from a publicly deidentified database, informed consent was waived. Briefly, this study was based on a famous public database. No reference number or ID was available for this public database. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data set analysed to generate the findings for this study is available from the corresponding author on reasonable request.

ORCID iD

Qi Guo http://orcid.org/0000-0003-3145-1309

REFERENCES

1 Rossaint J, Zarbock A. Acute kidney injury: definition, diagnosis and epidemiology. *Minerva Urol Nefrol* 2016;68:49–57.
2 Hoste EAJ, Bagshaw SM, Bellomo R, *et al*. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med* 2015;41:1411–23.
3 Abd ElHafeez S, Tripepi G, Quinn R, *et al*. Risk, predictors, and outcomes of acute kidney injury in patients admitted to intensive care units in Egypt. *Sci Rep* 2017;7:17163.
4 Jentzer JC, Breen T, Sidhu M, *et al*. Epidemiology and outcomes of acute kidney injury in cardiac intensive care unit patients. *J Crit Care* 2020;60:127–34.
5 Holland EM, Moss TJ. Acute Noncardiovascular illness in the cardiac intensive care unit. *J Am Coll Cardiol* 2017;69:1999–2007.
6 Greenberg JH, Zappitelli M, Jia Y, *et al*. Biomarkers of AKI progression after pediatric cardiac surgery. *J Am Soc Nephrol* 2018;29:1549–56.
7 Chang C-H, Yang C-H, Yang H-Y, *et al*. Urinary biomarkers improve the diagnosis of intrinsic acute kidney injury in coronary care units. *Medicine* 2015;94:e1703.
8 Iasonos A, Schrag D, Raj GV, *et al*. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008;26:1364–70.
9 Guan C, Li C, Xu L, *et al*. Risk factors of cardiac surgery-associated acute kidney injury: development and validation of a perioperative predictive nomogram. *J Nephrol* 2019;32:937–45.
10 Tseng P-Y, Chen Y-T, Wang C-H, *et al*. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care* 2020;24:478.
11 Le S, Allen A, Calvert J, *et al*. Convolutional neural network model for intensive care unit acute kidney injury prediction. *Kidney Int Rep* 2021;6:1289–98.
12 An S, Luo H, Wang J, *et al*. An acute kidney injury prediction nomogram based on neurosurgical intensive care unit profiles. *Ann Transl Med* 2020;8:194.
13 Johnson AEW, Pollard TJ, Shen L, *et al*. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
14 Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract* 2012;120:c179–84.
15 Ronco C, Bellomo R, Kellum JA. Acute kidney injury. *Lancet* 2019;394:1949–64.
16 Jentzer JC, Bennett C, Wiley BM, *et al*. Predictive value of the sequential organ failure assessment score for mortality in a contemporary cardiac intensive care unit population. *J Am Heart Assoc* 2018;7:1–11.
17 Yang L, Xing G, Wang L, *et al*. Acute kidney injury in China: a cross-sectional survey. *Lancet* 2015;386:1465–71.
18 Thakar CV, Arrigain S, Worley S, *et al*. A clinical score to predict acute renal failure after cardiac surgery. *J Am Soc Nephrol* 2005;16:162–8.
19 Mehta RH, Grab JD, O'Brien SM, *et al*. Bedside tool for predicting the risk of postoperative dialysis in patients undergoing cardiac surgery. *Circulation* 2006;114:2208–16.
20 Zhou X, Sun Z, Zhuang Y, *et al*. Development and validation of nomogram to predict acute kidney injury in patients with acute myocardial infarction treated Invasively. *Sci Rep* 2018;8:9769.
21 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
22 Li Y, Chen X, Wang Y, *et al*. Application of group LASSO regression based Bayesian networks in risk factors exploration and disease prediction for acute kidney injury in hospitalized patients with hematologic malignancies. *BMC Nephrol* 2020;21:162.
23 Zhou R, Zhang J, Zeng D, *et al*. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. *Cancer Immunol Immunother* 2019;68:433–42.
24 Huang Y, Liu Z, He L, *et al*. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* 2016;281:947–57.
25 Coulson T, Bailey M, Pilcher D, *et al*. Predicting acute kidney injury after cardiac surgery using a simpler model. *J Cardiothorac Vasc Anesth* 2021;35:866–73.
26 Søvik S, Isachsen MS, Nordhuus KM, *et al*. Acute kidney injury in trauma patients admitted to the ICU: a systematic review and meta-analysis. *Intensive Care Med* 2019;45:407–19.
27 Sengthavisouk N, Lumlertgul N, Keomany C, *et al*. Epidemiology and short-term outcomes of acute kidney injury among patients in the intensive care unit in Laos: a nationwide multicenter, prospective, and observational study. *BMC Med* 2020;18:180.
28 Panwar R, Tarvade S, Lanyon N, *et al*. Relative hypotension and adverse kidney-related outcomes among critically ill patients with shock. A multicenter, prospective cohort study. *Am J Respir Crit Care Med* 2020;202:1407–18.
29 Bouchard J, Soroko SB, Chertow GM, *et al*. Fluid accumulation, survival and recovery of kidney function in critically ill patients with acute kidney injury. *Kidney Int* 2009;76:422–7.
30 Hong S, Lynn HS. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, Non-linearity, and interaction. *BMC Med Res Methodol* 2020;20:199.