



Published in final edited form as:

Am J Kidney Dis. 2023 January ; 81(1): 36–47. doi:10.1053/j.ajkd.2022.06.004.

Prediction of Mortality and Major Adverse Kidney Events in Critically Ill Patients With Acute Kidney Injury

Javier A. Neyra, MD, MSCS^{1,4}, Victor Ortiz-Soriano, MD¹, Lucas J. Liu, MS^{2,3}, Taylor D. Smith, MS^{2,3}, Xilong Li, PhD⁴, Donglu Xie, MS⁵, Beverley Adams-Huet, MS^{4,5}, Orson W. Moe, MD^{4,6}, Robert D. Toto, MD^{4,5,6}, Jin Chen, PhD^{2,3}

¹Department of Internal Medicine, Division of Nephrology, Bone and Mineral Metabolism University of Kentucky, Lexington, KY, USA

²Department of Internal Medicine, Division of Biomedical Informatics, University of Kentucky, Lexington, KY, USA

³Department of Computer Science, University of Kentucky, Lexington, KY, USA

⁴Charles and Jane Park Center for Mineral Metabolism and Clinical Research, University of Texas Southwestern, Dallas, TX, USA

⁵Department of Population and Data Sciences, University of Texas Southwestern, Dallas, TX, USA

⁶Department of Internal Medicine, Division of Nephrology, University of Texas Southwestern, Dallas, TX, USA

Abstract

Rationale & Objective—Risk-prediction tools for assisting acute kidney injury (AKI) management have focused on AKI onset but have infrequently addressed kidney recovery. We developed clinical models for risk-stratification of mortality and major adverse kidney events in critically ill patients with incident AKI.

Study Design—Multicenter cohort study.

Setting & Participants—9,587 adult patients admitted to heterogenous ICUs (March 2009 to February 2017) who developed AKI within the first 3 days of their ICU stays.

Corresponding Author: Javier A. Neyra, MD, MSCS, University of Kentucky Medical Center, 800 Rose St, MN668, Lexington, KY, 40536, javier.neyra@uky.edu.

Authors' Contributions: Contributed to study concept and design: JAN, VOS, LJJ, TDS, JC; helped with data extraction, management, harmonization, and validation: VOS, LJJ, TDS, XL, DX, BA-H; performed data analysis: LJJ, TDS, XL, JC; performed model evaluation and interpretation: JAN, VOS, LJJ, XL, JC; supervised study procedures at UTSW: OWM, RDT; supervised study procedures at UKY: JC; supervised the overall study procedures and execution: JAN. Each author contributed important intellectual content during manuscript drafting or revision and agrees to be personally accountable for the individual's own contributions and to ensure that questions pertaining to the accuracy or integrity of any portion of the work, even one in which the author was not directly involved, are appropriately investigated and resolved, including with documentation in the literature if appropriate.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Financial Disclosure: The authors declare that they have no relevant financial interests.

Predictors—Multimodal clinical data consisting of 71 features collected in first 3 days of ICU stay.

Outcomes—1) Hospital mortality and 2) major adverse kidney events (MAKE), defined as the composite of death, dependence on renal replacement therapy and a drop in eGFR \geq 50% from baseline up to 120 days from hospital discharge.

Analytical Approach—Four machine learning algorithms (logistic regression, random forest, support vector machine, and extreme gradient boosting) and the SHapley Additive exPlanations (SHAP) framework were used for feature selection and interpretation. Model performance was evaluated by 10-fold cross validation and external validation.

Results—One developed model including 15 features outperformed the SOFA score for the prediction of hospital mortality: AUC (95%CI) 0.79 (0.79–0.80) vs. 0.71 (0.71–0.71) in the development cohort and 0.74 (0.73–0.74) vs. 0.71 (0.71–0.71) in the validation cohort, $p < 0.001$ for both. A second developed model including 14 features outperformed KDIGO AKI severity staging for the prediction of MAKE: 0.78 (0.78–0.78) vs. 0.66 (0.66–0.66) in the development cohort and 0.73 (0.72–0.74) vs. 0.67 (0.67–0.67) in the validation cohort, $p < 0.001$ for both.

Limitations—The models are only applicable to critically ill adult patients with incident AKI within the first 3 days of an ICU stay.

Conclusions—The reported clinical models exhibited better performance for mortality and kidney recovery prediction compared to standard scoring tools commonly used in critically ill patients with AKI in the ICU. Additional validation is needed to support the utility and implementation of these models.

PLAIN-LANGUAGE SUMMARY

Acute kidney injury (AKI) occurs commonly in critically ill patients admitted to the intensive care unit (ICU) and is associated with high morbidity and mortality. Prediction of mortality and recovery after an episode of AKI may assist bedside decision-making. In this manuscript, we described the development and validation of a clinical model using data from the first 3 days of an ICU stay to predict hospital mortality and major adverse kidney events occurring up to 120 days after hospital discharge among critically ill adult patients who developed AKI within the first 3 days of an ICU stay. The proposed clinical models exhibited good performance for outcome prediction and, if further validated, could enable risk-stratification for timely interventions that promote kidney recovery.

Keywords

Mortality; kidney recovery; acute kidney injury; critically ill patients; machine learning

INTRODUCTION

Acute kidney injury (AKI) occurs in up to 50% of critically ill patients admitted to Intensive Care Units (ICU)^{1–4}. Incident AKI increases risk of hospital mortality^{4, 5} and carries deleterious systemic effects in survivors, predisposing to cardiovascular disease^{6–8} and progressive kidney disease^{9–12}. Furthermore, the overall morbidity associated with AKI

increases resource utilization and healthcare costs¹³. Given the public health relevance of AKI, there is tremendous need for development of clinical tools that assist with risk-stratification of prognosis and post-AKI outcomes. These tools can potentially support interventions that can attenuate the burden of AKI and its complications, as well as augment precision in clinical decisions that could positively impact patient-centered outcomes after discharge.

Recent studies clearly showed that clinical risk-prediction models are helpful in predicting AKI onset in different critically ill populations^{14–20}. Even though these tools are of importance for the prevention, early diagnosis, and timely management of AKI, external validation of model performance and implementation of these tools for guiding interventions to prevent AKI or its progression are lacking. Notably, only few studies have focused on prediction of AKI recovery or post-AKI outcomes^{21–23}. Given the competing risk of death when evaluating kidney recovery in the critically ill population, the outcome of major adverse kidney events (MAKE), the composite of death, need of renal replacement therapy (RRT) and impaired kidney function, has gained recognition as a key metric of post-AKI prognosis²³. Therefore, the ability to identify which patients would develop MAKE after an AKI episode could be critical to guide clinical decisions and optimize delivery of healthcare services.

The aim of this study was to develop and validate useful clinical models for the risk-stratification of critically ill adult patients that suffered from AKI, for their individual risk of hospital mortality and MAKE. We hypothesized that clinical models that incorporate multimodal data beyond severity of AKI staging are informative in ICU settings and therefore provide paramount prognostic information.

METHODS

Study Design and Study Population

We conducted a multicenter, retrospective cohort study. The derivation cohort consisted of adult patients admitted to the ICU at the University of Kentucky (UKY) Hospital from March 2009 to February 2017. The validation cohort consisted of adult patients admitted to the ICU at the University of Texas Southwestern (UTSW) Medical Center from May 2009 to December 2015. Both cohorts were linked to the United States Renal Data System (USRDS). We included patients aged ≥ 18 years that stayed in the ICU for at least 1 day, had a minimum of two serum creatinine (sCr) measurements in the first 3 days of ICU admission, and had a diagnosis of AKI in the first 3 days of ICU admission based on sCr-KDIGO criteria²⁴. Patients were excluded if they had a diagnosis of end-stage kidney disease (ESKD) before or at the time of hospital admission, had undergone kidney transplant, had a baseline estimated glomerular filtration rate (eGFR) <15 mL/min/1.73m², or if they died within the first 3 days of ICU admission (Figure 1). If patients had multiple ICU admissions, only data from the first ICU admission were included in the analysis. If the time gap between consecutive ICU admissions was <12 hours, we considered them as a single ICU admission. The protocol was approved by the Institutional Review Boards (IRB) of both institutions (UKY: 17-0444-P1G and UTSW: STU 112015-069). Given the retrospective nature of the study, informed consent was waived.

Study Variables and Definitions

Data included demographics, comorbidities, hospital diagnoses and procedures, and critical illness parameters. Comorbidities were determined by ICD-9/10-CM codes and classified according to Charlson²⁵ and Elixhauser²⁶ scores. Chronic kidney disease (CKD) was defined using baseline eGFR <60 ml/min/1.73 m² or ICD-9/10-CM codes. Acuity of illness was determined by components of the Sequential Organ Failure Assessment (SOFA)²⁷ and the Acute Physiology and Chronic Health Evaluation II (APACHE II)²⁸ scores. Laboratory data, vital sign data, use of mechanical ventilation/extracorporeal organ support, and pressor/inotrope support were extracted from electronic health records (EHRs) flowsheets. Descriptive data of relevant clinical parameters are presented as mean ± standard deviation or median (interquartile range) depending on the distribution of continuous variables, and frequency (%) for categorical variables (Table 1).

Data Extraction, Validation, and Harmonization

Data were automatically extracted from EHRs and properly validated with a 5% review of individual EHRs for specific clinical parameters of interest. Data were extracted using the same definitions and rules across the two study sites. In total, there were 71 validated clinical variables included in the initial model, and they were all measured or determined during the first 3 days of ICU admission: from the time of ICU admission to 23:59:59 of the same day (Day 0) and the subsequent 3 days in the ICU (from 00:00:00 to 23:59:59 in Day 1, 2 and 3). Highest or lowest values of continuous parameters were selected for model classification. A detailed description and definitions, including missing data, of study variables are provided in Table S1. Missing data of SOFA or APACHE II individual components were imputed as the lowest attributable score. Missing data of baseline sCr were imputed by resolving the eGFR-EPI equation to 75. All other missing data were imputed using mean. Fluid overload (FO) was estimated as cumulative fluid balance (liters) in reference to ICU admission weight (kg) using the formula²⁹: $FO = (\text{fluid intake} - \text{fluid output}) / (\text{admission weight}) \times 100\%$.

Study Outcomes

Two clinical outcomes were evaluated: 1) hospital mortality and 2) MAKE in the 120 days following hospital discharge. MAKE was also evaluated in survivors only. MAKE was defined following 4-tiered criteria: (A) If the patient died within 120 days following hospital discharge (or prior to discharge); (B) If the patient survived but was identified as recipient of any type of RRT within the last 48 hours of hospital stay based on EHR data; (C) If the patient survived but developed ESKD within 120 days following hospital discharge based on EHR data and/or the USRDS; (D) If the patient survived and was not identified as ESKD or as recipient of RRT within the last 48 hours of hospital stay, the 120-day eGFR was used to compute the change in eGFR in reference to the baseline value. The 120-day eGFR was calculated by using the median of all sCr values within a window of 30 days closest but before 120 days after discharge. If the eGFR at 120 days was <50% of the baseline eGFR, this was considered MAKE. eGFR was computed using the EPI equation.³⁰

Model Feature Selection

We assessed all 71 features by training the models using the derivation cohort (UKY). Specifically, four machine learning algorithms: logistic regression (LR), random forest (RF), support vector machine (SVM) and extreme gradient boosting (XGBoost) were used for training the models independently for the prediction of hospital mortality and MAKE. Each feature was then ranked according to its correlation coefficient or relative importance in each trained model. In conjunction with the feature ranking, clinical reasoning was used to filter features according to explainability (ability of features to justify the results) and feasibility (access to data for reproducibility). The feature selection processes resulted in a total of 15 features in the developed clinical model for hospital mortality prediction and 14 features for MAKE prediction. The resulting feature ranking is provided in Table S2. The overall feature selection processes are illustrated in Supplementary Figure 1.

Model Predictive Performance Evaluation

Predictive performance of proposed clinical models was evaluated using 10-fold cross-validation in the derivation cohort (UKY), along with external validation in the UTSW cohort. The performances of proposed clinical models using four machine learning algorithms (LR, RF, SVM and XGBoost) are provided in Table S3. We selected RF as the main algorithm for our proposed clinical model due to its slightly superior performance compared to other algorithms in the derivation cohort. The proposed clinical models were compared with the corresponding baseline models (hospital mortality: SOFA; MAKE: maximum AKI KDIGO staging in ICU Day 0–3). We selected LR for the baseline models due to its superior performance than RF in single feature scenario. First, we computed ROC-AUC, accuracy (proportion of observations that are correctly predicted), precision/positive predictive value (PPV) (proportion of observations that are true positive among all predicted positives), sensitivity, specificity, F1 (harmonic mean of precision and sensitivity), negative predictive value (NPV) (proportion of observations that are true negative among all predicted negatives), calibration intercept and slope with 95% confidence intervals for proposed and baseline models³¹. The cut-off for binary prediction of all parameters was 0.5. We further compared the performance between proposed and baseline models by 1) the difference in AUC, whereas statistical significance was computed using Delong's test and 2) the categorical net reclassification improvement (NRI)^{32, 33}. We defined two risk categories for NRI using a cut-off value of 0.5 (<0.5 and ≥0.5) and provided reclassification tables for these categories stratified according to the occurrence of events (i.e., hospital mortality or MAKE). Performance metrics were computed on the validation data of UKY and the whole UTSW cohort without down-sampling. In addition, we examined the observed and predicted risk probabilities of outcomes graphically by bar charts in the two study cohorts.

Feature Interpretation and Visualization

We used the SHapley Additive exPlanations (SHAP) framework for interpreting specific feature importance in the proposed models³⁴. SHAP assigns each feature an importance score for the prediction of the outcome. A positive SHAP value for a particular feature indicates the feature increases the predicted risk of the outcome, while a negative value indicates the feature reduces the predicted risk of the outcome. We also illustrated SHAP

scores with specific feature data in selected patients for case-utility visualization. All analyses were conducted in SAS 9.4 (SAS Institute, Cary, NC), R programming (R Foundation for Statistical Computing, Vienna, Austria), and Python 3 (Python Software Foundation, Beaverton, OR).

RESULTS

Clinical Characteristics

A total of 7,354 AKI patients were included in the UKY and 2,233 in the UTSW cohort (Figure 1). In the UKY cohort, the median age [IQ1-IQ3] was 63 [51.7–73.6] years, 55.9% were male, 91.6% were white and 6.8% black, while in the UTSW cohort, the median age was 64 [53.8–73.7] years, 57.2% were male, 62.3% were white and 22.6% black. Prevalent hypertension (UKY 76.5% and UTSW 74.2%) and diabetes (UKY 41.6% and UTSW 41.5%) were frequent. Median SOFA scores were 6 [4–8] in both cohorts. More than 40% of patients in both cohorts required mechanical ventilation and more than 15% required pressor/inotrope support. A summary of patient characteristics, including AKI characteristics, is provided in Table 1. The overall hospital mortality rates were 21.9% and 9.9% in the UKY and UTSW cohorts, respectively. MAKE occurred in 32.4% of patients in the UKY cohort and in 25.7% of patients in the UTSW cohort (Table S4).

Prediction of Hospital Mortality

The developed clinical model consisted of 15 features that were interpreted in order of relevance according to SHAP (Figure 2). For example, exposure to pressor/inotrope, urine output, highest FiO₂, lowest platelet count and highest blood urea nitrogen were among the top predictors in both cohorts. Descriptive data of the 15 features are presented in Table S5. Selected case studies with specific feature values are summarized in Supplementary Figure 2.

The clinical model outperformed SOFA score for the prediction of hospital mortality in both the derivation (UKY) and validation (UTSW) cohorts: AUC (95%CI) 0.79 (0.79–0.80) vs. 0.71 (0.71–0.71), $p < 0.001$ in the UKY cohort and 0.74 (0.73–0.74) vs. 0.71 (0.71–0.71), $p < 0.001$ in the UTSW cohort. The clinical model improved the sensitivity and the NPV for hospital mortality prediction in both cohorts. The clinical model significantly improved hospital mortality risk-reclassification when compared to the SOFA score as per NRI in the UKY cohort (Table 2). Detailed reclassification of hospital mortality prediction is reported in Table S6.

In the UKY cohort, 474 patients had <10% predicted risk of hospital mortality and only 7 (1.5%) died. In contrast, among 2,851 patients with more than 50% predicted risk of mortality, 1,171 (41.1%) patients actually died. Similar risk-classification was observed in the UTSW cohort (Figure 3A). The clinical model exhibited a similar overall better performance when compared to the APACHE II score.

Prediction of MAKE

The developed clinical model consisted of 14 features that were interpreted in order of relevance according to SHAP (Figure 4). For example, last AKI KDIGO severity score at D0–D3, urine output, highest FiO₂, need of pressor/inotrope and highest blood urea nitrogen were among the top predictors in both cohorts. Descriptive data of the 14 features are presented in Table S7. Selected case studies with specific feature values are summarized in Supplementary Figure 3.

The clinical model outperformed maximum AKI KDIGO score for the prediction of MAKE: UKY cohort AUC of 0.78 (0.78–0.78) vs. 0.66 (0.66–0.66), $p<0.001$ and UTSW cohort 0.73 (0.72–0.74) vs. 0.67 (0.67–0.67), $p<0.001$. The clinical model improved the sensitivity and the NPV for MAKE prediction in both cohorts. The clinical model significantly improved MAKE risk-reclassification when compared to the maximum AKI KDIGO score as per NRI in both cohorts (Table 3). Detailed reclassification of MAKE prediction is reported in Table S8.

In the UKY cohort, 214 patients had <10% predicted risk of MAKE and only 10 (4.7%) of these patients truly developed the outcome. On the contrary, 2,984 patients had more than 50% predicted risk of MAKE, and 1,661 (55.7%) of these patients experienced the event. Similar risk-classification was observed in the UTSW cohort (Figure 3B).

Prediction of MAKE in Survivors

The clinical model had better performance than the maximum AKI KDIGO score to predict MAKE in survivors up to 120 days after hospital discharge (n=5,395 in UKY and 1,864 in UTSW cohorts): UKY cohort AUC of 0.84 (0.83–0.84) vs. 0.79 (0.79–0.80) for AKI KDIGO, $p<0.001$ and UTSW cohort 0.79 (0.79–0.80) vs. 0.76 (0.76–0.76) for AKI KDIGO, $p<0.001$. Risk reclassification by NRI was significantly improved in both cohorts (Table 4). Detailed reclassification of MAKE prediction in this subset of patients is reported in Table S9. Among UKY cohort survivors, 416 patients had <10% predicted risk of MAKE and only 6 (1.4%) of these patients developed the outcome. In contrast, 1,305 patients had >50% predicted risk of MAKE and 320 (24.5%) of these patients developed the outcome. Similar risk-classification was observed in the UTSW cohort.

Prediction of Hospital Mortality and MAKE stratified by CKD at baseline

Similar predictive performances were observed when the cohorts were stratified according to CKD status at baseline (Table S10).

Online Tool

We created an online tool that assists with computing the score for the proposed clinical models of hospital mortality and MAKE prediction. This tool will help to risk-stratify critically ill adult patients with incident AKI in the first 3 days of ICU admission and is available and accessible on-line for public use free-of-charge at <http://phenomics.uky.edu/taki/taki.html>.

DISCUSSION

The main deliverable of this study is the development and validation of pragmatic clinical models using data from the first 3 days of ICU stay to predict hospital mortality and major adverse kidney events up to 120 days post-hospital discharge in critically ill patients admitted to the ICU that had incident AKI within the first 3 days of ICU stay. Unique aspects of these clinical models are 1) the use of machine-learning algorithms and processes of explainability/clinical rationale for final variable selection and 2) the use of validated and harmonized ICU datasets from two large academic medical centers (UKY: development cohort and UTSW: external validation cohort). Our study provides an online platform available to the public for the bedside application of the proposed models and highlights the need to further develop artificial intelligence-based clinical models for enhanced and continuous prediction of post-AKI outcomes by addition of multiorgan, multimodal and dynamic clinical data.

Most clinical AKI studies using machine learning methodology to date, have focused on the development and validation of risk-classification models for continuous prediction of incident AKI in hospitalized patients^{31, 35–38}. Further, only few studies have tailored risk-prediction models for AKI onset in critically ill patients in the ICU^{14, 39}. Specific to the critically ill population, concerns have been raised for the feasibility of external validation, the lack of multimodal and multiorgan time-varying data, and the difficulties in expanding clinical data with biomarker or extracorporeal support device data for enhancement of the clinical models⁴⁰. Although still at infancy levels, the continuous evolution of artificial intelligence applied to biomedical research will enable the ability to analyze multimodal big data that were previously inaccessible with the ultimate goal of assisting clinical decisions at the bedside, particularly in vulnerable populations such as those with acute illnesses in the ICU⁴¹.

Specifically, risk-classification of mortality or major adverse kidney events post-AKI is an area with limited research. This is particularly important in the ICU setting as critically ill patients represent a vulnerable and growing population with high morbidity and mortality, and high incidence of AKI as it is estimated that about half of the 4 to 6 million patients admitted to ICUs every year in the U.S. suffer from AKI⁴². A randomized clinical trial including cardiac surgery patients at high risk of postoperative AKI showed that timely implementation of AKI bundles by a specialized team significantly decreased incident postoperative AKI when compared to standard of care⁴³. In contrast, education to clinicians and nurses about AKI bundles without implementation action or enrichment to include high-risk patients did not prevent progression of AKI in critically ill patients admitted to the ICU⁴⁴. It is therefore possible that identification of high-risk AKI patients using clinical risk-classification tools may better direct efforts for implementation of AKI bundles early in the course of AKI or ICU admission with considerable impact on clinical outcomes. Further, the use of these risk-classification tools may direct care of critically ill survivors of AKI to specialized post-discharge clinics that could assist in promotion of kidney and critical illness recovery and improvement of patient-centered outcomes⁴⁵.

Currently, prediction of outcome in critically ill patients with AKI relies solely on clinical rationale but development of robust risk prediction approaches using EHR data is possible. One should note that about two-thirds of the final features in the developed clinical models are currently part of risk-classification tools used in the general ICU population, supporting the generalizability of the models^{27, 28}. Importantly, clinical models such as the ones developed and validated in this study can be incorporated into the EHR as part of clinical decision support systems to guide bedside decisions to promote AKI recovery and/or use for enrichment of clinical trials to tailor high-risk groups at higher chances of benefit from the intervention.

Our study has some limitations. First, the retrospective nature of the study could carry information bias inherent to large databases generated through automated EHR extraction. Nonetheless, we validated ~5% of the data with emphasis in parameters used in the present study. Second, we did not use urine output criterion as part of the AKI definition to derive the cohort of patients with incident AKI within the first 3 days of ICU stay but the clinical models included urine output as a relevant parameter for post-AKI outcome prediction. Third, we propose clinical models that were carefully developed and externally validated only for prediction of mortality and MAKE in critically ill adult patients with incident AKI within the first 3 days of ICU stay. Therefore, our models may not be applicable for patients who develop AKI later in the course of ICU stay or for those who develop AKI outside the ICU setting. Nonetheless, one should recognize that about two-thirds of AKI events in the ICU are diagnosed either on ICU admission or within the first 3–7 days of ICU stay⁴². Fourth, our clinical models were compared to reference models such as SOFA, APACHE II and maximum AKI KDIGO score, which are risk-classification scores that have not been specifically developed or validated for outcome prediction in the subset of AKI patients included in this study or to specifically predict MAKE. Fifth, our clinical models did not include time-varying (dynamic) clinical data or novel kidney biomarker data, which may further enhance the prediction of post-AKI outcomes. This is work in the pipeline for our group and others.

We developed and validated readily applicable clinical models using data from the first 3 days of ICU stay to predict hospital mortality and major adverse kidney events up to 120 days post-hospital discharge in critically ill patients admitted to the ICU with incident AKI within the first 3 days of ICU stay. We used large, harmonized and validated ICU datasets from two academic medical centers that represent a heterogeneous ICU population. The proposed clinical models exhibited good performance for outcome prediction and could enable risk-stratification for timely interventions that promote or enhance kidney recovery, both during ICU stay (KDIGO AKI bundle, resource allocation) and after discharge from the hospital (specialized follow-up care). We provide an online platform available to the public for applicability of the clinical models at the bedside. Additional external validation is needed to warrant wide implementation of these clinical models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Support:

This study was supported in part by the University of Texas Southwestern Medical Center O'Brien Kidney Research Core Center (NIH, P30 DK079328-06) and the University of Kentucky Center for Clinical and Translational Science (NIH/NCATS, UL1TR001998). JAN is supported by grants from NIDDK (R01DK128208, R56 DK126930 and P30 DK079337) and NHLBI (R01 HL148448-01). Funders did not have a role in study design, data collection, analysis, reporting, or the decision to submit this publication.

REFERENCES

1. Susantitaphong P, Cruz DN, Cerda J, et al. World incidence of AKI: a meta-analysis. *Clinical journal of the American Society of Nephrology : CJASN*. 2013;8:1482–1493. [PubMed: 23744003]
2. Szczech LA, Harmon W, Hostetter TH, et al. World Kidney Day 2009: problems and challenges in the emerging epidemic of kidney disease. *Journal of the American Society of Nephrology : JASN*. 2009;20:453–455. [PubMed: 19244576]
3. Rewa O, Bagshaw SM. Acute kidney injury[mdash]epidemiology, outcomes and economics. *Nat Rev Nephrol*. 2014;10:193–207. [PubMed: 24445744]
4. Bouchard J, Acharya A, Cerda J, et al. A Prospective International Multicenter Study of AKI in the Intensive Care Unit. *Clinical journal of the American Society of Nephrology : CJASN*. 2015;10:1324–1331. [PubMed: 26195505]
5. Uchino S, Kellum JA, Bellomo R, et al. Acute renal failure in critically ill patients: a multinational, multicenter study. *Jama*. 2005;294:813–818. [PubMed: 16106006]
6. Wu VC, Wu CH, Huang TM, et al. Long-term risk of coronary events after AKI. *Journal of the American Society of Nephrology : JASN*. 2014;25:595–605. [PubMed: 24503241]
7. Wu VC, Wu PC, Wu CH, et al. The impact of acute kidney injury on the long-term risk of stroke. *Journal of the American Heart Association*. 2014;3 (4):e000933. doi: 10.1161/JAHA.114.000933. [PubMed: 25027018]
8. Odutayo A, Wong CX, Farkouh M, et al. AKI and Long-Term Risk for Cardiovascular Events and Mortality. *Journal of the American Society of Nephrology : JASN*. 2017;28:377–387. [PubMed: 27297949]
9. Chawla LS, Amdur RL, Amodeo S, Kimmel PL, Palant CE. The severity of acute kidney injury predicts progression to chronic kidney disease. *Kidney International*. 2011;79:1361–1369. [PubMed: 21430640]
10. Chawla LS, Eggers PW, Star RA, Kimmel PL. Acute kidney injury and chronic kidney disease as interconnected syndromes. *The New England journal of medicine*. 2014;371:58–66. [PubMed: 24988558]
11. Heung M, Steffick DE, Zivin K, et al. Acute Kidney Injury Recovery Pattern and Subsequent Risk of CKD: An Analysis of Veterans Health Administration Data. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2016;67:742–752. [PubMed: 26690912]
12. Thakar CV, Christianson A, Himmelfarb J, Leonard AC. Acute kidney injury episodes and chronic kidney disease risk in diabetes mellitus. *Clinical journal of the American Society of Nephrology : CJASN*. 2011;6:2567–2572. [PubMed: 21903988]
13. Silver SA, Long J, Zheng Y, Chertow GM. Cost of Acute Kidney Injury in Hospitalized Patients. *Journal of hospital medicine*. 2017;12:70–76. [PubMed: 28182800]
14. Malhotra R, Kashani KB, Macedo E, et al. A risk prediction score for acute kidney injury in the intensive care unit. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*. 2017;32:814–822. [PubMed: 28402551]
15. 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;114:2208–2216; quiz 2208. [PubMed: 17088458]
16. Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP. A clinical score to predict acute renal failure after cardiac surgery. *Journal of the American Society of Nephrology : JASN*. 2005;16:162–168. [PubMed: 15563569]

17. James MT, Pannu N, Hemmelgarn BR, et al. Derivation and External Validation of Prediction Models for Advanced Chronic Kidney Disease Following Acute Kidney Injury. *Jama*. 2017;318:1787–1797. [PubMed: 29136443]
18. Kheterpal S, Tremper Kevin K, Englesbe Michael J, et al. Predictors of Postoperative Acute Renal Failure after Noncardiac Surgery in Patients with Previously Normal Renal Function. *Anesthesiology*. 2007;107:892–902. [PubMed: 18043057]
19. Kiers HD, van den Boogaard M, Schoenmakers MC, et al. Comparison and clinical suitability of eight prediction models for cardiac surgery-related acute kidney injury. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*. 2013;28:345–351. [PubMed: 23222415]
20. Chawla LS, Abell L, Mazhari R, et al. Identifying critically ill patients at high risk for developing acute renal failure: a pilot study. *Kidney Int*. 2005;68:2274–2280. [PubMed: 16221229]
21. Itenov TS, Berthelsen RE, Jensen JU, et al. Predicting recovery from acute kidney injury in critically ill patients: development and validation of a prediction model. *Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine*. 2018;20:5460.
22. Lee BJ, Hsu CY, Parikh R, et al. Predicting Renal Recovery After Dialysis-Requiring Acute Kidney Injury. *Kidney international reports*. 2019;4:571–581. [PubMed: 30993232]
23. Grams ME, Sang Y, Coresh J, et al. Candidate Surrogate End Points for ESRD after AKI. *Journal of the American Society of Nephrology : JASN*. 2016;27:2851–2859. [PubMed: 26857682]
24. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney inter., Suppl* 2012; 2: 1–138.
25. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*. 1987;40:373–383. [PubMed: 3558716]
26. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical care*. 1998;36:8–27. [PubMed: 9431328]
27. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive care medicine*. 1996;22:707–710. [PubMed: 8844239]
28. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Critical care medicine*. 1985;13:818–829. [PubMed: 3928249]
29. Goldstein SL, Currier H, Graf C, Cosio CC, Brewer ED, Sachdeva R. Outcome in children receiving continuous venovenous hemofiltration. *Pediatrics*. 2001;107:1309–1312. [PubMed: 11389248]
30. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med*. 1999;130.
31. Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. *Critical care medicine*. 2018;46:1070–1077. [PubMed: 29596073]
32. 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. *Statistics in medicine*. 2008;27:157–172; discussion 207–112. [PubMed: 17569110]
33. Kundu Suman A. YSA Janssens Cecile J.W.. PredictABEL: Assessment of Risk Prediction Models 2020.
34. Lundberg SMA L Su-In. A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems: Curran Associates Inc.*; 2017:4768–4777.
35. Churpek MM, Carey KA, Edelson DP, et al. Internal and External Validation of a Machine Learning Risk Score for Acute Kidney Injury. *JAMA Network Open*. 2020;3:e2012892–e2012892. [PubMed: 32780123]

36. Hsu C-N, Liu C-L, Tain Y-L, Kuo C-Y, Lin Y-C. Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization From Electronic Health Records: Development and Validation Study. *J Med Internet Res*. Vol 222020:e16903.
37. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572:116–119. [PubMed: 31367026]
38. Simonov M, Ugwuowo U, Moreira E, et al. A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: A descriptive modeling study. *PLoS medicine*. 2019;16:e1002861. [PubMed: 31306408]
39. Flechet M, Guiza F, Schetz M, et al. AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. *Intensive care medicine*. 2017;43:764–773. [PubMed: 28130688]
40. Neyra JA, Leaf DE. Risk Prediction Models for Acute Kidney Injury in Critically Ill Patients: Opus in Progressu. *Nephron*. 2018;140:99–104. [PubMed: 29852504]
41. Chan L, Vaid A, Nadkarni GN. Applications of machine learning methods in kidney disease: hope or hype? *Curr Opin Nephrol Hypertens*. 2020;29:319–326. [PubMed: 32235273]
42. Hoste EA, Bagshaw SM, Bellomo R, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive care medicine*. 2015;41:1411–1423. [PubMed: 26162677]
43. Meersch M, Schmidt C, Hoffmeier A, et al. Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. *Intensive care medicine*. 2017;43:1551–1561. [PubMed: 28110412]
44. Koeze J, van der Horst ICC, Wiersema R, et al. Bundled care in acute kidney injury in critically ill patients, a before-after educational intervention study. *BMC nephrology*. 2020;21:381. [PubMed: 32883219]
45. Siew ED, Liu KD, Bonn J, et al. Improving Care for Patients after Hospitalization with Acute Kidney Injury. *Journal of the American Society of Nephrology : JASN*. 2020.

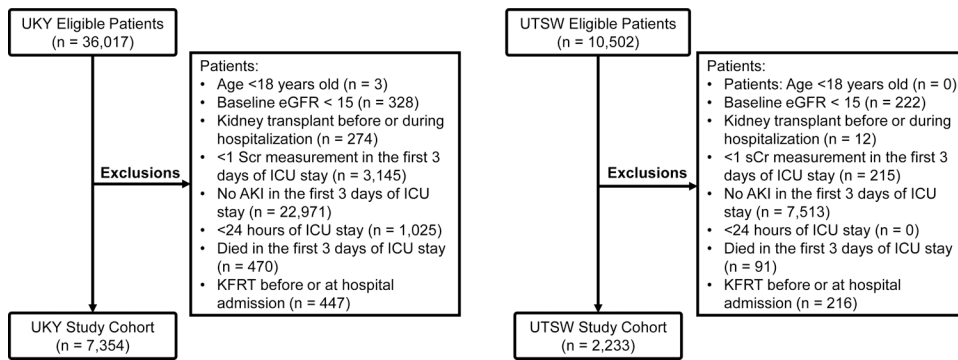


Figure 1.
Flowchart of derivation (UKY) and validation (UTSW) study cohorts.

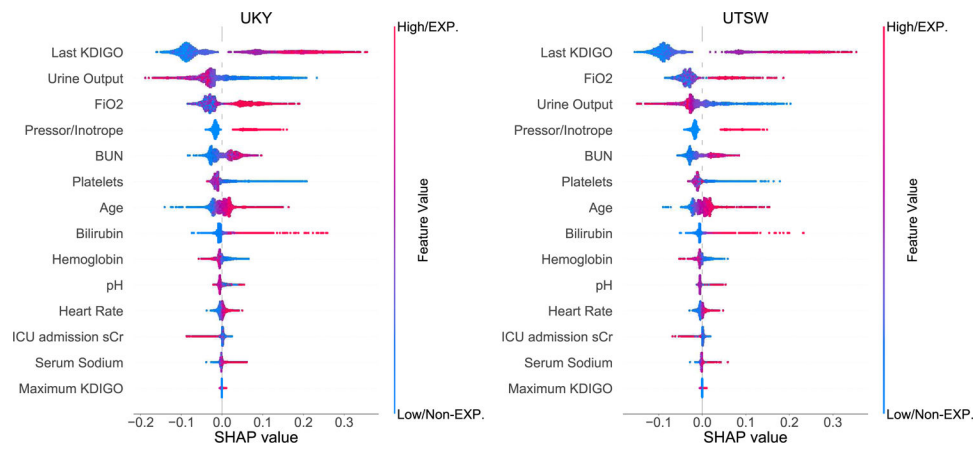


Figure 2. SHapley Additive exPlanations (SHAP) framework of top 15 final features for the prediction of hospital mortality. Each dot denotes a SHAP value for a feature in a particular individual. The dot color represents if the feature is present or absent (categorical data) or high vs. low (continuous data), whereas blue denotes absent/low and red denotes present/high. The X-axis represents the scale of SHAP values. A positive SHAP value for a particular feature indicates the feature increases the predicted risk of the outcome, while a negative value indicates the feature reduces the predicted risk of the outcome. The Y-axis includes features ordered from top to bottom by their impact on the model prediction based on mean absolute SHAP values. EXP =exposure; Non-EXP =non-exposure.

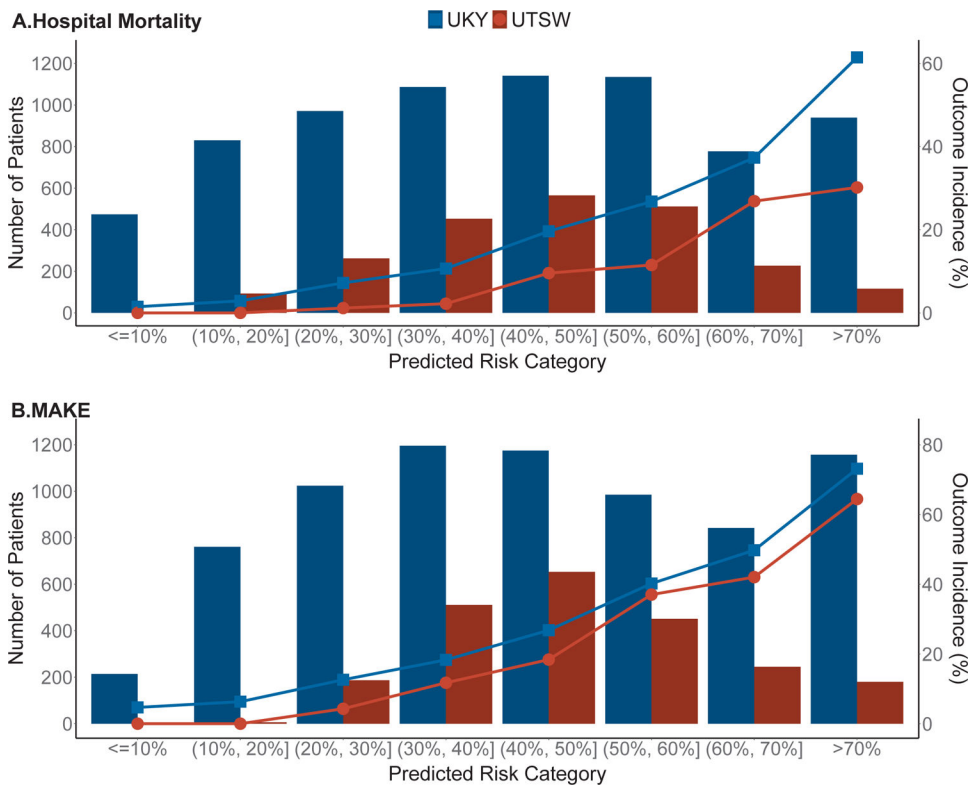


Figure 3. Predicted risk vs. outcome incidence for (A) hospital mortality and (B) MAKE in both study cohorts. The column bar denotes the number of patients in each category and the line chart denotes the outcome incidence.

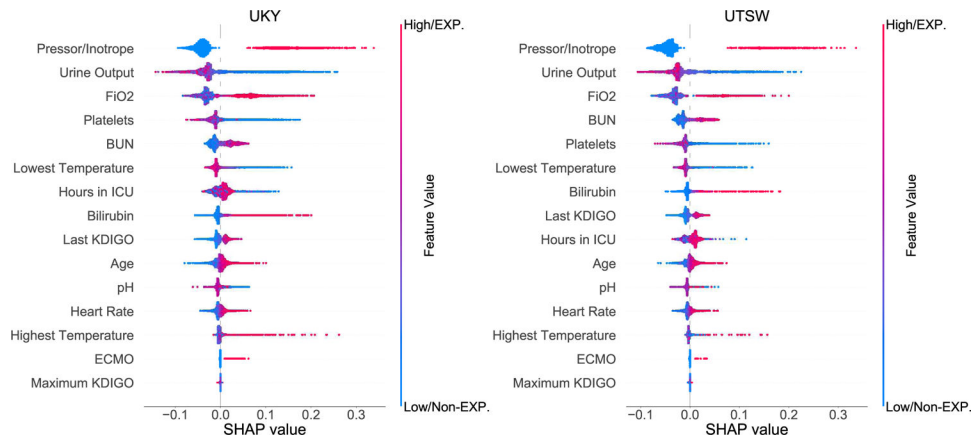


Figure 4. SHapley Additive exPlanations (SHAP) framework of top 14 final features for the prediction of MAKE. Each dot denotes a SHAP value for a feature in a particular individual. The dot color represents if the feature is present or absent (categorical data) or high vs. low (continuous data), whereas blue denotes absent/low and red denotes present/high. The X-axis represents the scale of SHAP values. A positive SHAP value for a particular feature indicates the feature increases the predicted risk of the outcome, while a negative value indicates the feature reduces the predicted risk of the outcome. The Y-axis includes features ordered from top to bottom by their impact on the model prediction based on mean absolute SHAP values. EXP =exposure; Non-EXP =non-exposure.

Table 1.

Patient characteristics in the study cohorts

Variables	UKY Cohort (Derivation) n = 7354	UTSW Cohort (Validation) n = 2233
Age, years, median [IQR25th-75 th]	63.02 [51.71–73.55]	63.98 [53.84–73.68]
Gender, male, n (%)	4111 (55.90)	1277 (57.19)
Race, n (%)		
- White	6734 (91.57)	1392 (62.34)
- Black	501 (6.81)	504 (22.57)
- Other	119 (1.62)	337 (15.09)
BMI, kg/m ² , median [IQR25th-75 th]	29.07 [24.62–34.95]	28.65 [24.44–34.25]
Comorbidity		
Charlson Score, median [IQR25th-75 th]	3.00 [1.00–5.00]	4.00 [2.00–6.00]
Elixhauser Score, median [IQR25th-75 th]	6.00 [4.00–8.00]	6.00 [4.00–8.00]
Diabetes, n (%)	3062 (41.64)	927 (41.51)
Hypertension, n (%)	5623 (76.46)	1657 (74.21)
Baseline eGFR, mL/min/1.73m ² , median [IQR25th-75 th] #	78.52 [54.53–99.21]	70.31 [47.29–95.13]
CKD, n (%) #	604 (8.21)	439 (19.66)
Critical Illness characteristics		
SOFA Score *, median [IQR25th-75 th]	6.00 [4.00–8.00]	6.00 [4.00–8.00]
APACHE II Score *, median [IQR25th-75 th]	19.00 [16.00–23.00]	17.00 [14.00–22.00]
Pressor/Inotrope *, n (%)	1568 (21.32)	327 (15.20)
Hospital LOS, days, median [IQR25th-75 th]	11.49 [6.44–21.29]	10.34 [6.43–18.27]
ICU LOS *, hours, median [IQR25th-75 th]	76.55 [61.17–83.6]	76.02 [54.43–81.45]
ECMO *, n (%)	101 (1.37)	12 (0.54)
Mechanical ventilation *, n (%)	3822 (51.97)	914 (40.93)
MV days *, median [IQR25th-75 th]	3.00 [1.27–3.43]	3.00 [1.00–3.36]
Sepsis on ICU admission, n (%)	1942 (26.41)	456 (20.42)

Variables	UKY Cohort (Derivation) n = 7354	UTSW Cohort (Validation) n = 2233
Urine output *, mL/h, median [IQR25th-75 th]	68.90 [39.67–110.42]	60.01 [39.42–90.13]
Urine flow *, mL/kg/h, median [IQR25th-75 th]	0.81 [0.45–1.37]	0.72 [0.47–1.08]
Fluid Overload *, %, median [IQR25th-75 th]	4.36 [0.01–10.76]	0.70 [–1.69 to 3.68]
Laboratory Data		
Bicarbonate *, mmol/L, median [IQR25th-75 th]	21.50 [18.50–24.50]	22.50 [19.50–25.00]
BUN *, mg/dL, median [IQR25th-75 th]	39.00 [27.00–58.00]	31.00 [21.00–48.00]
Hematocrit *, %, median [IQR25th-75 th]	30.45 [26.40–35.50]	30.20 [26.49–34.80]
Hemoglobin *, g/dL, median [IQR25th-75 th]	10.10 [8.75–11.85]	10.05 [8.80–11.60]
AKI Characteristics		
Baseline sCr, mg/dL, median [IQR25th-75 th] #	0.96 [0.75–1.25]	1.04 [0.78–1.40]
ICU admission sCr, mg/dL, median [IQR25th-75 th]	1.60 [1.15–2.31]	1.45 [1.00–2.15]
Peak sCr D0–D3, mg/dL, median [IQR25th-75 th]	1.94 [1.47–2.84]	1.86 [1.40–2.70]
Last sCr D0–D3, mg/dL, median [IQR25th-75 th]	1.47 [1.08–2.15]	1.46 [1.05–2.09]
Maximum KDIGO D0–D3, n (%)		
- Stage 1	3564 (48.46)	1291 (57.81)
- Stage 2	1786 (24.29)	457 (20.47)
- Stage 3	2004 (27.25)	485 (21.72)
Last KDIGO D0–D3, n (%)		
- No AKI	2641 (35.91)	872 (39.05)
- Stage 1	2089 (28.41)	720 (32.24)
- Stage 2	1085 (14.75)	265 (11.87)
- Stage 3	1539 (20.93)	376 (16.84)
Inpatient Dialysis Characteristics		
CRRT* only, n (%)	576 (7.83)	158 (7.08)
HD* only, n (%)	243 (3.30)	26 (1.16)
CRRT and HD both*, n (%)	5 (0.07)	3 (0.13)

Variables	UKY Cohort (Derivation) n = 7354	UTSW Cohort (Validation) n = 2233
Total days of CRRT [*] , median [IQR25th-75 th]	3.00 [2.00-4.00]	3.00 [2.00-3.00]
Total days of HD [*] , median [IQR25th-75 th]	2.46 [1.59-3.00]	2.00 [1.00-3.00]

* Measured or calculated with data within the first 3 days of ICU admission.

Reported only for patients with measured baseline sCr.

Abbreviations: AKI (acute kidney injury), APACHE (acute physiologic assessment and chronic health evaluation), BMI (body mass index), BUN (blood urea nitrogen), CKD (chronic kidney disease), CRRT (continuous renal replacement therapy), D0-D3 (from day 0 to day 3 of ICU admission), ECMO (extracorporeal membrane oxygenation), HD (hemodialysis), IABP (intra-aortic balloon pump), ICU (intensive care unit), IQR (interquartile range), KDIGO (Kidney Disease: Improving Global Outcomes), LOS (length of stay), MV (mechanical ventilation), SCr (serum creatinine), SOFA (sequential organ failure assessment), UKY (University of Kentucky), UTSW (University of Texas Southwestern).

Table 2. Predictive performance of models for hospital mortality prediction in the derivation (UKY) and validation (UTSW) cohorts.

Performance metrics (95% CI)	SOFA Score		APACHE II Score		Clinical model*	
	Derivation Cohort (UKY)	Validation Cohort (UTSW)	Derivation Cohort (UKY)	Validation Cohort (UTSW)	Derivation Cohort (UKY)	Validation Cohort (UTSW)
AUC	0.71(0.71-0.71)	0.71(0.71-0.71)	0.69(0.68-0.69)	0.67(0.67-0.67)	0.79 (0.79-0.80)	0.74 (0.73-0.74)
Difference in AUC (vs. SOFA)	-	-	-	-	0.08	0.03
- P value	-	-	-	-	<0.001	<0.001
Difference in AUC (vs. APACHE II)	-	-	-	-	0.10	0.07
- P value	-	-	-	-	<0.001	<0.001
Accuracy	0.65(0.64-0.65)	0.73(0.73-0.73)	0.65(0.64-0.65)	0.52(0.50-0.54)	0.71 (0.71-0.71)	0.65 (0.64-0.66)
Precision	0.34(0.34-0.35)	0.19(0.19-0.19)	0.34(0.33-0.34)	0.14(0.13-0.14)	0.41 (0.40-0.42)	0.18 (0.17-0.18)
Sensitivity	0.68(0.67-0.68)	0.55(0.55-0.55)	0.63(0.62-0.64)	0.73(0.70-0.75)	0.72 (0.72-0.73)	0.69 (0.67-0.71)
Specificity	0.64(0.63-0.64)	0.75(0.75-0.75)	0.65(0.65-0.65)	0.50(0.47-0.52)	0.71 (0.70-0.71)	0.64 (0.63-0.65)
F1	0.46(0.45-0.46)	0.29(0.29-0.29)	0.44(0.43-0.45)	0.23(0.23-0.23)	0.52 (0.52-0.53)	0.28 (0.27-0.29)
PPV	0.34(0.34-0.35)	0.19(0.19-0.19)	0.34(0.33-0.34)	0.14(0.13-0.14)	0.41 (0.40-0.42)	0.18 (0.17-0.18)
NPV	0.88(0.87-0.88)	0.94(0.94-0.94)	0.86(0.86-0.87)	0.94(0.94-0.94)	0.90 (0.90-0.90)	0.95 (0.95-0.95)
Calibration Intercept	-1.27(-1.29 to -1.25)	-2.03(-2.03 to -2.02)	-1.28(-1.29 to -1.26)	-2.47(-2.47 to -2.46)	-1.25 (-1.27 to -1.24)	-2.23 (-2.26 to -2.20)
Calibration Slope	1.02(1.00-1.04)	1.12(1.08-1.16)	0.99(0.95-1.03)	0.75(0.73-0.78)	1.13 (1.11-1.14)	1.17 (1.12-1.21)
NRI % (vs. SOFA)						
- Categorical	-	-	-	-	0.12 [0.09 to 0.15]	0.05 [-0.03 to 0.13]
- P value	-	-	-	-	<0.001	0.20
NRI % (vs. APACHE II)						
- Categorical	-	-	-	-	0.15 [0.12 to 0.18]	0.12 [0.04 to 0.20]
- P value	-	-	-	-	<0.001	<0.001

* The proposed clinical model included 15 features, refer to the Methods section for details. The machine learning method for the reported performance evaluation of SOFA and APACHE II is Logistic Regression and for the proposed clinical model is Random Forest.

Abbreviations: APACHE (acute physiologic assessment and chronic health evaluation), AUC (area under the curve), CI (confidence interval), F1 (F1 score), NPV (negative predictive value), NRI (net reclassification index), PPV (positive predictive value), SOFA (sequential organ failure assessment), UKY (University of Kentucky), UTSW (University of Texas Southwestern)

Predictive performance of models for MAKE prediction in the derivation (UKY) and validation (UTSW) cohorts.

Table 3.

Performance metrics (95% CI)	Maximum AKI KDIGO Severity Score		Clinical model *	
	Derivation Cohort (UKY)	Validation Cohort (UTSW)	Derivation Cohort (UKY)	Validation Cohort (UTSW)
AUC	0.66(0.66–0.66)	0.67(0.67–0.67)	0.78 (0.78–0.78)	0.73 (0.72–0.74)
Difference in AUC (vs. KDIGO score)	-	-	0.12	0.06
- P value	-	-	<0.001	<0.001
Accuracy	0.62(0.61–0.63)	0.68(0.64–0.72)	0.72 (0.71–0.72)	0.67 (0.66–0.69)
Precision	0.45(0.44–0.46)	0.43(0.38–0.47)	0.55 (0.54–0.56)	0.42 (0.40–0.43)
Sensitivity	0.61(0.58–0.63)	0.54(0.47–0.62)	0.69 (0.69–0.70)	0.67 (0.64–0.69)
Specificity	0.62(0.60–0.65)	0.73(0.65–0.81)	0.73 (0.72–0.73)	0.68 (0.64–0.71)
F1	0.51(0.50–0.51)	0.47(0.46–0.47)	0.61 (0.61–0.62)	0.51 (0.50–0.52)
PPV	0.45(0.44–0.46)	0.43(0.38–0.47)	0.55 (0.54–0.56)	0.42 (0.40–0.43)
NPV	0.77(0.77–0.78)	0.82(0.82–0.83)	0.83 (0.83–0.83)	0.85 (0.85–0.86)
Calibration Intercept	-0.74(-0.75 to -0.73)	-0.97(-0.98 to -0.96)	-0.72 (-0.73 to -0.70)	-1.10 (-1.17 to -1.02)
Calibration Slope	1.02(0.99–1.05)	1.25(1.20–1.29)	1.07 (1.05–1.09)	1.34 (1.27–1.41)
NRI % (vs. KDIGO score)				
- Categorical	-	-	0.20[0.17 to 0.22]	0.11[0.06 to 0.16]
- P value	-	-	<0.001	<0.001

* The proposed clinical model included 14 features, refer to the Methods section for details. The machine learning method for the reported performance evaluation of maximum AKI KDIGO severity score is Logistic Regression and for the proposed clinical model is Random Forest.

Abbreviations: APACHE (acute physiologic assessment and chronic health evaluation), AUC (area under the curve), CI (confidence interval), F1 (score), KDIGO (Kidney Disease: Improving Global Outcomes), MAKE (major adverse kidney events), NPV (negative predictive value), NRI (net reclassification index), PPV (positive predictive value), SOFA (sequential organ failure assessment), UKY (University of Kentucky), UTSW (University of Texas Southwestern)

Predictive performance of models for MAKE prediction in survivors up to 120 days following hospital discharge. These included 5395 AKI survivors (7.8% developed MAKE) in the UKY cohort and 1864 survivors (11% developed MAKE) in the UTSW cohort.

Table 4.

Performance metrics (95% CI)	Maximum AKI KDIGO Severity Score		Clinical model*	
	Derivation Cohort (UKY)	Validation Cohort (UTSW)	Derivation Cohort (UKY)	Validation Cohort (UTSW)
AUC	0.79(0.79–0.80)	0.76(0.76–0.76)	0.84 (0.83–0.84)	0.79 (0.79–0.80)
Difference in AUC (vs. KDIGO score)	-	-	0.05	0.03
- P value	-	-	<0.001	<0.001
Accuracy	0.80(0.79–0.80)	0.82(0.82–0.82)	0.79 (0.79–0.80)	0.83 (0.82–0.84)
Precision	0.23(0.23–0.23)	0.33(0.33–0.33)	0.24 (0.23–0.24)	0.34 (0.32–0.36)
Sensitivity	0.69(0.68–0.69)	0.58(0.58–0.58)	0.75 (0.74–0.76)	0.58 (0.55–0.61)
Specificity	0.81(0.80–0.81)	0.85(0.85–0.85)	0.80 (0.79–0.80)	0.86 (0.84–0.88)
F1	0.34(0.34–0.35)	0.42(0.42–0.42)	0.36 (0.36–0.37)	0.43 (0.42–0.44)
PPV	0.23(0.23–0.23)	0.33(0.33–0.33)	0.24 (0.23–0.24)	0.34 (0.32–0.36)
NPV	0.97(0.97–0.97)	0.94(0.94–0.94)	0.97 (0.97–0.98)	0.94 (0.94–0.95)
Calibration Intercept	-2.48(-2.51 to -2.45)	-1.82(-1.85 to -1.80)	-2.47 (-2.50 to -2.44)	-1.85 (-1.99 to -1.71)
Calibration Slope	1.08(1.05–1.11)	1.03(0.98–1.08)	0.97 (0.95–0.99)	1.28 (1.14–1.41)
NRI % (vs. KDIGO score)				
- Categorical	-	-	0.07 [0.04 to 0.10]	0.04 [0.00 to 0.07]
- P value	-	-	<0.001	0.03

* The proposed clinical model included 14 features, refer to the Methods section for details. The machine learning method for the reported performance evaluation of maximum AKI KDIGO severity score is Logistic Regression and for the proposed clinical model is Random Forest.

Abbreviations: APACHE (acute physiologic assessment and chronic health evaluation), AUC (area under the curve), CI (confidence interval), F1 (F1 score), KDIGO (Kidney Disease: Improving Global Outcomes), MAKE (major adverse kidney events), NPV (negative predictive value), NRI (net reclassification index), PPV (positive predictive value), SOFA (sequential organ failure assessment), UKY (University of Kentucky), UTSW (University of Texas Southwestern)