







Predictive Approaches for Acute Dialysis Requirement and Death in COVID-19

Akhil Vaid,^{1,2} Lili Chan,^{3,4} Kumardeep Chaudhary ^{1,3} Suraj K. Jaladanki ¹ Ishan Paranjpe ¹ Adam Russak ¹ Arash Kia,^{1,5} Prem Timsina,^{1,5} Matthew A. Levin,^{1,6,7} John Cijiang He,⁴ Erwin P. Böttinger,^{2,8} Alexander W. Charney,^{1,9,10,11} Zahi A. Fayad,^{1,12} Steven G. Coca ⁴ Benjamin S. Glicksberg ^{1,2,7} and Girish N. Nadkarni,^{1,2,3,4} on behalf of MSCIC*

Abstract

Background and objectives AKI treated with dialysis initiation is a common complication of coronavirus disease 2019 (COVID-19) among hospitalized patients. However, dialysis supplies and personnel are often limited.

Design, setting, participants, & measurements Using data from adult patients hospitalized with COVID-19 from five hospitals from the Mount Sinai Health System who were admitted between March 10 and December 26, 2020, we developed and validated several models (logistic regression, Least Absolute Shrinkage and Selection Operator (LASSO), random forest, and eXtreme GradientBoosting [XGBoost; with and without imputation]) for predicting treatment with dialysis or death at various time horizons (1, 3, 5, and 7 days) after hospital admission. Patients admitted to the Mount Sinai Hospital were used for internal validation, whereas the other hospitals formed part of the external validation cohort. Features included demographics, comorbidities, and laboratory and vital signs within 12 hours of hospital admission.

Results A total of 6093 patients (2442 in training and 3651 in external validation) were included in the final cohort. Of the different modeling approaches used, XGBoost without imputation had the highest area under the receiver operating characteristic (AUROC) curve on internal validation (range of 0.93–0.98) and area under the precision-recall curve (AUPRC; range of 0.78–0.82) for all time points. XGBoost without imputation also had the highest test parameters on external validation (AUROC range of 0.85–0.87, and AUPRC range of 0.27–0.54) across all time windows. XGBoost without imputation outperformed all models with higher precision and recall (mean difference in AUROC of 0.04; mean difference in AUPRC of 0.15). Features of creatinine, BUN, and red cell distribution width were major drivers of the model's prediction.

Conclusions An XGBoost model without imputation for prediction of a composite outcome of either death or dialysis in patients positive for COVID-19 had the best performance, as compared with standard and other machine learning models.

CJASN 16: 1158–1168, 2021. doi: <https://doi.org/10.2215/CJN.17311120>

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected >103 million people worldwide (1). Initial reports from China found low rates of AKI, but follow-up studies from Europe and the United States have found that AKI affects up to 40% of patients hospitalized with coronavirus disease 2019 (COVID-19), which is associated with significant morbidity and mortality (2–4). A large proportion of these patients, approximately 20%, will ultimately be treated with dialysis (3).

During the height of the pandemic in New York City, dialysis availability was limited due to shortages of dialysis nurses, dialysis machines and supplies, and personal protective equipment (5). Many hospitals had to reduce dialysis treatment times and use alternative strategies (e.g., acute peritoneal dialysis

and potassium binders between dialysis treatments) to manage the increased number of patients (6,7). Additionally, patients with AKI had significantly higher mortality than the patients who did not have AKI. Accurate prediction of acute dialysis and of those at risk of mortality early in the hospitalization course of patients with COVID-19 will allow for the establishment of kidney protective measures, the close monitoring of kidney function, and early and informed conversations with patients and family members regarding goals of care.

Machine learning models can harness the disparate data collected during clinical care in electronic health records (EHRs) for accurate outcome predictions. Several machine learned models have been published since the COVID-19 pandemic started, but they have not addressed AKI and dialysis (8). However, several

Due to the number of contributing authors, the affiliations are listed at the end of this article.

Correspondence:

Dr. Girish N. Nadkarni, Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1243, New York, NY 10128. Email: Girish.nadkarni@mountsinai.org

machine learning models for detection of AKI have been developed. Fletchet *et al.* (9) previously demonstrated that a machine learned model had similar discriminative performance as physicians; however, physicians tended to overestimate AKI compared with the machine learned model. Two additional studies have attempted to build models for continuous prediction of AKI, with some success (10,11). We aimed to develop and validate a machine-learning model to predict a composite end point of AKI treated with dialysis or death in patients hospitalized with COVID-19 early in the hospital course.

Materials and Methods

Data Sources and Inclusion Criteria

We used EHR data from patients with laboratory-confirmed SARS-CoV-2 infection between March 10 and December 26, 2020. Patient data were sourced from five hospitals in the Mount Sinai Health System (MSHS): Mount Sinai Hospital (MSH), Mount Sinai Morningside, Mount Sinai West, Mount Sinai Brooklyn, and Mount Sinai Queens. We described the rates of AKI and patient characteristics in a subset of these patients in our previous article (3). Patients were classified into those admitted to MSH (internal validation and training set), and patients from other hospitals (OH) formed the external validation set. The MSH hospital, located at the intersection of the Upper East Side and Harlem, is the largest hospital in the MSHS and serves as the referral center for the other hospitals. Therefore, patients at MSH tended to be in a more acute state than those in the other hospitals. However, the other MSHS hospitals are in diverse locations (Queens, Brooklyn, *etc.*) and, therefore, the internal and external hospital systems are ethnically, geographically, and socioeconomically diverse, improving generalizability.

We included all inpatients who were >18 years of age and had a laboratory-confirmed SARS-CoV-2 infection, as determined by a positive SARS-CoV-2 RT-PCR result. We included only patients who had a positive COVID-19 test performed 48 hours before or after hospital admission.

Feature Selection

Study data included patient demographics (age, sex, reported race, and ethnicity), comorbidities derived from International Classification of Diseases 10 (ICD-10) codes, first vital sign measurement, first laboratory value reported within the initial 12 hours of admission, and the use of vasopressors during the first 12 hours of hospital admission. We chose this 12-hour time window because the reporting of laboratory measures was often delayed during the start of the pandemic due to hospital resource limitations. To correctly contextualize pulse oximetry values, we created three features containing the lowest measured oxygen saturation, the highest measured fraction of inspired oxygen, and whether the patient was on mechanical ventilation within the initial 12 hours of admission. Features and outcomes (including mechanical ventilation and dialysis treatment) were extracted from the Clarity database, from Epic systems, and the COVID-19 registry built by the Clinical Data Science group at Mount Sinai. Procedures were derived from procedure orders and flow sheets.

Data collected after initiation of the first session of dialysis were dropped, and patients were excluded if they had a session of dialysis within the initial 12 hours of admission, because our first prediction time point was 1 day posthospitalization. We excluded features with >30% missing data across patients, followed by excluding patients with >30% missing data across the remaining features. We excluded patients who had kidney failure on dialysis or had a prior kidney transplant (Figure 1). For models unable to function in the presence of missing data, we performed k-nearest neighbor (k=5) imputation on the remaining dataset. The full list of features included in the model and the proportions of missing data are presented in Table 1.

Outcome

The primary outcome was a composite outcome of death or acute dialysis treatment within time periods of 1, 3, 5, or 7 days after admission.

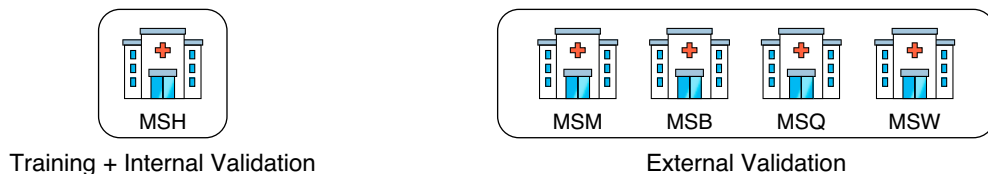
Model Development and Selection

We trained and tested several models, including a boosted decision tree model named eXtreme Gradient Boosting (XGBoost) (12), in addition to a popular ensemble machine-learning model called random forest. XGBoost was chosen for its resilience to missing data; resistance to overfitting in datasets with imbalanced feature/outcome ratios; availability of hyperparameters, which allow tuning for imbalanced datasets; and explainability of predictions using SHapley Additive exPlanations (SHAP) scores. SHAP scores are a game-theoretic approach to model interpretability that provide explanations of global model structure on the basis of combinations of local explanations for each prediction (13).

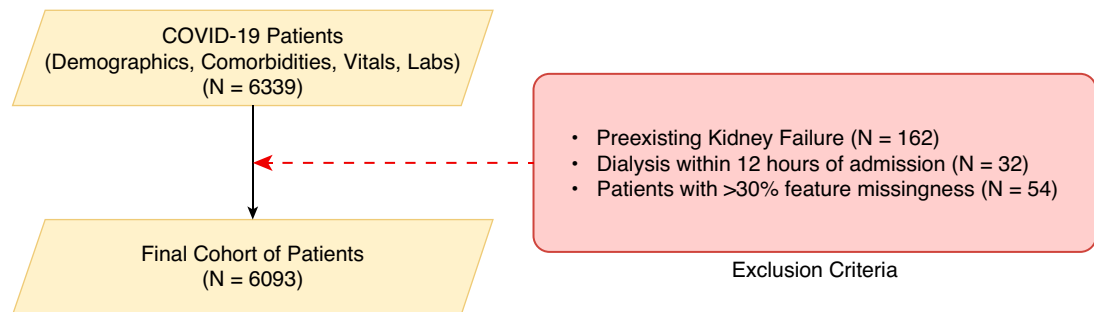
As a baseline comparison of model performance, we utilized logistic regression (using all of the features used in the model) and logistic regression with L1 regularization (LASSO) models within the training and testing pipeline. Additionally, to demonstrate the effect of imputation on XGBoost performance, we also trained and tested additional XGBoost models on the imputed dataset.

All models were internally validated on MSH data for each of the time points after admission (14). Internal validation signals that the algorithm is capable of discerning patterns within a single population, independently of a single training-testing data split. To demonstrate generalizability, all models were then trained on the entirety of data from MSH for the corresponding time frame, followed by external validation on patients from other hospitals. Model performance was evaluated using the area under the receiver operating characteristic (AUROC) curve, the area under the precision-recall curve (AUPRC), accuracy, sensitivity, specificity, and F-1 score. Whereas AUROC plots the true positive rate (sensitivity) on the *y* axis and the false positive rate (1-specificity) on the *x* axis, the AUPRC has the precision (also called positive predictive value) on the *y* axis and the recall (also called sensitivity) on the *x* axis. Although both AUROC and AUPRC provide model performance metrics across different probability thresholds, AUPRC is a better metric in imbalanced datasets, such as in our case, where the outcome (dialysis or death) occurred in a minority of patients and the focus is on identifying patients who had that outcome (15). When the dataset is imbalanced,

A. Facilities



B. Data



C. Timeline

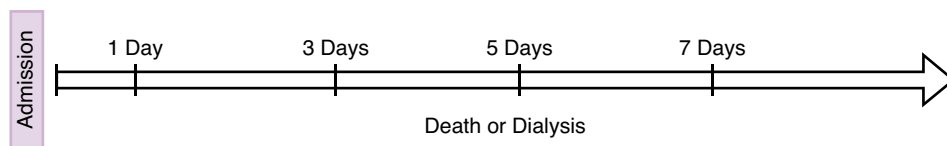


Figure 1. | Overall schema of study. (A) Mount Sinai Hospital (MSH) was used for training and internal validation, whereas the other hospitals were used for external validation. (B) Study flow diagram. (C) Model predictions were generated for hospital day 1, 3, 5, and 7. COVID-19, coronavirus disease 2019; MSB, Mount Sinai Brooklyn; MSM, Mount Sinai Morningside; MSQ, Mount Sinai Queens; MSW, Mount Sinai West.

meaning there are a large number of true negative outcomes, a change in the number of false positives will lead to small changes in the false positive rate used in the AUROC. However, because the positive predictive value used in the AUPRC compares the true positives with the false positives, it is, therefore, more sensitive to changes in the number of false positives than AUROC. Although the AUPRC can range from zero to one like in AUROC, the baseline depends on the fraction of positive cases where an AUPRC above this fraction would be considered a good model.

To calculate metrics that require a probability threshold, we calculated and averaged the optimal threshold derived using the Youden J statistic across all bootstrap iterations. Area under the curve values and reported 95% confidence intervals were generated through 500 bootstrap iterations, each with a unique random seed, using the normal bootstrap method (16).

Hyperparameter Optimization

Hyperparameter optimization for the XGBoost and random forest models was performed using randomized grid searching with 5000 discrete grid options, with each option subjected to a further ten-fold stratified crossvalidation. For both of the LASSO models, an exhaustive grid search was performed by varying the inverse regularization (C)

parameter for a total of 100 grid options, with each option subjected to a further ten-fold stratified crossvalidation. The logistic regression model was unpenalized and had no hyperparameters to optimize for.

Evaluation of Calibration

Calibration is a postprocessing technique that adjusts per-case probability predicted by a model according to the observed outcome, thereby improving error distribution and allowing for the use of model outputs as risk scores. We performed isotonic and sigmoid calibration on the probability estimates reported by each model, and evaluated calibration results on the basis of the Brier score, which measures the difference between the predicted and actual outcome (17). Isotonic calibration is nonparametric, but generally more data hungry; isotonic calibration is less data hungry and, therefore, less prone to overfit. The sigmoid calibration method is based on the Platt logistic model, which assumes that the calibration curve can be corrected by applying a sigmoid function to the raw predictions (18).

Model Interpretation

We evaluated which features were most responsible for the XGBoost model predictions by using SHAP scores (13). All analyses were performed using the pandas, scikit-learn,

Table 1. Patient characteristics stratified by acute dialysis treatment and death at hospital day 7

Feature	No Dialysis or Death		Dialysis or Death		Missing (%) ^a
	MSH (n=2244)	OH (n=3183)	MSH (n=198)	OH (n=468)	
Age (yr), median (IQR)	62 (47–72)	68 (55–79)	70 (59–79)	78 (67–86)	0
Male, n (%)	1279 (57)	1739 (54)	135 (62)	292 (61)	0
Race, n (%)					
Asian	91 (4)	167 (5)	5 (2)	26 (5)	0
Black	418 (19)	928 (29)	72 (33)	126 (26)	0
Not sure	1 (0)	1 (0)	0 (0)	0 (0)	0
Other	882 (39)	1163 (37)	79 (36)	167 (35)	0
Pacific Islander	10 (1)	3 (0)	0 (0)	0 (0)	0
Unknown	118 (5)	82 (3)	10 (5)	15 (3)	0
White	719 (32)	823 (26)	51 (23)	146 (30)	0
Comorbidities, n (%)					
Asthma	195 (9)	174 (6)	14 (7)	24 (5)	0
Atrial fibrillation	169 (8)	196 (6)	26 (13)	42 (9)	0
COPD	109 (5)	145 (5)	15 (8)	27 (6)	0
Cancer	272 (12)	199 (6)	31 (16)	24 (5)	0
Chronic viral hepatitis	63 (3)	40 (1)	7 (4)	7 (2)	0
Coronary artery disease	300 (13)	410 (13)	43 (22)	82 (18)	0
Diabetes	493 (22)	568 (18)	65 (33)	86 (18)	0
HIV	37 (2)	49 (2)	5 (3)	3 (1)	0
Hypertension	763 (34)	819 (26)	97 (49)	116 (25)	0
Heart failure	199 (9)	266 (8)	34 (17)	65 (14)	0
Obesity	320 (14)	246 (8)	43 (22)	31 (7)	0
Obstructive sleep apnea	159 (7)	75 (2)	18 (9)	14 (3)	0
CKD	81 (4)	123 (4)	38 (17)	28 (6)	0
Admission laboratory values, median (IQR)					
Anion gap (mEq/L)	11 (10–13)	12 (10–14)	14 (12–17)	14 (12–17)	3
Albumin (g/dl)	3.0 (2.6–3.4)	3.0 (2.6–3.4)	2.7 (2.3–3.0)	2.7 (2.4–3.05)	9
Alkaline phosphatase (U/L)	75 (59–101)	73 (58–98)	81 (61–112)	81 (62–109)	9
Alanine transaminase (U/L)	29 (18–50)	29 (18–52)	29 (16–53)	35 (22–66)	9
Aspartate transaminase (U/L)	37 (26–60)	39 (26–63)	48 (28–87)	64 (42–111)	9
Basophil (%)	0.2 (0.1–0.4)	0.3 (0.1–0.5)	0.2 (0.1–0.3)	0.2 (0.1–0.3)	13
Bicarbonate (mEq/L)	24 (22–27)	23 (20–25)	23 (19–26)	21 (17–24)	3
BUN (mg/dl)	16 (11–25)	17 (12–28)	44 (27–70)	40 (22–65)	3
Calcium (mEq/L)	8.4 (8.0–8.8)	8.2 (7.8–8.6)	8.2 (7.8–8.6)	7.9 (7.5–8.4)	3
Chloride (mEq/L)	103 (100–106)	104 (101–107)	103 (98–108)	106 (101–112)	3
Serum creatinine (mg/dl)	0.9 (0.7–1.1)	0.9 (0.7–1.3)	2.8 (1.4–5.8)	1.7 (1.1–2.9)	3
C-reactive protein (mg/L)	95 (41–174)	96 (42–186)	167 (96–253)	201 (121–280)	20
Eosinophil (%)	0.3 (0.1–1.0)	0.3 (0.2–0.9)	0.3 (0.1–0.7)	0.2 (0.1–0.4)	13
Blood glucose (mg/dl)	114 (96–156)	114 (91–160)	137 (104–187)	144 (108–211)	3
Hematocrit (%)	38.4 (34.1–42.4)	38.9 (34.6–42.5)	35.0 (29.0–39.65)	38.5 (33.8–43.5)	2
Hemoglobin (g/dl)	12.5 (11.1–13.7)	12.6 (11.2–13.8)	11.1 (9.48–12.92)	12.35 (10.8–13.9)	2
Lymphocyte (n)	0.9 (0.6–1.3)	1.0 (0.7–1.5)	0.7 (0.5–1.0)	0.9 (0.6–1.2)	9
Lymphocyte (%)	14.0 (8.4–21.0)	14.5 (8.9–22.3)	9.0 (5.0–16.3)	8.7 (5.3–14.0)	7
Mean corpuscular hemoglobin concentration (g/dl)	29 (28–31)	30 (28–31)	29 (27–31)	30 (28–32)	2
Mean corpuscular volume (fl)	89 (85–92)	92 (88–95)	89 (85–94)	94 (89–98)	3
Monocyte (n)	0.4 (0.3–0.6)	0.5 (0.3–0.7)	0.4 (0.2–0.6)	0.4 (0.3–0.7)	9
Monocyte (%)	6.4 (4.1–9.2)	6.75 (4.5–9.4)	4.6 (2.9–7.3)	4.4 (3.0–6.5)	6
Mean platelet volume (fl)	9.2 (8.1–10.4)	8.4 (7.6–9.4)	9.1 (8.2–10.0)	8.9 (8.0–9.8)	3
Neutrophil (n)	5.2 (3.3–7.97)	5.6 (3.8–8.8)	6.7 (4.0–11.4)	9.1 (5.9–13.3)	6
White blood cell count (10 ³ μl)	6.8 (4.8–9.8)	7.3 (5.4–10.3)	8.0 (5.4–12.4)	10.4 (7.0–14.4)	2
Admission vitals, median (IQR)					
Oxygen saturation (%)	92 (90–95)	93 (90–95)	89 (83–93)	90 (81–93)	0
FiO ₂	21 (21–21)	21 (21–21)	21 (21–50)	21 (21–21)	0
Pulse rate	86 (76–96)	85 (75–96)	88 (76–100)	91 (78–107)	0
Respiratory rate	19 (18–20)	18 (18–20)	20 (18–22)	19 (18–21)	0
Temperature (°F)	98.6 (97.8–99.7)	98.2 (97.5–98.9)	98.6 (97.5–99.9)	98.1 (97.5–99.0)	0
Diastolic BP (mm Hg)	70 (63–79)	73 (66–81)	70 (60–77)	70 (61–80)	0
Systolic BP (mm Hg)	126 (114–140)	128 (114–143)	132 (114–150)	124 (106–143)	0
Hospitalization events, n (%)					
Vasopressors	249 (11)	258 (8)	62 (28)	155 (32)	0
Intubation	151 (7)	181 (6)	31 (14)	38 (8)	0
Intensive care unit	246 (11)	239 (8)	57 (26)	80 (17)	0

MSH, Mount Sinai Hospital; OH, other hospitals; IQR, interquartile range; COPD, chronic obstructive pulmonary disease; FiO₂, fraction of inspired oxygen.

^aProportion with missing value from entire cohort.

Table 2. Performance (area under the receiver curve and area under the precision-recall curve) of different machine learning algorithms for prediction of death or dialysis need at hospital day 1, 3, 5, and 7

Algorithm	Day 1				Day 3				Day 5				Day 7			
	Internal Validation		External Validation		Internal Validation		External Validation		Internal Validation		External Validation		Internal Validation		External Validation	
	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)
LASSO	0.95 (0.95 to 0.95)	0.56 (0.55 to 0.57)	0.73 (0.72 to 0.73)	0.18 (0.17 to 0.18)	0.93 (0.93 to 0.94)	0.63 (0.63 to 0.64)	0.85 (0.85 to 0.86)	0.35 (0.35 to 0.36)	0.93 (0.93 to 0.93)	0.68 (0.68 to 0.69)	0.85 (0.85 to 0.85)	0.46 (0.46 to 0.46)	0.92 (0.92 to 0.92)	0.70 (0.70 to 0.71)	0.86 (0.86 to 0.86)	0.55 (0.55 to 0.55)
Logistic regression	0.87 (0.87 to 0.88)	0.16 (0.15 to 0.17)	0.71 (0.71 to 0.72)	0.06 (0.06 to 0.06)	0.86 (0.86 to 0.86)	0.24 (0.24 to 0.25)	0.77 (0.77 to 0.77)	0.17 (0.17 to 0.17)	0.85 (0.85 to 0.86)	0.28 (0.28 to 0.28)	0.78 (0.78 to 0.78)	0.27 (0.27 to 0.27)	0.84 (0.84 to 0.84)	0.33 (0.33 to 0.33)	0.81 (0.81 to 0.81)	0.36 (0.36 to 0.36)
Random forest	0.92 (0.92 to 0.93)	0.63 (0.62 to 0.64)	0.84 (0.84 to 0.84)	0.29 (0.29 to 0.29)	0.94 (0.94 to 0.94)	0.71 (0.71 to 0.72)	0.84 (0.84 to 0.84)	0.38 (0.38 to 0.39)	0.93 (0.93 to 0.93)	0.71 (0.71 to 0.72)	0.84 (0.84 to 0.84)	0.46 (0.46 to 0.46)	0.92 (0.92 to 0.92)	0.72 (0.72 to 0.73)	0.83 (0.83 to 0.83)	0.51 (0.51 to 0.51)
XGBoost (imputed)	0.96 (0.96 to 0.96)	0.67 (0.66 to 0.68)	0.82 (0.82 to 0.82)	0.21 (0.21 to 0.22)	0.95 (0.95 to 0.95)	0.73 (0.73 to 0.73)	0.86 (0.86 to 0.86)	0.39 (0.39 to 0.39)	0.93 (0.93 to 0.94)	0.73 (0.73 to 0.74)	0.85 (0.85 to 0.85)	0.47 (0.47 to 0.47)	0.92 (0.92 to 0.92)	0.73 (0.73 to 0.74)	0.84 (0.84 to 0.84)	0.52 (0.52 to 0.52)
XGBoost (nonimputed)	0.98 (0.98 to 0.98)	0.82 (0.81 to 0.83)	0.87 (0.87 to 0.87)	0.27 (0.27 to 0.27)	0.96 (0.96 to 0.96)	0.8 (0.8 to 0.81)	0.86 (0.86 to 0.86)	0.42 (0.42 to 0.42)	0.94 (0.94 to 0.95)	0.79 (0.79 to 0.79)	0.86 (0.86 to 0.86)	0.5 (0.5 to 0.5)	0.93 (0.93 to 0.93)	0.78 (0.77 to 0.78)	0.85 (0.85 to 0.85)	0.54 (0.54 to 0.54)

95% CI, 95% confidence interval; AUROC, area under the receiver operating characteristic; AUPRC, area under the precision-recall curve; LASSO, Least Absolute Shrinkage and Selection Operator; XGBoost, eXtreme Gradient Boosting.

Matplotlib, shap, SciPy, and XGBoost libraries, within a Python 3.9.1 environment (12, 18–22).

Results

The schema of the overall study design is shown in Figure 1. A total of 6093 patients (2442 in training and 3651 in external validation), admitted from March 10 to December 26, 2020, were included for analysis. Of the patients who were included for analysis, rates of death or dialysis treatment in the training dataset (MSH) were 1% within 1 day, 4% at 3 days, 6% at 5 days, and 8% at 7 days. Similarly, the rates of death or dialysis treatment in the external validation dataset (other hospitals within the MSHS) were 2% within 1 day, 5% at 3 days, 10% at 5 days, and 13% at 7 days (Supplemental Figure 1 and Supplemental Table 1). The baseline characteristics of these patients, stratified by treatment with dialysis or death at 7 days after hospital admission, are shown in Table 1.

Of the five machine learning models tested, the nonimputed XGBoost had the best performance, with an AUROC of 0.98 at 1 day, 0.96 at 3 days, 0.94 at 5 days, and 0.93 at 7 days on internal validation (Figure 2 and Table 2). The AUPRC was 0.82 at 1 day, 0.80 at 3 days, 0.79 at 5 days, and 0.78 at 7 days (Figure 3). XGBoost outperformed baseline regression models (logistic regression and LASSO), with values of between 0.84 and 0.95 for AUROC, and between 0.16 and 0.70 for AUPRC, on internal validation. Furthermore, the nonimputed XGBoost performed better than the imputed XGBoost models with regard to both AUROC (0.92–0.96) and AUPRC (0.67–0.73). The XGBoost models also outperformed the random forest models for both the AUROC (0.92–0.94) and AUPRC (0.63–0.72). In addition to the confidence intervals for these metrics being well separated, an ANOVA for the differences between AUROC and AUPRC between models was also significant ($P < 0.001$) across all time frames.

External validation on the dataset from the other hospitals confirmed that nonimputed XGBoost again outperformed all other models, with AUROC values ranging from 0.85 to 0.87, and AUPRC values ranging from 0.27 to 0.54, over all time windows (Figures 2 and 3 and Table 2). The imputed XGBoost performance was worse than nonimputed XGBoost, with an AUROC ranging between 0.82 and 0.86, and AUPRC ranging between 0.21 and 0.52. Random forest models also had worse performance than XGBoost, with the AUROC ranging between 0.83 and 0.84, and AUPRC ranging between 0.29 and 0.51 (Table 2). XGBoost without imputation outperformed all models, with a higher precision and recall (mean difference in AUROC of 0.04; mean difference in AUPRC of 0.15). In subgroup analysis of only patients who were admitted to the intensive care unit within 12 hours of hospital admission from the external validation cohort, the nonimputed XGBoost model had the best performance, with the AUROC ranging between 0.78 and 0.9 and AUPRC between 0.64 and 0.82 (Supplemental Table 2). Finally, as shown in Supplemental Figure 2, the Brier score is low and the statistical significance between predicted and actual proportions is not significant, indicating the nonimputed XGBoost model was well calibrated.

Detailed performance statistics for the models—including sensitivity, specificity, and positive and negative predictive values—are shown in Supplemental Figure 3 and

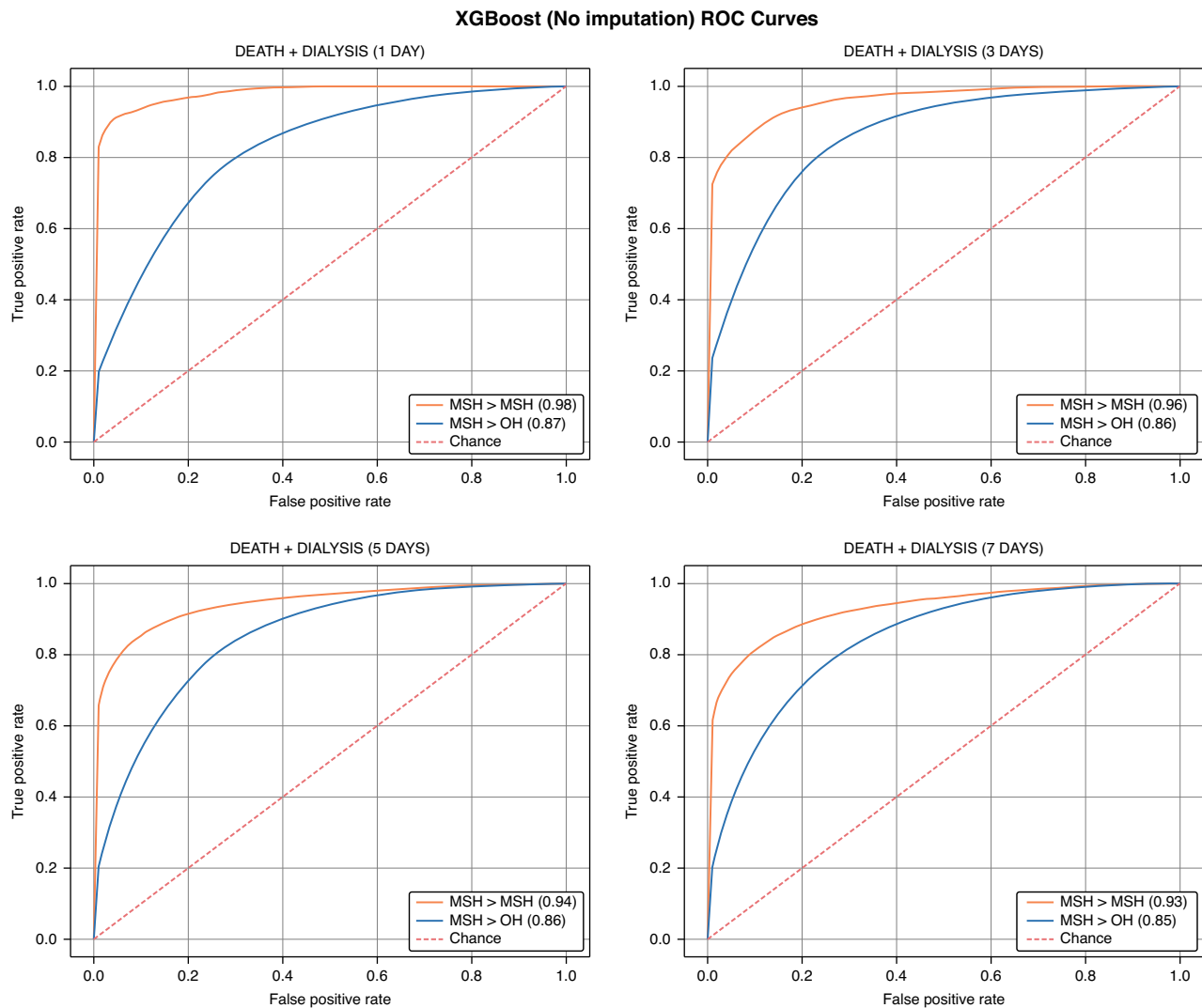


Figure 2. | XGBoost without imputation had high area under the receiver operating characteristic (AUROC) curves. AUROC curves for the eXtreme Gradient Boosting (XGBoost) model without imputation at hospital day 1, 3, 5, and 7. Orange line represents results in the training set, blue line represents external validation, and the dotted red line is chance. OH, other hospitals.

Supplemental Table 3. Using the optimal threshold derived using the Youden J statistic, nonimputed XGBoost had the highest sensitivity (ranging between 0.84 and 0.95) and specificity (0.9–0.96) across time points in the internal validation cohort. Results were similar in the external validation cohort, with sensitivity ranging between 0.82 and 0.90, and specificity ranging between 0.79 and 0.87. We calculated SHAP scores and created summary plots to illustrate feature importance with respect to model prediction for each time window (Figure 4 and Supplemental Figure 4). During all time horizons, serum creatinine at admission was one of the major features driving model predictions. Other clinically relevant features included BUN, systolic BP, age, and oxygen saturation.

Discussion

Using EHR data gathered within the first 12 hours of hospital admission from a demographically diverse

population of patients admitted to five hospitals in New York City during the initial COVID-19 surge, we developed, crossvalidated, and externally tested several predictive models for a composite outcome of acute dialysis treatment or death at various time windows after inpatient admission. We compared these models using well-established performance metrics and found that a boosted decision tree-based XGBoost model without imputation had superior performance. In particular, the precision (positive predictive value) and recall (sensitivity) were consistently significantly higher for a nonimputed XGBoost model than other approaches at most time points. This is important because a model with high positive predictive value minimizes false positives and can, therefore, help avoid clinician fatigue and alert burnout. Similarly, high sensitivity means the model will have fewer false negatives, thus maximizing the utility in identifying patients at need for dialysis or at risk for death. Finally, using SHAP values, we highlight several features that were important

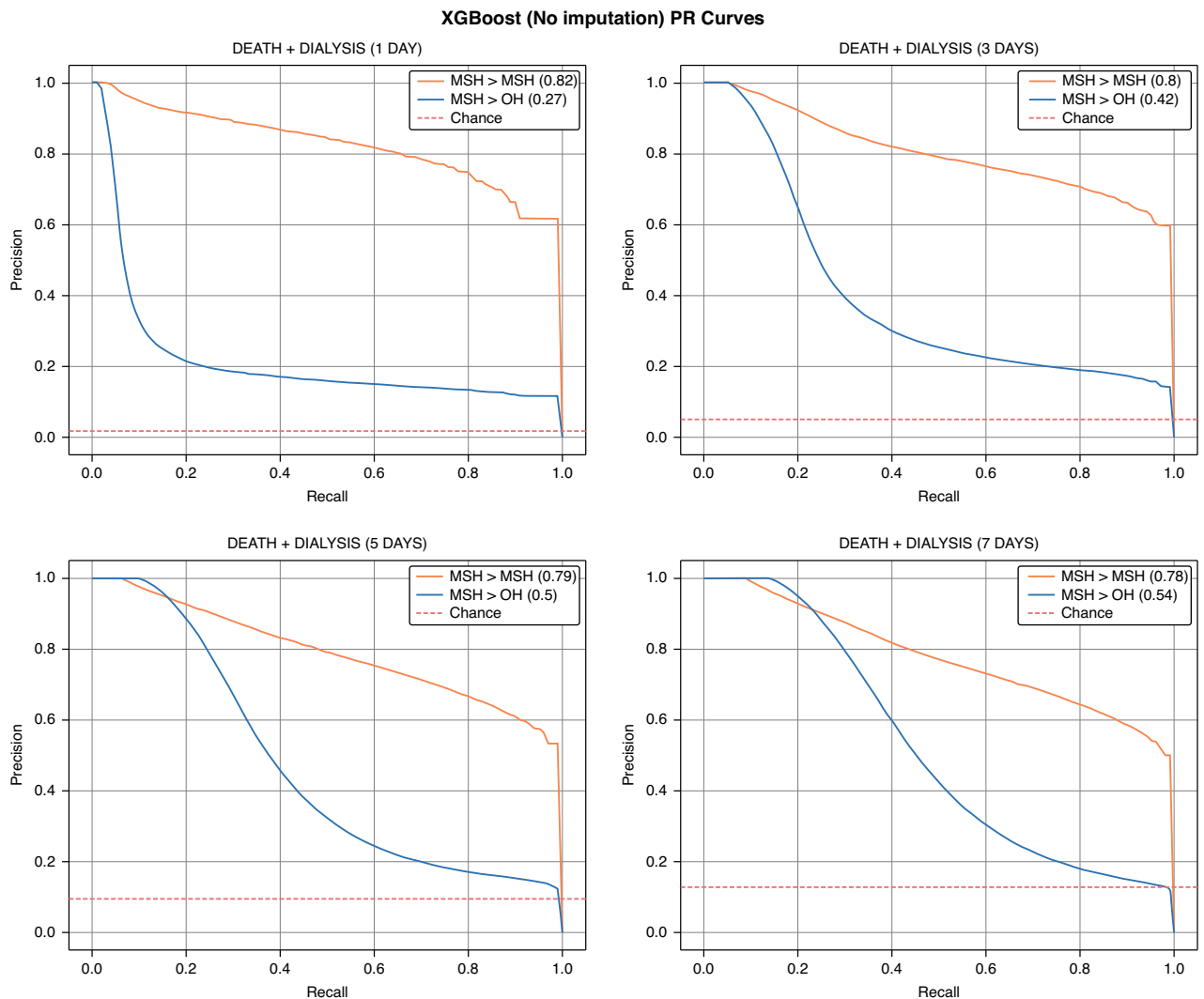


Figure 3. | XGBoost without imputation had high area under the precision-recall curves (AUPRC). AUPRC for the XGBoost model without imputation at hospital day 1, 3, 5, and 7. Orange line represents results in the training set, blue line represents external validation, and the dotted red line is chance. OH, other hospitals.

in driving the prediction of the XGBoost models across several time horizons.

Several machine learning prediction models have been developed in nephrology. Although there has been some variability in performance of these machine learned models, performance generally outperforms traditional statistical methods, such as logistic regression. The decision of which machine learning approach will perform best for the task is not possible to predict beforehand. Tree-based algorithms, such as random forest and XGBoost, are particularly popular because they have high accuracy, stability, and ability to map nonlinear relationships. Although both are ensemble techniques that build a varying number of decision trees on different subsamples and then take the average of the trees, XGBoost is graphics-processing-unit accelerated, supports second-order differentials, is able to handle missing data, and is considered better for unbalanced data.

The performance advantage of XGBoost without imputation persisted in external validation in a diverse group of

patients admitted to several different facilities. The lack of required imputation for optimal model performance is particularly helpful for model deployment in a hospital setting with relatively few patients with COVID-19. Although prospective validation is needed, these models may have a clinical effect in identifying patients at high risk of acute dialysis treatment and death. Such identification can inform providers regarding which patients require closer monitoring, the potential need for transfer to step-down or critical care units, and in identifying patients who would benefit from discussions regarding their clinical care. Before the COVID-19 pandemic, several groups had developed machine learned models, with good discrimination, for the prediction of AKI (10, 23). Building on these models, we chose to provide accurate predictions across a variety of time windows to allow for increased flexibility for hospital planning. We did not do a continuously updating model, although this is an area of future interest and would provide additional insight into clinical care.

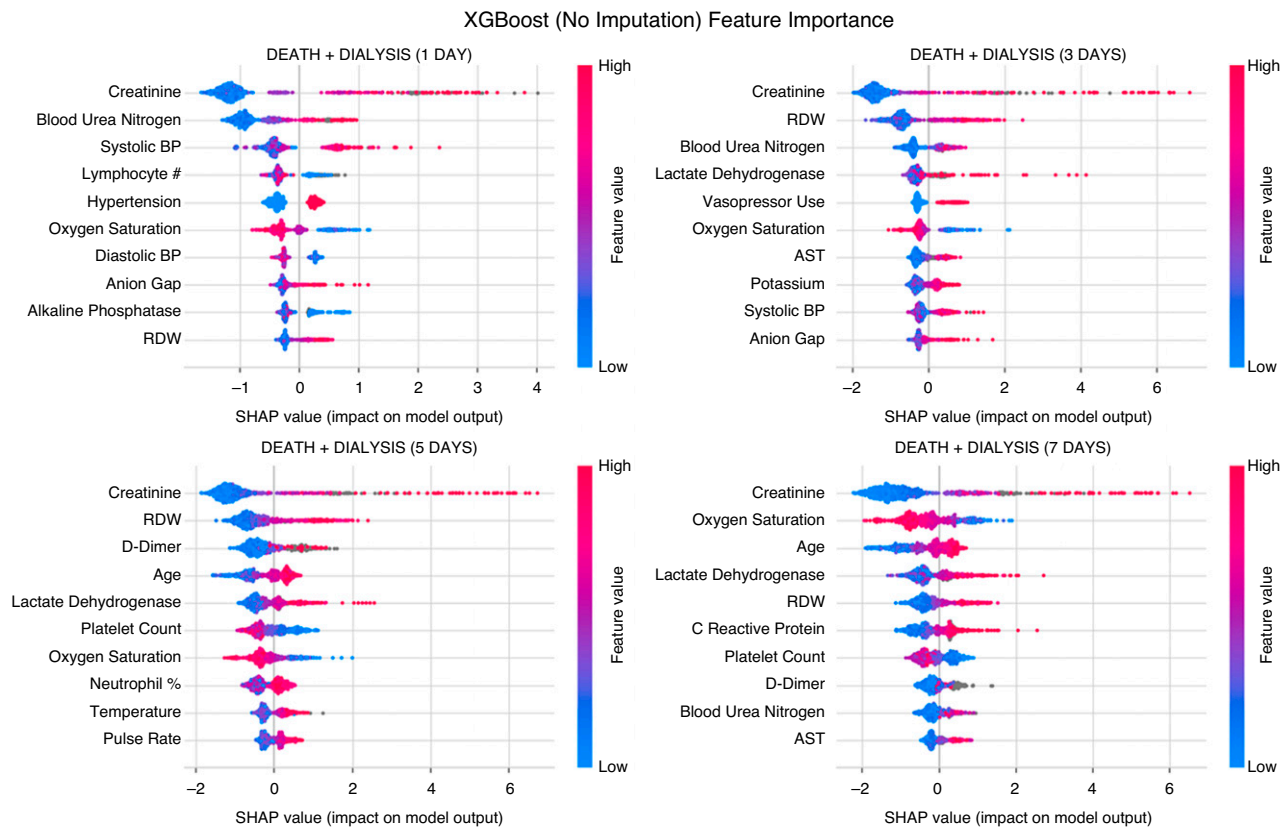


Figure 4. | Top features contributing to model prediction were similar across time frames. SHapley Additive exPlanations (SHAP) summary plots showing features driving model prediction toward death or dialysis at 1, 3, 5, and 7 days postadmission. Each dot represents a single patient. In each of the four plots, features are listed on the left in decreasing order of importance. Higher values of a feature are shown in red, and lower values are shown in blue. Positive SHAP values (plotted to the right of the zero on the x axis) push the model toward a prediction of death or dialysis, whereas negative SHAP values push the model away from a prediction of death or dialysis. For example, creatinine levels are the most important feature driving prediction. Because creatinine levels are in red on the right, higher values of creatinine drive model prediction toward the outcome. Similarly, higher oxygen saturation is plotted in red on the left and is associated with the model's prediction being driven away from death or dialysis. AST, aspartate aminotransferase; RDW, red cell distribution width.

Our framework allows for a clinically pertinent understanding of the XGBoost model's structure and the factors driving the predictions. Elevated serum creatinine concentrations and BUN were found to drive model predictions most strongly toward the composite outcome of death and dialysis—an expected finding (24, 25). Although the performance of the nonimputed XGBoost model decreased on external validation, this may be related to differences in the patient cohort demographics and disease severity, with the MSH hospital generally caring for patients who are multimorbid and in a more acute state than those in the other hospitals. Test parameters of sensitivity, specificity, positive predictive value, and negative predictive value will depend on the threshold used to classify patients as predicted to have the outcome. We chose to use a threshold derived from the Youden J statistic; however, depending on the circumstances, users can choose a threshold to optimize a different test parameter (26).

Given the near universal use of EHRs, implementation of machine learned models can be integrated into the EHR of different health care systems. Prior studies demonstrated that incorporation of systems for AKI prediction and alerting result in decreased length of stay and mortality (27, 28).

Although we have taken a vigorous approach of internal validation and external validation, we recognize that our approach will need additional validation in other health care systems to demonstrate generalizability.

There are limitations to this study. Our models make predictions on the basis of admission data alone, and events after admission may drive the course of a patient away from these predictions. In our view, a live, continuously updating, modeling approach is more appropriate for such cases, and this will be the focus of future work. However, using values from the time of admission has benefits, particularly in providing risk estimates at the earliest time point within a given patient's hospital course. The inclusion of patients from all hospitals into one dataset used for training and testing may boost performance; however, differences in patient care and distribution of missing data across hospitals may bias the models. This was the primary reason for training on one hospital and externally validating on other hospitals. We were encouraged by the similarities in performance across the internal and external validation experiments. Due to limited dialysis availability during the pandemic, patients who may have normally received dialysis may not have received it due to staffing

or equipment shortages. Patients who were discharged or lost to follow-up (e.g., transferred to another hospital) were censored. Although our prediction time window was short, this censoring may have biased our model. Finally, although variables that are more predictive of dialysis treatment and death may exist, not all values are drawn at admission for all patients. Accordingly, the presence of a threshold for missing data may have eliminated these variables from consideration for our models. We believe that future considerations for at-risk patients will merit the inclusion of a more comprehensive set of features (including biomarkers or imaging data) that may be used to further improve performance.

In conclusion, identification of patients at risk for acute dialysis and death in COVID-19 presents a variety of challenges. One such difficulty pertains to resource allocation in a potentially overcrowded hospital. Our models may assist with this challenge and are currently being prospectively validated and deployed in a real-world setting to aid in the management of hospitalized patients with COVID-19.

Disclosures

E.P. Böttinger reports receiving honoraria from Bayer, Bosch Health Campus, Sanofi, and Siemens; having consultancy agreements with Deloitte and Roland Berger; ownership interest in Digital Medicine E. Böttinger GmbH, EBCW GmbH, and Ontomics, Inc.; and serving as a scientific advisor for, or member of, Bosch Health Campus and Seer Biosciences Inc. L. Chan reports receiving honoraria from Fresenius Medical Care, being employed by Icahn School of Medicine at Mount Sinai, receiving research funding from the National Institutes of Health (NIH), and receiving financial compensation as a consultant for Vifor Pharma Inc. K. Chaudhary reports serving as a statistical advisor at *BMC Cancer* and as an associate editor of *BMC Medical Genomics*, and being employed by Icahn School of Medicine at Mount Sinai. S.G. Coca reports having consultancy agreements with Akebia, Bayer, Boehringer Ingelheim, CHF Solutions, Relypsa, RenalytixAI, Quark, and Takeda Pharmaceuticals; being supported by NIH grants U01DK106962, R01DK115562, R01HL085757, U01OH011326, R01DK112258, and KRTI UG 2019; serving on the editorial boards of *CJASN*, *JASN*, and *Kidney International* and as an associate editor of *Kidney360*; receiving consulting fees from Goldfinch Bio and inRegen; being employed by Icahn School of Medicine at Mount Sinai (Mount Sinai owns part of RenalytixAI); receiving research funding from inRegen and RenalytixAI; having ownership interest in pulseData and RenalytixAI; having patents and inventions with RenalytixAI; and serving as a scientific advisor or member of RenalytixAI. Z.A. Fayad reports receiving honoraria from Alexion and GlaxoSmithKline; receiving research funding from Amgen, Bristol Myers Squibb, Daiichi Sankyo, NIH, and Siemens Healthineers; being employed by Mount Sinai Medical Center; having ownership interest in, consultancy agreements with, and serving as a scientific advisor for, or member of, Trained Therapeutix Discovery; and having patents and inventions with Trained Therapeutix Discovery. B.S. Glicksberg, S.K. Jaladanki, A. Kia, M.A. Levin, A. Russak, and P. Timsina report being employed by Icahn School of Medicine at Mount Sinai. J.C. He reports serving on the editorial boards for *American Journal of Physiology*, *Diabetes*, *JASN*, and *Kidney International*; receiving honoraria (\$3400) from AstraZeneca; serving as a board member of the Chinese American Society of Nephrology and International Chinese Society of Nephrology;

being employed by Icahn School of Medicine at Mount Sinai; serving as an associate editor for *Kidney Disease* and section editor for *Nephron*; having consultancy agreements with, and owning equity in, Renalytix AI; and receiving research funding from Shanghai Innovation. G.N. Nadkarni reports receiving consulting fees from AstraZeneca, BioVie, GLG Consulting, and Reata; receiving research funding from Goldfinch Bio; being supported by National Institutes of Health (NIH) grants R01DK108803, U01HG007278, U01HG009610, and U01DK116100; and having ownership interest in, being employed by, having consultancy agreements with, and serving as a scientific advisor for, or member of, Pensieve Health and RenalytixAI. All remaining authors have nothing to disclose.

Funding

G.N. Nadkarni is supported by the NIH career development award grant K23DK107908 and is also supported by NIH grant R56DK126930. L. Chan is supported by the National Institute of Diabetes and Digestive and Kidney Diseases career development grant K23DK124645.

Acknowledgments

To all of the nurses, physicians, and providers who contributed to the care of these patients. To the patients and their family members who were affected by this pandemic.

Data Sharing Statement

Source code is available at <https://github.com/HPIMS/COVID-Dialysis-Prediction> and has been released under a GNU General Public License version 3. Data request can be made by contacting the corresponding author.

Supplemental Material

This article contains the following supplemental material online at <http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.17311120/-/DCSupplemental>.

Supplemental Figure 1. Cumulative incidence plots of death, dialysis, and death and dialysis.

Supplemental Figure 2. Calibration for the nonimputed XGBoost model over different time horizons. The difference for the calibration line over the model (under both isotonic and sigmoid activation) is not significantly different from the line of perfect calibration.

Supplemental Figure 3. Sensitivity, specificity, and positive predictive value of XGBoost without imputation model at various prediction thresholds.

Supplemental Figure 4. Ten features with highest SHAP scores in the XGBoost without imputation model at hospital day 1, 3, 5, and 7.

Supplemental Table 1. Number of death and dialysis in the MSH and OH cohort at prediction time windows.

Supplemental Table 2. Model performance in patients admitted to the intensive care unit in the external validation set.

Supplemental Table 3. Performance metrics of different models in internal and external validation.

References

1. Johns Hopkins Coronavirus Resource Center: COVID-19 map. Available at: <https://coronavirus.jhu.edu/map.html>. Accessed June 12, 2020
2. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC,

- Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS; China Medical Treatment Expert Group for Covid-19: Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 382: 1708–1720, 2020
3. Chan L, Chaudhary K, Saha A, Chauhan K, Vaid A, Zhao S, Paranjpe I, Somani S, Richter F, Miotto R, Lala A, Kia A, Tim-sina P, Li L, Freeman R, Chen R, Narula J, Just AC, Horowitz C, Fayad Z, Cordon-Cardo C, Schadt E, Levin MA, Reich DL, Fuster V, Murphy B, He JC, Charney AW, Böttinger EP, Glicksberg BS, Coca SG, Nadkarni GN; Mount Sinai COVID Informatics Center (MSCIC): AKI in hospitalized patients with COVID-19. *J Am Soc Nephrol* 32: 151–160, 2021
 4. Russo E, Esposito P, Taramasso L, Magnasco L, Saio M, Briano F, Russo C, Dettori S, Vena A, Di Biagio A, Garibotti G, Bassetti M, Viazzì F; GECOVID Working Group: Kidney disease and all-cause mortality in patients with COVID-19 hospitalized in Genoa, Northern Italy. *J Nephrol* 34: 173–183, 2021
 5. Reddy YNV, Walensky RP, Mendu ML, Green N, Reddy KP: Estimating shortages in capacity to deliver continuous kidney replacement therapy during the COVID-19 pandemic in the United States. *Am J Kidney Dis* 76: 696–709.e1, 2020
 6. Division of Nephrology, Columbia University Vagelos College of Physicians: Disaster response to the COVID-19 pandemic for patients with kidney disease in New York City. *J Am Soc Nephrol* 31: 1371–1379, 2020
 7. Sourial MY, Sourial MH, Dalsan R, Graham J, Ross M, Chen W, Golestaneh L: Urgent peritoneal dialysis in patients with COVID-19 and acute kidney injury: A single-center experience in a time of crisis in the United States. *Am J Kidney Dis* 76: 401–406, 2020
 8. Lalmuanawma S, Hussain J, Chhakchhuak L: Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals* 139: 110059, 2020
 9. Flechet M, Falini S, Bonetti C, Güiza F, Schetz M, Van den Berghe G, Meyfroidt G: Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: A prospective evaluation of the AKI predictor. *Crit Care* 23: 282, 2019
 10. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, LedSAM JR, Mohamed S: A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572: 116–119, 2019
 11. Kate RJ, Pearce N, Mazumdar D, Nilakantan V: A continual prediction model for inpatient acute kidney injury. *Comput Biol Med* 116: 103580, 2020
 12. Chen TQ, Guestrin C: XGBoost: A scalable tree boosting system. Presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13–17, 2016
 13. Lundberg SM, Lee SI: A unified approach to interpreting model predictions. Presented at the 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, December 4–9, 2017
 14. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Ver-gouwe Y, Habbema JD: Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 54: 774–781, 2001
 15. Davis J, Goadrich M: The relationship between precision-recall and ROC curves. Presented at the 23rd International Conference on Machine Learning, Pittsburgh, PA, June 25–29, 2006
 16. Oliphant T: *scipy.stats.t*. Available at: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html>. Accessed August 25, 2020
 17. Rufibach K: Use of Brier score to assess binary predictions. *J Clin Epidemiol* 63: 938–939, author reply 939, 2010
 18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E: Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12: 2825–2930, 2011
 19. Pilgrim M: *Dive Into Python 3*, New York, Apress, 2009
 20. The Pandas Development Team: *pandas-dev/pandas*: Pandas, Geneva, Switzerland, Zenodo, 2020
 21. Hunter JD: Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9: 90–95, 2007
 22. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P; SciPy 1.0 Contributors: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat Methods* 17: 261–272, 2020
 23. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, Winslow C, Shah N, Afshar M, Koyner JL: Internal and external validation of a machine learning risk score for acute kidney injury. *JAMA Netw Open* 3: e2012892, 2020
 24. Tartof SY, Qian L, Hong V, Wei R, Nadjafi RF, Fischer H, Li Z, Shaw SF, Caparosa SL, Nau CL, Saxena T, Rieg GK, Ackerson BK, Sharp AL, Skarbinski J, Naik TK, Murali SB: Obesity and mortality among patients diagnosed with COVID-19: Results from an integrated health care organization. *Ann Intern Med* 173: 773–781, 2020
 25. Klang E, Kassim G, Soffer S, Freeman R, Levin MA, Reich DL: Severe obesity as an independent risk factor for COVID-19 mortality in hospitalized patients younger than 50. *Obesity (Silver Spring)* 28: 1595–1599, 2020
 26. Youden WJ: Index for rating diagnostic tests. *Cancer* 3: 32–35, 1950
 27. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kel-lum JA: Clinical decision support for in-hospital AKI. *J Am Soc Nephrol* 29: 654–660, 2018
 28. Wilson FP, Greenberg JH: Acute kidney injury in real time: Prediction, alerts, and clinical decision support. *Nephron* 140: 116–119, 2018

Received: November 5, 2020 **Accepted:** April 28, 2021

A.V. and L.C. contributed equally to this work.

Published online ahead of print. Publication date available at www.cjasn.org.

*The MSCIC authors are Alex Charney, Allan C. Just, Benjamin Glicksberg, Girish Nadkarni, Laura Huckins, Paul O'Reilly, Riccardo Miotto, and Zahi Fayad (core investigators); Adam J. Ruskak, Adeeb Rahman, Akhil Vaid, Amanda Le Dobbyn, Andrew Leader, Arden Moscati, Arjun Kapoor, Christie Chang, Christopher Bellaire, Daniel Carrion, Fayzan Chaudhry, Felix Richter, Georgios Soultanidis, Ishan Paranjpe, Ismail Nabeel, Jessica De Freitas, Jiayi Xu, Johnathan Rush, Kipp Johnson, Krishna Vemuri, Kumardeep Chaudhary, Lauren Lepow, Liam Cotter, Lora Liharska, Marco Pereanez, Mesude Bicak, Nicholas DeFelice, Nidhi Naik, Noam Beckmann, Rajiv Nadu-kuru, Ross O'Hagan, Shan Zhao, Sulaiman Somani, Tielman T. Van Vleck, Tinaye Mutetwa, Tingyi Wanyan, Valentin Fauveau, Yang Yang, and Yonit Lavin (data scientists/analysts); Alona Lanksy, Ashish Atreja, Diane Del Valle, Dara Meyer, Eddy Golden, Farah Fasi-huddin, Hwei Hsun Wen, Jason Rogers, Jennifer Lilly Gutierrez, Laura Walker, Manbir Singh, Matteo Danieleto, Melissa A. Nieves, Micol Zweig, Renata Pyzik, Rima Fayad, Patricia Glowe, Sharlene Calorossi, Sparshdeep Kaur, Steven Ascolillo, and Yovanna Roa (operations/infrastructure); Anuradha Lala-Trindade, Steven G. Coca, Bethany Percha, Keith Sigel, Paz Polak, Robert Hirten, Talia Swartz, Ron Do, and Ruth J. F. Loos (other investigators); Dennis Charney, Eric Nestler, Barbara Murphy, David Reich, Erwin Böttinger,

Kumar Chatani, and Glenn Martin (leadership oversight board); and Eric Nestler, Patricia Kovatch, Joseph Finkelstein, Barbara Murphy, Joseph Buxbaum, Judy Cho, Andrew Kasarskis, Carol Horowitz, Carlos Cordon-Cardo, Monica Sohn, Glenn Martin, Adolfo Garcia-Sastre, Emilia Bagiella, Florian Krammer, Judith Aberg, Jagat Narula, Robert Wright, Erik Lium, Rosalind Wright, Annetine Gelijns, Valentin Fuster, and Miriam Merad (scientific advisory board).

AFFILIATIONS

¹The Mount Sinai Clinical Intelligence Center, Icahn School of Medicine at Mount Sinai, New York, New York

²The Hasso Plattner Institute of Digital Health, Icahn School of Medicine at Mount Sinai, New York, New York

³The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York

⁴The Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York

⁵Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, New York

⁶Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, New York

⁷Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York

⁸Digital Health Center, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

⁹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York

¹⁰The Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York

¹¹The Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York

¹²BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, BioMedical Engineering and Imaging Institute, Icahn School