Title
A Machine Learning Model for Predicting Mortality within 90 Days of Dialysis Initiation

Journal
Kidney360, 3(9)

ISSN
2641-7650

Authors
Rankin, Summer
Han, Lucy
Scherzer, Rebecca
et al.

Publication Date
2022-09-29

DOI
10.34067/kid.0007012021

Peer reviewed

Kidney360

A Machine Learning Model for Predicting Mortality within 90 Days of Dialysis Initiation

Summer Rankin,[1] *Lucy Han*,[1] *Rebecca Scherzer*,[2] *Susan Tenney*,[1] *Matthew Keating*,[1] *Kimberly Genberg*,[1] *Matthew Rahn*,[3] *Kenneth Wilkins*,[4] *Michael Shlipak*,[2] *and Michelle Estrella*[2]

Key Points

- This paper presents an eXtreme Gradient Boosting (XGBoost) model that predicted mortality in the first 90 days after dialysis initiation using data from the United States Renal Data System.
- Such a model could facilitate patient-clinician shared decision making on whether to initiate dialysis or pursue medical management.
- The XGBoost models discriminated mortality risk in both the nonimputed (c=0.826) and imputed (c=0.827) models.

Abstract

Background The first 90 days after dialysis initiation are associated with high morbidity and mortality in end-stage kidney disease (ESKD) patients. A machine learning–based tool for predicting mortality could inform patient-clinician shared decision making on whether to initiate dialysis or pursue medical management. We used the eXtreme Gradient Boosting (XGBoost) algorithm to predict mortality in the first 90 days after dialysis initiation in a nationally representative population from the United States Renal Data System.

Methods A cohort of adults initiating dialysis between 2008–2017 were studied for outcome of death within 90 days of dialysis initiation. The study dataset included 188 candidate predictors prognostic of early mortality that were known on or before the first day of dialysis and was partitioned into training (70%) and testing (30%) subsets. XGBoost modeling used a complete-case set and a dataset obtained from multiple imputation. Model performance was evaluated by c-statistics overall and stratified by subgroups of age, sex, race, and dialysis modality.

Results The analysis included 1,150,195 patients with ESKD, of whom 86,083 (8%) died in the first 90 days after dialysis initiation. The XGBoost models discriminated mortality risk in the nonimputed (c=0.826, 95% CI, 0.823 to 0.828) and imputed (c=0.827, 95% CI, 0.823 to 0.827) models and performed well across nearly every subgroup (race, age, sex, and dialysis modality) evaluated (c>0.75). Across predicted risk thresholds of 10%–50%, higher risk thresholds showed declining sensitivity (0.69–0.04) with improving specificity (0.79–0.99); similarly, positive likelihood ratio was highest at the 40% threshold, whereas the negative likelihood ratio was lowest at the 10% threshold. After calibration using isotonic regression, the model accurately estimated the probability of mortality across all ranges of predicted risk.

Conclusions The XGBoost-based model developed in this study discriminated risk of early mortality after dialysis initiation with excellent calibration and performed well across key subgroups.

KIDNEY360 3: 1556–1565, 2022. doi: https://doi.org/10.34067/KID.0007012021

Introduction

ESKD is associated with exceedingly high morbidity and mortality, especially within the first 90 days of dialysis initiation (1–3). During this vulnerable period of transition into dialysis, patients may experience adverse health events, including vascular access placement, fluid fluctuations that lead to either volume overload or hypotension, electrolyte derangements associated with increased risks of arrhythmia, and loss of residual kidney function. Such events present risk of further complications, particularly for the increasing number of patients who are initiating dialysis at an advanced age and have significant comorbidities such as diabetes, hypertension, and heart failure (4).

In light of these risks, there is a growing call to consider conservative medical management for ESKD during clinical decision making in multimorbid

[1]Booz Allen Hamilton, McLean, Virginia
[2]Kidney Health Research Collaborative (KHRC), University of California San Francisco (UCSF), San Francisco, California
[3]Office of the National Coordinator for Health Information Technology (ONC), Washington, DC
[4]National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), Bethesda, Maryland

Correspondence: Dr. Susan Tenney, Booz Allen Hamilton, 8283 Greensboro Drive, McLean, VA 22102. Email: Tenney_Susan@bah.com

patients (5–7). However, qualitative studies have shown that a patient's decision surrounding dialysis initiation relies on their intuition and the potential effect of treatment on their quality of life and survival (8,9). Conversely, clinicians tend to make their decisions largely on the basis of patient-related clinical factors (age, comorbidities, *etc.*) and default to chronic dialysis as the only option for management of ESKD (5). A predictive tool to estimate patient risk of early mortality after the initiation of dialysis could inform patient-clinician shared decision making on whether to initiate dialysis or to pursue medical management.

Although prediction models have been developed to estimate the probability of mortality after dialysis initiation, most have largely used conventional regression methods (10–19). Despite the ability of contemporary methods such as machine learning (ML) to integrate a rich array of clinically available data with the potential for broad generalizability, to our knowledge, few prior studies have leveraged ML to predict early mortality after dialysis initiation (20–22). This study sought to build an eXtreme Gradient Boosting (23) (XGBoost)-based model using data from the United States Renal Data System (USRDS) to: (*1*) optimize mortality prediction within the first 90 days of dialysis initiation in a nationally representative population, and (*2*) calibrate the model such that predicted mortality likelihoods are reasonably unbiased across the risk spectrum as defined below in the section on ML.

Materials and Methods

Study Design

All adults aged ≥18 years who initiated chronic hemodialysis or peritoneal dialysis between January 1, 2008, and December 31, 2017, were retrospectively identified from the USRDS national data registry maintained by the National Institute of Diabetes and Digestive and Kidney Diseases and containing data from Centers for Medicare & Medicaid Services, the United Network for Organ Sharing, and the ESKD networks. This study was approved by the University of California San Francisco Institutional Review Board and adhered to the Declaration of Helsinki. The selection criteria utilized for the USRDS tables (Supplemental Table 1)—PATIENTS, MEDEVID, pre-ESKD Medicare Claims, Kidney Transplant—resulted in a study cohort of 1,150,195 patients. The overall study design is shown in Figure 1.

Outcome Measure

The primary study outcome was all-cause mortality within 90 days of dialysis initiation. The date of death was ascertained from the USRDS PATIENT table. Outcome data were available for all patients in the selected study cohort through the entire 90-day assessment period.

Predictors

The study dataset was prepared using variables from the USRDS data that had clinical relevance and prognostic value for mortality in the first 90 days after dialysis initiation. To produce a high-quality study dataset for training a model, the following criteria were applied (Supplemental Table 2): cleaning and correctly labeling candidate predictors, structuring and curating to ensure that missing values and outliers were handled appropriately, splitting using random sampling into training (70%) and testing (30%) datasets, and preparing a data dictionary. The predictors in the study were limited to information that was known on or before the first day of dialysis. The study dataset consisted of 188 predictors, with one record per patient. Each variable used for building the model was assessed to determine if it should be excluded as an operational factor (24) (*i.e.*, a nuisance variable not related to overall health but present in the data, such as the day of a physician's signature, *etc.*). Variables that were true operational factors were removed from the dataset.

Two types of predictors were included in the study dataset: (*1*) predictors taken directly from the USRDS tables (*e.g.*, age, race, hemoglobin) and (2) predictors derived from variables in the USRDS data (*e.g.*, time on kidney transplant waitlist derived by subtracting dialysis date from the kidney transplant list date). The full list of predictors, including derivation methods, are shown in the Data Dictionary (Supplemental Tables 3 and 4).

Data Preprocessing

Clinical and laboratory variables that had missing values for >40% of patients were not included in the full list of predictors. For clinical and laboratory variables from the MEDEVID table used in the study dataset, M. Estrella and M. Shlipak defined the upper and lower bounds such that any values outside these bounds were considered clinically impossible (Supplemental Table 5); these outliers were set as missing values in the study dataset. Each record was additionally supplemented with distinct indicators of whether each such value was deemed an outlier (Supplemental Table 4, rows 8–14). The proportion of data considered to be outliers ranged from 0.5% to 2% of values across the clinical variables. Two datasets were then prepared for modeling: a nonimputed and a multiply imputed dataset. Within the nonimputed dataset, missing data were handled natively using XGBoost, as described below. To maximize reproducibility of the model, both the nonimputed and imputed study data were partitioned randomly into ten stratified nonoverlapping subsets (later referred to as subset 0, subset 1, … , subset 9). These ten partitions were further split into training subsets (70% of the whole study dataset, n=804,890) and testing subsets (30% of whole study dataset, n=345,305) to allow sufficient data both to train and to robustly evaluate the XGBoost models.

Partitioning the study dataset into ten subsets allowed for more efficiency in handling the missing values. The clinical and laboratory variables in the dataset were multiply imputed for each subset and included as predictors for the dataset used in the XGBoost imputed model (25). Imputed variables included height, weight, body mass index, serum creatinine, serum albumin, hemoglobin, and GFR estimated by the CKD-EPI equation (eGFR) (26). The missing values in these clinical and laboratory variables were imputed using multiple imputations by chained equations (27) (MICE) to create five imputations to target 95% relative efficiency.

Statistical Analyses

Cross-tabulation was used to examine unadjusted differences in baseline characteristics, stratified by train/test

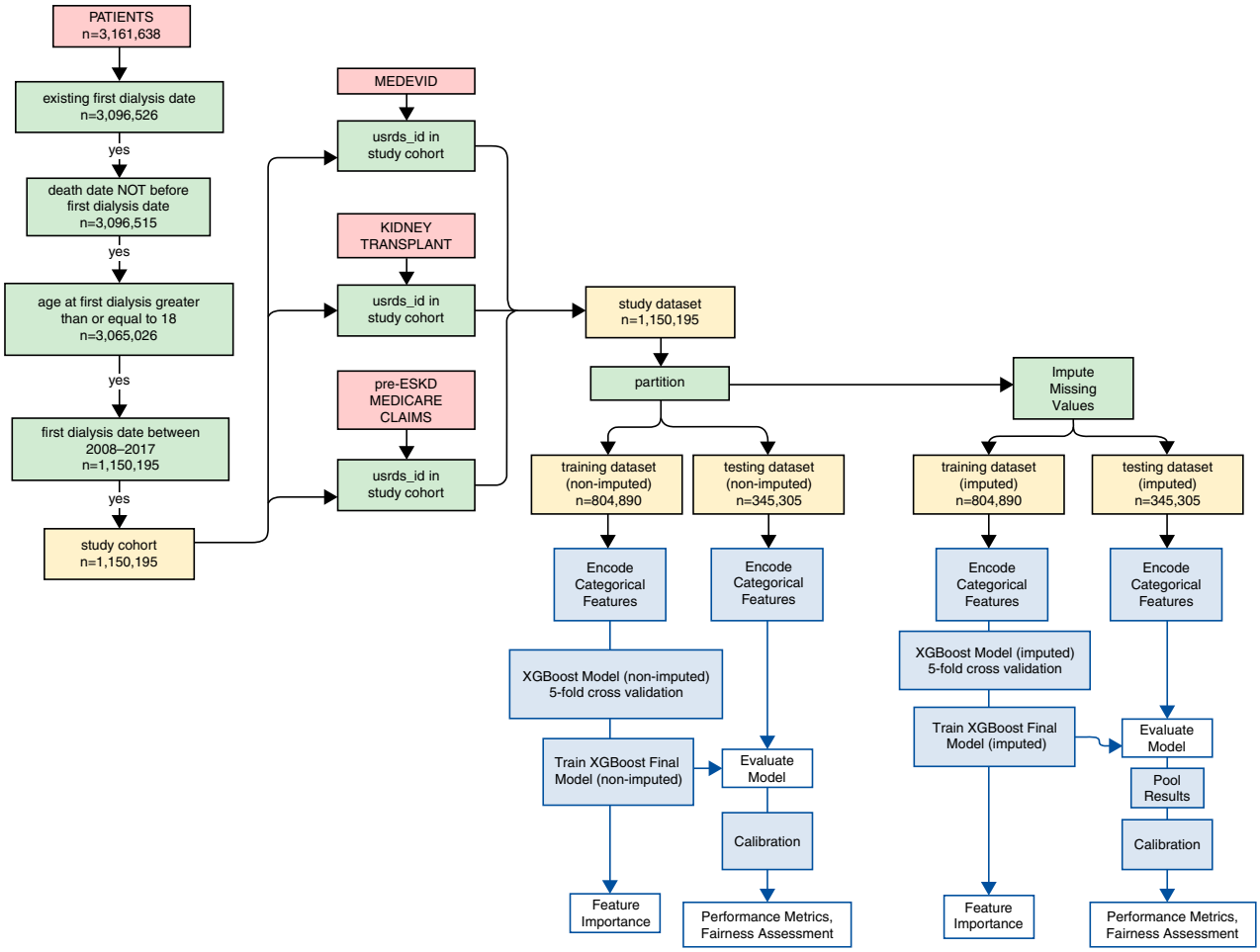

Figure 1. | **Study cohort criteria and analysis approach for predicting mortality within 90 days of dialysis in ESKD patients.** Pink, tables from United States Renal Data System (USRDS) database; green, cohort and dataset creation; yellow, constructed tables; blue, machine learning methods; white, evaluation. Usrds_id is the identification number for a single patient in the USRDS tables. XGBoost, eXtreme Gradient Boosting.

split. Categorical variables are summarized as frequencies and proportions, and continuous variables are summarized as medians and interquartile ranges or means±SDs, as appropriate.

Machine Learning

The XGBoost algorithm was selected to develop the prediction model for several reasons. First, XGBoost is a supervised learning model of gradient boosted decision trees that is widely used in classification tasks because it uses standard classification benchmarks, returns predictor ranking, and is scalable to large datasets due to its ability to parallel process. Second, it can be applied to a wide array of use cases, data types, and desired prediction outcomes. Third, it has shown superior performance relative to other ML models in previous studies of kidney disease (28,29). Fourth, it handles noninformatively missing values natively using a sparsity-aware split finding algorithm, which allows for the comparison of models with or without the use of imputed data.

Two XGBoost models were developed in this study: one on the nonimputed study dataset (missing values were handled natively by the XGBoost model) and the other on the imputed dataset. Before modeling, all categorical variables with more than two factors were one-hot encoded (*e.g.*, turning categorical variable factors into separate binary variables) in both datasets (see example in Figure 2) (30). The training data were used to tune model settings (*i.e.*, hyperparameters), which were optimized on the area under the receiver operating characteristic curve (AUC ROC or c-statistic) using Bayesian optimization and five-fold cross-validation. The range of hyperparameters that were tuned are shown in Supplemental Table 6. The final model was trained on the 70% training subset using the best hyperparameters from the five-fold cross-validation. For the imputed model, an XGBoost model was run for each imputed dataset. The resulting estimates (between 0 and 1) were combined by averaging the model prediction scores per patient across the five imputations. Calibration was performed using a nonparametric isotonic regressor (31) trained on 66% of the testing dataset (subsets 7 and 8, n=230,482) and evaluated on the remaining 33% of the testing dataset (subset 9, n=114,823). The final model was evaluated on the testing dataset using multiple metrics: (*1*) *c*-statistics; (2) the most influential predictors using gain to

Patient ID	Maturing AVF
1	yes
2	no
3	unknown
4	missing

Patient ID	Has Maturing AVF	Does Not Have Maturing AVF	Unknown Maturing AVF	Missing Maturing AVF
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1

Figure 2. | An example of a categorical variable before and after one-hot encoding. An example categorical variable (maturing arteriovenous fistula [AVF]) has four categories for four fictional patients (left table). The table on the right shows the resulting four variables after one-hot encoding.

reveal the underlying inputs that influence mortality risk; (*3*) assessment of model calibration by plotting the observed versus estimated risk by decile of predicted risk; (*4*) sensitivity and specificity at the predicted mortality risk cut points of 10%, 20%, 30%, 40%, and 50%, given the overall population risk of 8%, as candidate thresholds to denote high risk; and (*5*) the ability to discriminate risk across subgroups on the basis of age, sex, race, and dialysis modality.

Results

Cohort Characteristics

The final study cohort included 1,150,195 patients with ESKD, of whom 86,083 (8%) died in the first 90 days after dialysis initiation. Overall, the mean age at initiation of dialysis was 63 years, 27% were Black, 57% were men, and 98% had at least one comorbidity. Baseline demographic and clinical characteristics stratified by train/test split are presented in Table 1. The training and the test cohorts had comparable characteristics, suggesting that the train/test split was valid.

XGBoost Model Results

Discrimination of the XGBoost models was high and similar regardless of whether the missing data were handed natively (c=0.826, 95% CI, 0.823 to 0.828) or multiply imputed (c=0.827, 95% CI, 0.823 to 0.827), as shown in Figure 3. The top 20 predictors from the XGBoost nonimputed model on the basis of gain are shown in Table 2. The top 5 predictors by contribution to the model were age, total hospital days, time between first and last hospitalization, missing information on exogenous erythropoietin (EPO), and presence of a maturing arteriovenous fistula. Substantial overlap in selected predictors and their prediction rankings was also observed in the XGBoost model fit on the multiply imputed data (Supplemental Table 7). As a sensitivity analysis, we ran more limited models using only the ten most influential predictors on the basis of the feature importance analysis. These more limited models yielded a c-statistic of 0.78 (95% CI, 0.782 to 0.787) for the nonimputed model and 0.769 (95% CI, 0.765–0.77) for the imputed. Calibration of the model predictions using isotonic regression (31) showed close agreement between observed and expected event rates across the full range of predicted risk for the model fit on the nonimputed dataset (Figure 4) and on the imputed dataset (Supplemental Figure 1). Supplemental Table 8 and Table 3 and show the performance across predicted risk thresholds of 10% through 50% of the nonimputed and imputed model, respectively, assessed that best illustrates the trade-offs between the following metrics: sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio. With increasing risk thresholds, sensitivity progressively decreased, whereas specificity remained high and showed slight improvement. The positive likelihood ratio was highest at the 40% threshold, whereas the negative likelihood ratio was lowest at the 10% threshold.

Discrimination was compared across each race, age, sex, and dialysis modality categories, as shown in Table 4. Discrimination was sufficient (c>0.75) across all subgroups that were considered.

Discussion

The first 90 days after dialysis are a high-risk period, and yet existing prediction tools lack the ability to identify patients at high risk for early mortality. To address this gap, we constructed a risk prediction model using XGBoost on the USRDS data. The XGBoost model developed in this study achieved sufficient discrimination (c>0.75) for predicting mortality within the first 90 days of dialysis. Furthermore, the model was well calibrated, with little difference between the predicted and observed event rates across the risk spectrum.

The ability of our model to distinguish risk for early mortality among incident dialysis patients is significantly improved compared with previously developed risk scores for near-term mortality (11–13). The native XGBoost model with the nonimputed data and the model with imputed data both achieved an overall *c*-statistic of 0.826 (95% CI, 0.824 to 0.828) and 0.827 (95% CI, 0.823 to 0.827), respectively. In contrast, a prior study by Thamer and colleagues using logistic regression focused on predicting 3- and 6-month mortality among incident dialysis patients aged ≥65 years derived from the USRDS registry achieved a c-statistic of 0.69–0.72 (10). Other studies that used traditional regression modeling aimed to predict 6-month or 1-year mortality achieved similar discrimination as reported by Thamer *et al.* To our knowledge, only one prior study has used an ML-based approach to predict mortality within the first 90 days of dialysis initiation. Using a random forest approach, Akbilgic *et al.* obtained an overall *c*-statistic of 0.75 for prediction of 90-day mortality. The model performed well across most subgroups and had slightly better performance compared with Cox regression models (20). Although Akbilgic *et al.* utilized electronic health record data, which allows for a richer set of

Table 1. Demographic and clinical characteristics of the training and testing cohorts

Characteristics	Training Data (N=804,890)	Testing Data (N=345,305)
Demographic characteristics		
Age, yr	63±15	63±15
Race		
White	537,460 (67)	230,577 (67)
Black	218,237 (27)	93,560 (27)
American Indian or Alaska Native	7483 (0.9)	3225 (0.9)
Asian	30,030 (4)	12,965 (4)
Native Hawaiian or Pacific Islander	8810 (1)	3776 (1)
Other or multiracial	2088 (0.3)	881 (0.3)
Unknown	782 (0.1)	321 (0.1)
Sex[a]		
Men	463,183 (58)	198,347 (57)
Women	341,702 (42)	146,957 (43)
Comorbid characteristics		
Diabetes	452,424 (56)	193,697 (56)
Hypertension	688,465 (86)	295,806 (86)
Cardiovascular disease	283,715 (35)	121,685 (35)
Heart failure	240,728 (30)	102,863 (30)
Peripheral arterial disease	93,329 (12)	40,258 (12)
Underlying cause of ESKD		
Diabetes	372,162 (46)	159,048 (46)
Hypertension	234,353 (29)	100,873 (29)
Glomerulonephritis	59,758 (7)	25,856 (7)
Other	138,617 (17)	59,528 (17)
Laboratory characteristics		
Height, cm	168±12	168±12
Height missing	16,286 (2)	6935 (2)
Weight, kg	84±25	84±25
Weight missing	14,340 (2)	6120 (2)
BMI, kg/m^2	30±8	30±8
BMI missing	19,939 (2)	8500 (2)
Serum albumin, g/dl	3.2±0.7	3.2±0.7
Serum albumin missing	246,862 (30)	105,235 (30)
Hemoglobin, g/dl	9.6±1.64	9.6±1.6
Hemoglobin missing	122,654 (15)	52,018 (15)
Serum creatinine, mg/dl	6.4±3.5	6.4±3.52
Serum creatinine missing	14,762 (2)	6321 (2)
eGFR	10±5	10±5
eGFR missing	23,078 (3)	9910 (3)
Prior nephrology care characteristics		
Has maturing arteriovenous fistula	121,294 (15)	52,067 (15%)
Has maturing arteriovenous graft	16,645 (2)	6,932 (2%)
Received exogenous erythropoietin	140,046 (17)	59,882 (17%)
Under care of kidney dietician	60,946 (8)	26,370 (8%)
Had prior nephrology care	478,881 (60)	205,556 (60%)
Medicare pre-ESKD claims characteristics		
IP claims[b]	3 (2, 6)	3 (2, 6)
IP claims	417,523 (52)	178,968 (52)
OP claims[b]	16 (6, 39)	17 (6, 39)
OP claims	444,874 (55)	190,395 (55)
HS claims[b]	2 (1, 4)	2 (1, 4)
Missing HS claims	796,123 (99)	341,590 (99)
HH claims[b]	2 (1, 5)	2 (1, 5)
Missing HH claims	647,880 (80)	278,043 (81)
SN claims[b]	3 (2, 5)	3 (2, 5)
Missing SN claims	706,475 (88)	303,303 (88)

Results displayed either as mean±SD or *n* (%) unless otherwise indicated. eGFR calculated using the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) equation. IP, inpatient; OP, outpatient; HS, hospice; HH, home health; SN, skilled nursing.
[a]Missing values for sex are not reported as the aggregate count is under 11.
[b]Results displayed as median (Q1, Q3) per patient.

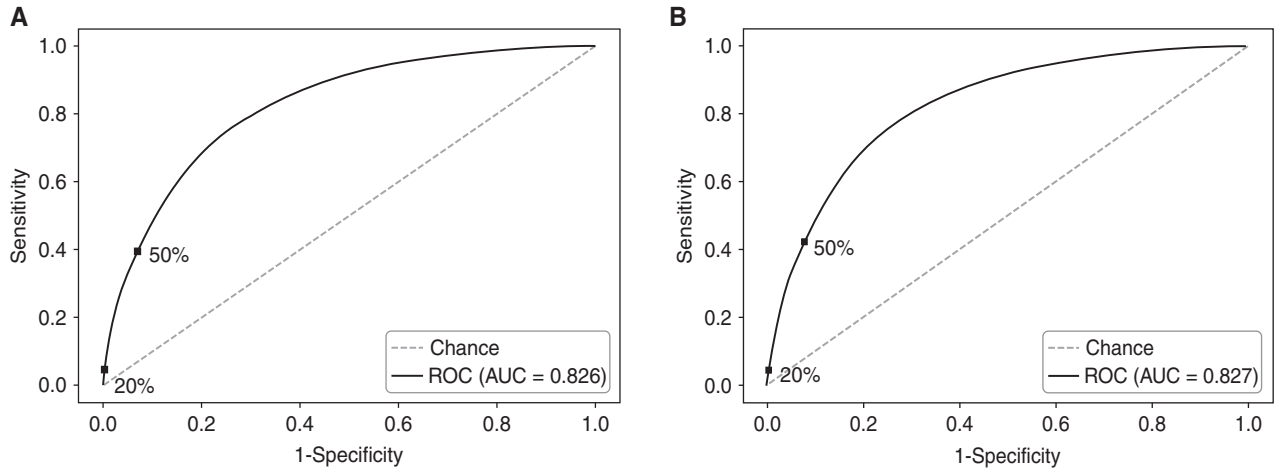

Figure 3. | **Area under the receiver operating characteristic curve (AUC ROC) plots for XGBoost models.** The AUC is 0.826 (95% CI, 0.823 to 0.828) for nonimputed (A) and 0.827 (95% CI, 0.823 to 0.827) for imputed (B). The 20% and 50% thresholds are plotted on each curve with a point on the solid line; the dashed diagonal line is the performance for chance prediction.

predialysis predictors, the study population was limited to veterans, who are predominantly men and older. The present study relied on USRDS data, which enabled inclusion of a broader study population representative of the US dialysis population. Aside from inherent differences in how traditional regression methods and ML-based methods incorporate candidate predictors, differences in model performance between our study and prior studies may also be due to differences in the study populations, mortality incidence across different study periods and clinical settings, and the spectrum of candidate predictors.

In this study, the XGBoost model identified the predictors most influential in mortality risk in the early phase after initiating dialysis. The majority of these variables were related to the patient's health status, including several features that indicated a greater likelihood of frailty (32): older age, frequent hospitalizations, institutionalization or nursing home occupancy, inability to ambulate, and a classification of being unfit for kidney transplantation. Laboratory indicators of health status included serum albumin, creatinine concentrations, and eGFR. Other predictors selected by the model are indicators of the length of time before ESKD and the quality of care delivered: arteriovenous fistula status, unknown receipt of erythropoiesis-stimulating agents, unknown cause of ESKD, and nutritionist care. Although it is reassuring that most of these predictors have face validity as determinants of early mortality risk, causality is not a requirement for inclusion into an ML prediction algorithm. More important are the availability of the predictors to clinicians and other researchers, the model's generalizability across groups of patients, and its ability to distinguish wide ranges of risk. Almost 200 variables from USRDS were included in the initial model; however, many of these variables may not be available to clinicians. As a sensitivity analysis, we restricted the model to the ten most influential features, which yielded a lower c-statistic compared with the full model (c=0.78 versus c=0.83). The slight decrease in performance when using only the ten most influential predictors illustrates the importance of these features for the prediction, even when many other features are available.

Early mortality prediction is challenging among patients newly diagnosed with ESKD because the overall mortality risk is relatively low (only 8% in the USRDS cohort); risk prediction is easiest for situations with a balance of cases and noncases. To account for the class imbalance, the positive class (died in the first 90 days) was weighted more heavily in the models, which applies a stronger penalty to the model when the minority class is incorrectly classified and a weaker penalty when the majority class is incorrectly classified. As shown by the XGBoost model results, there was no obvious threshold to balance the trade-offs of sensitivity and specificity for predicting mortality, although our model was well calibrated across the broad range of risks (as shown in Figure 4). At a predicted risk threshold of 10%, sensitivity was 69% and specificity was approximately 79%; in contrast, at a predicted risk threshold of 50%, specificity exceeded 99% but sensitivity was only 4%. This reflects the challenge of risk prediction in the ambulatory setting in clinical medicine; models are often excellent at placing patients into appropriate risk groups but are much weaker at identifying specific individual patient who will experience an adverse event, such as death, in the first 90 days of dialysis.

A strength of this study is that it uses data from the USRDS, which represents the largest and most representative population of ESKD patients. The USRDS offers nearly complete inclusion of ESKD patients within the United States and enables linkage to Medicare claims. This large sample size provides robust assessment of risk and will ensure reproducibility and generalizability of the results generated in this study. Limitations of USRDS include lack of specific prognostic data and high rates of missing data for predialysis features and other predictors of interest, including laboratory data (*e.g.*, urine biomarkers, phosphorous, and calcium), comorbidities, and cause of death (33). An additional strength of this study is the XGBoost algorithm—a flexible, interpretable ML method, which can natively handle noninformatively missing data while offering high predictive accuracy. Using XGBoost, we were able

Table 2. Top 20 predictors of mortality within 90 days of dialysis initiation and their ranking of importance for the nonimputed XGBoost model as measured through gain (the relative contribution of the predictor to the model)

Rank	Feature	Gain	Died in 90 Days, *N*=86,083	Survived in 90 Days, *N*=1,064,112
1	Age, yr	0.1454	71±12	62±14
2	Total inpatient hospital days	0.0743	40±49	29±44
3	Duration of time between first and most recent hospitalizations	0.0502	562±533	487±512
4	Missing information on EPO receipt (as compared with having information)	0.0371	20,744 (24)	270,825 (25)
5	Has maturing AVF	0.0356	8386 (10)	164,975 (15)
6	Serum albumin	0.0352	2.8±0.6	3.1±0.7
7	Institutionalized	0.0271	18,895 (21)	77,104 (7)
8	Serum creatinine	0.0251	5.2±2.8	6.4±3.5
9	Patient documented to be medically unfit for transplantation	0.0241	14,194 (16)	55,713 (5)
10	Underlying cause of ESKD categorized as other	0.0219	15,717 (18)	98,182 (9)
11	Number of days between first and last claim	0.0213	916±594	858±588
12	Missing information on whether a patient was under the care of kidney dietician (as compared with having information)	0.0198	4034 (4)	70,956 (6)
13	GFR-EPI	0.0192	11±5	9±4
14	Cause of ESKD	0.0190	86,083	1,064,112
15	Nursing home occupant	0.0174	17,124 (20)	65,410 (6)
16	Does not have maturing AVF	0.0156	67,416 (78)	629,017 (59)
17	Inability to ambulate	0.0142	15,759 (18)	64,544 (6)
18	Patient documented to be unsuitable for kidney transplant due to age	0.0124	7893 (9)	42,387 (4)
19	Duration of time between first and last outpatient claim	0.0122	891±592	848±583
20	Has maturing AVG	0.0115	1607 (2)	21,970 (2)

Results displayed either as mean±SD or *n* (%). EPO, exogenous erythropoietin; AVF, arteriovenous fistula; GFR-EPI, GFR calculated using the Chronic Kidney Disease Epidemiology Collaboration equation; AVG, arteriovenous graft; XGBoost, eXtreme Gradient Boosting.

to create a model with high specificity, discrimination, and calibration while identifying risk factors of clinical significance. Limitations of XGBoost are that it is computationally intensive when using a large dataset (more than one million rows) and that multiple hyperparameters must be tuned in order to achieve good model fit. Further, in contrast to traditional regression methods, XGBoost does not provide interpretable regression coefficients and confidence intervals, especially as there are many parameters that the model learns from the training data.

In summary, the XGBoost-based model developed in this study was able to predict risk of early mortality after dialysis initiation with high accuracy and with strong discrimination across key subgroups. Such an ML-based approach

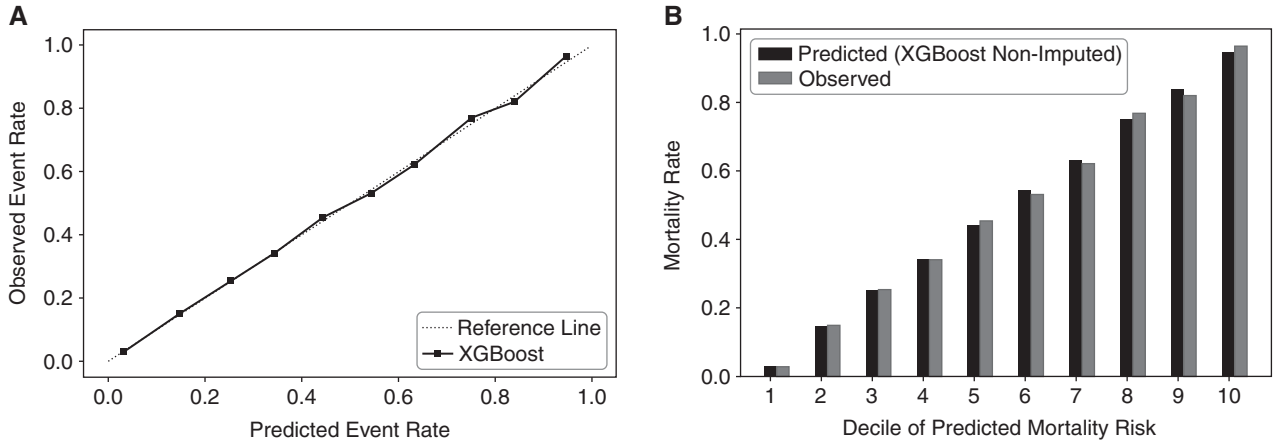

Figure 4. | Calibration plot for XGBoost nonimputed model predicted risks. (A) Predicted event rate on the *x*-axis and observed event rate on the *y*-axis. (B) Predicted and observed event rates by decile of predicted risk.

Table 3. Predicted risk of mortality with 90 days of dialysis initiation at 10%, 20%, 30%, 40%, and 50% thresholds for the XGBoost model on the nonimputed study dataset

Model Threshold	Sensitivity	Specificity	Likelihood Ratio (+)	Likelihood Ratio (–)	True Positive	False Positive	True Negative	False Negative
0.1	0.69	0.79	3.39	0.38	5947	21,712	84,546	2618
0.2	0.39	0.93	5.82	0.64	3394	7229	99,029	5171
0.3	0.19	0.97	9.22	0.81	1709	2299	103,959	6856
0.4	0.12	0.99	12.85	0.88	1036	1000	105,258	7529
0.5	0.04	0.99	12.04	0.95	397	234	106,024	8168

True positives, number of patients the model correctly predicted died in 90 days; false positives, number of patients the model incorrectly predicted died in 90 days; true negatives, number of patients the model correctly predicted survived in 90 days; false negatives, number of patients the model incorrectly predicted survived the first 90 days; sensitivity, true positives/(true positives+false negatives); specificity, true negatives/(true negatives+false positives); likelihood ratio (positive class), sensitivity/(1–specificity); likelihood ratio (negative class). (1–sensitivity)/specificity. XGBoost, eXtreme Gradient Boosting.

could facilitate shared decision making among patients and clinicians facing the complex decision of dialysis initiation versus conservative medical management of ESKD. To optimize the potential utility of ML-based algorithms in this clinical context, future efforts should consider assessing a broader set of options for ESKD management using additional pre-ESKD data sources that complement current USRDS data, including temporary trial of dialysis and palliative dialysis, and capturing additional patient-centered predictors.

Table 4. Comparison of discrimination by subgroup for the XGBoost nonimputed model

Category	Area Under the Curve
Race	
White (*N*=76,751)	0.819
Black (*N*=31,088)	0.826
American Indian (*N*=1042)	0.849
Asian (*N*=4308)	0.847
Native Hawaiian or Pacific Islander (*N*=1241)	0.840
Other or multiracial (*N*=295)	0.822
Unknown (*N*=98)	0.822
Age group, yr	
18–25 (*N*=1490)	0.799
26–35 (*N*=4269)	0.823
36–45 (*N*=8693)	0.838
46–55 (*N*=17,602)	0.818
56–65 (*N*=28,372)	0.795
66–75 (*N*=28,723)	0.789
76–85 (*N*=20,635)	0.770
86+ (*N*=5039)	0.753
Sex	
Men (*N*=66,033)	0.831
Women (*N*=48,769)	0.819
Dialysis modality	
Hemodialysis (*N*=103,242)	0.818
Continuous cycling peritoneal dialysis (*N*=5016)	0.822
Continuous ambulatory peritoneal dialysis (*N*=4440)	0.858
Other (*N*=31)	0.933
NA (*N*=2094)	0.778

Disclosures

M. Estrella reports being an employee of the University of California, San Francisco, and San Francisco VA Health Care System; consultancy for Eiland & Bonnin (PC); research funding from Bayer, Inc., and Booz Allen Hamilton; honoraria from the American Kidney Fund, AstraZeneca, Boehringer Ingelheim, and the National Kidney Foundation; and other interests or relationships with *American Journal of Kidney Diseases*, *CJASN*, and the National Kidney Foundation. K. Genberg reports being an employee of Booz Allen Hamilton and IBM, and ownership interest in Booz Allen and IBM. L. Han reports being an employee of Booz Allen Hamilton. M. Keating reports being an employee of Booz Allen Hamilton; consultancy for Booz Allen Hamilton; and ownership interest in Booz Allen Hamilton and Kimbell Royalty Partners. M. Rahn reports being an employee of HHS/Office of the National Coordinator for Health IT, and an advisory or leadership role for the Office of the National Coordinator for Health IT. S. Rankin reports being an employee of Booz Allen Hamilton. R. Scherzer reports being an employee of UCSF, and an advisory or leadership role (editorial board) for *CJASN*, *JAIDS*, and *Kidney360*. M.G. Shlipak reports consultancy agreements with Cricket Health; Intercept Pharmaceuticals, University of Washington—Cardiovascular Health Study, and Veterans Medical; research funding from Bayer Pharmaceuticals; honoraria from AstraZeneca, Bayer, and Boeringer Ingelheim; being a scientific advisor for or membership of the *American Journal of Kidney Disease*, *Circulation*, and *JASN*; and being a board member of the Northern California Institute for Research and Education. S. Tenney reports being an employee of Booz Allen Hamilton. K.J. Wilkins reports an advisory or leadership role for the *International Journal of Obesity* (editorial board [unpaid]); and has previously made commitments to involve members of the following kidney patient/advocacy organizations in kidney research methods-focused scientific conferences or technical expert panels: American Association of Kidney Patients (*via* board member Jenny Kitsen) and Renal Support Network (*via* President and Founder Lori Hartwell); the only one to have taken place within last 24 months is Voice of the Patient (*via* founder Kevin Fowler).

Funding

This project was funded by the Office of the National Coordinator for Health Information Technology (ONC) in the Office of the Secretary within the U.S. Department of Health and Human Services, through a contract awarded to Booz Allen Hamilton, Inc. (contract number: HHSP233201500132I).

Acknowledgments

This paper was authored by the study team from Booz Allen Hamilton and University of California at San Francisco. The authors would like to acknowledge the contributions of the following individuals from ONC: Jiuyi Hua, Adam Wong, Alda Yuan, Stephanie Garcia, and Diana Ciricean. The study team wishes to acknowledge the in-depth guidance and input provided throughout the course of this project by the Technical Expert Panel: Peter Chang (Co-Director, Center for AI in Diagnostic Medicine, University of California, Irvine School of Medicine), Mark DePristo (Founder & CEO, BigHat Biosciences), Kevin Fowler (President, The Voice of the Patient), James Hickman (Product Lead, Epic), Eileen Koski (Program Director for Health and Data Insights, IBM), and Jarcy Zee, (Assistant Professor of Biostatistics, University of Pennsylvania). The authors also thank the following data scientists from Booz Allen for their input and expert advice on ML methodology: Timothy Fries, Cecily Abraham, Edward Raff, Lauren Neal, and Julio Gonzalez.

The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the US government.

Author Contributions

M. Estrella, K. Genberg, L. Han, M. Keating, S. Rankin, R. Scherzer, M.G. Shlipak, S. Tenney, and K.J. Wilkins were responsible for the investigation; M. Estrella, K. Genberg, M. Keating, M.G. Shlipak, and S. Tenney were responsible for supervision; M. Estrella, L. Han, M. Keating, S. Rankin, R. Scherzer, M.G. Shlipak, S. Tenney, and K.J. Wilkins were responsible for conceptualization and methodology; M. Estrella, L. Han, S. Rankin, R. Scherzer, M.G. Shlipak, and S. Tenney wrote the original draft of the manuscript; K. Genberg, L. Han, M. Keating, M. Rahn, S. Rankin, and S. Tenney were responsible for resources; K. Genberg, L. Han, M. Keating, M. Rahn, and S. Tenney were responsible for project administration; K. Genberg, M. Keating, and M. Rahn were responsible for funding acquisition; S. Rankin was responsible for data curation, formal analysis, software, and validation; and all authors reviewed and edited the manuscript.

Data Sharing Statement

Partial restrictions to the data and/or materials apply: data used for this project were obtained from USRDS *via* a data use agreement. The training dataset generated for this project is expected to be hosted by USRDS at a future date.

Supplemental Material

This article contains the following supplemental material online at http://kidney360.asnjournals.org/lookup/suppl/doi:10.34067/KID.0007012021/-/DCSupplemental.

Study Dataset

- Study Dataset Preparation Methodology
- Supplemental Table 1
- Supplemental Table 2
- Data Dictionary
  - o Predictors taken directly from the USRDS data
  - o Supplemental Table 3
  - o Predictors derived from USRDS datasets
  - o Supplemental Table 4
- Outliers
- Supplemental Table 5

Machine Learning

- Hyperparameter Tuning of XGBoost Models
- Supplemental Table 6
- XGBoost Imputed Model Results
- Supplemental Table 7
- Supplemental Table 8
- Supplemental Figure 1

Project Resources as a Foundation for Future Work

References

1. Soucie JM, McClellan WM: Early death in dialysis patients: Risk factors and impact on incidence and mortality rates. *J Am Soc Nephrol* 7: 2169–2175, 1996 https://doi.org/10.1681/ASN.V7102169
2. Chan KE, Maddux FW, Tolkoff-Rubin N, Karumanchi SA, Thadhani R, Hakim RM: Early outcomes among those initiating chronic dialysis in the United States. *Clin J Am Soc Nephrol* 6: 2642–2649, 2011 https://doi.org/10.2215/CJN.03680411
3. Foley RN, Chen S-C, Solid CA, Gilbertson DT, Collins AJ: Early mortality in patients starting dialysis appears to go unregistered. *Kidney Int* 86: 392–398, 2014 https://doi.org/10.1038/ki.2014.15
4. United States Renal Data System: *2019 USRDS Annual Data Report: Epidemiology of Kidney Disease in the United States*, Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2019
5. O'Connor NR, Kumar P: Conservative management of end-stage renal disease without dialysis: A systematic review. *J Palliat Med* 15: 228–235, 2012 https://doi.org/10.1089/jpm.2011.0207
6. Wongrakpanich S, Susantitaphong P, Isaranuwatchai S, Chenbhanich J, Eiam-Ong S, Jaber BL: Dialysis therapy and conservative management of advanced chronic kidney disease in the elderly: A systematic review. *Nephron* 137: 178–189, 2017 https://doi.org/10.1159/000477361
7. Renal Physicians Association: Shared decision-making in the appropriate initiation of and withdrawal from dialysis clinical practice guideline. Available at: https://cdn.ymaws.com/www.renalmd.org/resource/resmgr/ESRD_Guidelines/Recommendations_Summary.pdf. Accessed May 20, 2022
8. Wong SPY, McFarland LV, Liu C-F, Laundry RJ, Hebert PL, O'Hare AM: Care practices for patients with advanced kidney disease who forgo maintenance dialysis. *JAMA Intern Med* 179: 305–313, 2019 https://doi.org/10.1001/jamainternmed.2018.6197
9. Hussain JA, Flemming K, Murtagh FE, Johnson MJ: Patient and health care professional decision-making to commence and withdraw from renal dialysis: A systematic review of qualitative research. *Clin J Am Soc Nephrol* 10: 1201–1215, 2015 https://doi.org/10.2215/CJN.11091114
10. Thamer M, Kaufman JS, Zhang Y, Zhang Q, Cotter DJ, Bang H: Predicting early death among elderly dialysis patients: Development and validation of a risk score to assist shared decision making for dialysis initiation. *Am J Kidney Dis* 66: 1024–1032, 2015 https://doi.org/10.1053/j.ajkd.2015.05.014
11. Doi T, Yamamoto S, Morinaga T, Sada KE, Kurita N, Onishi Y: Risk score to predict 1-year mortality after haemodialysis initiation in patients with stage 5 chronic kidney disease under predialysis nephrology care. *PLoS One* 10: e0129180, 2015 https://doi.org/10.1371/journal.pone.0129180
12. Floege J, Gillespie IA, Kronenberg F, Anker SD, Gioni I, Richards S, Pisoni RL, Robinson BM, Marcelli D, Froissart M, Eckardt KU: Development and validation of a predictive mortality risk score from a European hemodialysis cohort. *Kidney Int* 87: 996–1008, 2015 https://doi.org/10.1038/ki.2014.419
13. Cohen LM, Ruthazer R, Moss AH, Germain MJ: Predicting six-month mortality for patients who are on maintenance hemodialysis. *Clin J Am Soc Nephrol* 5: 72–79, 2010 https://doi.org/10.2215/CJN.03860609
14. Ramspek CL, Voskamp PW, van Ittersum FJ, Krediet RT, Dekker FW, van Diepen M: Prediction models for the mortality risk in chronic dialysis patients: A systematic review and independent external validation study. *Clin Epidemiol* 9: 451–464, 2017 https://doi.org/10.2147/CLEP.S139748

15. Barrett BJ, Parfrey PS, Morgan J, Barré P, Fine A, Goldstein MB, Handa SP, Jindal KK, Kjellstrand CM, Levin A, Mandin H, Muirhead N, Richardson RM: Prediction of early death in end-stage renal disease patients starting dialysis. *Am J Kidney Dis* 29: 214–222, 1997 https://doi.org/10.1016/S0272-6386(97)90032-9
16. Couchoud C, Labeeuw M, Moranne O, Allot V, Esnault V, Frimat L, Stengel B; French Renal Epidemiology and Information Network (REIN) registry: A clinical score to predict 6-month prognosis in elderly patients starting dialysis for end-stage renal disease. *Nephrol Dial Transplant* 24: 1553–1561, 2009 https://doi.org/10.1093/ndt/gfn698
17. Foley RN, Parfrey PS, Hefferton D, Singh I, Simms A, Barrett BJ: Advance prediction of early death in patients starting maintenance dialysis. *Am J Kidney Dis* 23: 836–845, 1994 https://doi.org/10.1016/S0272-6386(12)80137-5
18. Wagner M, Ansell D, Kent DM, Griffith JL, Naimark D, Wanner C, Tangri N: Predicting mortality in incident dialysis patients: An analysis of the United Kingdom Renal Registry. *Am J Kidney Dis* 57: 894–902, 2011 https://doi.org/10.1053/j.ajkd.2010.12.023
19. Obi Y, Nguyen DV, Zhou H, Soohoo M, Zhang L, Chen Y, Streja E, Sim JJ, Molnar MZ, Rhee CM: Development and validation of prediction scores for early mortality at transition to dialysis. *Mayo Clinic Proc* 93: 1224–1235, 2018, https://doi.org/10.1016/j.mayocp.2018.04.017
20. Akbilgic O, Obi Y, Potukuchi PK, Karabayir I, Nguyen DV, Soohoo M, Streja E, Molnar MZ, Rhee CM, Kalantar-Zadeh K, Kovesdy CP: Machine learning to identify dialysis patients at high death risk. *Kidney Int Rep* 4: 1219–1229, 2019 https://doi.org/10.1016/j.ekir.2019.06.009
21. Garcia-Montemayor V, Martin-Malo A, Barbieri C, Bellocchio F, Soriano S, Pendon-Ruiz de Mier V, Molina IR, Aljama P, Rodriguez M: Predicting mortality in hemodialysis patients using machine learning analysis. *Clin Kidney J* 14: 1388–1395, 2020 https://doi.org/10.1093/ckj/sfaa126
22. Sheng K, Zhang P, Yao X, Li J, He Y, Chen J: Prognostic machine learning models for first-year mortality in incident hemodialysis patients: Development and validation study. *JMIR Med Inform* 8: e20578, 2020 https://doi.org/10.2196/20578
23. Chen T, Guestrin C: Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13–17, 2016, pp 785–794 https://doi.org/10.1145/2939672.2939785
24. Salkind NJ: *Encyclopedia of research design*, Thousand Oaks, CA, Sage, 2010 https://doi.org/10.4135/9781412961288
25. Jakobsen JC, Gluud C, Wetterslev J, Winkel P: When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Med Res Methodol* 17: 162, 2017 https://doi.org/10.1186/s12874-017-0442-1
26. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration): A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009 https://doi.org/10.7326/0003-4819-150-9-200905050-00006
27. Buuren S, Groothuis-Oudshoorn K: Mice: Multivariate imputation by chained equations in R. *J Stat Softw* 45: 1–67, 2011
28. Zhang Z, Ho KM, Hong Y: Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care* 23: 112, 2019 https://doi.org/10.1186/s13054-019-2411-z
29. Tang C, Li J, Xu D, Liu X, Hou W, Lyu K, Xiao S, Xia Z: [Comparison of machine learning method and logistic regression model in prediction of acute kidney injury in severely burned patients]. *Zhonghua Shao Shang Za Zhi* 34: 343–348, 2018 https://doi.org/10.3760/cma.j.issn.1009-2587.2018.06.006
30. Tabassum S, Sampa MB, Islam R, Yokota F, Nakashima N, Ahmed A: A data enhancement approach to improve machine learning performance for predicting health status using remote healthcare data. Presented at the 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Dhaka, Bangladesh, November 28–29, 2020, pp 308–312 https://doi.org/10.1109/ICAICT51780.2020.9333506
31. Zadrozny B, Elkan C: Transforming classifier scores into accurate multiclass probability estimates. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada, July 23–26, 2002, pp 694–699 https://doi.org/10.1145/775047.775151
32. Sy J, Johansen KL: The impact of frailty on outcomes in dialysis. *Curr Opin Nephrol Hypertens* 26: 537–542, 2017 https://doi.org/10.1097/MNH.0000000000000364
33. Foley RN, Collins AJ: The USRDS: What you need to know about what it can and can't tell us about ESRD. *Clin J Am Soc Nephrol* 8: 845–851, 2013 https://doi.org/10.2215/CJN.06840712

Received: November 2, 2021 **Accepted:** July 15, 2022