



Development of a scoring tool for predicting prolonged length of hospital stay in peritoneal dialysis patients through data mining

Jingyi Wu^{1,2}, Guilan Kong^{1,2}, Yu Lin¹, Hong Chu³, Chao Yang³, Ying Shi⁴, Haibo Wang^{1,2,4}, Luxia Zhang^{1,2,3}; on behalf of China Kidney Disease Network Working Group

¹National Institute of Health Data Science, Peking University, Beijing, China; ²Advanced Institute of Information Technology, Peking University, Hangzhou, China; ³Renal Division, Department of Medicine, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China; ⁴China Standard Medical Information Research Center, Shenzhen, China

Contributions: (I) Conception and design: J Wu, G Kong, L Zhang; (II) Administrative support: L Zhang, G Kong, H Wang; (III) Provision of study materials or patients: G Kong, L Zhang, H Wang; (IV) Collection and assembly of data: Y Shi, H Wang; (V) Data analysis and interpretation: J Wu, Y Lin, C Yang, H Chu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Guilan Kong, PhD. National Institute of Health Data Science, Peking University, 38 Xueyuan Road, Haidian District, Beijing 100191, China. Email: guilan.kong@hsc.pku.edu.cn; Luxia Zhang, MD, MPH. Renal Division, Department of Medicine, Peking University First Hospital, 8 Xishiku St, Xicheng District, Beijing 100034, China. Email: zhanglx@bjmu.edu.cn.

Background: The hospital admission rate is high in patients treated with peritoneal dialysis (PD), and the length of stay (LOS) in the hospital is a key indicator of medical resource allocation. This study aimed to develop a scoring tool for predicting prolonged LOS (pLOS) in PD patients by combining machine learning and traditional logistic regression (LR).

Methods: This study was based on patient data collected using the Hospital Quality Monitoring System (HQMS) in China. Three machine learning methods, classification and regression tree (CART), random forest (RF), and gradient boosting decision tree (GBDT), were used to develop models to predict pLOS, which is longer than the average LOS, in PD patients. The model with the best prediction performance was used to identify predictive factors contributing to the outcome. A multivariate LR model based on the identified predictors was then built to derive the score assigned to each predictor. Finally, a scoring tool was developed, and it was tested by stratifying PD patients into different pLOS risk groups.

Results: A total of 22,859 PD patients were included in our study, with 25.2% having pLOS. Among the three machine learning models, the RF model achieved the best prediction performance and thus was used to identify the 10 most predictive variables for building the scoring system. The multivariate LR model based on the identified predictors showed good discrimination power with an AUROC of 0.721 in the test dataset, and its coefficients were used as a basis for scoring tool development. On the basis of the developed scoring tool, PD patients were divided into three groups: low risk (≤ 5), median risk [5–10], and high risk (> 10). The observed pLOS proportions in the low-risk, median-risk, and high-risk groups in the test dataset were 11.4%, 29.5%, and 54.7%, respectively.

Conclusions: This study developed a scoring tool to predict pLOS in PD patients. The scoring tool can effectively discriminate patients with different pLOS risks and be easily implemented in clinical practice. The pLOS scoring tool has a great potential to help physicians allocate medical resources optimally and achieve improved clinical outcomes.

Keywords: Length of stay (LOS); logistic regression (LR); machine learning; peritoneal dialysis (PD); scoring methods

Submitted Jan 23, 2020. Accepted for publication Aug 21, 2020.

doi: 10.21037/atm-20-1006

View this article at: <http://dx.doi.org/10.21037/atm-20-1006>

Introduction

The prevalence of chronic kidney disease (CKD) has steadily increased annually across all stages of CKD (1). CKD affects approximately 12.8% men and 14.6% women worldwide (2). Over 690 million people worldwide have CKD, which is nearly 1.5 times the number of people who have diabetes and approximately 20 times the number of people with human immunodeficiency virus infection or acquired immune deficiency syndrome (3). The prevalence of end-stage kidney disease (ESKD), the advanced stage of CKD, has been increasing due to the aging of the population and the increasing prevalence of diabetes and hypertension (3). Between 4.9 and 9.7 million people were estimated to need renal replacement therapy (RRT) in 2010, and Asia was the region having the largest number of people needing RRT (4). With the increase in ESKD prevalence, the number of patients needing RRT should also have increased in the past decade. CKD and ESKD have imposed heavy burdens on healthcare budgets. In the US, total Medicare spending on patients with CKD and ESKD was in excess of \$120 billion in 2017, and the spending for patients with ESKD totaled \$35.9 billion, accounting for 7.2% of the overall Medicare paid claims (1). In China, the prevalence of CKD was 10.8% in 2010 (5), and the prevalence of ESKD was 237.3 per million population in 2012 (6). The average inpatient spending on patients with CKD and ESKD was about \$3,750 and \$3,472, respectively, in 2015 (7). On the basis of these data, the total inpatient spending in China was in excess of \$465 billion for patients with CKD and \$945 million for patients with ESKD in 2015. Inpatient hospital care accounts for more than 40% of ESKD cost and thus is set as the target for cost reduction (8).

Length of stay (LOS) in the hospital is highly predictive of in-hospital cost among the factors contributing to the total cost of hospitalization (9). Accurate prediction of LOS can provide useful prognostic information that may help clinicians make optimal use of medical resources and produce better clinical decisions. Previous studies developed LOS prediction tools for conditions such as critical care (10-12), heart diseases (13-17), and liver diseases (12,18). Some of these tools were customized from traditional severity scoring tools (14,16), some were developed via logistic regression (LR) with or without severity scores as independent predictors (12,17), and some were developed using machine learning methods (10,15). Meanwhile, some new and specific scoring tools have been developed for LOS

prediction (18,19).

However, LOS prediction tools for ESKD are lacking. ESKD patients treated with peritoneal dialysis (PD) tend to spend more days in the hospital than patients treated with hemodialysis (20,21). In the present study, we aimed to develop a specific scoring tool to predict LOS for patients treated with PD by combining machine learning and traditional LR.

Methods

Study population

The Hospital Quality Monitoring System (HQMS) is a national database containing standardized electronic inpatient discharge records from 878 tertiary hospitals in 2015 (41.4% of tertiary hospitals in China) (22). As of December 2015, over 40 million inpatient discharge records in 31 provinces have been collected by the HQMS under the authority of the National Health Commission of China. As a part of standard practice in China, the standardized electronic inpatient discharge record of one patient must be completed by the physicians who have the most comprehensive understanding of the patient's medical condition to ensure legal validity of the records. Data in HQMS included patient demographics, clinical diagnosis, procedures and operations, and expenditure breakdowns. All personal information was deidentified, and no patient privacy data were identified. Data in HQMS can only be available with the written approval of the Bureau of Medical Administration and Medical Service Supervision, National Health Commission of China, and hence is not open to readers.

We extracted PD patients from the HQMS database between 2013 and 2015. Patients who met the following criteria were included: (I) age between 18 and 100 years and (II) condition meeting the definition of PD. Exclusion criteria were as follows: (I) diagnosis of acute kidney injury or kidney transplantation; (II) death in the hospital; (III) LOS longer than 30 days; and (IV) readmission within a day after previous hospital discharge. We excluded patients with LOS longer than 30 days for the following reasons. First, existing studies in the literature excluded extreme outliers (23,24). After consulting experienced clinicians, we considered LOS longer than 30 days as extreme outliers in our study. Second, the LOS pattern of patients with extreme LOS may be different from that of patients with normal LOS (25). For patients with more than one hospitalization in this study, we randomly selected one record to ensure

that all observations were independent and patients with varying severities were included in model development. In the literature, two methods have been used to deal with this problem: (I) selecting the first hospitalization record and (II) randomly selecting one hospitalization record. Compared with the first method, the second method may help include patients with varying severities (26). The patients treated with PD were identified from records of discharge diagnoses and in-hospital medical operations using International Classification of Diseases-10 (ICD-10) codes. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Ethics Committee of Peking University First Hospital [No. 2015[928]], and informed consent from patients was not required because it was a secondary use of deidentified patient data and the identifiable personal information was absent.

Data collection

After reviewing prior studies and consulting experienced clinicians, we determined 34 variables as candidate predictors. All selected predictor variables were available at admission. The most predictive variables among the 34 candidate predictors were further identified using a machine learning method for the development of our scoring tool. The candidate predictor variables included demographic characteristics, disease characteristics, and clinical characteristics (Table 1). The categories of admission reasons and comorbidities were determined by researchers together with experienced clinicians. Fewer than 15% of records had missing values for the variables of nationality and admission type, and the missing values were considered as a special category. Referring to previous studies (27), we retained the records with missing values for variables with data missing rates lower than 15%. The reasons were as follows. Considering the missing values as a special category for a variable will not affect the correlation between the variable and the outcome of our study. Instead, it enables us to keep more data and reduce the selection bias that may occur if these records are excluded.

In the present study, the clinical outcome of having a prolonged LOS (pLOS), which is defined as an LOS longer than the average LOS, 16 days for patients with ESKD in China (28), was used as the outcome event.

Machine learning methods

In the machine learning area, ensemble learning models

usually perform better than single learning models (29,30). In the present study, both models were employed to develop LOS prediction tools and identify the predictive factors with the most contribution to the outcome of pLOS. Three machine learning methods, classification and regression tree (CART) (31), random forest (RF) (32), and gradient boosting decision tree (GBDT) (33), were employed in this study. Grid search and stratified five-fold cross validation were combined to find the optimal parameters for the three machine learning models.

CART is a type of tree-like decision making model. A CART model can be derived from the training dataset using various algorithms. Assuming that a set of input variables (independent variables) and an output variable (dependent variable) were recorded for each case in the training dataset, the trained decision tree would be a classification tree if the output were categorical and a regression tree if the output were continuous. For classification tree training, Information Gain or the Gini Index can be employed as criteria for variable selection at each node. For the regression tree, the mean squared error or similar index can be employed as a criterion for variable selection in tree growth. If a trained CART model is used, an estimate can be obtained for a new case after traversing tree nodes from root to leaf by selecting the nodes that represent the category or value for the independent variable of the new case. We used the *DecisionTreeClassifier* package in Python 3.7 to construct the CART model. The optimal parameters of the CART model were found as follows: the maximal depth was 6, and the minimum number of samples required to split an internal node was 250.

RF is a type of ensemble learning model with decision trees acted as its basic learning models. In an RF model training, its basic decision trees should be trained independently and in parallel. Two randomness patterns occur in an RF model training. In one, the training dataset for each tree is a bootstrap sample of the whole training dataset. In the other, a fixed number of variables are selected at random for each tree growth. To classify a new object using an RF model, the input variables are placed down each of the trees in the forest, and each tree would give a classification for the object as a vote. The RF model would choose the classification having the most votes. We used the *RandomForestClassifier* package in Python 3.7 to construct the RF model. The optimal parameters of the RF model were found as follows: the number of decision trees was 300, the maximal depth of each decision tree was 28, the minimum number of samples required to split an internal

Table 1 Candidate predictor variables considered for the scoring system

Category	Variable name
Demographic characteristics	Age
	Sex
	Nationality
	Place of residence
	Insurance type
Disease characteristics [†]	Admission reason
	Specific cause of CKD
Comorbidities [†]	Diabetes
	Hypertension
	Heart failure
	Cardiac arrhythmia
	Coronary heart disease
	Stroke
	Pulmonary infections
	Infections except pulmonary infections
	Tumor
	Gastrointestinal hemorrhage
	Gastrointestinal inflammation/ulcer
	Gallbladder disease
	Liver disease
	Kidney stone
	Peripheral vascular disease
	Gout
	Hyperparathyroidism
	Hypoparathyroidism
	Hyperlipidemia
	Fracture
Clinical characteristics	Admission type
	Number of hospitalizations within 6 months
	Number of emergency admissions within 6 months
	Admission department
	Planned admission or not
	Admission day of the week
	Admission to the same hospital as the last or not

[†], disease characteristics and comorbidities were extracted using ICD-10 codes. CKD, chronic kidney disease.

node was 20, and the number of variables selected at each node was 10.

GBDT is also a type of ensemble learning models with decision trees as base models. However, a GBDT model is different from an RF model in two aspects. First, all decision trees in a GBDT model are trained in line other than in parallel as in the RF model training. A decision tree in a GBDT model is trained with the object of minimizing the residual between the trees-based prediction result and the observed outcome. Second, a GBDT model is a stage-wise additive model with the outputs of all decision trees as sequential inputs, whereas an RF model obtains its final classification by a majority voting system. We used the *GradientBoostingClassifier* package in Python 3.7 to construct the GBDT model. The optimal parameters of the GBDT model were found as follows: the number of decision trees was 250, the fraction of samples used for each decision tree was 0.8, the maximal depth was 5, the minimum number of samples required to split an internal node was 250, and the number of variables selected at each node was 8.

In addition to the abovementioned machine learning methods, the traditional LR model was employed as well to develop a LOS prediction model. This model was then used as the benchmark model in performance comparison. The independent variable set of the LR model was the same as that used for machine learning model development.

A stratified five-fold cross validation method was employed for model development and validation. The whole dataset was split into five folds, and each fold contained approximately the same percentage of samples of each class. Any four folds were used for model training, whereas the remaining fold was employed for the model test. The prediction performance of the four models was evaluated using Brier score (34), area under the receiver operation characteristic curve (AUROC) (34), and estimated calibration index (ECI) (35). The Brier score is an overall performance measure, which ranges from 0 for a perfect model to 1 for a perfect inaccurate model. A higher AUROC represents a stronger discrimination power of the corresponding model, and a lower ECI suggests a stronger calibration power. The model with the best prediction performance after five rounds of training and test should be the optimal model.

Combined approach for deriving a scoring tool

Traditionally, scoring tools are developed via LR (18). In particular, independent predictors contributing to

dependent outcome are identified through multivariate LR analysis, and then the score to be assigned to each predictor can be determined by the odds ratio associated with it.

In the present study, we proposed to combine machine learning and LR to develop a specific scoring tool for pLOS prediction in patients treated with PD. The diagram illustrating our proposed approach for building the scoring tool is shown in *Figure 1*. First, we employed machine learning methods to develop pLOS prediction tools for patients treated with PD, and the model with the best prediction performance was used to identify the predictive factors contributing to the outcome. Second, we constructed a multivariate LR model based on the identified 10 predictors and obtained the P values and odds ratios corresponding to each predictor. Finally, predictors with $P < 0.05$ were assigned relative scores according to their odds ratios, and a scoring tool was formed. The top 10 predictors were selected rather than the top 5 or less predictors to build the scoring tool for three reasons. First, as shown in a previous study (36), a scoring tool based on 10 variables or less could be easily implemented in clinical practice. Second, theoretically, 10 variables can provide more predictive information for LOS prediction than 5 or less predictors. Third, the 10 variables included in the final multivariate LR model were filtered again based on the P value of each variable to ensure that all the variables in the final scoring system are the most predictive for LOS. For scoring tool development, we split the whole dataset randomly into two parts, 80% as the training dataset and 20% as the test dataset. The multivariate LR model was built using the training dataset, and the scoring tool was evaluated by stratifying patients into different pLOS risk groups in the test dataset.

Results

A total of 22,859 patients treated with PD were included in our study. The average age of this cohort was 51.9 ± 14.9 years, and the proportion of male patients was 55.6%. The proportion of patients with pLOS was 25.2%. The baseline characteristics of patients treated with PD are shown in *Table 2*. Approximately 39.5% of the patients treated with PD were from eastern China, and 37.8% were covered by the urban employee basic medical insurance (UEBMI). The most frequent admission reasons included ESKD (58.3%), dialysis access (18.4%), dialysis complications (15.3%), surgery (2.3%), and hypertension (0.5%). The most frequent CKD causes were glomerulonephropathy

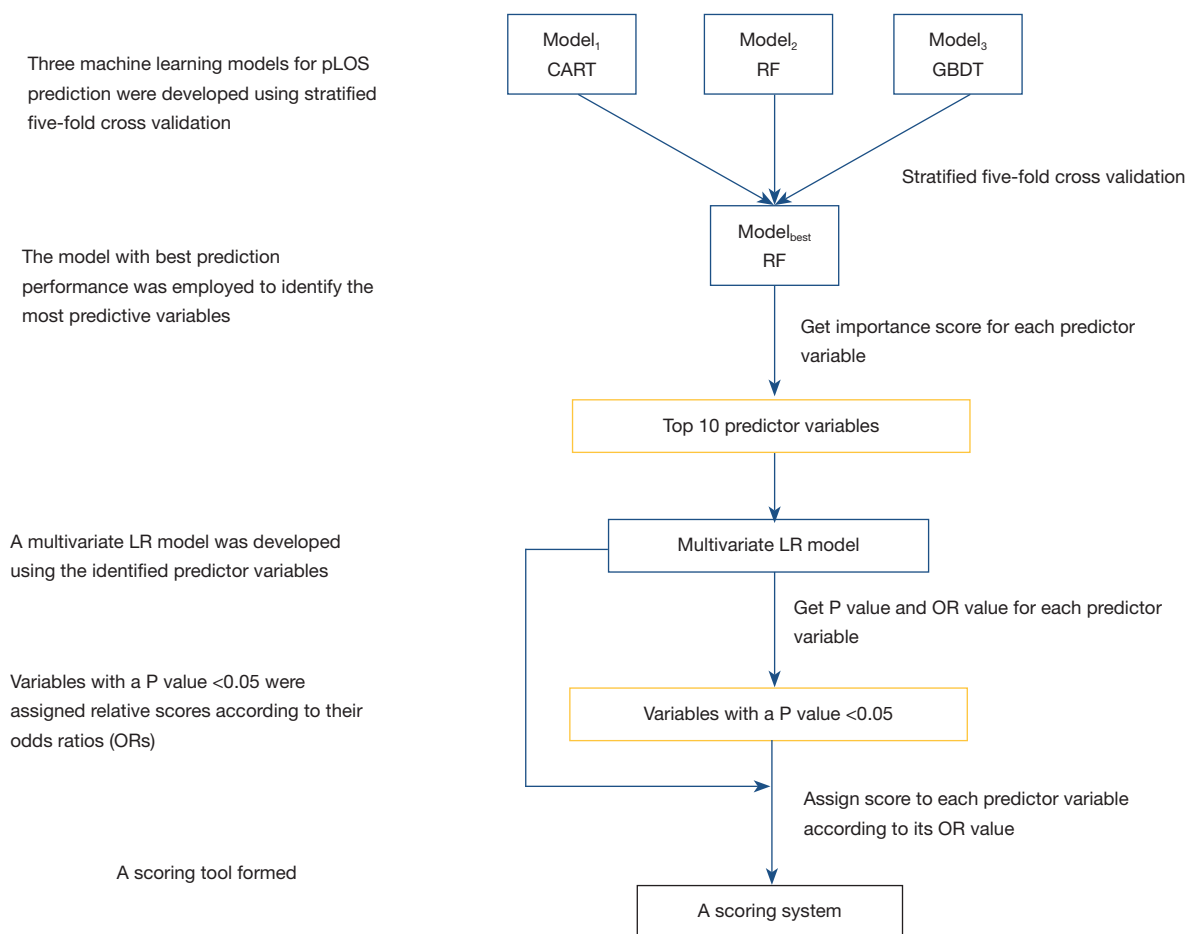


Figure 1 Diagram of our proposed approach for building a scoring tool. LR, logistic regression; RF, random forest; CART, classification and regression tree; GBDT, gradient boosting decision tree.

(GN) (22.1%), hypertensive nephropathy (HN) (12.5%), and diabetic nephropathy (DN) (11.6%). About half of the patients treated with PD had two or more comorbidities.

Table 3 shows the prediction performance of the four models, namely, LR, CART, RF, and GBDT, using five-fold cross validation. Among the four models, the RF model achieved the best prediction performance in terms of overall prediction performance (Brier score, 0.158), discrimination (AUROC, 0.756), and calibration (ECI, 7.883). The 10 most predictive factors identified by the RF model are listed in *Figure 2*.

The prediction performance of the multivariate LR model constructed using the 10 most predictive variables identified by the RF model is summarized in *Table 4*. The AUROC of the constructed multivariate LR model was 0.728 in the training dataset (80%) and 0.721 in the test dataset (20%), which demonstrated the good discrimination

power of the model. The scoring system built based on the multivariate LR model is illustrated in *Table 5*.

The total score of the scoring system ranged from 0 to 23 for each patient. According to the total score distribution in the study cohort, we divided all patients treated with PD into three groups: low risk (≤ 5), median risk [5–10], and high risk (> 10). Similar to traditional scoring tools (36), the scoring tool developed in this study was evaluated by testing the prediction performance of the LR model and the risk stratification performance of the scoring tool. Comparison between the predicted pLOS probability generated by the LR model and the observed pLOS proportion in different pLOS risk groups in the training and test datasets is shown in *Figure 3*. The mean predicted pLOS probabilities generated by the LR model were similar to the observed pLOS proportions in various risk groups, and this result demonstrated that the LR model had superior calibration

Table 2 Characteristics of patients treated with PD in the final cohort

Item	Total, n (%)	With pLOS, n (%)	Without pLOS, n (%)	Proportion difference (%) (with pLOS – without pLOS)
N	22,859	5,754 (25.2)	17,105 (74.8)	–
Age, year	51.9±14.9	52.7±15.0	51.6±14.9	–
Sex				
Female	10,142 (44.4)	2,597 (45.1)	7,545 (44.1)	1.0
Male	12,717 (55.6)	3,157 (54.9)	9,560 (55.9)	–1.0
Nationality				
Han	18,089 (79.1)	4,597 (79.9)	13,492 (78.9)	1.0
Others	931 (4.1)	225 (3.9)	706 (4.1)	–0.2
Unclear	3,839 (16.8)	932 (16.2)	2,907 (17.0)	–0.8
Place of residence				
Eastern China	9,023 (39.5)	1,992 (34.6)	7,031 (41.1)	–6.5
Northern China	2,138 (9.4)	680 (11.8)	1,458 (8.5)	3.3
Central China	3,267 (14.3)	909 (15.8)	2,358 (13.8)	2.0
Southern China	3,803 (16.6)	1,054 (18.3)	2,749 (16.1)	2.2
Southwestern China	2,788 (12.2)	670 (11.6)	2,118 (12.4)	–0.8
Northwestern China	1,043 (4.6)	188 (3.3)	855 (5.0)	–1.7
Northeastern China	797 (3.5)	261 (4.5)	536 (3.1)	1.4
Insurance type				
UEBMI	8,635 (37.8)	2,121 (36.9)	6,514 (38.1)	–1.2
URBMI	2,083 (9.1)	520 (9.0)	1,563 (9.1)	–0.1
NRCMS	5,821 (25.5)	1,544 (26.8)	4,277 (25.0)	1.8
Free medical care	312 (1.4)	80 (1.4)	232 (1.4)	0.0
Self-paid treatment	3,318 (14.5)	788 (13.7)	2,530 (14.8)	–1.1
Others	2,690 (11.8)	701 (12.2)	1,989 (11.6)	0.6
Admission reason				
ESKD [†]	13,329 (58.3)	2,003 (34.8)	11,326 (66.2)	–31.4
Dialysis access	4,213 (18.4)	2,271 (39.5)	1,942 (11.4)	28.1
Dialysis complications	3,502 (15.3)	902 (15.7)	2,600 (15.2)	0.5
Diabetes	188 (0.8)	50 (0.9)	138 (0.8)	0.1
Hypertension	349 (1.5)	85 (1.5)	264 (1.5)	0.0
Heart failure	35 (0.2)	10 (0.2)	25 (0.1)	0.1
Coronary heart disease	125 (0.5)	27 (0.5)	98 (0.6)	–0.1
Stroke	90 (0.4)	26 (0.5)	64 (0.4)	0.1
Infection	228 (1.0)	42 (0.7)	186 (1.1)	–0.4
Hypertension	114 (0.5)	36 (0.6)	78 (0.5)	0.1

Table 2 (continued)

Table 2 (continued)

Item	Total, n (%)	With pLOS, n (%)	Without pLOS, n (%)	Proportion difference (%) (with pLOS – without pLOS)
Gastrointestinal hemorrhage	20 (0.1)	6 (0.1)	14 (0.1)	0.0
Tumor	51 (0.2)	15 (0.3)	36 (0.2)	0.1
Severe anemia	84 (0.4)	51 (0.9)	33 (0.2)	0.7
Surgery	531 (2.3)	230 (4.0)	301 (1.8)	2.2
Specific cause of CKD				
Diabetic nephropathy	2,657 (11.6)	752 (13.1)	1,905 (11.1)	2.0
Hypertensive nephropathy	2,846 (12.5)	636 (11.1)	2,210 (12.9)	-1.8
Glomerulonephropathy	5,061 (22.1)	1,311 (22.8)	3,750 (21.9)	0.9
Tubulointerstitial nephropathy	394 (1.7)	106 (1.8)	288 (1.7)	0.1
Obstructive nephropathy	337 (1.5)	113 (2.0)	224 (1.3)	0.7
Others	11,564 (50.6)	2,836 (49.3)	8,728 (51.0)	-1.7
Number of comorbidities				
0	3,460 (15.1)	709 (12.3)	2,751 (16.1)	-3.8
1	7,853 (34.4)	1,741 (30.3)	6,112 (35.7)	-5.4
2	6,239 (27.3)	1,678 (29.2)	4,561 (26.7)	2.5
3	3,409 (14.9)	983 (17.1)	2,426 (14.2)	2.9
≥4	1,898 (8.3)	643 (11.2)	1,255 (7.3)	3.9

[†], admission reason was recorded as ESKD in the electronic inpatient discharge record. UEBMI, urban employee basic medical insurance; URBMI, urban resident basic medical insurance; NRCMS, new rural cooperative medical insurance; ESKD, end-stage kidney disease; pLOS, prolonged length of stay; CKD, chronic kidney disease.

power. Therefore, the scoring system derived from the LR model had a reliable base. Meanwhile, the observed pLOS proportions in the low-risk, median-risk, and high-risk groups in the test dataset were 11.4%, 29.5%, and 54.7%, respectively. This result showed a significant increasing tendency of the observed pLOS proportion in the patients from low to high risk, and the observed pLOS proportion in the group with high pLOS risk was around twice the pLOS proportion in the adjacent group with low pLOS risk. Moreover, on the basis of the distribution of average LOS across the different pLOS risk groups in the training and test datasets, the average LOS for the patients from the low- to high-risk groups also showed a significant increasing tendency (*Figure 4*). The increasing tendency in the observed pLOS proportion and the average LOS across the different pLOS risk groups, and the close similarity between the LR predicted pLOS probabilities and the observed pLOS proportions in the different pLOS

groups, demonstrated the effectiveness of the scoring tool in stratifying the pLOS risk of patients treated with PD.

Discussion

The hospital admission rate is high in patients treated with PD (37), and LOS is a key indicator for medical resource allocation in hospitals. A practical tool that could accurately and quickly predict LOS in patients treated with PD would be helpful for nephrologists to improve the efficiency of medical resource allocation and to achieve improved outcomes for patients. However, no effective LOS prediction tools for patients treated with PD are available in existing studies.

In this study, we developed and validated a scoring system for stratifying the risk of pLOS in patients treated with PD by combining machine learning methods and the traditional LR model. The newly developed scoring system

Table 3 Prediction performance of the four models

Model	Brier score	AUROC	ECI
LR	0.161	0.743	8.036
CART	0.163	0.731	8.173
RF	0.158	0.756	7.883
GBDT	0.158	0.755	7.891

LR, logistic regression; CART, classification and regression tree; RF, random forest; GBDT, gradient boosting decision tree; AUROC, area under the receiver operation characteristic curve; ECI, estimated calibration index.

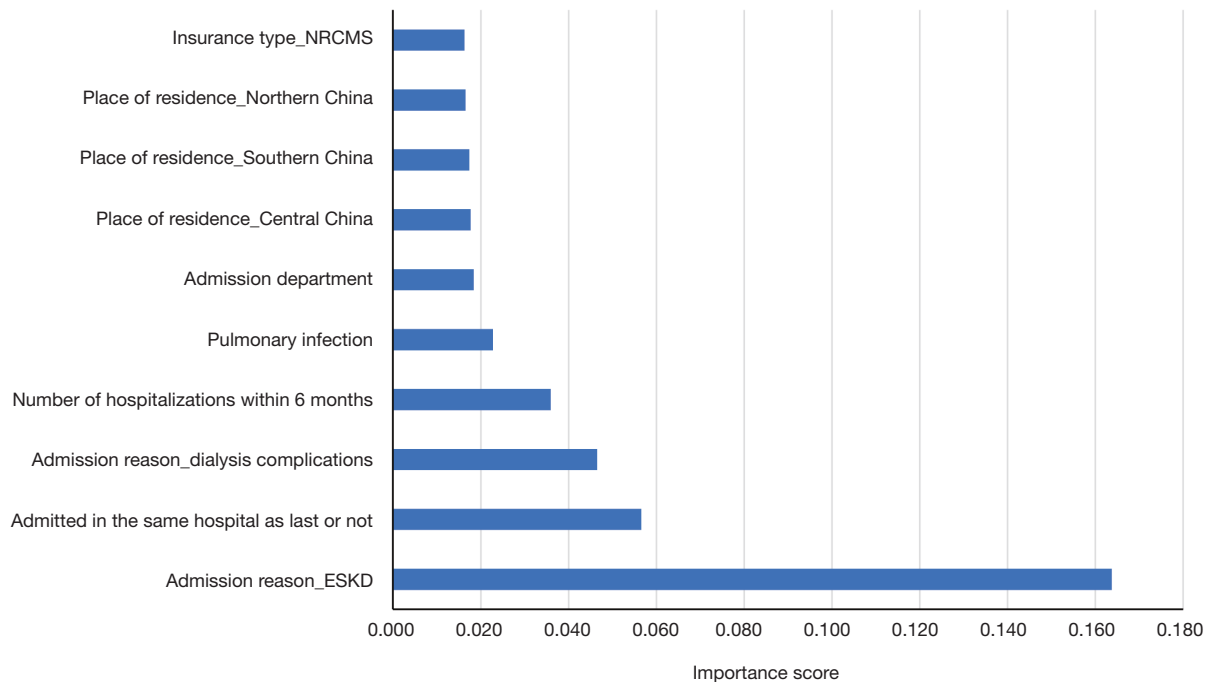


Figure 2 Ten most predictive factors identified by the RF model. ESKD, end-stage kidney disease; RF, random forest; NRCMS, new rural cooperative medical insurance.

effectively discriminated patients with different pLOS risks. The scoring system took advantage of the superior prediction performance of the machine learning model and the interpretability of the traditional LR model. The RF model had the best prediction performance among the three machine learning models in terms of overall prediction performance, discrimination, and calibration. Thus, it was employed to identify the most predictive variables contributing to outcomes. As an ensemble machine learning model, the RF model makes prediction by combining the outputs of a multitude of base models, thereby reducing the bias that may occur in single learning models (29).

Various LOS prediction tools have been developed for

other diseases. Some were scoring tools customized from existing clinical risk scoring tools. For example, Meadows *et al.* (13) customized the commonly used cardiac mortality risk scoring tool EuroSCORE to predict LOS in ICU for patients after cardiac surgery. Some were new scoring tools developed using traditional LR models. For example, Rana *et al.* (18) devised a scoring system based on an LR model to predict LOS for patients with liver transplantation and used univariate analysis to identify significant predictors for the scoring system. Others were LOS prediction models based on machine learning methods. For example, Chuang *et al.* (38) developed various machine learning models to predict LOS for patients who underwent general surgery

Table 4 The LR model derived using top 10 predictive variables identified by the RF model

Variable	Coefficient	P value	OR	95% confidence interval
Admission reason				
ESKD	-1.506	0.000*	0.222	0.206, 0.240
Dialysis complications	-0.861	0.000*	0.423	0.382, 0.468
Admission to the same hospital as the last or not				
No				
Yes	-0.552	0.000*	0.576	0.522, 0.636
Number of hospitalizations within 6 months				
0				
1	0.096	0.087	1.100	0.986, 1.228
2	-0.117	0.137	0.889	0.762, 1.038
3	-0.174	0.162	0.840	0.659, 1.073
≥4	-0.324	0.033*	0.723	0.537, 0.973
Comorbidity				
Pulmonary infections	0.619	0.000*	1.856	1.665, 2.071
Admission department				
Nephrology department	-0.155	0.000*	0.856	0.804, 0.911
Others				
Place of residence				
Northern China	0.126	0.033*	1.134	1.010, 1.273
Central China	0.433	0.000*	1.542	1.392, 1.707
Southern China	0.357	0.000*	1.429	1.297, 1.575
Insurance type				
NRCMS	-0.038	0.341	0.962	0.889, 1.042

*, $P < 0.05$. NRCMS, new rural cooperative medical insurance; LR, logistic regression; RF, random forest; ESKD, end-stage kidney disease.

and compared their prediction performances with those of traditional LR models. Their results showed that the ensemble machine learning model RF achieved much superior performance to the traditional LR model. In the literature, existing LOS prediction models could be classified into two types: (I) models derived from the LR model or customized from traditional severity scoring systems and (II) models developed using machine learning methods. The first type of models is easily interpretable for physicians. Meanwhile, the second type shows superior prediction performance to LR models; however, the internal reasoning process is difficult to express. In the present study, we developed a scoring tool for LOS

prediction by combining machine learning methods and the traditional LR model. Compared with the traditional approaches for building a scoring tool, we used a different method to identify the predictor variables. Traditionally, the predictive variables of a scoring tool are identified using univariate analysis or multivariate LR analysis (19,39). In our approach, three machine learning models for pLOS prediction were constructed initially, and then the model with the best prediction performance was used to assign importance scores to the included variables, thereby allowing the most predictive variables to be identified. Compared with existing LOS prediction models, our scoring tool for LOS prediction has several strengths.

Table 5 A scoring system for LOS prediction

Variable	Score
Admission reason	
ESKD	2
Dialysis complications	5
Others	7
Admission to the same hospital as the last or not	
No	2
Yes	0
Number of hospitalizations within 6 months	
<4	1
≥4	0
Comorbidity	
Pulmonary infections	2
Admission department	
Nephrology department	1
Others	0
Place of residence	
Northern China	1
Central China	2
Southern China	1

LOS, length of stay; ESKD, end-stage kidney disease.

First, the scoring tool took advantage of the interpretability of the traditional LR model and the superior prediction performance of machine learning methods. Unlike the traditional LR-based scoring systems, our newly developed scoring tool is based on a set of predictors identified by a machine learning method and is expected to demonstrate superior performance. Compared with prediction models developed directly through machine learning methods, our scoring tool for LOS prediction has better interpretability. Second, compared with prediction models in the form of complex equations or algorithms, a scoring tool for LOS prediction can be more easily implemented in clinical practice. Third, our scoring tool for LOS prediction is the first one specialized for patients treated with PD and it performed well on our national database. However, the prediction performance of our scoring tool is difficult to compare with those of existing LOS prediction models in the literature because they are specialized for different

patient groups. Nevertheless, it has great potential to support doctors in patient risk stratification and medical resource allocation.

Previous studies have attempted to build prediction models for patients treated with PD. Zhao *et al.* (40) used data from a single dialysis center in China to develop a Cox proportional hazards model for predicting 2-year mortality in patients treated with PD. Cao *et al.* (41) used data from a national multicenter cohort from China Peritoneal Dialysis Registry to establish a Cox proportional hazards model for predicting one-year mortality in patients treated with PD. Zhang *et al.* (42) used data from Henan Peritoneal Dialysis Registry to develop an LR model for predicting cerebrovascular disease mortality at 2 years for patients treated with PD. Tangri *et al.* (43) used data from a large, multicenter dataset, the United Kingdom Renal Registry, to predict PD technique failure by employing an artificial neural network model and compared its performance with that of the LR model. However, no models were developed for LOS prediction in patients with ESKD.

The most predictive variables for LOS prediction employed in our scoring system included admission reason, admission to the same hospital as the last or not, number of hospitalizations within 6 months, complications with pulmonary infections or not, admission department, and place of residence. After the RF model identified predictor variables, the two variables with missing data were eliminated from the final predictor set for the scoring tool. Admission reason was identified as the main factor affecting LOS in our study, which is consistent with findings of previous studies to the effect that admission reason was more important than other social characteristics in determining LOS (44,45). A patient's admission diagnosis leads to the initial course of treatment and accounts for the main differences in hospital care (46). Similar to previous studies (47,48), admission to a hospital different from the prior hospital is a risk factor for pLOS. The underlying mechanism behind this association may be insufficient systems for coordinating care information across hospitals (49,50). Patients treated with PD admitted to the nephrology department had a lower risk for pLOS than patients admitted to other departments, and the same impact has been verified in patients treated with hemodialysis (51). Previous studies also found that the number of hospitalizations within the past 6 months and place of residence were important indicators of LOS in patients (52-55).

The present study has several strengths. First, the

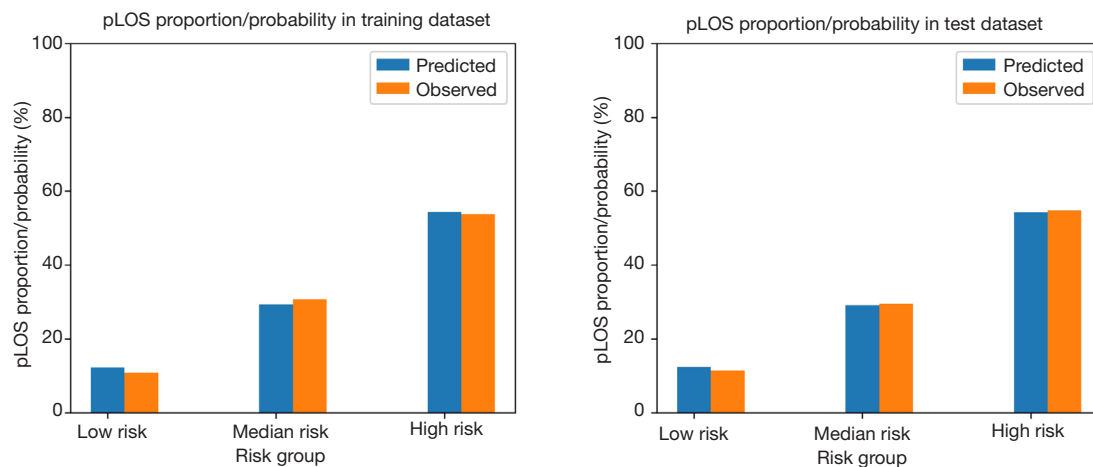


Figure 3 Comparison of the LR generated and observed pLOS probabilities in different pLOS risk groups. LR, logistic regression; pLOS, prolonged length of stay.

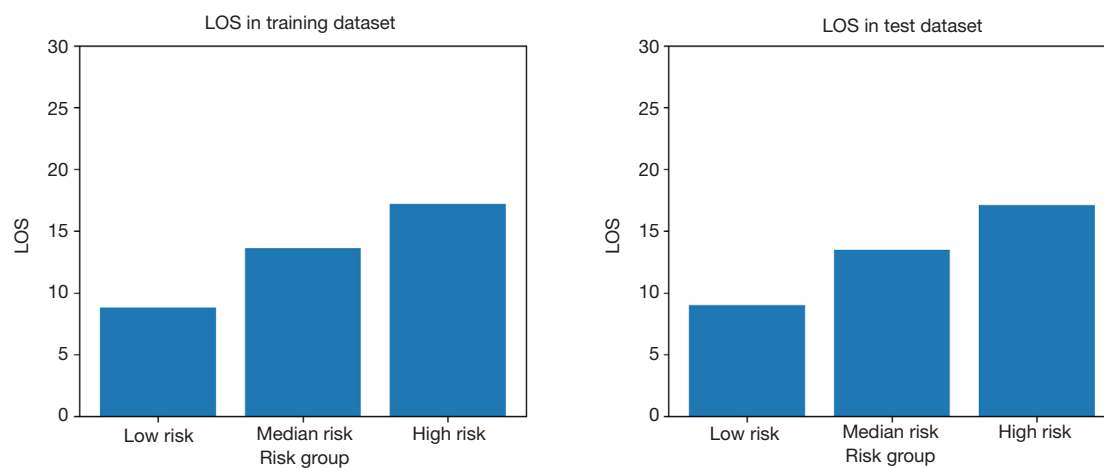


Figure 4 Distribution of averaged LOS across different pLOS risk groups. LOS, length of stay; pLOS prolonged LOS.

scoring system for LOS prediction developed in this study was based on a combination of a machine learning model with superior prediction performance and the traditional LR model, which considers not only model predictability but also model interpretability. Second, a large, multicenter dataset with a nationally representative population of China was used. Third, the scoring system was built using predictive variables that are available at admission time. These factors may help the scoring tool be easily adopted in practice.

However, this study also has limitations. First, other potentially valuable variables, including pathological and laboratory characteristics that could be predictive for LOS,

are not available in our dataset. Second, this scoring system was constructed only based on data from Chinese patients, and it was not externally validated.

In sum, this study developed a scoring tool for stratifying pLOS risk in patients treated with PD by combining machine learning methods and the traditional LR model. The newly developed scoring system can effectively discriminate patients with different pLOS risks. The tool has great potential to aid physicians to risk-stratify patients and perform optimal resource allocation. The performance of our scoring tool was validated using an internal test dataset, and it performed well on our national database. Given that a large, multicenter dataset with a

nationally representative population of China was used in this study, the scoring tool derived from it should have good generalizability to some extent. The developed scoring tool is purely academic thus far; however, we plan to integrate it into the information system of a pilot hospital for prospective validation.

Acknowledgments

The authors thank the Bureau of Medical Administration and Medical Service Supervision, National Health Commission of China for the support of this study.

Funding: This study was supported by Grants from the National Natural Science Foundation of China (Grant No. 81771938, 91846101), from Peking University (Grant No. BMU2018MX020, PKU2017LCX05), the National Key Technology R&D Program of the Ministry of Science and Technology of the People's Republic of China (2016YFC1305400), the University of Michigan Medical School-Peking University Health Science Center Joint Institute for Translational and Clinical Research (BMU2020JI011, BMU20160466), and Chinese Scientific and Technical Innovation Project 2030 (No. 2018AAA0102100).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-20-1006>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Ethics Committee of Peking University First Hospital [No. 2015[928]], and informed consent from patients was not required because it was a secondary use of deidentified patient data and the identifiable personal information was absent.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-

commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Saran R, Robinson B, Abbott KC, et al. US Renal Data System 2019 Annual Data Report: Epidemiology of Kidney Disease in the United States. *Am J Kidney Dis* 2020;75:A6-A7.
2. Hill NR, Fatoba ST, Oke JL, et al. Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. *PLoS One* 2016;11:e0158765.
3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392:1789-858.
4. Liyanage T, Ninomiya T, Jha V, et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. *Lancet* 2015;385:1975-82.
5. Zhang L, Wang F, Wang L, et al. Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet* 2012;379:815-22.
6. Zhang L, Zuo L. Current burden of end-stage kidney disease and its future trend in China. *Clin Nephrol* 2016;86:27-8.
7. Zhang L, Zhao MH, Zuo L, et al. China Kidney Disease Network (CK-NET) 2015 Annual Data Report. *Kidney Int Suppl* (2011) 2019;9:e1-e81.
8. Bruns FJ, Seddon P, Saul M, et al. The cost of caring for end-stage kidney disease patients: an analysis based on hospital financial transaction records. *J Am Soc Nephrol* 1998;9:884-90.
9. Chang KC, Tseng MC, Weng HH, et al. Prediction of Length of Stay of First-Ever Ischemic Stroke. *Stroke* 2002;33:2670-4.
10. Ma X, Si Y, Wang Z, et al. Length of stay prediction for ICU patients using individualized single classification algorithm. *Comput Methods Programs Biomed* 2020;186:105224.
11. Houthoofd R, Ruysinck J, van der Hertten J, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif Intell Med* 2015;63:191-207.

12. Klinzing S, Brandi G, Stehberger PA, et al. The combination of MELD score and ICG liver testing predicts length of stay in the ICU and hospital mortality in liver transplant recipients. *BMC Anesthesiol* 2014;14:103.
13. Meadows K, Gibbens R, Gerrard C, et al. Prediction of Patient Length of Stay on the Intensive Care Unit Following Cardiac Surgery: A Logistic Regression Analysis Based on the Cardiac Operative Mortality Risk Calculator, EuroSCORE. *J Cardiothorac Vasc Anesth* 2018;32:2676-82.
14. Kwok CS, Clark AB, Musgrave SD, et al. The SOAR stroke score predicts hospital length of stay in acute stroke: an external validation study. *Int J Clin Pract* 2015;69:659-65.
15. Daghistani TA, Elshawi R, Sakr S, et al. Predictors of in-hospital length of stay among cardiac patients: A machine learning approach. *Int J Cardiol* 2019;288:140-7.
16. Scherer B, Moser EAS, Brown JW, et al. Vasoactive-ventilation-renal score reliably predicts hospital length of stay after surgery for congenital heart disease. *J Thorac Cardiovasc Surg* 2016;152:1423-9.e1.
17. Burton BN, Abudu B, Danforth DJ, et al. Prediction Model for Extended Hospital Stay Among Medicare Beneficiaries After Percutaneous Coronary Intervention. *J Cardiothorac Vasc Anesth* 2019;33:3035-41.
18. Rana A, Witte ED, Halazun KJ, et al. Liver transplant length of stay (LOS) index: A novel predictive score for hospital length of stay following liver transplantation. *Clin Transplant* 2017;31:e13141.
19. DeLuzio MR, Keshava HB, Wang Z, et al. A model for predicting prolonged length of stay in patients undergoing anatomical lung resection: a National Surgical Quality Improvement Program (NSQIP) database study. *Interact Cardiovasc Thorac Surg* 2016;23:208-15.
20. Kumar VA, Ledezma ML, Idroos ML, et al. Hospitalization Rates in Daily Home Hemodialysis Versus Peritoneal Dialysis Patients in the United States. *Am J Kidney Dis* 2008;52:737-44.
21. Suri RS, Li L, Nesrallah GE. The risk of hospitalization and modality failure with home dialysis. *Kidney Int* 2015;88:360-8.
22. Zhang L, Wang H, Long J, et al. China Kidney Disease Network (CK-NET) 2014 Annual Data Report. *Am J Kidney Dis* 2017;69:A4.
23. Badawi O, Liu X, Hassan E, et al. Evaluation of ICU Risk Models Adapted for Use as Continuous Markers of Severity of Illness Throughout the ICU Stay. *Crit Care Med* 2018;46:361-7.
24. Liang Y, Chang C, Zhu H, et al. Correlation between decrease of CRP and resolution of airway inflammatory response, improvement of health status, and clinical outcomes during severe acute exacerbation of chronic obstructive pulmonary disease. *Intern Emerg Med* 2015;10:685-91.
25. Campbell OM, Cegolon L, Macleod D, et al. Length of Stay After Childbirth in 92 Countries and Associated Factors in 30 Low- and Middle-Income Countries: Compilation of Reported Data and a Cross-sectional Analysis from Nationally Representative Surveys. *PLoS Med* 2016;13:e1001972.
26. Harel Z, Wald R, McArthur E, et al. Rehospitalizations and Emergency Department Visits after Hospital Discharge in Patients Receiving Maintenance Hemodialysis. *J Am Soc Nephrol* 2015;26:3141-50.
27. Peng L, Lei L. A Review of Missing Data Treatment Methods. *Intelligent Information Management Systems and Technologies*; 2005;1:412-9.
28. National Health and Family Planning Commission. National Medical Service and Quality Safety Report. Beijing: People's Medical Publishing House; 2016.
29. Han X, Zheng XN, Wan Y, et al. Random forest can accurately predict the development of end-stage renal disease in immunoglobulin a nephropathy patients. *Ann Transl Med* 2019;7:234.
30. Wang R. Significantly improving the prediction of molecular atomization energies by an ensemble of machine learning algorithms and rescanning input space: A stacked generalization approach. *J Phys Chem C* 2018;122:8868-73.
31. Breiman L. Classification and regression trees. Routledge; 2017.
32. Breiman L. Random Forests. *Machine Learning* 2001;45:5-32.
33. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat* 2001;29:1189-232.
34. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the Performance of Prediction Models A Framework for Traditional and Novel Measures. *Epidemiology* 2010;21:128-38.
35. Van Hoorde K, Van Huffel S, Timmerman D, et al. A spline-based tool to assess and visualize the calibration of multiclass risk predictions. *J Biomed Inform* 2015;54:283-93.
36. Chen T, Li X, Li YX, et al. Prediction and Risk Stratification of Kidney Outcomes in Iga Nephropathy. *Am J Kidney Dis* 2019;74:300-9.
37. US Renal Data System. 2018 USRDS annual data report:

- Epidemiology of kidney disease in the United States.: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD; 2018.
38. Chuang MT, Hu YH, Lo CL. Predicting the prolonged length of stay of general surgery patients: A supervised learning approach. *Int Trans Oper Res* 2016;25:75-90.
 39. Crawford TC, Magruder JT, Grimm JC, et al. A Comprehensive Risk Score to Predict Prolonged Hospital Length of Stay After Heart Transplantation. *Ann Thorac Surg* 2018;105:83-90.
 40. Zhao C, Luo QM, Xia X, et al. Risk score to predict mortality in continuous ambulatory peritoneal dialysis patients. *Eur J Clin Invest* 2014;44:1095-103.
 41. Cao XY, Zhou JH, Cai GY, et al. Predicting One-Year Mortality in Peritoneal Dialysis Patients: An Analysis of the China Peritoneal Dialysis Registry. *Int J Med Sci* 2015;12:354-61.
 42. Zhang X, Yu D, Cai Y, et al. Derivation and Validation of Risk Scores to Predict Cerebrovascular Mortality Among Incident Peritoneal Dialysis Patients. *Kidney Blood Press Res* 2018;43:1141-8.
 43. Tangri N, Ansell D, Naimark D. Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression. *Nephrol Dial Transplant* 2008;23:2972-81.
 44. Maguire PA, Taylor IC, Stout RW. Elderly patients in acute medical wards - factors predicting length of stay in hospital. *Br Med J (Clin Res Ed)* 1986;292:1251-3.
 45. Draper B, Luscombe G. Quantification of factors contributing to length of stay in an acute psychogeriatric ward. *Int J Geriatr Psychiatry* 1998;13:1-7.
 46. Johnson T, McNutt R, Odwazny R, et al. Discrepancy Between Admission and Discharge Diagnoses as a Predictor of Hospital Length of Stay. *J Hosp Med* 2009;4:234-9.
 47. Burke RE, Jones CD, Hosokawa P, et al. Influence of Nonindex Hospital Readmission on Length of Stay and Mortality. *Med Care* 2018;56:85-90.
 48. Flaks-Manov N, Shadmi E, Bitterman H, et al. Readmission To A Different Hospital–Risk Factors And Impact On Length Of Stay. *Value Health* 2014;17:A148.
 49. Kash BA, Baek J, Davis E, et al. Review of successful hospital readmission reduction strategies and the role of health information exchange. *Int J Med Inform* 2017;104:97-104.
 50. Axon RN, Gebregziabher M, Everett CJ, et al. Dual health care system use is associated with higher rates of hospitalization and hospital readmission among veterans with heart failure. *Am Heart J* 2016;174:157-63.
 51. Kshirsagar AV, Hogan SL, Mandelkehr L, et al. Length of stay and costs for hospitalized hemodialysis patients: nephrologists versus internists. *J Am Soc Nephrol* 2000;11:1526-33.
 52. Liu CM, Li CS, Liu CC, et al. Determinants of psychogeriatric inpatient length of stay and direct medical costs: A 6-year longitudinal study using a national database in Taiwan. *Psychiatry Clin Neurosci* 2012;66:423-31.
 53. Xu Z, Xue C, Zhao F, et al. Hospitalization Costs and Length of Stay in Chinese Naval Hospitals Between 2008 and 2016 Based on Influencing Factors: A Longitudinal Comparison. *Mil Med* 2020;185:e282-e289.
 54. Jia H, Li LY, Li WG, et al. Impact of Healthcare-Associated Infections on Length of Stay: A Study in 68 Hospitals in China. *Biomed Res Int* 2019;2019:2590563.
 55. Li Q, Lin ZQ, Masoudi FA, et al. National trends in hospital length of stay for acute myocardial infarction in China. *BMC Cardiovasc Disord* 2015;15:9.

Cite this article as: Wu J, Kong G, Lin Y, Chu H, Yang C, Shi Y, Wang H, Zhang L; on behalf of China Kidney Disease Network Working Group. Development of a scoring tool for predicting prolonged length of hospital stay in peritoneal dialysis patients through data mining. *Ann Transl Med* 2020;8(21):1437. doi: 10.21037/atm-20-1006