BMC Medical Research Methodology

RESEARCH

Open Access

An algorithm to assess importance of predictors in systematic reviews of prediction models: a case study with simulations

Ruohua Yan[1†], Chen Wang[2†], Chao Zhang[1], Xiaohang Liu[1], Dong Zhang[3*] and Xiaoxia Peng[1*]

Abstract

Background How to assess the importance of predictors in systematic reviews (SR) of prediction models remains largely unknown. The commonly used indicators of importance for predictors in individual models include parameter estimates, information entropy, etc., but they cannot be quantitatively synthesized through meta-analysis.

Methods We explored the synthesis method of the importance indicators in a simulation study, which mainly solved the following four methodological issues: (1) whether to synthesize the original values of the importance indicators or the importance ranks; (2) whether to normalize the importance ranks to a same dimension; (3) whether and how to impute the missing values in importance ranks; and (4) whether to weight the importance indicators according to the sample size of the model during synthesis. Then we used an empirical SR to illustrate the feasibility and validity of the synthesis method.

Results According to the simulation experiments, we found that ranking or normalizing the values of the importance indicators had little impact on the synthesis results, while imputation of missing values in the importance ranks had a great impact on the synthesis results due to the incorporation of variable frequency. Moreover, the results of means and weighted means of the importance indicators were similar. In consideration of accuracy and interpretability, synthesis of the normalized importance ranks by weighted mean was recommended. The synthesis method was used in the SR of prediction models for acute kidney injury. The importance assessment results were approved by experienced nephrologists, which further verified the reliability of the synthesis method.

Conclusions An importance assessment of predictors should be included in SR of prediction models, using the weighted mean of importance ranks normalized to a same dimension in different models.

Keywords Importance, Systematic review, Prediction model, Predictor

[†]Ruohua Yan and Chen Wang contributed equally to this work.

*Correspondence:
Dong Zhang
pumczhangdong@126.com
Xiaoxia Peng
pengxiaoxia@bch.com.cn
[1] Center for Clinical Epidemiology and Evidence-based Medicine, Beijing Children's Hospital, Capital Medical University, National Center for Children Health, No.56 Nanlishi Road, Beijing 100045, China
[2] Outpatient Department, Beijing Children's Hospital, Capital Medical University, National Center for Children Health, No.56 Nanlishi Road, Beijing 100045, China
[3] Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, No.167 Beilishi Road, Beijing 100037, China

Background

Systematic review (SR) is one of the most important methods to obtain high-quality evidence from published or unpublished literature and to guide medical decision-making [1]. By qualitatively summarizing the characteristics of the individual studies and/or quantitatively synthesizing the effect size estimations, SR can combine the best available information and provide more convincing conclusions for a specific research question. The methodology for synthesizing results of treatment studies and diagnostic studies has been well developed [2, 3]. However, for prognostic studies, especially those aimed to develop prediction models for a certain clinical outcome, how to conduct SR and summarize the characteristics and performance of the existing models need further clarification.

Several guidance papers have been published to address the methodological issues involved in SR of prediction models. For example, checklists have been formulated for critical appraisal and data extraction in SR of prediction models [4], methods have been proposed to synthesize the model performance [5], and new tools have been developed to assess the risk of bias and applicability of the individual studies [6]. In some cases, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) score can be used to evaluate the reporting quality of the models [7, 8]. However, aside from the issues mentioned above, another important issue for SR of prediction models is to assess the importance of predictors. Traditionally, predictors are listed by the frequency of appearance [8, 9], with the assumption that the more frequent a predictor is included in the individual models, the more important it is. But demographic variables are more likely to be considered, even if sometimes they are not effective predictors of the outcome. On the contrary, some predictors with high impacts are included in only a few models because of the difficulties in measurement or the inaccessibility of measurements in terms of time. It is necessary to propose a more appropriate algorithm to assess the importance of predictors in SR of prediction models.

Meta-analysis of parameter estimates (e.g., odds ratio [OR]) of a predictor from several statistical models may be a good approach to evaluate the predictive value of the predictor to the outcome. However, with the advances in computer technology in recent years, more and more models are developed based on machine learning, and they usually provide a black-box calculator rather than parameter estimates [10, 11]. When conducting SR of machine learning models, researchers can only obtain information about which predictors are included in and how important are they evaluated by information indices such as information entropy [12]. Such information indices cannot be quantitatively combined with traditional statistical indices through meta-analysis to assess the overall importance of predictors.

For clarity, we defined the statistical or information indices that present the importance of predictors in individual models as the "importance indicator", the values of the importance indicators as the "importance value", and the ranks of the importance values as the "importance rank". The commonly used importance indicators include OR for logistic regression models [13, 14], split order for tree-based algorithms [13–15], and feature importance score for gradient boosting classifiers [16, 17], etc. A possible method to combine different importance indicators for predictors in SR of prediction models is to rank the predictors according to their importance in each model at first, and then synthesize the importance ranks rather than the original importance values. But different models often include different numbers of variables, and the dimension of ranks may be widely divergent. For example, ranking 5th in a model that includes only 5 variables or in a model that includes 100 variables are completely incomparable. Moreover, if a variable is not included in a specific model (either not measured or excluded at variable screening), whether to consider it as a missing value or the least important predictor during synthesis is also an issue to be solved.

Therefore, we explored the synthesis method of the importance indicators in SR of prediction models through a simulation study. By comparing the importance assessment results of different synthesis methods with the simulated "real importance", the best performed method would be determined. Then we used an empirical SR to illustrate the application of the synthesis method. Our study aimed to figure out how to assess the importance of predictors in SR of prediction models.

Methods

Simulation experiments

When synthesizing the importance indicators in SR of prediction models, the simplest way is to calculate the mean importance value for each predictor. However, four methodological concerns were involved in the synthesis process. First, as the values of different importance indicators were not combinable, should we rather use the importance ranks? Second, considering that different models included different numbers of variables, should we normalize the importance ranks to a same dimension? Third, for predictors that are not included in a specific model, should we impute the least important ranks for them? Fourth, during the synthesis analysis, should we weight the importance values or ranks according to the sample size of the model? We generated several simulation experiments to answer the above four questions.

Referring to the real-world conditions while not losing generality, we assumed that 30 prediction models (M1-M30) were screened in a SR, each of which included a number of predictors ranging from 5 to 100. All predictors formed a variable pool of 300 variables (V1-V300). The true importance ranks of the predictors were marked by the number, i.e., V1 means the most important predictor and V300 means the least important predictor.

In the simulation study, we firstly sampled an integer from 5 to 100 as the number of predictors in a model (p), and then randomly selected p predictors from the variable pool. We assumed that the importance ranks of these predictors varied around their true ranks, with a deviation of no more than $10\% \times p$. Therefore, we added a random integer between $-10\% \times p$ to $10\% \times p$ to the true rank of each predictor, and re-ranked the addition results to obtain the final importance ranks of all predictors in a specific model. If the addition results were the same for more than one predictor, the ranks among these predictors would be referred to their true ranks. The difference between the final ranks and the true ranks might be slightly exceed to $10\% \times p$ because of the interaction of the ranks, but this did not affect our overall assumption.

The importance values were subsequently simulated between 0 and 1 regardless of the corresponding importance indicators, assuming that all values were in prior scaled to a range of 0 to 1 to ensure their dimensional comparability. Specifically, in a model that included several predictors, the importance value of the predictor that ranked first was specified as 1, the importance value of the predictor that ranked last was specified as 0, and the importance values of other predictors were sampled from a uniform distribution, with more important predictors specifying greater important values. The above process was repeated 30 times, to represent the 30 models in a SR. Since large-scale studies tended to consider more predictors, we assumed that the sample size of each model was proportional to the number of predictors for simplicity.

Scenario settings

The first simulation experiment aimed to decide whether to synthesize the importance values or the importance ranks. Therefore, we compared the importance assessment results of scenario A which synthesized the importance values and scenario B which synthesized the importance ranks. We supposed that synthesis of the importance ranks might be more reasonable, and thus how to normalize and impute the importance ranks were the key problems to be solved in the second and third experiments. Taking scenario B as the reference, scenario C synthesized the importance ranks that were normalized to 1 to the maximum rank (r), with lower values indicating more important predictors. For a model with p variables, if a variable was ranked as k, the normalized rank would be $1 + (k-1)/(p-1) \times (r-1)$. Scenario D synthesized the importance ranks that were imputed for missing by the maximum rank, based on the assumption that predictors not mentioned by a model were considered as the least important. We did not consider the normalization or imputation method for the original importance values, because the range of importance values can vary widely depending on the importance indicator, and thus normalization to 0 to 1 in advance is important for the comparability of the importance values. On the other hand, imputation is not possible for the original importance values. In each scenario, the importance values or ranks were synthesized by mean (scenarios A, B, C, and D) and weighted mean (scenarios A', B', C', and D') respectively, where weight was the sample size of the model.

For example, model M3 included five predictors V49, V133, V215, V245, and V282. Since only ± 0.5 deviation from the true ranks was allowed, the importance ranks of these predictors were the same as the true ranks. The importance values were randomly simulated as 1, 0.68, 0.54, 0.04, and 0, respectively. In scenario A, the above values were extracted for the five predictors, and the importance values for other predictors in the variable pool were regarded as missing. In scenario B, the importance ranks of V49, V133, V215, V245, and V282 were assigned as 1, 2, 3, 4, and 5, respectively. The importance ranks of other predictors were regarded as missing. In scenario C, the importance ranks of V49, V133, V215, V245, and V282 were normalized to 1, 75, 149, 223, and 297, respectively (the maximum rank was 297 because only 297 of the 300 predictors were selected by the 30 models and formed the variable pool). The importance ranks of other predictors were regarded as missing. In scenario D, the importance ranks of V49, V133, V215, V245, and V282 were 1, 2, 3, 4, and 5, respectively. The importance ranks of other predictors were all imputed as 297. A more complicated example of model M1 (with 74 predictors) is provided in Table S1 in Additional file 1.

The accuracy of the synthesis methods was evaluated by the difference of the synthesized importance ranks from the true ranks. The method that obtained the most accurate synthesized ranks would be used to assess the importance of predictors in an empirical SR of prediction models.

Empirical systematic review

In order to illustrate how the synthesis method derived from the simulation study can be applied to empirical research, a SR of prediction models for acute kidney injury (AKI) in general hospital populations was taken as an example. The framing of the question, data extraction, appraisal, and reporting of this review followed the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [4] and the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) statement [3].

The search terms included controlled vocabulary and keywords for (acute kidney injury or acute renal failure) and (prediction model or predict). The following English and Chinese databases were systematically searched from January 1, 2010 to June 6, 2022, including Medline (PubMed), Embase, Web of Science, Cochrane Library, CNKI, Wan Fang, Sinomed, and VIP. Detailed search strategies are shown in Table S2 in Additional file 1. The literature was screened by two investigators based on titles and abstracts, and full articles were reviewed if eligible, with disagreements resolved by a third reviewer. The inclusion and exclusion criteria are shown in Table S3 in Additional file 1. Data extraction was performed by two investigators with disagreements resolved by a third investigator. The data extraction form was determined based on the CHARMS Checklist [4] and previous reviews [8] (shown in Table S4 in Additional file 1), which included information on study design, participants, settings, data collection timeframe, outcome with definitions, sample size, modeling method, variable selection method, indicators for variable importance, and number and type of predictors.

Importance indicators

Importance indicators of predictors reported by individual models included parameter estimates [13, 14, 17–29], importance rank [13–15, 30], importance score [16, 17], information gain [12], chi-square value [31], change of model accuracy [11], and selection frequency in several modeling iterations [14]. Values of importance indicators were extracted from each model, and were normalized to 0–1 using the min–max normalization method, i.e., the minimum and maximum values that the indictor attained in the model were set to 0 and 1, respectively, while the other values were set proportionally between 0 and 1. For models which provided parameter estimates, only predictors that showed statistical significance were extracted. If OR was provided [13, 14, 18–22], β was calculated according to the logarithm of OR, and was regarded as the importance value for further synthesis. Then, the importance values were processed as the importance ranks in each model if necessary. The overall importance was assessed according to the synthesis method recommended by the simulation study.

Evaluation of the algorithm

To verify the importance assessment results of the empirical SR, a questionnaire survey was conducted based on purposive sampling to collect Chinese nephrologists' opinions regarding whether the predictors of AKI were important or not. The variable pool consisting of all predictors mentioned in the SR of prediction models for AKI was stratified by the quartile of the synthesized importance ranks, and six predictors were randomly sampled from each stratum. A total of 24 predictors were included in the questionnaire for the question "Do you think this predictor is important for AKI prediction? -Important/Unimportant/Unknown". The proportion of nephrologists selecting "Important" for each predictor was calculated.

Results

Simulation results

The simulated importance values of predictors in 30 models are provided in Table S5 in Additional file 1. Figure 1 displays the importance heat map of predictors considered by seven or more models, and the importance heat map of all predictors is shown in Figure S1 in Additional file 1. The importance assessment results of scenarios A, B, and C were similar and close to the true rank, indicating that the synthesis results were hardly affected by the ranking of original importance values or the normalization of ranks. On the other hand, the results of scenario D showed substantial differences from other scenarios, suggesting that imputation might have a great impact on the synthesis results. Because predictors missing in a model would be assigned the maximum rank, the fewer a predictor was considered by models, the greater the rank mean and the lower the importance. In other words, the synthesized rank of a predictor would be largely determined by its frequency of appearance in scenario D. Therefore, predictors that appeared in at least seven models in Fig. 1 were almost ranked as Quartile 1 in scenario D, but not necessarily in terms of their true importance. As the results of scenarios with or without quotes were similar, whether weighting the sample size of each model in the synthesis process has little influence on the synthesis results.

Table 1 further compared the importance assessment results in different scenarios. Although synthesis of importance values, importance ranks, and normalized importance ranks could all obtain relatively accurate results, the performance of synthesizing normalized importance ranks by weighted mean (scenario C') was the best. The differences between the synthesized ranks and the true ranks ranged from −33 to 49, with a median (interquartile range) of 0 (−7, 8). More than 2/3 of the

Predictor	Freq	Scenario A	Scenario A'	Scenario B	Scenario B'	Scenario C	Scenario C'	Scenario D	Scenario D'
V9	7	0.96	0.96	3.50	3.68	11.95	12.40	228.52	190.86
V13	8	0.96	0.95	5.06	5.04	22.83	21.03	219.15	184.94
V29	7	0.90	0.90	7.50	8.99	32.63	35.31	229.45	206.80
V35	7	0.87	0.87	8.21	8.98	34.42	35.43	229.62	197.30
V42	9	0.86	0.87	10.67	11.63	49.07	49.76	211.10	176.71
V47	7	0.77	0.78	13.71	15.66	62.01	63.40	230.90	202.26
V48	7	0.88	0.88	6.21	8.01	37.91	42.28	229.15	236.21
V70	7	0.80	0.80	11.21	14.00	58.05	60.08	230.32	215.92
V72	7	0.74	0.74	18.43	21.24	86.84	88.00	232.00	205.66
V83	9	0.76	0.75	16.39	17.79	78.24	79.81	212.82	181.06
V98	9	0.65	0.67	20.33	23.27	110.17	107.03	214.00	190.65
V103	8	0.64	0.63	19.94	22.28	114.58	116.83	223.12	211.82
V117	8	0.61	0.61	24.00	26.33	105.93	108.39	224.20	187.59
V136	7	0.56	0.56	26.93	28.51	125.46	125.36	233.98	204.48
V140	9	0.54	0.56	23.50	27.81	134.52	134.21	214.95	200.02
V145	8	0.54	0.52	26.19	32.09	140.26	143.62	224.78	208.42
V147	7	0.54	0.57	25.21	27.71	148.86	143.24	233.58	220.91
V149	7	0.48	0.46	26.29	31.62	151.50	152.40	233.83	223.47
V153	7	0.50	0.50	30.36	31.01	148.19	146.55	234.78	208.27
V171	11	0.41	0.39	33.64	37.98	174.88	174.36	200.43	170.13
V183	7	0.33	0.33	38.57	40.01	187.86	186.81	236.70	211.07
V191	8	0.32	0.31	34.00	38.29	191.03	193.43	226.87	212.93
V200	7	0.32	0.32	46.21	47.23	200.94	200.76	238.48	203.68
V213	7	0.29	0.31	46.50	50.04	202.39	201.18	238.55	204.34
V216	7	0.28	0.23	37.50	45.59	207.49	217.50	236.45	226.75
V221	8	0.24	0.24	41.50	44.00	219.71	221.44	228.87	209.03
V234	9	0.24	0.20	46.44	55.45	231.81	238.72	221.83	198.79
V237	9	0.21	0.20	43.28	49.38	231.95	234.62	220.88	201.57
V241	7	0.20	0.22	35.93	42.35	220.64	218.38	236.08	229.04
V243	7	0.16	0.16	54.21	57.66	254.57	254.47	240.35	213.96
V254	8	0.18	0.19	45.19	51.69	237.60	235.13	229.85	209.39
V273	8	0.17	0.15	42.50	55.50	248.26	249.76	229.13	220.61
V278	8	0.11	0.13	46.50	54.49	269.11	263.26	230.20	217.24
V282	9	0.02	0.03	39.67	52.49	284.36	282.49	219.80	225.03
V299	9	0.04	0.03	49.83	57.75	284.25	285.64	222.85	209.49

Quartile 1 Quartile 2 Quartile 3 Quartile 4

Fig. 1 Importance heat map of the predictors in the simulated prediction models (Freq ≥ 7)

Scenario A synthesized the importance values (ranged from 0 to 1, with higher values indicating more important predictors). Scenario B synthesized the importance ranks (ranged from 1 to 93, with lower values indicating more important predictors). Scenario C synthesized the normalized importance ranks (ranged from 1 to 297, with lower values indicating more important predictors). Scenario D synthesized the imputed importance ranks (ranged from 1 to 297, with lower values indicating more important predictors). Scenarios without quotes calculated the means of importance values or ranks. Scenarios with quotes calculated the weighted means weighted by the sample size of each model. The true importance ranks were marked by the number of the respective variables (e.g., variable V5 has the true rank 5)

synthesized ranks were within 10 of the true ranks, and the proportion within 30 (10% of the total number of predictors in the variable pool) was 99.33%. By contrast, synthesis of imputed importance ranks showed the least accuracy. Based on the above simulation results, we synthesized the normalized importance ranks by weighted mean in the following empirical SR.

Study characteristics

In the SR of prediction models for AKI in general hospital populations, 8159 potentially eligible records were firstly identified. After screening titles and abstracts, full-text evaluation was carried out for 309 records, and 27 studies were finally included (the PRISMA study flow chart was shown in Fig. 2) [11–37]. The basic characteristics of the included studies are shown in Table 2 and Table S6 in Additional file 2. Of the 27 studies, only one was conducted prospectively. Twenty-two were cohort studies and five were case–control studies (including one nested case–control study). Nine studies were performed in multi-centers. The AKI incidence in general hospital populations ranged from 0.1% to 14.4%, varying with different AKI definitions and classifications.

A total of 41 prediction models were reported by the included studies (Table 3). Among them, 24 models used logistic regression or derived methods for modeling (such as ordinal logistic regression, Bayesian logistic regression, penalized logistic regression, least absolute shrinkage and selection operator [LASSO] regression, etc.), and 17 models used machine learning methods (such as random forest, gradient boosting machine [GBM], lightGBM, XGBoost, Adaboost, recurrent neural network [RNN], multilayer perceptron [MLP], etc.). The number of predictors considered by models ranged from 5 to 2048. Thirty models provided importance information of predictors. For statistical models, importance was

Table 1 Comparison of the importance assessment results in different scenarios[a]

	Scenario A	Scenario A′	Scenario B	Scenario B′	Scenario C	Scenario C′	Scenario D	Scenario D′
Difference from the true rank								
Mean	0	0	0	0	0	0	0	0
SD	14.95	14.58	32.70	29.94	11.80	11.92	113.71	112.71
Min	−42	−39	−140	−140	−31	−33	−286	−271
Q1	−10	−10	−14	−13	−7	−7	−73.5	−82
Median	−1	0	2	3	1	0	−8	−1
Q3	9	9	19.5	16	7	8	74	75
Max	54	50	74.5	68	49	49	276	284
Equal to the true rank (%)	2.69	3.03	1.68	1.68	2.69	5.05	0.67	0.67
Within 5 of the true rank (%)	33.00	31.65	19.53	23.23	37.04	38.72	3.03	4.71
Within 10 of the true rank (%)	55.56	56.23	34.01	36.70	67.68	67.00	8.08	8.42
Within 30 of the true rank (%)	96.30	96.30	73.40	76.09	98.99	99.33	21.21	22.56

Scenario A synthesized the importance values. Scenario B synthesized the importance ranks. Scenario C synthesized the normalized importance ranks. Scenario D synthesized the imputed importance ranks. Scenarios without quotes calculated the means of importance values or ranks. Scenarios with quotes calculated the weighted means weighted by the sample size of each model

[a] The table presents the mean (SD), the median (Q1, Q3), and the min/max value of the differences between the synthesized importance ranks and the true ranks for the 297 predictors, as well as the proportions that the synthesized importance ranks equal to, or within 5/10/30 of the true ranks among the 297 predictors

mainly assessed by β or OR. For machine learning models, importance was generally assessed by importance score or rank.

Importance assessment

Among the 30 models with importance indicators, 255 predictors were considered in total, including 6 demographic characteristics, 69 medications, 85 diagnoses, 70 laboratory indicators, and 25 other variables (Table S7 in Additional file 1). Age (Freq = 21), diabetes mellitus (Freq = 17), serum creatinine (SCr, Freq = 15), blood urea nitrogen (BUN, Freq = 14), and glomerular filtration rate (GFR, Freq = 14) were most frequently used for modeling.

After excluding 161 predictors only reported in one study that might affect the stability of overall ranking, the following 24 predictors were placed in the top quarter, including 5 medications (nonsteroidal anti-inflammatory drugs [NSAIDs], angiotensin-converting enzyme inhibitors [ACEI] or angiotensin receptor blockers [ARB], diuretics, vancomycin, aminoglycosides), 7 diagnoses (shock, heart failure [HF], respiratory failure [RF], chronic kidney disease [CKD], chronic urologic disease [CUD], hypertension [HTN], prior AKI), 9 laboratory indicators (serum potassium [K], serum phosphate [P], SCr, serum magnesium [Mg], urea acid [UA], heart rate [HR], BUN, GFR, respiratory rate [RR]), and 3 other variables (cardiovascular surgery, major surgery, contrast agent). The importance heat map of these predictors is shown in Fig. 3. No demographic characteristics were considered as the most important predictors for AKI in general hospital populations.

We further surveyed the approval of nephrologists of this ranking through a network questionnaire. A total of 66 Chinese nephrologists who engaged in the research of AKI participated in the survey. Of the six predictors ranked as Quartile 1 of importance that included in the questionnaire, five were considered as important by more than 90% experienced nephrologists (Table S8 in Additional file 1). The clinical cognition showed high consistency with the synthesis results.

Discussion

In this study, we used statistical simulations and a case study to explore methods for assessing the importance of predictors in SR of prediction models. We noticed that the weighted mean of importance ranks normalized to a same dimension in different models could provide the most accurate importance assessment results while maintaining interpretability. Illustration of an empirical SR showed that, although some variables were frequently modeled, their predictive values were limited. On the contrary, some variables of high importance were not widely recognized. Therefore, when evaluating the value of predictors, frequency and importance should be comprehensively considered.

SR of prediction models typically only report the frequency of predictors to indicate the degree of attention paid to the predictors [8, 38–40]. However, frequency is not equivalent to importance. As shown in our empirical SR, SCr, BUN, and GFR were important predictors for AKI in general hospital populations, and they were considered in 14 or more models, which suggested that their effects on AKI have been fully understood. By

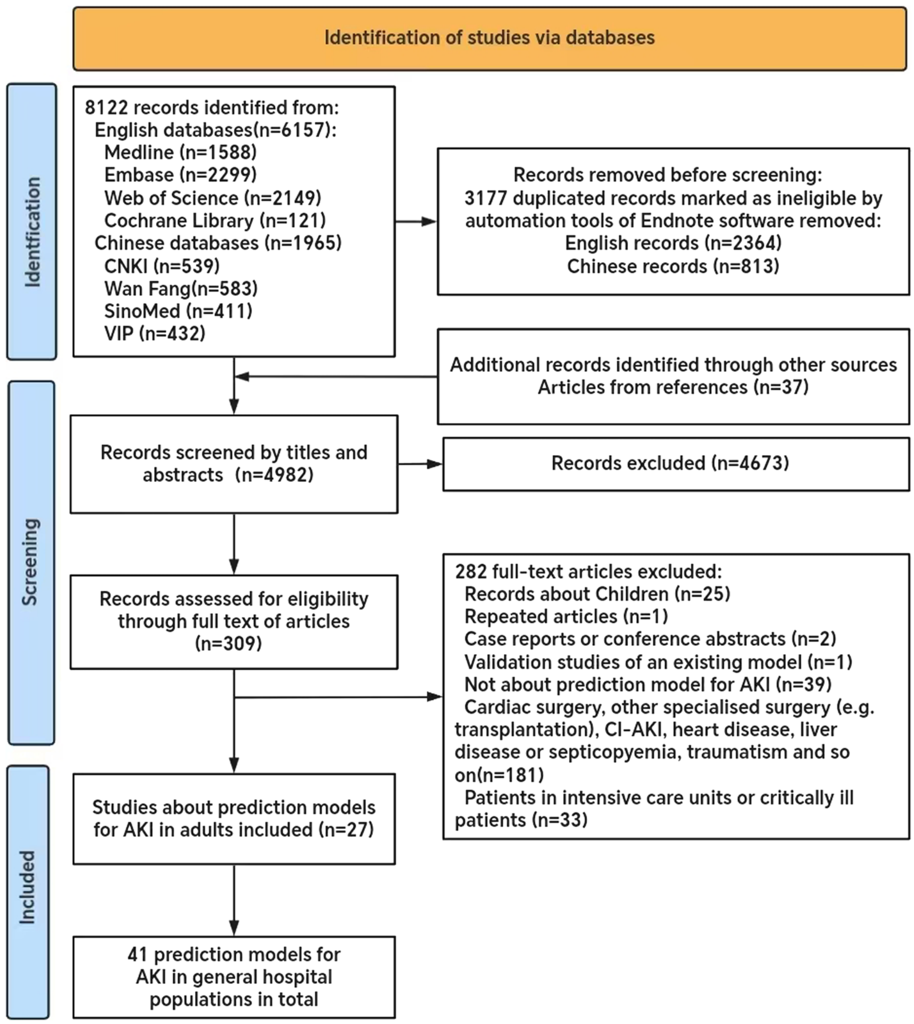

Fig. 2 PRISMA study flow chart

contrast, cardiovascular surgery and shock were the two most important predictors identified by our research, but they were included in only two models published in 2021, indicating that they still need further investigation and consideration in the establishment of future models. Although bicarbonate (HCO_3^-), hemoglobin (Hb), and albumin (ALB) were considered by 8 models among all 30 models, the synthesized importance ranks of these predictors were low, indicating that their actual impacts on AKI might be overestimated. Given the above inconsistencies between frequency and importance ranks, the importance assessment of predictors in SR of prediction models is as important, if not more important, than the frequency assessment.

The indicator of importance for predictors in traditional statistical models is usually their parameter estimates. With the development of machine learning, more and more indicators are used to reflect importance, including split order (for tree-based algorithms) [13–15], feature importance score (for gradient boosting classifiers) [16, 17], and change of model accuracy (for neural networks) [23]. However, few studies have focused on and evaluated these importance indicators. Sanchez-Pinto et al. have assessed the importance of variables by heat maps to compare different variable selection methods for clinical predictive modeling [41]. Till now, no studies have proposed methods to synthesize the importance indicators.

There are four major problems in the process of synthesizing importance indicators: (1) whether to synthesize the original importance values or the importance ranks; (2) whether to normalize the importance ranks to a same dimension; (3) whether and how to impute the missing values in importance ranks; and (4) whether to synthesize importance indicators by means or weighted means. According to our simulation experiments, we

Table 2 Basic characteristics of the included studies

Author, year	Country	Design	Multicenter	Data collection timeframe	Sample size	Number of events	Event rate	Event	AKI definition
Drawz, 2008 [18]	USA	R, case–control	Yes (3)	2003	540	180	/	any AKI	AKIN
Matheny, 2010 [23]	USA	R, cohort	Yes (EHR)	1999–2003	26,107	1352	5.2%	risk or injury	RIFLE
Forni, 2013 [24]	UK	R, cohort	No	/	1314	95	7.2%	any AKI	KDIGO
Cronin, 2015 [13]	USA	R, cohort	Yes (116)	2003–2012	1,620,898	128,457	7.9%	any AKI	KDIGO
Bedford, 2016 [19]	UK	R, cohort	No	2011	27,532	2514	9.1%	any AKI (24 h or 72 h)	KDIGO
Kate, 2016 [20]	USA	R, cohort	No	2013	25,045	1782	7.1%	any AKI	AKIN
Koyner, 2016 [31]	USA	R, cohort	Yes (5)	2008–2013	202,961	17,541	8.6%	any AKI	KDIGO
Cheng, 2017 [32]	USA	R, cohort	No	2007–2016	48,955	4405	9.0%	any AKI	KDIGO
Davis, 2017 [14]	USA	R, cohort	No	2003–2012	170,675	13,142	7.7%	any AKI	KDIGO
He, 2018 [33]	USA	R, cohort	No	2007–2016	76,957	7259	9.4%	any AKI	KDIGO
Koyner, 2018 [16]	USA	R, cohort	No	2008–2016	121,158	17,482	14.4%	any AKI	KDIGO
Mohamadlou, 2018 [34]	USA	R, cohort	Yes (2)	2001–2012, 2008–2017	68,319	1410	2.1%	stage 2 or 3 AKI	KDIGO
Weisenthal, 2018 [17]	USA	R, case–control	No	/	90,013	5618	/	any AKI	KDIGO
Li, 2019 [35]	China	R, case–control	No	2012	1698	840	/	any AKI	KDIGO
Zhang, 2019 [15]	China	R, cohort	No	2012–2016	90,780	7983	8.8%	any AKI	KDIGO
Tomasev, 2019 [11]	UK, USA	R, cohort	Yes (114)	2011–2015	703,782	94,307	13.4%	any AKI	KDIGO
Hsu, 2020 [30]	China (Taiwan)	R, nested case–control	No	2010–2017	224,867	19,448	8.6%	any CA-AKI	KDIGO
Kate, 2020 [12]	USA	R, cohort	Yes (15)	2013–2015	44,691	3786	8.5%	any AKI	AKIN
Safadi, 2020 [21]	USA	R, cohort	No	2013–2014	7048	765	10.9%	any AKI	KDIGO
Su, 2021 [22]	China	R, case–control	No	2014–2015	310	112	/	any AKI	KDIGO
Carpio, 2021 [25]	Spain	R, cohort	No	2011–2017	165,893	95	0.1%	stage 3 AKI	KDIGO
Chen, 2021 [26]	China	R, cohort	Yes (3)	2014–2015	80,091	4108	5.1%	any AKI	KDIGO
Elrewihby, 2021 [27]	Egypt	P, cohort	No	2019	10,243	107	1.0%	any AKI	KDIGO
Kim, 2021 [36]	Korea	R, cohort	Yes (2)	2013–2017	149,108	2791	1.9%	any AKI	KDIGO
Martin-Cleary, 2021 [28]	Spain	R, cohort	No	2015–2016	47,466	2385	5.0%	any AKI	KDIGO
Segarra, 2021 [29]	Spain	R, cohort	No	2017–2018	58,397	2260	3.9%	any AKI	KDIGO
Xu, 2021 [37]	China	R, cohort	No	2019	47,960	2694	5.6%	any AKI	KDIGO

Abbreviation: *R* retrospective, *P* prospective, *EHR* electronic health record, *AKI* acute kidney injury, *CA-AKI* community-acquired acute kidney injury, *AKIN* Acute Kidney Injury Network, *RIFLE* Risk, Injury, Failure, Loss, and End-stage kidney classification, *KDIGO* Kidney Disease: Improving Global Outcomes

Table 3 Technical details of the prediction models

Model ID	Author, year	Modeling method	Variable selection method	Importance indicator	Number of predictors
M1	Drawz, 2008 [18]	LR	Backward	OR	5
M2	Matheny, 2010 [23]	LR	/	β	27
M3	Forni, 2013 [24]	LR	/	β	7
M4	Cronin, 2015 [13]	LR	/	OR	96
M5	Cronin, 2015 [13]	LASSO regression	/	OR	83
M6	Cronin, 2015 [13]	RF	/	Importance rank	95
M7	Bedford, 2016 [19]	Ordinal LR	Backward	OR	26
M8	Bedford, 2016 [19]	Ordinal LR	Backward	OR	28
M9	Kate, 2016 [20]	LR	/	OR	6
M10	Koyner, 2016 [31]	LR	/	Chi-squared value	25
/	Cheng, 2017 [32]	LR	/	/	2048
/	Cheng, 2017 [32]	RF	/	/	2048
/	Cheng, 2017 [32]	AdaBoostM1	/	/	2048
M11	Davis, 2017 [14]	LR	Significant variables in univariate analyses	OR	23
M12	Davis, 2017 [14]	RF	Significant variables in univariate analyses	Importance rank	23
M13	Davis, 2017 [14]	Penalized LR	Significant variables in univariate analyses	L1% selected[a]	23
/	He, 2018 [33]	LR	The Remove Useless node in Weka to remove attributes that do not vary at all or vary too much in the training set with a default threshold of 99%	/	917
/	He, 2018 [33]	RF	The Remove Useless node in Weka to remove attributes that do not vary at all or vary too much in the training set with a default threshold of 99%	/	917
M14	Koyner, 2018 [16]	GBM	/	Importance score	20
/	Mohamadlou, 2018 [34]	XGBoost	/	/	6
M15	Weisenthal, 2018 [17]	GBM	All extremely sparse features (with fewer than 100 non-zero or non-missing elements) were removed	Importance score	21
M16	Weisenthal, 2018 [17]	Penalized LR	All extremely sparse features (with fewer than 100 non-zero or non-missing elements) were removed	Absolute L1 coefficient	18
/	Li, 2019 [35]	LR	/	/	14
/	Li, 2019 [35]	AdaBoost	/	/	14
/	Li, 2019 [35]	MLP	/	/	14
M17	Zhang, 2019 [15]	RF	/	Importance rank	25
M18	Zhang, 2019 [15]	LightGBM	/	Importance rank	25
M19	Tomasev, 2019 [11]	RNN	/	ΔAccuracy[b]	7
M20	Hsu, 2020 [30]	XGBoost	The top 10 important features	Importance rank	10
M21	Hsu, 2020 [30]	LASSO regression	The top 10 important features	Importance rank	10
M22	Kate, 2020 [12]	Continual prediction model	/	Information gain	66
M23	Safadi, 2020 [21]	LR	Backward	OR	8
M24	Su, 2021 [22]	LR	Backward	OR	4
M25	Carpio, 2021 [25]	LR	Stepwise	β	19
M26	Chen, 2021 [26]	LR	Stepwise	β	12
M27	Chen, 2021 [26]	LR	Stepwise	β	10
M28	Elrewihby, 2021 [27]	LR	The top 10 variables with the highest unstandardized coefficients	β	10
/	Kim, 2021 [36]	RNN	/	/	107
M29	Martin-Cleary, 2021 [28]	LR	Backward (using AIC) and Bayesian model averaging	β	23
M30	Segarra, 2021 [29]	LR	Forward	β	22
/	Xu, 2021 [37]	RNN	The top 50 lab investigation variables with the highest detection frequencies	/	50

Abbreviation: *LR* logistic regression, *LASSO* least absolute shrinkage and selection operator, *RF* random forest, *GBM* gradient boosting machine, *MLP* multilayer perceptron, *RNN* recurrent neural network, *AIC* Akaike information criterion, *OR* odds ratio

[a] L1% selected refers to the times a predictor was selected in 200 L-1 penalized logistic regression (L1) modeling iterations;

[b] ΔAccuracy refers to the change of model accuracy when a predictor's value was substantially increased

Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
Cardiovascular surgery															
Shock															
K							216.3			216.5					
P	1.0													242.8	
Combination drugs 2									59.7						
Combination drugs 1	92.5														
SCr	255.0	188.0								1.0				227.8	189.7
Major surgery															
HF															
Mg							206.7								
RF									1.0						
Loop Diuretics		167.1		120.8	88.8	1.0									217.4
CKD			225.9												
Contrast agent															
UA															
HR										124.2				253.1	
CUD															
BUN				252.3	251.7	3.0				88.2	223.0	9.6	1.0		182.6
HTN				174.6	155.4	29.0					224.1	109.4	1.0		
Vancomycin				170.6	153.0	33.0					238.1	229.3	1.0		
GFR				239.4	235.7	43.0	179.4	234.7			1.0	1.0	1.0		254.3
RR			248.2							196.0					
Aminoglycosides		204.0		161.0	143.4	151.0									
prior AKI									206.9						172.0

Variable	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	Synthesis result	Freq	Rank
Cardiovascular surgery											1.0			19.4		8.0	2	1
Shock										100.8					1.0	82.6	2	2
K		143.6	61.2	31.5								255.0		255.0		87.8	7	3
P				1.0	246.4	236.9										104.9	5	4
Combination drugs 2							132.8									106.2	2	5
Combination drugs 1							132.8									132.5	2	6
SCr	233.7	39.9	1.0	255.0	1.0	1.0		1.0	248.4			86.0	1.0			139.2	15	7
Major surgery										141.1				157.9	120.1	140.2	3	8
HF							152.3				207.7	67.3				141.9	3	9
Mg									1.0		128.6					154.8	3	10
RF							236.1			176.0					81.4	156.7	4	11
Loop Diuretics																157.9	5	12
CKD					241.8	219.7		216.6		1.0				134.2	134.5	161.6	7	13
Contrast agent		1.0	236.7							191.2					187.6	161.7	4	14
UA		146.1	137.2											247.0		162.3	3	15
HR							121.2									162.8	3	16
CUD										188.8					124.4	177.1	2	17
BUN	209.8			16.2	226.8	255.0	175.2				223.6					179.1	14	18
HTN							254.8			233.7			213.4		237.8	182.7	10	19
Vancomycin																184.8	6	20
GFR	251.0				208.8	63.1					246.5	1.0				184.9	14	21
RR							157.5									186.1	3	22
Aminoglycosides							251.3									188.2	5	23
prior AKI	216.6						164.9									190.3	4	24

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Not selected
Demographics					
Medications					
Diagnoses					
Labs					
Others					

Fig. 3 Importance heat map of the top 24 predictors for AKI in general hospital populations

Abbreviation: K, serum potassium; P, serum phosphate; Combination drugs 2, used of nonsteroidal anti-inflammatory drugs (NSAIDs), angiotensin-converting enzyme inhibitors (ACEI) and diuretics; Combination drugs 1, Use of NSAIDs, ACEI, angiotensin receptor blockers (ARB), or diuretics; SCr, serum creatinine; HF, heart failure; Mg, serum magnesium; RF, respiratory failure; CKD, chronic kidney disease; UA, uric acid; HR, heart rate; CUD, chronic urologic disease; BUN, blood urea nitrogen; HTN, hypertension; GFR, glomerular filtration rate; RR, respiratory rate; AKI, acute kidney injury

found that ranking or normalizing the importance values had little impact on the results, while imputation had a greater impact due to the incorporation of variable frequency. Moreover, the results of means and weighted means were similar. In consideration of the importance assessment accuracy and the interpretability, synthesis of the normalized importance ranks by weighted mean was recommended. The synthesis method was further used in the SR of prediction models for AKI in general hospital populations, which provided essential clues for future research of variables that had been rarely recognized but might have significant impacts on the outcome. For example, cardiovascular surgery is the second most common cause of AKI in the intensive care setting [42], but its importance was far understated if only in terms of the frequency. Aminoglycosides are common drugs with renal toxicity and associated with the adverse renal prognosis [43, 44], especially AKI [45]. The importance rank could represent the predictive value of aminoglycosides better than the frequency.

To our knowledge, this is the first study to explore how to assess the importance of predictors in SR of prediction models. Unlike previous SR which only used the frequency of predictors to represent importance, we considered more substantive measures and provided a convincing synthesis method for importance assessment. The rationality of the synthesis method was fully explained by both simulation and case analysis. In addition, the synthesized importance ranks of predictors of AKI were recognized by experienced nephrologists, which further manifested the reliability of the synthesis method.

There are some limitations of our study. First, in the simulation section, we assumed that the sample size of each model was proportional to the number of predictors. This is an unrealistic assumption, although there may be a high correlation between the number of predictors and the sample size in real situations. According to our simulation results, whether weighting by sample size has little influence on the synthesis results. Therefore, even if the assumption is relatively strong, it would not substantially affect our simulation findings. On the other hand, a larger sample size generally indicates more reliable research findings, and thus weighting by sample size has its theoretical rationality. Based on the above reasons, it is still recommended to calculate the weighted mean when synthesizing the importance of predictors. Second, in the case study section, considering that there were no unified diagnostic criteria for AKI before 2010, we restricted our search period from 2010 to 2022. Meanwhile, the multi-database search strategy was limited to titles, keywords, and abstracts, which may lead to missing detection. Nevertheless, the empirical SR was intended to illustrate the methods presented in the simulation study, so possible problems in the search process would not affect the research conclusions. Last but not least, the synthesis method proposed by our study was unable to evaluate and treat inter-study heterogeneity like meta-analysis. Hence, it is suggested to evaluate the quality of studies by the Prediction model Risk Of Bias

ASsessment Tool (PROBAST) [6] or the TRIPOD statement [7] and select those with high quality for synthesis.

Despite the above limitations, our study provided a noteworthy methodology for importance assessment, and for the first time demonstrated that importance assessment could not be equated with frequency assessment. We strongly recommend to include an importance assessment of predictors in future SR of prediction models, so as to draw higher-quality evidence.

Conclusions

In conclusion, to assess the importance of predictors in SR of prediction models, we recommend the following steps. First, after extracting the importance values of predictors from different models, a variable pool that contained all predictors should be generated. Then, the importance values of predictors should be ranked in each model and normalized to a range from 1 to the total number of the variable pool to ensure dimensional comparability. Finally, the synthesized rank should be calculated according to the mean of normalized importance ranks weighted by the sample size of the model. Such a method is highly interpretable on the premise of accuracy, and can provide information on the importance of predictors for a certain outcome of concern.

Abbreviations

SR	Systematic review
TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
OR	Odds ratio
AKI	Acute kidney injury
CHARMS	Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies
PRISMA	Preferred reporting items for systematic review and meta-analysis
LASSO	Least absolute shrinkage and selection operator
GBM	Gradient boosting machine
RNN	Recurrent neural network
MLP	Multilayer perceptron
SCr	Serum creatinine
BUN	Blood urea nitrogen
GFR	Glomerular filtration rate
NSAIDs	Nonsteroidal anti-inflammatory drugs
ACEI	Angiotensin-converting enzyme inhibitors
ARB	Angiotensin receptor blockers
HF	Heart failure
RF	Respiratory failure
CKD	Chronic kidney disease
CUD	Chronic urologic disease
HTN	Hypertension
K	Serum potassium
P	Serum phosphate
Mg	Serum magnesium
UA	Urea acid
HR	Heart rate
RR	Respiratory rate
HCO_3^-	Bicarbonate
Hb	Hemoglobin
ALB	Albumin
PROBAST	Prediction model risk of bias assessment tool

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-025-02492-7.

Additional file 1.

Additional file 2.

Acknowledgements
Not applicable.

Authors' contributions
RY and XP designed the study. RY made the simulation. CW, CZ, and XL collected the data of the systematic review. CW conducted the questionnaire survey. RY and CW analyzed the data. RY, CW, and DZ interpreted the data. RY and CW drafted the manuscript. XP and DZ revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Funding
This study was supported by Beijing Hospitals Authority Youth Programme (Grant No. QML20231208); National Natural Science Foundation of China (Grant No. 72404197 and 72174128); and High-Level Public Health Technical Talent Training Plan (Grant No. 2022-2-026). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Data availability
All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 October 2023 Accepted: 3 February 2025
Published online: 14 February 2025

References

1. Luo Y. Enhancing systematic review with machine learning modelling for global health research. Lancet Digit Health. 2023;5:e2-3.
2. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the PRISMA-DTA Group, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018;319:388–96.
3. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
4. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11:e1001744.
5. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356:i6460.
6. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170:51–8.
7. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for

Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73.
8. Hodgson LE, Sarnowski A, Roderick PJ, Dimitrov BD, Venn RM, Forni LG. Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. BMJ Open. 2017;7:e016591.
9. Hu Q, Li SJ, Chen QL, Chen H, Li Q, Wang M. Risk Factors for Acute Kidney Injury in Critically Ill Neonates: A Systematic Review and Meta-Analysis. Front Pediatr. 2021;9:666507.
10. Dong J, Feng T, Thapa-Chhetry B, Cho BG, Shum T, Inwald DP, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care. 2021;25:288.
11. Tomasev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 2019;572:116–9.
12. Kate RJ, Pearce N, Mazumdar D, Nilakantan V. A continual prediction model for inpatient acute kidney injury. Comput Biol Med. 2020;116:103580.
13. Cronin RM, VanHouten JP, Siew ED, Eden SK, Fihn SD, Nielson CD, et al. National Veterans Health Administration inpatient risk stratification models for hospital-acquired acute kidney injury. J Am Med Inform Assoc. 2015;22:1054–71.
14. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc. 2017;24:1052–61.
15. Zhang Q, Guo SW, Liang XL, Chen YH, Wu Y, Fu L. Risk Prediction and Key Influence Factors Analysis of Acute Kidney Injury in Inpatients. Chin J Biomed Eng. 2019;38:702–10.
16. Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. Crit Care Med. 2018;46:1070–7.
17. Weisenthal SJ, Quill C, Farooq S, Kautz H, Zand MS. Predicting acute kidney injury at hospital re-entry using high-dimensional electronic health record data. PLoS ONE. 2018;13:e0204920.
18. Drawz PE, Miller RT, Sehgal AR. Predicting hospital-acquired acute kidney injury–a case-controlled study. Ren Fail. 2008;30:848–55.
19. Bedford M, Stevens P, Coulton S, Billings J, Farr M, Wheeler T, et al. Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study. Southampton: NIHR Journals Library; 2016.
20. Kate RJ, Perez RM, Mazumdar D, Pasupathy KS, Nilakantan V. Prediction and detection models for acute kidney injury in hospitalized older adults. BMC Med Inform Decis Mak. 2016;16:39.
21. Safadi S, Hommos MS, Enders FT, Lieske JC, Kashani KB. Risk Factors for Acute Kidney Injury in Hospitalized Non-Critically Ill Patients: A Population-Based Study. Mayo Clin Proc. 2020;95:459–67.
22. Su YQ, Shen DQ, Wang YM, Xu XL, Teng J, Ding XQ. Establishment of a predictive model for acute kidney injury in emergency inpatients. Chin J Clin Med. 2021;28:562–7.
23. Matheny ME, Miller RA, Ikizler TA, Waitman LR, Denny JC, Schildcrout JS, et al. Development of inpatient risk stratification models of acute kidney injury for use in electronic health records. Med Decis Making. 2010;30:639–50.
24. Forni LG, Dawes T, Sinclair H, Cheek E, Bewick V, Dennis M, et al. Identifying the patient at risk of acute kidney injury: a predictive scoring system for the development of acute kidney injury in acute medical patients. Nephron Clin Pract. 2013;123:143–50.
25. Carpio JD, Marco MP, Martin ML, Ramos N, de la Torre J, Prat J, et al. Development and Validation of a Model to Predict Severe Hospital-Acquired Acute Kidney Injury in Non-Critically Ill Patients. J Clin Med. 2021;10:3959.
26. Chen X, Xu J, Li Y, Xu X, Shen B, Zou Z, et al. Clin Epidemiol. 2021;13:383–96.
27. Elrewihby W, Kasem H, Raghib A, Salah Y, Kora M. Developing and validating a risk score model for prediction of acute kidney injury in non-ICU hospitalized patients. Clin Nephrol. 2021;95:182–8.
28. Martin-Cleary C, Molinero-Casares LM, Ortiz A, Arce-Obieta JM. Development and internal validation of a prediction model for hospital-acquired acute kidney injury. Clin Kidney J. 2019;14:309–16.
29. Segarra A, Del Carpio J, Paz Marco M, Jatem E, Gonzalez J, Chang P, et al. Integrating electronic health data records to develop and validate a predictive model of hospital-acquired acute kidney injury in non-critically ill patients. Clin Kidney J. 2021;14:2524–33.
30. Hsu CN, Liu CL, Tain YL, Kuo CY, Lin YC. Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization From Electronic Health Records: Development and Validation Study. J Med Internet Res. 2020;22:e16903.
31. Koyner JL, Adhikari R, Edelson DP, Churpek MM. Development of a Multicenter Ward-Based AKI Prediction Model. Clin J Am Soc Nephrol. 2016;11:1935–43.
32. Cheng P, Waitman LR, Hu Y, Liu M. Predicting Inpatient Acute Kidney Injury over Different Time Horizons: How Early and Accurate? AMIA Annu Symp Proc. 2018;2017:565–74.
33. He J, Hu Y, Zhang X, Wu L, Waitman LR, Liu M. Multi-perspective predictive modeling for acute kidney injury in general hospital populations using electronic medical records. JAMIA Open. 2018;2:115–22.
34. Mohamadlou H, Lynn-Palevsky A, Barton C, Chettipally U, Shieh L, Calvert J, et al. Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data. Can J Kidney Health Dis. 2018;5:2054358118776326.
35. Li QH. Prediction of acute kidney injury and clinical application optimization based on machine learning. Beijing: Beijing Jiaotong University; 2019.
36. Kim K, Yang H, Yi J, Son HE, Ryu JY, Kim YC, et al. Real-Time Clinical Decision Support Based on Recurrent Neural Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation. J Med Internet Res. 2021;23:e24120.
37. Xu J, Hu Y, Liu H, Mi W, Li G, Guo J, et al. A novel multivariable time series prediction model for acute kidney injury in general hospitalization. Int J Med Inform. 2022;161:104729.
38. Raeisi-Dehkordi H, Kummer S, Francis Raguindin P, Dejanovic G, Eylul Taneri P, Cardona I, et al. Risk Prediction Models of Natural Menopause Onset: A Systematic Review. J Clin Endocrinol Metab. 2022;107:2934–44.
39. Wang J, Yin MJ, Wen HC. Prediction performance of the machine learning model in predicting mortality risk in patients with traumatic brain injuries: a systematic review and meta-analysis. BMC Med Inform Decis Mak. 2023;23:142.
40. Xu Z, Li F, Xin Y, Wang Y, Wang Y. Prognostic risk prediction model for patients with acute exacerbation of chronic obstructive pulmonary disease (AECOPD): a systematic review and meta-analysis. Respir Res. 2024;25:410.
41. Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for clinical predictive modeling. Int J Med Inform. 2018;116:10–7.
42. Wang Y, Bellomo R. Cardiac surgery-associated acute kidney injury: risk factors, pathophysiology and treatment. Nat Rev Nephrol. 2017;13:697–711.
43. Benoit SW, Goldstein SL, Dahale DS, Haslam DB, Nelson A, Truono K, et al. Reduction in Nephrotoxic Antimicrobial Exposure Decreases Associated Acute Kidney Injury in Pediatric Hematopoietic Stem Cell Transplant Patients. Biol Blood Marrow Transplant. 2019;25:1654–8.
44. Ehrmann S, Helms J, Joret A, Martin-Lefevre L, Quenot JP, Herbrecht JE, et al. Nephrotoxic drug burden among 1001 critically ill patients: impact on acute kidney injury. Ann Intensive Care. 2019;9:106.
45. Ryan L, Heed A, Foster J, Valappil M, Schmid ML, Duncan CJA. Acute kidney injury (AKI) associated with intravenous aciclovir in adults: Incidence and risk factors in clinical practice. Int J Infect Dis. 2018;74:97–9.

Publisher's Note