Journal of Clinical Epidemiology 123 (2020) 69−79

Journal of Clinical Epidemiology

REVIEW

A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients

Ype de Jong, MD MSc[a,b,*], Chava L. Ramspek[a], Vera H.W. van der Endt[a], Maarten B. Rookmaaker[c], Peter J. Blankestijn[c], Robin W.M. Vernooij[c,d], Marianne C. Verhaar[c], Willem Jan W. Bos[b,e], Friedo W. Dekker[a], Gurbey Ocak[c], Merel van Diepen[a]

[a]*Department of Clinical Epidemiology, Leiden University Medical Center (LUMC), Leiden, The Netherlands*
[b]*Department of Internal Medicine, Leiden University Medical Center (LUMC), Leiden, The Netherlands*
[c]*Department of Nephrology and Hypertension, University Medical Center Utrecht (UMCU), Utrecht, The Netherlands*
[d]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
[e]*Department of Internal Medicine, Sint Antonius Hospital, Nieuwegein, The Netherlands*

Accepted 19 March 2020; Published online 30 March 2020

Abstract

Objectives: The objective of this study was to systematically review and externally assess the predictive performance of models for ischemic stroke in incident dialysis patients.

Study Design and Setting: Two reviewers systematically searched and selected ischemic stroke models. Risk of bias was assessed with the PROBAST. Predictive performance was evaluated within The Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD), a large prospective multicenter cohort of incident dialysis patients. For discrimination, c-statistics were calculated; calibration was assessed by plotting predicted and observed probabilities for stroke, and calibration-in-the-large.

Results: Seventy-seven prediction models for stroke were identified, of which 15 were validated. Risk of bias was high, with all of these models scoring high risk in one or more domains. In NECOSAD, of the 1,955 patients, 127 (6.5%) suffered an ischemic stroke during the follow-up of 2.5 years. Compared with the original studies, most models performed worse with all models showing poor calibration and discriminative abilities (c-statistics ranging from 0.49 to 0.66). The Framingham showed reasonable calibration; however, with a c-statistic of 0.57 (95% CI 0.50−0.63), the discrimination was poor.

Conclusion: This external validation demonstrates the weak predictive performance of ischemic stroke models in incident dialysis patients. Instead of using these models in this fragile population, either existing models should be updated, or novel models should be developed and validated.

Keywords: External validation; Incident dialysis; Ischemic stroke; Prediction model; Systematic review; Predictive performance; Discrimination; Calibration

1. Introduction

Stroke is a leading cause of morbidity and mortality worldwide. While mortality rates are declining, incidence rates and disease burden have increased over the years [1]. Stroke rates increase with declining renal function and reach a fivefold to tenfold increase in end-stage renal disease patients on dialysis compared with the general population [2−5]. Furthermore, the prognosis in patients on dialysis suffering from a stroke is generally poor: hemodialysis patients have a 3-fold higher risk of death after acute stroke compared with nondialysis populations [5,6].

Identification of those dialysis patients at increased risk for stroke is thus of major importance. Prediction models that assess the risk of stroke, such as the commonly used CHA_2DS_2-$VASC_2$ [7] and $CHADS_2$ [8] have been developed and validated to efficiently allocate individualized anticoagulation therapy. External validation, a step which is essential before implementation of prediction models, shows reasonable predictive performance in independent cohorts with similar characteristics as the development

Conflict of interest: None of the authors has anything to disclose. None of the authors has any form of conflicted interests.

* Corresponding author. Department of Clinical Epidemiology, Leiden University Medical Center, PO Box 9600, Leiden, The Netherlands. Tel.: +31 (0)71 526 4037; fax: +31 (0)71 526 6994.

E-mail address: Y.de_jong@lumc.nl (Y. de Jong,).

https://doi.org/10.1016/j.jclinepi.2020.03.015

What is new?

Key findings

- Risk prediction models for ischemic stroke perform poorly in incident dialysis patients, both in discrimination and calibration.
- The risk of bias of the included 15 prediction models was high, with all models scoring high risk on one or more domains of the PROBAST.

What this adds to what was known?

- Predictive performance of multiple stroke risk prediction models at a clinically relevant moment in a well-defined incident dialysis population.
- Comparison of stroke risk models.

What is the implication and what should change now?

- Contrary to current practice, these models should not be used in these fragile patients in their present form.
- Novel models should be developed and validated, or exiting models updated in dialysis patients.

cohorts of these models. However, dialysis patients were not included in the development of these models, and predictive performance within this high-risk population is largely unknown: only the CHA_2DS_2-$VASC_2$ and $CHADS_2$ have been externally validated. One study reported modest discrimination in a prevalent cohort of dialysis patients [9], and another found good predictive performance in a small cohort of dialysis patients with atrial fibrillation [10]. However, many more prediction models exists and are commonly used in clinical practice, despite the uncertainties regarding predictive performance in this fragile population. In addition, as weighing the benefits of anticoagulation versus the increased risk of bleeding is essential, we have previously conducted an external validation of bleeding risk models which showed poor predictive performance in incident dialysis patients [11]. To further contribute to the ongoing discussion on stroke management in dialysis patients, the aim of the present study is to provide a systematic review and independent external validation of stroke risk models in incident dialysis patients.

2. Methods

2.1. Systematic review

The current review was designed to identify prediction models that assess the risk of ischemic stroke in any population. The PRISMA [12], TRIPOD [13], and CHARMS [14] guidelines were followed to ensure transparent reporting.

2.1.1. Study selection

Studies were included if they met the following predefined selection criteria: 1. The study developed a multivariable prognostic prediction model, with a prediction research question as aim, as opposed to an etiological or methodological goal. 2. The study outcome must be, or must contain the first event of ischemic stroke, and be assessed in a longitudinal design. 3. The study must present at least one measure to assess the predictive performance of the model. Studies in too distinct populations were excluded, such as studies on adverse outcomes of medical interventions and in-hospital stroke. Diagnostic algorithms and studies on genetic associations with ischemic stroke were excluded as well. The search strategy is explained in more detail in the Supplement.

2.1.2. Data extraction and risk of bias

Data extraction and quality assessment was conducted by Y.d.J. Included prediction models were assessed for risk of bias and applicability using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [15,16]. The PROBAST consists of 20 signaling questions for risk of bias within four domains (participant selection, predictors, outcome, and analysis) and three questions for applicability within the first three domains. In addition, we added the domain ‘‘usability’’, which describes whether the model could be used in the present form for risk prediction.

2.2. External validation

2.2.1. Study population and predictor definitions

The Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD) was a prospective, multicenter cohort study in which 38 dialysis centers participated. Between 1997 and 2007, patients older than 18 years without previous renal replacement therapy were included at initiation of dialysis, which was defined as the baseline. Patients were censored when they underwent renal transplantation, died, or withdrew from the study. Although information on death and transplantation of NECOSAD is updated biannually (last update on 04-2019), information on stroke was available until 06-2009, which was used as censoring date. Weight and blood pressure were measured after dialysis. Medication usage and medical history were taken from patients charts. Smoking behavior was recorded as never, ceased, or current smoker. Cholesterol levels were measured in venous blood, and proteinuria was measured in 24 h urine sampling. For the external validation, we used the original predictor definition of the included studies if possible or selected a proxy based on literature and clinical expertise. As the predictive performance is likely influenced by a less-stringent proxy selection, the model was

excluded for validation if more than one predictor was different in NECOSAD compared with the original study.

2.2.2. Outcome

Our outcome ischemic stroke was defined as an ischemic cerebrovascular accident (CVA) requiring hospitalization, or fatal ischemic stroke. This was recorded in the study follow-up forms as CVA, which included both ischemic and hemorrhagic events. To exclude other diagnoses than ischemic stroke, such as hemorrhagic stroke, transient ischemic attack, or thromboembolisms, we developed key word searches in free text entries that were associated with hospitalizations, surgeries, and reasons for dialysis abatement. Furthermore, we used information from a subset of NECOSAD that was chart-reviewed as part of a data quality check.

2.2.3. Statistical analysis

For discrimination, the area under the ROC curve and Harrell's c-statistic for logistic and Cox regression models, respectively, were calculated. For calibration, we calculated the observed risk within the study's original timeframe using Kaplan-Meier survival probabilities for Cox models. Calibration plots were calculated using observed versus predicted probabilities in 10 equal-sized groups, and by fitting a LOWESS curve on the observed and predicted probabilities [17]. For models presenting event rates, the cumulative incidence was approximated (method detailed in the Supplement). For models presenting only beta's without baseline risk, we estimated the constant by refitting the prognostic index (as these were all logistic models) [18]. Stroke prediction models were validated within the original timeframe if applicable, within the maximal follow-up if no timeframe was specified, or pragmatically within 10 years if both the timeframe and maximum follow-up were not specified. Missing data were assumed to be missing at random and were imputed using multiple imputation (detailed in Supplement). We conducted four sensitivity analyses: 1. to further differentiate between ischemic and hemorrhagic stroke, we conducted a chart review as part of a data quality check. Of the 38 participating centers in NECOSAD, data from a representative subset of six dialysis centers (four regional hospitals and two academic hospitals, with a total of 755 patients; 38.6% of whole study sample) were chart-reviewed and model performance was subsequently evaluated in this cohort. 2. As vitamin K antagonists (VKA) may be prescribed for prevention of ischemic stroke in patients with a high risk of ischemic stroke, we performed an analysis only on those patients without VKA. 3. To estimate the effect of competing risk, a "worst-case" analysis in which all patients that died were regarded as ischemic stroke was conducted as well. 4. Stratification on treatment modality, that is, hemodialysis and peritoneal dialysis. RStudio version 1.1.463 and IBM SPSS 25.0 were used.

3. Results

3.1. Systematic search and study selection

The search yielded 901 references, of which 61 studies were included. Cross-reference searching resulted in an additional 16 studies. Of these 77 studies, 15 studies were subsequently validated (Fig. 1). We validated 11 models with the exact same predictor definition as the original models; for the other four models, a proxy was used for one of the predictors in the model: gastrointestinal disease instead of history of bleeding for the GARFIELD-AF [19], a proxy that was used before [11]. For the model of Lip et al. [20], we used the whole follow-up time for the predictor "time within therapeutic range" if the patient used a VKA. Left ventricular hypertrophy diagnosed by ECG

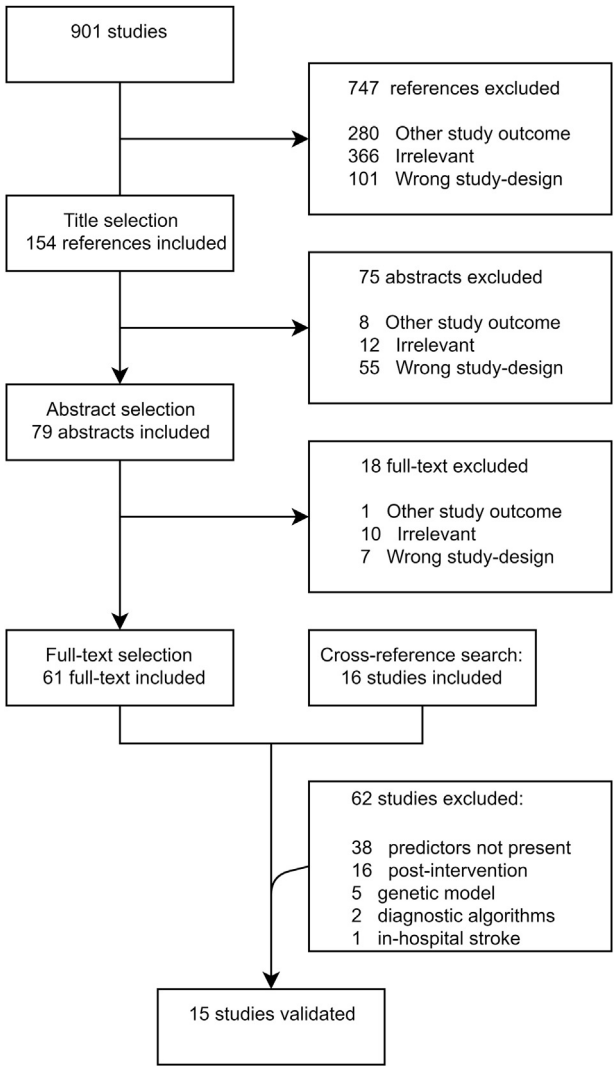

Fig. 1. Flow chart of study selection. The label "irrelevant" was used for studies that did not present a prediction model; "other study outcome" was used for prediction studies that were not on ischemic stroke; "wrong study design" was used for reviews on prediction studies, model updates, external validation, and implementation studies.

Table 1. Overview of the 15 included and externally validated studies

Study	Design	Outcome	Population	Male %	Age
AFI Investigators, 1994 [23] Name: AFI	RCT	Ischemic stroke TIA Systemic embolus	AF	-	69*
Gage, 2001 [8] Name: $CHADS_2$	Cohort (retrospective)	Ischemic stroke TIA	AF	42.0	81*
Wang, 2003 [24] Name: Framingham heart study	Cohort (prospective)	Ischemic stroke Hemorrhagic stroke TIA	AF	53.2	75* (SD 9)
Chambless, 2004 [25] Name: ARIC model	Cohort (prospective)	Ischemic stroke	Atherosclerosis	44.8	-
Zhang, 2005 [26] Name: -	Cohort (prospective)	Ischemic stroke TIA	General	100	45* (SD 7.98)
Diener, 2005 [27] Name: Essen stroke risk score	RCT	Stroke (undefined)	Cardiovascular	-	-
Wu, 2006 [28] Name: -	Cohort (prospective)	Ischemic stroke CHD	General	49.4	46* (SD 6)
Assmann, 2007 [29] Name: PROCAM Risk score	Cohort (prospective)	Stroke (undefined) TIA	General	72.6	45.7* (SD 6.8)
Rietbrock, 2008 [30] Name: Modified-$CHADS_2$	Case-control	Ischemic stroke Hemorrhagic stroke	AF	48.6	-
Lip, 2010 [7] Name: CHA_2DS_2-VASC	Cohort (prospective)	Ischemic stroke TIA Systemic embolus	AF	59.2	66* (SD 14)
Lip, 2013 [20] Name: -	RCT	Stroke (undefined) Major bleeding Systemic embolus	AF	65.0	70* (SD 9)
Singer, 2013 [31] Name: ATRIA	RCT	Ischemic stroke Systemic embolus	AF	-	-
Yatsuya, 2013 [32] Name: -	Cohort (prospective)	Ischemic stroke Hemorrhagic stroke	General	33.9	-
Ferket, 2014 [33] Name: -	Cohort (prospective)	Ischemic stroke	General	42.6	-
Fox, 2017 [19] Name: GARFIELD-AF	Cohort (prospective)	Ischemic stroke Systemic embolus TIA	AF	55.5	71† (63–78)

Abbreviations: RCT, randomized controlled trial; TIA, transient ischemic attack; AF, atrial fibrillation; CHD, coronary heart disease; SD, standard deviation; EPV, events per variable.
For age, the values are *mean, †median, or not stated.

was a predictor in two models [21,22]; in NECOSAD, this was based on the medical history. All validated models are presented in more detail in the Supplement.

3.2. Study characteristics

The characteristics of the included studies are shown in Table 1. Apart from two studies [8,30] with a retrospective or case-control design, all studies used a prospective study design, either as a randomized controlled trial or observational study. Five studies were conducted in the general population [21,26,28,29,32], eight studies in atrial fibrillation cohorts [7,8,19,20,23,24,30,31], one in a cardiovascular risk population [27], and one in patients with atherosclerosis [25]. A prediction timeframe was stated in ten studies and ranged between 1 and 10 years. In total, thirteen models used Cox and two used logistic regression. Models were presented as a point-based risk score in nine studies, a full formula or beta's with intercept in three studies, a calculator in one study, a decision rule in one study, and beta's without a constant in one study. There was substantial risk for overfitting in most models: eight studies did not perform internal validation or used split-sample validation. Events per variable, another indication

n total/*n* events (%)	*n* cand. pred (EPV)	Time Frame	Model method	Internal validation	Discrimination C-statistic	Calibration
3,706/51 (1.38)	15 (3.4)	-	Cox	-	-	-
1,733/94 (5.42)	5 (14.2)	-	Cox	Bootstrapping	0.82	-
868/111 (12.79)	11 (10.1)	5 yr	Cox	Bootstrapping	0.66	Nam and D'Agostino
14,685/434 (2.96)	16 (27.1)	10 yr	Cox	Bootstrapping	-	Calibration curve
3,000/49 (1.63)	8 (6.1)	10 yr	Cox	Split-sample	0.72	Hosmer-Lemeshow Observed vs. expected
19,099/775 (3.95)	32 (24.2)	-	Cox	-	-	-
9,903/371 (3.75)	6 (61.8)	10 yr	Cox	-	0.80	Hosmer-Lemeshow Observed vs. expected
8,130/85 (1.05)	57 (1.5)	10 yr	Cox	Cross-validation	0.78	Observed vs. expected
305,566/19,925 (6.52)	18 (1,106.9)	5 yr	Cox	-	0.72	-
1,084/25 (2.31)	9 (2.8)	1 yr	Logistic	-	0.61	-
2293/94 (4.1)	6 (15.7)	-	Logistic	-	0.73	-
10,927/685 (6.27)	12 (57.1)	-	Cox	Split-sample Bootstrapping	0.73	Nam and D'Agostino
15,672/790 (5.04)	10 (79)	10 yr	Cox	Bootstrapping	0.73	Gronnesby and Borgan Observed vs. expected
27,493/2,559 (9.31)	14 (182.8)	10 yr	Cox	Cross-validation Bootstrapping	0.76	Calibration curve Nam and D'Agostino
38,935/473 (1.21)	32 (11.9)	1 yr	Cox	Cross-validation	0.69	Calibration curve Nam and D'Agostino

of overfitting, ranged between 1.5 and 1,106.9 and was below 20 in eight of the fifteen studies. In most studies, model performance was good, with the original c-statistics (available for 12/15 studies) ranging between 0.61 for the CHA_2DS_2-VASC [7] and 0.82 for the $CHADS_2$ [8]. A measure of calibration was given in nine studies, generally showing good calibration.

3.3. Risk of bias

All included studies showed high risk of bias on at least one domain of the PROBAST tool, with three studies scoring a high risk of bias on three domains (Fig. 2A; details on individual studies are presented in Supplement Table S1). Thirteen studies scored poor on the domain ''outcome'', mainly because of the absence of a time interval, or the use of composite outcomes, such as ''stroke'' (which could include combinations of ischemic stroke, hemorrhagic stroke, or TIA). Of the 15 studies, fourteen scored poor on the domain ''analysis'', which included predictor selection, competing risks, overfitting, and model performance. Only one model accounted for competing risks for ischemic stroke, namely intracranial hemorrhage and death from other causes [33]. The applicability of the models was generally good (Fig. 2B; Supplement Table S1). Seven models were less or not applicable for individual risk prediction: for six models, this was because no predicted probability was given, but an observed event rate

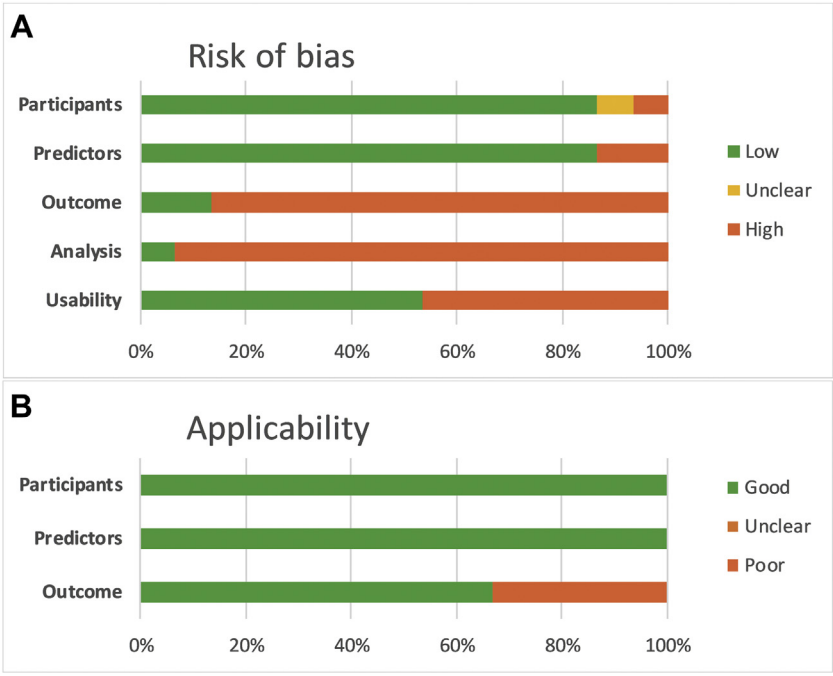

Fig. 2. PROBAST risk of bias summary, percentage of studies with a low, unclear, and high risk of bias (A) and the percentage of studies with a good, unclear, and poor applicability (B) per domain of the PROBAST tool. The domain ''usability'', which was added by the authors of the present study, consisted of one question: ''could the model be used in the present form for risk prediction?'' Abbreviations: PROBAST, Prediction model Risk Of Bias ASsessment Tool.

[7,8,23,27,30,31], for one model [20] beta's were given but no constant was provided.

3.4. Validation cohort

The baseline characteristics of NECOSAD are presented in Table 2. In total, 2,051 patients were enrolled, of which 1,955 (95.3%) were followed after the baseline measurements and subsequently used for the present study. The mean age was 59.98 year (SD 15.1) and 1,216 (62.2%) patients were male. Most patients were on hemodialysis (64.9%), the remainder on peritoneal dialysis. At the end of a median follow-up of 2.5 year, 127 (6.5%) patients suffered an ischemic stroke, 43 of which were fatal. A total of 846 (43.3%) patients died on other causes during follow-up, whereas 571 (29.2%) patients received a transplant.

3.5. Performance of stroke risk scores

While the discrimination of the original studies was moderate to good, it was poor in the validation cohort, with c-statistics ranging from 0.49 (95% CI 0.40–0.58) for the study by Wu [28] to 0.66 (0.59–0.74) for the GARFIELD-AF model [19]. Except for the CHA_2DS_2-VASC [7], all models that presented a c-statistic in the original study were less able to discriminate between low- and high-risk patients in the validation cohort (Table 3). These results were consistent in the sensitivity analyses: recoding all patients who died as ischemic stroke instead of censoring increased discrimination of all models slightly with c-statistics increasing on average with 0.04 (range 0 to 0.08). Stratifying between treatment modality increased discrimination marginally for hemodialysis (c-statistic average increase 0.02, range −0.01 to 0.07), and decreased for peritoneal dialysis (average −0.06, range −0.13 to −0.03). For the other sensitivity analyses, discrimination was consistent with the main analysis: in the chart-reviewed patients, the average difference with the main analysis was 0 (range −0.07 to 0.14); in non-VKA users, this was also 0 (−0.01 to 0.02), detailed in Supplement Table S6 and Fig. S2.

Calibration plots are shown in Fig. 3. Predicted probabilities for six models [7,8,23,27,30,31] were approximated, as only event rates were given in the original studies. The Framingham Heart Study [24] was the only study showing good calibration. For the other studies, calibration was poor both in respect to the actual agreement between observed and predicted probabilities (calibration-in-the-large, Table 3) as well as the calibration curves, which showed over- or underprediction, or a combination of both. The broadness of the range of predicted probabilities differed between studies: 0.05% to 6.61% for Zhang et al. [26] and 0.046% to 92.13% for Chambless et al. [25]. Calibration was comparable in the sensitivity analyses, but as the observed risk was notably higher in sensitivity analysis three in which death was recoded to ischemic stroke, calibration differed more substantially (Supplement Table S7, Figs. S3–S8). Models with a short timeframe did not perform differently, nor did models for which a proxy predictor was used [19–22].

Table 2. Characteristics of validation cohort NECOSAD before multiple imputation

Demographics	Total N = 1,955 (%)	Missing N (%)	No stroke N = 1,828 (%)	Stroke N = 127 (%)
Age (mean, SD)	59.98 (15.1)	3 (0.2)	59.59 (15.2)	65.56 (12.1)
Sex (Male, %)	1,216 (62.2)	4 (0.2)	1,147 (62.7)	69 (54.3)
Vitamin K antagonist use, %	221 (11.3)	210 (10.7)	200 (10.9)	21 (16.8)
Antiplatelet drug use, %	396 (20.3)	210 (10.7)	355 (19.4)	41 (32.3)
Antihypertensive drug use, %	1,439 (73.6)	210 (10.7)	1,335 (73.0)	104 (81.9)
Systolic blood pressure >140 mm Hg, %	1,090 (55.8)	20	1,024 (56.0)	66 (52.0)
Smoking		200 (10.2)		
Current	392 (20.1)		367 (20.1)	25 (19.7)
Ever	792 (40.5)		734 (40.2)	58 (45.7)
Comorbidities, %				
Prior stroke	146 (7.5)	193 (9.9)	126 (6.9)	20 (15.7)
Heart failure	201 (10.3)	193 (9.9)	182 (10.0)	19 (15.0)
Left ventricle hypertrophy	258 (13.2)	193 (9.9)	229 (12.5)	29 (22.8)
Peripheral artery disease	245 (12.5)	193 (9.9)	215 (11.8)	30 (23.6)
Coronary artery disease	193 (9.9)	193 (9.9)	176 (9.6)	17 (13.4)
Malignancy	169 (8.6)	194 (9.9)	161 (8.8)	8 (6.3)
Diabetes	387 (19.8)	193 (9.9)	354 (19.4)	33 (26.0)
Dialysis modality (%)		10 (0.5)		
Hemodialysis	1,268 (64.9)		1,177 (64.4)	91 (71.7)
Peritoneal dialysis	677 (34.6)		642 (35.1)	35 (27.6)
Primary kidney disease, (%)		192 (9.8)		
Diabetic nephropathy	281s (14.4)		254 (13.9)	27 (21.3)
Glomerulonephritis	240 (12.3)		229 (12.5)	11 (8.7)
Vascular	331 (16.9)		291 (15.9)	40 (31.5)
Other	911 (46.6)		863 (47.2)	48 (37.8)

Abbreviations: NECOSAD, The Netherlands Cooperative Study on the Adequacy of Dialysis.

4. Discussion

In this prospective cohort study of 1,955 incident dialysis patients, we externally and independently validated 15 predictive models for ischemic stroke. All studies showed poor predictive performance, both for discrimination and calibration. C-statistics ranged between 0.49 for the model by Wu et al. [28] and 0.66 for the GARFIELD-AF [19], where a c-statistic >0.80 is usually regarded as good. Apart from the Framingham Heart Study [24] which was well calibrated, calibration was also poor. External validation of the $CHADS_2$ and CHA_2DS_2-VASC has only twice been performed in prevalent dialysis patients, yielding comparable results with our study in a large study on 10.999 atrial fibrillation patients on dialysis [9]. However, this study was in prevalent dialysis patients only. The second study showed better predictive performance, but was conducted in a small sample of 141 atrial fibrillation patients on dialysis with only 15 events [10]. Both studies presented discrimination but offered no information on calibration, which, with regard to risk comparison between bleeding and ischemic stroke, could be argued to be of more importance than discrimination.

The poor predictive performance in dialysis patients could have several explanations. First, we demonstrated the high risk of bias. For example, while the rule of thumb of 10 or 20 events per variable is debated [34–37], and more nuanced methods exists [38,39], it is generally accepted that a lower number of events per predictor may result in overfitting and consequently reduce the external validity [15,40]. In our study, more than half of the included studies used less than 20 events per predictor. This observation is not unique to models on ischemic stroke [41], but is demonstrated in other fields as well, for example, for models predicting end-stage renal disease in patients with chronic kidney disease [42]. The TRIPOD guidelines [13] and the PROBAST tool [15,16], both recently published, can possibly aid authors developing new prediction models to avoid commonly encountered methodological errors. Second, differences in case-mix heterogeneity between the original development cohort and the external validation cohort may result in lower discriminative ability even if the fitted regression coefficients are correct [43]. As NECOSAD is likely a more homogeneous cohort than the development cohorts of the validated models, reduced discriminative ability may partly be explained by case-

Table 3. Predictive performance of the 15 included studies, in the original study and in the NECOSAD external validation cohort

			Discrimination		Calibration	
Study	**Model name**	**Model type**	**Original C-statistic**	**Validation C-statistic (95% CI)**	**Observed**	**Predicted**
AFI Investigators, 1994 [23]	AFI	Decision rules	-	0.61 (0.56–0.65)	0.1638	0.4154
Gage, 2001 [8]	$CHADS_2$	Risk score	0.82	0.61 (0.56–0.66)	0.1638	0.3058
Wang, 2003 [24]	Framingham Heart Study	Risk score	0.66	0.57 (0.50–0.63)	0.1379	0.1260
Chambless, 2004 [25]	ARIC	Formula	-	0.61 (0.56–0.66)	0.1638	0.1247
Zhang, 2005 [26]	-	Formula	0.72	0.53 (0.48–0.58)	0.1638	0.0243
Diener, 2005 [27]	Essen stroke risk score	Risk score	-	0.64 (0.59–0.70)	0.1638	0.0428
Wu, 2006 [28]	-	Risk score	0.80	0.49 (0.40–0.58)	0.0926	0.0759
Assmann, 2007 [29]	PROCAM	Risk score	0.78	0.61 (0.56–0.66)	0.1638	0.0515
Rietbrock, 2008 [30]	Modified-CHADS2	Risk score	0.72	0.62 (0.56–0.68)	0.1153	0.1032
Lip, 2010 [7]	CHA_2DS_2-VASC	Risk score	0.61	0.65 (0.57–0.73)	0.0256	0.0187
Lip, 2013 [20]	-	Betas (no constant)	0.73	0.60 (0.54–0.65)	NA[a]	NA[a]
Singer, 2013 [31]	ATRIA	Risk score	0.73	0.63 (0.58–0.69)	0.1215	0.1965
Yatsuya, 2013 [32]	-	Risk score	0.73	0.56 (0.50–0.63)	0.1113	0.0472
Ferket, 2014 [33]	-	Calculator	0.76	0.61 (0.56–0.66)	0.1638	0.0933
Fox, 2017 [19]	GARFIELD-AF	Formula	0.69	0.66 (0.59–0.74)	0.0277	0.0372

Abbreviations: NECOSAD, The Netherlands Cooperative Study on the Adequacy of Dialysis.

[a] The model by Lip et al. was provided without a constant and was recalibrated for the external validation, resulting in values for observed and predicted probabilities that are equal.

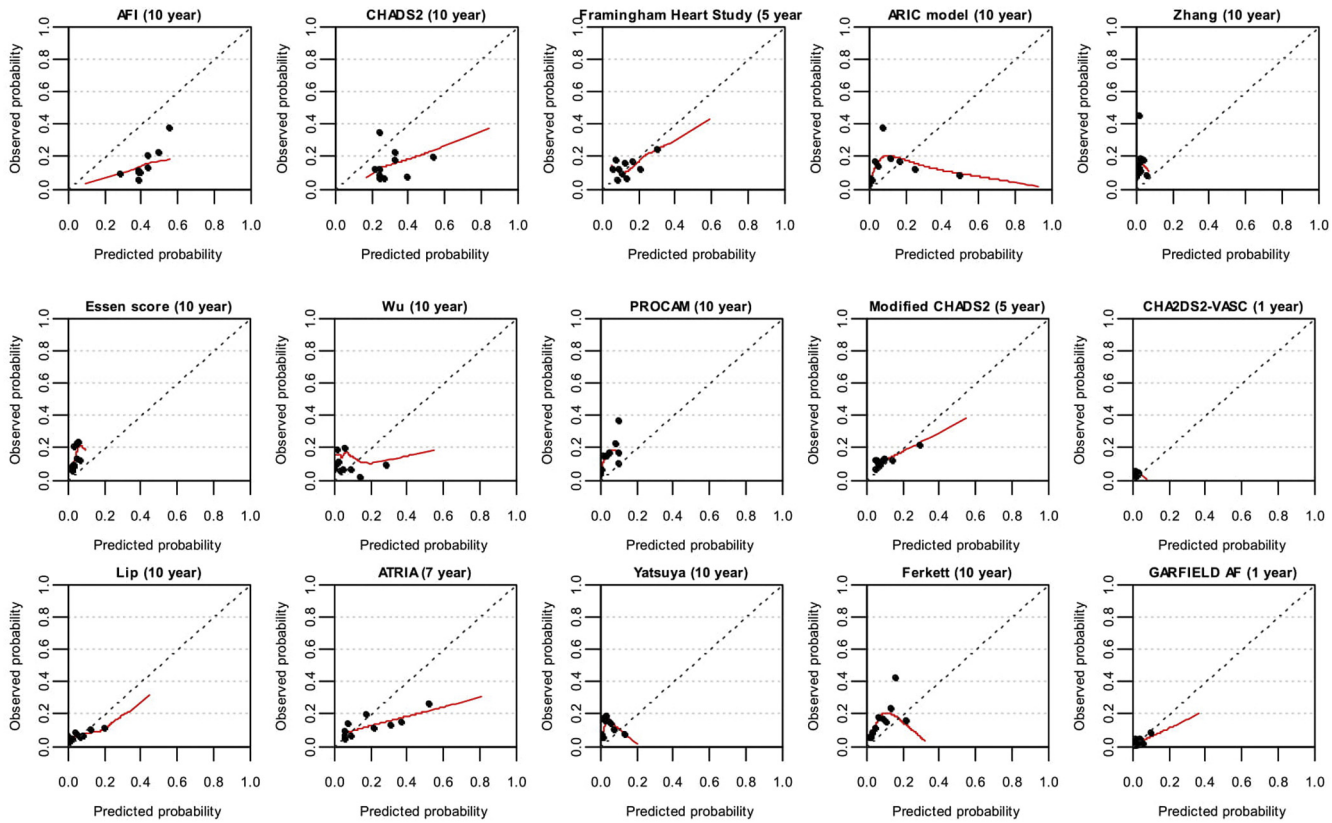

Fig. 3. Calibration plots of the included studies, showing observed and predicted probabilities for ischemic stroke in NECOSAD. Six studies (AFI, $CHADS_2$, Essen score, modified $CHADS_2$, and the CHA_2DS_2-VASC) provided event rates, which were recalculated to cumulative incidences. The model by Lip was presented without a constant, and was subsequently recalibrated. Abbreviations: NECOSAD, The Netherlands Cooperative Study on the Adequacy of Dialysis.

mix difference. Furthermore, predictors that have predictive value in the original development cohort may be of less value in dialysis cohorts due to patient characteristics. Other, more dialysis-specific predictors may be better able to discriminate in this relatively homogeneous population. However, it should be noted that these models are commonly used in incident dialysis patients and thus reflect the actual predictive performance in current clinical use. Third, competing risks (e.g. transplantation, cessation of dialysis therapy, death, or loss to follow-up) may play a major role and greatly impact predictive performance [44]. In our 15 validated studies, only one study accounted for such competing risks [33]. To demonstrate the possible effect of competing risk, we performed a ''worst case'' sensitivity analysis in which all patients who died were regarded as having the outcome as well. While the discrimination of all models showed a modest increase, calibration was off, as all models underpredicted this artificially increased risk.

The main strength of this study is the independent external validation of 15 different ischemic stroke risk models in the same population of incident dialysis patients, allowing comparison of models and increasing the number of validated models in this clinically relevant population substantially. Furthermore, the large and well-defined prospective cohort of 1,955 incident dialysis patients, with a substantial number of events allowed for well-powered analyses at a clinical relevant time point, namely the initiation of dialysis. Our study has several limitations. First, while CVA was recorded in NECOSAD, this included both hemorrhagic and ischemic events. We developed two strategies to overcome this problem: first, we searched for text entries that differentiated between hemorrhagic and ischemic CVA. Second, more than a third of the patients were chart-reviewed and used as data-quality check. Validating the models in this subset resulted in similar predictive performance, but with a higher degree of uncertainty due to the reduced sample size and lower number of events. Another limitation is the lack of information on cardiac arrhythmias, such as atrial fibrillation, reducing the number of possible models to validate. Using prescription of oral anticoagulation as proxy for atrial fibrillation seems reasonable, and was done in our previous study [11]. However, as oral anticoagulation is directly and protectively associated with ischemic stroke, we refrained from validating these models in our cohort. As atrial fibrillation is the main indication for VKA use in dialysis patients, we performed a sensitivity analysis in the non-VKA users, the results of which were similar to the main analysis. We considered to perform the same analysis in VKA users only, but refrained from doing so because of the low number of patients and events in this subgroup. Nevertheless, analyzing the performance in all dialysis patients rather than in a subgroup of patients with atrial fibrillation is clinically relevant because many dialysis patients have an atrial fibrillation event without being diagnosed [45]. Finally, we validated all models for the single-outcome ischemic stroke. As most models predicted a composite outcome, this may have reduced the predictive performance.

Kidney disease and stroke share common risk factors, such as hypertension, aging, diabetes mellitus, and dyslipidemia [46]. The occurrence of atrial fibrillation, also a major risk factor for ischemic stroke in dialysis patients, is more than ten times frequent compared with the general population [47−50], is increasing in prevalence [51] and is often unnoticed [45]. Other risk factors include volumetric changes associated with both end-stage renal disease and dialysis therapy [52−54], and the accelerated atherosclerotic cerebral vascular disease caused in part by the uremic process [55−57]. Apart from the increased risk of stroke in dialysis patients, the risk of hemorrhage is also increased. Until now, no randomized controlled trials on stroke prevention with any form of anticoagulation have been performed in dialysis patients [58]. No high-quality guidelines on stroke prevention in this population exists. Furthermore, we have previously shown that commonly used bleeding risk models have poor predicting performance in incident dialysis patients and should not be used in this population [11]. The poor predictive performance of 15 ischemic stroke risk models of this present study is complementary to these findings. Thus, while the use of these clinical decision aids are appealing as a method to standardize the allocation of care in a seemingly objective manner, clinicians should keep these limitations in mind when applying these models and also consider more dialysis specific variables.

In summary, we have demonstrated the poor predictive performance of ischemic stroke risk models in dialysis patients in addition to our recent external validation of bleeding risk scores. These notions warrant caution for risk stratification in dialysis patients and underline the urgent need for prediction model development specifically targeted at dialysis patients. Alternatively, promising existing models, such as the Framingham Heart Score, which showed good calibration but poor discrimination, could be updated by incorporating dialysis specific variables.

CRediT authorship contribution statement

Ype de Jong: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Visualization, Formal analysis, Writing - original draft, Writing - review & editing. **Chava L. Ramspek:** Methodology, Writing - review & editing. **Vera H.W. van der Endt:** Writing - review & editing. **Maarten B. Rookmaaker:** Writing - review & editing. **Peter J. Blankestijn:** Writing - review & editing. **Robin W.M. Vernooij:** Writing - review & editing. **Marianne C. Verhaar:** Writing - review & editing. **Willem Jan W. Bos:** Writing - review & editing. **Friedo W. Dekker:** Resources, Supervision, Writing - review & editing. **Gurbey Ocak:** Conceptualization, Writing - review & editing. **Merel van Diepen:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Acknowledgments

The Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD) was funded in part by an unrestricted grant from the Dutch Kidney Foundation. The funding source was involved in neither the collection, interpretation, and analysis of the data nor the decision on the writing and submission of this report for publication. This present study was supported by a grant from the Dutch Kidney Foundation (16OKG12).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.03.015.

References

[1] Feigin VL, Norrving B, Mensah GA. Global burden of stroke. Circ Res 2017;120:439−48.
[2] Kawamura M, Fijimoto S, Hisanaga S, Yamamoto Y, Eto T. Incidence, outcome, and risk factors of cerebrovascular events in patients undergoing maintenance hemodialysis. Am J Kidney Dis 1998;31:991−6.
[3] Seliger SL, Gillen DL, Longstreth WT Jr, Kestenbaum B, Stehman-Breen CO. Elevated risk of stroke among patients with end-stage renal disease. Kidney Int 2003;64:603−9.
[4] Toyoda K, Fujii K, Ando T, Kumai Y, Ibayashi S, Iida M. Incidence, etiology, and outcome of stroke in patients on continuous ambulatory peritoneal dialysis. Cerebrovasc Dis 2004;17:98−105.
[5] Power A, Chan K, Singh SK, Taube D, Duncan N. Appraising stroke risk in maintenance hemodialysis patients: a large single-center cohort study. Am J Kidney Dis 2012;59:249−57.
[6] Ovbiagele B. Chronic kidney disease and risk of death during hospitalization for stroke. J Neurol Sci 2011;301:46−50.
[7] Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 2010;137:263−72.
[8] Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA 2001;285:2864−70.
[9] Chao TF, Liu CJ, Wang KL, Lin YJ, Chang SL, Lo LW, et al. Incidence and prediction of ischemic stroke among atrial fibrillation patients with end-stage renal disease requiring dialysis. Heart Rhythm 2014;11:1752−9.
[10] Wang TK, Sathananthan J, Marshall M, Kerr A, Hood C. Relationships between anticoagulation, risk scores and adverse outcomes in dialysis patients with atrial fibrillation. Heart Lung Circ 2016;25:243−9.
[11] Ocak G, Ramspek C, Rookmaaker MB, Blankestijn PJ, Verhaar MC, Bos WJW, et al. Performance of bleeding risk scores in dialysis patients. Nephrol Dial Transplant 2019;34:1223−31.
[12] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700.
[13] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.
[14] 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.
[15] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1−33.
[16] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51−8.
[17] Cleveland WS. LOWESS: a program for smoothing scatterplots by robust locally weighted regression. Am Stat 1981;35:54.
[18] Janssen KJ, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KG. A simple method to adjust clinical prediction models to local circumstances. Can J Anaesth 2009;56:194−201.
[19] Fox KAA, Lucas JE, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, et al. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. BMJ Open 2017;7:e017157.
[20] Lip GYH, Lane DA, Buller H, Apostolakis S. Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: the AMADEUS Study. Chest 2013;144:1839−47.
[21] D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. Stroke 1994;25:40−3.
[22] Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. Stroke 1991;22:312−8.
[23] Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994;154:1449−57.
[24] Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. JAMA 2003;290:1049−56.
[25] Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the atherosclerosis risk in communities study. Am J Epidemiol 2004;160:259−69.
[26] Zhang XF, Attia J, D'Este C, Yu XH, Wu XG. A risk score predicted coronary heart disease and stroke in a Chinese cohort. J Clin Epidemiol 2005;58:951−8.
[27] Diener HC, Ringleb PA, Savi P. Clopidogrel for the secondary prevention of stroke. Expert Opin Pharmacother 2005;6:755−64.
[28] Wu Y, Liu X, Li X, Li Y, Zhao L, Chen Z, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation 2006;114:2217−25.
[29] Assmann G, Schulte H, Cullen P, Seedorf U. Assessing risk of myocardial infarction and stroke: new data from the prospective cardiovascular munster (PROCAM) study. Eur J Clin Invest 2007;37:925−32.
[30] Rietbrock S, Heeley E, Plumb J, van Staa T. Chronic atrial fibrillation: incidence, prevalence, and prediction of stroke using the Congestive heart failure, Hypertension, Age >75, Diabetes mellitus, and prior Stroke or transient ischemic attack (CHADS2) risk stratification scheme. Am Heart J 2008;156:57−64.
[31] Singer DE, Chang Y, Borowsky LH, Fang MC, Pomernacki NK, Udaltsova N, et al. A new risk scheme to predict ischemic stroke and other thromboembolism in atrial fibrillation: the ATRIA study stroke risk score. J Am Heart Assoc 2013;2:e000250.
[32] Yatsuya H, Iso H, Yamagishi K, Kokubo Y, Saito I, Suzuki K, et al. Development of a point-based prediction model for the incidence of total stroke: Japan public health center study. Stroke 2013;44:1295−302.

[33] Ferket BS, van Kempen BJ, Wieberdink RG, Steyerberg EW, Koudstaal PJ, Hofman A, et al. Separate prediction of intracerebral hemorrhage and ischemic stroke. Neurology 2014;82:1804−12.
[34] Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503−10.
[35] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007;165:710−8.
[36] van Smeden M, de Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol 2016;16:163.
[37] Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger TV. Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. J Clin Epidemiol 2011;64:993−1000.
[38] van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. Stat Methods Med Res 2019;28:2455−74.
[39] Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38:1276−96.
[40] Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res 2017;26:796−808.
[41] Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:1−12.
[42] Ramspek CL, de Jong Y, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. Nephrol Dial Transplant 2019;(gfz018). https://doi.org/10.1093/ndt/gfz018.
[43] Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol 2010;172:971−80.
[44] Shih CJ, Ou SM, Chao PW, Kuo SC, Lee YJ, Yang CY, et al. Risks of death and stroke in patients undergoing hemodialysis with new-onset atrial fibrillation: a competing-risk analysis of a nationwide cohort. Circulation 2016;133:265−72.
[45] Buiten MS, de Bie MK, Rotmans JI, Gabreels BA, van Dorp W, Wolterbeek R, et al. The dialysis procedure as a trigger for atrial fibrillation: new insights in the development of atrial fibrillation in dialysis patients. Heart 2014;100:685−90.
[46] Toyoda K, Ninomiya T. Stroke and cerebrovascular diseases in patients with chronic kidney disease. Lancet Neurol 2014;13:823−33.
[47] Konigsbrugge O, Posch F, Antlanger M, Kovarik J, Klauser-Braun R, Kletzmayr J, et al. Prevalence of atrial fibrillation and antithrombotic therapy in hemodialysis patients: cross-sectional results of the Vienna InVestigation of AtriaL fibrillation and thromboembolism in patients on HemoDIalysis (VIVALDI). PLoS One 2017;12:e0169400.
[48] Genovesi S, Pogliani D, Faini A, Valsecchi MG, Riva A, Stefani F, et al. Prevalence of atrial fibrillation and associated factors in a population of long-term hemodialysis patients. Am J Kidney Dis 2005;46:897−902.
[49] Vazquez E, Sanchez-Perales C, Garcia-Garcia F, Castellano P, Garcia-Cortes MJ, Liebana A, et al. Atrial fibrillation in incident dialysis patients. Kidney Int 2009;76:324−30.
[50] Winkelmayer WC, Patrick AR, Liu J, Brookhart MA, Setoguchi S. The increasing prevalence of atrial fibrillation among hemodialysis patients. J Am Soc Nephrol 2011;22:349−57.
[51] Molnar AO, Sood MM. Predicting in a predicament: stroke and hemorrhage risk prediction in dialysis patients with atrial fibrillation. Semin Dial 2018;31:37−47.
[52] Gottlieb D, Mildworf B, Rubinger D, Melamed E. The regional cerebral blood Flow in patients under chronic hemodyalytic treatment. J Cereb Blood Flow Metab 1987;7:659−61.
[53] Postiglione A, Faccenda F, Gallotta G, Rubba P, Federico S. Changes in middle cerebral artery blood velocity in uremic patients after hemodialysis. Stroke 1991;22:1508−11.
[54] Hata R, Matsumoto M, Handa N, Terakawa H, Sugitani Y, Kamada T. Effects of hemodialysis on cerebral circulation evaluated by transcranial Doppler ultrasonography. Stroke 1994;25:408−12.
[55] Kennedy R, Case C, Fathi R, Johnson D, Isbel N, Marwick TH. Does renal failure cause an atherosclerotic milieu in patients with end-stage renal disease? Am J Med 2001;110:198−204.
[56] Amann K, Neususs R, Ritz E, Irzyniec T, Wiest G, Mall G. Changes of vascular architecture independent of blood pressure in experimental uremia. Am J Hypertens 1995;8:409−17.
[57] Kawagishi T, Nishizawa Y, Konishi T, Kawasaki K, Emoto M, Shoji T, et al. High-resolution B-mode ultrasonography in evaluation of atherosclerosis in uremia. Kidney Int 1995;48:820−6.
[58] Chao TF, Chen SA. Risk of ischemic stroke and stroke prevention in patients with atrial fibrillation and renal dysfunction. J Atr Fibrillation 2015;8:1196.