

RESEARCH

Open Access



# External validation of six clinical models for prediction of chronic kidney disease in a German population

Susanne Stolpe<sup>1\*</sup>, Bernd Kowall<sup>1</sup>, Denise Zwanziger<sup>1</sup>, Mirjam Frank<sup>1</sup>, Karl-Heinz Jöckel<sup>1</sup>, Raimund Erbel<sup>1</sup> and Andreas Stang<sup>1,2</sup>

## Abstract

**Background:** Chronic kidney disease (CKD) is responsible for large personal health and societal burdens. Screening populations at higher risk for CKD is effective to initiate earlier treatment and decelerate disease progress. We externally validated clinical prediction models for unknown CKD that might be used in population screening.

**Methods:** We validated six risk models for prediction of CKD using only non-invasive parameters. Validation data came from 4,185 participants of the German Heinz-Nixdorf-Recall study (HNR), drawn in 2000 from a general population aged 45–75 years. We estimated discrimination and calibration using the full model information, and calculated the diagnostic properties applying the published scoring algorithms of the models using various thresholds for the sum of scores.

**Results:** The risk models used four to nine parameters. Age and hypertension were included in all models. Five out of six c-values ranged from 0.71 to 0.73, indicating fair discrimination. Positive predictive values ranged from 15 to 19%, negative predictive values were > 93% using score thresholds that resulted in values for sensitivity and specificity above 60%.

**Conclusions:** Most of the selected CKD prediction models show fair discrimination in a German general population. The estimated diagnostic properties indicate that the models are suitable for identifying persons at higher risk for unknown CKD without invasive procedures.

**Keywords:** Chronic Kidney Disease, Prediction Model Screening, External Validation Sensitivity, Specificity ROC Curve

## Background

Prevalence of chronic kidney disease (CKD), defined by a chronically impaired renal function, is growing worldwide and a challenge for public health. In Germany, prevalence of a decreased renal function is up to 11.5% [1]. Patients with CKD are at higher risk of cardiovascular comorbidities, hospitalization, kidney failure and premature death

[2]. As CKD cannot be cured, treatment aims at monitoring CKD risk factors – especially hypertension and blood glucose – to decelerate its progression and to prevent the incidence of secondary diseases [3]. Early diagnosis of a prevalent CKD can support these efforts. Despite the high prevalence and relevance for public health, public and patient awareness for CKD is low. The main reason for the low awareness is that CKD remains asymptomatic until reaching more serious stages. Moreover, a declining renal function may be a physiological sign of older age, which

\*Correspondence: Susanne.stolpe@uk-essen.de

<sup>1</sup> Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Hufelandstr. 55, 45147 Essen, Germany  
Full list of author information is available at the end of the article



often hinders physicians to designate an impaired renal function in older age as CKD [4, 5].

In Germany, fewer than 50% of patients with CKD—even with hypertension or cardiovascular disease—knew about their condition [6, 7]. In the USA, CKD unawareness is even more prevalent [8]. The high unawareness for CKD is astonishing, as regular monitoring of renal function should be mandatory for patients with hypertension, diabetes or of older age [3–9–11]. However, in general practice, even in patients with diabetes or hypertension, renal function, blood glucose or blood pressure are not regularly monitored [12, 13] and risk factors are inadequately controlled [14, 15].

According to Tonelli et al. most of the principles for population screening—formulated in a Delphi process among experts—fit to screening for CKD [16]. Screening for CKD has already been shown to be successful [17, 18] and cost-effective in detecting unknown cases in population subgroups at higher risk for CKD, such as in persons with diabetes, hypertension or aged 60 years and more [19–21]. In a simulation study, it was shown that CKD prediction scores can be cost-effectively used to initially identify people at higher risk for incident CKD, and to screen these subsequently for CKD by testing for

albuminuria [22]. Prediction models suitable for identifying people at higher risk for CKD should be easy to apply, preferentially using non-invasive parameters only. Many CKD risk models of different complexity have already been developed—regarding the prediction of incidence, prevalence and progression to kidney failure [23]. However, missing external validation might frequently hinder the implementation into practice [24, 25].

The aim of this study was, to externally validate prediction models, that estimate the probability of a prevalent CKD using non-invasive parameters only in a German general population.

## Material and methods

### CKD prediction models

Starting with a review on CKD prediction models from 2012 [23], we searched the literature for further models that comprise only clinical information for estimating the risk of prevalent CKD. We identified five models that meet these criteria [26–30]. Among these, Bang reported two versions of a model developed in the same population: SCORED and modified SCORED (Table 1). For these models, a scored version as self-completing questionnaire is published [27]. The intercepts for the

**Table 1** Identified prediction models for CKD: included parameters and their coefficients

	Bang SCORED	Bang „modified SCORED“	Kearns	Kshirsagar	Kwon	Thakkinian
No of parameters	9	7	5	8	7	4
Intercept	-5.40 <sup>a</sup>	-5.38 <sup>a</sup>	-3.63	-3.30	-6.53 <sup>a</sup>	-2.8 <sup>b</sup>
Age (yrs)	1.55 [50–59] 2.31 [60–69] 3.23 [≥ 70]	1.55 [50–59] 2.29 [60–69] 3.29 [≥ 70]	1.075 [per 10 yrs <sup>c</sup> ] -0.01 [age <sup>2</sup> /10 yrs <sup>c</sup> ] 0.104 [age < 50]	0.63 [50–59] 1.33 [60–69] 1.46 [≥ 70]	1.16 [50–59] 1.91 [60–69] 2.71 [≥ 70]	0.6 [50–59] 1.4 [60–69] 2.1 [≥ 70]
Sex – Female	0.29	0.34	0.73	0.13	0.40	
Anemia	0.93			0.48	0.94	
Hypertension	0.45	0.47	0.74 + age < 50: 0.56	0.55	0.48	0.80
Diabetes	0.44	0.47		0.33	0.73	0.90
Ischemic heart disease or stroke (Hx)	0.59	0.67		0.26	0.60	
Heart failure (hx)	0.45	0.51	0.86 CHF + age < 50: 0.29	0.50		
Ischemic heart disease (Hx)			0.51 + age < 50: 0.13			
Peripheral vascular disease (Hx)	0.74			0.41		
Proteinuria	0.83	0.88			0.48	
Kidney stones (Hx)						1

Intercept: the baseline prevalence of CKD in the cohort if all other predictors of a model are not present, meaning their coefficients are zero

Coefficient: The impact with which a predictor effects the estimation of CKD probability. The larger the coefficient, the stronger the impact; e.g. anemia (coefficient 0.93) has a higher impact in the SCORED model on the estimation of CKD probability than diabetes (0.44)

<sup>a</sup> Intercept according to personal information by H. Bang

<sup>b</sup> Intercept estimated using the prevalence of CKD in the validation population

<sup>c</sup> age per 10 yrs = (age - 46.72) / 10yrs

SCORED models and the models by Kwon and Thakkinstian had not been published. For the SCORED and the Kwon model, we were able to get the missing information about the intercept from the authors. For the Thakkinstian model we used the intercept published in a validation study [25]. As the Kearns model estimated an unrealistic high risk for CKD in our validation population, we contacted the author and learned that the age-parameter should have had been centered (by subtracting the value 46.72) prior to the division of age by 10). This had not been reported in the manuscript. Further, no cutoffs or scoring rule for the Kearns-Model had been described, although sensitivity and specificity information had been estimated in the paper. As we could get no information on the applied rules for scoring the model parameters, we were not able to estimate the diagnostic properties of the Kearns model in our validation cohort.

The number of parameters used in the CKD prediction models ranged from four (Thakkinstian) to nine (SCORED) (Table 1). Age and hypertension were the only predictors used in all models. The Kearns model relies heavily on age, using age as interaction term with other parameters as well. History of kidney stones and history of ischemic heart disease were used only once.

#### Validation population

We used the German Heinz Nixdorf Recall Study (HNR), a population based cohort, [31] for external validation (Table 2). Baseline data from 4,814 participants drawn from the general population aged 45–75 years in 2000 were available. We included all participants with a valid measurement of serum creatinine ( $N=4,789$ ).

#### Measurement of variables

In HNR, all laboratory data had been analyzed centrally in the laboratory of the university hospital of Essen. Serum creatinine (according to Jaffé) was determined on a Siemens Healthcare Diagnostics ADVIA Chemistry. Serum creatinine was not standardized to isotope dilution mass spectrometry. Hypertension was defined as either a blood pressure of at least 140 mmHg systolic or at least 90 mmHg diastolic or taking antihypertensive medication. Blood pressure cut-offs were selected according to the cut-offs used to define hypertension in the validated risk models. Diabetes was defined according to the respective definitions used in the risk models: either self-reported prevalent diabetes [27, 28] or using a combination of known diabetes or taking antidiabetic drugs [29, 30]. Albuminuria was defined as albumin/creatinine ratio (ACR)  $\geq 30$  mg/dl. In all models except the Kwon model, anemia was coded if hemoglobin levels were  $< 12$  g/dl. In the Kwon model, the threshold for hemoglobin was  $< 12$  g/dl for women and  $< 13$  g/dl for

men. Peripheral vascular artery disease was defined according to clinical information.

#### Definition of chronic kidney disease in development and validation populations

In all validated risk models, CKD was defined as an estimated glomerular filtration rate (eGFR)  $< 60$  ml/min/1.73m<sup>2</sup> calculated by the Modification of Diet in Renal Disease (MDRD) equation.

All models had been developed to estimate the risk of a prevalent CKD stage 3a or more. Therefore, we defined CKD as eGFR  $< 60$  ml/min/1.73m<sup>2</sup> accordingly. We used the CKD-Epi equation for calculating eGFR, as recommended by the Kidney Disease Improving Global Outcomes (KDIGO) [32]. In Germany, the CKD-Epi equation is widely used to report eGFR with laboratory results. As sensitivity analysis, we calculated the eGFR with the MDRD and the new Full-Age-Spectrum (FAS) equation (equations listed in Supplement S2 + S3) [33]. We used the respective creatinine-based equations, because cystatin c measurements are not widely available in general practice.

#### Handling of missing values

We did a complete case analysis regarding all predictors used in the identified models, leaving 4,185 participants in HNR. The Thakkinstian score was validated in a subsample of HNR with information on the parameter 'history of kidney stones' that is used in this model only ( $N=3,433$ ).

#### Statistical analysis

The models' discrimination was estimated by the c-statistic (Area Under the ROC Curve, AUC) [34] and the Tjur coefficient [35]. The Tjur coefficient is the difference between the mean predicted probability in cases and in non-cases. The higher this difference the better the discriminative ability of a score. Calibration was assessed graphically.

As measures for overall performance, we estimated the mean average prediction error (MAPE) and the Scaled Brier Score [25, 35]. MAPE averages the deviations between the prediction (ranging between 0 and 1) and the respective true value of zero or 1. The smaller the MAPE, the better the prediction. The Scaled Brier Score is calculated by the squared difference between the prediction and the true value of outcome (=Brier Score) divided by the product of the mean prediction value and 1- mean prediction value [36]. It ranges from 0 to 1 representing 0 to 100% and is similarly interpreted as

**Table 2** Characteristics of the validation and development populations of the CKD risk models (means and standard deviation (SD) or percent (%))

	Development populations for identified CKD prediction models					
	Heinz-Nixdorf-Recall-Study	Bang (SCORED, modified SCORED) 2007	Keams 2013	Kshirsagar 2008	Kwon 2011	Thakkinstian 2011
Country	Germany	USA, 72% white	UK, 30% white, 53% unknown	USA, 78% white	Korea	Thailand
Size of population (N)	4,185	8,530	743,935	9,470	6,565	3,459
CKD defined by	GFR < 60	GFR < 60	GFR < 60	GFR < 60	GFR < 60	KDIGO CKD1-5
Prevalence of CKD	8.6% (MDRD) 9.2% (CKD-Epi)	5.4% (MDRD)	6.8% (MDRD)	16.9% (MDRD)	4.6% (MDRD)	17.5 (CKD1-5, MDRD)
Handling of missings	Excluded	Excluded	Missing BP: no hypertension	Not reported	Excluded	Not reported
Female	50.5%	52%	50%	56%	50%	54.5%
Age, mean (SD)	59.6 (7.8)	46.0 (31.4)	46.7 (18.2)	57 (9)	44.2 (32.4)	45.2 (46.5)
Age, range	45–75	20–85	≥ 18	45–64	≥ 19	≥ 18
BMI, mean (SD)	27.8 (4.6)	28.0 (12.9)	-	27 (5)	23.6 (8.1)	24.0 (11.8)
Smoking	23.3%	20%	20%	10%	27.5 <sup>a</sup>	35.9%
Diabetes	7.9%	8%	5%	9%	8.3% <sup>b</sup>	11.9%
Hypertension	59.2%	34%	15%	36%	22.5%	27.5%
Ischemic heart disease or stroke	6.9%	4.9%	2% stroke, 4% IHD	8%	3.2% (IHD or stroke)	3.4% (heart disease)
Heart failure	3.5%	2.1%	1%	0.7%	-	-
Proteinuria	1.7%	10%	-	-	10.3%	-
PVD	2.3%	2.7%	1%	4%	-	-
Kidney stones	12.0% (N = 3,398)	-	-	-	-	5.0%
Serum creatinine mean (SD)	0.93 (0.2)	0.89 (0.4)	-	0.8 (0.2)	0.9 (8.1)	1.1 (1.18) male, 0.8 (1.18) female
eGFR, mean (SD)	79.1 (17.6) (CKD-Epi)	94.0 (48.9) (MDRD)	-	-	85.9 (56.7) (MDRD)	-
Anemia	2.1% (Hb < 12)	2.7% <sup>c</sup>	-	-	8.1%	-

BP: blood pressure; CKD: chronic kidney disease; CVD: cardiovascular disease; IHD: ischemic heart disease; KDIGO: Kidney Disease Improving Global Outcomes; MDRD: modification of Diet in Renal Disease; PVD: peripheral vascular disease

<sup>a</sup> smoking defined as: > 5 packs life time or current

<sup>b</sup> diabetes defined as: glucose ≥ 126 or anti-diabetic medication or insulin therapy

<sup>c</sup> anemia defined as: treatment of anemia

<sup>d</sup> history of bypass operation used as surrogate for heart failure

Pearson's  $R^2$ , indicating the rate of variability explained by the model.

We estimated sensitivity, specificity and predictive values of the models after scoring the model parameters using the cut-offs reported by the authors. The optimal threshold for the validation population was identified with the Youden index.

Additionally, we calculated the rate of expected to observed cases (E/O-proportion) for the thresholds used. An E/O-proportion close to 1 indicates agreement between the number of expected cases according to the models' cut-offs and observed cases, an E/O-proportion > 1 indicates overestimation of CKD risk.

All statistical calculations were done using SAS software Version 9.4 for Windows, SAS Institute Inc., Cary, NC, USA.

## Results

### Validation of identified risk models in the German HNR study

Compared to the development populations of the validated risk models, the participants in the HNR study were older (mean age 59.6 compared to 44.2 to 57 years), and reported hypertension more often (59% compared to 15 to 36%) (Table 2). Prevalence of CKD in HNR calculated by the MDRD-equation was 8.6%. It was higher than

in the development populations of the Bang (SCORED) (5.4%), Kearns (6.8%) and Kwon (4.6%) models. In the US based population used for deriving the Kshirsagar Score, CKD prevalence was 16.9%. Thakkinstian reported a prevalence of 17.5%, however, in contrast to the other studies, CKD here comprises CKD stages 1–5.

The mean estimated probability of prevalent CKD differed strongly and ranged from 0.025 (Kwon) to 0.317 (Thakkinstian) (Table 3).

### Discrimination

Discrimination of all models but in the Kshirsagar model was lower in HNR compared to the development data sets (Table 3). C-statistics ranged from 0.67 (Thakkinstian) to 0.73 (modified SCORED) using CKD-Epi equation for defining CKD. With MDRD equation, c-values were < 0.7 for all scores. The FAS equation yielded a better discrimination (c-statistics 0.74–0.80) (see supplement table S4). The Tjur-coefficient was largest and indicated best discrimination of prediction for the Thakkinstian score (0.124) and the Kearns model (0.116).

### Calibration

The calibration plots showed reasonable fit only with the Bang models (SCORED, modified SCORED); CKD risk was underestimated in persons with lower CKD risk and

**Table 3** External validation of identified CKD models; Measures presented with standard deviation (SD) or 95% confidence intervals (CI)

Characteristics	SCORED	Modified SCORED	Kearns	Kshirsagar	Kwon	Thakkinstian
Range of scoring points	[0–12]	[0–10]	n.a	[0–9]	[0–10]	[0–16]
eGFR equation in development	MDRD	MDRD	MDRD	MDRD	MDRD	MDRD
c-statistic in development (validation) set	0.88	0.87	0.90	0.69	0.83	0.77
<b>Results of validation in HNR</b>						
Participants in HNR	4,185	4,185	4,185	4,185	4,185	3,433
Mean prediction (SD)	0.084 (0.092)	0.087 (0.093)	0.148 (0.126)*	0.149 (0.081)	0.025 (0.029)	0.317 (0.189)
c-statistic (95% CI)	0.72 (0.70; 0.75)	0.73 (0.70; 0.75)	0.73 (0.71; 0.76)	0.71 (0.69; 0.74)	0.72 (0.70; 0.75)	0.67 (0.64; 0.70)
<b>Score performance in HNR</b>						
Tjur coefficient <sup>1</sup> (95%-CI)	0.063 (0.055- 0.071)	0.063 dec (0.055; 0.071)	0.116 (0.104; 0.129)	0.062 (0.053; 0.070)	0.020 (0.018; 0.023)	0.124 (0.102; 0.147)
MAPE <sup>2</sup> (SD)	0.148 (0.24)	0.150 (0.24)	0.193 (0.21)	0.203 (0.20)	0.108 (0.27)	0.329 (0.20)
Brier Scaled <sup>3</sup> (%)	12.0	11.6	2.7	7.3	38.0	4.2
<b>Examples of predicted risks for CKD (characteristics of fictitious persons)</b>						
w, 65 yrs with hypertension and IHD	30%	17%	39%	26%	4%	25%
w, 75 yrs with hypertension, DM, proteinuria and anemia	68%	68%	52%	15%	15%	62%
m, 57 yrs with hypertension, DM, HF	8%	8%	16%	2%	2%	27%
m, 72 yrs with DM and proteinuria	29%	29%	15%	18%	7%	43%

Validation data set: Heinz Nixdorf Recall study (Germany). CKD is defined as eGFR < 60 ml/min/1.73m<sup>2</sup> using CKD-Epi equation for calculating eGFR

<sup>1</sup> Tjur coefficient of discrimination = difference in the estimated mean probability of CKD for cases and non-cases; better with larger values

<sup>2</sup> MAPE = Mean average prediction error; better with small values

<sup>3</sup> Brier Scaled = comparable to  $R^2$ ; better with larger values

overestimated in persons with higher risk. Risk of prevalent CKD was generally overestimated by the Kearns, Kshirsagar and Thakkinstian model. The Kwon model underestimated the probability of prevalent CKD. Calibration to the HNR population was poor in all but the SCORED models (see supplementary file 1).

The Kwon score yielded the smallest mean average prediction error with 0.108 compared to the Thakkinstian score resulting in the largest MAPE with 0.329. The Brier Scaled which can be interpreted as  $R^2$  indicated that the Kwon score prediction fitted best to the population.

Applying the published scoring rules for the models to the HNR study, the E/O-proportion depended strongly on the eGFR equation used. For example, using a threshold of six points, the ‘modified SCORED’ had an E/O-proportion of 1.17 (95%-CI: 1.06; 1.28) with the CKD-Epi equation, 1.55 (1.42; 1.68) with the MDRD and 0.91 (0.84; 0.99) with the FAS equation (supplement Table 4, S4).

**Diagnostic criteria**

Thresholds of the scored models that resulted in a sensitivity > 60% and specificity > 60% (five points for the Bang and the Kwon models, four points for the Kshirsagar model and 7 points for the Thakkinstian model) resulted in positive predictive values (PPV) between 15 and 19%. Negative predictive values (NPV) were similar for all models independent of a chosen threshold and ranged between 93 and 95% (Table 4). Using the FAS-equation yielded the highest sensitivity and specificity for all thresholds compared to the CKD-Epi or MDRD equation (see supplement table S3a/b).

**Discussion**

We externally validated six prediction models which estimate the probability for prevalent CKD without any laboratory measurement. Most models yielded c-statistics of about 0.72 in the German HNR validation cohort. The calibration to the validation data set was reasonable only for the Bang (SCORED) models. All models but the Thakkinstian model, estimated a PPV of about 19% in a general population which indicates a good suitability as tools to identify patients at higher risk for whom further CKD diagnostic would be advisable.

**Current use of risk models in CKD**

Only few existing CKD prediction models are in use, such as the Kidney-Failure-Risk-Equation for prediction of progression of CKD to kidney failure [37, 38]. To prevent progression to serious consequences of CKD such as kidney failure or cardiovascular diseases, early diagnosis of CKD is necessary. Due to the asymptomatic progress of the disease even patients in later stages of CKD are often undiagnosed and therefore untreated [8].

**Table 4** Diagnostic criteria of validated CKD models for various thresholds: sensitivity, specificity and predictive values and proportions of expected to observed cases with 95%-confidence intervals

Model	Sensitivity	Specificity	Predictive values		Expected/Observed proportion (95% -CI)
			Positive	Negative	
SCORED					
4	84.9	44.4	13.4	96.7	6.33 (6.09; 6.59)
5	61.0	72.5	18.4	94.8	3.32 (3.14; 3.51)
6	30.6	90.3	24.3	92.8	1.26 (1.15; 1.38)
7	11.4	97.4	13.5	96.6	0.37 (0.32; 0.44)
Modified SCORED					
4	84.4	45.3	13.5	96.6	6.24 (6.00; 6.50)
5	60.8	73.4	18.8	94.9	3.24 (3.06; 3.42)
6	29.1	91.1	24.9	92.7	1.17 (1.06; 1.28)
7	9.1	97.9	31.0	91.4	0.29 (0.24; 0.33)
Kshirsagar					
3	84.2	44.4	13.3	96.5	6.33 (6.09; 6.59)
4	60.8	72.8	18.5	94.8	3.29 (3.12; 3.48)
5	29.9	90.7	24.6	92.7	1.22 (1.11; 1.33)
Kwon					
4	84.7	45.2	13.5	96.7	6.26 (6.01; 6.51)
5	59.7	73.7	18.7	94.8	3.20 (3.02; 3.38)
6	27.0	91.6	24.5	92.5	1.10 (1.00; 1.21)
7	6.2	98.6	30.4	91.2	0.21 (0.16; 0.26)
Thakkinstian					
6	82.9	45.7	13.4	96.3	6.19 (5.94; 6.44)
7	69.9	58.2	14.5	95.0	4.83 (4.61; 5.05)
8	43.9	81.0	18.9	93.4	2.32 (2.17; 2.48)
9	41.8	83.8	20.8	93.4	2.02 (1.88; 2.16)
10	38.7	86.6	22.6	93.3	1.71 (1.59; 1.85)
11	28.8	90.8	24.2	92.7	1.19 (1.09; 1.31)

CKD defined as eGFR < 60 ml/min/1.73m<sup>2</sup> calculated with CKD-Epi equation. CKD prevalence in validation population 9.2%

Contrasting to screening for CKD in the general population, case finding in populations at higher risk has been proven cost-effective [39]. Existing CKD risk models have not yet been applied in programs aiming at identifying persons at higher risk for CKD [17, 40]. The Kshirsagar and the SCORED model were used in a simulation study that proved cost-effectiveness in identifying persons for screening for early stages of CKD [22], SCORED was evaluated as screening tool for CKD in a small number of participants (N = 172) as alternative to regular CKD screening protocols [41].

Missing external validation of published CKD risk models can hinder clinical implementation. The SCORED, the Kwon, Khirsagar and the Thakkinstian

model have already been externally validated in 2016 in a UK population, but not with regard to the probability of prevalent CKD [24]. Also, models using parameters that are either unfamiliar, unusually scaled, complicated to calculate or costly to collect (e.g. genetic information) [42] or that do not reflect the general opinion in clinic about risk factors have a low probability of being used in clinical routine [43, 44]. Most of the identified models comprise familiar predictors which reflect the current knowledge about CKD risk factors. Only the Kearns and the Thakkinstian model do not use diabetes or sex for CKD prediction— parameters which usually would be regarded as relevant for estimating the risk for CKD.

### Validation results

Although the number and type of parameters used in the identified models differed, all -but the Thakkinstian model- showed fair discriminative properties in the HNR cohort with *c*-statistics ranging from 0.72 – 0.74. These *c*-statistics can be judged as satisfactory regarding the non-invasive and dichotomous nature of their predictors that facilitate potential implementations. Taking the small age range of the HNR cohort compared to the development populations of most of the validated scores into account, *c*-statistics could have been expected to be lower than in validation populations with full-age-spectrum [45] as, within a small age-range, it is more difficult to discriminate cases and non-cases when age is the most relevant prediction factor.

Calibration plots revealed a slight underestimation of CKD risk for the two Bang models. For all other models, calibration was poor. We recommend a re-calibration of the intercept [46] if the estimated probabilities are of interest in regard to a clinical implementation of a model. In our estimation of the diagnostic properties, we relied on the sum of scoring points for the parameters.

In our validation population, as well as in the populations used for developing the SCORED models, the Kshirsager and the Thakkinstian model, serum creatinine was measured with the Jaffe method. Compared with the more recent enzymatic methods, Jaffe measured creatinine values are in the mean higher, especially in the range of (enzymatically measured) creatinine values indicating CKD stages 1-3a [47]. Fewer people will be diagnosed with CKD in the early stages, if creatinine is measured enzymatically. This leads to a lower observed probability of CKD in the population. For those models underestimating the CKD probability, such as the SCORED scores, the use of enzymatically measured creatinine for diagnosing CKD could improve calibration.

### Potential implementation of the risk models

As the selected models do not use any laboratory or genetic information to estimate the probability for a prevalent CKD, these models can be used in screening scenarios where laboratory or genetic information would be too difficult or too expensive to get. Using the scoring rules for the models regarding the answers to the model parameters would enable to implement these models as self-completing check-list tool for patients which can easily be evaluated in a screening scenario [45]. However, acceptance of a prediction model is dependent on its face validity which means that the model parameters describe known risk factors. The validity of the Thakkinstian model without consideration of sex or of the Kearns model which does not imply diabetes might be questioned by physicians. On the other hand, proteinuria as a known risk factor is included in the SCORED and the Kwon model, but whether patients know for sure whether they had blood in their urine can be doubted.

Nevertheless, in a German general population, the modified SCORED and the Kwon model had good external validity and diagnostic properties. We think, that both models are suitable to identify people at higher risk for CKD at low cost if implemented as web based tool or distributed as paper questionnaire on information leaflets for example in public places or at health institutions. People who learn that they have a higher risk for CKD according their answers to the questionnaire may inform their GP who can decide to initiate further CKD diagnostic. Depending on the rationale behind a CKD screening or case finding effort and on the estimated prevalence of CKD in the population that is screened, the user of the models decides, whether sensitivity, specificity or PPV should be maximized and select the respective threshold value based upon this preference. We think this pragmatic approach can contribute to higher awareness for CKD, leading to earlier diagnosis and treatment.

### Strengths

The HNR cohort is of high data quality and has been the base for many publications so far. All relevant parameters of the models have been available for the external validation. To our knowledge we are the first to externally validate prediction models for prevalent CKD.

### Limitations

We did not intend to do a systematic review on all CKD models suitable for risk estimation for prevalent CKD. Therefore it might be possible that we did not include all existing models. However, to our knowledge, we were the first to evaluate all the selected models in regard to their ability to predict prevalent CKD. We think that we

herewith support potential implementations of these models.

The MDRD equation used in our sensitivity analyses has its weakness in a limited validity with eGFR levels  $\geq 60$  ml/min/1.73m<sup>2</sup>. This however, does not affect our validation results, as all models estimate the risk for CKD stage 3 or more which is defined by lower eGFR levels.

## Conclusions

Our external validation of risk models for prevalent CKD yielded fair discrimination in a German population-based cohort. Calibration to the data was satisfactorily only for some scores. Diagnostic properties show that the models can be useful in screening scenarios to identify people at higher risk for CKD. As only non-invasive parameters are used, they can easily be implemented as tool for patient self-assessment of CKD risk.

Further studies would need to estimate the impact of model based CKD screening on primary care in regard of risk factor control and health behavior in patients diagnosed with CKD - especially in people with diabetes, hypertension or of older age. However, it seems that a CKD diagnosis does not necessarily lead to a better control of risk factors such as hypertension or high blood glucose [7, 15]. The reason for this needs to be explored further to enhance health care for CKD patients.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12882-022-02899-0>.

**Additional file 1: S1.** Equations for eGFR estimation used in the analyses. **S2.** Equation of prediction models externally validated. **S3a.** Calibration plots for validated prediction models for CKD in the German Heinz-Nixdorf-Recall study (N=4,185). **S3b.** Calibration plots for validated prediction models for CKD in the German Heinz-Nixdorf-Recall study (N=4,185) – after re-calibration of the intercept. **Table S4.** Performance indicators for selected CKD risk models; CKD defined as eGFR <60ml/min/1.73m<sup>2</sup> calculated with FAS and MDRD equation. **Tab. S5.** Diagnostic criteria for prediction models for various threshold for CKD defined by MDRD and FAS equations (sensitivity, specificity and predictive values and the respective proportions of expected to observed cases and 95%-confidence intervals).

## Acknowledgements

The authors express their gratitude to all study participants of the Heinz Nixdorf Recall (HNR) Study, the personnel of the HNR study center and the EBT-scanner facilities, the investigative group and all former employees of the HNR study. The authors also thank the Advisory Board of the HNR Study: T. Meinertz, Hamburg, Germany (Chair); C. Bode, Freiburg, Germany; P.J. de Feyter, Rotterdam, Netherlands; B. Güntert, Hall i.T., Austria; F. Gutzwiller, Bern, Switzerland; H. Heinen, Bonn, Germany; O. Hess (†), Bern, Switzerland; B. Klein (†), Essen, Germany; H. Löwel, Neuherberg, Germany; M. Reiser, Munich, Germany; G. Schmidt (†), Essen, Germany; M. Schwaiger, Munich, Germany; C. Steinmüller, Bonn, Germany; T. Theorell, Stockholm, Sweden; and S.N Willich, Berlin, Germany.

## Authors' contributions

Research idea and study design: SS, AS; data acquisition and quality control: KJ, MF, DZ, RE; analysis/interpretation: SS, BK; supervision or mentorship: AS,

BK. All authors read the draft and provided comments. The author(s) read and approved the final manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL. The authors thank the Heinz Nixdorf Foundation (Chairman: Martin Nixdorf; Past Chairman: Dr jur. Gerhard Schmidt (†)), for their generous support of this study. Parts of the study were also supported by the German Research Council (DFG) (DFG project: El 969/2–3, ER 155/6–1;6–2, HO 3314/2–1;2–2;3;4–3, INST 58219/32–1, JO 170/8–1, KN 885/3–1, PE 2309/2–1, and SI 236/8–1;9–1;10–1), the German Ministry of Education and Science (BMBF project: 01EG0401, 01GI0856, 01GI0860, 01GS0820\_WB2-C, 01ER1001D, and 01GI0205), the Ministry of Innovation, Science, Research and Technology, North Rhine-Westphalia (MIWFT-NRW), the Else Kröner-Fresenius-Stiftung (project: 2015\_A119) and the German Social Accident Insurance (DGUV project: FF-FP295). Furthermore, the study was supported by the Competence Network for HIV/AIDS, the deanship of the University Hospital and IFORES of the University Duisburg-Essen, the European Union, the German Competence Network Heart Failure, Kulturstiftung Essen, the Protein Research Unit within Europe (PURE), the Dr Werner-Jackstädt Stiftung, and the following companies: Celgene GmbH München, Imatron/GE-Imatron, Janssen, Merck KG, Philips, ResMed Foundation, Roche Diagnostics, Sarstedt AG&Co, Siemens HealthCare Diagnostics, and Volkswagen Foundation. The study was funded by DFG within framework of TRR 296 LocoTact (DFG project: 424957847).

## Availability of data and materials

The corresponding author has full access to all data in the study and final responsibility for the submission of the article for publication. Due to data security reasons (i.e., data contain potentially participant identifying information), the HNR Study does not allow sharing data as a public use file. Data requests can be addressed to: recall@uk-essen.de.

## Declarations

### Ethics approval and consent to participate

The Heinz-Nixdorf-Study has been performed in accordance with the Declaration of Helsinki and has been approved by the ethic committee of the university Duisburg-Essen on 12. May 1999 (Approval-Nr. 99–69-1200). All participants gave their written informed consent before participating in the studies. Human gametes or stem cells were not involved.

### Consent for publication

Not applicable.

### Competing interests

The authors have declared that no competing interests exist.

### Author details

<sup>1</sup>Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Hufelandstr. 55, 45147 Essen, Germany. <sup>2</sup>School of Public Health, Department of Epidemiology, Boston University, Boston, USA.

Received: 21 December 2021 Accepted: 11 July 2022

Published online: 01 August 2022

## References

1. Bruck K, Stel VS, Gambaro G, et al. CKD Prevalence varies across the European general population. *J Am Soc Nephrol*. 2016;27(7):2135–47. <https://doi.org/10.1681/ASN.2015050542>.
2. Baumeister SE, Boger CA, Kramer BK, et al. Effect of chronic kidney disease and comorbid conditions on health care costs: A 10-year observational study in a general population. *Am J Nephrol*. 2010;31(3):222–9. <https://doi.org/10.1159/000272937>.
3. Stevens PE, Levin A. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med*. 2013;158(11):825–30. <https://doi.org/10.7326/0003-4819-158-11-201306040-00007>.

4. Delanaye P, Jager KJ, Bokenkamp A, et al. CKD: a call for an age-adapted definition. *J Am Soc Nephrol*. 2019;30(10):1785–805. <https://doi.org/10.1681/ASN.2019030238>.
5. Greer RC, Crews DC, Boulware LE. Challenges perceived by primary care providers to educating patients about chronic kidney disease. *J Ren Care*. 2012;38(4):174–81. <https://doi.org/10.1111/j.1755-6686.2012.00323.x>.
6. Wagner M M, Wanner C C, Schich M M, et al. Patient's and physician's awareness of kidney disease in coronary heart disease patients - a cross-sectional analysis of the German subset of the EUROASPIRE IV survey. *BMC Nephrol*. 2017;18(1):321. <https://doi.org/10.1186/s12882-017-0730-3>.
7. Stolpe S, Bock E, Scholz C, Stang A, Blume C. [Undiscovered impairment of renal function - a field with high relevance for the public health community.] 2018; <https://doi.org/10.3205/18gmds056>
8. Tuot DS, Plantinga LC, Hsu CY, et al. Chronic kidney disease awareness among individuals with clinical markers of kidney dysfunction. *Clin J Am Soc Nephrol*. 2011;6(8):1838–44. <https://doi.org/10.2215/cjn.00730111>.
9. Cosentino F, Grant PJ, Aboyans V, et al. 2019 ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J*. 2020;41(2):255–323. <https://doi.org/10.1093/eurheartj/ehz486>.
10. Guidelines AHATFoCP. Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American college of cardiology. *J Am Coll Cardiol*. 2018;71(19):e127–e248. <https://doi.org/10.1016/j.jacc.2017.11.006>.
11. NICE. Hypertension in adults: Diagnosis and management: National Institute for Health and Care Excellence. 2019.
12. Kostev K, Lucas A, Jacob L. Frequency of blood pressure and estimated glomerular filtration rate testing in type 2 diabetes mellitus: a retrospective study with 43,509 patients. *Exp Clin Endocrinol Diabetes*. 2019;127(7):455–60. <https://doi.org/10.1055/a-0581-4870>.
13. Jacob L, Seitz F, Kostev K. Frequency of blood pressure and estimated glomerular filtration rate monitoring in patients affected by hypertension: a retrospective study with 176 565 patients in Germany. *Blood Press Monit*. 2018;23(2):85–90. <https://doi.org/10.1097/mbp.0000000000000311>.
14. Rheinberger M, Jung B, Segiet T, et al. Poor risk factor control in outpatients with diabetes mellitus type 2 in Germany: The DIAbetes COhoRTE (DIACORE) study. *PLoS ONE*. 2019;14(3): e0213157. <https://doi.org/10.1371/journal.pone.0213157>.
15. Tuot DS, Plantinga LC, Judd SE, et al. Healthy behaviors, risk factor control and awareness of chronic kidney disease. *Am J Nephrol*. 2013;37(2):135–43. <https://doi.org/10.1159/000346712>.
16. Tonelli M, Dickinson JA. Early Detection of CKD: Implications for Low-Income, Middle-Income, and High-Income Countries. *J Am Soc Nephrol*. 2020;31(9):1931–40. <https://doi.org/10.1681/asn.2020030277>.
17. Galbraith LE, Ronksley PE, Barnieh LJ, et al. The see kidney disease targeted screening program for CKD. *Clin J Am Soc Nephrol*. 2016;11(6):964–72. <https://doi.org/10.2215/cjn.11961115>.
18. Hallan SI, Dahl K, Oien CM, et al. Screening strategies for chronic kidney disease in the general population: follow-up of cross sectional health survey. *BMJ*. 2006;333(7577):1047. <https://doi.org/10.1136/bmj.39001.657755.BE>.
19. Toussaint N. Screening for early chronic kidney disease 2012. Caring for Australian and New Zealanders with Kindey Impairment (CARI) Guidelines <https://www.cariguideelines.org/guidelines/chronic-kidney-disease/early-chronic-kidney-disease/screening-for-early-chronic-kidney-disease/>
20. Boersma C, Gansevoort RT, Pechlivanoglou P, et al. Screen-and-treat strategies for albuminuria to prevent cardiovascular and renal disease: cost-effectiveness of nationwide and targeted interventions based on analysis of cohort data from the Netherlands. *Clin Ther*. 2010;32(6):1103–21. <https://doi.org/10.1016/j.clinthera.2010.06.013>.
21. Manns B, Hemmelgarn B, Tonelli M, et al. Population based screening for chronic kidney disease: cost effectiveness study. *BMJ*. 2010;341:c5869. <https://doi.org/10.1136/bmj.c5869>.
22. Yarnoff BO, Hoerger TJ, Simpson SK, et al. The cost-effectiveness of using chronic kidney disease risk scores to screen for early-stage chronic kidney disease. *BMC Nephrol*. 2017;18(1):85. <https://doi.org/10.1186/s12882-017-0497-6>.
23. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med*. 2012;9(11): e1001344. <https://doi.org/10.1371/journal.pmed.1001344>.
24. Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford. *UK BMC Med*. 2016;14:104. <https://doi.org/10.1186/s12916-016-0650-2>.
25. Mogueo A, Echouffo-Tcheugui JB, Matsha TE, Erasmus RT, Kengne AP, Erasmus RT, Kengne AP. Validation of two prediction models of undiagnosed chronic kidney disease in mixed-ancestry South Africans. *BMC Nephrol*. 2015;16:94. <https://doi.org/10.1186/s12882-015-0093-6>.
26. Kearns B, Gallagher H, de Lusignan S. Predicting the prevalence of chronic kidney disease in the english population: a cross-sectional study. *BMC Nephrol*. 2013;14:49. <https://doi.org/10.1186/1471-2369-14-49>.
27. Bang H, Vupputuri S, Shoham DA, et al. SCreening for Occult REnal Disease (SCORED): a simple prediction model for chronic kidney disease. *Arch Intern Med*. 2007;167(4):374–81. <https://doi.org/10.1001/archinte.167.4.374>.
28. Kshirsagar AV, Bang H, Bomback AS, et al. A simple algorithm to predict incident kidney disease. *Arch Intern Med*. 2008;168(22):2466–73. <https://doi.org/10.1001/archinte.168.22.2466>.
29. Kwon KS, Bang H, Bomback AS, et al. A simple prediction score for kidney disease in the Korean population. *Nephrology (Carlton)*. 2012;17(3):278–84. <https://doi.org/10.1111/j.1440-1797.2011.01552.x>.
30. Thakkinstian A, Ingsathit A, Chairprasert A, et al. A simplified clinical prediction score of chronic kidney disease: a cross-sectional-survey study. *BMC Nephrol*. 2011;12:45. <https://doi.org/10.1186/1471-2369-12-45>.
31. Stang A, Moebus S, Dragano N, et al. Baseline recruitment and analyses of nonresponse of the Heinz Nixdorf Recall Study: identifiability of phone numbers as the major determinant of response. *Eur J Epidemiol*. 2005;20(6):489–96. <https://doi.org/10.1007/s10654-005-5529-z>.
32. Levey AS, Stevens LA. Estimating GFR using the CKD epidemiology collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD prevalence estimates, and better risk predictions. *Am J Kidney Dis*. 2010;55(4):622–7. <https://doi.org/10.1053/j.ajkd.2010.02.337>.
33. Trocchi P, Girdt M, Scheidt-Nave C, Markau S, Stang A. Impact of the estimation equation for GFR on population-based prevalence estimates of kidney dysfunction. *BMC Nephrol*. 2017;18(1):341. <https://doi.org/10.1186/s12882-017-0749-5>.
34. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? *Stat Methods Med Res*. 2018;27(11):3505–22. <https://doi.org/10.1177/0962280217705678>.
35. Tjur T. Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. *Am Stat*. 2009;63(4):366–72.
36. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>.
37. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011;305(15):1553–9. <https://doi.org/10.1001/jama.2011.451>.
38. Tangri N, Ferguson T, Komenda P. Pro: risk scores for chronic kidney disease progression are robust, powerful and ready for implementation. *Nephrol Dial Transplant*. 2017;32(5):748–51. <https://doi.org/10.1093/ndt/gfx067>.
39. Komenda P, Rigatto C, Tangri N. screening strategies for unrecognized CKD. *Clin J Am Soc Nephrol*. 2016;11(6):925–7. <https://doi.org/10.2215/CJN.04190416>.
40. Vart P, Reijneveld SA, Bultmann U, Gansevoort RT. Added value of screening for CKD among the elderly or persons with low socioeconomic status. *Clin J Am Soc Nephro*. 2015;10(4):562–70.
41. Harward DH, Bang H, Hu Y, Bomback AS, Kshirsagar AV. Evaluation of the scored questionnaire to identify individuals with chronic kidney disease in a community-based screening program in rural North Carolina. *J Community Med Health Educ*. 2014;4(Suppl 2):007. <https://doi.org/10.4172/2161-0711.S2-007>.
42. Lerner B, Desrochers S, Tangri N. Risk prediction models in CKD. *Semin Nephrol*. 2017;37(2):144–50. <https://doi.org/10.1016/j.semnephrol.2016.12.004>.
43. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606. <https://doi.org/10.1136/bmj.b606>.

44. Schuit E, Groenwold RH, Harrell FE Jr, et al. Unexpected predictor-outcome associations in clinical prediction research: causes and solutions. *CMAJ*. 2013;185(10):E499-505. <https://doi.org/10.1503/cmaj.120812>.
45. Bang H, Mazumdar M, Kern LM, Shoham DA, August PA, Kshirsagar AV. Validation and comparison of a novel screening guideline for kidney disease: keeping scored. *Arch Intern Med*. 2008;168(4):432–5. <https://doi.org/10.1001/archinternmed.2007.122>.
46. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. 2008;61(1):76–86. <https://doi.org/10.1016/j.jclinepi.2007.04.018>.
47. Syme NR, Stevens K, Stirling C, McMillan DC, Talwar D. Clinical and analytical impact of moving from jaffe to enzymatic serum creatinine methodology. *J Appl Lab Med*. 2020;5(4):631–42. <https://doi.org/10.1093/jalm/jfaa053>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

