

1 **External validation of six clinical models for prediction of unknown chronic kidney**
2 **disease in a German population**

3

4 Susanne Stolpe, MSc
5 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
6 Essen, Germany
7 Hufelandstr. 55, D-45147 Essen
8 Tel: +49 201 723 77231
9 Susanne.stolpe@uk-essen.de (corresponding author)

10

11 Bernd Kowall, PD
12 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
13 Essen, Germany

14

15 Denise Zwanziger, PhD
16 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
17 Essen, Germany

18

19 Mirjam Frank, PhD
20 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
21 Essen, Germany

22

23 Karl-Heinz Jöckel, Prof
24 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
25 Essen, Germany

26

27 Raimund Erbel, Prof
28 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
29 Essen, Germany

30

31 Andreas Stang, Prof
32 University Hospital Essen, Institute for medical informatics, biometry and epidemiology,
33 Essen, Germany
34 School of Public Health, Department of Epidemiology, Boston University, Boston, USA

35

36 Word count abstract: 209
37 Word count article 2.886

38

39 Abstract

40

41 **Background**

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

46

47 **Methods**

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

54

55 **Results**

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

60

61 **Conclusions**

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

65

66 **Background**

67 Prevalence of chronic kidney disease (CKD), defined by a chronically impaired renal function,
68 is growing worldwide and a challenge for public health. In Germany, prevalence of a decreased
69 renal function is up to 11.5 %¹ . Patients with CKD are at higher risk of cardiovascular
70 comorbidities, hospitalization, end stage renal disease (ESRD) and premature death². As a CKD
71 cannot be cured, treatment aims at monitoring CKD risk factors – especially hypertension and
72 blood glucose – to decelerate its progression and to prevent the incidence of secondary
73 diseases³. Early diagnosis of a prevalent CKD can support these efforts. Despite the high
74 prevalence and relevance for public health, public and patient awareness for CKD is low. The
75 main reason for the low awareness is that CKD remains asymptomatic until reaching more
76 serious stages. Moreover, a declining renal function is a physiological sign of older age, which
77 often hinders physicians to designate an impaired renal function in older age as CKD.^{4,5}

78 In Germany, fewer than 50% of patients with CKD -even with hypertension or cardiovascular
79 disease - knew about their condition ^{6,7}. In the USA, CKD unawareness is even more
80 prevalent⁸. The high unawareness for CKD is astonishing, as regular monitoring of renal
81 function should be mandatory for patients with hypertension, diabetes or of older age^{3,9-11}.
82 However, in general practice, even in patients with diabetes or hypertension, renal function,
83 blood glucose or blood pressure are not regularly monitored ^{12,13} and risk factors are
84 inadequately controlled ^{14,15}.

85 According to Tonelli et al. most of the principles for population screening -formulated in a
86 Delphi process among experts- fit to screening for CKD¹⁶. Screening for CKD has already been
87 shown to be cost-effective in detecting unknown cases in population subgroups at higher risk
88 for CKD ¹⁷⁻¹⁹. In a simulation study, it was shown that CKD prediction scores can be cost-
89 effectively used to initially identify people at higher risk for incident CKD, and to screen these
90 subsequently for CKD by testing for albuminuria ²⁰. Prediction models suitable for identifying

91 people at higher risk for CKD should be easy to apply, preferentially using non-invasive
92 parameters only. Many CKD risk models of different complexity have already been developed
93 – regarding the prediction of incidence, prevalence and progression to ESRD ²¹. However,
94 missing external validation might frequently hinder the implementation into practice ^{22,23}.

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

98 **Material and methods**

99 *CKD prediction models*

100 Starting with a review on CKD prediction models from 2012 ²¹, we searched the literature for
101 further models that comprise only clinical information for estimating the risk of prevalent CKD.
102 We identified five models that meet these criteria ²⁴⁻²⁸. Among these, Bang reported two
103 versions of a model developed in the same population: SCORED and modified SCORED
104 (Table 1). For these models, a scored version as self-completing questionnaire is published ²⁵.
105 The intercepts for the SCORED models and the models by Kwon and Thakkinstian had not
106 been published. For the SCORED and the Kwon model, we were able to get the missing
107 information about the intercept from the authors. For the Thakkinstian model we used the
108 intercept published in a validation study ²³. As the Kearns model estimated an unrealistic high
109 risk for CKD in our validation population, we contacted the author and learned that the age-
110 parameter should have had been centered (by subtracting the value 46.72) prior to the division
111 of age by 10). This had not been reported in the manuscript. Further, no cutoffs or scoring rule
112 for the Kearns-Model had been described, although sensitivity and specificity information had
113 been estimated in the paper. As we could get no information on the applied rules for scoring

114 the model parameters, we were not able to estimate the diagnostic properties of the Kearns
115 model in our validation cohort.

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

120

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

	Bang SCORED	Bang „modified SCORED“	Kearns	Kshirsagar	Kwon	Thakkinstian
No of parameters	9	7	5	8	7	4
Intercept	-5.40 ^a	-5.38 ^a	-3.63	-3.30	-6.53 ^a	-2.8 ^b
Age (yrs)	1.55 [50-59]	1.55 [50-59]	1.075 [Per 10 yrs ^c]	0.63 [50-59]	1.16 [50-59]	0.6 [50-59]
	2.31 [60-69]	2.29 [60-69]	-0.01 [age ² /10 yrs ^c]	1.33 [60-69]	1.91 [60-69]	1.4 [60-69]
	3.23 [\geq 70]	3.29 [\geq 70]	0.104 [age < 50]	1.46 [\geq 70]	2.71 [\geq 70]	2.1 [\geq 70]
Sex – Female	0.29	0.34	0.73	0.13	0.4	
Anemia	0.93			0.48	0.94	
Hypertension	0.45	0.47	0.74	0.55	0.48	0.8
			+ age <50: 0.56			
Diabetes	0.44	0.47		0.33	0.73	0.9
Ischemic heart disease or stroke (Hx)*	0.59	0.67		0.26	0.60	
Heart failure (hx)	0.45	0.51	0.86	0.50		
			CHF + age <50: 0.29			
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

122 a Intercept according to personal information by H. Bang.

123 b Intercept estimated using the prevalence of CKD in the validation population

124 c age per 10 yrs = (age -46.72) / 10yrs

125

126 *Validation population*

127 We used the German Heinz Nixdorf Recall Study (HNR), a population based cohort,²⁹ for
128 external validation. Baseline data from 4,814 participants drawn from the general population
129 aged 45-75 years in 2000 were available. We included all participants with a valid measurement
130 of serum creatinine (N=4,789).

131 *Measurement of variables*

132 In HNR, all laboratory data had been analyzed centrally in the laboratory of the university
133 hospital of Essen. Serum creatinine (according to Jaffé) was determined on a Siemens
134 Healthcare Diagnostics ADVIA Chemistry. Serum creatinine was not standardized to isotope
135 dilution mass spectrometry. Hypertension was defined as either a blood pressure of at least
136 140mmHg systolic or at least 90mmHg diastolic or taking antihypertensive medication. Blood
137 pressure cut-offs were selected according to the cut-offs used to define hypertension in the
138 validated risk models. Diabetes was defined according to the respective definitions used in the
139 risk models: either self-reported prevalent diabetes^{25,26} or using a combination of known
140 diabetes or taking antidiabetic drugs^{27,28}. Albuminuria was defined as albumin/creatinine ratio
141 (ACR) ≥ 30 mg/dl. In all models except the Kwon model, anemia was coded if hemoglobin levels
142 were < 12 g/dl. In the Kwon model, the threshold for hemoglobin was < 12 g/dl for women and
143 < 13 g/dl for men. Peripheral vascular artery disease was defined according to clinical
144 information.

145 *Definition of chronic kidney disease in development and validation populations*

146 In all validated risk models, CKD was defined as an estimated glomerular filtration rate (eGFR)
147 < 60 ml/min/1.73m² calculated by the Modification of Diet in Renal Disease (MDRD) equation.
148 All models had been developed to estimate the risk of unknown prevalent CKD stages 3 or
149 more. Therefore, we defined CKD as eGFR < 60 ml/min/1.73m² accordingly. We used the CKD-

150 Epi equation for calculating eGFR, as recommended by the Kidney Disease Improving Global
151 Outcomes (KDIGO)³⁰. In Germany, the CKD-Epi equation is widely used to report eGFR with
152 laboratory results. As sensitivity analysis, we calculated the eGFR with the MDRD and the new
153 Full-Age-Spectrum (FAS) equation (equations listed in supplement S2+S3)³¹. We used the
154 respective creatinine-based equations, because cystatin c measurements are not widely
155 available in general practice.

156 *Handling of missing values*

157 We did a complete case analysis regarding all predictors used in the identified models, leaving
158 4,185 participants in HNR. The Thakkinstian score was validated in a subsample of HNR with
159 information on the parameter ‘history of kidney stones’ that is used in this model only
160 (N=3,433).

161 *Statistical analysis*

162 The models’ discrimination was estimated by the c-value and the Tjur coefficient³². The Tjur
163 coefficient is the difference between the mean predicted probability in cases and in non-cases.
164 The higher this difference the better the discriminative ability of a score. Calibration was
165 assessed graphically.

166 As measures for overall performance, we estimated the mean average prediction error (MAPE)
167 and the Scaled Brier Score^{23,32}. MAPE averages the deviations between the prediction (ranging
168 between 0 and 1) and the respective true value of zero or 1. The smaller the MAPE, the better
169 the prediction. The Scaled Brier Score is calculated by the squared difference between the
170 prediction and the true value of outcome (=Brier Score) divided by the product of the mean
171 prediction value and 1 - mean prediction value³³. It ranges from 0 to 1 representing 0% to 100%
172 and is similarly interpreted as Pearson’s R², indicating the rate of variability explained by the
173 model.

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

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

181 All statistical calculations were done using SAS 9.4.

182

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

	Heinz-Nixdorf- Recall-Study	Development populations for identified CKD prediction models				
		Bang (SCORED, modified SCORED) 2007	Kearns 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 ^a	35.9%
Diabetes	7.9%	8%	5%	9%	8.3% ^b	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% ^c	-	-	8.1%	-

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

186 a: smoking defined as: >5 packs life time or current.

187 b: diabetes defined as: glucose ≥126 or anti-diabetic medication or insulin therapy.

188 c: anemia defined as: treatment of anemia.

189 d: history of bypass operation used as surrogate for heart failure.

190

191 **Results**

192 *Validation of identified risk models in the German HNR study*

193 Compared to the development populations of the validated risk models, the participants in the
194 HNR study were older (mean age 59.6 compared to 44.2 to 57 years), and reported hypertension
195 more often (59% compared to 15 to 36%) (Table 3). Prevalence of CKD in HNR calculated by
196 the MDRD-equation was 8.6%. It was higher than in the development populations of the Bang
197 (SCORED) (5.4%), Kearns (6.8%) and Kwon (4.6%) models. In the US based population used
198 for deriving the Kshirsagar Score, CKD prevalence was 16.9%. Thakkinstian reported a
199 prevalence of 17.5%, however, in contrast to the other studies, CKD here comprises CKD stages
200 1-5.

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

203

204 Table 3. External validation of identified models to predict unknown CKD; validation data set: Heinz Nixdorf Recall study (Germany). CKD is defined as eGFR
 205 <60ml/min/1.73m² using CKD-Epi equation for calculating eGFR. Measures presented with standard deviation (SD) or 95% confidence intervals (CI).

Characteristics	SCORED	Modified SCORED	Kearns	Kshirsagar	Kwon	Thakkinstian ^a
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
AUC in development (validation) set	0.88	0.87	0.90	0.69	0.83	0.77
Results of validation in HNR						
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)
Area under the curve (AUC, 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)
Score performance in HNR						
Tjur coefficient ¹ (95%-CI)	0.063 (0.055- 0.071)	0.063 (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 ² (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 ³ (%)	12.0	11.6	2.7	7.3	38.0	4.2
Examples of predicted risks for CKD (characteristics of fictitious persons)						
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%

206 1 Tjur coefficient of discrimination= difference in the mean prediction for cases and non-cases; better with larger values

207 2 MAPE = Mean average prediction error; better with small values

208 3 Brier Scaled = comparable to R²; better with larger values

209 *Discrimination*

210 Discrimination of all models but in the Kshirsagar model was lower in HNR compared to the
211 development data sets (Table 3). C-values ranged from 0.67 (Thakkinstian) to 0.73 (modified
212 SCORED) using CKD-EPI equation for defining CKD. With MDRD equation, c-values were
213 <0.7 for all scores. The FAS equation yielded a better discrimination (c-values 0.74-0.80) (see
214 supplement table S4). The Tjur-coefficient was largest and indicated best discrimination of
215 prediction for the Thakkinstian score (0.124) and the Kearns model (0.116).

216 *Calibration*

217 The calibration plots showed reasonable fit only with the Bang models (SCORED, modified
218 SCORED). CKD risk was overestimated in persons with lower CKD risk and overestimated in
219 persons with higher risk. Risk of unknown CKD was generally overestimated by the Kearns,
220 Kshirsagar and Thakkinstian model. The Kwon model underestimated the probability of
221 unknown CKD. Calibration to the HNR population was poor in all but the SCORED models
222 (see supplementary file).

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

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

231 *Diagnostic criteria*

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

239 Tab. 4 Diagnostic criteria of validated models for various thresholds (sensitivity, specificity and predictive values
240 as well as proportions of expected to observed cases with 95%-confidence intervals) for CKD. CKD defined as
241 eGFR <60m/min/1.73m² calculated with CKD-Epi equation; CKD prevalence 9.2%.

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)

242

243 **Discussion**

244 We externally validated six prediction models which estimate the probability for prevalent
245 unknown CKD without any laboratory measurement. Most models yielded c-values of about
246 0.72 in the German HNR validation cohort. The calibration to the validation data set was
247 reasonable only for the Bang (SCORED) models. A PPV of about 19% in a general population
248 as estimated with all models but the Thakkestian model indicates a good suitability as tools to
249 identify patients at higher risk for whom further CKD diagnostic would be advisable.

250 *Current use of risk models in CKD*

251 Only few existing CKD prediction models are in use, such as the Kidney-Failure-Risk-Equation
252 for prediction of renal failure.^{34,35} To prevent progression to serious consequences of CKD
253 such as ESRD or cardiovascular diseases, early diagnosis of CKD is necessary. Due to the
254 asymptomatic progress of the disease even patients in later stages of CKD are often
255 undiagnosed and therefore untreated⁸

256 Contrasting to screening for CKD in the general population, case finding in populations at
257 higher risk has been proven cost-effective^{36,37}. Existing CKD risk models have not yet been
258 applied in programs aiming at identifying persons at higher risk for CKD.^{18,38} The Khirsagar
259 and the SCORED model were used in a simulation study that proved cost-effectiveness in
260 identifying persons for screening for early stages of CKD²⁰, SCORED was evaluated as
261 screening tool for CKD in a small number of participants (N=172) as alternative to regular CKD
262 screening protocols³⁹.

263 Missing external validation of published CKD risk models can hinder clinical implementation.
264 The SCORED, the Kwon-, Khirsagar- and the Thakkestian- model have already been
265 externally validated in 2016 in a UK population, but not with regard to the probability of
266 prevalent CKD²². Also, models using parameters that are either unfamiliar, unusually scaled,

267 complicated to calculate or costly to collect (e.g. genetic information)⁴⁰ or that do not reflect
268 the general opinion in clinic about risk factors have a low probability of being used in clinical
269 routine.^{41,42} Most of the identified models comprise familiar predictors which reflect the
270 current knowledge about CKD risk factors. Only the Kearns and the Thakkinstian model do not
271 use diabetes or sex for CKD prediction– parameters which usually would be regarded as
272 relevant for estimating the risk for CKD.

273 *Validation results*

274 Although the number and type of parameters used in the identified models differed, all but the
275 Thakkinstian model showed fair discriminative properties in the HNR cohort with c-values
276 ranging from 0.72 – 0.74. These c-values can be judged as satisfactory regarding the non-
277 invasive and dichotomous nature of their predictors that facilitate potential implementations.
278 Taking the small age range of the HNR cohort compared to the development populations of
279 most of the validated scores into account, c-values could have been expected to be lower than
280 in validation populations with full-age-spectrum⁴³ as, within a small age-range, it is more
281 difficult to discriminate cases and non-cases when age is the most relevant prediction factor.

282 Calibration plots revealed a slight underestimation of CKD risk for the two Bang models. For
283 all other models, calibration was poor. We recommend a re-calibration of the intercept⁴⁴ if the
284 estimated probabilities are of interest in regard to a clinical implementation of a model. In our
285 estimation of the diagnostic properties we relied on the sum of scoring points for the parameters.

286 *Potential implementation of the risk models*

287 As the selected models do not use any laboratory or genetic information to estimate the risk for
288 a prevalent CKD, these models can be used in screening scenarios where laboratory or genetic
289 information would be too difficult or too expensive to get. Using the scoring rules for the models
290 regarding the answers to the model parameters would enable to implement these models as self-

291 completing check-list tool for patients which can easily be evaluated in a screening scenario⁴³.
292 However, acceptance of a prediction model is dependent on its face validity which means that
293 the model parameters describe known risk factors. The validity of the Thakkestian model
294 without consideration of sex or of the Kearns model which does not imply diabetes might be
295 questioned by physicians. On the other hand, proteinuria as a known risk factor is included in
296 the SCORED and the Kwon model, but whether patients know for sure whether they had blood
297 in their urine can be doubted.

298 Nevertheless, in a German general population, the modified SCORED and the Kwon model
299 had good external validity and diagnostic properties. We think, that both models are suitable to
300 identify people at higher risk for CKD at low cost if implemented as web based tool or
301 distributed as paper questionnaire on information leaflets for example in public places or at
302 health institutions. People who learn that they have a higher risk for CKD according their
303 answers to the questionnaire may inform their GP who can decide to initiate further CKD
304 diagnostic. We think this pragmatic approach can contribute to higher awareness for CKD,
305 leading to earlier diagnosis and treatment.

306 *Strengths*

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

310 *Limitations*

311 We did not intend to do a systematic review on all CKD models suitable for risk estimation for
312 unknown CKD. Therefore it might be possible that we did not include all existing model.
313 However, to our knowledge, we were the first to evaluate all the selected models in regard to

314 their ability to predict prevalent CKD. We think that we herewith support potential
315 implementations of these models.

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

319 **Conclusions**

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

325

326 **Funding**

327 The authors thank the Heinz Nixdorf Foundation (Chairman: Martin Nixdorf; Past Chairman:
328 Dr jur. Gerhard Schmidt (†)), for their generous support of this study. Parts of the study were
329 also supported by the German Research Council (DFG) (DFG project: EI 969/2-3, ER 155/6-
330 1;6-2, HO 3314/2-1;2-2;2-3;4-3, INST 58219/32-1, JO 170/8-1, KN 885/3-1, PE 2309/2-1,
331 and SI 236/8-1;9-1;10-1), the German Ministry of Education and Science (BMBF project:
332 01EG0401, 01GI0856, 01GI0860, 01GS0820_WB2-C, 01ER1001D, and 01GI0205), the
333 Ministry of Innovation, Science, Research and Technology, North Rhine-Westphalia
334 (MIWFT-NRW), the Else Kröner-Fresenius-Stiftung (project: 2015_A119) and the German
335 Social Accident Insurance (DGUV project: FF-FP295). Furthermore, the study was supported
336 by the Competence Network for HIV/AIDS, the deanship of the University Hospital and
337 IFORES of the University Duisburg-Essen, the European Union, the German Competence
338 Network Heart Failure, Kulturstiftung Essen, the Protein Research Unit within Europe
339 (PURE), the Dr Werner-Jackstädt Stiftung, and the following companies: Celgene GmbH
340 München, Imatron/GE-Imatron, Janssen, Merck KG, Philips, ResMed Foundation, Roche
341 Diagnostics, Sarstedt AG&Co, Siemens HealthCare Diagnostics, and Volkswagen
342 Foundation. The study was funded by DFG within framework of TRR 296 LocoTact (DFG
343 project: 424957847).

344 **Ethics approval and consent to participate**

345 The Heinz-Nixdorf-Study has been approved by the ethic committee of the university hospital
346 Essen on 12. May 1999. All participants gave their written informed consent before
347 participating in the studies.

348 **Participant protection**

349 Data is anonymized according to the data protection act, individual participants are not
350 identifiable.

351 **Competing interests**

352 The authors have declared that no competing interests exist.

353 **Authors' contributions**

354 Research idea and study design: SS, AS; data acquisition: KJ, MF, DZ, RE;
355 analysis/interpretation: SS, BK; statistical analysis: SS; supervision or mentorship: AS, BK.
356 Each author contributed important intellectual content during manuscript drafting or revision
357 and agrees to be personally accountable for the individual's own contributions and to ensure
358 that questions pertaining to the accuracy or integrity of any portion of the work, even one in
359 which the author was not directly involved, are appropriately investigated and resolved,
360 including with documentation in the literature if appropriate.

361 **Acknowledgement**

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

Literature:

1. Bruck K, Stel VS, Gambaro G, et al. CKD Prevalence Varies across the European General Population. *J Am Soc Nephrol*. 2016;27(7):2135-2147.
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):222-229.
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):825-830.
4. Delanaye P, Jager KJ, Bokenkamp A, et al. CKD: A Call for an Age-Adapted Definition. *J Am Soc Nephrol*. 2019;30(10):1785-1805.
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):174-181.
6. Wagner M, Wanner C, Schich 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.
7. Stolpe S, Bock E, Scholz C, Stang A, Blume C. Unentdeckte Beeinträchtigung der Nierenfunktion - ein Feld mit hoher Relevanz für die Public Health Community. 2018.
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):1838-1844.
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.
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. In. *J Am Coll Cardiol*. Vol 712018:e:127-e248.
11. NICE. Guideline. In. *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):455-460.
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.
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.
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):135-143.
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):1931-1940.
17. Toussaint N. *Screening for early chronic kidney disease*. 2012.
18. 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):964-972.
19. Manns B, Hemmelgarn B, Tonelli M, et al. Population based screening for chronic kidney disease: cost effectiveness study. *BMJ*. 2010;341:c5869.
20. 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.

21. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS medicine*. 2012;9(11):e1001344.
22. 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.
23. Mogueo A, Echouffo-Tcheugui JB, Matsha TE, Erasmus RT, Kengne AP. Validation of two prediction models of undiagnosed chronic kidney disease in mixed-ancestry South Africans. *BMC Nephrol*. 2015;16:94.
24. 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.
25. 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):374-381.
26. Kshirsagar AV, Bang H, Bombback AS, et al. A simple algorithm to predict incident kidney disease. *Arch Intern Med*. 2008;168(22):2466-2473.
27. Kwon KS, Bang H, Bombback AS, et al. A simple prediction score for kidney disease in the Korean population. *Nephrology (Carlton)*. 2012;17(3):278-284.
28. Thakkestian A, Ingsathit A, Chairasert A, et al. A simplified clinical prediction score of chronic kidney disease: a cross-sectional-survey study. *BMC Nephrol*. 2011;12:45.
29. 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):489-496.
30. 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):622-627.
31. Trocchi P, Girndt 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.
32. Tjur T. Coefficients of Determination in Logistic Regression Models-A New Proposal: The Coefficient of Discrimination. *Am Stat*. 2009;63(4):366-372.
33. 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):128-138.
34. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *Jama*. 2011;305(15):1553-1559.
35. 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):748-751.
36. Berns JS. Routine Screening for CKD Should Be Done in Asymptomatic Adults ... Selectively. *Clin J Am Soc Nephro*. 2014;9(11):1988-1992.
37. Komenda P, Rigatto C, Tangri N. Screening Strategies for Unrecognized CKD. *Clin J Am Soc Nephrol*. 2016;11(6):925-927.
38. 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):562-570.
39. Harward DH, Bang H, Hu Y, Bombback 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.
40. Lerner B, Desrochers S, Tangri N. Risk Prediction Models in CKD. *Semin Nephrol*. 2017;37(2):144-150.
41. 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.

42. 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.
43. 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):432-435.
44. 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.

Supporting information

Supporting file includes:

S1: Equations for eGFR estimation used in the analyses

S2: Equation of prediction models externally validated

S3: Fig S1: Calibration of the six models using CKD-Epi equation to define CKD 3-5

S4: Performance indicators for selected CKD risk models; CKD defined as eGFR <60ml/min/1.73m² calculated with FAS and MDRD equation

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).