

## Predicting the risk of chronic kidney disease in the UK:

an evaluation of QKidney® scores using a primary care database

### Abstract

#### Background

Chronic kidney disease is a major health concern that, if left untreated, may progress to end-stage kidney failure (ESKF). Identifying individuals at an increased risk of kidney disease and who might benefit from a therapeutic or preventive intervention is an important challenge.

#### Aim

To evaluate the performance of the QKidney® scores for predicting 5-year risk of developing moderate-severe kidney disease and ESKF in an independent UK cohort of patients from general practice records.

#### Design and setting

Prospective cohort study to evaluate the performance of two risk scores for kidney disease in 364 practices from the UK, contributing to The Health Improvement Network (THIN) database.

#### Method

Data were obtained from 1.6 million patients registered with a general practice surgery between 1 January 2002 and 1 July 2008, aged 35–74 years, with 43 186 incident cases of moderate-severe kidney disease and 2663 incident cases of ESKF. This is the first recorded evidence of moderate-severe chronic kidney and ESKF as recorded in general practice records.

#### Results

The results from this independent and external validation of QKidney scores indicate that both scores showed good performance data for both moderate-severe kidney disease and ESKF, on a large cohort of general practice patients. Discrimination and calibration statistics were better for models including serum creatinine; however, there were considerable amounts of missing data for serum creatinine. QKidney scores both with and without serum creatinine were well calibrated.

#### Conclusion

QKidney scores have been shown to be useful tools to predict the 5-year risk of moderate-severe kidney disease and ESKF in the UK.

#### Keywords

kidney disease; primary care; risk prediction; validation.

### INTRODUCTION

Chronic kidney disease (CKD) is a major health concern that, if left untreated, may progress to end-stage kidney failure (ESKF). While end-stage renal failure is relatively rare, the treatment, namely dialysis or transplantation, is expensive: in the UK, 2% of the total NHS budget is spent on renal replacement therapy.<sup>1</sup> Furthermore, those with CKD are at a greater risk of dying from cardiovascular disease than progressing to ESKF.<sup>2–5</sup> Identifying individuals who are at an increased risk of kidney disease, and who might benefit from a therapeutic or preventive intervention, is an important challenge. QKidney® scores are sex-specific multivariable risk scores that have recently been developed to predict the 5-year risk of moderate-severe CKD and the 5-year risk of developing ESKF in a primary care population.<sup>6</sup>

The risk scores were developed and validated on a large cohort of patients (2.4 million) from the QResearch® (www.qresearch.org) database; two-thirds of the cohort were randomly allocated for model development and one-third to model validation. The QResearch database is a large database comprising over 12 million anonymised health records from 602 practices throughout the UK, using the Egton Medical Information Systems (EMIS) computer system (used in 59% of general practices in England). QKidney scores were developed on 1.6 million patients aged 35–74 years, contributing 8 million person-years of observation, with 41 119 incident cases of moderate-severe CKD, and 2800

incident cases of ESKF between 1 January 2002 and 31 December 2008. QKidney scores were derived using a Cox proportional hazards model. Fractional polynomials were used to model non-linear risk relationships with continuous risk predictors, and the presence of interactions between risk predictors was tested.<sup>7</sup> Multiple imputation was used to replace missing values for key risk predictors (body mass index, systolic blood pressure, smoking status, and serum creatinine), to reduce the biases that can occur in a 'complete-case' analysis.<sup>8,9</sup> The final prediction models included 17 risk predictors for females and 15 risk predictors for males (Box 1). The open source code to calculate the QKidney scores is available from [www.qkidney.org](http://www.qkidney.org), released under the GNU Lesser General Public Licence (version 3).

This article describes the results from an independent evaluation assessing the performance of QKidney scores on a large dataset of general practice records in the UK. Due to concerns raised during the peer review process of using serum creatinine as a predictor of kidney disease, the authors of QKidney scores omitted it from the models.<sup>10</sup> This study thus sought to examine the added value and performance of including serum creatinine in the current QKidney scores.

### METHOD

#### Cohort selection

Study participants were patients registered between 1 January 2002 and 1 July 2008 and

**G Collins**, PhD, senior medical statistician;  
**D Altman**, DSc, professor of medical statistics,  
Centre for Statistics in Medicine, University of  
Oxford, Oxford.

#### Address for correspondence

Gary Collins, University of Oxford, Centre for  
Statistics in Medicine, Wolfson College Annexe,  
Linton Road, Oxford, OX2 6UD.

**E-mail:** [gary.collins@cs.m.ox.ac.uk](mailto:gary.collins@cs.m.ox.ac.uk)

**Submitted:** 20 September 2011; **Editor's response:**  
26 October 2011; **final acceptance:** 2 January 2012.

#### ©British Journal of General Practice

This is the full-length article (published online  
26 Mar 2012) of an abridged version published in  
print. Cite this article as: **Br J Gen Pract 2012;**  
**DOI: 10.3399/bjgp12X636065.**

## How this fits in

Chronic kidney disease (CKD) is a major health concern that, if left untreated, may progress to end-stage kidney failure (ESKF). While end-stage renal failure is relatively rare, the treatment, namely dialysis or transplantation, is expensive: in the UK, 2% of the total NHS budget is spent on renal replacement therapy. A new risk score has recently been developed to predict the 5-year risk of moderate-severe CKD and the 5-year risk of developing ESKF in a primary care population. The risk score has good discrimination and calibration and could be integrated into clinical computer systems to identify individuals who are at increased risk of developing moderate-severe CKD and ESKF.

recorded on The Health Improvement Network (THIN) database ([www.thin-uk.com](http://www.thin-uk.com)). Patients were eligible if they were aged 35–74 years, had no previously recorded evidence of CKD, were permanent residents in the UK, and had no interrupted periods of registration with a practice.

## Outcome measures

The two primary outcomes are recorded evidence of moderate-severe CKD and ESKF. Moderate-severe CKD was defined as the first occurrence of any of the following during follow-up: recorded kidney transplant, recorded kidney dialysis, recorded diagnosis of nephropathy, glomerular filtration rate  $<45 \text{ mL/min/1.73 m}^2$ , corresponding to stage 3B CKD,<sup>11</sup> or recorded diagnosis of

proteinuria. ESKF was defined as the first occurrence of any of the following during follow-up: recorded kidney transplant, recorded kidney dialysis, or glomerular filtration rate  $<15 \text{ mL/min/1.73 m}^2$ , corresponding to stage 5 CKD.<sup>11</sup> Glomerular filtration rates were calculated using the MDRD (modification of diet in renal disease) equation, using laboratory-reported creatinine values.<sup>12</sup>

## Statistical analysis

For each patient, an entry date was determined, which was the latest of the date of their 30th birthday, date of registration with the practice, date on which the practice computer system was installed plus 1 year, and the beginning of the study period (1 January 2002). Only patients with a minimum of 1 year's complete data in their medical record were included in the analysis. The observation time was calculated from the entry date to an exit date, which was defined as the earliest date of recorded CKD, date of death, date of deregistration with the practice, date of last upload of computerised data, or the study end date (1 July 2008).

Smoking status was derived from combining two risk factors; (1) whether the patient was a non-smoker, ex-smoker, or current smoker, and (2) the amount of cigarettes smoked, defined as light ( $<10$  cigarettes/day), moderate (10–19 cigarettes/day), or heavy ( $\geq 20$  cigarettes/day).

Multiple imputation using all predictors plus the outcome variable was used to replace missing values for smoking status, amount of cigarettes smoked, systolic blood pressure, body mass index, and serum creatinine. This involves creating multiple copies of the data and imputing the missing values with sensible values randomly selected from their predicted distribution. Twenty imputed datasets were generated and results from analyses on each of the imputed datasets were combined using Rubin's rules to produce estimates and confidence intervals that incorporate the uncertainty of imputed values.<sup>13</sup> The 5-year estimated risk of CKD (moderate-severe CKD and ESKF) for every patient in the THIN cohort was calculated using QKidney scores (with and without serum creatinine).

The predictive performance of QKidney scores for the THIN cohort was assessed by examining measures of calibration and discrimination. Calibration refers to how closely the predicted 5-year CKD risk agrees with the observed 5-year CKD risk. This was assessed for each tenth of predicted risk, and each 5-year age band, by

## Box 1. Summary of risk predictors in QKidney® scores

### Predictors for males and females

- Age (years)
- Body mass index ( $\text{kg/m}^2$ )
- Self-assigned ethnicity [white/not recorded, Indian, Pakistani, Bangladeshi, other Asian, black African, black Caribbean, other included mixed]
- Townsend deprivation score [output area level 2001 census data evaluated as a continuous variable]
- Smoking status [non-smoker, ex-smoker, light smoker ( $<10$  cigarettes/day), moderate smoker (10–19 cigarettes/day), heavy smoker ( $\geq 20$  cigarettes/day)]
- Systolic blood pressure (mmHg)
- Diagnosis of type 1 diabetes [yes/no]
- Diagnosis of type 2 diabetes [yes/no]
- Diagnosis of cardiovascular disease [yes/no]
- Diagnosis of congestive cardiac failure [yes/no]
- Diagnosis of peripheral vascular disease [yes/no]
- Diagnosis of rheumatoid arthritis [yes/no]
- Treated hypertension [yes/no]
- Two or more prescriptions of non-steroidal anti-inflammatory drugs 6 months before study entry [yes/no]
- Recorded family history of kidney disease including polycystic kidneys [yes/no]
- Serum creatinine ( $\mu\text{mol/litre}$ )

### Additional predictors for females

- Diagnosis of systemic lupus erythematosus [yes/no]
- Evidence of kidney stones based on diagnosis or operative procedure at baseline [yes/no]

**Table 1. Characteristics of patients aged 35–74 years and free of chronic kidney disease at baseline, in the THIN validation cohort**

	Females (n = 790 929)	Males (n = 802 577)
Median age (IQR), years	50 [42 to 60]	50 [41 to 59]
<b>Body mass index</b>		
Mean (SD), kg/m <sup>2</sup>	26.63 [5.14]	27.05 [4.24]
Not recorded, n (%)	128 394 [16.23]	200 734 [25.01]
<b>Systolic blood pressure</b>		
Mean (SD), mmHg	131.7 [20.27]	135.8 [18.70]
Not recorded, n (%)	40 390 [5.11]	103 564 [12.90]
<b>Serum creatinine</b>		
Mean (SD), µmol/litre	80.2 [16.63]	97.2 [20.59]
Not recorded, n (%)	671 258 [84.87]	695 092 [86.61]
<b>Ethnic group, n (%)</b>		
White/not recorded	773 961 [97.85]	786 349 [97.98]
Indian	4172 [0.53]	4221 [0.53]
Pakistani	1158 [0.15]	1233 [0.15]
Bangladeshi	353 [0.04]	460 [0.06]
Other Asian	1821 [0.23]	1848 [0.23]
Black Caribbean	2378 [0.30]	1772 [0.22]
Black African	2725 [0.34]	2509 [0.31]
Chinese	720 [0.09]	564 [0.07]
Other	3641 [0.46]	3621 [0.45]
<b>Smoking status</b>		
Non-smoker	360 380 [45.56]	285 736 [35.60]
Ex-smoker	103 038 [13.03]	136 372 [16.99]
Current smoker: light (<10 cigarettes/day)	48 864 [6.18]	48 258 [6.01]
Current smoker: moderate (10–19 cigarettes/day)	79 359 [10.03]	74 864 [9.33]
Current smoker: heavy (≥20 cigarettes/day)	60 944 [7.71]	90 001 [11.21]
Current smoker: amount not recorded	108 416 [13.71]	98 326 [12.25]
Smoking status not recorded	29 928 [3.78]	69 020 [8.60]
<b>Clinical conditions</b>		
Type 1 diabetes	2257 [0.29]	3049 [0.38]
Type 2 diabetes	21 950 [2.78]	31 000 [3.86]
Cardiovascular disease	30 263 [3.83]	54 278 [6.76]
Congestive cardiac failure	3635 [0.46]	5676 [0.71]
Peripheral vascular disease	10 305 [1.30]	10 684 [1.33]
Treated hypertension	78 917 [9.98]	69 360 [8.64]
Rheumatoid arthritis	10 069 [1.27]	4529 [0.56]
Systemic lupus erythematosus	1030 [0.13]	—
Kidney stones	3805 [0.48]	—
Current use of NSAIDs	211 163 [26.93]	175 924 [22.05]
Family history of kidney disease	627 [0.08]	534 [0.07]
5-year risk of moderate-severe kidney disease (95% CI)	3.19 [3.15 to 3.23]	2.43 [2.39 to 2.46]
5-year risk of end-stage kidney failure (95% CI)	0.15 [0.14 to 0.16]	0.19 [0.18 to 0.20]

IQR = interquartile range. NSAID = non-steroidal anti-inflammatory drug. SD = standard deviation.

calculating the ratio of predicted to observed CKD risk, separately for males and for females. Calibration was assessed by plotting the observed proportions versus the predicted risk. Observed 5-year CKD risks were obtained using the method of Kaplan–Meier.

Discrimination is the ability of the risk score to differentiate between patients who experience a CKD event during the study and those who do not. This measure is quantified by calculating the area under the receiver operating characteristic curve (AUROC) statistic; a value of 0.5 represents chance and 1 represents perfect

discrimination. The D statistic<sup>14</sup> and  $R^2$  statistic<sup>15</sup> were also calculated; these are measures of discrimination and explained variation respectively, and are tailored towards censored survival data. Higher values of D indicate greater discrimination, where an increase of 0.1 over other risk scores is a good indicator of improved prognostic separation.<sup>14</sup>

All statistical analyses were carried out in R (version 2.12.2), and the ICE procedure for multiple imputation in Stata (version 11.2).

## RESULTS

Between 1 January 2002 and 1 July 2008, 1 593 506 eligible patients from 364 general practices in the UK were registered in the THIN database. For the moderate-severe kidney disease end-point, 1 581 745 patients (49.6% females) were eligible for analysis, contributing 7 549 339 person-years of observation, among whom there were 43 186 incident cases of moderate-severe kidney disease. For the ESKF end-point, 1 593 506 patients were eligible for analysis, contributing 7 708 730 person-years of observation, among whom there were 2663 incident cases of ESKF. The median follow-up for the moderate-severe kidney disease end-point was 6.21 years (interquartile range [IQR] = 3.06 to 6.49 years) and for the ESKF end-point was 6.41 years (IQR = 3.24 to 6.49 years). In total, 939 718 patients (59%) were followed up for 5 years or more for the moderate-severe kidney disease end-point and 970 083 patients (61%) for the ESKF end-point.

The 5-year observed risk of moderate-severe kidney disease in females (24 371 incident cases of moderate-severe kidney disease) aged between 35 and 74 years was 3.19 [95% confidence interval (CI) = 3.15% to 3.23%] and in males (18 815 incident cases of moderate-severe kidney disease) was 2.43% [95% CI = 2.39 % to 2.46%]. The 5-year observed risk of ESKF in females (1171 incident cases of ESKF) aged 35–74 years was 0.15% [95% CI = 0.14% to 0.16%] and in males (1492 incident cases of ESKF) was 0.19% [95% CI = 0.18% to 0.20%]. Table 1 details the characteristics of the patients in the THIN cohort.

Complete data on smoking status, smoking category, body mass index, and systolic blood pressure were available for 70.1% of females (n = 549 661) and 62.4% of males (n = 497 451) for the models without serum creatinine. Most patients (n = 1 415 575; 89.5%) had no, or only one, missing risk factor (Table 2). For the models containing serum creatinine, complete data on smoking status, smoking category, body

**Table 2. Completeness of data**

Number of risk predictors not recorded (per patient)	Models without serum creatinine				Models including serum creatinine			
	Moderate-severe chronic kidney disease		End-stage kidney failure		Moderate-severe chronic kidney disease		End-stage kidney failure	
	Females (n = 784 005)	Males (n = 797 740)	Females (n = 790 929)	Males (n = 802 577)	Females (n = 784 005)	Males (n = 797 740)	Females (n = 790 929)	Males (n = 802 577)
0	549 661 (70.11)	497 451 (62.36)	554 747 (70.14)	500 988 (62.42)	83 096 (10.60)	73 713 (9.24)	87 229 (11.03)	76 394 (9.52)
1	182 448 (23.27)	186 015 (23.32)	184 132 (23.28)	187 137 (23.32)	494 580 (63.08)	450 176 (56.43)	496 953 (62.83)	451 894 (56.31)
2	23 597 (3.01)	45 621 (5.72)	23 686 (2.99)	45 721 (5.70)	156 173 (19.92)	161 683 (20.27)	156 509 (19.79)	162 016 (20.19)
3	7864 (1.00)	12 794 (1.60)	7904 (1.00)	12 839 (1.60)	22 387 (2.86)	44 233 (5.54)	22 436 (2.84)	44 292 (5.52)
4	20 435 (2.61)	55 859 (7.00)	20 460 (2.59)	55 892 (6.96)	7948 (1.01)	12 920 (1.62)	7975 (1.01)	12 951 (1.61)
5	—	—	—	—	19 821 (2.53)	55 015 (6.90)	19 827 (2.51)	55 030 (6.86)

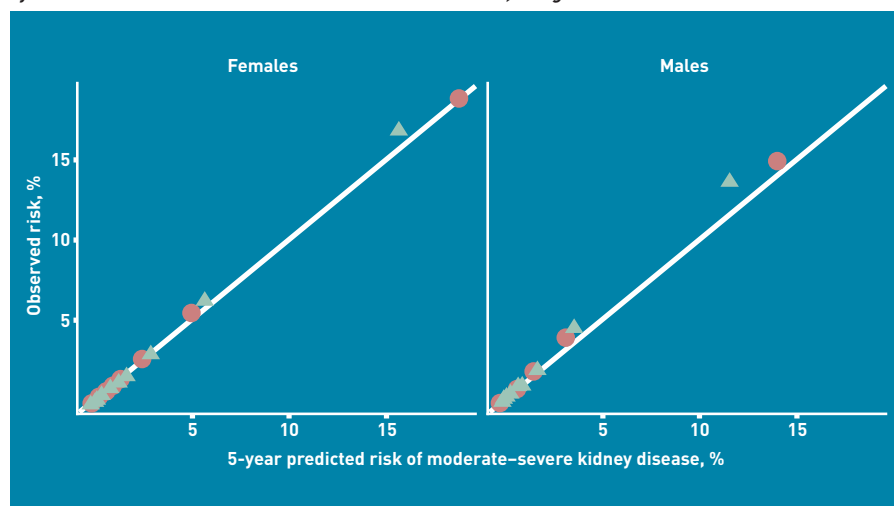
mass index, systolic blood pressure, and serum creatinine were available for 10.6% of females ( $n = 83\,906$ ) and 9.24% of males ( $n = 73\,713$ ). For the moderate-severe kidney disease model including serum creatinine, 1 101 565 patients (69.6%) had

no, or only one, missing risk factor (Table 2). There were considerably higher levels of missing data for serum creatinine, with only 227 156 patients (14.3%) having complete data recorded. Both males and females with serum creatinine recorded, tended to be older, have higher systolic blood pressure, have a diagnosis of type 2 diabetes or cardiovascular disease, and be treated for hypertension.

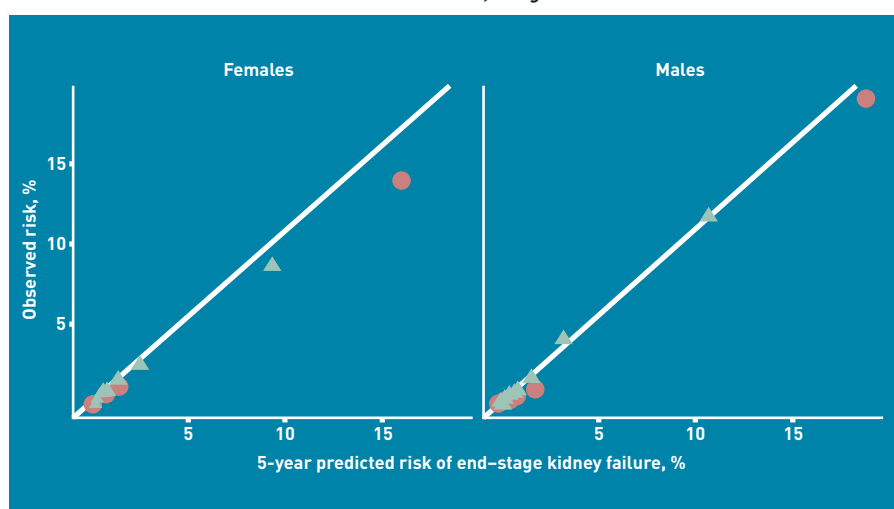
Table 3 shows the incidence rates (per 1000 person-years) for moderate-severe kidney disease and ESKF by age and sex in the THIN cohort. Females experienced 24 371 incident cases of moderate-severe kidney disease during 3 771 740 person-years of follow-up: incidence rate 6.46 per 1000 person-years (95% CI = 6.38 to 6.54). Males experienced 18 815 incident cases of moderate-severe kidney disease during 3 777 599 person-years of follow-up: incidence rate 4.98 per 1000 person-years (95% CI = 4.91 to 5.05). The rates of moderate-severe kidney disease were highest in those aged 70–74 years at baseline: incidence rate in females of 27.47 (95% CI = 26.91 to 28.04) and in males 24.86 (95% CI = 24.24 to 25.49) per 1000 person-years of observation. Overall, 1171 incident cases of ESKF were recorded for females during the 3 866 065 person-years of follow-up, giving an overall incidence rate of 0.30 per 1000 person-years (95% CI = 0.29 to 0.32) in females, whereas for males, 1492 incident cases of ESKF occurred during 3 842 665 person-years of follow-up, giving an incident rate of 0.39 per 1000 person-years (95% CI = 0.37 to 0.41)

Table 4 presents the performance data on discrimination and calibration for the QKidney scores in the THIN cohort. The  $R^2$  statistic (percentage of explained variation) was about 4% higher for the models predicting moderate-severe kidney disease including serum creatinine. For the models predicting ESKF, the  $R^2$  statistic was about

**Figure 1. Observed versus predicted risks of moderate-severe kidney disease with and without serum creatinine by tenth of risk. Circles denote model without serum creatinine; triangles denote model with serum creatinine.**



**Figure 2. Observed versus predicted risks of end-stage kidney failure with and without serum creatinine by tenth of risk. Circles denote model without serum creatinine; triangles denote model with serum creatinine.**



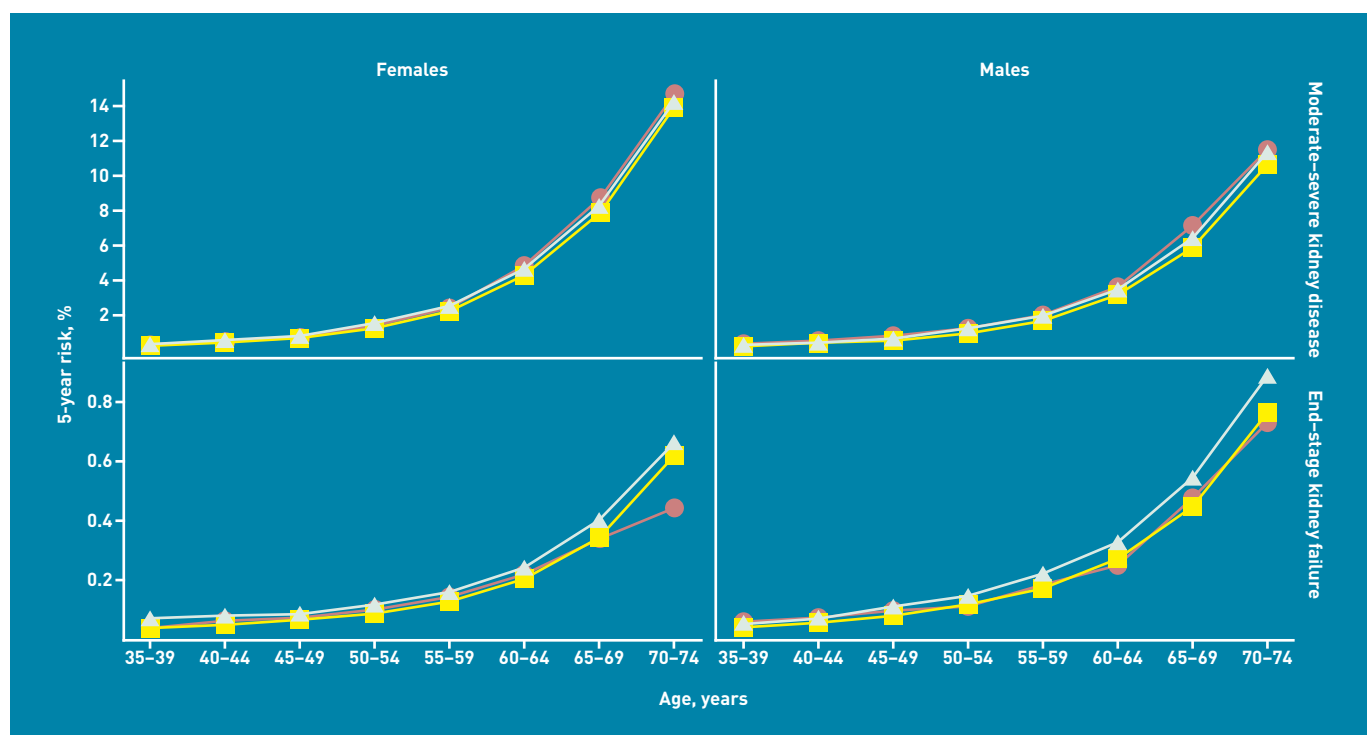
**Table 3. Incidence rates of moderate-severe chronic kidney disease and end-stage kidney failure in the THIN cohort, aged 35–74 years**

Age, years	Moderate-severe chronic kidney disease			End-stage kidney failure		
	Person-years of observation	Number of cases	Rate/1000 person years (95% CI)	Person-years of observation	Number of cases	Rate/1000 person years (95% CI)
<b>Females</b>						
35–39	644 881.3	359	0.56 [0.50 to 0.62]	646 620.9	45	0.07 [0.05 to 0.09]
40–44	590 913.5	552	0.93 [0.86 to 1.02]	593 277.1	70	0.12 [0.09 to 0.15]
45–49	520 478.1	761	1.46 [1.36 to 1.57]	523 547.9	74	0.14 [0.11 to 0.18]
50–54	506 988.9	1349	2.66 [2.52 to 2.81]	511 988.0	100	0.20 [0.16 to 0.24]
55–59	506 251.1	2429	4.80 [4.61 to 4.99]	515 158.2	154	0.30 [0.25 to 0.35]
60–64	385 147.9	3849	9.99 [9.68 to 10.31]	398 914.2	171	0.43 [0.37 to 0.50]
65–69	331 147.9	6057	18.29 [17.83 to 18.76]	354 377.1	248	0.70 [0.62 to 0.79]
70–74	331 411.0	9105	27.47 [26.91 to 28.04]	322 180.6	309	0.96 [0.86 to 1.07]
Total	3 771 740	24 371	6.46 [6.38 to 6.54]	3 866 065	1171	0.30 [0.29 to 0.32]
<b>Males</b>						
35–39	662 971.0	415	0.63 [0.57 to 0.69]	664 735.4	79	0.12 [0.09 to 0.15]
40–44	616 713.9	612	0.99 [0.92 to 1.07]	619 419.8	86	0.14 [0.11 to 0.17]
45–49	538 149.4	846	1.57 [1.47 to 1.68]	541 459.1	107	0.20 [0.16 to 0.24]
50–54	515 799.8	1241	2.41 [2.27 to 2.54]	520 266.3	124	0.24 [0.20 to 0.28]
55–59	509 148.5	2084	4.09 [3.92 to 4.27]	516 362.2	185	0.36 [0.31 to 0.41]
60–64	377 527.7	2850	7.55 [7.27 to 7.83]	386 830.6	202	0.52 [0.45 to 0.60]
65–69	313 758.5	4713	15.02 [14.60 to 15.46]	329 639.2	320	0.97 [0.87 to 1.08]
70–74	243 530.3	6054	24.86 [24.24 to 25.49]	263 952.7	379	1.44 [1.29 to 1.59]
Total	3 777 599	18 815	4.98 [4.91 to 5.05]	3 842 665	1492	0.39 [0.37 to 0.41]

**Figure 3. Observed versus predicted risks of moderate-severe kidney disease and end-stage kidney failure with and without serum creatinine by sex and age. Circles denote observed risk; triangles denote model with serum creatinine; squares denote model without serum creatinine.**

25% higher for the models including serum creatinine. The D statistic was noticeably higher for the models predicting ESKF including serum creatinine in both females and males (3.80 and 3.97, respectively) than for the models without serum creatinine (2.21 and 2.18, respectively). Values for the AUROC were comparable between females and males for the model predicting moderate-severe kidney disease without

serum creatinine (both 0.875) and for the models predicting ESKF (0.833 and 0.841, respectively). Corresponding values for the models including serum creatinine were noticeably higher, especially for the models predicting ESKF, where the values were 0.948 and 0.954 for females and males respectively. The performance of the models, restricting the analyses to those with available data to judge the adequacy of



**Table 4. Performance data of QKidney® scores with and without serum creatinine**

Model without serum creatinine	Multiple imputation (m = 20)		Complete case	
	Females	Males (n = 797 740)	Females	Males
<b>Moderate-severe kidney disease</b>	(n = 784 005)	(n = 797 740)	(n = 549 661)	(n = 497 451)
R <sup>2</sup> (95% CI)	56.23 (55.78 to 56.67)	57.83 (57.35 to 58.31)	55.68 (55.16 to 56.20)	55.11 (54.51 to 55.71)
D statistic (95% CI)	2.32 (2.30 to 2.34)	2.40 (2.37 to 2.42)	2.29 (2.27 to 2.32)	2.26 (2.24 to 2.30)
AUROC	0.875	0.875	0.872	0.861
<b>End-stage kidney failure</b>	(n = 790 929)	(n = 802 577)	(n = 554 747)	(n = 500 988)
R <sup>2</sup> (95% CI)	53.89 (51.81 to 55.94)	53.10 (51.24 to 54.97)	54.43 (51.95 to 56.77)	51.10 (48.70 to 53.40)
D statistic (95% CI)	2.21 (2.12 to 2.31)	2.18 (2.09 to 2.26)	2.23 (2.13 to 2.35)	2.09 (1.99 to 2.19)
AUROC	0.833	0.841	0.834	0.831
Model including serum creatinine	Multiple imputation (m = 20)		Complete case	
	Females	Males	Females	Males
<b>Moderate-severe kidney disease</b>	(n = 784 005)	(n = 797 740)	(n = 83 096)	(n = 73 713)
R <sup>2</sup> (95% CI)	60.44 (59.94 to 60.94)	62.25 (61.73 to 62.76)	51.33 (50.32 to 52.31)	51.25 (50.17 to 52.30)
D statistic (95% CI)	2.53 (2.50 to 2.56)	2.63 (2.60 to 2.66)	2.10 (2.06 to 2.14)	2.10 (2.06 to 2.14)
AUROC	0.893	0.892	0.851	0.837
<b>End-stage kidney failure</b>	(n = 790 929)	(n = 802 577)	(n = 87 229)	(n = 76 394)
R <sup>2</sup> (95% CI)	77.47 (76.43 to 78.50)	79.01 (78.23 to 79.80)	70.35 (67.99 to 72.49)	70.94 (68.80 to 72.90)
D statistic (95% CI)	3.80 (3.68 to 3.91)	3.97 (3.88 to 4.07)	3.15 (2.98 to 3.32)	3.19 (3.04 to 3.36)
AUROC	0.948	0.954	0.886	0.885

AUROC = area under the receiver operating characteristic curve.

missing data mechanism, shows clear differences for the ESKF model including serum creatinine, revealing a possible missing not-at-random mechanism. However, regardless of this, the results are noticeably higher both for those with available data and for those with multiple imputed approaches, indicating improved performance for the models including serum creatinine.

Figures 1 and 2 present the calibrations of QKidney scores for males and females by tenth of risk for the models with and without serum creatinine. Model calibration was good, with close agreement between the predicted and observed risk of moderate-severe kidney disease and ESKF across all tenths of risk, with no discernible over-prediction or under-prediction. Similarly, Figure 3 displays the calibration of QKidney scores for males and females by age group. Agreement between predicted and observed kidney disease risks was generally good across all age groups, with a slight over-prediction in the latter age groups for the ESKF model.

## DISCUSSION

### Summary

This independent evaluation of the performance of QKidney scores was carried out on the THIN database, based on general practices recording clinical data using the INPS Vision system, which is used in 20% of UK general practices. The dataset comprised 1.6 million patients registered between 1 January 2002 and 1 July 2008, contributing 7.5 million person-years of

observation. The performance data on the THIN cohort presented in this article provide strong evidence to support the external validity of QKidney scores in predicting the 5-year risk of developing moderate-severe kidney disease and ESKF. The differences between observed and predicted risks for the model to predict moderate-severe kidney disease (both with and without serum creatinine) were negligible across age groups for males and females.

### Strengths and limitations

A major strength of this study is the size and the representativeness of the cohort, achieved by including a large number of general practices using the EMIS computer system. Furthermore, the treatment of missing data was handled using the current recommended approaches to avoid the biases that occur when omitting individuals with incomplete data.<sup>8,16,17</sup> Both models for moderate-severe kidney disease and ESKF with or without serum creatinine had impressive discrimination properties and were well calibrated to the setting. When considering which model to use in routine clinical practice, with or without serum creatinine, the likelihood of the availability of risk factors is a major feature for consideration. There were substantial amounts of missing data for serum creatinine that should be taken into consideration when deciding which model to use. It is likely that serum creatinine was predominantly recorded in those who were suspected of being at an increased risk of kidney disease and thus is not routinely

### Funding

This study was not funded.

### Ethical approval

Trent Multicentre Research Ethics Committee.

### Provenance

Freely submitted; externally peer reviewed.

### Competing interests

The authors have declared no competing interests.

### Discuss this article

Contribute and read comments about this article on the Discussion Forum: <http://www.rcgp.org.uk/bjgp-discuss>



collected in those who are at low or negligible risk.

#### **Comparison with existing literature**

To date, QKidney scores are the first multivariable prediction models to predict the risk of moderate-severe kidney disease and ESKF for use in UK primary care. Existing prediction models for kidney disease have been developed in the US,<sup>18</sup> Canada,<sup>19</sup> the Netherlands,<sup>18,20</sup> and Taiwan;<sup>21</sup> however, the performance data to support these models are, at best, comparable<sup>19,20</sup> to the QKidney scores without serum creatinine. There are also no data to support their use in the UK. Furthermore, three of these prediction models include laboratory-based risk factors, making their use in routine clinical practice limited.<sup>19-21</sup>

#### **Implications for practice and research**

The performance data from this independent validation of QKidney scores support the case of using these scores to identify individuals who are at increased risk

of developing moderate-severe CKD and ESKF. The scores without serum creatinine have the potential to be used in clinical practice, as they only require information that can be obtained during a routine general practice appointment. While the addition of serum creatinine improved the predictive performance of the models, use of the models including serum creatinine as a population-based screening tool are limited, due to the laboratory-measured serum creatinine.

This study has provided an independent and external evaluation of the QKidney scores on a large cohort of patients in the UK to predict moderate-severe chronic kidney disease and ESKF. It has assessed the performance of QKidney scores with and without the inclusion of serum creatinine and has provided good evidence to confirm the predictive power of the risk scores without the need for recalibration. Given the large amounts of missing data for serum creatinine, the model without serum creatinine is the preferred model.

## REFERENCES

1. Ansell D, Feest T, Hodsman A, *et al.* *UK Renal Registry, The Renal Association, the ninth annual report*. Bristol: UK Renal Registry, 2006.
2. Drey N, Roderick P, Mullee M, Rogerson M. A population-based study of the incidence and outcomes of diagnosed chronic kidney disease. *Am J Kidney Dis* 2003; **42**(4): 677–684.
3. Go AS, Chertow GM, Fan D, *et al.* Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 2004; **351**(13): 1296–1305.
4. John R, Webb M, Young A, Stevens PE. Unreferred chronic kidney disease: a longitudinal study. *Am J Kidney Dis* 2004; **43**(5): 825–835.
5. Keith DS, Nichols GA, Gullion CM, *et al.* Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. *Arch Intern Med* 2004; **164**(6): 659–663.
6. Hippisley-Cox J, Coupland C. Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney® scores. *BMC Fam Pract* 2010; **11**(1): 49.
7. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999; **28**(5): 964–974.
8. Vergouwe Y, Royston P, Moons KGM, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* 2010; **63**(2): 205–214.
9. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999; **8**(1): 3–15.
10. BMC Family Practice. Pre-publication versions of this article and reviewers' reports. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney® Scores. <http://www.biomedcentral.com/1471-2296/11/49/prepub> (accessed 29 Feb 2012)
11. Crowe E, Halpin D, Stevens P. Early identification and management of chronic kidney disease: summary of NICE guidance. *BMJ* 2008; **337**: a1530.
12. Levey AS, Bosch JP, Lewis JB, *et al.* A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999; **130**(6): 461–470.
13. Rubin RB. *Multiple imputation for nonresponse in surveys*. New York, NY: J Wiley & Sons; 1987.
14. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004; **23**(5): 723–748.
15. Royston P. Explained variation for survival models. *Stata J* 2006; **6**(1): 83–96.
16. Janssen KJM, Donders ART, Harrell FE Jr, *et al.* Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol* 2010; **63**(7): 721–727.
17. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data with prognostic modelling studies: a simulation study. *BMC Med Res Methodol* 2010; **10**: 7.
18. Kshirsagar AV, Bang H, Bombardier AS, *et al.* A simple algorithm to predict incident kidney disease. *Arch Intern Med* 2008; **168**(22): 2466–2473.
19. 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.
20. Halbesma N, Jansen DF, Heymans MW, *et al.* Development and validation of a general population renal risk score. *Clin J Am Soc Nephrol* 2011; **6**(7): 1731–1738. Epub 2011/07/08.
21. Chien KL, Lin HJ, Lee BC, *et al.* A prediction model for the risk of incident chronic kidney disease. *Am J Med* 2010; **123**(9): 836–846.