BMJ

An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study

Gary S Collins, medical statistician,[1] Douglas G Altman, professor of statistics in medicine[1]

Centre for Statistics in Medicine, Wolfson College Annexe, University of Oxford, Oxford OX2 6UD

Correspondence to: G S Collins gary.collins@csm.ox.ac.uk

ABSTRACT

Objective To independently evaluate the performance of the QRISK score for predicting 10 year risk of cardiovascular disease in an independent UK cohort of patients from general practice and compare the performance with Framingham equations.

Design Prospective open cohort study.

Setting 274 practices from England and Wales contributing to the THIN database.

Participants 1.07 million patients, registered between 1 January 1995 and 1 April 2006, aged 35-74 years (5.4 million person years) with 43 990 cardiovascular events.

Main outcome measures First diagnosis of cardiovascular disease (myocardial infarction, coronary heart disease, stroke, and transient ischaemic attack) recorded in general practice records.

Results This independent validation indicated that QRISK offers an improved performance in predicting the 10 year risk of cardiovascular disease in a large cohort of UK patients over the Anderson Framingham equation. Discrimination and calibration statistics were better with QRISK. QRISK explained 32% of the variation in men and 37% in women, compared with 27% and 31% respectively for Anderson Framingham. QRISK underpredicted risk by 13% for men and 10% for women, whereas Anderson Framingham overpredicted risk by 32% for men and 10% for women. In total, 85 010 (8%) of patients would be reclassified from high risk (≥20%) with Anderson Framingham to low risk with QRISK, with an observed 10 year cardiovascular disease risk of 17.5% (95% confidence interval 16.9% to 18.1%) for men and 16.8% (15.7% to 18.0%) for women. The incidence rate of cardiovascular disease events among men was 30.5 per 1000 person years (95% confidence interval 29.9 to 31.2) in high risk patients identified with QRISK and 23.7 per 1000 person years (23.2 to 24.1) in high risk patients identified with Anderson Framingham. Similarly, the incidence rate of cardiovascular disease events among women was 26.7 per 1000 person years (25.8 to 27.7) in high risk patients identified with QRISK compared with 22.2 per 1000 person years (21.4 to 23.0) in high risk patients identified with Anderson Framingham.

Conclusions The QRISK cardiovascular disease risk equation offers an improvement over the long established Anderson Framingham equation in terms of identifying a high risk population for cardiovascular disease in the United Kingdom. QRISK underestimates 10 year cardiovascular disease risk, but the magnitude of underprediction is smaller than the overprediction with Anderson Framingham.

INTRODUCTION

Risk prediction models can play an important role in decision making and future management of individual or groups of patients with a particular medical condition.[1] Such models are designed, in principle, to estimate or predict the probability or risk of a patient developing some future clinical event based on a number of patient and disease characteristics.

Unfortunately, of the overwhelming plethora of risk score indices published every year most make no clinical impact, offer little in the design of future prognostic studies and subsequently disappear into the archives.[2] To date, the absence of explicit validation guidelines or indeed reporting guidelines for risk prediction has hampered the quality and clarity of published studies. Studies often have similar methodological problems to other types of studies: poor design, small sample size, incomplete data, inappropriate statistical analyses, and optimistic interpretation are concerns. In addition, there is a general lack of convincing external validation.[3]

The latter point, external validation, is an essential step in any risk model development in order to evaluate and show transportability of the model so that it could be applied with confidence on a cohort other than the derivation cohort.[4 5] Box and Draper aptly put it: "All models are wrong; the practical question is how wrong do they have to be to not be useful."[6] The act of validation is thus merely quantifying and judging how useful (or not) the model is in estimating the risk of developing a particular outcome.

The interpretation of the results from studies developing and validating risk prediction models is too often focused on classifying patients into risk groups, relying mainly on receiver operating characteristic curves, and neglects the accuracy of the actual risk prediction (calibration). Accurate prediction is crucial as any systematic overprediction would inevitably lead to a

disproportionate number of people being targeted for treatment, affecting healthcare resources and potentially exposing patients to unnecessary treatments. Similarly, any systematic underprediction of risk could potentially deny patients much needed treatment.

QRISK, a new multifactor cardiovascular disease risk prediction algorithm, was recently developed and validated for use in the United Kingdom. Initial results describing the derivation and internal validation were published in July 2007.[7] QRISK includes traditional cardiovascular disease risk factors—(1) age, (2) sex, (3) systolic blood pressure, (4) smoking status, and (5) serum cholesterol: high density lipoprotein ratio—that are incorporated in the long established Framingham risk equations,[8] but it also includes (6) body mass index, (7) family history of cardiovascular disease, (8) social deprivation (Townsend score),[9] and (9) the use of antihypertensive treatment. We note QRISK excludes patients with a pre-existing diagnosis of diabetes and does not include electrocardiogram assessment of left ventricular hypertrophy, both included in Framingham. The development of QRISK made use of elaborate statistical methods such as fractional polynomials[10] to model any non-linear risk relationships with continuous variables and multiple imputation methods[11] to avoid potentially biased estimates obtained from a reduced complete-case dataset.

In response to a number of critiques of the original *BMJ* papers,[12] the QRISK authors undertook a revision to account for statin use and implemented an improved approach to multiple imputation method to account for missing data. They then applied this revised model on an external dataset (THIN) in a second validation study to assess model performance.[13]

Despite the two published papers on the development and validation of QRISK, on impressively large cohorts, showing a more than competitive performance when compared with the Anderson Framingham and ASSIGN equations,[14] the National Institute for Health and Clinical Excellence (NICE) has not recommended QRISK and has recommended continued use of Framingham.[15] In addition, despite Framingham's limitations and known shortcomings,[16-19] NICE has recommended an unexplained adjustment factor to adjust for family history and ethnicity[20] (ethnicity is not a risk factor in the version of QRISK in this paper). However, this adjusted Anderson Framingham approach has, to date, not undergone any validation or peer reviewed consultation in the public domain. The QRISK developers have also been criticised for not making the QRISK algorithm publicly available so that head-to-head comparison can be made.[21] Finally, an unpublished report, attempting to revalidate the QRISK model on the THIN cohort, failed to replicate the results of the external validation and has further contributed doubts about the performance of QRISK.[22] This report has subsequently been shown to be incorrect and misleading and based on an incorrect and naive specification of a model, and to date it has not been made available in the public domain.[23]

This article describes and synthesises the results from an independent and external evaluation of QRISK commissioned by the Department of Health and compares the performance of QRISK against the Anderson Framingham equation[8] currently recommended by NICE.[20] In addition, we compare QRISK with a recently developed sex-specific Framingham risk equation.[24] This article presents additional material on the performance of QRISK currently not presented elsewhere. As the QRISK equation has not been made available in the public domain, we emphasise that we were granted full access by the QRISK authors to the QRISK algorithm and accompanying documentation, enabling an accurate implementation. In addition, we were granted permission to use the THIN dataset by EPIC Database Research Company.

METHODS
Study population
For this independent validation and verification analysis of QRISK, we used patients from the THIN database (www.thin-uk.com) as described by Hippisley-Cox et al[13] who were registered between 1 January 1995 and 31 March 2006. Patients were excluded if they had a prior diagnosis of cardiovascular disease, had invalid dates or invalid recorded risk factor values out of plausible range, were under the age of 35 years, were aged 75 years or over, were missing Townsend scores, had a diagnosis of pre-existing diabetes, or were prescribed statins at baseline.

Table 1 | Characteristics of 1 072 800 patients from the THIN database. Values are numbers (percentages) of patients unless stated otherwise

Characteristic	Men (n=529 813)	Women (n=542 987)
Median (interquartile range) age (years)	48 (40-57)	49 (41-59)
Mean (SD) systolic blood pressure (mm Hg)	135.6 (19.4)	132.1 (21.0)
Mean (SD) total serum cholesterol (mmol/l)	5.7 (1.1)	5.8 (1.2)
Mean (SD) high density lipoprotein cholesterol (mmol/l)	1.3 (0.4)	1.6 (0.4)
Mean (SD) total serum cholesterol: high density lipoprotein ratio	4.5 (1.3)	3.9 (1.2)
Mean (SD) body mass index (kg/m^2)	26.6 (4.0)	26.1 (4.9)
Current smoker	141 113 (26.6)	124 094 (22.9)
History of coronary heart disease in first degree relative <60 years old	18 638 (3.5)	22 922 (4.2)
Receiving antihypertensive treatment	35 066 (6.6)	56 886 (10.5)
Deprivation index (Townsend score fifths):		
1 (most affluent)	145 577 (27.5)	151 352 (27.9)
2	119 132 (22.5)	125 689 (23.1)
3	108 212 (20.4)	112 150 (20.7)
4	89 290 (16.9)	90 521 (16.7)
5 (most deprived)	67 602 (12.8)	63 275 (11.7)
No (%) of incident 10 year cardiovascular disease events	25 963 (4.9)	18 027 (3.3)
10 year risk (95% CI) of cardiovascular disease events	9.87 (9.71 to 10.03)	6.55 (6.43 to 6.68)
Total person years of observation (years)	2 603 294	2 753 924

Statistical analysis

Using QRISK, we calculated the 10 year estimated risk of cardiovascular disease for every patient in the THIN cohort. We obtained observed 10 year cardiovascular disease risks using the method of Kaplan-Meier. Missing data on three risk factors (total serum cholesterol: high density lipoprotein ratio, systolic blood pressure, and body mass index) were replaced with unpublished age-sex reference values from the QRESEARCH cohort used in the development of the QRISK risk algorithm for all risk prediction equations. The total serum cholesterol: high density lipoprotein ratios were replaced by reference values matched for age and sex, not the two individual components of this ratio. Where smoking status was not indicated, patients were assumed to be non-smokers.

Predictive performance of QRISK for the THIN cohort was assessed by examining measures of calibration and discrimination. Calibration measures how closely predicted 10 year cardiovascular disease risk agrees with observed 10 year cardiovascular disease risk. This was assessed for each tenth of predicted risk, ensuring 10 equally sized groups, and for each 5 year age band by calculating the ratio of predicted to observed cardiovascular disease risk, separately for men and for women. Calibration of the model predictions was assessed by plotting observed proportions versus predicted probabilities; where a 45° line denotes perfect calibration. The ratio of predicted to observed 10 year cardiovascular disease risk was calculated for each sex and overall, where a value of 1 is indicative of good agreement. The Brier score was also calculated, which is a measure of accuracy and is the average squared deviation between predicted and observed risk; a lower score represents higher accuracy.

Discrimination is the ability of the risk prediction model to differentiate between patients who experience a cardiovascular disease 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; where a value of 1 represents perfect discrimination. The cross classification of patients was tabulated for three risk groups (low <10%, intermediate 10% to <20%, and high ≥20%).

We calculated the D statistic[25] and R^2 statistic[26] (derived from the D statistic), which are measures of discrimination and explained variation respectively and are specific to censored survival data. Higher values of D indicate greater discrimination, and an increase of 0.1 over other risk prediction models is a good marker of improved prognostic separation.[25]

An important aspect when introducing a new risk prediction rule is the classification of patients into high and low risk and the number of patients that would be reclassified to a different risk category when compared to the standard risk prediction approach (here the Anderson Framingham equation). Patients are classified as being at high risk if their predicted risk is 20% or more.[27]

We compared the performance of QRISK with estimates of risk derived using the 1991 Framingham equation[8] (termed Anderson Framingham in this paper) and a recently developed sex-specific Framingham equation (termed Cox Framingham in this paper).[24] The Anderson Framingham equation[8] is based on a Weibull accelerated failure time model and is the current approach recommended in the UK
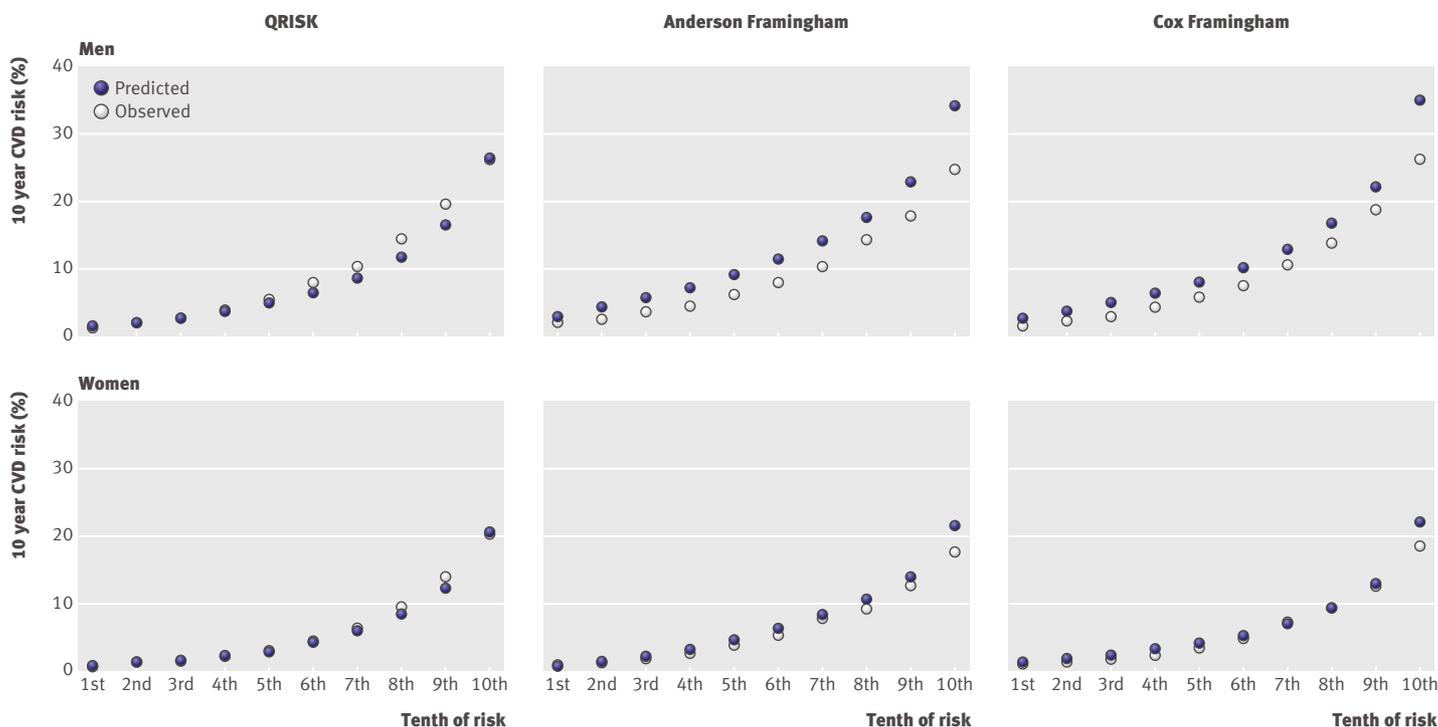

Fig 1 | Predicted versus observed 10 year risk of cardiovascular disease for QRISK and Framingham risk equations by tenth of risk
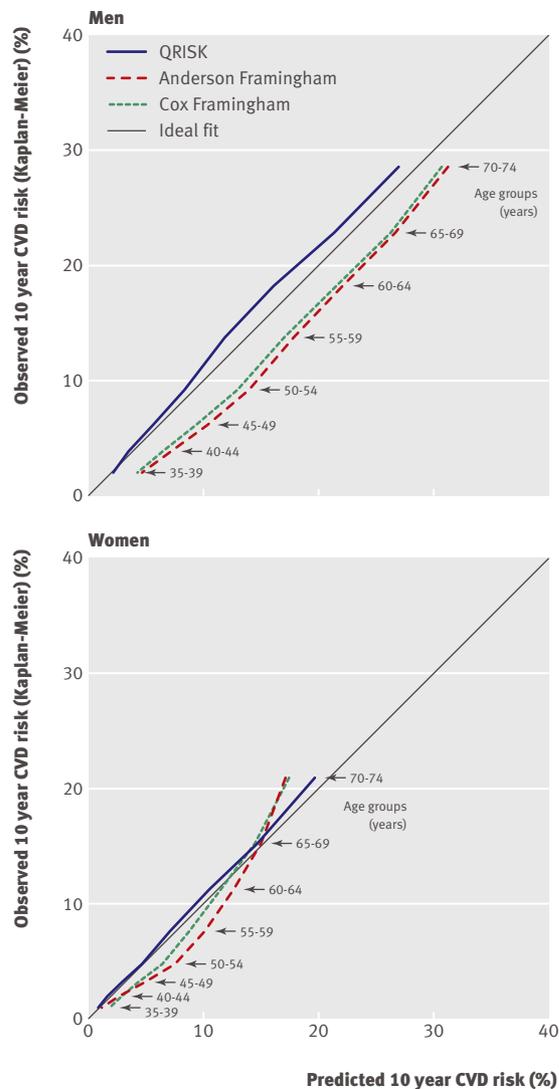
Fig 2 | Predicted versus observed 10 year risk of cardiovascular disease for QRISK and Framingham risk equations in 5 year age bands

and described in the Joint British Society guidelines.[28] The Anderson Framingham equation recommended by NICE was calculated by summing the risk from two individual equations for coronary heart disease and stroke (scores exceeding 100 are capped). The 2008 sex-specific version of Framingham is based on the Cox proportional hazards model and is a general cardiovascular risk score (coronary heart disease, stroke, peripheral artery disease, or heart failure).[8 24] Individual cardiovascular components (for coronary heart disease, which includes myocardial infarction and stroke including transient ischaemic stroke) were extracted by multiplying the general cardiovascular score by a calibration factor[24] and were summed to obtain a cardiovascular score (for myocardial infarction, coronary heart disease, stroke, and transient ischaemic attack endpoints).

All statistical analyses were carried out in *R* (version 2.8.0).[29]

RESULTS

Between 1 January 1995 and 31 March 2006, there were 1 787 169 patients from 288 practices registered in the THIN database. After sequentially excluding, as per the exclusion criteria, 120 281 patients with prior diagnosis of cardiovascular disease, 2253 with invalid dates, 439 740 aged <35 years or ≥75 years, 114 123 with missing Townsend scores, 28 148 patients with a pre-existing diabetes diagnosis, and 9824 patients with prior statin use, the analysed cohort consisted of 1 072 800 patients (see table 1). The median follow-up was 4.9 years (range 0 to 12 years), and 36 483 patients were followed for at least 10 years. The 10 year observed risk of a cardiovascular event in men aged 35-74 years was 9.87% (95% confidence interval 9.71% to 10.03%) and in women was 6.55% (6.43% to 6.68%).

Complete data for all risk factors considered were available for 26.9% of women and 25.5% of men. There were markedly high levels of missing data for total serum cholesterol (59.1% of women, 59.6% of men) and high density lipoprotein (70.6% of women, 71.4% of men). For most patients (63%), one or more of the three risk factors was missing (total serum cholesterol: high density lipoprotein ratio, systolic blood pressure, and body mass index); these had to be replaced with the QRESEARCH age-sex reference values. For 9% of the THIN cohort all three were not recorded. The observed 10 year cardiovascular disease risk was noticeably higher in those with complete data recorded on risk factors (19.9% (19.5% to 20.2%) for men, 12.8% (12.6% to 13.1%) for women) compared with those who had at least one missing risk factor (5.0% (4.9% to 5.2%) for men, 3.4% (3.2% to 3.5%) for women).

Discrimination and calibration

Fig 1 visually shows the agreement between mean observed risk and mean predicted risk grouped by tenths of predicted risk for all three models. For both men and women, the QRISK model gives a more accurate estimate of predicted risk compared with either Framingham equation. The accuracy of the Framingham equations deteriorates for those patients with higher risk for both men and women. Both Framingham equations consistently overestimate the risk for nearly all tenths of risk for men and women. In contrast, QRISK underestimates risk for both men and women.

Fig 2 shows the agreement between mean observed risk and mean predicted risk by 5 year age bands for both men and women. Both Framingham equations overestimate mean risk for all age groups for men and overestimate mean risk for women in all except the 65-69 and 70-74 year age groups, in which it underestimates risk, most notably in patients aged 70-74 years. QRISK underestimates cardiovascular disease risk in all age groups for men and to a lesser degree for women. QRISK provides more accurate cardiovascular disease risk estimates in all age groups

Table 2 | Ratio of predicted to observed risks of 10 year cardiovascular disease event for QRISK and Framingham risk equations across tenths of predicted risk and overall

Tenth of risk	QRISK			Anderson Framingham			Cox Framingham		
	Predicted risk (%)	Observed risk (%)	Ratio	Predicted risk (%)	Observed risk (%)	Ratio	Predicted risk (%)	Observed risk (%)	Ratio
Women:									
First	0.67	0.49	1.36	0.62	0.67	0.92	1.18	0.81	1.46
Second	1.00	1.02	0.98	1.23	1.08	1.14	1.76	1.06	1.66
Third	1.42	1.32	1.07	2.03	1.58	1.28	2.35	1.57	1.50
Fourth	2.03	2.00	1.01	3.15	2.65	1.19	3.08	2.31	1.33
Fifth	2.90	3.04	0.95	4.56	3.78	1.21	4.02	3.30	1.22
Sixth	4.09	4.30	0.95	6.25	5.32	1.18	5.28	4.78	1.10
Seventh	5.79	6.36	0.91	8.21	7.73	1.06	6.96	7.31	0.95
Eighth	8.32	9.44	0.88	10.55	9.06	1.16	9.29	9.28	1.00
Ninth	12.27	13.73	0.89	13.71	12.77	1.07	12.83	12.50	1.03
Tenth	20.46	20.19	1.01	21.55	17.49	1.23	21.96	18.42	1.19
Overall	5.89	6.55	0.90	7.19	6.55	1.10	6.87	6.55	1.04
Men:									
First	1.52	1.26	1.21	2.93	2.11	1.39	2.73	1.58	1.73
Second	2.12	2.17	0.97	4.34	2.59	1.67	3.87	2.26	1.71
Third	2.80	2.86	0.98	5.74	3.61	1.59	5.11	3.04	1.68
Fourth	3.72	3.93	0.95	7.31	4.61	1.59	6.50	4.44	1.46
Fifth	4.96	5.59	0.89	9.18	6.33	1.45	8.18	5.87	1.39
Sixth	6.59	8.08	0.82	11.43	8.24	1.39	10.29	7.71	1.33
Seventh	8.75	10.42	0.84	14.19	10.44	1.36	13.06	10.83	1.21
Eighth	11.87	14.56	0.82	17.82	14.36	1.24	16.83	13.91	1.21
Ninth	16.66	19.89	0.84	22.87	18.10	1.26	22.29	18.77	1.19
Tenth	26.62	26.26	1.01	34.27	24.83	1.38	35.18	26.17	1.34
Overall	8.56	9.87	0.87	13.01	9.87	1.32	12.39	9.87	1.25
All patients	7.21	8.16	0.88	10.10	8.16	1.23	9.60	8.16	1.18

compared with either Framingham model except for women aged 60-64 and 65-69 years.

Fig 3 shows the relationship between cardiovascular disease in relation to increasing age and predicted risk from QRISK and Framingham equations. The QRISK risk algorithm approximates well to the observed Kaplan-Meier cardiovascular disease estimates for both men and women across all age groups. The Anderson Framingham equation performs less well for men and women. Also, Anderson Framingham overestimates risk for women aged 40-64 then underestimates risk in women in the higher age groups, suggesting that it is not fully capturing the age-female component appropriately. Note that the Anderson Framingham equation is a single equation with a sex coefficient in the model, whereas QRISK comprises two sex-specific equations.

Table 2 shows the comparison of observed and predicted risk of cardiovascular disease at 10 years across each tenth of risk (the first tenth represents the lowest risk) for both Framingham equations and QRISK algorithm. Overall, the Framingham equations overpredicted risk at 10 years by 23% for the Anderson Framingham model and by 18% for the Cox Framingham model, whereas QRISK underpredicted risk by 12%. The Anderson Framingham equation performed similarly in women when compared with QRISK, with Anderson Framingham overpredicting risk by 10%

across the tenths of risk compared with an underprediction of 10% by QRISK. The Cox Framingham overpredicted the risk in women overall by only 4%, but this impressive performance is likely to be because the Cox Framingham overpredicts risk for women aged 35-64 years yet underpredicts risk for women aged ≥65, thus averaging out to an artificially good performance (fig 2). In men, both Framingham equations consistently overpredict risk in each tenth and overall by 32% and 25% for the Anderson Framingham and Cox Framingham model respectively. This compares with an overall underprediction of 13% by QRISK.

Table 3 shows discrimination and calibration performance data for QRISK and both Framingham equations. The R^2 statistic (percentage of explained variation) is higher for QRISK in both men and women (31.7% and 36.6% respectively) compared with the next best model, the Cox Framingham equation (29.5% and 32.3% respectively). The D discrimination statistic, where a higher value represents better discrimination, is higher in both men and women for QRISK (1.39 and 1.56 respectively). For the Anderson Framingham equation, the corresponding D statistic values for men and women were 1.26 and 1.38 respectively; values lower than the corresponding QRISK values by more than 0.1, indicating poorer discrimination of the Anderson Framingham. The Brier score, which is a measure of accuracy, was lower (that is,

more accurate) for QRISK in men (0.0470) compared with either the Anderson Framingham (0.0545) or Cox Framingham equation (0.0530). Similarly, for women, the Brier score was lower for QRISK (0.0321) compared with the Anderson Framingham (0.0334) and Cox Framingham equations (0.0330).

Risk classification

With a threshold of 20% to identify high risk patients, Anderson Framingham would identify 20% of the male cohort and 5% of the female cohort as being at high risk, compared with 10% and 4% respectively with QRISK. Table 4 shows the number of patients who would be reclassified from high risk (≥20%) with the Anderson Framingham equation to low risk (<20%) with QRISK and vice versa, and the average predicted and observed risks. In total, 85 010 patients (71% men, 29% women) would be reclassified (8%), of whom 57 199 men (67%) and 13 566 women (16%) would be downgraded from high risk with the Anderson Framingham to low risk with QRISK. In these patients, the observed 10 year cardiovascular disease risk was 17.5% (95% confidence interval 16.9% to 18.1%) for men and 16.8% (15.7% to 18.0%) for women. Conversely, 3548 men (4%) and 10 697 women (13%) would be reclassified from low risk with Anderson Framingham to high risk with QRISK, with observed 10 year cardiovascular disease risks of 25.5% (23.1% to 28.1%) and 23.1% (21.6% to 24.7%) respectively.

The incidence rate of cardiovascular events among men designated high risk with QRISK was 30.5 per 1000 person years (95% confidence interval 29.9 to 31.2), whereas it was 23.7 per 1000 person years (23.2 to 24.1) with the Anderson Framingham equation. Similarly, for women identified as high risk with QRISK, the incidence rate of cardiovascular events was 26.7 per 1000 person years (25.8 to 27.7) and was 22.2 per 1000 person years (21.4 to 23.0) with Anderson Framingham. Thus, using the 20% threshold to identify high risk patients, QRISK identified a group of patients at a higher risk of a cardiovascular event than those identified with Anderson Framingham. Conversely, the incidence rates among those not
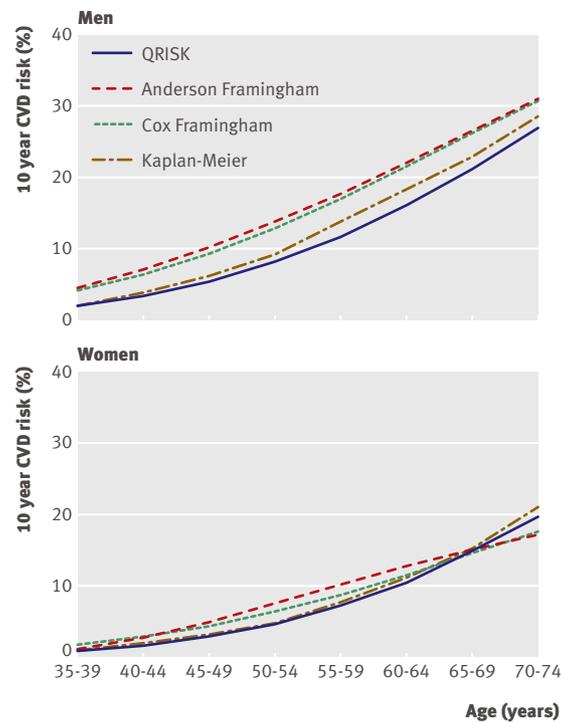

Fig 3 | Predicted versus observed (Kaplan-Meier) 10 year risk of cardiovascular disease for QRISK and Framingham risk equations by 5 year age bands

identified as being at high risk (<20%) was 7.8 per 1000 person years (7.6 to 7.8) for men and 6.5 per 1000 person years (6.4 to 6.6) for women with QRISK and 5.6 per 1000 person years (5.5 to 5.6) for men and 5.7 per 1000 person years (5.6 to 5.8) for women with Anderson Framingham.

In a similar manner to table 4, tables 5 and 6 show the cross classification of patients into three risk groups —<10% (low), 10 to <20% (intermediate), and ≥20% (high)—by QRISK and the Anderson Framingham equation, along with predicted and observed cardiovascular disease risk. Of the 106 265 men identified at being at high risk with Anderson Framingham, 57 199 (53.8%) would be reclassified as being at low risk (3%) or intermediate risk (50.9%). The observed risk of the

Table 3 | Discrimination and calibration statistics for predicted 10 year risk of cardiovascular disease by QRISK and Framingham risk equations

Statistic	QRISK		Anderson Framingham		Cox Framingham	
	Men	Women	Men	Women	Men	Women
AUROC	0.762	0.789	0.737	0.761	0.752	0.770
D statistic (95% CI)	1.39 (1.38 to 1.41)	1.56 (1.53 to 1.58)	1.26 (1.24 to 1.28)	1.38 (1.35 to 1.40)	1.33 (1.31 to 1.34)	1.41 (1.39 to 1.44)
R^2 statistic (95% CI)	31.71 (31.09 to 32.31)	36.64 (35.94 to 37.34)	27.31 (26.69 to 27.93)	31.18 (30.45 to 31.91)	29.52 (28.91 to 30.14)	32.32 (31.59 to 33.04)
Brier score*	0.0470	0.0321	0.0545	0.0334	0.0530	0.0330
Predicted/observed	0.87	0.90	1.32	1.10	1.25	1.04
Predicted/observed (overall)	0.88		1.23		1.18	

*A lower score indicates better accuracy of risk estimates

Table 4 | Comparison of QRISK and Anderson Framingham risk equation in classification of patients' 10 year risk of cardiovascular disease (low (<20%) or high (≥20%)) and observed and predicted risk

Anderson Framingham	QRISK			
	Men		Women	
	Low risk	High risk	Low risk	High risk
Low risk				
No of patients	420 000	3 548	506 478	10 697
No of cardiovascular events	12 975	566	13 486	1 446
% of patients classified	79.3	0.7	93.3	2.0
Average predicted risk of cardiovascular event (%):				
QRISK	5.4	24.7	4.8	23.8
Anderson Framingham	9.0	17.00	6.1	15.2
Observed risk % (95% CI)	6.5 (6.4 to 6.7)	25.5 (23.1 to 28.1)	5.3 (5.2 to 5.5)	23.1 (21.6 to 24.7)
High risk				
No of patients	57 199	49 066	13 566	12 246
No of cardiovascular events	5 243	7 179	1 324	1 771
% of patients classified	10.8	9.3	2.50	2.3
Average predicted risk of cardiovascular event (%):				
QRISK	15.2	26.8	15.3	25.9
Anderson Framingham	24.6	33.1	23.9	28.0
Observed risk % (95% CI)	17.5 (16.9 to 18.1)	26.3 (25.6 to 27.1)	16.8 (15.7 to 18)	24.2 (22.9 to 25.6)

patients downgraded to intermediate risk was 17.7% (17.1% to 18.3%). Similarly, for women, of the 25 812 patients identified at being at high risk with the Anderson Framingham equation, 13 566 patients (52.6%) would be downgraded into low (3.6%) and intermediate risk (48.9%). The observed risk of the patients downgraded to intermediate risk was 16.9% (15.8% to 18.1%). In contrast, 10 010 women identified as being at intermediate risk with Anderson Framingham were identified by QRISK as being at high risk, with an observed risk of 22.7% (95 21.2% to 24.3%).

Finally, the Framingham equations predicted that, of one million men identified at being at high risk, the number who would have a cardiovascular event over the next 10 years is 141 111 (Cox Framingham) and 116 896 (Anderson Framingham). Similarly, the number of high risk women expected to have a cardiovascular event over the next 10 years is 126 143 (Cox Framingham) and 119 905 (Anderson Framingham). However, in one million men and one million women identified as being at high risk with QRISK, the expected number that will go on to have a cardiovascular event in the next 10 years is higher at 147 204 and 140 217 respectively.

DISCUSSION
Principle findings
In this large cohort of patients, the Anderson Framingham equation[8] overestimated the 10 year risk of cardiovascular disease by 32% in men, by 10% in women, and by 23% overall. The newer Cox Framingham equation showed an improvement over the Anderson Framingham equation by overpredicting cardiovascular disease risk by 25% in men and by 18% overall, and it compared favourably with both

Anderson Framingham and QRISK in women by overpredicting risk for the entire female cohort by only 4%. However, this value gives a false impression by averaging out overprediction in younger women and underprediction in older women to an artificially good performance. With QRISK, 10 year cardiovascular disease risk was underpredicted by 13% for men, by 10% for women, and by 12% overall, but this model provides the most accurate predictions of 10 year cardiovascular disease risk in this large UK cohort. This finding is probably not surprising, as QRISK was developed on a separate but equally large cohort of patients in the UK and is thus more tailored to the UK population. In addition, QRISK contains additional risk factors—social deprivation, body mass index, family history, and current treatment with antihypertensives—which are known to affect cardiovascular disease risk[30] and that are not included in either of the Framingham equations.

On the evidence presented in this paper, the Framingham equations in their present form clearly overpredict 10 year cardiovascular disease risk in the United Kingdom, and this is more noticeable for men. In the THIN cohort, 20% of the male population would be identified at being at high risk of a cardiovascular disease event over the next 10 years, compared with just 10% if the QRISK algorithm had been used. Yet, if one million male patients and one million female patients identified as being at high risk were followed, then the number of cardiovascular disease events over the next 10 years will be 25% and 17% higher in men and women respectively with QRISK compared with Anderson Framingham, indicating that QRISK will target more high risk patients that would benefit from treatment.

Table 5 | Comparison of QRISK and Anderson Framingham risk equation in classification of male patients' 10 year risk of cardiovascular disease (low (<10%), intermediate (10% to <20%), high (≥20%)) and observed and predicted risk

Anderson Framingham	QRISK			Total No (%) of patients reclassified
	Low risk	Intermediate risk	High risk	
Low risk				
No of patients	257 315	1 213	46	
No of cardiovascular events	4 130	197	21	
% of patients classified	48.6	0.2	‹0.1	
Average predicted risk of cardiovascular event (%):				1 259 (0.2)
QRISK	3.1	12.3	23.4	
Anderson Framingham	5.8	8.5	8.6	
Observed risk % (95% CI)	3.7 (3.5 to 3.85)	28.3 (23.3 to 33.1)	—	
Intermediate risk				
No of patients	108 539	52 933	3 502	
No of cardiovascular events	4 113	4 535	545	
% of patients classified	20.5	10.0	0.7	
Average predicted risk of cardiovascular event (%):				112 041 (21.1)
QRISK	7.1	12.9	24.8	
Anderson Framingham	13.1	16.6	17.1	
Observed risk % (95% CI)	7.7 (7.4 to 8.0)	16.3 (15.7 to 17.0)	25.0 (22.4 to 27.4)	
High risk				
No of patients	3 162	54 037	49 066	
No of cardiovascular events	235	5 008	7 179	
% of patients classified	0.6	10.2	9.3	
Average predicted risk of cardiovascular event (%):				57 199 (9.8)
QRISK	8.7	15.6	26.8	
Anderson Framingham	22.8	24.7	33.1	
Observed risk % (95% CI)	14.3 (12.0 to 16.5)	17.7 (17.1 to 18.3)	26.3 (25.6 to 27.1)	

Strengths and weaknesses

There were high levels of missing data in this THIN cohort, especially for the risk factor total serum cholesterol: high density lipoprotein ratio. Most patients (79%) had at least one missing risk factor (from total serum cholesterol: high density lipoprotein ratio, systolic blood pressure, or body mass index) which required reference values (matched for age and sex) to be used. As one would expect, the observed 10 year cardiovascular disease risk was higher among patients with complete data on risk factors compared with those with at least one missing risk factor, as a consequence of those patients who visit their general practitioner for health-related problems being more likely to have these risk factors measured. The high amounts of missing data for total serum cholesterol and high density lipoprotein cholesterol were also observed in the development of QRISK from the QRESEARCH dataset, which uses the EMIS computer system.

Recent criticisms regarding missing data and questions about uncertainty on the accuracy of QRISK are misinformed and unfounded.[21] The original

development and validation,[7] the subsequent external validation,[13] and this commissioned external and independent validation and evaluation of QRISK have shown, on extremely large cohorts, that QRISK provides the scientific community with a risk prediction algorithm that performs better than the currently recommended Anderson Framingham risk score. Our independent evaluation was carried out using a different statistical software package (R version 2.8.0) from that used by the QRISK authors (Stata version 9.2), further supporting material presented in this article.

Readers may query whether the Framingham equations should be recalibrated to the UK population, so that the performance of the recalibrated model could then compare more favourably with QRISK.[21] Our brief was to independently compare QRISK with the current model recommended in the Joint British Society Guidelines,[28] which is the standard Anderson Framingham equation, a model developed on a comparatively small (n=5573), homogenous white sample from a single town in the US between 1968 and 1975. Undoubtedly, a recalibrated model could be expected to be more competitive, but at present no such recalibrated model exists in the public domain.

Another cardiovascular risk score, ASSIGN, has relatively recently been published, which in common with QRISK includes family history and social deprivation risk factors.[31] ASSIGN is a risk score developed using cohorts of Scottish men (n=6540) and women (n=6757) recruited in the 1980s and 1990s from the Scottish Heart Health Study,[32] when the incidence of cardiovascular disease was higher than in England.[33] However, comparison against ASSIGN was not possible in this validation study, as the social deprivation index used in ASSIGN (Scottish Index of Multiple Deprivation) is not recorded in the THIN database and there is no direct conversion to the Townsend index. Similarly, the number of cigarettes smoked (required to calculate the ASSIGN score) is not collected within the THIN database, only whether a patient smokes.

Future research

A modified version of QRISK (QRISK2) has recently been developed which includes ethnicity to account for the increased risk in south Asian men and women in the United Kingdom.[34] Initial results comparing this new model to the original QRISK and the Anderson Framingham equation recommended by NICE show an improvement in performance, but further independent validation on an external cohort of patients would be required.

Conclusions

In this study, we have provided an independent and external validation of the QRISK risk score on a large cohort of patients. We have assessed the performance of QRISK against the current NICE recommended model and have provided evidence to

Table 6 | Comparison of QRISK and Anderson Framingham risk equation in classification of female patients' 10 year risk of cardiovascular disease (low (<10%), intermediate (10% to <20%), high (≥20%)) and observed and predicted risk

Anderson Framingham	QRISK			Total No (%) of patients reclassified
	Low risk	Intermediate risk	High risk	
Low risk				
No of patients	384 923	10 901	687	
No of cardiovascular events	5 589	1 140	106	
% of patients classified	70.9	2.0	0.3	
Average predicted risk of cardiovascular event (%):				11 588 (2.1)
QRISK	2.7	12.7	23.2	
Anderson Framingham	3.8	8.4	8.4	
Observed risk % (95% CI)	3.1 (3.0 to 3.2)	18.6 (17.3 to 20.0)	29.0 (22.7 to 36.7)	
Intermediate risk				
No of patients	48 729	61 925	10 010	
No of cardiovascular events	1 943	4 814	1 340	
% of patients classified	9.0	11.4	1.8	
Average predicted risk of cardiovascular event (%):				58 739 (10.8)
QRISK	7.5	14.0	23.9	
Anderson Framingham	12.5	14.4	15.6	
Observed risk % (95% CI)	7.5 (7.0 to 7.9)	14.1 (13.6 to 14.6)	22.7 (21.2 to 24.3)	
High risk				
No of patients	940	12 626	12 246	
No of cardiovascular events	73	1 251	1 771	
% of patients classified	0.2	2.3	2.3	
Average predicted risk of cardiovascular event (%):				13 566 (2.5)
QRISK	8.5	15.8	25.9	
Anderson Framingham	22.8	24.0	28.0	
Observed risk % (95% CI)	15.4 (10.9 to 21.3)	16.9 (15.8 to 18.1)	24.2 (22.9 to 25.6)	

support the use of QRISK in favour of the Anderson Framingham equation.

Contributors: GSC conducted the analysis and prepared the first draft, which was revised according to comments and suggestions from DGA. GSC is guarantor for the paper.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Cardiovascular risk prediction in the United Kingdom is based on the US Framingham model that overpredicts risk

QRISK was developed using a large cohort of UK patients and published in 2007

Risk prediction models need to be independently and externally validated to objectively evaluate performance

WHAT THIS STUDY ADDS

Independent evaluation of QRISK showed an improvement in performance over the Framingham equations in a large external cohort of UK patients

QRISK identified a group of high risk patients who will go on to experience more cardiovascular events over the next 10 years than a similar high risk group identified by Framingham

1. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
2. Wyatt JC, Altman DG. Commentary: Prognostic models: clinical useful or quickly forgotten? *BMJ* 1995;311:1539-41.
3. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
4. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453-73.
5. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826-32.
6. Box GEP, Draper NR. *Empirical model-building and response surfaces*. New York: Wiley, 1987.
7. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335:136.
8. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 pt 2):293-8.
9. Tunstall-Pedoe H, Woodward M. By neglecting deprivation, cardiovascular risk scoring will exacerbate social gradients in disease. *Heart* 2006;92:307-10.
10. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;28:964-74.
11. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: J Wiley, 1987.
12. Rapid responses to: Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. www.bmj.com/cgi/eletters/335/7611/136 accessed 03/03/2009.
13. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. *Heart* 2008;94:34-9.
14. Woodward M, Brindle P, Tunstall-Pedoe H. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* 2007;93:172-6.
15. Cooper A, O'Flynn N, on behalf of the Guideline Development Group. Risk assessment and lipid assessment for primary and secondary prevention of cardiovascular disease: summary of NICE guidance. *BMJ* 2008;336:1246-8.
16. Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ* 2003;327:1267.
17. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart* 2006;92:1752-9.
18. Robson J. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. *Heart* 2008;94:1331-2.
19. Tooth PP. Making a case for quantitative assessment of cardiovascular risk. *Journal of Clinical Lipidology* 2007;1:234-41.
20. National Institute for Health and Clinical Excellence. Lipid modification: Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. London: NICE, 2008.
21. Neil HAW, Perera R, Armitage JM, Farmer AJ, Mant D, Durrington PN. Estimated 10-year cardiovascular risk in a British population: results of a national screening project. *International Journal of Clinical Practice* 2008;62:1322-31.
22. Robins M, Potter J, Poulter N, Skinner J, de la Iglesia B. Response to Section 4,3 of the Guideline on Cardiovascular Risk Assessment. 2008.
23. Collins GS, Altman DG. Report to the Department of Health: Independent validation of QRISK on the THIN database. Oxford: University of Oxford, 2008.
24. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743-53.
25. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;23:723-48.

26 Royston P. Explained variation for survival models. *Stata Journal* 2006;6:1-14.

27 National Institute for Health and Clinical Excellence. Statins for the prevention of cardiovascular events in patients at increased risk of developing cardiovascular disease or those with established cardiovascular disease. London: NICE, 2006.

28 JBS2. Joint British Society guidelines on prevention of cardiovascular disease in clinical practice. *Heart* 2005;91(suppl 5):1-52.

29 R Development Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, www.R-project.org.

30 Lloyd-Jones DM, Nam B-H, D'Agostino RB, Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* 2004;291:2202-11.

31 Woodward M, Brindle P, Tunstall-Pedoe H. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* 2007;93:172-6.

32 Tunstall-Pedoe H, Woodward M, Tavendale R, Brook RA, McCluskey MK. Comparison of the prediction by 27 different factors of coronary heart disease and death in men and women of the Scottish Heart Health Study: cohort study. *BMJ* 1997;315:722-9.

33 Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ* 2003;327:1267.

34 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:a332.