

An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study

Gary S Collins, senior medical statistician,¹ Douglas G Altman, director, professor of statistics in medicine¹

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

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

Cite this as: *BMJ* 2010;340:c2442
doi:10.1136/bmj.c2442

ABSTRACT

Objective To evaluate the performance of the QRISK2 score for predicting 10-year cardiovascular disease in an independent UK cohort of patients from general practice records and to compare it with the NICE version of the Framingham equation and QRISK1.

Design Prospective cohort study to validate a cardiovascular risk score.

Setting 365 practices from United Kingdom contributing to The Health Improvement Network (THIN) database.

Participants 1.58 million patients registered with a general practice between 1 January 1993 and 20 June 2008, aged 35-74 years (9.4 million person years) with 71 465 cardiovascular events.

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

Results QRISK2 offered improved prediction of a patient's 10-year risk of cardiovascular disease over the NICE version of the Framingham equation. Discrimination and calibration statistics were better with QRISK2. QRISK2 explained 33% of the variation in men and 40% for women, compared with 29% and 34% respectively for the NICE Framingham and 32% and 38% respectively for QRISK1. The incidence rate of cardiovascular events (per 1000 person years) among men in the high risk group was 27.8 (95% CI 27.4 to 28.2) with QRISK2, 21.9 (21.6 to 22.2) with NICE Framingham, and 24.8 (22.8 to 26.9) with QRISK1. Similarly, the incidence rate of cardiovascular events (per 1000 person years) among women in the high risk group was 24.3 (23.8 to 24.9) with QRISK2, 20.6 (20.1 to 21.0) with NICE Framingham, and 21.8 (18.9 to 24.6) with QRISK1.

Conclusions QRISK2 is more accurate in identifying a high risk population for cardiovascular disease in the United Kingdom than the NICE version of the Framingham equation. Differences in performance between QRISK2 and QRISK1 were marginal.

INTRODUCTION

Cardiovascular disease is an important health concern globally, with just under a third of all deaths attributed to cardiovascular disease in 2004 (www.who.int, fact

sheet No 317). In the United Kingdom, there are almost 200 000 deaths each year relating to diseases of the heart and circulatory system, with more than one in three deaths associated with cardiovascular disease (www.heartstats.org). General practitioners need an accurate and reliable tool to help them identify patients at high risk of having a cardiovascular event. Numerous multivariable risk scores have been developed to estimate a patient's 10 year risk of cardiovascular disease based on certain key known risk factors,^{1,2} including the Framingham risk score³ and the Reynolds risk score,⁴ both developed using patient data from US, the SCORE system using patients from multiple European countries,⁵ and ASSIGN using patients from Scotland.⁶ In the United Kingdom, until recently the National Institute for Health and Clinical Excellence (NICE) recommended use of the long established Framingham equation to inform a patient treatment plan about cardiovascular risk.³ NICE have now ceased recommending any single risk score equation, leaving healthcare professionals to choose the tool they consider to be most appropriate.⁷

The QRISK1 risk score, derived by the QRESEARCH organisation of the University of Nottingham, is a model to predict the 10 year risk of developing cardiovascular disease. It was developed⁸ and validated^{9,10} on large general practice databases in the United Kingdom using data from three million patients with 17.5 million person years of observation. QRISK1 includes traditional risk factors such as age, sex, systolic blood pressure, smoking status, and total serum cholesterol:high density lipoprotein ratio that are included in the long established Framingham equations, but it also includes body mass index, family history of cardiovascular disease, social deprivation (Townsend score), and use of antihypertensive treatment. Performance data comparing QRISK1 with the Framingham equation indicated that QRISK1 is a more accurate tool to predict the development of cardiovascular disease in the United Kingdom.^{9,11}

QRISK2, the successor to QRISK1, is a new multivariable risk score that contains all the risk factors that are in QRISK1 but also includes self assigned ethnicity

Summary of risk factors in QRISK1, QRISK2, and Framingham equation

QRISK1

Age (continuous)
 Ratio of total serum cholesterol:high density lipoprotein (continuous)
 Systolic blood pressure (continuous)
 Smoking status (current smoker/non-smoker (including former smoker))
 Body mass index (continuous)
 Family history of coronary heart disease in first degree relative under 60 years (yes/no)
 Townsend deprivation score (output area level 2001 census data evaluated as continuous variable)
 Receiving treatment for blood pressure at baseline (at least one current prescription of at least one antihypertensive agent) (yes/no)
 Systolic blood pressure × Receiving treatment for blood pressure at baseline

QRISK2

Age (continuous)
 Ratio of total serum cholesterol:high density lipoprotein (continuous)
 Systolic blood pressure (continuous)
 Smoking status (current smoker/non-smoker (including former smoker))
 Body mass index (continuous)
 Family history of coronary heart disease in first degree relative under 60 years (yes/no)
 Townsend deprivation score (output area level 2001 census data evaluated as continuous variable)
 Treated hypertension (diagnosis of hypertension and at least one current prescription of at least one antihypertensive agent) (yes/no)
 Self assigned ethnicity (white (or not recorded)/Indian/Pakistani/Bangladeshi/other Asian/black African/black Caribbean/other (including mixed))
 Type 2 diabetes (yes/no)
 Rheumatoid arthritis (yes/no)
 Atrial fibrillation (yes/no)
 Renal disease (yes/no)
 Age × body mass index
 Age × Townsend score
 Age × systolic blood pressure
 Age × family history of cardiovascular disease
 Age × smoking current
 Age × treated hypertension
 Age × type 2 diabetes
 Age × atrial fibrillation

Framingham equation (version recommended by NICE)

Age (continuous)
 Ratio of total serum cholesterol/high density lipoprotein (continuous)
 Systolic blood pressure (continuous)
 Smoking status (current smoker (or quit within last year)/non-smoker)
 Sex (male/female)
 Left ventricular hypertrophy (yes/no)
 Type 2 diabetes (yes/no)
 Age × type 2 diabetes
 Left ventricular hypertrophy × age
 Age × sex

and conditions associated with cardiovascular risk (including diagnosed type 2 diabetes, treated hypertension, rheumatoid arthritis, renal disease, and atrial fibrillation).¹² QRISK2 also contains interactions between age and Townsend score, body mass index, systolic blood pressure, family history, smoking status, treated hypertension, diagnosis of type 2 diabetes, and atrial fibrillation (see box for variables included in QRISK1, QRISK2 and NICE Framingham). All continuous risk factors were carefully handled and kept continuous throughout the model building process, fractional polynomials were used to model nonlinear risk relations where appropriate.¹³ Technical details of the actual QRISK2 model can be found on the QRESEARCH website (www.qresearch.org).

After the development of a new multivariable risk score, it needs to be subjected to external validation.^{14,15} The performance of a risk score is typically overestimated in the original data used to develop the score.¹⁴ External validation is a crucial step to provide sufficient evidence about the performance of the risk score on a cohort not used in the score's development.^{2,14-16} It is disappointing, however, that most risk scores that are published fail to undergo this extra step. Studies that develop and validate risk scores are far too often of low quality, poorly designed, use inappropriate statistical techniques, and are based on small selective cohorts with insufficient numbers of events.^{14,15,17-19} Furthermore, it is likely that many of the risk scores published each year are opportunistically produced to maximise the output from a clinical study for which developing a risk score was not declared a priori before collecting the data.²⁰ Guiding efficient clinical decision making requires that the assessment of risk be accurate, yet many risk scores in use have not adequately shown this essential quality. Low quality risk scores may be used clinically to stratify patients into risk groups or as inclusion or exclusion criteria for randomised controlled trials.

This article describes the results from an independent and external validation of QRISK2 and compares the performance of QRISK2 against QRISK1⁸ and an adjusted version of the Framingham equation³ previously recommended by NICE.²¹

METHODS

Study population

Study participants were patients registered between 1 January 1993 and 20 June 2008 and recorded on the THIN database (www.thin-uk.com). 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 74 years or over, had missing Townsend scores (social deprivation), or were prescribed statins at baseline.

Cardiovascular disease outcomes

The primary outcome measure was the first diagnosis of cardiovascular disease (myocardial infarction, angina, coronary heart disease, stroke, and transient

Table 1 | Characteristics of patients from THIN database who were included in study. Values are numbers (percentages) unless stated otherwise

	Women	Men
Total patients	797 373 (50.37)	785 733 (49.63)
Total person years of observation	4 832 294	4 567 097
Median (IQR) age (years)	49 (41-59)	48 (40-57)
Country:		
England	688 015 (86.29)	678 375 (86.34)
Scotland	43 276 (2.77)	42 471 (5.41)
Wales	44 033 (5.43)	44 065 (5.61)
Northern Ireland	22 049 (5.52)	20 822 (2.65)
Ethnicity:		
White or not recorded	780 117 (97.84)	769 409 (97.92)
Indian	4007 (0.50)	3928 (0.50)
Pakistani	1138 (0.15)	1181 (0.15)
Bangladeshi	356 (0.04)	447 (0.06)
Other Asian	1953 (0.24)	1900 (0.24)
Black Caribbean	2270 (0.28)	1762 (0.22)
Black African	2921 (0.37)	2728 (0.35)
Chinese	782 (0.10)	615 (0.08)
Other including mixed	3829 (0.48)	3763 (0.48)
Mean (SD) systolic blood pressure (mm HG)	131.54 (20.73)	135.22 (19.01)
Mean (SD) total serum cholesterol concentration (mmol/l)	5.76 (1.16)	5.58 (1.12)
Mean (SD) HDL cholesterol (mmol/l)	1.58 (0.43)	1.31 (0.35)
Mean (SD) total serum cholesterol:HDL ratio	3.91 (1.25)	4.54 (1.38)
Mean (SD) body mass index (kg/m ²)	26.27 (4.99)	26.71 (4.11)
Positive family history of coronary heart disease	38 562 (4.84)	31 744 (4.04)
Current smoker	180 774 (22.67)	215 650 (27.45)
Treated hypertension	50 758 (6.37)	37 529 (4.78)
Type 2 diabetes	13 800 (1.73)	18 470 (2.35)
Rheumatoid arthritis	8479 (1.06)	3555 (0.45)
Atrial fibrillation	3035 (0.38)	4937 (0.63)
Chronic kidney disease	1187 (0.15)	1087 (0.14)
Deprivation index (Townsend score) fifth:		
1 (most affluent)	219 287 (27.50)	212 487 (27.04)
2	183 335 (22.99)	175 250 (22.30)
3	165 391 (20.74)	161 349 (20.53)
4	138 226 (17.34)	138 201 (17.59)
5 (most deprived)	91 134 (11.43)	98 446 (12.53)
Incident CVD events in 10 years	29 057 (3.64)	42 408 (5.4)
Observed 10 year risk (95% CI) of CVD events	5.89 (5.81 to 5.96)	9.00 (8.90 to 9.10)

HDL=high density lipoprotein. CVD=cardiovascular disease.

ischaemic stroke) recorded on the general practice clinical computer system.

Statistical analysis

Ten-year estimated cardiovascular disease risk for every patient in the THIN cohort was calculated using QRISK2. Observed 10-year cardiovascular disease risks were obtained using the method of Kaplan-Meier. Multiple imputation was used to replace missing values for body mass index, systolic blood pressure, total serum cholesterol:high density lipoprotein ratio, and smoking status. Multiple imputation is a powerful technique that offers substantial improvements over the biased and flawed value replacement

approaches based on complete cases or cases matched for age and sex.^{22,23} It involves creating multiple copies of the data and imputing the missing values with sensible values randomly selected from their predicted distribution. We used the *MICE* (Multivariate Imputation by Chained Equations) library in *R* software to create 20 imputed datasets and then combined the results from analyses on each of the imputed datasets to produce estimates and confidence intervals that incorporate the uncertainty of imputed values.

Predictive performance of QRISK2 for the THIN cohort was assessed by examining measures of calibration and discrimination. Calibration refers to how closely the predicted 10 year risk of cardiovascular disease agrees with the observed 10 year 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 risk of cardiovascular disease separately for men and for women. Calibration of the risk score predictions was assessed by plotting observed proportions versus predicted probabilities. The Brier score for censored survival data was also calculated,²⁴ 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 score to differentiate between patients who experience a cardiovascular event during the study and those who do not. This measure is quantified by calculating the area under the receiver operating characteristics curve (AUROC) statistic; a value of 0.5 represents chance, and 1 represents perfect discrimination. We also calculated the *D* statistic²⁵ and *R*² statistic²⁶ (derived from the *D* statistic), which 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. Scaled rectangular diagrams are presented to illustrate the discrimination performance of QRISK2 and NICE Framingham.²⁷

Although predicted risk varies across a continuum, clinical decisions require creation of risk groups. An important aspect, therefore, when considering adopting a new risk prediction rule is the classification of patients into high and low risk and the number of patients who would be reclassified to a different risk category when compared with the standard means of risk prediction²⁸ (here the NICE Framingham equation). Patients were identified as high risk if their 10 year predicted cardiovascular disease risk was $\geq 20\%$, as per the guidelines set out by NICE.²¹

We compared the performance of QRISK2 with its predecessor QRISK1⁸ and with a modified version of the Framingham equation³ recommended by NICE. There is an increased risk of developing cardiovascular disease in people with a family history of premature disease, and risk varies between ethnic groups in the United Kingdom. Until recently, the approach recommended by NICE was to apply adjustment factors to

Table 2 Crude and age adjusted incidence rates (per 1000 person years) for cardiovascular disease by sex, age, country, ethnicity, and social deprivation

	Women				Men			
	Total person years	No of incident cases	Crude incidence rate (95% CI)	Age standardised rates (95% CI)	Total person years	No of incident cases	Crude incidence rate (95% CI)	Age standardised rates (95% CI)
Total	4 832 294	29 057	6.01 (5.94 to 6.08)	—	4 567 097	42 408	9.29 (9.20 to 9.37)	—
Age (years):								
35-44	1 664 619	1965	1.18 (1.13 to 1.23)	—	1 718 707	3999	2.33 (2.26 to 2.40)	—
45-54	1 414 936	5029	3.55 (3.46 to 3.65)	—	1 396 812	10 013	7.17 (7.03 to 7.31)	—
55-64	1 033 376	8936	8.65 (8.47 to 8.83)	—	924 993	14 284	15.44 (15.19 to 15.70)	—
65-74	719 363	13 127	18.25 (17.94 to 18.56)	—	526 585	14 112	26.80 (26.36 to 27.24)	—
Country:								
England	4 168 032	24 501	5.88 (5.80 to 5.95)	5.75 (5.68 to 5.82)	3 943 200	36 125	9.16 (9.07 to 9.26)	9.06 (8.97 to 9.16)
Scotland	265 125.6	2026	7.64 (7.31 to 7.98)	7.97 (7.63 to 8.33)	248 023	2652	10.69 (10.29 to 11.11)	11.15 (10.73 to 11.59)
Wales	258 111.9	1613	6.25 (5.95 to 6.56)	6.11 (5.81 to 6.41)	247 922	2346	9.46 (9.08 to 9.85)	9.21 (8.84 to 9.59)
Northern Ireland	141 025.1	917	6.50 (6.09 to 6.94)	6.77 (6.33 to 7.22)	127 951	1285	10.40 (9.50 to 10.61)	10.42 (9.85 to 11.01)
Ethnic group:								
White or not recorded	4 766 861	28 625	6.00 (5.94 to 6.07)	5.89 (5.82 to 5.96)	4 509 833	41 856	9.28 (9.19 to 9.37)	9.20 (9.11 to 9.29)
Indian	19 175	150	7.82 (6.62 to 9.18)	9.33 (7.86 to 10.99)	16 915	219	12.95 (11.29 to 14.78)	13.62 (11.87 to 15.56)
Pakistani	4376	45	10.28 (7.50 to 13.76)	13.22 (9.47 to 17.95)	3936	68	17.28 (13.42 to 21.90)	19.55 (15.04 to 24.99)
Bangladeshi	1236	11	8.90 (4.44 to 15.92)	8.66 (4.30 to 15.57)	1326	16	12.07 (6.89 to 19.60)	16.81 (9.43 to 27.67)
Other Asian	6311	35	5.55 (3.86 to 7.71)	6.67 (4.58 to 9.38)	5745	64	11.14 (8.58 to 14.23)	13.89 (10.60 to 17.90)
Black Caribbean	11 238	90	8.01 (6.44 to 9.84)	8.94 (7.16 to 11.02)	8150	64	7.85 (6.05 to 10.03)	6.93 (5.31 to 8.88)
Black African	8102	20	2.47 (1.51 to 3.81)	4.26 (2.55 to 6.69)	7476	22	2.94 (1.84 to 4.46)	4.44 (2.66 to 6.96)
Chinese	2399	9	3.75 (1.71 to 7.12)	6.41 (2.75 to 12.68)	1787	10	5.60 (2.68 to 10.29)	6.42 (3.05 to 11.88)
Other	12 596	72	5.72 (4.47 to 7.20)	7.33 (5.69 to 9.28)	11 929	89	7.46 (5.99 to 9.18)	8.94 (7.14 to 11.05)
Deprivation index (Townsend score) fifth:								
1 (most affluent)	1 387 251	6078	4.38 (4.27 to 4.49)	4.52 (4.41 to 4.64)	1 303 775	10 334	7.93 (7.77 to 8.08)	7.84 (7.69 to 7.99)
2	1 140 383	6168	5.41 (5.27 to 5.55)	5.23 (5.10 to 5.37)	1 053 009	9542	9.06 (8.88 to 9.25)	8.65 (8.48 to 8.83)
3	999 210	6183	6.19 (6.03 to 6.34)	6.04 (5.89 to 6.20)	934 476	8851	9.47 (9.28 to 9.67)	9.48 (9.29 to 9.68)
4	801 680	6060	7.56 (7.37 to 7.75)	7.30 (7.11 to 7.48)	762 139	7841	10.29 (10.06 to 10.52)	10.40 (10.18 to 10.64)
5 (most deprived)	503 770	4568	9.07 (8.81 to 9.33)	8.81 (8.55 to 9.07)	513 698	5840	11.37 (11.08 to 11.66)	11.82 (11.52 to 12.13)

the Framingham equation for ethnicity and family history, which are not captured in the Framingham equation. The Framingham score is multiplied by 1.4 for South Asian men (no adjustment for South Asian women) and by 1.5 for people with a history of coronary heart disease in a first degree relative. For South Asian men with a family history of coronary heart disease in a first relative, both adjustments are applied. To date there has been no detailed published scientific evaluation and validation for the choice of adjustment factor.

All statistical analyses were carried out in *R* (version 2.9.1).²⁹

RESULTS

The THIN database system included 3 587 306 eligible patients registered at 382 practices in the United Kingdom. After sequentially excluding, as per the exclusion criteria, 625 454 patients who left the general practice before 1 January 1993, 1 107 612 aged <35 years or ≥74 years, 161 746 with missing Townsend scores, 86 693 with a prior diagnosis of cardiovascular disease, and 23 019 with prior statin use, the analysed cohort consisted of 1 583 106 patients registered between 1 January 1993 and 20 June 2008 at 365 general

practices. The median follow-up was 6.2 years, and 241 859 (15.28%) patients were followed for at least 10 years. The 10 year observed risk of a cardiovascular event in men aged 35-74 years was 9.00% (95% confidence interval 8.90% to 9.10%) and in women was 5.89% (5.81% to 5.96%). Table 1 details the characteristics of eligible patients.

Complete data for all risk factors (ratio of total serum cholesterol to high density lipoprotein cholesterol, systolic blood pressure, body mass index, and smoking status) considered were available for 18.4% of women (n=146 651) and 16.0% of men (n=125 978). Most patients (n=1 066 926 (67.4%)) had none or only one missing risk factor. There were markedly high levels of missing data for the total serum cholesterol:high density lipoprotein ratio (74.7% for women and 74.6% for men). For other risk factors, the levels of missing data were, for body mass index, 19.4% for women, 28.5% for men; for systolic blood pressure, 7.4% for women, 16.5% for men; and for smoking status, 20.1% for women, 29.2% for men.

Table 2 shows the incidence rates (per 1000 person years) for cardiovascular disease by age, sex, region of UK, ethnicity, and deprivation index. In total there were 71 465 incident cases of cardiovascular disease

Table 3 | Discrimination and model performance statistics for QRISK2, QRISK1, and NICE Framingham (version of the Framingham equation recommended by NICE) in estimating 10-year risk of a cardiovascular event in the THIN cohort

	QRISK2	QRISK1	NICE Framingham
Women			
AUROC statistic	0.801	0.799	0.774
<i>D</i> statistic (95% CI)	1.66 (1.56 to 1.76)	1.61 (1.50 to 1.71)	1.47 (1.29 to 1.64)
<i>R</i> ² statistic (95% CI)	39.5 (36.6 to 42.4)	38.2 (35.1 to 41.3)	33.8 (28.5 to 39.2)
Brier score* (95% CI)	0.052 (0.050 to 0.054)	0.052 (0.050 to 0.054)	0.054 (0.051 to 0.057)
Men			
AUROC statistic	0.773	0.771	0.750
<i>D</i> statistic (95% CI)	1.45 (1.31 to 1.59)	1.42 (1.28 to 1.55)	1.30 (1.12 to 1.48)
<i>R</i> ² statistic (95% CI)	33.3 (28.9 to 37.8)	32.3 (28.3 to 36.4)	28.7 (23.1 to 34.3)
Brier score* (95% CI)	0.076 (0.074 to 0.078)	0.076 (0.074 to 0.079)	0.082 (0.079 to 0.085)

AUROC = area under the receiver operating characteristics curve.
*Lower score indicates better accuracy of risk estimates.

during the study period from 9.4 million person years of observation. The age adjusted incidence rate (per 1000 person years) for cardiovascular disease was highest in Scottish women and men (7.97 and 12.04) and lowest in English women and men (5.75 and 9.89).

The incidence of cardiovascular disease varied widely between different ethnic groups. The age standardised rates (per 1000 person years) for the white reference group were 5.89 for women and 9.20 for men. The highest rates were among the South Asian groups—for example, in Pakistani women the rate was 13.22 (per 1000 person years) and in Pakistani men it was 19.55.

Discrimination and calibration

For an accurate risk score the predicted and observed risks will agree. Fig 1 shows calibration plots for the three risk scores. Both QRISK2 and its predecessor, QRISK1, show much better agreement between observed risk and predicted risk grouped by tenth of risk compared with the NICE adjusted Framingham equation.

Table 3 presents discrimination and calibration performance data for QRISK2, QRISK1, and the NICE adjusted Framingham equation. The *R*² statistic (percentage of explained variation) is approximately 5% higher for QRISK2 in both men and women compared with the NICE adjusted Framingham equation. The difference in *R*² between QRISK2 and QRISK1 is less pronounced, with QRISK2 explaining around only 1% more of the variation in outcome in both men and women. The *D* discrimination statistic, where a higher score represents better discrimination, is higher for QRISK2 in both men and women compared with the NICE adjusted Framingham equation. As with the *R*² statistic, the difference in the *D* statistics between QRISK2 and QRISK1 is small. The Brier score (adjusted for censoring), a measure of prediction accuracy, was lower for QRISK2 in both men and women compared with the NICE adjusted Framingham equation. There was no difference between QRISK2 and QRISK1 in Brier score (0.076).

Risk classification

Using a 20% threshold for high risk of having a cardiovascular event, we calculated how many patients would be reclassified from low risk to high risk (and vice versa) using QRISK2 compared with the NICE adjusted version of Framingham.

In total, 90 823 male patients (11.6%) would be reclassified, with 1.8% (11 231) upgraded from low risk with NICE Framingham to high risk with QRISK2 (table 4). The observed risk in these patients was 20.02% (95% confidence interval 16.98% to 23.06%). The average predicted risk with the NICE Framingham equation was 14.73% whereas it was 24.08% with QRISK2. Nearly half the patients assessed as high risk with NICE Framingham (79 592 (45%)) would be downgraded to low risk with QRISK2. The observed risk in these patients was 14.00% (12.28% to 15.72%), compared with a mean predicted risk of 25.14% with NICE Framingham and 14.98% with QRISK2.

Similarly, 41 126 female patients (5.2%) would be reclassified, with 15 748 upgraded from low risk with NICE Framingham to high risk with QRISK2 (table 5). The mean observed risk in these patients was 20.07% (18.84% to 21.30%). The average predicted risk with the NICE Framingham equation was 15.19% and with QRISK2 was 23.70%. Likewise, 25 478 patients would be downgraded from high risk with NICE Framingham to low risk with QRISK2, with a mean observed risk of 13.36% (10.72% to 16.00%). The corresponding mean predicted risk was 24.24% with NICE Framingham and 15.32% with QRISK2.

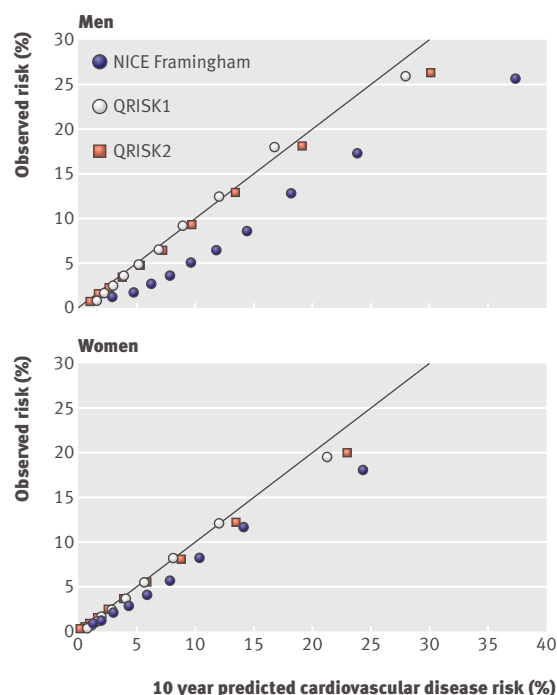


Fig 1 | Plot of observed versus predicted risks of cardiovascular disease for QRISK2, QRISK1, and the NICE adjusted Framingham equation.

Table 4 | Comparison of QRISK2 and NICE Framingham (version of the Framingham equation recommended by NICE) in classification of men in the THIN cohort into low or high 10-year risk of cardiovascular events and observed and predicted risk

	QRISK2		Total	No (%) of men reclassified
	Low risk (<20%)	High risk (≥20%)		
NICE Framingham, low risk (<20%)				
No (range) of men	599 179 (581 620–617 650)	11 231 (10 311–12 409)	610 410 (592 304–630 059)	
No (range) of events	18 258 (16 692–19 744)	1335 (1148–1507)	19 593 (17 845–21 251)	
Observed risk (95% CI)	5.32 (4.82 to 5.81)	20.02 (16.98 to 23.06)	5.57 (5.04 to 6.11)	11 231 (1.8)
Mean risk QRISK2 (95% CI)	5.28 (5.23 to 5.34)	24.08 (23.72 to 24.44)	5.62 (5.54 to 5.72)	
Mean risk NICE (95% CI)	8.64 (8.39 to 8.90)	14.73 (14.35 to 15.11)	8.76 (8.50 to 9.01)	
NICE Framingham, high risk (≥20%)				
No (range) of men	79 592 (69 790–88 184)	95 731 (85 884–10 5245)	175 323 (155 674–193 429)	
No (range) of events	7125 (6470–7549)	15 690 (14 410–17 014)	22 815 (21 157–24 563)	
Observed risk (95% CI)	14.00 (12.28 to 15.72)	24.52 (22.74 to 26.31)	19.76 (17.97 to 21.55)	79 592 (45.4)
Mean risk QRISK2 (95% CI)	14.98 (14.85 to 15.11)	28.87 (28.45 to 29.29)	22.57 (22.22 to 22.92)	
Mean risk NICE (95% CI)	25.14 (24.75 to 25.54)	35.47 (33.59 to 37.36)	30.79 (29.60 to 31.97)	
Total				
No (range) of men	678 771 (669 804–687 440)	106 962 (98 292–115 929)	785 733	
No (range) of events	25 383 (24 241–26 491)	17 025 (15 917–18 167)	42 408	
Observed risk (95% CI)	6.43 (6.12 to 6.74)	24.12 (22.68 to 25.56)	9.0 (8.9 to 9.1)	90 823 (11.6)
Mean risk QRISK2 (95% CI)	6.42 (6.21 to 6.63)	28.37 (27.89 to 28.85)	9.41 (8.88 to 9.94)	
Mean risk NICE (95% CI)	10.58 (10.02 to 11.14)	33.30 (31.21 to 35.39)	13.68 (12.63 to 14.72)	

Thus for both men and women, the mean predicted risks in patients reclassified from high to low risk (and vice versa) with QRISK2 were more accurate compared with the mean observed risk than mean predicted risks with NICE Framingham (see online appendix for low, intermediate, and high risk classification). Those patients who were reclassified as high risk with QRISK2 from low risk with NICE Framingham tended to be older, have been treated for hypertension, have a family history of coronary heart disease, and have a diagnosis of type 2 diabetes, rheumatoid arthritis, or atrial fibrillation.

The proportion of men and women classified as high risk by QRISK2 and NICE Framingham who had a subsequent cardiovascular event are displayed in fig 2, a scaled rectangular diagram. At current recommended treatment thresholds of 20%, the figure shows the modest discrimination performance of both QRISK2 and the NICE Framingham model. With QRISK2, 14% of the male cohort would be identified as being at high risk and would capture 40% of the cardiovascular events; NICE Framingham would identify 22% of the male cohort and 54% of the cardiovascular events. Similarly, for women, QRISK2 would identify 6% of the cohort as being at high risk and 26% of the cardiovascular events, whereas NICE Framingham would identify 7% of the cohort as high risk and 26% of all cardiovascular events.

Incidence of cardiovascular events in high risk groups

Using the 20% threshold to identify high risk patients, QRISK2 identified a group of patients at a higher risk of cardiovascular events than those identified with NICE Framingham. The incidence rate of cardiovascular events among men designated high risk with

QRISK2 was 27.8 per 1000 person years (95% confidence interval 27.4 to 28.2), whereas it was 21.9 (21.6 to 22.2) with NICE Framingham and 24.8 (22.8 to 26.9) with QRISK1. For women, the incidence rate of cardiovascular events in those designated high risk was 24.3 (23.8 to 24.9) with QRISK2, 20.6 (20.1 to 21.0) with NICE Framingham, and 21.8 (18.9 to 24.6) with QRISK1. Table 6 shows regional variations in the incidence of cardiovascular events in high risk groups.

DISCUSSION

Principal findings

We independently evaluated the performance of the QRISK2 risk score, in comparison with the risk prediction approach (NICE Framingham) that was until recently recommended by the National Institute for Health and Clinical Excellence in the United Kingdom, for predicting 10-year cardiovascular disease in

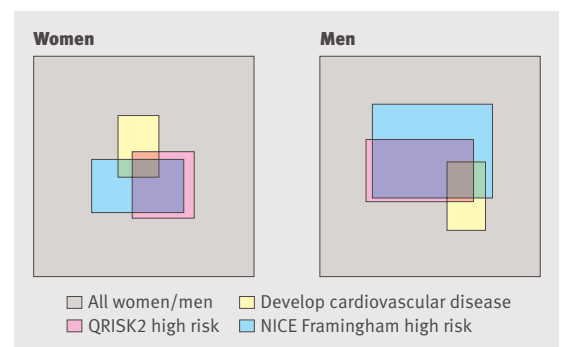


Fig 2 | Proportion of men and women classified as high 10-year risk of cardiovascular events (≥20%) by QRISK2 and the NICE version of the Framingham equation who also had a subsequent cardiovascular event

Table 5 | Comparison of QRISK2 and NICE Framingham (version of the Framingham equation recommended by NICE) in classification of women in the THIN cohort into low or high 10-year risk of cardiovascular events and observed and predicted risk

	QRISK2		Total	No (%) of women reclassified
	Low risk (<20%)	High risk (≥20%)		
NICE Framingham, low risk (<20%)				
No (range) of women	722 367 (716 767–729 126)	15 748 (14 802–16 926)	738 115 (733 041–744 762)	
No (range) of events	19 242 (18 390–19 865)	2141 (2063–2260)	21 383 (20 513–21 962)	
Observed risk (95% CI)	4.39 (4.15 to 4.63)	20.07 (18.84 to 21.30)	4.74 (4.51 to 4.97)	15 748 (2.1)
Mean risk QRISK2 (95% CI)	4.48 (4.38 to 4.58)	23.70 (23.61 to 23.79)	4.89 (4.76 to 5.03)	
Mean risk NICE (95% CI)	5.56 (5.38 to 5.74)	15.19 (14.96 to 15.41)	5.76 (5.57 to 5.96)	
NICE Framingham, high risk (≥20%)				
No (range) of women	25 478 (22 188–28 644)	33 780 (30 423–26 205)	59 258 (52 611–64 332)	
No (range) of events	2346 (2113–2584)	5328 (4888–6000)	7674 (7095–8544)	
Observed risk (95% CI)	13.36 (10.72 to 16.00)	22.36 (18.94 to 25.78)	18.53 (15.46 to 21.60)	25 478 (43.0)
Mean risk QRISK2 (95% CI)	15.32 (15.15 to 15.48)	28.19 (27.79 to 28.59)	22.66 (22.05 to 23.26)	
Mean risk NICE (95% CI)	24.24 (23.89 to 24.58)	30.97 (30.17 to 31.76)	28.07 (27.51 to 28.64)	
Total				
No (range) of women	747 845 (744 747–751 314)	49 528 (46 059–52 626)	797 373	
No (range) of events	21 588 (20 934–22 010)	7469 (7047–8123)	29 057	
Observed risk (95% CI)	4.73 (4.56 to 4.89)	21.67 (19.39 to 23.94)	5.89 (5.81 to 5.96)	41 126 (5.2)
Mean risk QRISK2 (95% CI)	4.85 (4.72 to 4.98)	26.76 (26.46 to 27.06)	6.21 (6.01 to 6.42)	
Mean risk NICE (95% CI)	6.19 (5.95 to 6.43)	25.95 (24.98 to 26.91)	7.42 (7.10 to 7.75)	

an independent UK cohort of patients from general practice. In this large cohort of 1.6 million patients, the NICE Framingham equation had inferior performance compared with either QRISK2 or its predecessor, QRISK1. The NICE Framingham risk score over-predicted 10-year risk of cardiovascular disease, compared with the more accurate QRISK2 and QRISK1 scores. The difference in performance between QRISK2 and QRISK1 was slight, with QRISK2 marginally outperforming QRISK1. QRISK2 includes five extra risk factors (self assigned ethnicity, type 2 diabetes, rheumatoid arthritis, atrial fibrillation, and chronic renal disease) as well as eight interaction terms. These additional risk factors do not require any considerable effort to collect; ethnicity would be recorded as white if this item were missing or not recorded. Absence of a recorded diagnosis of type 2 diabetes, rheumatoid arthritis, atrial fibrillation, or chronic renal disease is assumed to indicate that the person did not have that factor.

The development cohort (1.5 million patients), internal validation cohort (0.75 million patients), and this external validation cohort (1.6 million patients) included in total nearly 3.9 million patients (about 14% of the UK population aged between 34 and 74 years) with a total of 21 million person years of observation and 211 580 recorded cardiovascular events. This constitutes one of the largest groups of patients used to develop and externally validate a risk score before its recommendation and implementation in clinical practice, and, given the source and combined size of all cohorts, it is likely to be a fair representation of the population for which the equations are to be used.

The superior performance of the QRISK risk scores is not surprising as both QRISK risk scores were developed (and internally and externally validated) on large cohorts of general practice patients in the United Kingdom, the population for which the risk predictions were targeted and designed. This includes accounting for social deprivation, family history of coronary heart disease, and ethnicity, all known to increase the risk of developing cardiovascular disease. The Framingham score, by contrast, was developed on a comparatively small ($n=5573$), homogeneous white, though treatment-naive, sample from a single town in the US between 1968 and 1975. We evaluated the Framingham risk score with the NICE designated adjustment factors for family history and, for men, being of South Asian origin.

Arguably, QRISK2 would be more aptly compared with a Framingham equation recalibrated for the UK population, but we compared it with the Framingham risk score recommended by NICE, which is without reference to recalibration. Furthermore, the NICE Framingham is a single equation with a sex coefficient, whereas the QRISK approach has separate equations for men and women. Separate equations permit risk factors to be weighted differently for men and women.

Although some have welcomed the introduction of the QRISK equations^{30,31} and the debate about improving risk prediction, others have cautioned against their use for two main reasons; (a) the cohort used to develop both QRISK equations included patients who were not treatment naive, unlike the Framingham cohort,³² and (b) there are large amounts of missing cholesterol data.³³ With regard to the cohorts including patients who may have started additional treatments, it is now neither practically nor ethically

Table 6 | Incidence rates of cardiovascular disease (per 1000 person years) in high risk groups* identified by QRISK2 and NICE Framingham (version of the Framingham equation recommended by NICE) across regions of the United Kingdom

Country	Women		Men	
	QRISK2	NICE Framingham	QRISK2	NICE Framingham
England	23.5 (22.9 to 24.1)	18.9 (18.5 to 19.4)	25.4 (25.0 to 25.8)	19.5 (19.2 to 19.7)
Northern Ireland	24.7 (21.5 to 28.3)	21.2 (18.4 to 24.2)	27.3 (24.8 to 29.9)	21.5 (19.8 to 23.2)
Scotland	27.5 (25.2 to 30.0)	23.5 (21.5 to 25.6)	28.2 (26.4 to 30.0)	21.7 (20.5 to 23.0)
Wales	24.7 (22.4 to 27.2)	20.2 (18.3 to 22.2)	26.0 (24.4 to 27.7)	19.0 (18.0 to 20.1)

*High risk defined as $\geq 20\%$ predicted 10-year risk of cardiovascular disease.

possible to obtain large treatment-naive cohorts. Furthermore, although natural history is important, it is not clear that prognosis is best assessed from an untreated population.

Risk scores will inevitably become outdated with improvements in clinical outcomes and data recording and changes in population demographics. Thus, ensuring a risk score retains its usefulness to reflect current conditions is crucial.³⁴ QRISK2 will undergo annual updates to account for changes in population characteristics and improvements in data quality, with its most recent update having been released on 1 April 2010 (see www.qrisk.org for details of this). This entails re-fitting QRISK2 to the latest version of the QRESEARCH database to obtain updated regression coefficients.

Strengths and weaknesses

All the cohorts used in the development and validation of QRISK1 and QRISK2 had high levels of missing data for the total serum cholesterol:high density lipoprotein ratio, so it can be assumed to be a population feature. QRISK1 and QRISK2 were developed using established methods of multiple imputation for missing data to address this problem. Our external validation study also used multiple imputation, with 20 imputed datasets, to deal with the missing data. We note that problems of missing data in developing and validating risk scores are rarely considered when validating risk scores.³⁵⁻³⁸ Omitting patients with missing data in developing the risk score and conducting a complete-case analysis would introduce bias and

produce a poorly performing risk score. Even in validation, discarding missing observations will result in performance data that is biased, and so that practice should be avoided.^{37 38}

Conclusions

In this study, we have provided an independent and external validation of the QRISK2 risk score on a large cohort of patients in the United Kingdom. We have assessed the performance of QRISK2 against the NICE version of the Framingham equation and have provided evidence to support the use of QRISK2 in favour of the NICE 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.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: All authors have completed the Unified Competing Interest form at http://www.icmje.org/doi_disclosure.pdf (available on request from the corresponding author) and declare that GSC and DGA have no non-financial interests that may be relevant to the submitted work.

Ethical approval: Trent multicentre research ethics committee.

- 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.
- Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MSV, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009;119:2408-16.
- Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121(1 pt 2):293-8.
- Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds risk score. *JAMA* 2007;297:611-9.
- Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, de Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24:987-1003.
- 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.
- Mayor S. Doctors no longer have to use Framingham equation to assess heart disease risk, NICE says. *BMJ* 2010;340:c1774.
- 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.
- Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. *BMJ* 2009;339:b2584.
- 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.
- Collins GS, Altman DG. *Report to the Department of Health: independent validation of QRISK on the THIN database*. University of Oxford, 2008.
- 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.
- Royston P, Sauerbrei W. *Multivariable modelling: a pragmatic approach to fractional polynomials for continuous variables*. Wiley, 2008.
- Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453-73.
- Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
- McGeachan K, Macaskill P, Inwig L, Liew G, Wong TY. Assessing new biomarkers and predictive models for use in clinical practice: a clinician's guide. *Arch Intern Med* 2008;168:2304-10.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Cardiovascular risk prediction in the United Kingdom has until recently been based on a NICE adjusted version of the US Framingham model that has been shown to over-predict risk

QRISK2 was developed using a large cohort of UK patients and published in 2008

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

WHAT THIS STUDY ADDS

Independent evaluation of QRISK2 showed an improvement in performance over NICE Framingham in a large external cohort of UK patients

QRISK2 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 NICE Framingham

- 17 Omar RZ, Ambler G, Royston P, Eliahoo J, Taylor KM. Cardiac surgery risk modeling for mortality: a review of current practice and suggestions for improvement. *Ann Thorac Surg* 2004;77:2232-7.
- 18 Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med* 2010;8:21.
- 19 Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med* 2010;8:20.
- 20 Hemmingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009;339:b4184.
- 21 NICE Guideline. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease (2008). Available at <http://guidance.nice.org.uk/CG67> (accessed 8 Oct 2009).
- 22 Janssen KJM, Vergouwe Y, Donders ART, Harrell FE, Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem* 2009;55:994-1001.
- 23 Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3-15.
- 24 Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med* 1999;18:2529-45.
- 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 Marshall R. Cardiovascular risk can be represented by scaled rectangle diagrams. *J Clin Epidemiol* 2009;62:998-1000.
- 28 Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med* 2008;149:751-60.
- 29 R Development Core Team. R: a language and environment for statistical computing. www.R-project.org.
- 30 Jackson R. Cardiovascular risk prediction: are we there yet? *Heart* 2008;94:1-3.
- 31 Jackson R, Marshall R, Kerr AJ, Riddell T, Wells S. QRISK or Framingham for predicting cardiovascular risk? *BMJ* 2009;339:b2673.
- 32 Liew SM, Glasziou P. QRISK validation and evaluation. QRISK may be less useful. *BMJ* 2009;339:b3485.
- 33 Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk. A review for clinicians. *J Am Coll Cardiol* 2009;54:1209-27.
- 34 Tsang VT, Brown KL, Synnergren MJ, Kang N, de Leval MR, Gallivan S, et al. Monitoring risk-adjusted outcomes in congenital heart surgery: does the appropriateness of a risk model change with time? *Ann Thorac Surg* 2009;87:584-8.
- 35 Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003;56:28-37.
- 36 Janssen KJM, Vergouwe Y, Donders ART, Harrell FE, Chen Q, Grobbee DE, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem* 2009;55:994-1001.
- 37 Marshall A, Altman DG, Holder R, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Meth* 2009;9:57.
- 38 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;63:205-14.

Accepted: 23 April 2010