

# Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study

Julia Hippisley-Cox, professor of clinical epidemiology and general practice,<sup>1</sup> Carol Coupland, senior lecturer in medical statistics,<sup>1</sup> Yana Vinogradova, research fellow in medical statistics,<sup>1</sup> John Robson, senior lecturer in general practice,<sup>2</sup> Margaret May, research fellow in medical statistics,<sup>3</sup> Peter Brindle research and development strategy lead<sup>4</sup>

<sup>1</sup>Tower Building, University Park, Nottingham NG2 7RD

<sup>2</sup>Centre for Health Sciences, Queen Mary's School of Medicine and Dentistry, London

<sup>3</sup>Department of Social Medicine, University of Bristol

<sup>4</sup>Avon Primary Care Research Collaborative, Bristol Primary Care Trust

Correspondence to: J Hippisley-Cox  
julia.hippisley-cox@nottingham.ac.uk

doi:10.1136/bmj.39261.471806.55

## ABSTRACT

**Objective** To derive a new cardiovascular disease risk score (QRISK) for the United Kingdom and to validate its performance against the established Framingham cardiovascular disease algorithm and a newly developed Scottish score (ASSIGN).

**Design** Prospective open cohort study using routinely collected data from general practice.

**Setting** UK practices contributing to the QRESEARCH database.

**Participants** The derivation cohort consisted of 1.28 million patients, aged 35-74 years, registered at 318 practices between 1 January 1995 and 1 April 2007 and who were free of diabetes and existing cardiovascular disease. The validation cohort consisted of 0.61 million patients from 160 practices.

**Main outcome measures** First recorded diagnosis of cardiovascular disease (incident diagnosis between 1 January 1995 and 1 April 2007): myocardial infarction, coronary heart disease, stroke, and transient ischaemic attacks. Risk factors were age, sex, smoking status, systolic blood pressure, ratio of total serum cholesterol to high density lipoprotein, body mass index, family history of coronary heart disease in first degree relative aged less than 60, area measure of deprivation, and existing treatment with antihypertensive agent.

**Results** A cardiovascular disease risk algorithm (QRISK) was developed in the derivation cohort. In the validation cohort the observed 10 year risk of a cardiovascular event was 6.60% (95% confidence interval 6.48% to 6.72%) in women and 9.28% (9.14% to 9.43%) in men. Overall the Framingham algorithm over-predicted cardiovascular disease risk at 10 years by 35%, ASSIGN by 36%, and QRISK by 0.4%. Measures of discrimination tended to be higher for QRISK than for the Framingham algorithm and it was better calibrated to the UK population than either the Framingham or ASSIGN models. Using QRISK 8.5% of patients aged 35-74 are at high risk (20% risk or higher over 10 years) compared with 13% when using the Framingham algorithm and 14% when using ASSIGN. Using QRISK 34% of women and 73% of men aged 64-75 would be at high risk compared with 24% and 86%

according to the Framingham algorithm. UK estimates for 2005 based on QRISK give 3.2 million patients aged 35-74 at high risk, with the Framingham algorithm predicting 4.7 million and ASSIGN 5.1 million. Overall, 53 668 patients in the validation dataset (9% of the total) would be reclassified from high to low risk or vice versa using QRISK compared with the Framingham algorithm.

**Conclusion** QRISK performed at least as well as the Framingham model for discrimination and was better calibrated to the UK population than either the Framingham model or ASSIGN. QRISK is likely to provide more appropriate risk estimates to help identify high risk patients on the basis of age, sex, and social deprivation. It is therefore likely to be a more equitable tool to inform management decisions and help ensure treatments are directed towards those most likely to benefit. It includes additional variables which improve risk estimates for patients with a positive family history or those on antihypertensive treatment. However, since the validation was performed in a similar population to the population from which the algorithm was derived, it potentially has a "home advantage." Further validation in other populations is therefore required.

## INTRODUCTION

Cardiovascular disease is the leading cause of premature death and a major cause of disability in the United Kingdom.<sup>1</sup> Asymptomatic patients thought to be at high risk of cardiovascular disease need to be identified so they can be offered advice about lifestyle changes, such as smoking cessation, physical activity, and diet, about treatment to lower blood pressure and modify cholesterol levels, and about use of aspirin when appropriate.

Many guidelines recommend that the risk of cardiovascular disease is estimated by combining different risk factors into a numeric estimate of risk. A variety of risk calculators are available, as charts, tables, computer programs, and web based tools.<sup>2-4</sup> Equations derived from the American Framingham cohort study are the most widely used in the United Kingdom.<sup>5</sup>

Although the Framingham risk equations have been the most used method for many years they have major limitations. The Framingham cohort is almost entirely white and recalibration may be needed in more ethnically diverse populations.<sup>6</sup> The Framingham risk equations were developed during the peak incidence of cardiovascular disease in America. They perform well in similar populations but may over-estimate risk by up to 50% in contemporary northern European populations where the incidence of cardiovascular disease is lower.<sup>7</sup> The confidence intervals for estimates produced from the Framingham algorithm have been difficult to quantify and any estimate is uncertain for estimation of an individual's risk. The equations may also underestimate risk in some high risk subgroups, such as patients from deprived areas, potentially exacerbating health inequalities.<sup>8,9</sup> Lastly, the Framingham algorithm does not include factors such as social deprivation, body mass index, family history of cardiovascular disease, and current treatment with antihypertensives. The evidence supporting the utility of cardiovascular risk scores for primary prevention in the United Kingdom is scarce.<sup>10</sup>

In a major initiative to improve public health in the United Kingdom the National Institute for Health and Clinical Excellence has lowered the threshold for primary prevention with statins from a 10 year

cardiovascular disease risk of 40% to that of 20%.<sup>11,12</sup> We refer to a 10 year cardiovascular disease risk of 20% or more as high risk. Any systematic over-estimation of risk results in an excessive number of people being identified for treatment. Not only would this have a major impact on prescribing costs and NHS resources supporting lifelong treatment, it potentially also exposes patients to unnecessary treatments and hence possible side effects. It is important that this new public health programme includes social deprivation to reduce rather than to exacerbate existing social inequalities in cardiovascular disease and to target those at greatest risk.<sup>13</sup> In recognition of these problems Scotland has adopted ASSIGN equations that include social deprivation.<sup>14</sup>

For these reasons the underlying methods of risk prediction should be fit for purpose, targeting those likely to benefit most and avoiding inappropriate treatment of those at lower risk. The equation underpinning the estimate of cardiovascular disease risk needs to be revised and calibrated for the contemporary UK population, with an appropriate weighting for social deprivation and current treatment with antihypertensives. We therefore carried out a study to derive and validate a new cardiovascular disease risk score and compared this with the Framingham<sup>5</sup> and the new ASSIGN equations.<sup>15</sup>

**Table 1** | Computer recorded incidence of cardiovascular disease per 1000 person years with 95% confidence intervals in derivation and validation cohorts by age, sex, and deprivation fifth

	Derivation cohort			Validation cohort		
	No of incident cases of cardiovascular disease	Person years	Incidence rate per 1000 person years (95% CI)	No of incident cases of cardiovascular disease	Person years	Incidence rate per 1000 person years (95% CI)
No of women	27 828	4 238 309	6.57 (6.49 to 6.64)	13 107	2 025 193	6.47 (6.36 to 6.58)
Age group:						
35-44	1791	1 449 107	1.24 (1.18 to 1.29)	958	693 806	1.38 (1.30 to 1.47)
45-54	4861	1 249 132	3.89 (3.78 to 4.00)	2347	592 108	3.96 (3.81 to 4.13)
55-64	8597	878 314	9.79 (9.58 to 10.00)	3984	420 752	9.47 (9.18 to 9.77)
65-74	12 579	661 756	19.01 (18.68 to 19.34)	5818	318 527	18.27 (17.80 to 18.74)
Townsend fifth:						
First	4851	931 167	5.21 (5.07 to 5.36)	2324	440 671	5.27 (5.06 to 5.49)
Second	5134	921 744	5.57 (5.42 to 5.72)	2366	429 554	5.51 (5.29 to 5.73)
Third	5698	865 316	6.58 (6.42 to 6.76)	2559	410 564	6.23 (6.00 to 6.48)
Fourth	6187	804 175	7.69 (7.50 to 7.89)	2948	392 190	7.52 (7.25 to 7.79)
Fifth	5958	715 907	8.32 (8.11 to 8.54)	2910	352 215	8.26 (7.97 to 8.57)
No of men	37 843	4 001 632	9.46 (9.36 to 9.55)	17 705	1 909 991	9.27 (9.13 to 9.41)
Age group:						
35-44	3851	1 521 521	2.53 (2.45 to 2.61)	1865	738 492	2.53 (2.41 to 2.64)
45-54	9197	1 237 808	7.43 (7.28 to 7.58)	4238	584 812	7.25 (7.03 to 7.47)
55-64	12 591	777 769	16.19 (15.91 to 16.47)	5877	365 856	16.06 (15.66 to 16.48)
65-74	12 204	464 535	26.27 (25.81 to 26.74)	5725	220 831	25.92 (25.26 to 26.61)
Townsend fifth:						
First	7440	864 786	8.6 (8.41 to 8.80)	3496	404 789	8.64 (8.36 to 8.93)
Second	7797	856 868	9.1 (8.90 to 9.30)	3515	396 201	8.87 (8.58 to 9.17)
Third	7727	800 876	9.65 (9.44 to 9.87)	3566	382 090	9.33 (9.03 to 9.64)
Fourth	7671	755 090	10.16 (9.93 to 10.39)	3615	372 434	9.71 (9.40 to 10.03)
Fifth	7208	724 013	9.96 (9.73 to 10.19)	3513	354 477	9.91 (9.59 to 10.24)

## METHODS

Our prospective cohort study was carried out within a large UK primary care population. We used version 14 of the QRESEARCH database ([www.qresearch.org](http://www.qresearch.org)). This is a large, validated electronic database representative of primary care and containing the health records of 10 million patients over a 17 year period from 529 general practices using the EMIS computer system. It covers about 7% of the current UK population and the contributing practices are similar to over 8000 practices nationally for a wide variety of measures.<sup>16</sup> It has been used for a wide range of health service analyses and research studies, including studies of the incidence, risk, and treatment of cardiovascular disease.<sup>17-21</sup> The database contains area measures of ethnicity and deprivation evaluated at output area on the basis of the 2001 census and linked to every patient's record. The Townsend score<sup>22</sup> is a good area measure of material deprivation based on four variables (unemployment, overcrowding, non-car ownership, and non-home ownership) and has been widely used in medical research. It is available at output area, about 125 households. Most recently QRESEARCH has been linked to cause of death data from the Office for National Statistics enabling sensitivity analyses including certified cause of death. QRESEARCH is sufficiently large to allow modelling on two thirds of the database, with validation using the remaining third.

### Practices and cohorts

We included all the contributing practices in the United Kingdom once they had been using their EMIS system for at least a year. We randomly allocated two thirds of eligible practices to the derivation dataset, saving the remaining third for the validation dataset. The derivation dataset was used to derive a new risk

equation for estimating cardiovascular risk, validated using the validation dataset.

We identified an open cohort of patients aged 35-74 at the date of study entry, drawn from patients registered with eligible practices from 1 January 1995 to 1 April 2007. We excluded those with a diagnosis of cardiovascular disease or diabetes before their entry date. We also excluded temporary residents and patients with interrupted periods of registration with the practice. We also excluded 4% of patients who did not have a valid postcode related Townsend score.

For each patient we determined an entry date to the cohort analysis, which was the latest of the following dates: 35th birthday, date of registration with the practice, date on which the practice computer system was installed, and the beginning of the study period (1 January 1995). In addition we only included patients in the analysis once they had a minimum of one year's complete data in their medical record.

For each patient we determined the right censor date, which was the earliest date of the dates on which they developed the outcome of interest, the study period ended (1 April 2007), date of death, date of deregistration with the practice, or date of last upload of computerised data. We determined the person years at risk—the difference between the entry date and the right censor date. We used person years at risk as the denominator term for the incidence rates.

### Cardiovascular disease outcomes

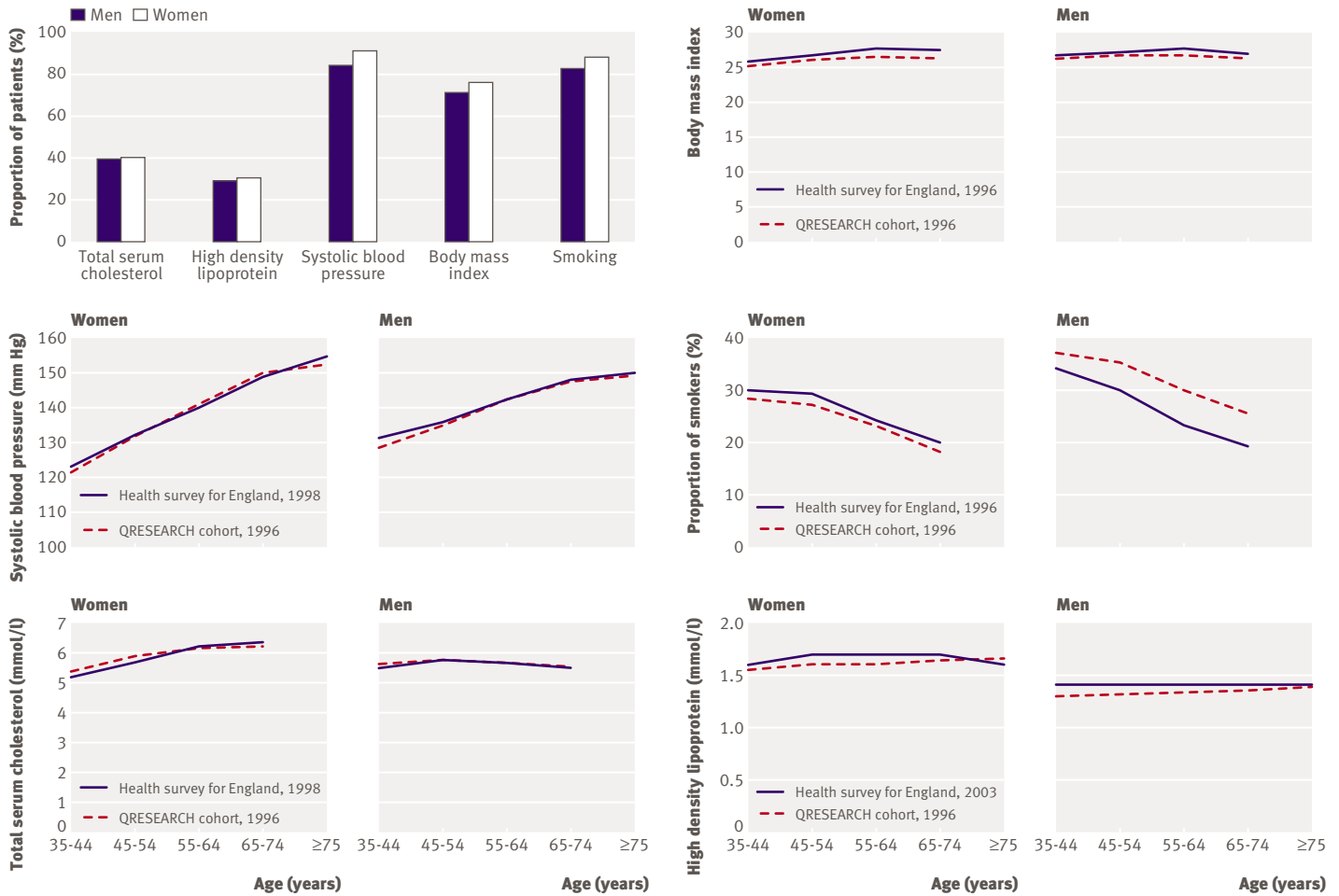
The primary outcome was the first recorded diagnosis of cardiovascular disease on the general practice's clinical computer system either before or at death occurring between 1 January 1995 and 1 April 2007. Our definition included myocardial infarction, coronary heart disease, stroke, and transient ischaemic attack.

**Table 2 | Baseline clinical characteristics for men and women aged 35-74 initially free from cardiovascular disease and diabetes in derivation and validation datasets. Values are means (standard deviations) unless stated otherwise**

Characteristics	Derivation cohort		Validation cohort	
	Men (n=636 753)	Women (n=646 421)	Men (n=305 140)	Women (n=309 413)
Median (interquartile range) age (years)	48 (40-57)	49 (41-59)	47 (40-57)	49 (41-59)
Median (interquartile range) Townsend score	-1.1 (-3.1-2.3)	-1.2 (-3.1-2.1)	-0.9 (-3.0-2.4)	-1.0 (-3.0-2.1)
Body mass index	26.5 (4.0)	26 (4.8)	26.5 (4.0)	26 (4.8)
Systolic blood pressure (mm Hg)	135.7 (19.6)	132.6 (21.5)	135.4 (19.7)	132.4 (21.6)
Total serum cholesterol level (mmol/l)	5.7 (1.1)	5.9 (1.2)	5.7 (1.1)	5.9 (1.2)
High density lipoprotein cholesterol level (mmol/l)	1.3 (0.4)	1.6 (0.4)	1.3 (0.4)	1.6 (0.4)
Ratio of total serum cholesterol to high density lipoprotein	4.6 (1.3)	4.0 (1.3)	4.7 (0.7)	4.0 (0.6)
No (%) of current smokers	179 144 (28.1)	149 089 (23.1)	85 834 (28.1)	70 879 (22.9)
No (%) with family history of cardiovascular disease*	57 708 (9.1)	78 442 (12.1)	28 610 (9.4)	38 852 (12.6)
Current treatment:				
Antihypertensives†	53 742 (8.4)	80 066 (12.4)	26 080 (8.5)	38 868 (12.6)
Angiotensin converting enzyme inhibitors†	15 969 (2.5)	16 101 (2.5)	7714 (2.5)	7787 (2.5)
β blockers†	25 084 (3.9)	38 536 (6.0)	12 109 (4.0)	18 720 (6.1)
Calcium channel blockers†	16 566 (2.6)	18 618 (2.9)	8019 (2.6)	9107 (2.9)
Thiazides†	15 476 (2.4)	31 919 (4.9)	7389 (2.4)	15 331 (5.0)

\*In first degree relative aged less than 60.

†At entry to cohort.



Comparison of mean values for variables by age and sex in the derivation cohort compared with health survey for England (year nearest to 1996 with risk factor measured). Mean high density lipoprotein levels from health survey were not available until 2003

We used nationally agreed definitions—that is, computer recorded Read codes—from the General Medical Services contract quality and outcomes framework.

We have used the general practitioner’s clinical diagnosis of cardiovascular disease for our main analysis. At the time of writing, however, patient level data for cause of death were available from the Office for National Statistics for the five years between 2002 and 2007. We were able to link these data at patient level within the general practices’ clinical computer systems and extract the resulting data onto the QRESEARCH database. We were therefore able to identify patients who had a primary or underlying certified cause of death as either coronary heart disease (codes I20-I25, international classification of diseases, 10th revision), stroke, or transient ischaemic attack and to use these data to examine the completeness of outcome data on QRESEARCH.

**Risk factors for cardiovascular disease**

We included the following risk factors in our analysis, using the value closest to the entry date to the cohort for each patient and imputing missing values when necessary: age (in single years); sex (men *v* women); smoking

status (current smoker, non-smoker—including former smoker); systolic blood pressure (continuous); ratio of total serum cholesterol to high density lipoprotein levels (continuous); left ventricular hypertrophy recorded in clinical records (yes or no); body mass index (continuous); family history of cardiovascular disease in a first degree relative aged less than 60 (yes or no); Townsend deprivation score (2001 census data at output area level evaluated as a continuous variable); percentage of South Asian residents at output areas (2001 census data evaluated at output areas as a continuous variable); current prescription of at least one anti-hypertensive—thiazide,  $\beta$  blocker, calcium channel blocker, or angiotensin converting enzyme inhibitor (yes or no).

**Model derivation and development**

We used the Cox proportional hazards model in the derivation dataset to estimate the coefficients associated with each potential risk factor for the first ever recorded diagnosis of cardiovascular disease for men and women separately. A priori we specified the variables we intended to include in the model on the basis of traditional risk scores (for example, Framingham

equation). We compared models using the Bayes information criterion. This is a likelihood measure in which lower values indicate better fit and in which a penalty is paid for increasing the number of variables. We examined the strength of the association between one unit increases in each continuous variable and we compared categories for other variables—for example, current smoking compared with non-smoking. Continuous variables were centred for analysis. We checked the assumptions of the proportional hazards model for each variable and tested for any non-linear relation between continuous independent variables and the outcome. We used fractional polynomials to model non-linear risk relations with continuous variables when appropriate.<sup>23,24</sup> Modelling using fractional polynomials is a flexible approach for modelling non-linear risk relations with continuous variables. We tested for interactions between systolic blood pressure and antihypertensive treatment and between smoking and deprivation.

Initially our models were fitted using patients without any missing data (complete case analysis), and the fractional polynomial terms were obtained from the complete case analysis using a multivariable approach. However, because patients with complete data might have a different health status and risk of cardiovascular disease compared with those with missing data, we fitted our principal models on the basis of multiple imputed datasets using Rubin's rules to combine effect estimates and estimate standard errors.<sup>24,25</sup> We used the ICE procedure in Stata<sup>26</sup> to obtain five imputed datasets.

We took the log of the hazard ratio for each of the risk factors (that is, the coefficients from the Cox regression) from the final model and used these as weights for the new cardiovascular disease risk equation. We then estimated each patient's probability of experiencing a cardiovascular event within 10 years by combining these weights with the characteristics of the patient and also using the baseline survivor function for all participants, as in other studies.<sup>15,27</sup> The baseline survivor function was estimated from the Cox regression model centred on the means of continuous risk factors and we extracted the value for 10 years' follow-up.

#### Validation of the new cardiovascular disease risk equation

Having obtained the risk equation in the derivation dataset (the QRISK score) we tested its performance (calibration and discrimination) in the validation dataset. We calculated the 10 year estimated cardiovascular disease risk for each patient in the validation dataset, replacing missing values for continuous variables with mean values obtained from the derivation dataset by five year age-sex bands and assuming patients were non-smokers when smoking status was not recorded.

To assess calibration (the degree of similarity between observed and predicted risks) we calculated the mean predicted risk of cardiovascular disease at 10 years and the observed risk at 10 years obtained using the 10 year Kaplan-Meier estimate. We then compared the ratio of the predicted to the observed cardiovascular disease risk for patients in the validation cohort in each tenth of predicted risk. We also

**Table 3 | Adjusted hazard ratios with 95% confidence intervals for preferred QRISK model for men and women aged 35-74**

Variables	Adjusted hazard ratio	Lower 95% confidence limit	Upper 95% confidence limit
<b>Women:</b>			
Log (age/10)	87.75	81.34	94.66
Ratio of total serum cholesterol to high density lipoprotein cholesterol levels	1.001	0.999	1.002
Body mass index (kg/m <sup>2</sup> )	1.015	1.013	1.018
Family history of premature cardiovascular disease	1.229	1.187	1.273
Smoking status (current smoker)	1.530	1.487	1.574
Townsend score of output area	1.035	1.031	1.038
Systolic blood pressure (mm Hg)	1.005	1.004	1.005
Receiving treatment for blood pressure at baseline	1.734	1.674	1.796
Interaction terms for systolic blood pressure×blood pressure treatment	0.996	0.995	0.997
<b>Men:</b>			
Log (age/10)	50.634	47.792	53.646
Ratio of total serum cholesterol to high density lipoprotein cholesterol levels	1.001	0.999	1.003
Body mass index (kg/m <sup>2</sup> )	1.022	1.019	1.025
Family history of premature cardiovascular disease	1.300	1.257	1.344
Smoking status (current smoker)	1.417	1.385	1.449
Townsend score of output area	1.017	1.014	1.020
Systolic blood pressure (mm Hg)	1.004	1.004	1.005
Receiving treatment for blood pressure at baseline	1.847	1.788	1.908
Interaction terms for systolic blood pressure×blood pressure treatment	0.993	0.992	0.995

Variables simultaneously adjusted for other variables in table. Continuous variables were centred. Hazard ratios correspond to increase of 1 unit for log (age/10), 1 mm Hg for systolic blood pressure, 1 kg/m<sup>2</sup> for body mass index, 1 unit for Townsend score, and 1 unit for ratio of total serum cholesterol to high density lipoprotein cholesterol levels.

**Table 4** | Ratio of predicted to observed risks of cardiovascular disease event at 10 years across tenths of estimated risks

Tenth*	QRISK			Framingham			ASSIGN		
	Predicted risk (%)	Observed risk (%)	Ratio	Predicted risk (%)	Observed risk (%)	Ratio	Predicted risk (%)	Observed risk (%)	Ratio
Women:									
First	0.71	0.44	1.61	0.59	0.53	1.12	1.31	0.34	3.81
Second	1.06	0.95	1.12	1.17	0.96	1.22	1.83	0.91	2.00
Third	1.49	1.45	1.03	1.95	1.59	1.22	2.41	1.49	1.61
Fourth	2.11	2.11	1	3.02	2.19	1.38	3.18	2.14	1.48
Fifth	2.97	3.06	0.97	4.41	3.02	1.46	4.23	3.08	1.37
Sixth	4.22	4.5	0.94	6.15	4.63	1.33	5.76	4.17	1.38
Seventh	6.06	6.16	0.99	8.21	6.97	1.18	8.05	6.60	1.22
Eighth	8.9	9.65	0.92	10.76	10.29	1.05	11.61	9.79	1.19
Ninth	13.36	13.95	0.96	14.15	14.63	0.97	17.28	13.86	1.25
Tenth	22.52	20.19	1.12	22.47	16.86	1.33	29.29	19.35	1.51
Overall	6.34	6.25	1.02	7.28	6.16	1.18	8.49	6.17	1.38
Men:									
First	1.64	0.93	1.77	2.86	1.03	2.79	3.37	0.92	3.65
Second	2.24	1.6	1.39	4.19	1.73	2.43	4.36	1.63	2.68
Third	2.92	2.6	1.12	5.54	2.53	2.19	5.31	2.56	2.07
Fourth	3.81	3.86	0.99	7.04	3.71	1.90	6.43	3.39	1.90
Fifth	5	4.86	1.03	8.86	5.43	1.63	7.83	5.22	1.50
Sixth	6.61	6.67	0.99	11.13	7.38	1.51	9.65	7.16	1.35
Seventh	8.88	9.88	0.9	14.00	10.11	1.38	12.16	10.19	1.19
Eighth	12.24	14.24	0.86	17.78	13.90	1.28	15.76	13.97	1.13
Ninth	17.4	18.46	0.94	23.19	19.05	1.22	21.30	18.79	1.13
Tenth	27.86	25.72	1.08	35.24	23.42	1.51	33.34	24.49	1.36
Overall	8.86	8.88	1.00	12.96	8.81	1.47	11.95	8.83	1.35
All patients	7.59	7.56	1.004	10.10	7.48	1.35	10.21	7.49	1.36

Comparison between QRISK, Framingham, and ASSIGN in validation cohort in patients aged 35-74.

\*First tenth is lowest tenth of risk.

compared predicted and observed risks overall for men and women by age band and by fifth of Townsend score. We calculated the area under the receiver operating curve to assess discrimination—that is, the ability of a risk prediction equation to distinguish between those who do and those who do not have a cardiovascular event during the follow-up period. We also calculated the D statistic<sup>28</sup> and an R<sup>2</sup> statistic,<sup>29</sup> which are measures of discrimination and explained variation for survival models.

#### Comparison with the Framingham and ASSIGN equations

We compared the performance of QRISK against the risk estimates derived from Framingham equations.<sup>5</sup> We used an equation for cardiovascular disease risk that was computed by summing the coronary risk (including myocardial infarction and coronary heart disease death plus angina plus coronary insufficiency) and stroke risk (including transient ischaemic attack) as these outcomes are closest to those used in randomised trials of drug effectiveness. They are also the outcomes used to define cardiovascular disease in the current Joint British Society Guidelines.<sup>30</sup>

We calculated the area under the receiver operating curve and the D statistic for the Framingham estimates in the validation dataset and compared them with the corresponding QRISK values. We also compared the

performance of QRISK against the ASSIGN equation.<sup>15</sup> ASSIGN contains similar variables to QRISK except it uses the number of cigarettes smoked per day instead of a categorical variable for smoking and the index of multiple deprivation as the measure of deprivation.

Finally, we calculated the proportion of patients in the validation sample who have a 20% or more risk of cardiovascular disease at 10 years, by age, sex, and deprivation according to QRISK and the Framingham and ASSIGN equations. We determined the proportion of patients who would be reclassified into a higher or lower risk category using the new risk equations. We estimated the number of people in the United Kingdom with a 10 year cardiovascular disease risk of 20% or more for each score, using UK population estimates for 2005. Stata (version 9.2) was used for all analyses.

#### RESULTS

Overall, 478 UK practices met our inclusion criteria of which 318 were randomly assigned to the derivation dataset and 160 to the validation dataset.

In the derivation cohort 3 374 617 patients of all ages were registered with 318 practices at some point during the study period. Of these, 1 283 174 were aged 35-74 years and were free of cardiovascular disease

and diabetes at baseline. These patients were included in the study. Of these, 50.4% were women. During the study period 65 671 incident cases of cardiovascular disease from 8.2 million person years of observation occurred giving a crude incidence rate for cardiovascular disease of 7.96 per 1000 person years. Incidence rates were higher in men than in women and increased steeply with age (table 1). The computer recorded incidence of cardiovascular disease was also significantly higher among men and women from deprived areas, the effect being more noticeable for women. The median follow-up was 6.5 years (range 0-12 years) and 306 259 patients were followed-up for at least 10 years. The 10 year observed risk of a cardiovascular event in women aged 35-74 was 6.69% (95% confidence interval 6.61% to 6.78%) and in men was 9.46% (9.36% to 9.56%).

In the validation dataset 614 553 eligible patients were aged 35-74 and of these 50.3% were women. Incidence rates among patients in the validation cohort were similar to those in the derivation cohort (table 1). The 10 year observed risk of a cardiovascular event in women aged 35-74 was 6.60% (6.48% to 6.72%) and in men was 9.28% (9.14% to 9.43%).

Table 2 compares the characteristics of eligible patients in the cohorts. Although the validation cohort was drawn from an independent group of practices, the baseline characteristics were similar to those for the derivation cohort.

#### Outcome comparison with data from the Office for National Statistics

A subgroup analysis was undertaken in the period 2002-7 for practices with linked data on death to determine the degree of under-ascertainment of deaths from cardiovascular disease on the general practice's clinical computer record—that is, patients with a cardiovascular disease death recorded on Office for National Statistics data who had not been identified through the general practice's clinical record. Of the 39 507 patients with a cardiovascular disease outcome on either record between 2002 and 2007, 37 074

(93.8%) had already been identified on the basis of their general practice record alone—that is, an additional 6.2% of patients were identified by virtue of their Office for National Statistics cause of death linked record.

#### Baseline characteristics of derivation cohort

At baseline 8.4% of men and 12.4% of women were receiving treatment for blood pressure. A family history of premature coronary heart disease in a first degree relative aged less than 60 was recorded in 9.1% of men and 12.1% of women and 0.4% had a recorded diagnosis of left ventricular hypertrophy.

The figure shows the percentage of patients with recorded values. It also shows mean values for body mass index, systolic blood pressure, total serum cholesterol and high density lipoprotein, and the proportion of smokers at baseline compared with data from the health survey for England for the year closest to 1996 that included measurement of the risk factor. Values by age for most variables were similar to those in the health survey suggesting that the study cohort is representative of the general population. The exception was smoking where the survey rates were slightly higher than the study rates for women but lower than the study rates for men.

We compared characteristics of patients with and without recorded values. Women with missing data on smoking status were less likely to be taking blood pressure treatment or to have a family history of premature coronary heart disease. Women with missing values for body mass index were less likely to be receiving treatment for blood pressure, less likely to smoke, and less likely to have a family history of premature coronary heart disease. Women with missing blood pressure measurements were less likely to be receiving treatment for blood pressure, less likely to smoke, less likely to have a family history of premature coronary heart disease, and were slightly younger. Women with missing data on the ratio of total serum cholesterol to high density lipoprotein were less likely to be receiving treatment for blood pressure or to have a family history of premature coronary heart disease, were more likely to be smokers, were younger, had lower systolic blood pressure, and had a slightly lower body mass index. A similar pattern was observed for men. In general patients with missing data had significantly different survival rates—for example, women with a cholesterol ratio recorded had a 10 year observed risk of a cardiovascular event of 4.0% compared with 7.9% for those with missing values. For men the values were 4.9% and 10.9%.

#### Model derivation

Table 3 shows the results of the Cox regression analysis for the final model. A log transformation was used for age but otherwise variables were fitted as linear terms as this best fitted the data according to the fractional polynomial analysis. Three main models were compared for fit and interpretability. The final model included the logarithm of age, ratio of serum

**Table 5** | Summary statistics comparing QRISK equation for 10 year risk of cardiovascular disease to predictions based on Framingham and ASSIGN equations applied to validation cohort in patients aged 35-74

Summary statistics	QRISK	Framingham	ASSIGN
<b>Women:</b>			
Receiver operating curve statistic*	0.7879	0.7744	0.7841
D statistic* (SE)	1.549 (0.014)	1.393 (0.014)	1.472 (0.014)
R <sup>2</sup> statistic† (SE)	36.4%(0.43)	31.7%(0.44)	34.1% (0.43)
<b>Men:</b>			
Receiver operating curve statistic*	0.7674	0.7598	0.7644
D statistic* (SE)	1.447 (0.013)	1.31 (0.012)	1.357 (0.012)
R <sup>2</sup> statistic† (SE)	33.3% (0.39)	29.1% (0.38%)	30.5% (0.38)

Although QRISK validation sample was drawn from independent practices all use same clinical computer system (EMIS) therefore score needs to be validated in practices using other clinical systems.

\*Higher scores indicate better discrimination.

†Indicates percentage of variation explained.

**Table 6** | Numbers (percentages) of patients with a cardiovascular disease risk score  $\geq 20\%$  by sex and Townsend fifth in patients aged 35-74, estimated using three risk scores

Variables	Total population	QRISK		Framingham		ASSIGN	
		No of patients	%	No of patients	%	No of patients	%
Women							
Townsend fifth*:							
First	63 102	1871	2.97	2878	4.56	5451	8.64
Second	62 641	2408	3.84	3014	4.81	5699	9.10
Third	62 566	3054	4.88	3309	5.29	6661	10.65
Fourth	61 561	4598	7.47	3887	6.31	8168	13.27
Fifth	59 543	5910	9.93	3755	6.31	8873	14.90
Men							
Townsend fifth*:							
First	60 074	5770	9.60	12 324	20.51	9504	15.82
Second	60 020	6145	10.24	12 210	20.34	9546	15.90
Third	60 337	6569	10.89	12 175	20.18	9798	16.24
Fourth	61 346	7607	12.40	12 654	20.63	10 738	17.50
Fifth	63 363	7976	12.59	12 382	19.54	11 430	18.04
Total	614 553	51 908	8.45	78 588	12.79	85 868	13.97

\*First fifth is most affluent and last fifth most deprived.

cholesterol to high density lipoprotein levels, systolic blood pressure, body mass index, family history of premature cardiovascular disease, smoking status, Townsend deprivation score, and use of at least one blood pressure treatment. The final model also included an interaction term between systolic blood pressure and blood pressure treatment. Left ventricular hypertrophy was omitted from the final model as the prevalence of recording was low and its inclusion made little difference to the overall model. The area measure of ethnicity was also omitted from the final model since it did not improve the model fit over and above the inclusion of the area measure of deprivation and interpretation of the results was difficult. Overall the final model was superior to the other models according to the Bayes information criterion statistic.

Hazard ratios for the final model show that the risk of a cardiovascular disease event increased with increasing age, body mass index, and Townsend deprivation score. The risk was higher in patients who smoked, had a family history of premature cardiovascular disease, and were receiving blood pressure treatment at baseline. Risk increased more steeply with systolic blood pressure in those not receiving treatment for blood pressure. The hazard ratio associated with the ratio for total serum cholesterol to high density lipoprotein level was just above and close to one for both sexes. However, we decided a priori to include this risk factor in the model. The model was repeated omitting the cholesterol ratio term and a small change was found to the coefficients for some of the other variables.

#### Calibration and discrimination of QRISK v Framingham v ASSIGN

Table 4 compares predicted and observed risks of a cardiovascular disease event at 10 years across each tenth of risk (the first tenth representing the lowest risk) for the Framingham, ASSIGN, and QRISK

equations. Overall, the Framingham equation over-predicted risk at 10 years by 35%, ASSIGN by 36%, and QRISK by 0.4%. All three equations tended to over predict risk in the lowest three tenths for men and women. The most noticeable over-prediction occurred with ASSIGN, followed by the Framingham equation and then QRISK. The Framingham equation, however, consistently over-predicted risk in men across every tenth, with a 51% over-prediction in the top tenth compared with 36% for ASSIGN and 8% for QRISK

Table 5 shows the validation statistics for QRISK compared with the Framingham and ASSIGN equations. The receiver operating curve statistic indicates that the final QRISK score has at least as good as, if not slightly better, discrimination than the Framingham and ASSIGN equations.

The  $R^2$  statistic (standard error) for QRISK in women was 36.4% (0.43%), which is higher than the corresponding value for the Framingham equation which was 31.7% (0.44%). Similarly, the  $R^2$  statistic for the model for men was higher for QRISK than for the Framingham equation. The D statistic for QRISK was 1.45 for men and 1.55 for women, both higher than the corresponding values for the Framingham equations which were 1.31 and 1.39 respectively, indicating better discrimination for QRISK.

#### Predictions with age, sex, and social deprivation

Table 6 shows the proportion of patients with a cardiovascular disease risk score of  $\geq 20\%$  (high risk) by Townsend fifths and sex according to each of the three risk prediction scores.

The most noticeable differences were observed for women. QRISK predicts 9.9% of women aged 35-74 from the most deprived fifth to be at high risk compared with 3.0% of women from the most affluent fifth. The corresponding values for the Framingham



equation are 6.3% (most deprived) and 4.6% (most affluent).

QRISK predicts 12.6% of men from the most deprived areas to be at high risk compared with 9.6% of those from the most affluent areas. The corresponding values for the Framingham equation were 19.5% (most deprived) compared with 20.5% (most affluent).

Overall QRISK predicted 8.5% of patients aged 35-74 to be at high risk compared with 12.8% for the Framingham equation and 14.0% for ASSIGN (table 7). Using QRISK 34.5% of women and 72.9% of men aged 64-75 would be at high risk compared with 24.1% and 86.0% using the Framingham equation. Estimates based on QRISK give 3.2 million patients aged 35-74 at high risk in the United Kingdom for 2005, with the Framingham equation predicting 4.7 million and ASSIGN 5.1 million.

Overall, 53 668 patients in the validation dataset (9% of total) would be reclassified from high to low risk or vice versa using QRISK compared with using the Framingham equation. Of these, 13 494 patients were classified as low risk using the Framingham equation but high risk using QRISK, and their observed 10 year risk was 22.32% (95% confidence interval 21.36% to 23.32%). Conversely, of those 40 174 at high risk using the Framingham equation but at low risk using QRISK, the observed 10 year risk was 16.45% (15.98% to 16.93%).

## DISCUSSION

We derived and validated a new cardiovascular disease risk equation (QRISK) in the United Kingdom in readiness for a major change in national policy on the identification of patients at high risk of cardiovascular disease. It is the largest such study to have ever been undertaken and the first time that routine data in a UK general practice population have been used rather than an observational study in a predefined cohort. The study was validated on an independent sample of practices (although data capture methods are uniform in the

derivation and validation sample) and performs as well if not better than the Framingham algorithm in the UK population. However, since the validation was carried out in a similar population to the population from which the algorithm was derived, it potentially has a "home advantage." Further validation in other populations is therefore required.

QRISK includes standard as well as additional risk factors for cardiovascular disease, such as deprivation, family history of premature coronary heart disease, body mass index, and the effect of existing antihypertensive treatment. Inclusion of antihypertensive treatment at baseline is of relevance to 1 in 10 of the population and failure to take this into account would lead to a significant underestimation of cardiovascular disease risk in the population. The positive associations with antihypertensive treatment and cardiovascular disease risk were expected as these patients might have several factors as well as blood pressure associated with increased risk. Also these patients might have subclinical disease.

Unlike the existing Framingham equations and the European SCORE equation,<sup>31</sup> QRISK identifies and includes deprivation in the estimation of cardiovascular disease risk. This will be an important step in supporting national initiatives to reduce health inequalities in cardiovascular disease<sup>32</sup> and is likely to be an improvement on the Framingham algorithm, which tends to over-estimate risk in affluent areas and underestimate risk in deprived areas.<sup>8</sup> Also, a weighting for social deprivation might help to minimise health inequalities, which may increase when new interventions are introduced because of the inverse equity hypothesis.<sup>33</sup> The inclusion of a family history of premature coronary heart disease is important because observational studies indicate that such disease in a first degree relative increases risk by 50% or more,<sup>34,35</sup> although our estimates were lower. Further work is needed to identify the utility of using self reported ethnicity rather than an area based measure,

**Table 7 | Numbers (percentages) of patients by age and sex with cardiovascular disease risk  $\geq 20\%$  in validation cohort and estimated numbers in UK population in 2005 according to QRISK and Framingham scores**

Age band	Total No in validation cohort	QRISK			Framingham		
		No at high risk in validation cohort	Proportion at high risk (%)	Estimated No at high risk in UK, 2005	No at high risk in validation cohort	Proportion at high risk (%)	Estimated No at high risk in UK, 2005
<b>Women</b>							
Age group:							
35-44	113 725	0	0	0	15	0.01	616
45-54	85 938	1	0	45	688	0.80	31 189
55-64	61 729	1277	2.07	74 008	4572	7.41	264 970
65-74	48 021	16 563	34.49	917 189	11 568	24.09	640 587
<b>Men</b>							
Age group:							
35-44	125 035	0	0	0	381	0.30	13 952
45-54	86 932	243	0.28	10 669	6815	7.84	299 201
55-64	56 690	7228	12.75	439 660	23 164	40.86	1 409 004
65-74	36 483	26 596	72.9	1 741 720	31 385	86.03	2 055 342
All patients	614 553	51 908	8.45	3 183 291	78 588	12.79	4 714 860

because the inclusion of area measures of ethnicity in QRISK did not improve the utility of the score possibly as a result of the correlation with the area based measure of deprivation. QRISK tackles components not included in SCORE, which was designed for European use. SCORE does not include cardiovascular disease morbidity, which is an integral part of both trial data (non-fatal myocardial infarction and stroke) and cost effectiveness reviews. The health technology assessment of statins in the United Kingdom has identified a cardiovascular disease event of 20% in 10 years as the threshold for treatment, of which most are non-fatal events.<sup>36</sup>

Although QRISK includes more variables than the Framingham algorithm and therefore seems more complex, QRISK has been derived from routinely collected clinical data from general practice. Not only is QRISK likely to have a face validity in the setting where it is likely to be applied, but software can be developed within the general practice clinical system to automatically calculate scores. Therefore this could be implemented simply in clinical practice to flag high risk patients for further assessment. Furthermore, the case for inclusion of antihypertensive treatment has already been made with respect to Framingham equations for risk prediction of stroke, a component of this new equation.<sup>37</sup>

This study is based on a large, representative and contemporary UK population with data obtained from a validated research database. We used an open cohort study design, which allowed patients to enter the study population throughout the study period. This was done so that patients who had more recent data could be included in the estimation of the model coefficients. The disadvantage of this approach is that not all our patients had the opportunity to contribute 10 years worth of follow-up data although we did have over 300 000 patients with at least 10 years of follow-up data. The median follow-up time was therefore relatively short compared with the long established Framingham cohort and the more recent cohort used for ASSIGN. This could introduce some imprecision in the estimate of survival at 10 years. To assess this we restricted the analysis to the closed cohort and found minimal difference in the baseline survivor function. As incidence of cardiovascular disease is changing over time it is important to update this algorithm periodically as planned. The cohort is socially, ethnically, and geographically diverse and over 250 times the size of the original Framingham cohorts from 30-50 years ago. It includes both men and women.

The utility of any tool for risk prediction depends on the quality of the data used to derive it; both risk factors and outcome measure. Although levels of completeness for some risk factors were good (notably body mass index, smoking, and blood pressure), recording levels were low for serum cholesterol. This is a likely source of bias and may have contributed to the attenuation of the hazard ratio of the cholesterol ratio term towards one. However, comparison with the mean values for systolic blood pressure, cholesterol, and

body mass index from the health survey for England shows that this population is likely to be representative of the population in England. This means our results should be generalisable. The exception was smoking where the survey rates were higher than the study rates for women but lower than the study rates for men. This could reflect a differential bias in self reported smoking status either within the survey, within the patients' electronic health record, or within both. A strength of our study, however, is the use of multiple imputation of missing values, a technique that is designed to increase the power of the analysis and produce models that are more statistically reliable and applicable within clinical practice.<sup>24 25</sup> The pattern of missing values, including differences in survival for patients with and without missing data, supported the use of this approach as compared with a complete case analysis. We used a single baseline measure for systolic blood pressure, which might lead to underestimation of its association with cardiovascular disease risk owing to large variation within individuals. Also we included a relatively crude measure of blood pressure treatment rather than undertaking a more complex time varying approach.

The outcome measures in our study were not formally adjudicated—that is, some patients with a diagnosis of cardiovascular disease may have been misdiagnosed. Validation studies of similar general practice databases show that false positive diagnoses occur in fewer than 10% of cases.<sup>38</sup> Similarly, the information for some patients with cardiovascular disease may not be recorded on the clinical records. Such misclassification of outcomes, if present and if non-differential for risk factors, will result in hazard ratios tending towards one, which would reduce the discriminatory power of a risk score. Our subgroup analysis between 2002 and 2007, however, showed a 94% ascertainment rate for cardiovascular disease outcomes based on the general practice clinical record alone compared with outcomes including death certification data.

Under-ascertainment is more likely to affect patients from the most deprived areas and so it is possible that our algorithm will underestimate risk in these patients. This underlines the need to include measures of deprivation in both the development and the application of risk prediction scores as well as the need for more and better data linkage across the health service. It is worth noting that this is the first time that cause of death data from the Office for National Statistics have been linked to general practice data across 60% of the United Kingdom. Only a five year subset of data were ready in time for this analysis, data back to 1993 will be used to update algorithms for coronary heart disease and stroke.

Lastly, although we have validated QRISK on a separate sample of one third of practices from QRESEARCH (that is, a different population from the population used to develop the algorithm), it could be argued that this is not an entirely independent population because all the practices were using the EMIS clinical system. This potentially gives a “home advantage”

### WHAT IS ALREADY KNOWN ON THIS TOPIC

Current cardiovascular risk prediction algorithms based on the Framingham model tend to over-predict risk and are not well calibrated for the UK population

The Framingham model does not include measures of deprivation, family history, body mass index, or current treatment with antihypertensives

### WHAT THIS STUDY ADDS

A new risk score, QRISK, performed at least as well as the Framingham model for discrimination and was better calibrated to the UK population than either the Framingham model or ASSIGN

QRISK would identify a different high risk group of patients than the Framingham equation, with one in ten patients being reclassified into high or low risk

QRISK includes additional variables that allow better tailoring of management to the individual patient and will help to minimise health inequalities

to QRISK as prediction algorithms tend to perform better in populations similar to the ones in which they were derived. A second validation is therefore underway using a different general practice database (THIN) derived from practices using a different clinical computer system. This analysis, which will be a more severe test of the performance of QRISK, will be reported separately when complete.

#### Comparison with Framingham and ASSIGN algorithms

The Framingham cohort is an invaluable resource for epidemiology and aetiology owing to the completeness of data. The weakness is that it was based on exceptionally high rates of cardiovascular disease at the height of an epidemic but is now used as a tool to describe risks in contemporary European populations for the purpose of drug treatment. It has repeatedly been shown to be seriously inaccurate. QRISK explains a higher percentage of the variation than either the Framingham or ASSIGN algorithms in both men and women. Discrimination (the ability to differentiate between high and low risk patients) with QRISK was better than with the Framingham equation. Improved discrimination using QRISK results in a steeper gradient of cardiovascular disease risk for social deprivation compared with estimates using the Framingham equation. QRISK tends to increase cardiovascular disease risk estimates in people living in deprived areas relative to those in more affluent areas. This effect is most noticeable for women from deprived areas. Inclusion of area based deprivation measures within the QRISK prediction algorithm is likely to predict more accurately which patients are at high or low risk and so target interventions more appropriately on the basis of likely clinical need.

Our analysis shows that neither the Framingham nor ASSIGN equations is well calibrated for this UK population, with both scores tending to over-predict risk. If the patient's absolute risk is lower than predicted, the absolute benefits of an intervention will be smaller and the balance of risks and benefits will be less favourable. Additionally, overestimation of risk prediction will adversely affect direct prescribing costs as well as the costs associated with monitoring for, and managing,

side effects. The degree of over-prediction for the Framingham algorithm is consistent with that reported in other validation studies of the Framingham equation in the United Kingdom.<sup>39</sup> We think that the over-prediction is probably the result of a combination of differences in risk factor means and prevalences and baseline survival function. In ASSIGN, for example, 41.5% of patients smoked compared with 26% of the study population; 26% of patients in the ASSIGN cohort had a family history of premature coronary heart disease in a first degree relative compared with 11% recorded in the study population. The mean serum cholesterol level in the ASSIGN cohort was 6.23 mmol/l compared with 5.8 mmol/l in our population. ASSIGN is calibrated to the incidence of cardiovascular disease in Scotland, which is higher than that in England. Although much less noticeable, some over-prediction occurred in the top tenth using QRISK, which might reflect an effect of increasing age when the effect of risk factors diminish and survivors are more resilient. Alternatively it could be due to competing risks of death, such as cancer, which become more important at higher ages.

#### Conclusion

We have derived and validated a new cardiovascular disease risk prediction score that is better calibrated for the UK population and has better discrimination. Overall in patients aged 35-74, the Framingham algorithm over-predicted cardiovascular disease risk at 10 years by 35%, ASSIGN by 36%, and QRISK by 0.4%. QRISK predicted 9% of patients aged 35-74 to be at high risk compared with 13% for the Framingham equation and 14% for ASSIGN. Using QRISK we estimate that 34% of women and 73% of men aged 64-75 would be at high risk compared with 24% and 86% according to the Framingham equation. UK estimates on the basis of QRISK give 3.2 million patients at high risk in 2005, compared with 4.71 million for the Framingham equation and 5.1 million for ASSIGN. Overall, QRISK would reclassify about 1 in 10 patients from high to low risk or vice versa compared with the Framingham algorithm.

We think that QRISK is likely to provide more appropriate estimates of cardiovascular disease risk in contemporary UK populations and better discriminate those at high risk on the basis of their age, sex, and social deprivation as well as existing antihypertensive treatment. It is likely therefore to be a more equitable tool to inform patient management decisions.

We thank EMIS and the EMIS practices contributing to QRESEARCH; Patrick Royston for advice on multiple imputation, fractional polynomials, and the D statistic; Rod Jackson, Stuart Pocock, Philip Bath, Hugh Tunstall-Pedoe, and Mark Woodward for reviewing the protocol; and the following who enabled the project, contributed to discussions, or facilitated the linkage of cause of deaths data with general practice data—Nirupa Dattani and Sir Peter Goldblatt (Office for National Statistics), Steve Daniels (Connecting for Health), John Fox and Dave Roberts (Information Centre), Andy Whitwam and David Stables (EMIS), Bill Kirkup, Maggie Rae, Michael Soljak, and Roger Boyle (Department of Health).

**Contributors:** JHC initiated and designed the study, obtained approvals, prepared the data, undertook the analysis and interpretation, and wrote the first draft of the paper. CC, YV, and MM contributed to the development of the protocol, design, analysis and interpretation, and drafting of the paper. CC and YV undertook some of the primary analyses with JHC. JR and PB contributed to the conception, design, analysis, interpretation, and drafting of the article and approved the final version.

**Funding:** The study received no external funding.

**Competing interests:** JR chairs, and PB is a member of, the NICE guideline development group on cardiovascular risk assessment "The modification of blood lipids for the primary and secondary prevention of cardiovascular disease." JHC is codirector of QRESEARCH, a not for profit organisation that is a joint partnership between the University of Nottingham and EMIS (leading supplier of information technology for 60% of UK general practices). QRESEARCH undertakes analyses for the Department of Health and other government organisations. All research using the database is peer reviewed and published. QRESEARCH is not used for studies for the pharmaceutical industry.

**Ethical approval:** This study was approved by the Trent multicentre research ethics committee.

- 1 British Heart Foundation. *Coronary heart disease statistics*. London: BHF, 2007.
- 2 Department of Health. *National service framework for coronary heart disease*. London: DoH, 2000.
- 3 NCEP Expert Panel. *Detection, evaluation and treatment of high blood cholesterol in adults (adult treatment panel III): executive summary*. 3, 1-30. Report of the national cholesterol education program (NCEP). Bethesda, MD: National Institute of Health/National Heart, Lung and Blood Institute, 2001.
- 4 Mann J, Crooke M, Fear H, Hay D, Jackson R, Neutze J, et al. Guidelines for detection and management of dyslipidaemia. Scientific Committee of the National Heart Foundation of New Zealand. *NZ Med J* 1993;106:133-41.
- 5 Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121:293-8.
- 6 Asia Pacific Cohort Studies Collaboration. Cardiovascular risk prediction tools for populations in Asia. *J Epidemiol Community Health* 2007;61:115-21.
- 7 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;327:1-6.
- 8 Brindle P, McConnachie A, Upton M, Hart C, Davey-Smith G, Watt G. The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *Br J Gen Pract* 2005;55:838-45.
- 9 Tunstall-Pedoe H, Woodward M. By neglecting deprivation, cardiovascular risk scoring will exacerbate social gradients in disease. *Heart* 2006;92:307-10.
- 10 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.
- 11 North of England Hypertension Guideline Development Group. *Essential hypertension: managing adult patients in primary care (NICE guideline)*. University of Newcastle upon Tyne: Centre for Health Services Research, 2004.
- 12 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 guidance type: Technology appraisal*. In: Excellence NifHaC, editor, 2006.
- 13 White C, Galen FV, Chow YH. Trends in social class differences in mortality by cause. *Health Stats Q* 2003;20:25-36.
- 14 Scottish Intercollegiate Guidelines Network. *Risk estimation and the prevention of cardiovascular disease. A national clinical guideline*. Edinburgh: SIGN, 2007.
- 15 Woodward M, Brindle P, Tunstall-Pedoe H. Adding social deprivation and family history to the cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* 2007;2:172-6.
- 16 Hippisley-Cox J, Vinogradova Y, Coupland C, Pringle M. *Comparison of key practice characteristics between general practices in England and Wales and general practices in the QRESEARCH data*. Report to the Health and Social Care Information Centre: University of Nottingham, 2005.
- 17 Hippisley-Cox J, Parker C, Coupland C, Vinogradova Y. Inequalities in the identification and treatment of hyperlipidaemia in coronary heart disease in patients with serious mental health problems: cross-sectional study. *Heart* 2007 (in press).
- 18 Hippisley-Cox J, Coupland C. Effect of statins on the mortality of patients with ischaemic heart disease: population based cohort study with nested case-control analysis. *Heart* 2006;92:752-8.
- 19 Hippisley-Cox J. *Trends and variations in GMS indicators for coronary heart disease. Report to the Department of Health*. Nottingham: University of Nottingham, 2005.
- 20 Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. *BMJ* 2005;330:1366-74.
- 21 Hippisley-Cox J, Coupland C. Effect of combinations of drugs on all cause mortality in patients with ischaemic heart disease: nested case-control analysis. *BMJ* 2005;330:1059-63.
- 22 Townsend P, Davidson N. *The Black report*. London: Penguin, 1982.
- 23 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.
- 24 Clark T, Altman D. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2002;56:28-37.
- 25 Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227-41.
- 26 Royston P. Multiple imputation of missing values: update of ice. *Stata J* 2005;5:527-36.
- 27 Wilson P, D'Agostino R, Levy D, Belanger A, Silbershatz H, Kannel W. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837-47.
- 28 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stats Med* 2004;23:723-48.
- 29 Royston P. Explained variation for survival models. *Stata J* 2006;6:1-14.
- 30 Joint British Society Guidelines 2. Joint British Society guidelines on prevention of cardiovascular disease in clinical practice. *Heart* 2005;91(Suppl 5):1-52.
- 31 De Becker G, Ambrosioni E, Borch-Johnsen K, Brotons C, Cifkova R, Dallongeville J, et al. European guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J* 2003;24:1601-10.
- 32 Department of Health. *Tackling health inequalities: a programme for action*. London: DoH, 2003:83.
- 33 Victora C, Vaughan J, Barros F, Anamaria C, Tomasi E. Explaining trends in inequities: evidence from Brazilian child health studies. *Lancet* 2000;356:1093-8.
- 34 Lloyd Jones X, Lloyd-Jones D, Nam B, D'Agostino RS, Levy D, Murabito J, 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;291:2204-11.
- 35 Nilsson P, Nilsson J, Berglund G. Family burden of cardiovascular mortality: risk implications for offspring in a national register linkage study based upon the Malmo Preventive Project. *J Intern Med* 2004;255:229-35.
- 36 Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, et al. A systematic review and economic evaluation of statins for the prevention of coronary events. In: HTA, ed. *Health Technology Assessment*, 2007.
- 37 D'Agostino R, Wolf P, Belanger A, Kannel W. Stroke risk profile: adjustment for antihypertensive medication. *Stroke* 1994;25:40-3.
- 38 Jick H, Derby L, Gurewich V, Vasilakis C. The risk of myocardial infarction associated with antihypertensive drug treatment in persons with uncomplicated essential hypertension. *Pharmacotherapy* 1996;16:321-6.
- 39 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;327:1267-73.

Accepted: 27 June 2007