Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2

BMJ

RESEARCH

Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2

Julia Hippisley-Cox, professor of clinical epidemiology and general practice,[1] Carol Coupland, senior lecturer in medical statistics,[1] Yana Vinogradova, research fellow in medical statistics,[1] John Robson, senior lecturer in general practice,[2] Rubin Minhas, coronary heart disease lead,[3] Aziz Sheikh, professor of primary care research and development,[4] Peter Brindle, research and development strategy lead[5]

[1]Division of Primary Care, Tower Building, University Park, Nottingham NG2 7RD
[2]Centre for Health Sciences, Queen Mary's School of Medicine and Dentistry, London E1 2AT
[3]Medway Primary Care Trust, Unit 2, Gillingham, Kent ME7 0NJ
[4]Division of Community Health Sciences: GP Section, University of Edinburgh, Edinburgh EH8 9DX
[5]Avon Primary Care Research Collaborative, Bristol Primary Care Trust, Bristol BS2 8EE
Correspondence to: J Hippisley-Cox Julia.hippisley-cox@nottingham.ac.uk

Cite this as: *BMJ* 2008;336:a332
doi:10.1136/bmj.39609.449676.25

ABSTRACT

Objective To develop and validate version two of the QRISK cardiovascular disease risk algorithm (QRISK2) to provide accurate estimates of cardiovascular risk in patients from different ethnic groups in England and Wales and to compare its performance with the modified version of Framingham score recommended by the National Institute for Health and Clinical Excellence (NICE).

Design Prospective open cohort study with routinely collected data from general practice, 1 January 1993 to 31 March 2008.

Setting 531 practices in England and Wales contributing to the national QRESEARCH database.

Participants 2.3 million patients aged 35-74 (over 16 million person years) with 140 000 cardiovascular events. Overall population (derivation and validation cohorts) comprised 2.22 million people who were white or whose ethnic group was not recorded, 22 013 south Asian, 11 595 black African, 10 402 black Caribbean, and 19 792 from Chinese or other Asian or other ethnic groups.

Main outcome measures First (incident) diagnosis of cardiovascular disease (coronary heart disease, stroke, and transient ischaemic attack) recorded in general practice records or linked Office for National Statistics death certificates. Risk factors included self assigned ethnicity, age, sex, smoking status, systolic blood pressure, ratio of total serum cholesterol:high density lipoprotein cholesterol, body mass index, family history of coronary heart disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, type 2 diabetes, renal disease, atrial fibrillation, and rheumatoid arthritis.

Results The validation statistics indicated that QRISK2 had improved discrimination and calibration compared with the modified Framingham score. The QRISK2 algorithm explained 43% of the variation in women and 38% in men compared with 39% and 35%, respectively, by the modified Framingham score. Of the 112 156 patients classified as high risk (that is, ≥20% risk over 10 years) by the modified Framingham score, 46 094 (41.1%) would be reclassified at low risk with QRISK2. The 10 year observed risk among these reclassified patients was 16.6% (95% confidence interval 16.1% to 17.0%)—that is, below the 20% treatment threshold. Of the 78 024 patients classified at high risk on QRISK2, 11 962 (15.3%) would be reclassified at low risk by the modified Framingham score. The 10 year observed risk among these patients was 23.3% (22.2% to 24.4%)—that is, above the 20% threshold. In the validation cohort, the annual incidence rate of cardiovascular events among those with a QRISK2 score of ≥20% was 30.6 per 1000 person years (29.8 to 31.5) for women and 32.5 per 1000 person years (31.9 to 33.1) for men. The corresponding figures for the modified Framingham equation were 25.7 per 1000 person years (25.0 to 26.3) for women and 26.4 (26.0 to 26.8) for men). At the 20% threshold, the population identified by QRISK2 was at higher risk of a CV event than the population identified by the Framingham score.

Conclusions Incorporating ethnicity, deprivation, and other clinical conditions into the QRISK2 algorithm for risk of cardiovascular disease improves the accuracy of identification of those at high risk in a nationally representative population. At the 20% threshold, QRISK2 is likely to be a more efficient and equitable tool for treatment decisions for the primary prevention of cardiovascular disease. As 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 advised.

INTRODUCTION

Cardiovascular disease is the leading cause of premature death and a major cause of disability in the United Kingdom.[1] Evidence from randomised controlled trials supports the effectiveness of statins in reducing cardiovascular risk and the National Institute for Health and Clinical Excellence (NICE) has lowered the threshold for intervention for primary prevention with statins from a 10 year risk of cardiovascular disease of 40% to 20%.[2 3] In April 2008, the UK government announced a new major initiative to reduce the risk of vascular disease.[4] This will build on new NICE guidelines for lipid modification to be published later this year.[5] It is important that this major

public health programme targets those at greatest risk and reduces, rather than exacerbates, existing persistent and widening ethnic and social inequalities in risk of cardiovascular disease.[6-8] A broader approach to preventative cardiovascular medicine is required that recognises the evidence for the role of the biological, socioeconomic, and ethnic determinants of health[9 10] and is responsive to changes in secular trends in the incidence of coronary heart disease.[11]

Recent advances in the development of models to assess risk of cardiovascular disease means these can now recognise and take account of the increased risk associated with social deprivation in the UK.[12 13] Rates of cardiovascular disease, however, vary considerably between ethnic groups, which might reflect increased susceptibility and differential exposure to risk factors.[14-17] While several risk prediction scores derived from prospective studies can be used to identify and prioritise people for risk reducing interventions,[12 13 18] they do not include a variable for self assigned ethnicity. One cross sectional study used ethnicity specific levels of risk of cardiovascular disease and risks factors to estimate 10 year risk. The use of this tool, however, excluded diabetes, lacked precision because of small numbers, and has not been validated.[19]

In May 2008, NICE recommended multiplying the results of a modified version of the US Framingham score ("modified Framingham") by a correction factor of 1.4 for south Asian men in the UK.[5] This does not reflect the heterogeneity in risk of cardiovascular disease between south Asian populations, the increased risk in women, confounding by deprivation,[20] and the possibility of double counting through adjustments for both ethnicity and family history. An appropriate estimation of risk by ethnic group is important to improve cardiovascular outcomes, avoid the potential for further deterioration in health inequalities,[21] and ensure the efficient allocation of resources used to support cardiovascular disease prevention programmes.

Because of the lack of prospective outcome data on black and minority ethnic groups,[22] a contemporary and specific algorithm is needed to accurately quantify risk among such patients and to identify the independent or interacting contributions of factors including deprivation, family history, and diabetes.[23-25] The QRESEARCH database contains longitudinal data, individual risk factors, demographic data, measures of social deprivation, and, increasingly, records of self assigned ethnicity, which provide a unique opportunity to model all these factors together.

We built on our previous risk prediction algorithm (QRISK1)[12] to develop a revised algorithm that incorporates self assigned ethnicity as well as a range of other potentially relevant conditions associated with cardiovascular risk such as type 2 diabetes, treated hypertension, rheumatoid arthritis, renal disease, and atrial fibrillation (QRISK2). By including an increased range of potential risk factors, we hypothesised that we would be better able to personalise risk to the individual patient.

METHODS

Study design and data source

We conducted a prospective cohort study in a large UK primary care population using a similar method to our original analysis.[12] We used version 19 of the QRESEARCH database (www.qresearch.org). This is a large validated primary care electronic database containing the health records of 11 million patients registered from 551 general practices using the Egton Medical Information System (EMIS) computer system.[12] Practices and patients on the database are nationally representative[26] and similar to those on other primary care databases that use other clinical software systems.[27]

The QRESEARCH database now contains information on the cause of death as recorded on the patient's Office for National Statistics (ONS) death certificate. This data linkage, which is based on NHS number, has now been successfully completed back to 1993. A recorded cause of death is now linked for over 97% of patients on the QRESEARCH database who have died.

Practice selection

We included all QRESEARCH practices in England and Wales once they had been using their current EMIS system for at least a year (to ensure completeness of recording of morbidity and prescribing data), randomly allocating two thirds of practices to the derivation dataset with one third to the validation dataset. We used the simple random sampling utility in Stata to assign practices to the derivation or validation cohort.

Cohort selection

We identified an open cohort of patients aged 35-74 at the study entry date, drawn from patients registered with eligible practices during the 15 years from 1 January 1993 to 31 March 2008. We used an open cohort design as this allows patients to enter the population throughout the whole study period rather than requiring registration on 1 January 1993, thus better reflecting the realities of routine general practice.

We excluded patients with a prior recorded diagnosis of cardiovascular or cerebrovascular disease, temporary residents, patients with interrupted periods of registration with the practice, and those who did not have a valid Townsend deprivation score. We also excluded patients who were taking statins at baseline.

For each patient we determined an entry date to the cohort, 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 plus one year, and the beginning of the study period (1 January 1993). In addition we included patients in the analysis only once they had a minimum of one year's complete data in their medical record.

Coding of ethnicity

We used Read codes for self assigned ethnicity. The codes were grouped into the NHS standard 16+1 categories[28] for the initial descriptive analysis. The 16+1 categories were then further grouped into the

final nine reporting groups to ensure sufficient numbers of events to enable a meaningful analysis. The white ethnic group was combined with the group where ethnicity was not recorded since, assuming the study population is comparable with the UK population, 93% or more of people without ethnicity recorded would be expected to be from a white ethnic group. The category "other including mixed" comprised white and black Caribbean, white and black African, white and Asian, other mixed, other black, and other ethnic group. The "white or not recorded" category comprised British, Irish, and other white background as well as not recorded. This was designated as the reference category. The category of other Asian included Read codes for east African Asian, Indo-Caribbean, Punjabi, Kashmiri, Sri Lankan, Tamil, Sinhalese, Caribbean Asian, British Asian, mixed Asian, or Asian unspecified.

Cardiovascular disease outcomes

The primary outcome measure was the first recorded diagnosis of cardiovascular disease recorded on the general practice clinical computer system or their linked ONS death certificate during the study period. For this study, we included coronary heart disease (angina and myocardial infarction), stroke, or transient ischaemic attacks in the term cardiovascular disease but not peripheral vascular disease.

The Read codes used for case identification on the computer record were nationally agreed ones used in the quality and outcomes framework for general practice for coronary heart disease and cerebrovascular disease. The ICD-10 codes used for case identification on the ONS death certificate were: angina pectoris (I20); acute myocardial infarction (I22); complications following acute myocardial infarction (I23); other acute ischaemic heart disease (I24); chronic ischaemic heart disease (I25); cerebral infarction (I63); and stroke, not specified as haemorrhage or infarction (I64).

Included variables

- Self assigned ethnicity (white/not recorded, Indian, Pakistani, Bangladeshi, other Asian, black African, black Caribbean, Chinese, other including mixed)
- Age (years)
- Sex (males *v* females)
- Smoking status (current smoker, non-smoker (including ex-smoker))
- Systolic blood pressure[18] (continuous)
- Ratio of total serum cholesterol/high density lipoprotein cholesterol[18] (continuous)
- Body mass index (BMI)[12] (continuous)
- Family history of coronary heart disease in first degree relative under 60 years[12] (yes/no)
- Townsend deprivation score[12] (output area level 2001 census data evaluated as a continuous variable)
- Treated hypertension[12] (diagnosis of hypertension and at least one current prescription of at least one antihypertensive agent)
- Rheumatoid arthritis[29] (yes/no)
- Chronic renal disease[30] (yes/no)
- Type 2 diabetes[18] (yes/no)
- Atrial fibrillation[31 32] (yes/no)

Risk factors for cardiovascular disease

We included variables in our analysis that are known or thought to affect cardiovascular risk (box). We used the value closest to the entry date to the cohort for each patient, imputing missing values where necessary, as described below.

Model derivation and development

We calculated crude incidence rates of cardiovascular disease according to age, ethnic group, and deprivation in fifths. We directly age standardised the incidence rates by ethnic group and deprivation using the age distribution in five year bands of the entire derivation cohort as the standard population. We also age standardised the means of continuous variables and proportions with risk factors by ethnic group using the same method.

We used Cox proportional hazards models in the derivation dataset to estimate the coefficients and hazard ratios associated with each potential risk factor for the first ever recorded diagnosis of cardiovascular disease for men and women separately. As in our previous paper, we compared models using the Bayesian information criteria (BIC).[33] We used fractional polynomials to model non-linear risk relations with continuous variables where appropriate.[34 35] We tested for interactions between each variable and age and between diabetes and deprivation and included significant interactions in the final model. Continuous variables were centred for analysis.

Our main analyses used multiple imputation to replace missing values for systolic blood pressure, cholesterol/HDL ratio, smoking status, and body mass index. Our final model was fitted based on multiply imputed datasets using Rubin's rules to combine effect estimates and estimate standard errors to allow for the uncertainty caused by missing data.[35 36] Multiple imputation is a statistical technique designed to reduce the biases that can occur in "complete case" analysis along with a substantial loss of power and precision.[37-40] Multiple imputation allows patients with incomplete data to still be included in analyses and makes full use of all the available data, increasing power and precision.[41] The imputation technique involves creating multiple copies of the data and replaces missing values with imputed values based on a suitable random sample from their predicted distribution. We used the ICE procedure in Stata[42] to obtain five imputed datasets (further details are available from the corresponding author).

We took the log of the hazard ratio for each variable from the final model and used these as weights for the new cardiovascular disease risk equations. We combined these weights with the baseline survivor function centred on the means of continuous risk factors to derive a risk equation for 10 years' follow-up.

Validation of new equation

We tested the performance of the new model (QRISK2) in the validation dataset and compared it against both the original model (QRISK1) and the modified Framingham equation recommended by NICE.[43] This

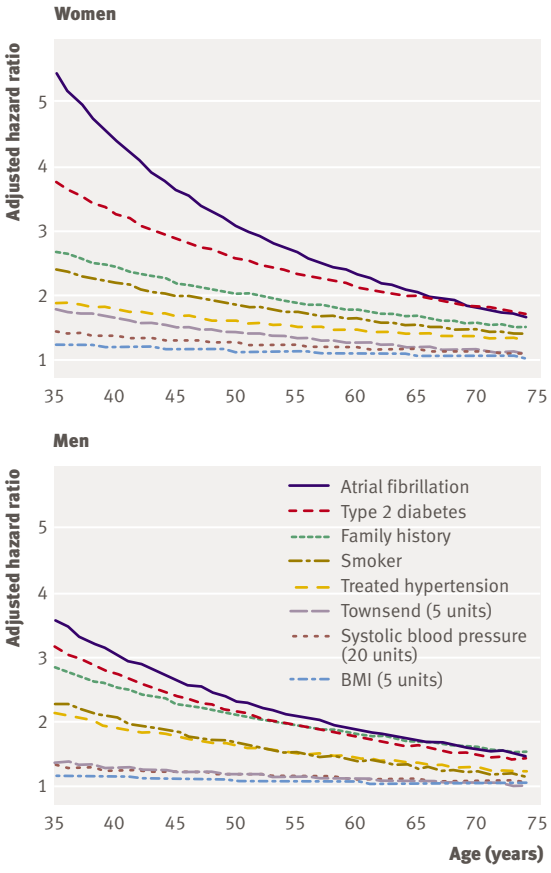

Fig 1 | Impact of age on hazard ratios for cardiovascular disease risk factors using the QRISK2 model

modified equation is based on one of the original Anderson equations[18] and is used to derive separate risks for coronary heart disease and stroke for an individual. The two risks are then added together (if these combined risks exceed 100%, the risk is then set to 100%). For south Asian men, NICE advises multiplying the resulting Framingham score by 1.4. For people with a family history of coronary heart disease in a first degree relative, then the risk is multiplied by 1.5. For south Asian men with a family history of coronary heart disease both multipliers are applied to the individual.

We calculated the 10 year estimated risk of cardiovascular disease for each patient in the validation dataset using multiple imputation to replace missing values as in the derivation dataset.

We calculated the mean predicted and observed cardiovascular disease risk at 10 years[12] and compared these by 10th of predicted risk for each score. The observed risk at 10 years was obtained by using the 10 year Kaplan-Meier estimate. We calculated the Brier score (a measure of goodness of fit where lower values indicate better accuracy)[44] using the censoring adjusted version adapted for survival data,[45] D statistic (a measure of discrimination where higher values indicate better discrimination),[46] and an R^2 statistic. The R^2 statistic is a measure of explained variation where higher values indicate more explained variation.[47] We also calculated the area under the receiver operator curve (ROC), where higher values indicate better discrimination.

We calculated the proportion of patients in the validation sample with an estimated 10 year risk of cardiovascular disease of 20% or more by age, sex, ethnicity, and deprivation according to the QRISK2 algorithm compared with the modified Framingham score. We determined the proportion of patients who would be reclassified into a higher or lower risk category using the new risk equations at the 20% thresholds and determined the observed 10 year risks among those patients who would be reclassified.

As we used all the available data on the QRESEARCH database we did not calculate required sample size before the study. Analyses were conducted using Stata (version 10), with a significance level of 0.01 (two tailed).

RESULTS

Derivation and validation datasets

Practices and patients

Overall, 531 UK practices met our inclusion criteria, of which 355 were randomly assigned to the derivation dataset and 176 to the validation dataset. We excluded 20 practices that did not have complete data for the relevant study period (four practices) or were from Scotland (seven practices) or Northern Ireland (nine practices).

We studied 2.29 million patients with over 16 million person years and 140 115 cardiovascular events. There were 1 591 209 patients in the derivation cohort, of whom 55 626 had cardiovascular disease before the start of the study leaving 1 535 583 patients (773 291 women, 50.4%) aged 35-74 and free of cardiovascular disease. Table 1 shows the numbers of patients in each ethnic group.

Baseline characteristics of derivation and validation cohort

Table 1 compares the characteristics of eligible patients in both cohorts. Ethnicity was recorded in 209 214 (27.1%) women and 181 110 (23.8%) men. Among patients with ethnicity recorded 89.3% were from a white ethnic group. The mean follow-up was 7.3 years for women and 6.9 for men. Some 437 676 patients (232 306 women and 205 370 men) had more than 10 years of follow-up data.

While this validation cohort was drawn from an independent group of practices, the baseline characteristics were similar to those for the derivation cohort.

Incidence of cardiovascular disease

Table 2 shows the incidence rates of cardiovascular disease by age, sex, deprivation, and ethnicity in the derivation cohort. There were 96 709 incident cases of cardiovascular disease (41 042 in women) during the study period from 10.9 million person years of observation. Of all events, 7.4% in women and 7.8% in men were identified with the ONS linked death data (that is, were not identified with the general practice data alone). The crude incidence rate for cardiovascular disease was slightly higher than in our original study with a rate of 7.3

per 1000 person years for women and 10.5 per 1000 person years for men. In the validation dataset there were 750 232 eligible patients aged 35 to 74, and, of these, 50.1% were women and the incidence rates were similar to the derivation dataset (data not shown, but available from the corresponding author).

The incidence of crude and age standardised cardiovascular disease varied widely between ethnic groups (table 2). The age standardised rates for the white reference group were 10.5 per 1000 person years (95% confidence interval 10.4 to 10.6) for men and 7.3 per 1000 person years (7.2 to 7.3) for women. The highest age standardised rates were among south Asians groups—for example, for Bangladeshi people the rate was 24.4 per 1000 person years (19.8 to 29.0) for men and 11.3 per 1000 person years (8.5 to 14.1) for women. Age standardised rates were also high for Indian and Pakistani men and women compared with the white reference group. They were also higher for black Caribbean women and men from the "other Asian" group. In contrast, black African, Chinese, and black Caribbean men tended to have lower rates, as did black African women (table 2).

Characteristics of events

Table 3 shows the characteristics of events among men and women by ethnic group. Overall, 30.8% of events were stroke or transient ischaemic attacks, but this varied between ethnic groups. For example in the derivation dataset, 48.9% of first events among black Caribbean men and 36.4% among black African men were stroke or transient ischaemic attacks; the corresponding figures for women were 33.5% and 24.2%.

Prevalence of risk factors by ethnicity

Table 4 shows the distribution of risk factors, standardised for age, among each of the main ethnic groups. There was substantial heterogeneity across the ethnic groups in risk factors for cardiovascular disease and this also differed between men and women within an ethnic group. The notable results include differences in the age standardised prevalence of smoking among men of Bangladeshi (53.2%, 50.2% to 56.2%), Caribbean (40.6%, 38.9% to 42.4%), Pakistani (32.9%, 30.8% to 35.1%), white/not recorded (32.2%, 32.1% to 32.3%), Chinese (28.0%, 24.6% to 31.4%), Indian (23.7%, 22.3% to 25.1%), and black African (16.6%, 15.1% to 18.2%) origin. Current smoking rates were all lower for women in each ethnic group compared with men but varied widely between women from different groups.

There were also substantial differences in the age standardised prevalence of type 2 diabetes between ethnic groups with highest rates among Bangladeshis

Table 1 | Characteristics of patients aged 35-74 in derivation and validation cohorts in QRESEARCH database (version 19) 1993-2008

	Derivation cohort		Validation cohort	
	No (%) of women	No (%) of men	No (%) of women	No (%) of men
No of patients	773 291	762 292	375 763	374 469
Total person years observation	5 645 104	5 280 571	2 594 842	2 470 729
Median age (IQR)	49 (41-60)	48 (40-58)	49 (41-59)	47 (40-57)
Ethnicity:				
White or not recorded	752 241 (97.3)	743 159 (97.5)	363 516 (96.7)	363 097 (97.0)
Indian	3635 (0.47)	3693 (0.48)	2241 (0.60)	2200 (0.59)
Pakistani	2035 (0.26)	2033 (0.27)	1114 (0.30)	1246 (0.33)
Bangladeshi	1213 (0.26)	1269 (0.17)	611 (0.16)	723 (0.19)
Other Asian	1802 (0.16)	1422 (0.19)	1086 (0.29)	988 (0.26)
Black Caribbean	3928 (0.51)	3109 (0.41)	1870 (0.50)	1495 (0.40)
Black African	3655 (0.47)	3316 (0.44)	2423 (0.64)	2201 (0.59)
Chinese	1128 (0.15)	859 (0.11)	675 (0.18)	478 (0.13)
Other including mixed	3654 (0.47)	3432 (0.45)	2227 (0.59)	2041 (0.55)
Risk factors:				
Ethnicity recorded	209 214 (27.1)	181 110 (23.8)	108 540 (28.9)	94 522 (25.2)
BMI recorded	622 741(80.5)	562 278 (73.8)	304 084 (80.9)	274 403 (73.3)
Smoking recorded	703 574 (91.0)	650 460 (85.3)	344 194 (91.6)	319 800 (85.4)
Cholesterol/HDL ratio recorded	265 402 (34.3)	247 116 (32.4)	210 638 (56.1)	125 037 (33.4)
Systolic blood pressure recorded	711 935 (92.1)	647 782 (85.0)	344 967 (91.8)	313 125 (83.6)
Complete BMI and smoking	615 301 (79.6)	554 070 (72.7)	301 016 (80.1)	270 956 (72.4)
Positive family history of CHD	97 448 (12.6)	73 740 (9.7)	48 610 (12.9)	36 761 (9.8)
Current smoker	176 202 (22.8)	208 913 (27.4)	88 672 (23.6)	104 829 (28.0)
Treated hypertension	55 069 (7.12)	42 607 (5.59)	25 953 (6.91)	20 083 (5.36)
Type 2 diabetes	13 127 (1.70)	17 107 (2.24)	6186 (1.65)	8179 (2.18)
Rheumatoid arthritis	7187 (0.93)	2996 (0.39)	3310 (0.88)	1380 (0.37)
Atrial fibrillation	2692 (0.35)	1880 (0.25)	1242 (0.33)	2155 (0.58)
Chronic kidney disease	1227 (0.16)	1117 (0.15)	621 (0.17)	498 (0.13)

IQR=interquartile range; BMI=body mass index; HDL=high density lipoprotein cholesterol; CHD=coronary heart disease.

Table 2 | Crude and age standardised cardiovascular disease incidence rate per 1000 person years with 95% confidence intervals by age, sex, deprivation, and ethnicity in derivation dataset

	Women				Men			
	Total person years	No of incident cases	Crude incidence rate	Age standardised rates (95% CI)	Total person years	No of incident cases	Crude incidence rate	Age standardised rates (95% CI)
Total	5 645 105	41 042	7.27	—	5 280 571	55 667	10.54	—
Age (years):								
35-44	1 902 715	2590	1.36	—	1 971 609	5472	2.78	—
45-54	1 648 885	6823	4.14	—	1 621 901	13 076	8.06	—
55-64	1 192 905	12 438	10.43	—	1 047 287	18 281	17.46	—
65-74	900 599	19 191	21.31	—	639 775	18 838	29.44	—
Ethnic group:								
White/not recorded	5 537 244	40 278	7.27	7.25 (7.18 to 7.32)	5 190 709	54 705	10.54	10.53 (10.44 to 10.62)
Indian	21 654	186	8.59	10.88 (9.23 to 12.52)	20 150	285	14.14	16.88 (14.84 to 18.91)
Pakistani	10 981	115	10.47	13.24 (10.63 to 15.85)	9726	175	17.99	20.94 (17.75 to 24.13)
Bangladeshi	6707	67	9.99	11.30 (8.46 to 14.14)	5976	119	19.91	24.43 (19.83 to 29.03)
Other Asian	8097	45	5.56	8.41 (5.55 to 11.27)	5725	75	13.10	15.44 (11.84 to 19.03)
Black Caribbean	25 126	209	8.32	9.72 (8.35 to 11.09)	18 888	141	7.46	7.02 (5.80 to 8.24)
Black African	12 869	33	2.56	3.78 (2.25 to 5.31)	11 014	44	3.99	6.21 (4.12 to 8.30)
Chinese	5863	18	3.07	4.92 (2.52 to 7.31)	4585	17	3.71	5.40 (2.51 to 8.29)
Other	16563	91	5.49	8.43 (6.56 to 10.29)	13 798	106	7.68	10.27 (8.26 to 12.27)
Fifth of Townsend deprivation score:								
1	1 487 418	8488	5.71	6.01 (5.88 to 6.13)	1 374 821	13 181	9.59	9.43 (9.27 to 9.59)
2	1 283 177	7972	6.21	6.43 (6.29 to 6.57)	1 186 443	11 845	9.98	9.89 (9.72 to 10.07)
3	1 148 111	8599	7.49	7.35 (7.19 to 7.50)	1 060 665	11 456	10.80	10.70 (10.51 to 10.90)
4	970 900	8709	8.97	8.57 (8.39 to 8.75)	900 720	10 381	11.53	11.61 (11.39 to 11.83)
5*	755 499	7274	9.63	9.60 (9.38 to 9.82)	757 923	8804	11.62	12.38 (12.13 to 12.64)

*Most deprived.

(14.4% women, 16.8% men), Pakistanis (14.2% women, 12.0% men), and Indians (11.7% women, 13.3% men) and lowest among the white reference group (1.5% women, 2.1% men).

Treated hypertension was highest among Caribbean and black African men and women. Recorded family history of coronary heart disease in a first degree relative was highest among Indian men and women and lowest among black African men and women.

Model development

Table 5 shows the results of the Cox regression analysis for the QRISK2 model. We used a log transformation for age but otherwise fitted variables as linear terms as this provided a better fit with the data according to the fractional polynomial analysis. The table shows variables that had significant interactions with age and these indicate increased hazard ratios for the risk factors among younger patients compared with older patients (fig 1).

Calibration and discrimination of QRISK2

The QRISK2 model was marginally superior to the original QRISK1 equation and both models were superior to the modification of the Framingham score for the D statistic, ROC statistic, and the R^2 value—for both men and women (table 6). For example, the QRISK2 algorithm explained 43% of the variation in women and 38% in men. The figures for modified Framingham score were 39% and 35%, respectively. Also, as an example, the D statistic was 1.79 (1.77 to 1.82) in women and 1.62 (1.59 to 1.64) in men for the QRISK2 model compared with 1.63 (1.61 to 1.66) and 1.50 (1.47 to 1.52) for the modified Framingham score. All three scores performed better in women than in men.

Figure 2 compares predicted and observed risks of a cardiovascular disease event at 10 years across each 10th of predicted risk (first 10th representing the lowest risk). This shows that the QRISK2 model is better calibrated than the modified Framingham score.

Predictions with age, sex, deprivation, and ethnicity

Table 7 shows the breakdown of patients by age and sex with a predicted 10 year risk of 20% or more with the QRISK2 model and the modified Framingham score. Overall, the QRISK2 model would predict 10.4% of patients as high risk compared with 14.9% for the modified Framingham score.

Figure 3 shows the proportion of patients estimated to be at high risk with QRISK2 and the Framingham score within each ethnic group. QRISK2 would identify 14.2% (11.5% to 17.0%) of Bangladeshi and 10.1% (8.8% to 11.3%) of Indian women at high estimated risk compared with 7.2% (5.1% to 9.3%) and 4.6% (3.7% to 5.5%) with the Framingham score, respectively. QRISK2 would identify 14.0% (13.9% to

Table 3 | Characteristics of first cardiovascular events in derivation and validation cohort. Figures are numbers of events (percentage of stroke or transient ischaemic attacks)

Ethnicity	Derivation cohort	Validation cohort
Women		
White/not recorded	40 278 (35.3)	17 677 (34.8)
Indian	186 (21.0)	112 (23.2)
Pakistani	115 (18.3)	76 (34.2)
Bangladeshi	67 (31.3)	36 (19.4)
Other Asian	45 (28.9)	26 (26.9)
Black Caribbean	209 (33.5)	73 (42.5)
Black African	33 (24.2)	32 (40.6)
Chinese	18 (11.1)	14 (57.1)
Other	91 (31.9)	55 (29.1)
Men		
White/not recorded	54 705 (27.6)	24 626 (27.2)
Indian	285 (19.6)	186 (22.6)
Pakistani	175 (16.6)	115 (20.9)
Bangladeshi	119 (13.4)	83 (16.9)
Other Asian	75 (20.0)	56 (23.2)
Black Caribbean	141 (48.9)	78 (35.9)
Black African	44 (36.4)	43 (27.9)
Chinese	17 (23.5)	13 (46.1)
Other	106 (24.5)	95 (28.1)

14.1%) of white men at high risk compared with 22.0% (21.9% to 22.1%) with the Framingham score.

Reclassification statistics

Of the 112 156 patients classified as high risk (risk of ≥20% over 10 years) with the Framingham score, 46 094 (41.1%) would be reclassified at low risk with QRISK2. The 10 year observed risk among these reclassified patients was 16.6% (16.1% to 17.0%)—that is, below the 20% threshold for high risk.

Of the 78 024 patients classified at high risk with QRISK2, 11 962 (15.3%) would be reclassified as low risk with the Framingham score. The 10 year observed risk among these patients predicted to be at high risk with QRISK2 was 23.3% (22.2% to 24.4%)—that is, above the 20% threshold for high risk.

The annual incidence rate of cardiovascular events among those with a QRISK2 score of ≥20% was 30.6 per 1000 person years (95% confidence interval 29.8 to 31.5) for women and 32.5 per 1000 person years (31.9 to 33.1) for men. Both these figures are higher than the annual incidence rate for patients identified as high risk with the modified Framingham score. The annual incidence rate for these patients was 25.7 per 1000 person years (25.0 to 26.3) for women with 26.4 (26.0 to 26.8) for men. In other words, at the 20% threshold, the population identified by QRISK2 was at higher risk of a CV event than the population identified by the Framingham score.

Clinical examples

Table 8 shows some clinical examples for patients from different ethnic groups who would be reclassified with QRISK2 compared with the modified Framingham score. We have calculated 95% confidence intervals around the QRISK2 score. For example, a 64 year old Indian woman from a moderately deprived area with a systolic blood pressure of 130, BMI of 23.1, treated

Table 4 | Distribution of cardiovascular disease risk factors by ethnic group in men and women in derivation cohort; figures are age standardised

	Means					Percentages						
	Age at entry	Townsend score	Systolic blood pressure	BMI	Total cholesterol/ HDL ratio	Family history of CHD	Current smoker	Treated hyperten-sion	Type 2 diabetes	Rheumatoid arthritis	Atrial fibrillation	Chronic renal disease
Women												
White/not recorded	54	−0.65	133	26.1	3.9	12.7	25.4	7.0	1.5	0.9	0.4	0.2
Indian	50	1.06	130	26.4	4.0	18.5	8.1	9.6	11.7	1.3	0.2	0.2
Pakistani	49	2.58	129	28.4	4.3	12.1	5.8	7.7	14.2	1.4	0.4	0.5
Bangladeshi	49	5.77	126	26.1	4.6	6.3	12.8	9.5	14.4	0.9	0.2	0.3
Other Asian	49	2.09	130	25.4	4.0	11.2	10.5	9.6	7.1	0.4	0.2	0.4
Black Caribbean	51	3.66	136	28.7	3.5	9.1	15.9	24.8	10.3	1.0	0.2	0.3
Black African	48	4.19	138	29.4	3.6	4.5	4.4	17.3	6.8	0.3	0.2	0.1
Chinese	49	1.95	125	23.4	3.7	6.2	6.1	9.8	5.2	0.8	0.1	0.4
Other	49	2.80	131	27.0	3.8	11.3	18.2	13.6	6.4	1.0	0.1	0.4
Men												
White/not recorded	52	−0.5	136	26.6	4.5	9.7	32.2	5.5	2.1	0.4	0.6	0.1
Indian	48	1.1	133	25.7	4.6	16.8	23.7	12.0	13.3	0.4	0.3	0.4
Pakistani	48	2.6	131	26.3	4.9	13.0	32.9	5.7	12.0	0.2	0.2	0.2
Bangladeshi	48	5.5	126	25.0	5.2	5.9	53.2	6.4	16.8	0.8	0.1	0.2
Other Asian	48	2.2	132	25.5	4.7	9.1	28.3	8.3	10.3	0.2	0.3	0.1
Black Caribbean	53	3.7	136	26.7	3.9	6.4	40.6	14.0	9.0	0.3	0.3	0.4
Black African	47	4.3	139	26.6	4.0	3.2	16.6	15.6	8.6	0.3	0.3	0.6
Chinese	50	2.5	127	24.0	4.3	4.2	28.0	6.1	3.8	0.0	0.3	0.3
Other	49	3.0	134	26.5	4.4	8.5	34.4	9.3	7.9	0.4	0.3	0.3

BMI=body mass index; HDL=high density lipoprotein cholesterol; CHD=coronary heart disease.

Table 5 | Adjusted hazard ratios (95% CI) for cardiovascular disease for QRISK2 model in derivation cohort (see figure 1 for effect of age on relevant hazard ratios where there are age interactions)

	Women	Men
White/not recorded	1	1
Indian	1.43 (1.24 to 1.65)	1.45 (1.29 to 1.63)
Pakistani	1.80 (1.5 to 2.17)	1.97 (1.70 to 2.29)
Bangladeshi	1.35 (1.06 to 1.72)	1.67 (1.40 to 2.01)
Other Asian	1.15 (0.86 to 1.54)	1.37 (1.09 to 1.72)
Black Caribbean	1.08 (0.94 to 1.24)	0.62 (0.53 to 0.73)
Black African	0.58 (0.42 to 0.82)	0.63 (0.47 to 0.85)
Chinese	0.69 (0.44 to 1.10)	0.51 (0.32 to 0.83)
Other	1.04 (0.85 to 1.28)	0.91 (0.75 to 1.10)
Age (10% increase)*	1.66 (1.65 to 1.68)	1.59 (1.58 to 1.60)
BMI (5 unit increase)	1.08 (1.06 to 1.10)	1.09 (1.07 to 1.11)
Townsend score (5 unit increase)	1.37 (1.34 to 1.40)	1.18 (1.16 to 1.20)
Systolic blood pressure (mm Hg) (20 unit increase)	1.20 (1.18 to 1.22)	1.19 (1.17 to 1.20)
Cholesterol/HDL ratio	1.17 (1.16 to 1.18)	1.19 (1.18 to 1.20)
Family history coronary heart disease	1.99 (1.92 to 2.05)	2.14 (2.08 to 2.20)
Current smoker	1.80 (1.75 to 1.86)	1.65 (1.60 to 1.70)
Treated hypertension	1.54 (1.45 to 1.63)	1.68 (1.60 to 1.77)
Type 2 diabetes	2.54 (2.33 to 2.77)	2.20 (2.06 to 2.35)
Rheumatoid arthritis	1.50 (1.39 to 1.61)	1.38 (1.25 to 1.52)
Atrial fibrillation	3.06 (2.39 to 3.93)	2.40 (2.07 to 2.79)
Renal disease	1.70 (1.43 to 2.03)	1.75 (1.51 to 2.02)
Age* BMI interaction	0.976 (0.970 to 0.982)	0.985 (0.979 to 0.991)
Age* Townsend interaction (5 unit increase in score)	0.938 (0.930 to 0.946)	0.973 (0.967 to 0.98)
Age* systolic blood pressure interaction (20 unit increase in systolic blood pressure)	0.966 (0.961 to 0.971)	0.964 (0.96 to 0.969)
Age* family history interaction	0.927 (0.914 to 0.94)	0.923 (0.912 to 0.935)
Age* smoking interaction	0.931 (0.920 to 0.943)	0.932 (0.922 to 0.942)
Age* treated hypertension interaction	0.952 (0.934 to 0.971)	0.916 (0.901 to 0.931)
Age* type 2 diabetes interaction	0.904 (0.877 to 0.931)	0.902 (0.881 to 0.924)
Age* atrial fibrillation interaction	0.858 (0.795 to 0.926)	0.893 (0.852 to 0.935)

BMI=body mass index; HDL=high density lipoprotein cholesterol.
*All age terms expressed as 10% increase in age (for example, 50 to 55 years).

hypertension, and cholesterol/HDL ratio of 5.3 would have a modified Framingham score of 12.0%, but a QRISK2 score of 24.7% (24.4% to 25.0%) at 10 years. A 54 year old Bangladeshi man who is a non-smoker and has treated hypertension, a systolic blood pressure of 142 mm Hg, a BMI of 27.0, and a cholesterol ratio of 4.2 and lives in one the most deprived areas would have a modified Framingham score of 17.0% (including the adjustment for being south Asian) but a 10 year QRISK2 score of 23.5% (22.8% to 24.1%).

DISCUSSION

We developed and validated a cardiovascular risk algorithm that simultaneously takes account of ethnicity and deprivation. The algorithm has face validity in the setting in which it will be used and had good discrimination and calibration. There are three main reasons why this study is likely to make an important impact on the decisions of doctors, patients, and commissioners. Firstly, in this prospective study we developed and validated a risk prediction algorithm that provides an individualised estimate of cardiovascular risk and includes the independent contributions of ethnicity and deprivation. This permits identification of those individuals and groups likely to be most disadvantaged by use of existing treatment algorithms. Such patients include south Asian women, who would otherwise be less likely to be identified. This information will, if acted on, help to reduce health inequalities.

Secondly, it extends and improves on our original equation for cardiovascular risk[12] by incorporating important additional clinical conditions (such as rheumatoid arthritis, chronic kidney disease, and atrial fibrillation), allowing more accurate quantification of risks for individual patients. This information should be considered in the context of specific treatment guidelines. Knowledge of cardiovascular risk might be useful in assessing response efficacy and concordance with recommended healthcare interventions for these specific conditions.

Thirdly, it also allows better quantification of risk of cardiovascular disease for patients with type 2 diabetes, which is especially prevalent among south Asian patients. Though there are alternative cardiovascular risk algorithms for patients with diabetes,[18 48] none is based on a large nationally representative primary care cohort, has large numbers of incident events, and also simultaneously takes account of other important risk factors such as deprivation and ethnicity. Although current guidelines might indicate statins for people with diabetes, knowledge of cardiovascular risk can be useful in helping to identify patients at particularly low risk for whom a statin might not be needed.

Strengths and limitations

The strengths and limitations of using this approach and the QRESEARCH database to develop and validate a new risk prediction algorithm have been discussed previously.[12 27]

We included more sophisticated modelling of the effect of age on risk factors, which results in greater weighting of some risk factors in younger patients, such as smoking status, family history of coronary heart disease, type 2 diabetes, systolic blood pressure, treated hypertension, BMI, deprivation, and atrial fibrillation. This also has the effect that in people without the risk factors the increase in risk with age will be steeper than with QRISK1. The inclusion of patients with type 2 diabetes in the main study population will have tended to increase the overall level of risk in the study population and this will also have tended to increase the risk for an individual, as can be seen from the hazard ratios (table 5).

We updated the analysis to include data until March 2008, increasing the number of patients with at least 10 years of follow-up data to almost 440 000 patients. We have furthermore included the linked cause of death as recorded by the Office for National Statistics (ONS). Death linkage increased cases of cardiovascular disease by about 7% across the entire study period, as

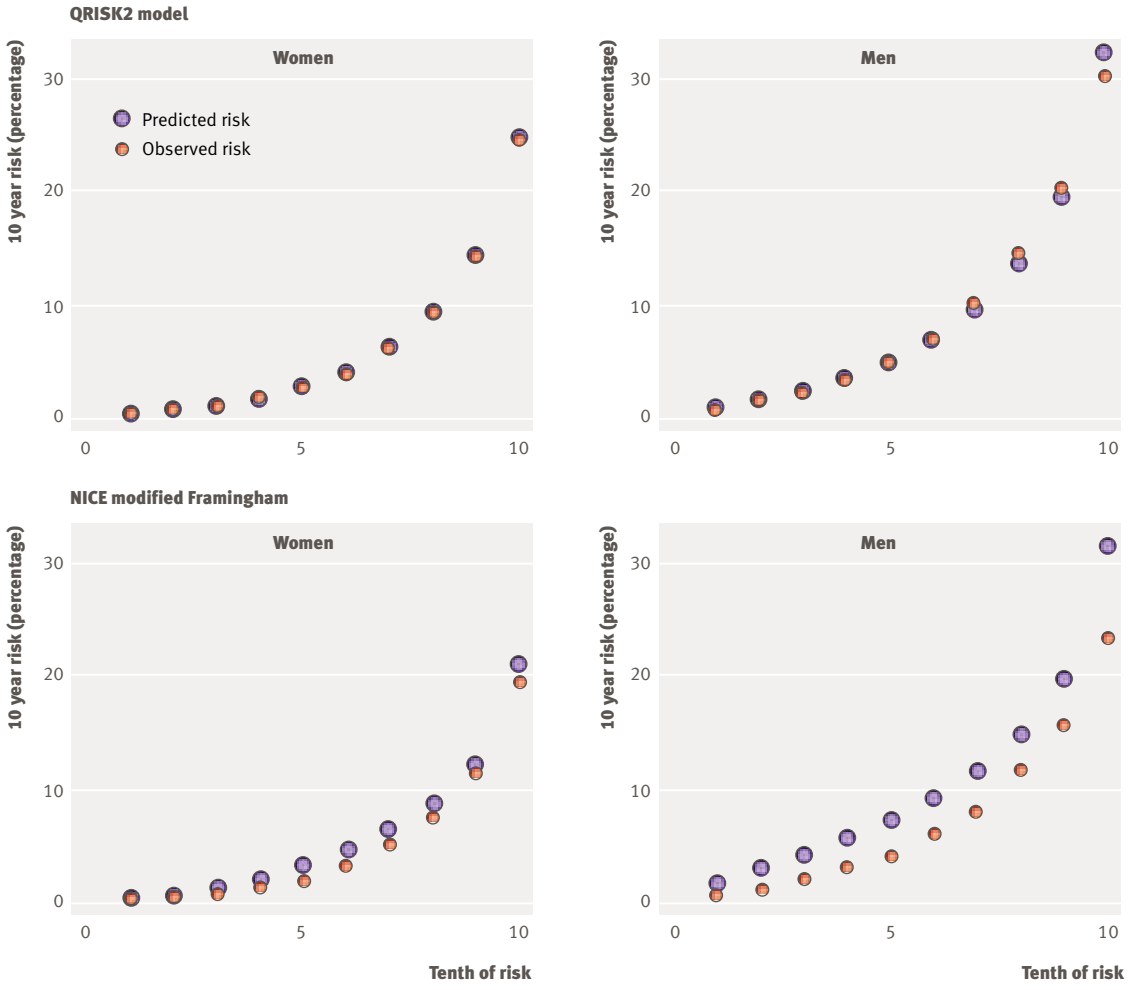

Fig 2 | Predicted and observed risk by 10th of predicted risk for QRISK2 model and NICE modification of score in validation dataset

the data from ONS were not available for the full study period at the time of the original study.[27]

We used self assigned ethnicity as reported by the patient to their general practice; this has advantages over analyses where ethnicity is assigned by an informant rather than the patient or is imputed geographically or is related to country of birth. The latter is particularly problematic with increasing numbers of people from ethnic minorities now being born in the UK.[49] We also disaggregated the south Asian groups and reported on them separately, which addresses concerns with studies that tend to combine them into one group when there are differences in exposure to risk factors and rates and outcomes of diseases.[25] Though only a quarter of patients had self assigned ethnicity recorded, we think it is reasonable to assume that where patients have self assigned ethnicity recorded as Bangladeshi (for example) that this is accurate and the patient was indeed Bangladeshi. Misclassification would most affect the reference category of "white or not recorded," but because of the mix of the populations of England and Wales less than 10% of such patients were probably from a non-white ethnic group. This misclassification would therefore, if anything, tend to underestimate the relative effect of ethnicity on cardiovascular risk.

Just fewer than 3% of our total sample were classified as belonging to a minority ethnic group compared with the national proportion in this age group of 6.6% (based on projections for 2006[50]). The comparison, however, is not "like for like" as national estimates are for 2006 and migration patterns and population demographics have probably changed over the 15 year period of our study. None the less, the lower percentage of patients

Table 6 | Validation statistics for new QRISK2 model compared with modified NICE equation in validation cohort. Figures are means (95% confidence intervals)

	QRISK2 model	QRISK1 model	Modified Framingham equation
Women			
R^2	43.47 (42.78 to 44.16)	42.94 (42.23 to 43.66)	38.87 (38.12 to 39.62)
D statistic	1.795 (1.769 to 1.820)	1.776 (1.750 to 1.801)	1.632 (1.606 to 1.658)
ROC statistic	0.817 (0.814 to 0.820)	0.814 (0.811 to 0.817)	0.800 (0.797 to 0.803)
Brier score	0.086 (0.083 to 0.089)	0.081 (0.078 to 0.084)	0.093 (0.090 to 0.096)
Men			
R^2	38.38 (37.75 to 39.01)	37.63 (36.99 to 38.27)	34.78 (34.12 to 35.45)
D statistic	1.615 (1.594 to 1.637)	1.590 (1.568 to 1.612)	1.495 (1.473 to 1.517)
ROC statistic	0.792 (0.789 to 0.794)	0.788 (0.786 to 0.791)	0.779 (0.776 to 0.782)
Brier score	0.136 (0.134 to 0.139)	0.128 (0.125 to 0.131)	0.177 (0.174 to 0.180)

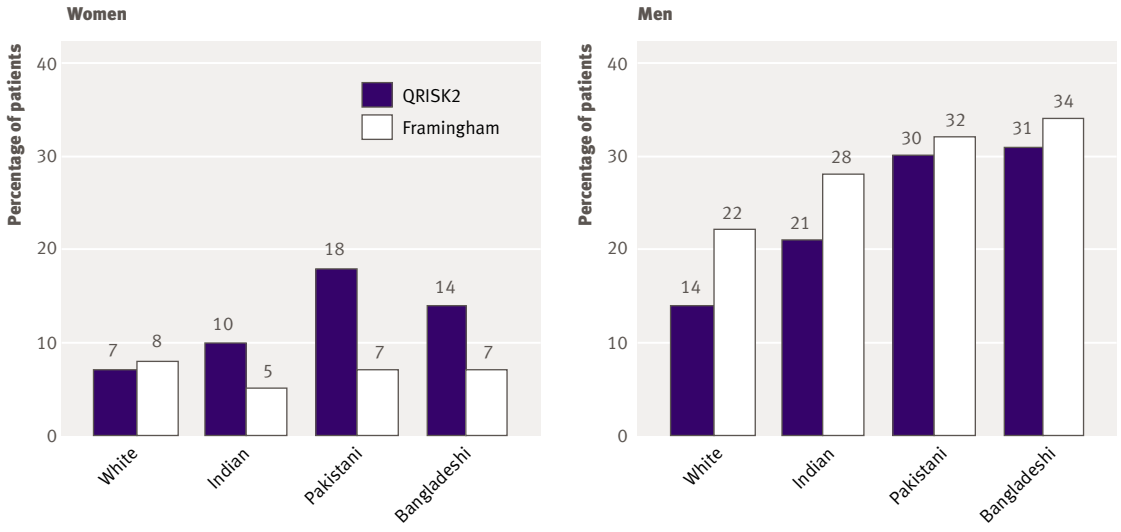

Fig 3 | **Percentage of white and south Asian patients at high risk in validation dataset with QRISK2 and Framingham score in QRESEARCH database (in modified Framingham score, inflation factor of 1.4 is applied to south Asian men but not south Asian women)**

from minority groups raises concerns about the possible under-representativeness of practices from ethnically diverse inner city areas or misclassification error, or both. We think under-representativeness of practices from ethnically diverse areas is unlikely as QRESEARCH practices are drawn from across England and Wales and have been shown to be similar to practices nationally for a range of measures.[26] In fact, QRESEARCH has proportionately more practices in areas of higher ethnicity such as the East Midlands, Yorkshire, and Humberside (fig 4. Also, table 1 shows that among patients from both cohorts, when ethnicity was recorded 11.7% were from a minority group. This is higher than from census estimates for 2006, indicating either over-representation of practices from ethnically diverse areas or that practices in ethnically diverse areas are more likely to record ethnicity, or both. Therefore, the reason for the apparent under-representation of people from black and minority ethnic groups has arisen is probably because we combined thenot recorded and the white groups. This combined group will contain additional patients from groups classified as other than white. This would, if non-differential, result in a bias towards the null hypothesis of no difference in risk between ethnic groups. The net consequence of this would be, if anything, to underestimate hazard ratios in the minority ethnic populations in question rather than generate spurious associations.

With a number of policy and legislative drivers co-aligning, ethnicity coding is likely to improve exponentially in the UK, and this evolving picture will therefore allow us to continue to monitor the impact of incorporating more complete ethnicity data into our models. But for the present, even though it is imperfect, incorporating ethnicity into our disease risk algorithm has, we believe, clearly been an important advance in understanding risk of disease in ethnically diverse populations. Furthermore, it is unlikely that a better estimate could be obtained for England and Wales given the difficulties of assembling a sufficiently large prospective cohort for follow-up over 10 or more years.

Another potential limitation of our study is that we have assumed that the absence of a recorded diagnosis of diabetes (or family history, for example) is equivalent to the person not having that factor. This is probably valid for diabetes as there have been consistent efforts in general practice over the past 15 years to develop and validate diabetes registers (including comparisons against prescribed medication for diabetes), though we accept there will additionally be large numbers of cases not yet diagnosed by clinicians. Recording of family history is less systematic in primary care and might be more susceptible to recording bias. As recording of risk factors becomes more complete over time, then better estimates of the relevant hazard ratios will be possible.

Also relevant is that we have calculated 95% confidence intervals around the QRISK2 scores to give a better idea of precision. We have improved on the method for validation by using multiple imputation for missing values in the validation set rather than mean values by age and sex derived from the derivation dataset as in our original study and independent validation.[12 27] One important limitation, though, is that while we have validated the results in a physically discrete group of practices, these practices all use the same EMIS clinical system and hence there is a potential "home advantage" that might reduce the generalisability to other systems, although, conversely, it is ideally suited for use in the EMIS system. In other words, any comparison done in the one third sample of practices in QRESEARCH will tend to favour QRISK2 compared with other prognostic scores. Our previous study[27] was additionally validated in a database (THIN, "The Health Improvement

Network") derived from a set of practices using a different clinical system (In Practice Systems) and gave similar results (apart from the prevalence of family history, which was lower in the THINdatabase). This suggests that our findings are probably generalisable to the 20% of practices in England and Wales that use In Practice Systems in addition to the 60% of practices that already use the EMIS clinical system from which the equation is derived. Further validation of QRISK2 is not currently possible on the THIN database as the database does not have the linked ONS death certificate data and recording of ethnicity is too low (personal communication, THIN, 2008). The validation we have presented constitutes the best currently possible given the extent and nature of comparable datasets. The results should generalise to at least 80% of practices nationally. None the less, it is important that QRISK2 is validated by another team on external populations and an international version of QRISK2 is being developed to allow this and will be reported in due course. In particular, we are working with another primary care database (THIN) to link their data to ONS death certificate data so that this can be used as a data source for further validation. Ethnicity recording could be improved on primary care databases by linkage of individual level data on self assigned ethnicity from the 2001 census, and this will be undertaken and reported, assuming access to these data is granted.

Comparisons with the modified Framingham score

This study improves on our original equation for cardiovascular risk in terms of its potential application as outlined above and also because the more complex model has slightly better discrimination (that is, greater ability to separate patients at high and low risk) than our original model. The QRISK1 equation improved on other equations in use in the UK by including additional readily available risk factors such as deprivation, family history, BMI, and blood pressure treatment. With QRISK2, the improvement in discrimination and calibration compared with the modified Framingham score remains significant, although this is probably partly because the modelling was undertaken on a more contemporaneous population from England and Wales and we used a more sophisticated approach for modelling and included additional variables. We have not compared QRISK2 with the most recently published Framingham score as this uses a much broader definition of cardiovascular disease that is less relevant to UK guidelines.[51] QRISK2 seems to improve on the Framingham score based Ethrisk,[19] perhaps because of its greater precision, larger sample, and prospective study design.

In contrast to our previous study, we compared QRISK2 with the modified Framingham risk score recently recommended by NICE. The modified score, in common with the risk equation advocated by the Joint British Societies, involves summing risks from two risk equations for coronary heart disease and stroke, which is mathematically incorrect because these are not independent outcomes and therefore will give an invalid result. This addition of the two separate and non-independent risks results in some patients having an estimated risk of more than 100% and would also result in overestimation of risk for other individuals at lower estimates of risk. This might have accounted for some of the overprediction. The inflation factors of 1.4 for south Asian men and 1.5 for those with a family history coronary heart disease, which have been developed by consensus rather than a mathematical model based on individual patient data, might also have accounted for some of the overprediction, although this was still present on our previous analysis where the inflation factors had not been applied.[12 27]

Comparisons with the literature

We found substantial heterogeneity between risk factors within south Asian populations and our prevalence figures for risk factors are comparable with the literature,[19 20] which increases the face validity of our findings. For example, as others have found, Bangladeshi men have higher rates of smoking but lower mean systolic blood pressure levels than Pakistani or Indian men.[20] Indian and Pakistani men and women have higher mean BMI than Bangladeshis.[20] Prevalence of type 2 diabetes was higher in Bangladeshis and Pakistanis than Indians.[20] Similarly, cholesterol/HDL ratio was higher among each of the south Asian groups compared with the white reference category.[20] Our findings also confirm

Table 7 | Number and percentage of patients in validation cohort with estimated cardiovascular risk of ≥20% by five year age band with QRISK2 compared with modified Framingham equation

Age (years)	Total population	QRISK2 model (%)	Modified Framingham equation
Women			
35-39	78 887	16 (0.02)	4 (0.01)
40-44	62 153	56 (0.09)	88 (0.14)
45-49	55 879	193 (0.35)	636 (1.14)
50-54	49 392	517 (1.05)	2048 (4.15)
55-59	40 183	1228 (3.06)	3892 (9.69)
60-64	33 831	2971 (8.78)	5952 (17.59)
65-69	29 045	6657 (22.92)	7870 (27.10)
70-74	26 393	13 988 (53.00)	9581 (36.30)
Men			
35-39	85 424	97 (0.11)	289 (0.34)
40-44	67 911	345 (0.51)	1570 (2.31)
45-49	58 085	910 (1.57)	4703 (8.10)
50-54	49 171	2336 (4.75)	9635 (19.59)
55-59	38 955	4986 (12.90)	14 330 (36.79)
60-64	30 849	9649 (31.28)	17 008 (55.13)
65-69	24 924	15 905 (63.81)	18 117 (72.69)
70-74	19 150	18 170 (94.88)	16 433 (85.61)
Totals			
Women	375 763	25 626 (6.82)	30 071 (8.00)
Men	374 469	52 398 (13.99)	82 085 (21.92)
Both	750 232	78 024 (10.40)	112 156 (14.95)

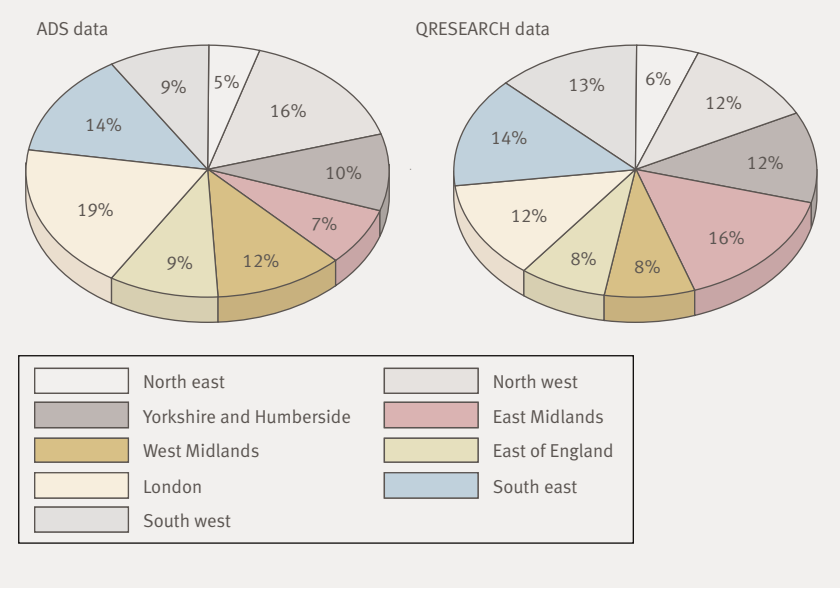

Fig 4 | **Proportion of practices in England by geographical region in national attribution dataset (Department of Health) and in QRESEARCH on 31 March 2004**

Nazroo's observations[52] and the findings of the Whitehall II study[53] of the independent effects of both ethnicity and deprivation. Overall, the results of our study add to a growing body of evidence that combining people of south Asian origin into one category is potentially misleading.

The magnitude of the increased cardiovascular risk among south Asians compared with white patients seems to be higher than the 40% previously thought in the absence of prospective incidence data.[22 24] For example, in our study, compared with the white reference group the adjusted risk is 45% higher (29% to 63%) among Indian men, 67% higher (40% to 101%) among Bangladeshi men, and 97% higher (70% to 129%) among Pakistani men, even after adjustment for multiple confounders including deprivation and diabetes. Similarly, the adjusted risks for Indian, Pakistani, and Bangladeshi women are all increased compared with the white reference population. Our results also suggest that the increased cardiovascular risks observed for Pakistani men are significantly higher than those for Indian men. The difference between these two groups for women is similar, although of borderline significance when a direct comparison is made, probably because of a lack of power.

There were also differences in the proportion of events that were stroke or transient ischaemic attacks rather than coronary heart disease. For example, a high proportion of first events among black Caribbean and black Africans was stroke or transient ischaemic attacks, which is consistent with the literature.[54 55] Other studies have found differences in mortality between different ethnic groups, such as the unexplained persistent higher mortality among Bangladeshis.[56] This deserves further study as to the underlying causes and potential missed opportunities for care.

Clinical implementation

QRISK2 has been designed to estimate cardiovascular risk for an entire population of patients in primary care by using data already collected within the patient's electronic health record and by using default values for body mass index, cholesterol concentration, and systolic blood pressure where these data have not been recorded in the past five years. Computer generated risk scores have been integrated within routine clinical use of computers in UK primary care for the past 10 years, and, with QRISK2 embedded within computer applications, a rank ordered recall list can be generated so that those at greatest clinical need can be recalled first. Once such patients have been recalled, the individual can have a full clinical cardiovascular check to calculate an actual QRISK2 based on the most up to date data that are then used to guide decisions about treatment.

The only item in QRISK2 that is not already routinely collected and recorded electronically is the Townsend deprivation score, which is linked to an individual postcode. This score has already been integrated into the EMIS clinical system and linked to the records of over 32 million patients. The mapping of postcode to deprivation score will also be made available, together with the supporting reference tables and algorithm itself. QRISK2 can then be integrated within clinical management systems so that it can be used on an ongoing basis to generate an estimated score based on existing data. QRISK2 will be updated as improved analytical techniques are developed for application to the QRESEARCH database. QRISK will evolve as data quality and completeness improves and population characteristics change (obesity is increasing, while incidence of cardiovascular, for example). This will ensure that future versions of QRISK remain well calibrated to the population of England and Wales and makes best use of technical developments. Lastly, the NHS' electronic health record(NHS Care Record Service) is central to the NHS Connecting for Health's national programme for information technology and this will, within a relatively short space of time, result in electronic health records replacing paper based records in hospitals in England.[57] The plan is for these eventually to incorporate computerised decision support tools and so this will allow disease risk algorithms such as QRISK2 to be largely automatically populated with routine electronically coded data as is already possible in primary care in the UK.

These estimates, like any predictive score, are an aid but not a replacement for judgment in individual clinical circumstances. We have specifically identified atrial fibrillation and rheumatoid arthritis for consideration as both are known to be associated with increased risk[31 32 58 59] and knowledge of them might inform clinical management for an individual patient. We recognise that the likely age and comorbidity of these individuals, however, might place them at being at high risk of cardiovascular disease and therefore not appropriate for a primary prevention tool such as

QRISK2. Nevertheless, if we had omitted rheumatoid arthritis and atrial fibrillation, the effect would be to underestimate risk for individuals with either of these two conditions who did not yet have concurrent cardiovascular disease. The prevalence of rheumatoid arthritis and atrial fibrillation is low so this will have a minimal impact on the overall precision of the model or its application at a population level, but we believe the additional complexity of the model is justified as no additional data entry will be required from most users, while it also provides relevant information to the individual patient with one or either of these conditions and their clinicians.

QRISK2 provides a mechanism for estimating absolute risk among individuals. Use of this information, however, should be tightly coupled with suitable guidelines. There are some patients in whom a QRISK2 score should not be calculated, including those with pre-existing cardiovascular disease (who we excluded from this study). Risk estimation should not be used for people with conditions such as peripheral vascular disease, heart failure, familial hypercholesterolaemia, or other conditions not specifically identified in the algorithm that are known to be associated with high risks of cardiovascular events.[5] We have not added further to the exclusions in this dataset as to do so would have added complexity with no appreciable gain in precision for people in whom we do not recommend the use of this score.

Clinical impacts and health inequalities

A risk prediction algorithm that does not include deprivation or ethnicity is likely to result in the inequitable definition of risk for affluent and deprived communities and also substantially underestimate the risk in south Asian people, especially women, in whom, like men, it is the commonest cause of premature death. Primary prevention programmes that do not take these variables into account risk exacerbating rather than reducing existing health inequalities,[6-8] especially as the evidence suggests that health inequalities naturally widen at the start of new health initiatives.[21] Other research highlights additional difficulties with accessing effective health promotion, including lack of risk awareness, influences of culture and lifestyle, time restrictions, and language difficulties[60] and this needs to be addressed once patients have been identified to improve clinical outcomes.

The QRISK2 algorithm, like its predecessor, has better calibration and is a better discriminator of risk of cardiovascular disease than the modified Framingham score. A major advantage of QRISK2 is the ability of the algorithm to be updated as population demographics, ethnic composition, prevalence of risk factors, and incidence of cardiovascular diseasechange. It also demonstrates the utility of linked electronic data for research to develop tools that can help doctors to make better decisions. The marked gradient with deprivation has already been demonstrated with QRISK1. The further identification of ethnicity as an independent factor additional to deprivation is an important consideration, particularly for south Asian women at high risk. A broader range of important clinical conditions included in QRISK2 but not in the modified Framingham score make it a more clinically relevant tool. Highlighting risks of conditions including type 2 diabetes and chronic renal disease supports further integration of vascular strategies and informs individual assessment.

The modified Framingham score underestimates risk in south Asian women. Like the earlier version, QRISK2 includes BMI and treatment for hypertension, neither of which are included in the Framingham score; in QRISK2, family history contributes an important additional weighting particularly at younger ages. The clinical relevance, superior performance, and equitable assignment of QRISK2 make it an appropriate tool to assist in the delivery of public health

Table 8 | Clinical examples for patients who would be reclassified with QRISK2 instead of NICE modified Framingham equation

Age (years)	Ethnic group	Family history	Systolic blood pressure	BMI	Cholesterol/HDL ratio	Smoker	Treated hypertension	Type 2 diabetes*	Chronic kidney disease	Townsend score†	Framingham score 10 year risk (%)	QRISK2 10 year risk (%) (95% CI)
Men												
65	Indian	Yes	100	24.7	3.3	No	No	No	No	5	17	31.3 (30.9 to 31.7)
54	Bangladeshi	No	142	27.0	4.2	No	Yes	No	No	10	17	23.5 (22.8 to 24.1)
54	Black African	No	150	21.0	7.3	No	No	No	No	4	23	9.0 (7.7 to 10.3)
55	Indian	No	156	27.0	4.7	No	No	No	No	−4	24	12.7 (12.2 to 13.2)
65	Caribbean	No	146	29.1	5.4	No	No	No	No	4	26	14.8 (14.2 to 15.5)
42	White	Yes	132	36.0	5.3	Yes	Yes	No	No	11	17	35.2 (34.9 to 35.5)
Women												
64	Indian	No	130	23.1	5.3	No	Yes	No	No	5	12	24.7 (24.4 to 25.0)
60	Bangladeshi	No	132	36.0	4.3	No	Yes	No	No	11	9	21.1 (20.6 to 21.6)
48	Pakistani	Yes	140	33.2	4.5	No	Yes	No	No	8	9	26.1 (25.7 to 26.4)
58	White	No	154	34.0	3.4	Yes	Yes	No	No	10	16	21.4 (21.3 to 21.5)

BMI=body mass index; HDL=high density lipoprotein cholesterol.
*NICE lipid modification guideline does not include diabetes so this is for illustrative purposes only.
†Interval score ranges between −6 (most affluent) and 11 (most deprived).

WHAT IS ALREADY KNOWN ON THIS TOPIC

A 10 year cardiovascular disease risk threshold of 20% is recommended for intervention with statins for the primary prevention of cardiovascular disease

Current algorithms for risk of cardiovascular disease do not adequately account for the combined effect of socioeconomic status and ethnicity, leading to an underestimate of risk in high risk populations that might potentially exacerbate existing health inequalities

WHAT THIS STUDY ADDS

Compared with a white reference population, there is a substantially increased risk of cardiovascular disease in south Asian men and women that is independent of social deprivation, diabetes, and family history

The results of the calibration and discrimination statistics for QRISK2 were significantly better than those for the modified Framingham score in the validation sample

At the 10 year risk threshold of 20%, the population identified by QRISK2 was at higher risk of a CV event than the population identified by the modified algorithm

programmes that recognise the broader determinants of cardiovascular health, such as ethnicity and deprivation. This has particular relevance to equity of delivery of health care to the UK's south Asian communities and might help to reduce widening health inequalities.

We acknowledge the contribution of David Stables (EMIS) and EMIS practices contributing to the QRESEARCH database. In particular we acknowledge his contribution in linking the ONS death certificate data to individual records held within EMIS clinical systems so that it could be extracted on to the QResearch database and used for this project. We thank Aneez Esmail (University of Manchester), Ruthie Birger and Chris Millett (Imperial College London), and Nadeem Qureshi (University of Nottingham) for ethnicity coding.

Contributors: JH-C initiated and designed the study, obtained approvals, prepared the data, undertook the analysis and interpretation, and wrote the first draft paper. CC and YV contributed to the development of the protocol, design, and analysis and interpretation and drafting of the paper. CC also undertook some of the primary analyses with JHC. JR and PB contributed to the conception, design, analysis, interpretation, and drafting of article and approved the final draft. RM and AS contributed to suggestions for analysis, drafting, interpretation, and approved the final draft. JH-C is the guarantor.

Funding: No external funding. The authors were funded as part of their clinical or academic positions and meeting expenses were met by the University of Nottingham.

Competing interests: JR chaired and PB and RM were members of the NICE guideline development group on cardiovascular risk assessment. JHC is codirector of QRESEARCH—a not for profit organisation that is a joint partnership between the University of Nottingham and EMIS. EMIS is the leading commercial supplier of IT systems for 56% of general practices in England and Wales and it is likely to implement QRISK2 into its clinical management system. EMIS is likely to also distribute the software package for those using it for academic research or other organisations interesting in implementing QRISK2 into practice or (www.qresearch.org/Public/qriskInformationforClinicians.aspx). RM is a 2008 Harkness Fellow in healthcare policy and practice and is the chair of the cardiovascular working group of the South Asian Health Foundation (SAHF), which receives unrestricted funding from the Department of Health and BHF and unrestricted grants from the pharmaceutical industry. AS chairs the equality and diversity forum of the National Clinical Assessment Service. AS is PI on NHS Connecting for Health's evaluation of the implementation of the NHS Care Record Service. QRESEARCH undertakes analyses for the Department of Health and other government organisations.

Ethical approval: Trent multicentre research ethics committee.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 British Heart Foundation. *Coronary heart disease statistics*. London: British Heart Foundation, 2007.
2 North of England Hypertension Guideline Development Group. *Essential hypertension: managing adult patients in primary care (NICE guideline)*. Newcastle, Centre for Health Services Research: University of Newcastle upon Tyne, 2004.
3 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*. London: NICE, 2006.
4 Department of Health. *Putting prevention first: vascular checks: risk assessment and management*. London: DH, 2008:15.
5 National Collaborating Centre for Primary Care. *Section 4.3 of the guideline on cardiovascular risk assessment: the modification of blood lipids for the primary and secondary prevention of cardiovascular disease*. London: NICE, 2008:43.
6 Ramsay S, Morris R, Lennon L, Wannamethee S, Whincup P. Are social inequalities in mortality in Britain narrowing? Time trends from 1978 to 2005 in a population-based study of older men. *J Epidemiol Community Health* 2008;62:75-80.
7 Ward P, Noyce P, AS SL. How equitable are GP practice prescribing rates for statins? An ecological study in four primary care trusts in North West England? *Int J Equity Health* 2007;6:2.
8 Department of Health. *Tackling health inequalities: a programme for action*. London: Department of Health, 2003:83.
9 Zaman J, Brunner E. Social inequalities and cardiovascular disease in South Asians. *Heart* 2008;94:406-7.
10 Yusuf S, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:937-52.
11 O'Flaherty M, Ford E, Allender S, Scarborough P, Capewell S. Coronary heart disease trends in England and Wales from 1984 to 2004: concealed levelling of mortality rates among young adults. *Heart* 2008;94:178-81.
12 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 doi:10.1136/bmj.39261.471806.55.
13 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.
14 Gill PS, Kai J, Bhopal R, Wild S. Health care needs assessment: black and minority ethnic groups (http://hcna.radcliffe-oxford.com/bemgframe.htm). In: Raftery J, Stevens A, Mant J, eds. *Health care needs assessment: the epidemiological based needs assessment reviews, 3rd series*. Abingdon: Radcliffe Medical Press, 2008 (in press).
15 Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med* 2003;348:1166-70.
16 Chaturvedi N. Ethnic differences in cardiovascular disease. *Heart* 2003;89:681-6.
17 McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in south Asians. *Lancet* 1991;337:382-6.
18 Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;121:293-8.
19 Brindle P, May M, Gill PS, Cappuccio F, D'Agostino Snr R, Fischbacher C, et al. Primary prevention of cardiovascular disease: a web-based risk score for seven British black and minority ethnic groups. *Heart* 2006 doi:10.1136/hrt.2006.092346.
20 Bhopal R, Unwin N, White M, Yallop J, Walker L, Alberti KGMM, et al. Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. *BMJ* 1999;319:215-20.
21 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.
22 Ranganathan M, Bhopal R. Exclusion and inclusion of nonwhite ethnic minority groups in 72 North American and European cardiovascular cohort studies. *PLoS* 2006;3(3):e44.
23 Bhopal R, Fischbacher C, Vartiainen E, Unwin N, White M, Alberti G. Predicted and observed cardiovascular disease in South Asians: application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data. *J Public Health* 2005;27:93-100.
24 Bhopal R. What is the risk of coronary heart disease in South Asians? A review of UK research. *J Public Health* 2000;22:375-85.
25 Patel K, Bhopal R. *The epidemic of coronary heart disease in South Asian populations: causes and consequences*. Birmingham: South Asian Health Foundation, 2004:164.
26 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*. Nottingham: University of Nottingham, 2005.
27 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.
28 Department of Health. *A practical guide to ethnic monitoring in the NHS and social care*. London: DH, 2005:61.
29 Maradit-Kremers H, Crowson CS, Nicola PJ, Ballman KV, Roger VL, Jacobsen SJ, et al. Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: a population-based cohort study. *Arthritis Rheum* 2005;52:402-11.
30 Sarnak MJ, Levey AS, Schoolwerth AC, Coresh J, Culleton B, Hamm LL, et al. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association councils on kidney in cardiovascular disease, high blood pressure research, clinical cardiology, and epidemiology and prevention. *Circulation* 2003;108:2154-69.
31 Wang TJ, Larson MG, Levy D, Vasan RS, Leip EP, Wolf PA, et al. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham heart study. *Circulation* 2003;107:2920-5.
32 Wolf P, D'Agostino R, Belanger A, Kannel W. Probability of stroke: a risk profile from the Framingham study. *Stroke* 1991;22:312-8.
33 Weakliem DL. A critique of the Bayesian information criterion for model selection. *Sociol Methods Res* 1999;27:359-97.
34 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.
35 Gray A, Clarke P, Farmer A, Holman R, United Kingdom Prospective Diabetes Study G. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). *BMJ* 2002;325:860.
36 Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227-41.
37 Schafer J, Graham J. Missing data: our view of the state of the art. *Psychol Methods* 2002;7:147-77.
38 Group TAM. Academic medicine: problems and solutions. *BMJ* 1989;298:573-9.
39 Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. *J Clin Epidemiol* 2007;60:979.
40 Moons KGM, Donders RART, Stijnen T, Harrell FJ. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006;59:1092.
41 Clark T, Altman D. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003;56:28-37.
42 Royston P. Multiple imputation of missing values: update of ICE. *Stata J* 2005;5:527-36.
43 National Collaborating Centre for Primary Care. *Cardiovascular risk assessment: the modification of blood lipids for the primary and secondary prevention of cardiovascular disease: full guideline, consultation draft*. London: National Collaborating Centre for Primary Care, 2007:216.
44 Gail M, Pfeiffer R. On evaluating models of absolute risk. *Biostatistics* 2005;6:227-39.
45 Erika Graf CSWSMS. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med* 1999;18:2529-45.
46 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;23:723-48.
47 Royston P. Explained variation for survival models. *Stata J* 2006;6:1-14.
48 Stevens RJ, Kothari V, Adler AI, Stratton IM. The UKPDS risk engine: a model for the risk of coronary heart disease in type II diabetes (UKPDS 56). *Clin Sci (Lond)* 2001;101:671-9.
49 Gill PS, Bhopal R, Wild S, Kai J. Limitations and potential of country of birth as proxy for ethnic group. *BMJ* 2005 doi:10.1136/bmj.330.7484.196-a.
50 Large P, Ghosh K. *Population estimates by ethnic group—methodology paper*. 2006. www.statistics.gov.uk/downloads/theme_population/MethodologyforPEEG.pdf.
51 D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham heart study. *Circulation* 2008;117:743-53.
52 Nazroo J. South Asian people and heart disease: an assessment of the importance of socioeconomic position. *Ethn Dis* 2001;11:401-11.
53 Whitty C, Brunner E, Shipley M, Hemingway H, Marmot M. Differences in biological risk factors for cardiovascular disease between three ethnic groups in the Whitehall II study. *Atherosclerosis* 1999;142:279-86.
54 Gaines K, Burke G. Differences in stroke: black-white differences in the United States population. *Neuroepidemiology* 1995;14:209-23.
55 Stewart J, Dundas R, Howard RS, Rudd AG, Wolfe CD. Ethnic differences in incidence of stroke: prospective study with stroke register. *BMJ* 1999;318:967-71.
56 Bhopal R, Rahemtulla T, Sheikh A. Persistent high stroke mortality in Bangladeshi populations. *BMJ* 2005;331:1096-7.
57 Car J, Black A, Anandan C, Cresswell K, Pagliari C, McKinstry B, et al. *The impact of eHealth on the quality and safety of healthcare*. Birmingham: NHS Connecting for Health Evaluation Programme. 2008:672. www.pcpoh.bham.ac.uk/publichealth/cfhep/NHS_CFHEP001_eHealth_report_Full_version.pdf.
58 Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham heart study. *Circulation* 1998;98:946-52.
59 Maradit-Kremers H, Crowson C, Nicola P, Ballman K, Roger V, Jacobsend S, et al. Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: a population-based cohort study. *Arthritis Rheum* 2005;52:402-11.
60 Molokhia M, Oakeshott P. A pilot study of cardiovascular risk assessment in Afro-Caribbean patients attending an inner city general practice. *Fam Pract* 2000;17:60-2.

Accepted: 28 May 2008