

# Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database

Julia Hippisley-Cox, professor of clinical epidemiology and general practice,<sup>1</sup> Carol Coupland, associate professor in medical statistics,<sup>1</sup> John Robson, senior lecturer general practice,<sup>2</sup> Peter Brindle, research and evaluation programme director<sup>3</sup>

<sup>1</sup>Division of Primary Care, University Park, Nottingham NG2 7RD, UK

<sup>2</sup>Centre for Health Sciences, Queen Mary's School of Medicine and Dentistry, London E1 2AT, UK

<sup>3</sup>Avon Primary Care Research Collaborative, NHS Bristol, South Plaza, Bristol BS1 3NX, UK

Correspondence to: Julia Hippisley-Cox [juliahippisleycx@gmail.com](mailto:juliahippisleycx@gmail.com)

Cite this as: *BMJ* 2010;341:c6624  
doi:10.1136/bmj.c6624

## ABSTRACT

**Objective** To develop, validate, and evaluate a new QRISK model to estimate lifetime risk of cardiovascular disease.

**Design** Prospective cohort study with routinely collected data from general practice. Cox proportional hazards models in the derivation cohort to derive risk equations accounting for competing risks. Measures of calibration and discrimination in the validation cohort.

**Setting** 563 general practices in England and Wales contributing to the QResearch database.

**Subjects** Patients aged 30–84 years who were free of cardiovascular disease and not taking statins between 1 January 1994 and 30 April 2010: 2 343 759 in the derivation dataset, and 1 267 159 in the validation dataset.

**Main outcomes measures** Individualised estimate of lifetime risk of cardiovascular disease accounting for smoking status, ethnic group, systolic blood pressure, ratio of total cholesterol:high density lipoprotein cholesterol, body mass index, family history of coronary heart disease in first degree relative aged <60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, and atrial fibrillation. Age-sex centile values for lifetime cardiovascular risk compared with 10 year risk estimated using QRISK2 (2010).

**Results** Across all the 1 267 159 patients in the validation dataset, the 50th, 75th, 90th, and 95th centile values for lifetime risk were 31%, 39%, 50%, and 57% respectively. Of the 10% of patients in the validation cohort classified at highest risk with either the lifetime risk model or the 10 year risk model, only 18 385 (14.5%) were at high risk on both measures. Patients identified as high risk with the lifetime risk approach were more likely to be younger, male, from ethnic minority groups, and have a positive family history of premature coronary heart disease than those identified with the 10 year QRISK2 score. The lifetime risk calculator is available at [www.qrisk.org/lifetime/](http://www.qrisk.org/lifetime/).

**Conclusions** Compared with using a 10 year QRISK2 score, a lifetime risk score will tend to identify patients for intervention at a younger age. Although lifestyle

interventions at an earlier age could be advantageous, there would be small gains under the age of 65, and medical interventions carry risks as soon as they are initiated. Research is needed to examine closely the cost effectiveness and acceptability of such an approach.

## INTRODUCTION

Cardiovascular disease is the leading cause of premature death and a major cause of disability in the UK.<sup>1</sup> National policies now support targeting of interventions to reduce risk of cardiovascular disease among high risk patients.<sup>2–5</sup> Validated risk prediction algorithms, such as QRISK2,<sup>6–8</sup> are used in such programmes to identify patients for intervention when they are at high risk—defined by a 10 year cardiovascular disease threshold of  $\geq 20\%$ .<sup>3</sup>

Applying this 20% risk threshold for intervention may not identify younger patients who, because of their age, have a low absolute 10 year risk but who have a high relative risk compared with their peers. This is because age has such a dominant effect in calculating absolute cardiovascular risk. Some argue that younger patients with an adverse risk profile may have more to gain during their lifetime if interventions are started at a younger age rather than waiting until they cross the 20% threshold.<sup>9–12</sup> Lifetime risks which measure the cumulative risk of developing a disease during the remainder of an individual's life<sup>13</sup> would reflect this relatively high risk and, given that lifetime risk estimates provide assessment over the full life course, they may provide a more appropriate assessment of future risks than estimates limited to 10 years, particularly at younger ages.<sup>9–11</sup>

There are currently no published algorithms that estimate lifetime risk of cardiovascular disease derived from contemporaneous UK data, and none which incorporates social deprivation or ethnicity.<sup>12</sup> We therefore developed, validated, and evaluated a new QRISK model to estimate individualised lifetime risk of cardiovascular disease using routinely collected data from UK general practice.

## METHODS

### Study design and data source

We conducted a prospective cohort study in a large population of primary care patients from an open cohort study using the QResearch database (version 29). We included all participating practices in England and Wales who had been using their EMIS (Egton Medical Information System) computer system for at least a year. We randomly allocated two thirds of practices to a derivation dataset and retained a third for a validation dataset. We identified an open cohort of patients aged 30–84 years drawn from patients registered with practices between 1 January 1994 and 30 April 2010. We excluded patients who did not have a postcode related Townsend deprivation score (5.2% of patients), those who had been prescribed statins before the study start date (3.0% of patients), and those with pre-existing cardiovascular disease (3.6%). Entry to the cohort was the latest date of the study start date (1 January 1994), the date the patient became 30 years old, or 12 months after the patient registered with the practice.

### Clinical outcomes

We identified incident cases of cardiovascular disease based on the first recorded diagnosis of cardiovascular disease recorded on the general practice computer system or their linked death certificate during the study period.<sup>6–8</sup> The term cardiovascular includes coronary heart disease (angina and myocardial infarction), stroke, or transient ischaemic attacks, but not peripheral vascular disease. We defined other causes of death as deaths in patients without recorded evidence of cardiovascular disease as defined above.

### Predictor and exposure variables

We used the same predictor variables as QRISK2,<sup>6</sup> with the exception of smoking status (which we categorised as a five level variable) and age (which we included as the underlying time function rather than as a predictor variable). The following variables were included in the final models for men and women separately:

- Smoking status (heavy smoker ( $\geq 20$  cigarettes/day), moderate smoker (10–19/day), light smoker ( $< 10$ /day), former smoker, non-smoker)
- Self-assigned ethnicity (white (or not recorded), Indian, Pakistani, Bangladeshi, other Asian, black African, black Caribbean, Chinese, other (including mixed))
- Systolic blood pressure (continuous)
- Ratio of total serum cholesterol to high density lipoprotein (HDL) cholesterol (continuous)
- Body mass index (weight (kg)/(height (m)<sup>2</sup>) (continuous)
- Family history of coronary heart disease in first degree relative aged  $< 60$  years (yes/no)
- Townsend deprivation score<sup>8</sup> (output area level 2001 census data evaluated as a continuous variable)

- Treated hypertension (diagnosis of hypertension and at least one current prescription of at least one antihypertensive agent)
- Rheumatoid arthritis (yes/no)
- Atrial fibrillation (yes/no)
- Type 2 diabetes (yes/no)
- Chronic renal disease (yes/no), based on presence of diagnostic codes as in QRISK2<sup>14</sup> rather than defined by glomerular filtration rates.

### Statistical modelling

We used multiple imputation to replace missing values for systolic blood pressure, total cholesterol:HDL cholesterol ratio, smoking status, and body mass index. We used the ICE procedure in STATA for imputation and included all the predictor variables listed above plus the survival outcome variables for cardiovascular disease and non-cardiovascular death. We carried out conditional imputation for smoking status. This allowed us to impute a smoking quantity for current smokers when the quantity smoked was not recorded. Our final model was based on five multiply imputed datasets using Rubin's rules to combine effect estimates and estimate standard errors to allow for the uncertainty due to missing data.<sup>15 16</sup> We used the same fractional polynomial terms for body mass index as in QRISK2 (2010).

We developed the model to estimate the lifetime risk of cardiovascular disease, with death (non-cardiovascular) accounted for as a competing risk. Lifetime risk conveys the cumulative risk of developing a disease during the remainder of an individual's life.<sup>13</sup> Competing risk analyses are recommended among elderly populations where a significant proportion of patients may not experience the outcome of interest (such as cardiovascular disease) because they have already experienced a "competing event" such as death.<sup>17</sup> Failure to account for death as a competing risk will tend to overestimate cardiovascular risk among the elderly. We used cause-specific hazard models to account for competing risks, which involved fitting two separate Cox models—one for cardiovascular disease and one for deaths from other causes—including the same predictor variables in both models. Patients who didn't die or have cardiovascular disease were censored at the earliest date of deregistration with the practice, last upload of computerised data, or the study end date (30 April 2010).

In previous QRISK models, we used time since cohort entry as the underlying time function in the Cox regression and incorporated age and its interactions as predictor variables in the final model. In this analysis, we used age as the underlying time function in the Cox regression by setting the origin as the patient's date of birth, as done elsewhere,<sup>18</sup> and defining a delayed entry date as the study entry date. This allows estimation of cause-specific hazard rates across the age range from the youngest age at study entry to the latest age at study exit.

We used a published formula<sup>19</sup> to derive the cumulative incidence function for cardiovascular disease,

**Table 1** | Baseline characteristics of the derivation and validation cohorts. Patients are free from cardiovascular disease and not prescribed statins at baseline. Values are numbers (percentages) of patients unless otherwise stated.

	Derivation cohort (n=2 343 759)	Validation cohort (n=1 267 159)
Women	1 189 845 (50.8)	645 012 (50.9)
Mean (SD) age (years)	48.1 (14.3)	48.0 (14.2)
Mean (SD) Townsend score	-0.2 (3.4)	-0.3 (3.5)
Smoking status:		
Non-smoker	1 176 386 (50.2)	631 545 (49.8)
Former smoker	356 697 (15.2)	193 974 (15.3)
Current smoker (amount not recorded)	99 100 (4.2)	59 178 (4.7)
Light smoker (<10 cigarettes/day)	142 369 (6.1)	71 037 (5.6)
Moderate smoker (10-19/day)	175 419 (7.5)	91 679 (7.2)
Heavy smoker (≥20/day)	136 202 (5.8)	74 056 (5.8)
Smoking status not recorded	257 586 (11.0)	145 690 (11.5)
Ethnic group:		
White or not recorded	2 229 834 (95.1)	1 219 987 (96.3)
Indian	22 598 (1.0)	7 577 (0.6)
Pakistani	11 137 (0.5)	3 663 (0.3)
Bangladeshi	6 432 (0.3)	2 632 (0.2)
Other Asian	12 581 (0.5)	5 032 (0.4)
Caribbean	13 454 (0.6)	4 666 (0.4)
Black African	20 801 (0.8)	9 471 (0.8)
Chinese	5 915 (0.3)	3 068 (0.2)
Other	21 007 (0.9)	11 063 (0.8)
Clinical conditions:		
Treated hypertension*	132 585 (5.7)	67 986 (5.4)
Type 2 diabetes	40 504 (1.7)	20 868 (1.7)
Family history of early coronary heart disease†	247 981 (10.6)	143 593 (11.3)
Atrial fibrillation	12 031 (0.5)	6 589 (0.5)
Chronic renal disease	3 594 (0.2)	1 917 (0.2)
Clinical values:		
Systolic blood pressure recorded	2 027 470 (86.5)	1 081 944 (85.4)
Mean (SD) systolic blood pressure (mm Hg)	131.9 (20.5)	131.7 (20.5)
BMI recorded	1 773 567 (75.7)	949 434 (74.9)
Mean (SD) BMI (kg/m <sup>2</sup> )	26.1 (4.5)	26.1 (4.5)
Smoking status and BMI recorded	1 754 250 (74.9)	937 808 (74.0)
Serum total and HDL cholesterol recorded	692 590 (29.6)	354 853 (28.0)
Mean (SD) total cholesterol:HDL cholesterol ratio	4.2 (1.3)	4.2 (1.3)

BMI=body mass index; HDL=high density lipoprotein.

\*A recorded diagnosis of hypertension and treatment that could include angiotensin converting enzyme inhibitors, angiotensin receptor blockers, aldosterone antagonists,  $\beta$  blockers, thiazides, or calcium channel blockers.

†Heart disease in a first degree relative aged <60 years.

accounting for competing events, using estimates obtained from the two Cox models. This formula multiplies the hazard contribution for cardiovascular disease at a given age by the probability of being both alive and free from cardiovascular disease at that age, and then sums these values across the age range of interest. It can be used to calculate the cumulative incidence function for a patient based on his or her age at baseline and up to any age less than the latest age at study exit, and can also incorporate the patient's covariate values to estimate individual cumulative incidence values. We used this method to estimate the lifetime risk for each patient including events up to the age of 95 years. We used the cut point of 95 years

in our definition for lifetime risk as this is consistent with the value used in other studies of lifetime risk.<sup>11</sup> Also few patients live beyond this age. This calculation incorporates estimates of the probability each patient will be alive and free from cardiovascular disease for ages up to 95, accounting for his or her individual risk factors rather than applying population values of life expectancy as in some other studies.<sup>18</sup> We used the same method to evaluate the 10 year risk for each patient.

In order to validate the performance of the lifetime model at 10 years, we applied the algorithms to our validation cohort and calculated measures of discrimination ( $R^2$  statistic for survival data<sup>20</sup> and area under the receiver operating characteristic curve (ROC statistic)). In order to determine the calibration of the lifetime risk model, we compared observed with predicted lifetime risks by 10th of predicted risk, taking account of competing risks in the calculation of observed risks.

We examined the distribution of lifetime risk estimates and calculated age-sex centile values using methods developed by Royston<sup>21</sup> and compared the distribution with the 10 year risk estimates derived from QRISK2 (2010). We also compared characteristics of the top 10% of patients with the highest lifetime risk with those for the top 10% of patients with highest 10 year risk based on QRISK2 (2010).

We used all the available data in the derivation cohort to develop the model and all the available data from the validation cohort to test its performance. We used STATA (version 11) for all analyses.

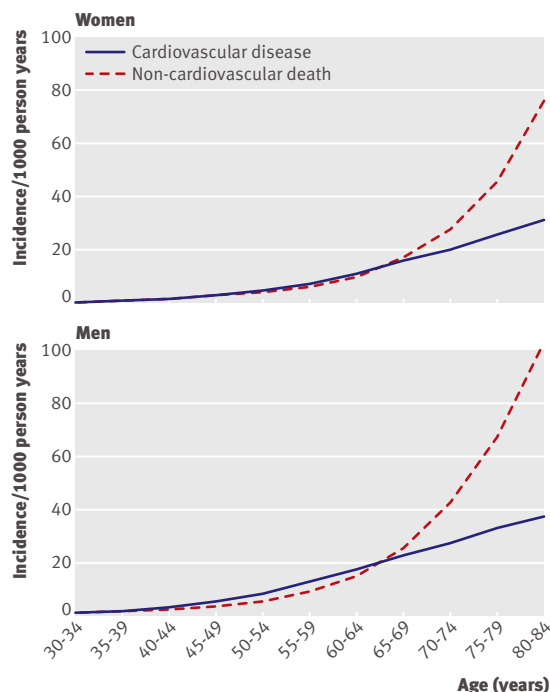
## RESULTS

### Characteristics of the study population

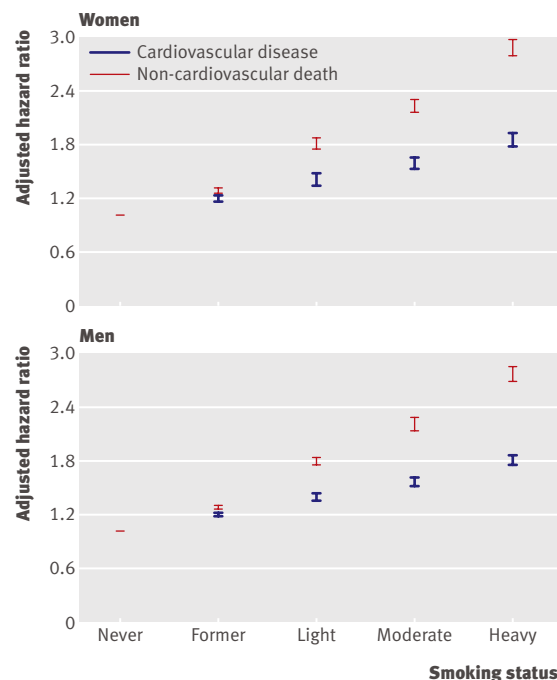
Overall, 563 QRResearch practices in England and Wales met our inclusion criteria: 374 were randomly assigned to the derivation dataset, and the remainder became the validation dataset. We identified 2 509 517 patients aged 30–84 years at study entry in the derivation cohort with complete Townsend scores, of whom 78 619 were taking statins and a further 87 139 had pre-existing cardiovascular disease, leaving 2 343 759 patients for analysis.

We identified 1 353 435 patients aged 30–84 at study entry in the validation cohort with complete Townsend scores, of whom 38 047 were taking statins and a further 48 229 had pre-existing cardiovascular disease, leaving 1 267 159 patients for analysis. Table 1 compares the baseline characteristics of the patients in the derivation and validation cohorts, which were similar, and similar to those reported in previous studies.<sup>6,8</sup>

Overall in the derivation dataset we identified 121 623 incident cases of cardiovascular disease (including cardiovascular events before death and death due to cardiovascular disease) and 148 671 deaths from other causes, arising from a total of 16 485 396 person years of observation. Figure 1 shows the incidence of cardiovascular disease and deaths from other causes by age and sex. The death rates from other causes rise more steeply than



**Fig 1** | Incidence of cardiovascular disease and deaths from other causes per 1000 person years by age and sex in the derivation cohort of 2 343 759 patients



**Fig 2** | Adjusted hazard ratios for cardiovascular disease and death from other causes by smoking status in men and women in the derivation cohort of 2 343 759 patients

cardiovascular event rates in men aged >65 and women >70 years.

#### Adjusted hazard ratios for risk factors in the derivation cohort

Table 2 shows the hazard ratios for men and women for the cardiovascular lifetime risk model. The adjusted hazard ratios for each predictor variable were similar in size to those reported previously.<sup>6</sup>

Smoking showed a dose-response relationship, with the highest risk of cardiovascular disease among heavy smokers. Smoking status was more closely associated with the risk of death from other causes than risk of cardiovascular disease in both men and women, as shown in fig 2.

#### Assessment of performance of the lifetime model in the validation cohort

We determined the performance of the lifetime model evaluated at 10 years. The ROC (receiver operating characteristics) statistic in women was 0.842 (95% confidence interval 0.840 to 0.844) and in men was 0.828 (0.826 to 0.830). The  $R^2$  statistic was 47.0% (46.5 to 47.5) in women and 43.4% (42.9 to 43.9) in men.

Table 3 shows the predicted and observed lifetime risks and the ratio of predicted to observed risk, taking account of death as a competing risk. The results show a small degree of under-prediction in those at low predicted risk—for example, the ratio of predicted to observed lifetime risk in the lowest 10th was 0.90 for men and 0.82 for women. In the highest 10th of risk, the predicted:observed ratio was 1.01 for men and 1.02 for women, suggesting good calibration in this group.

#### Distribution of lifetime and 10 year cardiovascular risk in the validation cohort

Figure 3 shows the 50th centiles for lifetime risk of cardiovascular disease and for QRISK2 10 year risk at each age in men and women. The 50th centile for lifetime risk remains fairly constant, increasing until 60 years in men and 65 years in women, after which it begins to decline. In contrast, the 50th centile for QRISK2 10 year risk increases steeply with age in both men and women, crossing the 20% threshold for risk at age 72 in women and age 65 in men. Across all these patients, the 50th, 75th, 90th, and 95th centile values for lifetime risk were 31%, 39%, 50%, and 57% respectively.

#### Characteristic of patients in top 10% of lifetime risk and 10 year risk

Since there are no accepted thresholds for lifetime risk, we decided to define a high lifetime risk as a value above the 90th centile (that is, a lifetime risk of cardiovascular disease >50%) and for comparative purposes defined a high 10 year risk as a QRISK2 score above the 90th centile (that is, 10 year risk >23.4%). We used these values to identify and compare the 10% of the 1 267 159 patients in the validation dataset with the highest lifetime risk and the 10% with the highest 10 year risk.

Of the 10% classified at high risk with either the lifetime risk model ( $n=126\,716$ ) or the 10 year risk model ( $n=126\,715$ ), only 18 385 (14.5%) were high risk on both measures. Table 4 shows the characteristics of the patients with a high 10 year risk compared with those with a high lifetime risk. Among those with a



**Table 2** | Adjusted hazard ratios\* for cardiovascular disease for individual predictor variables in the derivation cohort of 2 343 759 patients

Variables	Adjusted hazard ratio (95% CI)	
	Women	Men
Body mass index†	1.32 (1.22 to 1.44)	1.54 (1.45 to 1.63)
Systolic blood pressure (per 20 mm Hg increase)	1.13 (1.12 to 1.14)	1.11 (1.10 to 1.12)
Total cholesterol:HDL cholesterol ratio (per unit increase)	1.17 (1.16 to 1.18)	1.18 (1.17 to 1.18)
Townsend score (per 5 unit increase)‡	1.13 (1.11 to 1.14)	1.06 (1.05 to 1.07)
Smoking status:		
Non-smoker	1.00	1.00
Former smoker	1.17 (1.14 to 1.21)	1.18 (1.16 to 1.21)
Light smoker (<10 cigarettes/day)	1.39 (1.33 to 1.45)	1.38 (1.34 to 1.43)
Moderate smoker (10-19/day)	1.57 (1.52 to 1.63)	1.55 (1.51 to 1.60)
Heavy smoker (≥20/day)	1.84 (1.77 to 1.91)	1.79 (1.74 to 1.84)
Ethnic group:		
White or not recorded	1.00	1.00
Indian	1.42 (1.28 to 1.58)	1.50 (1.38 to 1.63)
Pakistani	2.04 (1.78 to 2.34)	2.05 (1.84 to 2.28)
Bangladeshi	1.61 (1.30 to 1.98)	2.14 (1.85 to 2.46)
Other Asian	1.14 (0.92 to 1.40)	1.32 (1.12 to 1.56)
Caribbean	1.03 (0.91 to 1.16)	0.71 (0.63 to 0.81)
Black African	0.69 (0.54 to 0.89)	0.70 (0.56 to 0.86)
Chinese	0.77 (0.55 to 1.08)	0.79 (0.58 to 1.06)
Other	0.99 (0.85 to 1.16)	0.90 (0.78 to 1.04)
Clinical conditions:		
Family history of early coronary heart disease§	1.67 (1.63 to 1.71)	1.84 (1.80 to 1.88)
Type 2 diabetes	1.67 (1.60 to 1.73)	1.60 (1.55 to 1.66)
Treated hypertension	1.33 (1.30 to 1.36)	1.37 (1.34 to 1.40)
Rheumatoid arthritis	1.43 (1.35 to 1.53)	1.37 (1.26 to 1.50)
Atrial fibrillation	1.89 (1.78 to 2.01)	1.63 (1.54 to 1.72)
Chronic renal disease	1.67 (1.44 to 1.95)	1.59 (1.39 to 1.83)

HDL=high density lipoprotein.

\*Hazard ratios were adjusted for all other variables listed in the table.

†Fractional polynomial terms for body mass index: for women, (body mass index/10)<sup>0.5</sup>; for men, ln(body mass index/10).

‡Increasing Townsend scores indicate increasing levels of deprivation.

§Heart disease in a first degree relative aged &lt;60 years.

high lifetime risk, 93 426 (73.7%) were men, compared with 69 794 (55.1%) of those with a high 10 year risk. Overall, the patients with the high lifetime risk were more likely to be younger, male, and much more likely to have a positive family history of coronary heart disease than those with a high 10 year risk. There were also more patients from South Asian ethnic groups and current smokers in the high lifetime risk group.

#### Worked examples of individual lifetime cardiovascular risk

##### *Clinical example 1*

A 54 year old white woman who is a non-smoker, from an affluent area, and with a family history of premature coronary heart disease in a first degree relative, body mass index of 33.2, systolic blood pressure of 150 mm Hg; and a total cholesterol:HDL cholesterol ratio of 5.3, has a 10 year risk of cardiovascular disease of 14%, placing her below the current 10 year intervention threshold of 20%. However, her lifetime risk of cardiovascular disease is 62%, which places her above the 95th centile for lifetime risk (fig 4). A similar

person with better risk factors—that is, body mass index 28, systolic blood pressure 128 mm Hg, total: HDL cholesterol ratio 4—has a lifetime risk of 49%.

The age at which the cardiovascular disease risk reaches 50% is the expected age for a cardiovascular event to occur. In this example, with current risk factors, the expected age is 81 years (that is, in 27 years time), but with better risk factor control this would be delayed to around 95 years (14 years later).

##### *Clinical example 2*

A 31 year old Indian woman who is a non-smoker, from a deprived area, and with a family history of premature coronary heart disease in a first degree relative, body mass index 27.2, systolic blood pressure 128 mm Hg, total cholesterol:HDL cholesterol ratio of 5, has a 10 year risk of 1% and a lifetime risk of 74%. With a 10 year risk of 1%, this woman falls below the 10 year intervention threshold of 20%, but an intervention might be recommended if a lifetime risk threshold of 50% was used instead (fig 5). A similar patient with a total cholesterol:HDL cholesterol ratio of 4 has a lifetime risk of 69%.

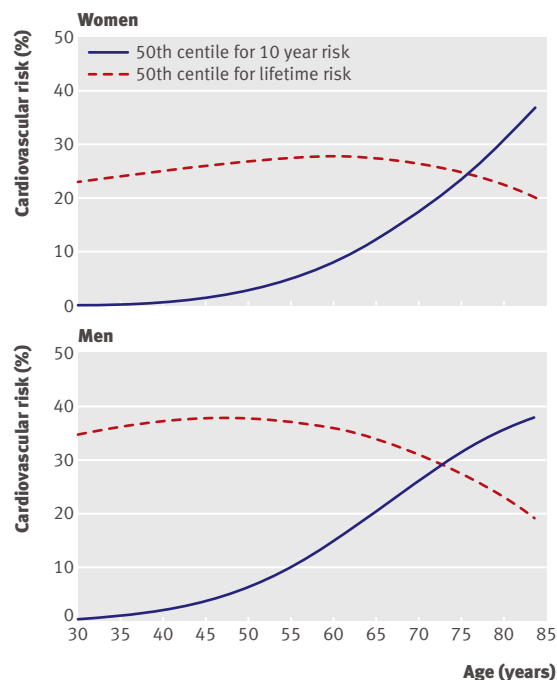
With current risk factors, the expected age for a cardiovascular event to occur is 80 years (in 49 years time), but with better risk factor control this would be around 83 years (three years later).

##### *Clinical example 3*

A 60 year old white man who is a heavy smoker with diabetes, who has a total cholesterol:HDL cholesterol

**Table 3** | Predicted and observed lifetime risk of cardiovascular disease by 10th of predicted lifetime risk in the validation cohort of 1 267 159 patients

Model decile	Mean lifetime risk (%)		Ratio of predicted to observed
	Predicted	Observed	
Women:			
1	18.5	22.4	0.8
2	21.3	25.9	0.8
3	22.9	27.3	0.8
4	24.4	28.5	0.9
5	26.0	29.4	0.9
6	27.8	31.9	0.9
7	30.2	34.8	0.9
8	33.7	36.8	0.9
9	39.5	41.3	1.0
10	51.9	50.8	1.0
Men:			
1	22.5	25.0	0.9
2	27.2	32.1	0.9
3	29.8	34.9	0.9
4	32.0	37.3	0.9
5	34.2	39.3	0.9
6	36.6	42.1	0.9
7	39.5	44.9	0.9
8	43.5	47.5	0.9
9	49.9	51.0	1.0
10	64.4	63.7	1.0



**Fig 3** | Comparison of 50th centile of risk of cardiovascular disease: lifetime risk versus 10 year risk using QRISK2 (2010) by age and sex

ratio of 5, systolic blood pressure of 140 mm Hg, and body mass index of 27.5 would have a 10 year risk of 24% and a lifetime risk of 41%. A similar patient who was a non-smoker would have a 10 year risk score of 16% and a lifetime risk of 44%.

These values might seem counterintuitive in that, while we would expect the 10 year risk estimate to be lower for a non-smoker than a heavy smoker, we would not expect the lifetime risk to increase for the non-smoker. However, a non-smoker has a lower risk of dying from other causes (such as lung cancer) than a heavy smoker, and this increases his or her chances of living longer and thus increases the lifetime risk of developing cardiovascular disease.

## DISCUSSION

### Summary of main findings

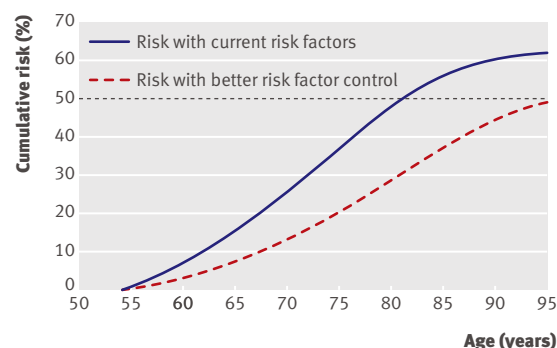
We have developed, validated, and evaluated a new QRISK model to estimate lifetime risk of cardiovascular disease using routinely collected data from electronic health records. Compared with using a 10 year QRISK2 score, a lifetime approach will lead to patients being identified for intervention at a younger age, with a higher proportion of men, more people from non-white ethnic groups, and more with a family history of premature coronary heart disease.

### Comparison with other studies

Although other studies have reported methods to estimate 30 year or lifetime risk,<sup>10 11</sup> ours is the first to estimate lifetime risk of cardiovascular disease based on contemporaneous UK data from primary care. We think ours is also the first study to include a wider

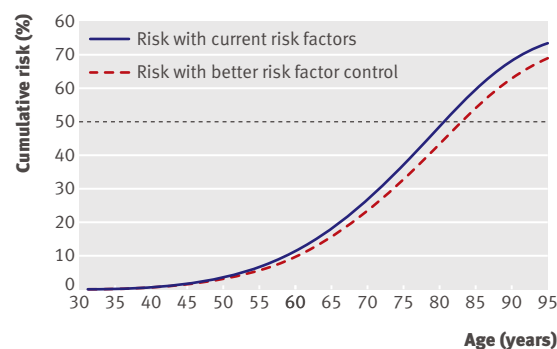
range of risk factors such as ethnicity, social deprivation, rheumatoid arthritis, chronic renal disease, atrial fibrillation, and family history of premature coronary heart disease.<sup>6</sup>

The Framingham investigators found that 10 year risks were better short term predictors of cardiovascular disease incidence than a lifetime equation but underestimated long term risk, particularly in younger people.<sup>22</sup> A recent Framingham study by Lloyd-Jones et al estimated lifetime risks at 50 years of age and reported risks of 52% for men and 39% for women.<sup>11</sup> These are considerably higher than the 50th centile for lifetime risk in our study for patients aged 50 years, which was 38% for men and 29% for women. The higher average risks derived from the Framingham cohort reflect the known tendency of algorithms based on older US data to over-predict risk in contemporary European settings.<sup>23</sup> The reasons for this may include a broader definition of outcomes<sup>11 22</sup> and falling disease incidence since the peak of the epidemic in the 1970s.<sup>11 22</sup>



Horizontal dashed line=90th centile for lifetime risk (values above this are for the top 10% of the population). To obtain this woman's current lifetime risk, read along horizontal axis to age 95 and select value on vertical axis for the blue line (62%). This is the cumulative cardiovascular risk. For the woman's lifetime risk with better risk factor control, select the value of the red line for age of 95 years (49%). To determine the woman's current 10 year risk, add 10 years to the age and read the value on the vertical axis corresponding to age 64 years.

**Fig 4** | A 54 year old woman's individual risk of cardiovascular disease over remaining lifetime—current risk versus risk with better risk factor control (see text for details).



See fig 4 for details of how to interpret graph.

**Fig 5** | A 31 year old woman's individual risk of cardiovascular disease over remaining lifetime—current risk versus risk with better risk factor control (see text for details)

**Table 4** | Characteristics of the 10% of patients in the validation cohort who were at highest risk of cardiovascular disease based on QRISK2 10 year risk score ( $\geq 23.4\%$ ) and on lifetime risk ( $\geq 50\%$ ). Values are numbers (percentages) of patients unless otherwise stated

	10 year risk		Lifetime risk	
	Men (n=69 794)	Women (n=56 921)	Men (n=93 426)	Women (n=33 290)
Mean (SD) age (years)	71.7 (7.6)	76.7 (6.0)	45.2 (10.6)	50.1 (11.7)
Mean (SD) Townsend score	-0.1 (3.5)	0.3 (3.5)	-0.5 (3.4)	0.5 (3.5)
Current smoker	20 300 (29.1)	11 845 (20.8)	30 466 (32.6)	10 639 (32.0)
Age band (years):				
30–44	178 (0.3)	76 (0.1)	49 009 (52.5)	11 593 (34.8)
45–64	11 523 (16.5)	2 376 (4.2)	39 570 (42.4)	17 490 (52.5)
65–74	30 753 (44.1)	13 728 (24.1)	4 469 (4.8)	3 725 (11.2)
75–84	27 340 (39.2)	40 741 (71.6)	378 (0.4)	482 (1.4)
Ethnic group:				
White or not recorded	68 652 (98.4)	56 326 (99.0)	83 411 (89.3)	29 054 (87.3)
Indian	325 (0.5)	149 (0.3)	2 872 (3.1)	1 135 (3.4)
Pakistani	236 (0.3)	153 (0.3)	1 870 (2.0)	1 162 (3.5)
Bangladeshi	209 (0.3)	67 (0.1)	1 504 (1.6)	568 (1.7)
Other Asian	110 (0.2)	38 (0.1)	1 784 (1.9)	342 (1.0)
Caribbean	90 (0.1)	89 (0.2)	171 (0.2)	344 (1.0)
Black African	26 (0.0)	8 (0.0)	170 (0.2)	39 (0.1)
Chinese	16 (0.0)	10 (0.0)	142 (0.2)	30 (0.1)
Other	130 (0.2)	81 (0.1)	1 502 (1.6)	616 (1.9)
Clinical conditions:				
Treated hypertension	14 196 (20.3)	14 833 (26.1)	10 066 (10.8)	7 865 (23.6)
Type 2 diabetes	6 907 (9.9)	5 369 (9.4)	3 425 (3.7)	1 784 (5.4)
Family history of early coronary heart disease*	9 037 (12.9)	6 005 (10.5)	53 600 (57.4)	27 601 (82.9)
Atrial fibrillation	2 712 (3.9)	2 327 (4.1)	792 (0.8)	381 (1.1)
Chronic renal disease	417 (0.6)	270 (0.5)	155 (0.2)	91 (0.3)
Clinical values				
Mean (SD) BMI ( $\text{kg}/\text{m}^2$ )	26.8 (20.8)	26.5 (4.7)	28.5 (4.5)	29.4 (5.4)
Mean (SD) total:HDL cholesterol ratio	4.7 (1.5)	4.4 (1.4)	5.6 (1.6)	5.1 (1.4)
Mean (SD) systolic blood pressure (mm Hg)	150 (21)	154 (22)	138 (20)	144 (22)

BMI=body mass index. HDL=high density lipoprotein

\*Heart disease in a first degree relative aged <60 years.

Other studies have questioned the clinical usefulness of estimating lifetime risk of cardiovascular disease and have developed an alternative approach to communicating short term absolute risk that incorporates the benefits of relative risk and long term risk measures.<sup>24</sup> This approach may be more intuitive for patients and clinicians, but there are concerns that most individuals have high lifetime risks and that this would lead to similar management for everyone.<sup>24</sup> Our study, however, suggests that it is possible to define centile values which could be used as thresholds for classifying additional high risk patients, although our values are clearly applicable only to patients from a UK setting.

### Strengths of study

Our lifetime model is based entirely on clinically available data recorded in NHS electronic healthcare records as part of routine clinical care. The advantages of using the QResearch database include the potential for continued updating.<sup>25 26</sup> Thus, our model is derived from a large, ethnically diverse, contemporaneous population which is representative of those patients with whom the risk estimation method is likely to be

used. The model can also be updated to take account of improvements in data quality (such as increasing numbers of patients with ethnicity or smoking status recorded) or refined over time to reflect trends in population characteristics, changes in clinical requirements such as a need to estimate cardiovascular risk over a broader age range,<sup>27</sup> or to accommodate improved methods for communicating cardiovascular risk to patients.<sup>9</sup>

Our lifetime model estimates risks for non-smokers, former smokers, and light, moderate, and heavy smokers rather than simply for current smokers versus current non-smokers as in the 10 year QRISK2 algorithm.<sup>6</sup> As others have reported, the risks associated with deaths from other causes were also marked.<sup>11</sup> There is a differential effect of smoking status on cardiovascular risk and risk of death from other causes, with a stronger effect on the risk of death from other causes. This probably explains the observation that, when competing risks are being accounted for, there may be cases when the lifetime risks of cardiovascular disease in smokers are lower than those for similar patients who are non-smokers (such as with

clinical example 3 above). This is likely to occur because the non-smokers have a reduced risk of death from other causes and so are likely to live longer, which increases the length of time over which they may develop cardiovascular disease. Our results have good face validity, and the finer grading of smoking can now be incorporated into the QRISK2 10 year risk algorithm.<sup>6</sup>

Our lifetime model is flexible and can be used to calculate risk over different age ranges: for example, it could be used to estimate risk up to the age of 75 rather than 95, since cardiovascular events before age 75 could be considered as premature events.

Lastly, an important advantage of our approach is that the calculation incorporates estimates of the probability that a patient will be alive and free from cardiovascular disease for any age up to 95, accounting for individual risk factors rather than applying population values of life expectancy as in some other studies.<sup>18</sup>

#### Limitations of study

As with all observational studies, our study is subject to bias and confounding, including the effects of missing data. These limitations and the analyses to minimise their effect have been discussed in detail in previous papers.<sup>6-8 28-32</sup> Although we have presented a model for lifetime risk, this has been derived from a contemporaneous UK population with data collected and recorded over a follow-up period of up to 16 years. While this may be a limitation, data collection over a longer follow-up period could also be problematic given temporal changes in incidence of cardiovascular disease and associated risk factors.<sup>11</sup> There have also been temporal changes in use of statins, although these changes have been most marked in the last few years of our study period and previous sensitivity analyses suggest that our models remain robust.<sup>33</sup> Our approach, however, uses age as the underlying time function and has enabled us to derive lifetime risk of cardiovascular disease until the age of 95. The model is based on contemporaneous data and minimises the potential impact of bias due to temporal changes in populations, risk factors, and interventions. Nevertheless, the models do not account for secular trends across the age range.

Although our validation has been undertaken in a separate set of patients and practices from the ones used to develop the models, they all use the same clinical computer system (Egton Medical Information Systems, EMIS). This potentially gives our algorithm some home advantage, though other studies have validated QRISK on practices using a different clinical computer system and have found strikingly similar results.<sup>7 34 35</sup> In addition, EMIS is used in 59% of practices in England and provides the clinical records for over 39 million patients nationally.

#### Lifetime risk for assessment at individual level

The QRISK lifetime model could be used at the individual level to communicate cardiovascular risk to patients, especially younger patients with family

histories of cardiovascular disease, for whom early lifestyle changes or pharmacological intervention could lead to significant gains over their lifetime.<sup>9</sup> A web based calculator ([www.qrisk.org/lifetime](http://www.qrisk.org/lifetime)) could be used to display to both patient and clinician, the current risk trajectory alongside the risk for the same patient with better control of modifiable risk factors such as smoking, body mass index, systolic blood pressure, and total cholesterol:HDL cholesterol ratio. Although our study did not assess the potential benefits of interventions, our first clinical examples suggest marked differences in lifetime risk for patients with good control of risk factors, especially those at younger ages, compared with those with more adverse profiles. This is consistent with other studies, which have shown that individuals with a high 30 year risk may not be identified using a 10 year risk prediction.<sup>10 11 36</sup>

While there is good evidence for lifestyle modification at all ages,<sup>37 38</sup> it is not clear whether individual treatment decisions should be based on lifetime risk instead of 10 year risk (see worked clinical examples above). If interventions were focused on patients above the 90th centile of lifetime risk, then a person at high lifetime risk might receive treatment at an early age (such as the 31 year old woman in clinical example 2) even though cumulative absolute risk would remain well below 20% for the next 30 years (see fig 3). In other words, younger people would tend to have long term treatment with few short term gains, and, as fig 3 shows, most lifetime risk accrues after the age of 65 years. In primary prevention of cardiovascular disease in people at lower levels of absolute risk, there is currently a lack of trial evidence showing benefits from statin treatment in terms of total mortality.<sup>39</sup> The exposure of large numbers of people at low risk to potential harm<sup>28 29 40</sup> over a prolonged period in the absence of such evidence is a major issue. However, there may be more benefits from lifestyle change at a younger age without the potential harms. Similarly, it is far from certain whether the most elderly people (those over the age of 90) with a high lifetime risk would be a patient group that clinicians would wish to treat.

#### Future research

Our study leaves many unanswered questions, including whether early intervention in people with a high lifetime risk but low 10 year risk would have a greater clinical benefit than later intervention; whether people at low absolute risk would value long term treatments with little short term gain; determining the appropriate threshold for lifetime risk to balance the expected benefits (such as event-free years) against the potential adverse effects of interventions such as statins<sup>28 29 40</sup>; and determining the most appropriate and effective methods for communicating absolute and relative risk to patients.

#### Conclusion

We have developed and validated a new model for estimating lifetime risk of cardiovascular disease. This approach identifies different people being at high risk



## WHAT IS ALREADY KNOWN ON THIS TOPIC

Validated risk prediction algorithms such as QRISK2 usually use a 10 year absolute risk of cardiovascular disease of  $\geq 20\%$  to identify patients at high risk

Applying this 20% risk threshold may miss people at younger ages who, despite a low absolute 10 year risk, have a high risk relative to their peers

Presentation of lifetime risk may provide further information, particularly at younger ages, to support management decisions and lifestyle changes

## WHAT THIS STUDY ADDS

A QRISK lifetime approach identified younger patients with a high lifetime risk who were not identified using 10 year risk estimates. Such patients were more likely to be men, from non-white ethnic groups, and have a family history of premature coronary heart disease

While lifestyle interventions at an earlier age may be advantageous, medical interventions carry risks as soon as they are initiated, so the net potential benefit, acceptability, and cost effectiveness of such an approach is unclear

compared with estimating 10 year risk. Compared with the 10 year QRISK2 score, the lifetime approach identified patients for intervention at a younger age, with a higher proportion of men, more people from ethnic minority groups, and more with a family history of premature coronary heart disease. Although lifestyle interventions at an earlier age could be advantageous, medical interventions carry risks as soon as they are initiated, so the net potential benefit remains uncertain. Additional research is needed to closely examine the cost effectiveness and acceptability of such an approach.

The lifetime risk calculator is available at [www.qrisk.org/lifetime/](http://www.qrisk.org/lifetime/) (free for non-commercial research, educational, and personal use).

We thank the EMIS practices which contribute to the QResearch database, and EMIS for expertise in establishing, developing, and supporting the database.

**Contributors:** JH-C initiated the study; undertook the literature review, data extraction, data manipulation, and primary data analysis; and wrote the first draft of the paper. CC contributed to the design, analysis, interpretation, and drafting of the paper. JR and PB contributed to the development of core ideas, the analysis plan, interpretation of the results, and the drafting of the paper.

**Funding:** There was no external funding.

**Competing interests:** JH-C is professor of clinical epidemiology at the University of Nottingham and codirector of QResearch—a not-for-profit organisation that is a joint partnership between the University of Nottingham and EMIS (leading commercial supplier of information technology for 60% of general practices in the UK). JH-C is also director of ClinRisk, which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of medical statistics at the University of Nottingham and a consultant statistician for ClinRisk. JR and PB have received no financial support for undertaking this work. JR and PB were previously members of the NICE Guideline Development Group for Lipid Modification, of which JR was chair. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations. There are no other relationships or activities that could have influenced the submitted work.

**Ethical approval:** The project was reviewed in accordance with the QResearch agreement with Trent Multi-Centre Research Ethics Committee.

**Data sharing:** The patient level data from QResearch are specifically licensed according to its governance framework. See [www.qresearch.org](http://www.qresearch.org) for further details. The Read codes groups used are available from the authors on request. The lifetime risk algorithm will be published as open source software under the GNU Lesser Public Licence.

- 1 British Heart Foundation. *Coronary heart disease statistics*. British Heart Foundation, 2007.
- 2 Department of Health. *Putting prevention first—vascular checks: risk assessment and management*. DoH, 2008:15.
- 3 NICE clinical guideline 64. *Lipid modification—Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease*. NICE, 2008.
- 4 Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106:3140-1.
- 5 Graham I, Atar D, Borch-Johnsen K, Boysen G, Burell G, Cifkova R, et al. European guidelines on cardiovascular disease prevention in clinical practice: executive summary. *Eur Heart J* 2007;28:2375-414.
- 6 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475-82, doi:10.1136/bmj.39609.449676.25.
- 7 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.
- 8 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335:136, doi:10.1136/bmj.39261.471806.55.
- 9 Elward K, Simpson R, Mendy P. Improving cardiovascular risk reduction for primary prevention—utility of lifetime risk assessment. *Postgrad Med J* 2010;122:192-9.
- 10 Pencina MJ, D'Agostino RB, Sr, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the Framingham Heart Study. *Circulation* 2009;119:3078-84.
- 11 Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PWF, et al. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation* 2006;113:791-8.
- 12 Berger JS, Jordan CO, Lloyd-Jones D, Blumenthal RS. Screening for cardiovascular risk in asymptomatic patients. *J Am Coll Cardiol* 2010;55:1169-77.
- 13 Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *Lancet* 1999;353:89-92.
- 14 Hippisley-Cox J. QRISK Search Definition (v2.8). University of Nottingham, 2010.
- 15 Gray A, Clarke P, Farmer A, Holman R, for the United Kingdom Prospective Diabetes Study (UKPDS) Group. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). *BMJ* 2002;325:860.
- 16 Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227-41.
- 17 Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009;20:555-61, doi:10.1097/EDE.0b013e3181a39056.
- 18 Vasan RS, Beiser A, Seshadri S, Larson MG, Kannel WB, D'Agostino RB, et al. Residual lifetime risk for developing hypertension in middle-aged women and men: the Framingham Heart Study. *JAMA* 2002;287:1003-10.
- 19 Kalbfleisch J, Prentice R. *The statistical analysis of failure time data*. Hoboken, 2002.
- 20 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;23:723-48.
- 21 Royston P, Wright EM. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. *J R Stat Soc: A* 1998;161:79-101.
- 22 Lloyd-Jones DM, Wilson PWF, Larson MG, Beiser A, Leip EP, D'Agostino RB, et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. *Am J Cardiol* 2004;94:20-4.
- 23 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.
- 24 Wells S, Kerr A, Eadie S, Wiltshire C, Jackson R. "Your heart forecast": a new approach for describing and communicating cardiovascular risk? *Heart* 2010;96:708-13.
- 25 Cooney MT, Dudina A, D'Agostino R, Graham IM. Cardiovascular risk estimation systems in primary prevention: do they differ? Do they make a difference? Can we see the future? *Circulation* 2010;122:300-10.
- 26 Dent THS. Predicting the risk of coronary heart disease. II: The role of novel molecular biomarkers and genetics in estimating risk, and the future of risk prediction. *Atherosclerosis* (forthcoming).
- 27 Hippisley-Cox J, Coupland C, Robson J, Brindle P. QRISK2 (2010) annual update information. University of Nottingham, 2010:4.
- 28 Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database. *BMJ* 2010;340:c2197.

- 29 Hippisley-Cox J, Coupland C. Individualising the risks of statins in men and women in England and Wales: population-based cohort study. *Heart* 2010;96:939-47.
- 30 Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ* 2009;338:b880.
- 31 Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *BMJ* 2009;339:b4229.
- 32 Hippisley-Cox J, Coupland C. Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the QKidney(R) scores. *BMC Fam Pract* 2010;11:49.
- 33 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. QRISK—authors response. [electronic response to Hippisley-Cox et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study]. *BMJ* 2007. [www.bmj.com/cgi/eletters/335/7611/136](http://www.bmj.com/cgi/eletters/335/7611/136)
- 34 Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. *BMJ* 2009;339:b2584.
- 35 Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. *BMJ* 2010;340:c2442.
- 36 Daviglus ML, Liu K, Pirzada A, Yan LL, Garside DB, Feinglass J, et al. Favorable cardiovascular risk profile in middle age and health-related quality of life in older age. *Arch Intern Med* 2003;163:2460-8.
- 37 Ebrahim S, Beswick A, Burke M, Davey Smith G. Multiple risk factor interventions for primary prevention of coronary heart disease. *Cochrane Database Syst Rev* 2006;(4):CD001561.
- 38 Gotto AM Jr. The Multiple Risk Factor Intervention Trial (MRFIT): a return to a landmark trial. *JAMA* 1997;277:595-7.
- 39 Ray KK, Seshasai SRK, Erqou S, Sever P, Jukema JW, Ford I, et al. Statins and all-cause mortality in high-risk primary prevention: a meta-analysis of 11 randomized controlled trials involving 65 229 participants. *Arch Intern Med* 2010;170:1024-31.
- 40 Naveed S, David P, Heather MM, Paul W, Brendan MB, Anton JMdC, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* 2010;375:735-42.

Accepted: 15 November 2010