Derivation and validation of a new Cardiovascular Disease Risk Score for the UK.

Final Protocol (version 2.0)

10 May 2007

Author: Julia Hippisley-Cox, 14 May 2007

Table of contents

1 Purpose of document 1-3

2 Introduction 2-4

3 Aims 3-5

4 Description of the data source 4-5

5 Methods 5-6

5.1 Study design 5-6

5.2 Practice and patient study population 5-6
5.2.1 Practice selection 5-6
5.2.2 Cohort selection 5-6
5.2.3 Exclusion criteria 5-7

5.3 Cardiovascular disease outcomes 5-7

5.4 Measurement of outcomes 5-7

5.5 Cardiovascular disease risk factors 5-8

5.6 Model derivation and development 5-8
5.6.1 Missing values 5-9

5.7 Derivation of the new CVD risk equation 5-10

5.8 Validation of the new CVD risk equation 5-10
5.8.1 Calibration 5-10
5.8.2 Discrimination 5-11
5.8.3 Comparison with Framingham and ASSIGN 5-11

6 Discussion of methodology 6-12

6.1 Bias and confounding 6-12
6.1.1 Misclassification bias of outcome measures 6-12
6.1.2 Information or recording bias 6-12
6.1.3 Unmeasured confounding 6-12

6.2 Risks and benefits for statins 6-12

6.3 Scientific and ethical approval 6-13

6.4 Funding 6-13

6.5 Acknowledgments 6-13

7 References 7-14

8 Appendix 1: Background to Framingham 8-16

8.1 What is Framingham, what is it based on and how is it used? 8-16

8.2 Framingham cohort description 8-16

8.3 Cardiovascular disease (CVD) outcomes included in Framingham 8-16

9 Appendix 2 – Health Survey for England data, 2003 9-17

10 Revision History 10-18

Author: Julia Hippisley-Cox, 14 May 2007

1 Purpose of document

This is a protocol for the development of a new UK cardiovascular risk equation (QRISK) using the QRESEARCH database and validation of its performance on independent data. It has been written for the Information Centre at the request of the Department of Health and the National Collaborating Centre for Primary Care. cardiovascular risk guideline development group.

A new CVD risk score is needed because the current CVD risk score (which is based on the Framingham cohort) is not well calibrated for UK population and in general tends to over predict risk. Over-prediction of risk may lead to inappropriate treatment for patients and increased costs for the NHS. A new risk algorithm which can more reliably identify people at high risk is now needed to underpin new guidance due from NICE in July 2007. This guidance will recommend that the threshold for treatment with statins is lowered from 40% to 20% risk of CVD over ten years.

Unlike the existing Framingham equations, the new CVD risk equation will also allow the NHS to identify social deprivation and include this as a factor in the estimation of CVD risk. This will be a significant step in reducing health inequalities due to cardiovascular disease; the leading cause of premature death, disease and disability.

The development of an integrated approach to vascular risk assessment is seen as a priority within the Department of Health, not least by the Chief Medical Officer and the National Director for Heart Disease, and this work is central to developing an appropriate risk assessment model for the UK population

Lastly, the Scottish Intercollegiate Guidelines Network (SIGN) has developed a risk score called ASSIGN[1] using a dataset from a Scottish cohort. It is not known how this performs in other UK countries as this cohort was recruited in the 1980's when Scotland had the highest CVD mortality in the world.

The research team undertaking this work includes:

Julia Hippisley-Cox	Peter Brindle
Carol Coupland	Margaret May
Yana Vinogradova	John Robson

Author: Julia Hippisley-Cox, 14 May 2007

2 Introduction

Cardiovascular disease (CVD) is a major cause of death and disability in the developed world[2]. Identification of asymptomatic individuals who are at high risk of developing CVD is the accepted method for the primary prevention of CVD. Asymptomatic patients thought to be at high risk of CVD need to be identified so they can be offered blood pressure and cholesterol lowering treatment together with aspirin and advice about exercise, diet and smoking cessation..

The identification of patients at high risk of CVD is fundamental to many national screening and prevention programmes. Many guidelines recommend that CVD risk is estimated by combining different risk factors into a numeric risk score. There are a variety of risk calculators available as charts, tables, computer programs and web-based tools[3-5] most of which are based on equations derived from the American Framingham Heart Study which is described in detail elsewhere[6].

Whilst the Framingham risk equation has been the 'gold standard' for many years, it has significant limitations. The Framingham cohort is almost entirely white, which limits its application to other more ethnically mixed populations where recalibration is needed[7]. The number of outcome events in Framingham is low which gives its predictions a considerable degree of uncertainty. The Framingham risk equations were developed during the peak of incidence of CVD in America and apply well to similar populations but tend to over-estimate risk by up to 50% in Northern European populations where CVD incidence is lower[8]. They may also under estimate risk in some high risk subgroups such as patients from deprived areas potentially exacerbating health inequalities[9 10]. Lastly, Framingham does not take account of factors likely to affect risk such as deprivation, body mass index, family history and current treatment with antihypertensives. The evidence supporting the use of cardiovascular risk scores for primary prevention in the UK is scarce[11].

New guidance issued by the National Institute of Health and Clinical Excellence (NICE) will lower the threshold for intervention with statins and other drug treatments from a ten year CVD risk of 40% to 20% [12 13]. With this lowering of threshold, any systematic over-estimation of risk, is likely to result in many more people being identified for inappropriate treatment.. Over-identification and over treatment of patients will also impact on prescribing costs and this is likely to have a long term impact since treatment tends to be recommended as life long. For these reasons, it is critical that the underlying methods of risk prediction are fit for purpose so that treatment can be targeted to those who are most likely to benefit. In other words, the equation underpinning the estimate of CVD risk needs to be revised and calibrated for the UK population with an appropriate weighting for social deprivation to avoid exacerbating social inequality.

Author: Julia Hippisley-Cox, 14 May 2007

3 Aims

The overall aim of this research is to derive a new CVD risk score which can be used in clinical practice to better identify high risk patients and target interventions. The new CVD risk score will be derived from a very large representative population of patients contributing data to the UK QRESEARCH database. It will include in the risk prediction traditional risk factors as well as explore additional risk factors such as deprivation, family history, body mass index, ethnicity and the effect of existing antihypertensive treatment. It will be validated on independent sample and predictions compared against Framingham, the current gold standard.

We will also estimate the numbers of patients in the UK likely to need further assessment and treatment in the UK based on a 20% CVD risk threshold measured using the new QR CVD risk compared with the yield based on the existing Framingham equation.

4 Description of the data source

The QRESEARCH database will be used for this study (http://www.qresearch.org). This is a very large validated and representative primary care electronic database containing the health records of 10 million patients over a 17 year period. It has been used for a used for a wide range of research studies including studies of cardiovascular disease incidence, risk and treatment[14-18]. QRESEARCH contains deprivation scores (e.g. Townsend score evaluated at output area) linked to every patient record based on the 2001 census. Most recently, the database has been linked to ONS cause of death data thus allowing sensitivity analyses involving different CVD endpoints. The size and representativeness of the population in QRESEARCH means the results are likely to be generalisable to the UK population and CVD risk estimated with good precision. QRESEARCH is sufficiently large to allow the modelling on two thirds of the database with validation using the remaining one third.

The results of this research should allow the NHS to target the effects of deprivation on vascular risk directly, a significant step in reducing health inequalities due to this group of conditions that are the leading cause of premature death, disease and disability. In addition, this will be the first time that a measure of deprivation has been included in a risk prediction tool to help implement national guidelines.

Author: Julia Hippisley-Cox, 14 May 2007

5 Methods

5.1 Study design

This will be a prospective cohort study conducted within a very large UK primary care population derived from the QRESEARCH database which will determine new risk prediction equations for cardiovascular disease and then test out its performance.

5.2 Practice and patient study population

5.2.1 Practice selection

We will use QRESEARCH version 14 for the analysis. This database is complete until 01 April 2007. Practices will be eligible for inclusion in the analysis a year after the date on which their EMIS system was installed and will stop being included in the analysis on the date of the last recorded upload of data. All eligible practices in the UK contributing to the QRESEARCH database will be included.

We will divide the full QRESEARCH database into two parts –a development dataset and a validation dataset. The development dataset will consist of a random selection of two thirds of the practices contributing to the whole database. The **validation dataset** will consist of the remaining one third of practices. The **development dataset** will be used to derive a new coronary risk equation risk score and it will be validated using the validation dataset.

5.2.2 Cohort selection

We will identify an open cohort of patients aged 35 and over, drawn from patients registered with eligible practices from 01 Jan 1995 onwards until 01 April 2007. For each patient we determine an **entry date** to the cohort analysis which will be the latest of the following dates: 35th birthday, date of registration with the practice, date on which the practice computer system was installed and the beginning of the study period. In addition, patients will only be included in the analysis once they have a minimum of one year of complete medical record data.

For each patient, we will the right censor date which will be the earliest date on which they developed the outcome of interest, the study period ended, date of death, date of de-registration with the practice or date of last upload of computerised data or 100th birthday.

We will then determine the person years at risk which will be the difference between their entry date and their right censor date. Person years at risk will then be used as the denominator term for the incidence rates.

Author: Julia Hippisley-Cox, 14 May 2007

5.2.3 Exclusion criteria

As this is a study of primary prevention, patients with a diagnosis of cardiovascular disease or diabetes prior to their entry date will be excluded from the analysis. We will also exclude patients under the age of 35 years or over 100 years of age, temporary residents and patients with interrupted periods of registration with the practice.

5.3 Cardiovascular disease outcomes

The primary outcome *(outcome 1)* for this main analysis is the first ever diagnosis of cardiovascular disease (CVD) i.e. incident diagnosis of CVD.

CVD will include myocardial infarction, coronary heart disease, stroke and transient ischaemic attacks. Post mortem diagnoses of these outcomes will be included where available (linked cause of death data requested from ONS).

Additional endpoints undertaken will be reported separately including.

- ***Outcome 2***: Coronary heart disease (myocardial infarction, angina) including post-mortem diagnosis.
- ***Outcome 3:*** Stroke or transient ischaemic attack including post mortem diagnosis.
- ***Outcome 4***: All cause mortality.

5.4 Measurement of outcomes

A patient will be considered to have the primary outcome of interest if they have either a computer recorded Read codes for CHD, stroke, Transient ischaemic attack (as defined in the GMS Contract Quality and Outcomes Framework) within their primary care record or relevant ICD 10 codes for post mortem diagnosis within their linked ONS death record.

For each patient, we will determine the date of the first event (either CHD, myocardial infarction, stroke or TIA) prior to death will be the earliest recorded date of the relevant diagnosis based on the computerised Read codes entered within the patient's medical record. A patient will be considered to have coronary heart disease as a cause of death if they have an ICD code I20-25 in the primary cause of death or the final underlying cause of death. A patient will be considered to have stroke or TIA if they have an ICD10 code of I63 or I64 as the primary cause of death or the final underlying cause of death in their linked deaths data.

5.5 Cardiovascular disease risk factors

Using the derivation dataset, we will describe the following variables for our study population and include them as potential risk factors in the analysis.

- Age in single years
- Sex (males vs females)
- Smoking status (current smoker, non smoker, not recorded)
- Systolic blood pressure (continuous)
- Total serum cholesterol (continuous)
- High density lipoprotein (continuous)
- Left ventricular hypertrophy recorded in clinical records (binary yes/no)
- Body mass index (continuous)
- Family history of cardiovascular disease in first degree relative under 60 years (binary variable yes/no).
- Townsend deprivation score (2001 census data evaluated at output areas as a continuous variable)
- % South Asian evaluated at output areas as a continuous variable (2001 census data evaluated at output areas as a continuous variable)

For the above variables we will use values recorded closest to the study entry date in the analysis. We will impute missing values where data are missing as described below.

We will also investigate the effect of treatments with statins and antihypertensive (beta blockers, thiazides, ACE inhibitors or calcium channel blockers following drugs) in subsequent analyses which will be reported separately. The analysis will include prescription for these drugs at the entry date and also incorporate information on their use in the follow-up period until the right censor date. If there is evidence that treatment modifies or attenuates the association between the risk factor and CVD risk, then it might be possible to incorporate a measure of treatment into the final CVD risk score.

5.6 Model derivation and development

We will use the Cox proportional hazards model in the derivation dataset to estimate the unadjusted and adjusted hazard ratios associated with each potential risk factor for the first ever recorded diagnosis of CVD (ie our primary outcome measure) for males and females separately.

A priori, we specified the variables we intend to include in the model based on traditional risk scores (ie Framingham). We will eliminate variables that are missing or poorly recorded in a large number of subjects in general practice (such as left ventricular hypertrophy. We will also exclude variables which aren't statistically significant and compare models using the Bayes Information Criteria (BIC). This is a likelihood measure in which lower values indicate better fit and in which a penalty is

Author: Julia Hippisley-Cox, 14 May 2007

paid for increasing the number of variables. Thus the variables selected for inclusion should not only provide the best fit but also a parsimonious prediction model.

We will examine the strength of the association between one unit increases in each continuous variable (e.g. 1 mmol increase in total serum cholesterol) and compare categories for other variables (e.g. compare current smoking compared with non smoking).

We will test the assumptions of the proportional hazards model for each variable. We will test for any non linear relationship between independent variables and the outcome. We will use fractional polynomials where there are non-linear relationships.

We will undertake selected testing for interactions between variables included in the final model based on interactions which have been cited in published reports (for example interaction between deprivation).

Subsequent analyses will include some variables as time varying covariates e.g. use of antihypertensives and statins and examine for an interaction between systolic blood pressure and treated hypertension. These analyses will be reported separately.

5.6.1 Missing values

For clinical values such as systolic blood pressure, cholesterol, HDL, body mass index, family history and smoking status, the best clinical value would be a value evaluated at the point of entry to the cohort. However it is highly likely that a significant proportion of patients will not have a value measured on the exact entry date itself but measured some time before or after. Therefore we will utilise the values closest to the entry date for patients.

Our first model will be fitted using patients without any missing data (ie complete case analysis). However, since patients with complete data might have a different health status to those with missing data, the principal model which will be fitted will be based on multiply imputed datasets using Rubin's rules to combine effect estimates and correctly estimate standard errors to allow for the uncertainty due to missing data.

We will re-fit the final model using two other methods of imputing values to allow for missing data. First we will impute missing values by age and sex based on data from the QRESEARCH population itself. Secondly, we will use the mean values reported by age and sex from the Health Survey for England data. For consistency and comparisons, we will use the same age bands will be < 35; 35-44; 45-54; 55-64; 65-74; 75 plus. However, since this method does not capitalise on the correlation between the variables at patient level and will artificially narrow standard errors our final models will be based on the multiple imputed datasets.

5.7 Derivation of the new CVD risk equation

The key outcome from this research is a utility which will be able to evaluate an individual's absolute risk of developing one of the four key outcomes over a specified period of time given that they were free from the disease at baseline entry into the cohort. This will take into account the individuals characteristics and existing treatment.

In order to determine this, we will take the log of the hazard ratio for each the risk factors (ie the coefficients from the Cox regression) from the final model and use these as weights for the new CVD risk equation. We can then estimate each individual's probability of experiencing a cardiovascular event within 10 years by combining these weights with the characteristics of the individual and also using the baseline survivor function for all study participants estimated using the Kaplan-Meier method. We will also examine a method similar to the one described by Wilson et al[19] and used for the derivation of the ASSIGN score[1] and other scores[7].

5.8 Validation of the new CVD risk equation

Once the final CVD risk equations have been developed using the derivation dataset, we will prospectively test their performance using the validation dataset. We will use two measures of performance, calibration (ie degree of similarity between observed and predicted events) and discrimination (i.e. ability of a risk prediction equation to distinguish between those who do and do not have a cardiovascular event during the follow up period).

We will also undertake comparisons between the new CVD risk equation and the existing gold standard CVD risk score based on Framingham. We will calculate the 10 year CVD risk for each patient in the validation dataset for QRISK and Framingham imputing missing values as described above.

We will separately report on comparisons between QRISK and the newly developed ASSIGN CVD equation[1].

5.8.1 Calibration

In order to assess calibration, we will compare observed and predicted CVD risk for patients in the validation cohort. We will report the Brier score which computes the sum of squared differences between the observed outcome and fitted probabilities and for which smaller values indicate better concordance between predicted and observed outcomes[20].We will rank patients into deciles according to their predicted risk score and compare observed and predicted risk for males and females separately and for each quintile of Townsend score.

5.8.2 Discrimination

We will assess discrimination of the new CVD risk equation in the validation cohort by using an adaptation of the C statistic (Harrell's C concordance statistic) for censored survival data. This is the equivalent of the Receiver Operator Curves for survival analyses using Cox regression as it allows for the variable length of follow up for an individual patient. The C statistic is defined as the proportion of all usable subject pairs in which the predictions and outcomes are concordant. We will consider a C statistic of greater than or equal to 0.7 as indicating good discrimination, 0.8 would be very good and 0.9 would be excellent.

In addition, we will also present sensitivity, specificity, positive and negative predictive powers of the new CVD risk equation.

5.8.3 Comparison with Framingham and ASSIGN

The existing gold standard CVD risk prediction tool is based on the Framingham risk equation (Anderson 1991). We will therefore compare performance of the new CVD risk equation with Framingham. We define CVD risk as the sum of CHD risk (including MI and CHD death plus angina plus coronary insufficiency) and stroke risk (including TIA) rather than using the CVD risk estimate from Framingham which also includes CCF and peripheral vascular disease. This will mean that we have comparable outcomes.

In addition, we will undertake an additional comparison between the new CVD risk equation and ASSIGN (which will be reported separately).

In order to do this we will apply the Framingham equation and the ASSIGN equation to individual patient level data in the validation dataset. For the ASSIGN score we will need to estimate cigarettes per day where this is not recorded using either the relevant read code (e.g. heavy smoker 40 plus) or values from the assign website by age and sex.

We will then calculate each of the above summary statistics to determine discrimination and calibration.

We will compare the scores for all patients aged 35 plus and also for all patients aged 35 to 74 years as in the original studies.

We will calculate the proportion of patients in the validation sample who have a ten year risk of CVD or 20% or more according to QRISK, ASSIGN and Framingham. We will describe the proportion of patients who are reclassified into a higher or lower risk category.

Analyses will be conducted using Stata (version 9.2).

Author: Julia Hippisley-Cox, 14 May 2007

6 Discussion of methodology

6.1 Bias and confounding

6.1.1 Misclassification bias of outcome measures

This analysis is based on routinely collected data from clinical general practice in the UK. This is in contrast with other CVD risk scores which have been based on tightly controlled and defined cohorts. The disadvantage of the QR cohort is that the outcome measures have not been formally adjudicated i.e. there may be patients with a diagnosis of coronary heart disease who have been misdiagnosed. Similarly, there may be patients with coronary heart disease where the information is not recorded on the clinical records. Such misclassification of outcomes, if present and if non-differential with respect to the risk factors, will tend to result in hazards ratios which tend towards one. However, the inclusion of linked cause of death data is likely to minimise this misclassification.

A key advantage of this analysis is that it is based on a very large representative population which is in the setting where the risk prediction score will later be applied. The results from the analysis should therefore be generalisable to the setting in which it will be used.

6.1.2 Information or recording bias

The main source of information bias is likely to be absent recording of serum cholesterol and smoking status. We will only have data on people that have had contact with their GP surgery. These are a biased population and may not be considered representative of the general healthy population. We will explore the likely impact of this by fitting models with and without imputed values as described above.

6.1.3 Unmeasured confounding

There is likely to be some residual confounding due to unmeasured variables as with all observational studies of this kind.

6.2 Risks and benefits for statins

A separate series of analyses, outside the scope of the current protocol, will address the issues on risks and benefits of statins, ACE and aspirin per se as different samples are likely to be needed. We will compare outcomes for patients on statins, aspirin and ACE compared with non users of each. For statins, we will examine risk of cancer, Parkinson's disease, neuropathy, myopathy. For aspirin, we need to examine risk of renal impairment, gastrointestinal ulcer and haemorrhage. For ACE we need to look at renal impairment.

Author: Julia Hippisley-Cox, 14 May 2007

6.3 Scientific and ethical approval

The proposal has been approved by the Trent Multi Centre Research Ethics Committee.

6.4 Funding

This project has been commissioned by the Information Centre on behalf of the Department of Health.

6.5 Acknowledgments

We acknowledge the contribution of EMIS and EMIS practices contributing to the QRESEARCH database. We acknowledge the following who have facilitated the linkage of cause of deaths data with general practice data to enable this project to take place: Nirupa Dattani and Sir Peter Goldblatt (Office of National Statistics), Steve Daniels (Connecting for Health), John Fox and Dave Roberts (Information Centre), Andy Whitwam and David Stables (EMIS), Julia Hippisley-Cox (University of Nottingham), John Robson (NICE), Bill Kirkup, Michael Soljak, Roger Boyle (Department of Health).

7 References

1. 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;1:1-8.
2. Petersen S, Rayner M. Coronary heart disease statistics: 2002 edition. 1-164.: British Heart Foundation, 2002.
3. Department of Health. National Service Framework for Coronary Heart Disease. London: Department of Health, 2000.
4. NCEP Expert Panel. Detection, evaluation and treatment of high blood cholesterol in adults (Adult treatment panel III): Executive summary. 3, 1-30. *Report of the National Cholesterol Education Program (NCEP)*: National Institute of Health/National Heart, Lung and Blood Institute, 2001.
5. Mann J, Crooke M, Fear H, Hay D, Jackson R, Neutze J, et al. Guidelines for detection and management of dyslipidaemia. Scientific Committee of the National Heart Foundation of New Zealand. *New Zealand Medical Journal.* 1993;106(953):133-141.
6. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *American Heart Journal* 1991;121:293-298.
7. Asia Pacific Cohort Studies Collaboration. Cardiovascular risk prediction tools for populations in Asia. *Journal of Epidemiology & Community Health* 2007;61:115-121.
8. Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ* 2003;327:1-6.
9. Brindle P, McConnachie A, Upton M, Hart C, Davey-Smith G, Watt G. The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *Br J Gen Pract* 2005;55(520):838-45.
10. Tunstall-Pedoe H, Woodward M. By neglecting deprivation, cardiovascular risk scoring will exacerbate social gradients in disease. *Heart* 2006;92:307-310.
11. 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;92:1752-1759.
12. North of England Hypertension Guideline Development Group. Essential hypertension: managing adult patients in primary care (NICE guideline). University of Newcastle upon Tyne, Crown., 2004.
13. NICE. Statins for the prevention of cardiovascular events in patients at increased risk of developing cardiovascular disease or those with established cardiovascular disease guidance type: Technology appraisal. In: Excellence NIfHaC, editor, 2006.
14. Hippisley-Cox J, Parker C, Coupland C, Vinogradova Y. Inequalities in the identification and treatment of hyperlipidaemia in coronary heart disease in patients with serious mental health problems: cross-sectional study. *Heart* 2007 (in press).

Author: Julia Hippisley-Cox, 14 May 2007

15. Hippisley-Cox J, Coupland C. Effect of statins on the mortality of patients with ischaemic heart disease: population based cohort study with nested case–control analysis. *Heart* 2006;92:752-758.
16. Hippisley-Cox J. Trends and variations in GMS indicators for coronary heart disease. Report to the Department of Health. Nottingham: University of Nottingham, 2005.
17. Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients on Cox 2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case control analysis. *British Medical Journal* 2005;330:1366-1374.
18. Hippisley-Cox J, Coupland C. Effect of combinations of drugs on all cause mortality in patients with ischaemic heart disease: nested case control analysis. *British Medical Journal* 2005;330:1059-1063.
19. Wilson P, D'Agostino R, Levy D, Belanger A, Silbershatz H, Kannel W. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97(18):1837-47.
20. Gail M, Pfeiffer R. On evaluating models of absolute risk. *Biostatistics* 2005;6:227-239.
21. Hosmer D, Lemeshow S. *Applied Logistic Regressopm*. New York: John Wiley & Sons, Inc., 1989.

Author: Julia Hippisley-Cox, 14 May 2007

8 Appendix 1: Background to Framingham

8.1 What is Framingham, what is it based on and how is it used?

The Framingham risk equation predicts 10 year risk of fatal and non-fatal cardiovascular events using age; smoking status; diagnosis of diabetes; systolic blood pressure; total cholesterol; HDL; left ventricular hypertrophy. The original Framingham cohort was recruited from the town of Framingham in the United States in 1948. It consisted of 5,209 men and women aged 30-62 who have been followed up every two years. A second cohort (the Framingham offspring cohort) was recruited in 1971 and consisted of 5124 children of the original cohort. A third cohort (the Framingham grandchild cohort) has 3,500 people

8.2 Framingham cohort description

The paper by Anderson[6] (American Heart Journal) paper presents the analysis for six different end points over 4-12 years. The analysis is based on a cohort of 5573 men and women aged 30-74 years from the Framingham Heart Study and the Framingham Offspring study. Baseline values measured between 1968 and 1975. Patients were included if free of cancer (except basal cell carcinoma) and pre-existing CVD.

8.3 Cardiovascular disease (CVD) outcomes included in Framingham

There were six different end points over 4-12 years included in the Framingham analysis:

1. Myocardial infarction (including silent and unrecognised MI)
2. Death from CHD (sudden or non-sudden)
3. CHD (including MI and CHD death plus angina plus coronary insufficiency)
4. Stroke (including TIA)
5. CVD (all of the above plus CCF and peripheral vascular disease)
6. CVD death (ie death from CVD)

The significant risk factors included in the Framingham equation are

- Age
- Sex
- Systolic blood pressure
- Total serum cholesterol
- HDL cholesterol
- Current smoker (or quit within the last year)
- Diabetes (ie on hypyglcaemics or raised fasting of 140mg/DL or above)
- Left ventricular hypertrophy on ECG in men only.

9 Appendix 2 – Health Survey for England data, 2003

The following data are from the Health Survey for England (2003).

sex	ageband	Total serum cholesterol	Low HDL cholesterol	Systolic blood pressure	Diastolic blood pressure	Body mass index	% smokers
females	35-44	5.4	1.6	122	71	26.8	27
females	45-54	5.8	1.7	128	74	27.4	25
females	55-64	6.3	1.7	135	75	28.2	20
females	65-74	6.2	1.7	143	74	28	13
females	75-100	6.1	1.6	148	71	26.9	9
Males	35-44	5.8	1.4	131	74	27.8	26
Males	45-54	5.9	1.4	135	77	28.2	25
Males	55-64	5.8	1.4	139	77	28.3	19
Males	65-74	5.5	1.4	141	73	28	10
Males	75-100	5.3	1.4	145	70	26.9	7

Notes: adults aged 16 and over
Source Department of Health (2004) Health Survey for England 2003. The stationary Office: London

10 Revision History

Revision date	Version	Summary of Changes
09 Oct 2006	1-0	Original version by JHC
18 Feb 2007	1-1	Updated following comments and telecom involving Peter Brindle, Carol Coupland, Margaret May, Yana Vinogradova
01 March 2007	1-2	Updated after further more detailed discussion with Carol Coupland
10 March 2007	1-3	Further JHC edits w.r.t. missing values
15 March 2007	1-4	JHC and CC edits w.r.t. fractional polynomials
21st March 2007	1-5	Additional comments from JR. additional information on ICD cause of death. Final version for peer review by NICE and DH
21st March 2007	1-6	Amended ICD codes for CVD outcomes
28th April 2007	1-7	Final version of protocol