RESEARCH REPORT

Prediction of stroke in the general population in Europe (EUROSTROKE): Is there a role for fibrinogen and electrocardiography?

K G M Moons, M L Bots, J T Salonen, P C Elwood, A Freire de Concalves, Y Nikitin, J Sivenius, D Inzitari, V Benetou, J Tuomilehto, P J Koudstaal, D E Grobbee

....

See end of article for authors' affiliations
....

Correspondence to:
Dr M L Bots, Julius Centre for General Practice and Patient Oriented Research, University Medical Centre Utrecht, room D01.335, PO Box 85500, 3508 GA Utrecht, the Netherlands; M.L.Bots@jc.azu.nl

Accepted for publication 19th September 2001
....

Background: To decide whether a person with certain characteristics should be given any kind of intervention to prevent a cardiovascular event, it would be helpful to classify subjects in low, medium and high risk categories. The study evaluated which well known cerebrovascular and cardiovascular correlates, in particular fibrinogen level and ECG characteristics, are able to predict the occurrence of stroke in men of the general population using data from three European cohorts participating in EUROSTROKE.

Methods: EUROSTROKE is a collaborative project among ongoing European population based cohort studies and designed as a prospective nested case-control study. For each stroke case two controls were sampled. Strokes were classified according to MONICA criteria or reviewed by a panel of four neurologists. Complete data were available of 698 men (219 stroke events) from cohorts in Cardiff (84 cases/200 controls), Kuopio (74/148) and Rotterdam (61/131). Multivariable logistic regression modeling was used to evaluate which information from history, physical examination (for example, blood pressure), blood lipids, and fibrinogen and ECG measurements independently contributed to the prediction of stroke. The area under receiver operating characteristic curve (ROC area) was used to estimate the predictive ability of models.

Results: Independent predictors from medical history and physical examination were age, stroke history, medically treated hypertension, smoking, diabetes mellitus and diastolic blood pressure. The ROC area of this model was 0.69. After validating and transforming this model to an easy applicable rule, 40% of all future stroke cases could be predicted. Adding pulse rate, body mass index, blood lipids, fibrinogen level and ECG parameters did not improve the classification of subjects in low, medium and high risk. Results were similar when fibrinogen was dichotomised at the upper tertile or quintile.

Conclusion: In the general male population the future occurrence of stroke may be predicted using easy obtainable information from medical history and physical examination. Measurement of pulse rate, body mass index, blood lipids, fibrinogen level and ECG characteristics do not contribute to the risk stratification of stroke and have no value in the screening for stroke in the general male population.

Coronary heart disease and cerebrovascular disease are still major causes of morbidity, disability and death in most industrialised countries. Over the years several modifiable cardiovascular risk factors have been identified that are associated with an increased risk of coronary heart disease or cerebrovascular disease. This has offered ways for treatment and prevention. For example, intervention studies showed that treatment of mild to moderate hypertension or lowering serum total cholesterol by drug treatment in subjects with raised levels, reduces the risk of coronary heart disease and cerebrovascular disease. However, it has been repeatedly shown that subjects at the highest absolute risk of future cardiovascular events benefit more from these interventions than those at low or intermediate risk.[1–5]

Practising physicians regularly have to decide whether a particular patient with certain cardiovascular characteristics should be put on (drug) treatment. For this decision the potential benefits and hazards of the intervention should be carefully weighted. To provide "tailored care", a number of guidelines for practising physicians have been issued in which the start of treatment in an individual patient is related to the individuals absolute risk of future coronary heart disease and cerebrovascular disease.[6 7] An established approach for estimation of individual risks (risk stratification) is the use of risk functions or prediction rules, in which the probability of occurrence of disease is estimated as a combined function of the levels of various cardiovascular risk factors. Several of such risk function are present to predict coronary heart disease,[8 9] stroke[10–12] and all cause mortality.[13] None of those available function for stroke have focused on ECG based myocardial infarction or fibrinogen.

Increased levels of fibrinogen have consistently been associated with increased risk of coronary heart disease and stroke.[14–16] In addition, fibrinogen has been shown to be related to presence and extent of atherosclerosis.[17 18] Based on these findings, measurement of fibrinogen has been implicated as a tool to help distinguishing subjects at high cardiovascular risk from those with lower risk.[19] Signs on an ECG of a previous myocardial infarction is an easily obtainable measurement for indication of an individual risk of cardiovascular disease. Misclassification of previous myocardial infarction is much less using an ECG based diagnosis compared with questionnaire based diagnosis, and thus prognostic value of the first is in general increased.[20]

....

Abbreviations: MI, myocardial infarction; LVH, left ventricular hypertrophy; MEANS, Means Electrocardiogram Analysis System

This study evaluated which well known correlates of stroke contribute to the prediction of stroke occurrence in men of the general population using data from three European cohorts participating in EUROSTROKE. In particular, we determined the added value of fibrinogen level and electrocardiogram (ECG) characteristics to these established correlates for the classification of subjects in low, medium and high risk categories. We aimed to define an easily applicable scoring or prediction rule that can be used by physicians as a screening tool to early estimate the risk of future stroke in men.

METHODS

EUROSTROKE is a collaborative study among 10 European research centres to investigate (1) the variation in incidence of fatal and non-fatal ischaemic and haemorrhagic stroke among populations in different European countries; (2) whether the observed differences in stroke incidence across countries can be explained by differences in prevalence of established cardiovascular risk factors; (3) the relative importance of smoking, some selected dietary factors (potassium intake, alcohol consumption), haemostatic disturbances (fibrinogen) and comorbidity (rheumatic heart disease, atrial fibrillation) as determinants of the occurrence of ischaemic and haemorrhagic stroke. The EUROSTROKE database is drawn from 10 ongoing European population-based prospective follow up studies (cohorts) and is designed as a case-control study nested within these 10 ongoing cohort studies. Because EUROSTROKE is based on ongoing studies, information on cardiovascular risk factors in each of the participating centres was already collected before the formal start of the EUROSTROKE project. An exhaustive attempt was made to further harmonise the collected information to make comparison across studies possible. At each moment that a case occurred the corresponding cohort was sampled for two control subjects. Controls were matched on day of baseline examination only. At present data, from three cohort studies (Finland, the Netherlands and United Kingdom) 698 male subjects, 219 stroke events and 479 controls, were used in the present analyses. The mean follow up time of the 698 patients was 7.3 years (range 0.9–20.9 years). The rationale and design of EUROSTROKE have been described in detail elsewhere.[21] The EUROSTROKE project formally started on 1 January 1994.

Outcome (stroke) definition: EUROSTROKE case review board

A case reviewing panel consisting of four neurologists classified all stroke events occurring in EUROSTROKE based on the information that was obtainable such as patients' history, results from lumbar puncture, computed tomography or magnetic resonance imaging, findings from necropsy report, and final codes from existing and operating registries such as for example the FINMONICA registry. Stroke was defined as a clinical event of rapid onset consisting of neurological deficit lasting more than 24 hours unless death supervenes, or if it lasts less than 24 hours, an appropriate lesion to explain the deficit is seen in a brain image. The event could not be directly caused by trauma to the brain, tumour, or infection. Based on the information present, the neurologist classified the event into first and recurrent stroke, and into subarachnoid haemorrhage, intracranial haemorrhage, intracerebral infarction, or unspecified stroke. Cerebral infarction was classified according to internationally accepted criteria.[22 23] In addition, the certainty of the diagnosis was assessed in definite, probable, possible and no stroke. The present analysis is restricted to definite and probable strokes. A stroke was considered to have occurred when (1) the event had led to a hospitalisation and the hospital discharge record indicated a diagnosis of a new stroke. The clinical diagnosis was based on signs and symptoms, and neuroimaging investigations during hospital stay (definite stroke); or (2) in case of no hospitalisation, signs and symptoms associated with the event obtained

from the general practitioner's records were highly suggestive of a stroke according to the neurologists (probable stroke) or (3) in case of out hospital death, when the general practitioner reported that the cause of death was a cerebrovascular accident and a cardiac cause was judged by the general practitioner to be highly unlikely (probable stroke).

Finland

The Finnish contribution to EUROSTROKE comes from the Kuopio Ischaemic Heart Disease Risk Factor study, which is a population-based prospective cohort study comprised of an age stratified random sample of 2682 men aged 42, 48, 54 and 60 years. The baseline examination was performed between 1984 and 1989.[24] Fatal and non-fatal stroke cases were collected through the national mortality statistics and the FINMONICA stroke registries. Stroke was defined according to FINMONICA criteria and definitions.[25] Presently, case ascertainment revealed 74 (definite or probable) stroke cases. Altogether 148 control subjects were randomly drawn from the cohort (sampling fraction 1:17.6).

The Netherlands

The Dutch contribution to EUROSTROKE comes from the Rotterdam Study, which is a population-based prospective follow up study among 7983 subjects (3130 men; 4853 women), aged 55 years or over, living in the suburb of Ommoord in Rotterdam, the Netherlands.[26] Baseline data were collected from March 1990 to July 1993. In the Rotterdam Study, information on incident fatal and non-fatal events is obtained from the general practitioners (GPs) working in the study district of Ommoord as has been detailed elsewhere.[27] Fatal and non-fatal stroke cases were collected through a GP based registry. In short, the GPs involved report all possible cases of stroke to the Rotterdam research centre. All suspected cerebrovascular events reported by the GPs were submitted for review to the EUROSTROKE case review board. Presently, 61 events among men were classified as definite or probable strokes and a total of 131 control subjects were randomly drawn from the male cohort (sampling fraction 1:23.4).

United Kingdom

The British contribution to EUROSTROKE comes from the Caerphilly Heart Disease study in Wales, United Kingdom, in which 2512 men, aged 45 to 59 years are participating.[28] Baseline examinations took place from 1979 to 1983. Follow up examinations were performed from 1984 to 1988 (phase II) and from 1989 to 1993 (phase III). Stroke events were registered through national mortality statistics, hospital discharge records, self report and family report. Of the registered events, additional information on signs and symptoms, neuroimaging, necropsy and a copy of the discharge records were collected when available. When complete, stroke cases were submitted for review to the EUROSTROKE case review board. At present, 100 stroke events had been submitted for review and a total of 200 controls subjects were drawn (sampling fraction = 1:12.1). Eighty four of the 100 reviewed cases were classified as definite/probable stroke by the EUROSTROKE review board.

Potential predictors

Although exhaustive attempts were made to harmonise the collected information, measurements could not be further standardised beyond the attempts done in each individual study. In each centre, baseline information on medical history was obtained by questionnaire. Smoking behaviour was categorised into current, former or never, and alcohol consumption into current drinkers and non-current drinkers (former and never). History of hypertension was considered present if the patient used antihypertensive drugs. Presence of diabetes mellitus was generally based on direct questioning "Do you suffer from diabetes mellitus" although in the

Rotterdam study diabetes mellitus was considered present when subjects used blood sugar lowering drugs. Information on personal and familial history of stroke and myocardial infarction (MI), was based on direct questioning (for example, "Did you ever suffer from a stroke?"). Presence of angina pectoris was based on either the cardiovascular Rose questionnaire or direct questioning. Height and weight were measured and body mass index (kg/m^2) calculated.

In general, systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured at baseline twice (at one occasion) in sitting position. In Cardiff, only one blood pressure measurement was performed. A fasting (apart from Rotterdam) blood sample was taken at baseline for determination of serum lipids (total and HDL cholesterol). Fibrinogen level was measured at baseline according to the Von Clauss method in Kuopio, according to the nephelometric method in Cardiff and in Rotterdam fibrinogen levels were derived from the clotting curve of the prothrombin time assay using Thromborel S as reagent on an Automated Coagulation Laboratory. The latter method correlates well with the Von Clauss method. In all centres an ECG was made at baseline of which three characteristics were evaluated: left ventricular hypertrophy (LVH), presence of MI and arrythmia (defined as presence of atrial fibrillation, atrial flutter or sinus arrythmia). In Kuopio and Cardiff LVH was assessed according to the Minnesota classification system (codes 3.1, 3.3, 3.4).[29] In Rotterdam, an automated diagnostic classification system of the Modular Electrocardiogram Analysis System (MEANS) was used in which LVH diagnosis was assessed on the parameters voltage, shape and repolarisation, as detailed elsewhere,[30 31] and MI as described by de Brunye and coworkers.[32]

Data analysis

Firstly, within each country, the association (odds ratio (OR) and 95% confidence intervals (95% CI)) between the occurrence of stroke and each potential predictor was quantified using univariable logistic regression analyses in order to evaluate whether there were major differences between the three countries. Subsequently, if a test for homogeneity showed no major heterogeneity (p value >0.10), the three ORs were pooled using the Mantel-Haenszel method to estimate an overall (pooled) estimate of the association of each predictor with the outcome. Continuous variables were initially analysed without categorisation but various cut off values and transformations (square root, log) were evaluated as well.[33 34]

Then, multivariable logistic regression modelling was applied to assess the independent contribution of variables in the prediction of stroke. Models were constructed in accordance to the chronology in which predictors are to be documented in clinical practice. Hence, we first included all variables from patient history into an overall "history model". Model reduction was performed by excluding variables with p values >0.10. The reduced "history model" was consecutively extended with body mass index, blood pressure parameters, pulse rate, blood lipids, fibrinogen level and ECG characteristics to evaluate their added value in the prediction of stroke. Of each model, the reliability (goodness of fit) was quantified using the Hosmer and Lemeshow test[35] and their ability to discriminate between patients with and without stroke using the area under the receiver operating characteristic curve (ROC area).[36] The ROC area is a suitable parameter to summarise the discriminative or predictive value and can range from 0.5 (no discrimination, like a coin flip) to 1.0 (perfect discrimination). Differences in predictive value between different (extended) models was estimated by comparison of ROC areas taking into account the correlation between the models as they were based on the same subjects.[37 38]

Besides evaluation of the added value of various tests such as fibrinogen level and ECG characteristics, we also aimed to define one final model that predicts stroke best. To adjust for

overfitting—that is, to adjust for too optimistic estimates of the odds ratio (or regression coefficient) of the predictors in the final model—bootstrapping techniques were used as a method of model validation.[34 39] Subsequently, to obtain an easy applicable prediction rule, the adjusted regression coefficients of the validated model were multiplied by a factor 10 and rounded to the nearest integer.

In multivariable analyses 235 subjects had missing values for one or more predictors. The average number of missings per predictor was 4%, ranging from 0.2% to a maximum of 21%. Missing data are seldom at random. It has been shown that deleting subjects with a missing value on one of the predictors included in the multivariable model (so called complete case analyses) commonly leads to biased results and surely to a loss of power.[34 40 41] To decrease bias and to increase statistical efficiency, it is better to impute these missings rather than doing a complete case analyses.[34 40 41] Accordingly, we imputed per country (that is, before pooling) our missing data using the expectation maximisation method available in SPSS (version 9.0) software. Such imputation is based on the correlation between each variable with missing values and all other variables as estimated from the (463) complete subjects. Furthermore, as the different countries showed different distributions of various predictors such as age, blood lipids and ECG characteristics, we repeated the analyses after equalising the variances and symmetrising the distributions of the predictors. Two methods were used: (1) per country, dividing each predictor by its corresponding standard deviation and (2) per country, transforming each predictor to a standard normal distribution score (Z score)—that is, (value–mean)/standard deviation. These analyses also yielded similar results such that only the results from the untransformed predictors is presented.

RESULTS

Table 1 shows the distribution of the potential predictors per country and in the pooled data. Differences across countries were present for age, smoking, history of MI, hypertension and angina pectoris, and ECG characteristics. However, for all predictors the odds ratios were reasonably homogenous between the different countries: all tests for homogeneity had a p value >0.10. Table 2 shows the stroke characteristics per country and in the pooled dataset. Of the 219 stroke events, 187 (85%) were first strokes.

Table 3 shows the results of the univariable analyses of the pooled data. When fibrinogen level was dichotomised at the upper tertile or at the upper quintile, it was also significantly associated with stroke with odds ratios of 2.1 (95% CI: 1.4 to 3.0) and 1.9 (95% CI: 1.3 to 2.7), respectively.

The overall history model (that is, first 10 variables of table 3) had a ROC area of 0.66. The reduced history model (model 1, table 4) had the same ROC area. Further exclusions or other combinations of predictors significantly decreased the ROC area. Body mass index, pulse rate and blood lipids did not have added value to model 1. When adding blood pressure measurements to model 1, DBP was, and SBP was not significant. Hence, we retained only DBP (model 2, table 4). None of the various definitions of hypertension, such as, SBP>160 mm

Table 1 General characteristics of the participating centres and the overall study population

Characteristic	Cardiff, UK (n=284)	Kuopio, FIN (n=222)	Rotterdam, NL (n=192)	All (n=698)
Case/control	84/200	74/148	61/131	219/479
Age (y)	58 (6)	55 (4)	72 (9)	61 (9)
Current alcohol use (%)	90.3	81.0	75.8	86.5
Current smoking (%)	47.2	27.9	30.8	36.6
Body mass index (kg/m^2)	26.6 (3.6)	27.0 (3.4)	25.9 (3.3)	26.5 (3.5)
Body mass index ⩾25 kg/m^2 (%)	66	70	58	65
Family history of stroke (%)	30.6	19.8	24.5	25.5
Family history of MI (%)	35.6	49.5	36.5	40.3
Diabetes mellitus (%)	4.2	6.3	7.9	5.9
History of stroke (%)	3.2	5.4	8.1	5.2
History of MI (%)	3.9	11.7	16.5	9.9
History of hypertension (%)	18.3	22.5	14.6	18.6
Angina pectoris (%)	13.7	32.4	13.6	19.7
Pulse rate (beats/minute)	60 (18)	44 (13)	69 (18)	57 (19)
Systolic pressure (mm Hg)	146 (22)	133 (18)	145 (23)	141 (22)
Diastolic pressure (mm Hg)	86 (13)	89 (10)	77 (13)	85 (13)
Total cholesterol (mmol/l)	5.8 (1.2)	6.1 (1.2)	6.3 (1.2)	6.0 (1.2)
HDL cholesterol (mmol/l)	1.05 (0.33)	1.26 (0.28)	1.20 (0.31)	1.17 (0.32)
Fibrinogen level (g/l)	4.04 (0.95)	3.20 (0.62)	2.84 (0.71)	3.56 (0.95)
MI on ECG (%)	5.3	4.1	17.9	8.2
Arrythmia on ECG (%)	3.2	8.1	12.5	7.3
LVH on ECG (%)	3.5	4.0	6.8	4.2

Values are unadjusted proportions or means with standard deviations in parentheses.
UK, United Kingdom; FIN, Finland; NL, The Netherlands; MI, myocardial infarction; ECG, electrocardiogram; LVH, left ventricular hypertrophy.

Hg, DBP>90 mm Hg or current blood pressure lowering drug use; SBP>140 mm Hg, DBP>95 mm Hg or current blood pressure lowering drug use; SBP-DBP, yielded a higher predictive accuracy than model 2. Addition of the three ECG characteristics to model 2 marginally increased the ROC area from 0.69 to 0.70. Of these three, only the presence of MI was independently associated with stroke (OR=2.5; 95% CI:1.4 to 5.0). Fibrinogen when added to model 2 was marginally associated with stroke (model 3, table 4). Excluding smoking from table 3, the ROC area remained 0.70 and the ORs of the other predictors were essentially unchanged. Also, when fibrinogen level was included as a dichotomous term, the performance of model 3 was not improved. Combined addition of fibrinogen level and the three ECG characteristics to model 2, yielded a ROC area of 0.70. The fit of all models was good: Hosmer and Lemeshow tests were far from significant (p values >0.50).

As fibrinogen level nor ECG variables showed added value, model 2 was defined as the final model that was validated using bootstrapping (table 5). The validated model was then transformed to an easy applicable scoring rule: 0.3 × age + 19 × history of stroke + 8 × diabetes + 5 × current smoking + 3 × history of hypertension + 0.2 × diastolic blood pressure (last column table 5). Such scoring rule can be considered as one overall predictor or screening test (including several individual predictor variables). The score can be considered as its (test) result and was estimated for each subject by assigning the points for each predictor present and adding these points. For instance, a subject of 60 years, who has a history of stroke and diabetes, does not smoke, has no history of hypertension and with a diastolic blood pressure of 80, receives a score of 61 (18 + 19 + 8 + 0 + 0 + 16). In our data, the score ranged from 27 to 72 (mean 39) and the ROC area of the rule was 0.68 (95% CI: 0.63 to 0.73).

To obtain an estimate of the absolute incidence of stroke across categories of the score in the total cohort, all control subjects were given a weight that was the inverse of the sampling fraction from their corresponding cohort (that is, 17.6 for the controls of the Finland cohort, 23.4 for the Dutch

Table 2 Stroke characteristics of the participating centres and the overall study population

Characteristic	Cardiff, UK (n=284)	Kuopio, FIN (n=222)	Rotterdam, NL (n=192)	All (n=698)
Total stroke	84 (38)	74 (34)	61 (28)	219
Fatal	24	14	10	48
Non-fatal	60	60	51	171
Haemorrhagic	11	15	5	31
Ischaemic	57	54	42	153
Unspecified	16	5	14	35
First stroke	74 (40)	66 (35)	47 (25)	187
Fatal	20	11	7	38
Non-fatal	54	55	40	149
Haemorrhagic	10	14	5	29
Ischaemic	52	47	32	131
Unspecified	12	5	10	27

Values are absolute numbers and percentages between parentheses.
UK, United Kingdom; FIN, Finland; NL, The Netherlands.

Table 3 Association of each potential predictor with stroke occurrence in the overall study population (n=698)

Potential predictor	Stroke present (n=219)	Stroke absent (n=479)	Odds ratio (95% CI)
Age (per year)*	62	60	1.1 (1.0 to 1.1)
Current smoking (%)	43.3	33.5	1.7 (1.2 to 2.3)
Current alcohol use (%)	86.5	86.5	1.1 (0.6 to 1.8)
Diabetes mellitus (%)	10.6	3.8	2.8 (1.5 to 5.4)
History of stroke (%)	13.4	1.5	10.0 (4.5 to 25.3)
History of MI (%)	12.9	8.4	1.5 (0.9 to 2.6)
History of hypertension (%)	26.9	14.8	2.1 (1.4 to 3.1)
Angina pectoris (%)	19.7	19.6	0.9 (0.6 to 1.4)
Family history of stroke (%)	24.8	21.1	1.2 (0.7 to 2.0)
Family history of MI (%)	41.0	40.0	1.1 (0.8 to 1.5)
Pulse rate (per beat/minute)*	59	56	1.0 (1.0 to 1.0)
Body mass index (per kg/m^2)*	26.6	26.4	1.0 (0.9 to 1.1)
Systolic BP (per 10 mm Hg)*	146	139	1.2 (1.1 to 1.3)
Diastolic BP (per 10 mm Hg)*	87	84	1.3 (1.1 to 1.4)
Total cholesterol (per mmol/l)*	6.0	6.0	1.0 (0.9 to 1.2)
HDL level (per mmol/l)*	1.15	1.18	0.6 (0.3 to 1.1)
Fibrinogen level (per g/l)*	3.78	3.46	1.5 (1.2 to 1.9)
MI on ECG (%)	14.0	5.8	2.7 (1.5 to 5.1)
Arrythmia on ECG (%)	7.1	7.8	1.0 (0.5 to 1.8)
LVH on ECG (%)	6.4	3.7	1.8 (0.8 to 3.9)

The odds ratios are adjusted for age (except the odds ratio for age).
MI, myocardial infarction; ECG, electrocardiogram; LVH, left ventricular hypertrophy.
*Values are means.

Table 4 Independent predictors of stroke (odds ratio with 95% CI between parentheses)

Predictor	Model 1 Patient history	Model 2 Patient history + blood pressure	Model 3 Patient history + blood pressure + fibrinogen
Age (per year)	1.02 (1.00 to 1.04)	1.03 (1.01 to 1.05)	1.03 (1.01 to 1.05)
History of stroke	7.68 (3.21 to 18.40)	7.67 (3.19 to 18.41)	7.46 (3.10 to 17.92)
Diabetes	2.2 1 (1.11 to 4.43)	2.38 (1.18 to 4.78)	2.38 (1.18 to 4.80)
Current smoking	1.67 (1.18 to 2.36)	1.69 (1.19 to 2.39)	1.57 (1.10 to 2.24)
History of hypertension	1.58 (1.03 to 2.43)	1.43 (0.92 to 2.20)	1.41 (0.91 to 2.18)
Diastolic blood pressure (per mm Hg)	–	1.02 (1.01 to 1.04)	1.02 (1.01 to 1.04)
Fibrinogen level (per g/l)	–	–	1.16 (0.98 to 1.38)
ROC area (95% CI)	0.66 (0.62 to 0.70)	0.69 (0.64 to 0.73)	0.70 (0.65 to 0.75)

ROC area = area under receiver operating characteristic curve.

controls and 12.1 for the British controls).[42][43] Hence, a new dataset was created that included all 219 cases and the weighted control group resembling the entire cohort. Table 6 shows the incidence of subjects across selected score categories. From this table one can directly obtain the predictive value for presence or absence of stroke per score category (reading horizontally). For example, of all 232 subjects with a score >50, 14.2% (n=33) developed a stroke while this was only 1% (n=9) in the 762 subjects with a score <30. Reading the table vertically provides estimates of the sensitivity and specificity at different thresholds. For example, 1316 (640+444+232) subjects (8%+5%+3% = 16% of total) received a score ≥41. Of these, 86 (33+20+33) indeed developed a stroke, correctly predicting 39% of all future stroke patients (that is, the sensitivity or true positive rate) and 1230 did not develop a stroke, 16% of all "non-stroke" subjects predicted as future stroke cases (that is, the 1– specificity or false positive rate); a specificity of 84%. Using a threshold of ≥46, the sensitivity was 24% with only 8% false positives (specificity=92%).

DISCUSSION

With the aim to provide more tailored care in the future, this study evaluated which patient information may predict the presence of future stroke occurrence in men of the general population (aged 40 to 85) using data from three European countries participating in EUROSTROKE. We have shown that easy obtainable patient information can reasonably predict the occurrence of stroke within seven years. Using these parameters as a simple screening instrument facilitates the early prediction of about 40% of all future stroke cases (sensitivity) with only about 15% false positive predictions (subjects incorrectly predicted as stroke cases). Pulse rate, body mass index, blood lipids, nor fibrinogen level and ECG characteristics show added value to the simple, non-invasive characteristics in the risk stratification.

These results are in agreement with a large number of studies reporting that higher age, smoking, diabetes mellitus, previous stroke, hypertension history and increased diastolic blood pressure are associated with an increased risk of stroke.[44–47] The fact, however, that a risk factor is significantly associated with stroke does not automatically imply that this factor is of use for (early) identification of high or low risk subjects. This is very evident from our analyses, which indeed showed that additional to the medical history plus blood pressure, well known correlates such as pulse rate, body mass index, fibrinogen level and ECG characteristics do not (considerably) add to the prediction of future stroke events. Hence, we do not recommend using these latter measurements as screening tools in the general population to detect early stroke events. As fibrinogen level and ECG characteristics did marginally contribute to the prediction of stroke, it would be interesting, for example, to evaluate whether the (marginal) benefit from measuring fibrinogen levels in all aged, diabetic and smoking men with a history of stroke, outweigh the corresponding measurement costs and patient burden. Further research seems necessary in this area.

The major advantage of a pooled analyses or meta-analyses is to quantify associations with much more precision as compared with individual studies. However, to appreciate the results of our pooled analyses some aspects need to be discussed. Firstly, because of the EUROSTROKE design, the classification of stroke was not done uniformly. The same applies to the definition of the ECG characteristics. Nevertheless, the associations of the investigated factors did not show major differences across the different countries such that a pooled analyses was indicated. Moreover, in daily practice similar differences between countries and medical

Table 5 Odds ratio, regression coefficient and score of each predictor from Model 2 (table 4) after validation

Predictor	Odds ratio	Regression coefficient*	Score†
Age (per year)	1.03	0.03	0.3
History of stroke	6.36	1.85	19
Diabetes	2.26	0.80	8
Current smoking	1.60	0.47	5
History of hypertension	1.39	0.33	3
Diastolic blood pressure (per mm Hg)	1.02	0.02	0.2

The constant of the model after bootstrapping was –4.64.
*Regression coeffcient = ln (odds ratio); †the score per predictor is obtained by multiplying the regression coefficient by 10, and then rounded to nearest integer.

Table 6 Estimated distribution of patients with and without stroke in the original cohort (of the three countries combined) according to the risk score of model 2 (table 4 and 5)

Risk score	Total*	Stroke incidence (%)†	Stroke‡	No stroke§
<30	762	1.0	9 (4)	753 (9)
30–35	2712	1.4	38 (18)	2674 (33)
36–40	3519	2.4	86 (39)	3433 (42)
41–45	640	5.2	33 (15)	607 (8)
46–50	444	4.5	20 (9)	424 (5)
>50	232	14.2	33 (15)	199 (3)
Total	8309		219 (100)	8090 (100)

Values represent absolute number of patients and percentages of the column total between parentheses (except for stroke incidence).
*Total number of patients per score category; †actual (observed) stroke incidence per score category; ‡number of subjects with stroke per score category; §number of subjects without a stroke per score category.

practices will be present and risk estimates should be sufficiently robust to overcome this heterogeneity. Secondly, the frequency of arryhthmia on the ECG was dominated by sinus arrythmia and the prevalence of atrial fibrillation was low. We could therefore not adequately study the predictive value of atrial fibrillation. Thirdly, differences in distributions of continuous predictors across centres may also have biased the findings. When the analyses were repeated using distribution independent techniques we found similar results. Fourthly, the inception of the Rotterdam study was about 10 years later (in 1990) than the other two studies. Diagnostic procedures to detect stroke (for example, computed tomography and magentic resonance imaging) had been introduced during the mid-eighties. This introduction might have improved the detection of minor stroke cases and a better subclassification of detected stroke cases (for example, better distinction between haemorraghic and non-haemorraghic strokes). The latter could not affect the present results because our outcome was the presence or absence of any stroke. But, in the first years of the Cardiff and Kuopio study, a number of minor stroke cases could have been missed. As all sampled controls were extensively screened on the absence of any stroke (minor and major), the missed stroke cases could not have affected the specificity but only the sensitivity of our prediction model. We believe, however, that this bias is small as the problem could only occur in the first years of the Kuopio and Cardiff study. Fifthly, because of the case-control design of this study the odds ratios (regression coefficients) of the predictors are correctly estimated whereas a baseline risk of stroke (or the intercept of a prediction model) could not be directly estimated. Therefore, we reported in table 6 a transformed scoring rule that does not use the intercept (as it is equal for each subject anyway) but still enables classification of subjects according to their absolute risk of stroke. Such scoring rule is easier to apply in practice. To estimate the absolute risk for future subjects in the population at hand, one can use the regression coefficients in table 5. If the incidence in the population at hand is different from ours (2.6%), the intercept in table 5 should be adjusted using standard techniques.[48 49] Finally, although we used bootstrapping techniques to validate the scoring rule external validation studies are necessary before its application to practice.

We suggested to use an arbitrary threshold of 41 in the risk score above which patients are classified as future cases of stroke which yielded a sensitivity of about 40% and specificity of 84%. It should be noted that these estimates apply to a prognostic setting and should not be confused with (usually much higher) estimates obtained from a diagnostic setting. Furthermore, with our choice of threshold we assumed that 16% false positive predictions would be acceptable merely because the consequences of this misclassification (that is, closer attention by the physician) do not cause much harm. Moreover, as these false predictions do have higher risk scores and probably worse prognosis than those with low risk scores, the former will benefit from the closer attention anyway. Nevertheless, using table 5 one could choose other thresholds in the risk score leading to other percentages of misclassifications. Proper definition of thresholds requires information about acceptable "costs" related to misclassifications and was beyond the scope of this paper. This also requires further research.

In conclusion, in the general male population the occurrence of stroke may be predicted early using six simple characterisics. Pulse rate, body mass index, blood lipids, fibrinogen level and ECG characteristics do not seem valuable in the screening for future stroke cases. Additional research is necessary to validate the proposed scoring rule in new subjects and to evaluate its cost effectiveness.

ACKNOWLEDGEMENTS

. . . .

Authors' affiliations

K G M Moons, M L Bots, D E Grobbee, Julius Centre for General Practice and Patient Oriented Research, University Medical Centre Utrecht, the Netherlands
M L Bots, D E Grobbee, Epidemiology and Biostatistics, Erasmus University Medical School, Rotterdam, the Netherlands
J T Salonen, Research Institute of Public Health, University of Kuopio, Kuopio, Finland
P C Elwood, MRC Epidemiology Unit, Llandough Hospital, Penarth, South Glamorgan, UK
A Freire de Concalves, Neurology, Hospitais da Universidade de Coimbra, Coimbra, Portugal
Y Nikitin, Russian Academy of Medical Sciences Siberian Branch, Institute of Internal Medicine, Novosibirsk, Russia
J Sivenius, Department of Neurology, University of Kuopio, Kuopio, Finland
D Inzitari, Neurological and Psychiatric Sciences, University of Florence, Florence, Italy
V Benetou, Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece
J Tuomilehto, Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland
P J Koudstaal, Neurology, University Hospital Rotterdam Dijkzigt, Rotterdam, the Netherlands

REFERENCES

1 **Davey Smith G**, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering the initial level of risk. *BMJ* 1993;**306**:1367–73.
2 **Hoes AW**, Grobbee DE, Lubsen J. Does drug treatment improve survival? Reconciling the trials in mild-to-moderate hypertension. *J Hypertens* 1995;**13**:805–11.
3 **West of Scotland Coronary Pevention Group**. West of Scotland Coronary Prevention Study: identification of highrisk groups and comparison with other cardiovascular intervention trials. *Lancet* 1996;**348**:1339–42.
4 **Sever P**, Beevers G, Bulpitt C, et al. Management guidelines in essential hypertension: report of the second working party of the British Hypertension Society. *BMJ* 1993;**306**:983–7.
5 **Haq IU**, Jackson PR, Yeo WW, et al. Sheffield risk and treatment table for cholesterol lowering for primary prevention of coronary heart disease. *Lancet* 1995;**346**:1467–71.
6 **Pyorala K**, De Backer G, Graham I, et al. Prevention of coronary heart disease in clinical practice; recommendation of the Task Force of the European Society of Cardiology, European Atherosclerosis Society and European Society of Hypertension. *Eur Heart J* 1994;**15**:1300–31.
7 **Core Services Committee**. *Guidelines for the management of mildly raised blood pressure in New Zealand*. Wellington: Ministry of Health, 1995.
8 **Anderson KM**, Wilson PWF, Odell PM, et al. An updated coronary risk profile. A statement for health professionals. *Circulation* 1991;**83**:356–62.
9 **Haq IU**, Ramsay LE, Yeo WW, et al. Is the Framingham risk function valid for northern European populations? A comparison of methods for estimating absolute coronary risk in high risk men. *Heart* 1999;**81**:40–6.
10 **Wolf PA**, D'Agostino RG, Belanger AJ, et al. Probability of stroke: a risk profile from the Framingham Study. *Stroke* 1991;**22**:312–18.
11 **Dippel DW**, Koudstaal PJ. We need stronger predictors of major vascular events in patients with a recent transient ischemic attack or nondisabling stroke. Dutch TIA Trial Study Group. *Stroke* 1997;**28**:774–76.
12 **Truelsen T**, Lindenstrom E, Boysen G. Comparison of probability of stroke between the Copenhagen City Heart Study and the Framingham Study. *Stroke* 1994;**25**:802–7.
13 **Hoes AW**, Grobbee DE, Valkenburg HA, et al. Cardiovascular risk and all cause mortality; a 12-year follow-up study in The Netherlands. *Eur J Epidemiol* 1993;**9**:285–92.
14 **Qizilbash N**, Jones L, Warlow C, et al. Fibrinogen and lipid concentrations as risk factors for transient ischaemic attacks and minor ischaemic strokes. *BMJ* 1991;**303**:605–9.
15 **Di Minno G**, Mancini M. Measuring plasma fibrinogen to predict stroke and myocardial infarction. *Arteriosclerosis* 1990;**10**:1–7.
16 **Wilhelmsen L**, Svardsudd K, Korsan-Bengtsen K, et al. Fibrinogen as a risk factor for stroke and myocardial infarction. *N Engl J Med* 1984;**311**:501–5.
17 **Leng GC**, Papacosta O, Whincup P, et al. Femoral atherosclerosis in an older British population: prevalence and risk factors. *Atherosclerosis* 2000;**152**:167–74.

18 **Ridker PM**, Stampfer MJ, Rifai N. Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. *JAMA* 2001;**285**:2481–5.

19 **Elisaf M**. The treatment of coronary heart disease: an update. Part 1: An overview of the risk factors for cardiovascular disease. *Curr Med Res Opin* 2001;**17**:18–26.

20 **de Bruyne MC**, Mosterd A, Hoes AW, *et al*. Prevalence, determinants, and misclassification of myocardial infarction in the elderly. *Epidemiology* 1997;**5**:495–500.

21 **Grobbee DE**, Koudstaal PJ, Bots ML, *et al*. Incidence and risk factors for ischaemic and haemorrhagic stroke in Europa. EUROSTROKE: A Collaborative study among research centres in Europa: Rationale and design. *Neuroepidemiology* 1996;**15**:291–300.

22 **Bamford J**, Sandercock P, Dennis M, *et al*. Classification and natural history of clinically identifiable subtypes of cerbral infarction. *Lancet* 1991;i:1521–6.

23 **European Atrial Fibrillation Trial Study Group**. Secondary prevention in non-rheumatic atrial fibrillation after transcient ischeamic attack or minor stroke. *Lancet* 1993;**342**:1255–62.

24 **Salonen JT**, Salonen R, SeppänenK, *et al*. High density lipoprotein, HDL2 and HDL3 subfractions and the risk of acute myocardial infarction: a prospective population study in Eastern Finnish men. *Circulation* 1992;**84**:129–39.

25 **Tuomilehto J**, Sarti C, Marva EV, *et al*. The FinMonica stroke register. Community-based stroke registration and analysis of stroke incidence in Finland, 1983–1985. *Am J Epidemiol* 1992;**135**:1259–70.

26 **Hoffman A**, Grobbee DE, Jong PTVM de, *et al*. Determinants of disease and disability in the elderly. The Rotterdam Elderly Study. *Eur J Epidemiol* 1991;**7**:403–22.

27 **Bots ML**, Hoes AW, Koudstaal PJ, *et al*. Common carotid intima-media thickness and risk of stroke and myocardial infarction: The Rotterdam Study. *Circulation* 1997;**96**:1432–7.

28 **Caerphilly and Speedwell collaborative heart disease studies**. The Caerphilly and Speedwell Collaborative Group. *J Epidemiol Community Health* 1984;**38**:259–62.

29 **Rose GA**, Blackburn H, Gillum RF, *et al*, editors. *Cardiovascular survey methods*. Geneva: World Health Organisation, 1968:123–43.

30 **Van Bemmel JH**, Kors JA, van Herpen G. Methodology for the modular electriocardiogram analysis system (MEANS). *Methods Inf Med* 1990;**29**:346–53.

31 **Willems JL**, Abreu-Lima C, Arnaud P, *et al*. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991;**325**:1767–73.

32 **de Bruyne MC**, Mosterd A, Hoes AW, *et al*. Prevalence, determinants, and misclassification of myocardial infarction in the elderly. *Epidemiology* 1997;**8**:495–500.

33 **Harrell FE**, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 1988;**80**:1198–202.

34 **Harrell FE**, Lee KL, Mark DB. Multivarable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–87.

35 **Hosmer DW**, Lemeshow S. Applied logistic regression. New York: Wiley, 1989:140–5.

36 **Hanley JA**, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.

37 **Delong ER**, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non parametric approach. *Biometrics* 1988;**44**:817–45.

38 **Hanley JA**, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;**148**:839–43.

39 **Effron B**. Estimating the error of a prediction rule: improvement on cross-validation. *J Am Stat Assoc* 1983;**78**:316–31.

40 **Greenland S**, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analysis. *Am J Epidemiol* 1995;**142**:1255–64.

41 **Little RJA**. Regression with missing X's: a review. *J Am Stat Assoc* 1992;**87**:1227–37.

42 **Mietinen OS**. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol* 1974;**99**:325–33.

43 **Roest M**, Banga JD, Tempelman MJ, *et al*. The factor V Arg506Gln mutation is not associated with cardiovascular mortality in older women. *Am J Epid* 1998;**149**:665–70.

44 **Boyesen G**, Nyboe J, Appleyard M, *et al*. Stroke incidence and risk factors for stroke in Copenhagen, Denmark. *Stroke* 1988;**19**:1345–53.

45 **Donnan GA**, Adena MA, O'Malley HM, *et al*. Smoking as a risk factor for cerebral ische-mia. *Lancet* 1989;ii:643–6.

46 **Ellekjaer EF**, Wyller TB, Svere JM, *et al*. Lifestyle factors and risk of cerebral infarction. *Stroke* 1992;**23**:829–34.

47 **Welin L**, Svardsudd K, Wilhelmsen L, *et al*. Analysis of risk factors for stroke in a cohort of men born in 1913. *N Engl J Med* 1987;**317**:521–6.

48 **Poses RM**, Cebul RD, Collins lM, *et al*. The importance of disease prevalence in transporting clinical prediction rules. *Ann Intern Med* 1986;**105**:586–91.

49 **Wigton RS**, Connor JL, Centor RM. Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. *Arch Intern Med* 1986;**146**:81–3.