Framingham risk prediction equations for incidence of cardiovascular disease using detailed measures for smoking

Haider Mannan, Chris Stevenson, Anna Peeters, Helen Walls, John McNeil

Dept. of Epidemiology and Preventive Medicine, Monash University, Victoria, Australia

Abstract

Current prediction models for risk of cardiovascular disease (CVD) incidence incorporate smoking as a dichotomous yes/no measure. However, the risk of CVD associated with smoking also varies with the intensity and duration of smoking and there is a strong association between time since quitting and the risk of disease onset. This study aims to develop improved risk prediction equations for CVD incidence incorporating intensity and duration of smoking and time since quitting.

The risk of developing a first CVD event was evaluated using a Cox's model for participants in the Framingham offspring cohort who attended the fourth examination (1988-92) between the ages of 30 and 74 years and were free of CVD (n=3751). The full models based on the smoking variables and other risk factors, and reduced models based on the smoking variables and non-laboratory risk factors demonstrated good discrimination, calibration and global fit. The incorporation of both time since quitting among past smokers and pack-years among current smokers resulted in better predictive performance as compared to a dichotomous current/non-smoker measure and a current/quitter/never smoker measure. Compared to never smokers, the risk of CVD incidence increased with pack-years. Risk among those quitting more than five years prior to the baseline exam and within five years prior to the baseline exam were similar and twice as high as that of never smokers. A CVD risk equation incorporating the effects of pack-years and time since quitting provides an improved tool to quantify risk and guide preventive care.

Correspondence: Haider Mannan, Department of Epidemiology and Preventive Medicine, Monash University, Alfred Centre, 99 Commercial Road, Melbourne 3004, Victoria, Australia.
E-mail: haider.mannan@monash.edu

Key words: coronary heart disease, predictive equation, detailed smoking measures, other risk factors, reduced equation.

Acknowledgments: the Framingham Heart Study - Offspring (FHS-O) is conducted and supported by the NHLBI in collaboration with the FHS-O Study Investigators. This manuscript was prepared using a limited access dataset obtained from the NHLBI and does not necessarily reflect the opinions or views of the FHS-O or the NHLBI. This research was supported by an NHMRC health services research grant (n. 465130), an NHMRC/NHF PhD scholarship and a Vichealth Fellowship.

Contributions: HRM designed the whole study, analyzed the data and wrote the draft of the manuscript; CS, AP and HW helped improve the draft while AP also conceptualized the paper; JM first conceptualized the paper as part of an NHMRC grant for which he is the Principal Investigator.

Conflict of interest: the authors report no conflicts of interest.

Received for publication: 6 July 2010.
Accepted for publication: 2 September 2010.

Heart International 2010; 5:e11
doi:10.4081/hi.2010.e11

Introduction

Cigarette smoking is an important risk factor for developing cardiovascular disease (CVD).[1-4] The risk of CVD increases with the number of cigarettes smoked and with duration of smoking and there is a substantial increase in CVD mortality for continuing smokers compared with ex-smokers.[5] Also, there is a lag between smoking behavior among ex-smokers and disease onset, so that past smoking behavior will influence health status in current non-smokers.[6] For ex-smokers, increasing intervals since quitting are associated with a progressively lower CVD mortality rate, risk of myocardial infarction and ischemic stroke until it levels off.[7-9]

Smoking acts synergistically with other major risk factors of CVD such as age, sex, high blood pressure, dyslipidemia and diabetes.[10] This has led to the development and validation of a number of multivariable risk models which can be used by primary care physicians to assess the risk in individual patients of developing all atherosclerotic CVD[11-20] or specific types of CVD, namely, coronary heart disease,[10,20-25] stroke,[26] peripheral vascular disease,[27] or heart failure.[28] Multivariable assessment has been advocated to estimate absolute CVD risk and guide treatment of risk factors,[2,29] and the Framingham CVD risk assessment tool and other similar risk assessment tools have been validated[22,23,25] and also re-calibrated in other ethnically diverse populations.[10,17,20,26-29]

This study seeks to examine the role of cigarette smoking as a risk factor in risk models in greater detail than the simple dichotomous yes/no variable[1-3] that has been incorporated in all the existing risk prediction models. Risk equations will be developed incorporating the effect of pack-years, time since quitting, and combined effects of intensity and duration of smoking, the latter variables being more readily available to general practitioners (GPs) than pack-years. This study will also develop reduced equations based on the detailed smoking measures and risk factors readily available to GPs such as non-laboratory risk factors. In this study, we use a slightly modified Framingham risk equation as the reference equation. Alcohol intake is included as it is a possible confounder of cigarette smoking with CVD risk, and systolic (SBP) and diastolic (DBP) blood pressures are included since a recent Framingham based risk model for CVD risk[20] demonstrated improved predictive performance when both these measures were included in the risk model.

Materials and Methods

The details for design, selection criteria, examination procedures and an operative definition of CVD for the Framingham Offspring Heart Study have been provided

elsewhere.[20,26-29] To provide current estimates of CVD risk over a broad age range, smoking status information from the first examination was carried forward to the fourth examination (1988 to 1992). Participants were eligible for the present study if they were free of CVD, aged 30-74 with complete data on covariates. Participants aged 75 years and older were excluded because of possible differences in risk factors in this older group and its potentially large influence in the algorithm determination. After exclusions, the sample consisted of 3,751 participants (mean age 51.61 years; 1,937 women).

Measurement of developing cardiovascular disease risk factors

The Framingham Offspring study collected data on smoking status, the number of cigarettes smoked per day, age at starting regular smoking, and age at quitting smoking. The risk factors included in the models were smoking status (defined according to the particular study objective), both components of blood pressure (SBP and DBP), either total cholesterol/HDL ratio or both HDL and total cholesterol depending on which provides better prediction, age, sex, diabetes status, body mass index (BMI) and alcohol consumption.

In the Framingham Offspring study, at each examination participants underwent a physical examination, anthropometry, blood pressure determination, and phlebotomy for vascular risk factors. The examination blood pressure was calculated as the average of 2 physician-obtained measures. Standardized enzymatic methods were used to determine serum total and HDL cholesterol levels. The various measures of cigarette smoking were obtained by self-report. Diabetes was defined as a fasting glucose of above 126 mg/dL.

Examination 4 was considered the baseline for our study as this had CVD data for a broad age range. Information on time since quitting, age at quitting and duration of smoking were not available at this examination. First, age at quitting at examination 4 was estimated from information of this variable from examination 1 and smoking history from examinations 1 through to 4. For those who were quitters at both examinations 1 and 4 but resumed smoking in between the two examinations, their age at quitting at examination 1 was carried forward to examination 4 by adding their age at quitting at examination 1 to the time interval they remained smokers after examination 1. Similarly, for those who were smokers at examination 1 but quitters at examination 4, this was estimated by taking the average of the age at quitting and the preceding age. Time since quitting at examination 4 was estimated by subtracting age at quitting at examination 4 from current age at examination 4.

For those who were smokers at both examinations 1 and 4, their duration of smoking at examination 1 was carried forward to examination 4 by adding their duration of smoking at examination 1 to the interval they remained smokers after examination 1. Similarly, for those who were non-smokers at examination 1 but smokers at examination 4, this was estimated as the interval they remained smokers after examination 1.

Categories of smoking

Current pack-years of smoking were calculated by dividing the number of cigarettes being smoked per day by 20 to obtain an estimate of "packs" and multiplying this by the number of years a person is a smoker. To separate the effect of past smoking and current pack-years of smoking and assess a dose-response pattern, the smoking status and current pack-years were combined into a five-category "pack-years" variable defined as: never smokers, former smokers and current smokers with under 20, 20-39, and 40 or more pack-years. Other categorizations for pack-years were found to be less effective in terms of predictive ability. To separate the effect of current smoking and time since quitting among past smokers and assess a dose-response pattern, the smoking status and time since quitting were combined into a four-category variable defined as: never smokers, former smokers with five years or less and over five years since quitting, and current smokers. Compared to the cut-off of five years, other categorizations for time since quitting did not predict the risk of CVD incidence as well. To assess the joint effects of intensity and duration of smoking, current smokers were classified into 9 groups with three levels of intensity (<20, 20-39 and 40+ pack-years) at each of three levels of duration (<30, 30-39 and 40+ years).

Development and assessment of predictive models

Three groups of models were fitted, each incorporating the range of risk factors listed above. The first group focused on the effect of pack-years of smoking (among current smokers). First a model with a simple binary current/non-smoker measure for smoking status was compared to an equivalent model with the pack-years variable defined above. Then, a model with a current/past/never smoker measure for smoking status was compared to the model containing the pack-years variable. A similar approach was adopted for a second group of models focussed on time since quitting, and for a third group including the joint effects of intensity and duration of smoking among current smokers in the smoking status variable. While developing the risk equations, only those risk factors which were significantly related to the risk of CVD incidence were included in the model.

For each of the three groups of models, we also fitted reduced models incorporating the smoking variables and non-laboratory risk factors. The non-laboratory risk factors included age, sex, SBP, DBP, BMI and alcohol consumption. The full and reduced models were compared to examine whether use of only non-laboratory risk factors instead of all the risk factors resulted in a significant loss in prediction or not. In some of the risk equations, the ratio of total cholesterol to HDL cholesterol was used because no improved fit was found when the covariates were used separately.

We used Cox's proportional-hazards regressions[30] to relate risk factors to the incidence of a CVD event during a follow up from exam 4 to exam 7 after confirming that the assumption of proportionality of hazards was met. For testing proportionality of hazards a non-significant interaction between a covariate and log (survival time) would support the assumption for the covariate.[31] This was complemented by the plotting of Schoenfeld residuals against survival time.

The discriminative ability of a model was assessed by Harrell's c statistic.[32,33] Harrell recommends using bootstrap techniques to assess the degree of 'overoptimism' in the c statistic due to using the same data for model fitting and assessment. We applied this technique and found negligible overoptimism with the bias in all the c-statistic estimates less than 0.006. To assess the improvement in discrimination between two nested models, we used a test for difference in two correlated c statistics.[34] But for models containing standard risk factors and possessing reasonably good discrimination, very large "independent" associations of the new covariate with the outcome are required to provide a meaningfully larger c statistic.[35-37] To overcome this limitation of c statistic, the NRI[38] and IDI[38] provide valuable supplements to the c-statistic analyses.[38,39] These attempt to quantify in different ways the level of shift in the distribution of absolute risk after a new covariate is included in the model. For calculating NRI, we assessed risk reclassification[38] by sorting the predicted risk for each model into 3 categories (<6%, 6% to <20%, and ≥20%). The benefit and cost of using a new model compared to the baseline model was measured by the proportions of subjects with and without subsequent events, respectively, who are classified as high-risk (e.g. ≥20%) according to the model.[40] To assess calibration of the fitted model, we computed the Hosmer-Lemeshow statistic and its modification for the reclassification tables.[38] A significant value for this statistic indicates poor calibration. We also

performed likelihood ratio tests to evaluate whether the global model fit improved after the inclusion of additional covariates. Finally, we assessed the overall goodness of fit for each of the models using BIC and AIC. These allow us to rank the models' goodness of fit adjusting for model parsimony.

Results

The risk factor characteristics in our sample at the baseline examination 4 are shown in Table 1. The sample consists of 26.7% never smokers, 48.44% quitters (15.06% quit within five years and 33.38% prior to five years), 14.5% current smokers (2.8%, 5.46% and 16.56% have pack-years <20, 20-39 and ≥40). Also, 24.85%, 5.81% and 13.39% had a duration of <30 years, 30-39 years and ≥40 years while 4.11%, 8.21% and 12.53% had smoking intensity of less than 20 cigarettes per day, 20-39 cigarettes per day and 40 or more cigarettes per day.

Table 2a shows that the inclusion of pack-years among current smokers in our measure of smoking status (model 3) improved predictive ability significantly compared to the models which included smoking status as a simple dichotomous variable (model 1) and current/past/never smoker variable (model 2). The models without control (Table 2a) and with control (Table 2b) for pack-years indicated excellent calibration.

The inclusion of time since quitting in our measure of smoking status (model 4) performed well in terms of model discrimination and overall fit but not so well in terms of calibration (Table 2b). Model 4 improved predictive ability significantly based on global fit and reclassification statistics compared to model 1 and model 2 (Table 2c).

The models incorporating pack-years and time since quitting (model 5), and the joint effect of both intensity and duration of smoking, and time since quitting (model 6), performed well in terms of all model performance indicators (Table 2d). These models showed significantly improved prediction performance in terms of global fit and reclassification measures compared to the models with a simple current/non-smoker measure of smoking status and current/past/never smoker measure of smoking status, respectively (Tables 2e and f). Table 2e further shows that model 5 predicts significantly better than models 3 and 4.

The best risk prediction equation

Table 2e further shows that, as expected, improvement in discrimination through c statistic for model 5 relative to models 1 through 4 is moderate as the c statistic for the latter (reference) models was already fairly high (this is a limitation of the c statistic as discussed in Materials and Methods). Based on AIC and BIC, model 5 predicts better than models 1 through 4 and model 6 when model parsimony is taken into consideration (Table 2g). (Since models 5 and 6 are not nested, they can only be ranked using AIC and BIC). Thus, model 5 is the preferred model.

Table 2b shows that, compared to never smokers, the risk of CVD incidence increased with past smoking status and pack-years among current smokers for which the increase in risk seemed to slow down for those with exposure to 40 or more pack-years. Those who quit within five years prior to the baseline exam had a risk of developing a CVD event almost 2 times higher than those who never smoked, while those who quit prior to five years had a similar risk to never smokers.

Table 1. Summary of statistics for risk factors (at exam 4) used in risk models for total population characteristics.

	Summary statistic
Sex, N (%)	
Females	1937 (51.64)
Males	1814 (48.36)
Age (years), mean (SD)	51.61 (9.63)
Total-C (mg/dL), mean (SD)	206.30 (39.17)
HDL-C (mg/dL), mean (SD)	49.52 (14.80)
Systolic blood pressure (mmHg), mean (SD)	127.11 (18.89)
Diastolic blood pressure (mmHg), mean (SD)	79.25 (10.01)
Triglycerides (meq/liter), mean (SD)	125.92 (101.34)
Alcohol (ounce), mean (SD)	2.89 (4.39)
Total cholesterol (mg/dL), mean (SD)	206.31 (39.17)
HDL cholesterol (mg/dL), mean (SD)	49.52 (14.80)
Total/HDL cholesterol ratio, mean (SD)	4.54 (1.65)
Body mass index (kg/m^2), mean (SD)	26.86 (4.81)
Never smoking, N (%)	1002 (26.71)
Past smoking, N (%)	1817 (48.44)
Time since quitting ≤5 years, N (%)	565 (15.06)
Time since quitting >5 years, N (%)	1252 (33.38)
Current smoking, N (%)	932 (24.85)
Pack-years <20, N (%)	105 (2.80)
Pack-years 20-39, N (%)	206 (5.49)
Pack-years 40+, N (%)	621 (16.56)
Duration of smoking <30 years, N (%)	212 (5.65)
Duration of smoking 30-39 years, N (%)	218 (5.81)
Duration of smoking 40+ years, N (%)	502 (13.39)
Intensity of smoking <20 cigarettes (per day), N(%)	154 (4.11)
Intensity of smoking 20-39 cigarettes (per day), N(%)	308 (8.21)
Intensity of smoking 40+ cigarettes (per day), N (%)	470 (12.53)
Diabetes, N (%)	163 (4.35)

Reclassification of subjects

In this section, we describe how many subjects were reclassified overall and with respect to 'high risk' category of more than 20% when we used our preferred full model (model 5) as compared to model 1. Based on Table 3a, for 383 subjects who experienced a CVD event, using model 5 as compared to model 1, the net gain in reclassification proportion was significantly different from zero (P=0.0001). This statistic for subjects who did not experience an event was not significant. Overall, for all 3,751 subjects, the net gain in reclassification proportion was 0.0711, significantly different from zero.

If we had used model 5 rather than model 1, 19.3% of those who develop CVD would be appropriately assessed for their cardiovascular risk while only 11.4% of those who do not develop CVD would be falsely assessed for their cardiovascular risk, the difference of which is highly significant (P<0.0001).

Simpler developing cardiovascular disease risk prediction equations using non-laboratory risk factors

The simple office or non-laboratory risk factor-based CVD risk prediction equation

incorporating the joint effects of pack-years and time since quitting and non-laboratory risk factors (model 11) performed reasonably well (*data not shown*). Compared to this model, the corresponding full model (model 5) predicts better both in terms of global fit and discrimination (Table 3b). The simple office-based CVD risk prediction model incorporating the effects of intensity and duration of smoking, time since quitting and non-laboratory risk factors (model 12) also performed reasonably well (*data not shown*). The improvement in global fit and discrimination for the corresponding full model was also significant (Table 3c).

The best reduced risk prediction equation

Table 3d shows that, based on the global fit statistic and the measures of reclassification, model 11 predicts the risk of CVD incidence significantly better than models with non-laboratory risk factors and a yes/no measure of current smoking status (model 7), current/past/never smoker measure of smoking status (model 8), pack-years incorporated into the smoking status variable (model 9) and time since quitting incorporated into the smoking status variable (model 10). Again as

Table 2a. Model performance statistics for the model with current/non-smoker measure of smoking (model 1) and improvement in prediction due to including pack-years in model 3.

Statistic	Value	Degrees of freedom		P
Hosmer Lemeshow	4.6310	9		0.8652
Modified HL	4.3962	9		0.8835
Likeliood ratio	312.0313	8		<0.0001
C (SE)	0.8041 (0.0109)			
95% CI	0.7827-0.8255			
Likelihood ratio				
Model 3 vs. Model 1	12.9552	3		0.0047
Model 3 vs. Model 2	9.1547	1		0.0025
Difference between two correlated C	**Estimate (SE)**	**95% CI**	χ^2	**P**
Model 3 vs. Model 1	0.0023 (0.0022)	-0.0020, 0.0066	1.0678	0.3014
Model 3 vs. Model 2	0.0018 (0.0018)	-0.0017, 0.0053	1.0086	0.3152
	Estimate	**95% CI**	**Z**	**P**
NRI				
Model 3 vs. Model 1	0.0033	0.0012, 0.0054	3.0628	0.0022
Model 3 vs. Model 2	0.0160	-0.0034, 0.0354	1.6160	0.1061
IDI				
Model 3 vs. Model 1	0.0332	0.0029, 0.0636	2.1406	0.0323
Model 3 vs. Model 2	0.0021	0.0006, 0.0037	2.7307	0.0063

Table 2b. Risk equations of models 3 and 4.

	Model 3 incorporating pack-years and other risk factors		Model 4 incorporating time since quitting in past smoking status and other risk factors	
	Estimate (SE)	Hazard Ratio (95% CI)	Estimate (SE)	Hazard Ratio (95% CI)
Smoking variables				
Smoking status (Never smoker)				
Past smoker	0.3167**** (0.1306)	1.373 (1.063, 1.773)		
Pack-years				
<20	0.3320 (0.2734)	1.394 (0.816, 2.382)		
20-39	0.9049*(0.2271)	2.472 (1.584, 3.857)		
40+	0.9617*(0.1789)	2.616 (1.842, 3.714)		
Time since quitting				
≤5 years			0.7744* (0.1823)	2.169 (1.518,3.101)
>5 years			0.0163 (0.1310)	1.016 (0.786, 1.314)
Current smoker			0.7003* (0.1373)	2.014 (1.539,2.637)
Total	3751		3751	
Event	383		383	
Censored	3368		3368	
% censored	89.79		89.79	
	χ^2	P	χ^2	P
Test				
Likelihood ratio	455.5912	<0.0001	456.578	<0.0001
Hosmer lemeshow	9.21449	0.4177	19.0085	0.0251
Modified HL	8.56713	0.4782	17.9658	0.0355
C-statistic				
Estimate (SE)	0.8047 (0.0110)		0.8082 (0.0109)	
95% CI	(0.7832, 0.8263)		(0.7869, 0.8296)	

The coefficients for the smoking variables have been adjusted for the effects of age, sex, SBP, DBP, total cholesterol/HDL ratio, triglycerides, diabetes and alcohol. The reference category for a categorical variable is in the parentheses. *, **, *** and **** indicate P<0.0001, 0.001, 0.01 and 0.05, respectively. The statistical tests presented at the bottom of the table are for the full model versus the null model.

expected, improvement in c statistic for model 11 relative to models 7 through 10 is moderate as the c statistic for the latter (reference) models was already fairly high. AIC and BIC indicate that when model parsimony is taken into consideration, model 11 fits better than model 12 and models 7 through 10 (Table 3e). (Since models 11 and 12 are not nested, they have only been ranked using AIC and BIC.) Thus, model 11 is the best model.

Reclassification of subjects

Table 3f shows that when our best reduced model (model 11) was used compared to the reduced model with a current smoker/non-smoker measure (model 7), for all 3,751 subjects the net gain in reclassification proportion was significantly different from zero (P<0.0001). For comparing model 5 against model 11, reclassifications for subjects with and without events are summarized in Table

Table 2c. Improvement in prediction due to including time since quitting among past smokers in model 4.

Likelihood ratio	Value	Degrees of freedom		P
Vs. model 1	15.9421	2		0.0003
Vs. model 2	11.3852	1		0.0007
Difference between two correlated C	**Estimate (SE)**	**95% CI**	**χ^2**	**P**
Vs. model 1	0.0041 (0.0027)	-0.0011, 0.0094	2.3715	0.1236
Vs. model 2	0.0036 (0.0026)	-0.0014, 0.0087	2.0005	0.1572
Estimate	**95% CI**	**Z**	**P**	
NRI				
Vs. model 1	0.0394	0.0122, 0.0666	2.8332	0.0046
Vs. model 2	0.0222	-0.0092, 0.0537	1.3832	0.1666
IDI				
Vs. model 1	0.0020	-0.0004, 0.0046	1.5924	0.1112
Vs. model 2	0.0009	-0.0017, 0.0036	0.6823	0.4950

Table 2d. Risk equations of models 5 and 6.

	Model 5 incorporating both pack-years and time since quitting and other risk factors		Model 6 incorporating both intensity and duration of smoking and time since quitting and other risk factors	
	Estimate (SE)	Hazard Ratio (95% CI)	Estimate (SE)	Hazard Ratio (95% CI)
Smoking variables				
Smoking status (Never smoker)				
Past smoker				
Pack-years				
<20	0.1487 (0.3437)	1.160 (0.592, 2.276)		
20-39	0.8254**(0.2181)	2.283 (1.489, 3.501)		
40+	0.8038*(0.1646)	2.234 (1.618, 3.085)		
Time since quitting				
≤5 years	0.7813*(0.1822)	2.184 (1.528, 3.122)	0.7808*(0.1824)	2.183 (1.527, 3.122)
>5 years	0.0147 (0.1312)	1.015 (0.785, 1.312)	0.0179 (0.1311)	1.018 (0.787, 1.316)
Intensity<20 and				
duration <30			0.7112 (0.5091)	2.037 (0.751, 5.524)
30-39			0.7757****(0.3449)	2.172 (1.105, 4.271)
40+			0.3254 (0.3472)	1.385 (0.701, 2.735)
Intensity 20-39 and				
duration <30			0.4986 (0.4652)	1.646(0.661, 4.098)
30-39			0.7926***(0.2740)	2.209 (1.291, 3.780)
40+			0.5652****(0.2298)	1.760 (1.122, 2.761)
Intensity 40+ and				
duration <30			1.3344***(0.4654)	3.798 (1.525, 9.457)
30-39			1.2824** (0.3477)	3.605 (1.824, 7.127)
40+			1.0216****(0.4200)	2.778 (1.220, 6.327)
Total	3751		3751	
Event	383		383	
Censored	3368		3368	
% Censored	89.8		89.8	
	χ^2	P	χ^2	P
Likelihood ratio	463.7085	<0.0001	466.1424	<0.0001
Hosmer Lemeshow	16.0902	0.0650	13.9224	0.1251
Modified HL	15.2299	0.0848	13.1900	0.1541
C-statistic				
Estimate (SE)	0.8084 (0.0109)	0.8092 (0.0109)		
95% CI	(0.7869, 0.8298)	(0.7878, 0.8305)		

The coefficients for the smoking variables have been adjusted for the effects of age, sex, SBP, DBP, total cholesterol/HDL ratio, triglycerides, diabetes and alcohol. The reference category for a categorical variable is in the parentheses. *, **, *** and **** indicate P<0.0001, 0.001, 0.01 and 0.05, respectively. The statistical tests presented at the bottom of the table are for the full model versus the null model.

Table 2e. Comparison of model 5 with models 1 through 4.

Likelihood ratio	Value	Degrees of freedom		P
Vs. model 1	21.0725	4		0.0003
Vs. model 2	16.5156	3		0.0008
Vs. model 3	11.5018	1		0.0006
Vs. model 4	6.1304	2		0.0466
Difference between two correlated C	**Estimate (SE)**	**95% CI**	**χ^2**	**P**
Vs. model 1	0.0042 (0.0030)	-0.0016, 0.0101	2.0284	0.1544
Vs model 2	0.0037 (0.0029)	-0.0019, 0.0094	1.6927	0.1932
Vs. model 3	0.0036 (0.0026)	-0.0014, 0.0086	1.9864	0.1587
Vs. model 4	0.0001 (0.0013)	-0.0024, 0.0026	0.0075	0.9310
	Estimate	**95% CI**	**Z**	**P**
NRI				
Vs. model 1	0.0711	0.0347, 0.1075	3.8260	0.0001
Vs. model 2	0.0438	0.0131, 0.0745	2.7904	0.0052
Vs. model 3	0.0213	-0.0093, 0.0520	1.3630	0.1728
Vs. model 4	0.0043	-0.0099, 0.0187	0.5983	0.5496
IDI				
Vs. model 1	0.0037	0.0007, 0.0066	2.4830	0.0130
Vs. model 2	0.0025	-0.0004, 0.0056	1.6543	0.0980
Vs. model 3	0.0010	-0.0017, 0.0037	0.7267	0.4673
Vs. model 4	0.0016	0.0002, 0.0030	2.2511	0.0243

Table 2f. Improvement in prediction due to including both intensity and duration of smoking among current smokers in model 6.

Likelihood ratio	Value	Degrees of freedom		P
Vs. model 1	23.5064	10		0.0090
Vs. model 2	18.9495	9		0.0256
Difference between two correlated C	**Estimate (SE)**	**95% CI**	**χ^2**	**P**
Vs. model 1	0.0050 (0.0030)	-0.0009, 0.0110	2.7338	0.0982
Vs. model 2	0.0045 (0.0029)	-0.0012, 0.0103	2.3760	0.1232
	Estimate	**95% CI**	**Z**	**P**
NRI				
Vs. model 1	0.0440	0.0120, 0.0761	2.6845	0.0072
Vs. model 2	0.0268	0.0073, 0.0611	-1.5365	0.1244
IDI				
Vs. model 1	0.0031	-0.00008, 0.0062	1.9101	0.0561
Vs. model 2	0.0019	-0.0013, 0.0052	1.1773	0.2390

Table 3a. Reclassification table between the best model with a detailed smoking measure (model 5) and the model with a current/non-smoker smoking measure (model 1) as the reference model.

Model 1 Frequency (row per cent)	Model 5 <6 %	6-20 %	>20%	Total
Participants who experience a CVD event				
<6 %	148	27	0	175
6-20%	8	123	10	141
>20%	0	3	64	67
Total	156	153	74	383
Net gain in reclassification proportion (P)	0.0678 (0.0001)			
Participants who do not experience a CVD event				
<6 %	1746	56	0	1802
6-20%	105	1072	44	1221
>20%	0	6	339	345
Total	1851	1134	383	3368
Net gain in reclassification proportion (P)	0.0032 (0.4488)			
NRI (P)	0.0711 (0.0001)			

Table 2g. Information indices for models 1 through 6.

Model	BIC	AIC
1	5908.260	5836.753
2	5880.559	5834.197
3	5873.677	5833.183
4	5875.005	5824.811
5	5868.240	5823.681
6	5872.286	5833.247

3g. For all 3,751 subjects, the net gain in reclassification proportion was significantly different from zero (P=0.034). For 3,368 subjects who did not experience a CVD event, the net gain in reclassification proportion was significantly different from zero (P<0.0001) while for 383 subjects who experienced a CVD event this was not (P=0.0943).

If we had used model 5 rather than model 11, 18.02% of those who develop CVD would be appropriately assessed for their cardiovascular risk while only 10.2% of those who do not develop CVD would be falsely assessed for their cardiovascular risk, the difference of which is highly significant (P<0.0001).

Strengths and limitations

The equations derived from this study have several advantages over previous versions. The inclusion of detailed smoking measures rather than a crude dichotomous yes/no smoking measure in the risk equations is novel. We have also used a more recent population cohort to develop the risk equations. The influence of high density lipoprotein (HDL) cholesterol, which has been measured in the Framingham Heart Study since 1968, is also reflected in our equations. Also, the influence of alcohol intake which has been measured in the Framingham Heart Study since its initiation is reflected in the risk equations. Alcohol intake was not included in any of the previously developed risk equations for predicting CVD events based on the Framingham data.

One limitation of the present study is that the Framingham sample is predominantly Caucasian, so the generalizability of the CVD risk function in other ethnic groups is uncertain. Other Framingham risk functions have shown themselves to recalibrate well in other samples.[10,11,28,29]

Discussion

This study demonstrated the benefits of adding detailed smoking measures to CVD risk prediction equations. Incorporating pack-years among current smokers significantly improved prediction and so did time since

quitting among past smokers. When both these variables were incorporated in the smoking variable compared to using a simple current smoker/non-smoker variable, 7.1% of subjects changed their risk categories which was statistically significant.

We observed that the risk of CVD incidence first increased with pack-years but then slightly declined at high exposure to pack-years. Although in some other cohorts the beneficial effect of smoking cessation has been found to be as early as 1-3 years,[8,9] we observed that any major reduction in the risk of CVD incidence did not occur until after five years since quitting. Compared to never smokers, the risk of CVD incidence was almost twice as high for those who had quit within five years prior to baseline while this risk reduced to the level of never smokers for those who had quit more than five years ago. We observed that the risk of CVD incidence first increased with both intensity and duration of smoking but then declined at high duration and not at high intensities.

We also showed that if we had incorporated time since quitting and pack-years in smoking status instead of using a simple current/non-smoker measure, a significantly higher proportion of those developing a CVD event would have moved up to the 'high risk' category compared to those not having a CVD event who moved up to this category. The result was similar when we excluded laboratory based risk factors from the models. In each case, the former participants would be appropriately assessed while the latter would be falsely assessed if we incorporated time since quitting and pack-years in smoking status. Those appropriately assessed can benefit from additional screening for CVD risk and would require more aggressive intervention for smoking cessation,[41] and this would assist in preventing more deaths. However, this benefit would be at the cost of falsely identifying people who do not develop CVD as high risk. These people may unnecessarily receive additional screening, and it may even cause undue stress and burden to various smoking cessation programs. The performance of the model which incorporates time since quitting and pack-years, when considered in this perspective, relies on the relative importance one places on these benefits and costs. From a CVD prevention perspective, the benefits associated with smoking cessation for prevention of CVD clearly outweigh the costs associated with CVD screening and smoking cessation programs.

Our risk prediction equations are unlikely to be seriously affected by confounding bias as we have adjusted for all major confounders of smoking, including age, sex, alcohol consumption and the blood pressure components. We conducted further analyses of the offspring cohort to explore the possibility of any effect of distortion medical treatments may have on risk of CVD incidence, and found that the regression coefficients of the risk equations were fairly robust after adjustment for cardioactive medications (*data not shown*). We also explored the possibility of the effect of reverse causation, as there may be some reduction in the apparent risk of CVD incidence among current smokers (and hence among different categories of pack-years and their components) because of a tendency for people to give up smoking after they begin to be influenced by some life threatening condi-

Table 3b. Lack of fit statistics due to excluding laboratory risk factors from the equation with pack-years, time since quitting and other risk factors (models 11 vs. 5).

Test	Value	Degrees of freedom	P	
Likelihood ratio	56.4307	3	<0.0001	
Difference between two correlated C	**Estimate (SE)**	**95% CI**	**χ^2**	**P**
	0.0173 (0.0039)	0.0096, 0.0250	19.3872	<0.0001
	Estimate	**95% CI**	**Z**	**P**
NRI	0.0662	0.0220, 0.1104	2.9248	0.0034
IDI	0.0149	0.0079, 0.0218	4.2177	<0.0001

Table 3c. Lack of fit statistics due to excluding laboratory risk factors from the equation with intensity, duration of smoking, time since quitting and other risk factors (models 12 vs. 6).

Test	Value	Degrees of freedom	P	
Likelihood ratio	56.6429	3	<0.0001	
Difference between two correlated C	**Estimate (SE)**	**95% CI**	**χ^2**	**P**
	0.0173 (0.0040)	0.0095, 0.0250	18.8621	<0.0001
	Estimate	**95% CI**	**Z**	**P**
NRI	0.0625	0.0171, 0.1078	2.9248	0.0034
IDI	0.0151	0.0082, 0.0220	4.3264	<0.0001

*Indicates conditional likelihood ratio test for model 6 versus model 5.

Table 3d. Comparison of model 11 with models 7 through 10.

Likelihood ratio	Value	Degrees of freedom	P	
Vs. model 7	25.0982	4	<0.0001	
Vs. model 8	20.4029	3	0.0001	
Vs. model 9	15.6708	1	<0.0001	
Vs. model 10	6.7224	2	0.0346	
Difference between two correlated C	**Estimate (SE)**	**95% CI**	**χ^2**	**P**
Vs. model 7	0.0055 (0.0033)	-0.0010, 0.0121	2.7246	0.0988
Vs. model 8	0.0048 (0.0032)	-0.0015 , 0.0111	2.2025	0.1378
Vs. model 9	0.0043 (0.0029)	-0.0014, 0.0101	2.1770	0.1401
Vs. model 10	0.0004 (0.0013)	-0.0021, 0.0029	0.1125	0.7373
	Estimate	**95% CI**	**Z**	**P**
NRI				
Vs. model 7	0.0767	0.0398, 0.11361	4.0786	<0.0001
Vs. model 8	0.0435	0.0112, 0.0758	2.6323	0.0084
Vs. model 9	0.0023	0.0293, 0.0340	0.1431	0.8862
Vs. model 10	0.0140	-0.00002, 0.0281	1.9534	0.0507
IDI				
Vs. model 7	0.0048	0.0016, 0.0080	3.0020	0.0026
Vs. model 8	0.0035	0.0003, 0.0067	2.1472	0.0317
Vs. model 9	0.0020	-0.0009, 0.0049	1.3443	0.1788
Vs. model 10	0.0015	0.0001, 0.0028	2.1944	0.0282

Table 3e. Information indices for models 7 through 12.

Model	BIC	AIC
7	5914.898	5891.210
8	5916.151	5888.515
9	5923.315	5887.783
10	5913.592	5874.834
11	5906.418	5874.112
12	5923.315	5883.890

tion, such that their risk of CVD incidence may or may not be caused by smoking. To reduce this bias, we conducted a sub-analysis where we excluded from our baseline cohort those with a history of cancer and other non-CVD conditions (*data not shown*). This exclusion did not substantially affect the results. Sub-analyses conducted whereby we excluded from our baseline cohort those smokers who gave up smoking after examination 4 and those quitters who took up smoking after examination 4, and later those current smokers from our baseline cohort whose smoking intensity or duration changed dramatically in subsequent examinations, did not substantially alter our results.

Incorporating pack-years, time since quitting and individual components of pack-years in the smoking status variable, and adjusting for other risk factors in the risk equations, produced very good discrimination and calibration. However, since the c statistic for the CVD risk equations with the full set of risk factors ranged from 0.8041 to 0.8100, this suggested that additional risk factors (e.g. age at quitting, types of cigarettes, environmental smoking, abdominal obesity, family history of CVD, etc.) may further improve model discrimination.

In conclusion, we have demonstrated that CVD risk equations including a simple current smoker/non-smoker variable underestimate the overall effect of cigarette smoking, as risk prediction improves significantly when pack-years and time since quitting are incorporated. We advocate the use of improved risk prediction equations incorporating more detailed smoking measures in order to ensure current prevention strategies are optimally effective and cost-effective, and that ex-smokers are not consistently undertreated.

Table 3f. Reclassification table between the best reduced model with a detailed smoking measure (model 11) and the reduced model with a current/non-smoker smoking measure (model 7) as the reference model.

Model 7 Frequency (row per cent)	Model 11 <6%	6-20%	>20%	Total
Participants who experience a CVD event				
<6 %	149	29	0	178
6-20%	8	131	10	149
>20%	0	5	51	56
Total	157	165	61	383
Net gain in reclassification proportion (P)	0.0678 (0.0001)			
Participants who do not experience a CVD Event				
<6 %	1658	57	0	1715
6-20%	119	1136	41	1296
>20%	0	9	348	357
Total	1777	1202	389	3368
Net gain in reclassification proportion (P)	0.0089 (0.0459)			
NRI (P)	0.0767 (<0.0001)			

Table 3g. Reclassification table between the best full model with a detailed smoking measure (model 5) and the best reduced model with a detailed smoking measure (model 11) as the reference model.

Model 7 Frequency (row per cent)	Model 11 <6 %	6-20 %	>20%	Total
Participants who experience a CVD event				
<6 %	148	19	0	167
6-20%	18	119	23	160
>20%	0	10	46	56
Total	166	148	69	383
Net gain in reclassification proportion (P)	0.0366 (0.0943)			
Participants who do not experience a CVD Event				
<6 %	1691	86	0	1777
6-20%	160	995	67	1222
>20%	0	93	276	369
Total	1851	1174	343	3368
Net gain in reclassification proportion (P)	0.0297 (<0.0001)			
NRI (P)	0.0662 (0.034)			

References

1. US Dept of Health and Human Services. Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General. US Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 1989. DHHS Publication (CDC) 89-8411.
2. US Dept of Health and Human Services. The Health Benefits of Smoking Cessation. A Report of the Surgeon General. USDHHS, Centers for Disease Control. Office of Smoking and Health; 1990. DHHS Publication (CDC) 90-8416.
3. Shinton R, Beevers G. Meta-analysis of relation between cigarette smoking and stroke. BMJ 1989;298:789-94.
4. US Department of Health and Human Services, State Cardiovascular Disease Highlights, 1997, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. For copies of this CDC document, contact the Cardiovascular Health Branch, National Center for Chronic Disease Prevention and Health Promotion, GA, USA.
5. Wolf PA, D'Agostino RB, Kannel WB, et al. Cigarette smoking as a risk factor for stroke: the Framingham study. JAMA 1988;259:1025-9.
6. Hoogenveen R, van Baal P, Boshuizen H, et al. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. Cost Effectiveness and Resource Allocation 2008;6:1.
7. Ockene JK, Kuller LH, Svendson KH, et al. The relationship of smoking cessation to coronary heart disease and lung cancer in the Multiple Risk Factor Intervention Trial (MRFIT). Am J Public Health 1990;80:954-8.
8. Teo KK, Ounpuu S, Hawken S, et al; INTERHEART Study Investigators. Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. Lancet 2006;368:647-58.

9. Lightwood JM, Glantz SA. Short-term economic and health benefits of smoking cessation: myocardial infarction and stroke. Circulation 1997;96:1089-96.
10. Executive summary of the 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). JAMA 2001;285: 2486-97.
11. Anderson KM, Odell PM, Wilson PW, et al. Cardiovascular disease risk profiles. Am Heart J 1991;121:293-8.
12. British Cardiac Society. JBS 2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. Heart 2005;91:v1-v52.
13. Conroy RM, Pyorala K, Fitzgerald AP, et al;, for the SCORE Project Group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE Project. Eur Heart J 2003;24:987-1003.
14. De Backer G, Ambrosioni E, Borch-Johnsen K, et al; for the Third Joint Task Force of European and Other Societies on Cardiovascular Disease Prevention in Clinical Practice. European guidelines on cardiovascular disease prevention in clinical practice: Third Joint Task Force of European and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of eight societies and by invited experts). Eur Heart J 2003;24:1601-10.
15. Jackson R. Updated New Zealand cardiovascular disease risk-benefit prediction guide. BMJ 2000;320:709-10.
16. Menotti A, Lanti M, Gabiti-Rosei E, et al. New tools for prediction of cardiovascular disease risk derived from Italian population studies. Nutr Metab Cardiovasc Dis 2005;15:426-40.
17. Ridker PM, Buring JE, Rifai N, et al. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds risk score. JAMA 2007;297:611-9.
18. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335:136.
19. Woodward M, Brindle P, Tunstall-Pedoe H; for the SIGN Group on Risk Estimation. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). Heart 2007;93:172-6.
20. D'Agostino RB, Sr,Vasan RS, Pencina MJ, et al. General Cardiovascular Risk Profile for Use in Primary Care. Circulation 2008;117:743-53.
21. Anderson KM, Wilson PW, Odell PM, et al. An updated coronary risk profile: a statement for health professionals. Circulation 1991;83:356-62.
22. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Munster (PROCAM) Study. Circulation 2002;105:310-5.
23. Ferrario M, Chiodini P, Chambless LE, et al. Prediction of coronary events in a low incidence population: assessing accuracy of the CUORE Cohort Study prediction equation. Int J Epidemiol 2005;34:413-21.
24. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837-47.
25. Zhang XF, Attia J, D'Este C, et al. A risk score predicted coronary heart disease and stroke in a Chinese cohort. J Clin Epidemiol 2005;58:951-8.
26. Wolf PA, D'Agostino RB, Belanger AJ, et al. Probability of stroke: a risk profile from the Framingham study. Stroke 1991;22:312-8.
27. Murabito JM, D'Agostino RB, Silbershatz H, et al. Intermittent claudication: a risk profile from the Framingham Heart Study. Circulation 1997;96:44-9.
28. Kannel WB, D'Agostino RB, Silbershatz H, et al. Profile for estimating risk of heart failure. Arch Intern Med 1999;159:1197-204.
29. Doll R, Peto R, Boreham J, et al. Mortality in relation to smoking: 50 year's observations on male British doctors. BMJ 2004;328:1519.
30. Cox DR. Regression models and life tables. J Royal Stat Soc 1972;34(series B):187-220.
31. Hosmer DW Jr, Lemeshow S. Applied Survival Analysis: Regression Modeling of Time to Event Data. Wiley, New York, 1999.
32. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.
33. D'Agostino R, Nam BH. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Handbook of Statistics. Amsterdam, The Netherlands: Elsevier 2004:1-25.
34. Antolini L, Nam BH, Agostino RB. Inference on correlated discrimination measures in survival analysis: a nonparametric approach. Commun Stat Theory Methods 2004;33:2117-35.
35. Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. American Journal of Epidemiology 2004;159:882-90.
36. Greenland P, O'Malley PG. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase A2 and C-reactive protein for stroke risk. Arch Intern Med 2005;165:2454-6.
37. Ware JH. The limitations of risk factors as prognostic tools. New Eng J Med 2006; 355:2615-7.
38. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27: 157-72.
39. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 2004;23:2109-23.
40. Pepe MS, Feng Z, Huang Y, et al. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol 2008;167:362-8.
41. Executive summary of the 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). JAMA 2001;285: 2486-97.