Cardiovascular disease risk profiles

Keaven M. Anderson, PhD, Patricia M. Odell, PhD, Peter W. F. Wilson, MD, and William B. Kannel, MD, MPH *Framingham and Boston, Mass.*

This article presents prediction equations for several cardiovascular disease endpoints, which are based on measurements of several known risk factors. Subjects (n = 5573) were original and offspring subjects in the Framingham Heart Study, aged 30 to 74 years, and initially free of cardiovascular disease. Equations to predict risk for the following were developed: myocardial infarction, coronary heart disease (CHD), death from CHD, stroke, cardiovascular disease, and death from cardiovascular disease. The equations demonstrated the potential importance of controlling multiple risk factors (blood pressure, total cholesterol, high-density lipoprotein cholesterol, smoking, glucose intolerance, and left ventricular hypertrophy) as opposed to focusing on one single risk factor. The parametric model used was seen to have several advantages over existing standard regression models. Unlike logistic regression, it can provide predictions for different lengths of time, and probabilities can be expressed in a more straightforward way than the Cox proportional hazards model. (AM HEART J 1990;121:293-8.)

The Framingham Heart Study has been operational for more than 40 years and has identified a number of risk factors that interact in a deleterious manner to have a cumulative impact on cardiovascular disease (CVD). Experience has shown that a multifactorial approach, one that takes into consideration all the risk factors, is probably the best strategy for the prevention of coronary heart disease (CHD).

This article presents prediction equations for several CVD endpoints based on measurements of several known risk factors. They may be used for estimating outcome probabilities over a range of 4 to 12 years for persons aged 30 to 74 years. These equations are then compared with a recently developed equation for predicting CHD[1] to see if that single profile predicts the related endpoints equally well. Separate profiles with diastolic (DBP) and systolic blood pressure (SBP) are included for all outcomes. A method for developing confidence intervals for predicted probabilities, hazard ratios, and excess risk estimates[2] is also presented. Examples illustrate how to use the equations and calculate the confidence intervals.

METHODS

The population studied consisted of 5573 members of the Framingham Heart Study and Framingham Offspring Study cohorts, who ranged in age from 30 to 74 years. Baseline characteristics were measured from 1968 through 1975, and 12 years of follow-up were included. Only persons free of CVD and cancer (other than basal cell carcinomas) were included in the study. (For further details, see Anderson et al.,[1] who detail the development of the equation for CHD that is presented here.)

Equations were developed for the following outcomes: myocardial infarction (MI, including silent and unrecognized MI); death from CHD (sudden or nonsudden); CHD (consisting of MI and CHD death plus angina pectoris and coronary insufficiency); stroke, including transient ischemia; CVD (including all the above plus congestive heart failure and peripheral vascular disease); and death from CVD (CVD death).

A parametric statistical model[2] was used to provide predicted probabilities for each of the outcomes. This modeling is based on risk factor levels and (possibly censored) times until events. Let T denote the time until the event of interest. Assume $x_1, x_2 \ldots x_k$ represents the risk factor measurements for an individual. For example, x_1 might be age in years, x_2 systolic blood pressure, and so forth. The coefficients $\beta_0, \beta_1 \ldots \beta_k$, as well as θ_0 and θ_1, will represent the parameters that we will estimate. The value $\mu = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$ is assumed to be a linear function of the risk factors and $\log \sigma = \theta + \theta_1 \mu$ is considered to be a linear function of μ.

To compute the probability that time until event is less than some arbitrary time t for given values of μ and σ, let

$$u = \frac{\log(t) - \mu}{\sigma} \quad \text{(Equation 1).}$$

Assume

$$P(T > t) = P\left\{\frac{\log(T) - \mu}{\sigma} > u\right\}$$

$$= 1 - \exp(-\exp(u)) \quad \text{(Equation 2).}$$

Equation 2 is the predicted probability of an event by

From the National Heart, Lung, and Blood Institute, and Boston University.

Reprint requests: Keaven Anderson, PhD, Framingham Heart Study, 5 Thurber St., Framingham, MA 01701.

4/0/24902

Table I. SBP prediction equation coefficients for all outcomes studied

Coefficients	*CHD*	*MI*	*CHD death*	*Stroke*	*CVD*	*CVD death*
θ_0	0.9145	3.4064	2.9851	−0.4312	0.6536	0.8207
θ_1	−0.2784	−0.8584	−0.9142	—	−0.2402	−0.4346
β_0	15.5305	11.4712	11.2889	26.5116	18.8144	−5.0385
female	28.4441	10.5109	0.2332	0.2019	−1.2146	0.2243
log(age)	−1.4792	−0.7965	−0.9440	−2.3741	−1.8443	8.2370
$(\log(age))^2$	—	—	—	—	—	−1.2109
log(age) × female	−14.4588	−5.4216	—	—	0.3668	—
$(\log(age))^2$ × female	1.8515	0.7101	—	—	—	—
log(SBP)	−0.9119	−0.6623	−0.5880	−2.4643	−1.4032	−0.8383
cigarettes (Y/N)	−0.2767	−0.2675	−0.1367	−0.3914	−0.3899	−0.1618
log(total-C ÷ HDL-C)	−0.7181	−0.4277	−0.3448	−0.0229	−0.5390	−0.3493
diabetes	−0.1759	−0.1534	−0.0474	−0.3087	−0.3036	−0.0833
diabetes × female	−0.1999	−0.1165	−0.2233	−0.2627	−0.1697	−0.2067
ECG-LVH	−0.5865	—	−0.1237	−0.2355	−0.3362	−0.2946
ECG-LVH × male	—	−0.1588	—	—	—	—

Table II. DBP prediction equation coefficients for all outcomes studied

Coefficients	*CHD*	*MI*	*CHD death*	*Stroke*	*CVD*	*CVD death*
θ_0	0.9341	3.4587	2.1249	−0.4212	0.6761	0.9076
θ_1	−0.2825	−0.8647	−0.6860	—	−0.2421	−0.4528
β_0	15.5222	11.0436	12.0963	25.1067	17.5392	−9.0211
female	32.4811	5.1559	0.2619	0.1558	−0.8019	0.2102
log(age)	−1.6346	−0.9302	−1.3025	−3.0997	−2.1231	9.5223
$(\log(age))^2$	—	—	—	—	—	−1.3999
log(age) × female	−16.4933	−2.6310	—	—	0.2584	—
$(\log(age))^2$ × female	2.1059	0.3472	—	—	—	—
log(DBP)	−0.8670	−0.5132	−0.4762	−1.7556	−1.0117	−0.5073
cigarettes (Y/N)	−0.2789	−0.2721	−0.1553	−0.3975	−0.3900	−0.1548
log(total-C ÷ HDL-C)	−0.7142	−0.4228	−0.4056	0.0297	−0.5365	−0.3423
diabetes	−0.2082	−0.1764	−0.0860	−0.4047	−0.3575	−0.1178
diabetes × female	−0.1973	−0.1184	−0.2539	−0.2506	−0.1661	−0.1982
ECG-LVH	−0.7195	—	−0.1591	−0.2801	−0.3847	−0.3181
ECG-LVH × male	—	−0.1702	—	—	—	—

time t. This implies T follows a Weibull distribution. In general, a negative β coefficient for a variable means that a high value of that variable is associated with high risk.[2]

Each model was estimated in two steps. First, covariates were chosen separately for each sex that appeared to model age well. For different models this may involve a quadratic age term or interactions between the age covariates and sex. Then covariates representing additional risk factors were added. Separate equations were developed with the use of SBP and DBP; except for the blood pressure covariate, the models are identical. The maximum likelihood method was used to estimate parameters.

In addition to differences in age covariates, there are two deviations from the general model: (1) The model for stroke does not contain a θ_1 parameter, because its inclusion results in little improvement in the log likelihood. (2) The model for MI includes ECG–left ventricular hypertrophy (ECG-LVH) only for men, because for woman its coefficient is positive and nonsignificant.

A hazard ratio is similar to a risk or odds ratio. Again, using a Weibull distribution as in equation 2, assume two persons have predicted probabilities p_1 and p_2, respectively. Then the hazard ratio of individual one relative to individual two at time t is

$$\frac{\log(1 - p_1)}{\log(1 - p_2)}.$$

The excess risk of individual one relative to individual two after time t is the difference in the predicted probabilities of disease for the two individuals $(p_1 - p_2)$.

The delta method described by Anderson[2] may be used to provide confidence intervals, predicted probabilities, hazard ratios, and excess risk. Examples of confidence intervals are shown in the following sections. Details of the calculations may be found in the Appendix.

Table III. The 10-year CHD risk prediction for 65-year-old nonsmoking, nondiabetic men without ECG-LVH (except where noted); excess risks and hazard ratios relate to SBP 120 mm Hg, total cholesterol 180 mg/dl (4.66 mmol/L), and HDL cholesterol 45 mg/dl (1.17 mmol/L)

SBP (mm Hg)	*Total cholesterol mg/dl (mmol/L)*	*HDL cholesterol mg/dl (mmol/L)*	*10-yr predicted risk*	*Hazard ratio*	*Excess risk*
160	240 (6.22)	38 (0.98)	27.4% (23.8%, 31.5%)*	2.5 (2.1, 3.1)	15.5% (12.1%, 18.8%)
140	250 (6.48)	35 (0.91)	26.4% (22.8%, 30.4%)	2.4 (2.0, 2.9)	14.4% (11.5%, 17.3%)
140	220 (5.70)	42 (1.09)	19.9% (17.1%, 23.0%)	1.7 (1.5, 2.0)	7.9% (6.4%, 9.5%)
120	240 (6.22)	38 (0.98)	19.7% (16.5%, 23.4%)	1.7 (1.5, 1.9)	7.8% (5.9%, 9.6%)
110	250 (6.48)	35 (0.91)	20.0% (16.3%, 24.4%)	1.7 (1.5, 2.0)	8.0% (5.4%, 10.6%)
160 (ECG-LVH yes)	240 (6.22)	38 (0.98)	47.1% (33.1%, 63.5%)	5.0 (3.0, 8.2)	35.1% (19.6%, 50.6%)

*The 95% confidence intervals are shown in parentheses.

The quantity u, computed in equation 1, provides a useful unit for comparison of risks for different persons. Although it is a simple function of the predicted probability of disease, it is more normally distributed since it is not restricted to the interval (0,1). To compare models for different end points, u values were computed for each model for each of the 5573 persons in the population. All models are heavily dependent on age, certainly an unmodifiable risk factor. To clarify relationships among models based on the remaining risk factors, u values were corrected for age effects for each model. (This was done by taking residuals from a linear regression with u as the dependent variable and log(age) and [log(age)]^2 as the independent variables.) Correlations of these age-corrected residuals of u values were then obtained to measure the strengths of association between the different models.

RESULTS

Tables I and II present coefficients for the estimated equations. The risk factors are denoted as follows: age—age in years; female—1 if female, 0 if male (female sex is a protective risk factor); male—0 if male, 1 if female; SBP—average of two office measurements of SBP (mm Hg); DBP—average of two office measurements of DBP (mm Hg); total cholesterol—total serum cholesterol (mg/dl) as measured by the Abell-Kendall method[3]; high-density lipoprotein (HDL) cholesterol (mg/dl) determined after heparin manganese precipitation; cigarettes—1 if cigarette smoker (or quit within last year), 0 otherwise; diabetes—1 if diabetes, 0 otherwise (diabetes is defined as under treatment with insulin or oral agents or a fasting glucose of 140 mg/dl or above[4]); and ECG-LVH—1 if definite ECG-LVH, 0 otherwise. SBP is used in Table I; DBP is used in Table II. Time intervals of 4 to 12 years are recommended.

Table IV. Estimated percentiles of 6-year CHD predicted probabilities for each age group in the population studied

Men				*Women*			
(yr)	*10%*	*50%*	*90%*	*(yr)*	*10%*	*50%*	*90%*
30-34	0.3%	1.0%	2.9%	30-34	<0.1%	<0.1%	0.3%
35-39	0.6%	2.0%	5.1%	35-39	<0.1%	0.2%	1.5%
40-44	1.3%	3.4%	7.5%	40-44	0.2%	0.7%	2.3%
45-49	1.9%	5.2%	11.0%	45-49	0.6%	1.6%	4.7%
50-54	3.3%	7.0%	14.2%	50-54	1.0%	2.9%	8.2%
55-59	4.3%	9.1%	17.5%	55-59	1.9%	4.6%	11.8%
60-64	6.0%	11.7%	20.8%	60-64	2.3%	5.5%	12.4%
65-69	8.5%	14.7%	25.6%	65-69	3.0%	6.2%	13.4%
70-74	9.2%	15.1%	26.5%	70-74	3.3%	7.4%	14.6%

Predicted probability: An example. As an example of how to compute a predicted probability, consider the CHD equation for a 55-year-old woman with diabetes who smokes, has an SBP of 135 mm Hg, total cholesterol of 230 mg/dl (5.96 mmol/L), HDL cholesterol of 48 mg/dl, and no ECG-LVH. Using Table I, begin by computing $\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{female} + \hat{\beta}_2 \times \log(\text{age}) + \hat{\beta}_3 \times \log(\text{age}) \times \text{female} + \hat{\beta}_4 \times [\log(\text{age})]^2 \times \text{female} + \hat{\beta}_5 \times \log(\text{SBP}) + \hat{\beta}_6 \times \text{cigarettes} + \hat{\beta}_7 \times \log(\text{total cholesterol} \div \text{HDL cholesterol}) + \hat{\beta}_8 \times \text{diabetes} + \hat{\beta}_9 \times \text{diabetes} \times \text{female} = 15.5305 + 28.4441 - (1.479 + 14.4588) \times \log(55) + 1.8515 \times [\log(55)]^2 - 0.9119 \times \log(135) - 0.2767 - 0.7181 \times \log(230/48) - 0.1759 - 0.1999 = 3.588$.

Next compute: $\log(\hat{\sigma}) = \theta_0 + \theta_1\hat{\mu} = 0.9145 - 0.2784 \times \hat{\mu} = -0.08430$, so $\hat{\sigma} = e^{-0.0843} = 0.9192$. If $t = 10$ years, we have

$$\hat{u} = \frac{\log(10) - \hat{\mu}}{\hat{\sigma}} = -1.398.$$

Table V. Parameters, covariate means, and parameter covariance matrix needed to compute delta method estimates of standard errors

Model coefficients							
θ_0	β_0	β_1 female	β_2 male × log(age)	β_3 female × log(age ÷ 74)^2	β_4 log(SBP)	β_5 cigarettes	β_6 log(total-C ÷ HDL-C)
−0.31546	4.41815	−5.85489	−1.47921	1.85148	−0.91192	−0.27667	−0.71811
Location	$(\overline{X})$						
		0.53526	1.78404	0.13875	4.85349	0.39727	1.44776
Parameter covariance matrix (Cov or C)							
θ_0	β_0	female	male × log (age)	female × log(age ÷ 74)^2	log(SBP)	cigarettes	log (total-C ÷ HDL-C)
0.00341							
0.00684	0.01629						
−0.03807	−0.09178	1.00413					
−0.00956	−0.02289	0.25204	0.06351				
0.01483	0.41039	−0.30124	−0.07325	0.18982			
−0.00499	−0.01170	0.05898	0.01510	−0.02402	0.04211		
−0.00158	−0.00394	0.03119	0.00776	−0.01326	0.00421	0.00349	
−0.00476	−0.01153	0.08615	0.02137	−0.03395	0.01017	0.00328	0.01609
−0.00111	−0.00255	0.00931	0.00198	−0.00894	0.00102	0.00096	0.00202
−0.00108	−0.00237	0.02574	0.00728	−0.00316	0.00205	0.00048	0.00189
−0.00362	−0.00816	0.05805	0.01440	−0.02800	0.00286	0.00274	0.00829
0.00464	0.01082	−0.09000	−0.02253	0.03900	−0.01139	−0.00374	−0.01136

Thus the 10-year predicted probability for CHD is $1 - \exp(-\exp(-1.398)) = 0.22$.

Multiple risks: An example. Table III presents an example of predicted risk as a function of both SBP and lipoprotein values. It is meant to suggest that controlling multiple risk factors rather than just blood pressure should be considered as a risk reduction strategy. All examples in the table assume a 65-year-old, nonsmoking, nondiabetic male without ECG-LVH. To begin, we assume that SBP is 160 mm Hg, total cholesterol 240 mg/dl (6.22 mmol/L), and HDL cholesterol 38 mg/dl (0.98 mmol/L). In the second line, we consider lowering SBP to 140 mm Hg, but worsening both lipoprotein values by approximately 10%, which results in about the same predicted risk as the original profile. In the third line, we again consider SBP to be 140 mm Hg and improve the two lipoprotein values by about 10%, which results in a greater than 20% reduction in predicted CHD risk. In the fourth and fifth examples of the table, we show that with lipoprotein values unchanged or slightly worsened, larger changes in SBP are required to obtain a lowering of predicted risk of an amount comparable with the third line of the table. Note that this example is not meant to imply that changing risk factors in persons by these amounts would result in the stated changes in CHD risk. These figures represent the differences between persons in an epidemiologic study. Smaller differences would probably be expected from risk factor improvement because of residual effects of previously higher levels.

Note that most of the 95% confidence intervals in this table are relatively small. Risk associated with the less frequent covariates of ECG-LVH and diabetes is much less reliably estimated. The last line of the table is comparable with the first except that ECG-LVH is added. This results in a large increase in the estimated risk, but a much larger increase in the estimated confidence intervals. Thus although we know this person is at elevated risk, it is difficult to know to what extent.

Typical risk levels. To help clarify the meaning of the equations, predicted probabilities of CHD within 6 years were obtained for each person in the population from which the equations were estimated. This was done for each 5-year age group, separately for men and women. The tenth and ninetieth percentiles of these distributions are offered as guidelines for high and low risks within each group and are shown in Table IV. This is suggested as more realistic than the arbitrary risk factor levels that were previously used to illustrate high and low risk.[5] In that publication estimates of high risk assumed the presence of both diabetes and ECG-LVH. These conditions are

β_7 Diabetes	β_8 female × diabetes	β_9 ECG-LVH	θ_1
−0.17591	−0.19987	−0.58653	−0.27843
0.06047	0.027633	0.007716	

Diabetes	female × diabetes	ECG-LVH	θ_1
0.01226			
−0.01110	0.02613		
0.00191	0.00156	0.05525	
−0.00262	−0.00232	−0.00898	0.01285

relatively rare, especially in combination. Thus the assumption resulted in unrealistically elevated high-risk estimates, which could lead to unwarranted complacency in those with somewhat lower calculated risks. For example, in that article 35-year-old men had a high 6-year risk of 16.9%, which is much higher than scores for the vast majority of such men. The present equation gives a ninetieth percentile risk of 5.1%. Thus those persons whose risks are near 5% are, in fact, at relatively high risk compared with other men of their age. With no improvement in covariate values, this risk will become much higher as they become older. It would also appear more alarming if the projection was for a longer time period.

Although the CHD equation has been used in each of the preceding examples, any of the other equations might have been used instead if different endpoints were of interest. The procedure is the same in each case; only the values of the coefficients vary.

Any of the equations that use DBP may also be used. For most outcomes (CHD, MI, CHD death in particular) differences in predictive probabilities are slight. Because the log likelihoods are slightly higher when SBP is used, we recommend this be done if convenient. The differences in outcome are not statistically significant when SBP as opposed to DBP is used. The situation changes for stroke and to a lesser extent CVD (which includes stroke). Here the differences are substantial, with SBP producing considerably higher log likelihoods.

Correlations of the age-corrected u residuals for CHD with those for other endpoints were above 0.95 with all endpoints except stroke. These high correlations might be expected, because the endpoints have many events in common. Because these correlations are high except for stroke, there appears to be little practical difference between the weighting of risk factors in the different equations. However, risk factor weightings are statistically significant for different equations.

At a glance, Tables I and II show that with regard to risk factors, stroke is substantially different from the "heart" diseases in two ways. First, blood pressure is even more strongly associated with stroke than with the other endpoints; second, total cholesterol and HDL cholesterol are of little statistical significance. Because of these differences, it might be expected that an equation developed to predict CHD would not be particularly effective in estimating the risk of stroke. To some extent, this proves to be the case. However, even for stroke, the CHD equation is a fairly good predictor (for the two u residuals, $r = 0.64$ with SBP; $r = 0.58$ with DBP).

DISCUSSION

Evaluation of the CHD equation. A major aim of this article has been to evaluate further the CHD equation[1] and, if possible, to extend its use to other endpoints. The confidence intervals that have been developed indicate that most estimates of predicted probabilities are fairly accurate. For ECG-LVH, this is not the case because the low prevalence of this condition in the population studied does not provide much information on the degree of its association with CHD. To a lesser extent, the same is true for diabetes.

Comparisons of the various equations suggest that the CHD equation does reasonably well at discriminating between relatively high- and low-risk persons for all the endpoints studied. This is not surprising since, except for stroke, the various endpoints have much in common with CHD. The association between the CHD and stroke equations is weaker than that for CHD with other outcomes. Because of the differences in origin, a separate equation for predicting stroke may be desirable. (For such an equation incorporating the additional risk factors, see Wolf et al.[6])

Advantages of methodology. The parametric model used in this article has several advantages over other standard regression models. Unlike logistic regression, it can provide predictions for different lengths of time. The predictive probabilities can be expressed in a way that is more straightforward than for the Cox proportional hazards model. The latter, which has a nonparametric component, is less simply summarized. The assumptions of proportional hazards (constant σ) is not required for this model, as it is for both the Cox and standard Weibull. Except for the stroke model, allowing σ to vary (i.e., not requiring that $\theta_1 = 0$) permits a better fit.

REFERENCES

1. Anderson KM, Wilson PWF, Odell PM, Kannel WB. An updated coronary risk profile. Circulation, Jan. 1991;83:357-63.
2. Anderson KM. A non-proportional hazards Weibull accelerated failure time model. Biometrics (in press).
3. Abell LL, Levy BB, Brodie BB, Kendall FE. A simplified method for the estimation of total cholesterol in serum and demonstration of its specificity. J Biol Chem 1952;195:357-66.
4. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. Diabetes 1983;28:1039-57.
5. McGee D. The Framingham Study: an Epidemiologic Investigation of Cardiovascular Disease, Section 27. Bethesda, Md.: U.S. Government Printing Office, 1973.
6. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke; a risk profile from the Framingham Study. Stroke (in press).

APPENDIX

In this section we give details of the computation of the confidence intervals presented in Table III. Because the numbers in Table V are rounded, recalculating the computations in Table III produces some differences of 1 in the last decimal place shown. Again we refer to the CHD equation with SBP. First, a reparameterization of the problem is presented. The average of each covariate in the population is subtracted from the individual values before the equation is estimated. The reparameterization results in different values for θ_0, β_0, and the coefficient for female. The first two sections of Table V give the coefficients and means, respectively. The estimated covariance matrix for the parameter estimates is presented in the bottom part of the table. We will use β to denote the vector of coefficients $(\beta_0, \beta_1 \ldots \beta_9)$ in Table V. Similarly, θ will represent (θ_0, θ_1). In the following, t will denote some fixed length of time, such as 10 years. The vector X will correspond to a vector with 1, followed by the covariate values for an individual denoted underneath the symbols β_1 to β_9 in Table V. Finally, $\overline{X}$ and c are as denoted in Table V.

The first value for which a confidence interval is computed is

$$u(\beta,\theta) = \frac{\log(t) - \beta'(X - \overline{X})}{\exp(\theta_0 + \theta_1\beta'[X - \overline{X}])}.$$

We let $sd(\hat{u})$ denote the delta method standard deviation estimate of $\hat{u} = \hat{u}(\hat{\beta}, \hat{\theta})$. A 95% confidence interval for u is computed as $\hat{u} + 1.96sd(\hat{u})$, labeled (u_L, u_U). The 95% confidence interval for the predicted probability of an event by time t given the covariate vector X is then $(F(u_L), F(u_U))$, where $F(u) = 1 - \exp(-\exp u)$. To compute $sd(\hat{u})$ we need first the vector of partial derivatives.

$D_u = (\partial u/\partial\theta_0, \partial u/\partial\beta_0, \partial u/\partial\beta_1 \ldots \partial u/\partial\beta_K, \partial u/\partial\theta_1)$. The values in D_u may be computed as follows: $\partial u/\partial\theta_o = -u$; $\partial u/\partial\beta_0 = -1/\sigma$; $\partial u/\partial\beta_i = -x_i(1/\sigma + \theta_1 u)$, $i = 1, 2, \ldots, k$; $\partial u/\partial\theta_1 = u(\beta_1\chi_1 + \ldots + \beta_k\chi_k)$. We substituted $\hat{\beta}$, $\hat{\theta}_0$, and $\hat{\theta}_1$ for the true values β, θ_0, and θ_1 in the above equations to provide estimates.

Once D_u is computed, let C be the covariance matrix in Table V and let $sd(\hat{u}) = (D'_u C D_u)^{1/2}$. The confidence interval for the hazard ratio is computed as described above. Let $w = u_1 - u_2$, where u_1 and u_2 correspond to the two profiles for which you are computing a hazard ratio. Compute D_{u1} and D_{u2} as above and then let $D_w = D_{u1} - D_{u2}$. Next compute $sd(\hat{w}) = (D'_w C D_w)^{1/2}$, $w_L = \hat{w} - 1.96sd(\hat{w})$, and $w_U = \hat{w} + 1.96sd(\hat{w})$. Finally, the hazard ratio estimate is $\hat{R} = e^{\hat{w}}$, and its 95% confidence interval is from $R_L = exp(w_L)$ *to* $R_U = exp(w_U)$.

To compute a confidence interval for the absolute difference in two estimated probabilities, let u_1 and u_2 be as above. Then let $D = \exp(u_1 - \exp(u_1))D_{u1} + \exp(u_2 - \exp(u_2))D_{u2}$. Next compute $\Delta = u_1 - u_2$ and $sd(\hat{\Delta}) = (D'CD)^{1/2}$. Finally, the estimate of the difference in the predicted probabilities is $\Delta = \exp(-\exp(u_2)) - \exp(-\exp(u_1))$, the lower limit of the 95% confidence interval is $\Delta - 1.96sd(\hat{\Delta})$, and the upper confidence limit is $\Delta + 1.96sd(\hat{\Delta})$.