STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:1631–1660 (DOI: 10.1002/sim.1742)

TUTORIAL IN BIOSTATISTICS
Presentation of multivariate data for clinical use: The Framingham Study risk score functions

Lisa M. Sullivan[1,2,*,†], Joseph M. Massaro[1,2] and Ralph B. D'Agostino Sr.[1,2]

[1]*Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, Boston, MA 02115, U.S.A.*
[2]*Boston University Statistics and Consulting Unit, Department of Mathematics and Statistics, 111 Cummington Street, Boston, MA 02215, U.S.A.*

SUMMARY

The Framingham Heart Study has been a leader in the development and dissemination of multivariable statistical models to estimate the risk of coronary heart disease. These models quantify the impact of measurable and modifiable risk factors on the development of coronary heart disease and can be used to generate estimates of risk of coronary heart disease over a predetermined period, for example the next 10 years. We developed a system, which we call a points system, for making these complex statistical models useful to practitioners. The system is easy to use, it does not require a calculator or computer and it simplifies the estimation of risk based on complex statistical models. This system represents an effort to make available a tool for clinicians to aid in their decision-making process regarding treatment and to assist them in motivating patients toward healthy behaviours. The system is also readily available to patients who can easily estimate their own coronary heart disease risk and monitor this risk over time.

KEY WORDS: regression models; points system; multivariable analysis; risk scores; prediction functions

1. INTRODUCTION

The Framingham Heart Study has been a leader in the development and dissemination of multivariable models to estimate the risk of coronary heart disease [1, 2]. These models quantify the impact of measurable and modifiable risk factors on the development of coronary heart disease. The models can be used to generate estimates of risk of coronary heart disease and are useful for determining appropriate treatments and for motivating patients to change behaviours.

The multivariable models essentially contain weights that are associated with risk factors such as age, blood pressure, cholesterol, treatment for hypertension or hypercholesterolaemia,

*Correspondence to: Lisa M. Sullivan, Statistics and Consulting Unit, Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215, U.S.A.
†E-mail: lsull@bu.edu

Received April 2003
Accepted July 2003

smoking status and diabetes. By entering a particular individual's risk factor profile, a risk of coronary heart disease over a specified time frame (e.g. 10 years) can be generated. The risk factors that are considered in the model are those that are considered to be significantly associated with coronary heart disease and are in addition readily available in clinical practice. There are other risk factors that might warrant inclusion but because they can only be measured by time-consuming, costly or invasive testing procedures they are not generally considered. There are still other risk factors such as exercise and nutrition that are important but can be very difficult to measure with adequate precision. To make the models as easy as possible to use in clinical practice and to minimize noise, we restrict our attention to those risk factors that are generally accepted as readily available in clinical practice and precisely measured. Since these models contain what are considered the clinically important risk factors, they are viewed as generally applicable to other populations. The issues that affect transportability are the distributions of the risk factors and the incidence rate of the outcome event, these, however, can be handled with minor adjustments to the model (described below).

One of the initial primary hypotheses of the Framingham Study was the identification of a single 'cause' for coronary heart disease. It quickly became apparent that the disease process was multifactorial and multivariable models were necessary and appropriate for quantifying the impact of multiple risk factors. The Framingham Study has a long history of producing these multivariable models and initial models date back to the 1960s. The first models were based on logistic regression analysis and discriminant function analysis [3–5]. As more data were accumulated (i.e. serial assessments of the risk factors and longer follow-up for events), survival analysis techniques were used and models were updated [1, 2, 6]. The Framingham Study has also produced multivariable models for specific events such as stroke [7, 8], peripheral vascular disease [9] and congestive heart failure [10]. Still other multivariable models have been produced for subsequent events; these models quantify the effects of risk factors on repeat events in persons who have a history of coronary disease [11]. The underlying objective in each of these models was to determine the function that best predicted the likelihood of an event (e.g. initial coronary heart disease, subsequent event) based on readily available and measurable risk factors. While Framingham has a long history of developing and disseminating these models, there have been concerns over the years related to their application to populations that differ ethnically, racially, according to risk factor prevalence or event incidence.

In 1999, the National Heart Lung and Blood Institute convened a workshop to assess the validity of the Framingham coronary heart disease functions in ethnically and racially diverse populations. The populations included whites, blacks, Native Americans, Japanese American men and Hispanic men. The performance of the Framingham sex-specific coronary heart disease models was evaluated and contrasted with the performance of sex-specific models developed on each population. Performance was assessed by three criteria: similarity of the weights assigned to each risk factor, discrimination of the model (i.e. its ability to correctly distinguish individuals who do and do not suffer coronary heart disease) and calibration (i.e. agreement between predicted event rates and actual event rates). The Framingham models performed well in whites and blacks, and with some minor recalibration adjustment can be applied to other ethnic groups [12]. This assessment provided clinicians with evidence that the Framingham models could be transported to other settings and in those different settings the models performed well.

In this paper, we summarize our approach to disseminating these multivariable models for routine use in clinical practice. The multivariable mathematical models are fairly complex. The strategy we propose for simplifying the computations to produce risk estimates can be applied without a calculator or computer. In Section 2, we motivate the problem with a recent application. In Section 3, we describe the logic of our approach for estimating risk, which we call a points system. In Section 4, we provide a specific example using multiple logistic regression analysis. In Section 5, we detail the general step-by-step algorithm for setting up a points system and in Section 6, we provide a second example using Cox proportional hazards regression analysis. In Section 7 we discuss interpretation issues.

2. MOTIVATION

The National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) was released in 2001 and updated the existing guidelines for clinical management of high cholesterol. The guidelines were meant to supplement clinical judgement and identified individuals as candidates for intensive preventative management based on their absolute 10-year risk of coronary heart disease [13]. Models to estimate the absolute risk of coronary heart disease were developed based on the Framingham data and are included in the NCEP ATPIII guidelines. The models were developed as follows.

Since persons with existing coronary heart disease were considered at sufficiently high risk for intensive management of cholesterol, as were persons with diabetes, the population at risk included persons aged 30–79 years who were free of coronary heart disease and diabetes at the baseline examination when risk factors were measured. Samples of $n = 4261$ men and 5182 women who were free of coronary heart disease were followed for 12 years for the development of coronary heart disease (CHD), defined as myocardial infarction or coronary death. Sex-specific Cox proportional hazards regression models were developed relating development of CHD to age, systolic blood pressure, total cholesterol, high-density lipoprotein (HDL), smoking status and treatment for hypertension. Interactions between age and total cholesterol and between age and smoking were also included in the sex-specific models.

The treatment guidelines match specific treatment strategies to the absolute 10-year risk of coronary disease. If the 10-year risk exceeds 20 per cent patients are indicated for aggressive treatment, if the 10-year risk is between 10 and 20 per cent patients are indicated for more moderated treatment. The exact treatment strategies are described in detail in the Executive Report of the NCEP ATP III available at http://www.nhlbi.nih.gov/guidelines/cholesterol/index.htm. The score sheets (or points systems) that enable clinicians to determine the 10-year risk of coronary disease are contained in the Executive Report and are displayed in Table I. Separate score sheets were produced for men and women. To use the score sheets a clinician simply totals the points associated with the patient's age, their cholesterol and smoking status, which depend on age, HDL and systolic blood pressure either treated or untreated. A point total is computed by summing all the points for the risk factor profile. The 10-year coronary risk is then determined based on the point total as indicated in the table at the bottom of the score sheet.

These score sheets are based on sex-specific Cox regression models. Without the score sheets it is possible to generate an estimate of 10-year risk using the Cox models directly.

Table I. NCEP ATPIII score sheets.

Men					
Age					Points
30–34					−9
35–39					−4
40–44					0
45–49					3
50–54					6
55–59					8
60–64					10
65–69					11
70–74					12
75–79					13
Total cholesterol	Points at age 30–39	Points at age 40–49	Points at age 50–59	Points at age 60–69	Points at age 70–79
<160	0	0	0	0	0
160–199	4	3	2	1	0
200–239	7	5	3	1	0
240–279	9	6	4	2	1
280+	11	8	5	3	1
Smoking status					
Non-smoker	0	0	0	0	0
Smoker	8	5	3	1	1
HDL					Points
60+					−1
50–59					0
40–49					1
<40					2
Systolic blood pressure			If untreated		If treated
<120			0		0
120–129			0		1
130–139			1		2
140–159			1		2
160+			2		3

Point total	10-year risk (per cent)	Point total	10-year risk (per cent)
<0	<1	5	2
0	1	6	2
1	1	7	3
2	1	8	4
3	1	9	5
4	1	10	6

Table I. Continued.

Point total	10-year risk (per cent)	Point total	10-year risk (per cent)
11	8	17 or more	⩾30
12	10		
13	12		
14	16		
15	20		
16	25		

Women	
Age	Points
30–34	−7
35–39	−3
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	12
70–74	14
75–79	16

Total cholesterol	Points at age 30–39	Points at age 40–49	Points at age 50–59	Points at age 60–69	Points at age 70–79
<160	0	0	0	0	0
160–199	4	3	2	1	1
200–239	8	6	4	2	1
240–279	11	8	5	3	2
280+	13	10	7	4	2
Smoking status					
Non-smoker	0	0	0	0	0
Smoker	9	7	4	2	1

HDL	Points
60+	−1
50–59	0
40–49	1
<40	2

Systolic blood pressure	If untreated	If treated
<120	0	0
120–129	1	3
130–139	2	4
140–159	3	5
160+	4	6

Table I. Continued

Point total	10-year risk (per cent)	Point total	10-year risk (per cent)
<9	<1	20	11
9	1	21	14
10	1	22	17
11	1	23	22
12	1	24	27
13	2	25 or more	$\geqslant 30$
14	2		
15	3		
16	4		
17	5		
18	6		
19	8		

The model form is as follows:

$$\text{Risk estimate} = 1 - S_0(t)^{\exp(\Sigma\beta X - \Sigma\beta\bar{X})} \quad (1)$$

where $S_0(t)$ is the average survival at time t (e.g. $t = 10$ years) or the survival rate at the mean values of the risk factors, β's are the Cox regression coefficients, X's are the individual's values on the risk factors and $\bar{X}$'s are the means of the risk factors.

To generate the risk for an individual, it is necessary to compute $\Sigma\beta X$ by multiplying the coefficients associated with each risk factor (β) by the specific values, X, that comprise the risk profile (e.g. age, systolic blood pressure, total cholesterol, smoking status and so on). The other terms, $\Sigma\beta\bar{X}$ and $S_0(10)$, are not subject specific. The computational difficulty in using the function is in computing the portion of the equation that captures the individual's risk profile, specifically $\Sigma\beta X$. The points system is a means to easily generate this term for each individual risk profile. The logic of the system is outlined in the next section.

3. LOGIC OF THE POINTS SYSTEM

The Cox model (1) can be used directly to estimate the 10-year risk for a given risk profile. The computations using the model can be tedious. It is particularly tedious to compute the required $\Sigma\beta X$ for a given risk factor profile. The points system is a system that simplifies the computation of $\Sigma\beta X$. This is achieved by assigning integer points to each level of each risk factor so that a clinician can easily approximate $\Sigma\beta X$ for a specific risk factor profile by summing integer points. The risk estimate is then determined from a reference table which provides risk estimates for each point total. While the function itself can accommodate distinct values for the continuous risk factors (e.g. age, total cholesterol, HDL), the points system is organized around categories. The categories are designed to mirror clinically meaningful risk factor states. For example, we use the Joint National Committee's (JNC VI) [14] blood pressure categories to organize the blood pressure points. To aid in the interpretation of risk estimates, we also provide tables of comparative risks. Specifically, we display the risks for persons of the same age and sex with the lowest (or optimal) levels of each risk factor and low levels of each risk factor (which might be more achievable). These comparative risks

Table II. NCEP ATPIII comparative risks for men and women.

Men	Lowest risk	Low risk
Age group	Total cholesterol <160, HDL 60+, untreated systolic blood pressure, <120 non-smoker (per cent)	Total cholesterol 160–199, HDL 50–59, untreated systolic blood pressure, 120–129 non-smoker (per cent)
30–34	0	0
35–39	0	1
40–44	0	1
45–49	1	2
50–54	2	4
55–59	3	6
60–64	5	8
65–69	7	10
70–74	9	13
75–79	12	16
Women		
30–34	0	0
35–39	0	0
40–44	0	0
45–49	0	0
50–54	0	1
55–59	0	1
60–64	1	2
65–69	1	2
70–74	2	4
75–79	3	5

can be used for motivating patients to change risk factors so as to reduce their coronary heart disease risk. The tables of comparative risks for men and women that accompanied the NCEP ATP III score sheets (in Table I) are shown in Table II.

In the next section, we illustrate the procedure for setting up a points system using a multiple logistic regression model. We then present the general algorithm for setting up a points system, followed by an example using a Cox regression model.

4. EXAMPLE 1: MULTIPLE LOGISTIC REGRESSION MODEL

We now illustrate the development of a points system for a multiple logistic regression model in which we consider as risk factors age, sex, systolic blood pressure and current smoking status. The outcome of interest is the development of CHD during a 5-year follow-up period.

In this example, we consider a situation in which we have a sample of $n = 9443$ individuals from the Framingham Heart Study between the ages of 30 and 79 who are free of cardiovascular disease. We follow each individual for 5 years for the development of 'hard' coronary heart disease which includes myocardial infarction or coronary death. Some investigations of coronary heart disease also include angina in the definition of the outcome. We wish to avoid any subjectivity and therefore restrict our attention to the more objective outcome. With

Table III. Summary statistics on risk factors and outcome: Example 5.1.

Risk factor	Total ($n = 9443$)	Men ($n = 5182$)	Women ($n = 4261$)
Sex: per cent male	55 per cent	—	—
Age, years	51.4 (12.7)	50.6 (12.4)	52.1 (12.9)
Systolic blood pressure, mm Hg	128.9 (19.5)	130.4 (17.8)	127.8 (20.6)
Per cent current smoker	33 per cent	34 per cent	32 per cent
Number of events (per cent) over 5 Years	237 (2.5 per cent)	171 (3.3 per cent)	66 (1.5 per cent)

complete follow-up, the outcome of interest is dichotomous. Suppose we consider as risk factors age, sex, systolic blood pressure and current smoking status (yes or no). We develop a multiple logistic regression model relating the four risk factors to the development of coronary heart disease. As part of model checking, we examined interactions between the risk factors and none were clinically or statistically significant, therefore a single model was estimated using the combined sample. Summary statistics on the risk factors are shown in Table III for the combined or total sample and then for men and women, considered separately. Also included are details on the outcome of interest. We identify each step involved in developing a points system for this model in the following sections.

4.1. Estimate the parameters of the multivariable model

The estimates of the regression coefficients of the multiple logistic regression model are shown below. Odds ratios (ORs) and 95 per cent confidence intervals (CIs) are also included to enhance interpretability.

Risk factor	Regression coefficient	P	OR	95 per cent CI for OR
Intercept	−10.5161	0.0001	—	
Age	0.0575	0.0001	1.059	1.046–1.073
Male sex	1.3078	0.0001	3.698	2.579–4.956
Systolic blood pressure	0.0185	0.0001	1.109	1.012–1.025
Current smoker	0.9456	0.0001	2.574	1.958–3.384

4.2. Organize the risk factors into categories and determine reference values for each

We now organize the risk factors into meaningful categories and determine a reference value (e.g. mid-point) for each category.

Risk factor	Categories	Reference value (W_{ij})
Age*	30–39	34.5
	40–49	44.5
	50–59	54.5
	60–69	64.5
	70–79	74.5

Continued

Risk factor	Categories	Reference value (W_{ij})
Sex	Female	0
	Male	1
Systolic blood pressure**	<120	107
	120–129	125
	130–139	135
	140–159	150
	$\geqslant 160$	175
Current smoker	No	0
	Yes	1

*The age range in the sample is 30–79.
**The range of systolic blood pressures is 78–240. To determine the reference values for the first and last categories, we use the 1st percentile (94) and the 99th percentile (190) to minimize the influence of extreme values.

4.3. *Determine the referent risk factor profile* ($W_{i\mathrm{REF}}, i = 1, \ldots, 4$)

We now select a referent risk factor profile by choosing a base category for each risk factor. The base category is the category assigned 0 points in the scoring system. Less healthy risk factor states are assigned positive points so that a higher point total conveys more risk. We consider a 30–39 year old, non-smoking female with systolic blood pressure between 120 and 129 as the referent risk factor profile. The base category for each risk factor is shown in boldface type below.

Risk factor	Categories	Reference value (W_{ij})
Age*	**30–39**	$\mathbf{34.5 = W_{1REF}}$
	40–49	44.5
	50–59	54.5
	60–69	64.5
	70–79	74.5
Sex	**Female**	$\mathbf{0 = W_{2REF}}$
	Male	1
Systolic blood pressure**	<120	107
	120–129	$\mathbf{125 = W_{3REF}}$
	130–139	135
	140–159	150
	$\geqslant 160$	170
Current smoker	**No**	$\mathbf{0 = W_{4REF}}$
	Yes	1

4.4. *Determine how far each category is from the base category in regression units*

We now compute how far each category of each risk factor is from the base category in terms of regression units, i.e. for each risk factor we compute $\beta_i(W_{ij} - W_{i\mathrm{REF}})$.

Risk factor	Categories	Reference value (W_{ij})	β_i	$\beta_i(W_{ij} - W_{i\mathrm{REF}})$
Age			0.0575	
	30–39	$34.5 = W_{1\mathrm{REF}}$		0
	40–49	44.5		0.5750
	50–59	54.5		1.1500
	60–69	64.5		1.7250
	70–79	74.5		2.3000
Sex			1.3078	
	Female	$0 = W_{2\mathrm{REF}}$		0
	Male	1		1.3078
Systolic blood pressure			0.0185	
	<120	107		−0.3330
	120–129	$125 = W_{3\mathrm{REF}}$		0
	130–139	135		0.1850
	140–159	150		0.4625
	$\geqslant 160$	170		0.8325
Current smoker			0.9456	
	No	$0 = W_{4\mathrm{REF}}$		0
	Yes	1		0.9456

4.5. *Set the constant, B*

We now define the constant for the points system, or the number of regression units that will correspond to one point. Here, we let B reflect the increase in risk associated with a 5-year increase in age:

$$B = 5(0.0575) = 0.2875$$

4.6. *Determine points associated with each of the categories of the risk factors*

Points associated with each category of each risk factor are computed by: $\mathrm{Points}_{ij} = \beta_i(W_{ij} - W_{i\mathrm{REF}})/B$. The points are rounded to the nearest integer.

Risk factor	Categories	Reference value (W_{ij})	β_i	$\beta_i(W_{ij} - W_{i\mathrm{REF}})$	$\mathrm{Points}_{ij} = \beta_i(W_{ij} - W_{i\mathrm{REF}})/B$
Age			0.0575		
	30–39	$34.5 = W_{1\mathrm{REF}}$		0	0
	40–49	44.5		0.5750	2

Continued

Risk factor	Categories	Reference value (W_{ij})	β_i	$\beta_i(W_{ij} - W_{i\mathrm{REF}})$	Points$_{ij} = \beta_1(W_{ij} - W_{i\mathrm{REF}})/B$
	50–59	54.5		1.1500	4
	60–69	64.5		1.7250	6
	70–79	74.5		2.3000	8
Sex			1.3078		
	Female	$0 = W_{2\mathrm{REF}}$		0	0
	Male	1		1.3078	5
Systolic blood pressure			0.0185		
	<120	107		−0.3330	−1
	120–129	$125 = W_{3\mathrm{REF}}$		0	0
	130–139	135		0.1850	1
	140–159	150		0.4625	2
	$\geqslant 160$	170		0.8325	3
Current smoker			0.9456		
	No	$0 = W_{4\mathrm{REF}}$		0	0
	Yes	1		0.9456	3

4.7. *Determine risks associated with point totals*

We now determine the risks that are associated with each point total. The first step is to determine the theoretical range of the point totals based on the point system computed in Section 4.6. In this system, the theoretical range of point totals is −1 to 19. We now need to attach a risk estimate to each point total using the multiple logistic regression equation:

$$\hat{p} = \frac{1}{1 + \exp\left(-\sum_{i=0}^{p} \beta_i X_i\right)}$$

The point total, when multiplied by the constant ($B = 0.2875$) approximates $\sum_{i=1}^{p} \beta_i X_i$. The model calls for $\sum_{i=0}^{p} \beta_i X_i$ so we need to add the estimate of the intercept $\beta_0 = -10.5161$ as well as 'add back' the values we considered the base values for the continuous risk factors, age and systolic blood pressure. For this model, we approximate $\sum_{i=0}^{p} \beta_i X_i$ as follows:

$$\sum_{i=0}^{p} \beta_i X_i \approx -10.5161 + 0.0575(34.5) + 0.0185(125) + B(\text{Point total})$$

Substituting the above into the logistic model, we generate the following table:

Point total	Estimate of risk	Point total	Estimate of risk
−1	0.0015	10	0.0341
0	0.0020	11	0.0449
1	0.0026	12	0.0590
2	0.0035	13	0.0771
3	0.0047	14	0.1002
4	0.0062	15	0.1293
5	0.0083	16	0.1652
6	0.0110	17	0.2088
7	0.0147	18	0.2602
8	0.0195	19	0.3192
9	0.0258		

We would package this points system as follows:

Risk factor	Categories	Points
Age		
	30–39	0
	40–49	2
	50–59	4
	60–69	6
	70–79	8
Sex		
	Female	0
	Male	5
Systolic blood pressure		
	<120	−1
	120–129	0
	130–139	1
	140–159	2
	$\geqslant 160$	3
Current smoker		
	No	0
	Yes	3

Point total	Estimate of risk
−1	0.0015
0	0.0020
1	0.0026
2	0.0035
3	0.0047

Continued

Point total	Estimate of risk
4	0.0062
5	0.0083
6	0.0110
7	0.0147
8	0.0195
9	0.0258
10	0.0341
11	0.0449
12	0.0590
13	0.0771
14	0.1002
15	0.1293
16	0.1652
17	0.2088
18	0.2602
19	0.3192

The following examples illustrate the correspondence between risks estimated by the logistic model and those approximated by the points system.

Case 1: A 55-year-old, non-smoking male with a systolic blood pressure of 135.

Risk factor	Value	Points
Age	55	4
Sex	Male	5
Systolic blood pressure	135	1
Current smoker	No	0
	Point total	10
	Estimate of risk	0.0338

The risk estimate based on the logistic model is

$$\sum_{i=0}^{p} \beta_i X_i = -10.5161 + 0.0575(55) + 1.3078(1) + 0.0185(135) + 0.9456(0) = -3.5483$$

$$\hat{p} = \frac{1}{1 + \exp(3.5483)} = 0.0280$$

The points system gives a 5-year risk estimate of 3 per cent, employing the model directly also gives 3 per cent.

Case 2: A 72-year-old, smoking female with a systolic blood pressure of 150.

Risk factor	Value	Points
Age	72	8
Sex	Female	0
Systolic blood pressure	150	2
Current smoker	Yes	3
	Point total	13
	Estimate of risk	0.0764

The risk estimate based on the logistic model is

$$\sum_{i=0}^{p} \beta_i X_i = -10.5161 + 0.0575(72) + 1.3078(0) + 0.0185(150) + 0.9456(1) = -2.656$$

$$\hat{p} = \frac{1}{1 + \exp(2.656)} = 0.0657$$

The points system gives a 5-year risk estimate of 8 per cent, employing the model directly gives 7 per cent.

Case 3: A 75-year-old, smoking male with a systolic blood pressure of 160.

Risk factor	Value	Points
Age	75	8
Sex	Male	5
Systolic blood pressure	160	3
Current smoker	Yes	3
	Point total	19
	Estimate of risk	0.3172

The risk estimate based on the logistic model is

$$\sum_{i=0}^{p} \beta_i X_i = -10.5161 + 0.0575(75) + 1.3078(1) + 0.0185(160) + 0.9456(1) = -0.9902$$

$$\hat{p} = \frac{1}{1 + \exp(0.9920)} = 0.2709$$

The points system gives a 5-year risk estimate of 32 per cent, employing the model directly gives 27 per cent. While the estimates produced by the points system and those produced by the model are more disparate for this case, both estimates are sufficiently large to warrant

action toward modification of risk factors. In addition, if a confidence interval estimate was produced for this case, it would likely be fairly wide.

5. ALGORITHM FOR POINTS SYSTEM

We now describe the general approach for developing a points system. Some of the steps are specific to the type of model used to estimate the multivariable function (e.g. multiple linear or multiple logistic regression, Cox proportional hazards regression). In the following, we indicate which steps are specific to the type of model used to estimate the function and which are not.

5.1. Estimate the parameters of the multivariable model

Consider the model $f(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ where Y is the dependent or outcome variable (e.g. $Y = 1$ indicates the presence of a particular event and $Y = 0$ indicates the absence of the event, or Y is a continuous outcome score), $f(Y)$ is a function of Y that can be represented as a combination of the risk factors X_i, and $X_1, \ldots, X_p$ are the candidate risk factors (X_i $i = 1, \ldots, p$, can be continuous or indicator/dummy variables reflecting dichotomous risk factors or categories of risk factors), and $\beta_0, \beta_1, \ldots, \beta_p$ are the estimates of the regression coefficients based on the appropriate regression model (e.g. multiple linear regression analysis, multiple logistic regression analysis, Cox proportional hazards regression).

5.2. Organize the risk factors into categories and determine reference values

If a risk factor is *continuous*—set up contiguous classes and determine reference values for each. For example if $X_1 =$ age in years and the range is 30–79, we use the age categories 30–39, 40–49, 50–59, 60–69, 70–79. In order to determine points for each category, we also need to specify a reference value for each category. The mid-points are generally acceptable reference values, for example using the age categories indicated we use 34.5, 44.5, 54.5, 64.5 and 74.5, respectively. There are exceptions, for example suppose $X_2 =$ systolic blood pressure (SBP) in mmHg and the range is 80–210. We might consider the following five categories <120, 120–129, 130–139, 140–159, $\geqslant 160$. The mid-points for the three middle risk factor categories are 125, 135 and 150, respectively. The reference value for the last category should reflect the mid-point of SBPs 160 or more. Since there may be some extreme values in the blood pressure distribution (e.g., the maximum is 210), the mid-point between 160 and the 99th percentile of the observed systolic blood pressures is a reasonable mid-point for the last category. Suppose the 99th percentile is 190, the mid-point for the last category is 175. A similar strategy can be used to determine the reference value for the first systolic blood pressure category (SBP <120). Suppose the 1st percentile in the sample is 89. The mid-point between the 1st percentile and 119 is 104 and considered to be the referent value for the first category.

If a risk factor is modelled by a *set of dummy variables* (each coded as $0 =$ absent or $1 =$ present) reflecting distinct categories of the risk factor, then the reference value is simply 0 or 1.

If a risk factor is dichotomous and modelled as an *indicator* variable (e.g. 0 = absent, 1 = present), the reference value is again either 0 or 1 and nothing more needs to be performed at this step.

Let W_{ij} denote the reference value (e.g. mid-points for continuous risk factors organized into categories, or values 0 or 1 for risk factors modelled by a set of dummy variables or a single indicator) for the jth category of the ith risk factor, where $i = 1, \ldots, p$, and $j = 1, \ldots, c_i$, where c_i = the total number of categories for risk factor i.

5.3. Determine the referent risk factor profile

Next we determine the appropriate category for each risk factor to serve as the base category. The base category for each risk factor is the category assigned 0 points in the scoring system. Categories reflecting worse (less healthy) states of the risk factor will be assigned positive points, while categories reflecting healthier states will be assigned negative points.

The reference value of the base category is denoted $W_{i\mathrm{REF}}$, for each of the i risk factors, $i = 1, \ldots, p$.

5.4. Determine how far each category is from the base category in regression units

For each risk factor, we next determine how far each category is from the base category $W_{i\mathrm{REF}}$, in terms of regression units. Specifically, we determine the following for each category j of each risk factor i: $\beta_i(W_{ij} - W_{i\mathrm{REF}})$, $i = 1, \ldots, p$ and $j = 1, \ldots, c_i$.

5.5. Set the fixed multiplier or constant B

The constant, B, is the number of regression units that reflect 1 point in the final points system.

In the Framingham points systems, we often base the constant on age which has been shown to be important and significant in an increasing fashion in most risk score functions developed in the Framingham Study. Framingham investigators often set up the constant (or one point) to be equivalent to the increase in risk associated with a 5-year increase in age. For example, suppose X_1 = age in years and $\beta_1 = 0.05$. We set the constant $B = 0.05 (5) = 0.25$.

5.6. Determine the number of points for each of the categories of each risk factor

The points for each category of each risk factor are determined by the following:

$$\mathrm{Points}_{ij} = \beta_i(W_{ij} - W_{i\mathrm{REF}})/B$$

Note that the base category of each risk factor is assigned 0 points using this formula.

5.7. Determine risks associated with point totals

For each risk factor profile a point total is computed. The final step in setting up the point system is the determination of the estimates of risk (or probability of developing an event over the predetermined time frame) associated with each point total. This step is specific to the type of multivariable model employed in Section 5.1 as it requires the use of the exact

model to determine the estimates of risk. The following are typical models used for risk estimation along with their specific formulations:

Model	Risk estimate
Multiple logistic regression	$\hat{p} = \dfrac{1}{1 + \exp\left(-\sum_{i=0}^{p} \beta_i X_i\right)}$
Cox proportional hazards regression	$\hat{p} = 1 - S_0(t)^{\exp\left(\sum_{i=1}^{p} \beta_i X_i - \sum_{i=1}^{p} \beta_i \bar{X}_i\right)}$
Weibull regression	$\hat{p} = 1 - \exp\left(-\exp\left(\dfrac{\ln(t) - \sum_{i=1}^{p} \beta_i X_i}{\sigma}\right)\right)$

The basic idea of the points system is to approximate the contribution of the risk factors in the estimate of risk, specifically to estimate $\sum_{i=1}^{p} \beta_i X_i$ which is the component of each model shown above that depends on the specific risk factor profile under consideration.

The risk estimates in the points system associated with specific risk factor profiles are computed by substituting the product of the total number of points and the constant, B, which approximates $\sum_{i=1}^{p} \beta_i X_i$, into the appropriate formula (e.g. logistic regression equation) to estimate the risk. There are several issues that require special attention at this stage. These include the presence of an intercept term and handling of continuous risk factors. Each issue is discussed separately below.

Intercept term: It is important to note that the points system does not include a separate point allocation for an intercept. For example, with the logistic model the product of the point total and the constant, B, approximates $\sum_{i=1}^{p} \beta_i X_i$. In order to estimate $\sum_{i=0}^{p} \beta_i X_i$, we need to include the estimate of the intercept, β_0, to estimate $\sum_{i=0}^{p} \beta_i X_i$ which is then used to compute the estimate of risk.

Continuous risk factors: In the points system, we set up categories for the continuous risk factors and assigned a reference value to each (Section 5.2). We then determined a base risk factor category and assigned 0 points to that category. When we total all the points we are essentially adding up how far a particular individual's risk factor profile is from the referent profile. The $\sum_{i=1}^{p} \beta_i X_i$ term in the multivariable model reflects a particular risk profile and not the distance from the referent risk factor profile. Therefore, we need to include both the referent risk factor profile and the distance from that profile in order to produce the appropriate $\sum_{i=1}^{p} \beta_i X_i$ for the risk estimate. For example, suppose $X_1 =$ age in years and we consider age 40 as the referent age. If we wish to determine the $\sum_{i=1}^{p} \beta_i X_i$ for an individual who is age 50 we can think of that person as 10 years older than the referent age (as we do when we set up the points system). To determine the age component of $\sum_{i=1}^{p} \beta_i X_i$ we need to sum $\beta_1(40)$ and $\beta_1(50-40)$. The latter is approximated in the points system, and the former needs to be 'added back' when we compute the corresponding risk estimates.

We illustrate the implementation of this algorithm in the next section using a Cox proportional hazards regression model.

6. EXAMPLE 2: COX PROPORTIONAL HAZARDS MODEL

In this example, we consider an application in which we have a sample of males between the ages of 30 and 74 who are free of cardiovascular disease. We follow each individual for 10 years for the development of hard coronary heart disease (i.e. myocardial infarction or coronary death). With follow-up extending to 10 years, we use survival analysis methods to take into account incomplete follow-up information. The risk factors include age in years, categories of blood pressure, total cholesterol and HDL cholesterol, current smoking status (yes or no) and diabetes status (yes or no). Sex-specific models were developed for the purposes of a validation study, which is described in detail in D'Agostino *et al.* [12]. To illustrate the method, we use the model developed for men. We again use the same numbering system as that presented in the previous section to identify the distinct steps in setting up the points system.

6.1. Estimate the parameters of the multivariable model

Table IV below was reported in D'Agostino *et al.* [12] and represents the coefficients of the Cox proportional hazards model for men along with the means (or proportions of men in each risk-factor category) of the risk factors and the average 10-year survival which are needed for the computations.

Table IV. Cox proportional hazards regression coefficients and means of the risk factors [12]: Example 5.2.

Risk factor	Regression coefficient	Mean or proportion
Age, years	0.0533	48.3
Optimal BP (DBP† <80 and SBP‡ <120)	0.0948	0.20
Normal BP (DBP 80–84 or SBP 120–129)	Base	0.24
Hi Normal BP (DBP 85–89 or SBP 130–139)	0.4225	0.20
Stage I Hypertension (DBP 90–99 or SBP 140–159)	0.6596	0.23
Stage II+ Hypertension (DBP $\geqslant 100$ or SBP $\geqslant 160$)	0.8964	0.13
Total cholesterol <160	-0.3781	0.07
Total cholesterol 160–199	Base	0.31
Total cholesterol 200–239	0.5696	0.39
Total cholesterol 240–279	0.7438	0.17
Total cholesterol $\geqslant 280$	0.8284	0.06
HDL <35	0.6074	0.19
HDL 35–44	0.3684	0.36
HDL 45–49	Base	0.15
HDL 50–59	0.0000	0.19
HDL $\geqslant 60$	-0.4608	0.11
Current smoker	0.7277	0.40
Diabetes	0.5252	0.05
Average 10-year survival $= 0.943$*		

*Kaplan–Meier estimate of the survival rate at the mean values of the risk factors.
†DBP = Diastolic blood pressure.
‡SBP = Systolic blood pressure.

6.2. Organize the risk factors into categories and determine reference values for each

We now organize the only continuous risk factor, age, into categories and determine a reference value (in this case the mid-point) for each category. The remaining risk factors are modelled by sets of dummy variables or indicator variables.

Risk factor	Categories	Reference value (W_{ij})
Age*	30–34	32
	35–39	37
	40–44	42
	45–49	47
	50–54	52
	55–59	57
	60–64	62
	65–69	67
	70–74	72
Optimal BP (DBP <80 and SBP <120)	No	0
	Yes	1
High normal BP (DBP 85–89 or SBP 130–139)	No	0
	Yes	1
Stage I hypertension (DBP 90–99 or SBP 140–159)	No	0
	Yes	1
Stage II+ hypertension (DBP $\geqslant 100$ or SBP $\geqslant 160$)	No	0
	Yes	1
Total cholesterol <160	No	0
	Yes	1
Total cholesterol 200–239	No	0
	Yes	1
Total cholesterol 240–279	No	0
	Yes	1
Total cholesterol $\geqslant 280$	No	0
	Yes	1
HDL <35	No	0
	Yes	1
HDL 35–44	No	0
	Yes	1
HDL 50–59	No	0
	Yes	1
HDL $\geqslant 60$	No	0
	Yes	1
Current smoker	No	0
	Yes	1
Diabetes	No	0
	Yes	1

*The age range in the sample is 30–74.

6.3. Determine the referent risk factor profile ($W_{i\mathrm{REF}},\ i=1,\ldots,6$)

We now select a referent risk factor profile by choosing a base category for each risk factor. We consider a non-smoking, non-diabetic 42-year-old male with normal blood pressure

(DBP 80–84 or SBP120–129), total cholesterol between 160 and 199 and HDL between 45 and 49 as the referent risk factor profile. The base categories are indicated in boldface type below.

Risk factor	Categories	Reference value (W_{ij})
Age	30–34	32
	35–39	37
	40–44	**42 = W_{1REF}**
	45–49	47
	50–54	52
	55–59	57
	60–64	62
	65–69	67
	70–74	72
Optimal BP (DBP <80 and SBP <120)	**No**	**0 = W_{2REF}**
	Yes	1
High normal BP (DBP 85–89 or SBP 130–139)	**No**	**0 = W_{2REF}**
	Yes	1
Stage I hypertension (DBP 90–99 or SBP 140–159)	**No**	**0 = W_{2REF}**
	Yes	1
Stage II+ hypertension (DBP $\geqslant 100$ or SBP $\geqslant 160$)	**No**	**0 = W_{2REF}**
	Yes	1
Total cholesterol <160	**No**	**0 = W_{3REF}**
	Yes	1
Total cholesterol 200–239	**No**	**0 = W_{3REF}**
	Yes	1
Total cholesterol 240–279	**No**	**0 = W_{3REF}**
	Yes	1
Total cholesterol $\geqslant 280$	**No**	**0 = W_{3REF}**
	Yes	1
HDL <35	**No**	**0 = W_{4REF}**
	Yes	1
HDL 35–44	**No**	**0 = W_{4REF}**
	Yes	1
HDL 50–59	**No**	**0 = W_{4REF}**
	Yes	1
HDL $\geqslant 60$	**No**	**0 = W_{4REF}**
	Yes	1
Current smoker	**No**	**0 = W_{5REF}**
	Yes	1
Diabetes	**No**	**0 = W_{6REF}**
	Yes	1

Note: Blood pressure, total cholesterol and HDL are modelled as categories. The base category is indicated when the remaining categories are coded 0 (see Section 6.1).

6.4. Determine how far each category is from the base category in regression units

We now compute how far each category of each risk factor is from the base category in terms of regression units using $\beta_i(W_{ij} - W_{i\mathrm{REF}})$.

Risk factor	Categories	Reference value (W_{ij})	β_i	$\beta_i(W_{ij} - W_{i\mathrm{REF}})$
Age			0.0533	
	30–34	32		−0.5330
	35–39	37		−0.2665
	40–44	$42 = W_{1\mathrm{REF}}$		0
	45–49	47		0.2665
	50–54	52		0.5330
	55–59	57		0.7995
	60–64	62		1.0660
	65–69	67		1.3325
	70–74	72		1.5990
[-6.6pt] Optimal BP			0.0948	
	No	$0 = W_{2\mathrm{REF}}$		0
	Yes	1		0.0948
High normal BP			0.4225	
	No	$0 = W_{2\mathrm{REF}}$		0
	Yes	1		0.4225
Stage I hypertension			0.6596	
	No	$0 = W_{2\mathrm{REF}}$		0
	Yes	1		0.6596
Stage II+ hypertension			0.8964	
	No	$0 = W_{2\mathrm{REF}}$		0
	Yes	1		0.8964
[-6.6pt] Total cholesterol <160			−0.3781	
	No	$0 = W_{3\mathrm{REF}}$		0
	Yes	1		−0.3781
Total cholesterol 200–239			0.5696	
	No	$0 = W_{3\mathrm{REF}}$		0
	Yes	1		0.5696
Total cholesterol 240–279			0.7438	
	No	$0 = W_{3\mathrm{REF}}$		0
	Yes	1		0.7438
Total cholesterol $\geqslant 280$			0.8284	
	No	$0 = W_{3\mathrm{REF}}$		0
	Yes	1		0.8284
HDL <35			0.6074	
	No	$0 = W_{4\mathrm{REF}}$		0
	Yes	1		0.6074
HDL 35–44			0.3684	
	No	$0 = W_{4\mathrm{REF}}$		0
	Yes	1		0.3684
HDL 50–54			0	
	No	$0 = W_{4\mathrm{REF}}$		0
	Yes	1		0
HDL $\geqslant 60$			−0.4608	
	No	$0 = W_{4\mathrm{REF}}$		0
	Yes	1		−0.4608
Current smoker			0.7277	
	No	$0 = W_{5\mathrm{REF}}$		0
	Yes	1		0.7277
Diabetes			0.5252	
	No	$0 = W_{6\mathrm{REF}}$		0
	Yes	1		0.5252

6.5. Set the constant, B

We now define the constant for the points system, or the number of regression units that will correspond to one point. Here we let B reflect the increase in risk associated with a 5-year increase in age:

$$B = 5(0.0533) = 0.2665$$

6.6. Determine points associated with each of the categories of the risk factors

We now determine the points associated with each category of each risk factor using $\text{Points}_{ij} = \beta_i(W_{ij} - W_{i\text{REF}})/B$.

Risk factor	Categories	Reference value (W_{ij})	β_i	$\beta_i(W_{ij} - W_{i\text{REF}})$	$\text{Points}_{ij} = \beta_{\text{I}}(W_{ij} - W_{i\text{REF}})/B$
Age			0.0533		
	30–34	32		−0.5330	−2
	35–39	37		−0.2665	−1
	40–44	$42 = W_{1\text{REF}}$		0	0
	45–49	47		0.2665	1
	50–54	52		0.5330	2
	55–59	57		0.7995	3
	60–64	62		1.0660	4
	65–69	67		1.3325	5
	70–74	72		1.5990	6
Optimal BP			0.0948		
	No	$0 = W_{2\text{REF}}$		0	0
	Yes	1		0.0948	0
High normal BP			0.4225		
	No	$0 = W_{2\text{REF}}$		0	0
	Yes	1		0.4225	2
Stage I hypertension			0.6596		
	No	$0 = W_{2\text{REF}}$		0	0
	Yes	1		0.6596	2
Stage II+ hypertension			0.8964		
	No	$0 = W_{2\text{REF}}$		0	0
	Yes	1		0.8964	3
Total cholesterol <160			−0.3781		
	No	$0 = W_{3\text{REF}}$		0	0
	Yes	1		−0.3781	−1
Total cholesterol 200–239			0.5696		
	No	$0 = W_{3\text{REF}}$		0	0
	Yes	1		0.5696	2

Continued

Risk factor	Categories	Reference value (W_{ij})	β_i	$\beta_i(W_{ij} - W_{i\mathrm{REF}})$	Points$_{ij}$ = $\beta_{\mathrm{I}}(W_{ij} - W_{i\mathrm{REF}})/B$
Total cholesterol 240–279			0.7438		
	No	$0 = W_{3\mathrm{REF}}$		0	0
	Yes	1		0.7438	3
Total cholesterol $\geqslant 280$			0.8284		
	No	$0 = W_{3\mathrm{REF}}$		0	0
	Yes	1		0.8284	3
HDL <35			0.6074		
	No	$0 = W_{4\mathrm{REF}}$		0	0
	Yes	1		0.6074	2
HDL 35–44			0.3684		
	No	$0 = W_{4\mathrm{REF}}$		0	0
	Yes	1		0.3684	1
HDL 50–54			0		
	No	$0 = W_{4\mathrm{REF}}$		0	0
	Yes	1		0	0
HDL $\geqslant 60$			−0.4608		
	No	$0 = W_{4\mathrm{REF}}$		0	0
	Yes	1		−0.4608	−2
Current smoker			0.7277		
	No	$0 = W_{5\mathrm{REF}}$		0	0
	Yes	1		0.7277	3
Diabetes			0.5252		
	No	$0 = W_{6\mathrm{REF}}$		0	0
	Yes	1		0.5252	2

$B = 0.2665$.

6.7. *Determine risks associated with point totals*

We now determine the risks that are associated with each point total. The first step is to determine the theoretical range of the point totals based on the point system computed in Section 6.6. In this system, the theoretical range of point totals is −4 to 19. We now need to attach a risk estimate to each point total using the Cox Proportional Hazards model: $\hat{p} = 1 - S_0(t)^{\exp\left(\sum_{i=1}^{p} \beta_i X_i - \sum_{i=1}^{p} \beta_i \bar{X}_i\right)}$. The point total, when multiplied by the constant ($B = 0.2665$) approximates $\sum_{i=1}^{p} \beta_i X_i$. We need to add back the age value we considered the base (age 42), so for this model, we approximate $\sum_{i=1}^{p} \beta_i X_i$ as follows:

$$\sum_{i=1}^{p} \beta_i X_i \approx 0.0533(42) + B(\text{Point total})$$

We also need to compute $\sum_{i=1}^{p} \beta_i \bar{X}_i$ which can be done once using the regression coefficients and the means (or proportions) of the risk factors given in Table IV. For this example, $\sum_{i=1}^{p} \beta_i \bar{X}_i = 3.840037$. Substituting these components along with the estimate of the average 10-year survival (also shown in Table IV) into the Cox model, we generate the following table:

Point total	Estimate of risk	Point total	Estimate of risk
−5	0.0032	8	0.0950
−4	0.0041	9	0.1221
−3	0.0053	10	0.1564
−2	0.0069	11	0.1991
−1	0.0090	12	0.2516
0	0.0118	13	0.3150
1	0.0153	14	0.3897
2	0.0200	15	0.4751
3	0.0260	16	0.5689
4	0.0338	17	0.6666
5	0.0439	18	0.7616
6	0.0569	19	0.8462
7	0.0736		

We would package this points system as follows:

Risk factor	Categories	Points
Age		
	30–34	−2
	35–39	−1
	40–44	0
	45–49	1
	50–54	2
	55–59	3
	60–64	4
	65–69	5
	70–74	6
Blood pressure		
	Optimal	0
	Normal	0
	High normal	2
	Stage I hypertension	2
	Stage II+ hypertension	3

Continued

Risk factor	Categories	Points
Total cholesterol		
	<160	-1
	160–199	0
	200–239	2
	240–279	3
	$\geqslant 280$	3
HDL		
	<35	2
	35–44	1
	45–49	0
	50–59	0
	>60	-2
Current smoker		
	No	0
	Yes	3
Diabetes		
	No	0
	Yes	2

Point total	Estimate of risk
-5	0.0032
-4	0.0041
-3	0.0053
-2	0.0069
-1	0.0090
0	0.0118
1	0.0153
2	0.0200
3	0.0260
4	0.0338
5	0.0439
6	0.0569
7	0.0736
8	0.0950
9	0.1221
10	0.1564
11	0.1991
12 or more*	>0.20

*Since there are so few individuals in the upper ranges of the distribution, we cut off the risk table so as not to overstate the precision in the risk estimates. This issue is addressed further in Section 7.

The following examples illustrate the correspondence between the risks estimated by the Cox model and those approximated by the points system.

Case 1: A 54-year-old male with normal blood pressure, total cholesterol between 200 and 239, HDL between 50 and 59, a current smoker, but not diabetic.

Risk factor	Value	Points
Age	54	2
Blood pressure	SBP = 125, DBP = 82	0
Total cholesterol	210	2
HDL	55	0
Current smoker	Yes	3
Diabetes	No	0
	Point total	7
	Estimate of risk	0.0736

The risk estimate based on the Cox model is computed as follows:

$$\sum_{i=1}^{p} \beta_i X_i = 0.0533(54) + 0.0948(0) + 0.4225(0) + 0.6595(0) + 0.8964(0) - 0.3781(0)$$
$$+ 0.5696(1) + 0.7438(0) + 0.8284(0) + 0.6074(0) + 0.3684(0) + 0.0000(1)$$
$$- 0.4608(0) + 0.7277(1) + 0.5252(0) = 4.1755$$

$$\sum_{i=1}^{p} \beta_i \bar{X}_i = 0.0533(48.3) + 0.0948(0.20) + 0.4225(0.20) + 0.6595(0.23) + 0.8964(0.13)$$
$$- 0.3781(0.07) + 0.5696(0.39) + 0.7438(0.17) + 0.8284(0.06) + 0.6074(0.19)$$
$$+ 0.3684(0.36) + 0.0000(0.19) - 0.4608(0.11) + 0.7277(0.40) + 0.5252(0.05)$$
$$= 3.83258$$

$$\hat{p} = 1 - S_0(t)^{\exp\left(\sum_{i=1}^{p} \beta_i X_i - \sum_{i=1}^{p} \beta_i \bar{X}_i\right)} = 1 - 0.94298^{\exp(4.1755 - 3.83258)}$$
$$= 0.0788$$

The points system gives a 10-year estimate of risk of 7 per cent, employing the Cox model directly gives 8 per cent.

Case 2: A 62-year-old male with high normal blood pressure, total cholesterol between 200 and 239, HDL between 45 and 49, a non-smoker with diabetes.

Risk factor	Value	Points
Age	62	4
Blood pressure	SBP = 130, DBP = 85	2
Total cholesterol	215	2
HDL	48	0
Current smoker	No	0
Diabetes	Yes	2
	Point total	10
	Estimate of risk	0.1564

The risk estimate based on Cox model is computed as follows:

$$\sum_{i=1}^{p} \beta_i X_i = 0.0533(62) + 0.0948(0) + 0.4225(1) + 0.6596(0) + 0.8964(0) - 0.3781(0)$$

$$+ 0.5696(1) + 0.7438(0) + 0.8284(0) + 0.6074(0) + 0.3684(0) + 0.0000(0)$$

$$- 0.4608(0) + 0.7277(0) + 0.5252(1) = 4.8219$$

$$\sum_{i=1}^{p} \beta_i \bar{X}_i = 3.840037$$

$$\hat{p} = 1 - S_0(t)^{\exp\left(\sum_{i=1}^{p} \beta_i X_i - \sum_{i=1}^{p} \beta_i \bar{X}_i\right)} = 1 - 0.94298^{\exp(4.8219 - 3.840037)}$$

$$= 0.1451$$

The points system gives a 10-year estimate of risk of 16 per cent, employing the Cox model directly gives 15 per cent.

Case 3: A 65-year-old male with stage I hypertension, total cholesterol between 240 and 279, HDL between 35 and 44, a non-smoker and not diabetic.

Risk factor	Value	Points
Age	65	5
Blood pressure	SBP = 150, DBP = 90	2
Total cholesterol	250	3
HDL	40	1
Current smoker	No	0
Diabetes	No	0
	Point total	11
	Estimate of risk	0.1991

The risk estimate based on Cox model is computed as follows:

$$\sum_{i=1}^{p} \beta_i X_i = 0.0533(65) + 0.0948(0) + 0.4225(0) + 0.6596(1) + 0.8964(0) - 0.3781(0) + 0.5696(0) + 0.7438(1) + 0.8284(0) + 0.6074(0) + 0.3684(1) + 0.0000(0) - 0.4608(0) + 0.7277(0) + 0.5252(0) = 5.2363$$

$$\sum_{i=1}^{p} \beta_i \bar{X}_i = 3.840037$$

$$\hat{p} = 1 - S_0(t)^{\exp\left(\sum_{i=1}^{p} \beta_i X_i - \sum_{i=1}^{p} \beta_i \bar{X}_i\right)} = 1 - 0.94298^{\exp(5.2363 - 3.840037)} = 0.2112$$

The points system gives a 10-year estimate of risk of 20 per cent, employing the Cox model directly gives 21 per cent.

7. INTERPRETATION ISSUES

The examples in the previous sections illustrate the correspondence between the risk estimates produced by the points system and those produced by the multivariable models directly. For most risk profiles, there is very good agreement between the estimates produced by the points system and those produced by the models. There are instances, however, when there is divergence and these occur primarily at the extremes (e.g. in an individual with the most unhealthy risk factor levels in every risk factor).

There are limitations with the points systems. In order to achieve the simplicity in utilization we lose some of the information that is only captured by using the function directly. The loss of information is slightly more pronounced when the risk factors are modelled as continuous variables because the points system is based on categories. We have assessed the degree of agreement between the risk estimates based on the points system, and those based on the functions and the intraclass correlations generally exceed 0.90.

For Examples 1 and 2, there is very good agreement between the risk estimates based on the points system and those based on the multiple logistic regression model and Cox proportional hazards models, respectively (See Table V). The weighted kappa for Example 1 is 0.85 (95 per cent CI: 0.83–0.86) and for Example 2 is 0.87 (95 per cent CI: 0.85–0.88). To facilitate presentation, the 5- and 10-year risk estimates are organized into clinically meaningful categories.

It is also important to recognize that both the points system and the function produce point estimates of the absolute risk of event over a specific time period. Confidence interval estimates around the point estimates are important to address precision. Since many of the treatment guidelines are based on absolute levels of risk, it will then become necessary to modify the guidelines to accommodate the confidence interval estimates of risk. The addition of the confidence intervals will make it clear that the risks are less stable at the upper end. We generally truncate the risk tables at values that reflect reasonable limits of the data.

Table V. Agreement between risk estimates based on points system and risk estimates based on multivariable model.

Example 1	Risk estimate based on points system		
Risk estimate based on multiple logistic regression model (per cent)	<5 per cent	5–10 per cent	>10 per cent
<5	7911	230	0
5–10	99	763	73
>10	2	66	299

Weighted $\kappa = 0.85$ (95 per cent CI: 0.83–0.86).

Example 2	Risk estimate based on points system		
Risk estimate based on Cox proportional hazards regression model (per cent)	<10 per cent	10–20 per cent	>20 per cent
<10	1642	10	0
10–20	110	410	569
>20	0	69	193

Weighted $\kappa = 0.87$ (95 per cent CI: 0.85–0.88).

In summary, the points system is an approach for making complex statistical models useful to practitioners. Clinicians are extremely busy in everyday practice, most make intense efforts to stay abreast of the most recent advances in medical research. Incorporating effective evidence-based approaches into everyday clinical practice can be difficult. This system is one means of simplifying the assessment of the multi-factorial nature of coronary disease risk. It represents an effort to make available a tool for clinicians to aid in their decision-making process regarding treatment and to assist them in motivating patients toward healthy behaviours.

REFERENCES

1. Anderson KM, Wilson PWF, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. *Circulation* 1991; **83**:356–362.
2. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; **97**:1837–1847.
3. Truett J, Cornfield J, Kannel WB. A Multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases* 1967; **20**:511–524.
4. Cornfield J, Gordon T, Smith W. Quantal response curves for experimentally uncontrolled variables. *Bulletin of International Statistical Institute* 1961; **38**:97–115.
5. Walker S, Duncan D. Estimation of the probability of an event as a function of several independent variables. *Biometrika* 1967; **54**:167–179.
6. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *American Journal of Cardiology* 1976; **38**:46–51.
7. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke* 1991; **3**:312–318.
8. D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. *Stroke* 1994; **25**:40–43.

9. Murabito JM, D'Agostino RB, Silbershatz H *et al*. Intermittent claudication: a risk profile from the Framingham Study. *Circulation* 1997; **96**: 44–49.
10. Kannel WB, D'Agostino RB, Silbershatz H, Belanger AJ, Wilson PF, Levy D. Profile for estimating risk of heart failure. *Archives of Internal Medicine* 1999; **159**:1197–1204.
11. D'Agostino RB, Russell MW, Huse DM *et al*. Primary and subsequent coronary risk appraisal: new results from the Framingham Study. *American Heart Journal* 2000; **139**:272–281.
12. D'Agostino RB, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *Journal of Anti-aging Medicine* 2001; **296**:180–187.
13. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment in Adults (Adult Treatment Panel III) Executive Summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment in Adults (Adult Treatment Panel III). *Journal of Anti-aging Medicine* 2001; **285**:2486–2497.
14. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. *Archives of International Medicine* 1997; **157**:2413–2446.

