HealthLink Risk Profile Analytic Model

Dr. Joseph Boffa

Background

The HealthLink screening results are in themselves very informative and provide valuable feed back to participants. How do these results compare with national norms as it relates to possible disease outcomes? Recently the Framingham Heart Study (FHS) [1] reported on the development of a statistically derived mathematical model of coronary heart disease risk based on the following risk factors:

- **Age**
- **Gender**
- **Total Cholesterol**
- **High Density Lipoproteins**
- **Blood Pressure**
- **Diabetes Status**
- **Smoking Status**

The model is derived from a population-based sample, which included 2489 men and 2856 women 30 to 74 years old at the time of their Framingham Heart Study examination in 1971 to 1974. Participants attended either the 11th examination of the original Framingham cohort or the initial examination of the Framingham Offspring Study. Similar research protocols were used in each study. Persons with overt CHD at the baseline examination were excluded. The statistical tests employed included age-adjusted linear regression or logistic regression to test for trends across blood pressure, TC, LDL-C, and HDL-C categories. Equations were computed in the Framingham data by fitting Cox proportional hazards regression models [5], the most commonly used procedure for estimating the probability of an event given information on a risk profile for that event. In these equations, each risk factor is linked to a number that acts as does a slope in the equation of a line (the "β" in the equation being the slope or regression coefficient and the "X" being the risk factor, $Y = A + \beta X$). An equation therefore includes a set of slopes that can be multiplied by the risk factors to give the sum, Y, for each participant. Each slope is in effect a weight to be applied to each risk factor. In the Cox Proportional Hazard model the linear function is part of an exponential function of the general form: $\text{Hazard} = H_0 e^{(\beta X)}$. Where H_0 is a baseline hazard, e is the base of the natural logarithm and the exponential component is the linear sum of the risk factors and slopes that was estimated from the FHS data. The linear function in the Cox model has no intercept because the intercept is subsumed into the baseline hazard[3]. As will also be pointed out the baseline hazard is generally left unspecified and will be canceled out when Cox Hazard Ratios are computed to calculate relative risk.

In the Framingham study two gender specific risk models were developed:

Let **Hazard (t)** be the hazard measurement at time t for an individual with Hazard factor vector **X** $(X_1 \ldots .X_n)$ and baseline hazard **Hazard (0):**

> **Males:**
>
> $$\textbf{Hazard (t)} = \textbf{Hazard (0)} \, e^{(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \ldots)}$$
>
> **Females:**
>
> $$\textbf{Hazard (t)} = \textbf{Hazard (0)} \, e^{(\beta_1 X_1 + \beta_1 X^2_1 + \beta_2 X_2 \ldots .)}$$

where formula coefficient weighting vector (slopes) β $(\beta_1 \ldots \ldots .\beta_n)$ are derived from FHS risk assessment data. Each risk model is very similar with the exception for females. The exponential risk formula contains a square term for age. The exponential component for males is a straightforward linear additive function of risk factors while for females it is a quadratic function.

HealthLink collects all the relevant input risk factors required for the FHS CHD risk assessment model using total cholesterol (TC) as the cholesterol input factor. With this model there is now the ability to compare HealthLink screening derived CHD risk assessments. The input variables collected by HealthLink that represent the hazard vectors are:

- X1- Age
- $X1^2$- Age^2 for females only
- X2- Hypertension
- X3- Total Cholesterol (TC)
- X4- High Density Lipoproteins (HDL)
- X5- Fasting Blood Glucose
- X6- Smoking

The input hazard factor covariate vectors for TC, HDL and blood glucose were coded to conform to input factor break down required of the FHS CHD risk model[1]. The Cox model is often called a proportional hazard model because, given two individuals with hazard factor vector values $\mathbf{X_1}$ and $\mathbf{X_2}$ and baseline hazard **H (0)**, the ratio of their hazard rates[2] is:

$$\frac{H(t|X_1)}{H(t|X_2)} = \frac{H(0) \, e^{(\beta X_1)}}{H(0) \, e^{(\beta X_2)}} = e^{\{\beta X_1 - \beta X_2\}}$$

This ratio is constant over time and is independent from time t; hence, the hazard ratio of any two cases is proportional. Their constant hazard ratio measures the relative risk of an individual with risk factor vector $\mathbf{X_1}$ compared to the other individual's risk factor vector $\mathbf{X_2}$. The nice thing about the Cox Hazard Ratio model in the case given above is that the baseline hazard **H(0)** is given no particular parametric criteria and can be left undefined.[3,4]

HealthLink Risk Profile Index Analysis

HealthLink's primary goal is to monitor and enhance changes in retiree risk due to participation in the screening and wellness programs. Each retiree's medical screening and smoking history input factors represent **H(t)** or hazard at time t. The central issue of the project is to measure improvement in hazard as a consequence of Healthlink's interventions. What would **H(t)** be if all medical screenings and smoking history are at recommended levels? Do input factors clinically improve over time such that Cox Hazard Ratios measure this improvement? In comparing retiree participant risks, each retiree has an observed hazard risk **H(t)$_o$** assessment based on his/her observed recorded risk factor vector **Xo** at the time of each screening. For that individual we can also impute an expected or "normal level of risk" **H(t)$_e$** risk factor vector **Xe** in which all screening inputs are normal**.** A relative risk calculation can be expressed as:

$$\textbf{Cox Hazard Ratio} = \frac{\textbf{H(t|X}_{\textbf{observed}}\textbf{)}}{\textbf{H(t|X}_{\textbf{expected}}\textbf{)}} = \textbf{Relative Risk} = e^{\{\,\beta Xo - \beta Xe\,\}}$$

The observed risk factors (**Xo**) are the participant's actual screening results at the time of each screening session. The expected risk factors (**Xe**) are input factors that are imputed normal values as defined by JNC and other recognized criteria <u>for that same individual</u>. Each individual then can be assessed as to the ratio of observed results with an ideal or expected normal screening input. If a particular retiree's observed screening risk factors are all normal and he is also a non smoker then the exponential vectors $\beta \textbf{X}_o$ and $\beta \textbf{X}_e$ would be equal. The relative risk would then equal $\textbf{e}^{(0)}$ or 1. Any observed screening scores (**Xo**) in the elevated range would then lead to an increase in relative risk. For example if another individual is a male with normal medical screenings but is a smoker then the relative risk for that individual would be $\textbf{e}^{(.52337)}$ or 1.7 because the only difference in the two vectors is the FHS β vector weight for a smoker which is .52337. This represents a net increment of risk due to the smoking covariate. The smoking risk factor represents an overall 70% increase in the risk for CHD over the expected or targeted risk. For the purposes of the HealthLink Project, successful outcomes would indicate a measurable improvement of each individual's relative risk over time.

Another feature of using this model is that each individual is his/her own control. Because of this, age for the observed risk vector $\beta \textbf{Xo}$ is identical to age in the expected risk vector $\beta \textbf{Xe}$. Since both vectors contain the same age covariate, age as an input factor does not contribute to incremental changes in relative risk. For Example:

If vector $\beta_1 X_1$ is a weighted covariate for age and $\beta_2 X_2$ is a weighted covariate for smoking where X_2 is 0 for non smokers and 1 for smokers then if an individual male is a smoker the Relative Risk for that individual based on FHS weights would be:

$$RR = e^{\text{(observed vector (smoker) – expected vector (non smoker))}}$$

Or

$$RR = e^{(\beta_1 Xo_1 + \beta_2 Xo_2) - (\beta_1 Xe_1 + \beta_2 Xe_2)}$$

Where:

X_1 is age and $Xo_1 = Xe_1$
X_2 is 1 for smoker and 0 for non-smoker

The ratio of the two hazards reduces to:

$$e^{(\beta 2)} \text{ or } e^{(.52337)} \text{ or } 1.7$$

Each individual is his/her own expected comparison. Incremental assessment of the Cox Hazard Ratio then simplifies to changes in medical screening and smoking status covariates.[3,4] The hazard ratio then can be properly interpreted as relative risk for that specific individual with the standard of comparison being that individual's observed medical tests and smoking status with an ideal non smoker with normal medical screenings and of the same age. Using each individual as a comparison, relative risk becomes a standard by which individuals can judge their personal net screening improvement over time. This measure differs from the 10 year CHD probability calculations presented in the Wilson et. al. paper. With this methodology the calculated relative risk is based only on the screening vectors i.e. blood measures and smoking status. It in effect is a tool that monitors those vectors over time. Its goal is to help individuals set up targets for screening improvements and not for statistical comparisons of population risk.

Participant Feedback Adjustments

The purpose of the risk profile is to develop a composite score that retirees can use to monitor their progress. It is very difficult for the lay public to readily assess relative risk measures with a risk of 1 considered normal. In order to make this number more understandable to the participant, all relative risks were categorized into 5 groups or "codes":

Code	Relative Risk range
0	1.00 or smaller
1	Greater than 1.00 to 1.70
2	Greater than 1.70 to 2.20
3	Greater than 2.20 to 2.80
4	Greater than 2.80

For further clarification of the composite risk score, a description of the relative importance of each category was developed and described in the newsletter that went out to all HealthLink members. The guidelines as published are:

0 - Normal Risk

> If you are in this category then the combined risk of all your screening tests are in the normal range in term of impact on heart disease. It can be considered the average risk for some one your age. Though there may be some variation in specific tests, the combined risk impact is at a level appropriate for your age.

1 - Above Average Risk

> Individuals in this risk category usually indicate a level of risk that one or possibly two of the screening results are in the elevated range. It is a level of risk that extra care in exercise, diet, smoking cessation and a healthy lifestyle can have a great impact.

2 - Elevated Risk; 3 - Elevated Risk, Moderate; 4 - Elevated Risk, Severe

> Individuals in these three categories are at such a risk level that a concerted effort should be made to improve their screening results. In addition to lifestyle changes, medication may be required to improve screening results. As with any wellness program an ongoing dialog and guidance with an individual's primary care physician is of extreme importance, especially as one's risk profile get into these elevated ranges.

Ongoing feedback is a key element of the HealthLink project. The summary of screening results into a well defined set of targets will help focus each individual's efforts and also hopefully provide additional motivation to change their risk status.

References

1. Peter W. F. Wilson; R. B. D'Agostino; D. Levy; A. M. Belanger; H. Silbershatz; W. B. Kannel
   <u>Prediction of Coronary Heart Disease using Risk Factor Categories</u>
   Circulation. 1998;97:1837-1847.)

2. Zhu, Zhiwei; Hoag, Michael; Julien,Stehane;Sufang, Cui.
   <u>Estimating Mortality of Insured Advanced-age Populations with Cox Regression Model;</u> Risk Management, Transamerica Reinsurance, 2001

3. Cleves, M. A.; Gould, W.W.; Gutierrez, R.G. <u>Introduction to Survival Analysis Using Stata ;</u>Stata Press, College Station Texas, 2002

4. Kleinbaum, D.G.; <u>Survival Analysis;</u> Springer-Verlag, New York, 1995

5. Cox, D.R. Regression models and life tables.
   J R Stat Soc B. 1972; 34: 187-220.