CHAPTER THREE

PROPORTIONAL HAZARDS MODEL

The model for comparing the hazard between two groups of individuals can be expressed as

$$h_N(t) = \psi h_S(t) \tag{3.1}$$

where t is non-negative values and ψ is a constant, and $h_S(t)$ and $h_N(t)$ are the hazards of standard treatment and new treatment respectively. The model can be easily generalized. Suppose that the survival data are available for n individuals and the hazard function for the ith observations is denoted by $h_i(t)$, $i = 1,2,\ldots, n$. Let $h_0(t)$ be the hazard function for an individual on the standard treatment at time t. Therefore, we have $h_i(t) = \psi h_0(t)$ and, consequently, $\frac{h_i(t)}{h_0(t)} = \psi$ is defined as the hazard ratio. Since the relative hazard ψ cannot be negative, it is convenient to set $\psi = exp(\beta)$ where β is the coefficient parameter. Further, we may rewrite the parameter β as a logarithm of the hazard ratio, that is, $\beta = \log(\psi)$. Any value of β in the range $(-\infty, \infty)$ will lead to a positive value of ψ. Note that positive values of β is obtained when hazard ratio ψ is greater than unity, that is when the new treatment is inferior to the standard.

Now, consider a survival data set with an explanatory variable X. Assume that X is an indicator variable which takes the value zero if an individual is on standard treatment and unity if an individual is on the new treatment. If x_i is the value of X for the ith individual in the study, $i = 1,2,\ldots, n$, the hazard function for this individual can be written as

$$h_i(t) = \exp(\beta x_i) h_0(t) \tag{3.2}$$

where $x_i = 1$ if the *i*th individual is on the new treatment and $x_i = 0$ if the *i*th individual is on the standard treatment. This is the proportional hazard model for comparing two groups of treatment.

Cox (1972) considered the situation where the hazard of death at a particular time depends on the values $x_1, x_2, ..., x_p$ of p explanatory variables $X_1, X_2, ..., X_p$ respectively. The values are assumed to have been recorded at the time origin of the study. The set of values for the *i*th observation in the proportional hazard model will be represented by a vector $\boldsymbol{x}_i$, such that, $\boldsymbol{x_i} = (x_1, x_2, ..., x_p)'$. Let $h_0(t)$ be the hazard function for an individual for whom the values of all the explanatory variables make up the vector $\boldsymbol{x}$ are zero and $h_0(t)$ is called the baseline hazard function. Hence, the hazard function for the *i*th observation can then be written as

$$h_i(t) = \psi(\boldsymbol{x}_i) h_0(t) \tag{3.3}$$

Note that the $\psi(\boldsymbol{x}_i)$ in equation (3.3) cannot be negative values. Therefore, it is convenient to write equation (3.3) in term of $\exp(\eta_i)$, where η_i is a linear combination of the p explanatory variables, that is $\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}$ where $\beta_1, \beta_2, ..., \beta_p$ are the coefficients of explanatory variables in the model. η_i is known as risk score or prognostic index for the *i*th individual. The general proportional hazard model then becomes

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}) h_o(t) \tag{3.4}$$

This model is a semi-parametric model in terms of unspecified probability of the hazard function and also the unknown probability distribution for survival times.

3.1 Estimating the PHM Parameters

Let $t_1, t_2 \cdots t_n$ denote the survival times for *n* observed survival times. Suppose that there are *r* distinct death times and $n - r$ right-censored survival times. We assume that there are no ties in the data. Let $t_{(j)}$ be the *j*th order of death such that $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$. Let $R(t_{(j)})$ be the set of individuals who are alive and uncensored at a time just prior to $t_{(j)}$. If

$$\text{P [individual with variables } \boldsymbol{x}_j \text{ dies at } t_{(j)} \mid \text{ one death at } t_{(j)}]$$

is considered, it gives us

$$\frac{\text{P[individual with variables } \boldsymbol{x}_j \text{ dies at } t_{(j)}]}{\text{P[one death at } t_{(j)}]},$$

and therefore $\dfrac{h_i(t_{(j)})}{\sum_{\ell \in R(t_{(j)})} h_i(t_{(\ell)})}$.

Consequently, as shown by Cox (1972), the relevant likelihood function for the PHM model is given by

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_{(j)})}{\sum_{\ell \in R(t_{(j)})} \exp(\boldsymbol{\beta}' \boldsymbol{x}_{(\ell)})} \tag{3.5}$$

where $\boldsymbol{x}_j$ is the vector of covariates for the individual who dies at the *j*th ordered death time, $t_{(j)}$. It is usually referred to as a partial likelihood function. Further, let

$$\delta_i = \begin{cases} 0 & \text{the } i\text{th time is right – censored} \\ 1 & \text{otherwise} \end{cases}$$

Then, equation (3.5) can be written as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \left[\frac{\exp\left(\boldsymbol{\beta}' \boldsymbol{x}_{(i)}\right)}{\sum_{\ell \in R(t_i)} \exp\left(\boldsymbol{\beta}' \boldsymbol{x}_{(\ell)}\right)} \right]^{\delta_i} \tag{3.6}$$

where $R(t_i)$ is the risk set at time t_i. The log-likelihood function is then

$$\log\ L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta}' \boldsymbol{x}_{(i)} - \log \sum_{\ell \in R(t_i)} \exp\left(\boldsymbol{\beta}' \boldsymbol{x}_{(\ell)}\right) \right]$$

Using the Newton-Raphson procedure, the maximum likelihood estimate of parameter $\boldsymbol{\beta}$ can be obtained. The estimates at the $(s+1)$th iteration, $\hat{\boldsymbol{\beta}}_{s+1}$, is given by

$$\hat{\boldsymbol{\beta}}_{s+1} = \hat{\boldsymbol{\beta}}_s + \boldsymbol{I}^{-1}\left(\hat{\boldsymbol{\beta}}_s\right)\boldsymbol{u}\left(\hat{\boldsymbol{\beta}}_s\right) \text{ for } s = 0,1,...$$

where $\boldsymbol{u}\left(\hat{\boldsymbol{\beta}}_s\right)$ is the vector of efficient score and $\boldsymbol{I}^{-1}\left(\hat{\boldsymbol{\beta}}_i\right)$ is the inverse of the information matrix evaluated at $\hat{\boldsymbol{\beta}}_s$. The procedure can start with $\hat{\boldsymbol{\beta}}_0 = 0$. The process is terminated when the change in the log-likelihood function is sufficiently small, or when the largest of the relative changes in the values of the parameter estimates is sufficiently small. When the iterative procedure has converged, the variance-covariance matrix of the parameter estimates can be approximated by the inverse of the information matrix evaluated at $\hat{\boldsymbol{\beta}}$, that is, $\boldsymbol{I}^{-1}\left(\hat{\boldsymbol{\beta}}_i\right)$. The square root of the diagonal elements of this matrix is then the standard error of the estimated values of $\beta_1, \beta_2, ..., \beta_p$ respectively.

3.2 Estimating the Baseline Hazard Functions of PHM

So far, we have only considered the estimation of parameter $\boldsymbol{\beta}$ in the linear component of a proportional hazards model. Once a suitable model for a set of survival data has been identified, the hazard function and the corresponding survivor function can be

estimated. These estimations can then be used to summarize the survival experience of individuals in the study.

Estimates of the baseline hazard function can be derived using an approach based on the method of maximum likelihood (see Kalbfleisch and Prentice (1973)). Collet (1994) described the estimation of baseline hazard in detail and described here. Suppose that there are r distinct death times arranged in increasing order, $t_{(1)} < t_{(2)} < ... < t_{(r)}$, and there are d_j deaths and n_j individual at risk at time $t_{(j)}$. The estimated baseline hazard function at time $t_{(j)}$ is then given by

$$\hat{h}_0(t_{(j)}) = 1 - \hat{\xi}_j \tag{3.7}$$

where $\hat{\xi}_j$ is the solution of the equation

$$\sum_{l \in D(t_{(j)})} \frac{\exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)}{1 - \hat{\xi}_j^{\exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)}} = \sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l) \tag{3.8}$$

for j = 1, 2, ..., r. In equation (3.8), $D(t_{(j)})$ is the set of all d_j individuals who die at the jth ordered death time, $t_{(j)}$, and $R(t_{(j)})$ is the set of all n_j individual at risk of death at time $t_{(j)}$. The estimates of the β's which form the vector $\hat{\boldsymbol{\beta}}$ are those which maximize the likelihood function in equation (3.5).

3.2.1 No Tied Death Time

In the particular case where there are no tied death times, that is, where $d_j = 1$ for $j = 1, 2, ..., r$, the left-hand side of equation (3.8) is a single term for every j. This equation can then be solved to give

$$\hat{\xi}_j = \left(1 - \frac{\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)}\right)^{\exp(-\hat{\boldsymbol{\beta}}'\boldsymbol{x}_{(j)})}$$

where $\boldsymbol{x}_{(j)}$ is the vector of explanatory variables for the individual who dies at times $t_{(j)}$. When there are ties, one or more of d_j are greater than unity. As a result, equation (3.8) cannot be solved explicitly and an iterative scheme is required.

3.2.2 Tied Death Time

Assume that the hazard of death is constant between adjacent death times. An appropriate estimate of the baseline hazard function in this interval is then obtained by dividing the estimated hazard in equation (3.7) by the time interval, to give the step function

$$\hat{h}_0(t) = \frac{1 - \xi_j}{t_{(j+1)} - t_{(j)}} \tag{3.9}$$

for $t_{(j)} \le t < t_{(k+j)}$, j = 1, 2, ..., r - 1, with $\hat{h}_0(t) = 0$ for $t < t_{(1)}$.

The quantity $\hat{\xi}_j$ can be regarded as an estimate of the probability that an individual survives through the interval from $t_{(j)}$ to $t_{(j+1)}$. The baseline survivor function can then be estimated by

$$\hat{S}_0(t) = \prod_{j=1}^{k} \hat{\xi}_j \tag{3.10}$$

for $t_{(k)} \le t < t_{(k+1)}$, k = 1, 2, ..., r – 1. This function is a step function. The estimated value of the baseline survivor is unity for $t < t_{(1)}$, and zero for $t \ge t_{(r)}$, unless there are censored survival times greater than $t_{(r)}$.

From equation (1.4), the baseline cumulative hazard function is given by $H_0(t) = \log S_0(t)$. Thus, the estimate of $H_0(t)$ can be approximated by

$$\hat{H}_0(t) = -\log \hat{S}_0(t) = -\sum_{j=1}^{k} \log \hat{\xi}_j \tag{3.11}$$

for $t_{(k)} \le t < t_{(k+1)}$, $k = 1,2,...,r-1$, with $\hat{H}_0(t) = 0$ for $t < t_{(1)}$.

3.2.3 Alternative Method

Revisiting equation (3.8), the term $\hat{\xi}_j^{\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)}$ in the denominator of the left-hand side of the equation can be written as $\exp\{\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)\log\hat{\xi}_j\}$. Taking the first two terms in the expansion of the exponent gives

$$\exp\{\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)\log\hat{\xi}_j\} \approx 1 + \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)\log\hat{\xi}_j .$$

Writing $1-\tilde{\xi}_j$ for the estimated baseline hazard at time $t_{(j)}$ obtained using this approximation, and substituting $1 + \exp\left\{\exp \hat{\boldsymbol{\beta}}'\boldsymbol{x}_l \log \hat{\xi}_j\right\}$ for $\hat{\xi}_j^{\exp(\hat{\beta}_j x_l)}$ in equation (3.8), we find $\tilde{\xi}_j$ such that

$$-\sum_{l\in D(t_{(j)})} \frac{1}{\log \xi_j} = \sum_{l\in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l).$$

Therefore, we have $\dfrac{-d_j}{\log \xi_j} = \sum_{l\in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)$ since d_j is the number of deaths at the jth ordered death time, $t_{(j)}$. Hence, we have

$$\tilde{\xi}_j = \exp\left(\frac{-d_j}{\sum_{l\in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)}\right) \tag{3.12}$$

From equation (3.10), an estimate of the survivor function, based on the values of $\tilde{\xi}_j$, is given by

$$\tilde{S}_0(t) = \prod_{j=1}^{k} \exp\left(\frac{-d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)} \right) \tag{3.13}$$

for $t_{(k)} \le t < t_{(k+1)}$, k = 1, 2, ..., r - 1. The estimate of the baseline cumulative hazard function derived from $\tilde{S}_0(t)$ is

$$\tilde{H}_0(t) = -\log \tilde{S}_0(t) = \sum_{j=1}^{k} \frac{d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)} \tag{3.14}$$

for $t_{(k)} \le t < t_{(k+1)}$, k = 1, 2, ..., r – 1. This estimate is often referred to as the Nelson-Aalen estimate or the Breslow estimate of the baseline cumulative hazard function.

When there are no covariates, the estimated baseline survivor function in equation (3.13) becomes

$$\prod_{j=1}^{k} \exp(-d_j / n_j) \tag{3.15}$$

since n_j is the number of individuals at risk at time $t_{(j)}$. This is the Nelson-Aalen estimate of the survivor function given in equation $\tilde{S}(t) = \prod_{j=1}^{k} \exp\left(\frac{-d_j}{n_j} \right)$, and the corresponding estimate of the baseline cumulative hazard function is $\sum_{j=1}^{k} d_j / n_j$.

A further approximation is found by noting that the expression $\frac{-d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}' x_l)}$, in the exponent of equation (3.12) tend to be small, unless there are large numbers of ties at particular death times. Taking the first two terms of the expansion of this exponent, and denoting this new approximation to ξ_j by ξ_j^*, gives

$\xi_j^* = 1 - \dfrac{d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\beta}' x_l)}$. Adapting equation (3.9), the estimated baseline hazard function in the interval from t_j by $t_{(j+1)}$ is then given by

$$h_0^*(t) = 1 - \frac{d_j}{(t_{(j+1)} - t_{(j)}) \sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)} \tag{3.16}$$

for $t_{(j)} \le t < t_{(j+1)}$, j = 1, 2, ..., r - 1. Using ξ_j^* in place of $\tilde{\xi}_j$ in equation (3.10), the corresponding estimated baseline survivor function is

$$S_0^*(t) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{\sum_{l \in R(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)} \right),$$

and a further approximate estimate of the baseline cumulative hazard function is $H_0^*(t) = -\log S_0^*(t)$. Notice that the cumulative hazard function in equation (3.14) at time *t* can be expressed in the form $\tilde{H}_0(t) = \sum_{j=1}^{k} (t_{(j+1)} - t_{(j)} h_0^*(t)$ where $h_0^*(t)$ is given in equation (3.16). Consequently, differences in successive values of the estimated baseline cumulative hazard function in equation (3.14) provide an approximation to the baseline hazard function, at times $t_{(1)}$, $t_{(2)}$, ..., $t_{(r)}$, that can easily be computed.

In practice, it will often be computationally advantageous to use either $\tilde{S}_0(t)$ or $S_0^*(t)$ in place of $\hat{S}_0(t)$. When the number of tied survival times is small, all three estimates will be very similar.

3.3 Variable Selection Procedures

Determination on selection of any explanatory variables is carried out by comparing

any two or more models based on the statistics, $-2\log\hat{L}$. This statistic is always positive and the smaller the value is the better. Two models can then be compared based on the difference between the values of their respective statistics values, say $-2\log\hat{L}(1)$ and $-2\log\hat{L}(2)$, giving

$$-2\log\hat{L}(1)-(-2\log\hat{L}(2))=-2\log\left\{\frac{\hat{L}(1)}{\hat{L}(2)}\right\} \tag{3.20}$$

Equation (3.20) is the log-likelihood ratio statistics for testing H_0: extra terms in the model 2 are all zero. The log-likelihood ratio statistics has an asymptotic chi-squared distribution with the degrees of freedom is given by the number of extra terms considered in Model 2 when compared to Model 1.

For example, begin modeling the survival data set with no explanatory variable included; known as *null* model, then the value of $-2\log\hat{L}$ is recorded. Then fit models that contain each of the variables one at a time. The values of $-2\log\hat{L}$ for these models are recorded. Then obtain the difference between the new $-2\log\hat{L}$ with the $-2\log\hat{L}$ of the null model. Based on the reductions observed for each model, we can find the *p*-value and compare with the specified significant level. In order to include or omit a term, the significant level recommended is less than 10%.

Than variable that appears to be the most important from previous step is included first into the model. Then we repeat the similar procedure by including one variable at a time into the current model in order to identify the next variable to be included in the model. The process is repeated until none of the variables are significant.