Generalized competing event model

Let *n, k,* and *p* be the number of observations, covariates, and mutually exclusive event types, respectively. Let *z* be the number cause-specific events, and *p-z* be the number of competing events. Let **d** represent the *k* x 1 vector of covariate values, and $\mathbf{1_m}$ represent a *m* x 1 vector of 1's. Let *i* be an index of natural numbers ranging from 1 to p. Let λ_{0i} represent the cause-specific hazard for event *i*, and $\lambda_0 = \Sigma\, \lambda_{0i}$ represent the hazard for any event, under a given set of experimental conditions.

We model the cause-specific hazard for event *i*, under an alternative set of conditions as $\lambda_{1i} = g(\mathbf{X}\boldsymbol{\beta}_i)\, \lambda_{0i}$, for an invertible function g(·), an *n* x *k* data matrix **X**, and a *k* x 1 vector of effect coefficients $\boldsymbol{\beta}_i$. The hazard for any event under the alternative set of conditions is $\lambda_1 = \Sigma\, \lambda_{1i} = \Sigma\, g(\mathbf{X}\boldsymbol{\beta}_i)\, \lambda_{0i}$ and the hazard ratio is expressed as:

$$\lambda_1 / \lambda_0 = \Sigma\, g(\mathbf{X}\boldsymbol{\beta}_i)\, \lambda_{0i} / \Sigma\, \lambda_{0i} \qquad (2)$$

in other words, the hazard ratio is a weighted average of the effects on the cause-specific hazards under the initial conditions. Here $\boldsymbol{\beta}$ is the *k* x *p* coefficient matrix, with each element $\beta_{v,w}$ representing the effect of covariate *v* on event *w*. Note that under the assumption of effect homogeneity with respect to the cause-specific events, $\boldsymbol{\beta}_j=\boldsymbol{\beta}_k=\beta$ for all j, k ε {1,…,p}, therefore:

$$\lambda_1 / \lambda_0 = \Sigma\, g(\mathbf{X}\boldsymbol{\beta}_i)\, \lambda_{0i} / \Sigma\, \lambda_{0i} = \Sigma\, g(\mathbf{X}\beta)\, \lambda_{0i} / \Sigma\, \lambda_{0i} = g(\mathbf{X}\beta)\, \Sigma\, \lambda_{0i} / \Sigma\, \lambda_{0i} = g(\mathbf{X}\beta).$$

Let $\mathbf{b}_i$ be a maximum (partial) likelihood estimator for $\boldsymbol{\beta}_i$ (e.g., using $g(x) = e^x$);[1] alternatively, we can let $\mathbf{b}_i$ represent an analogous maximum partial likelihood estimator for sub-distribution hazards.[2,3] Let $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2 \ldots \mathbf{b}_p]$ be the k x p matrix of coefficients, with each element $b_{v,w}$ of $\mathbf{B}$ representing the estimated effect of covariate v on event w. Since columns of $\mathbf{B}$ are interchangeable, we can order the elements of $\mathbf{B}$ such that the first z vectors correspond to events of interest and the remaining p-z vectors correspond to competing events, i.e. $\mathbf{B}_{1,z} = [\mathbf{b}_1\ \mathbf{b}_2 \ldots \mathbf{b}_z]$ and $\mathbf{B}_{z,p} = [\mathbf{b}_{z+1}\ \mathbf{b}_{z+2} \ldots \mathbf{b}_p]$, so $\mathbf{B} = [\mathbf{B}_{1,z}\ \mathbf{B}_{z,p}]$. Now using the data vector $\mathbf{d}$, we construct an individual risk score as follows:

$$R = (\mathbf{d}^T \mathbf{B}_{z,p})\ \mathbf{1}_{p-z} - (\mathbf{d}^T \mathbf{B}_{1,z})\ \mathbf{1}_z \qquad (3)$$

Note that under the assumption of effect homogeneity with respect to the cause-specific events, $\mathbf{b}_j=\mathbf{b}_k=\mathbf{b}$ for all j, k ε {1,…,p}, so $R = c\mathbf{d}^T\mathbf{b}$ for some constant c.

References

1. Cox DR. Regression models and life tables. *J R Stat Soc Series B Stat Methodol* 1972;B34:187-220.
2. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496-509.
3. Jeong JH, Fine JP. Parametric regression on cumulative incidence function. *Biostatistics* 2007;8:184-196.