

Statistical models used in MORGAM stroke risk analyses

Sangita Kulahtinal*and Juha Karvanen†

December 9, 2008

This document is part of the supplement for the article “Relative risks for stroke by age, sex and population based on follow-up of 18 European populations in the MORGAM project” available at <http://www.ktl.fi/publications/morgam/stroke/>

1 Proportional Hazards model

The data are grouped into five regions: Nordic, Central, UK, Italy and France. Each region consists of one or more populations. Let k denote the region, $k = 1, 2, 3, 4, 5$ and j denote the population. Three sets of analyses are carried out: (a) men only, (b) women only, and (c) both men and women combined. The hazard of having a stroke event for an individual i in the population j and region k with a vector of covariate values, x_i (which are age at the baseline examination, blood pressure, body mass index, smoking status and HDL cholesterol) at time t_i is

$$\lambda_{ijk}(t_i | x_i) = \lambda_{jk}(t_i) \exp(\beta'_k x_i), \quad (1)$$

where $\beta'_k = (\beta_{k1}, \dots, \beta_{k5})$ are the effects of covariates. Following three hypotheses are of particular interest:

H_{01} : The covariate effects are the same between regions.

*The Indic Society for Education and Development (INSEED), Nashik, India. e-mail: sangita.kulathinal@inseed.org

†Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Mannerheimintie 166, 00300 Helsinki, Finland. e-mail: juha.karvanen@ktl.fi

H_{02} : The covariate effects are the same between men and women.

H_{03} : The covariate effects are the same between age-groups.

Note that the covariate effects are assumed to be the same across the populations within the region. However, the baseline hazard is allowed to vary between the populations within the region. The hazard ratio for m unit increase in a covariate value, say x_{il} at a given time t is $\exp(\beta_{kl}m)$ for the region k . The above model is first fitted to evaluate the effects of covariates in each region. Here we consider data from individuals without history of disease at baseline (defined in the main article) and also when all the covariate values are available.

2 Regional comparison

The hypothesis of interest is whether the effects of covariates differ between the regions or not. Three sets of analyses are carried out: (a) men only, (b) women only, and (c) both men and women combined.

The global hypothesis of equality of all five covariate effects between five regions is stated as

$$\beta_{1l} = \beta_{2l} = \beta_{3l} = \beta_{4l} = \beta_{5l} = \beta_l, \quad \forall l = 1, 2, 3, 4, 5.$$

Under this hypothesis, the model (1) can be written as

$$\lambda_{ijk}(t_i | x_i) = \lambda_{jk}(t_i) \exp(\beta' x_i), \quad (2)$$

where $\beta' = (\beta_1, \dots, \beta_5)$ are the common values of the covariate effects. This model is fitted by pooling the data from all the regions and then the hypothesis is tested using the likelihood ratio test using the chi-squared distribution with $(25 - 5) = 20$ degrees of freedom. The likelihood ratio statistic is defined as $-2(L_1 - L_0)$ where L_1 is the log-likelihood under the model (2) and L_0 is the log-likelihood under the model (1).

If this global hypothesis is rejected then the hypothesis of equality of effects of each covariate between the regions is tested separately. For example, the hypothesis of equal effect of age at the baseline can be stated as

$$\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41} = \beta_{51} = \beta_1.$$

Under this hypothesis, the model (1) can be written as

$$\lambda_{ijk}(t_i | x_i) = \lambda_{jk}(t_i) \exp(\beta'_{k1} x_i), \quad (3)$$

where $\beta'_{k1} = (\beta_1, \beta_{k2}, \beta_{k3}, \beta_{k4}, \beta_{k5})$, the coefficient corresponding to age does not vary with k . This hypothesis can be tested using the likelihood ratio test using the chi-square distribution with $(25 - 21) = 4$ degrees of freedom.

Note that region France does not include women and hence the global hypothesis for women involve only four regions. The degrees of freedom are adjusted accordingly to $(20 - 5) = 15$ and $(20 - 17) = 3$, respectively. The hazard ratios and corresponding confidence intervals are plotted in the figures.

3 Comparison between men and women

To evaluate the effect of sex on stroke events, a proportional hazard model is fitted by pooling the data for men and women separately from all the regions to ensure the reasonable cohort size and number of events in each group. Here, covariate effects are compared separately for men and women.

The hazard model for the sex s can be specified as

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta'_s x_i), \quad (4)$$

where $\beta'_s = (\beta_{s1}, \dots, \beta_{s5})$, and $s = 1$ if men and $s = 2$ if women. Note that the covariate effects do not depend on the region but the baseline hazard is allowed to vary with population within each region. The model (4) is fitted for men and women to evaluate covariate effects. A global hypothesis of equality of covariate effects among men and women can be stated as

$$\beta_{1l} = \beta_{2l} = \beta_l, \quad \forall l = 1, 2, 3, 4, 5.$$

Under this hypothesis, the hazard model (4) reduces to

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta' x_i), \quad (5)$$

where $\beta' = (\beta_1, \dots, \beta_5)$. The hypothesis is tested using the likelihood ratio test using chi-square distribution with $(10 - 5) = 5$ degrees of freedom.

If the global hypothesis is rejected then the hypothesis of equality of covariate effects for each covariate is tested. For example, the hypothesis of equal effect of age at the baseline can be stated as

$$\beta_{11} = \beta_{21} = \beta_1.$$

Under this hypothesis, the model (4) can be written as

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta'_{s1} x_i), \quad (6)$$

where $\beta'_{s1} = (\beta_1, \beta_{s2}, \beta_{s3}, \beta_{s4}, \beta_{s5})$, the coefficient corresponding to age does not vary with s . This hypothesis can be tested using the likelihood ratio test using the chi-square distribution with $(10 - 9) = 1$ degree of freedom.

4 Comparison between age-groups

Four age-groups were created to compare the covariate effects. The age-groups were defined using the age at the time of the baseline examination as $[-, 45)$, $[45, 55)$, $[55, 65)$, and $[65, -]$. These four groups can be compared using the method described in section 3 which is for two groups. The data were analysed separately for men and women.

The hazard model for the age-group s can be specified as

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta'_s x_i), \quad (7)$$

where $\beta'_s = (\beta_{s1}, \dots, \beta_{s5})$, and $s = 1, \dots, 4$. Note that the covariate effects do not depend on the region but the baseline hazard is allowed to vary with population within each region. The model (7) is fitted for each age-group to evaluate covariate effects. A global hypothesis of equality of covariate effects can be stated as

$$\beta_{1l} = \beta_{2l} = \beta_{3l} = \beta_{4l} = \beta_{5l} = \beta_l, \quad \forall l = 1, 2, 3, 4, 5.$$

Under this hypothesis, the hazard model (7) reduces to

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta' x_i), \quad (8)$$

where $\beta' = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$. The hypothesis is tested using the likelihood ratio test using chi-square distribution with $20 - 5 = 15$ degrees of freedom. If this hypothesis is rejected then the hypothesis of equal covariate effects for each covariate is tested separately. The model under the equal effect of blood pressure across the age-groups can be specified as

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta'_{s2} x_i), \quad (9)$$

where $\beta'_{s2} = (\beta_{s1}, \beta_2, \beta_{s3}, \beta_{s4}, \beta_{s5})$. The likelihood ratio statistic has a chi-square distribution with $(20 - 17) = 3$ degrees of freedom.

Note that the hypothesis of equality of age effects across the age-groups corresponds to the linearity assumption for age in the proportional hazards model. Here, we expect that this hypothesis is true and the hypothesis of equality of the other covariate effect can be tested by using the same age effect across the age-groups. The model under the equal effect of blood pressure across the age-groups under the assumption of the same effect of age across the age-groups can be specified as

$$\lambda_{ijk}^s(t_i | x_i) = \lambda_{jk}^s(t_i) \exp(\beta'_{s12} x_i), \quad (10)$$

where $\beta'_{s12} = (\beta_1, \beta_2, \beta_{s3}, \beta_{s4}, \beta_{s5})$. The likelihood ratio statistic has a chi-square distribution with $(20 - 14) = 6$ degrees of freedom.

5 Implementation in R

The analyses are carried out in R using `coxph` function from `survival` package. The models in R are parameterized as the models described above, i.e. variables for the interactions are explicitly created. The results are plotted using a modification of `metaplot` function from `rmeta` package. Tables are created directly in R as HTML tables.