

# Chapter 4

## Modelling Covariate Information

In this chapter, a novel model for incorporating covariate information based upon a modification of the generalized linear model using profile likelihood approach is presented.

### 4.1 A Generalized Linear Model

The theory of generalized linear models was first developed by Nelder and Wedderburn (1972). In simple terms, the components of a generalized linear model consist of three elements: the error distribution, the systematic component, and the link function [13, 15, 23, 25].

Firstly, the error distribution need to be chosen for a random response variable  $Y_i$ ,  $i = 1, \dots, n$ . Typically,  $Y_i$  is assumed to come from an exponential family, for example, a Gaussian, binomial, or Poisson distribution.

Secondly, the systematic component is a linear predictor  $\eta$  which combines linearly a set of parameters  $\beta_1, \dots, \beta_p$  and covariates  $z_{i1}, \dots, z_{ip}$  as

$$\eta_i = \beta_1 z_{i1} + \dots + \beta_p z_{ip}. \quad (4.1)$$

Thirdly, the link function is a monotone function  $g$  which relates the mean  $E(Y_i) = \mu_i$  of the random variable to the linear predictor:

$$g(\mu_i) = \eta_i. \quad (4.2)$$

Now a novel modification of the generalized linear model (as described above) is applied to model the variation of treatment effect by means of covariate information. Let  $z_{ij}$

be the value of the  $j$ -th covariate in the  $i$ -th center, for  $i = 1, \dots, k$  and  $j = 1, \dots, p$ . The linear predictor  $\eta_i$  for covariate information in the  $i$ -th center becomes

$$\eta_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \dots + \beta_p z_{ip} \quad (4.3)$$

where  $\beta_0, \beta_1, \dots, \beta_p$  are the parameters of the model whose values are usually unknown and have to be estimated from the data. We consider again the profile log-likelihood for the relative risk estimator

$$L^*(\theta) = \sum_{i=1}^k x_i^T \log(\theta_i) - (x_i^C + x_i^T) \log(n_i^C + \theta_i n_i^T). \quad (4.4)$$

The form of (4.4) suggests to use the canonical link  $\theta_i = \exp(\eta_i)$  to link the linear predictor (4.3) to the relative risk parameter  $\theta_i$  which guarantees that  $\theta_i \geq 0$ , an essential requirement for a relative risk. With the canonical link, the profile log-likelihood for covariate information becomes

$$L^*(\beta) = \sum_{i=1}^k x_i^T \eta_i - (x_i^C + x_i^T) \log(n_i^C + \exp(\eta_i) n_i^T) \quad (4.5)$$

with  $\eta_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \dots + \beta_p z_{ip}$ .

## 4.2 Finding Maximum Likelihood Estimates

We now estimate the parameters of model  $\beta_j$ , for  $j = 0, 1, \dots, p$ . To find the maximum likelihood estimate of  $\beta_j$  we need to maximize the profile log-likelihood (4.5). For this purpose, consider the partial derivative w.r.t  $\beta_j$

$$\frac{\partial L^*}{\partial \beta_j}(\beta) = \sum_{i=1}^k x_i^T z_{ij} - x_i n_i^T \frac{\exp(\eta_i)}{n_i^C + \exp(\eta_i) n_i^T} z_{ij} \quad (4.6)$$

where  $x_i = x_i^T + x_i^C$ , and the corresponding vector of partial derivatives, the gradient is

$$\nabla L^*(\beta) = \left( \frac{\partial L^*}{\partial \beta_0}, \dots, \frac{\partial L^*}{\partial \beta_p} \right)'. \quad (4.7)$$

Furthermore, *Hesse* matrix of second derivatives is

$$\frac{\partial^2 L^*}{\partial \beta_h \partial \beta_j}(\beta) = - \sum_{i=1}^k \frac{x_i n_i^T n_i^C \exp(\eta_i)}{(n_i^C + \exp(\eta_i) n_i^T)^2} z_{ij} z_{ih} \quad (4.8)$$

so that (4.8) becomes in matrix form

$$\nabla^2 L^*(\beta) = \left( \frac{\partial^2 L^*}{\partial \beta_h \partial \beta_j}(\beta) \right) = -Z'W(\beta)Z \quad (4.9)$$

where  $Z$  is the design matrix of covariate information defined as

$$Z = \begin{pmatrix} z_{10} & z_{11} & z_{12} & \dots & z_{1p} \\ z_{20} & z_{21} & z_{22} & \dots & z_{2p} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ z_{k0} & z_{k1} & z_{k2} & \dots & z_{kp} \end{pmatrix}_{k \times (p+1)}$$

with

- $k$  is the number of centers,
- $p$  is the number of covariates,
- $z_{i0}$  is the constant value of coefficient in the  $i$ -th center,  $z_{i0} = 1$ ,
- $z_{i1}, \dots, z_{ip}$  is value of covariates in the  $i$ -th center,

and  $W(\beta)$  is a diagonal matrix defined as

$$W(\beta) = \begin{pmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1k} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2k} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ w_{k1} & w_{k2} & w_{k3} & \dots & w_{kk} \end{pmatrix}_{k \times k} \quad (4.10)$$

with  $w_{ij} = 0$ , if  $i \neq j$  and

$$w_{ii} = \frac{x_i n_i^T n_i^C \exp(\eta_i)}{(n_i^C + \exp(\eta_i) n_i^T)^2}.$$

Then, we used the Newton-Raphson procedure which is a traditional method for finding maximum likelihood estimate, to iteratively construct the maximum likelihood estimates of  $\beta_j$ . Choose some  $\beta^{(0)}$  as initial values (for example  $\beta^{(0)} = 0$ ) and then update  $\beta$  according to

$$\beta^{(n+1)} = \beta^{(n)} - \nabla^2 L^*(\beta^{(n)})^{-1} \nabla L^*(\beta^{(n)}) \quad (4.11)$$

until convergence.

### 4.3 Finding Standard Errors of Effect Estimates

We can estimate variances of maximum likelihood estimates from the negative inverse of the information matrix (4.9) [15, 23]. The variances estimate equals

$$\hat{v}ar(\hat{\beta}_j) = (Z'W(\hat{\beta})Z)_{jj}^{-1} \quad (4.12)$$

so that the standard errors become

$$s.\hat{e}.\hat{(\beta}_j) = \sqrt{\hat{v}ar(\hat{\beta}_j)} \quad (4.13)$$

and significance of individual effects can be consequently obtained by means of a Wald-test defined as

$$T_j = \frac{\hat{\beta}_j}{s.\hat{e}.\hat{(\beta}_j)} \quad (4.14)$$

so that the  $P$ -value of Wald test under the null hypothesis of no effect of the  $j$ -th covariate can be found as

$$P\text{-value} = 1 - F(T_j) \quad , T_j \geq 0 \quad (4.15)$$

with  $F(T_j)$  as the cumulative function of the standard normal distribution.

### 4.4 Finding Relative Risk and 95% CI for Covariate Information

After finding standard errors of effect estimates, we compute the relative risk and 95% confidence interval for covariate information as follows:

We consider again the linear predictor model (4.3)

$$\eta_i = \beta_0 + \beta_1 z_{i1} + \dots + \beta_p z_{ip}. \quad (4.16)$$

If  $z_{i1}, z_{i2}, \dots, z_{ip} = 0$ , then  $\eta_i = \beta_0$ , and the relative risk for the  $i$ -th center is estimated as

$$\hat{\theta}_i = \exp(\hat{\beta}_0) \quad (4.17)$$

and the associated 95% CI can be found as

$$\exp\{\hat{\beta}_0 \pm 1.96 \times s.\hat{e}.\hat{(\beta}_0)\}. \quad (4.18)$$

If  $z_{i1}, z_{i2}, \dots, z_{ip} \neq 0$ , then  $\eta_i = \beta_0 + \beta_1 z_{i1} + \dots + \beta_p z_{ip}$ , and the relative risk for the  $i$ -th center is estimated as

$$\hat{\theta}_i = \exp(\hat{\eta}_i) \quad (4.19)$$

where  $\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \dots + \hat{\beta}_p z_{ip}$ , and the associated 95% CI can be defined as

$$\exp\{\hat{\eta}_i \pm 1.96 \times s.\hat{e}.\hat{\eta}_i\} \quad (4.20)$$

where  $s.\hat{e}.\hat{\eta}_i$  can be found as

$$s.\hat{e}.\hat{\eta}_i = \sqrt{z_i' \widehat{COV}(\hat{\beta}) z_i} \quad (4.21)$$

where  $z_i = (z_{i0}, z_{i1}, z_{i2}, \dots, z_{ip})'$ , with  $z_{i0} = 1$ .

## 4.5 Elements of the Developed Software Tool

The profile log-likelihood for covariate information (4.5) is a non-standard log-likelihood. It is neither a Poisson log-likelihood nor any of the log-likelihoods available in the standard generalized linear model family. This makes it less attractive to use one of the existing statistical packages such as STATA, S-plus, MINITAB, or any other package which offers macro-like programming. Instead, it was decided to use the Microsoft Fortran Power Station to develop a software tool to accomplish modelling relative risk based upon the profile likelihood, since the Microsoft Fortran Power Station has many features that make development easy and efficient. In the following, we describe some of the features of this developed tool.

- The first feature is to calculate relative risk based upon modelling of unobserved heterogeneity. This feature allows to calculate relative risk based upon multi-level model, fixed-effect model, and profile likelihood model.
- The second feature is to calculate relative risk based upon modelling of covariate information. This feature allows to incorporate covariate information based upon a modification of the generalized linear model using profile likelihood approach in order to estimate the relative risk based upon the significant covariates.
- The third feature is to provide model diagnostics such as directional-derivative-type diagnostics based upon the multi-level model, fixed-effect model and profile likelihood model.

- The fourth feature is to allow center allocation based on the maximum posterior distribution.

For example, we consider the multicenter study of 59 trials that evaluates the effect of nicotine replacement therapy (NRT) on the quitting smoking. There are two different forms of NRT (patch and gum) and two different types of support (high support and low support). The modelling of covariate information using profile likelihood approach has been applied to see if the quitting smoking is dependent on the form of NRT, and/or on the type of support. The results of calculation relative risk based upon the modelling of covariate information using the profile likelihood approach from the developed software is displayed in the Figure 4.1. All examples and analyses presented in the next chapters were done using this software.

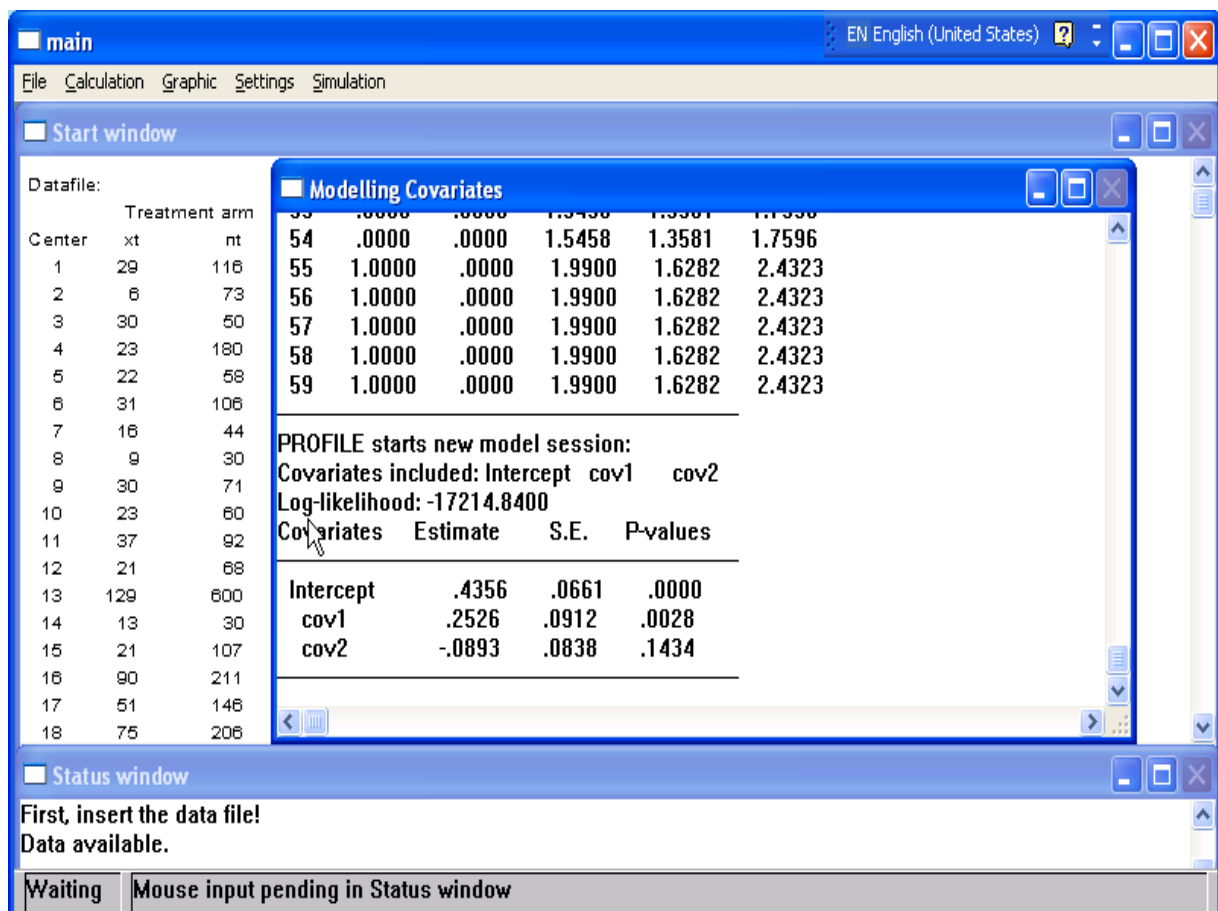


Figure 4.1: An Example of Output Window based upon the Developed Software Tool for the Quitting Smoking Study

