Chapter 2

The Classical Methods

2.1 Weighted Regression Methods

These methods use the idea of a generalized linear model to explain variation in treatment effects by covariates [4, 17, 29, 32, 36].

Let y_i be the observed treatment effect in the i-th study, for $i = 1, \ldots, k$. The y_i can be the observed log-odds ratio or log-relative risk in a trial with binary outcome or the observed mean difference in a trial with continuous outcome. In our case, y_i is the observed log-relative risk. It is assumed that y_i are independently distributed as

$$y_i \sim N(\theta_i, v_i) \tag{2.1}$$

where θ_i is the true treatment effect in the i-th study, and v_i is the variance of the log-relative risk in the i-th study.

For the fixed-effects model, it is supposed that there are p known covariates $z_1, \ldots, z_p$ which are presumed to account completely for variation in the true effects, so θ_i is specified by $\beta'\mathbf{z}_i$, where β is a column vector of regression coefficients $(\beta_0, \beta_1, \ldots, \beta_p)'$ and $\mathbf{z}_i$ is a column vector that contains the values of p covariates for study i.

$$y_i \sim N(\beta'\mathbf{z}_i, v_i) \tag{2.2}$$

and the fixed effects regression model for treatment effect estimate becomes

$$y_i = \beta'\mathbf{z}_i + e_i \tag{2.3}$$

where e_i is the error of estimation of study i. Each e_i is statistically indenpendent with a mean of zero and variance v_i.

Estimation of regression coefficients is usually carried out via weighted least squares algorithms. The analysis can be conducted using standard computer programs (e.g., in SAS, SPSSX, or BMDP) that compute weighted multiple regression analyses. The weights can be defined by the reciprocal of the sampling variances as $w_i = 1/v_i$. The standard errors of regression coefficient estimates can be computed as the square roots of the diagonal elements of the inverse of the weighted sum of squares and cross products matrix.

For the random-effects model, it is assumed that the true treatment effects θ_i vary randomly across studies and are independently distributed as

$$\theta_i \sim N(\mu, \tau^2) \tag{2.4}$$

where μ is the mean of the distribution of θ_i across studies, and τ^2 is the variance of the distribution of θ_i across studies.
To incorporate the covariates and thus account for heterogeneity among studies, μ can be specified by $\beta'\mathbf{z}_i$

$$\theta_i \sim N(\beta'\mathbf{z}_i, \tau^2) \tag{2.5}$$

and the random effects regression model for treatment effect estimate becomes

$$y_i = \beta'\mathbf{z}_i + \delta_i + e_i \tag{2.6}$$

where δ_i is the random effect of study i, that is the deviation of study i's true treatment effect from the true mean of all studies having the same covariate values. Each random effect, δ_i, is assumed to be independent with a mean of zero and variance τ^2, and e_i is the error of estimation of study i. Each e_i is statistically independent with a mean of zero and variance v_i.

The design vector $\mathbf{z}_i$ and within-study variance v_i are assumed to be known, and regression coefficient vector β and between-study variance τ^2 are estimated from the data. Notice that equation (2.6) has two components in its error term, $\delta_i + e_i$, which are assumed to be independent, leading to a covariance equal zero, so that the marginal variance of y_i is

$$v_i^* = Var(\delta_i + e_i) = \tau^2 + v_i \tag{2.7}$$

Estimation of regression coefficients are also obtained by weighted least squares algorithm with weights $w_i^* = 1/v_i^*$ and τ^2 must be explicitly estimated in order to undertake the weighted regression. The details and discussion of the estimation of the between-study variance can be reviewed in detail in Berkey and Thompson. [4, 36]. The statisti-

cal literature describes equation (2.6) as a mixed effects linear model with fixed effects β and random effects δ_i.

Notice that equation (2.6) is identical to the fixed effects regression model given in equation (2.3), with one important exception: the addition of the random effects δ_i. The treatment effect estimates in the random effects regression model correspond to the mean value of log-relative risk for all studies with a specified combination of covariates. The treatment effect estimates from a fixed effects model, in contrast, correspond to the single true value for that combination of covariates because between-study variance τ^2 equals zero [4].

A particular disadvantage of this modelling is the inappropriate identity link which is used to link covariate information to relative risk, since it does not guarantee that the relative risk estimates are positive, which would be an essential requirement for a relative risk. This problem is overcome using the canonical link which guarantees that relative risk estimates are positive.

The second disadvantage is a potential violation of the normality assumption of both the observed treatment effects and the random effects. For example, it has been assumed that the log-relative risk is normally distributed, however, this may not be appropriate for small studies or small number of events. Moreover, in practice, the v_i are rather estimated from the data than known, so the correlation between estimates of log-relative risk and their variance estimates may produce bias in the estimates of regression coefficients [4]. These problems are overcome by directly using the structure of the binary data, binomial or poisson.

2.2 Logistic Regression Methods

These methods use directly the binomial structure for the binary data. Let y_{ij} be the number of events in the j-th group ($j = 0$ control, $j = 1$ treated) of study i and n_{ij} be the number of subjects in the j-th group of study i. Also let π_{ij} denote the risk (probability) of an event in the j-th group of study i. It is assumed that y_{ij} is independently distributed as

$$y_{ij} \sim \text{Binomial}(\pi_{ij}, n_{ij}) \tag{2.8}$$

Suppose there are p known covariates $z_1, \ldots, z_p$ which might be the sources of variation between studies. Let $\mathbf{z}_{ij}$ be a column vector that contains the values of p covariates in

the j-th group of study i, and u_j be an indicator variable for the treatment group (0 for control, 1 for treated).

The conventional logistic regression model can be written as

$$\text{logit}(\pi_{ij}) = \alpha_i + \beta^* u_j + \beta' \mathbf{z}_{ij} \tag{2.9}$$

where α_i is the intercept parameter in study i, β^* is the overall average value of log-odds ratio adjusted for covariates, and β is a column vector that contains the log-odds ratio per unit change for p covariates.

For the conventional logistic regression model, it is assumed that α_i is a fixed parameter and β^* is a fixed effects parameter. However, there is no allowance for effect heterogeneity in this model.

One method to incorporate effect heterogeneity into the models is the multi-level approach [3, 36, 37]. An appropriate model can be written as

$$\text{logit}(\pi_{ij}) = \alpha_i + \beta_i^* u_j + \beta' \mathbf{z}_{ij} \tag{2.10}$$

where α_i is the fixed intercept parameter in study i, and β_i^* is the log-odds ratio in study i which varies randomly across studies and has an independent normal distribution as

$$\beta_i^* \sim N(\beta^*, \tau^2) \tag{2.11}$$

Notice that in equations (2.9) and (2.10), it is assumed that α_i is a fixed parameter. However, the number of α_i parameters increases with the number of centers, leading to the Neyman-Scott problem [26].

An alternative multi-level model in which intercept parameter α_i are regarded as random rather than fixed is written as follows:

$$\text{logit}(\pi_{ij}) = \alpha_i + \beta_i^* u_j + \beta' \mathbf{z}_{ij} \tag{2.12}$$

where the α_i are independently distributed as

$$\alpha_i \sim N(\alpha, \tau_\alpha^2) \tag{2.13}$$

and where the β_i^* are independently distributed as

$$\beta_i^* \sim N(\beta^*, \tau_\beta^2). \tag{2.14}$$

and also $cov(\alpha_i, \beta_i^*) = \rho \tau_\alpha \tau_\beta$, where ρ is the correlation coefficient.

An alternative formula of equation (2.12) can be written as

$$\text{logit}(\pi_{ij}) = \alpha + \gamma_i + \beta^* u_j + \delta_i u_j + \beta' \mathbf{z}_{ij} \tag{2.15}$$

where the γ_i are independently distributed as

$$\gamma_i \sim N(0, \tau_\alpha^2) \tag{2.16}$$

and where the δ_i are independently distributed as

$$\delta_i \sim N(0, \tau_\beta^2). \tag{2.17}$$

This is the simplest and most conventional multi-level model together with random effects model for β_i^* which could lead to an extension of the univariate random effects model to a bivariate normal model. It is important to consider the covariance between the α_i and β_i^* in a bivariate normal model. If $cov(\alpha_i, \beta_i^*)$ is assumed to be zero, the between-study variance of the log-odds across control groups is equal to τ_α^2, while that across treatment groups is equal to $\tau_\beta^2 + \tau_\alpha^2$. The between-study variation in control groups is thereby forced to be less than or equal to the between-study variation in treatment groups. This assumption may not be appropriate for the general situation. When $cov(\alpha_i, \beta_i^*)$ is rather estimated than assumed to be zero, the variance-covariance matrix of the bivariate log-odds parameter estimates is modelled by a combination of the three parameters τ_α, τ_β and ρ. This non-zero covariance assumption allows the model to investigate a relation between baseline risk and treatment effect. However, this alternative model presents an extended complexity in the multi-level approach. This issue has been discussed in Turner *et al.* [37].

Notice that in equations (2.9), (2.10), and (2.12), two kinds of parameters occur. The first type is the parameter of interest, that is the coefficient of indicator variable β^* and the coefficients of covariates β. The second type is the nuisance parameter, that is the intercept parameter α_i. However, the nuisance parameter is not our major interest parameter, but it is necessary for a complete description of the model. Also, it complicates the inference. All these models, based on the binomial structure of data, suffer from dealing with the nuisance parameter α_i. Therefore, it appears attractive to investigate the profile likelihood approach which eliminates the nuisance parameter before dealing with the inference for parameter of interest, and thereby, keeping the dimensionality of the approach low.