8
Generalized Estimating Equations

8.1 Introduction

The main issue with full likelihood approaches for marginal models is the computational complexity they entail. The net benefit can be efficiency gain, but this comes at the cost of an increased risk for model misspecification. Of course, full likelihood methods clearly allow the researcher to calculate joint or union probabilities (such as in the POPS data, Section 7.10) and to make, perhaps subtle, inferences about the association structure. The latter was exemplified in Section 7.7.7. Chapter 7 also made it clear that there is no unambiguous choice for a full distributional specification. For example, while the Bahadur model (Section 7.2) is easy to generate, it suffers from severe restrictions on the parameter space. Other models may become unwieldy in computational terms when the number of repeated measures increases beyond a moderate number.

For all of these reasons, when we are mainly interested in first-order marginal mean parameters and pairwise interactions, a full likelihood procedure can be replaced by quasi-likelihood based methods (McCullagh and Nelder 1989). In quasi-likelihood, the mean response is expressed as a parametric function of covariates, and the variance is assumed to be a function of the mean up to possibly unknown scale parameters. Wedderburn (1974) first noted that likelihood and quasi-likelihood theories coincide for exponential families and that the quasi-likelihood estimating equations provide consistent estimates of the regression parameters $\boldsymbol{\beta}$ in any generalized linear model, even for choices of link and variance functions that do not correspond to exponential families.

For clustered and repeated data, Liang and Zeger (1986) proposed so-called *generalized estimating equations* (GEE or GEE1) which require only the correct specification of the univariate marginal distributions provided one is willing to adopt 'working' assumptions about the association structure. These models are a direct extension of basic quasi-likelihood theory from cross-sectional to repeated or otherwise correlated measurements. They estimate the parameters associated with the expected value of an individual's vector of binary responses and phrase the working assumptions about the association between pairs of outcomes in terms of marginal correlations. The method combines estimating equations for the regression parameters $\boldsymbol{\beta}$ with moment-based estimation for the correlation parameters entering the working assumptions.

Although Liang and Zeger's (1986) original proposal is undoubtedly the best known one, not in the least due to its implementation in a number of standard software packages, including the SAS procedure GENMOD, a number of alternative proposals have been made as well. Prentice (1988) extended their results to allow joint estimation of probabilities and pairwise correlations. Lipsitz, Laird, and Harrington (1991) modified the estimating equations of Prentice (1988) to allow modeling of the association through marginal odds ratios rather than marginal correlations. When adopting GEE1, one does not use information of the association structure to estimate the main effect parameters. As a result, it can be shown that GEE1 yields consistent main effect estimators, even when the association structure is misspecified. However, severe misspecification may affect the efficiency of the GEE1 estimators. In addition, GEE1 is less adequate when some scientific interest is placed on the association parameters.

Second-order extensions of these estimating equations (GEE2) that include the marginal pairwise association as well, have been studied by Zhao and Prentice (1990), using correlations, and Liang, Zeger, and Qaqish (1992), using odds ratios. They note that GEE2 is nearly fully efficient, as compared to a full likelihood approach, though bias may occur in the estimation of the main effect parameters when the association structure is misspecified. A variation to this theme, using conditional probability ideas, has been proposed by Carey, Zeger, and Diggle (1993). It is referred to as *alternating logistic regressions* and is studied in Section 8.6, alongside second-order GEE. In the same spirit, in Section 8.7, we will show how the hybrid model, combining elements of a marginal and a conditional formulation, introduced in Section 7.8, can be used as the basis for another GEE approach, maintaining computational ease (Fitzmaurice and Laird 1993, Fitzmaurice, Laird, and Rotnitzky 1993).

In Section 8.2, we present the basic GEE theory, while extensions and variations to the theme are the topic of Section 8.3. Some of these are then developed in sections to follow. Prentice's method is reviewed in Section 8.4. Second-order generalized estimating equations are introduced in Section 8.5. GEE based on odds ratios and alternating logistic regressions are discussed in Section 8.6. GEE based on the hybrid marginal-conditional formulation is given in Section 8.7, from which some of the other methods follow as special cases. An alternative approach, based on linearization, is given in Section 8.8. Next, three case studies are analyzed: the NTP data (Section 8.9), the heatshock study, a developmental toxicity study (Section 8.10), and the sports injuries trail (Section 8.11).

8.2 Standard GEE Theory

Let us adopt the regression notation, as outlined in Section 7.1.

In many longitudinal applications, inferences based on mean parameters (e.g., dose effect) are of primary interest. Specifying the full joint distribution would then be unnecessarily cumbersome. When inferences for the parameters in the mean model $\mathrm{E}(\boldsymbol{Y}_i)$ are based on classical maximum likelihood theory, full specification of the joint distribution for the vector $\boldsymbol{Y}_i$ of repeated measurements within each unit i is necessary. For discrete data, this implies specification of the first-order moments, as well as of all higher-order moments. For Gaussian data, full-model specification reduces to modeling the first- and second-order moments only, a situation much simpler than in the non-Gaussian case. However, even then can the choice of inappropriate covariance models seriously invalidate inferences for the mean structure.

A technique enabling the researcher to restrict modeling to the first moment only is based on so-called *generalized estimating equations* (GEEs, Liang and Zeger 1986, Zeger and Liang 1986, Diggle *et al* 2002). One way to approach the methodology is by making two observations. First, the score equations for a multivariate marginal normal model $\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, V_i)$ (Chapter 4; see also Verbeke and Molenberghs 2000, Chapter 5) are given by

$$\sum_{i=1}^{N} X_i'(A_i^{1/2}R_iA_i^{1/2})^{-1}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}) \quad = \quad \mathbf{0}, \tag{8.1}$$

in which the marginal covariance matrix V_i has been decomposed in the form

$$V_i = A_i^{1/2}R_iA_i^{1/2}, \tag{8.2}$$

with A_i the matrix with the marginal variances on the main diagonal and zeros elsewhere, and with R_i equal to the marginal correlation matrix. Decomposition (8.2) is a little unusual in this context, although it is easy to see what it would look like for such structures as, for example, compound symmetry and AR(1). A common decomposition is in terms of a marginalized hierarchical model: $V_i = \Sigma_i + Z_i D Z_i'$ (see Chapter 4). The motivation will become clear before too long.

As a second observation, the score equations to be solved when computing maximum likelihood estimates under a marginal generalized linear model, (Chapter 3) assuming independence of the responses within units (either ignoring the correlation in the repeated measures structure or when truly dealing with uncorrelated measures), takes the form

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'} (A_i^{1/2} I_{n_i} A_i^{1/2})^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu_i}) \quad = \quad \mathbf{0}, \tag{8.3}$$

where, again, A_i is again the diagonal matrix with the marginal variances along the main diagonal. The mean $\boldsymbol{\mu_i}$ follows from a vector of generalized linear models, specified for each component of the outcome vector. For example, a logistic regression can be specified for each of the components.

Note that expression (8.1) is of the form (8.3) but with the correlations between repeated measures taken into account. A key distinction between both is that A_i (and V_i as a whole) in (8.1) is usually parameterized by a set of parameters, functionally independent of the marginal regression parameters β. On the other hand, A_i in (8.3) is fully specified by the marginal regression parameters β, through the mean-variance link, common to most commonly used generalized linear models, as outlined in Chapter 3. Thus, when measurements are truly uncorrelated, one can restrict model specification to the marginal mean function, as the variance will automatically follow, perhaps up to an overdispersion parameter. These observations are crucial in what follows.

A seemingly straightforward extension of (8.3) that would account for the correlation structure is

$$S(\boldsymbol{\beta}) \quad = \quad \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'} (A_i^{1/2} R_i A_i^{1/2})^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu_i}) \ = \ \mathbf{0}, \tag{8.4}$$

obtained from replacing the identity matrix I_{n_i} by a correlation matrix R_i. Now, even though (8.4) seems to follow from combining the most general aspects of (8.1) with those of (8.3), matters are not this simple. Although $A_i = A_i(\boldsymbol{\beta})$ follows directly from the marginal mean model, $\boldsymbol{\beta}$ commonly contains absolutely no information about R_i, whence R_i is to be parameterized by an additional parameter vector: $R_i = R_i(\boldsymbol{\alpha})$. Thus, while the first moment completely specified the second (and higher order) moments in the univariate case, this is only partially so in the correlated data setting, the variances are still specified by the marginal means, but the correlations are not. This sets the repeated measures and other correlated data settings fundamentally apart from their univariate counterpart. Simply adding model components (and hence score equations) for the correlation parameters does not solve the problem. To see this, recall that we wanted to restrict model specification to the first moments only, but are faced with the second moments. If we would model the second moments, we would have to address the third and fourth moments as well. Eventually, a full specification of the joint distribution would be obtained, precisely what we wanted to avoid. These observations also underscore the difference between the Gaussian and non-Gaussian settings, as (8.1) is sufficient for the Gaussian case: given the first- and second-order moments, *and assuming multivariate normality*, the joint distribution is fully specified. Thus, in summary, it is too simple to state that the repeated non-Gaussian case is simply a combination of elements from the Gaussian repeated measures case with elements from univariate generalized linear models.

Liang and Zeger (1986) provide a nice way out of this apparent gridlock. While still acknowledging the need for $R_i(\boldsymbol{\alpha})$ in V_i and (8.4), they allowed the modeler to *specify an incorrect structure* or so-called working correlation matrix. Using method of moments concepts, they showed that, when the marginal mean $\boldsymbol{\mu_i}$ has been correctly specified as $h(\boldsymbol{\mu_i}) = X_i\boldsymbol{\beta}$ and when mild regularity conditions hold, the estimator $\widehat{\boldsymbol{\beta}}$ obtained from solving (8.4) is consistent and asymptotically normally distributed with mean $\boldsymbol{\beta}$ and asymptotic variance-covariance matrix covariance matrix

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = I_0^{-1} I_1 I_0^{-1}, \tag{8.5}$$

where

$$I_0 = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}, \tag{8.6}$$

$$I_1 = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}'}{\partial \boldsymbol{\beta}} V_i^{-1} \mathrm{Var}(\boldsymbol{Y}_i) V_i^{-1} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}. \tag{8.7}$$

Consistent estimates can be obtained by replacing all unknown quantities in (8.5) by consistent estimates. Apart from a working correlation matrix, it is possible to incorporate an overdispersion parameter as well, whence $A_i^{1/2} R_i A_i^{1/2}$ in (8.4) would be replaced by

$$V_i = V_i(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi) = \phi A_i(\boldsymbol{\beta})^{1/2} R_i(\boldsymbol{\alpha}) A_i(\boldsymbol{\beta})^{1/2}, \tag{8.8}$$

ϕ being the additional overdispersion parameter.

Observe that, when R_i would be correctly specified, $\mathrm{Var}(\boldsymbol{Y}_i) = V_i$ in (8.7) and then $I_1 = I_0$. As a result, (8.5) would reduce to I_0^{-1}, corresponding to full likelihood, i.e., when the first and second moment assumptions would be correct. Thus, (8.5) reduces to full likelihood when the working correlation structure is correctly specified but generally differs from it. There is no price to pay in terms of consistency of asymptotic normality, but there may be efficiency loss when the working correlation structure differs strongly from the true underlying structure.

Thus, whether or not the working correlation structure is correct, point estimates and standard errors based on (8.5) are asymptotically correct. Such standard errors were called 'robust' by Liang and Zeger (1986), while the variance estimator (8.5) is sometimes referred to as the 'sandwich estimator,' for obvious reasons. In the meantime, the terms 'empirically corrected' variance and standard errors found their way to common use, to avoid confusion with methods from robust statistics. In contrast, I_0^{-1} was initially referred to as the 'naive' estimator, but currently the more neutral '(purely) model based' estimator is more common. Note that estimates and standard errors resulting from GEE are often reported in the format 'estimate (empirically corrected standard error; model-based standard error),' in line with the convention used by Liang and Zeger (1986) in their original article. Unless when used for didactical purposes, or when the model-based standard error would be of some scientific interest, this is not necessary. The empirically corrected standard error is the one to be used, the other one generally incorrect. At best, it can be seen as an indication of the 'distance' between the working assumptions for the correlation and the true structure. When both standard errors are far apart, this can be seen as an indication for a poor choice of working assumptions. Once again, a poor working assumption is not wrong, but may hamper efficiency and, when at all possible, it may be of interest to then try alternative working assumptions. The term 'empirical correction' stems from the fact that the data $\boldsymbol{Y}_i$ are used in I_1, not directly following from the likelihood function.

Two further specifications are needed before GEE is operational: $\text{Var}(\boldsymbol{Y}_i)$ on the one hand and $R_i(\boldsymbol{\alpha})$, with in particular estimation of $\boldsymbol{\alpha}$, on the other hand. Full modeling will not be an option, since we would then be forced to do what we want to avoid. First, modeling $\text{Var}(\boldsymbol{Y}_i)$ would imply modeling all components of (8.8) correctly, which we wanted to avoid. Second, fully modeling $R_i(\alpha)$ would, once again, bring in the need to address third and fourth order moments, which we wanted to avoid as well. Let us discuss the pragmatic solutions found to both of these issues in turn.

Turning attention to the empirical covariance of the outcome vector, $\text{Var}(\boldsymbol{Y}_i)$ in (8.5) is typically replaced by

$$(\boldsymbol{y}_i - \boldsymbol{\mu_i})(\boldsymbol{y}_i - \boldsymbol{\mu_i})'. \tag{8.9}$$

Although this may seem a natural choice at first sight, also because it is an unbiased estimate at the sole condition that the mean is correctly specified, it is perhaps less so when one realizes it has rank at most one! However, while a poor estimator for $\text{Var}(\boldsymbol{Y}_i)$, it is adequate to estimate I_1 and ultimately (8.5), given the summation over N units in I_1. The deficient rank poses no problems since no inversion takes place within I_1 and, as an extra safety, I_1 does not need to be inverted. It has been reported

TABLE 8.1. *Common choices for the working correlation assumptions in standard generalized estimating equations and moment-based estimators thereof.*

Structure	$\mathrm{Corr}(Y_{ij}, Y_{ik})$	Estimator
Independence	0	—
Exchangeable	α	$\widehat{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i(n_i-1)}\sum_{j\neq k} e_{ij}e_{ik}$
AR(1)	$\alpha^{\lvert j-k \rvert}$	$\widehat{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i-1}\sum_{j\leq n_i-1} e_{ij}e_{i,j+1}$
Unstructured	α_{jk}	$\widehat{\alpha}_{jk} = \frac{1}{N}\sum_{i=1}^{N} e_{ij}e_{ik}$

that replacing (8.9) by $(\boldsymbol{y}_i - \widehat{\boldsymbol{\mu_i}})(\boldsymbol{y}_i - \widehat{\boldsymbol{\mu_i}})'$ may induce some bias into the procedure (Crowder 1995).

Next, regarding the working correlation parameters $\boldsymbol{\alpha}$ and the overdispersion parameter ϕ, Liang and Zeger (1986) proposed moment-based estimates. To this end, first define residuals

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}} \tag{8.10}$$

in line with (7.7), introduced for the Bahadur model. Note that $e_{ij} = e_{ij}(\boldsymbol{\beta})$ through $\mu_{ij} = \mu_{ij}(\boldsymbol{\beta})$ and therefore also through $v(\mu_{ij})$, the variance at time j, and hence the jth diagonal element of A_i. We still assume the variance is decomposed as (8.8). Common choices for the working assumptions are presented in Table 8.1. Similarly, the dispersion parameter can be estimated by

$$\widehat{\phi} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i} e_{ij}^2. \tag{8.11}$$

Note that the independence structure brings about no additional parameters $\boldsymbol{\alpha}$ and hence, when there is no overdispersion, parameter estimates $\widehat{\boldsymbol{\beta}}$ will not differ from those obtained from logistic regression. Even then, the asymptotic variance covariance matrix, obtained from (8.5), and hence the standard errors, will differ from the ones obtained with logistic regression, the latter stemming from the model-based but incorrect I_1^{-1}. Independence and exchangeable working assumptions can be used in virtually all applications, whether longitudinal, clustered, multivariate, or otherwise correlated. Clearly, AR(1) and unstructured are less relevant for clustered data, longitudinal studies with unequally spaced measurements and/or sequences with differing lengths, etc. However, even though it seems less advisable to use such structures in cases where they are not supported by the study's design, it is strictly speaking not a mistake as, once again, working assumptions *are allowed to be wrong!* Note that the AR(1) parameter is estimated using adjacent pairs of measurements only, in contrast to the exchangeable correlation, for which all pairs within a sequence are employed. This is not wrong, but may be somewhat inefficient as, for example, pairs two occasions apart contribute information to α^2 and hence to α. Of course, incorporating such information clutters the moment-based estimators and most implementations still follow Table 8.1, as in Liang and Zeger (1986).

Now, (8.4), conceived to estimate $\boldsymbol{\beta}$, are in need of $\boldsymbol{\alpha}$ and ϕ, while the moment-based estimates for $\boldsymbol{\alpha}$ (Table 8.1) and expression (8.11) for ϕ depend on $\boldsymbol{\beta}$. This circularity is the final stumbling block in the way, but can be circumvented by an iterative procedure. The standard iterative procedure to fit GEE, based on Liang and Zeger (1986), is then as follows:

1. Compute initial estimates for $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(0)}$ say, using a univariate GLM, i.e., assuming independence or, in other words, using conventional logistic regression.

2. Compute the quantities needed in the estimating equation, i.e., Pearson residuals e_{ij} from (8.10), $\boldsymbol{\alpha}$ from Table 8.1, and ϕ from (8.11).

3. Based on these, $R_i(\boldsymbol{\alpha})$ can be computed, as well as V_i from (8.8).

4. Then, given the current estimate of $\boldsymbol{\beta}$ after t iterations, $\boldsymbol{\beta}^{(t)}$ say, update the estimate for $\boldsymbol{\beta}$:

$$\begin{aligned} \boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} - \left[\sum_{i=1}^{N}\left(\frac{\partial\boldsymbol{\mu_i}}{\partial\boldsymbol{\beta}'}\right)V_i^{-1}\left(\frac{\partial\boldsymbol{\mu_i}}{\partial\boldsymbol{\beta}'}\right)'\right]^{-1} \\ &\quad \times\left[\sum_{i=1}^{N}\left(\frac{\partial\boldsymbol{\mu_i}}{\partial\boldsymbol{\beta}'}\right)V_i^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_i)\right]. \end{aligned} \tag{8.12}$$

The second, third, and fourth steps need to be iterated until convergence.

In conclusion, we have a method at our disposition to obtain valid inferences about a marginal regression model for repeated and otherwise clustered data, without the need to fully specify the joint distribution of the outcomes. This is most useful when the outcomes are of a non-Gaussian nature, as the linear mixed-effects model provides a flexible framework in the latter case (Chapter 4). However, it would still be possible to apply robust inference in the Gaussian case as well (Verbeke and Molenberghs 2000, Section 6.2.4), in case interest is confined to the marginal regression parameters $\boldsymbol{\beta}$, and there is doubt about a correct specification of the covariance structure and/or the random-effects structure. Indeed, the usual estimate for $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} X_i'W_iX_i\right)^{-1}\sum_{i=1}^{N} X_i'W_i\boldsymbol{Y_i},$$

with $\boldsymbol{\alpha}$ replaced by its ML or REML estimate. Conditional on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\beta}}$ has mean

$$\begin{aligned} E\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right] &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \sum_{i=1}^{N} X_i' W_i E(\boldsymbol{Y_i}) \\ &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \sum_{i=1}^{N} X_i' W_i X_i \boldsymbol{\beta} \\ &= \boldsymbol{\beta}, \end{aligned}$$

provided that $E(\boldsymbol{Y_i}) = X_i\boldsymbol{\beta}$. Hence, for $\widehat{\boldsymbol{\beta}}$ to be unbiased, it is sufficient that the mean of the response is correctly specified. Conditional on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\beta}}$ has covariance

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\beta}}) &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \\ &\quad \times \left(\sum_{i=1}^{N} X_i' W_i \mathrm{Var}(\boldsymbol{Y}_i) W_i X_i\right) \qquad (8.13) \\ &\quad \times \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \\ &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1}. \qquad (8.14) \end{aligned}$$

Note that (8.14) assumes that the covariance matrix $\mathrm{Var}(\boldsymbol{Y}_i)$ is correctly modeled as $V_i = Z_i D Z_i' + \Sigma_i$, which then again plays the role of the purely model-based estimate. The empirically corrected estimate for $\mathrm{Var}(\widehat{\boldsymbol{\beta}})$, which does not assume the covariance matrix to be correctly specified is obtained from replacing $\mathrm{Var}(\boldsymbol{Y}_i)$ in (8.13) by

$$\left(\boldsymbol{Y}_i - X_i\widehat{\boldsymbol{\beta}}\right)\left(\boldsymbol{Y}_i - X_i\widehat{\boldsymbol{\beta}}\right)', \qquad (8.15)$$

rather than V_i. The sole condition for (8.15) to be unbiased for $\mathrm{Var}(\boldsymbol{Y}_i)$ is that the mean is again correctly specified.

In spite of this potential use for Gaussian outcomes, GEE is most commonly used for non-Gaussian measurement sequences. The need is avoided to specify third- and higher-order moments or, more precisely, third- and higher-order correlations, and two-way correlations are allowed to be misspecified. Should they be correctly specified, and should a set of appropriate third- and higher-order correlations be chosen, together with marginal logit links for binary outcomes, then the Bahadur model (Section 7.2) would follow. Thus, standard GEE can be seen a moment-based version of the Bahadur model. After choosing the marginal response functions, there is always at least one, trivial, Bahadur model that corresponds to the estimating equations, found by setting all correlations to zero, i.e., independence. In general, the working correlations, found upon convergence of GEE, may not necessarily correspond to a valid joint probability mass function, given the severe constraints on the Bahadur model (Section 7.2.2). This need not be a drawback, as the working correlations are merely a device to provide consistent and asymptotically normal point estimates for the marginal regression parameters and, if well chosen, also reasonably efficient. They should not be made a part of formal inference.

The previous statement implies that, strictly speaking, the following two questions should remain unanswered or at least approached cautiously:

- Are particular working correlation values large, moderate, or small?
- Among a set of working correlation matrices under correlation, which one is best?

The first question is a natural one to ask. However, an answer does not come easily, since ordinarily no standard errors are given alongside the working correlations, and neither should they. Indeed, as stated above, they are only devices to support estimation of the regression parameters, with a status almost below the one of nuisance parameter. One can interpret them, informally and with great caution, when the empirically corrected and model-based standard errors are close, for then there usually is good evidence that the working correlation structure has been chosen in line with the true structure (Drum and McCullagh 1993, who present a critical view on the methodology). This may be the case, in particular, when the working correlation structure is fairly general, such as 'unstructured' in the cases of balanced data (with corresponding measurements for different subjects taken at the same time or approximately the same time). Of course, an unstructured covariance matrix is no guarantee for a correct specification since the covariance structure may further depend, for example, on certain covariates.

Turning to the second question, it ought to be clear that there are no formal model comparison tools for the correlation parameters. Because there are no standard errors, Wald-type tests are not possible, and also likelihood-ratio and score tests are not easy to use. Although some model comparison and goodness-of-fit tools have been proposed (Rotnitzky and Jewell 1990), they are for the mean model and not for the association structure, as they should be as, once again, the association is mere nuisance in the GEE philosophy. The worst possible, in fact unscientific, approach that can be taken is to base one's choice for working assumptions on the outcome (significance) for the regression parameters.

Thus, in conclusion, the working correlation structure ought to be left alone or at most used in a very informal way. It is best to specify a single working correlation structure upfront when the need exists to specify a primary analysis. Perhaps some others can be used by way of sensitivity analysis for the regression parameters. It seems best to specify the working correlation structure in agreement with the design of the study (counterexamples being exchangeability for multivariate outcomes or AR(1) for unequally spaced longitudinal measurements), and as general as the data support. The latter is usually a function of the number of subjects in a study, as well as the number of measurements per subject.

When GEE is deemed unsatisfactory in the sense that there is some scientific interest is the association structure, then one should turn to some of the extensions of GEE, reviewed in Section 8.3, in particular to GEE2 (Section 8.5), GEE methods combining a marginal and conditional specification (Section 8.7), or even to alternating logistic regressions (Section 8.6) or pseudo-likelihood (Chapter 9).

Some theoretical considerations regarding problems that may occur with GEE are presented in Crowder (1995), Sutradhar and Das (1999), and Vonesh, Wang, and Majumdar (2001).

8.3 Alternative GEE Methods

In the previous section, standard GEE, as introduced by Liang and Zeger (1986), was discussed. A number of alternatives have been proposed. Prentice (1988) replaced the moment-based estimation for the working correlation parameters by a second set of estimating equations. By making the working assumption that both sets are independent, computational complexity is avoided and, again, the correlation model need not be correctly specified for the marginal regression parameters to be consistent and asymptotically normal. Prentice's method is discussed in Section 8.4. As soon as the two sets of estimating equations are assumed to be correlated, one obtains GEE2, in the sense that the first and second moments are then fully modeled, with working assumptions made about the third and fourth order moments. This method, which is one step up from Prentice's method, is discussed in Section 8.5.

Lipsitz, Laird, and Harrington (1991) adapted Prentice's method to switch from marginal correlation coefficients to marginal odds ratios. These are but two of the association choices from Table 7.3. Thus, while standard GEE and Prentice's method can be seen as derived from the Bahadur model (Section 7.2), the method by Lipsitz, Laird, and Harrington (1991) derives from the multivariate Dale model (Sections 7.3 and 7.7, see also Chapter 6). Of course, GEE2 can be formulated not only with correlations but also based on odds ratios (Liang, Zeger, and Qaqish 1992). GEE with odds ratios (Lipsitz, Laird, and Harrington 1991, Liang, Zeger, and Qaqish 1992), and the link to alternating logistic regression (Carey, Zeger, and Diggle 1993), is discussed in Section 8.6.

Another method, close in spirit to GEE as it also derives from quasi-likelihood ideas, is based on linearizing the link function. It is presented in Section 8.8. A nice feature is that it can be fitted using the SAS procedure GLIMMIX. The method is in fact a special case of a more general approach, that allows the inclusion of random effects into a generalized linear model (Chapter 14).

8.4 Prentice's GEE Method

Prentice (1988) amended the basic GEE or GEE1 of Liang and Zeger (1986), described in Section 8.2. This method allows for estimation of both parameters vectors, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, in the marginal response model and the pair-wise correlations, respectively. The key difference with the original GEE is that for both sets of parameters, estimating equations are proposed. Thus, this GEE estimator for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ may be defined as a solution to:

$$\sum_{i=1}^{N} \boldsymbol{D}_i' V_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) \quad = \quad \mathbf{0}, \qquad (8.16)$$

$$\sum_{i=1}^{N} \boldsymbol{E}_i' W_i^{-1}(\boldsymbol{Z}_i - \boldsymbol{\delta}_i) \quad = \quad \mathbf{0}, \qquad (8.17)$$

where $\boldsymbol{Z}_i$ consists of components, doubly indexed by (j_1, j_2) and taking the form:

$$Z_{ij_1j_2} = \frac{(Y_{ij_1} - \mu_{ij_1})(Y_{ij_2} - \mu_{ij_2})}{\sqrt{\mu_{ij_1}(1-\mu_{ij_1})\mu_{ij_2}(1-\mu_{ij_2})}}.$$

The terms carry information about the correlation between measures Y_{ij_1} and Y_{ij_2} on the same subject. In summary,

$$\boldsymbol{Z}_i = (Z_{i12}, Z_{i13}, \ldots, Z_{i,n_i-1,n_i}). \qquad (8.18)$$

Further, $\delta_{ij_1j_2} = E(Z_{ij_1j_2})$,

$$D_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}, \qquad E_i = \frac{\partial \boldsymbol{\delta}_i}{\partial \boldsymbol{\alpha}},$$

V_i is the variance-covariance matrix of $\boldsymbol{Y}_i$, and W_i is the working variance-covariance matrix of $\boldsymbol{Z}_i$. Strictly speaking, V_i is no working covariance matrix, since the second moments are specified by (8.17). In contrast, W_i does contain working assumptions, usually being that the third- and fourth-order correlations, defined by (7.8), are equal to zero. We will return to these in Section 8.7.

The assumption is made that (8.16) and (8.17) are independent. This would entail a price in terms of efficiency, but has the advantage that, just as in Section 8.2, misspecifying the correlation structure does not hamper consistency and asymptotic normality of the marginal regression parameters. Each set of parameters comes with precision estimates, whence formal inference is possible about the set of parameters for one is prepared to believe the equations have been correctly specified. This could be (8.16), (8.17), or both. The option to make formal inferences about the correlation parameters is a net increase of capabilities over standard GEE1.

The joint asymptotic distribution of $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{N}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is Gaussian with mean zero and with variance-covariance matrix consistently estimated by

$$N \cdot \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}' \\ \boldsymbol{0} & \boldsymbol{C} \end{pmatrix},$$

where

$$\boldsymbol{A} = \left(\sum_{i=1}^{N} \boldsymbol{D}_i' V_i^{-1} \boldsymbol{D}_i \right)^{-1}, \tag{8.19}$$

$$\boldsymbol{B} = \left(\sum_{i=1}^{N} \boldsymbol{E}_i' W_i^{-1} \boldsymbol{E}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{E}_i' W_i^{-1} \frac{\partial \boldsymbol{Z}_i}{\partial \boldsymbol{\beta}} \right) \tag{8.20}$$

$$\times \left(\sum_{i=1}^{N} \boldsymbol{D}_i' V_i^{-1} \boldsymbol{D}_i \right)^{-1}, \tag{8.21}$$

$$\boldsymbol{C} = \left(\sum_{i=1}^{N} \boldsymbol{E}_i' W_i^{-1} \boldsymbol{E}_i \right)^{-1}, \tag{8.22}$$

$$\Lambda_{11} = \sum_{i=1}^{N} \boldsymbol{D}_i' V_i^{-1} \mathrm{Cov}(\boldsymbol{Y}_i) V_i^{-1} \boldsymbol{D}_i, \tag{8.23}$$

$$\Lambda_{12} = \sum_{i=1}^{N} \boldsymbol{D}_i' V_i^{-1} \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Z}_i) W_i^{-1} \boldsymbol{E}_i, \tag{8.24}$$

$$\Lambda_{21} = \Lambda_{12}, \tag{8.25}$$

$$\Lambda_{22} = \sum_{i=1}^{N} \boldsymbol{E}_i' W_i^{-1} \mathrm{Cov}(\boldsymbol{Z}_i) W_i^{-1} \boldsymbol{E}_i, \tag{8.26}$$

and $\mathrm{Var}(\boldsymbol{Y}_i)$, $\mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Z}_i)$, and $\mathrm{Var}(\boldsymbol{Z}_i)$ are estimated by the quantities

$$(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)', (\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Z}_i - \boldsymbol{\delta}_i)', (\boldsymbol{Z}_i - \boldsymbol{\delta}_i)(\boldsymbol{Z}_i - \boldsymbol{\delta}_i)',$$

respectively, in analogy with GEE1. One may wonder why there is no need to go back and forth between solving the estimating equations and moment-based estimation, as in Section 8.2. In this case, this would mean solving

(8.16) and (8.17), and then switch to moment based estimation for higher moments. However, as stated before, one typically assumes the third and fourth moments are zero. One could call these 'higher-order independence' working assumption, obviating the need for additional parameters.

The above model has a close resemblance with the Bahadur model, as it is based on its first to fourth moments. Williamson, Lipsitz, and Kim (1997) wrote a SAS macro for Prentice's method.

The updating method for Prentice's GEE iterates between solving each of the equations (8.16) and (8.17).

8.5 Second-order Generalized Estimating Equations (GEE2)

Second-order GEE have been proposed by Zhao and Prentice (1990), using correlations, and by Liang, Zeger, and Qaqish (1992), using odds ratios. They are a simple extension of Prentice's (1988) method, described in Section 8.4, by combining the outcome vector $\boldsymbol{Y}_i$ and the pairwise cross-products, $\boldsymbol{Z}_i$, as in (8.18), into a single outcome vector:

$$\boldsymbol{W}_i = (\boldsymbol{Y}_i', \boldsymbol{Z}_i')'. \tag{8.27}$$

The vector $\boldsymbol{W}_i$ has $n_i + \binom{n_i}{2}$ components. Further, let

$$\boldsymbol{\Theta}_i = (\boldsymbol{\mu}_i', \boldsymbol{\delta}_i')',$$

the corresponding mean vector, obtained by assembling the means from (8.16) and (8.17). Assuming $\boldsymbol{\delta}_i$ depends on a vector of regression parameters $\boldsymbol{\beta}$, which now combines the $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ from Section 8.4, the vector $\boldsymbol{\beta}$ can be estimated by solving the second-order generalized estimating equations:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{N} D_i' V_i^{-1} [\boldsymbol{W}_i - E(\boldsymbol{W}_i)] = \mathbf{0}, \tag{8.28}$$

where

$$D_i = \frac{\partial \boldsymbol{\Theta}_i'}{\partial \boldsymbol{\beta}}.$$

As usual, $V_i = \text{Cov}(\boldsymbol{W}_i)$. Calculation of all matrices involved is straightforward with the exception of the covariance matrix V_i, which contains third- and fourth-order probabilities. Again, as in Section 8.4, the three-way and higher order correlations are set equal to zero. As before, the parameter estimates $\widehat{\boldsymbol{\beta}}$ can then be calculated using, for example, a Fisher scoring algorithm. Provided the first- and second-order models have been correctly specified, $\widehat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ and has an asymptotic multivariate normal distribution with mean vector $\boldsymbol{\beta}$ and variance-covariance matrix consistently estimated by:

$$\widehat{V}(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{N} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{D}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{U}_i(\widehat{\boldsymbol{\beta}}) \boldsymbol{U}_i(\widehat{\boldsymbol{\beta}})'\right) \times \left(\sum_{i=1}^{N} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{D}_i\right)^{-1},$$

the usual sandwich estimator.

In principle, there is no reason why one should stop at GEE2. Higher-order GEE is perfectly conceivable. When, moments 1 up to K would be modeled, working assumptions of order $K+1$ up to $2K$ would be needed. Obviously, this will become increasingly cumbersome, not only algebraically, also regarding implementation and computation time. As the order increases, the relative gain will also decrease, as less and less information would be contained in higher moments. When K becomes equal to the length of the response vector $\boldsymbol{Y}_i$, full likelihood is recovered and our specification, carried through to order n_i, would produce the Bahadur model. When higher orders are of interest, this is usually in situations where the joint probabilities need to be calculated and then full likelihood effectively is the only option. Thus, most commonly encountered are GEE1 and GEE2 on the one hand, and then full likelihood on the other hand.

8.6 GEE with Odds Ratios and Alternating Logistic Regression

The GEE versions discussed in Sections 8.2, 8.4, and 8.5 all used correlation as a measure to capture association, either as moment estimated working assumptions, or as part of the estimating equations. Thus, as indicated earlier, all can be seen as deriving from the Bahadur model. The advantage of correlations is that the estimating equations, such as (8.4), include the covariance matrix V_i as in (8.5), and (working) correlation parameters can be used in a particularly straightforward fashion to compose the matrix $R_i(\boldsymbol{\alpha})$. However, many authors have stated that the odds ratio is a particularly straightforward measure to capture association between binary or categorical outcomes (Molenberghs and Lesaffre 1994, 1999, Fitzmaurice, Laird, and Ware 2004, p. 298, see also Chapters 6 and 7). In the context of GEE, the same observation has been made. Lipsitz, Laird, and Harrington (1991) considered GEE1 for binary data with odds ratios, while Liang, Zeger, and Qaqish (1992) did the same for GEE2. The Bahadur-based correlation, expressed as (7.5) and leading to bivariate joint probabilities (7.6), needs to be replaced by the Dale-based odds ratio (7.39), leading to bivariate joint probabilities (7.40). Still focusing on binary outcomes and based on the bivariate probabilities, calculation of V_i in (8.5) is straightforward and follows from observing that

$$\begin{array}{rcccl} V_{i,jj} & = & \mathrm{Var}(Y_{ij}) & = & \mu_{ij}(1-\mu_{ij}), \\ V_{i,j_1,j_2} & = & \mathrm{Cov}(Y_{ij_1}, Y_{ij_2}) & = & \mu_{ij_1j_2} - \mu_{ij_1}\mu_{ij_2}. \end{array}$$

Note that the expectation of a component of $\boldsymbol{Z}_i$, $Z_{ij_1j_2}$ say, equals $\mu_{ij_1j_2}$, the probability of a success at occasions j_1 and j_2 at the same time. Assuming a model for the pairwise odds ratios as in (7.39), and working assumptions for the third- and fourth-order log odds ratios (usually by setting them equal to zero), the model specification is complete. Lipsitz, Laird, and Harrington (1991) assumed, for simplicity, W_i in (8.17) to be diagonal, avoiding working assumptions about the third and fourth order; it even avoids calculating the third- and fourth-order probabilities altogether. It is then simple to solve (8.16) and (8.17) with the vector $\boldsymbol{Y}_i$ still equal to the response vector, and with $\boldsymbol{Z}_i$ in (8.18) changed to

$$\boldsymbol{Z}_i = (Y_{i1}Y_{i2}, Y_{i1}Y_{i3}, \ldots, Y_{i,n_i-1}Y_{in_i}). \tag{8.29}$$

The same principles as outlined above can be applied to second-order GEE (8.28). This idea was followed by Liang, Zeger, and Qaqish (1992). While they set the third- and fourth-order log odds ratios equal to zero, obviating the need to invoke additional (moment-based) estimation, they still needed to compute third- and fourth-order probabilities for GEE2, following one of the methods associated with the Dale model, e.g., using the IPF algorithm or the Plackett polynomials (Sections 7.4 and 7.7). This can be computationally less than straightforward, but luckily there is another alternative, termed *alternating logistic regressions* (ALR) and proposed by Carey, Zeger, and Diggle (1993). The method is different from all of the GEE methods considered so far, but has communality with both GEE1 and GEE2 based on odds ratios. In particular, it is almost as efficient as GEE2, and shares the computational ease of conventional GEE1.

Let us first introduce the method, and then provide some further perspective on its advantages. Let μ_{ij} be as before, described by

$$\mathrm{logit} P(Y_{ij} = 1) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}, \tag{8.30}$$

and let $\alpha_{ij_1j_2} = \ln(\psi_{ij_1j_2})$ be the marginal log odds ratio. Then,

$$\begin{array}{rl} & \mathrm{logit} P(Y_{ij_1} = 1 | Y_{ij_2} = y_{ij_2}) \\ = & \ln\left(\dfrac{\mu_{ij_1} - \mu_{ij_1j_2}}{1 - \mu_{ij_1} - \mu_{ij_2} + \mu_{ij_1j_2}}\right) + \alpha_{ij_1j_2} y_{ij_2}. \end{array} \tag{8.31}$$

The marginal logistic regression (8.30) is in line with the Bahadur model, the Dale model with logistic margins, and all of the GEEs discussed in this chapter. However, rather than further specifying the models by additional marginal description of the pairwise association, a logistic model for an outcome *conditional upon another outcome* is presented, which derives trivially from the expression for the log odds ratio. Logistic regression (8.31) is a little unconventional in the sense that instead of an intercept, there is an offset, i.e., a constant term free of unknown parameters, given the mean model. An example of an offset can be found in Section 3.7. The principle of ALR is to iterate between solving (8.30) and (8.31). Iteration is indeed required because solving (8.30) requires the covariance matrix V_i of $\boldsymbol{Y}_i - \boldsymbol{\mu}_i$, which depends on both β and α, while also (8.31) depends on both. The updating problem can be phrased in terms of simultaneously solving two sets of estimating equations, the first one being exactly equal in form to (8.16), the second one being

$$\sum_{i=1}^{N} \widetilde{\boldsymbol{E}}_i' \widetilde{W}_i^{-1} \boldsymbol{R}_i = \boldsymbol{0}, \tag{8.32}$$

where

$$\widetilde{\boldsymbol{E}}_i = \frac{\partial \boldsymbol{\zeta}_i}{\partial \boldsymbol{\alpha}_i'},$$

$\boldsymbol{\zeta}_i$ is a vector with elements $\zeta_{ij_1j_2} = P(Y_{ij_1} = 1 | Y_{ij_2} = y_{ij_2})$, $\widetilde{W}_i$ is a diagonal matrix with elements $\zeta_{ij_1j_2}(1 - \zeta_{ij_1j_2})$, and $\boldsymbol{R}_i$ a vector with elements $Y_{ij_1j_2} - \zeta_{ij_1j_2}$.

Note that ALR extends beyond classical GEE, in the sense that precision estimates follow for both the $\boldsymbol{\beta}$ and the $\boldsymbol{\alpha}$ parameters. However, unlike with GEE2, and even with Prentice's (1988) and Prentice and Zhao (1991) GEE, no working assumptions about the third- and fourth-order odds ratios are required. Thanks to the clever combination of a marginal and a conditional specification, addressing the third and fourth moments is avoided all together, which is strictly different from setting them equal to zero.

In (8.31), arbitrary structures for the log odds ratio parameters $\alpha_{ij_2j_2}$ can be assumed. The odds ratio equivalent of exchangeability would set them all equal to the same constant α. When measurements are taken at fixed time points, an unstructured specification is possible. Further, when measurements are equally spaced, banded structures or other equivalents of autoregressive correlation structures can be entertained.

ALR has been implemented in the SAS procedure GENMOD. More detail is given in Section 10.4.

As was seen here, a combination of marginal and conditional specification can be advantageous. ALR is not the only instance to confirm this. In the next section, a family of hybrid marginal and conditional model specifications is considered.

8.7 GEE2 Based on a Hybrid Marginal-conditional Model

In the previous section, alternating logistic regression combined marginal and conditional aspects of model specification. The hybrid model, combining marginal and conditional aspects, presented in Section 7.8 can be used as a basis for GEE just as easily as the Bahadur and Dale models studied before.

A set of GEE2, proposed also by Heagerty and Zeger (1996), can be derived by specifying only the first and second moments that derive from (7.50):

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right) \boldsymbol{M}_i^{-1}(\boldsymbol{v}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{8.33}$$

with notation as in Section 7.8.1. Observe that these score equations assume the same form for any fixed value of $\boldsymbol{\Omega}_i$, with $\boldsymbol{\Omega}_i = \mathbf{0}$ as a special but important case. However, this leaves $\boldsymbol{M}_i$ partly unspecified. A standard procedure is to replace it by a working covariance matrix, depending on a set of (nuisance) parameters $\boldsymbol{\alpha}$. Heagerty and Zeger (1996) advocated setting the higher order conditional association parameters equal to zero (or, more generally, to a fixed constant). This particular set of GEE2 does not require estimation of extra parameters, a property shared with the GEE2 methods described in Section 8.5 and 8.6. Expression (8.33) can also be seen as the score equations for the likelihood specified by the following member of the quadratic exponential family of Zhao and Prentice (1990):

$$f(y_i|\boldsymbol{\Psi}_i) = \exp\left\{ \boldsymbol{\Psi}_i' \boldsymbol{v}_i - A(\boldsymbol{\Psi}_i) \right\}. \tag{8.34}$$

Another, slightly different set of GEE2, which also does not require estimation of nuisance parameters, is found by setting all three and higher order *marginal* log odds ratios equal to zero, in agreement with GEE2 in Section 8.6 and Liang, Zeger, and Qaqish (1992).

Computing the covariance $\boldsymbol{M}_i$ in (8.33) involves the third and fourth order probabilities. With conditional constraints, they are easily computed using the IPF algorithm, as outlined in Section 7.8.1. To proceed with marginal working assumptions, we first need to define the three- and four-way marginal odds ratios. They can also be introduced using conditional lower order odds ratios. If $\psi_{ij_1j_2|j_3}(y)$ is the conditional odds ratio of outcomes Y_{ij_1} and Y_{ij_2}, given $Y_{ij_3} = y$, then

$$\psi_{ij_1j_2j_3} = \frac{\psi_{ij_1j_2|j_2}(1)}{\psi_{ij_1j_2|j_3}(0)}, \qquad \psi_{ij_1j_2j_3j_4} = \frac{\psi_{ij_1j_2j_3|j_4}(1)}{\psi_{ij_1j_2j_3|j_4}(0)}.$$

To compute the probabilities, again, the IPF algorithm as presented in Section 7.12.3 or the polynomial method of Section 7.7.4 can be used.

When the outcomes are categorical rather than binary, the likelihood presented in Section 7.8.2 can be used and, given the above, GEE2 follows in a straightforward fashion.

8.8 A Method Based on Linearization

All versions of GEE studied sofar can be seen as deriving from the score equations of corresponding likelihood methods, such as the Bahadur model (Section 7.2), the Dale model (Section 7.7), or the hybrid model (Section 7.8). In a sense, GEE results from considering only a subvector of the full vector of scores, corresponding to either the first moments only (the outcomes themselves), or the first and second moments (outcomes and cross-products thereof). On the other hand, they can be seen as an extension of the quasi-likelihood principles, where appropriate modifications are made to the scores to be sufficiently flexible and "work" at the same time. A classical modification is the inclusion of an overdispersion parameter, while in GEE also (nuisance) correlation parameters are introduced.

An alternative approach consists of linearizing the outcome, in the sense of Nelder and Wedderburn (1972), to construct a working variate, to which then weighted least squares is applied. In other words, iteratively reweighted least squares (IRLS) can be used (McCullagh and Nelder 1989). Within each step, the approximation produces all elements typically encountered in a multivariate normal model, and hence corresponding software tools can be used. In case our models would contain random effects as well (Section 14.4), the core of the IRLS could be approached using linear mixed models tools. The SAS procedure GLIMMIX is such a tool and the general case will be taken up in Chapter 14. Here, we restrict attention to the marginal-model situation. Nevertheless, it is important to note that the tools developed here can be approached using the SAS procedure GLIMMIX, as well as with the GLIMMIX macro.

Write the outcome vector in a classical (multivariate) generalized linear models fashion:

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i \tag{8.35}$$

where, as usual, $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i)$ is the systematic component and $\boldsymbol{\varepsilon}_i$ is the random component, typically following from a multinomial distribution. We assume that

$$\text{Var}(\boldsymbol{Y}_i) = \text{Var}(\boldsymbol{\varepsilon}_i) = \Sigma_i. \tag{8.36}$$

The model is further specified by assuming

$$\begin{aligned} \boldsymbol{\eta}_i &= g(\boldsymbol{\mu}_i), \\ \boldsymbol{\eta}_i &= X_i\boldsymbol{\beta}, \end{aligned}$$

where $\boldsymbol{\eta}_i$ is the usual set of linear predictors, $g(.)$ is an inverse vector link function, typically made up of logit components, X_i is a design matrix and $\boldsymbol{\beta}$ are the regression parameters.

Estimation proceeds by iteratively solving

$$\sum_{i=1}^{N} X_i' W_i X_i \boldsymbol{\beta} = \sum_{i=1}^{N} W_i \boldsymbol{Y}_i^*, \tag{8.37}$$

where a working variate $\boldsymbol{y}_i^*$ has been defined, following from a first-order Taylor series expansion of $\boldsymbol{\eta}_i$ around $\boldsymbol{\mu}_i$:

$$\begin{aligned} \boldsymbol{Y}_i^* &= \widehat{\boldsymbol{\eta}}_i + (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i) F_i^{-1}, \\ F_i &= \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}. \end{aligned} \tag{8.38}$$

The weights in (8.37) are specified as

$$W_i = F_i' \Sigma_i^{-1} F_i. \tag{8.39}$$

Note that in the specific case of an identity link, $\boldsymbol{\eta}_i = \boldsymbol{\mu}_i$, $F_i = I_{n_i}$ and $\boldsymbol{Y}_i = \boldsymbol{Y}_i^*$, whence a standard multivariate regression follows.

8.9 Analysis of the NTP Data

The NTP data, introduced in Section 2.7, have been analyzed in Section 7.2.3, by means of the Bahadur model specialized to clustered data (Section 7.2.2). Table 7.1 presented estimates and standard errors for a simple model, with marginal logits linear in dose, and a common correlation parameter, fitted the external, visceral, skeletal, and collapsed outcomes in the DEHP, EG, and DYME studies.

Here, we will consider the same model, but then from the GEE angle. We will apply standard GEE (Section 8.2), Prentice's modification (Section 8.4), and the linearization method (Section 8.8). The first approach was fitted using the SAS procedure GENMOD, the second one with a SAS macro developed by Stuart Lipsitz (Williamson, Lipsitz, and Kim 1997), and the third one using the SAS macro GLIMMIX or, equivalently, with the SAS procedure GLIMMIX. More details on software are deferred to Chapter 10. For all of these analyses, both independence (Table 8.2) and exchangeable (Table 8.3) working assumptions were considered. Other working assumptions, such as AR(1) and unstructured, are less sensible here, given the clustered nature of the data. Several models include, in addition to working assumptions, an overdispersion parameter ϕ.

In addition to these analysis, GEE2 estimates are provided in Table 8.4, based on the same models as in the Bahadur analysis, described by (7.14) and (7.15).

TABLE 8.2. *NTP Data. Parameter estimates (model-based standard errors; empirically corrected standard errors) for GEE1 with independence working assumptions, fitted to various outcomes in the DEHP study. β_0 and β_d are the marginal intercept and dose effect, respectively; ϕ is the overdispersion parameter.*

Outcome	Par.	Standard	Prentice	Linearized
External	β_0	-5.06(0.30;0.38)	-5.06(0.33;0.38)	-5.06(0.28;0.38)
	β_d	5.31(0.44;0.57)	5.31(0.48;0.57)	5.31(0.42;0.57)
	ϕ	0.90		0.74
Visceral	β_0	-4.47(0.28;0.36)	-4.47(0.28;0.36)	-4.47(0.28;0.36)
	β_d	4.40(0.43;0.58)	4.40(0.43;0.58)	4.40(0.43;0.58)
	ϕ	1.00		1.00
Skeletal	β_0	-4.87(0.31;0.47)	-4.87(0.31;0.47)	-4.87(0.32;0.47)
	β_d	4.89(0.46;0.65)	4.90(0.47;0.65)	4.90(0.47;0.65)
	ϕ	0.99		1.02
Collapsed	β_0	-3.98(0.22;0.30)	-3.98(0.22;0.30)	-3.98(0.22;0.30)
	β_d	5.56(0.40;0.61)	5.56(0.40;0.61)	5.56(0.41;0.61)
	ϕ	0.99		1.04

For a given outcome in a given study, results from the Bahadur model, the various GEE1 versions, and GEE2, are very similar. Even though for some parameters the estimated values differ a bit between analyses, they preserve the directionality and, roughly, the magnitude of the effect. This is not surprising, given that all can be seen as deriving from Bahadur's model. However, just as in, for example, Section 7.10, we observe a mild shrinkage. This is, again, due to the parameter constraints on the Bahadur model and, to a lesser extent, on GEE2. For the parameters in the Bahadur model to be allowable, all higher-order probabilities need to be valid, while for GEE2 this is necessary only up to the fourth order, the farthest the working assumptions reach. For GEE1, it is sufficient for the pairwise probabilities to be valid. Thus, it is possible for GEE to provide a valid parameter combination that cannot be reconciled with a Bahadur model, having the same lower order parameters. This does not mean there would be *no* fully specified model corresponding to it. Given the orthogonality properties of the hybrid marginal-conditional model, presented in Section 7.8, there is always a model of this type encompassing the GEE-based parameters.

The constraints on the Bahadur model are very severe indeed. For instance, the allowable range of β_a for the external outcome in the DEHP data is $(-0.0164; 0.1610)$ when β_0 and β_d are fixed at their MLE. This range translates to the very narrow $(-0.0082; 0.0803)$ on the correlation scale, excluding the GEE based values for the correlation ρ.

TABLE 8.3. *NTP Data. Parameter estimates (model-based standard errors; empirically corrected standard errors) for GEE1 with exchangeable working assumptions, fitted various outcomes in the DEHP study.* β_0 *and* β_d *are the marginal intercept and dose effect, respectively;* ρ *is the correlation;* ϕ *is the overdispersion parameter.*

Outcome	Par.	Standard	Prentice	Linearized
External	β_0	-4.98(0.40;0.37)	-4.99(0.46;0.37)	-5.00(0.36;0.37)
	β_d	5.33(0.57;0.55)	5.32(0.65;0.55)	5.32(0.51;0.55)
	ϕ	0.88		0.65
	ρ	0.11	0.11(0.04)	0.06
Visceral	β_0	-4.50(0.37;0.37)	-4.51(0.40;0.37)	-4.50(0.36;0.37)
	β_d	4.55(0.55;0.59)	4.59(0.58;0.59)	4.55(0.54;0.59)
	ϕ	1.00		0.92
	ρ	0.08	0.11(0.05)	0.08
Skeletal	β_0	-4.83(0.44;0.45)	-4.82(0.47;0.44)	-4.82(0.46;0.45)
	β_d	4.84(0.62;0.63)	4.84(0.67;0.63)	4.84(0.65;0.63)
	ϕ	0.98		0.86
	ρ	0.12	0.14(0.06)	0.13
Collapsed	β_0	-4.05(0.32;0.31)	-4.06(0.35;0.31)	-4.04(0.33;0.31)
	β_d	5.84(0.57;0.61)	5.89(0.62;0.61)	5.82(0.58;0.61)
	ϕ	1.00		0.96
	ρ	0.11	0.15(0.05)	0.11

Comparing model-based and empirically corrected standard errors, there is a clear difference in the case of independence working assumptions, but less so in the exchangeable case. Comparing both analyses is a case in point that the choice of working assumptions, whether right or wrong, is not important for the method's consistency and asymptotic normality. The impact on efficiency is minor. The statement about efficiency continues to hold when comparing all marginal analyses. In case where one is merely interested in assessing the effect of dose, GEE1, being the simplest of all methods, will do fine. When there is additional interest in the association, care is needed with GEE1. Table 8.2 provides no association parameter at all. The correlation in Table 8.3 should be approached cautiously, as the exchangeable correlation is, at best, a nuisance parameter, for which no formal inference is possible. Moreover, because we are allowed to misspecify our association model, there is no *a priori* guarantee that the parameter is trustworthy. However, in this particular case, exchangeability seems reasonable, both on biological grounds and given the design of the study. When more formal inferences about the correlation parameters

TABLE 8.4. *NTP Data. Parameter estimates (empirically corrected standard errors) for GEE2 with exchangeable correlation, fitted to various outcomes in three studies.* β_0 *and* β_d *are the marginal intercept and dose effect, respectively;* β_a *is the Fisher* z *transformed correlation;* ρ *is the correlation.*

Outcome	Parameter	DEHP	EG	DYME
External	β_0	-4.98(0.37)	-5.63(0.67)	-7.45(0.73)
	β_d	5.29(0.55)	3.10(0.81)	8.15(0.83)
	β_a	0.15(0.05)	0.15(0.05)	0.13(0.05)
	ρ	0.07(0.02)	0.07(0.02)	0.06(0.02)
Visceral	β_0	-4.49(0.36)	-7.50(1.05)	-6.89(0.75)
	β_d	4.52(0.59)	4.37(1.14)	5.51(0.89)
	β_a	0.15(0.06)	0.02(0.02)	0.11(0.07)
	ρ	0.07(0.03)	0.01(0.01)	0.05(0.03)
Skeletal	β_0	-5.23(0.40)	-4.05(0.33)	
	β_d	5.35(0.60)	4.77(0.43)	
	β_a	0.18(0.02)	0.30(0.03)	
	ρ	0.09(0.01)	0.15(0.01)	
Collapsed	β_0	-5.23(0.40)	-4.07(0.71)	-5.75(0.48)
	β_d	5.35(0.60)	4.89(0.90)	8.82(0.91)
	β_a	0.18(0.02)	0.26(0.14)	0.18(0.12)
	ρ	0.09(0.01)	0.13(0.07)	0.09(0.06)

are required, GEE2 is a viable alternative. This may be less so with the Bahadur model, given the strong parameter space restrictions.

An alternative when the association is of interest is provided by alternating logistic regressions (Section 8.6). Results of fitting ALR to the NTP data are summarized in Table 8.5. The association is in terms of log odds ratios α, as in (8.31). For convenience, we also present the odds ratios ψ. As it is a sensible choice in our case, and for ease of comparison with Tables 8.3 and 8.4, an exchangeable odds ratio structure is chosen, in the sense that all odds ratios are equal. Again, parameter estimates are similar to the ones obtained in Tables 8.2–8.4, and this holds for the standard errors as well. Of course, the association being in terms of (log) odds ratios, comparison with the correlations of the earlier analyses is not straightforward, although the relative magnitudes are roughly preserved. An advantage of the ALR analyses, apart from its implementation in standard software (the SAS procedure GENMOD, see Chapter 10), is that standard errors are provided for the association parameters. In fact, the asymptotic covariance matrix for all estimates together can be obtained.

TABLE 8.5. *NTP Data. Parameter estimates (empirically corrected standard errors) for alternating logistic regression with exchangeable odds ratio, fitted to various outcomes in three studies. β_0 and β_d are the marginal intercept and dose effect, respectively; α is the log odds ratio; ψ is the log odds ratio.*

Outcome	Parameter	DEHP	EG	DYME
External	β_0	-5.16(0.35)	-5.72(0.64)	-7.48(0.75)
	β_d	5.64(0.52)	3.28(0.72)	8.25(0.87)
	α	0.96(0.30)	1.45(0.45)	0.79(0.31)
	ψ	2.61(0.78)	4.26(1.92)	2.20(0.68)
Visceral	β_0	-4.54(0.36)	-7.61(1.06)	-7.24(0.88)
	β_d	4.72(0.57)	4.50(1.13)	6.05(1.04)
	α	1.12(0.30)	0.49(0.42)	1.76(0.59)
	ψ	3.06(0.92)	1.63(0.69)	5.81(3.43)
Skeletal	β_0	-4.87(0.49)	-3.28(0.22)	-4.92(0.34)
	β_d	4.90(0.70)	3.85(0.39)	6.73(0.65)
	α	1.05(0.40)	1.43(0.22)	1.62(0.37)
	ψ	2.86(1.14)	4.18(0.92)	5.05(1.87)
Collapsed	β_0	-4.04(0.31)	-3.19(0.22)	-5.08(0.37)
	β_d	5.93(0.63)	3.86(0.40)	7.98(0.75)
	α	1.17(0.29)	1.40(0.22)	1.26(0.31)
	ψ	3.22(0.93)	4.06(0.89)	3.53(1.09)

8.10 The Heatshock Study

A unique type of developmental toxicity study was originally developed by Brown and Fabro (1981) to assess the impact of heat stress on embryonic development, and adapted by Kimmel *et al* (1993) to investigate effects of both temperature and duration of exposure. In these heatshock experiments, the embryos are explanted from the uterus of the maternal dam during the gestation period and cultured *in vitro*. Each individual embryo is subjected to a short period of heat stress by placing the culture vial into a water bath, involving an increase over body temperature of 3 to 5°C for a duration of 5 to 60 minutes. The embryos are examined 24 hours later for signs of impaired or accelerated development.

This type of developmental toxicity test system has several advantages over the standard Segment II design. First, the exposure is administered directly to the embryo, so controversial issues regarding the unknown (and often non-linear) relationship between the level of exposure to the maternal dam and that received by the developing embryo need not be addressed. While genetic factors are still expected to exert an influence on the vulnerability to injury of embryos from a common dam, direct exposure to

TABLE 8.6. *Heatshock Study. Hybrid marginal-conditional parameter estimates (model-based standard errors; empirically corrected standard errors) for models fitted to the outcomes MBN, FBN, OLF, and BRB. Covariate effects are allowed to differ across outcomes, and a different association parameter is assumed for each pair. Model 1 presents the estimates under conditional constraints for the higher order association; Model 2 uses marginal constraints. Higher order associations are included in Model 3. Models 1 and 2 are at the same time maximum likelihood and GEE2. Model 3 is full likelihood. Part I: Marginal parameters.*

Parameter	Model 1	Model 2	Model 3
Marginal parameters			
Midbrain (MBN)			
Intercept	-1.81(0.23;0.24)	-1.81(0.23;0.24)	-1.83(0.23;0.24)
‘posdur’	-0.12(0.04;0.04)	-0.12(0.04;0.04)	-0.10(0.04;0.04)
‘durtemp’	0.04(0.01;0.01)	0.04(0.01;0.01)	0.04(0.01;0.01)
Forebrain (FBN)			
Intercept	-1.73(0.23;0.23)	-1.73(0.23;0.23)	-1.71(0.23;0.22)
‘posdur’	-0.09(0.04;0.04)	-0.09(0.04;0.04)	-0.09(0.04;0.04)
‘durtemp’	0.04(0.01;0.01)	0.04(0.01;0.01)	0.04(0.01;0.01)
Olfactory system (OLF)			
intercept	-1.43(0.22;0.21)	-1.44(0.22;0.21)	-1.46(0.20;0.21)
‘posdur’	-0.21(0.04;0.05)	-0.21(0.04;0.05)	-0.21(0.04;0.05)
‘durtemp’	0.07(0.01;0.01)	0.07(0.01;0.01)	0.07(0.01;0.01)
Branchial bars (BRB)			
intercept	-1.19(0.20;0.20)	-1.18(0.20;0.20)	-1.09(0.20;0.20)
‘posdur’	-0.13(0.04;0.04)	-0.13(0.04;0.04)	-0.13(0.04;0.04)
‘durtemp’	0.04(0.01;0.01)	0.04(0.01;0.01)	0.04(0.01;0.01)

individual embryos reduces the need to account for such litter effects. Thus, the clustering induced by litter effects are not considered in our analysis. A detailed analysis of the clustering aspect can be found in Aerts *et al* (2002). Second, the exposure pattern can be much more easily controlled than in most developmental toxicity studies, as it is possible to achieve target temperature levels in the water bath within one to two minutes. Whereas the typical Segment II study requires waiting eight to twelve days after exposure to assess its impact, information regarding the effects of exposure are quickly obtained in heatshock studies. Finally, this animal test system provides a convenient mechanism for examining the joint effects of both duration of exposure and exposure levels, which until recently have received little attention. The actual study design for the set of experiments is shown in Kimmel *et al* (1994). Of the 327 embryos exposed, 50 did not

TABLE 8.7. *Heatshock Study. Hybrid marginal-conditional parameter estimates (model-based standard errors; empirically corrected standard errors) for models fitted to the outcomes MBN, FBN, OLF, and BRB. Covariate effects are allowed to differ across outcomes, and a different association parameter is assumed for each pair. Model 1 presents the estimates under conditional constraints for the higher order association; Model 2 uses marginal constraints. Higher order associations are included in Model 3. Models 1 and 2 are at the same time maximum likelihood and GEE2. Model 3 is full likelihood. Part II: Association parameters.*

Parameter	Model 1	Model 2	Model 3
Pairwise association			
MBN FBN	3.22(0.38;0.39)	3.22(0.38;0.40)	3.13(0.37;0.38)
MBN OLF	2.69(0.36;0.38)	2.69(0.36;0.37)	2.77(0.36;0.38)
MBN BRB	2.10(0.32;0.33)	2.10(0.32;0.33)	2.17(0.33;0.33)
FBN OLF	3.58(0.41;0.42)	3.59(0.41;0.42)	3.62(0.42;0.44)
FBN BRB	2.54(0.34;0.34)	2.55(0.34;0.34)	2.60(0.34;0.34)
OLF BRB	2.52(0.33;0.34)	2.53(0.33;0.34)	2.61(0.33;0.34)
Higher order association			
MBN FBN OLF			1.30(1.34;1.42)
MBN FBN BRB			0.96(1.19;1.17)
MBN OLF BRB			0.22(1.30;1.38)
FBN OLF BRB			2.12(1.48;1.51)
MBN FBN OLF BRB			3.18(1.77;1.80)
Deviance	946.05	945.15	937.80

survive the heat stress exposure and were excluded from further analysis. The remaining 277 animals have complete data.

Historically, the strategy for comparing responses among exposures of different durations to a variety of environmental agents has relied on a conjecture called Haber's law, which states that adverse response levels should be the same for any equivalent level of dose times duration (Haber 1924). Clearly, the appropriateness of applying Haber's law depends on the pharmacokinetics of the particular agent, the route of administration, the target organ, and the dose/duration patterns under consideration. Although much attention has been focused on documenting exceptions to this rule, it is often used as a simplifying assumption in view of limited testing resources and the multitude of exposure scenarios. However, given the current desire to develop regulatory standards for a range of exposure durations, models flexible enough to describe the response patterns over varying levels of both exposure concentration and duration are greatly needed.

Although a wide variety of statistical methods have been developed for cancer risk assessment, the issue of multiple endpoints does not present

TABLE 8.8. *Heatshock Study. Empirically corrected (e.c.) and model-based (m.b.) Wald test statistics based on Models 1–3. Apart from tests for common covariate effects and common pairwise association, tests for common covariate effects among MBN, FBN, and BRB [indicated by (**)] are presented, as well as a test whether the association splits into two groups: pairs including versus excluding BRB.* * *indicates* $p < 0.05$.

Hypothesis	df	Model 1 e.c.	Model 1 m.b.	Model 2 e.c.	Model 2 m.b.	Model 3 e.c.	Model 3 m.b.
Common 'posdur'	3	*7.95	*9.28	*8.47	*9.85	*9.87	*11.31
Common 'posdur' (**)	2	0.16	0.19	0.17	0.19	0.10	0.07
Common 'durtemp'	3	5.73	7.68	6.13	*8.21	7.39	*9.92
Common 'durtemp' (**)	2	1.36	1.08	1.49	1.14	1.92	1.18
Common pairwise assoc.	5	*12.58	*13.63	*12.55	*13.55	10.16	*11.92
Two groups of pairwise assoc.	4	6.49	6.18	6.52	6.19	4.99	5.65

quite the degree of complexity in this area as it does for developmental toxicity studies. The endpoint of interest in an animal cancer bioassay is typically the occurrence of a particular type of tumor, whereas in developmental toxicity studies there is no clear choice for a single type of adverse outcome. In fact, an entire array of outcomes are needed to define certain birth defect syndromes (Khoury *et al* 1987, Holmes 1988).

The data have been analyzed before by Williams, Molenberghs, and Lipsitz (1996). In line with Molenberghs and Ritter (1996), we will consider a multivariate analysis on four binary morphological parameters: Midbrain (MBN), Forebrain (FBN), Olfactory System (OLF), and Branchial Bars (BRB). They are coded as affected versus normal. If Haber's law is satisfied, the main covariate is 'durtemp,' the product of duration and dose (temperature increase). We found that the main effect 'duration' is also important. However, we expect duration to have no effect at the control dose, therefore it was recoded as 'posdur,' which is equal to 'duration' in the exposed groups and zero in the control group. Including the main effect 'temperature' does not significantly improve the fit.

All of our analyses in this section will be conducted by means of the hybrid between a marginal and conditional model, for which the full likelihood version was given in Section 7.8, with a GEE2 version introduced in Section 8.7. Tables 8.6 and 8.7 show three models fitted to these data. Given the orthogonality between lower-order and higher-order parameters, the estimates can be considered both as stemming from maximum likelihood, as well as from GEE2, depending on whether one views the higher-order association is set equal to zero because this is believed to be the correct structure, or rather merely as a working assumption. In a few models,

TABLE 8.9. *Heatshock Study. Hybrid marginal-conditional parameter estimates (empirically corrected standard errors) for models fitted to the outcomes MBN, FBN, OLF, and BRB. Covariate effects are allowed to differ across outcomes. Common covariate effects are assumed for MBN, FBN, and BRB. Pairwise associations are grouped in: (1) Group 1, containing all pairs formed from MBN, FBN, and OLF, and (2) Group 2, all pairs containing BRB. Models 4, 5, and 7 are at the same time maximum likelihood and GEE2. Model 6 is full likelihood.*

Parameter	Model 4	Model 5	Model 6	Model 7
Marginal parameters				
Intercepts				
MBN	-1.75(0.20)	-1.75(0.20)	-1.70(0.20)	-1.74(0.20)
FBN	-1.49(0.19)	-1.49(0.19)	-1.47(0.19)	-1.49(0.19)
OLF	-1.40(0.21)	-1.41(0.21)	-1.39(0.21)	-1.45(0.21)
BRB	-1.41(0.19)	-1.41(0.19)	-1.40(0.19)	-1.42(0.19)
Covariates (MBN, FBN, BRB)				
‘posdur’	-0.12(0.03)	-0.12(0.03)	-0.12(0.03)	-0.13(0.03)
‘durtemp’	0.04(0.01)	0.04(0.01)	0.04(0.01)	0.04(0.01)
Covariates (OLF)				
‘posdur’	-0.22(0.05)	-0.22(0.05)	-0.22(0.05)	-0.22(0.04)
‘durtemp’	0.07(0.01)	0.07(0.01)	0.07(0.01)	0.07(0.01)
Pairwise association				
Intercepts				
Group 1	3.10(0.29)	3.11(0.29)	3.11(0.29)	3.50(0.40)
Group 2	2.32(0.26)	2.33(0.26)	2.32(0.26)	2.73(0.38)
Covariates				
‘posdur’				0.16(0.07)
‘durtemp’				-0.05(0.02)
Higher order association				
MBN FBN OLF			0.47(1.35)	
MBN FBN BRB			0.96(1.06)	
MBN OLF BRB			0.19(1.44)	
FBN OLF BRB			1.87(1.50)	
MBN FBN OLF BRB			2.23(1.76)	
Deviance	959.31	959.73	954.24	951.79

higher-order association parameters are included as well (Models 3 and 6), implying they are full likelihood.

Models 1 and 2 do not include higher order associations. Model 1 applies conditional constraints, whereas Model 2 considers its marginal counter-

TABLE 8.10. *Heatshock Study. Estimated odds ratios for Model 7 in Table 8.9. Entries marked with a * correspond to a duration-temperature combination not present in the data.*

	Duration		
Temp.	5	30	60
0.0	33.1	33.1	33.1
3.0	36.1	55.6	93.5
3.5	32.0	26.9	21.8
4.0	28.3	13.0	* 5.1
4.5	25.1	6.3	* 1.2
5.0	22.2	3.0	* 0.3

parts. Model 3 includes the higher-order associations as well. Parameter estimates and standard errors are shown. Clearly, the marginal parameters are virtually the same across models, with the same holding true for the standard errors. Further, it is clear that some of the covariate effects are very similar, and some of the pairwise association parameters are very close to each other.

Table 8.8 presents test statistics based on the model based and robust variance estimators, obtained for Models 1–3. A common 'posdur' effect is clearly not tenable. A common 'durtemp' effect gives p-values that are borderline, as $\chi^2 = 7.68$ corresponds to $p = 0.053$ and $\chi^2 = 7.39$ to 0.061. From the model parameters we observe that the effects of 'posdur' and 'durtemp' are virtually the same for MBN, FBN, and BRB, whereas OLF differs slightly. The test statistics presented in Table 8.8 support these hypotheses. A common pairwise association parameter is not supported, but if the association is divided into two groups (pairs with and without BRB) a simplification which is consistent with the data is achieved.

Reduced models are presented in Table 8.9, where only robust standard errors are shown. Observe that the similarities across Models 4–6 are even greater. Comparing models from Tables 8.6 and 8.7 with their corresponding ones in Table 8.9 using a likelihood ratio statistics yields: 13.26 (Model 4 versus Model 1), 14.58 (Model 5 versus Model 2), and 16.42 (Model 6 versus Model 3), all on 8 degrees of freedom. Only the last one is above the 5 % critical level.

We gathered some evidence for a dependence of pairwise association on the level of exposure. Model 7 in Table 8.9 shows an extension of Model 4, where a common linear effect of 'posdur' and 'durtemp' is included for the pairwise odds ratios. Allowing for a quadratic effect shows no significant improvement. The pairwise association for pairs excluding BRB is described by a log odds ratio of $3.5 + 0.16 *$ 'posdur' $- 0.049 *$ 'durtemp'. The effect of 'temperature' is not significant. The association increases (slightly) with

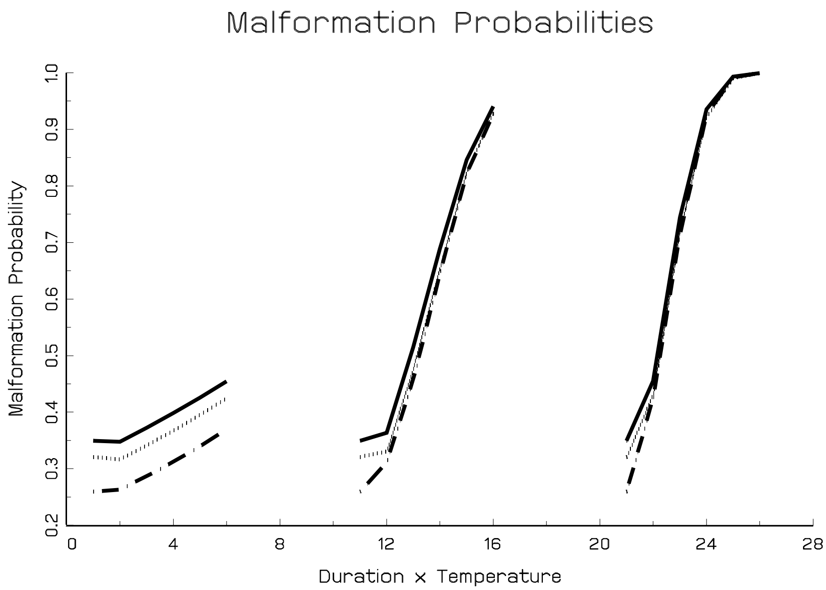

FIGURE 8.1. *Heatshock Study. Malformation probabilities, based on models including three and four outcomes. Ranges of 'durtemp' at three levels of 'duration' are presented. Solid line: MBN, FBN, OLF, BRB; dotted line: FBN, OLF, BRB; dots and dashes: MBN, FBN, OLF.*

'posdur,' but a dramatic decrease is seen with 'durtemp.' A selection of the estimated odds ratios are shown in Table 8.10.

The pairwise associations are important as a tool used to reduce the length of the outcome vector. Indeed, observe that the association between MBN and FBN is very high, and that 'posdur' and 'durtemp' have a similar effect on both. This might imply that considering, e.g., FBN, OLF, and BRB only might yield a similar predicted probability of *any* malformation. In Figure 8.1, we show the malformation probability for a range of 'durtemp' values, at duration levels 5, 30, and 60 minutes. The malformation probabilities are estimated based on three models: Model 1, including all four outcomes, the three-way version with FBN, OLF, and BRB, and the three-way version with MBN, FBN, and OLF. In the latter case, BRB has been omitted. As the association between pairs including BRB is observed to be smaller, it is not surprising that the latter model underestimates the malformation probability as it ignores important independent information. This is best seen at smaller doses, which is important if the models are used for low dose extrapolation.

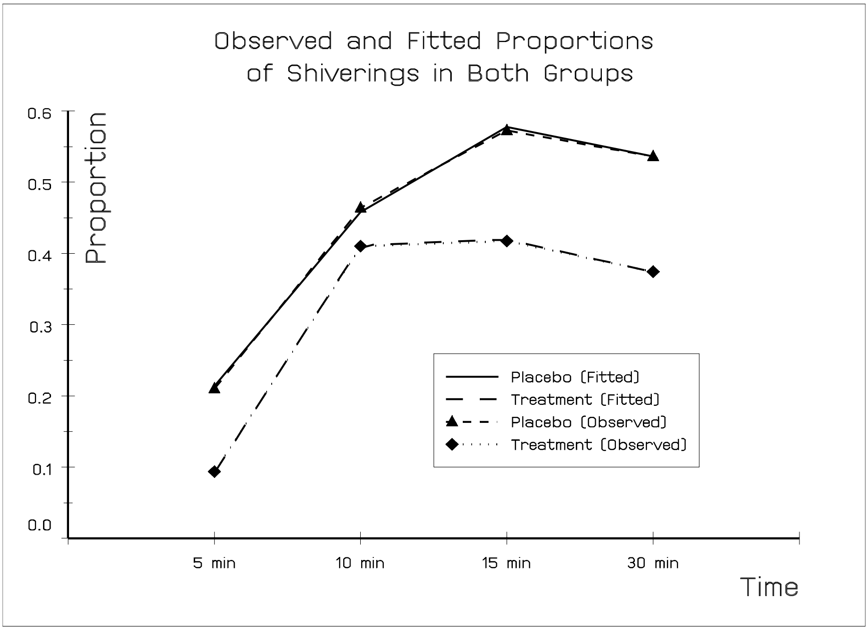

FIGURE 8.2. *Sports Injuries Trial. Observed and fitted proportions of shivering in both arms. Fitted proportions are based on Model 2 in Table 8.11.*

8.11 The Sports Injuries Trial

The sports injuries trial has been introduced in Section 2.8.

We will apply the hybrid marginal and conditional model, introduced in Sections 7.8 and 8.7, in the context of both the longitudinal outcome as well as with repeated measures on the two outcomes, shivering and awakeness. Note that, just as in Section 8.10, two perspectives on the parameter estimates obtained from the hybrid model are possible, maximum likelihood and GEE2. The first one applies when the higher-order association is modeled explicitly or considered to be zero, in line with the working assumptions. The second one applies when the higher-order parameters are set equal to zero by way of working assumption only.

8.11.1 Longitudinal Analysis

The first analysis considers four binary measurements of shivering (at 5, 10, 15, and 20 minutes). Data are presented in Table 2.12. We are interested in a treatment difference and its evolution over time. First, the profiles show a quadratic time trend, as can be seen in Figure 8.2. Next, we need a cubic polynomial to describe the difference between treatment and placebo

TABLE 8.11. *Sports Injuries Trial. Hybrid marginal-conditional model parameter estimates (model based standard errors; empirically corrected standard errors) for models fitted to four binary shivering responses (at 5, 10, 15, and 20 minutes). Model 1 presents unrestricted and Model 2 presents restricted association parameters.*

Parameter	Model 1	Model 2
	Marginal parameters	
Intercept	0.15(0.16;0.16)	0.15(0.16;0.16)
Time effect:		
Linear	0.48(0.08;0.08)	0.48(0.08;0.08)
Quadratic	-0.33(0.06;0.06)	-0.33(0.06;0.06)
Treatment effect:		
Main effect	-0.36(0.23;0.23)	-0.36(0.22;0.23)
Linear interaction	-0.52(0.19;0.18)	-0.52(0.19;0.18)
Quadratic interaction	-0.20(0.10;0.10)	-0.20(0.10;0.10)
Cubic interaction	0.28(0.09;0.08)	0.28(0.09;0.08)
	Association	
$(1,2)$	3.67(0.73;0.71)	3.71(0.72;0.70)
$(1,3)$	2.54(0.55;0.57)	
$(1,4)$	1.45(0.38;0.39)	
$(2,3)$	2.69(0.31;0.31)	
$(2,4)$	1.48(0.26;0.26)	
$(3,4)$	2.61(0.30;0.30)	
$(1,3) = (2,3) = (3,4)$		2.64(0.21;0.21)
$(1,4) = (2,4)$		1.47(0.25;0.25)
Deviance	1104.09	1104.20

profiles. Not surprisingly, the difference is more marked at later times. Observed an fitted profiles are plotted in Figure 8.2.

Next, we study the association structure. There are six pairwise association parameters, one for each pair of measurement times. There is an extraordinary strong association between the first and second time, the odds ratio equals 39.2. This is explained by the relatively small number of changes in shivering state at the beginning of the trial. Then, association decreases with distance between time points. For the five remaining associations, we consider measurements 1 and 2 to occur virtually together and group the parameters by the difference in time between both measurements: (1,3), (2,3), and (3,4) on the one hand and (1,4) and (2,4) on the other. This reduces the number of association parameters to three, while

TABLE 8.12. *Sports Injuries Trial. Hybrid marginal-conditional model parameter estimates (standard errors) for models fitted to four ordinal shivering responses (at 5, 10, 15, and 20 minutes).*

Parameter	Estimate
Marginal parameters	
First cutpoint:	
Intercept	0.25(0.15)
Linear time	0.47(0.16)
Quadratic time	-0.40(0.06)
Cubic time	0.02(0.07)
Second cutpoint:	
Intercept	-1.42(0.18)
Linear time	0.86(0.22)
Quadratic time	-0.47(0.12)
Cubic time	0.01(0.11)
Treatment effect:	
Main effect	-0.49(0.21)
Linear interaction	-0.46(0.22)
Quadratic interaction	-0.05(0.08)
Cubic interaction	0.22(0.10)
Association	
$(1,2)$	3.81(0.45)
$(1,3)$	3.29(0.39)
$(1,4)$	1.59(0.30)
$(2,3)$	2.69(0.29)
$(2,4)$	1.49(0.26)
$(3,4)$	2.52(0.29)

virtually not changing the quality of the fit. Table 8.11 presents parameter estimates (standard errors) for both unrestricted (Model 1) and restricted (Model 2) associations.

Taking a likelihood perspective, the overall deviance goodness-of-fit statistic is 7.66 on 20 degrees of freedom, providing evidence that there is no need for higher order association. This means that, while a GEE2 perspective is still possible, assuming the higher-order association is left unspecified and replaced by working assumptions, it is fine too to adopt a likelihood point of view, where the first-order and second-order moments have been modeled correctly, and the higher-order associations vanish.

TABLE 8.13. *Sports Injuries Trial. Cross-classification of two pairs of dichotomized shivering and awakeness measurements (at 10 and 20 minutes).*

Shivering	Awakeness			
	(0,0)	(0,1)	(1,0)	(1,1)
	Placebo arm			
(0,0)	14	12	0	20
(0,1)	3	17	0	8
(1,0)	0	12	0	6
(1,1)	3	15	0	28
	Treatment arm			
(0,0)	12	23	0	28
(0,1)	5	9	0	5
(1,0)	2	13	0	9
(1,1)	3	24	0	6

TABLE 8.14. *Sports Injuries Trial. Hybrid marginal-conditional model parameter estimates (model-based standard errors; empirically corrected standard errors) for models fitted to two pairs of shivering/awakeness measurements, at 10 and 20 minutes. When the two sets of standard errors coincide, only one is shown. Part I: Marginal parameters.*

Parameter	Model 1	Model 2	Model 3	Model 4
	Marginal parameters			
Shivering at 10 minutes:				
Intercept	-0.15(0.17;0.17)	-0.15(0.17)	-0.15(0.17)	-0.16(0.17)
Treatment	-0.21(0.24;0.24)	-0.22(0.24)	-0.22(0.24)	-0.21(0.24)
Shivering at 20 Minutes:				
Intercept	0.15(0.17;0.17)	0.15(0.17)	0.15(0.17)	0.14(0.17)
Treatment	-0.67(0.24;0.25)	-0.66(0.24)	-0.66(0.24)	-0.67(0.24)
Awakeness at 10 Minutes:				
Intercept	-0.20(0.17;0.17)	-0.20(0.17)	-0.20(0.17)	-0.20(0.17)
Treatment	-0.45(0.25;0.25)	-0.44(0.25)	-0.44(0.25)	-0.43(0.25)
Awakeness at 20 Minutes:				
Intercept	1.86(0.25;0.26)	1.78(0.24)	1.78(0.24)	1.85(0.24)
Treatment	-0.26(0.33;0.33)	-0.10(0.34)	-0.10(0.34)	-0.23(0.33)

Finally, we reconsidered this analysis, but now on ordinal endpoints. Because category 3 is either empty or very sparse for the four shivering

TABLE 8.15. *Sports Injuries Trial. Hybrid marginal-conditional model parameter estimates (model-based standard errors; empirically corrected standard errors) for models fitted to two pairs of shivering/awakeness measurements, at 10 and 20 minutes. When the two sets of standard errors coincide, only one is shown. Part II: Association parameters.*

Parameter	Model 1	Model 2	Model 3	Model 4
Pairwise association				
Shivering 1/Shivering 2:				
Intercept	1.48(0.26;0.26)	1.43(0.37)	1.44(0.37)	1.32(0.25)
Treatment		0.08(0.52)	0.08(0.52)	
Shivering 1/Awakeness 1:				
Intercept	0.03(0.25;0.25)	0.62(0.35)	0.62(0.35)	
Treatment		-1.26(0.51)	-1.25(0.51)	
Shivering 1/Awakeness 2:				
Intercept	1.36(0.41;0.42)	1.80(0.65)	1.81(0.64)	1.16(0.36)
Treatment		-0.80(0.85)	-0.81(0.84)	
Shivering 2/Awakeness 1:				
Intercept	-0.28(0.25;0.25)	0.33(0.34)	0.33(0.34)	
Treatment		-1.34(0.53)	-1.34(0.53)	-0.83(0.36)
Shivering 2/Awakeness 2:				
Intercept	0.58(0.36;0.37)	1.15(0.52)	1.15(0.52)	
Treatment		-1.10(0.71)	-1.10(0.71)	
Awakeness 1/Awakeness 2:				
Intercept	$+\infty$	$+\infty$	$+\infty$	$+\infty$
Higher-order Association				
Shivering 1/Shivering 2/Awakeness 1:				
Intercept			2.58(0.82)	
Treatment			-2.81(1.20)	
Shivering 1/Shivering 2/Awakeness 2:				
Intercept			$-\infty$	
Treatment			0.27(1.18)	
Deviance	1260.03	1249.67	1234.72	1260.82

measures being studied, it is combined with category 2. Consequently, we have two sets of profiles. Potentially, both time and treatment effects can differ depending on the cutpoint. There is evidence for such a difference in the time trend in the form of a Wald test of 10.26 on 3 degrees of freedom ($p = 0.017$). On the other hand, it is plausible to consider a single treatment profile, common to both cutpoints (Wald test of 6.85 on 4 degrees of freedom; $p = 0.14$). Estimates for the corresponding model are presented in Table 8.12.

8.11.2 A Bivariate Longitudinal Analysis

The second analysis considers two pairs of shivering and awakeness outcomes, at 10 and 20 minutes. Data are given in Table 8.13. The main interest lies in the effect of treatment for each outcome, as well as in the association between the outcomes. There is a complication with the association between the two awakeness measures, due to the structural zeros described earlier. Indeed, the corresponding log odds ratio is equal to infinity. If this parameter is estimated along with the others, we obtain a solution on the boundary of the parameter space, invalidating inference. One way out is to set this parameter equal to zero or another arbitrary (finite) value. However, this is unsatisfactory form a theoretical point of view, as we assume independence, knowing that there is an infinitely large association. Alternatively, we can incorporate a log odds ratio of $+\infty$ as a structural feature of the model. Some straightforward technical modifications are required to the fitting program, such as replacing (6.16) by $\mu_{ij_1j_2} = \min(\mu_{ij_1}, \mu_{ij_2})$. The parameter estimates are given in Tables 8.14 and 8.15 (Model 1).

The effect of treatment is clearly seen at the second shivering measurement and only marginally at the first awakeness measurement. Only two of the estimated pairwise associations are strong: between both shivering measurements, and between the first shivering and the second awakeness measurement. Because shivering often occurs as the patient abruptly changes levels of consciousness, this could explain the association.

When computing the goodness-of-fit, one has to take into account that in each $2 \times 2 \times 2 \times 2$ table (one for each treatment group), there are 4 zero cells by design, reducing the data degrees of freedom to 22. Model 1 yields a deviance G^2 statistic of 25.32 on 9 degrees of freedom, which is clearly unacceptable. First, the two-way association can be extended by allowing for differences in association for the two treatment groups. The G^2 statistic reduces to 14.95 on 4 degrees of freedom, which still leaves room for improvement. To extend the model, the higher order associations need to be modeled as well. Recall that, due to the orthogonality of marginal and conditional parameters, this model (Model 2 in Tables 8.14 and 8.15) can be considered satisfactory as it is saturated in the marginal parameters, and the model-based and empirically corrected standard errors coincide (hence only one entry is shown).

As we allowed the pairwise interactions to depend on treatment, a more detailed picture than the one from Model 1 emerges. Apart from a structural $+\infty$ for the association parameter between both awakeness measures, we find a relatively strong odds ratio for the two shiverings (consistent with other analyses), without evidence for a treatment dependence. These two odds ratios describe the longitudinal part of the association, pertaining to two measurements of the same variable at different occasions.

Alternatively, one can seek to estimate the higher order association parameters as well. Here too, we have to take into account the zero cells. It suffices to leave out all higher order interactions containing awakeness measures simultaneously. This leaves two three-way conditional odds ratios to estimate: (shivering 1, shivering 2, awakeness 1) and (shivering 1, shivering 2, awakeness 2). Assuming these are constant yields a G^2 statistic of 9.13 on 2 degrees of freedom. This implies that also the higher order interactions are treatment dependent. Due to a sampling zero, the second one of these log odds ratios is zero. Setting it equal to zero, and estimating the value only in the treatment group, then corresponds to the saturated model (Model 3 in Tables 8.14 and 8.15). It is interesting to note that the first of the three-way interactions (shivering 1, shivering 2, awakeness 1) is significant, at least in the placebo group.

Problems with sampling zeros occur less frequently when the higher order association is described via marginal odds ratios (Molenberghs and Lesaffre 1994). Comparing Models 2 and 3, it might be argued that setting the higher association parameters equal to zero is a sensible choice, especially when scientific interest is limited to the first two moments.

To interpret the two-way association, we observe that some of the associations in Models 2 and 3 do not attain statistical significance. Hence it is useful to consider a more parsimonious model. We simplify Model 2 such that only the following pairwise associations are included: a common log odds ratio for the (shivering 1, shivering 2) and (shivering 1, awakeness 2) pairs and association between shivering 2 and awakeness 1 in the treatment group only. Comparing this model to Model 2 with a likelihood ratio test, of course taking the likelihood perspective on the model, we obtain a G^2 test statistic value of 11.15 on 7 degrees of freedom ($p = 0.13$). Note that the main effect parameters all change less than 0.01 except for awakeness at 20 minutes.