Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2009   *Paper* 51

Correlated Binary Regression Using Orthogonalized Residuals

Richard C. Zink[*]   Bahjat F. Qaqish[†]

[*]Department of Biostatistics and Data Management, Inspire Pharmaceuticals, Richard.Zink@jmp.com

[†]University of North Carolina at Chapel Hill, qaqish@bios.unc.edu

Correlated Binary Regression Using Orthogonalized Residuals

Richard C. Zink and Bahjat F. Qaqish

Abstract

This paper focuses on marginal regression models for correlated binary responses when estimation of the association structure is of primary interest. A new estimating function approach based on orthogonalized residuals is proposed. This procedure allows a new representation and addresses some of the difficulties of the conditional-residual formulation of alternating logistic regressions of Carey, Zeger & Diggle (1993). The new method is illustrated with an analysis of data on impaired pulmonary function.

Correlated Binary Regression Using Orthogonalized Residuals

Richard C. Zink

Department of Biostatistics and Data Management
Inspire Pharmaceuticals
Durham, North Carolina, U.S.A.
(rzink@inspirepharm.com)

Bahjat F. Qaqish

Department of Biostatistics, University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, U.S.A.
(qaqish@bios.unc.edu)

Abstract

This paper focuses on marginal regression models for correlated binary responses when estimation of the association structure is of primary interest. A new estimating function approach based on orthogonalized residuals is proposed. This procedure allows a new representation and addresses some of the difficulties of the conditional-residual formulation of alternating logistic regressions of Carey, Zeger & Diggle (1993). The new method is illustrated with an analysis of data on impaired pulmonary function.

Keywords: Alternating logistic regressions; Correlated binary observations; Clustered data; Generalized estimating equations; Marginal models.

1
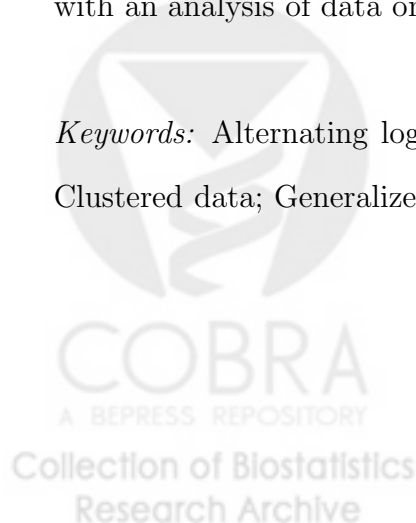
1 Introduction

This paper focuses on marginal regression models for correlated binary responses when estimation of the association structure is of primary interest. Throughout, all vectors are column vectors. Suppose data are available on K independent subjects, families, pedigrees or clusters. Let i identify a cluster and j and k index observations within a cluster. The triple index ijk references observations j and k of cluster i, $1 \leq j < k \leq n_i$, where n_i is the cluster sample size. For cluster i, the response vector is $Y_i = (Y_{i1}, \ldots, Y_{in_i})^\top$, where each Y_{ij} is a Bernoulli random variable with mean $\mu_{ij} = \mathrm{pr}(Y_{ij} = 1)$. Define also $\mu_{ijk} = E[Y_{ij}Y_{ik}] = \mathrm{pr}(Y_{ij} = Y_{ik} = 1)$. The dependence or association between Y_{ij} and Y_{ik} can be represented by the odds ratio

$$\psi_{ijk} = \frac{\mu_{ijk}(1 - \mu_{ij} - \mu_{ik} + \mu_{ijk})}{(\mu_{ij} - \mu_{ijk})(\mu_{ik} - \mu_{ijk})},$$

the correlation coefficient, $\rho_{ijk} = \mathrm{corr}(Y_{ij}, Y_{ik})$, or by other measures such as the kappa coefficient.

Dependence of the mean on covariates is modeled through a link function g_1, $g_1(\mu_{ij}) = x_{ij}^\top \beta$, where x_{ij} is a covariate p-vector associated with Y_{ij} and the components of β are the mean parameters. Dependence of the pairwise association on covariates is modeled through a second link function g_2, $g_2(\mu_{ij}, \mu_{ij}, \mu_{ijk}) = z_{ijk}^\top \alpha$, where z_{ijk} is a covariate q-vector associated with the pair (Y_{ij}, Y_{ik}) and the components of α are the association parameters. Common choices for link functions include logit and probit for the mean

structure and log odds ratio and Fisher's z-transformation of the correlation coefficient for the association structure. Most of the discussions below are not dependent on any particular choice of link functions. Finally, define θ be the $(p+q)$-vector $(\beta^\top, \alpha^\top)^\top$ and note that the covariance matrix $\Sigma_i = \mathrm{cov}(Y_i)$ is completely determined by θ.

The regression model described above is a marginal model because the expectations involved in μ_{ij} and μ_{ijk} are not conditional on other responses or on latent random effects. Differences in interpretation and applicability of marginal, conditional and random-effects models have been elaborated by Zeger, Liang & Albert (1988), Neuhaus, Kalbfleish & Hauk (1991) and Heagerty & Zeger (2000). For $n_i > 2$, the marginal model parameters θ do not fully specify the joint distribution of Y_i so that maximum-likelihood estimation is not possible without further assumptions. Because the joint distribution of Y_i is determined by 2^{n_i} probabilities, except for small n_i, computation of maximum-likelihood estimates becomes very demanding.

To reduce this burden, second-order generalized estimating equations were developed by Liang, Zeger & Qaqish (1992) for estimation of θ with minimal further assumptions. The basic idea is to append to Y_i the $m_i = n_i(n_i - 1)/2$ products $W_{ijk} = Y_{ij}Y_{ik}$, then develop an estimating equation based on the extended vector. The second-order generalized estimating equa-

3

tions are

$$U_{\theta,GEE2} = \sum_{i=1}^{K} \begin{pmatrix} D_i & 0 \\ A_i & C_i \end{pmatrix}^{\top} (\Sigma_i^*)^{-1} \begin{pmatrix} Y_i - \mu_i \\ W_i - \delta_i \end{pmatrix}, \tag{1}$$

where $W_i = (W_{i12}, \cdots, W_{i,n_i-1,n_i})^{\top}$, $\delta_i = E[W_i]$, $A_i = \partial \delta_i / \partial \beta$, $C_i = \partial \delta_i / \partial \alpha$ and $\Sigma_i^* = \mathrm{cov}((Y_i^{\top}, W_i^{\top})^{\top})$. The matrix Σ_i^* involves third- and fourth-order cross-moments not specified by the marginal model. A working version of Σ_i^* is obtained by assuming that third- and fourth-order logistic contrasts are zero (Liang et al., 1992).

For future reference we define the first-order generalized estimating equations for β (Liang & Zeger, 1986),

$$U_{\beta,GEE1} = \sum_{i=1}^{K} D_i^{\top} \Sigma_i^{-1} (Y_i - \mu_i) = 0, \tag{2}$$

where

$$\Sigma_i = \mathrm{cov}(Y_i) = \mathrm{diag}(\sigma_{ijj}^{\frac{1}{2}}) \, R_i \, \mathrm{diag}(\sigma_{ijj}^{\frac{1}{2}}),$$

$\sigma_{ijj} = \mathrm{var}(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$ and $R_i = \mathrm{corr}(Y_i)$.

A practical difficulty in implementing (1) for large clusters is that the computational effort grows very quickly with n_i. Computing (1) requires solving a linear system in $n_i(n_i + 1)/2$ unknowns with effort $O(n_i^6)$ floating point operations. Besides computational complexity, another reason for seeking alternatives to second-order generalized estimating equations is the sensitivity of the β estimates to misspecification of the association model.

4

Several alternatives to (1) combine $U_{\beta,GEE1}$ with a pairwise kernel, κ_{ijk}, whose sum over all pairs defines the cluster's contribution to the estimating function for α,

$$U_\alpha = \sum_{i=1}^{K} \sum_{j<k} \kappa_{ijk}.$$

Prentice (1988) suggested the pairwise kernel

$$\kappa_{ijk} = -\frac{\partial T_{ijk}}{\partial \alpha}^\top \frac{T_{ijk}}{\mathrm{var}(T_{ijk})} \tag{3}$$

where

$$T_{ijk} = \frac{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})}{(\sigma_{ijj}\sigma_{ikk})^{\frac{1}{2}}} - \rho_{ijk}.$$

Lipsitz, Laird & Harrington (1991) developed an estimating function using the kernel

$$\kappa_{ijk} = \frac{\partial \delta_{ijk}}{\partial \alpha}^\top \frac{W_{ijk} - \delta_{ijk}}{\mathrm{var}(W_{ijk})}. \tag{4}$$

One final method is alternating logistic regressions (Carey, Zeger & Diggle, 1993) which, along with $U_{\beta,ALR} = U_{\beta,GEE1}$, defines $U_{\alpha,ALR}$ using a pairwise kernel based on conditional residuals

$$\kappa_{ijk} = \frac{\partial \zeta_{ijk}}{\partial \alpha}^\top \frac{M_{ijk}}{S_{ijk}}, \tag{5}$$

5

where

$$\zeta_{ijk} \quad = \quad E[Y_{ij}|Y_{ik}] = \mu_{ij} + \frac{\sigma_{ijk}}{\sigma_{ikk}}(Y_{ik} - \mu_{ik}),$$

$\sigma_{ijk} = \text{cov}(Y_{ij}, Y_{ik}) = \mu_{ijk} - \mu_{ij}\mu_{ik}$, $M_{ijk} = Y_{ij} - \zeta_{ijk}$, and $S_{ijk} = \text{var}(Y_{ij}|Y_{ik}) = \zeta_{ijk}(1 - \zeta_{ijk})$.

Lipsitz & Fitzmaurice (1996) showed that (5) can be used to model correlations while Klar, Lipsitz & Ibrahim (1993) used it to estimate models for rater agreement through the kappa coefficient.

In what follows, let M_i denote the vector with components M_{ijk} and S_i denote the diagonal matrix with diagonal elements S_{ijk}.

Note that the matrix S_i is stochastic and does not consist of the diagonal elements of any genuine covariance matrix; clearly $\text{var}(M_{ijk}) \neq \zeta_{ijk}(1 - \zeta_{ijk})$. Stochastic covariance matrices in estimating equations are feasible (Heyde, 1997, section 2.6) in the context of nested sigma fields leading to a martingale structure, but that is not the case with (5). The stochastic nature of S_i and $\partial\zeta_i/\partial\alpha$ makes theoretical investigation of (5) through standard estimating equation theory not possible. Another point is that although $U_{\alpha,ALR}$ is invariant to permutations of the Y_i vector (Carey, 1992; Kuk, 2004) the associated robust variance estimator is not. In the current SAS software, version 9.2, the robust variance estimator is averaged over estimators obtained from the original y_i and a reversed version of y_i (personal communication with Vincent Carey and with Gordon Johnston at SAS Institute). While

6

this approach achieves invariance to the permutations of Y_i, it is unclear if this approximation is appropriate.

Asymptotic efficiency calculations reported by Carey et al. (1993) show $U_{\alpha,ALR}$ to be nearly as efficient as $U_{\alpha,GEE2}$. The calculations were limited to equal size clusters, $n_i = 4$, with a common covariate pattern used for all clusters. Lipsitz & Fitzmaurice (1996) found that $U_{\alpha,ALR}$ is more efficient than methods that rely on (3) or (4), especially when the pairwise correlation is high or when cluster size is variable. However, their efficiency calculations were limited to the case $n_i \leq 3$.

2 Orthogonalized Residuals

2.1 Genesis

The orthogonalized residuals approach is based on two ideas. First, pairwise residuals are developed via a projection argument. Second, a weighted combination of these residuals is formed using an approximate covariance matrix that is still computationally feasible for larger clusters. Let $R_{iYW} = \text{corr}((Y_i^\top, W_i^\top)^\top)$ and $R_{iWW} = \text{corr}(W_i)$. The matrix R_{iYW} has elements of the form $\text{corr}(Y_{ij'}, W_{ijk})$. It is natural to expect elements with $j' = j$ or $j' = k$ to be largest in magnitude. To eliminate these correlations, the orthogonalized residuals approach utilizes the residuals from the linear regressions of

7

the W_{ijk} on Y_{ij} and Y_{ik}. Specifically, $Q_{ijk} =$

$$W_{ijk} - \{\mu_{ijk} + b_{ijk:j}(Y_{ij} - \mu_{ij}) + b_{ijk:k}(Y_{ik} - \mu_{ik})\},\qquad(6)$$

where

$$
\begin{aligned}
b_{ijk:j} &= \mu_{ijk}(1 - \mu_{ik})(\mu_{ik} - \mu_{ijk})/d_{ijk}, \\
b_{ijk:k} &= \mu_{ijk}(1 - \mu_{ij})(\mu_{ij} - \mu_{ijk})/d_{ijk}, \\
d_{ijk} &= \sigma_{ijj}\sigma_{ikk} - \sigma_{ijk}^2.
\end{aligned}
$$

It follows that $\mathrm{corr}(Y_{ij}, Q_{ijk}) = \mathrm{corr}(Y_{ik}, Q_{ijk}) = 0$, so this definition of Q_{ijk} introduces $n_i - 1$ zeros into each row of $R_{iYQ} = \mathrm{corr}((Y_i^\top, Q_i^\top)^\top)$, where Q_i is an m_i-vector with elements Q_{ijk}. In addition, we have observed that this construction tends to reduce the magnitude of the other entries in R_{iYQ} as compared to R_{iYW}, and also the magnitude of the off-diagonal elements in $R_{iQQ} = \mathrm{corr}(Q_i)$ as compared R_{iWW}. A numerical example is given below.

The orthogonalized residuals estimating equation for the marginal association parameters is

$$U_{\alpha,ORTH} = \sum_{i=1}^{K} E\left[\frac{-\partial Q_i}{\partial\alpha}^\top\right] P_i^{-1} Q_i = \sum_{i=1}^{K} C_i^\top P_i^{-1} Q_i,\qquad(7)$$

where P_i is a diagonal matrix with elements $v_{ijk} = \mathrm{var}(Q_{ijk}) =$

$$\frac{\mu_{ijk}(\mu_{ij} - \mu_{ijk})(\mu_{ik} - \mu_{ijk})(1 - \mu_{ij} - \mu_{ik} + \mu_{ijk})}{\mu_{ij}\mu_{ik}(1 - \mu_{ij} - \mu_{ik} + 2\mu_{ijk}) - \mu_{ijk}^2}.$$

8

The estimating equation for β is $U_{\beta,ORTH} = U_{\beta,GEE1}$. The computational advantage of the diagonal structure for P_i is that a simple explicit inverse exists, and matrices of dimension $m_i \times m_i$ need never be formed in computer memory. The computational effort is identical to that for (3), (4) and (5). Computation proceeds by iteratively reweighted least squares.

Following arguments similar to Prentice (1988) and Liang & Zeger (1986), the asymptotic distribution of $K^{\frac{1}{2}}(\hat{\theta} - \theta)$ is multivariate Gaussian with mean zero and covariance matrix consistently estimated by $KL^{-1}\Lambda L^{-\top}$ where L and Λ consist of the following blocks

$$
\begin{aligned}
L_{11} &= \sum_{i=1}^{K} \hat{D}_i^\top \hat{V}_i^{-1} \hat{D}_i, \\
L_{12} &= 0, \\
L_{21} &= -\sum_{i=1}^{K} \hat{C}_i^\top \hat{P}_i^{-1} \hat{E}\left[\frac{\partial Q_i}{\partial \beta}\right], \\
L_{22} &= \sum_{i=1}^{K} \hat{C}_i^\top \hat{P}_i^{-1} \hat{C}_i, \\
\Lambda_{11} &= \sum_{i=1}^{K} \hat{D}_i^\top \hat{V}_i^{-1} \tilde{\text{cov}}(Y_i) \hat{V}_i^{-1} \hat{D}_i, \\
\Lambda_{12} &= \sum_{i=1}^{K} \hat{D}_i^\top \hat{V}_i^{-1} \tilde{\text{cov}}((Y_i^\top, Q_i^\top)^\top) \hat{P}_i^{-1} \hat{C}_i, \\
\Lambda_{21} &= \Lambda_{12}^\top, \\
\Lambda_{22} &= \sum_{i=1}^{K} \hat{C}_i^\top \hat{P}_i^{-1} \tilde{\text{cov}}(Q_i) \hat{P}_i^{-1} \hat{C}_i,
\end{aligned}
$$

where hats denote evaluation at $\hat{\theta}$, $\tilde{\text{cov}}(Y_i) = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^\top$, $\tilde{\text{cov}}((Y_i^\top, Q_i^\top)^\top) =$

9

$(Y_i - \hat{\mu}_i)\hat{Q}_i^\top$, and $\tilde{\text{cov}}(Q_i) = \hat{Q}_i\hat{Q}_i^\top$. It is clear from (6) that $Q_{ijk} = Q_{ikj}$ which implies that both $U_{\alpha,ORTH}$ and its associated robust variance estimator $KL^{-1}\Lambda L^{-\top}$, are invariant to permutations of the data y_i.

It is shown in the appendix that orthogonalized residuals is equivalent to alternating logistic regressions, and is true for all link functions g_1 and g_2. However, the formulation of (7) offers the advantage that it follows a standard estimating equation approach. Thus it resolves the difficulties mentioned above with the formulation of alternating logistic regressions and offers insight into their efficiency behaviour. A practical advantage of $U_{\alpha,ORTH}$ is that the associated robust variance estimator is invariant to permutations of y_i.

2.2 Example

The effectiveness of orthogonalization is illustrated using data from the 6-City Study (Ware et al., 1984). The response vector Y_i consists of $n_i = 4$ binary observations per child, indicating respiratory illness at ages 7–10. Only data from the 350 children with non-smoking mothers are used. The data are summarized in Table 1.

The 16 observed proportions, $(237/350, \cdots, 11/350)$, are used as the true distribution under which the correlation matrices presented below are calcu-

10

Table 1: Summary of 6-City Outcomes

y_1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
y_2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
y_3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
y_4	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Count	237	24	16	6	15	3	7	5	10	3	2	2	4	2	3	11

lated. The correlation between the residuals $Y_i - \mu_i$ and $W_i - \delta_i$ is

$$
R_{iYW} = \begin{bmatrix}
\mathbf{0.62} & \mathbf{0.58} & \mathbf{0.53} & 0.35 & 0.36 & 0.33 \\
\mathbf{0.65} & 0.44 & 0.38 & \mathbf{0.68} & \mathbf{0.56} & 0.38 \\
0.41 & \mathbf{0.62} & 0.39 & \mathbf{0.69} & 0.42 & \mathbf{0.60} \\
0.38 & 0.42 & \mathbf{0.68} & 0.40 & \mathbf{0.68} & \mathbf{0.72}
\end{bmatrix}
$$

The largest entries, bolded, are those of the type $\mathrm{corr}(Y_{ij'}, W_{ijk})$ where $j' = j$ or k with an average of 0.63. The average of the remaining correlations is 0.39. In contrast, the correlation between $Y_i - \mu_i$ and the orthogonalized residuals Q_i is

$$
R_{iYQ} = \begin{bmatrix}
\mathbf{0} & \mathbf{0} & \mathbf{0} & 0.08 & 0.09 & 0.08 \\
\mathbf{0} & 0.11 & 0.10 & \mathbf{0} & \mathbf{0} & 0.06 \\
0.10 & \mathbf{0} & 0.09 & \mathbf{0} & 0.07 & \mathbf{0} \\
0.14 & 0.14 & \mathbf{0} & 0.09 & \mathbf{0} & \mathbf{0}
\end{bmatrix}
$$

The construction of Q_i introduces $n_i - 1 = 3$ zeros into each row of R_{iYQ}.

11

Remarkably, the other correlations have gone down considerably; from an average of 0.39 to 0.10. Overall, the average entry has gone down from 0.51 to 0.05. This shows that orthogonalization is quite effective in achieving approximate orthogonality between the two sets of residuals. An added benefit occurs in R_{iQQ}. The matrix R_{iWW} has off-diagonal elements ranging from 0.47 to 0.72, and averaging 0.62. By comparison, for R_{iQQ}, the range is 0.15 to 0.44, and the average is 0.30.

3   Application

The orthogonalized residuals approach was applied to data from parents and siblings of subjects with chronic obstructive pulmonary disease (COPD) and their controls (Cohen, 1980). The outcome of interest is impaired pulmonary function. The model for the marginal mean is the same as that used in Qaqish & Liang (1992) and includes the covariates: intercept, sex, race, age centered at 50, smoking staus and an indicator as to whether the subject was a relative of someone with COPD or a control. Associations are modeled through log odds ratios with distinct parameters for each familial relationship: parent-parent (α_{pp}), sibling-sibling (α_{ss}) or parent-sibling (α_{ps}).

Table 2 displays estimates obtained from GEE2 (Qaqish & Liang, 1992), alternating logistic regressions from SAS GENMOD (ALR; SAS Institute Inc.) and by orthogonalized residuals(ORTH). Results for the parameter estimates for ALR and ORTH are identical by definition; however, the stan-

12

Table 2: Association Parameter Estimates and Empirical Standard Errors

Model	Parent-Parent(α_{pp})	Sibling-Sibling(α_{ss})	Parent-Sibling(α_{ps})
GEE2	-1.09	0.873	0.984
	(1.09)	(0.565)	(0.519)
ORTH	-1.177	1.108	0.986
	(1.138)	(0.764)	(0.673)
ALR	-1.177	1.108	0.986
	(1.385)	(0.603)	(0.717)

Results are parameter (stderr). GEE2 from Qaqish & Liang (1992).

dard errors for α_{pp} and α_{ss} are quite different due to the approximation used within GENMOD. In fact, ALR under- or overestimates the standard errors computed by ORTH by as much as 21% or 22%, respectively.

4   Discussion

This paper described a new method for estimation of association parameters for correlated binary responses. Orthogonalized residuals offer a new representation of alternating logistic regressions through marginal residuals, and thus provide insight into the alternating logistic regressions estimating equations. Further, orthogonalized residuals eliminate the need for the approximation to the robust variance estimator currently used for alternating logistic regressions.

An additional consequence of the formulation of alternating logistic regressions in Carey et al. (1993) is that it is not clear how to allow a non-

13

diagonal S_i to improve efficiency for the marginal association parameters. However, orthogonalized residuals does not have this difficulty. It is possible to expand the diagonal structure of P_i by assuming a working correlation matrix $R^*_{iQQ}(\lambda)$ for R_{iQQ} and approximating $\text{cov}(Q_i)$ using

$$P_i := \text{diag}(v_{ijk}^{\frac{1}{2}}) R^*_{iQQ}(\lambda) \text{diag}(v_{ijk}^{\frac{1}{2}}),$$

where λ is an r-vector of nuisance parameters to be estimated. One particular structure of interest would have two correlation parameters; one for pairs (j, k) and (j', k') that share an index and another for the case where all indices are distinct. However, complex formulations of $R^*_{iQQ}(\lambda)$ will be limited to clusters of small to moderate size since it is necessary to invert P_i. A possibility for clusters of any size is an exchangeable structure of $R^*_{iQQ}(\lambda)$. Exploring these methods is a topic of future research.

Acknowledgements

14
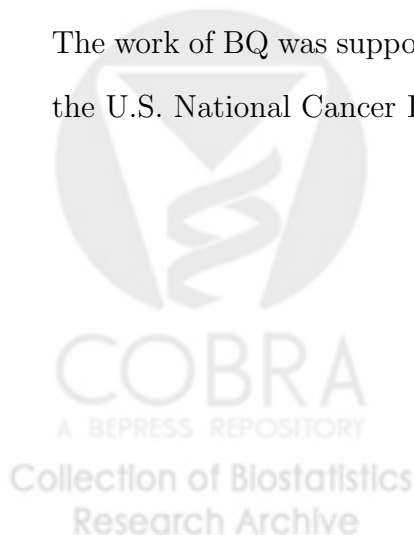
References

Carey, V. (1992). *Regression Analysis of Large Binary Clusters.* Ph.D. Dissertation, The Johns Hopkins University.

Carey, V., Zeger, S.L. & Diggle, P. (1993). Modeling multivariate binary data with alternating logistic regressions. *Biometrika* **80**, 517–26.

Cohen, B.H. (1980). Chronic pulmonary disease: a challenge in genetic epidemiology. Am. J. Epidemiol. **112**, 274–88.

Heagerty, P.J., & Zeger, S.L. (2000). Marginalized multilevel models and likelihood inference. *Statist. Sci.* **15**, 1–19.

Heyde, C.C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation.* Springer Verlag, New York.

Klar, N., Lipsitz, S.R. & Ibrahim, J.G. (2000). An estimating equations approach for modeling kappa. *Biomet. J.* **42**, 45–58.

Kuk, A.Y.C. (2004). Permutation invariance of alternating logistic regression for multivariate binary data. *Biometrika* **91**, 758–61.

Liang, K.Y. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Liang, K.Y., Zeger, S.L. & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *J. R. Statist. Soc.* B **54**, 3–40.

15

Lipsitz, S.R., Laird, N.M. & Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* **78**, 153–60.

Lipsitz, S.R. & Fitzmaurice, G.M. (1996). Estimating equations for measures of association between repeated binary responses. *Biometrics* **52**, 903–12.

Neuhaus, J. M., Kalbfleish, J. D. & Hauk, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int. Statist. Rev.* **59**, 25-35.

Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**, 1033–48.

Qaqish B.F. & Liang, K.Y.(1992). Marginal models for correlated binary responses with multiple classes and multiple levels of nesting. *Biometrics* **48**, 939–50.

Ware, J.H., Dockery, D.W., Spiro, A., Speizer, F.E. & Ferris, B.G. (1984). Passive smoking, gas cooking, and respiratory health of children living in six cities. *Am. Rev. Resp. Dis.* **129**, 366–74.

Zeger, S.L. Liang, K.Y. & Albert, P. (1988). Models for longitudinal data: A generalized estimating equations approach. *Biometrics* **44**, 1049–60.

Appendix

Proof that the orthogonalized residuals estimating equation is equivalent to alternating logistic regressions. Since U_β is the same for both approaches, it suffices to show that $U_{\alpha,ALR} = U_{\alpha,ORTH}$ where

$$U_{\alpha,ALR} = \sum_{i=1}^{K} \sum_{j<k} \frac{\partial \zeta_{ijk}}{\partial \alpha}^\top \frac{Y_{ij} - \zeta_{ijk}}{\zeta_{ijk}(1 - \zeta_{ijk})}$$

and

$$U_{\alpha,ORTH} = \sum_{i=1}^{K} \sum_{j<k} \frac{\partial \mu_{ijk}}{\partial \alpha}^\top \frac{Q_{ijk}}{v_{ijk}}.$$

Further, it suffices to show that the contributions from each (j, k) pair are equal. Note that $\mu_{ijk} = \mu_{ij}\mu_{ik} + \rho_{ijk}(\sigma_{ijj}\sigma_{ikk})^{\frac{1}{2}}$ and

$$\zeta_{ijk} = \mu_{ij} + \frac{\sigma_{ijk}}{\sigma_{ikk}}(Y_{ik} - \mu_{ik}) = \mu_{ij} + \rho_{ijk}(\sigma_{ijj}\sigma_{ikk})^{\frac{1}{2}} \frac{(Y_{ik} - \mu_{ik})}{\sigma_{ikk}}.$$

Writing

$$
\begin{aligned}
\frac{\partial \zeta_{ijk}}{\partial \alpha} &= \frac{\partial \rho_{ijk}}{\partial \alpha} \frac{\partial \zeta_{ijk}}{\partial \rho_{ijk}} = \frac{\partial \rho_{ijk}}{\partial \alpha}(\sigma_{ijj}\sigma_{ikk})^{\frac{1}{2}} \frac{(Y_{ik} - \mu_{ik})}{\sigma_{ikk}} \\
&= \frac{\partial \rho_{ijk}}{\partial \alpha} \frac{\partial \mu_{ijk}}{\partial \rho_{ijk}} \frac{(Y_{ik} - \mu_{ik})}{\sigma_{ikk}} = \frac{\partial \mu_{ijk}}{\partial \alpha} \frac{(Y_{ik} - \mu_{ik})}{\sigma_{ikk}},
\end{aligned}
$$

it follows that the contributions are equal if

$$\frac{Y_{ik} - \mu_{ik}}{\sigma_{ikk}} \frac{Y_{ij} - \zeta_{ijk}}{\zeta_{ijk}(1 - \zeta_{ijk})} = \frac{Q_{ijk}}{v_{ijk}}.$$

17

Straightforward algebra shows that the above equality is true for each of

the four possible patterns of (Y_{ij}, Y_{ik}), that is, for

$(Y_{ij}, Y_{ik}) \in \{(0,0),(0,1),(1,0),(1,1)\}.$

18