NIH Public Access
Author Manuscript
Scand Stat Theory Appl. Author manuscript; available in PMC 2013 September 01.

Published in final edited form as:
Scand Stat Theory Appl. 2012 September 1; 39(3): 515–527. doi:10.1111/j.1467-9469.2012.00802.x.

Orthogonalized residuals for estimation of marginally specified association parameters in multivariate binary data

BAHJAT F. QAQISH,
Department of Biostatistics, University of North Carolina at Chapel Hill

RICHARD C. ZINK, and
JMP Life Sciences, SAS Institute, Inc

JOHN S. PREISSER
Department of Biostatistics, University of North Carolina at Chapel Hill

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. A special case of the proposed procedure allows a new representation of the alternating logistic regressions method through marginal residuals. The connections between second-order generalized estimating equations, alternating logistic regressions, pseudo-likelihood and other methods are explored. Efficiency comparisons are presented, with emphasis on variable cluster size and on the role of higher-order assumptions. 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; pairwise pseudo-likelihood

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, $1 \le j < k \le n_i$, 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

Corresponding Author: Bahjat F. Qaqish, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. qaqish@bios.unc.edu.

Supporting Information
Additional Supporting Information may be found in the online version of this article. The materials include a description of the efficient computation of orthogonalized residuals, a table of binomial probabilities, and figures describing additional asymptotic efficiency results. The first figure describes efficiency for marginal mean parameters while the second figure is a version of Figure 1 from the manuscript restricting ρ to [0,0.15] for better resolution.

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

the correlation coefficient, ρ_{ijk} = 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. Finally, define θ be the $(p + q)$-vector $(\beta^{\top}, \alpha^{\top})^{\top}$ and note that the covariance matrix Σ_i = 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 et al. (1988), Neuhaus et al. (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 (Liang et al., 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 equations are

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

where $W_i = (W_{i12}, \cdots, W_{i;n_i-1;n_i})^{\top}$, $\delta_i = E[W_i]$, $D_i = \partial\mu_i/\partial\beta$, $A_i = \partial\delta_i/\partial\beta$, $C_i = \partial\delta_i/\partial\alpha$ and $\sum_i^{*}=\mathrm{cov}((Y_i^{\top}, W_i^{\top})^{\top})$. The matrix $\sum_i^{*}$ involves third- and fourth-order cross-moments not specified by the marginal model. A working version of $\sum_i^{*}$ is obtained by assuming that third- and fourth-order logistic contrasts are zero (Liang et al., 1992). Prentice & Zhao (1990) presented an alternative formulation to (1) for maximum likelihood estimation under a quadratic exponential model based on fixing higher-order moments.

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

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

where

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

σ_{ijj} = var(Y_{ij}) = $\mu_{ij}(1 - \mu_{ij})$ and R_i = 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.

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,

$$U_\alpha=\sum_{i=1}^{K}\sum_{j<k}\kappa_{ijk}. \quad (3)$$

Prentice (1988) suggested the pairwise kernel

$$\kappa_{ijk}=-\frac{\partial T_{ijk}}{\partial\alpha}^{\top}\frac{T_{ijk}}{\text{var}(T_{ijk})} \quad (4)$$

where

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

The resulting estimating function will be denoted $U_{\alpha;P}$.

Lipsitz et al. (1991) developed an estimating function, here denoted $U_{\alpha;L}$, using the kernel

$$\kappa_{ijk}=\frac{\partial\delta_{ijk}}{\partial\alpha}^{\top}\frac{W_{ijk}-\delta_{ijk}}{\text{var}(W_{ijk})}. \quad (5)$$

As an alternative to generalized estimating equations, since the joint distribution of any pair (Y_{ij}, Y_{ik}) is completely determined by θ, it is possible to define a log pseudo-likelihood for the i-th cluster as $l_i(\theta) = \Sigma_{j<k}\, l_{ijk}(\theta)$, where $l_{ijk}(\theta)$ = log pr($Y_{ij} = y_{ij}$, $Y_{ik} = y_{ik}$; θ). Kuk & Nott (2000) suggested the kernel $\kappa_{ijk} = \partial l_{ijk}/\partial\alpha$ while le Cessie & van Houwelingen (1994) suggested a weighted version, $\kappa_{ijk} = \{1/(n_i - 1)\}\ \partial l_{ijk}/\partial\alpha$. We denote these estimating functions as $U_{\alpha;KN}$ and $U_{\alpha;CH}$, respectively. Both Kuk & Nott (2000) and le Cessie & van Houwelingen (1994) suggested

$$U_{\beta,KN} = U_{\beta,CH} = \sum_{i=1}^{K} \frac{1}{n_i - 1} \sum_{j<k} \frac{\partial l_{ijk}}{\partial \beta}$$

for estimation of β. Based on efficiency studies with $n_i \le 6$, Geys et al. (1998) seem to favor $U_{\alpha;KN}$ over $U_{\alpha;CH}$.

One final method, along with $U_{\beta;ALR} = U_{\beta;GEE1}$, defines $U_{\alpha;ALR}$ using a pairwise kernel in (3) based on conditional residuals

$$\kappa_{ijk} = \frac{\partial \zeta_{ijk}}{\partial \alpha}^{\top} \frac{M_{ijk}}{S_{ijk}}, \quad (6)$$

where

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

$\sigma_{ijk} = \mathrm{cov}(Y_{ij}, Y_{ik}) = \mu_{ijk} - \mu_{ij}\mu_{ik}$, $M_{ijk} = Y_{ij} - \zeta_{ijk}$, and $S_{ijk} = \mathrm{var}(Y_{ij} | Y_{ik}) = \zeta_{ijk}(1 - \zeta_{ijk})$. The original formulation of $U_{\alpha;ALR}$, known as alternating logistic regressions (Carey et al., 1993), models pairwise associations in terms of $g_2(\psi_{ijk}) = \log(\psi_{ijk})$. 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 $\mathrm{var}(M_{ijk}) \ne \zeta_{ijk}(1 - \zeta_{ijk})$. 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 (6). An important consequence is that it is not clear how to allow a non-diagonal S_i in order to improve efficiency. Further, the stochastic nature of S_i and $\partial \zeta_i / \partial \alpha$ makes theoretical investigation of (6) through standard estimating equation theory not possible. Another point is that while $U_{\alpha;ALR}$ is invariant to permutations of the Y_i vector (Kuk, 2004) the associated robust variance estimator is not. In SAS 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).

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), who modeled pairwise associations in terms of ρ_{ijk}, found that $U_{\alpha;ALR}$ is more efficient than methods that rely on (4) or (5), especially when the pairwise correlation is high or when cluster size is variable. However, their efficiency calculations were limited to the case $n_i \le 3$.

Before concluding this section, it is worth noting a connection among the methods. By expressing l_{ijk} as $l_{ijk}(\theta) = \log \mathrm{pr}(Y_{ik} = y_{ik}; \beta) + \log \mathrm{pr}(Y_{ij} = y_{ij} | Y_{ik} = y_{ik}; \beta, \alpha)$ and differentiating with respect to α it becomes clear that $U_{\alpha;KN}$ is identical to $U_{\alpha;ALR}$. However, note that $U_{\beta;ALR} = U_{\beta;GEE1} \ne U_{\beta;CH} = U_{\beta;KN}$. Because of the relation of ALR to pairwise likelihood, Kuk (2004, 2007) refers to ALR as a hybrid pairwise likelihood.

2. Orthogonalized residuals

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

$$\mathrm{corr}\left(\left(Y_i^\top, W_i^\top\right)^\top\right)=\begin{pmatrix} R_i & R_{iYW} \\ R_{iYW}^\top & R_{iWW}\end{pmatrix}, \quad (7)$$

where R_{iWW} = corr(W_i) and R_{iYW} has elements of the form 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 the W_{ijk} on Y_{ij} and Y_{ik}. Specifically,

$$Q_{ijk}=W_{ijk}-\{\mu_{ijk}+b_{ijk:j}(Y_{ij}-\mu_{ij})+b_{ijk:k}(Y_{ik}-\mu_{ik})\}, \quad (8)$$

where $b_{ijk:j} = \mu_{ijk}(1 - \mu_{ik})(\mu_{ik} - \mu_{ijk})/d_{ijk}$, $b_{ijk:k} = \mu_{ijk}(1 - \mu_{ij})(\mu_{ij} - \mu_{ijk})/d_{ijk}$, $d_{ijk}=\sigma_{ijj}\sigma_{ikk}-\sigma_{ijk}^2$. It follows that corr($Y_{ij}$, Q_{ijk}) = corr(Y_{ik}, Q_{ijk}) = 0, so this definition of Q_{ijk} introduces $n_i - 1$ zeros into each row of the matrix R_{iYQ} which has elements of the form corr($Y_{ij'}$, Q_{ijk}), where Q_i is an m_i-vector with elements Q_{ijk} taking the place of W_i in (7). 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} = corr(Q_i) as compared R_{iWW}. A numerical example is given below.

The second aspect of the orthogonalized residuals approach is to approximate R_{iQQ} by an exchangeable working correlation matrix

$$R_{iQQ}^*(\lambda)=\lambda J+(1-\lambda)I, \quad (9)$$

where λ is a nuisance parameter to be estimated, I is the identity matrix and J is a matrix of 1's, both of order $m_i \times m_i$. Thus cov(Q_i) is approximated by

$$P_i=\mathrm{diag}(v_{ijk}^{\frac{1}{2}})R_{iQQ}^*(\lambda)\mathrm{diag}(v_{ijk}^{\frac{1}{2}}),$$

where

$$v_{ijk}=\mathrm{var}(Q_{ijk})=\frac{\mu_{ijk}(\mu_{ij}-\mu_{ijk})(\mu_{ik}-\mu_{ijk})(1-\mu_{ij}-\mu_{ik}+\mu_{ijk})}{\mu_{ij}\mu_{ik}(1-\mu_{ij}+\mu_{ik}+2\mu_{ijk})-\mu_{ijk}^2}.$$

With the above definitions of Q_i and P_i, the orthogonalized residuals estimating equation for the marginal association parameters is

$$U_{\alpha,ORTH}=\sum_{i=1}^{K} E\left[\frac{-\partial Q_i^{\top}}{\partial \alpha}\right] P_i^{-1} Q_i=\sum_{i=1}^{K} C_i^{\top} P_i^{-1}, Q_i \quad (10)$$

where C_i is as defined in (1). The estimating equation for β is $U_{\beta;ORTH} = U_{\beta;GEE1}$. The computational advantage of the exchangeable structure in (9) 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 virtually identical to that for (4), (5) and (6). Details are given in the Supplementary Appendix.

Computation proceeds by iteratively reweighted least squares with the estimate of λ updated in each iteration. A simple moment estimator of λ is

$$\widehat{\lambda}=\widehat{\lambda}(\theta)=\frac{\sum_{i=1}^{K}\left\{\left(\sum_{j<k}\frac{Q_{ijk}}{v_{ijk}^{\frac{1}{2}}}\right)^2-\sum_{j<k}\frac{Q_{ijk}^2}{v_{ijk}}\right\}}{\sum_{i=1}^{K} m_i(m_i-1)}. \quad (11)$$

Following arguments similar to Prentice (1988) and Liang & Zeger (1986), the asymptotic distribution of $K^{\frac{1}{2}}(\widehat{\theta}-\theta)$ 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

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

where hats denote evaluation at $(\theta\mathcal{O}, \hat{\lambda}(\hat{\theta}))$, $\tilde{\text{cov}}(Y_i) = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^{\top}$, $\tilde{\text{cov}}(Y_i,Q_i)=(Y_i-\widehat{\mu}_i)\widehat{Q}_i^{\top}$ and $\tilde{\text{cov}}(Q_i)=\widehat{Q}_i\widehat{Q}_i^{\top}$. The requirement is that $\hat{\lambda}$ is a $K^{\frac{1}{2}}$-consistent estimator of λ^*, the limiting value of the average off-diagonal element of R_{iQQ}, assumed to exist. It is clear from (8) 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.

A special case, to be denoted $U_{\alpha;ORTH0}$, ensues if λ in (10) is not estimated, but rather fixed at zero, so that P_i becomes a diagonal matrix. It is shown in Appendix 1 that, for any pair of link functions (g_1, g_2), this special case is equivalent to $U_{\alpha;ALR}$. However, the formulation of (10) 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 $U_{\alpha;ORTH0}$ is that the associated robust variance estimator is invariant to permutations of y_i.

By et al. (2011) describe software for $U_{\beta;GEE1}$ in combination with $U_{\alpha;ORTH}$ or $U_{\alpha;ORTH0}$ for both R (By et al., 2008) and SAS (macro available at http://www.bios.unc.edu/~qaqish/software.htm).

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, ⋯, 11/350), are used as the true distribution under which the correlation matrices presented below are calculated.

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 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}. 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. Finally, the estimated value of λ from (11) is $\hat{\lambda} = 0.2805$.

3. Efficiency Comparisons

The asymptotic efficiency of several estimating functions relative to U_{GEE2} is evaluated for the simple model of a common mean and a common pairwise correlation and unequal cluster size:

$$\begin{aligned} E[Y_{ij}]&=\mu \quad (1 \le j \le n_i), \\ \mathrm{corr}(Y_{ij}, Y_{ik})&=\rho \quad (1 \le j<k \le n_i). \end{aligned} \tag{12}$$

The model implies a common pairwise odds ratio ψ. The alternating logistic regressions estimating equations are

$$U_{\mu,ALR} = U_{\mu,GEE1} = \sum_{i=1}^{K} \frac{n_i}{1+\rho(n_i-1)}(\overline{Y}_i - \mu),$$
$$U_{\rho,ALR} = \sum_{i=1}^{K} U_{i,\rho,ALR} = \sum_{i=1}^{K} \sum_{j<k} \frac{\{Y_{ij}-\mu-\rho(Y_{ik}-\mu)\}\{Y_{ik}-\mu\}}{\{\mu+\rho(Y_{ik}-\mu)\}\{1-\mu-\rho(Y_{ik}-\mu)\}}$$
$$= \sum_{i=1}^{K} \left\{ \frac{G_{i0}\mu}{1-\mu(1-\rho)} - \frac{G_{i1}}{1-\rho} + \frac{G_{i2}(1-\mu)}{1-(1-\mu)(1-\rho)} \right\},$$

where $\overline{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $G_{it} = \#\{(j, k) : 1 \le j < k \le n_i,\ Y_{ij} + Y_{ik} = t\}$. Note that the G_{it} are simple quadratic functions of cluster totals $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$, $G_{i0} = (n_i - y_{i.})(n_i - y_{i.} - 1)/2$, $G_{i1} = y_{i.}(n_i - y_{i.} - 1)$ and $G_{i2} = y_{i.}(y_{i.} - 1)/2$.

For example, $U_{\rho;ALR} = a_0 G_{i0} + a_1 G_{i1} + a_2 G_{i2}$, where the coefficients (a_0, a_1, a_2) are the kernels for $U_{\rho;ALR}$ under model (12) given in Table 2. Hence, $U_{\rho;ALR}$ can be expressed as $U_{\rho;ALR} = b_0 + b_1 y_{i.} + b_2 y_{i.}^2$, where the coefficients (b_0, b_1, b_2) are functions of μ, ρ and n_i. Similarly, the U_μ component can be expressed in the form $c_0 + c_1 y_{i.} + c_2 y_{i.}^2$, with coefficients depending on the specific procedure. For example, in $U_{\mu;ALR}$, $c_0 = -n_i\mu/\{1 + \rho\,(n_i - 1)\}$, $c_1 = 1/\{1 + \rho\,(n_i - 1)\}$ and $c_2 = 0$.

All the estimating functions discussed in this section have a similar structure, but with different choices of the coefficients (b_0, b_1, b_2). The estimating equations for orthogonalized residuals are $U_{\mu;ORTH} = U_{\mu;ALR}$ and $U_{\rho,ORTH} = \sum_{i=1}^{K} (1+(m_i-1)\lambda)^{-1} U_{i,\rho,ALR}$. The pairwise log pseudo-likelihood $l_i(\mu, \rho) = G_{i0} \log\{(1-\mu)^2 + \rho\mu\,(1-\mu)\} + G_{i1} \log\{\mu\,(1-\mu\,(1-\rho)\} + G_{i2} \log\{\mu^2 + \rho\mu\,(1-\mu)\}$ is used to derive the weighted estimating functions of le Cessie & van Houwelingen (1994), Kuk & Nott (2000) and the unweighted version, U_{PL}.

The shared structure of the estimating equations facilitates calculation of asymptotic efficiencies. Efficiency in relation to estimation of the log odds ratio log(Ψ), which can be expressed as a function of μ and ρ, is considered below; efficiency for estimation of β is relegated to the Supplementary Appendix. Since (1) is the optimal quadratic estimating function it will be used as the reference. For each estimation procedure, the asymptotic variance matrix is computed as the inverse of the efficiency matrix (Morton, 1981) and used to compute the relative efficiencies (additional details are provided in Appendix 2). Since all estimating functions under consideration are quadratic in $Y_{i:}$, the efficiency matrices must involve third and fourth moments of $Y_{i:}$. As these moments are not specified by the model (12), additional assumptions are required. Efficiency calculations are done under three forms for the distribution of $Y_{i:}$. The first is the beta-binomial (Skellam, 1948)

$$\mathrm{pr}(Y_{i.} = t; \mu, \rho) = \binom{n_i}{t} \prod_{j=0}^{t-1} (\mu + j\tau) \prod_{j=0}^{n_i-t-1} (1-\mu+j\tau) \Big/ \prod_{j=0}^{n_i-1} (1+j\tau),$$

where $\tau = \rho/(1-\rho)$. The second form is the mixture (Morel & Neerchal, 1997)

$$Y_{i.} \sim \begin{cases} \mathrm{Bin}(n_i, \rho^{\frac{1}{2}} + \mu(1-\rho^{\frac{1}{2}})) & \text{with probability } \mu, \\ \mathrm{Bin}(n_i, \mu(1-\rho^{\frac{1}{2}})) & \text{with probability } 1-\mu. \end{cases}$$

The third form is the mixture (Madsen, 1993)

$$Y_{i.} \sim \begin{cases} \mathrm{Bin}(n_i, \mu) & \text{with probability } 1-\rho, \\ n_i\, \mathrm{Bin}(1, \mu) & \text{with probability } \rho. \end{cases}$$

All three distributions reduce to the binomial when $\rho = 0$. As an example, the table in the Supplementary Appendix shows $P(Y_{i\cdot} = t)$ for the binomial and the three distributions with $n_i = 10$, $\mu = 0.3$, $\rho = 0.3$. Note that since we assume a constant μ, the upper bound on ρ is 1 (Chaganty & Joe, 2006).

Cluster size, n_i, is taken to be a 1:1 mix of $n_i = 5$ and $n_i = 25$. The overall efficiency pattern was similar for different values of μ, so only results for $\mu = 0.2$ are presented.

Figure 1 shows the efficiency for the pairwise log odds ratio. It shows that $U_{\alpha;ORTH}$, equation (10), is nearly fully efficient for all values of ρ. Figure 1 shows clearly that among the remaining procedures, no single one uniformly dominates the others for all values of ρ. It also shows that, by incorporating a weight matrix, $U_{\alpha;ORTH}$ gains considerable efficiency over $U_{\alpha;ALR}$. All procedures except $U_{\alpha;CH}$ and $U_{\alpha;L}$ are fully efficient at $\rho = 0$, but their efficiencies drop precipitously for even moderate values of ρ. Procedure $U_{\alpha;CH}$ stands out by not being fully efficient at $\rho = 0$. Its efficiency peaks quickly to drop again under the beta-binomial and Madsen models, but it tracks $U_{\alpha;ORTH}$ closely for $\rho > 0.4$ under the Morel-Neerchal model. A figure restricting $\rho \in [0, 0.15]$ for better resolution is left to the Supplementary Appendix.

One criticism of Figure 1 is that the value of λ that goes into $U_{\alpha;ORTH}$ depends on the true distribution, which in practical applications is unknown. For this reason, Figure 2 shows the efficiency curves for $U_{\alpha;ORTH}$ under the nine combinations of true and assumed models. The three plots in the middle column correspond to the Morel-Neerchal working model. Figure 2 shows that under that working model, $U_{\alpha;ORTH}$ achieves considerable efficiency gains over $U_{\alpha;ALR}$.

4. An Application

The orthogonalized residuals approach was applied to data from $n = 407$ parents and siblings of subjects with chronic obstructive pulmonary disease (COPD) and their controls (Cohen, 1980). The binary outcome of interest is impaired pulmonary function and the number of families is $K = 184$ with family size (n_i) ranging from 1 to 10. 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 status 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 3 shows within-cluster association parameter estimates obtained from GEE2 (Qaqish & Liang, 1992), alternating logistic regressions from SAS GENMOD (SAS Institute, Inc.) and orthogonalized residuals where λ is fixed at 0 or estimated using equation (11). Three other estimation methods compute cluster-specific λ_i under various distributional assumptions (Madsen, 1993; Morel & Neerchal, 1997; Skellam, 1948). For each of these three estimators, the mean and correlation parameters for each cluster are estimated using the averages of the cluster means μ_{ij} or off-diagonal elements of corr(Y_i), respectively. Details for estimation of λ under different models are described in Appendix 3.

Results for the parameter estimates for ALR from SAS GENMOD and ORTH_{ALR} are identical by definition; however, the standard errors for α_{pp} and α_{ss} are quite different due to the approximation used within GENMOD. In general, there is a fair amount of variation in a given parameter estimate across models, though the models ORTH_{MOMENT} and ORTH_{MAD} have estimates that are very similar to each other as do ORTH_{BB} and ORTH_{MN}. Further, there is no one model that has parameter estimates that are comparable to those in GEE2, though some models show similarity for subsets of the parameters.

An explanation for the large standard errors may be due to the limited information in the data to estimate association. There are 14 clusters with both parents present (for estimating α_{pp}), and 42 clusters with one parent present (a total of 56 clusters contributing to the estimation of α_{ps}). Further, there are 86 clusters with only a single sibling, and these do not contribute to the estimation of α_{ss}.

5. Conclusions

This paper addressed efficiency in estimation of association for binary responses. It was shown that in a simple model with varying cluster sample sizes, alternating logistic regressions can be fairly inefficient. This result contrasts sharply with that of Carey et al. (1993) who found alternating logistic regressions to be nearly as efficient as second-order generalized estimating equations, albeit in a special case of equal cluster sizes ($n_i = 4$). A new estimating equation based on orthogonalized residuals was developed and shown to have appreciable efficiency gain over alternating logistic regressions in the presence of unequal cluster sizes with minimal additional computational cost. This is an important finding, since it is unlikely that cluster sizes will be equal in most practical applications.

The approach based on orthogonalized residuals has other important features. First, its derivation follows a projection argument, and this provides insight into how the methodology maintains efficiency for association parameters. Second, a special case of orthogonalized residuals reformulates alternating logistic regressions into an estimating function using marginal residuals with standard weighting (i.e., dependence on data only through estimated parameters). This feature allows straightforward computation of the robust covariance estimate, alleviating the need for the approximation currently in use in popular software packages. Further, it opens the door for further developments of alternating logistic regressions including computation in mind.

For estimation of association parameters, efficiency will depend on higher moments, and no single procedure that is based only on the first two moments will be uniformly more efficient. As the true distribution will not be known, we currently suggest using orthogonalized residuals with λ computed according to the moment estimator (11) or a Morel-Neerchal working model. An attractive property of (11) is that it is a consistent estimator of λ under any true model while Morel-Neerchal was shown to be robust to misspecification of higher order moments in our efficiency calculations. However, it is important to check the sensitivity of results to other assumed models. A simulation study reported elsewhere (Zink, 2003) was undertaken to investigate the finite sample performance ($K = 100$) of ORTH with λ estimated as in (11), ALR ($U_{\alpha;ORTH0}$) and pseudo-likelihood methods for marginal mean and association models that contain observation and cluster-level covariates. Correlated binary data were generated for unequal cluster sizes (n_i) that ranged in a systematic way from 2 to 20 (or 2 to 50). Overall, the methods examined performed similarly with percent relative bias for within-cluster association parameter estimates being low (though tending to be negative), and coverage probabilities being close to the nominal level of 0.95 (though tending to undercoverage). Orthogonal residuals estimating λ as in (11) had greater finite sample efficiency (based on ratio of estimated mean squared errors) than ALR, but it also tended to estimate association parameters with slightly more bias. Further research is needed to understand this phenomenon. Replacing (9) with more complex structures and extending λ to a vector parameter may improve finite sample performance. 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 the indices are distinct. Finally, the ORTH procedures reported in this paper based on distributional assumptions allowing cluster-specific λ_i warrant further study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank the editor, two referees for their constructive comments and suggestions. The work of Bahjat Qaqish was supported in part by grant 1-R01-CA101901-01A1 from the U. S. National Cancer Institute. Richard Zink was supported by a pre-doctoral training grant from the National Institute of Environmental Health Sciences while a student at UNC-Chapel Hill.

References

By, K.; Qaqish, BF.; Preisser, JS. R package version 1.5. 2008. orth: Multivariate logistic regression using orthogonalized residuals.

By, K.; Qaqish, BF.; Preisser, JS.; Perin, J.; Zink, RC. Department of Biostatistics Technical Report Series. Working Paper 22. The University of North Carolina; Chapel Hill: 2011. ORTH: R and SAS software for correlated binary data based on orthogonalized residuals and alternating logistic regressions. Available at: http://biostats.bepress.com/uncbiostat/papers/art22

Carey V, Zeger SL, Diggle P. Modeling multivariate binary data with alternating logistic regressions. Biometrika. 1993; 80:517–26.

Chaganty NR, Joe H. Range of correlation matrices for dependent Bernoulli random variables. Biometrika. 2006; 93:197–206.

Cohen BH. Chronic pulmonary disease: a challenge in genetic epidemiology. American Journal of Epidemiology. 1980; 112:274–88. [PubMed: 6968157]

Geys H, Molenberghs G, Lipsitz S. A note on the comparison of pseudo-likelihood and generalized estimating equations for marginally specified odds ratio models with exchangeable association structure. J Statist Comp Sim. 1998; 62:45–71.

Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference. Statist Sci. 2000; 15:1–19.

Heyde, CC. Quasi-likelihood and its application: a general approach to optimal parameter estimation. Springer Verlag; New York: 1997.

Kuk AYC. Permutation invariance of alternating logistic regression for multivariate binary data. Biometrika. 2004; 91:758–61.

Kuk AYC. A hybrid pairwise likelihood method. Biometrika. 2007; 94:939–52.

Kuk AYC, Nott DJ. A pairwise likelihood approach to analyzing correlated binary data. Statist Probab Lett. 2000; 47:329–35.

le Cessie S, van Houwelingen JC. Logistic regression for correlated binary data. Applied Statistics. 1994; 43:95–108.

Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

Liang KY, Zeger SL, Qaqish B. Multivariate regression analyses for categorical data. J R Stat Soc Ser B Stat Methodol. 1992; 54:3–40.

Lipsitz SR, Laird NM, Harrington DP. Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. Biometrika. 1991; 78:153–60.

Lipsitz SR, Fitzmaurice GM. Estimating equations for measures of association between repeated binary responses. Biometrics. 1996; 52:903–12. [PubMed: 8805760]

Madsen RW. Generalized binomial distributions. Comm Statist Theory Methods. 1993; 22:3065–86.

Morel JG, Neerchal NK. Clustered binary logistic regression in teratology data using a finite mixture distribution. Stat Med. 1997; 16:2843–53. [PubMed: 9483718]

Morton R. Efficiency of estimating equations and the use of pivots. Biometrika. 1981; 68:227–33.

Neuhaus JM, Kalbfleish JD, Hauk WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. International Statistical Review. 1991; 59:25–35.

Prentice RL. Correlated binary regression with covariates specific to each binary observation. Biometrics. 1988; 44:1033–48. [PubMed: 3233244]

Qaqish BF, Liang KY. Marginal models for correlated binary responses with multiple classes and multiple levels of nesting. Biometrics. 1992; 48:939–50. [PubMed: 1420848]

Skellam JG. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. J R Stat Soc Ser B Stat Methodol. 1948; 10:257–61.

Ware JH, Dockery DW, Spiro A, Speizer FE, Ferris BG. Passive smoking, gas cooking, and respiratory health of children living in six cities. American Review of Respiratory Disease. 1984; 129:366–74. [PubMed: 6703495]

Zeger SL, Liang KY, Albert P. Models for longitudinal data: A generalized estimating equations approach. Biometrics. 1988; 44:1049–60. [PubMed: 3233245]

Zhao LP, Prentice RL. Correlated binary regression using a quadratic exponential model. Biometrika. 1990; 77:642–48.

Zink, RC. PhD doctoral dissertation. Chapel Hill, NC: The University of North Carolina at Chapel Hill; 2003. Correlated Binary Regression Using Orthogonalized Residuals. Available at: http://www.sph.unc.edu/bios/theses_and_dissertations_836_6180.html

Appendix

1. Proof that the orthogonalized residuals estimating equation with $\lambda = 0$ is equivalent to alternating logistic regressions. Since U_β is the same for both approaches, it suffices to show that $U_{\alpha;ALR} = U_{\alpha;ORTH0}$ where

$$U_{\alpha,ALR}=\sum_{i=1}^{K}\sum_{j<k}\frac{\partial\zeta_{ijk}}{\partial\alpha}^{\top}\frac{Y_{ij}-\zeta_{ijk}}{\zeta_{ijk}(1-\zeta_{ijk})} \quad \text{and} \quad U_{\alpha,ORTH0}=\sum_{i=1}^{K}\sum_{j<k}\frac{\partial\mu_{ijk}}{\partial\alpha}^{\top}\frac{Q_{ijk}}{v_{ijk}}.$$

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

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

Straightforward, but tedious, algebra shows that the above equality is true for each of the four possible patterns of (Y_{ij}, Y_{ik}), that is, for $(Y_{ij}, Y_{ik}) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

2. Details of asymptotic efficiency calculations. For all the estimating functions under study, the two components of the estimating function for (μ, ρ) are sums of cluster contributions. The i-th cluster contributes the 2×1 vector $U_i=(c_0+c_1y_{i.}+c_2y_{i.}^2, b_0+b_1y_{i.}+b_2y_{i.}^2)^\top$, where the coefficients are known functions of (n_i, μ, ρ). The form of the coefficients depends on the specific procedure.

The estimating function is the 2-vector $\sum_{i=1}^{K} U_i$. Define $D_i = E[\partial U_i/\partial\theta]$ and $V_i = \text{cov}(U_i)$. The efficiency matrix (Morton, 1981) is $(\Sigma_i D_i)^\top(\Sigma_i V_i)^{-1}(\Sigma_i D_i)$ and the asymptotic covariance matrix of $(\hat{\mu}, \hat{\rho})^\top$ is its inverse. Clearly, D_i involves the first two moments of $Y_{i:}$, which are completely determined by θ. However, V_i involves the third and fourth moments of $Y_{i:}$, which involve additional parameters. Those third and fourth moments were computed depending on the specific model for $Y_{i:}$. Numerically, since U_i, being a function of $Y_{i:}$, is a discrete random vector, computing V_i can be done via a simple sum over the range of $Y_{i:}$, i.e. 0 to n_i. The form of $P(Y_{i:} = y_{i:})$ depends on the specific distribution of $Y_{i:}$.

We note that GEE2 is the optimal quadratic estimating function, and hence is given by $\sum_{i=1}^{K} D_i^\top V_i^{-1} U_i$ and its asymptotic variance is $\left(\sum_i D_i^\top V_i^{-1} D_i\right)^{-1}$. Efficiency of the various procedures was computed relative to GEE2, with the form of the GEE2 estimating equations in (1) determined by the true model.

The transformation from the asymptotic covariance matrix of $(\hat{\mu}, \hat{\rho})^\top$ to the asymptotic covariance matrix of $(\hat{\mu}, \log\hat{\psi})^\top$ is done via the delta method. This is straightforward since the odds ratio, ψ, is easily expressed as an explicit function of μ and ρ. Code for calculating and plotting efficiencies is available at http://www.bios.unc.edu/~qaqish/software.htm).

3. Calculation of λ_i. The asymptotic efficiency calculations in Section 3 rely on the fact that λ_i is, by definition, the average correlation among the m_i residuals $\{Q_{ijk}, j < k\}$. Under model (12), all Q_{ijk} have the same variance, i.e. the ν_{ijk}'s are all equal, say $\nu_{ijk} = \nu$, where ν is a function of μ and ρ only (the general expression for ν_{ijk} is given in Section 2). Let $\tau_i^2=\text{var}\left(\sum_{j<k} Q_{ijk}\right)$. Hence, $\tau_i^2==m_i\nu\{1+(m_i-1)\lambda_i\}$. Because $\Sigma_{j<k} Q_{ijk}$ is a discrete random variable, being a (quadratic) function of $Y_{i:}$, τ_i^2 is computed via a simple summation over the range of $Y_{i:}$. Then λ_i is computed as $\lambda_i=\{\tau_i^2/(m_i\nu)-1\}/(m_i-1)$..

The mathematical expressions for λ_i are quite lengthy, except in the beta-binomial case, for which

$$\lambda_i=\frac{2\rho(2+\rho+n_i\rho)}{(n_i+1)(1+\rho)(1+2\rho)}$$

In the beta-binomial model, $\lambda_i = 0$ if $\rho = 0$; λ_i approaches $(n_i + 3)/\{3(n_i + 1)\}$ if n_i is fixed and $\rho \to 1$; approaches 1/3 as $\rho \to 1$ and n_i gets large. Modeling software that includes computations for λ_i under BB, MN, and MAD models (e.g., Table 3) as well as for λ as in equation (11) is available at http://www.bios.unc.edu/~qaqish/software.htm).

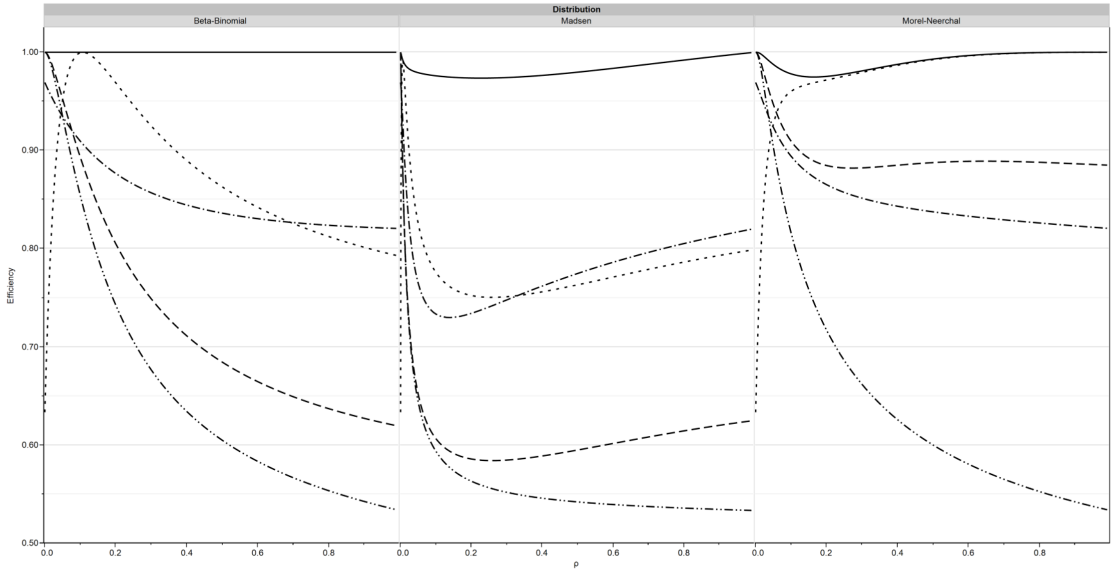

Figure 1.

Asymptotic efficiency of estimating equations for pairwise log odds ratio implied by model (12) relative to second-order generalized estimating equations under three higher-order model assumptions. $\mu = 0.2$, 1:1 mix of $n_i = 5$ and $n_i = 25$. $U_{\rho;ORTH}$ solid, $U_{\rho;ALR} = U_{\rho;KN}$ dashed, $U_{\rho;CH}$ dotted, $U_{\rho;L}$ dotdash, $U_{\rho;P}$ dotdotdash.

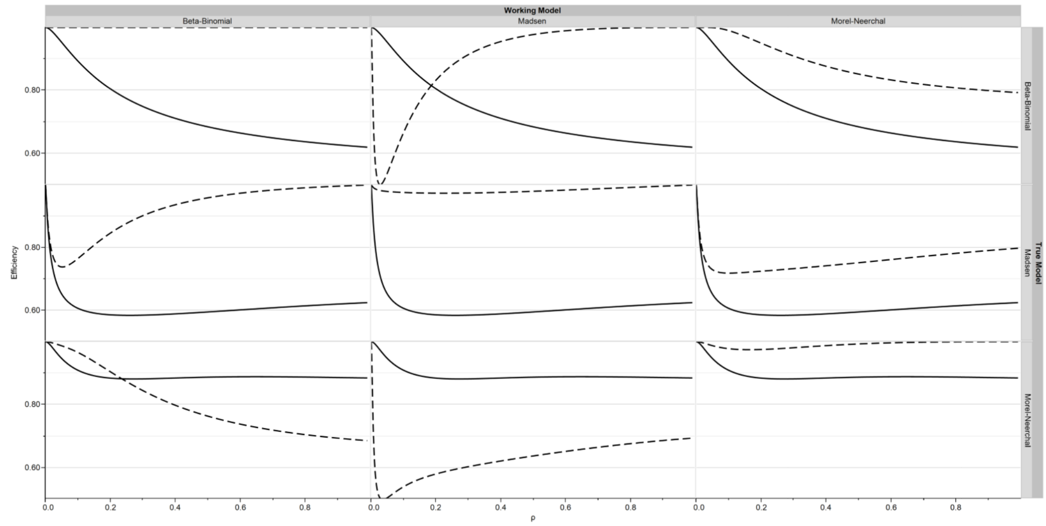

Figure 2.
Asymptotic efficiency of orthogonalized residuals and alternating logistic regressions relative to second-order generalized estimating equations for estimation of the pairwise log odds ratio under different true and working models. Rows are indexed by the true model, and columns by the working model. $U_{\rho;ORTH}$ solid, $U_{\rho;ALR}$ dashed.

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

$watermark-text

$watermark-text

Table 2

Pairwise kernels based on (y_{ij}, y_{ik}) for estimating ρ in model (12)

Estimating function	$y_{ij} + y_{ik} = 0$	$y_{ij} + y_{ik} = 1$	$y_{ij} + y_{ik} = 2$
$U_{\rho;ALR}$	$(\gamma^{-1} + \rho)^{-1}$	$-(1 - \rho)^{-1}$	$(\gamma + \rho)^{-1}$
$U_{\rho;P}$	$\gamma - \rho$	$-1 - \rho$	$\gamma^{-1} - \rho$
$U_{\rho;L}$	$- \mu^2(1 + \rho\gamma^{-1})$	$- \mu^2(1 + \rho\gamma^{-1})$	$1 - \mu^2(1 + \rho\gamma^{-1})$

Note that $U_{\rho;KN} = U_{\rho;ALR}$, $U_{\rho,CH} = \sum_{i=1}^{K} (n_i - 1)^{-1} U_{i,\rho,ALR}$, $U_{\rho;P}$ is from (4), $U_{\rho;L}$ is from (5), $U_{\rho,ORTH} = \sum_{i=1}^{K} (1 + (m_i - 1)\lambda)^{-1} U_{i,\rho,ALR}$ and $\gamma = \mu/(1 - \mu)$.

Table 3

Log Pairwise Odds Ratio Association Parameter Estimates and Empirical Standard Errors

	Parent-Parent(α_{pp})	**Sibling-Sibling(α_{ss})**	**Parent-Sibling(α_{ps})**
Model			
GEE2	−1.090 (1.090)	0.873 (0.565)	0.984 (0.519)
SAS GENMOD ALR	−1.177 (1.385)	1.108 (0.603)	0.986 (0.717)
ORTH$_{ALR}$	−1.177 (1.138)	1.108 (0.764)	0.986 (0.673)
ORTH$_{MOMENT}$	−1.264 (1.185)	0.795 (0.502)	0.782 (0.661)
ORTH$_{BB}$	−1.184 (1.131)	0.887 (0.570)	0.842 (0.627)
ORTH$_{MN}$	−1.178 (1.129)	0.911 (0.578)	0.868 (0.640)
ORTH$_{MAD}$	−1.275 (1.134)	0.788 (0.493)	0.761 (0.657)

Values are estimate (standard error). ORTH$_{ALR}$ assumes $\lambda = 0$. ORTH$_{MOMENT}$ estimates λ using (11), which for these data equals 0.2060. ORTH methods BB, MN and MAD estimates cluster-specific λ_i assuming the clusters were generated from beta-binomial, Morel-Neerchal and Madsen distributions, respectively. GEE2 results from Qaqish & Liang, 1992.