Biometrika (1993), **80**, 3, *pp.* 517–26
Printed in Great Britain

Modelling multivariate binary data with alternating logistic regressions

BY VINCENT CAREY, SCOTT L. ZEGER
Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health, 615 N. Wolfe Street, Baltimore, Maryland 21205, U.S.A.

AND PETER DIGGLE
School of Engineering, Computing, and Mathematical Sciences, Lancaster University, Lancaster, U.K.

SUMMARY

Marginal models for multivariate binary data permit separate modelling of the relationship of the response with explanatory variables, and the association between pairs of responses. When the former is the scientific focus, a first-order generalized estimating equation method (Liang & Zeger, 1986) is easy to implement and gives efficient estimates of regression coefficients, although estimates of the association among the binary outcomes can be inefficient. When the association model is a focus, simultaneous modelling of the responses and all pairwise products (Prentice, 1988) using second-order estimating equations gives more efficient estimates of association parameters as well. However, this procedure can become computationally infeasible as the cluster size gets large. This paper proposes an alternative approach, alternating logistic regressions, for simultaneously regressing the response on explanatory variables as well as modelling the association among responses in terms of pairwise odds ratios. This algorithm iterates between a logistic regression using first-order generalized estimating equations to estimate regression coefficients and a logistic regression of each response on others from the same cluster using an appropriate offset to update the odds ratio parameters. For clusters of size n, alternating logistic regression involves evaluation and inversion of matrices of order n^2 rather than n^4 as required for second-order generalized estimating equations. The alternating logistic regression estimates are shown to be reasonably efficient relative to solutions of second-order equations in a few problems. The new method is illustrated with an analysis of neuropsychological tests on patients with epileptic seizures.

Some key words: Clustered data; Generalized estimating equation; Logistic regression.

1. INTRODUCTION

Multivariate binary responses are common in the biological and social sciences. For example, we might observe the presence/absence of a disease in each of two eyes for an individual, for all members of a household, or repeatedly over time. The objectives of statistical analysis of such data include (i) describing the dependence of each binary response on explanatory variables, and (ii) characterizing the degree of association between pairs of outcomes as well as the dependence of this association on covariates.

Dale (1986), McCullagh & Nelder (1989), Prentice (1988) and Liang, Zeger & Qaqish

(1992) have discussed the class of 'marginal models' for regression analysis of multivariate binary data. The central idea is to model separately the marginal expectation of each binary variable as well as the association between pairs of outcomes in terms of explanatory variables. In this paper, we will use odds ratios to measure association. Consider binary data obtained in m clusters. For cluster $i = 1, \ldots, m$, let $Y_i = (Y_{i1}, \ldots, Y_{in_i})^{\mathrm{T}}$ be an $n_i \times 1$ response vector with mean $E(Y_i) = \mu_i$ and let ψ_{ijk} be the odds ratio between responses Y_{ij} and Y_{ik} $(1 \leqslant j < k \leqslant n_i)$ defined by

$$\psi_{ijk} = \frac{\mathrm{pr}\,(Y_{ij} = 1, Y_{ik} = 1)\,\mathrm{pr}\,(Y_{ij} = 0, Y_{ik} = 0)}{\mathrm{pr}\,(Y_{ij} = 1, Y_{ik} = 0)\,\mathrm{pr}\,(Y_{ij} = 0, Y_{ik} = 1)}. \tag{1}$$

A marginal model can be specified as follows:

1. $h(\mu_{ij}) = x_{ij}^{\mathrm{T}}\beta_j$, where $h(\,.\,)$ is a known link function (McCullagh & Nelder, 1989, p. 27), x_{ij} is a $p \times 1$ vector of explanatory variables associated with Y_{ij} and β_j are regression coefficients to be estimated;
2. $\log \psi_{ijk} = z_{ijk}^{\mathrm{T}}\alpha$, where z_{ijk} is a $q \times 1$ vector of covariates which specifies the form of the association between Y_{ij} and Y_{ik}, and α is a $q \times 1$ vector of association parameters to be estimated.

Throughout this paper we treat explanatory variables as constants and suppress them in lists of conditioning variables, using for example $E(Y_{ij})$ rather than $E(Y_{ij} \mid x_{ij})$. The term 'marginal' refers to modelling $E(Y_{ij})$ as opposed to $E(Y_{ij} \mid Y_{ik}, j \neq k)$ or to $E(Y_{ij} \mid Y_{ik}, k < j)$ which is commonly modelled in the time series context.

To further clarify the ideas underlying the marginal model, consider a family study with binary outcome, Y_{ij}, indicating whether or not member j in family i has pulmonary disease. The study objectives might include (a) identifying risk factors for disease such as smoking, presence of a gas stove in the home, age and sex; and (b) assessing whether the disease tends to aggregate in families after accounting for common environmental factors. In this example, a marginal model could be used to regress the disease outcome on the vector x_{ij} of risk factors and to model the odds ratios among family members to test the familial aggregation hypothesis. Notice that, under a genetic model for familial aggregation, the odds ratios would be expected to differ for two spouses, a parent and sib, and two sibs. Hence the vector z_{ijk} would, among other variables, include an indicator for the relationship of family members j and k.

Liang & Zeger (1986) introduced the use of 'generalized estimating equations', multivariate analogues of quasi-likelihood estimating equations, for estimating β in situations when α is a nuisance parameter. They originally modelled correlations; Lipsitz, Laird & Harrington (1991) used odds ratios to measure association. Fitzmaurice & Laird (1993) have used conditional odds ratios given all other responses in the vector to measure association. Liang et al. (1992) have shown that solutions to these first-order equations often have high efficiency for estimating β.

In problems such as the pulmonary disease example above, the association among responses is one of the primary scientific focuses. Hence the odds ratio parameters, α, are no longer nuisance parameters. Liang et al. (1992) have shown that estimates of α obtained by solving first-order estimating equations can be seriously inefficient. Prentice (1988) and Zhao & Prentice (1990) proposed quadratic estimating equations for correlation parameters. Asymptotic results from Liang et al. (1992) show that these estimates are reasonably efficient in simple cases considered.

Estimating pairwise odds ratio parameters using quadratic estimating equations requires the calculation and inversion of a weighting matrix with order n^4 elements as detailed in § 2.

Evaluating the elements of these matrices requires solution of cubic and seventh-degree polynomials. For small n_i, for example $n_i < 5$, this presents no obstacle, but for even moderate values of n_i as routinely arise in longitudinal and survey sampling applications, solution of second-order equations becomes computationally impractical.

This paper proposes another approach to the estimation of β and α which can be reasonably efficient for both sets of parameters and which avoids the computational burdens of the second-order methods. In the simplest case, with $\log \psi_{ijk} = \alpha$, the approach is to alternate between two steps:

(i) for a given α, estimate β as a parameter in a marginal logistic regression using a first-order generalized estimating equation;

(ii) for a given β, estimate the odds ratio parameter α using a logistic regression of Y_{ij} on each Y_{ik} $(k > j)$ with offset that involves μ_{ij} and $\nu_{ijk} = E(Y_{ij} Y_{ik})$.

We therefore refer to this algorithm as alternating logistic regressions. Section 2 briefly reviews the generalized estimating equation approaches to regression with multivariate binary responses. Section 3 details the new methodology and presents asymptotic properties of alternating logistic regression estimates. We find that alternating logistic regression estimates of α have efficiency comparable to those of second-order estimates in a few different problems with clusters of size four. The method is illustrated with a brief example in § 4 which is followed by discussion.

2. Generalized estimating equations

Generalized estimating equations are multivariate analogues of quasi-likelihood estimating equations proposed by Wedderburn (1974). With scalar responses, there is usually a unique integral of the estimating function analogous to the likelihood function, hence the term 'quasi-likelihood'. In the multivariate case, it is more common that the integral is not uniquely defined. In addition, there are often additional nuisance parameters to be estimated. If we let $V_i(\mu_i; \alpha)$ be an $n_i \times n_i$ weighting matrix which approximates the covariance matrix of Y_i, the generalized estimating equation for β has the form

$$U_1(\beta) = \sum_{i=1}^{m} \left(\frac{\partial \mu_i}{\partial \beta}\right)^{\mathrm{T}} V_i(\mu_i; \alpha)^{-1}(Y_i - \mu_i) = 0. \qquad (2)$$

Liang & Zeger (1986) show that the solution $\hat{\beta}_1$ of (2) with α replaced by a $m^{\frac{1}{2}}$-consistent estimate $\hat{\alpha}$ is asymptotically Gaussian with mean 0 and variance given by

$$\operatorname{cov}(\hat{\beta}_1) = I_0^{-1} I_1 I_0^{-1},$$

where

$$I_0 = m^{-1} \sum \frac{\partial \mu_i^{\mathrm{T}}}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta}, \quad I_1 = m^{-1} \sum \frac{\partial \mu_i^{\mathrm{T}}}{\partial \beta} V_i^{-1} \operatorname{cov}(Y_i) V_i^{-1} \frac{\partial \mu_i}{\partial \beta}.$$

Note that if $V_i = \operatorname{cov}(Y_i)$, then $\operatorname{cov}(\hat{\beta}_1) = I_0^{-1}$. Liang & Zeger used moment estimates of α in (2). An advantage of this approach is that in many problems, inferences about β are robust to misspecification of V_i for large m. Furthermore, $\hat{\beta}_1$ is reasonably efficient (Liang et al., 1992) when $\operatorname{cov}(Y_i)$ is well approximated. Fitzmaurice & Laird (1993) have shown that (2) is the score equation for β corresponding to a log linear model for Y_i when $\operatorname{cov}(Y_i)$ is correctly specified.

When α is the scientific focus, Prentice (1988) proposed expanding the estimating equations to model simultaneously the response Y_i and ${}_{n_i}C_2$ cross-products

$$W_i = (Y_{i1}Y_{i2}, Y_{i1}Y_{i3}, \ldots, Y_{i1}Y_{in_i}, \ldots, Y_{i,n_i-1}Y_{in_i})^{\mathrm{T}}.$$

The resulting estimating equations have the form

$$U_2(\beta, \alpha) = \sum_{i=1}^{m} \left\{ \frac{\partial(\mu_i, \nu_i)^{\mathrm{T}}}{\partial(\beta, \alpha)^{\mathrm{T}}} \right\}^{\mathrm{T}} \operatorname{cov}^{-1}(Y_i, W_i)(y_i - \mu_i, w_i - \nu_i)^{\mathrm{T}} = 0, \tag{3}$$

where $\nu_i = E(W_i)$. Liang et al. (1992) show that $(\hat{\beta}_2, \hat{\alpha}_2)$, the solution of $U_2(\beta, \alpha) = 0$, is highly efficient for both β and α in a variety of problems. The problem facing us with (3) is that, for a cluster of size n, the matrix $B = \operatorname{cov}(Y, W)$ has dimensions $(n + {}_nC_2) \times (n + {}_nC_2)$. For $n = 4$, B is 10×10; for $n = 10$, B is 55×55; for $n = 50$, still a small size for many applications, $B = 1275 \times 1275$, making solution of (3) computationally difficult. Furthermore, when pairwise odds ratios are used to measure association, evaluation of $\operatorname{cov}(Y, W)$ requires solving $O(n^3)$ cubic and $O(n^2)$ seventh-degree polynomial equations for each n-cluster at each iteration of a Newton–Raphson algorithm. In short, solutions of second-order estimating equations quickly become impractical for moderate or large n.

3. Alternating logistic regressions

3·1. *General*

The alternating logistic regressions procedure combines the first-order generalized estimating equations for β with new logistic regression equations for estimating α. We retain the first-order approach for β because it gives robust and reasonably efficient estimates when the assumed form of $\operatorname{cov}(Y_i)$ is close to the true covariance matrix. The new equations for α are designed to avoid the computational burden of second-order equations that results from evaluating and inverting the $({}_{n_i}C_2 + n_i) \times ({}_{n_i}C_2 + n_i)$ matrix, $\operatorname{cov}(Y_i, W_i)$. Our strategy is to estimate α using the ${}_{n_i}C_2$ conditional events, Y_{ij} given $Y_{ik} = y_{ik}$. In the simple case with $\log \psi_{ijk} \equiv \alpha$, we estimate α by regressing Y_{ij} on Y_{ik}, for $1 \leqslant j < k \leqslant n_i$ with an appropriate offset. Our prior hypothesis is that weighting the conditional elements as if independent of one another and of Y_{ij} will yield reasonably efficient estimates of α in many problems.

The alternating logistic regression strategy follows from suggestions by Firth (1992) and Diggle (1992) in the discussion of Liang et al. (1992). Before detailing the algorithm, we motivate the approach to estimating α. Let γ_{ijk} be the log odds ratio between outcomes Y_{ij} and Y_{ik}, let $\mu_{ij} = \operatorname{pr}(Y_{ij} = 1)$ and $\nu_{ijk} = \operatorname{pr}(Y_{ij} = 1, Y_{ik} = 1)$. Then, following Diggle (1992),

$$\operatorname{logit} \operatorname{pr}(Y_{ij} = 1 \mid Y_{ik} = y_{ik}) = \gamma_{ijk} y_{ik} + \log\left(\frac{\mu_{ij} - \nu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \nu_{ijk}} \right). \tag{4}$$

Suppose we assume that $\gamma_{ijk} = \alpha$. Then the pairwise log odds ratio α is the regression coefficient in a logistic regression of Y_{ij} on Y_{ik} as long as the second term on the right-hand side in (4) is used as an 'offset'. Note that the offset depends on the current value of $\delta = (\beta^{\mathrm{T}}, \alpha^{\mathrm{T}})^{\mathrm{T}}$ so that iteration is required.

More generally, we assume $\gamma_{ijk} = z_{ijk}^{\mathrm{T}}\alpha$ where the vector z_{ijk} is a known set of the pair-specific covariates. For example, in the pulmonary disease problem above, z_{ijk} would encode the type of family relationship for Y_{ij} and Y_{ik}: husband-wife, parent-sib, or sib-sib. Here, α can be estimated by regressing y_{ij} on $z_{ijk} y_{ik}$ with the same offset as above.

To be specific, the alternating logistic regression procedure iterates between the following two steps until convergence.

Step 1. Given the current values of $\hat{\delta}^{(r)}$, calculate $\hat{V}^{(r)}$ and solve the estimating equation (2) for an updated $\hat{\beta}^{(r+1)}$.

Step 2. Given $\hat{\beta}^{(r+1)}$ and $\hat{\alpha}^{(r)}$, evaluate the offset in equation (4) and perform the offset logistic regression of y_{ij} on $z_{ijk} y_{ik}$ with a total of $\sum {}_{n_i}C_2$ observations, where the summation is over the range $i = 1, \ldots, m$, to obtain $\hat{\alpha}^{(r+1)}$.

The details are provided in § 3·2.

3·2. *Alternating logistic regression estimating equations*

We describe the procedure for a collection of m clusters of differing sizes n_i. Let $\nu_i = E(W_i)$ denote an ${}_{n_i}C_2$-vector whose elements are $\nu_{ijk} = E(Y_{ij} Y_{ik})$ for $1 \leqslant j < k \leqslant n_i$. The elements of ν_i are ordered so that the rightmost indices of ν_{ijk} vary fastest. Let ζ_i be the ${}_{n_i}C_2$-vector with elements

$$\zeta_{ijk} = E(Y_{ij} \mid Y_{ik} = y_{ik}) = \operatorname{logit}^{-1}\left\{\gamma_{ijk} y_{ik} + \log\left(\frac{\mu_{ij} - \nu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \nu_{ijk}}\right)\right\}, \qquad (5)$$

and let R_i be the vector of residuals with elements

$$R_{ijk} = Y_{ij} - E(Y_{ij} \mid Y_{ik} = y_{ik}) = Y_{ij} - \zeta_{ijk}.$$

We let S_i denote the ${}_{n_i}C_2 \times {}_{n_i}C_2$ diagonal matrix with diagonal element $\zeta_{ijk}(1 - \zeta_{ijk})$, and let T_i denote the ${}_{n_i}C_2 \times q$ matrix $\partial \zeta_i / \partial \alpha$. Finally, we let

$$A_i = Y_i - \mu_i, \quad B_i = \operatorname{cov}(Y_i), \quad C_i = \frac{\partial \mu_i}{\partial \beta}.$$

Note that, when the elements of Y_i are permuted, so are the elements of W_i, given our pairwise indexing scheme. In addition, some elements of the vectors ζ_i and R_i will change values when elements of Y_i are permuted. Despite this, the proposed estimation procedure is invariant to permutation of the elements of Y_i as shown by V. Carey in an unpublished Ph.D. thesis of Johns Hopkins University.

The alternating logistic regression estimate of δ is the simultaneous solution of the following unbiased estimating equations:

$$U_\beta = \sum_{i=1}^{m} C_i^{\mathrm{T}} B_i^{-1} A_i = 0, \qquad (6)$$

$$U_\alpha = \sum_{i=1}^{m} T_i^{\mathrm{T}} S_i^{-1} R_i = 0. \qquad (7)$$

We solve the estimating equations (6) and (7) for β and α using the nonlinear Gauss–Seidel algorithm (Thisted, 1986, p.181). The updating sequence and the formulae linking product-moments and pairwise odds ratios are as in formulae (8), (9) and (6) of Lipsitz et al. (1991). As a Gauss–Seidel procedure with positive-definite expected derivative matrix, alternating logistic regression converges given starting values sufficiently close to the solution. In practice, this algorithm converges very quickly when ordinary logistic regression estimates are used as starting values for β, and 0 is used for α.

Alternating logistic regression is computationally feasible for very large clusters; equation (3) is not. For each cluster of size n, the calculation of the diagonal ‘weighting matrix' S_i in alternating logistic regression involves $O(n^2)$ closed form computations of the expression (5), whereas the calculation of the weighting matrix cov (Y_i, W_i) of equation (3) with associations measured in terms of pairwise odds ratios requires solution of $O(n^3)$ cubic and $O(n^2)$ seventh-degree polynomials. The weighting matrix in (3) has dimension of order $O(n^2)$ so that $O(n^6)$ computations are required for its inversion. In alternating logistic regression, only the weighting matrix, cov (Y_i), which has dimension n must be inverted using $O(n^3)$ operations. To illustrate the computational savings of alternating logistic regression, we have fitted a simple logistic regression with $z_{ijk}^{\mathrm{T}}\alpha = \alpha_0$ to a simulated data set comprising of 100 clusters of sizes $n_i \equiv 2$, 4 and 8. Each procedure was run 10 times on each data set; computations were performed on a SunSparcStation 2 in single-user mode. The average 'user' times over the 10 runs are in Table 1. For $n_i = 2$, alternating logistic regression provides little saving. For $n_i = 4$, alternating logistic regression is 5 times faster than solving equation (3); for $n_i = 8$, alternating logistic regression is 40 times faster.

Table 1. *Average time (seconds) to convergence of alternating logistic regression* (ALR) *and a second-order estimating equation procedure* (GEE2)

Model	Cluster size 2	4	8
ALR	1·6	3·2	8·3
GEE2	1·7	16·3	347·5

3·3. *Asymptotic properties*

The estimates $\hat{\delta}_{\mathrm{ALR}}$ obtained from the alternating logistic regression algorithm are consistent and asymptotically follow a Gaussian distribution. The following notation simplifies the statement of the asymptotic results:

$$U_.(\delta) = \Sigma_i U_i(\delta) = \begin{pmatrix} \Sigma_i C_i^{\mathrm{T}} B_i^{-1} A_i \\ \Sigma_i T_i^{\mathrm{T}} S_i^{-1} R_i \end{pmatrix}, \quad U_.^*(\delta) = \begin{pmatrix} \Sigma_i C_i^{\mathrm{T}} B_i^{-1} C_i & 0 \\ \Sigma_i T_i^{\mathrm{T}} S_i^{-1} D_i & \Sigma_i T_i^{\mathrm{T}} S_i^{-1} T_i \end{pmatrix},$$

with $D_i = \partial\zeta_i/\partial\beta$. Here $U_.^*(\delta)$ is an approximation to $\partial U_./\partial\delta$, ignoring terms of size $o_p(1)$.

Following results of Fahrmeir & Kaufmann (1985) and Liang & Zeger (1986), $m^{\frac{1}{2}}(\hat{\delta}_{\mathrm{ALR}} - \delta)$ is asymptotically $(p+q)$-variate Gaussian as $m \to \infty$, with covariance matrix given by

$$V_{\mathrm{ALR}} = \lim_{m\to\infty} m[E\{U_.^*(\delta)\}]^{-1} E\{U_.(\delta)U_.(\delta)^{\mathrm{T}}\}[E\{U_.^*(\delta)\}^{\mathrm{T}}]^{-1}. \tag{8}$$

A consistent estimate of the variance of $\hat{\delta}_{\mathrm{ALR}}$ is

$$\hat{V}_{\mathrm{ALR}} = \{U_.^*(\hat{\delta})\}^{-1}\left\{\sum U_i(\hat{\delta})U_i(\hat{\delta})^{\mathrm{T}}\right\}\{U_.^*(\hat{\delta})^{\mathrm{T}}\}^{-1}. \tag{9}$$

3·4. *Efficiency*

In this section, we compare the asymptotic efficiency of $\hat{\alpha}$ as estimated by alternating logistic regression and second-order generalized estimating equations following Liang et al. (1992) for a few simple designs. To evaluate asymptotic variances, we must compute the probabilities of each of the 2^n possible configurations of the n-vector, Y. We restrict attention to the case $n = 4$ where to specify the full distribution we require, in addition to δ, third and fourth moment parameter values. We adopt the convention used by Liang et al. (1992, § 3.3) of setting contrasts of log odds ratios equal to 0. With these additional assumptions, it is possible to evaluate the asymptotic variance of $\hat{\alpha}_{\text{ALR}}$ given in (8).

To examine the asymptotic efficiency of alternating logistic regression relative to the solution of the second-order estimating equation, we consider two regression designs.

Two-sample design: $\text{logit}\, \mu_{ij} = \beta_0 + \beta_1 1_{\{j>2\}} \quad (j = 1, \ldots, 4)$,

where $1_{\{.\}}$ is 1 if $\{.\}$ is true, and is 0 otherwise.

Trend design: $\text{logit}\, \mu_{ij} = \beta_0 + \beta_1 x_j, \quad x_j = -2, -1, 1, 2 \quad (j = 1, \ldots, 4)$.

In each model, we set $\beta_0 = -1{\cdot}386$ and $\beta_1 = 0{\cdot}69$. The odds ratio regression was assumed to be

$$\log \psi_{ijk} = \alpha_0 + \alpha_1 1_{\{j>1\,\&\,k>3\}} \quad (1 \leqslant j < k \leqslant 4).$$

Table 2 presents the asymptotic efficiencies for estimation of (α_0, α_1) for these two models with $\exp(\alpha_0)$ and $\exp(\alpha_1) = 1$, 2 and 5. Table 2 indicates that, for both designs, the efficiency of $\hat{\alpha}_0$ from alternating logistic regression is 90% or better relative to the solution of (3) throughout the parameter range considered. The efficiency of $\hat{\alpha}_1$ is also high except when α_1 becomes large in the two-sample design.

Table 2. *Asymptotic efficiency of estimates of α from alternating logistic regression relative to second-order estimating equation*

	Two-sample problem			Trend		
		$\exp(\alpha_1)$			$\exp(\alpha_1)$	
$\exp(\alpha_0)$	1	2	5	1	2	5
1	1000, 1000	991, 960	958, 771	998, 999	990, 974	964, 905
2	995, 989	986, 913	959, 675	983, 982	962, 932	938, 881
5	985, 956	979, 847	958, 578	944, 940	927, 899	915, 893

Cell entries are asymptotic efficiencies of $(\hat{\alpha}_0, \hat{\alpha}_1)$ times 1000.

4. Example

We illustrate the alternating logistic regression procedure with an analysis of tests of cognitive function in patients with medically intractable epileptic seizures. Partial lobectomy, removal of a portion of the brain (Engel, 1987), is an effective treatment for such seizures. Prior to such a radical intervention, the consequences to neuropsychological function must be assessed. 'Wada testing' (Wada & Rasmussen, 1960) is a method of identifying brain regions whose absence will impair specific cognitive functions. In Wada testing, sodium amobarbital is injected into the carotid artery to deaden temporarily a hemisphere of the brain. Cognitive function tests are administered to assess the likely effects of removing individual regions.

We analyze test data collected on 52 individuals at Johns Hopkins Hospital (Hart et al., 1993). The patients received amobarbital in the right hemisphere. After drug effect was established, each patient was given a sequence of up to 22 language test items, falling into six broad classes. The classes are, in order of administration: object-naming (abbreviated to OBJ; six items), word-reading (WOR; two items), picture-naming (PIN; two items), semantic picture-word matching (PIW; four items), concrete and abstract word-reading (COA; four items), and categorization (CAT; four items). All subjects performed perfectly on the categorization items, and these tests are omitted from analysis, reducing the size of a complete cluster to 18. The total number of test responses under analysis is 930.

Table 3. *Marginal regression analyses of Wada test results for* GLIM (*independence model*) *and alternating logistic regression with three association models*

Variable	Model A Est.	Model A Z	Model B Est.	Model B Rob. Z	Model C Est.	Model C Rob. Z	Model D Est.	Model D Rob. Z
Intercept	3·31	6·12	3·13	4·28	3·19	4·36	2·69	4·11
1 (right handed)	0·82	3·11	0·80	2·24	0·86	2·48	0·85	2·66
1 (right seizure focus)	0·26	1·03	0·23	0·79	0·18	0·65	0·26	0·86
log (age of onset)	−0·14	−1·78	−0·12	−1·42	−0·13	−1·64	−0·11	−1·39
1 (lesion)	−0·16	−0·69	−0·15	−0·53	−0·08	−0·28	−0·09	−0·32
1 (log time to mid-test)	−2·41	−6·89	−2·24	−4·31	−2·28	−4·46	−1·89	−4·30
1 (WOR)	0·44	1·20	0·43	1·74	0·42	1·58	0·62	2·64
1 (PIW)	0·24	0·68	0·25	0·68	0·19	0·51	0·28	0·78
1 (PIN)	0·33	1·17	0·35	0·96	0·19	0·53	0·25	0·76
1 (COA)	1·14	3·36	1·14	2·12	1·07	2·19	0·96	2·34
Within-subject	–	–	0·26	1·18	–	–	–	–
Within-class	–	–	–	–	1·22	4·34	–	–
OBJ	–	–	–	–	–	–	1·38	3·23
WOR	–	–	–	–	–	–	2·12	2·20
PIN	–	–	–	–	–	–	1·05	1·07
PIW	–	–	–	–	–	–	0·65	1·50
COA	–	–	–	–	–	–	1·80	1·89
Between-class	–	–	–	–	−0·04	−0·13	–	–
OBJ * WOR	–	–	–	–	–	–	1·93	3·59
OBJ * PIN	–	–	–	–	–	–	0·57	1·02
OBJ * PIW	–	–	–	–	–	–	−0·06	−0·14
OBJ * COA	–	–	–	–	–	–	−1·03	−0·28
WOR * PIN	–	–	–	–	–	–	0·75	1·15
WOR * PIW	–	–	–	–	–	–	−1·01	−0·25
WOR * COA	–	–	–	–	–	–	−1·68	−0·43
PIN * PIW	–	–	–	–	–	–	−0·32	−0·21
PIN * COA	–	–	–	–	–	–	−0·32	−0·17
PIW * COA	–	–	–	–	–	–	0·38	0·96

In Table 3 we present four marginal models: model A, a model which ignores association of test responses within a subject, fitted using GLIM; model B, an alternating logistic regression model assuming a constant pairwise log odds ratio; model C, an alternating logistic regression model assuming different within-test-class and between-test-class pairwise log odds ratios; model D, a model in which the 15 separate within- and between-test-class association parameters are fitted using alternating logistic regression. In model A we use naive standard errors based on the independence assumption to demonstrate the potential for incorrect inferences. Note that Z-statistics from model A tend to be too large as can be seen by comparing model A with any of the others in Table 3.

The regression results for models B, C and D are qualitatively similar. For a right-handed individual with right-hemisphere seizure focus, no lesion, and mean values for age of seizure onset (5·6 years), and time to mid-test (2·2 minutes), the probability of correct response to an object-identification question is estimated to be 0·89 in each of the models. This high success probability is consistent with prevailing theory that language functions are concentrated in the left hemisphere, and thus are relatively unimpaired by a right-hemisphere amobarbital injection. The reduction in success probability for left-handed patients may indicate presence of right-hemisphere language dominance for these patients; as 85% of the sample was right-handed, this interpretation is in need of further supporting evidence. That success probability is reduced in the presence of longer times elapsed to mid-test is to be expected, as accumulated delays in responding to test items would be indicative of language impairment. Higher success probabilities for items on concrete/abstract word reading probably reflect the fact that the amobarbital effect has diminished by this point in the testing scheme.

Turning to the association model B, the common within-person odds ratio does not appear to be significant. In model C, however, the within-test-class odds ratio is found to be highly significant, while there is little evidence of common between-class association. In the expanded model D, we allow for class-specific within-class association. Significant positive pairwise association was observed on responses within the object-naming (OBJ) and word-reading (WOR) test classes. Positive association was also detected for pairs of responses in which one element of the pair is an OBJ response, and the other is a WOR response. No other significant between-class associations were observed. Detection of test-classes such as OBJ and WOR which are highly concordant may aid in the pursuit of a more parsimonious testing scheme.

5. Discussion

First-order and second-order estimating equations permit estimation of first- and second-moment parameters in regression models for multivariate binary data. When association among the observations is of scientific importance and is measured using marginal odds ratios, the second-order methods are more efficient, but the computations required will preclude their application in studies with large clusters. This paper has proposed an alternative approach which overcomes the computational limitations encountered in many problems. The new method, alternating logistic regressions, involves matrices whose dimension is $O(n)$ as opposed to $O(n^2)$ for second-order estimating equations. Alternating logistic regression is reasonably efficient relative to solving equation (3) in the cases considered. In alternating logistic regression, we estimate the association parameters by modelling the conditional distribution of one response given another. Here, use of a diagonal weighting matrix in the odds ratio regression of one response on another gives a reasonably optimal weighting. In alternating logistic regression, we do not include $w_i - \nu_i$ in the estimating equation for β so that the solution β remains consistent in many problems even when $E(W_i) \neq \nu_i$.

Zhao & Prentice (1990) and Prentice & Zhao (1991) have studied estimating equations for (β, α) where the pairwise association parameters α describe correlations rather than odds ratios. Here, the computations are simpler but there are more complicated constraints on the values of α. Further work is necessary to compare this approach with alternating logistic regression. Fitzmaurice & Laird (1993) parameterize the log linear model for multivariate binary data in terms of marginal means and conditional pairwise odds ratios given all other responses in the vector. Here, the first-order generalized estimating equation for β is the maximum likelihood score equation. One limitation of this approach is that it applies only when $n_i = n$ for all clusters.

We illustrated our procedure with an application to neuropsychological test data in which clusters were typically of size 18. To confirm informally the finite sample validity of alternating logistic regression inferences, we conducted a simulation study in which samples of 52 clusters (individuals) were taken with replacement from the original example data. For each of 1086 such samples, we computed the alternating logistic regression estimates of model B and determined whether the nominal 95% confidence interval covered the parameter values estimated on the original data set. The actual coverage probabilities for the nominal 95% intervals ranged from 0·84 to 0·98 over the 12 parameters and had median value 0·90. Hence there is only a small degree of undercoverage with $m = 52$.

Self contained software implementing alternating logistic regression in the C programming language, interfaced to S (Becker, Chambers & Wilks, 1988), is to be made available through the STATLIB public domain distribution service.

6. ACKNOWLEDGMENTS

The authors thank Karen Bandeen-Roche and John Hart, Jr., for helpful discussions, and John Hart, Jr., and Barry Gordon for permission to use the Wada test data. We also thank the referees and associate editor for suggestions which have improved upon an earlier draft. Scott L. Zeger gratefully acknowledges support from the US National Institutes of Health grant AI 25529 and from the Merck, Sharp and Dohme Research Laboratory.

REFERENCES

BECKER, R., CHAMBERS, J. & WILKS, A. (1988). *The New S Language*. Pacific Grove: Wadsworth and Brooks.

DALE, J. R. (1986). Global cross-ratio models for bivariate, discrete ordered responses. *Biometrics* **42**, 909–17.

DIGGLE, P. (1992). Discussion of paper by K.-Y. Liang, S. L. Zeger and B. Qaqish. *J. R. Statist. Soc.* B **45**, 28–9.

ENGEL, J., JR. (1987). Comparison between electrophysiological approaches for presurgical evaluation. In *Presurgical Evaluation of Epileptics: Basics, Techniques, Implications*, Ed. H. G. Wieser and C. E. Elger, pp. 205–8. New York: Springer-Verlag.

FAHRMEIR, L. & KAUFMANN, S. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13**, 342–68.

FIRTH, D. (1992). Discussion of paper by K.-Y. Liang, S. L. Zeger, and B. Qaqish. *J. R. Statist. Soc.* B **45**, 24–6.

FITZMAURICE, G. M. & LAIRD, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80**, 141–51.

HART, J., LESSER, R. P., FISHER, R. S., SCHWERDT, P., BRYAN, R. N. & GORDON, B. (1993). Dominant-side intracarotid amobarbital spares comprehension of word-meaning. *Arch. Neurol.* To appear.

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 (with discussion). *J. R. Statist. Soc.* B **45**, 3–40.

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.

MARDIA, K. V. (1967). Some contributions to the contingency-type bivariate distributions. *Biometrika* **54**, 235–49.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

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

PRENTICE, R. L. & ZHAO, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–40.

THISTED, R. (1986). *Elements of Statistical Computing*. London: Chapman-Hall.

WADA, S. & RASMUSSEN, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance. *J. Neurosurg.* **17**, 266–82.

WEDDERBURN, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**, 439–47.

ZHAO, L. P. & PRENTICE, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77**, 642–8.

[*Received October* 1991. *Revised March* 1993]