BIOMETRICS 56, 73–80
March 2000

Random Effects Modeling of Multiple Binomial Responses Using the Multivariate Binomial Logit-Normal Distribution

Brent A. Coull

Department of Biostatistics, Harvard School of Public Health,
655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.
email: BCOULL@HSPH.HARVARD.EDU

and

Alan Agresti

Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, U.S.A.

SUMMARY. The multivariate binomial logit-normal distribution is a mixture distribution for which, (i) conditional on a set of success probabilities and sample size indices, a vector of counts is independent binomial variates, and (ii) the vector of logits of the parameters has a multivariate normal distribution. We use this distribution to model multivariate binomial-type responses using a vector of random effects. The vector of logits of parameters has a mean that is a linear function of explanatory variables and has an unspecified or partly specified covariance matrix. The model generalizes and provides greater flexibility than the univariate model that uses a normal random effect to account for positive correlations in clustered data. The multivariate model is useful when different elements of the response vector refer to different characteristics, each of which may naturally have its own random effect. It is also useful for repeated binary measurement of a single response when there is a nonexchangeable association structure, such as one often expects with longitudinal data or when negative association exists for at least one pair of responses. We apply the model to an influenza study with repeated responses in which some pairs are negatively associated and to a developmental toxicity study with continuation-ratio logits applied to an ordinal response with clustered observations.

KEY WORDS: Continuation-ratio logit; Generalized estimating equations (GEE); Generalized linear mixed model; Marginal model; Mixture model; Ordinal data; Overdispersion.

1. Introduction

This paper presents models for vectors $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_R)$ of binomial-type responses. Interest focuses on modeling the R response distributions, such as comparing those distributions at fixed values of covariates. When $\mathbf{Y}$ results from repeated measurement of a single binary variable, a common way to account for the correlations is with a random effect term for each subject or cluster. By contrast, the model discussed in this paper refers to a set of R separate response variables with a distinct random effect for each variable.

The models discussed are special cases of the generalized linear mixed model (GLMM), which has a linear predictor consisting of fixed effects relating to observed covariates and random effects relating to unobserved variables. For many repeated measurement problems, a simple random effects structure consisting of a random intercept is sufficient to account for correlations among different observations on the same subject or among observations on different subjects in the same cluster (e.g., Breslow and Clayton [1993] and references therein). For subject (or cluster) s, $s = 1, \ldots, N$, and binary response Y_{sr} that can take values zero and one, $r = 1, \ldots, R$, let $\pi_{sr} = P(Y_{sr} = 1)$ and let $\mathbf{x}_{sr}$ be a fixed covariate row vector. Most common is the logit link, for which the random-intercept form of model is

$$\text{logit}\,(\pi_{sr}) = \alpha_s + \mathbf{x}_{sr}\boldsymbol{\beta}, \tag{1}$$

where $\{\alpha_s\}$ are i.i.d. random variables with distribution function F in some parametric family and where $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector. Typically, one assumes that $\{\alpha_s\}$ are $\text{N}(0, \sigma^2)$ with unknown σ^2 and that the R responses (given α_s) are mutually independent. This case for (1) is often called a logistic-normal model. One can obtain maximum likelihood (ML) estimates of $(\sigma, \boldsymbol{\beta})$ using various methods, including Newton–Raphson maximization following Gauss–Hermite quadrature to numerically integrate out the random effects.

While model (1) has enjoyed much success for certain types of data, it suffers from two inflexibilities. First, the model implies nonnegative marginal log odds ratios among the R responses, averaged over subjects. To illustrate such inadequacy, we use Table 1, taken from Haber (1986) and also

Table 1
Observed and fitted counts of infection profiles for influenza data

$(Y_1, \ldots, Y_4)$	Observed count	Fitted values		
		Model (1)	BLN model (6)	BLN model (7)
0 0 0 0	140	138.4	138.4	138.7
0 0 0 1	31	20.8	23.9	31.1
0 0 1 0	16	19.5	16.9	16.2
0 0 1 1	3	4.1	5.4	2.0
0 1 0 0	17	22.1	19.2	18.6
0 1 0 1	2	4.6	6.1	2.3
0 1 1 0	5	4.3	3.8	6.0
0 1 1 1	1	1.2	2.1	.4
1 0 0 0	20	26.2	28.2	22.6
1 0 0 1	2	5.5	1.4	2.8
1 0 1 0	9	5.1	6.3	7.2
1 0 1 1	0	1.5	.6	.5
1 1 0 0	12	5.8	7.1	8.1
1 1 0 1	1	1.7	.7	.6
1 1 1 0	4	1.6	2.4	5.6
1 1 1 1	0	.6	.5	.3
Deviance (d.f.)		27.7 (10)	18.2 (8)	6.3 (8)

analyzed by Darroch and McCloud (1990). In this 2^4 contingency table, counts refer to infection profiles of a sample of 263 individuals for four influenza outbreaks occurring in the winters 1977/1978 to 1980/1981 in Tecumseh, Michigan. Each subject has four binary responses. Model form (1) with $\mathbf{x}_{s1} = (1,0,0,0)$, $\mathbf{x}_{s2} = (0,1,0,0)$, $\mathbf{x}_{s3} = (0,0,1,0)$, $\mathbf{x}_{s4} = (0,0,0,1)$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$ is appealing, with differences among $\{\beta_r\}$ providing comparisons of the log odds of catching the flu at the various outbreaks. However, the outbreaks in the first 3 years were caused by different viruses, while the outbreak in the fourth year was apparently caused by the same virus as in the first year. One expects responses for the first and fourth outbreaks to be negatively correlated because contracting influenza during one outbreak provides a stronger immunity against a subsequent outbreak of that type. Section 3 discusses this example further.

Second, the model implies an exchangeability in the conditional odds ratios for the marginal distribution. For responses $\{y_{sr}, r = 1, \ldots, R\}$ for an arbitrary subject, let $t_{ab} = \Sigma_r\, y_{sr} - y_{sa} - y_{sb}$. For the marginal distribution (averaged over subjects) at covariate values $\{\mathbf{x}_r = \mathbf{x}_{sr}\}$, the odds ratio between responses a and b, given the other responses, is the same for all pairs (a, b) with a common value of t_{ab}, and this is true regardless of the form of the random effects distribution. Thus, unlike in generalized estimating equations (GEE) analyses for repeated measurement (Liang and Zeger, 1986), model (1) does not provide a way of specifying correlation patterns that treat the responses in an asymmetric manner.

This article presents a generalization of model (1), called the multivariate binomial logit-normal model. It provides flexibility in modeling the correlation structure among R response variables that are each binomial, conditional on the random effects, by using a separate random effect for each variable. The model form itself is not new, as other authors have introduced logit models with multivariate normal random effects (e.g., Stiratelli, Laird, and Ware, 1984; Breslow and Clayton, 1993). In traditional mixed logistic models with multivariate normal random effects, however, the random effects vector usually consists of a subject-specific intercept and slope. The multivariate binomial logit-normal model differs from this in that the random terms are not regression coefficients common to all observations within a cluster but rather are a separate random intercept unique to each observation. This formulation is useful when the separate binomial variables refer to different classifications or when the R repeated responses are not positively correlated. Section 2 presents the model and a method for fitting it. Section 3 presents two examples of its application; Section 4 explores some properties of the model; and Section 5 discusses alternative models and alternative fitting procedures.

2. The Multivariate Binomial Logit-Normal Model

For subject s, conditional on $\boldsymbol{\pi}_s = (\pi_{s1}, \ldots, \pi_{sR})$, we let $\mathbf{Y}_s = (Y_{s1}, \ldots, Y_{sR})$ denote R independent binomial random variables with index vector $\mathbf{n}_s = (n_{s1}, \ldots, n_{sR})$ and parameters $\boldsymbol{\pi}_s$. Let $\text{logit}(\boldsymbol{\pi}_s)$ denote the logit operation applied componentwise to $\boldsymbol{\pi}_s$. We now generalize model (1) by incorporating a separate random effect for each of the R binomial responses, such that $\text{logit}(\boldsymbol{\pi}_s)$ is a multivariate normal random variable. Specifically,

$$\text{logit}(\boldsymbol{\pi}_s) = \boldsymbol{\alpha}_s + \mathbf{X}_s\boldsymbol{\beta}, \tag{2}$$

$s = 1, \ldots, N$, where $\mathbf{X}_s$ is the $R \times p$ covariate matrix whose rth row is $\mathbf{x}_{sr}$ and $\boldsymbol{\alpha}_s \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma})$. The parameters $\boldsymbol{\beta}$ describe the effects of the explanatory variables, while $\boldsymbol{\Sigma}$ contains parameters that reflect the heterogeneity among subjects as well as within-subject dependencies among the R variables. Model (1) is a special case of this model that restricts the R random effects to be identical (i.e., they have equal variances and all $R(R-1)/2$ pairwise correlations equal 1.0).

We refer to the mixture model in which, (i) conditional on $\boldsymbol{\pi}_s$, $\mathbf{Y}_s$ is a vector of R independent binomial random variables with indices $\mathbf{n}_s$ and parameters $\boldsymbol{\pi}_s$ and (ii) the $\{\text{logit}(\boldsymbol{\pi}_s)\}$ are i.i.d. from a $\text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution as a multivariate binomial logit-normal distribution. We denote it by $\text{BLN}^R(\mathbf{n}_s, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, or BLN for short. For instance, as just described, $\mathbf{Y}_s \sim \text{BLN}^R(\mathbf{n}_s, \mathbf{X}_s\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Let $f(\boldsymbol{\pi}; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ denote the probability density function for $\{\boldsymbol{\pi}_s\}$. Then

$$
\begin{aligned}
&f(\boldsymbol{\pi}; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \\
&\quad = (2\pi)^{-R/2} |\boldsymbol{\Sigma}|^{-1/2} \left\{ \prod_{r=1}^{R} [\pi_r(1-\pi_r)] \right\}^{-1} \\
&\qquad \times \exp\left\{ -\frac{1}{2} (\text{logit}(\boldsymbol{\pi}) - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\text{logit}(\boldsymbol{\pi}) - \mathbf{X}\boldsymbol{\beta}) \right\}, \\
&\qquad\qquad 0 \le \pi_r \le 1,
\end{aligned}
$$

for $r = 1, \ldots, R$. The $\text{BLN}^R(\mathbf{n}, \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ mixture has probability mass function

$$
p(\mathbf{y}; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int_{[0,1]^R} \left[\prod_{r=1}^{R} b(y_r \mid \pi_r; n_r) \right] f(\boldsymbol{\pi}; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\pi}, \quad y_r = 0, 1, \ldots, n_r, \; r = 1, \ldots, R, \tag{3}
$$

where $b(y_r \mid \pi_r; n_r)$ denotes the binomial probability mass function with n_r trials and parameter π_r.

Maximum likelihood estimation of the parameters $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is more complex computationally for this model than the usual univariate generalized linear mixed model because of the multivariate nature of the integral in (3). In our examples, we approximated this integral using multidimensional Gauss–Hermite quadrature. Consider the transformation (Aitchison and Ho, 1989) $\text{logit}(\boldsymbol{\pi}) - \mathbf{X}\boldsymbol{\beta} = \mathbf{Q}\mathbf{z}$, where $\mathbf{Q}$ is the unique lower triangular $R \times R$ matrix with nonnegative diagonal elements such that $\boldsymbol{\Sigma} = \mathbf{Q}\mathbf{Q}'$. The ML estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{Q}}\hat{\mathbf{Q}}'$ are obtained by maximizing the approximation to the log likelihood,

$$
l(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{N} \log \left[\sum_{\mathbf{k}} h(\mathbf{y}, \mathbf{X}_i\boldsymbol{\beta}, \mathbf{Q}, \mathbf{z}_{\mathbf{k}}) \nu_{\mathbf{k}}^R \right], \tag{4}
$$

where

$$
\begin{aligned}
&h(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{Q}, \mathbf{z}) \\
&\quad = \exp[\mathbf{y}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{z}) - \mathbf{1}' \log\{\mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{z})\}],
\end{aligned}
$$

$\mathbf{k} = (k_1, \ldots, k_R)$, $\mathbf{z}_{k_1 \cdots k_R} = (z_{k_1}, \ldots, z_{k_R})$ is a vector of univariate quadrature nodes, and the multivariate quadrature weights $\nu^R_{k_1 \cdots k_R}$ are products of the appropriate univariate quadrature weights.

We maximized (4) using FSQP (Zhao and Tits, 1994), a set of FORTRAN subroutines for numerically optimizing an objective function subject to linear and nonlinear constraints on the variables. One can fit the model with an unstructured $\boldsymbol{\Sigma}$ or with relevant special cases, such as a common variance with certain correlations equal to zero, others equal to one, and others equal to some unknown ρ. After fitting the model, one can make inferences in the usual manner using the model's log likelihood function. We used $I^{-1}(\hat{\boldsymbol{\theta}})$, where $I(\boldsymbol{\theta})$ is the observed information matrix, to estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$. We recommend plotting estimates and standard errors by the number of quadrature points q used in each dimension to ensure that this approximation has stabilized.

3. Examples of Multivariate Binomial Logit-Normal Models

We now provide two examples demonstrating the generality of the multivariate binomial logit-normal (BLN) model. First we analyze Table 1, for which multivariate random effects can account for some negative associations. Second, we analyze a developmental toxicity study, with logits applied to an ordinal outcome, in which a vector of random effects account for litter clustering.

3.1 *Repeated Univariate Response Permitting Negative Dependence: Influenza Data*

For the influenza data in Table 1, each subject has four binary responses (Y_1, Y_2, Y_3, Y_4), and we consider BLN models with $n = 1$. Let π_{sr} denote the probability that subject s gets the flu in year r and consider the BLN model

$$
\text{logit}(\pi_{sr}) = \alpha_{sr} + \beta_r. \tag{5}
$$

We first used a random effect covariance structure based on a sum of two components, one corresponding to subject heterogeneity and one corresponding to dependence between the first and fourth outbreaks because of the apparently common virus in years 1 and 4. Namely, we assumed that $\boldsymbol{\alpha}_s = \boldsymbol{\alpha}_{sH} + \boldsymbol{\alpha}_{sW}$, where $\boldsymbol{\alpha}_{sH} \overset{\text{i.i.d.}}{\sim} N^4(\mathbf{0}, \boldsymbol{\Sigma}_H)$ and $\boldsymbol{\alpha}_{sW} \overset{\text{i.i.d.}}{\sim} N^4(\mathbf{0}, \boldsymbol{\Sigma}_W)$ and $\boldsymbol{\alpha}_{sH}$ and $\boldsymbol{\alpha}_{sW}$ are independent, with

$$
\boldsymbol{\Sigma}_H = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \end{bmatrix}
$$

and

$$
\boldsymbol{\Sigma}_W = \begin{bmatrix} \sigma_2^2 & 0 & 0 & \rho\sigma_2^2 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_2^2 & 0 \\ \rho\sigma_2^2 & 0 & 0 & \sigma_2^2 \end{bmatrix}.
$$

This suggests fitting model (5) with $\boldsymbol{\alpha}_s$ covariance structure

$$
\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho_1\sigma^2 & \rho_1\sigma^2 & \rho_2\sigma^2 \\ \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 & \rho_1\sigma^2 \\ \rho_1\sigma^2 & \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 \\ \rho_2\sigma^2 & \rho_1\sigma^2 & \rho_1\sigma^2 & \sigma^2 \end{bmatrix}. \tag{6}
$$

Table 1 shows the BLN model fit with this correlation structure. Inspection of the parameter estimates as a function of q shows that the quadrature approximation stabilizes for $q > 17$. All fits reported here are based on $q = 20$ quadrature points. This model fit has likelihood-ratio goodness-of-fit statistic (deviance) $G^2 = 18.2$ with d.f. $= 8$, showing evidence of lack of fit. Model (5) with univariate random intercept is the special case of (6) in which $\rho_1 = \rho_2 = 1$, i.e., in which $\boldsymbol{\alpha}_s$ has only the $\boldsymbol{\alpha}_{sH}$ component and the random effect is the same for each response. This fit has $G^2 = 27.7$ with d.f. $= 10$. Not unexpectedly, this model also fits poorly since it forces all pairwise marginal log odds ratios to be nonnegative; e.g., the fitted marginal odds ratio for Y_1 and Y_4 is 1.44 compared with the sample value of .32. Table 1 also shows this model fit.

We next fitted BLN model (5) with unconstrained $\boldsymbol{\Sigma}$ to investigate why the model with structure (6) fits poorly. This model estimates that all correlations between outbreak four and the other three outbreaks are negative and suggests that the BLN model with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho_1\sigma^2 & \rho_1\sigma^2 & \rho_2\sigma^2 \\ \rho_1\sigma^2 & \sigma^2 & \rho_1\sigma^2 & \rho_2\sigma^2 \\ \rho_1\sigma^2 & \rho_1\sigma^2 & \sigma^2 & \rho_2\sigma^2 \\ \rho_2\sigma^2 & \rho_2\sigma^2 & \rho_2\sigma^2 & \sigma^2 \end{bmatrix} \tag{7}$$

may be reasonable. This would make biological sense if the outbreak in the fourth year shared viruses with the first three years. Table 1 shows the BLN fit with this covariance structure. This model fits the data well ($G^2 = 6.3$, d.f. $= 8$), and we gain 7 d.f. compared to the unstructured model by constraining the variance components to be equal and the correlations to have only two distinct values. The model yields $\hat{\boldsymbol{\beta}} = (-4.0, -4.4, -4.7, -4.5)$, $\hat{\sigma} = 4.05$, $\hat{\rho}_1 = .43$, and $\hat{\rho}_2 = -.25$. The $\hat{\boldsymbol{\beta}}$ components order the estimated probability of flu at the four outbreaks for individuals having common susceptibility α at each outbreak. For instance, for subject a at time 1 and subject b at time 2 for whom $\alpha_{a1} = \alpha_{b2}$, we estimate that the odds of succumbing to the first outbreak were $\exp[-4.0 - (-4.4)] = 1.5$ times the odds of succumbing to the second outbreak.

Section 4 shows equivalence in the sign of the correlation of the random effects and of the corresponding marginal responses. It follows that, for this model, the fitted log odds ratios for the marginal associations between Y_4 and each of Y_1, Y_2, and Y_3 are negative; in fact, the sample odds ratios are .32, .54, and .62, respectively. We also fitted the model that allows a possibly different random effects correlation between outbreaks 1 and 4 than between 2 and 4 and between 3 and 4, but this did not provide a significantly improved fit ($G^2 = 5.4$, d.f. $= 7$).

For these data, Darroch and McCloud (1990) separately measured two sources of dependence—population heterogeneity and negative within-subject $(1, 4)$ correlation. Their model specifies, for a given subject, mutual independence of the first three responses and, conditional on those responses, dependence of the fourth response on the first alone. This model also fits well ($G^2 = 5.5$, d.f. $= 5$). It provides an estimate of -2.27 for the log odds ratio relating the conditional association between Y_1 and Y_4, describing the within-subject immunity to outbreak 4 acquired by succumbing to outbreak 1. Their model, however, implies positive conditional associations between Y_2 and Y_4 given Y_1 and between Y_3 and Y_4 given Y_1, whereas the sample log odds ratios for these four cases are negative.

3.2 *Continuation-Ratio Logits For Clustered Ordinal Outcomes: Toxicity Data*

The BLN model also has application to continuation-ratio logit modeling of repeated or clustered ordinal responses since, for that multinomial model, the logits refer to sets of independent binomial variates (Fienberg, 1980). For the application discussed in this subsection, the dependence results from measurement on subjects in clusters of related subjects rather than from repeated measurement for each subject. Suppose there are N clusters and let π_{sj} denote the probability that a subject in cluster s makes response j, $j = 1, \ldots, J$, where for simplicity we suppress notation for explanatory variables. For cluster s, $\omega_{sj} = \pi_{sj}/(\pi_{sj} + \cdots + \pi_{sJ})$ is the conditional probability of response j, given response in category j or higher. The continuation-ratio logits are $\{\text{logit}(\omega_{sj}),\ j = 1, \ldots, J-1\}$.

Let $\{y_{sj},\ j = 1, \ldots, J\}$ be the number of subjects in cluster s having response j and let $m_s = \Sigma_{j=1}^{J} y_{sj}$ and $N_{sj} = m_s - \Sigma_{i<j} y_{si}$. For a given cluster in a continuation-ratio logit model, the multinomial response is equivalently a set of $J-1$ independent binomial counts $(y_{s1}, \ldots, y_{s(J-1)})$ out of $(N_{s1}, \ldots, N_{s(J-1)})$ trials with probability vector $\boldsymbol{\omega}_s = (\omega_{s1}, \ldots, \omega_{s(J-1)})$. Because of the independence of the $(J-1)$ responses for this model, for a given cluster effect, one can construct random effects models for the continuation-ratio logits as special cases of BLN models. The model has the form

$$\text{logit}\,(\boldsymbol{\omega}_s) = \boldsymbol{\alpha}_s + \mathbf{X}_s\boldsymbol{\beta}.$$

Here, for cluster s, $\mathbf{n}_s = (N_{s1}, N_{s2}, \ldots, N_{s(J-1)})$, unlike the application in Section 3.1 in which all binomial indices equal one.

We now apply this model to a developmental toxicity study conducted under the U.S. National Toxicology Program (Price et al., 1985). This study examined the developmental effects of ethylene glycol (EG) by administering one of four doses (0, .75, 1.50, 3.00 g/kg) to pregnant rodents, with (25, 24, 22, 23) animals in the four dose groups. The clusters are litters of mice. The three possible outcomes (dead/resorption, malformation, normal) for each fetus are ordered, normal being the most desirable result. The continuation-ratio logit is natural here since categories are hierarchically related; an animal must survive before a malformation can take place. A Dirichlet-trinomial model (Chen et al., 1991) and GEE methods (Ryan, 1992; Catalano, Ryan, and Scharfstein, 1994) estimate the population averaged effect of dose on a multinomial response in the presence of the litter effect. The use of random effect models to account for litter effects seems unexplored in the multinomial case.

Here, for litter s with dose i, $\text{logit}(\omega_{s(i)1})$ is the continuation-ratio logit for the probability of death and $\text{logit}(\omega_{s(i)2})$ is the continuation-ratio logit for the conditional probability of malformation, given survival. We account for the litter effect by the inclusion of litter-specific terms $\boldsymbol{\alpha}_{s(i)} = (\alpha_{s(i)1}, \alpha_{s(i)2})$ in the BLN model,

$$\text{logit}\left(\omega_{s(i)j}\right) = \alpha_{s(i)j} + \gamma_j + \beta_j x_i, \tag{8}$$

where x_i is the dose associated with the ith dose group and $\boldsymbol{\alpha}_{s(i)} \overset{\text{i.i.d.}}{\sim} \text{N}(0, \boldsymbol{\Sigma}_i)$. The multivariate random effect allows for differing amounts of litter effect overdispersion for the probability of death and for the probability of malformations, given survival.

Table 2 reports the maximized log likelihoods for model (8) and various special cases. The parameter estimates again stabilized for $q > 17$ quadrature points, and all fits reported are based on $q = 20$. The special cases are as follows:

(a) BLN^2 model with dose-specific random effects but common intercept and slope for the two logits, i.e., $\gamma_1 = \gamma_2$ and $\beta_1 = \beta_2$ in model (8);

Table 2
Maximized log likelihoods under the multivariate binomial logit-normal model and submodels for the ethylene glycol data of Price et al. (1985)

Multivariate binomial-logit normal model	Parameter dimension	Maximized log likelihood
BLN^2 with dose-specific $\boldsymbol{\Sigma}_i$ (model 8)	16	−457.3
BLN^2 with $\boldsymbol{\Sigma}_i$, common γ, β	14	−485.7
BLN^2 with common $\boldsymbol{\Sigma}$	7	−464.7
BLN^2 with common $\boldsymbol{\Sigma}$, $\rho = 0$	6	−464.7
Univariate σ^2	5	−474.0

(b) BLN^2 model with constant covariance matrix across the four doses, i.e., $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_4$;
(c) BLN^2 model with constant covariance matrix across dose and independent random effects, i.e., $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_4$ and $\rho = .0$;
(d) univariate dose-specific random effects, i.e., $\alpha_{s(i)1} = \alpha_{s(i)2}$ so that each entry of $\boldsymbol{\Sigma}_i$ is σ_i^2 and the correlation between random effects for the two logits equals one;
(e) the model with univariate constant variance component across dose, i.e., $\alpha_{s(i)1} = \alpha_{s(i)2}$ and $\sigma_i = \sigma$.

We can formally test the fit of the first three special cases against the general model (8) using likelihood-ratio tests. Since comparing the univariate models to their multivariate counterparts involves testing that the correlation parameters are on the boundary, we cannot formally compare these models using likelihood-ratio tests, but an informal analysis of maximized likelihoods seems adequate here. For these data, the independent multivariate random effects model with homogeneous variance structure (i.e., the fourth model listed in Table 2) appears to describe the data well. The parameter estimates are (with asymptotic standard errors for the dose effects in parentheses)

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} .3 & .0 \\ .0 & 2.5 \end{bmatrix},$$

$\hat{\gamma}_1 = -4.19$, $\hat{\gamma}_2 = -4.36$, $\hat{\beta}_1 = .08$ (.208), $\hat{\beta}_2 = 1.79$ (.225). The variance component estimates suggest a strong litter effect for the malformation outcome (given survival) but not for death. For a given cluster, there is no evidence of a dose effect on the death rate, but the estimated odds of malformation, given survival, are multiplied by exp(1.79) = 6.0 for every additional gram per kilogram of ethylene glycol that a mouse receives. The model, which corresponds to a separate univariate logistic-normal model for each conditional binomial outcome, specifies that the proportion of dead pups and the proportion of malformed pups (given survival) are independent both within litter and marginally.

3.3 *Other Applications of the BLN Model*

Coull and Agresti (1998) showed other applications of the BLN model. The model is often natural when two or more binary variables are measured repeatedly, such as in a longitudinal study. Here, different elements of the response vector refer to different characteristics. To illustrate, Coull and Agresti (1998) reanalyzed the leading-crowd data of Coleman (1964) that appears often in the discrete data literature (e.g., Agresti, 1997). The data consist of matched-pairs responses for two binary variables, and they used the BLN model with a bivariate normal random effect. A potential application with a repeated univariate response is in capture–recapture modeling for estimating population size. Coull and Agresti (1999) used a univariate model with random effects to permit animal heterogeneity. This model, however, cannot handle trap avoidance, resulting in negative associations between the capture responses at different occasions.

4. Characteristics of the Multivariate Binomial Logit-Normal Distribution

We next study some properties of the BLN distribution, which forms the basis of the models in this paper. Unfortunately, there are no closed-form expressions for its moments since such expressions do not exist for the moments of a multivariate logit-normal random variable. Let $\mathbf{Y}$ denote a $BLN^R(\mathbf{n}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ variate. Then

$$\mathrm{E}(Y_r) = \mathrm{E}\{\mathrm{E}(Y_r \mid \pi_r)\} = n_r \mathrm{E}(\pi_r)$$

and

$$\mathrm{cov}(Y_r, Y_{r'}) = n_r n_{r'} \mathrm{cov}(\pi_r, \pi_{r'}) + I(r = r') n_r \mathrm{E}\{\pi_r(1-\pi_r)\}, \tag{9}$$

where $I(\cdot)$ is an indicator function. Let $\pi_r^* = \exp(\mu_r)/[1 + \exp(\mu_r)]$. Taylor series expansions show that, for small $\{\sigma_{rr}\}$,

$$\begin{aligned} \mathrm{E}(\pi_r) &\approx \pi_r^*, \\ \mathrm{var}(\pi_r) &\approx \sigma_{rr}\left[\pi_r^*\left(1-\pi_r^*\right)\right]^2, \\ \mathrm{cov}(\pi_r, \pi_{r'}) &\approx \sigma_{rr'}\pi_r^*\left(1-\pi_r^*\right)\pi_{r'}^*\left(1-\pi_{r'}^*\right). \end{aligned}$$

We conducted a simulation study of various properties of the BLN distribution. This indicates that these approximations tend to break down when $\sigma_{rr} > .6$. For $\sigma_{rr} \leq .6$,

$$\begin{aligned} \mathrm{E}(Y_r) &\approx n_r \pi_r^* \\ \mathrm{cov}(Y_r, Y_{r'}) &\approx n_r n_{r'} \sigma_{rr'} \pi_r^*(1-\pi_r^*)\pi_{r'}^*(1-\pi_{r'}^*) \\ &\quad + I(r = r') n_r [\pi_r^*(1-\pi_r^*)][1 - \sigma_{rr'}\pi_r^*(1-\pi_r^*)]. \end{aligned}$$

In particular, for small normal dispersion,

$$\begin{aligned} &\mathrm{corr}(Y_r, Y_{r'}) \\ &\quad \approx n_r n_{r'} \sigma_{rr'} \pi_r^*(1-\pi_r^*)\pi_{r'}^*(1-\pi_{r'}^*) \\ &\qquad \div \left\{ \left[\left(n_r^2 - n_r\right) \sigma_{rr} \left[\pi_r^*(1-\pi_r^*)\right]^2 + n_r \pi_r^*(1-\pi_{r'}^*) \right] \right. \\ &\qquad\quad \times \left[\left(n_{r'}^2 - n_{r'}\right) \sigma_{r'r'} \left[\pi_{r'}^*(1-\pi_r^*)\right]^2 \right. \\ &\qquad\qquad \left. \left. + n_{r'} \pi_{r'}^*(1-\pi_{r'}^*) \right] \right\}^{1/2}. \end{aligned} \tag{10}$$

Although it is not possible to provide general properties of the BLN distribution, results for some special cases are apparent. If each $\sigma_{rr} = 0$, then the R marginal variables are independent binomials. If the normal correlations $\{\rho_{rr'} = 0\}$, then the R counts are independent but have extra-binomial dispersion, the degree of overdispersion governed by $\{\sigma_{rr}\}$. Also, for $r \neq r'$, (9) yields $\mathrm{cov}(Y_r, Y_{r'}) = n_r n_{r'} \mathrm{cov}(\pi_r, \pi_{r'})$. Since $\{\pi_r\}$ are monotone increasing functions of the normal components, the pairwise correlations $\mathrm{corr}(Y_r, Y_{r'})$ and $\mathrm{corr}(\pi_r, \pi_{r'})$ have the same sign as $\rho_{rr'}$. In fact, from

expression (10), for small and fixed normal variation and fixed normal means, $\text{corr}(Y_r, Y_{r'})$ increases as $\rho_{rr'}$ does. For small normal variation and extremely large n_r and $n_{r'}$, expression (10) also implies that $\text{corr}(Y_r, Y_{r'})$ is approximately equal to $\rho_{rr'}$ and that $\text{corr}(Y_r, Y_{r'})$ is increasing in σ_{rr} and in $\sigma_{r'r'}$ for fixed means and $\rho_{rr'}$.

Finally, we briefly mention the special case

$$\text{BLN}^2\left(\begin{pmatrix} n \\ n \end{pmatrix}, \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{bmatrix} \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 \end{bmatrix}\right).$$

Simulations indicate that $\text{corr}(Y_1, Y_2)$ depends strongly on σ, ρ, and n but only weakly on μ. For fixed ρ, not surprisingly, this correlation is stronger as σ increases for fixed n and as n increases for fixed σ. For further details, see Coull and Agresti (1998).

5. Comments

This final section discusses some alternative models and alternative ways of fitting models for multiple-response data.

5.1 *Alternative Models for Multiple-Response Data*

The models described in this article assumed a multivariate normal distribution for the random effects. An obvious question concerns the effect of misspecification. For the univariate model (1), Neuhaus, Hauck, and Kalbfleisch (1992) showed that, if the normal distribution can induce an intracluster correlation approximately equal to that for the actual mixing distribution, there is little bias in the estimation of fixed effects or the standard errors; when the actual distribution is highly skewed, there may be greater bias in estimating the overall intercept. Since a wide variety of mixing distributions can produce similar marginal distributions, an incorrect choice for the mixing distribution usually yields only slight bias. Research is now needed to investigate these issues for multivariate models.

To check empirically the consequence of choosing normality, one could compare results to those using a nonparametric random effects approach (Aitkin, 1996; Agresti, 1997). For the developmental toxicity data of Section 3.2, e.g., the litter-specific effects might follow a highly skewed distribution, so we applied Aitkin's approach. The slope estimates (with standard errors in parentheses) of .09 (.206) and 2.30 (.258) are somewhat different from the ones (.08, 1.79) under normality, but a comparison of them with standard errors provides the same substantive conclusions. For some multivariate logit models, it is possible to use a conditional ML approach with within-cluster effects (e.g., Agresti, 1997), and widely discrepant values of estimates between that and the normal random effects approach would also suggest caution.

An alternative approach for handling correlated binomial-type variates is with marginal models, for which effects of explanatory variables refer to the overall population rather than individual subjects. Best known for marginal modeling is the GEE methodology (Liang and Zeger, 1986). For marginal models based on a fully specified joint distribution and hence a likelihood function, ML fitting is possible but can be awkward; it has been used mainly for some relatively simple problems (e.g., Fitzmaurice, Laird, and Rotnitzky, 1993; Lang and Agresti, 1994). The GEE approach has the advantage of simplicity. When the models of this paper or marginal ML models are feasible, an advantage of them over GEE methods is that they provide a complete model: they have a likelihood function, making likelihood-ratio methods of inference possible. Otherwise, one's choice of model may reflect whether subject-specific descriptions have more relevance than population-averaged ones.

In some cases, analyses with similar effect as those of this paper result from using a univariate random effects model. Consider, e.g., the model

$$\text{logit}\,(\pi_{sr}) = \tau_r\alpha_s + \mathbf{x}_{sr}\boldsymbol{\beta}, \tag{11}$$

where $\alpha_s \overset{\text{i.i.d.}}{\sim} \text{N}(0,1)$. This is the special case of model (2) that specifies $\rho_{rr'} = 1$ if τ_r and $\tau_{r'}$ have the same sign and $\rho_{rr'} = -1$ otherwise. The discrimination parameters $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_R)$, which allow the R binary outcomes to be associated differently with the latent variable α_s, have been used in educational testing settings and in teratology studies (Legler and Ryan, 1997). This model requires only one-dimensional quadrature for fitting while still being capable of describing both positive and negative association structures among the R responses. It provides a good fit to the influenza data ($G^2 = 4.1$, d.f. $= 7$). It would be interesting to analyze both the bias incurred by fitting the simpler model (11) when the more general model (2) holds under a variety of correlation structures and the increase in variance incurred by fitting model (2) when the simpler model (11) holds.

5.2 *Alternative Fitting Methods for BLN Models*

In comparing integral approximation techniques, Evans and Schwartz (1995) recommended multiple quadrature over other methods when $R \leq 6$, but they pointed out that quadrature becomes computationally impractical for larger R. In fact, with the BLN model, we had computational difficulties when $R \geq 5$, particularly since stable estimates of variance components and standard errors may require large values of q. Alternative methods are then more appropriate. Recent research on GLMMs has provided Monte Carlo EM algorithms for cases in which numerical integration is infeasible (McCulloch, 1997; Booth and Hobert, 1999). The M step in this EM algorithm maximizes a Monte Carlo approximation of the expected value of the complete log likelihood given previous values of the parameter estimates and the observed data. The above articles take different approaches to generating the E-step Monte Carlo samples, with the Booth and Hobert approach being noteworthy in that it provides an estimate of the potential Monte Carlo error. Alternatively, one might use Breslow and Clayton's (1993) penalized quasi-likelihood approach, based on a Laplace approximation for the integral. This yields, however, a rougher approximation for the ML estimates, especially when variance components are large or data are far from normal (e.g., binary). Finally, Stiratelli et al. (1984) used an empirical Bayes approach with a diffuse prior for the regression parameters.

The models presented in this paper generalize to a vector of multinomial variates, i.e., to a multivariate multinomial logit-normal model. Also, our binomial mixture model and the Poisson mixture model in Aitchison and Ho (1989) are special cases of a multivariate generalized linear mixed model. More generally, different elements in $\mathbf{Y}$ could have different

conditional distributions. Finally, a key component of the BLN models is the use of a normal distribution for the logit transform of the binomial parameters. A similar approach is used in models for compositional data (Aitchison and Shen, 1980), which are random vectors on the simplex.

Acknowledgements

This work was supported by NIEHS grants ES07142 and ES05860 (BAC) and grants from NIH and NSF (AA). The authors would like to thank Ralitza Gueorguieva and Jim Hobert for helpful comments.

Résumé

La distribution binomiale logit-multinormale est une distribution de mélange pour laquelle (1) conditionnellement à un jeu de probabilités de succès et des indices de taille d'échantillon, le vecteur des fréquences de succès est constitué de variables binomiales indépendantes, et (2) le vecteur des logits des paramètres a une distribution multinormale. Nous utilisons cette distribution pour modéliser des réponses multivariées de type binomial en utilisant un vecteur d'effets aléatoires. Le vecteur des logits des paramètres a une moyenne qui est une fonction linéaire des variables explicatives et a une matrice de covariance non ou partiellement spécifiée. Le modèle généralise et fournit une plus grande flexibilité que le modèle univarié qui utilise un effet gaussien aléatoire pour prendre en compte une corrélation positive dans des données en grappe. Le modèle multivarié est utile quand des éléments différents du vecteur de réponse réfèrent à des caractéristiques différentes, chacune d'elle peut naturellement avoir son propre effet aléatoire. Il est aussi utile pour les mesures répétées binaires d'un simple réponse quand il n'y a une structure d'association non échangeable., tel qu'on le suspecte souvent dans les données longitudinales ou quand il existe une association négative pour au moins une paire de réponses. Nous appliquons le modèle a une étude sur la grippe avec des mesures répétées dans lesquelles certaines paires sont négativement associées et à une étude de toxicité sur le développement avec des logits de ratio continus appliqués à une réponse ordinale et des données en grappe.

References

Agresti, A. (1997). A model for repeated measurements of a multivariate binary response. *Journal of the American Statistical Association* **92,** 315–321.

Aitchison, J. and Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika* **76,** 643–653.

Aitchison, J. and Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* **67,** 261–272.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* **6,** 251–262.

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61,** 265–285.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Catalano, P. J., Ryan, L. M., and Scharfstein, D. O. (1994). Modeling fetal death and malformation in developmental toxicity. *Risk Analysis* **14,** 629–637.

Chen, J. J., Kodell, R. L., Howe, R. B., and Gaylor, D. W. (1991). Analysis of trinomial responses from reproductive and developmental toxicity experiments. *Biometrics* **47,** 1049–1058.

Coleman, J. S. (1964). *Introduction to Mathematical Sociology.* London: Free Press of Glencoe.

Coull, B. A. and Agresti, A. (1998). *Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution.* Technical Report 572, Department of Statistics, University of Florida, Gainesville.

Coull, B. A. and Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture–recapture studies. *Biometrics* **55,** 294–301.

Darroch, J. N. and McCloud, P. I. (1990). Separating two sources of dependence in repeated influenza outbreaks. *Biometrika* **77,** 237–243.

Evans, M. and Schwartz, T. (1995). Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science* **10,** 254–272.

Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data,* 2nd edition. Cambridge, Massachusetts: MIT Press.

Fitzmaurice, G. M., Laird, N. M., and Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statistical Science* **8,** 284–309.

Haber, M. (1986). Testing for pairwise independence. *Biometrics* **42,** 429–435.

Lang, J. B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* **89,** 625–632.

Legler, J. M. and Ryan, L. M. (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association* **92,** 13–20.

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

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92,** 162–170.

Neuhaus, J. M., Hauck, W. W., and Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* **79,** 755–762.

Price, C. J., Kimmel, C. A., Tyl, R. W., and Marr, M. C. (1985). The developmental toxicity of ethylene glycol in rats and mice. *Toxicology and Applied Pharmacology* **81,** 113–127.

Ryan, L. M. (1992). Quantitative risk assessment for developmental toxicity. *Biometrics* **48,** 163–174.

Stiratelli, R., Laird, N., and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40,** 961–971.

Zhou, J. L. and Tits, A. L. (1994). *User's Guide for FSQP,* Version 3.4, *A FORTRAN Code for Solving Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality and Linear Constraints.* Technical Report TR-92-107r4, Systems Research Center, University of Maryland, College Park.

Received July 1998. *Revised March* 1999. *Accepted May* 1999.