Journal of Official Statistics, Vol. 17, No. 1, 2001, pp. 149–161

Estimation of Interviewer Effects on Multivariate Binary Responses in a Community Based Survey

Sujuan Gao[1]

Estimation of interviewer effect in a survey is important in providing means to control nonsampling errors. It is often desirable to estimate the overall interviewer effect when multiple responses are obtained by one interviewer, especially in surveys where the interviewers' subjective judgment is used in determining the outcomes of the response. In this article I propose a two-stage logistic normal model for estimating interviewer effects on multivariate binary responses from surveys. The model is an extension to Anderson and Aitkin (1985) in univariate binary cases. A maximum likelihood estimation method using Gauss-quadrature points are presented with discussions on computational aspects of the approach. Data from a community based dementia survey are used to motivate and illustrate the proposed approach.

Key words: Multivariate binary outcomes; logistic normal model; Gauss quadrature.

1. Introduction

In sample surveys it is known that measurement errors such as interviewer errors from the data collection process may exceed sampling errors and become the major components in total survey errors. The presence of these nonsampling errors can inflate the variance of the sample mean by the relationship $\sigma^2(1 + (\bar{n} - 1)\rho)$, where σ^2 is the random error associated with each sampling unit, $\bar{n}$ is the average number of units processed by the same interviewer, and ρ is the intra-cluster correlation induced by the interviewer. With large $\bar{n}$ even a small ρ can increase the variance of the sample mean considerably. Therefore, substantial efforts have been devoted to first estimating and then controlling these nonsampling errors in surveys.

Estimation of interviewer effects is also used as a quality control tool in training interviewers and in designing and selecting questionnaire items. Many authors have studied the problem of estimating interviewers' variability on a continuous response from interpenetrated survey designs. See, for example, Fellegi (1974), Biemer and Stokes (1985), and Kleffe, Prasad, and Rao (1991). Anderson and Aitkin (1985) considered estimating interviewer effects from a binary item and their method is extended by Stokes (1988) to a categorical item. In these two studies it is assumed that there is no confounding between interviewers and sampling units. The Anderson and Aitkin (1985) approach for univariate binary response has been extended by Aitkin (1999) to a nonparametric

[1] Division of Biostatistics, Department of Medicine, Indiana University School of Medicine, 1050 Wishard Blvd, RG4101, Indianapolis, IN 46202-2872, U.S.A. Email: sgao@iupui.edu
Acknowledgments: This research was supported in part by NIH grants R01 AG15813, R01 AG09956 and P30 AG10133.

maximum likelihood method for generalized linear models. The extension focused on variance component models where different random effects are assumed to be independent.

In practice multiple binary responses are often obtained by interviewers in a survey. Although separate estimation of interviewer variation on each item may provide information on item selection, such an approach does not provide an overall estimate of interviewer variability. For quality control purposes, it may be desirable to estimate the overall interviewer effects on these multiple binary responses.

Many previous publications on multivariate binary models have considered regression models by specifying the multivariate binary distribution either entirely or partially. These include Lipsitz and Fitzmaurice (1994), Fitzmaurice (1995), Ekholm, Smith and McDonald (1995), and Molenberghs and Ritter (1996), to name just a few recent publications. In these models the correlations between multivariate variables are usually considered nuisance parameters and the main focus is on the estimation of mean regression parameters. McDonald (1994) considered multivariate binary responses using a ''discrete'' random effect model approach where the random effect parameter can take only one of two values so that closed form expressions can be obtained from the maximum likelihood approach.

In this article I consider the problem of estimating interviewer effects from multivariate binary responses. The binary responses are assumed to follow a two-stage logistic normal model. Interviewer effects are estimated by a maximum likelihood method using Gauss-quadrature points. A community based dementia survey is described in the following section to motivate the research. The data are used in Section 5 to illustrate the proposed method.

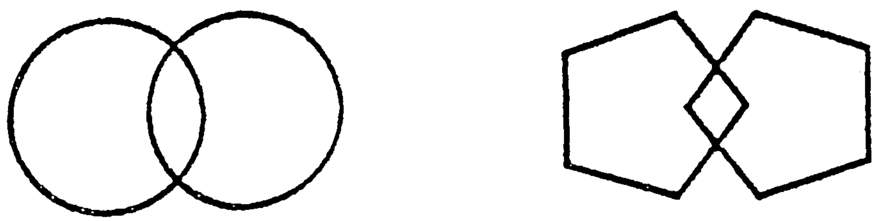
2. A Community Based Dementia Survey

In a community based study concerning dementia, African Americans age 65 and older in Indianapolis, Indiana, U.S.A., were randomly selected from the community to enter a longitudinal study on dementia and Alzheimer's disease. Twenty-nine census tracts in Indianapolis with more than 80 percent African Americans in the population were identified. A random list of 60 percent of household addresses within these tracts was generated by the local water company. Interviewers were sent to these addresses to locate subjects eligible for the study. Upon agreeing to join the study, the elderly subject was interviewed in his or her home. Part of the interview is to assess the participant's cognitive functioning with a series of questionnaire items adopted from several well-known neuropsychological tests. A total score from the cognitive test is derived as the total number of correct answers given by the subject. This score is subsequently used to determine if the person suffers from cognitive impairment and whether he or she should be invited to undergo more elaborate clinical assessment for dementia and Alzheimers's disease. Details of the study design have been published (Hendrie et al. 1995; Hall et al. 1996).

Most of the items in the cognitive interview involve memory testing where it is clear to the interviewer whether an answer is correct. However, there is also a section where the goal is to test the participant's praxis responses. In two questionnaire items, the study participant was asked to copy two interlocking circles and two interlocking pentagons,

Praxis - Copying
Now I would like for you to take my pencil and copy these figures in this space.

45. 46.



Incorrect.....0
Correct.......1

Incorrect.....0
Correct.......1

Fig. 1. The praxis questions

respectively (Figure 1). The interviewer then recorded the responses as correct or incorrect using the following instructions: score the two interlocking circles as correct if the two circles are closed with a visible intersection area; score the interlocking pentagons as correct if two five-sided figures overlap to form a four sided figure. Out of all questions in the cognitive part of the questionnaire these two items showed considerably larger interviewer variation than the other items.

Fifteen interviewers were used to obtain 2,212 interviews. Of these interviews 2,148 participants had complete information on both praxis items. Missing responses are mainly due to some subjects being too sick to complete the interview. In Table 1 I present the average percentage of correct answers as well as the number of interviews conducted by interviewers. The average percentages of correctly copying the two interlocking circles varied from 82 percent to 100 percent and the average percentages of correctly copying the two interlocking pentagons varied from 41 percent to 74 percent. It is also worth noting that the interviewers recording higher percentage correct for the first item also recorded higher percentage correct for the second item. For training and quality control purposes it is of interest to determine if there is significant interviewer variation in judging the two items and to estimate the extent of such variations.

For large-scale surveys, a inter-penetration scheme is often used to break the confounding factor between sampling units and interviewers. However, in this dementia survey, since the community is already a small geographical area, we assume that there is no interaction between interviewer effect and sampling units.

3. The General Multivariate Model

Suppose that m interviewers are used in a survey to obtain answer on K binary items. The ith interviewer conducts n_i interviews, $i = 1, \ldots, m$. Let y_{ijk} be the kth response obtained from the jth subject interviewed by the ith interviewer, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$ and $k = 1, \ldots, K$. Let x_{ij} be a $1 \times p$ covariate vector for the jth subject

Table 1. Numbers of interviews conducted by each interviewer and average percentage of correct answers on the two cognitive test items. Data from the Indianapolis-Ibadan Dementia Project

Interviewer	No. Interviewed	Circles (%)	Pentagons (%)
1	259	94.59	60.23
2	182	95.60	56.59
3	188	91.49	45.74
4	302	92.43	54.93
5	280	90.71	56.07
6	356	91.01	52.25
7	60	96.67	55.00
8	70	88.57	41.43
9	34	100	73.53
10	40	95.00	67.50
11	39	89.74	61.54
12	42	85.71	50.00
13	184	88.59	48.91
14	60	91.67	66.67
15	50	82.00	44.00

interviewed by the ith interviewer. We assume the following two-stage model for the response variable y_{ijk}.

Stage 1. y_{ijk} follows a Bernoulli distribution:

$$y_{ijk} \sim \text{Bernoulli}\,(p_{ijk})$$

Stage 2. The logits of p_{ijk} follow a mixed effect linear model:

$$\log \frac{p_{ijk}}{1-p_{ijk}} = \alpha_k + x_{ij}\beta + \gamma_{ik} \tag{1}$$

where α_k is the intercept term for the logit of the kth marginal mean parameter, β is a $p \times 1$ vector of fixed effect parameters, the vector $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{ik})'$ is multivariate normal with mean 0 and variance-covariance matrix D, where D is a $K \times K$ positive-definite matrix. We further assume conditional independence of y_{ijk} given random effect parameters.

The above model includes fixed effects so that covariates that are related to the probability of marginal response can be adjusted for. It assumes different marginal probabilities for the K binary items by including K intercepts in the model, reflecting the varying degree of difficulty on different items. The model can be easily modified to other forms of fixed effects. For example, different covariate effects can be modeled on different items. Here we focus on the modeling of interviewer effects instead. The variance-covariance matrix D is of general structure and can have the following as special cases:

Case 1. $D = 0$. No random effect due to interviewers is assumed on the logit scale. A regular logistic model is used to fit the data.

Case 2. $D = \sigma^2 I$, where I is a $K \times K$ identity matrix. This model assumes that an interviewer imposes the same effect on all K items.

Case 3.

$$D = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k^2 \end{bmatrix}$$

This model assumes that interviewer effects may vary from item to item.

We are interested in estimation and inference on the variance-covariance matrix D as well as on the fixed effect parameters α_k and β. This is in contrast to the approaches taken by many previous studies where fixed effect parameter estimates are the sole focus. Several estimation approaches can be used for parameter estimation from the two-stage model. One is the maximum likelihood approach. The advantage of a likelihood approach is that the likelihood ratio test can be used to compare models with different variance-covariance structure for D. However, models within the exponential family, except the normal distribution, with a normal random effect can be difficult to fit with the maximum likelihood approach because the resulting likelihood does not have a closed form.

The second approach is an approximate likelihood method such as the penalized quasi-likelihood (PQL, Breslow and Clayton 1993) and similar approaches (Goldstein 1995; Longford 1993). Disadvantages of the PQL method are that hypothesis tests on the variance components of interviewer effect are not in the model framework, although there are test procedures proposed in the literature (Lin 1996), and comparisons between nested models are not well established. The biggest concern for the PQL approach, however, was its seemingly lacking performance for binary response data, as shown in Rodriguez and Goldman (1995).

A Bayesian approach using Gibbs sampling for a two-stage logistic normal model has been proposed by Zeger and Karim (1991). However, the method is computationally intensive. Waclawiw and Liang (1993) proposed an empirical Bayes and estimating equation method for the logistic normal model. But their main focus is on the prediction of random effect parameters.

Since we are interested in both parameter estimation and model comparisons between various structure in the variance-covariance matrix D we choose to use the maximum likelihood approach here.

The marginal log-likelihood function under conditional independence assumption is:

$$l = \sum_{i=1}^{m} \log\left\{ \int \cdots \int \left[\prod_{j=1}^{n_i} \prod_{k=1}^{K} f_k(y_{ijk}|x_{ij}; \alpha_k, \beta, \gamma_{ik}) \right] g(\gamma_{i1}, \ldots, \gamma_{iK}|D) d\gamma_{i1} \ldots d\gamma_{iK} \right\} \tag{2}$$

where $f_k(y_{ijk}|x_{ij}; \alpha_k, \beta, \gamma_{ik})$ is the conditional distribution of y_{ijk}:

$$f_k(y_{ijk}|x_{ij}; \alpha_k, \beta, \gamma_{ik}) = \frac{e^{y_{ijk}(\alpha_k + x_{ij}\beta + \gamma_{ik})}}{1 + e^{\alpha_k + x_{ij}\beta + \gamma_{ik}}} \tag{3}$$

and $g(\gamma_{i1}, \ldots, \gamma_{iK}|D)$ is the probability distribution function of the multivariate normal distribution, $N(0, D)$.

In the univariate case considered by Anderson and Aitkin (1985) and Stokes (1988) Gauss-quadrature points were used to approximate the integration over the standard normal distribution. A similar approach is used here to avoid multi-dimension integration for the derivation of maximum likelihood estimates of α_k, β and D. First we introduce linear transformations of γ_i so that the integration is over standard normal distributions.

Let B be the triangular decomposition matrix of D, i.e., $BB' = D$. Now let $\delta_i = (\delta_{i1}, \ldots, \delta_{iK})' = B^{-1}\gamma_i$. We have $E(\delta_i) = 0$, $\text{var}(\delta_i) = B^{-1}$ $\text{var}(\gamma_i)$ $(B^{-1})' = B^{-1}BB'$ $(B^{-1})' = \text{I}$. Hence δ_i follows standard multivariate normal distribution. The log-likelihood function can then be rewritten as:

$$l = \sum_{i=1}^{m} \log\left\{ \int \cdots \int \left[\prod_{j=1}^{n_i} \prod_{k=1}^{K} f_k(y_{ijk}|x_{ij}; \alpha_k, \beta, \delta_{ik}) \right] \Phi(\delta_{i1}, \ldots, \delta_{iK}|D)\mathrm{d}\delta_{i1} \ldots \mathrm{d}\delta_{iK} \right\} \tag{4}$$

where $\Phi(\delta_{i1}, \ldots, \delta_{iK}|D)$ is the probability density function of the K-dimension standard normal distribution.

Let z_q be the qth Gauss-quadrature points and w_q be the weight factor associated with it (Abramowitz and Stegun 1972). In the general univariate case, integration of function f over the standard normal distribution can be approximated as:

$$\int f(x|\gamma)\Phi(\gamma)d\gamma \approx \sum_{q} f(x|z_q)w_q$$

In the multivariate case, the log-likelihood function can be approximated by

$$l = \sum_{i=1}^{m} \log\left\{ \sum_{q_1} \cdots \sum_{q_K} \left[\prod_{j=1}^{n_i} \prod_{k=1}^{K} f_k(y_{ijk}|x_{ij}; \alpha_k, \beta, z_{q_1}, \ldots, z_{q_K}) \right] w_{q1} \ldots w_{qK} \right\} \tag{5}$$

The Anderson and Aitkin (1985) approach for univariate binary response has been extended by Aitkin (1999) to a nonparametric maximum likelihood method for generalized linear models. However, the extension focused on variance component models where different random effects are assumed to be independent. The model discussed in this article allows a general form of correlation between random effects and maybe more realistic to use in the situation we described for the community survey data.

We discuss estimation using the maximum likelihood method in more details for the bivariate model in the following section.

4. The Bivariate Model

When $K = 2$, the log-likelihood function is

$$l = \sum_{i=1}^{m} \log\left\{ \int \int \left[\prod_{j=1}^{n_i} \frac{e^{y_{ij1}(\alpha_1 + x_{ij}\beta + \gamma_{i1})}}{1 + e^{\alpha_1 + x_{ij}\beta + \gamma_{i1}}} \frac{e^{y_{ij2}(\alpha_2 + x_{ij}\beta + \gamma_{i2})}}{1 + e^{a_2 + x_{ij}\beta + \gamma_{i2}}} \right] g\,(\gamma_{i1}, \gamma_{i2})\mathrm{d}\gamma_{i1}\mathrm{d}\gamma_{i2} \right\} \tag{6}$$

γ_{i1} and γ_{i2} have bivariate normal distribution with mean 0 and a general variance-covariance matrix

$$D = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma 2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Therefore,

$$B = \begin{bmatrix} \sigma_1\sqrt{1-\rho^2} & \rho\sigma_1 \\ 0 & \sigma_2 \end{bmatrix}$$

The following transformation is used in the log-likelihood function so that the joint distribution of δ_{i1} and δ_{i2} are standard normal:

$$\gamma_{i1} = \sigma_1\sqrt{1-\rho^2}\delta_{i1} + \rho\sigma_1\delta_{i2} \tag{7}$$

$$\gamma_{i2} = \sigma_2\delta_{i2} \tag{8}$$

Using Q Gauss quadrature points the log-likelihood function can be approximated by:

$$l = \sum_{i=1}^{m} \log\left\{ \sum_{q_1=1}^{Q}\sum_{q_2=1}^{Q}\left(\prod_{j=1}^{n_i} f_{ij1}f_{ij2}\right) w_{q_1}w_{q_2} \right\} \tag{9}$$

where

$$f_{ij1} = \frac{e^{y_{ij1}(\alpha_1 + x_{ij}\beta + \sigma_1\sqrt{1-\rho^2}q_1 + p\sigma_1 q_2)}}{1 + e^{\alpha_1 + x_{ij}\beta + \sigma_1\sqrt{1-\rho^2}q_1 + p\sigma_1 q_2}} \tag{10}$$

$$f_{ij2} = \frac{e^{y_{ij2}(\alpha_2 + x_{ij}\beta + \sigma_2 q_2)}}{1 + e^{\alpha_2 + x_{ij}\beta + \sigma_2 q_2}} \tag{11}$$

For univariate binary response Anderson and Aitkin (1985) have shown that the score equations are the same as those of weighted logistic regression models where the weights contain unknown parameters. Anderson and Aitkin suggested an EM algorithm between weighting and solving the score equation via logistic regression until parameter convergence. This method, as noted by Stokes (1988), increases computational load significantly when the number of quadrature points increases.

For the bivariate case and for the general multivariate cases the score equations from the log-likelihood function are generally not those from logistic regression, except in the situation where f_{ij1} and f_{ij2} do not contain any common parameters, which is equivalent to assuming complete independence between the two binary responses. When the model contains unrelated regression coefficients, the estimation can be done using a slight extension of the Anderson and Aitkin EM algorithm: two separate logistic regressions can be performed using two sets of quadrature points with given weights in the M step and then the weights are recomputed using both log-likelihood functions in the E step. However, in general the iterative algorithm of Anderson and Aitkin (1985) is not applicable to the general multivariate cases.

To derive maximum likelihood estimates from the log-likelihood function (9), we use a nonlinear numerical programming technique implemented in SAS procedure NLP (Non-Linear Programming). The NLP procedure is designed to maximize or minimize a non-linear function using numerical algorithms such as the Newton-Raphson ridge method. The nonlinear programming procedure is minimally affected by the increase in the number of quadrature points. The SAS technical report on this procedure offers detailed description of the algorithm and examples (SAS/OR Technical Report 1997).

One practical problem of using this procedure on data with large numbers of interviews conducted by each interviewer is numerical underflow. The numerical values inside the parentheses for log transformation may become too small for the algorithm to handle. One solution is to multiply $f_{ij1} f_{ij2}$ with a constant to enlarge the numerical values. Suppose a constant c is used in the log-likelihood function so that a different function l^* is obtained:

$$l^* = \sum_{i=1}^{m} \log \left\{ \sum_{q_1} \sum_{q_2} \left(\prod_{j=1}^{n_i} c f_{ij1} f_{ij2} \right) w_{q_1} w_{q_2} \right\} \tag{12}$$

It can be shown that l^* and the original log-likelihood function are related in the following way:

$$l = l^* - \sum_{i=1}^{m} n_i \log c$$

Therefore, the constant c can be determined in cases of numerical underflow problems with the original log-likelihood function. One can then seek maximum likelihood estimates by maximizing l^* instead of l and obtain the estimates of the log-likelihood function using the above equation.

5. Analysis of the Dementia Data

In this section I present results analyzing the dementia data described in Section 2. Since cognitive testing results have been shown to be affected by several other factors such as subjects' age, gender, years of education and whether they have suffered a stroke, I include these four variables in our bivariate model of interviewer effects so that variations explained by these covariates are accounted for. In addition, the marginal percentages of correct answers for circles and pentagons are quite different, reflecting perhaps a difference in the difficulty of, or subjects' familiarity with, the two tasks. I therefore include two separate intercept terms for the two items.

We fit a series of nested models to the dementia data varying only in the structure of variance covariance matrix D.

Model 1. $D = 0$

Model 2. $D = \sigma^2 I$.

Model 3. $D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

Model 4. $D = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1^2 \\ \rho\sigma_1^2 & \sigma_1^2 \end{bmatrix}$

Model 5. $D = \begin{bmatrix} \sigma_1^2 & \rho\sigma_2\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

For the first model we assume $D = 0$, which is equivalent to a fixed effect logistic model. In the second model we assume that interviewers have the same effect on both responses. In the third model we assume a different interviewer effect for the two items. In Model 4, a correlation structure is considered. Model 5 represents the most general variance-covariance structure.

The analysis was programmed in SAS using 10 quadrature points from Abramowitz and Stegun (1972). The 10 quadrature points offered improvement over 5 points and gave results not very different from those obtained using 20 quadrature points on a subset of models. Maximum likelihood estimates are obtained using the SAS procedure NLP. In addition to parameter estimation, the NLP procedure provides standard error estimates by using the numerical second derivative matrix derived from the parameter estimates. Since the first model is a regular logistic regression model without random effects, we fit the model using both NLP procedure and the regular logistic procedure. The results using the two procedures are identical. Analyses for all models are conducted running SAS on a shared UNIX station in the lowest priority batch mode. Computing times varied depending on the number of parameters in the model. The fitting of the fixed effect model took less than one minute while the fitting of Model 5 took about one and a half hours computing time.

We first consider models with common covariate effects for the two responses. Parameter estimates and likelihood ratio tests for comparison of models are presented in Table 2. As we can see from this table, parameter estimates for the fixed effects, α and β, changed little among the models. The likelihood ratio statistic comparing Model 2 to Model 1, a test for $\sigma^2 = 0$, was shown to follow a 50 : 50 mixture of two chi-squared distributions (Liang and Self 1987). The test indicates significant interviewer effects on both responses. The test comparing Model 3 to Model 2 suggests that interviewer effects on the first resposne are not significantly different from those on the second response. The log-likelihood values from Models 3, 4 and 5 indicate no significant improvement in the likelihood function from that of Model 2. Our analysis result suggests that there is a significant overall interviewer effect on these two responses.

In Table 3 we report symbolically the analysis results assuming different regression coefficients, but only for Models 1 and 2, because Model 3 represents a complete separation for the two responses. Thus two separate logistic normal models can be fitted to the data. Our focus here is again on the estimation of the interviewer effect. Two sets of likelihood ratio test are included in Table 3. The first set compares two nested models with the same set of fixed effect parameters, but with different variance-covariance structure (Model 1 versus Model 2 in Table 3). The second set compares the models with different regression coefficients, assuming the same variance-covariance structure. Hence the second comparison is of the respective models in Table 2 with those in Table 3. The results demonstrate again that the model with interviewer variance is a more appropriate model than the fixed effect model. They also indicate that the model assuming common regression coefficients is appropriate for the data. Note that these tests are performed for the collective group of covariates. It is foreseeable to allow different regression coefficients for certain covariates (for example, the gender variable in this data) and enforce common regression coefficients for the rest of the covariates.

To compare the proposed approach to the penalized quasi-likelihood estimation procedure, we present results in Table 4 for variance-covariance structure in the form of Model 2. The SAS macro GLIMMIX is used, which permits the random effect to enter

Table 2. *Maximum likelihood estimates and likelihood ratio test results for five models of interviewers' effect and common regression coefficients for the fixed effects estimates are in parentheses*

Variable Names	Parameter Estimates	Model 1 $\sigma_1^2=\sigma_2^2=0$ $\rho=0$	Model 2 $\sigma_1^2=\sigma_2^2$ $\rho=0$	Model 3 $\sigma_1^2\neq\sigma_2^2$ $\rho=0$	Model 4 $\sigma_1^2=\sigma_2^2$ $\rho\neq 0$	Mode $\sigma_1^2\neq$ $\rho\neq 0$
Intercept circles	$\hat{\alpha}_1$	3.8118 (0.5020)	3.8256 (0.5085)	3.8290 (0.5116)	3.8263 (0.5062)	3.79
Intercept pentagon	$\hat{\alpha}_2$	1.2255 (0.4928)	1.2292 (0.4990)	1.2263 (0.4985)	1.2290 (0.4961)	1.2
Age	$\hat{\beta}_1$	−0.0462 (0.0060)	−0.0462 (0.0060)	−0.0462 (0.0060)	−0.0458 (0.0060)	−0.04
Male	$\hat{\beta}_2$	0.1923 (0.0875)	0.1738 (0.0896)	0.1744 (0.0896)	0.2038 (0.0881)	0.20
Years of education	$\hat{\beta}_3$	0.2432 (0.0144)	0.2448 (0.0145)	0.2450 (0.0145)	0.2430 (0.0145)	0.24
Stroke	$\hat{\beta}_4$	−0.5753 (0.1266)	−0.5761 (0.1275)	−0.5773 (0.1276)	−0.5821 (0.1272)	−0.58
Covariance parameters	$\hat{\sigma}_1^2$	–	0.1123 (0.0766)	0.1902 (0.2103)	0.0851 (0.0560)	0.10
	$\hat{\sigma}_2^2$	–	–	0.0956 (0.0760)	–	0.02
	$\hat{\rho}$				0.7581 (0.5901)	0.73
Log likelihood	l	−1845.47	−1822.08	−1821.97	−1821.30	−182
Likelihood ratio statistics	$-2\Delta l$	–	46.78^{2-1}	0.22^{3-2}	1.57^{4-2}	2.10
Likelihood ratio test	p-value	–	< 0.0001	0.6390	0.2102	0.34

Table 3. Maximum likelihood estimates and likelihood ratio test results for two models of interviewers' effect and different regression coefficients for fixed effects. Standard error estimates are in parentheses

Variable Names	Parameter Estimates	Model 1 $\sigma_1^2 = \sigma_2^2 = 0$ $\rho = 0$	Model 2 $\sigma_1^2 = \sigma_2^2$ $\rho = 0$
Intercept circles	$\hat{\alpha}_1$	4.4822 (0.9348)	4.5017 (0.9385)
Age	$\hat{\beta}_{11}$	−0.0486 (0.0111)	−0.0487 (0.0111)
Male	$\hat{\beta}_{12}$	−0.0442 (0.1713)	−0.0357 (0.1718)
Years of education	$\hat{\beta}_{13}$	0.1907 (0.0252)	0.1904 (0.0253)
Stroke	$\hat{\beta}_{14}$	−0.4366 (0.2326)	−0.4338 (0.2333)
Intercept pentagon	$\hat{\alpha}_2$	0.9521 (0.5873)	0.9297 (0.5899)
Age	$\hat{\beta}_{21}$	−0.0458 (0.0072)	−0.0454 (0.0072)
Male	$\hat{\beta}_{22}$	0.2709 (0.1022)	0.2806 (0.1027)
Years of education	$\hat{\beta}_{23}$	0.2676 (0.0179)	0.2680 (0.0180)
Stroke	$\hat{\beta}_{24}$	−0.6402 (0.1538)	−0.6489 (0.1545)
	$\hat{\sigma}_1^2$	–	0.0768 (0.0497)
Log likelihood	l	−1841.47	−1817.92
Likelihood ratio statistics (I)	$-2\Delta l$	–	23.55^{2-1}
Likelihood ratio test (I)[a]	p-value	–	<0.0001
Likelihood ratio statistics (II)	$-2\Delta l$	8.01	8.32
Likelihood ratio test (II)[b]	p-value	0.0916	0.0806

[a]Likelihood ratio test (1) compares Model 1 with Model 2 in this table.
[b]Likelihood ratio test (II) compares each model in Table 3 to the respective models in Table 2.

a generalized linear model in the form of variance components with the assumption of mutual independence. The two estimates of interviewer effect for both models assuming common regression coefficients and assuming different regression coefficients are smaller than those from the maximum likelihood estimates in Tables 2 and 3, respectively. Previous simulation studies have shown that variance components estimates by PQL tend to underestimate the true parameters (Breslow and Clayton 1993; Rodriguez and Goldman 1995).

To compare the proposed maximum likelihood approach using Gauss quadrature to the penalized quasi-likelihood approach, simulation studies should be performed where the true parameter values are known. New computer programs need to be developed to handle general variance covariance structure for the PQL approach and for testing of variance covariance parameters. It remains an interesting topic for more research.

6. Discussion

In this article I have proposed the use of a two-stage model for estimating interviewer effects from multivariate binary responses, extending the results of Anderson and Aitkin (1985) from the univariate binary case to multivariate binary responses. Parameter estimation for the model can be achieved by a maximum likelihood approach using Gauss quadrature. Inference regarding interviewer effects can be drawn from the results of the likelihood ratio test. I have illustrated the approach using a community based dementia survey and discussed the computational aspects of the approach.

Table 4. Penalized quasi-likelihood estimates with common and different regression coefficients for fixed effects using the variance-covariance structure in Model 2. Standard error estimates are in parentheses

Variable Names	Parameter Estimates	Common regression coefficient $\sigma_1^2 = \sigma_2^2$ $\rho = 0$	Different regression coefficient $\sigma_1^2 = \sigma_2^2$ $\rho = 0$
Intercept circles	$\hat{\alpha}_1$	3.9692 (0.9705)	3.5783 (1.2718)
Age	$\hat{\beta}_{11}$	−0.0486 (0.0109)	−0.0487 (0.0109)
Male	$\hat{\beta}_{12}$	−0.0606 (0.1706)	−0.0601 (0.1709)
Years of education	$\hat{\beta}_{13}$	0.1916 (0.0246)	0.1916 (0.0249)
Stroke	$\hat{\beta}_{14}$	−0.4424 (0.2296)	−0.4421 (0.2297)
Intercept pentagon	$\hat{\alpha}_2$	−0.3121 (0.2932)	0.0802 (0.6500)
Age	$\hat{\beta}_{21}$	–	−0.0459 (0.0071)
Male	$\hat{\beta}_{22}$	–	0.2442 (0.1028)
Years of education	$\hat{\beta}_{23}$	–	0.2687 (0.0176)
Stroke	$\hat{\beta}_{24}$	–	−0.6360 (0.1524)
	$\hat{\sigma}_1^2$	0.0550 (0.0375)	0.0523 (0.0359)

The two-stage model proposed here can be extended to include more sources of non-sampling errors. However, a maximum likelihood method with Gauss quadrature points may become too complicated to use. As I mentioned in Section 3, other approaches are to be pursued for estimation and inference from the proposed model. For example, the penalized quasi-likelihood method (Breslow and Clayton 1993) can be applied and compared to the maximum likelihood approach taken here. Furthermore, Lin's test (1997) for variance components from generalized linear models may be used for inferences on interviewer effects. However, although the penalized quasi-likelihood method for the variance component model is implemented in some software (Wolfinger 1995), Lin's test is yet to be implemented in standard statistical software.

It is known that maximum likelihood estimates of variance components for linear mixed effect models are biased. Therefore, simulation studies are needed to assess the performance of the maximum likelihood estimator of interviewer effects in terms of bias and efficiency for the two stage logistic normal model used here. Further studies are needed to compare the performance of various approaches in a survey setting for the estimation of interviewer effects.

7. References

Abramowitz, M. and Stegun, I. (1972). Handbooks of Mathematical Functions. New York: Dover.

Aitkin, M. (1999). A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models. Biometrics, 55, 117–128.

Anderson, D. A. and Aitkin, M. (1985). Variance Component Models with Binary Response: Interviewer Variability. Journal of the Royal Statistical Society, B. 47, 203–210.

Biemer, P. P. and Stokes S. L. (1985). Optimal Design of Interviewer Variance Experiment in Complex Surveys. Journal of the American Statistical Association, 80, 158–166.

Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association, 88, 9–25.

Ekholm, A., Smith, P. W. F., and McDonald, J. W. (1995). Marginal Regression Analysis of a Multivariate Binary Response. Biometrika, 82, 847–854.

Fellegi, I. P. (1974). An Improved Method of Estimating the Correlated Response Variance. Journal of the American Statistical Association, 69, 496–501.

Fitzmaurice, G. M. (1995). A Caveat Concerning Independence Estimating Equations with Multivariate Binary Data. Biometrics, 51, 309–317.

Goldstein, H. (1995). Multilevel Statistical Models, 2nd edition. London: Edward Arnold.

Hall, K. S., Ogunniyi, A. O., Hendrie, H. C. et al. (1996). A Cross-Cultural Community Based Study of Dementias: Methods and Performance of the Survey Instrument: Indianapolis, U.S.A. and Ibadan, Nigeria. International Journal of Methods in Psychiatric Research, 6, 129–142.

Hendrie, H. C., Osuntokun, B. O., Hall, K. S. et al. (1995). The Prevalence of Alzheimer's Disease and Dementia in two Communities: Nigerian Africans and African Americans. American Journal of Psychiatry, 3, 1–28.

Kleffe, J., Prasad, N. G. N., and Rao, J. N. K. (1991). 'Optimal' Estimation of Correlated Response Variance under Additive Models. Journal of the American Statistical Association, 86, 144–150.

Liang, K. Y. and Self, S. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. Journal of the American Statistical Association, 82, 605–610.

Lin, X. (1997). Variance Component Testing in Generalised Linear Models with Random Effects. Biometrika, 84, 309–326.

Lipsitz, S. R. and Fitzmaurice, G. (1994). An Extension of Yule's Q to Multivariate Binary data. Biometrics, 50, 847–852.

Longford, N. T. (1993). Random Coefficient Models. Oxford: Clarendon Press.

McDonald, B. W. (1994). Two Random Effect Models for Multivariate Binary Data. Biometrics, 50, 164–172.

Molenberghs, G. and Ritter, L. L. (1996). Methods for Analyzing Multivariate Binary Data, with Association Between Outcomes of Interest. Biometrics, 52, 1121–1133.

Rodriguez, G. and Goldman, N. (1995). An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. Journal of the Royal Statistical Society, A, 158, 73–89.

SAS/OR Technical Report: The NLP Procedure, (1997). SAS Institute Inc., Cary, NC, U.S.A.

Stokes, L. (1988). Estimation of Interviewer Effects of Categorical Items in a Random Digit Dial Telephone Survey. Journal of the American Statistical Association, 83, 623–630.

Waclawiw, M. A. and Liang, K. Y. (1993). Prediction of Random Effects in the Generalized Linear Model. Journal of the American Statistical Association, 88, 171–178.

Wolfinger, R. (1995). The GLIMMIX SAS Macro. SAS Institute Inc., Cary, NL, U.S.A.

Zeger, S. L. and Karim, M. R. (1991). Generalized Linear Models with Random Effects: A Gibbs Sampling Approach. Journal of the American Statistical Association, 86, 79–86.

Received April 1999
Revised September 2000