

# Testing the Correlation for Clustered Categorical and Censored Discrete Time-to-Event Data When Covariates Are Measured without/with Errors

Yi Li

Department of Biostatistics, Harvard School of Public Health and the Dana Farber  
Cancer Institute, Boston, Massachusetts 02115, U.S.A.

*email:* yili@jimmy.harvard.edu

and

Xihong Lin

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

*email:* xlin@umich.edu

**SUMMARY.** In the analysis of clustered categorical data, it is of common interest to test for the correlation within clusters, and the heterogeneity across different clusters. We address this problem by proposing a class of score tests for the null hypothesis that the variance components are zero in random effects models, for clustered nominal and ordinal categorical responses. We extend the results to accommodate clustered censored discrete time-to-event data. We next consider such tests in the situation where covariates are measured with errors. We propose using the SIMEX method to construct the score tests for the null hypothesis that the variance components are zero. Key advantages of the proposed score tests are that they can be easily implemented by fitting standard polytomous regression models and discrete failure time models, and that they are robust in the sense that no assumptions need to be made regarding the distributions of the random effects and the unobserved covariates. The asymptotic properties of the proposed tests are studied. We illustrate these tests by analyzing two data sets and evaluate their performance with simulations.

**KEY WORDS:** Clustered data; Frailty models; Grouped survival data; Longitudinal data; Polytomous responses; Random effects; Score test; SIMEX; Variance components.

## 1. Introduction

Clustered nominal and ordinal categorical responses are observed frequently in biomedical and social sciences studies. For instance, in multicenter cancer clinical trials, several levels of tumor responses are often recorded; in school asthma intervention studies, asthma severity is often classified into multiple categories (e.g., mild, moderate, and severe). Analysis of such data requires accounting for the within-cluster correlation and the multinomial nature of the data. Random effects models are commonly used for analyzing clustered nominal and ordinal categorical data (Harville and Mee, 1984; Hedeker and Gibbons, 1994). In some applications, clustered censored discrete time-to-event data are observed. See Ten Have and Uttal (1994) for an example. These authors proposed frailty discrete proportional hazard models for such data.

In the analysis of such clustered categorical and censored discrete time-to-event data, it is of common interest to test for the correlation within clusters and the heterogeneity among clusters. Such tests have been proposed by using score statistics for the null hypothesis that variance components are zero for clustered continuous, binary, and Poisson outcomes within

the random effects model framework (Commenges et al., 1994; Lin, 1997), and for clustered censored continuous survival data using frailty models (Gray, 1995). However, such tests have not been developed for clustered polytomous data, nor for clustered censored discrete time-to-event data.

In the first part of this article (Sections 2 and 3), we develop tests for the within-cluster correlation for clustered polytomous and censored discrete time-to-event data by deriving score tests for the null hypothesis that variance components are zero in random effects models. Since the null hypothesis is on the boundary of the parameter space—unlike the Wald and likelihood ratio tests whose asymptotic distributions are mixtures of chi-squares—an advantage of the score tests is that their asymptotic distributions are still chi-square. Another advantage of the score tests is that no distribution on the random effects needs to be assumed—except for their first two moments. Hence, they are robust to misspecification of the distributions of the random effects. Further, the Wald tests and the LR tests require fitting random effects models that involve numerical integration; the score tests, however, only require fitting standard models under the null hypothesis

using existing standard software, and do not require numerical integration.

A common problem in the analysis of clustered data is the presence of covariate measurement errors. For example, in nutritional studies, fat intake is often measured with error using a 24-hour food recall (Carroll et al., 1995); in AIDS studies, CD4 counts are often measured with error (Tsiatis, Degruttola, and Wulfsohn, 1995). Valid statistical inference needs to account for measurement errors in covariates. The second part of this paper (Section 4) will extend the proposed score tests for variance components to the situation where covariates are measured with errors. We propose using the SIMEX method (Cook and Stefanski, 1994) to correct for measurement errors and develop SIMEX score tests for variance components. These tests are an extension of the SIMEX score test of Lin and Carroll (1999) to clustered polytomous and discrete time-to-event data with covariate measurement error.

We evaluate the finite sample performance of the proposed tests using simulations, and apply the proposed tests to two data sets: a longitudinal study on nasal drainage (Macknin, Mathew, and Medendorp, 1990) and a Kenya parasitaemia study (McElroy et al., 1997).

## 2. The Variance Component Score Tests for Clustered Categorical Data

Random effects generalized logistic models and cumulative probability models have been proposed to model clustered nominal and ordinal categorical data (Harville and Mee, 1984; Hedeker and Gibbons, 1994). Our main interest in this section is to develop score tests for the null hypothesis that the variance components are zero in such models, to test for the within-cluster correlation.

### 2.1 The Variance Component Score Test in Random Effects Generalized Logistic Models

Suppose that, for the  $j$ th ( $j = 1, \dots, n_i$ ) subject in the  $i$ th ( $i = 1, \dots, m$ ) cluster, a categorical response  $Y_{ij}$  belongs to one of  $N$  categories indexed by  $1, \dots, N$ . Conditional on the cluster-level random effect  $b_i$ , the observations  $Y_{ij}$  are independent, and the conditional probability  $P_{ij,k} = P(Y_{ij} = k | b_i)$  depends on a  $p \times 1$  covariate vector  $X_{ij}$ , through a generalized logistic model

$$\log \left( \frac{P_{ij,k}}{P_{ij,N}} \right) = \alpha_k + X'_{ij}\beta_k + b_i = X'_{ij,k}\beta + b_i, \quad k = 1, \dots, N-1, \quad (1)$$

where  $\beta_k$  is a  $p \times 1$  vector of fixed effects,  $b_i \sim F(b_i; \theta)$  for some distribution function  $F$  that has mean 0 and variance  $\theta$ ,  $X'_{ij,k} = e'_k \otimes (1, X'_{ij})$ , where  $\otimes$  denotes a Kronecker product,  $e_k$  is an  $(N-1) \times 1$  vector with the  $k$ th component equal to 1 and the rest of the components equal to 0, and  $\beta = (\alpha_1, \beta'_1, \dots, \alpha_{N-1}, \beta'_{N-1})'$ .

The marginal log-likelihood function for  $(\beta, \theta)$  is

$$\ell(\beta, \theta) = \sum_{i=1}^m \log \int \exp\{\ell_i(\beta, b_i)\} dF(b_i; \theta), \quad (2)$$

where  $\ell_i(\beta, b_i) = \sum_{j=1}^{n_i} \sum_{k=1}^N y_{ij,k} \log P_{ij,k}$ ,  $y_{ij,k} = I(Y_{ij} = k)$ , and  $I(\cdot)$  is an indicator function. The magnitude of  $\theta$  measures the degree of the within-cluster correlation. We are in-

terested in testing  $H_0: \theta = 0$  vs.  $H_1: \theta > 0$ , where  $H_0: \theta = 0$  implies no within-cluster correlation. Since the null hypothesis is on the boundary of the parameter space, the likelihood ratio test and the Wald test do not follow a chi-square distribution asymptotically (Self and Liang, 1987). We consider a score test for  $H_0$ , and show that it still follows a chi-square distribution asymptotically.

Using L'Hôpital's rule, some calculations show that the score statistic of  $\theta$  evaluated under  $H_0: \theta = 0$  is

$$\begin{aligned} U_\theta(\beta) &= \left. \frac{\partial \ell(\beta, \theta)}{\partial \theta} \right|_{\theta=0} \\ &= \sum_{i=1}^m \frac{1}{2} \left[ \frac{\partial^2 \ell_i(\beta, b_i)}{\partial b_i^2} + \left\{ \frac{\partial \ell_i(\beta, b_i)}{\partial b_i} \right\}^2 \right] \Bigg|_{b_i=0} \\ &= \frac{1}{2} \sum_{i=1}^m \left[ \left\{ \sum_{j=1}^{n_i} (\tilde{Y}_{ij} - \tilde{P}_{ij}) \right\}^2 - \sum_{j=1}^{n_i} \tilde{P}_{ij}(1 - \tilde{P}_{ij}) \right], \end{aligned} \quad (3)$$

where  $\tilde{Y}_{ij} = \sum_{k=1}^{N-1} y_{ij,k} = I(Y_{ij} \leq N-1)$ , and  $\tilde{P}_{ij} = \sum_{k=1}^{N-1} \exp(X'_{ij,k}\beta) / \{1 + \sum_{k=1}^{N-1} \exp(X'_{ij,k}\beta)\}$  is the mean of  $\tilde{Y}_{ij}$  under  $H_0$ . The detailed derivation of (4) is given in Appendix A.1. It is interesting to note that the form of (4) resembles the variance component score statistic for clustered binary data (Commenges et al., 1994). It can easily be shown that, under  $H_0: \theta = 0$ ,  $E\{U_\theta(\beta)\} = 0$  and  $m^{-1/2}U_\theta(\beta)$  is asymptotically normal  $N(0, I_{\theta\theta})$ , where  $I_{\theta\theta}$  is given in (6).

To study the properties of  $U_\theta(\beta)$  under  $H_1: \theta > 0$ , we expand  $E(\tilde{Y}_{ij} | b_i)$  as a quadratic function of  $b_i$ . Some calculations show that, under  $H_1: \theta > 0$ ,

$$E\{U_\theta(\beta)\} \approx \frac{1}{2} \sum_{i=1}^m \left\{ \sum_{j=1}^{n_i} \sum_{k \neq j}^{n_i} a_{ij} a_{ik} + \frac{1}{2} \sum_{j=1}^{n_i} a_{ij} (a'_{ij})^2 \right\} \theta,$$

where  $a_{ij} = \tilde{P}_{ij}(1 - \tilde{P}_{ij})$ , and  $a'_{ij} = 1 - 2\tilde{P}_{ij}$ . It follows that  $E\{U_\theta(\beta)\}$  is an increasing function of  $\theta$ . The test is consistent, and one would expect a large value of  $U_\theta(\beta)$  for a large value of  $\theta$ .

Since  $\beta$  is unknown under  $H_0$  and needs to be estimated, the score statistic for testing  $H_0$  is

$$S = U_\theta(\hat{\beta}) / \tilde{I}_{\theta\theta}^{1/2}(\hat{\beta}), \quad (5)$$

where  $\hat{\beta}$  is the MLE of  $\beta$  under  $H_0$ . It can be easily obtained by fitting the generalized logistic model  $\log(P_{ij,k}/P_{ij,N}) = X'_{ij,k}\beta$ .  $\tilde{I}_{\theta\theta} = I_{\theta\theta} - I_{\theta\beta'} I_{\beta\beta'}^{-1} I_{\beta\theta}$  is the efficient information of  $\theta$  evaluated under  $H_0: \theta = 0$ . Using L'Hôpital's rule, some calculations show that

$$\begin{aligned} I_{\theta\theta} &= E \left\{ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right\} \\ &= \frac{1}{4} \sum_{i=1}^m \left\{ \sum_{j=1}^{n_i} \tilde{P}_{ij} \tilde{Q}_{ij} (1 - 6\tilde{P}_{ij} \tilde{Q}_{ij}) + 2 \left( \sum_{j=1}^{n_i} \tilde{P}_{ij} \tilde{Q}_{ij} \right)^2 \right\}, \end{aligned} \quad (6)$$

$$I_{\beta\beta'} = \sum_{i=1}^m E \left( \frac{\partial \ell_i}{\partial \beta} \frac{\partial \ell_i}{\partial \beta'} \right) = \sum_{i=1}^m X'_i \Sigma_i X_i, \quad (7)$$

$$I_{\theta\beta'} = \sum_{i=1}^m E \left\{ \frac{\partial \ell_i}{\partial \theta} \frac{\partial \ell_i}{\partial \beta'} \right\} = \frac{1}{2} \sum_{i=1}^m P_i'(I_{N-1} \otimes G_i) X_i, \quad (8)$$

where the expectations are taken under  $H_0$ ,  $I_{N-1}$  denotes an  $(N-1) \times (N-1)$  identity matrix, and  $X_i$ ,  $G_i$ ,  $\Sigma_i$ ,  $P_i$  are given in Appendix A.1. Standard asymptotic calculations show that  $S$  is asymptotically  $N(0, 1)$  under  $H_0$ , and one rejects  $H_0$  if  $S$  is large and the test is one-sided.

The score test  $S$  for  $H_0: \theta = 0$  has several attractive features. It can be easily obtained by fitting the generalized logistic model  $\log(P_{ij,k}/P_{ij,N}) = X'_{ij,k}\beta$ , which is model (1) under  $H_0$ , using standard software, e.g., SAS PROC CATMOD. Hence, calculations of  $S$  do not involve any numerical integration. Second, it can be shown that it is the locally most powerful test. Finally, it is robust in the sense that no distribution is assumed for the random effect  $b_i$ .

## 2.2 The Variance Component Score Test in Random Effects Cumulative Probability Models

For clustered ordinal data, a widely used model is the cumulative probability random effects model, by modeling the cumulative probabilities  $r_{ij,k} = P(Y_{ij} \leq k)$  as

$$g(r_{ij,k}) = \alpha_k + X'_{ij}\beta_x + b_i = X'_{ij,k}\beta + b_i, \quad k = 1, \dots, N-1 \quad (9)$$

where  $g(\cdot)$  is a link function,  $X_{ij,k} = (e'_k, X'_{ij})'$ ,  $\beta = (\alpha_1, \dots, \alpha_{N-1}, \beta'_x)'$ , and  $b_i \sim F(\cdot, \theta)$  for some distribution function  $F$  with mean 0 and variance  $\theta$ . For  $g(r) = \text{logit}(r)$  and  $g(r) = \log\{-\log(1-r)\}$ , we have proportional odds and complementary log-log models.

Define  $z_{ij,k} = I(Y_{ij} \leq k)$ . Denote  $r_{ij} = (r_{ij,1}, \dots, r_{ij,N-1})'$ ,  $R_i = (r'_{i1}, \dots, r'_{in_i})'$ , and  $z_{ij}$ ,  $Z_i$  similarly. Some calculations show that the score statistic of  $\theta$  under  $H_0: \theta = 0$  is

$$U_{\theta}(\beta) = \frac{1}{2} \sum_{i=1}^m \left\{ (Z_i - R_i)' \Gamma_i^{-1} H_i 1_i 1_i' H_i \Gamma_i^{-1} (Z_i - R_i) - 1_i' \widetilde{W}_i 1_i \right\}, \quad (10)$$

where  $1_i$  is an  $n_i(N-1) \times 1$  vector of ones, and  $H_i$ ,  $\Gamma_i$  and  $\widetilde{W}_i$  are given in Appendix A.2. Note that (10) compares the empirical variance of the weighted responses to its nominal variance.

The score statistic for testing  $H_0: \theta = 0$  is  $S = U_{\theta}(\widehat{\beta}) / \widetilde{I}_{\theta\theta}^{1/2}(\widehat{\beta})$ , where  $\widehat{\beta}$  is the MLE of  $\beta$  under  $H_0$ , which can be easily obtained by fitting the standard cumulative probability model  $g(r_{ij,k}) = X'_{ij,k}\beta$ , and  $\widetilde{I}_{\theta\theta}(\widehat{\beta})$  is the efficient information of  $\theta$ , defined similarly to that in Section 2.1. The explicit expressions of the information matrices are rather complicated, since their calculations involve the third and fourth cumulants of a multinomial distribution. Specifically, it is shown in Appendix A.2 that, under  $H_0: \theta = 0$ ,

$$I_{\theta\theta} = \frac{1}{4} \sum_{i=1}^m \left\{ \sum_{j=1}^{n_i} \sum_{k,l,m,q=1}^{N-1} \xi_{ij,k} \xi_{ij,l} \xi_{ij,m} \xi_{ij,q} \right. \\ \times \kappa_4(z_{ij,k}, z_{ij,l}, z_{ij,m}, z_{ij,q}) \\ \left. + 2 \sum_{j=1}^{n_i} \sum_{k,l,m=1}^{N-1} \xi_{ij,k} \xi_{ij,l} v_{ij,m} \kappa_3(z_{ij,k}, z_{ij,l}, z_{ij,m}) \right. \\ \left. + V_i' \Gamma_i V_i + 2T_i^2 \right\},$$

$$I_{\beta\beta'} = \sum_{i=1}^m X_i' H_i \Gamma_i^{-1} H_i X_i,$$

$$I_{\theta\beta'} = \frac{1}{2} \sum_{i=1}^m (O_i' \Gamma_i^{-1} H_i X_i + V_i' H_i X_i),$$

where  $\kappa_3$  and  $\kappa_4$  are the third and fourth cumulants of  $z_{ij}$  (see equation (5.13) of McCullagh and Nelder (1989)), and  $\xi_{ij,k}$ ,  $T_i$ ,  $O_i$ , and  $V_i$  are given in Appendix A.2.

Standard asymptotic calculations show that the score statistic  $S$  follows  $N(0, 1)$  asymptotically under  $H_0$ , and has the same optimality and robustness properties stated at the end of Section 2.1. It can be easily calculated by fitting the standard cumulative probability model  $g(r_{ij,k}) = X'_{ij,k}\beta$  using existing software, e.g., SAS PROC CATMOD, and does not require any numerical integration. Again, a one-sided test is used, and  $H_0$  is rejected for a large value of  $S$ .

## 3. The Variance Component Tests for Clustered Censored Discrete Time-to-Event Data

In this section, first we introduce frailty models for clustered censored discrete time-to-event data, which are an extension of the random effects models for clustered categorical data, discussed in Section 2, to accommodate censoring. We then derive a score test for the within-cluster correlation. We assume that survival times are discrete and subject to right censoring, and that censoring is noninformative. Let  $t_{ij} = \min(v_{ij}, d_{ij})$  be the observed time for the  $j$ th subject ( $j = 1, \dots, n_i$ ) in the  $i$ th cluster (e.g., family) ( $i = 1, \dots, m$ ), where  $v_{ij}$  is the true survival time and  $d_{ij}$  is the censoring time. Let  $\delta_{ij} = I(v_{ij} \leq d_{ij})$  be the noncensoring indicator, which takes value 1 if a failure is observed and 0 otherwise. The survival time  $v_{ij}$  and the censoring time  $d_{ij}$  take discrete values at the  $\tau_k$  ( $1 \leq k \leq N$ ), where  $0 \leq \tau_1 < \tau_2 < \dots < \tau_N \leq \infty$ .

Conditional on the cluster-level frailty  $b_i$ , the observations  $(t_{ij}, \delta_{ij})$  are independent, with conditional hazard  $\lambda_{ij,k} = P(v_{ij} = \tau_k | v_{ij} \geq \tau_k, b_i)$  satisfying

$$g(\lambda_{ij,k}) = \alpha_k + X'_{ij}\beta_x + b_i = X'_{ij,k}\beta + b_i, \quad (11)$$

for  $k = 1, \dots, N-1$ , where  $g(\cdot)$  is a link function, and  $b_i \sim F(b_i; \theta)$  for some distribution  $F(\cdot)$  with mean 0 and variance  $\theta$ . For  $g(x) = \text{logit}(x)$ , equation (11) corresponds to a discrete logistic hazard model; for  $g(x) = \log\{\log(1-x)\}$ , it corresponds to a discrete proportional hazard model (Kalbfleisch and Prentice, 1980). In the absence of censoring, these two special cases are reduced to the logistic and complementary log-log models discussed in Section 2.2. Ten Have and Uttal (1994) discussed parameter estimation using the MLE in (11) by assuming  $F(\cdot)$  to be a normal distribution. The log-likelihood function of  $(t_{ij}, \delta_{ij})$  conditional on the frailty  $b_i$  is

$$\ell(t_{ij}, \delta_{ij} | b_i) = \sum_{k=1}^{N-1} \left\{ \delta_{ij} y_{ij,k} \log \left( \frac{\lambda_{ij,k}}{1 - \lambda_{ij,k}} \right) \right. \\ \left. + Y_{ij,k} \log(1 - \lambda_{ij,k}) \right\}, \quad (12)$$

where  $y_{ij,k} = I(t_{ij} = \tau_k)$ , and  $Y_{ij,k} = I(t_{ij} \geq \tau_k)$  is the at-risk indicator.

For the sake of convenience in expressing the score and the information matrix of  $\theta$ , we rewrite model (11) as

$$\log \left( \frac{\lambda_{ij,k}}{1 - \lambda_{ij,k}} \right) = h(X'_{ij,k}\beta + b_i), \tag{13}$$

where there is a one-to-one correspondence between  $g(\cdot)$  and  $h(\cdot)$ . Notice that  $y_{ij,k} = Y_{ij,k} - Y_{ij,k+1}$ , and  $Y_{ij,1} \equiv 1$ . It follows that (12) can be rewritten as

$$\ell(t_{ij}, \delta_{ij} | b_i) = \sum_{k=1}^N \{ \delta_{ij} Y_{ij,k} (h_{ij,k} - h_{ij,k-1}) - Y_{ij,k} s_{ij,k} \},$$

where  $h_{ij,k} = h(X'_{ij,k}\beta + b_i)$  when  $k = 1, \dots, N - 1$ ,  $h_{ij,k} = 0$  when  $k \in \{0, N\}$ ,  $s_{ij,k} = \log \{1 + \exp(h_{ij,k})\}$  when  $k = 1, \dots, N - 1$ , and  $s_{ij,N} = 0$ . The marginal log-likelihood of  $(\beta, \theta)$  can be obtained by integrating out the random effects  $b_i$  as

$$\begin{aligned} \ell(\beta, \theta) &= \sum_{i=1}^m \ell_i(\beta, \theta) \\ &= \sum_{i=1}^m \log \int \exp \left\{ \sum_{j=1}^{n_i} \ell(t_{ij}, \delta_{ij} | b_i) \right\} dF(b_i; \theta). \end{aligned} \tag{14}$$

Using equation (3), some calculations show that the score of  $\theta$  under  $H_0: \theta = 0$  is

$$\begin{aligned} U_\theta(\beta) &= \sum_{i=1}^m \frac{\partial \ell_i}{\partial \theta} = \frac{1}{2} \sum_{i=1}^m \left[ \left\{ \sum_{j=1}^{n_i} \sum_{k=1}^N Y_{ij,k} (\phi_{ij,k} - \omega_{ij,k}) \right\}^2 \right. \\ &\quad \left. - \sum_{j=1}^{n_i} \sum_{k=1}^N Y_{ij,k} (\psi_{ij,k} - \pi_{ij,k}) \right] \end{aligned} \tag{15}$$

where  $h_{ij,k}^{(l)} = h^{(l)}(X'_{ij,k}\beta)$ ,  $\phi_{ij,k} = \delta_{ij} \{h_{ij,k}^{(1)} - h_{ij,k-1}^{(1)}\}$ ,  $\pi_{ij,k} = \delta_{ij} \{h_{ij,k}^{(2)} - h_{ij,k-1}^{(2)}\}$ ,  $\omega_{ij,k} = h_{ij,k}^{(1)} \exp(h_{ij,k}) / \{1 + \exp(h_{ij,k})\}$  when  $k = 1, \dots, N - 1$  and  $\omega_{ij,N} = 0$ , and

$$\psi_{ij,k} = \frac{[h_{ij,k}^{(2)} \{1 + \exp(h_{ij,k})\} + \{h_{ij,k}^{(1)}\}^2] \exp(h_{ij,k})}{\{1 + \exp(h_{ij,k})\}^2},$$

when  $k = 1, \dots, N - 1$  and  $\psi_{ij,N} = 0$ . Here,  $h^{(l)}(\cdot)$  denotes the  $l$ th derivative of  $h(\cdot)$ . It can be shown that  $E \{U_\theta(\beta)\} = 0$  under  $H_0$ . When there is no censoring, i.e.,  $\delta_{ij} \equiv 1$ , noticing  $Y_{ij,k+1} = 1 - z_{ij,k}$  for  $k = 1, \dots, N - 1$ , it can be shown that (15) reduces to (10).

The score statistic for testing  $H_0: \theta = 0$  is  $S = U_\theta(\hat{\beta}) / \tilde{I}_{\theta\theta}^{1/2}(\hat{\beta})$ , and the test should be one-sided. Here,  $\tilde{I}_{\theta\theta}$  is the efficient information of  $\theta$  evaluated under  $H_0$ .  $\hat{\beta}$  is the MLE of  $\beta$  under  $H_0$ , and can be easily calculated by fitting the standard discrete failure time model  $g(\lambda_{ij,k}) = X'_{ij,k}\beta$ . Calculations of the expected information matrices are difficult, because they involve the censoring distribution. Hence, we propose using the observed information calculated under  $H_0$ :

$$\hat{I}_{\theta\theta} = \sum_{i=1}^m \left( \frac{\partial \ell_i}{\partial \theta} \right)^2, \quad \hat{I}_{\theta\beta} = \sum_{i=1}^m \frac{\partial \ell_i}{\partial \theta} \frac{\partial \ell_i}{\partial \beta}, \quad \hat{I}_{\beta\beta} = \sum_{i=1}^m \frac{\partial \ell_i}{\partial \beta} \frac{\partial \ell_i}{\partial \beta},$$

where  $\partial \ell_i / \partial \theta$  is given in (15), and  $\partial \ell_i / \partial \beta = \sum_{j=1}^{n_i} \sum_{k=1}^N Y_{ij,k} (\nu_{ij,k} - \rho_{ij,k})$ . Here  $\nu_{ij,k} = \delta_{ij} \{h_{ij,k}^{(1)} X_{ij,k} - h_{ij,k-1}^{(1)} X_{ij,k-1}\}$ , and  $\rho_{ij,k} = h_{ij,k}^{(1)} \exp(h_{ij,k}) X_{ij,k} / \{1 + \exp(h_{ij,k})\}$  when  $1 \leq k \leq N - 1$  and  $\rho_{ij,N} = 0$ .

Standard asymptotic calculations show that  $S$  follows  $N(0, 1)$  asymptotically under  $H_0$ . The score statistic  $S$  has the same attractive optimality and robustness properties as those laid out at the end of Section 2.1. It can be easily calculated by fitting standard discrete failure time models  $g(\lambda_{ij,k}) = X'_{ij,k}\beta$ , using, e.g., SAS PHREG, and requires no numerical integration.

#### 4. The Variance Component Tests in the Presence of Measurement Errors in Covariates

In this section, we extend the variance component score tests developed in Sections 2 and 3 to the situation when covariates are measured with error. Denote by  $X_{ij}$  a vector of unobserved covariates (e.g., the true CD4 count), and by  $C_{ij}$  other accurately measured covariates. For simplicity, we concentrate on random effects cumulative probability models (Section 2.2), and frailty discrete failure time models (Section 3). The results in this section apply directly to random effects generalized logistic models (Section 2.1).

The random effects cumulative probability model (9) and the frailty discrete failure time model (11) can be written in a unified form

$$g(p_{ij,k}) = \alpha_k + X'_{ij}\beta_x + C'_{ij}\beta_c + b_i, \tag{16}$$

where  $b_i$  follows some distribution  $F(\cdot, \theta)$ , with mean 0 and variance  $\theta$ . For the random effects cumulative probability model (9),  $p_{ij,k} = r_{ij,k}$ . For the frailty discrete failure time model (11),  $p_{ij,k} = \lambda_{ij,k}$ .

Suppose the observed covariates  $W_{ij}$  (e.g., the observed CD4 count) measure  $X_{ij}$  with error. We postulate a nondifferential additive measurement error model for  $W_{ij}$  (Carroll et al., 1995),

$$W_{ij} = X_{ij} + U_{ij}, \tag{17}$$

where the  $U_{ij}$  are independent measurement errors following  $N(0, \Sigma_u)$ . Suppose that the measurement error covariance  $\Sigma_u$  is known, or is estimated as  $\hat{\Sigma}_u$ , e.g., using replicates or validation data. We are interested in testing for no within-cluster correlation  $H_0: \theta = 0$  in the random effects measurement error models (16) and (17). We propose using the SIMEX method, by extending the results in Sections 2 and 3 to construct score tests for  $H_0$  to account for measurement errors.

Simulation extrapolation (SIMEX) is a simulation-based functional method for inference on model parameters in measurement error problems (Cook and Stefanski, 1994), where no distributional assumption is made about the unobserved covariates  $X_{ij}$ . We first briefly describe parameter estimation in random effects measurement error models (16)–(17) using

the SIMEX method, then discuss how to use the SIMEX idea to develop SIMEX score tests for  $H_0: \theta = 0$ .

The SIMEX method consists of two steps: the simulation step and the extrapolation step. In the simulation step, one generates data  $W_{ij}^*$  by adding to  $W_{ij}$  a random error following  $N(0, \eta \Sigma_u)$ , for some constant  $\eta > 0$ . One then calculates naive parameter estimates by fitting (16) with  $X_{ij}$  replaced by  $W_{ij}^*$ . This would give the naive estimates if the measurement error covariance is  $(1 + \eta)\Sigma_u$ . This procedure is repeated for a large number  $B$  times (e.g.,  $B = 100$ ), and the mean of the resulting  $B$  naive parameter estimates is calculated. One does this for a series of values of  $\eta$  (e.g.,  $\eta = 0.5, 1, 1.5, 2$ ). In the extrapolation step, a regression (e.g., quadratic) model is fit to the means of these naive estimates as a function of  $\eta$ , and is extrapolated to  $\eta = -1$ , which corresponds to zero measurement error variance. These extrapolated estimates give the SIMEX estimates of the model parameters. For details of the SIMEX method, see Cook and Stefanski (1994) and Carroll et al. (1995).

We now apply the SIMEX idea to constructing score tests for  $H_0: \theta = 0$  in the random effects measurement error models (16) and (17), by extending the results in Sections 2 and 3. The proposed SIMEX score tests are an extension of the work of Lin and Carroll (1999) to random effects measurement error models for clustered polytomous and censored discrete failure time data.

When there is no measurement error, the score statistics for testing  $H_0: \theta = 0$  under (16) take the same form  $U_\theta(\hat{\beta})/\tilde{I}_{\theta\theta}^{1/2}(\hat{\beta})$ , where  $U_\theta(\hat{\beta})$  is given in (10) for random effects cumulative probability models, and in (15) for frailty discrete time-to-event models. The denominator  $\tilde{I}_{\theta\theta}(\hat{\beta})$  is in fact the variance of  $U_\theta(\hat{\beta})$ . The main idea of the SIMEX variance component score test is to treat the score statistic in the numerator  $U_\theta(\cdot)$  as if it were a parameter estimator, and then use the SIMEX variance method (Section 4.3.5 of Carroll et al., 1995) to calculate the variance of this ‘‘estimator.’’ Specifically, in the SIMEX simulation step, one simply calculates naive score statistics using the score formula (10) and (15), by replacing  $X_{ij}$  with the simulated data  $W_{ij}^*$ . The rest of the steps parallel those in the standard SIMEX method for parameter estimation. Denoting the results by  $U_{simex}(\cdot)$  and  $\tilde{I}_{\theta\theta, simex}$ , respectively, the SIMEX score statistic is simply

$$S_{simex} = U_{simex} / \tilde{I}_{\theta\theta, simex}^{1/2}, \quad (18)$$

which follows  $N(0, 1)$  asymptotically when the true extrapolation function is used. Since in practice the true extrapolation function is unknown, an approximation (e.g., quadratic) is used. Our simulation study in Section 5 shows that the proposed SIMEX score tests perform well. The theoretical justification of the SIMEX score tests is similar to that in Lin and Carroll (1999), and is omitted.

The proposed SIMEX score test has several key features. First, it can be easily calculated by fitting standard cumulative probability models, or discrete failure time models, using available software, such as SAS PROC CATMOD and SAS PROC PHREG. Second, it is robust, in the sense that no distribution needs to be assumed for the frailty  $b_i$ , nor for the unobserved covariates  $X$ .

## 5. Simulation Studies

### 5.1 The Random Effects Generalized Logistic and Cumulative Probability Models

We performed simulation studies to evaluate the sizes and the powers of the variance component score tests proposed in Sections 2.1 and 2.2. We assumed the number of clusters  $m = 100$  and the cluster size  $n_i = 5$ . We consider a trichotomous response variable ( $N = 3$ ) and two covariates  $X_{ij} = (X_{ij1}, X_{ij2})^T$ , where  $X_{ij1}$  is 0 for half of the clusters and 1 for the rest, and mimics a treatment indicator, and  $X_{ij2} = j - 3$  ( $j = 1, \dots, 5$ ). The true parameter values are  $\alpha_1 = \alpha_2 = -1$ ,  $\beta_1 = (0.5, 1)$ , and  $\beta_2 = (2, 1)$  for model (1); and  $\alpha_1 = 0.5$ ,  $\alpha_2 = 1$ , and  $\beta_x = (1, 2)$  for model (9).

Following Lin (1997), we generated  $b_i$  from a normal mixture

$$F = \pi N\{-(1 - \pi)\mu, \sigma^2\} + (1 - \pi)N(\pi\mu, \sigma^2), \quad (19)$$

which has mean 0 and variance  $\theta = \pi(1 - \pi)\mu^2 + \sigma^2$ . Three cases for  $F$  were considered: 1) normal:  $\mu = 0$ ,  $\sigma^2 = 1$ ; 2) unimodal normal mixture:  $\pi = 0.25$ ,  $\mu = 0.5$ ,  $\sigma^2 = \theta - \pi(1 - \pi)\mu^2$ ; and 3) bimodal normal mixture:  $\pi = 0.50$ ,  $\mu = 1$ ,  $\sigma^2 = \theta - \pi(1 - \pi)\mu^2$ . In each simulation, we generated 500 trichotomous observations from model (1) or (9). We varied  $\theta$  from 0 to 1.0, to study the sizes and the powers of the score tests for the variance components. The experiment was replicated 2000 times for each parameter configuration. The nominal size was set to be 0.05. The empirical sizes and powers of the tests under the two models are given in Table 1.

The results in Table 1 show that the empirical sizes of the tests are very close to the nominal level of 0.05. As the variance component increases, the powers quickly approach 1. Greater powers are associated with the distribution deviating further from normality. We speculate that higher power is obtained because the curvature of the score function might be larger when the distribution of the random effect deviates from normality.

### 5.2 The Frailty Discrete Failure Time Model

We next conducted simulation studies to assess the sizes and powers of the variance component score tests in frailty discrete time-to-event models. We considered frailty discrete logistic and proportional hazard time-to-event models, i.e.,  $g(x) = \text{logit}(x)$  and  $g(x) = \log\{-\log(1 - x)\}$  in (11). We assumed the number of distinct failure times  $N = 5$ . The covariate  $X_{ij}$  in (11) was assumed to be a scalar, and was simulated

**Table 1**

*Empirical sizes and powers of the variance component score test in the random effects generalized logistic and proportional odds models observed in 2000 simulations*

Model	Distribution of $b_i$	$\theta$						
		0.0	0.3	0.5	0.6	0.8	0.9	1.0
Logistic	Normal	0.052	0.10	0.20	0.33	0.64	0.77	0.87
	Unimodal normal mixture	0.052	0.13	0.23	0.35	0.67	0.80	0.90
	Bimodal normal mixture	0.052	0.15	0.25	0.37	0.68	0.82	0.92
Prop. odds	Normal	0.046	0.14	0.48	0.70	0.81	0.88	0.96
	Unimodal normal mixture	0.046	0.16	0.51	0.73	0.82	0.92	0.97
	Bimodal normal mixture	0.046	0.17	0.54	0.75	0.85	0.94	0.98

**Table 2**

*Empirical sizes and powers of the variance component score test in the frailty discrete logistic and proportional hazard models observed in 2000 simulations*

Model	Dist. of $b_i$	# of clusters	$\theta$						
			0.0	0.1	0.3	0.5	0.6	0.8	1.0
Logistic	Normal	$m = 40$	0.047	0.08	0.12	0.21	0.27	0.35	0.68
		$m = 80$	0.058	0.09	0.14	0.31	0.42	0.52	0.81
		$m = 160$	0.054	0.10	0.24	0.64	0.81	0.90	1.00
	Bimodal	$m = 40$	0.053	0.09	0.17	0.28	0.37	0.46	0.75
		$m = 80$	0.058	0.10	0.21	0.45	0.58	0.69	0.90
		$m = 160$	0.054	0.13	0.32	0.73	0.86	0.94	1.00
PH	Normal	$m = 40$	0.054	0.06	0.13	0.37	0.45	0.57	0.83
		$m = 80$	0.056	0.07	0.16	0.49	0.63	0.74	0.95
		$m = 160$	0.052	0.15	0.52	0.94	0.95	1.00	1.00
	Bimodal	$m = 40$	0.055	0.10	0.20	0.45	0.56	0.81	0.92
		$m = 80$	0.056	0.14	0.36	0.74	0.83	0.90	0.99
		$m = 160$	0.052	0.23	0.63	0.97	0.99	1.00	1.00

independently from  $N(0, 1)$ ,  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (-0.2, -0.3, -0.4, -0.5)$ , and  $\beta_x = 0.5$ . We generated the frailty  $b_i$  from (19), and considered  $F$  to be normal (case (1)) and bimodal normal mixture (case (3)). We varied  $\theta$  from 0 to 1 to study the sizes and powers of the score tests.

In our experiments, we considered the following combinations: the number of clusters:  $m = 40, 80, 160$ ; the cluster size:  $n = 3$ ; and the censoring proportion = 20%. For each setting, 2000 simulated data sets were generated. The nominal size was set to be 0.05. Table 2 displays the empirical sizes and powers of the variance component score tests under the two models.

The proposed score tests perform well. The actual sizes of the tests are close to 0.05, and their performance improves

as the number of clusters increases from 40 to 160. As the variance component increases, the powers of the tests quickly approach 1. When the number of clusters is smaller, e.g., 40, the power is lower. Similar results are obtained when we allow the number of clusters  $n_i$  to vary across clusters, ranging from 1 to 5, with an average of 2 observations per cluster. This setting mimics the Kenya data.

### 5.3 The Frailty Measurement Error Discrete Failure Time Model

We next performed simulation studies to assess the sizes and powers of the SIMEX variance component score tests in frailty discrete failure time measurement error models (16) and (17). We considered both logistic and proportional hazard models. The parameter configurations were the same as in the frailty discrete failure time models with no measurement error, in Section 5.2, except that we assumed  $X_{ij}$  was not observed, and was measured with error by  $W_{ij} = X_{ij} + U_{ij}$ , where  $(U_{ij}, X_{ij})$  were generated independently from  $N(0, 0.5)$  and  $N(0, 1)$ , respectively.

The results of the empirical sizes and powers, based on 2000 simulations assuming  $b_i$  to be normal or normal mixture, are displayed in Table 3. For the purpose of comparison, we studied the performance of the naive score tests, which ignore the measurement error by replacing  $X_{ij}$  with  $W_{ij}$ , and the performance of the SIMEX score tests. A quadratic extrapolation function was used in the SIMEX extrapolation step. Our simulation results show that the levels of the naive score tests, calculated by ignoring the measurement error, are too high—and almost double the nominal value when the number of clusters is large. In contrast, the SIMEX score tests perform well, and their levels are very close to the nominal value. As

**Table 3**

*Empirical sizes and powers of naive and SIMEX score tests for variance component observed in 2000 simulations in the frailty discrete logistic and proportional hazard (PH) measurement error model*

Model	Dist. of $b_i$	# of clusters	Method	$\theta$						
				0.0	0.1	0.3	0.5	0.6	0.8	1.0
Logistic	Normal	$m = 40$	Naive	0.063	0.11	0.16	0.28	0.33	0.55	0.82
			SIMEX	0.052	0.09	0.14	0.24	0.29	0.51	0.77
		$m = 80$	Naive	0.073	0.12	0.20	0.43	0.52	0.64	0.91
			SIMEX	0.054	0.10	0.18	0.38	0.47	0.57	0.87
		$m = 160$	Naive	0.103	0.15	0.26	0.71	0.85	0.93	1.00
			SIMEX	0.053	0.14	0.24	0.65	0.83	0.91	1.00
	Bimodal	$m = 40$	Naive	0.071	0.13	0.18	0.29	0.36	0.58	0.87
			SIMEX	0.055	0.11	0.15	0.26	0.33	0.53	0.81
		$m = 80$	Naive	0.087	0.14	0.23	0.44	0.57	0.66	0.91
			SIMEX	0.058	0.12	0.21	0.43	0.54	0.64	0.87
		$m = 160$	Naive	0.098	0.16	0.37	0.71	0.87	0.94	1.00
			SIMEX	0.057	0.14	0.34	0.67	0.83	0.92	1.00
PH	Normal	$m = 40$	Naive	0.061	0.08	0.17	0.37	0.54	0.65	0.92
			SIMEX	0.052	0.07	0.15	0.33	0.51	0.61	0.85
		$m = 80$	Naive	0.065	0.09	0.21	0.52	0.65	0.77	0.97
			SIMEX	0.053	0.07	0.18	0.45	0.62	0.72	0.93
		$m = 160$	Naive	0.102	0.16	0.50	0.94	0.96	1.00	1.00
			SIMEX	0.053	0.13	0.46	0.93	0.94	0.98	1.00
	Bimodal	$m = 40$	Naive	0.082	0.16	0.24	0.58	0.68	0.79	0.90
			SIMEX	0.056	0.11	0.21	0.53	0.62	0.73	0.87
		$m = 80$	Naive	0.095	0.17	0.37	0.72	0.82	0.90	0.99
			SIMEX	0.054	0.11	0.33	0.69	0.79	0.86	0.97
		$m = 160$	Naive	0.093	0.29	0.60	0.95	0.97	1.00	1.00
			SIMEX	0.052	0.22	0.59	0.93	0.94	1.00	1.00

the variance component  $\theta$  increases, the powers of the SIMEX score tests increase and quickly approach 1. One may also note that the powers of the SIMEX test are lower than are the naive tests; however, this could be due to the inflated size of the naive tests. As a referee pointed out, this could also be a consequence of the fact that SIMEX shrinks the variance component estimate, while increasing its variance. Similar results are obtained when we vary the measurement error variance  $\sigma_u^2$ , e.g., from 1/9 to 1.

## 6. Applications

### 6.1 Application 1: The Nasal Drainage Study

To illustrate the variance component score tests for clustered polytomous data discussed in Section 2, we examined data from a longitudinal study on the efficacy of steam inhalation for treating common cold symptoms conducted by Macknin et al. (1990). This study involved 30 patients who had colds of recent onset. For each patient, two 20-minute steam inhalation treatments, spaced 60–90 minutes apart, were administered at the time of enrollment. Assessment of subjective response was made on an individual daily score card by the patient from day 1 (baseline) to day 4. On each day, the severity of nasal drainage was calibrated into 4 ordered categories (no symptoms, mild, moderate, and severe). One was interested in examining whether the severity improved following the treatment, and in testing whether the observations over time, for each subject, were likely to be correlated.

We considered models (1) and (9), with the time from the baseline as a covariate. When the random effects logistic model (1) was assumed, the variance component score statistic was 5.32 ( $p$ -value  $< 0.001$ ), which provided strong evidence for within-subject correlation over time. Similar results were found when the random effect proportional odds model (9) was fit (score statistic = 9.70,  $p$ -value  $< 0.001$ ). Note that these two tests assumed no distribution for the random effect  $b_i$ .

To examine the time effect, we fit (9) by further assuming that the random effect  $b_i$  followed  $N(0, \theta)$ . The MLE of the coefficient of time was  $-0.33$  (SE = 0.21). This suggests that the severity improved following the treatment, but the improvement was not statistically significant ( $p$ -value = 0.11). The estimated variance component was 2.31 (SE = 0.45). This result was consistent with the score test result.

### 6.2 Application 2: The Kenya Parasitaemia Study

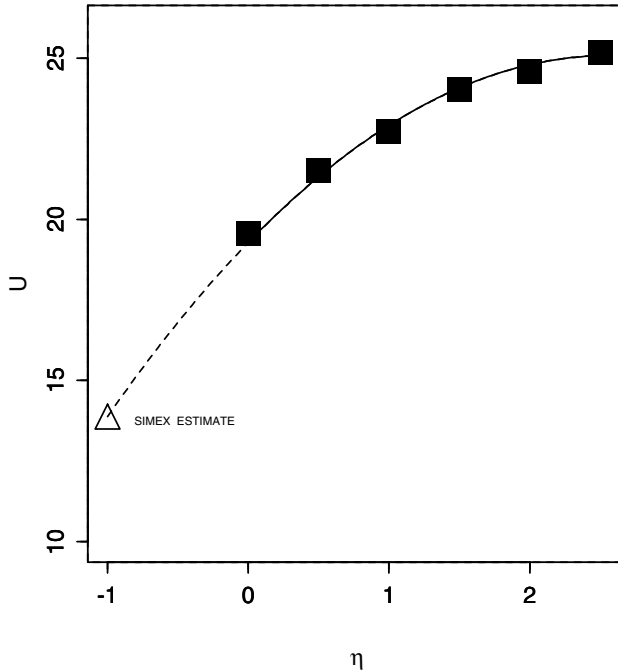
To illustrate the SIMEX score test for within-cluster correlation in the presence of measurement error in covariate discussed in Section 4, we considered a study for the effect of the infective mosquito bite exposure on the risk of parasitaemia among children in Western Kenya (McElroy et al., 1997). A total of 542 children, aged from six months to six years, from 309 households, were enrolled into the study between February 1986 and July 1987. Upon entry into the study, each child, regardless of his/her parasitaemia status, was treated to eliminate blood-stage infection. Their blood films were examined two weeks later to check whether or not their parasitaemia infections were eliminated. Twenty-three children, whose observation times were less than two weeks, were excluded from the analysis. The blood films of the rest of the children were found to be negative. These children were then followed for up to 10 weeks for the first recurrence of parasitaemia. The ob-

servations from different children within the same household were likely to be correlated.

The investigators were interested in studying the effects of cumulative infective mosquito bites in the first two-week period on the risk of recurrent parasitaemia. It was also of interest to test whether the observations from different children within the same household were likely to be correlated. The number of infective mosquito bites was measured only one night in each of the first two weeks. The cumulative infective bites in the first two weeks were calculated using the sum of the two night measures times 7. Hence, they were measured with substantial error. The other covariates included sex (1 = F, 0 = M), age, and baseline parasitaemia density. About 90% of the children experienced recurrent parasitaemia during the follow-up. The average number of total two-week bites was 12. As children were scheduled to be visited every other week after accrual, the observed event times were categorized into five categories: 1 (2 weeks), 2 (3–4 weeks), 3 (5–6 weeks), 4 (7–8 weeks), and 5 (8+weeks). There were 86 failures and 4 censored subjects at 2 weeks, 149 failures and 3 censored subjects between 3 and 4 weeks, 119 failures and 3 censored subjects between 5 and 6 weeks, 73 failures and 3 censored subjects between 7 and 8 weeks, and 33 failures and 46 censored subjects after 8 weeks. The baseline parasitaemia density was log transformed (LNBPDP), and the cumulative two-week bites were quartic-root transformed (BITE).

We considered the frailty discrete logistic, and proportional hazard measurement error, models (16) and (17) with  $X$  = true BITE and  $\mathbf{C}$  = (AGE, GENDER, LNBPDP). Since there were no validation data available, and there were no replicates, the measurement error variance  $\sigma_u^2$  cannot be estimated from the available data. We fit a linear random intercept model to the observed BITE ( $W$ ). This only allowed us to estimate the sum of  $\sigma_x^2$  and  $\sigma_u^2$ , which was 0.20. We then conducted a sensitivity analysis, by varying  $\sigma_u^2$  from 0 (naive analysis), to moderate ( $\sigma_u^2 = 0.08$ ), to severe ( $\sigma_u^2 = 0.20$ ) measurement error.

To test for the within-household correlation, we calculated the SIMEX score test for the variance component equal to 0 under the frailty discrete logistic and proportional hazard measurement error models. Under the frailty logistic measurement error model, the naive score statistic for  $H_0: \theta = 0$  calculated by ignoring the measurement error ( $\sigma_u^2 = 0$ ) was 0.93 ( $p$ -value = 0.18). The SIMEX score statistics, by assuming that  $\sigma_u^2 = 0.08$  and 0.20, were 0.63 ( $p$ -value = 0.26) and 0.34 ( $p$ -value = 0.37), respectively. Under the frailty proportional hazard measurement error model, the naive score statistic was 1.02 ( $p$ -value = 0.15), while the SIMEX score statistics, by assuming  $\sigma_u^2 = 0.08$  and 0.20, were 0.66 ( $p$ -value = 0.25) and 0.43 ( $p$ -value = 0.33), respectively. These results suggest that the within-household correlation is not significant. Note that no distributional assumptions about the frailty and the unobserved true BITE are made in the SIMEX score tests. A quadratic extrapolation function is used in the extrapolation step. To examine whether this is appropriate, we plot in Figure 1 the naive scores in the simulation step against  $\eta$  in squares, and then superimpose the quadratic extrapolation function. Figure 1 shows that the quadratic extrapolation function works well. Our simulation study also supports this finding.



**Figure 1.** SIMEX extrapolations for the score  $U$  when  $\sigma_u^2 = 0.08$  in the Kenya malaria study.

To evaluate the effect of the cumulative infective mosquito bites in the first two-week period on the risk of recurrent parasitaemia, we fit models (16) and (17) using the MLE, by further assuming the frailty  $b_i$  was normally distributed as  $N(0, \theta)$ . The results are presented in Table 4. Our analyses show that a higher level of exposure to infective mosquito bites significantly increases the risk of parasitaemia. Ignoring the measurement error attenuates the regression coefficient estimates and inflates the variance component estimates.

**7. Discussion**

We have proposed in this paper testing the within-cluster correlation, using the score tests for the variance components in random effects generalized logistic and cumulative probability models for clustered polytomous data. The results are ex-

tended to clustered censored discrete failure time data. A key feature of the proposed score tests is that they can be easily calculated by fitting standard generalized logistic and cumulative probability models or standard discrete failure time models using existing software, such as SAS PROC CATMOD and PROC PHREG. Further, they are robust in the sense that no distribution for the random effects needs to be assumed. Our simulation studies show that the proposed score tests perform well in terms of both sizes and powers.

In this paper, we have also considered testing the within-cluster correlation in the presence of covariate measurement error, using the SIMEX score tests. The SIMEX score tests enjoy all the properties of the variance component score tests in the no measurement error case. Furthermore, no distribution for the unobserved error-prone covariate is assumed. Our simulation studies show that the proposed SIMEX score tests perform well in terms of sizes and powers.

Although our score tests are robust to misspecification of the distributions of the random effects and the distribution of the unobserved error-prone covariates, they might be less powerful compared to fully parametric tests (e.g., Wald or likelihood ratio tests) calculated by assuming parametric distributions for the random effects and for the unobserved error-prone covariates. Unlike the score tests, these parametric tests do not have closed-form expressions and require numerical integration. Further research is needed to compare the robustness and the powers of these two classes of tests.

As pointed out by a referee, even when the null hypothesis of homogeneity is not rejected, the data may lend support to varying degrees of clustering. Hence, when the parametric assumption of the random effects is reasonable and there is some evidence of clustering, instead of fully relying on the variance component score test, it is a good practice to fit the random effects models to estimate the variance component—which will provide information about the magnitude of clustering.

RÉSUMÉ

Dans l’analyse des données catégorielles en grappe, on cherche souvent à tester la corrélation intra-grappe, et l’hétérogénéité entre grappes. Nous proposons une classe de « score tests »

**Table 4**

*Parameter estimates (SEs in parentheses) for the Kenya parasitaemia data using frailty discrete logistic and proportional hazard (PH) measurement error models. Naive means assuming  $\sigma_u^2 = 0.0$ . Moderate error means assuming the measurement error variance  $\sigma_u^2 = 0.08$ , while severe error means assuming  $\sigma_u^2 = 0.20$ .*

Parameter	Naive		Moderate Error		Severe Error	
	Logistic	PH	Logistic	PH	Logistic	PH
INTERCEPT1	-2.59 (0.31)	-2.45 (0.25)	-2.67 (0.32)	-2.50 (0.26)	-2.89 (0.33)	-2.65 (0.27)
INTERCEPT2	-1.52 (0.29)	-1.55 (0.24)	-1.61 (0.30)	-1.60 (0.24)	-1.81 (0.31)	-1.73 (0.25)
INTERCEPT3	-1.09 (0.29)	-1.21 (0.23)	-1.19 (0.30)	-1.27 (0.24)	-1.39 (0.31)	-1.40 (0.25)
INTERCEPT4	-0.86 (0.31)	-1.04 (0.25)	-0.96 (0.32)	-1.10 (0.26)	-1.16 (0.33)	-1.24 (0.26)
INTERCEPT5	-1.05 (0.36)	-1.16 (0.28)	-1.17 (0.37)	-1.24 (0.29)	-1.39 (0.38)	-1.40 (0.30)
BITE	0.47 (0.14)	0.34 (0.11)	0.57 (0.17)	0.41 (0.14)	0.88 (0.21)	0.63 (0.17)
AGE	0.05 (0.04)	0.04 (0.03)	0.06 (0.04)	0.05 (0.03)	0.07 (0.04)	0.05 (0.03)
GENDER	0.09 (0.12)	0.07 (0.10)	0.09 (0.12)	0.08 (0.10)	0.10 (0.13)	0.08 (0.10)
LNBPD	0.12 (0.06)	0.09 (0.05)	0.13 (0.06)	0.10 (0.05)	0.13 (0.06)	0.10 (0.05)
$\theta$	0.12 (0.11)	0.09 (0.07)	0.09 (0.13)	0.07 (0.09)	0.05 (0.17)	0.04 (0.16)



dont l'hypothèse nulle est la nullité des composantes de la variance dans des modèles à effet aléatoire, pour des données catégorielles nominales ou ordinales en grappe. Nous généralisons les résultats pour prendre en compte des données de survie censurées à pas de temps discrets. Nous considérons ensuite le cas de covariables avec erreur de mesure. Nous proposons l'utilisation de la méthode du SIMEX pour calculer le score test de l'hypothèse nulle de nullité des composantes de la variance. Les avantages principaux des tests proposés sont leur facilité d'implémentation, par ajustement de modèles de régression polytomique standard, et leur robustesse, puisque aucune hypothèse sur la distribution des effets aléatoires ou des covariables non observées n'est nécessaire. Les propriétés asymptotiques des tests proposés sont étudiées. Nous appliquons ces tests à deux jeux de données, et évaluons leur performance par simulations.

## REFERENCES

- Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.
- Commenges, D., Letenneur, L., Jacqmin, H., Moreau, J., and Dartigues, J. (1994). Test of homogeneity of binary data with explanatory variables. *Biometrics* **50**, 613–620.
- Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89**, 1314–1328.
- Gray, R. (1995). Tests for variation over groups in survival data. *Journal of the American Statistical Association* **90**, 198–203.
- Harville, D. and Mee, R. (1984). A mixed-model procedure for analyzing ordered categorical data. *Biometrics* **40**, 393–408.
- Hedeker, D. and Gibbons, R. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–945.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Lin, X. (1997). Variance component testing in generalized linear models with random effects. *Biometrika* **84**, 309–26.
- Lin, X. and Carroll, R. J. (1999). SIMEX variance component tests in generalized linear mixed measurement error models. *Biometrics* **55**, 613–619.
- Macknin, M. L., Mathew, S., Medendorp, S. V. (1990). Effect of inhaling heated vapor on symptoms of the common cold. *Journal of the American Medical Association* **264**, 989–991.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. New York: Chapman & Hall.
- McElroy, P. D., Beier, J. C., Oster, C. N., Onyango, F. K., Oloo, A. J., Lin, X., Beedle, C., and Hoffman, S. L. (1997). Dose- and time-dependent relations between infective anopheles inoculation and outcomes of plasmodium falciparum parasitaemia among children in western Kenya. *American Journal of Epidemiology* **145**, 945–956.
- Self, S. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests un-

der nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

- Ten Have, T. R. and Uttal, D. H. (1994). Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival data. *Applied Statistics* **43**, 371–384.
- Tsiatis, A. A., Degruetola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27–37.

Received July 2001. Revised August 2002.

Accepted August 2002.

## APPENDIX

## A.1 Derivation of the Variance Component Score Test in the Random Intercept Generalized Logistic Model

Since

$$\begin{aligned} \ell_i(\beta, b_i) &= \sum_{j=1}^{n_i} \sum_{k=1}^N y_{ij,k} \log P_{ij,k} \\ &= \sum_{j=1}^{n_i} \left[ \sum_{k=1}^{N-1} y_{ij,k} (X'_{ij,k} \beta + b_i) \right. \\ &\quad \left. - \log \left\{ 1 + \sum_{k=1}^{N-1} \exp(X'_{ij,k} \beta + b_i) \right\} \right], \end{aligned}$$

it follows that

$$\begin{aligned} \frac{\partial \ell_i(\beta, b_i)}{\partial b_i} \Big|_{b_i=0} &= \sum_{j=1}^{n_i} \left\{ \sum_{k=1}^{N-1} y_{ij,k} \right. \\ &\quad \left. - \frac{\sum_{k=1}^{N-1} \exp(X'_{ij,k} \beta)}{1 + \sum_{k=1}^{N-1} \exp(X'_{ij,k} \beta)} \right\} \\ &= \sum_{j=1}^{n_i} (\tilde{Y}_{ij} - \tilde{P}_{ij}) \\ \frac{\partial^2 \ell_i(\beta, b_i)}{\partial b_i^2} \Big|_{b_i=0} &= \sum_{j=1}^{n_i} - \frac{\sum_{k=1}^{N-1} \exp(X'_{ij,k} \beta)}{\left\{ 1 + \sum_{k=1}^{N-1} \exp(X'_{ij,k} \beta) \right\}^2} \\ &= \sum_{j=1}^{n_i} -\tilde{P}_{ij} (1 - \tilde{P}_{ij}). \end{aligned}$$

Equation (4) follows immediately.

We next calculate the information matrices  $I_{\beta\beta'}$ ,  $I_{\theta\beta'}$ ,  $I_{\theta\theta}$  under  $H_0: \theta = 0$ . First,

$$U_\beta = \left. \frac{\partial \ell(\beta, \theta)}{\partial \beta} \right|_{\theta=0} = \sum_{i=1}^m \sum_{j=1}^{n_i} \sum_{k=1}^{N-1} (y_{ij,k} - P_{ij,k}) X_{ij,k}.$$

We next define  $X_i = (X'_{i1}, \dots, X'_{in_i})'$ , where  $X_{ij} = (X_{ij,1}, \dots, X_{ij,N-1})'$ ,  $\tilde{Q}_{ij} = 1 - \tilde{P}_{ij}$ , and  $\Sigma_i = \{\Sigma_{i,r,l}\}$ , which is an  $(N-1) \times (N-1)$  block matrix whose  $(r, l)$ th block is

$$\Sigma_{i,rr} = \text{diag}\{P_{i1,r}(1 - P_{i1,r}), \dots, P_{in_i,r}(1 - P_{in_i,r})\},$$

$$\Sigma_{i,rl} = \text{diag}(-P_{i1,r}P_{i1,l}, \dots, -P_{in_i,r}P_{in_i,l}), r \neq l.$$

Let  $G_i = \text{diag}(2\tilde{P}_{ij}^2 - 3\tilde{P}_{ij} + 1, \dots, 2\tilde{P}_{in_i}^2 - 3\tilde{P}_{in_i} + 1)$  and  $P_i = (P'_{i,1}, \dots, P'_{i,N-1})'$ , where  $P_{i,r} = (P_{ij,r}, \dots, P_{in_i,r})'$ . Then under  $H_0: \theta = 0$ , we have

$$I_{\beta\beta'} = E(U_\beta U'_\beta) = \sum_{i=1}^m \sum_{j=1}^{n_i} \left\{ \sum_{k=1}^{N-1} P_{ij,k}(1 - P_{ij,k}) X_{ij,k} X'_{ij,k} - \sum_{l=1}^{N-1} \sum_{l \neq m} P_{ij,l} P_{ij,m} X_{ij,l} X'_{ij,m} \right\},$$

$$\begin{aligned} I_{\theta\beta'} &= E\{U_\theta U'_\beta\} \\ &= \frac{1}{2} \sum_{i=1}^m E \left( \left[ \left\{ \sum_{j=1}^{n_i} (\tilde{Y}_{ij} - \tilde{P}_{ij}) \right\}^2 - \sum_{j=1}^{n_i} \tilde{P}_{ij} (1 - \tilde{P}_{ij}) \right] \right. \\ &\quad \left. \times \left\{ \sum_{j=1}^{n_i} \sum_{k=1}^{N-1} (y_{ij,k} - P_{ij,k}) X'_{ij,k} \right\} \right) \\ &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^{n_i} (2\tilde{P}_{ij}^2 - 3\tilde{P}_{ij} + 1) \sum_{k=1}^{N-1} P_{ij,k} X'_{ij,k}, \end{aligned}$$

which can be written in the matrix form given in (7) and (8). The term  $I_{\theta\theta}$  can be derived in a similar way and its derivation is omitted.

*A.2 Derivation of the Score Tests in the Random Intercept Cumulative Probability Models*

Using the results in McCullagh and Nelder (1989), we have  $\partial \ell_{ij}(\beta, b_i) / \partial r_{ij} = \Gamma_{ij}^{-1}(z_{ij} - r_{ij})$  and  $\partial^2 \ell_{ij}(\beta, b_i) / \partial r_{ij} \partial r'_{ij} = \Omega_{ij} - \Gamma_{ij}^{-1}$ , where the  $k$ th column of  $\Omega_{ij}$  is  $\partial \Gamma_{ij}^{-1} / \partial r_{ij,k}(z_{ij} - r_{ij})$ , and  $\Gamma_{ij} = \text{cov}(z_{ij})$ . Using equation (3), the score of  $\theta$  under  $H_0: \theta = 0$  is

$$\begin{aligned} U_\theta(\beta) &= \frac{1}{2} \sum_{i=1}^m \left[ \sum_{j=1}^{n_i} \left\{ \frac{\partial^2 r'_{ij}}{\partial b_i^2} \Gamma_{ij}^{-1}(z_{ij} - r_{ij}) \right. \right. \\ &\quad \left. \left. + \sum_{k=1}^{N-1} \frac{\partial r_{ij,k}}{\partial b_i} \frac{\partial r'_{ij}}{\partial b_i} \frac{\partial \Gamma_{ij}^{-1}}{\partial r_{ij,k}}(z_{ij} - r_{ij}) \right. \right. \\ &\quad \left. \left. - \frac{\partial r'_{ij}}{\partial b_i} \Gamma_{ij}^{-1} \frac{\partial r_{ij}}{\partial b_i} \right\} \right. \\ &\quad \left. + \left\{ \sum_{j=1}^{n_i} \frac{\partial r'_{ij}}{\partial b_i} \Gamma_{ij}^{-1}(z_{ij} - r_{ij}) \right\}^2 \right] \Bigg|_{b_i=0}. \end{aligned}$$

Let  $h(\cdot) = g^{-1}(\cdot)$ ,  $h^{(j)}(\cdot)$  denote the  $j$ th derivative of  $h(\cdot)$ , and

$$H_i = \text{diag} \left\{ h^{(1)}(X'_{i1,1}\beta), \dots, h^{(1)}(X'_{i1,N-1}\beta), \dots, h^{(1)}(X'_{in_i,1}\beta), \dots, h^{(1)}(X'_{in_i,N-1}\beta) \right\},$$

$$h_i^{(2)} = \left\{ h^{(2)}(X'_{i1,1}\beta), \dots, h^{(2)}(X'_{i1,N-1}\beta), \dots, h^{(2)}(X'_{in_i,1}\beta), \dots, h^{(2)}(X'_{in_i,N-1}\beta) \right\}' ,$$

and  $1_i$  be an  $n_i(N-1) \times 1$  vector with ones. Let  $\Gamma_i = \text{diag}\{\Gamma_{i1}, \dots, \Gamma_{in_i}\}$  and

$$V_i = \Gamma_i^{-1} h_i^{(2)} + \text{diag} \left\{ - \sum_{k=1}^{N-1} h^{(1)}(X'_{i1,t}\beta) \Gamma_{i1}^{-1} \frac{\partial \Gamma_{i1}}{\partial r_{i1,t}} \Gamma_{i1}^{-1}, \dots, - \sum_{k=1}^{N-1} h^{(1)}(X'_{in_i,t}\beta) \Gamma_{in_i}^{-1} \frac{\partial \Gamma_{in_i}}{\partial r_{in_i,t}} \Gamma_{in_i}^{-1} \right\} H_i 1_i.$$

Then  $U_\theta(\beta)$  can be written in a simple matrix form

$$U_\theta(\beta) = \frac{1}{2} \sum_{i=1}^m \left\{ (Z_i - R_i)' \Gamma_i^{-1} H_i 1_i 1_i' H_i \Gamma_i^{-1} (Z_i - R_i) - 1_i' \tilde{W}_i 1_i \right\},$$

where  $\tilde{W}_i = H_i \Gamma_i^{-1} H_i - \text{diag}(V_i) \text{diag}(Z_i - R_i)$ .

We next calculate the information matrices. Let

$$M_i = \Gamma_i^{-1} H_i 1_i 1_i' H_i \Gamma_i^{-1}, \quad T_i = 1_i' H_i \Gamma_i^{-1} H_i 1_i,$$

$$\xi_i = \Gamma_i^{-1} H_i 1_i = (\xi'_{i1}, \dots, \xi'_{in_i})',$$

where  $\xi'_{ij} = (\xi_{ij,1}, \dots, \xi_{ij,N-1}) = \{h^{(1)}(X'_{ij1}\beta), \dots, h^{(1)}(X'_{ij,N-1}\beta)\} \Gamma_{ij}^{-1}$ . Under  $H_0: \theta = 0$ , we have

$$\begin{aligned} I_{\theta\theta} &= E\{U_\theta^2(\beta)\} = \frac{1}{4} \sum_{i=1}^m \left[ E\{\xi'_i(Z_i - R_i)\}^4 - T_i^2 \right. \\ &\quad \left. + E\{V'_i(Z_i - R_i)\}^2 \right. \\ &\quad \left. + 2E\{\xi'_i(Z_i - R_i) \times \xi'_i(Z_i - R_i) \times V'_i(Z_i - R_i)\} \right] \\ &= \frac{1}{4} \sum_{i=1}^m \left\{ \sum_{j=1}^{n_i} \sum_{t,l,m,q} \xi_{ij,k} \xi_{ij,t} \xi_{ij,m} \xi_{ij,q} \kappa_4 \right. \\ &\quad \left. \times (z_{ij,k}, z_{ij,l}, z_{ij,m}, z_{ij,q}) \right. \\ &\quad \left. + 2 \sum_{j=1}^{n_i} \sum_{t,l,m} \xi_{ij,k} \xi_{ij,l} v_{ij,m} \kappa_3 \right. \\ &\quad \left. \times (z_{ij,k}, z_{ij,l}, z_{ij,m}) \right. \\ &\quad \left. + V'_i \Gamma_i V_i + 2T_i^2 \right\}. \end{aligned}$$

It can be easily shown that under  $H_0$ ,  $U_\beta(\beta) = \sum_{i=1}^m X_i H_i \Gamma_i^{-1} (Z_i - R_i)$ . It follows that

$$\begin{aligned} I_{\beta\beta'} &= E\{U_\theta(\beta)U'_\beta(\beta)\} = \sum_{i=1}^m X_i' H_i \Gamma_i^{-1} H_i X_i, \\ I_{\theta\beta'} &= E\{U_\theta(\theta)U'_\beta(\beta)\} \\ &= \frac{1}{2} \sum_{i=1}^m E\{(Z_i - R_i)' M_i (Z_i - R_i) \\ &\quad \times (Z_i - R_i)' \Gamma_i^{-1} H_i X_i\} \\ &\quad + \frac{1}{2} \sum_{i=1}^m E\{V_i'(Z_i - R_i)(Z_i - R_i)' \Gamma_i^{-1} H_i X_i\} \quad (\text{A.1}) \end{aligned}$$

The second term in (A.1) is  $\frac{1}{2} V_i' H_i X_i$ . For the first term, calculate

$$\begin{aligned} o_{ij,k} &= E\{(Z_i - R_i)' M_i (Z_i - R_i)(z_{ij,k} - r_{ij,k})\} \\ &= E\left[\{\xi'_{ij}(z_{ij} - r_{ij})\}^2 e'_t(z_{ij} - r_{ij})\right] \\ &= \sum_{u,l} \xi_{ij,u} \xi_{ij,l} \kappa_3(z_{ij,u}, z_{ij,l}, z_{ij,k}). \end{aligned}$$

Let  $O_i = (o_{i1,1}, \dots, o_{i1,N-1}, \dots, o_{in_i,1}, \dots, o_{in_i,N-1})'$ ; then the first term in (A.1) is  $\frac{1}{2} \sum_{i=1}^m O_i' \Gamma_i^{-1} H_i X_i$ .

