

## A likelihood approach to estimating sensitivity and specificity for binocular data: Application in ophthalmology

A. R. de Leon<sup>1,\*</sup>, M. Guo<sup>1</sup>, C. J. Rudnisky<sup>2</sup> and G. Singh<sup>3</sup>

<sup>1</sup>*Department of Mathematics and Statistics, University of Calgary, Alta., Canada T2N 1N4*

<sup>2</sup>*Royal Alexandra Hospital and Department of Ophthalmology, University of Alberta, Alta., Canada*

<sup>3</sup>*Leong & Associates Actuaries and Consultants, Inc., Vancouver, BC, Canada V6E 4A6*

### SUMMARY

Binocular data typically arise in ophthalmology where pairs of eyes are evaluated, through some diagnostic procedure, for the presence of certain diseases or pathologies. Treating eyes as independent and adopting the usual approach in estimating the sensitivity and specificity of a diagnostic test ignores the correlation between eyes. This may consequently yield incorrect estimates, especially of the standard errors. The paper proposes a likelihood-based method of accounting for the correlations between eyes and estimating sensitivity and specificity using a model for binocular or paired binary outcomes. Estimation of model parameters *via* maximum likelihood is outlined and approximate tests are provided. The efficiency of the estimates is assessed in a simulation study. An extension of the methodology to the case of several diagnostic tests, or the same test measured on several occasions, which arises in multi-reader studies, is given. A further extension to the case of multiple diseases is outlined as well. Data from a study on diabetic retinopathy are analysed to illustrate the methodology. Copyright © 2007 John Wiley & Sons, Ltd.

**KEY WORDS:** correlated binary data; coverage probability; generalized linear mixed model; maximum likelihood estimation; random effects model

### 1. INTRODUCTION

This paper is motivated by a study conducted in Alberta, Canada, on the use of high-resolution stereoscopic digital photography in a teleophthalmology application, to evaluate diabetic patients

\*Correspondence to: A. R. de Leon, Department of Mathematics and Statistics, University of Calgary, Calgary, Alta., Canada T2N 1N4.

†E-mail: adeleon@math.ucalgary.ca

Contract/grant sponsor: Natural Sciences and Engineering Research Council of Canada; contract/grant number: 76-1060

Contract/grant sponsor: University of Calgary Research Grants Committee; contract/grant number: 78-1045

for treatable diabetic retinopathy [1]. The study performed evaluations at a distance which is akin to screening. However, unlike screening, which identifies a large group of patients who need further assessment, evaluation is more efficient in that it identifies only those patients requiring treatment. The popularity of evaluation and screening programmes has led to improved early intervention for and treatment of diseases resulting in substantial health-care cost savings [2]. In countries like Canada, where distances are great, the cost of travel necessitates evaluation at a distance, thereby allowing only those patients in need of treatment to travel to a specialist. A teleophthalmology system allowing for distance evaluation of retinopathy-related pathologies based on digital images of diabetic patients' eyes is a potentially cost-effective alternative to clinical examination. The purpose of the study was thus to determine whether diabetic retinopathy can be identified with high-resolution stereoscopic digital photography and whether this identification correlates well with the accepted gold standard of clinical examination.

The accuracy of a medical test for diagnosing the presence or absence of a disease can be described by its sensitivity and specificity with respect to a traditionally used and accepted test regarded as a 'gold standard.' Sensitivity is the probability that the new test indicates presence of the disease when the gold standard indicates that it is present while specificity is the probability that the new test indicates absence of the disease when the gold standard indicates that it is absent. Denoting by  $Y$  and  $D$ , the respective binary variables representing test result and disease status as determined by the gold standard, the test's sensitivity and specificity are then given by  $P(Y = 1|D = 1)$  and  $P(Y = 0|D = 0)$ , respectively (see, for example, Reference [3]).

It is commonplace in diagnostic studies to have patients undergo several diagnostic tests or be subjected to the same test on repeated occasions. While test results from different patients are still independent, those from the same patient are now correlated. A statistical problem facing researchers involved in such studies concerns the proper accounting in the analysis, of the correlation among measurements taken from the same patient. Treating the eyes as independent in the example described previously, and adopting the usual crude approach in estimating the sensitivity and specificity of the diagnostic test ignores the correlation between the eyes, and may consequently yield incorrect estimates, especially of the standard errors [4, 5]. Previous work that addressed this problem on the estimation of sensitivity and specificity and their standard errors in the context of clustered binary diagnostic data includes simple adjustments to standard errors [6–8], a generalized estimating equations (GEE) approach [9], and a likelihood approach [10] based on a commonly used correlated binomial model [11]. Ahn [12] reported that for moderate to large samples and moderate intra-cluster correlation, the GEE estimator outperforms those proposed by Lee and Dubin [6], Donner and Klar [7], and Rao and Scott [8].

In this paper, we develop a likelihood approach based on Prentice's [13] model for the joint distribution of binocular and paired binary data to estimate the sensitivity and specificity of binocular binary diagnostic tests. We describe the model and the consequent maximum likelihood estimation and inference in Section 2. Strengths as well as limitations of the model are highlighted and its performance relative to the crude independence approach, which ignores the intra-pair correlation and the GEE, is investigated. Section 3 outlines an extension of the model to multi-reader diagnostic studies. Data from a diabetic retinopathy study [1] are used to illustrate the methodology in Section 4. Finally, a summary of the paper as well as a further extension of the methodology to multi-reader multi-disease diagnostic studies is presented in Section 5.

## 2. A MODEL FOR BINOCULAR DIAGNOSTIC DATA

If  $\rho$  is the correlation between the left and right eyes of a patient, a model for the joint distribution of the binocular binary outcome  $(Y_L, Y_R)^\top$  is given by

$$P(Y_L = y_L, Y_R = y_R) = \left[ 1 + \frac{\rho \prod_{j=L,R} (y_j - p_j)}{\sqrt{p_L q_L p_R q_R}} \right] \prod_{j=L,R} p_j^{y_j} q_j^{1-y_j} \quad (1)$$

where  $q_j = 1 - p_j$  [13]. Note that model (1) is completely specified by the marginal probabilities  $p_j = P(Y_j = 1)$ ,  $j = L, R$ , and the inter-eye correlation  $\rho$ . To ensure that (1) is a proper joint probability distribution, the inter-eye correlation  $\rho$  needs to satisfy the following restriction:

$$\max \left\{ -\sqrt{\frac{p_L p_R}{q_L q_R}}, -\sqrt{\frac{q_L q_R}{p_L p_R}} \right\} < \rho < \min \left\{ \sqrt{\frac{p_R q_L}{p_L q_R}}, \sqrt{\frac{p_L q_R}{p_R q_L}} \right\} \quad (2)$$

From (1), we can see that when  $\rho = 0$ , the model reduces to the independent Bernoulli model, where  $Y_L$  and  $Y_R$  are assumed independent and each follows a Bernoulli distribution with success probability  $p_j$ ,  $j = L, R$ . Thus, (1) generalizes the independent Bernoulli model to the case of binocular binary outcomes.

For eye  $j = L, R$ , we define a marginal model for  $p_j$  with disease status  $D_j$  and covariates  $\mathbf{x}_j$  as  $p_j = h^{-1}[g(\beta_0, \beta_1, \boldsymbol{\beta}; D_j, \mathbf{x}_j)]$ , with  $h(\cdot)$  defined as some link function, usually taken as the logistic or probit link, and  $g(\cdot; D_j, \mathbf{x}_j)$  is some function specifying the manner by which  $D_j$  and  $\mathbf{x}_j$  are incorporated in the link function. For example, if  $g(\beta_0, \beta_1, \boldsymbol{\beta}; D_j, \mathbf{x}_j) = \beta_0 + \beta_1 D_j + \boldsymbol{\beta} \mathbf{x}_j$ , then we have a generalized linear model for which the eye-specific sensitivity  $\pi_{1j} = P(Y_j = 1 | D_j = 1, \mathbf{x}_j)$  and specificity  $\pi_{0j} = 1 - P(Y_j = 1 | D_j = 0, \mathbf{x}_j)$  of the test for eye  $j$  are given by

$$\pi_{1j} = h^{-1}(\beta_0 + \beta_1 + \boldsymbol{\beta}^\top \mathbf{x}_j) \quad (3)$$

$$\pi_{0j} = 1 - h^{-1}(\beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_j) \quad (4)$$

In many applications, the covariate  $\mathbf{x}_j$  is usually measured at the patient level, so that  $\mathbf{x}_L = \mathbf{x}_R$ . This implies that  $\pi_{1L} = \pi_{1R} = \pi_1$  and  $\pi_{0L} = \pi_{0R} = \pi_0$ . That is, the subject-specific sensitivity and specificity of the diagnostic test is independent of the particular eye under consideration. We assume that this is the case in what follows.

### 2.1. Likelihood estimation

Suppose  $N$  patients undergo diagnostic testing on both left and right eyes for some pathology. Let  $\{y_{iL}, y_{iR}, D_{iL}, D_{iR}, \mathbf{x}_i\}$ ,  $i = 1, \dots, N$ , denote the observed data. Assuming  $\rho_i = \rho$ , the likelihood function is given by

$$L(\boldsymbol{\eta}) = \prod_{i=1}^N \left[ 1 + \frac{\rho \prod_{j=L,R} (y_{ij} - p_{ij})}{\sqrt{p_{iL} q_{iL} p_{iR} q_{iR}}} \right] \prod_{j=L,R} p_{ij}^{y_{ij}} q_{ij}^{1-y_{ij}} \quad (5)$$

where  $\boldsymbol{\eta}^\top = (\beta_0, \beta_1, \boldsymbol{\beta}^\top, \rho)$ ,  $p_{ij} = h_{ij}^{-1} = h^{-1}[g(\beta_0, \beta_1, \boldsymbol{\beta}; D_{ij}, \mathbf{x}_{ij})]$ ,  $j = L, R, i = 1, \dots, N$ . The log-likelihood function is then

$$\ell(\boldsymbol{\eta}) = \sum_{r,r'=0,1} \sum_{m,m'=0,1} \sum_{i=1}^{n_{rr'}^{mm'}} \ell_{rr'}^{mm'}(\boldsymbol{\eta}|\mathbf{x}_i) \tag{6}$$

where  $n_{rr'}^{mm'}$  is the number of observations for which  $Y_L = r, Y_R = r', D_L = m$ , and  $D_R = m'$ , with  $\ell_{rr'}^{mm'}$  their corresponding log-likelihood contribution. The log-likelihood function (6) simplifies to

$$\ell(\beta_0, \beta_1, \rho) = \sum_{r,r'=0,1} \sum_{m,m'=0,1} n_{rr'}^{mm'} \ell_{rr'}^{mm'}(\beta_0, \beta_1, \rho) \tag{7}$$

in the absence of covariates. Table I displays the cross-classification of the binocular data according to the results of the test and the actual disease status for the left and right eyes. Putting  $\dot{\ell}(\boldsymbol{\eta}) = \partial\ell/\partial\boldsymbol{\eta}^\top$  as the score function and  $\ddot{\ell}(\boldsymbol{\eta}) = \partial^2\ell/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top$  as the Hessian matrix, the maximum likelihood estimate (MLE)  $\widehat{\boldsymbol{\eta}}^\top = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\beta}}^\top, \widehat{\rho})$  is obtained by solving the likelihood equations  $\dot{\ell}(\boldsymbol{\eta}) = \mathbf{0}^\top$  iteratively via a Newton–Raphson updating scheme. With  $g(\beta_0, \beta_1; D) = \beta_0 + \beta_1 D$ , the respective MLEs of  $\pi_1$  and  $\pi_0$  are then  $\widehat{\pi}_1 = h^{-1}(\widehat{\beta}_0 + \widehat{\beta}_1)$  and  $\widehat{\pi}_0 = 1 - h^{-1}(\widehat{\beta}_0)$ , with large-sample standard errors obtained via the delta method. It can be easily verified that  $\widehat{\boldsymbol{\eta}}$  is consistent and asymptotically multi-variate normal with mean  $\boldsymbol{\eta}$  and covariance matrix given by the inverse of the Fisher information matrix  $\mathbf{I}(\boldsymbol{\eta}) = E\boldsymbol{\eta}[-\ddot{\ell}(\boldsymbol{\eta})]$ . Standard asymptotic methods for testing hypotheses concerning  $\boldsymbol{\eta}$  readily apply.

A strength of model (1) is that the joint distribution of the binocular data is completely determined by the marginal probabilities and the intra-pair correlation. In addition, model (1) has convenient marginal and conditional distributions. The approach of completely specifying the joint distribution of  $Y_L$  and  $Y_R$  leads to straightforward likelihood estimation upon specification of parametric forms for the marginal probabilities. This affords us a whole battery of likelihood-based procedures for model inference, model checking, and validation. Moreover, a fully specified likelihood function for the binocular diagnostic data can be easily adapted to non-random sampling schemes like case–control data, a common occurrence in clinical and epidemiological studies in ophthalmology. Recent extensions of GEE have likewise been proposed to handle non-randomly sampled data [14].

However, model (1) has some drawbacks that may render it unsuitable for other applications. The requirement that probabilities be non-negative places constraints on the range of the

Table I. Assessment of presence (+) or absence (–) of a pathology in the left (L) and right (R) eyes of  $N$  patients.

Disease status	Test				Total
	L + R+	L – R+	L + R–	L – R–	
L + R+	$n_{11}^{11}$	$n_{01}^{11}$	$n_{11}^{10}$	$n_{00}^{11}$	$n^{11}$
L – R+	$n_{11}^{01}$	$n_{01}^{01}$	$n_{01}^{10}$	$n_{00}^{01}$	$n^{01}$
L + R–	$n_{11}^{10}$	$n_{01}^{10}$	$n_{10}^{10}$	$n_{00}^{10}$	$n^{10}$
L – R–	$n_{11}^{00}$	$n_{01}^{00}$	$n_{00}^{10}$	$n_{00}^{00}$	$n^{00}$
Total	$n_{11}$	$n_{01}$	$n_{10}$	$n_{00}$	$N$

intra-pair correlation  $\rho$ . Allowing the marginal probabilities to depend on patient-level covariates may severely restrict  $\rho$  into admitting mostly positive values. A possible remedy to this is to allow the intra-pair correlation to depend on disease status  $D_L$  and  $D_R$  of the eyes by modelling the covariance  $\text{cov}(Y_L, Y_R)$  through, say, an exponential link function.

While model (1) can be readily extended to the general clustered binary data setting [13], the resulting expression leads to a number of issues in estimation. Aside from stringent constraints on correlation parameters, likelihood estimation of regression coefficients becomes computationally infeasible. However, a generalization of model (1) to binocular data from multi-test diagnostic studies is still possible. By introducing random effects in (3), we can account for inter-test and inter-disease correlations among the clustered binocular data. This is adopted in Sections 3 and 4. Moreover, as outlined in Section 5, model (1) can be further generalized to multi-reader multi-disease studies, where disease-specific sensitivity and specificity estimates are desired, providing an alternative to the GEE approach.

In summary, model (1) is most appropriate for analysing binocular binary data, like those that arise in ophthalmology. It provides a useful alternative to the commonly used method which assumes independence of the eyes. This latter approach can be very inefficient in applications, as we show in the next section.

## 2.2. Simulation study

To examine the performance in finite samples of the MLEs of  $\pi_0$  and  $\pi_1$  based on model (1) and assess the accuracy in finite samples of the estimates and their standard errors described in Section 2.1, we carried out a series of simulations using binocular diagnostic data generated from model (1) with logistic link and  $g(\beta_0, \beta_1; D) = \beta_0 + \beta_1 D$ . The performance of the estimates was evaluated using the empirical coverage rates of 95 per cent confidence intervals that jointly combine the estimates and associated standard errors. We compare these with those obtained from the crude method and the GEE approach.

The results of three such simulations illustrate the performance of the estimates for different sample sizes. Tables II–IV are based on 2000 simulated data sets with sample sizes  $N = 100$  and 200 from model (1). Table II corresponds to the case with  $\rho = 0.1$  and  $\pi_0 = \pi_1 = 0.9$ . Table III corresponds to the case with  $\rho = 0.4$  and  $\pi_0 = \pi_1 = 0.7$ , while Table IV to  $\rho = 0.8$ , and  $\pi_0 = \pi_1 = 0.55$ . In each case, the disease status  $D_j$  ( $j = L, R$ ) was generated from model (1) with  $P(D_j = 1) = 0.5$  and correlation  $\rho$ . The means of the 2000 sets of estimates and standard errors were calculated. In addition, the empirical standard deviations of the estimates were computed and the empirical coverage rates, defined as the proportion of 95 per cent confidence intervals enclosing the true values of the sensitivity and specificity, were determined. The same was done for the GEE estimates (based on an independent ‘working’ correlation) and those from the crude method based on  $2N = 200, 400$  observations. The standard errors for the GEE estimates were calculated using the robust ‘sandwich’ variance estimator. The crude method estimates were obtained from an ordinary logistic regression on  $2N$  observations, incorrectly assumed to be independent.

Based on the central limit theorem, we constructed approximate 95 per cent confidence intervals ( $cr_e \pm 1.96[cr_e(1 - cr_e)/2000]^{1/2}$ ) to check if the nominal 95 per cent coverage rate is attained by the estimates, where  $cr_e$  is the empirical coverage rate.

These simulations suggest that the MLEs based on model (1) perform well in finite samples. We find little bias in the estimates and their standard errors were able to capture the true variability of the estimates. The empirical coverage rates reveal that, given large enough samples, confidence

Table II. Coverage probabilities of ML, crude method, and GEE estimates based on 2000 simulated data sets from model (1) with  $\rho = 0.1, \pi_0 = \pi_1 = 0.9$ .

		Estimate (mean)	SE (mean)	SD (empirical)	Coverage probability
<i>N</i> = 100					
Model (1)	$\pi_1$	0.9023	0.0292	0.0289	0.9365*
	$\pi_0$	0.9030	0.0292	0.0289	0.9355*
GEE approach	$\pi_1$	0.9000	0.0302	0.0315	0.9265
	$\pi_0$	0.9006	0.0301	0.0306	0.9220
Crude method	$\pi_1$	0.8999	0.0277	0.0297	0.9275
	$\pi_0$	0.9007	0.0277	0.0291	0.9210
<i>N</i> = 200					
Model (1)	$\pi_1$	0.9024	0.0205	0.0208	0.9435*
	$\pi_0$	0.9021	0.0205	0.0208	0.9405*
GEE approach	$\pi_1$	0.9000	0.0216	0.0222	0.9430*
	$\pi_0$	0.8993	0.0216	0.0222	0.9373*
Crude method	$\pi_1$	0.8999	0.0196	0.0210	0.9430*
	$\pi_0$	0.8997	0.0196	0.0209	0.9371*

\*Not significantly different from nominal 95 per cent coverage.

Table III. Coverage probabilities of ML, crude method, and GEE estimates based on 2000 simulated data sets from model (1) with  $\rho = 0.4, \pi_0 = \pi_1 = 0.7$ .

		Estimate (mean)	SE (mean)	SD (empirical)	Coverage probability
<i>N</i> = 100					
Model (1)	$\pi_1$	0.7113	0.0473	0.0482	0.9380*
	$\pi_0$	0.7111	0.0473	0.0482	0.9375*
GEE approach	$\pi_1$	0.6984	0.0499	0.0521	0.9275
	$\pi_0$	0.6989	0.0498	0.0526	0.9290
Crude method	$\pi_1$	0.6996	0.0424	0.0494	0.9125
	$\pi_0$	0.6996	0.0423	0.0499	0.9030
<i>N</i> = 200					
Model (1)	$\pi_1$	0.7100	0.0320	0.0335	0.9390*
	$\pi_0$	0.7116	0.0320	0.0335	0.9440*
GEE approach	$\pi_1$	0.6987	0.0354	0.0371	0.9275
	$\pi_0$	0.7007	0.0353	0.0368	0.9385*
Crude method	$\pi_1$	0.6992	0.0288	0.0345	0.9035
	$\pi_0$	0.7009	0.0287	0.0337	0.9155

\*Not significantly different from nominal 95 per cent coverage.

Table IV. Coverage probability of ML, crude method, and GEE estimates based on 2000 simulated data sets from model (1) with  $\rho = 0.8$ ,  $\pi_0 = \pi_1 = 0.55$ .

		Estimate (mean)	SE (mean)	SD (empirical)	Coverage probability
<i>N</i> = 100					
Model (1)	$\pi_1$	0.5628	0.0565	0.0549	0.9375*
	$\pi_0$	0.5644	0.0565	0.0549	0.9475*
GEE approach	$\pi_1$	0.5524	0.0584	0.0615	0.9160
	$\pi_0$	0.5521	0.0584	0.0611	0.9245
Crude method	$\pi_1$	0.5517	0.0443	0.0591	0.8505
	$\pi_0$	0.5530	0.0443	0.0571	0.8605
<i>N</i> = 200					
Model (1)	$\pi_1$	0.5616	0.0429	0.0437	0.9400*
	$\pi_0$	0.5603	0.0429	0.0437	0.9375*
GEE approach	$\pi_1$	0.5491	0.0415	0.0462	0.9170
	$\pi_0$	0.5479	0.0414	0.0465	0.9135
Crude method	$\pi_1$	0.5496	0.0340	0.0450	0.8580
	$\pi_0$	0.5485	0.0338	0.0452	0.8580

\*Not significantly different from nominal 95 per cent coverage.

intervals based on estimates from model (1) are able to attain the nominal 95 per cent level. While those based on GEE estimates are relatively close to the nominal level, almost all of the cases considered yielded rates that are significantly below 95 per cent. Not surprisingly, the crude method performed the worst, failing to attain the nominal 95 per cent level in all the cases. Note that model (1) is not any less robust than GEE as it, like GEE, is completely specified by the marginal probabilities and the inter-eye correlation. However, the simulations suggest that MLEs based on model (1) possess better finite-sample properties than their GEE counterparts, in terms of standard error estimates and coverage probabilities.

Note that the disparities between crude method coverage rates and the nominal 95 per cent level are generally higher for moderate to high intra-pair correlation  $\rho$  while no such deterioration is apparent in the coverage rates based on model (1) MLEs and GEE estimates. This is to be expected as model (1) and the GEE approach take account of the correlation between the eyes. Observe as well that when  $\rho$  becomes large, coverage rates based on the crude method and the GEE approach tend to deflate, drastically in the case of the former and slightly in the latter, indicating that associated standard errors are underestimated.

The implication of ignoring the inter-eye correlation is clear: failure to adjust for this correlation in any statistical analysis may lead to potentially incorrect inferences.

### 3. EXTENSION TO MULTI-READER DIAGNOSTIC STUDIES

Many diagnostic studies involve either subjecting patients to a number of tests or to a single test on several occasions. This is the situation when, for example, the diagnosis depends on the subjective

assessment of a so-called reader, in which case, the study protocol requires that at least two readers diagnose a patient to avoid reader bias. Another scenario when this occurs involves subjecting patients to a battery of tests, as is done in screening programmes.

Consider the diabetic retinopathy study described in Section 1, where left and right eyes of patients are evaluated by several readers for a number of retinopathy-related pathologies based on the same images. In addition to the inter-eye correlation induced by the binocular nature of the data, another source of correlation is present in this case. Because readers rely on the same image of the eye, their diagnoses are potentially correlated. One approach to estimating the sensitivity and specificity of a diagnostic procedure in this case is to ignore the inter-reader correlation and simply average the results from separate analyses of data from each of the readers *via* model (1). Another alternative is to adopt a GEE approach to incorporate inter-eye and inter-reader correlations.

In this section, we extend model (1) to the multi-reader setting by including a random effect to account for the other source of correlation. Diagnoses by several readers of the same digital image of a patient’s eye could well be correlated due to the similarity of reader diagnoses caused by certain characteristics, besides disease status, inherent to the subject [15]. This correlation can be explained by the addition to model (1) of an unobserved random variable which varies from patient to patient.

Let  $T \sim N(0, 1)$  be a latent standard normal variable independent of disease status and which varies across patients. Assuming no other covariates are available, the diagnosis  $Y_{jk}$  for eye  $j = L, R$ , by reader  $k = 1, \dots, K$ , is then allowed to depend on the disease status  $D_j$  and the latent variable  $T$  through a generalized linear model  $p_{jk}(t) = P(Y_{jk} = 1 | D_j, t) = h^{-1}(\beta_0 + \beta_1 D_j + \beta_{2k} t)$ . Note that while this assumes that  $g(\cdot; D_j, t)$  incorporates  $D_j$  and  $t$  additively, this is not necessary and  $g(\cdot; D_j, t)$  could easily be described without this assumption. To define the joint distribution of  $(Y_{L1}, Y_{R1})^\top, \dots, (Y_{LK}, Y_{RK})^\top$ , we assume that the diagnoses of the readers are independent of each other, conditional on  $(D_L, D_R)^\top$  and  $T$ . It follows that the unconditional joint distribution of  $(Y_{Lk}, Y_{Rk})^\top, k = 1, \dots, K$ , is then

$$P(Y_{Lk} = y_{Lk}, Y_{Rk} = y_{Rk}, \forall k | D_L, D_R) = \int_{-\infty}^{+\infty} \prod_{k=1}^K \left[ 1 + \frac{\rho\{y_{Lk} - p_{Lk}(t)\}\{y_{Rk} - p_{Rk}(t)\}}{\sqrt{p_{Lk}(t)q_{Lk}(t)p_{Rk}(t)q_{Rk}(t)}} \right] \times \prod_{j=L,R} [p_{jk}(t)]^{y_{jk}} [q_{jk}(t)]^{1-y_{jk}} d\Phi(t) \tag{8}$$

where  $\rho$  is the conditional inter-eye correlation and  $\Phi(\cdot)$  is the standard normal cumulative distribution function. We can use the joint distribution in (8) to get the inter-reader pairwise correlations. With  $\mu_{jk} = \int_{-\infty}^{+\infty} p_{jk}(t) d\Phi(t)$ ,  $\sigma_{jk}^2 = \int_{-\infty}^{+\infty} p_{jk}(t)q_{jk}(t) d\Phi(t) + \text{var}[p_{jk}(t)]$ , and

$$\mu_{Lk,j'k'} = \begin{cases} \int_{-\infty}^{+\infty} p_{Lk}(t)p_{Rk'}(t) \left[ 1 + \frac{\rho\{1 - p_{Lk}(t)\}\{1 - p_{Rk'}(t)\}}{\sqrt{p_{Lk}(t)q_{Lk}(t)p_{Rk'}(t)q_{Rk'}(t)}} \right] d\Phi(t) & \text{for } j' \neq L, k = k' \\ \int_{-\infty}^{+\infty} p_{Lk}(t)p_{j'k'}(t) d\Phi(t) & \text{for } k \neq k' \end{cases}$$

these correlations are as follows:

$$\rho(Y_{Lk}, Y_{j'k'} | D_L, D_{j'}) = \begin{cases} (\mu_{Lk, j'k'} - \mu_{Lk} \mu_{j'k'}) (\sigma_{Lk} \sigma_{j'k'})^{-1} & \text{for } k \neq k' \\ (\mu_{Lk, Rk} - \mu_{Lk} \mu_{Rk}) (\sigma_{Lk} \sigma_{Rk})^{-1} & \text{for } j' \neq L, k = k' \\ 1 & \text{for } j' = L, k = k' \end{cases} \quad (9)$$

Given the disease status, note that  $\rho(Y_{Lk}, Y_{j'k'} | D_L, D_{j'})$  represents the unconditional correlation between readings by two different readers ( $k \neq k'$ ) and  $\rho(Y_{Lk}, Y_{Rk} | D_L, D_R)$  the unconditional correlation between the left and right eye readings by the same reader. Note that by symmetry,  $\rho(Y_{L1}, Y_{L2} | D_L, D_R) = \rho(Y_{R1}, Y_{R2} | D_L, D_R)$  and  $\rho(Y_{L1}, Y_{L2} | D_L) = \rho(Y_{R1}, Y_{R2} | D_R)$ .

The unconditional probability  $p_{jk} = P(Y_{jk} = 1 | D_j)$  is the average value of  $p_{jk}(t)$  over  $T$  given by  $p_{jk} = \int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_1 D_j + \beta_2 t) d\Phi(t)$ . Because ophthalmologists are generally interested in coming up with measures of accuracy independent of the particular reader conducting the diagnostic procedure, we can assume  $\beta_{2k} = \beta_2$  for all  $k = 1, \dots, K$ . The population-averaged sensitivity and specificity are thus

$$\pi_1 = \int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_1 + \beta_2 t) d\Phi(t) \quad (10)$$

$$\pi_0 = 1 - \int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_2 t) d\Phi(t) \quad (11)$$

Note that our approach parallels that adopted by Hadgu and Qu [16] and Qu *et al.* [17] in analysing diagnostic data with an imperfect gold standard.

### 3.1. Likelihood estimation

Suppose  $N$  patients undergo diagnostic testing on both left and right eyes for some pathology by several readers. Let  $\{y_{iLk}, y_{iRk}, D_{iL}, D_{iR}\}$ , for  $k = 1, \dots, K$  and  $i = 1, \dots, N$ , denote the observed data. The likelihood contribution of patient  $i$  is

$$L_i = \int_{-\infty}^{\infty} \prod_{k,j} [p_{ij}(t)]^{y_{ijk}} [q_{ij}(t)]^{1-y_{ijk}} \left[ 1 + \frac{\rho\{y_{iLk} - p_{iL}(t)\}\{y_{iRk} - p_{iR}(t)\}}{\sqrt{p_{iL}(t)q_{iL}(t)p_{iR}(t)q_{iR}(t)}} \right] d\Phi(t) \quad (12)$$

The log-likelihood function is then  $\ell(\beta_0, \beta_1, \beta_2, \rho) = \sum_{i=1}^N \log L_i$ .

The integrals in (12) may be evaluated by Gauss–Hermite quadrature techniques [18]. The method simply replaces the integration by a summation over a finite number  $Q$  of Gaussian quadrature points usually taken to be 10 or 20. To get the MLE of  $\boldsymbol{\eta}^\top = (\beta_0, \beta_1, \beta_2, \rho)$ , we first evaluate the log-likelihood function via Gauss–Hermite quadrature method, then numerically maximize it with respect to  $\boldsymbol{\eta}$ . Let  $t_{i1}, \dots, t_{iQ}$ ,  $i = 1, \dots, N$ , be the Gaussian quadrature points. Then,

$$\begin{aligned} \ell(\beta_0, \beta_1, \beta_2, \rho) &\approx \sum_{i=1}^N \log \left( \sum_{q=1}^Q w_{iq} \prod_{j,r} [p_{ij}(t_{iq})]^{y_{ijr}} [q_{ij}(t_{iq})]^{1-y_{ijr}} \right. \\ &\quad \left. \times \left[ 1 + \frac{\rho\{y_{iLr} - p_{iL}(t_{iq})\}\{y_{iRr} - p_{iR}(t_{iq})\}}{\sqrt{p_{iL}(t_{iq})q_{iL}(t_{iq})p_{iR}(t_{iq})q_{iR}(t_{iq})}} \right] \right) \end{aligned} \quad (13)$$

where the weights  $w_{iq}$ ,  $q = 1, \dots, Q$ ;  $i = 1, \dots, N$ , depend only on  $Q$  and the standard normal density.

If  $\ell^*$  is the right-hand side of (13), the (approximate) MLE  $\hat{\boldsymbol{\eta}}$  of  $\boldsymbol{\eta}$  is obtained by solving the (approximate) score equations  $\dot{\ell}^*(\boldsymbol{\eta}) = \partial \ell^* / \partial \boldsymbol{\eta}^\top = \mathbf{0}^\top$  via the Newton–Raphson algorithm. The usual large-sample normal theory-based inferences still apply, except for the random effect  $\beta_2$ , as the asymptotic normality fails in this case. McCulloch and Searle [19] recommend a simple adjustment for the  $p$ -value when testing  $H_0 : \beta_2 = 0$ . We use this in the example in Section 4.

One question that naturally arises regarding model (8) concerns the possible misspecification of the random effects distribution and its impact on the estimates. Posterior predictive model checking [20] provides a Bayesian route for validating the normal random effects distribution; alternatively, a mixture of normals affords a more flexible class of distributions for the random effects [21]. While a misspecified random effects distribution in generalized linear mixed models has been observed to produce biased estimates of fixed effects [22], which are usually of primary interest in practice, tests of hypotheses concerning the model are still reliable but caution is needed in interpreting the estimates, as these are no longer consistent.

As pointed out by a referee, model (8) is a ‘hybrid’ between a random effects model and that introduced by Prentice [13]. This differs from the usual approach in accounting for between-eye and between-reader correlations through a generalized linear mixed model with crossed random effects. Such a model engenders computational difficulties [23] mainly due to the high-dimensional integration involved in likelihood evaluation. In comparison, model (8) involves only a single random effect and thus, is computationally less demanding. Moreover, the conditional inter-eye correlation  $\rho$  in model (8) can be interpreted as representing shared biological characteristics between the eyes which may affect the diagnoses besides the presence or absence of the disease. While a model with crossed random subject and eye effects is able to induce correlation between eyes, the correlation  $\rho$  provides a natural way of accounting for such dependence.

#### 4. APPLICATION TO DIABETIC RETINOPATHY DATA

The study involved  $N = 92$  diabetic patients in Alberta, Canada, who were referred to a group retina practice in Edmonton, Alberta [1]. The study protocol required that patients be clinically examined on the same day they underwent stereoscopic digital photography by a trained ophthalmic photographer using a high-resolution digital camera. The digital images were stored uncompressed and then graded by experienced readers at least two months after they were taken. They were assessed in random order, with a minimum of two months in between review of the left eye images and those of the right eyes to minimize reader recall.

In order to evaluate the patients for treatable diabetic retinopathy, retinal thickening as well as several pathologies that are indicative of retinal thickening were identified as either present (positive) or absent (negative). The pathologies identified included clinically significant macular oedema (CSME), microaneurysms, intra-retinal haemorrhage, hard exudates, and other diseases of note. Contact lens biomicroscopy (CLBM), the clinical examination considered to be the ‘gold standard’ for most, but not all, of the pathologies considered, was performed on all the patients by retinal specialists to determine disease status. Digital images of the patients’ eyes were graded by two specialists and patients were diagnosed as either positive or negative for the pathologies.

The objective of the study was to compare stereoscopic digital photography to CLBM in the evaluation of treatable retinopathy among diabetic patients. While digital photography provides a

cost-effective teleophthalmology system, adequate identification of potentially treatable retinopathy should be ensured before widespread implementation.

#### 4.1. Case of single reader

We illustrate the methodology described in Section 2 on the pathology microaneurysm. Table V shows the data concerning the presence (+) or absence (−) of microaneurysm in the left and right eyes of 92 diabetic patients as evaluated by Reader 1. Model (1) was applied to these data with  $p(D_{ij}) = P(Y_{ij} = 1|D_{ij})$  specified as a binary logistic regression model with parameters  $\beta_0$  and  $\beta_1$ , and with inter-eye correlation  $\rho$ .

Table VI displays the MLEs of these parameters and their standard errors along with those from the crude method and GEE. We note that the inter-eye correlation estimates of 0.1059 and 0.1702 from maximum likelihood and GEE, respectively, suggest a weak association between the left and right eye diagnoses for microaneurysm. Because of this, the estimates based on the crude method are very close to those from model (1) and GEE. For example, all three approaches yielded estimates of sensitivity and specificity for the test of about 84 and 95 per cent, respectively. Note, however, that the standard errors of estimates from model (1) are generally larger than those of estimates using the crude method and GEE. This is to be expected in view of the results reported in Sections 2.2 and 2.3.

Our proposed approach based on model (1) of analysing the data is more general than the crude method, and reduces to the crude method when the eyes are independent. If, by our method,

Table V. Evaluations by Readers 1 and 2 (in parentheses) of presence (+) or absence (−) of microaneurysm in the left (*L*) and right (*R*) eyes of  $N = 92$  diabetic patients.

Disease status	Test				Total
	<i>L</i> + <i>R</i> +	<i>L</i> − <i>R</i> +	<i>L</i> + <i>R</i> −	<i>L</i> − <i>R</i> −	
<i>L</i> + <i>R</i> +	38 (45)	8 (2)	3 (1)	2 (3)	51
<i>L</i> − <i>R</i> +	1 (1)	1 (1)	0 (0)	1 (1)	3
<i>L</i> + <i>R</i> −	2 (2)	0 (0)	2 (4)	2 (0)	6
<i>L</i> − <i>R</i> −	0 (1)	1 (2)	0 (1)	31 (28)	32
Total	41 (49)	10 (5)	5 (6)	36 (32)	92

Table VI. ML, crude method, and GEE estimates of parameters from model (1) for microaneurysm based on Reader 1 data.

Parameter	Model (1)		Crude		GEE	
	Estimate	SE	Estimate	SE	Estimate	SE
$\beta_0$	−2.8452	0.5400	−2.8478	0.5143	−2.8478	0.5143
$\beta_1$	4.4729	0.6044	4.4900	0.5752	4.4900	0.5752
$\rho$	0.1059	0.2407	—	—	0.1702	—
$\pi_1$	0.8359	0.0380	0.8378	0.0350	0.8378	0.0367
$\pi_0$	0.9451	0.0280	0.9452	0.0266	0.9452	0.0264

the inter-eye correlation  $\rho$  turns out to be not very different from 0, like in the case of microaneurysm above, then we may assume independence of the eyes, and the crude method will suffice.

4.2. Case of two readers

We now consider data on microaneurysm from two readers and analyse them *via* the proposed methodology. Table V displays, in parentheses, the left- and right-eye microaneurysm evaluations by another reader, Reader 2.

We apply model (8) to the two sets of readings to account for the correlation between the two readers, using a logistic link. The MLEs are  $\hat{\beta}_0 = -5.2561$  (SE = 1.3584),  $\hat{\beta}_1 = 9.6060$  (SE = 2.2866),  $\hat{\beta}_2 = 3.6834$  (SE = 0.9933), and  $\hat{\rho} = -0.0581$  (SE = 0.2673). Following McCulloch and Searle [19], we get the large-sample likelihood ratio statistic  $-2[\ell(\hat{\beta}_0, \hat{\beta}_1, 0, \hat{\rho}) - \ell(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\rho})] = 35.6$ , and upon comparison with the adjusted critical value  $\chi^2_{1,0.9} = 2.7055$ , we conclude that  $\beta_2$  is significant at the 5 per cent level, and that there does exist inter-reader correlation. Note as well the estimated conditional inter-eye correlation  $\hat{\rho} = -0.0581$ , indicating a weak association between the left and right eye diagnoses. We also calculated the (unconditional) correlations in (9) as follows:

$$\hat{\rho}(Y_{Lk}, Y_{Rk} | D_L, D_R) = \begin{cases} 0.5059 & \text{for } D_L = D_R = 0 \\ 0.4900 & \text{for } D_L = D_R = 1 \\ 0.1254 & \text{for } D_L = 0, D_R = 1 \text{ or } D_L = 1, D_R = 0 \end{cases}$$

for  $k = 1, 2$ . Similarly, we can get

$$\hat{\rho}(Y_{L1}, Y_{R2} | D_L, D_R) = \begin{cases} 0.5330 & \text{for } D_L = D_R = 0 \\ 0.5180 & \text{for } D_L = D_R = 1 \\ 0.1290 & \text{for } D_L = 0, D_R = 1 \text{ or } D_L = 1, D_R = 0 \end{cases}$$

and

$$\hat{\rho}(Y_{L1}, Y_{L2} | D_L) = \begin{cases} 0.5330 & \text{for } D_L = 0 \\ 0.5180 & \text{for } D_L = 1 \end{cases}$$

Using estimates of the marginal probabilities  $P(D_L = m, D_R = m')$  and  $P(D_L = m)$ , we obtain  $\hat{\rho}(Y_{L1}, Y_{R1}) = \hat{\rho}(Y_{L2}, Y_{R2}) = 0.4599$ ,  $\hat{\rho}(Y_{L1}, Y_{R2}) = 0.4852$ , and  $\hat{\rho}(Y_{L1}, Y_{L2}) = 0.5237$ .

We can now compare the results from the crude method, GEE, model (1), and model (8) concerning the sensitivity and specificity of the test. For the crude method and model (1), the average value of the estimates for the two readers was calculated. For the GEE, an independent ‘working’ correlation matrix containing the correlations  $\rho(Y_{jk}, Y_{j'k'})$  ( $j, j' = L, R; k, k' = 1, 2$ ) was specified for  $(Y_{L1}, Y_{R1}, Y_{L2}, Y_{R2})^T$ . The correlations based on the GEE estimates are  $\hat{\rho}_{GEE}(Y_{L1}, Y_{R1}) = \hat{\rho}_{GEE}(Y_{L2}, Y_{R2}) = 0.1854$ ,  $\hat{\rho}_{GEE}(Y_{L1}, Y_{R2}) = 0.2705$ , and  $\hat{\rho}_{GEE}(Y_{L1}, Y_{L2}) = 0.5041$ .

Results from Table VII indicate the close correspondence between estimates from the crude method and those from model (1) and GEE, which are slightly higher than those from model (8). This is to be expected as  $\hat{\rho}$  is quite small. Observe, however, that the standard errors from model (8) are larger than those from the other methods. This is not surprising because model (8), like the

Table VII. Comparison of estimates of sensitivity and specificity of test for diagnosing microaneurysm using the four approaches.

	Parameter	Estimate	SE	95 per cent CI	
				Lower	Upper
Model (1)	$\pi_1$	0.8784	0.0240	0.8314	0.9254
	$\pi_0$	0.9193	0.0235	0.8732	0.9654
Model (8)	$\pi_1$	0.8616	0.0674	0.7295	0.9937
	$\pi_0$	0.9054	0.0620	0.7839	1.0269
GEE approach	$\pi_1$	0.8739	0.0293	0.8165	0.9313
	$\pi_0$	0.9178	0.0305	0.8580	0.9775
Crude method	$\pi_1$	0.8739	0.0222	0.8304	0.9174
	$\pi_0$	0.9178	0.0226	0.8735	0.9621

GEE, accounts for the correlation between the readers, thus adding another source of variation in the analysis.

Therefore, treating the readers as independent ignores the correlation between the readers, and may consequently yield incorrect estimates, especially of the standard errors. Model (8) avoids this by accounting for this source of variation in the analysis.

## 5. DISCUSSION

This paper focuses on two main issues arising in multi-reader binocular diagnostic studies: how to account for inter-reader correlation while at the same time incorporating the correlation between the binocular outcomes. The general approach taken in the paper was a model-based one that relies on specifying a model for the joint distribution of the outcomes. Inferences are then developed for the parameters of the model. The approach is motivated by the need to account for the different sources of correlations in the data such as those between the binocular outcomes and between readers. This approach should be preferred to one that carries out separate analyses for the binary variables, as it provides a systematic and non-*ad hoc* way of analysing the data and results in substantial gains in efficiency.

The approach is illustrated with data from a study concerning retinopathy-related pathologies among diabetic patients as evaluated by several readers. Estimates derived from the approach are shown, empirically *via* simulations, to be more efficient than those from the crude method, which ignores the intra-pair correlation, and the GEE.

The proposed approach provides as well an alternative to the GEE method in multi-reader multi-disease diagnostic studies, where sensitivity and specificity estimates need to be obtained for each disease using data from several readers. Model (8) can be extended to the case of several pathologies in the diabetic retinopathy study by including another latent variable which varies across pathologies. Specifically, given  $K$  readers and  $V$  pathologies, we can define the joint distribution of  $(Y_{Lkv}, Y_{Rkv})^\top$ ,  $v = 1, \dots, V$ ;  $k = 1, \dots, K$ , by assuming the diagnoses across readers and across pathologies are independent of each other, conditional on  $(D_{Lv}, D_{Rv})^\top$ ,  $T$ , and  $U_k$ ,  $v = 1, \dots, V$ ;

$k = 1, \dots, K$ . The conditional joint distribution is then

$$P \left( \begin{array}{c} Y_{Lkv} = y_{Lkv}, Y_{Rkv} = y_{Rkv}, \\ \forall k, v \end{array} \middle| \begin{array}{c} D_{Lv}, D_{Rv}, \forall v; \\ t, u_k, \forall k \end{array} \right) = \prod_{k,v} [p_{Lkv}(t, u_k)]^{y_{Lkv}} [q_{Lkv}(t, u_k)]^{1-y_{Lkv}} \\ \times [p_{Rkv}(t, u_k)]^{y_{Rkv}} [q_{Rkv}(t, u_k)]^{1-y_{Rkv}} \\ \times \left[ 1 + \frac{\rho \prod_j \{y_{jkv} - p_{jkv}(t, u_k)\}}{\prod_j \sqrt{p_{jkv}(t, u_k)q_{jkv}(t, u_k)}} \right]$$

where  $\rho$  is the (conditional) inter-eye correlation, and a generalized linear mixed model with nested random effects  $T$  and  $U_k$  may be assumed for  $p_{jkv}$ ,  $k = L, R$ . Estimation and inference in this case are made more computationally cumbersome by the multi-dimensional integration involved in evaluating the likelihood function. We investigate this in another article.

While the methodology in the paper can easily accommodate unbalanced data (e.g. subjects missing data on an eye or a reader), its performance needs to be studied. Finally, many pathologies studied in the diabetic retinopathy study considered in the paper do not have perfect gold standards for determining true disease status. Further research in this direction seems to be needed.

#### ACKNOWLEDGEMENTS

The research of A. R. de Leon is supported by the Natural Sciences and Engineering Research Council of Canada contract/grant number 76-1060 and the University of Calgary Research Grants Committee grant number 78-1045. The authors are grateful to an associate editor and the referees for very helpful comments which greatly improved the form and content of the paper.

#### REFERENCES

1. Rudnisky CJ, Hinz BJ, Tennant MTS, de Leon AR, Greve MDJ. High-resolution stereoscopic digital fundus photography versus contact-lens biomicroscopy for the detection of clinically significant macular edema. *Ophthalmology* 2001; **109**:267–274.
2. Maberley D, Walker H, Koushnik A, Cruess A. Screening for diabetic retinopathy in James Bay, Ontario: a cost-effectiveness analysis. *Canadian Medical Association Journal* 2003; **168**(2):160–164.
3. Zhou X-H, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. Wiley: New York, 2002.
4. Leite MLC, Nicolosi A. Statistical analysis of correlated binary data in ophthalmology: a weighted logistic regression approach. *Ophthalmic Epidemiology* 1998; **5**:117–131.
5. Glynn RJ, Rosner B. Accounting for the correlation between fellow eyes in regression analysis. *Archives of Ophthalmology* 1992; **110**:381–387.
6. Lee E, Dubin N. Estimation and sample size considerations for clustered binary responses. *Statistics in Medicine* 1994; **13**:1241–1252.
7. Donner A, Klar N. Confidence interval construction for effect measures arising from cluster randomization trials. *Journal of Clinical Epidemiology* 1993; **46**:123–131.
8. Rao JNK, Scott A. A simple method for the analysis of clustered binary data. *Biometrics* 1992; **48**:577–585.
9. Smith PJ, Hadgu A. Sensitivity and specificity for correlated observations. *Statistics in Medicine* 1992; **11**: 1503–1509.
10. Hujuel PP, Moulton LH, Loesche WJ. Estimation of sensitivity and specificity of site-specific diagnostic tests. *Journal of Periodontal Research* 1990; **25**:193–196.
11. Bahadur RR. A representation of the joint distribution of responses to  $n$  dichotomous items. In *Studies in Item Analysis and Prediction*, Solomon H (ed.). Stanford University Press: Stanford, 1962; 158–168.

12. Ahn C. An evaluation of methods for the estimation of sensitivity and specificity of site-specific diagnostic tests. *Biometrical Journal* 1997; **39**:793–807.
13. Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics* 1988; **44**:1033–1048.
14. Zheng Y, Barlow WE, Cutter G. Assessing accuracy of mammography in the presence of verification bias and intrareader correlation. *Biometrics* 2005; **61**:259–268.
15. Boyko EJ, Alderman BW, Baron AE. Reference test errors bias the evaluation of diagnostic tests for ischemic heart disease. *Journal of General Internal Medicine* 1988; **3**:476–481.
16. Hadgu A, Qu Y. A biomedical application of latent class models with random effects. *Applied Statistics* 1998; **47**:603–616.
17. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996; **52**:797–810.
18. Lesaffre E, Spiessens B. On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics* 2001; **50**:325–335.
19. McCulloch CE, Searle SR. *Generalized, Linear, and Mixed Models*. Wiley: New York, 2001.
20. Carlin JB, Wolfe R, Hendricks Brown C, Gelman A. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics* 2001; **2**:397–416.
21. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer: Berlin, 2005.
22. Litière S, Alonso A, Molenberghs G, Geys H. Impact of misspecified random-effects distribution on maximum likelihood estimation in generalized linear mixed models. *Journal of Statistical Computation and Simulation* 2005, submitted for publication.
23. Bellio R, Varin C. A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling* 2005; **5**:217–227.