Joint estimation of disease-specific sensitivities and specificities in reader-based multi-disease diagnostic studies of paired organs

N. Withanage\textsuperscript{a,b}, A.R. de Leon\textsuperscript{b,*} and C.J. Rudnisky\textsuperscript{c}

\textsuperscript{a}Department of Economics & Statistics, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka; \textsuperscript{b}Department of Mathematics & Statistics, University of Calgary, Calgary, Canada; \textsuperscript{c}Department of Ophthalmology, University of Alberta, Edmonton, Canada

(Received 24 July 2013; accepted 26 March 2014)

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 fellow eyes. This may consequently yield incorrect estimates, especially of the standard errors. The paper is concerned with diagnostic studies wherein several diagnostic tests, or the same test read by several readers, are administered to identify one or more diseases. A likelihood-based method of estimating disease-specific sensitivities and specificities via hierarchical generalized linear mixed models is proposed to meaningfully delineate the various correlations in the data. The efficiency of the estimates is assessed in a simulation study. Data from a study on diabetic retinopathy are analyzed to illustrate the methodology.

Keywords: binocular data; correlated binary outcomes; pairwise likelihood; generalized linear mixed model; latent variable models

1. Introduction

Subjecting patients to several tests or to the same test on several occasions is a common protocol in many diagnostic studies in medicine and health. This is especially true in situations where test outcomes are reader-based assessments, in which case two or more readers are necessary to minimize, if not eliminate, so-called reader bias. Simply assuming that the test results are independent is unwise, since associations among these observations often exist; for example, test results from the same patient are frequently correlated. The proper accounting of these associations, which

\*Corresponding author. Email: adeleon@ucalgary.ca
intuitively exist between measurements from a patient, is an interesting statistical question. Failure to account for such correlations by treating measurements from the patient as independent may consequently yield incorrect inferences. Because valid information contained in correlated observations is less than expected from independent observations, standard errors (SEs) calculated by incorrectly assuming correlated observations to be independent tend to underestimate the true sampling variability, consequently yielding inflated type I errors of significance tests [18,22].

This paper is motivated by a reader-based diagnostic study of diabetic retinopathy [28] in Alberta, Canada, in which at least two readers were used to diagnose the presence or absence of certain pathologies (e.g. clinically significant macular edema (CSME), hard exudates (HEX), etc.) among diabetic patients who suffer from treatable diabetic retinopathy. In Canada, where a disproportionate number of diabetic patients are First Nations Canadians living on reserves in far-flung rural areas, sending ophthalmologists on remote clinics can be costly and inefficient. Due to advances in digital imaging in recent years, a possible alternative is distance evaluation wherein patients undergo stereoscopic digital photography using a high-resolution digital camera. In this approach, digital images of patients’ eyes are read by at least two ophthalmologists and patients are diagnosed as either positive (i.e. disease is present) or negative (i.e. disease is absent) for the pathologies. This cost-effective tele-ophthalmologic technique has the potential to increase rural accessibility to specialist eye care [20], allowing for early detection and treatment of diabetic retinopathy. Only patients who need treatments would have to travel to a specialist; the transportation cost is thus also reduced. However, before wide implementation of any potential new diagnostic methodology, its accuracy must first be examined. 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. In such reader-based diagnostic studies, it is often of interest to estimate the overall sensitivity and specificity, independently of readers. Making the usual assumption that the test results for different patients are independent, there is a complex correlation structure to a patient’s diagnostic data that needs to be incorporated in the analysis.

The accuracy of a diagnostic or screening test can be described by several measures, the most common of which are given by the test’s sensitivity and specificity with respect to the true disease status as determined by the ‘gold standard’. Sensitivity is the probability of a diseased patient having a positive test result and specificity is the probability of a non-diseased patient having a negative test result. There has been previous work on the estimation of sensitivity and specificity and the calculation of the estimates’ SEs in the context of clustered binary diagnostic data. These include simple adjustments to SEs introduced by Rao and Scott [24] and Donner and Klar [4] to account for the intra-cluster correlation, and a weighted estimate proposed by Lee and Dubin [12] for handling unbalanced cluster sizes. A similar approach based on weighting was discussed by Leite and Nicolosi [15] in the context of logistic regression analysis of binocular ophthalmologic data. Traditional estimation methods that ignore correlations in clustered diagnostic data have been shown to be inadequate as they underestimate SEs and lead to incorrect inferences [6]. Methods that account for these correlations are thus needed. Approaches to handling this problem include the generalized estimating equations (GEE) approach of Smith and Hadgu [29] (see also Sternberg and Hadgu [31] and Leisenring et al. [14]) and those based on likelihood methods studied by Huijoe et al. [11] and Rosner [27], among others (see also Sutradhar and Das [32] and Lefkopoulou et al. [13]). More recently, Hsiao et al. [8] discussed the assessment of inter- and intra-rater reliability for data structures with more than one level of nesting from a Bayesian point of view. One shortcoming of their model is that associations are restricted to be positive. A number of authors have also discussed estimating diagnostic accuracy without a gold standard; a latent class approach is common in this situation (see Hui and Walter [9], Hui and Zhou [10] and Zhou et al. [36]). Such an approach assumes that multiple readers/tests are
independent, conditional on the true disease status. However, this assumption cannot always be justified and may not be valid in practice. Alternatively, Qu et al. [23] developed a general latent class model with normally distributed subject-specific random effects to model the conditional dependence among multiple diagnostic tests. In a further extension, Wang and Zhou [34] recently incorporated normal subject-specific random effects while assuming fixed effects for the raters. Recent references on correlated ophthalmologic data include de Leon et al. [16,17].

A generalized linear mixed model (GLMM) for estimating disease-specific sensitivities and specificities using correlated binocular binary data from reader-based multi-disease diagnostic studies is discussed in this paper. Such an approach allows for the flexible incorporation of the specificities using correlated binocular binary data from reader-based multi-disease diagnostic studies. In a further extension, Wang and Zhou [34] recently incorporated normal subject-specific random effects while assuming fixed effects for the raters. We assume that

\[ Y_{ijkv} = \gamma^T x_{ij} + \beta_v(0) + B_{0v} + (\beta_{v1} + B_{11v})D_{ijv} + B_{ik} + B_{ij}^{(e)} + \epsilon_{ijkv}, \]

for all \( i, j, k \) and \( v \), with \( \gamma \) the vector of regression coefficients, \( \beta_v(0) \) and \( \beta_{v1} \) the pathology-specific fixed effects,

\[
\begin{pmatrix} B_{00}^{(p)} \\ \vdots \\ B_{0V}^{(p)} \\ B_{11}^{p} \\ \vdots \\ B_{1V}^{p} \end{pmatrix} \sim \text{iid } N_{2V}(0, \Sigma^{(p)}), \quad B_{ik}^{(r)} \sim \text{iid } N_K(0, \Sigma^{(r)}), \\
B_{ij}^{(e)} \sim \text{iid } N_2(0, \Sigma^{(e)}),
\]

where \( B_{0v}^{(p)} \) and \( B_{1v}^{(p)} \) are respective pathology-specific random intercepts and slopes for disease status, \( B_{ij}^{(e)} \) represents reader-specific random effects, and \( B_{ij}^{(e)} \) represents eye-specific random effects and where we assumed independence among the different random effects. We also have


\[ \epsilon_{ijkv} \overset{iid}{\sim} N(0, 1) \] as the errors, assumed to be independent of the random effects. It follows that the \( Y_{ijkv} \)'s have a joint latent Gaussian distribution, which implies a multivariate probit model for the binary outcomes. That is, the model given by Equation (1) yields a probit mixed model for \( Y_{ijkv} \) given by

\[
P(Y_{ijkv} = 1 | \beta_0, B_{i0v}, B_{1i1v}, B_{ik}, B_{ij}^{(e)}) = \Phi(\gamma^\top x_{ij} + \beta_0 + B_{i0v} + (\beta_{v1} + B_{1i1v})D_{ijv} + B_{ik} + B_{ij}^{(e)}),
\]

where \( \Phi(\cdot) \) is the standard Gaussian cumulative distribution function (CDF). Note that we assumed the cutpoints or thresholds to be all 0 for identifiability, since we included the intercept \( \beta_{v0} \) in the GLMM given by Equation (2). In addition, we made the conventional assumption that \( \text{var}(Y_{ijkv} | \beta_0, B_{i0v}, B_{1i1v}, B_{ik}, B_{ij}^{(e)}) = \text{var}(\epsilon_{ijkv} | \beta_0, B_{i0v}, B_{1i1v}, B_{ik}, B_{ij}^{(e)}) = 1 \), for all \( i, j, k, v \).

Note that the model given by Equation (1) is quite flexible. The random intercepts \( B_{i0}^{(p)} \) and random slopes \( B_{iki}^{(e)} \) of true disease status induce between-pathology correlations via their covariance matrix \( \Sigma^{(p)} \). The random effects \( B_{iki}^{(e)} \) represent between-reader heterogeneity and thus, account for correlations between readers arising from the fact that the readers base their pathology diagnoses on the same image of a patient’s eye, while the random effects \( B_{ij}^{(e)} \) account for correlations between fellow eyes. Observe that the associations incorporated in the model given by Equation (1) are not restricted only to positive correlations.

Define the stacked vector \( \mathbf{Y}_i = (Y_{i11}, Y_{i12}, \ldots, Y_{iM})^\top \) = \( (Y_{i1}, \ldots, Y_{iM})^\top \), where \( M = 2kV \), correspondingly, let \( \mathbf{Y}_i^* = (Y_{i11}^*, Y_{i12}^*, \ldots, Y_{iM}^*)^\top \) = \( (Y_{i1}^*, \ldots, Y_{iM}^*)^\top \). The multivariate LMM given by Equation (1) (i.e. the multivariate GLMM for the binary outcomes) then takes the form

\[
\mathbf{Y}_i^* = \mathbf{X}_i \beta + \mathbf{Z}_i B_i^{(p)} + \mathbf{Z}_2 B_i^{(r)} + \mathbf{Z}_3 B_i^{(e)} + \epsilon_i^*,
\]

where the stacked errors \( \epsilon_i^* = (\epsilon_{i11}^*, \epsilon_{i12}^*, \ldots, \epsilon_{iM}^*)^\top \) are iid according to an \( M \)-dimensional centered Gaussian distribution (i.e. with zero means) and covariance matrix \( \Sigma = \mathbf{I}_M \), the \( M \times M \) identity matrix. Here, \( \mathbf{X}_i \) is the matrix containing the disease status and eye-specific covariates for patient \( i \), and \( \beta = (\gamma^\top \beta_{10}, \beta_{11}, \ldots, \beta_{V0}, \beta_{V1})^\top \); \( \mathbf{Z}_{il}, \mathbf{Z}_{2i} \) and \( \mathbf{Z}_{3i} \) are the design matrices for the respective random effects \( B_i^{(p)} = ((B_{i0}^{(p)})^\top, (B_{1i1}^{(p)})^\top, \ldots, B_{ik}^{(p)})^\top \) and \( B_i^{(e)} \), representing pathology-specific, reader-specific and eye-specific random effects, respectively; note that \( \mathbf{Z}_{il} \) contains the disease statuses \( D_{ijv} \). In addition, \( B_i^{(p)}, B_i^{(r)}, B_i^{(e)} \) and \( \epsilon_i^* \) are assumed to be independent. It is then easy to see that the marginal distribution of \( \mathbf{Y}_i^* \) is an \( M \)-dimensional Gaussian distribution with mean \( \mu_i^* = \mathbf{E}(\mathbf{Y}_i^*) = \mathbf{X}_i \beta \) and covariance matrix

\[
\Sigma_i^* = \mathbf{Z}_1 \Sigma^{(p)} \mathbf{Z}_1^\top + \mathbf{Z}_2 \Sigma^{(r)} \mathbf{Z}_2^\top + \mathbf{Z}_3 \Sigma^{(e)} \mathbf{Z}_3^\top + \mathbf{I}_M = (\sigma_{hh}^*)_{h,h'=1,\ldots,M}.
\]

Note that although \( \Sigma_i^* \) varies with \( i \), its elements do not depend on \( i \). We thus get a (marginal) multivariate probit model for \( \mathbf{Y}_i \) given by

\[
P(Y_{i1} = y_{i1}, \ldots, Y_{iM} = y_{iM}; \mathbf{D}_i, \mathbf{X}_i) = \int_{A_{1M} \times \cdots \times A_{11}} \phi_M^{(c)}(\mathbf{y}_i^* - \mu_i^*; \Sigma_i^*) \, dy_i^*,
\]

where \( \phi_M^{(c)}(\cdot; \Sigma_i^*) \) is the centered \( M \)-dimensional Gaussian density with covariance matrix \( \Sigma_i^* \), with the interval \( A_h \) either \((\infty, 0] \) or \([0, +\infty) \) accordingly as whether \( Y_{ih} \) is 0 or 1, and with \( \mathbf{D}_i \) the vector of true disease statuses for patient \( i \).

Suppressing the indices, it follows that

\[
P(Y_h = 1; D, \mathbf{x}) = P(Y_h^* > 0; D, \mathbf{x}) = \Phi \left( \frac{\beta_0 + \beta_1 D + \gamma^\top \mathbf{x}}{\sqrt{\sigma_{hh}^*}} \right).
\]
We thus have eye-specific sensitivities $\text{Sen}_{jv}$ and specificities $\text{Spc}_{jv}$ for pathology $v$ as

$$\text{Sen}_{jv} = \Phi \left( \frac{\beta_{v0} + \beta_{v1} + \gamma^T x_j}{\sqrt{\sigma^2_{hh}}} \right),$$

$$\text{Spc}_{jv} = 1 - \Phi \left( \frac{\beta_{v0} + \gamma^T x_j}{\sqrt{\sigma^2_{hh}}} \right),$$

for $v = 1, \ldots, V$ and $j = L, R$. In many applications, the covariates $x_L$ and $x_R$ are usually measured at the patient level, so that $x_L = x_R = x$. This implies that $\text{Sen}_{Lv} = \text{Sen}_{Rv} = \text{Sen}_{v}$ and $\text{Spc}_{Lv} = \text{Spc}_{Rv} = \text{Spc}_{v}$. That is, the sensitivity and specificity of a diagnostic test for a pathology are independent of the particular eye under consideration. Note, however, that even in the absence of eye-level covariates, it is still possible to obtain eye-specific sensitivities and specificities by varying the parameters $\beta_{v0}$ and $\beta_{v1}$ with the eyes. Finally, because ophthalmologists are generally interested in measures of accuracy that are independent of the particular reader conducting the diagnostic procedures, $\text{Sen}_v$ and $\text{Spc}_v$ are conveniently free of $k$.

It can be shown that for any diagonal matrix $G$ with positive diagonal elements,

$$P(Y = y; X, \beta, \Sigma^*) = P(Y = y; X, G\beta, G\Sigma^*G).$$

This implies that the variances in the covariance matrix $\Sigma^*$ cannot be estimated based on the likelihood function [35] and hence only the correlations are identifiable (see appendix for a detailed discussion). For this reason, we assume that $\Sigma^{(p)}$ and $\Sigma^{(e)}$ in Equation (3) are correlation matrices. In this setting, by including different random effects for eyes and readers, fixing variances to unity does not unduly constrain the model. By allowing for an unstructured $\Sigma^{(p)}$, the researcher has the flexibility of having different correlations for different pathologies.

A noteworthy strength of the joint model given by Equation (3) lies in the flexibility in the way it accounts for and delineates the various correlations in the data through the introduction of reader-specific, pathology-specific and eye-specific random effects. The following section discusses the various marginal correlations induced by our model. As discussed in Section 3, the joint model also lends itself to computationally straightforward estimation of its parameters via PL.

### 2.1 Marginal correlations

Under the Gaussian latent specification of the binary outcomes in $Y_i$, the model in Equation (3) yields a correlated probit model for $Y_i$. This model is very flexible in accounting for the correlation structure of clustered multiple binary variables because of the underlying multivariate LMM for the Gaussian latent vector $Y^*_i$. Based on this, the associations among the binary outcomes are measured by the correlations between the corresponding latent continuous variables. These correlations are called the tetrachoric correlations between the binary outcomes and are commonplace in psychometrics. Because test results for different patients are assumed to be independent, the associations between the outcomes for a specific patient are captured solely by the three different random effects. For simplicity, we assume in what follows that only random intercepts $B^{(p)}_i$ for the pathologies are included in the joint model. Using $\text{corr}(Y^*_{ih}, Y^*_{ih'}) = \sigma^*_h/\sqrt{\sigma^*_h\sigma^*_{hh'}}$, the marginal tetrachoric correlations are then found to be

$$\text{corr}(Y^*_{ijkv}, Y^*_{ij'k'v}) = \frac{1 + \sigma_v^{(p)} + \rho^{(e)}}{3 + \sigma_v^{(p)}},$$

(8)
where \( \rho^{(p)} \), \( \rho^{(r)}(r) \) and \( \rho^{(e)} \) are the correlations between the pathology-specific random effects (i.e. the off-diagonal elements of \( \Sigma^{(p)} \)), the reader-specific random effects (i.e. the off-diagonal elements of \( \Sigma^{(r)} \)) and the eye-specific random effects (i.e. the off-diagonal element of \( \Sigma^{(e)} \)), respectively, with \( \sigma_{pp}^{(p)} = \text{var}(B_{ij}^{(p)}) \) (i.e. the diagonal elements of \( \Sigma^{(p)} \)). In practice, we may assume a common correlation between the reader-specific random effects (i.e. exchangeability of \( B_{i1}^{(r)}, \ldots, B_{ik}^{(r)} \)), in which case we have \( \rho^{(r)} = \rho^{(r)}(r) \), for all \( k \neq k' \).

Based on Equations (8)–(14), it is clear that our model provides a flexible specification of the dependence structure in the data.

### 3. Likelihood estimation

Marginally, \( Y_i^* \) follows an \( M \)-dimensional Gaussian distribution with mean \( \mu_i^* \) and covariance matrix \( \Sigma_i^* \). As shown in Section 2.1, \( \Sigma_i^* \) can accommodate a complex correlation structure for the data. However, although the full joint model for \( Y_i^* \) is completely specified, the evaluation of the corresponding multivariate Gaussian orthant probabilities is difficult in practice. For instance, the case of two readers and two pathologies necessitates the evaluation of \( 2 \times 2 \times 2 = 8 \) multivariate Gaussian orthant probabilities. To obviate the computational complexity involved, an alternative is to use a pseudo- or composite likelihood function obtained from low-dimensional margins [1,3]. Selection of low-dimensional margins is based on computational savings. Here we adopt the PL method outlined in [25,26].

#### 3.1 PL estimation

The PL contribution of subject \( i \) is obtained by replacing the full likelihood function (i.e. the density of \( Y_i^* \)) by the product of the pairwise probabilities for the pairs \( Y_{ih}^* \) and \( Y_{ih'}^* \), for \( h' < h \). Since

\[
\begin{pmatrix}
Y_{ih}^* \\
Y_{ih'}^*
\end{pmatrix}
\sim N_2
\begin{pmatrix}
\mu_{ih}^* \\
\mu_{ih'}^*
\end{pmatrix},
\begin{pmatrix}
\sigma_{hh}^* & \sigma_{hh'}^* \\
\sigma_{hh'}^* & \sigma_{h'h'}^*
\end{pmatrix},
\]


for any pair $h' < h$, then the contribution $p\ell_i$ of patient $i$ to the log-PL can be written as

$$p\ell_i = \sum_{h=1}^{M} \sum_{h'=1}^{h-1} \left( \delta_{ihh'}^{(11)} \log p_{ihh'}^{(11)} + \delta_{ihh'}^{(10)} \log p_{ihh'}^{(10)} + \delta_{ihh'}^{(01)} \log p_{ihh'}^{(01)} + \delta_{ihh'}^{(00)} \log p_{ihh'}^{(00)} \right)$$

$$= \sum_{h=1}^{M} \sum_{h'=1}^{h-1} p\ell_{ihh'},$$

(15)

where $\delta_{ihh'}^{(\ell r)} = 1$, if $Y_{ih} = \ell$ and $Y_{ih'} = r$, and $\delta_{ihh'}^{(\ell r)} = 0$, otherwise, with

$$p_{ihh'}^{(\ell r)} = P(Y_{ih} = \ell, Y_{ih'} = r) = \begin{cases} P(Y_{ih}^* > 0, Y_{ih'}^* > 0) & \text{if } \ell = r = 1, \\ P(Y_{ih}^* > 0, Y_{ih'}^* \leq 0) & \text{if } \ell = 1, r = 0, \\ P(Y_{ih}^* \leq 0, Y_{ih'}^* > 0) & \text{if } \ell = 0, r = 1, \\ P(Y_{ih}^* \leq 0, Y_{ih'}^* \leq 0) & \text{if } \ell = r = 0. \end{cases}$$

Note that

$$p_{ihh'}^{(11)} = \Phi \left( \frac{\mu_{ih}}{\sigma_{hh}} \right) - \Phi \left( -\frac{\mu_{ih}}{\sigma_{hh}} \right) + \Phi_2 \left( -\frac{\mu_{ih}}{\sigma_{hh}}, -\frac{\mu_{ih}}{\sigma_{hh'}} ; \rho_{hh'} \right),$$

(17)

$$p_{ihh'}^{(10)} = \Phi \left( -\frac{\mu_{ih}}{\sigma_{hh'}} \right) - \Phi \left( -\frac{\mu_{ih}}{\sigma_{hh}} \right) - \Phi_2 \left( -\frac{\mu_{ih}}{\sigma_{hh}}, -\frac{\mu_{ih}}{\sigma_{hh'}} ; \rho_{hh'} \right),$$

(18)

$$p_{ihh'}^{(01)} = \Phi \left( -\frac{\mu_{ih}}{\sigma_{hh}} \right) - \Phi_2 \left( -\frac{\mu_{ih}}{\sigma_{hh}}, -\frac{\mu_{ih}}{\sigma_{hh'}} ; \rho_{hh'} \right),$$

(19)

$$p_{ihh'}^{(00)} = \Phi_2 \left( -\frac{\mu_{ih}}{\sigma_{hh}}, -\frac{\mu_{ih}}{\sigma_{hh'}} ; \rho_{hh'} \right),$$

where $\Phi_2(\cdot; \tau)$ is the standardized bivariate Gaussian CDF with correlation $\tau$, and $\rho_{hh'}$ is the pairwise correlation between $Y_{ih}^*$ and $Y_{ih'}^*$, obtained by selecting the appropriate $2 \times 2$ submatrix of $\Sigma^*_i$.

The PL estimate $\hat{\Theta}$ of $\Theta$, which contains $\beta$ and the unique parameters in $\Sigma^{(p)}$, $\Sigma^{(r)}$ and $\Sigma^{(e)}$, is obtained by maximizing log PL $= \sum_{i=1}^{N} p\ell_i$ with respect to $\Theta$. This is accomplished by solving the pairwise score equations $0 = U_{p\ell}(\Theta) = \partial \log PL / \partial \Theta$. Since the pairwise score is a linear combination of valid likelihood score functions, then its unbiasedness follows under the usual regularity conditions. See also Fieuws and Verbeke [5] for a variant of this method.

### 3.2 Calculation of SEs of $\hat{\Theta}$

Following Renard et al. [26], the variances of the PL estimates can be obtained from the inverse of the Godambe information matrix $G(\Theta) = H(\Theta) J^{-1}(\Theta) H(\Theta)$ (also known as the sandwich information matrix; see, e.g. Song [30]), where

$$H(\Theta) = E \left\{ -\frac{\partial}{\partial \Theta} U_{p\ell}(\Theta) \right\} \quad \text{and} \quad J(\Theta) = E [ U_{p\ell}(\Theta) U_{p\ell}^T(\Theta) ] .$$
For clustered data, $H(\Theta)$ and $J(\Theta)$ are estimated by

$$
\hat{H}(\Theta) = H(\hat{\Theta}) = -\sum_{i=1}^{N} \frac{\partial}{\partial \Theta} U_{p\ell_i}(\Theta) \bigg|_{\Theta=\hat{\Theta}},
$$

$$
\hat{J}(\Theta) = J(\hat{\Theta}) = \sum_{i=1}^{N} U_{p\ell_i}(\Theta) U_{p\ell_i}^T(\Theta) \bigg|_{\Theta=\hat{\Theta}},
$$

where $U_{p\ell_i}(\Theta) = \partial p_{\ell_i}/\partial \Theta$ is the pairwise score equation for patient $i$. The SEs for the estimates in $\hat{\Theta}$ are obtained from the estimated covariance matrix $\hat{\Sigma}_\Theta = G^{-1}(\hat{\Theta}) = H^{-1}(\hat{\Theta}) J(\hat{\Theta}) H^{-1}(\hat{\Theta})$. Under standard regularity conditions, $\hat{\Theta}$ is consistent and has an asymptotic multivariate Gaussian distribution [2] with mean $\Theta$ and covariance matrix $G^{-1}(\Theta)$. The respective estimates $\hat{\text{Sen}}$, and $\hat{\text{Spc}}_v$ of $\text{Sen}$, and $\text{Spc}_v$ are directly obtained by the plug-in principle, and their SEs are given by the delta method. That is,

$$
\text{cov} \left( \begin{pmatrix} \hat{\text{Sen}}_v \\ \hat{\text{Spc}}_v \end{pmatrix} \right) = \begin{pmatrix} \frac{\partial}{\partial \Theta} \hat{\text{Sen}}_v \\ \frac{\partial}{\partial \Theta} \hat{\text{Spc}}_v \end{pmatrix} \hat{\Sigma}_\Theta \begin{pmatrix} \frac{\partial}{\partial \Theta} \hat{\text{Sen}}_v \\ \frac{\partial}{\partial \Theta} \hat{\text{Spc}}_v \end{pmatrix} \bigg|_{\Theta=\hat{\Theta}}.
$$

4. Simulation study

The goal of this simulation study is to examine the PL approach to estimation for the correlated probit model based on Equation (3) in terms of the finite-sample properties of the resulting estimates, including those for the sensitivities and specificities of the pathologies. Further, we also compare the estimates of pathology-specific sensitivities and specificities based on the PL method with those obtained by GEE, with a user-defined marginal working-correlation structure (similar to that for the correlated probit model) as well as with the ‘independence’ working-correlation matrix, which we refer to as the ‘crude’ method. The performance of the estimates $\hat{\theta}_g$ of $\theta_g$ were evaluated based on their relative bias $\text{RB} = 100 \times \{\text{mean of } \hat{\theta}_g - \theta_k\}/\theta_g$ and their relative efficiency $\text{RE} = \{\text{mean of SEs}\}/\{\text{empirical SD of } \hat{\theta}_g\}$, where SD denotes standard deviation. We also investigated the empirical coverage properties of the estimates by constructing the 95% confidence intervals (CIs) $\hat{\theta}_g \pm 1.96 \times \text{SE}(\hat{\theta}_g)$ for each parameter $\theta_g$, and obtaining the proportion of simulation repeats for which the corresponding CI captured $\theta_g$. For the simulations, we considered the case of two readers and two pathologies (i.e. $V = K = 2$). The underlying LMMs from which the data were simulated are given by

$$
Y_{ijkv}^* = \beta_{v0} + \beta_{v1} D_{ij} + B_{iv}^{(p)} + B_{ik}^{(r)} + B_{ij}^{(e)} + \epsilon_{ijkv}^*,
$$

with $\epsilon_{ijkv}^* \sim N(0, 1)$, and

$$
B_i^{(p)} = \begin{pmatrix} B_{i1}^{(p)} \\ B_{i2}^{(p)} \end{pmatrix} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma^{(p)} \right) = \begin{pmatrix} \sigma_{11}^{(p)} & \rho^{(p)} \sqrt{\sigma_{11}^{(p)} \sigma_{22}^{(p)}} \\ \rho^{(p)} \sqrt{\sigma_{11}^{(p)} \sigma_{22}^{(p)}} & \sigma_{22}^{(p)} \end{pmatrix},
$$

$$
B_i^{(r)} = \begin{pmatrix} B_{i1}^{(r)} \\ B_{i2}^{(r)} \end{pmatrix} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma^{(r)} \right) = \begin{pmatrix} 1 & \rho^{(r)} \\ \rho^{(r)} & 1 \end{pmatrix},
$$

where

$$
\rho^{(p)} = \rho^{(r)}, \quad \sigma_{11}^{(p)} = \sigma_{22}^{(p)}.
$$
effects. In the former, we marginally regress a reader’s diagnosis on the corresponding disease for models specified marginally; in contrast, GLMMs are specified conditionally, given random distributional assumption regarding the binocular data. Furthermore, GEEs work most naturally for the GEE approach, Sen

\[ \eta = R \] via the delta method. The GEE approach was implemented in \texttt{geepack} with independence working-correlation structure (i.e. \texttt{B})

estimates \[ [21,29] \]; these have the significant advantage over the model-based or naive estimate \[ (23) \] is discussed in appendix.

\[ \rho(\beta) \] were obtained using the GEE approach with independence working-correlation structure (i.e. \texttt{B})

Simulation results are presented in Table 1. These results generally suggest that PL estimates for the model perform quite well in finite samples. The RBs generally decreased with increasing sample size. These results appear to confirm those previously reported by Renard \textit{et al.} [26], who showed that PL estimates yield only slight efficiency loss compared to maximum likelihood estimates in random intercept models for clustered data. Overall, it appears that PL estimation provides reasonably good estimates with little efficiency loss.

We next compared the estimated pathology-specific sensitivities and specificities based on PL approach for the joint model based on Equation (23), against those from the crude method and from GEE, with a working-correlation structure similar to (but not the same as) the marginal correlation structure in the joint model; see appendix for details. The robust SEs of the GEE estimates of \( \beta = (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}) \) can be obtained using the so-called sandwich variance estimates [21,29]; these have the significant advantage over the model-based or naive estimate of being robust in the sense that they give consistent SEs even when the correlation structure is misspecified, provided that the marginal mean model is correctly specified. Note that, for the GEE approach, \( \text{Sen}_v = \Phi(\beta_0 + 1) \) and \( \text{Spc}_v = 1 - \Phi(\beta_0) \). Estimates of \( \text{Sen}_v \) and \( \text{Spc}_v \) are thus obtained directly by the plug-in method, and corresponding SEs are computed via the delta method. The GEE approach was implemented in \texttt{R} using the package \texttt{geese} in the \texttt{geepack} library [7]. Interestingly, the package \texttt{geese} can accommodate the user-defined working-correlation matrix \( \text{R} \), given in the appendix. Results from the crude method were obtained using the GEE approach with independence working-correlation structure (i.e. \( \text{R}_i \) is the identity matrix), with SEs calculated from the model-based variance estimate [19].

Note that the GEE method [29,31] is a non-likelihood-based approach that does not rely on any distributional assumption regarding the binocular data. Furthermore, GEEs work most naturally for models specified marginally; in contrast, GLMMs are specified conditionally, given random effects. In the former, we marginally regress a reader’s diagnosis on the corresponding disease.

\[ B^{(e)}_i = \left( B^{(e)}_{il}, B^{(e)}_{ir} \right) \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma^{(e)} = \begin{pmatrix} 1 & \rho^{(e)} \\ \rho^{(e)} & 1 \end{pmatrix} \right), \] (26)
Table 1. Mean of estimates (Ave), RB, RE, and CP of PL estimates for joint model based on Equation (23), with \( R = 500 \) repeated samples of sizes \( N = 50, 100, 200 \).

| Parameter | \( N = 50 \) | | | | \( N = 100 \) | | | | \( N = 200 \) | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| \( \beta_{10} = -1.25 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| -1.2820 | 2.5611 | 1.0310 | 95.2 | -1.2429 | -0.5715 | 1.0331 | 94.4 | -1.2467 | -0.2600 | 1.0247 | 96.0 |
| \( \beta_{20} = -1.25 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| -1.2919 | 3.3515 | 1.0028 | 95.4 | -1.2316 | -1.4735 | 0.9744 | 93.8 | -1.2369 | -1.0455 | 1.0362 | 95.3 |
| \( \beta_{11} = 4 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| 4.1280 | 3.2000 | 1.2048 | 96.8 | 3.9492 | -1.2710 | 1.0598 | 94.4 | 4.0015 | 0.0366 | 0.9847 | 94.5 |
| \( \beta_{22} = 4 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| 4.1180 | 2.9493 | 1.3568 | 99.0 | 3.9552 | -1.1195 | 1.0284 | 93.8 | 3.9668 | -0.8303 | 1.0127 | 93.8 |
| \( \rho(p) = 0.5 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| 0.5243 | 4.8604 | 1.1126 | 96.0 | 0.4918 | -1.6317 | 1.0471 | 94.8 | 0.5020 | 0.4030 | 1.0220 | 95.0 |
| \( \rho(r) = 0.5 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| 0.4445 | -11.0950 | 1.0676 | 97.0 | 0.5321 | 6.4170 | 1.0549 | 98.0 | 0.5204 | 4.0838 | 1.0172 | 93.8 |
| \( \rho(e) = 0.5 \) | | | | | | | | | | | | | |
| Ave | RB | RE | CP | Ave | RB | RE | CP | Ave | RB | RE | CP |
| 0.4762 | -4.7672 | 1.1480 | 96.4 | 0.5205 | 4.1046 | 1.0852 | 96.0 | 0.5166 | 3.3229 | 1.0044 | 95.3 |
Table 2. Mean of estimates (Ave), RB, RE, and CP of estimated sensitivities $\text{Sen}_v$ and specificities $\text{Spc}_v$ from joint model based on Equation (23), estimated via PL, and from a marginal probit model, estimated via the crude method and GEE, with (marginal) working correlation structure similar to that in joint model based on Equation (23).

| Method          | Parameter | $N = 50$       |          |          | $N = 100$       |          |          | $N = 200$       |          |
|-----------------|-----------|----------------|----------|----------|----------------|----------|----------|----------------|----------|
|                 |           | Ave  | RB    | RE     | CP     | Ave  | RB    | RE     | CP     | Ave  | RB    | RE     | CP     |
| Model (23)      | $\text{Sen}_1 = 0.9154$ | 0.9161 | 0.0754 | 1.0705 | 94.6 | 0.9161 | 0.0739 | 1.0116 | 95.5 | 0.9138 | -0.1714 | 0.9725 | 94.8 |
|                 | $\text{Sen}_2 = 0.9154$ | 0.9138 | -0.1696 | 1.0446 | 94.4 | 0.9147 | -0.0712 | 1.0266 | 94.8 | 0.9156 | 0.0188 | 1.0528 | 95.5 |
|                 | $\text{Spc}_1 = 0.7340$ | 0.7360 | 0.2679 | 1.0279 | 96.0 | 0.7335 | -0.0624 | 1.0220 | 94.6 | 0.7337 | -0.0417 | 1.0502 | 95.7 |
|                 | $\text{Spc}_2 = 0.7340$ | 0.7377 | 0.5037 | 0.9933 | 95.0 | 0.7355 | 0.2107 | 0.9697 | 94.2 | 0.7353 | 0.1787 | 1.0250 | 94.8 |
| Crude method    | $\text{Sen}_1 = 0.9154$ | 0.9160 | 0.0629 | 1.3225 | 95.2 | 0.9162 | 0.0858 | 1.2600 | 95.6 | 0.9138 | -0.1701 | 1.2224 | 94.5 |
|                 | $\text{Sen}_2 = 0.9154$ | 0.9137 | -0.1850 | 1.2809 | 95.0 | 0.9147 | -0.0792 | 1.2813 | 94.8 | 0.9155 | 0.0074 | 1.3232 | 95.5 |
|                 | $\text{Spc}_1 = 0.7340$ | 0.7362 | 0.2953 | 1.3212 | 95.8 | 0.7336 | -0.0553 | 1.3331 | 94.2 | 0.7338 | -0.0329 | 1.3849 | 96.0 |
|                 | $\text{Spc}_2 = 0.7340$ | 0.7374 | 0.4585 | 1.2898 | 95.5 | 0.7354 | 0.1901 | 1.2750 | 95.2 | 0.7351 | 0.1457 | 1.3469 | 95.3 |
| GEE             | $\text{Sen}_1 = 0.9154$ | 0.9160 | 0.0628 | 1.0564 | 94.8 | 0.9160 | 0.0624 | 1.0097 | 95.8 | 0.9147 | -0.0805 | 1.0416 | 94.8 |
|                 | $\text{Sen}_2 = 0.9154$ | 0.9142 | -0.1304 | 1.0427 | 94.8 | 0.9150 | -0.0424 | 0.9721 | 94.2 | 0.9148 | -0.0629 | 1.0184 | 95.3 |
|                 | $\text{Spc}_1 = 0.7340$ | 0.7361 | 0.2873 | 1.0392 | 96.0 | 0.7333 | -0.1007 | 0.9756 | 94.8 | 0.7324 | -0.2123 | 1.0214 | 95.3 |
|                 | $\text{Spc}_2 = 0.7340$ | 0.7381 | 0.5544 | 0.9996 | 95.4 | 0.7339 | 0.2651 | 1.0286 | 95.0 | 0.7347 | 0.1015 | 1.0402 | 95.3 |

Note: Here, $N = 50, 100, 200$ and $R = 500$. 
status, using the probit link function, as \( P(Y_{ijkv} = 1; D_{ijv}) = \Phi^{-1}(\beta_{v0} + \beta_{v1}D_{ijv}) \), for \( j = L, R \), \( k = 1, 2 \), \( v = 1, 2 \) and \( i = 1, \ldots, N \). Note that the estimates from the marginal model (under GEE) and those from a GLMM such as Equation (23), are different, since the marginal probit model in Equation (5) obtained by averaging Equation (23) with respect to the random effects is not the same marginal probit model under GEE. As a consequence, estimates obtained under GEE (with a marginal probit model) and under the joint model based on Equation (23) are different and not comparable (i.e. \( \beta_{v0} \) and \( \beta_{v1} \) are different for the two models). However, the pathology-specific sensitivities and specificities are comparable.

Results are presented in Table 2. Observe that the estimates from the three approaches are relatively close to the true values, with CP all close to the nominal 95% level. This is not surprising in light of the robustness of GEE estimates to mis specification of the marginal covariance structure. Not surprisingly, a significant efficiency loss can be observed for estimates from the crude method. The implication of ignoring the between-eyes, between-readers and between-pathologies correlations is thus clear: failure to account for these correlations in any statistical analysis may lead to potentially incorrect inferences.

Note that while the GEE estimates compare favorably with the PL estimates for the joint model, this is accomplished with a careful specification of the working-correlation matrix \( R \), which is hardly known in practice. Moreover, this necessitates the specification of an \( 8 \times 8 \) matrix; in general, the GEE method requires specifying an \( M \times M \) working-correlation matrix, which renders the GEE method infeasible in applications involving many readers (i.e. large \( K \)) and many pathologies (i.e. large \( V \)).

5. Application to diabetic retinopathy data

We now illustrate the proposed methodology described in Section 2 on data from the diabetic retinopathy study [16,28] described in Section 1. The study involved \( N = 94 \) diabetic patients in Alberta, Canada, who were referred to a comprehensive retina practice in Edmonton. The study protocol required that patients be clinically examined on the same day they underwent digital photography by a trained ophthalmic photographer using a high-resolution digital camera. The digital images were stored uncompressed and then graded by experienced ophthalmologists 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 and right eye images to minimize reader recall. In order to evaluate treatable diabetic retinopathy among the patients, a number of pathologies were identified as either present (positive) or absent (negative). The pathologies identified included CSME, microaneurysms, intra-retinal haemorrhage, HEX, and other diseases of note. Contact lens biomicroscopy, the clinical examination considered to be the ‘gold standard’ for most, but not all, of the pathologies considered, was performed on all patients by retinal specialists to determine disease status. Digital images of the patients’ eyes were graded by two ophthalmologists and patients were diagnosed as either positive or negative for the pathologies.

In what follows, we consider the pathologies CSME and HEX for which ‘gold standard’ values are available. The summarized data concerning the presence (+) or absence (−) of CSME and HEX in the left and right eyes of the patients as evaluated by two readers along with the true disease status are presented in Table 3. Pathology CSME pertains to the thickening and swelling of an eye’s macula due to fluid and protein deposits while pathology HEX involves the leakage of fluid and lipoprotein into the retina of the eye.

With \( K = V = 2 \), we adopt the same joint model from Equation (23), with only random intercepts for the pathologies. Furthermore, to ensure that the estimate of \( \Sigma^{(p)} \) is positive definite, we maximize the pairwise log-likelihood function indirectly with respect to the Cholesky factors of \( \Sigma^{(p)} \) instead of its direct elements [33]; estimates of its direct elements are then obtained by
Table 3. Evaluations by Reader 1 and Reader 2 (in parentheses; i.e. $K = 2$) of the presence (+) or absence (−) of pathologies CSME and HEX (i.e. $V = 2$) in left ($L$) and right ($R$) eyes of $N = 94$ diabetic patients.

| Disease status | (+, +)  | (−, +)  | (+, −)  | (−, −)  | Total     |
|----------------|---------|---------|---------|---------|-----------|
| (+, +)         | 17(19)/24(26) | 0(1)/0(0) | 1(2)/3(3) | 1(0)/1(1) | 19(22)/28(30) |
| (−, +)         | 3(2)/3(4)     | 3(5)/3(3) | 0(0)/1(2) | 4(5)/3(4) | 10(12)/10(13) |
| (+, −)         | 1(1)/2(0)     | 0(0)/0(0) | 3(5)/3(3) | 2(2)/2(3) | 6(8)/7(6)   |
| (−, −)         | 2(1)/2(1)     | 4(1)/3(3) | 3(0)/2(1) | 50(50)/42(40) | 59(52)/49(45) |
| Total          | 23(23)/31(31) | 7(7)/6(6) | 7(7)/9(9) | 57(57)/48(48) | 94         |

Note: Here, (+, +), for example, indicates ($L, R$) = (+, +).

Table 4. PL estimates and their SEs for pathologies CSME ($v = 1$) and HEX ($v = 2$) with $K = 2$ readers for joint model based on Equation (23) with $N = 94$.

| Effect          | Parameter | Est  | SE    | $z$     |
|-----------------|-----------|------|-------|---------|
| CSME Intercept  | $\beta_{10}$ | −2.9975 | 0.6403 | −4.6814 |
| Disease status  | $\beta_{11}$ | 4.8259 | 0.9512 | 5.0735  |
| HEX Intercept   | $\beta_{20}$ | −2.4756 | 0.5689 | −4.516  |
| Disease status  | $\beta_{21}$ | 4.3992 | 0.9479 | 4.6410  |
| Correlations    | Between-pathologies $\rho^{(p)}$ | 0.4006 | 0.4325 | −       |
|                 | Between-readers $\rho^{(r)}$     | 0.9937 | 0.0104 | −       |
|                 | Between-eyes $\rho^{(e)}$        | −0.7219 | 0.7709 | −       |
| Variance components | CSME $\sigma_{11}^{(p)}$ | 1.2413 | 0.6442 | 1.9269  |
|                 | HEX $\sigma_{22}^{(p)}$          | 1.2226 | 0.7314 | 1.6764  |

The table displays the estimates and their (large-sample) SEs based on the PL method. The results for the corresponding accuracy measures are given in Table 5, both based on the PL approach and the GEE method with three different working-correlation structures, namely, independence (which is equivalent to the crude method), unstructured and that based on the marginal correlation structure of the joint model. Note that the estimates for sensitivities and specificities are slightly different according to the working-correlation structure assumed under GEE.

From Table 5, the crude method estimates for sensitivities and specificities are slightly closer to those from the PL method, with the discrepancy between the estimates lying within 2%, a negligible difference in practice. The other two sets of estimates (i.e. GEE with unstructured and structured working correlations) are considerably different from those from the PL method, with a maximum discrepancy between the estimates of at most 8%.

Comparing the individual values for the accuracy measures, we note that the sensitivity of HEX is slightly higher than that of CSME, while the reverse is true for the specificities. The estimated sensitivities and specificities for both pathologies range between 73% and 93%. It should be noted that the crude method yields comparatively smaller SEs. This can be misleading since
Table 5. PL estimates for joint model based on Equation (23) of sensitivities and specificities and their SEs for pathologies CSME ($v = 1$) and HEX ($v = 2$) along with those obtained by GEE, with independence (i.e. crude method), unstructured, and structured working correlations (similar to that in joint model based on Equation (23)).

| Measure | Model (23) | Independence | Unstructured | Structured |
|---------|------------|--------------|--------------|------------|
|         | Est | SE  | z    | Est | SE  | z    | Est | SE  | z    |
| Sen1    | 0.8045 | 0.0532 | 15.1222 | 0.7260 | 0.0169 | 42.9586 | 0.7376 | 0.0166 | 44.4337 |
| Sen2    | 0.8176 | 0.0464 | 17.6207 | 0.7390 | 0.0221 | 32.6244 | 0.7557 | 0.0233 | 32.4334 |
| Spc1    | 0.9202 | 0.0199 | 46.2412 | 0.8917 | 0.0210 | 42.4619 | 0.8413 | 0.0308 | 27.3149 |
| Spc2    | 0.8784 | 0.0313 | 28.0639 | 0.8413 | 0.0308 | 27.3149 | 0.8473 | 0.0334 | 25.3683 |

Table 6. Estimated marginal tetrachoric correlations and their SEs for joint model based on Equation (23) for pathologies CSME ($v = 1$) and HEX ($v = 2$).

| Correlation | Est | SE  |
|-------------|-----|-----|
| CSME        |     |     |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk1}^*)$ | 0.7784 | 0.0777 |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk1}^*)$ | 0.4006 | 0.1571 |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk1}^*)$ | 0.3992 | 0.2126 |
| HEX         |     |     |
| $\text{corr}(Y_{ijk2}^*, Y_{ijk2}^*)$ | 0.7765 | 0.0874 |
| $\text{corr}(Y_{ijk2}^*, Y_{ijk2}^*)$ | 0.3956 | 0.1743 |
| $\text{corr}(Y_{ijk2}^*, Y_{ijk2}^*)$ | 0.3942 | 0.1735 |
| Between     |     |     |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk2}^*)$ | 0.5771 | 0.1006 |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk2}^*)$ | 0.5757 | 0.0999 |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk2}^*)$ | 0.1963 | 0.1680 |
| $\text{corr}(Y_{ijk1}^*, Y_{ijk2}^*)$ | 0.1949 | 0.1677 |

correlated observations contain less information than independent data, and when the former are incorrectly assumed to be the latter, underestimation of SEs may result, possibly leading to large Wald test statistics and inflated type I error rates [27]. This is evident from Table 2.

Estimates of the marginal tetrachoric correlations from the joint model based on Equation (23) are shown in Table 6. Note that these correlations pertain to the correlations between the underlying latent variables. The SEs of the estimated marginal correlations were computed via the delta method. Not surprisingly, the largest estimated correlations correspond to the case with different readers diagnosing the same eye for the same pathology, given by 0.7784 for CSME and 0.7765 for HEX. This implies that the between-readers agreement is quite strong. The estimated correlation 0.5771 between CSME and HEX for the same reader and the same eye is moderate (i.e. 0.4) while that between fellow eyes for the same reader and the same pathology is not.

Finally, note that the estimated marginal tetrachoric correlations are all positive although the correlation between the eye-specific random effects is negative (see Table 4).

6. Discussion

This paper focused on an important issue arising in reader-based multi-disease binocular diagnostic studies: how to account for correlations between readers’ diagnoses and between diseases/
pathologies, while at the same time incorporating the intrinsic correlation between patients’ fellow eyes. The general approach taken in the paper was a model-based one that relies on specifying a multivariate probit model for the joint distribution of the binary outcomes. The latent variable formulation of the binary diagnostic data should be appealing to ophthalmologists, as it provides a natural description of the biological process underlying retinopathy-related pathologies. A PL approach is adopted to overcome the computational complexities in computing multivariate orthant probabilities required in a full likelihood analysis of the probit model. Estimates derived from the approach are shown, empirically via simulations, to be more efficient than those from the crude method. While GEE estimates yielded comparable performance in terms of bias and efficiency, they require the specification of a working-correlation matrix whose dimension increases with the numbers of readers and pathologies. This can become impractical, even infeasible, in practice.

The methodology outlined in the paper is illustrated with data from a study concerning the diagnoses of two retinopathy-related pathologies among diabetic patients as evaluated by two ophthalmologists based on digital images of the eyes. Compared to GEE, our model has the advantage of being able to delineate inferences about various correlations – such as the binocular correlation between fellow eyes, the same-eye/different-pathologies correlation and the different-readers/same-pathology/same-eye correlation – without the need to specify a large working correlation matrix. Such a matrix can be cumbersome to construct when many pathologies and many readers exist.

Acknowledgments

This work was partially supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Part of this work was carried out at the Department of Statistics, Middle East Technical University (METU), while A.R. de Leon was a TÜBİTAK Visiting Scientist. He thanks the faculty, students and staff of METU, especially his host Dr Özlem İlkg-Dağ, for their hospitality.

Note

1. Supplementary Content may be viewed online at http://dx.doi.org/10.1080/02664763.2014.909790.

References

[1] B.C. Arnold and D. Strauss, Pseudolikelihood estimation: Some examples, Sankhya–B 53 (1991), pp. 233–243.
[2] D.D. Boos, On generalized score tests, Am. Stat. 46 (1992), pp. 327–333.
[3] D.R. Cox and N. Reid, A note on pseudolikelihood constructed from marginal densities, Biometrika 91 (2004), pp. 729–737.
[4] A. Donner and N. Klar, Confidence interval construction for effect measures arising from cluster randomization trials, J. Clin. Epidemiol. 46 (1993), pp. 123–131.
[5] S. Fieuws and G. Verbeke, Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles, Biometrics 62 (2006), pp. 424–431.
[6] R.J. Glynn and B. Rosner, Accounting for correlation between fellow eyes in regression analysis, Arch. Ophthalmol. 110 (1992), pp. 381–387.
[7] S. Højsgaard, U. Halekoh, and J. Yan, The R Package geepack for generalized estimating equations, J. Stat. Softw. 15 (2006), pp. 1–11.
[8] C.K. Hsiao, P.-C. Chen, and W.H. Kao, Bayesian random effects for interrater and test–retest reliability with nested clinical observations, J. Clin. Epidemiol. 64 (2011), pp. 808–814.
[9] S.L. Hui and S.D. Walter, Estimating the error rates of diagnostic tests, Biometrics 36 (1980), pp. 167–171.
[10] S.L. Hui and X.-H. Zhou, Evaluation of diagnostic tests without a gold standard, Stat. Methods Med. Res. 7 (1998), pp. 354–370.
[11] P.P. Hujoel, L.H. Moulton, and W.J. Loesche, Estimation of sensitivity and specificity of site-specific diagnostic tests, J. Periodontal Res. 25 (1990), pp. 193–196.
[12] E.W. Lee and N. Dubin, Estimation and sample size considerations for clustered binary responses, Stat. Med. 13 (1994), pp. 1241–1252.
[13] M. Lefkopoulou, D. Moore, and L. Ryan, The analysis of multiple correlated binary outcomes, J. Am. Stat. Assoc. 84 (1989), pp. 810–815.
[14] W. Leisenring, M.S. Pepe, and G. Longton, A marginal regression modelling framework for evaluating medical diagnostic tests, Stat. Med. 16 (1997), pp. 1263–1281.
[15] M.L.C. Leite and A. Nicolosi, Statistical analysis of correlated binary data in ophthalmology: A weighted logistic regression approach, Ophthal. Epidemiol. 5 (1998), pp. 117–131.
[16] A.R. de Leon, M. Guo, G. Singh, and C.J. Rudnisky, A likelihood approach to estimating sensitivity and specificity for binocular diagnostic data: Application in ophthalmology, Stat. Med. 26 (2007), pp. 3300–3314.
[17] A.R. de Leon, A. Soo, D. Bonzo, and C.J. Rudnisky, Joint estimation of diagnostic accuracy measures for paired organs – Application in ophthalmology, Biom. J. 51 (2009), pp. 837–850.
[18] A.R. de Leon and Y. Zhu, ANOVA extensions for mixed discrete and continuous data, Comput. Stat. Data Anal. 52 (2008), pp. 2218–2227.
[19] K.Y. Liang and S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986), pp. 13–22.
[20] D. Maberley, H. Walker, A. Koushnik, and A. Cruess, Screening for diabetic retinopathy in James Bay, Ontario: A cost-effectiveness analysis, Can. Med. Assoc. J. 168 (2003), pp. 160–164.
[21] G. Molenberghs and G. Verbeke, Models for Discrete Longitudinal Data, Springer, New York, 2005.
[22] P.C. O’Brien, Procedures for comparing multiple endpoints, Biometrics 40 (1984), pp. 1079–1087.
[23] Y. Qu, M. Tan, and M.H. Kutner, Random effects models in latent class analysis for evaluating accuracy of diagnostic tests, Biometrics 52 (1996), pp. 797–810.
[24] J.N.K. Rao and A.J. Scott, A simple method for the analysis of clustered binary data, Biometrics 48 (1992), pp. 577–585.
[25] D. Renard, H. Geys, G. Molenberghs, T. Burzykowski, and M. Buyse, Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes, Biom. J. 44 (2002), pp. 921–935.
[26] D. Renard, G. Molenberghs, and H. Geys, A pairwise likelihood approach to estimation in multilevel probit models, Comput. Stat. Data Anal. 44 (2004), pp. 649–667.
[27] B. Rosner, Multivariate methods for clustered binary data with more than one level of nesting, J. Am. Stat. Assoc. 84 (1989), pp. 373–380.
[28] C.J. Rudnisky, B.J. Hinz, M.T.S. Tennant, A.R. de Leon, and M.D.J. Greve, High-resolution stereoscopic digital fundus photography versus contact-lens biomicroscopy for the detection of clinically significant macular edema, Ophthalmology 109 (2002), pp. 267–274.
[29] P.J. Smith and A. Hadgu, Sensitivity and specificity for correlated observations, Stat. Med. 11 (1992), pp. 1503–1509.
[30] P.X.-K. Song, Correlated Data Analysis: Modeling, Analytics and Applications, Springer, New York, 2007.
[31] M.R. Sternberg and A. Hadgu, A GEE approach to estimating sensitivity and specificity and coverage properties of the confidence intervals, Stat. Med. 20 (2001), pp. 1529–1539.
[32] B.C. Sutradhar and K. Das, Generalized linear models for beta correlated binary longitudinal data, Commun. Stat. Theory Methods 26 (1997), pp. 617–635.
[33] D. Todem, K. Kim, and E. Lesaffre, Latent-variable models for longitudinal data with bivariate ordinal outcomes, Stat. Med. 26 (2007), pp. 1034–1054.
[34] Z. Wang and X.-H. Zhou, Random effects models for assessing diagnostic accuracy of traditional Chinese doctors in absence of a gold standard, Stat. Med. 31 (2012), pp. 661–671.
[35] H. Xu and B.A. Craig, Likelihood analysis of multivariate probit models using a parameter expanded MCEM algorithm, Technometrics 52 (2010), pp. 340–348.
[36] X.-H. Zhou, P. Castelluccio, and C. Zhou, Nonparametric estimation of ROC curves in the absence of a gold standard, Biometrics 61 (2005), pp. 600–609.