THE UNIVERSITY OF CALGARY

A Likelihood-Based Approach to Estimating Sensitivity and Specificity with Binocular Diagnostic Data—Application in Ophthalmology

by

Meijie Guo

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

NOVEMBER, 2005

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "A Likelihood-Based Approach to Estimating Sensitivity and Specificity with Binocular Diagnostic Data – Application in Ophthalmology " submitted by Meijie Guo in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

Supervisor, Dr. A.R. de Leon
Department of Mathematics and Statistics

Dr. X. Lu
Department of Mathematics and Statistics

Dr. R.J. Hilsden
Department of Community Health Sciences

Nov. 1, 2005

Date

Abstract

Binocular data typically arise in ophthalmology, where pairs of eyes are screened, through some diagnostic procedure, for the presence of certain diseases or pathologies. Treating the eyes as independent and adopting the usual approach in estimating the sensitivity and specificity of a diagnostic test ignores the correlation between the eyes, and may consequently yield incorrect estimates, especially of the standard errors.

This thesis 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 model is assessed both theoretically and by 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 analyzed to illustrate the methodology.

Acknowledgements

First of all, I would like to sincerely thank my supervisor, Dr. Alexander R. de Leon, for his encouragement, stimulating suggestions, and support during the research. He is not only an excellent statistician with deep vision but also, and most importantly, a kind person. I am glad to have worked with him. Without his time and effort, this thesis would not have been such a success.

I would also like to express my gratitude to the examiners of the thesis, Drs. Xuewen Lu and Robert J. Hilsden, for their comments and suggestions.

As always, it is impossible to mention everybody who had an impact on this work. I would like to thank everyone who had supported me throughout the course of the thesis, particularly, James Xu, Jia Shen, Lei Xiong, Hua Li, and Jian Kang.

Finally, my special thanks to my husband, Yongsheng Zhang, for his patience, tolerance, and spiritual support during those busy days.

For

Meixia Guo,

Mr. & Mrs. Xuqiang Zhao,

Xuming Zhao,

Ping Zhang,

and

all those who have supported me in my study

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of the Thesis

While early treatment of diseases contribute significantly to controlling health care costs, many of the diagnostic tests that allow for early disease detection can be enormously expensive, preventing health care providers from adopting them on a routine basis. However, with recent advances in technology, more and more cost-effective diagnostic testing procedures have been made available to practitioners. The evaluation of the accuracy of these methods has become a major undertaking for medical researchers, as inaccurate diagnoses can be financially disastrous.

This thesis is motivated by a study conducted in Alberta, Canada, on the use of high-resolution digital photography to screen diabetic patients for treatable retinopathy. Screening is a special case of diagnostic testing that enables medical practitioners to detect diseases at their early stages before patients manifest full-blown clinical symptoms. The popularity of screening programmes has led to improved early intervention for and treatment of diseases resulting in substantial health care cost savings. In countries like Canada, where distances are great, the cost of travel necessitates screening at a distance, thereby allowing only those patients in need of treatment to travel to a specialist. A teleophthalmology system allowing for distance screening of diabetic retinopathy based on digital images of patients' eyes is a potentially cost-effective alternative to clinical examination. The purpose of the study was thus to

evaluate whether diabetic retinopathy can be identified with high-resolution digital photography and whether this identification correlates well with the accepted gold standard of clinical examination.

We summarize in section 1.2 the commonly used methodologies for investigating the accuracies of binary diagnostic tests. Because patients in many efficacy studies may contribute several test results, as when both eyes are screened for retinopathy, there is a need to adopt appropriate methodologies that account for the intra-patient correlation. A number of regression techniques have recently been developed for correlated diagnostic data arising from multiple tests on the same patient. We review these regression approaches for clustered binary data in section 1.2. We also describe in section 1.3 data from a diabetic retinopathy study (Rudnisky et al., 2002) used to illustrate the methods developed in the thesis. The chapter concludes with a brief description and overview of the thesis.

1.2 Review of Literature

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 (Zhou et al., 2002).

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.

In the diabetic retinopathy study, for example, digital images of both left and right eyes of patients are screened for retinopathy. The binocular structure of the data impacts on the analysis, as an eye tends to have a greater correspondence with the fellow eye than with eyes of another patient. While it is possible to estimate a test's sensitivity and specificity on an eye-specific basis, thereby effectively ignoring the inter-eye correlation, incorrect inferences are likely to result from underestimated standard errors (Glynn and Rosner, 1992).

There has been previous work on the estimation of sensitivity and specificity and their standard errors in the context of clustered binary diagnostic data. These include simple adjustments to standard errors introduced by Rao and Scott (1992) and Donner and Klar (1993) to account for the intra-cluster correlation, and a weighted estimator proposed by Lee and Dubin (1994) for handling unbalanced cluster sizes. A similar approach based on weighting was recently discussed by Leite and Nicolosi (1998) in the context of logistic regression analysis of binocular ophthalmologic data.

Smith and Hadgu (1992) described a regression method based on the generalized estimating equations (GEE) approach (Liang and Zeger, 1986) to deal with clustered binary diagnostic data. Ahn (1997) reported that for moderate to large samples,

and moderate intra-cluster correlation, Smith and Hadgu's (1992) GEE estimator outperforms those proposed by Lee and Dubin (1994), Donner and Klar (1993), and Rao and Scott (1992).

A number of likelihood-based approaches have also been proposed in the literature for handling clustered binary data. Hujoel et al. (1990) adopted a commonly used correlated binomial model (Prentice, 1988; Bahadur, 1961) and applied it to model oral site-specific outcomes in a periodontal disease diagnostic study. Bonney (1987) considered a regressive logistic model for ordered binary data, such as those encountered in longitudinal studies. Qu et al. (1988) and Connolly and Liang (1988) introduced a class of conditional logistic regression models, which includes the polychotomous logistic regression model of Rosner (1984) as a special case. This was later extended by Rosner (1989) to the case of clustered binary data with several levels of nesting. See also Qaqish and Liang (1992) and Lefkopoulou et al. (1989).

More recently, Betensky and Whittemore (1996) generalized the quadratic exponential model (Zhao and Prentice, 1990) to analyze clustered multivariate binary data on familial cancers of the ovary and breast.

Several authors have also considered the beta-binomial model (Haseman and Kupper, 1979) for correlated binary data. Prentice (1986) modeled the joint distribution of correlated binary data in the presence of covariates using a generalization of the beta-binomial model. Sutradhar and Das (1997) proposed another generalization for multivariate longitudinal binary data based on generalized linear models.

Prentice (1988) gave a simple joint distribution for binocular and paired binary data which is completely determined by specification of its marginal probabilities and correlation. The model was first given by Bahadur (1961) and was recently used

by Sutradhar and Sutradhar (2001) in the context of classification. Lipsitz et al. (1990) likewise proposed maximum likelihood methods for analyzing paired binary data.

A likelihood-based approach based on Prentice's (1988) model is developed in the thesis to estimate the sensitivity and specificity of binocular binary diagnostic tests. The model has a simple form which is completely specified by the marginal probabilities of the binary outcomes and their correlation, and has a neat extension to the multi-reader multi-disease setting. Because of this, straightforward likelihood-based estimation and inference readily apply. While Zhao and Prentice's (1990) quadratic exponential model can be used as well to model the joint distribution of the binocular binary outcomes, their model considers conditional log-odds ratios as measures of associations and requires the computation of a normalizing constant. The number of odds ratios and the computational demands of calculating the normalizing constant can be prohibitive when the numbers of readers and diseases are large, so that a fully-likelihood based approach becomes infeasible. In contrast, the model we adopt in the thesis can reasonably easily handle this case via likelihood theory. The model of Prentice (1988) is also closely linked to the beta-binomial model.

The proposed method is applied to data from the diabetic retinopathy study (Rudnisky et al., 2002), which is discussed in the next section.

1.3 Diabetic Retinopathy Data

The study involved about a hundred diabetic patients in Alberta, Canada, who were referred to a comprehensive retina practice in Edmonton. The study protocol re-

quired 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 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 screen for treatable diabetic retinopathy among the patients, several pathologies that are indicative of retinal thickening were identified as either present (positive) or absent (negative). The pathologies considered included clinically significant macular edema (CSME), microaneurysms, intra-retinal hemorrhage, 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 at least two specialists and patients were diagnosed as either positive or negative for the pathologies.

Table 1.1: Data Set-up for One Pathology and Two Readers

	Reader 1		Reader 2	
Patient	left eye	right eye	left eye	right eye
1	Y_{1L1}	Y_{1R1}	Y_{1L2}	Y_{1R2}
2	Y_{2L1}	Y_{2R1}	Y_{2L2}	Y_{2R2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	Y_{NL1}	Y_{NR1}	Y_{NL2}	Y_{NR2}

The data set-up for a single pathology with two readers is displayed in Table 1.1, where Y_{iLk} and Y_{iRk} denote the binary test results for the left and right eyes of patient $i = 1, \cdots, N$, respectively, as graded by reader $k = 1, 2$. The data set-up for the case of two pathologies with two readers is presented in Table 1.2, where Y_{iLkv} and Y_{iRkv} represent the binary test results for the left and right eyes of patient $i = 1, \cdots, N$, respectively, as graded by reader $k = 1, 2$ for pathology $v = 1, 2$. The design can be considered as a full paired-patient-paired-reader design, whereby all digital images of patients' left and right eyes underwent grading by every reader.

Table 1.2: Data Set-up for Two Pathologies and Two Readers

	Reader 1		Reader 2	
Patient	left eye	right eye	left eye	right eye
1	Y_{1L11}	Y_{1R11}	Y_{1L21}	Y_{1R21}
	Y_{1L12}	Y_{1R12}	Y_{1L22}	Y_{1R22}
2	Y_{2L11}	Y_{2R11}	Y_{2L21}	Y_{2R21}
	Y_{2L12}	Y_{2R12}	Y_{2L22}	Y_{2R22}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	Y_{NL11}	Y_{NR11}	Y_{NL21}	Y_{NR21}
	Y_{NL12}	Y_{NR12}	Y_{NL22}	Y_{NR22}

The objective of the study was to compare digital photography to CLBM in the screening of treatable retinopathy among diabetic patients. While digital photography provides a cost-effective distance screening system, adequate identification of potentially treatable retinopathy should be ensured before widespread implementation.

Data from the study are used to illustrate the proposed methods for estimating sensitivity and specificity for binocular binary diagnostic data.

1.4 Overview of the Thesis

The objective of this thesis is to develop a likelihood-based approach for analyzing binocular data in the context of diagnostic testing.

Chapter 2 describes a model for the joint distribution of single-reader binocular diagnostic data and discusses maximum likelihood estimation of sensitivity and specificity along with their standard errors. Strengths as well as limitations of the model are highlighted and its performance relative to the crude approach, which ignores the intra-pair correlation, is investigated. The methodology is applied to data from the diabetic retinopathy study to illustrate its utility.

Chapter 3 gives an extension of the model to the case of multiple readers. The extended model introduces a random effect to account for the inter-reader correlation. Likelihood estimation is discussed and an extension to the case of multiple pathologies is outlined. The methodology is again illustrated using data from the diabetic retinopathy study.

Finally, a summary of the results of the thesis is presented in Chapter 4. Promising areas for future research are identified as well.

Chapter 2

Analysis of Binocular Diagnostic Data: Case of Single Reader

2.1 Introduction

Binocular binary diagnostic data typically arise in ophthalmology, where pairs of eyes are screened, through some diagnostic test, for the presence of certain abnormalities or pathologies. In the diabetic retinopathy study described in Chapter 1, for example, the diagnostic test involves digital images of patients' eyes that are 'diagnosed' by a trained reader for certain pathologies. Treating the eyes as independent 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 (Leite and Nicolosi, 1998).

In this chapter, we focus on the simplest case of one pathology and one reader and develop a model (Prentice, 1988) for the joint distribution of the binocular binary diagnostic data to estimate the sensitivity and specificity of the diagnostic test. The model is completely determined by its marginal probabilities and the inter-eye correlation, allowing for ordinary likelihood-based estimation in arbitrary parametric settings.

We also propose a marginal regression approach to model the sensitivity and specificity of the binary diagnostic test (Leisenring et al., 1997; Smith and Hadgu,

1992). Consider a binary test that screens for the presence (positive) or absence (negative) in a patient's left and right eyes for a pathology. Define Y_{iL} and Y_{iR} as the respective binary variables representing the test results for the left and right eyes, with $Y_{ij} = 1$ or 0 according as the result is positive or negative, $j = L, R$. For patient i, let D_{ij} and $\mathbf{x}_{ij}$ denote disease status, as determined by the gold standard, and covariates for eye $j = L, R$, respectively. A general marginal regression model for sensitivity π_{0j} and specificity π_{1j} based on eye j is then

$$\mathrm{P}\left(Y_{ij} = 1 | D_{ij}, \mathbf{x}_{ij}\right) = g(\beta, D_{ij}, \mathbf{x}_{ij}) = \begin{cases} \text{sensitivity} & \text{if } D_{ij} = 1 \\ 1 - \text{specificity} & \text{if } D_{ij} = 0 \end{cases},$$

where β is a vector of regression coefficients. This marginal regression model can now be embedded into a model for the joint distribution of Y_{iL} and Y_{iR} to facilitate a full likelihood-based approach to estimation and inference. We define such a model in the next section.

2.2 Single-Reader Model for Binocular Diagnostic Data

If ρ 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

$$\mathrm{P}(Y_L = y_L, Y_R = y_R) = \left[1 + \frac{\rho(y_L - p_L)(y_R - p_R)}{\sqrt{p_L q_L p_R q_R}}\right] \prod_{j=L,R} p_j^{y_j} q_j^{1-y_j} \qquad (2.1)$$

where $q_j = 1 - p_j$ (Prentice, 1988). Note that model (2.1) arises from Bahadur's model (Aerts et al., 2002; Bahadur, 1961) and is completely specified by the marginal

probabilities $p_j = \mathrm{P}(Y_j = 1)$, $j = L, R$, and the inter-eye correlation ρ . To ensure that (2.1) is a proper joint probability distribution, the inter-eye correlation ρ needs to satisfy the following restriction:

$$\max\left\{-\sqrt{\frac{p_L p_R}{q_L q_R}}, -\sqrt{\frac{q_L q_R}{p_L p_R}}\right\} < \rho < \min\left\{\sqrt{\frac{p_R q_L}{p_L q_R}}, \sqrt{\frac{p_L q_R}{p_R q_R}}\right\}. \tag{2.2}$$

From (2.1), 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 follow a Bernoulli distribution with success probability p_j, $j = L, R$. Thus, (2.1) generalizes the independent Bernoulli model to the case of binocular binary outcomes.

For eye $j = L, R$, we define a generalized linear model for p_j with covariates D_j and $\mathbf{x}_j$ as

$$\begin{aligned} p_j &= p(D_j, \mathbf{x}_j) \\ &= h^{-1}(\beta_0 + \beta_1 D_j + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j), \end{aligned} \tag{2.3}$$

with $h(\cdot)$ defined as some link function, usually taken as the logistic or probit link. Model (2.3) then allows us to model the sensitivity $\pi_{1j} = \mathrm{P}(Y_j = 1 | D_j = 1, \mathbf{x}_j)$ and specificity $\pi_{0j} = 1 - \mathrm{P}(Y_j = 1 | D_j = 0, \mathbf{x}_j)$ of the test for eye j as follows:

$$\begin{aligned} \pi_{1j} &= p(D_j = 1, \mathbf{x}_j) \\ &= h^{-1}(\beta_0 + \beta_1 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j), \end{aligned} \tag{2.4}$$

$$\begin{aligned} \pi_{0j} &= q(D_j = 0, \mathbf{x}_j) \\ &= 1 - h^{-1}(\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j). \end{aligned} \tag{2.5}$$

In the case of a logistic link function, (2.4) and (2.5) reduce to

$$\begin{aligned}\pi_{1j} &= \frac{\exp(\beta_0+\beta_1+\boldsymbol{\beta}^\top \mathbf{x}_j)}{1+\exp(\beta_0+\beta_1+\boldsymbol{\beta}^\top \mathbf{x}_j)},\\ \pi_{0j} &= \frac{1}{1+\exp(\beta_0+\boldsymbol{\beta}^\top \mathbf{x}_j)}.\end{aligned}$$

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 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.

2.2.1 Likelihood Representation

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

$$L = \prod_{i=1}^{N}\left[1+\frac{\rho\prod_{j=L,R}(y_{ij}+\theta_{ij}\{y_{ij}-1\})}{\sqrt{\theta_{iL}\theta_{iR}}}\right]\prod_{j=L,R}\left(\frac{\theta_{ij}^{y_{ij}}}{1+\theta_{ij}}\right), \tag{2.6}$$

where $\theta_{ij} = \theta\left(h_{ij}^{-1}\right) = p_{ij}/(1-p_{ij})$ is the odds for eye $j = L, R$ of patient i, and $h_{ij}^{-1} = h^{-1}(\beta_0 + \beta_1 D_{ij} + \boldsymbol{\beta}^\top \mathbf{x}_i)$. The log-likelihood function is then

$$\ell = \sum_{k,k'=0,1}\sum_{m,m'=0,1}\sum_{i=1}^{n_{kk'}^{mm'}}\ell_{kk'}^{mm'}(\beta_0,\beta_1,\boldsymbol{\beta},\rho|\mathbf{x}_i), \tag{2.7}$$

where $n_{kk'}^{mm'}$ is the number of observations for which $Y_L = k, Y_R = k', D_L = m$, and $D_R = m'$, with $\ell_{kk'}^{mm'}$ their corresponding log-likelihood contribution. Table 2.1

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, with $\beta = \mathbf{0}$.

Table 2.1: 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

A natural link function for binary data is the logistic link function. For this link function and assuming there are no other covariates except disease status D_{ij}, the log-likelihood function can be expressed in terms only of the parameters $\theta_0 = e^{\beta_0}, \theta_1 = e^{\beta_0+\beta_1}$, and ρ. In this case, $\pi_{1i} = \pi_1 = \theta_1/(1+\theta_1)$ and $\pi_{0i} = \pi_0 = 1/\theta_0$, for $i = 1, \cdots, N$.

2.2.2 Parameter Estimation and Inference

Since the joint distribution is completely specified, we outline in this section a full likelihood-based estimation method that yields maximum likelihood estimates (MLE) of the parameters.

Let $\eta^\top = (\alpha^\top, \rho)$, where $\alpha^\top = (\beta_0, \beta_1, \beta^\top)$. Define $\dot{\ell}(\eta) = \partial\ell/\partial\eta^\top$ as the score function and $\ddot{\ell}(\eta) = \partial^2\ell/\partial\eta\partial\eta^\top$ as the Hessian matrix. The MLE $\widehat{\eta}^\top = (\widehat{\alpha}^\top, \widehat{\rho})$ is the solution of the likelihood equations $\dot{\ell}(\eta) = \mathbf{0}^\top$. We solve these iteratively via the

Newton-Raphson updating scheme

$$\widehat{\eta}^{(t+1)} = \widehat{\eta}^{(t)} - \left[\ddot{\ell}\left(\widehat{\eta}^{(t)}\right)\right]^{-1}\left[\dot{\ell}\left(\widehat{\eta}^{(t)}\right)\right]^{\mathsf{T}}, \tag{2.8}$$

where $\widehat{\eta}^{(t)}$ is the estimate of η at iteration $t = 1, 2, \cdots$. The respective MLEs of π_1 and π_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 $\widehat{\eta}$ is consistent and asymptotically multivariate normal with mean η and covariance matrix given by the inverse of the information matrix $\mathbf{I}(\eta) = E_\eta\left[-\ddot{\ell}(\eta)\right]$. Standard asymptotic methods to perform hypothesis tests concerning η readily apply. In particular, suppose we wish to test the hypothesis $H_0' : \pi_1 \geq \pi_1^0$ (i.e., the test has sensitivity of at least $\pi_1^0 100\%$). Upon assuming $h(\cdot)$ is the logistic link and $\boldsymbol{\beta} = \mathbf{0}$, it is clear from the parametrization in the previous section that this hypothesis is equivalent to $H_0 : \mathbf{a}^{\mathsf{T}}\eta \geq \log\left(\frac{\pi_1^0}{1-\pi_1^0}\right)$, with $\mathbf{a} = (1, 1, 0)^{\mathsf{T}}$. This can be easily tested with Wald's statistic given by

$$Z = \frac{\mathbf{a}^{\mathsf{T}}\widehat{\eta} - \log\left(\frac{\pi_1^0}{1-\pi_1^0}\right)}{\sqrt{\mathbf{a}^{\mathsf{T}}\mathbf{I}^{-1}(\widehat{\eta})\mathbf{a}}}.$$

Under H_0, we have $Z \xrightarrow{a} N(0,1)$ and we reject if $Z < z_\alpha$, the $100(1-\alpha)$th percentile of the standard normal distribution.

2.2.3 Strengths and Weaknesses of Model

A strength of model (2.1) is that the joint distribution of the binocular data is completely determined by the marginal probabilities and the intra-pair correlation.

Hence, the regression parameters in (2.3) have marginal interpretations. This implies that the resulting parameter estimates have the same regression parameter interpretations as they would if each binary outcome is analyzed separately. In addition, model (2.1) 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 and validation. Moreover, unlike GEE (Qaqish and Liang, 1992; Lefkopoulou et al., 1989) and other quasi-likelihood approaches which cannot handle non-randomly sampled data, 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.

However, model (2.1) has some drawbacks that may render it unsuitable for other applications. The requirement that probabilities be nonnegative places constraints on the range of the intra-pair correlation ρ. Allowing the marginal probabilities to depend on patient-level covariates may severely restrict ρ into admitting mostly positive values. This, however, is not a serious issue in ophthalmological studies, as the inter-eye correlation is generally positive.

While model (2.1) can be readily extended to the general clustered binary data setting (Prentice, 1988), 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 (2.1) to the multi-reader binocular data structure displayed

in Tables 1.1 and 1.2 is still possible. By introducing random effects in (2.3), we can account for inter-reader and inter-pathology correlations among the clustered binocular data. This is adopted in Chapter 3.

In summary, model (2.1) is most appropriate for analyzing 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.

2.3 Asymptotic Relative Efficiency

In this section, we examine the potential gains in efficiency of model (2.1) over the crude approach which ignores the correlation ρ. For simplicity, we assume a logistic link and consider the case with $\beta_0 = 0$ and $\boldsymbol{\beta} = \mathbf{0}$.

For the crude method, we fit an ordinary logistic regression given by $\log[p/(1-p)] = \beta_1 D$, where $D = 0, 1$, is disease status and $p = P(Y = 1|D)$. Note that in this case, the eyes are assumed to be independent (i.e., $\rho = 0$). The asymptotic variance of the MLE $\widehat{\beta}_1^{\,c}$ of β_1 is given by $\text{avar}(\widehat{\beta}_1^{\,c}) = [n_1\pi_1(1-\pi_1)]^{-1}$, where n_1 is the number of observations such that $D = 1$, and $\pi_1 = \exp(\beta_1)/[1+\exp(\beta_1)]$. Letting $N^* = 2N = n_1 + n_0$, with n_0 the number of observations such that $D = 0$, we get

$$N^* \text{avar}\left(\widehat{\beta}_1^c\right) \to \frac{1}{\lambda\pi_1(1-\pi_1)}, \tag{2.9}$$

as $N^* \to \infty$, where $\lambda = \text{P}(D = 1)$, the incidence rate. Assuming $\lambda = \pi_1$ (i.e., the sensitivity of the test is high for a common disease and low for a rare disease) and

taking $\beta_1 = 1$, we get

$$\lim_{N^* \to \infty} N^* \text{avar}\left(\widehat{\beta}_1^c\right) = \frac{(1+e)^3}{e^2}. \tag{2.10}$$

Consider next model (2.1). Note that, conditional on $n^{mm'}$, $n_{kk'}^{mm'}$ follows a binomial distribution with parameters $n^{mm'}$ and $\mathrm{P}(Y_L = 1, Y_R = 1 | D_L = m, D_R = m')$. The asymptotic covariance matrix of β_1 and ρ given by $\mathbf{I}_{\beta_1,\rho}^{-1}$ involves only the marginal row counts $n^{mm'}$, such that

$$\frac{n^{mm'}}{N} \to \mathrm{P}(D_L = m, D_R = m'),$$

as $N \to \infty$. Assuming D_L and D_R have joint distribution given by (2.1) with marginal probabilities $\mathrm{P}(D_j = 1) = \pi_1$ for $j = L, R$, and correlation $\delta = \rho$, and taking $\beta_1 = 1$, we have for the MLE $\widehat{\beta}_1^m$ of β_1

$$\lim_{N \to \infty} N\text{avar}\left(\widehat{\beta}_1^m\right) = \begin{cases} \frac{c_1\rho^4 + c_2\rho^3 + c_3\rho^2 + c_4\rho + c_5}{c_6\rho^5 + c_7\rho^4 + c_8\rho^3 + c_9\rho^2 + c_{10}\rho + c_{11}} & \text{for } \rho \neq 0 \\ \frac{(1+e)^3}{e^2} & \text{for } \rho = 0 \end{cases}, \tag{2.11}$$

where

$$\begin{aligned} c_1 &= 4e^{-3/2}(1+e)^4(e^2 + 6e^3 + e^4 - 4e^{7/2} - 4e^{5/2}) \\ c_2 &= 4e^{-3/2}(1+e)^4(e + 2e^2 + 5e^3 + e^{9/2} + e^{5/2} - 4e^{3/2} + 10e^{7/2}) \\ c_3 &= 4e^{-3/2}(1+e)^4(9e^{7/2} + 10e^{5/2} + 5e^{3/2} - e^3 - 6e^2 - e) \\ c_4 &= 4e^{-3/2}(1+e)^4(3e^{3/2} - 6^{5/2} - 5e^{7/2} - 5e^2 - 2e - 1) \\ c_5 &= -4e^{-3/2}(1+e)^4(e^{1/2} - 9e^{5/2} - 6e^{3/2}) \end{aligned}$$

$$
\begin{aligned}
c_6 &= e + 5e^2 - 6e^3 - 6e^4 + 5e^5 + e^6 - 4e^{3/2} + 8e^{7/2} - 4e^{11/2} \\
c_7 &= e - 34e^2 - 6e^3 + 5e^4 - 27e^5 - 4e^6 + 8e^{3/2} + 8e^{5/2} + 32e^{7/2} + 8e^{9/2} + 8e^{11/2} \\
c_8 &= 3e + 27e^2 + 18e^3 + 42e^4 + 27e^5 + 11e^6 + 4e^{1/2} - 16e^{5/2} - 16e^{7/2} + 28e^{9/2} \\
c_9 &= 11e + 82e^2 + 26e^3 + 59e^4 + 139e^5 - 8e^{1/2} - 24e^{3/2} - 48e^{5/2} + 8e^{7/2} + 3 \\
c_{10} &= 8e^{3/2} + 16e^{5/2} - 24e^{7/2} - 8e - 28e^2 - 12e^3 - 20e^4 - 60e^5 \\
c_{11} &= -4e - 36e^2 - 108e^3 - 108e^4.
\end{aligned}
$$

Since $N^* = 2N$ and taking $\rho = \delta$, the asymptotic relative efficiency of $\widehat{\beta}_1^m$, the MLE of β_1 based on model (2.1), relative to the crude MLE $\widehat{\beta}_1^c$, is

$$
\begin{aligned}
ARE(\rho) &= \lim_{N\to\infty} \frac{\text{avar}\left(\widehat{\beta}_1^m\right)}{\text{avar}\left(\widehat{\beta}_1^c\right)} \\
&= \begin{cases} 2\lim_{N\to\infty} \frac{N\text{avar}\left(\widehat{\beta}_1^m\right)}{N^*\text{avar}\left(\widehat{\beta}_1^c\right)} & \text{for } \rho \neq 0 \\ 1 & \text{for } \rho = 0 \end{cases}.
\end{aligned} \tag{2.12}
$$

Note that (2.12) is a function of ρ, the correlation between Y_L and Y_R. Expression (2.12) simplifies to the ratio of (2.11) to (2.10) in the special case $\beta_1 = 1$.

Figures 2.1 to 2.3 plots (2.12) for a range of positive values of ρ at various fixed values of β_1. The choice of positive ρ values is in accordance with the common scenario encountered in ophthalmology, where the left and right eyes are generally positively correlated. It is clear that $ARE(\rho)$ is an increasing function of ρ, implying that $\widehat{\beta}_1^m$, the MLE based on model (2.1), is always more efficient than the crude MLE $\widehat{\beta}_1^c$. The plots indicate that the gains in efficiency will be greatest when the binocular outcomes are highly correlated. Thus, failure to account for the inter-eye correlation

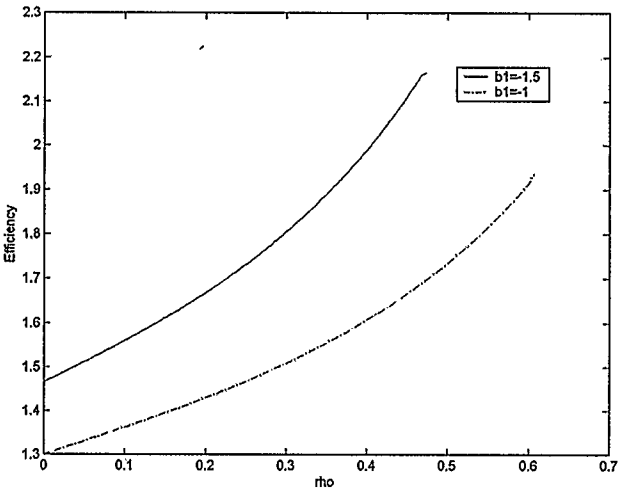

Figure 2.1: Plot of $ARE(\rho)$ against ρ for $\beta_1 = -1, -1.5$, and $\beta_0 = 0$

can result in quite substantial losses in efficiency in estimating the test's sensitivity π_1. This could then lead to invalid inferences in the form of underestimated p-values and confidence intervals that are too narrow (Leite and Nicolosi, 1998).

2.4 Simulation Study

To examine the performance in finite samples of the MLEs of π_0 and π_1 based on model (2.1) and assess the accuracy in finite samples of the large-sample standard error estimates described in section 2.2.2, we carried out a series of simulation exercises using binocular diagnostic data generated from model (2.1). We compare these with those obtained using the crude method.

The results of three such simulations illustrate the performance of the estimates for different sample sizes. Tables 2.2 to 2.4 are based on 500 simulated data with

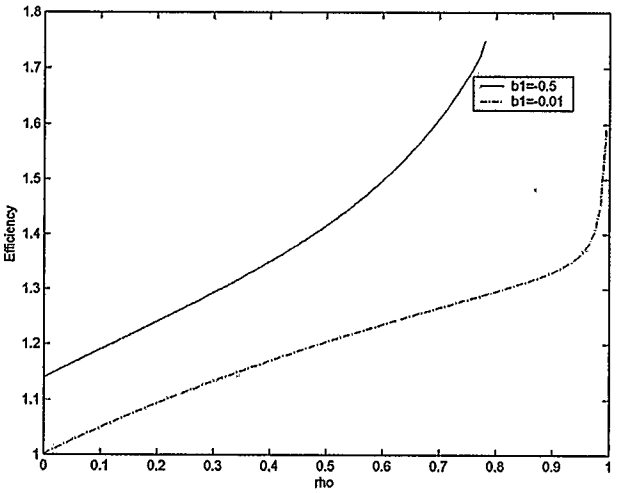

Figure 2.2: Plot of $ARE(\rho)$ against ρ for $\beta_1 = -0.01, -0.5$, and $\beta_0 = 0$

sample sizes $N = 100$ and 200 from the model (2.1). Table 2.2 corresponds to the case with $\rho = 0.1$ and $\pi_0 = \pi_1 = 0.9$. Table 2.3 corresponds to the case with $\rho = 0.4$ and $\pi_0 = \pi_1 = 0.7$, while Table 2.4 to $\rho = 0.8$ and $\pi_0 = \pi_1 = 0.55$. The data are generated using the following algorithm:

1 Fix ρ. Calculate $p^D_{d_L d_R} = \mathrm{P}(D_L = d_L, D_R = d_R)$ using model (2.1) with $\mathrm{P}(D_L = 1) = \mathrm{P}(D_R = 1) = 0.5$ and correlation ρ.

2 Generate a $U(0,1)$ random variable U, and

$$(D_L, D_R) = \begin{cases} (0,0) & \text{if } U < p^D_{00} \\ (0,1) & \text{if } p^D_{00} \le U < p^D_{00} + p^D_{01} \\ (1,0) & \text{if } p^D_{00} + p^D_{01} \le U < p^D_{00} + p^D_{01} + p^D_{10} \\ (1,1) & \text{if } p^D_{00} + p^D_{01} + p^D_{10} \le U < 1 \end{cases}$$

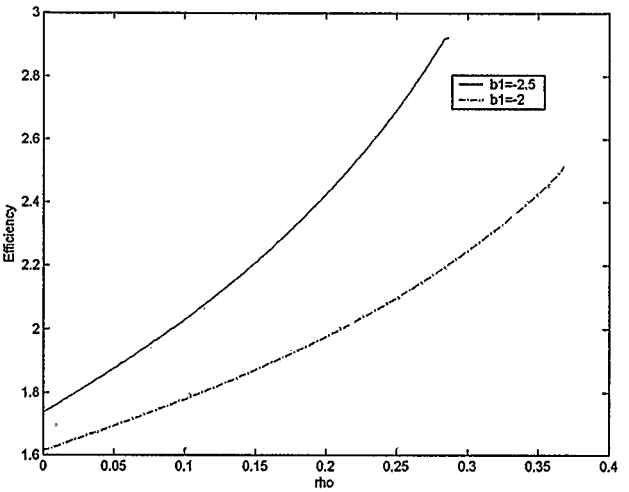

Figure 2.3: Plot of $ARE(\rho)$ against ρ for $\beta_1 = -2, -2.5$, and $\beta_0 = 0$

3 Fix β_0 and β_1 (to get $\pi_0 = \pi_1$). Calculate $p_{y_L y_R} = \mathrm{P}(Y_L = y_L, Y_R = y_R)$ using model (2.1) with $\mathrm{P}(Y_j = 1) = \frac{e^{\beta_0+\beta_1 D_j}}{1+e^{\beta_0+\beta_1 D_j}}$, $j = L, R$, and correlation ρ.

4 Generate a $U(0,1)$ random variable U, and

$$(Y_L, Y_R) = \begin{cases} (0,0) & \text{if } U < p_{00} \\ (0,1) & \text{if } p_{00} \le U < p_{00} + p_{01} \\ (1,0) & \text{if } p_{00} + p_{01} \le U < p_{00} + p_{01} + p_{10} \\ (1,1) & \text{if } p_{00} + p_{01} + p_{10} \le U < 1 \end{cases}$$

5 Go to **(1)** and repeat N times.

A check for condition (2.2) was built into the program codes to ensure that the ρ estimates satisfy the bound. When a ρ estimate failed to lie within the bounds, it is replaced by the midpoint of the interval. A more efficient approach would be to use

Fisher's z-transform $\tau = \log[(1+\rho)/(1-\rho)]$ to remove the constraints.

The means of the 500 sets of estimates and standard errors were calculated. In addition, the empirical standard deviation of the estimates were computed and the relative efficiency, defined as the ratio of the average standard error to the empirical standard deviation, was determined. The same thing was done for the crude method based on $2N = 200,400$ observations.

These simulations suggest that the MLEs based on model (2.1) perform well in finite samples. Except for the intra-pair correlation estimate $\widehat{\rho}$, we find little bias in the estimates and the standard error estimates were able to capture the true variability of the estimates. This can be seen in the uniformly high relative efficiencies in all cases. Note that the relative efficiencies of estimates based on model (2.1) are higher than those of the crude method. As ρ increases, the relative efficiencies from the crude method decrease while those from model (2.1) generally increase. This is to be expected as model (2.1) takes account of the correlation between the eyes. The crude method tends to underestimate the standard errors because it effectively assumes that there are more subjects providing independent information than is in fact the case.

While some bias appears in the estimates $\widehat{\rho}$, the bias generally decreased with increasing ρ. For instance, in the case $\rho = 0.1$, the ρ estimates yielded a relative bias of $(bias/\rho) \times 100 = 81.5\%$ for $N = 100$, a relatively high value. However, the relative bias went down to 12.5% for $\rho = 0.4$ and -0.075% for $\rho = 0.8$, with $N = 100$. The same observation holds for $N = 200$. Finally, we note that the bias in $\widehat{\rho}$ appears in other applications as well (Heagerty and Lele, 1998) and could be minimized by working with Fisher's z-transform $\tau = \log[(1+\rho)/(1-\rho)]$ instead of ρ. This needs

to be studied further for the proposed methodology.

Table 2.2: Relative Efficiency of Estimates based on 500 Simulated Datasets using Model (2.1) and using Crude Method with $\rho = 0.1, \pi_0 = \pi_1 = 0.9$

$N = 100$		Estimate (Mean)	SE (Mean)	SD (Empirical)	Relative Efficiency
Model	π_1	0.9048	0.0290	0.0310	0.9355
	π_0	0.9032	0.0290	0.0310	0.9355
	ρ	0.1815	0.2080	0.1768	1.1765
Crude	π_1	0.9025	0.0280	0.0315	0.8762
	π_0	0.9008	0.0272	0.0291	0.9347
$N = 200$					
Model	π_1	0.9027	0.0210	0.0200	1.0350
	π_0	0.9021	0.0210	0.0200	1.0350
	ρ	0.1990	0.1198	0.1343	0.8920
Crude	π_1	0.9003	0.0200	0.0201	0.9801
	π_0	0.8998	0.0200	0.0203	0.9704

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.

2.5 Application to Diabetic Retinopathy Data

We illustrate the methodology described in this chapter on the diabetic retinopathy data. Specifically, we consider the pathology microaneurysm. Table 2.5 shows the data concerning the presence (+) or absence (−) of microaneurysm in the left and right eyes of 92 diabetic patients.

The parametric model was applied to these data with $p(D_{ij}) = \mathrm{P}(Y_{ij} = 1|D_{ij})$ specified as a binary logistic regression model with parameters β_0 and β_1, and with

Table 2.3: Relative Efficiency of Estimates based on 500 Simulated Datasets using Model (2.1) and using Crude Method with $\rho = 0.4, \pi_0 = \pi_1 = 0.7$

$N = 100$		Estimate (Mean)	SE (Mean)	SD (Empirical)	Relative Efficiency
Model	π_1	0.7116	0.0470	0.0493	0.9533
	π_0	0.7091	0.0470	0.0493	0.9533
	ρ	0.4531	0.0923	0.1069	0.8634
Crude	π_1	0.7006	0.0420	0.0506	0.8300
	π_0	0.6977	0.0420	0.0516	0.8198
$N = 200$					
Model	π_1	0.7098	0.0310	0.0314	0.9968
	π_0	0.7114	0.0310	0.0314	0.9968
	ρ	0.4402	0.0605	0.0767	0.7888
Crude	π_1	0.6996	0.0280	0.0308	0.9123
	π_0	0.7015	0.0280	0.0323	0.8731

inter-eye correlation ρ.

Table 2.6 displays the MLEs of these parameters and their standard errors along with those from the crude method. We note that the inter-eye correlation $\widehat{\rho} = 0.1059$, while statistically significant, suggests 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 (2.1). For example, both approaches yielded estimates of sensitivity and specificity for the test of about 84% and 95%, respectively. Note however, that the standard errors of estimates from model (2.1) are generally larger than those of estimates using the crude method. This is to be expected in view of the results reported in sections 2.3 and 2.4.

Our proposed approach based on model (2.1) of analyzing the data is more general than the crude method, and reduces to the crude method when the eyes are

Table 2.4: Relative Efficiency of Estimates based on 500 Simulated Datasets using Model (2.1) and using Crude Method with $\rho = 0.8, \pi_0 = \pi_1 = 0.55$

$N = 100$		Estimate (Mean)	SE (Mean)	SD (Empirical)	Relative Efficiency
Model	π_1	0.5588	0.0560	0.0582	0.9588
	π_0	0.5542	0.0560	0.0582	0.9588
	ρ	0.7994	0.0545	0.0511	1.0665
Crude	π_1	0.5479	0.0440	0.0598	0.7324
	π_0	0.5431	0.0440	0.0573	0.7679
$N = 200$					
Model	π_1	0.5592	0.0430	0.0418	1.0263
	π_0	0.5598	0.0430	0.0418	1.0263
	ρ	0.8037	0.0410	0.0412	0.9951
Crude	π_1	0.5466	0.0340	0.0433	0.7852
	π_0	0.5474	0.0340	0.0442	0.7647

independent. If, by our method, the inter-eye correlation ρ 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.

2.6 Discussion

Clinical trials and epidemiological studies in ophthalmology often deal with data regarding the presence or absence of binocular findings. These data are taken from patients who usually contribute measurements from both eyes. In this chapter, we investigate the accuracy of a binary diagnostic test as determined by its sensitivity and specificity.

We propose a likelihood-based method of estimating sensitivity and specificity

Table 2.5: Assessment of Presence (+) or Absence (–) of Microaneurysm in the Left (L) and Right (R) Eyes of $N = 92$ Diabetic Patients

Disease Status	Test $L+R+$	$L-R+$	$L+R-$	$L-R-$	Total
$L+R+$	38	8	3	2	51
$L-R+$	1	1	0	1	3
$L+R-$	2	0	2	2	6
$L-R-$	0	1	0	31	32
Total	41	10	5	36	92

Table 2.6: Maximum Likelihood Estimates of Parameters from Model (2.1) and the Crude Method for Microaneurysm

	Model		Crude	
Parameter	Estimate	SE	Estimate	SE
β_0	-2.8452	0.5400	-2.8478	0.5143
β_1	4.4729	0.6044	4.4900	0.5752
ρ	0.1059	0.2407	—	—
π_1	0.8359	0.0380	0.8378	0.0350
π_0	0.9451	0.0280	0.9452	0.0266

using a parametric model that accounts for the inter-eye correlation. Estimation of model parameters via MLE is outlined and approximate hypothesis tests concerning the parameters are provided. The relative efficiencies of estimates based on the model was assessed theoretically and by simulation. Results reveal that treating the eyes as independent may lead to incorrect estimation of the standard errors, thereby resulting to invalid inferences. An application in a diabetic retinopathy study is used to illustrate the method.

Chapter 3

Analysis of Binocular Diagnostic Data: Case of Multiple Readers

3.1 Introduction

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 programs.

Consider the diabetic retinopathy study described in Chapter 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, two other sources of correlation are present in this case. Because readers rely on the same image of the eye, their diagnoses are potentially correlated. Moreover, because the pathologies are all related to retinopathy, it is very likely that the presence or absence of one influences the presence or absence of another. Thus, diagnoses for the pathologies based on the same eye are correlated. Note that the diagnoses in this setting are nested within pathologies and within readers.

One approach to estimating the sensitivity and specificity of a diagnostic procedure like the one described above is to ignore the inter-reader correlation and simply average the results from separate analyses via model (2.1) of data from each of the readers. Similarly, one can pretend that the pathologies are independent, and carry out separate analyses based on model (2.1) of each of the pathologies.

These approaches, however, can prove to be inefficient and hence, inadequate in practice. An obvious remedy would be to use a multivariate model for nested binocular binary data (e.g., Rosner, 1989). In this chapter, we instead extend model (2.1) to this situation by including random effects to account for the other sources of correlation, as is done in measurement reliability studies (Dunn, 1992). For simplicity, we assume in the next section that we have only one pathology and several readers; the case with more than one pathology is straightforward but more computationally involved.

3.2 Extension of Single-Reader Model to Several Readers

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. This correlation can be explained by the addition to model (2.1) of an unobserved random variable which varies from patient to patient. This is done in what follows.

Let $T \sim N(0,1)$ be a latent standard normal variable which varies across patients. Then the jth eye ($j = L, R$) diagnosis Y_{jk} from the kth reader ($k = 1, \cdots, K$) is assumed to depend on the disease status D_j, the latent variable T, and the covariates

$\mathbf{x}_j$, through a generalized linear model

$$\begin{aligned} p_{jk}(t) &= \mathrm{P}(Y_{jk}=1|D_j,t,\mathbf{x}_j) \\ &= p(D_j,t,\mathbf{x}_j) \\ &= h^{-1}(\beta_0+\beta_1 D_j+\beta_{2k}t+\boldsymbol{\beta}^{\top}\mathbf{x}_j). \end{aligned} \tag{3.1}$$

To define the joint distribution of $(Y_{L1},Y_{R1})^{\top},\cdots,(Y_{LK},Y_{RK})^{\top}$, we assume that the diagnoses of the readers are independent of each other, conditional on $(D_L,D_R)^{\top}$, $\mathbf{x}=\{\mathbf{x}_L,\mathbf{x}_R\}$, and T. Hence, the conditional joint distribution of $(Y_{Lk},Y_{Rk})^{\top}$, $k=1,\cdots,K$, is given by

$$\begin{aligned} \mathrm{P}(Y_{Lk}=y_{Lk},Y_{Rk}=y_{Rk},\forall k|D_L,D_R,t,\mathbf{x}) &= \prod_{k=1}^{K}\mathrm{P}(Y_{Lk}=y_{Lk},Y_{Rk}=y_{Rk}|D_L,D_R,t,\mathbf{x}) \\ &= \prod_{k=1}^{K}\left[1+\frac{\rho\{y_{Lk}-p_{Lk}(t)\}\{y_{Rk}-p_{Rk}(t)\}}{\sqrt{p_{Lk}(t)q_{Lk}(t)p_{Rk}(t)q_{Rk}(t)}}\right] \\ &\quad\times\prod_{j=L,R}[p_{jk}(t)]^{y_{jk}}[q_{jk}(t)]^{1-y_{jk}}. \end{aligned} \tag{3.2}$$

The unconditional joint distribution of $(Y_{Lk},Y_{Rk})^{\top}$, $k=1,\cdots,K$, is the average of (3.2) over T, that is,

$$\begin{aligned} \mathrm{P}(Y_{Lk}=y_{Lk},Y_{Rk}=y_{Rk},\forall k|D_L,D_R,\mathbf{x}) &= \int_{-\infty}^{+\infty}\prod_{k=1}^{K}\left[1+\frac{\rho\{y_{Lk}-p_{Lk}(t)\}\{y_{Rk}-p_{Rk}(t)\}}{\sqrt{p_{Lk}(t)q_{Lk}(t)p_{Rk}(t)q_{Rk}(t)}}\right] \\ &\quad\times\prod_{j=L,R}[p_{jk}(t)]^{y_{jk}}[q_{jk}(t)]^{1-y_{jk}}\,\mathrm{d}\Phi(t), \end{aligned} \tag{3.3}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. We can use the joint distribution in (3.3) to get the inter-reader pairwise correlations. With

$\mu_{jk} = \int_{-\infty}^{+\infty} p_{jk}(t)\mathrm{d}\Phi(t)$, $\sigma_{jk}^2 = \int_{-\infty}^{+\infty} p_{jk}(t)q_{jk}(t)\mathrm{d}\Phi(t) + \mathrm{var}[p_{jk}(t)]$, and

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

these correlations are as follows:

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

Note that $\rho(Y_{Lk}, Y_{j'k'})$ represents the unconditional correlation between readings by two different readers ($k \neq k'$) and $\rho(Y_{Lk}, Y_{Rk})$ the unconditional correlation between the left and right eye readings by the same reader.

Assuming T is independent of D_j and $\mathbf{x}_L = \mathbf{x}_R = \mathbf{x}$, then the unconditional probability $p_{jk} = \mathrm{P}(Y_{jk} = 1|D_j, \mathbf{x}_j) = p(D_j, \mathbf{x})$ is the average value of $p_{jk}(t)$ over T given by

$$p_{jk} = \int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_1 D_j + \beta_{2k}t + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{x})\mathrm{d}\Phi(t). \tag{3.4}$$

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 $k = 1, \cdots, K$. The sensitivity and specificity are

$$\pi_1 = \int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_1 + \beta_2 t + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{x})\mathrm{d}\Phi(t), \tag{3.5}$$

and

$$\pi_0 = 1 - \int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_2 t + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\mathrm{d}\Phi(t), \tag{3.6}$$

Note that our approach parallels that adopted by Hadgu and Qu (1998) and Qu et al. (1996) in analyzing diagnostic data with an imperfect gold standard.

3.2.1 Likelihood Representation

Suppose N patients undergo diagnostic testing on both left and right eyes for some pathology by several readers. Let $\{y_{iLk}, y_{iRk}, D_{iL}, D_{iR}, \mathbf{x}_i\}$, for $k = 1, \cdots, K$ and $i = 1, \cdots, N$, denote the observed data.

The likelihood contribution of patient i is

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

The log-likelihood function is then

$$\begin{aligned} \ell = \sum_{i=1}^{N} \log \Bigg(& \int_{-\infty}^{\infty} \prod_{j,k} [p_{ij}(t_i)]^{y_{ijk}} [q_{ij}(t_i)]^{1-y_{ijk}} \\ & \times \left[1 + \frac{\rho\{y_{iLk} - p_{iL}(t_i)\}\{y_{iRk} - p_{iR}(t_i)\}}{\sqrt{p_{iL}(t_i)q_{iL}(t_i)p_{iR}(t_i)q_{iR}(t_i)}}\right] d\Phi(t_i) \Bigg) \end{aligned} \tag{3.7}$$

The integrals in (3.7) may be evaluated by Gauss-Hermite quadrature techniques (Lesaffre and Spiessens, 2001). 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. Maximum likelihood for (3.7) is outlined in section 3.2.2.

3.2.2 Parameter Estimation

Define $\boldsymbol{\eta}^\top = (\boldsymbol{\alpha}^\top, \rho)$, where $\boldsymbol{\alpha}^\top = (\beta_0, \beta_1, \beta_2, \boldsymbol{\beta}^\top)$. To get the MLE of $\boldsymbol{\eta}$, we first evaluate the log-likelihood function (3.7) via Gauss-Hermite quadrature method, then numerically maximize it with respect to $\boldsymbol{\eta}$.

Let $t_{i1}, \cdots, t_{iQ}$, $i = 1, \cdots, N$, be the Gaussian quadrature points. Then,

$$\begin{aligned} \ell \;\approx\; & \sum_{i=1}^{N} \log \left(\sum_{q=1}^{Q} w_{iq} \prod_{j,r} \left[p_{ij}(t_{iq})\right]^{y_{ijr}} \left[q_{ij}(t_{iq})\right]^{1-y_{ijr}} \right. \\ & \left. \times \left[1 + \frac{\rho \left\{ y_{iLr} - p_{iL}(t_{iq}) \right\} \left\{ y_{iRr} - p_{iR}(t_{iq}) \right\}}{\sqrt{p_{iL}(t_{iq}) q_{iL}(t_{iq}) p_{iR}(t_{iq}) q_{iR}(t_{iq})}} \right] \right), \end{aligned} \tag{3.8}$$

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

If ℓ^* is the right-hand side of (3.8), the (approximate) MLE $\widehat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ is obtained by solving the (approximate) score equations $\dot{\boldsymbol{\ell}}^*(\boldsymbol{\eta}) = \partial \ell^* / \partial \boldsymbol{\eta}^\top = \mathbf{0}^\top$ via the Newton-Raphson algorithm. The usual properties of MLEs still apply to $\widehat{\boldsymbol{\eta}}$ in this case, and large-sample inference is carried out in standard fashion.

3.3 Extension to Several Pathologies

It is possible to extend model (3.3) further to the case of several pathologies as follows. To account for inter-pathology correlation, define another latent variable $U \sim N(0, 1)$, independent of T, which varies across patients. Then, given D_{jv}, the disease status for pathology $v = 1, \cdots, V$, T, and U, we have $p_{jv}(t, u) = h^{-1}(\beta_0 + \beta_1 D_{jv} + \beta_2 t + \beta_{3v} u)$, where we assumed $\beta_2 = \beta_{2k}$ for all $k = 1, \cdots, K$, and that no other covariates are available. The sensitivity and specificity of the test for pathology

$v = 1, \cdots, V$, are then given by

$$\pi_{1v} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_1 + \beta_2 t + \beta_{3v} u)\mathrm{d}\Phi(t)\mathrm{d}\Phi(u), \tag{3.9}$$

$$\pi_{0v} = 1 - \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^{-1}(\beta_0 + \beta_2 t + \beta_{3v} u)\mathrm{d}\Phi(t)\mathrm{d}\Phi(u). \tag{3.10}$$

To define the joint distribution of $(Y_{Lkv}, Y_{Rkv})^\top$, $v = 1, \cdots, V; k = 1, \cdots, K$, we assume the diagnoses across readers and across pathologies are independent of each other, conditional on $(D_{Lv}, D_{Rv})^\top$, T, and U. Hence, we get the conditional joint distribution as

$$\mathrm{P}(Y_{Lkv} = y_{Lkv}, Y_{Rkv} = y_{Rkv}, \forall k, v | D_{Lv}, D_{Rv}, \forall v; t, u) = \prod_{k,v}\prod_{j=L,R} [p_{jv}(t,u)]^{y_{jkv}} [q_{jkv}(t,u)]^{1-y_{jkv}}$$
$$\times \left[1 + \frac{\rho_v\{y_{Lkv} - p_{Lkv}(t,u)\}\{y_{Rkv} - p_{Rkv}(t,u)\}}{\sqrt{p_{Lkv}(t,u)q_{Lkv}(t,u)p_{Rkv}(t,u)q_{Rkv}(t,u)}}\right].$$

where ρ_v is the inter-eye correlation for pathology $v = 1, \cdots, V$. With μ_{jv} and σ^2_{jv} defined similarly as in section 3.2, and

$$\mu_{L1v,j'k'v'} = \begin{cases} \int_R p_{Lv}p_{Rv}\left[1 + \frac{\rho_v\{1-p_{Lv}\}\{1-p_{Rv}\}}{\sqrt{p_{Lv}q_{Lv}p_{Rv}q_{Rv}}}\right]\mathrm{d}\Phi(t)\mathrm{d}\Phi(u) & \text{for } j' \neq L, k' = 1, v = v' \\ \int_R p_{jv}p_{j'v'}\mathrm{d}\Phi(t)\mathrm{d}\Phi(u) & \text{for } j' \neq L \text{ or } k' \neq 1 \text{ or } v \neq v' \end{cases},$$

where $p_{jv} = p_{jv}(t,u)$ and $R = (-\infty, +\infty) \times (-\infty, +\infty)$, we have

$$\rho(Y_{L1v}, Y_{j'k'v'}) = \begin{cases} (\mu_{L1v,R1v} - \mu_{L1v}\mu_{R1v})(\sigma_{Lv}\sigma_{Rv})^{-1} & \text{for } j' \neq L, k' = 1, v = v' \\ (\mu_{L1v,j'k'v'} - \mu_{Lv}\mu_{j'v'})(\sigma_{Lv}\sigma_{j'v'})^{-1} & \text{for } j' \neq L \text{ or } k' \neq 1 \text{ or } v \neq v' \\ 1 & \text{for } j' = L, k' = 1, v = v' \end{cases}.$$

Note that $\rho(Y_{Lkv}, Y_{Rkv})$, $\forall k$, denotes the unconditional correlation between the left- and right-eye readings for the same pathology by the same reader while $\rho(Y_{L1v}, Y_{j'k'v'})$ denotes the unconditional correlation between readings for the same or different pathologies by the same or different readers.

The log-likelihood function becomes

$$\ell = \sum_{i=1}^{N} \log \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{j,k,v} [p_{ijv}(t_i, u_i)]^{y_{ijkv}} [q_{ijv}(t_i, u_i)]^{1-y_{ijkv}} \times \left[1 + \frac{\rho_v \{y_{iLkv} - p_{iLv}(t_i, u_i)\} \{y_{iRkv} - p_{iRv}(t_i, u_i)\}}{\sqrt{p_{iLv}(t_i, u_i) q_{iLv}(t_i, u_i) p_{iRv}(t_i, u_i) q_{iRv}(t_i, u_i)}} \right] d\Phi(t_i) d\Phi(u_i) \right).$$

Gaussian-Hermite quadrature methods can again be used to approximate ℓ and a numerical algorithm (e.g., Newton-Raphson algorithm) can be used to numerically solve the resulting (approximate) score equations to obtain the (approximate) MLEs $\widehat{\eta}^{\top} = (\widehat{\alpha}^{\top}, \widehat{\rho})$, where $\widehat{\alpha}^{\top} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_{31}, \cdots, \widehat{\beta}_{3V})$.

3.4 Application to Diabetic Retinopathy Data

We now apply the proposed methodology in section 3.2 to data from the diabetic retinopathy study. Specifically, we consider the pathology microaneurysm, and consider data from two readers.

Table 2.5 in Chapter 2 shows the left- and right-eye microaneurysm evaluations by Reader 1. The following Table 3.1 displays the evaluations by another reader, Reader 2.

We apply model (3.3) to the two sets of readings to account for the correlation between the two readers, using a logistic link. The MLEs are presented in Table 3.2

Table 3.1: Assessment by Reader 2 of Presence (+) or Absence (−) of Microaneurysm in the Left (L) and Right (R) Eyes of $N = 92$ Diabetic Patients

Disease	Test				
Status	$L+R+$	$L-R+$	$L+R-$	$L-R-$	Total
$L+R+$	45	2	1	3	51
$L-R+$	1	1	0	1	3
$L+R-$	2	0	4	0	6
$L-R-$	1	2	1	28	32
Total	49	5	6	32	92

along with their standard errors. From the Z values, it is clear that β_2 is significantly different from zero, which means that there does exist inter-rater correlation.

Table 3.2: MLEs of Parameters of Model (3.3) for Microaneurysm based on 2 Readers

Parameter	Estimate	SE	Z
β_0	-5.2561	1.3584	-3.8693
β_1	9.6060	2.2866	4.2010
β_2	3.5834	0.9933	3.6076
ρ	-0.0581	0.2673	-0.2174
π_1	0.8616	0.0674	12.7834
π_0	0.9054	0.0620	14.6032

We can now compare the results from the crude method, model (2.1), and model (3.3) concerning the sensitivity and specificity of the test. For the crude method and model (2.1), the average value of the estimates for the two readers were calculated. The estimates are presented in Table 3.3.

Results from Table 3.3 indicate the close correspondence between estimates from

Table 3.3: Comparison of the MLEs of Sensitivity and Specificity of Test for Diagnosing Microaneurysm Using the Three Methods

				95% CI	
	Parameter	Estimate	SE	Lower	Upper
Crude	π_1	0.8739	0.0222	0.8304	0.9174
Method	π_0	0.9178	0.0226	0.8735	0.9621
Model	π_1	0.8784	0.0240	0.8314	0.9254
(2.1)	π_0	0.9193	0.0235	0.8732	0.9654
Model	π_1	0.8616	0.0674	0.7295	0.9937
(3.3)	π_0	0.9054	0.0620	0.7839	1.0269

the crude method and those from model (2.1). This is to be expected as $\widehat{\rho}$ is quite small. Observe that the standard errors from model (3.3) are larger than those from the other two methods. This is not surprising because model (3.3) accounts for the correlation between the readers, thus adding another source of variation in the analysis. Note as well the wider confidence intervals for π_1 and π_0 based on model (3.3).

Therefore, treating the readers as independent ignores the correlation between the readers, and may consequently yield incorrect estimates, especially of the standard errors. Model (3.3) avoids this by accounting for this source of variation in the analysis.

3.5 Discussion

Clinical trials and epidemiological studies in ophthalmology often deal with data regarding the presence or absence of binocular findings on a number of eye pathologies

as determined by one or more trained specialists. These data are taken from patients who usually contribute multiple measurements on several pathologies from both eyes. In this chapter, we investigate the sensitivity and specificity of binary diagnostic tests when the diagnoses are determined by several readers.

We propose a likelihood-based method of estimating sensitivity and specificity using an extension of model (2.1) described in Chapter 2 that accounts for correlations besides the inter-eye correlation. The extension was achieved by including random effects to account for additional sources of correlation in the data, such as those between readers and those between pathologies. Estimation of model parameters was carried out via maximum likelihood estimation.

An application to data from the diabetic retinopathy study on one pathology and two readers is used to illustrate the method. Results reveal that treating the readers as independent may lead to incorrect estimation of the standard errors, thereby resulting in invalid inferences.

Chapter 4

Conclusion

4.1 Summary

The analysis of clustered binary diagnostic data is not straightforward because of a lack of standard models for the joint distribution of the variables. Besides the ad-hoc approach of carrying out separate analyses for the binary variables in the data, which are clearly deficient in many applications, a number of model-based alternatives have been previously proposed. We propose another model-based approach in this thesis which is particularly suited to binocular diagnostic data that arise in clinical and epidemiological studies in ophthalmology.

This thesis focuses on two main issues arising in multi-reader and multi-disease binocular diagnostic studies: how to account for inter-reader and inter-pathology correlations while at the same time incorporating the correlation between the binocular outcomes. The general approach taken in this thesis was a model-based one that relies on specifying a model for the joint distribution of the variables. 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 readers and between diseases or pathologies. This approach should be preferred to those that carry out separate analyses for the binary variables, as it provides a systematic and non-ad hoc way of analyzing the data and results in substantial gains in efficiency.

A model, previously discussed by Prentice (1988), is adopted in Chapter 2 to model the joint distribution of binocular binary diagnostic data. The model is completely determined by the marginal probabilities and the correlation between the binocular outcomes. Maximum likelihood estimation is outlined for the model, and large-sample inferential techniques are briefly discussed. The approach is illustrated by an application using data from a diabetic retinopathy study concerning a certain retinopathy-related pathology among diabetic patients as evaluated by a single reader. Estimates derived from the approach are shown, both theoretically and empirically via simulations, to be more efficient than those from the crude method, which ignores the intra-pair correlation.

In Chapter 3, an extension of the approach is introduced. The extension is accomplished by the addition of a random effect to account for the inter-reader correlation in multi-reader studies. A further extension to the case of several pathologies is also outlined, showing the flexibility of the method. Data from the diabetic retinopathy study involving two readers are again used to illustrate the usefulness of the method.

4.2 Future Research

Several issues still need to be addressed concerning the proposed method.

It would be interesting to investigate the performance of the methodology in cases where the sensitivity and specificity of the binocular binary diagnostic test are quite different, as when one is high and the other low. The simulations conducted only considered those cases where the two are the same, which is not very common in practice. Additional simulation studies thus need to be carried out to provide a

clearer picture of the performance of the methodology in various scenarios.

While complete specification of the likelihood function can be advantageous in many situations, there is always the danger of misspecification that can lead to invalid inferences. A way of checking the validity of our model thus needs to be developed. It is important that the robustness properties of the model be investigated.

There is also the matter of computations. The model can easily become unwieldy as the numbers of readers and pathologies increase. In cases where there are quite a number of readers and pathologies under consideration, evaluating the likelihood can prove to be computationally demanding. More efficient and easily implementable algorithms need to be developed for such cases.

Finally, many pathologies studied in the diabetic retinopathy study considered in the thesis do not have perfect gold standards for determining true disease status. A further extension of the method for analyzing multi-reader and multi-pathology studies wherein some of the pathologies lack a perfect gold standard, should be a welcome contribution to this area.

Bibliography

[1] Aerts, M., Geys, H., Molenberghs, G. and Ryan, L.M., *Topics in Modelling of Clustered Data*, Chapman & Hall/CRC, 2002.

[2] Ahn, C., An evaluation of methods for the estimation of sensitivity and specificity of site-specific diagnostic tests, *Biometrical Journal*, **39** (1997), 793-807.

[3] Bahadur, R.R., A representation of the joint distribution of responses to n dichotomous items, In *Studies in Item Analysis and Prediction*, H. Solomon (ed)., Stanford University Press, 158-168, 1961.

[4] Betensky, R.A. and Whittemore, A.S., An analysis of correlated multivariate binary data: Application to familial cancers of the ovary and breast, *Applied Statistics*, **45** (1996), 411-429.

[5] Bonney, G.E., Logistic regression for dependent binary observations, *Biometrics*, **35** (1987), 951-973.

[6] Connolly, M.A. and Liang, K-Y., Conditional logistic regression models for correlated binary data, *Biometrika*, **75** (1988), 501-506.

[7] Donner, A. and Klar, N., Confidence interval construction for effect measures arising from cluster randomization trials, *Journal of Clinical Epidemiology*, **46** (1993), 123-131.

[8] Dunn, G., Design and analysis of reliability studies, *Statistical Methods in Medical Research*, **1** (1992), 123-157.

[9] Glynn, R.J. and Rosner, B., Accounting for the correlation between fellow eyes in regression analysis, *Archives of Ophthalmology*, **110** (1992), 381-387.

[10] Hadgu, A. and Qu, Y., A biomedical application of latent class models with random effects, *Applied Statistics*, **47** (1998), 603-616.

[11] Haseman, J.K. and Kupper, L.L., Analysis of dichotomous response data from certain toxicological experiments, *Biometrics*, **35** (1979), 281-293.

[12] Heagerty, P.J. and Lele, S.R., A composite likelihood approach to binary spatial data, *Journal of the American Statistical Association*, **93** (1998), 1099-1111.

[13] Hujoel, P.P., Moulton, L.H. and Loesche, W.J., Estimation of sensitivity and specificity of site-specific diagnostic tests, *Journal of Periodontal Research*, **25** (1990), 193-196.

[14] Lee, E. and Dubin, N., Estimation and sample size considerations for clustered binary responses, *Statistics in Medicine*, **13** (1994), 1241-1252.

[15] Lefkopoulou, M., Moore, D. and Ryan, L., The analysis of multiple correlated binary outcomes, *Journal of the American Statistical Association*, **84** (1989), 810-815.

[16] Leisenring, W., Pepe, M.S. and Longton, G., A marginal regression modelling framework for evaluating medical diagnostic tests, *Statistics in Medicine*, **16** (1997), 1263-1281.

[17] Leite, M.L.C. and Nicolosi, A., Statistical analysis of correlated binary data in

ophthalmology: A weighted logistic regression approach, *Ophthalmic Epidemiology*, **5** (1998), 117-131.

[18] Lesaffre, E. and Spiessens, B., On the effect of the number of quadrature points in a logistic random-effects model: An example, *Applied Statistics*, **50** (2001), 325-335.

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

[20] Lipsitz, S.R., Laird, N.M. and Harrington, D.P., Maximum likelihood regression methods for paired binary data, *Statistics in Medicine*, **9** (1990), 1517-1525.

[21] Prentice, R.L., Correlated binary regression with covariates specific to each binary observation, *Biometrics*, **44** (1988), 1033-1048.

[22] Prentice, R.L., Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors, *Journal of the American Statistical Association*, **81** (1986), 321-327.

[23] Qaqish, B.F. and Liang, K-Y., Marginal models for correlated binary responses with multiple classes and multiple levels of nesting, *Biometrics*, **48** (1992), 939-950.

[24] Qu, Y., Tan, M. and Kutner, M.H., Random effects models in latent class analysis for evaluating accuracy of diagnostic tests, *Biometrics*, **52** (1996), 797-810.

[25] Rao, J.N.K. and Scott, A., A simple method for the analysis of clustered binary data, *Biometrics*, **48** (1992), 577-585.

[26] Rosner, B., Multivariate methods for clustered binary data with more than one level of nesting, *Journal of the American Statistical Association*, **84** (1989), 373-380.

[27] Rosner, B., Multivariate methods in ophthalmology with application to other paired-data situations, *Biometrics*, **40** (1984), 1025-1035.

[28] Rosner, B., Statistical methods in ophthalmology: An adjustment for the intraclass correlation between eyes, *Biometrics*, **38** (1982), 105-114.

[29] Rudnisky, C.J., Hinz, B.J., Tennant, M.T.S., de Leon, A.R., and Greve, M.D.J., High-resolution sterescopic digital fundus photography versus contact-lens biomicroscopy for the detection of clinically significant macular edema, *Ophthalmology*, **109** (2002), 267-274.

[30] Smith, P.J. and Hadgu, A., Sensitivity and specificity for correlated observations, *Statistics in Medicine*, **11** (1992), 1503-1509.

[31] Sutradhar, S.C. and Sutradhar, B., On classification of a bivariate binary observation, *Communications in Statistics-Theory & Methods*, **30** (2001), 2259-2279.

[32] Sutradhar, B.C. and Das, K., Generalized linear models for beta correlated binary longitudinal data, *Communications in Statistics-Theory & Methods*, **26** (1997), 617-635.

[33] Zhao, L.P. and Prentice, R.L., Correlated binary regression using a quadratic exponential model, *Biometrika*, **77** (1990), 642-648.

[34] Zhou, X-H., Obuchowski, N.A. and McClish, D.K., *Statistical Methods in Diagnostic Medicine*, Wiley & Sons, 2002.